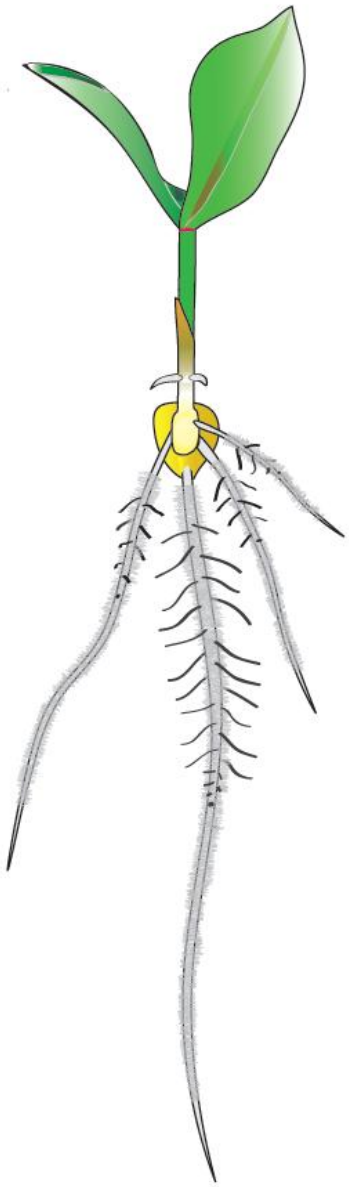
Funktionelle Genomik der Keimlingswurzeln von Mais (*Zea mays* L.)



Kumulative Habilitationsschrift

zur

Erlangung der *venia legendi*

für das Fach Genetik

an der

Landwirtschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

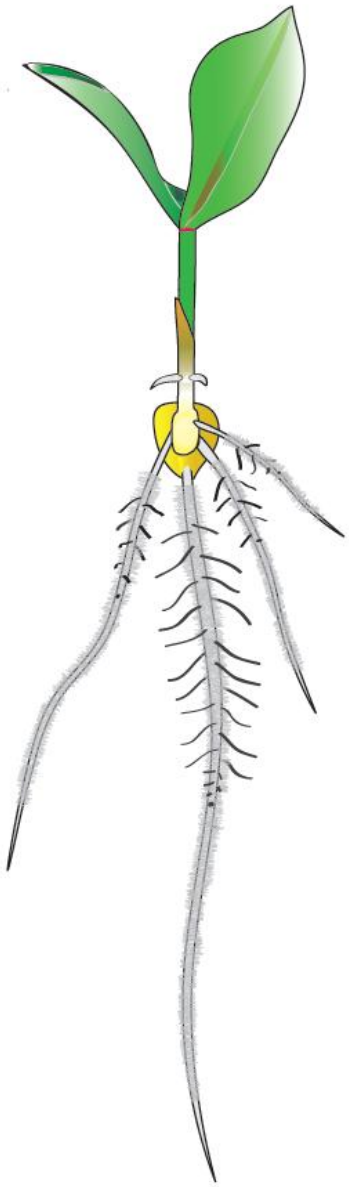
vorgelegt von

Dr. Caroline Marcon

Erscheinungsjahr: 2026

Institut für Nutzpflanzenwissenschaften und Ressourcenschutz
Professur für Crop Functional Genomics

Funktionelle Genomik der Keimlingswurzeln von Mais (*Zea mays* L.)



Kumulative Habilitationsschrift

zur

Erlangung der *venia legendi*

für das Fach Genetik

an der

Landwirtschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Dr. Caroline Marcon

Tag der Habilitation: 24.05.2023

Institut für Nutzpflanzenwissenschaften und Ressourcenschutz

Professur für Crop Functional Genomics

Inhaltsverzeichnis

1 Lebenslauf 4

2 Promotionsurkunde 6

3 Publikationslisten 7

3.1 Publikationen nach Abschluss der Promotion 7

3.2 Publikationen bis zur Promotion 10

4 Vorträge 11

4.1 Eingeladene Vorträge 11

4.2 Sonstige Vorträge 11

5 Poster-Präsentationen 12

6 Angaben zur Lehre 15

6.1 Lehrveranstaltungen 15

6.2 Betreute Abschlussarbeiten 16

6.2.1 Doktorarbeiten (Dr. agr.) 16

6.2.2 Masterarbeiten (M. Sc.) 17

6.2.3 Bachelorarbeiten (B. Sc.) 17

7.1 Eingeworbene Drittmittel 18

7.2 Internationale Forschungserfahrung 18

7.3 Gutachtertätigkeiten 19

7.4 Akademische Selbstverwaltung 19

7.5 Weiterbildung 19

7.6 Preise 19

8 Erklärungen gemäß § 3 der Habilitationsordnung 20

9 Zusammenfassung der wissenschaftlichen Arbeiten 21

9.1 Einleitung 21

9.1.1 Mais als Modellpflanze für funktionelle Genomanalysen 21

9.1.2 Das Wurzelsystem von Mais 23

9.2 Neue Studien zur funktionellen Genomik des jungen Maiswurzelsystems 25

9.2.1 Molekulargenetische Analysen der Wurzelentwicklung 25

9.2.1.1 Klonierung und Charakterisierung der *rth5* und *rth6* Gene 26

9.2.1.2 Eigenschaften der homeologen Gene *rum1* und *rul1* 28

9.2.2 Transkriptomanalysen der Wurzelentwicklung von Mais 30

9.2.2.1 Transkriptomprofile verschiedener Wurzeltypen und zellspezifische RNA-Seq-Analysen 31

9.2.2.2 Differentielle Genexpression in der Wurzelmutante *rum1* versus WT 32

9.2.2.3 Reaktion des Wurzel-Transkriptoms auf Wasserdefizitstress 33

9.2.2.4 Heterosis-assoziierte Transkriptomanalysen der Keimlingswurzeln von Maishybriden 35

9.2.2.5 Spezielle Expressionsmuster formen das Transkriptom reziproker Hybride 36

9.2.2.6 Single-parent expression als genereller Mechanismus der Komplementation in verschiedenen Wurzelgeweben, unter Wasserdefizitstress und in diversen Maishybriden 38

9.2.2.7 Gewebespezifische Dynamik der Genexpression in Hybriden 38

9.2.2.8 Stabilität der SPE-Gene bei Wasserdefizitstress 39

9.2.2.9 Genexpressionsprofile genetisch unterschiedlicher Inzuchtlinien und deren Hybride während der Primärwurzelentwicklung 41

9.2.2.10 Komplementation der Genexpression in evolutionär jungen nicht-syntänischen Genen 42

9.2.3 Proteom-Atlas der Primärwurzel von Mais 44

9.2.4 Insertionsmutagenese zur Erzeugung neuer Mutationen in Mais 48

9.2.4.1 Etablierung der europäischen Maismutanten-Ressource *BonnMu* 49

9.2.4.2 Automatisierte Identifizierung von Transposoninsertionen mit der *Mutant-Seq Workflow Utility* (*MuWU*) 54

9.2.4.3 Nutzung der *BonnMu*-Ressource für funktionelle Genanalysen 55

9.2.4.4 Ausweitung der *BonnMu*-Ressource 56

9.3 Zusammenfassung und Ausblick....56
9.4 Literaturverzeichnis....58
10 Anhang....65

1 Lebenslauf

Persönliche Daten

Name	Dr. Caroline Marcon
Universitätsanschrift	Rheinische Friedrich-Wilhelms-Universität Bonn
	Institut für Nutzpflanzenwissenschaften und Ressourcenschutz (INRES)
	Professur für Crop Functional Genomics (CFG)
	Friedrich-Ebert-Allee 144
	53113 Bonn
	Tel.: 0228-73 54269
	E-Mail: marcon@uni-bonn.de

Akademischer Werdegang

Seit 03/2019	Projektleiterin am Institut für Nutzpflanzenwissenschaften und Ressourcenschutz (INRES), Professur für Crop Functional Genomics (CFG), Rheinische Friedrich-Wilhelms-Universität Bonn
07/2012 – 03/2019	Wissenschaftliche Assistentin bei Prof. Dr. Frank Hochholdinger am Institut für Nutzpflanzenwissenschaften und Ressourcenschutz (INRES), Professur für Crop Functional Genomics (CFG), Rheinische Friedrich-Wilhelms-Universität Bonn

1 Lebenslauf

06/2007 – 06/2012	Promotion (Dr. rer. nat.) bei Prof. Dr. Frank Hochholdinger am Zentrum für Molekularbiologie der Pflanzen (ZMBP), Eberhard Karls Universität Tübingen Thema: „Heterosis-assoziierte Transkriptom- und Proteom-Komplexität in Embryonen und Keimlingswurzeln von Mais (*Zea mays* L.) Hybriden“
04/2006 – 04/2007	Diplomarbeit bei Prof. Dr. Claudia Oecking am Zentrum für Molekularbiologie der Pflanzen (ZMBP), Eberhard Karls Universität Tübingen Thema: „Regulation der pflanzlichen H^+-ATPase durch ihre autoinhibitorische Domäne und 14-3-3 Proteine“
10/2001 – 03/2006	Biologiestudium an der Eberhard Karls Universität Tübingen Hauptfach: Pflanzenphysiologie Nebenfächer: Genetik und Biochemie

Bildungsweg

08/2000 – 07/2001	Freiwilliges Ökologisches Jahr (FÖJ), Ökomobil des Regierungspräsidiums Stuttgart
09/1997 – 06/2000	Berufliches Gymnasium der Mildred-Scheel-Schule, Böblingen Abschluss: Abitur
09/1991 – 07/1997	Otto-Rommel-Realschule, Holzgerlingen Abschluss: Mittlere Reife

2 Promotionsurkunde

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Doktor

der Mathematisch-Naturwissenschaftlichen Fakultät
Promotionsfach Biologie

Frau Caroline Marcon
geboren am 22. Januar 1981 in Böblingen

wird aufgrund der mit sehr gut beurteilten Dissertation

„Heterosis-assoziierte Transkriptom- und Proteom-Komplexität in Embryonen und Keimlingswurzeln von Mais (*Zea mays* L.) Hybriden“

und der mit sehr gut bewerteten mündlichen Prüfung
am 27. Juni 2012

hiermit der akademische Grad

Doktor der Naturwissenschaften
(Dr. rer. nat.)

mit dem Gesamturteil
„sehr gut“ (magna cum laude) verliehen.

Tübingen, den 5. September 2012

Prof. Dr. Bernd Engler
Der Rektor

Prof. Dr. Wolfgang Rosenstiel
Der Dekan

3 Publikationslisten

In diesem Kapitel sind alle Publikationen mit Eigenbeteiligung aufgelistet. Mit diesen Arbeiten habe ich einen h-Index (Google Scholar) von 21 erreicht. Die Anzahl der Zitierungen beläuft sich auf 1380 (Stand: 31/03/2022).

3.1 Publikationen nach Abschluss der Promotion

Publikation	Impact Factor:
1. Stöcker T, Altrogge L, **Marcon C**, Win YN, Hochholdinger F, Schoof H (2021) *MuWU*: *Mutant-Seq* library analysis and annotation. Bioinformatics: btab679. doi.org/10.1093/bioinformatics/btab679.	**6.9**
2. **Marcon C***, Altrogge L, Win YN, Stöcker T, Gardiner JM, Portwood JL, Opitz N, Kortz A, Baldauf JA, Hunter CT, McCarty DR, Koch KE, Schoof H, Hochholdinger F (2020) *BonnMu*: a sequence-indexed resource of transposon induced maize mutations for functional genomics studies. Plant Physiol 184: 620-631. doi.org/10.1104/pp.20.00478. *****Seniorautorin.	**6.9**
3. Haberer G, Kamal N, Bauer E, Gundlach H, Fischer I, Seidel MA, Spannagl M, **Marcon C**, Ruban A, Urbany C, Nemri A, Hochholdinger, F, Ouzunova, M, Houben A, Schön C-C, Mayer KFX (2020) European maize genomes highlight intra-species variation in repeat and gene content. Nature Genet 52: 950-957. doi.org/10.1038/s41588-020-0671-9.	**27.6**
4. Vetterlein D, Lippold E, Schreiter S, Phalempin M, Fahrenkampf T, Hochholdinger F, **Marcon C**, Tarkka M, Oburger E, Ahmed M, Javaux M, Schlüter S (2020) Experimental platforms for the investigation of spatiotemporal patterns in the rhizosphere - laboratory and field scale. J Plant Nutr Soil Sc 000: 1-16. doi.org/10.1002/jpln.202000079.	**2.1**
5. Haberer G, Bauer E, Kamal N, Grundlach H, Fischer I, Seidel MA, Spannagl M, **Marcon C**, Ruban A, Urbany C, Nemri A, Hochholdinger F, Ouzunova M, Houben A, Schön C-C, Mayer KFX (2019) European maize genomes unveil pan-genomic dynamics of repeats and genes. bioRxiv 766444. doi.org/10.1101/766444.	**-**

6. Baldauf JA, **Marcon C**, Lithio A, Vedder L, Altrogge L, Piepho H-P, Schoof H, Nettleton D, Hochholdinger F (2018) Single-parent expression is a general mechanism driving extensive complementation of non-syntenic genes in maize (*Zea mays* L.) hybrids. Current Biol 28: 431-437. doi.org/10.1016/j.cub.2017.12.027. **9.6**
7. Hochholdinger F, **Marcon C**, Baldauf JA, Frey FP, Yu P (2018) Proteomics of maize root development. Front Plant Sci 9: 143. doi.org/10.3389/fpls.2018.00143. **4.4**
8. Hochholdinger F, Yu P, **Marcon C** (2018) Genetic control of root system development in maize. Trends Plant Sci 23:79-88. doi.org/10.1016/j.tplants.2017.10.004. **9.9**
9. Yu P, **Marcon C**, Baldauf JA, Frey FP, Baer M, Hochholdinger F (2018) Transcriptomic dissection of maize root system development. Compendium of plant genomes: The *Zea mays* Genome. 1st ed. (eds. J. Bennetzen, S. Flint-Garcia, C. Hirsch and R. Tuberosa) pp. 247-257. Springer Nature. doi.org/10.1007/978-3-319-97427-9_15. **-**
10. **Marcon C**, Paschold A, Malik WA, Lithio A, Baldauf JA, Altrogge L, Opitz N, Lanz C, Schoof H, Nettleton D, Piepho H-P, Hochholdinger F (2017) Stability of single-parent gene expression complementation in maize hybrids upon water deficit stress. Plant Physiol 173: 1247-1257. doi.org/10.1104/pp.16.01045. **6.4**
11. Unterseer S, Seidel MA, Bauer E, Haberer G, Hochholdinger F, Opitz N, **Marcon C**, Baruch K, Spannagl M, Mayer KFX, Schön C-C (2017) European Flint reference sequences complement the maize pan-genome. bioRxiv 103747. doi.org/10.1101/103747. **-**
12. **Marcon C**, Baldauf JA, Hochholdinger F (2016) Komplementation von Genexpressionsmustern in Maishybriden. Biospektrum 22: 603-605. doi.org/10.1007/s12268-016-0737-5. **-**
13. Baldauf JA, **Marcon C**, Paschold A, Hochholdinger F (2016) Nonsyntenic genes drive tissue-specific dynamics of differential, nonadditive and allelic expression patterns in maize hybrids. Plant Physiol 171: 159-167. doi.org/10.1104/pp.16.00262. **6.4**
14. Yu P, Baldauf JA, Lithio A, **Marcon C**, Nettleton D, Li C, Hochholdinger F (2016) Root type-specific reprogramming of maize pericycle transcriptomes by local high nitrate results in disparate lateral root branching patterns. Plant Physiol 170: 1783-1798. doi.org/ 10.1104/pp.15.01885. **6.4**

15. Li L, Hey S, Liu S, Liu Q, McNinch C, Hu HC, Wen TJ, **Marcon C**, Paschold A, Bruce W, Schnable PS, Hochholdinger F (2016) Characterization of maize *roothairless6* which encodes a D-type cellulose synthase and controls the switch from bulge formation to tip growth. Sci Rep 6: 34395. doi.org/10.1038/srep34395. **4.1**
16. Zhang Y, **Marcon C**, Tai H, von Behrens I, Ludwig Y, Hey S, Berendzen KW, Hochholdinger F (2016) Conserved and unique features of the homeologous maize Aux/IAA proteins ROOTLESS WITH UNDETECTABLE MERISTEM 1 and RUM1-like 1. J Exp Bot 67: 1137-1147. doi.org/10.1093/jxb/erv519. **5.9**
17. Tai H, Lu X, Opitz N, **Marcon C**, Paschold A, Lithio A, Nettleton D, Hochholdinger F (2016) Transcriptomic and anatomical complexity of primary, seminal, and crown roots highlight root type-specific functional diversity in maize (*Zea mays* L.). J Exp Bot 67: 1123-1135. doi.org/10.1093/jxb/erv513. **5.9**
18. Opitz N, **Marcon C**, Paschold A, Malik WA, Lithio A, Brandt R, Piepho H-P, Nettleton D, Hochholdinger F (2016) Extensive tissue-specific transcriptomic plasticity in maize primary roots upon water deficit. J Exp Bot 67: 1095-1107. doi.org/10.1093/jxb/erv453. **5.9**
19. **Marcon C**, Malik WA, Walley JW, Shen Z, Paschold A, Smith LG, Piepho H-P, Briggs SP, Hochholdinger F (2015) A high-resolution tissue-specific proteome and phosphoproteome atlas of maize primary roots reveals functional gradients along the root axes. Plant Physiol 168: 233-246. doi.org/10.1104/pp.15.00138. **6.4**
20. Paschold A, Larson NB, **Marcon C**, Schnable JC, Yeh CT, Lanz C, Nettleton D, Piepho H-P, Schnable PS, Hochholdinger F (2014) Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. Plant Cell 26: 3939-3948. doi.org/10.1105/tpc.114.130948. **9.6**
21. Opitz N, Paschold A, **Marcon C**, Malik WA, Lanz C, Piepho H-P, Hochholdinger F (2014) Transcriptomic complexity in young maize primary roots in response to low water potentials. BMC Genomics 15: 741. doi.org/10.1186/1471-2164-15-741. **3.5**
22. Nestler J, Liu S, Wen T-J, Paschold A, **Marcon C**, Tang HM, Li D, Li L, Meeley R, Sakai H, Bruce W, Schnable PS, Hochholdinger F (2014) *Roothairless5*, which functions in maize (*Zea mays* L.) root hair **6.1**

initiation and elongation encodes a monocot-specific NADPH oxidase. Plant J 79: 729-740. doi.org/10.1111/tpj.12578.

23. Zhang Y, Paschold A, **Marcon C**, Liu S, Tai H, Nestler J, Yeh CT, Opitz N, Lanz C, Schnable PS, Hochholdinger F (2014) The Aux/IAA gene *rum1* involved in seminal and lateral root formation controls vascular patterning in maize (*Zea mays* L.) primary roots. J Exp Bot 65: 4919-4930. doi.org/10.1093/jxb/eru249. **5.9**
24. Hochholdinger F, Paschold A, **Marcon C** (2013) Heterosis: Maishybride exprimieren mehr Gene als ihre Eltern. Biologie in unserer Zeit 43:11-12. doi.org/10.1002/biuz.201390001. **-**

3.2 Publikationen bis zur Promotion

Publikation	Impact Factor:
1. **Marcon C**, Paschold A, Hochholdinger F (2013) Genetic control of root organogenesis in cereals. In: De Smet I. (eds) Plant Organogenesis. Methods in Molecular Biology (Methods and Protocols), vol 959. Humana Press, Totowa, NJ. doi.org/10.1007/978-1-62703-221-6_4.	**-**
2. **Marcon C**, Lamkemeyer T, Malik WA, Ungrue D, Piepho H-P, Hochholdinger F (2013) Heterosis-associated proteome analyses of maize (*Zea mays* L.) seminal roots by quantitative label-free LC-MS. J Proteomics 93: 295-302. doi.org/10.1016/j.jprot.2013.04.015.	**3.5**
3. Paschold A, Jia Y, **Marcon C**, Lund S, Larson NB, Yeh CT, Ossowski S, Lanz C, Nettleton D, Schnable PS, Hochholdinger F (2012) Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents. Genome Res 22: 2445-2454. doi.org/10.1101/gr.138461.112.	**11.0**
4. **Marcon C**, Schützenmeister A, Schütz W, Madlung J, Piepho H-P, Hochholdinger F (2010) Nonadditive protein accumulation patterns in maize (*Zea mays* L.) hybrids during embryo development. J Proteome Res 9: 6511-6522. doi.org/10.1021/pr100718d.	**4.0**

5. Paschold A, **Marcon C**, Hoecker N, Hochholdinger F (2010) Molecular dissection of heterosis manifestation during early maize root development. Theor Appl Genet 120: 383-388. doi.org/10.1007/s00122-009-1082-6. **4.4**
6. Speth C, Jaspert N, **Marcon C**, Oecking C (2010) Regulation of the plant plasma membrane H^+-ATPase by its C-terminal domain: what do we know for sure? Eur J Cell Biol 89: 145-151. doi.org/10.1016/j.ejcb.2009.10.015. **3.0**
7. Ottmann C, Marco S, Jaspert N, **Marcon C**, Schauer N, Weyand M, Vandermeeren C, Duby G, Boutry M, Wittinghofer A, Rigaud JL, Oecking C (2007) Structure of a 14-3-3 coordinated hexamer of the plant plasma membrane H^+-ATPase by combining X-ray crystallography and electron cryomicroscopy. Mol Cell 25: 427-440. doi.org/10.1016/j.molcel.2006.12.017. **15.5**

4 Vorträge

4.1 Eingeladene Vorträge

06/2021 CiBreed – Center for Integrated Breeding Research, Universität Göttingen, Online:
„*BonnMu* – a mutant collection for functional genetics in maize“

10/2020 MAZE Phase 2 – Kick-Off Meeting 2020, Helmholtz Zentrum München, Online:
„*BonnMu* – a mutant collection for functional genetics in maize“

09/2020 KWS Saat SE & Co. KgaA, St. Louis, USA, Online:
„*BonnMu* – a mutant collection for functional genetics in maize“

4.2 Sonstige Vorträge

04/2022 Pflanzenwissenschaftliches Kolloquium, Bonn:
„*BonnMu* - a resource of transposon-induced maize mutations for functional genetics studies“

03/2016 Gesellschaft für Pflanzenzüchtung (GPZ) Haupttagung 2016, Bonn:

„Conservation of nonsyntenic transcriptomic variation in maize hybrids upon water deficit stress"

03/2014 Minisymposium, China Agricultural University, Peking:

„Transcriptome and proteome complexity in maize root tissues"

10/2013 Minisymposium – maize meets rice, Paris:

„Heterosis in roots of maize (*Zea mays* L.) hybrids"

5 Poster-Präsentationen

03/2021 63rd Maize Genetics Conference, Online:

Win YN, **Marcon C**, Stöcker T, Haberer G, Gardiner JM, Portwood JL, Schoof H, Hochholdinger F

„Expanding *BonnMu* – The transposon-induced mutant resource in the European flint maize line F7"

03/2019 GPZ Konferenz Genome Research, Universität Hohenheim:

Marcon C, Altrogge L, Win YN, Schoof H, Hochholdinger F

„A (European) reverse genetics resource in maize

01/2019 Plant and Animal Genome XXVII Conference (PAG), San Diego, USA:

Kamal N, Fischer I, Ruban A, Opitz N, Seidel M, **Marcon C**, Grundlach H, Spannagl M, Houben A, Presterl T, Ouzunova M, Hochholdinger F, Mayer KFX, Bauer E, Schön C-C, Haberer G

„European flint genomes complementing the maize pan- and core-genome"

03/2018 60th Maize Genetics Conference, Saint-Malo, Frankreich:

1. **Marcon C**, Altrogge L, Win YN, Kortz A, Schoof H, Hochholdinger F
   „Identification of novel maize (root) mutants by *Mu-Seq*"
2. Bauer E, Haberer G, Seidel MA, Grundlach H, Hochholdinger F, **Marcon C**, Baruch K, Spannagl M, Mayer KFX, Schön C-C
   „Four European flint reference sequences complement the maize pan-genome"

5 Poster-Präsentationen

03/2017 — 59th Maize Genetics Conference, St. Louis, USA:

1. Unterseer S, Seidel MA, Bauer E, Haberer G, Hochholdinger F, Opitz N, **Marcon C**, Baruch K, Spannagl M, Mayer KFX, Schön C-C
   „European Flint reference sequences complement the maize pan-genome“
2. Baldauf JA, **Marcon C**, Lithio A, Nettleton D, Hochholdinger F
   „Transcriptomic dissection of heterosis in maize primary roots at the interface of genotype and development“

03/2016 — 58th Maize Genetics Conference, Jacksonville, USA:

Baldauf JA, **Marcon C**, Paschold A, Hochholdinger F
„Nonsyntenic genes drive tissue-specific dynamics of differential, nonadditive and allelic expression patterns in heterotic maize hybrids“

06/2015 — Royal Society of Biology (SEB) Conference, Prag, Tschechien:

Marcon C, Walley JW, Shen Z, Malik WA, Paschold A, Piepho H-P, Smith LG, Briggs SP, Hochholdinger F
„Tissue-specific proteome and phosphoproteome atlas of maize primary roots reveals functional gradients along the root axes“

09/2014 — 7th International Symposium on Root Development, Weimar:

Marcon C, Walley JW, Shen Z, Malik WA, Paschold A, Piepho H-P, Smith LG, Briggs SP, Hochholdinger F
„Large-scale proteomic and phosphoproteomic analyses of maize root tissues“

03/2014 — 56th Maize Genetics Conference, Peking, China:

Marcon C, Walley JW, Shen Z, Smith LG, Briggs SP, Hochholdinger F
„Large-scale proteomic and phosphoproteomic analyses of maize root tissues“

01/2014 — Plant and Animal Genome XXII Conference (PAG), San Diego, USA:

Paschold A, Larson NB, **Marcon C**, Schnable JC, Yeh C-T, Lanz C, Nettleton D, Piepho H-P, Schnable PS, Hochholdinger F
„Transcriptomic complementation in maize (*Zea mays* L.) hybrid tissues relative to their inbred parents“

03/2011 53th Maize Genetics Conference, St. Charles, USA:

Marcon C, Hochholdinger F

„Non-additive protein accumulation patterns in maize hybrids during embryo development“

03/2010 52th Maize Genetics Conference, Riva del Garda, Italien:

1. **Marcon C**, Hochholdinger F
   „Non-additive protein accumulation patterns in maize hybrids during embryo development“
2. Paschold A, **Marcon C**, Hochholdinger F
   „Transcriptome-wide analysis of the manifestation of heterosis in young seedling roots of maize (*Zea mays* L.)“
   during embryo development

03/2009 51th Maize Genetics Conference, St. Charles, USA:

Marcon C, Hochholdinger F
„Molecular dissection of heterosis manifestation in maize roots and embryos“

07/2008 XX International Congress of Genetics, Berlin:

Marcon C, Hoecker N, Lamkemeyer T, Nordheim A, Hochholdinger F
„Proteomic dissection of heterosis manifestation in maize (*Zea mays* L.) roots and embryos“

6 Angaben zur Lehre

6.1 Lehrveranstaltungen

Leitung und Durchführung studentischer Praktika				
Semester	Lehrveranstaltung (Modulnummer)	Turnus	SWS (gesamt)	Beteiligung, SWS (anteilig)
Ab Sommersemester (SoSe) 2023	Grundlagen molekular-genetischer Praxis (B.Sc.) AGR-070	SoSe	6	**5**
Seit SoSe 2015	Molecular crop science I MA-M-01-PM	SoSe	6	**3**
SoSe 2015 – 2018	Heterosis in crops MA-E,M,P-22-FW	SoSe	6	**5**
Seit Wintersemester (WiSe) 2013/2014	Molecular aspects of cereal development MA-E,M,P-18-FW	WiSe	6	**4**
	Molecular crop science II MA-M-02-PM	WiSe	6	**3**

Leitung und Durchführung studentischer Seminare mit Übungen				
Semester	Lehrveranstaltung (Modulnummer)	Turnus	SWS (gesamt)	Beteiligung, SWS (anteilig)
Ab WiSe 2022/2023	Klassische und molekulare Genetik AGR-054	WiSe	6	**3**
Seit WiSe 2016/2017	Molecular crop science MA-04-P	WiSe	6	**1**
	Theoretical and applied aspects of genomic, transcriptomic, and proteomic analyses MA-E,M,P-21-FW	SoSe, WiSe	3	**2**
Seit SoSe 2015	Concepts of genetics MA-E,M,P-20-FW	SoSe, WiSe	3	**2**
Betreuung von Bachelor- und Masterarbeiten				
Seit SoSe 2015	Bachelorarbeit B-601	SoSe und WiSe	0,6	**0,6**
	Masterarbeit M-401	SoSe und WiSe	0,6	**0,6**

6.2 Betreute Abschlussarbeiten

6.2.1 Doktorarbeiten (Dr. agr.)

Seit 04/2020 Frau Magda Alejandra Guateque Alba (DAAD-Stipendium)

„Combinatory approach of forward and reverse genetics to identify and characterize novel maize root mutants from the European transposon insertional library *BonnMu*“

Seit 05/2019	Herr Yan Naing Win (DFG-Projekt MA 8427/1-1) „Development of a sequence-indexed collection of transposon tagged maize mutants as a novel public resource and its application in root developmental biology“

6.2.2 Masterarbeiten (M. Sc.)

Seit 05/2021	Herr Mahboob Ur Rahman Safi „Co-segregation analysis and phenotypic characterization of 11 *BonnMu*-stocks displaying irregular lateral root development“
09/2021	Herr Jonas Haberkorn „Phenotypic characterization and co-segregation analysis of the *magenta root dwarf 1* (*mrd1*) mutant of maize (*Zea mays* L.)”
01/2019	Herr Yan Naing Win „Development of a sequence-indexed collection of *Mutator*-tagged maize mutants and its application in root developmental biology“
07/2018	Frau Verena Müller „Identifizierung von potenziellen Interaktionspartnern des *lateral rootless 1* (*lrt1*) Kandidatengens von Mais“
12/2015	Herr Marcel Baer „Charakterisierung eines *lateral rootless 1* Kandidatengens von Mais“

6.2.3 Bachelorarbeiten (B. Sc.)

Seit 11/2021	Frau Linnéa Lukas „Positional cloning of the *diz1* gene in maize (*Zea mays* L.) underlying a mutant with severely twisted primary root“
09/2021	Herr Ben Wilhelm Braun „Charakterisierung der neuartigen *magenta root dwarf 1* (*mrd1*) Mutante in Mais (*Zea mays* L.) und die Klonierung des *mrd1*-Gens mittels "Bulked-Segregant"-Analyse“

01/2021	Frau Rebecca Reuss „Validation of *inositol pentakisphosphate 2-kinase* (*ipk1A*) mutations in maize using transposon-induced families of the *BonnMu*-resource“
12/2019	Frau Alina Leuschen „Novel *Mu*-tagged maize mutants: Forward genetics screen and validation“
12/2015	Herr Philipp Zimmermann „Charakterisierung ausgewählter Aux/IAA-Gene in der frühen Embryonalentwicklung von Mais (*Zea mays* L.)“
09/2015	Frau Eva Schäfer „Morphological and histological analysis of heterosis in maize (*Zea mays* L.) primary roots“

7 Sonstige Angaben

7.1 Eingeworbene Drittmittel

03/2019	DFG-Sachbeihilfe, MA 8427/1-1: „Development of a sequence-indexed collection of transposon tagged maize mutants as a novel public resource and its application in root developmental biology“ Laufzeit: 10/2019 – 09/2022 Förderbetrag: 272.300,- €

7.2 Internationale Forschungserfahrung

03/2011	Forschungsaufenthalt bei Dr. Laurie G. Smith und Dr. Steven P. Briggs, an der University of California San Diego (UCSD), San Diego USA: Proteomanalysen der Keimlingswurzeln von Mais

06/2008 – 07/2008	Forschungsaufenthalt bei Dr. Patrick S. Schnable, an der Iowa State University in Ames, Iowa, USA: Kartierung der „*roothair defective* (*rth*)“ Gene 4, 5 und 6

7.3 Gutachtertätigkeiten

Antioxidants (Impact Factor: 6.3)

BMC Genomics (Impact Factor: 3.9)

Frontiers in Plant Science (Impact Factor: 4.4)

International Journal of Molecular Sciences (Impact Factor: 5.9)

Plants (Impact Factor: 3.9)

Plant Cell Reports (Impact Factor: 4.5)

Plant Physiology and Biochemistry (Impact Factor: 4.1)

Remote Sensing (Impact Factor: 4.5)

7.4 Akademische Selbstverwaltung

Seit 10/2021	Mitglied der Studienkommission B.Sc. Agrarwissenschaften
	Vertreterin der wissenschaftlichen Mitarbeiterinnen und Mitarbeiter im INRES-Vorstand
Seit 09/2017	Beauftragte für Biologische Sicherheit (BBS) für genetische Anlagen der Universität Bonn

7.5 Weiterbildung

2017	Workshop: „Führung und Zusammenarbeit“
2015	Orientierungstag: „Gute Lehre – Schwerpunkt Landwirtschaft“

7.6 Preise

11/2012	Förderpreis des Deutschen Maiskomitees für Nachwuchs-wissenschaftler*innen, verliehen für die Dissertation

8 Erklärungen gemäß § 3 der Habilitationsordnung

Erklärungen nach § 3 Ziffer 9 und 12 der Habilitationsordnung der landwirtschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn vom 4. Dezember 2000.

1. Hiermit versichere ich, dass ich nicht gerichtlich vorbestraft bin und gegen mich kein gerichtliches Strafverfahren oder Ermittlungsverfahren der Staatsanwaltschaft anhängig ist.

2. Hiermit versichere ich, dass ich bisher noch kein Habilitationsverfahren beantragt habe.

_____ _____

Ort, Datum Unterschrift

9 Zusammenfassung der wissenschaftlichen Arbeiten

Die in der vorliegenden Habilitationsschrift zusammengefassten Forschungsarbeiten umfassen erstens molekulargenetische Untersuchungen, sowie Transkriptom- und Proteomanalysen der jungen Keimlingswurzeln von Mais (*Zea mays* L.) und zweitens die Herstellung einer Kollektion von Maismutanten für funktionell genomische Analysen. In den nachfolgenden Kapiteln sind die Referenzen mit eigener Publikationsbeteiligung farblich hervorgehoben. Detaillierte Beschreibungen der hier in den Kontext gestellten Forschungsergebnisse und deren Diskussion, in Form von Original-Publikationen, finden sich im Anhang (Kapitel 10).

9.1 Einleitung

9.1.1 Mais als Modellpflanze für funktionelle Genomanalysen

Mais gehört zur Familie der Süßgräser (*Poaceae*) und ist eine der wichtigsten und ältesten Kulturpflanzen der Erde. Bereits vor über 9000 Jahren wurde die Getreidepflanze im südlichen Mexiko domestiziert (Piperno et al., 2009). Mais wird heute weltweit auf einer Fläche von ungefähr 200 Millionen Hektar angebaut, was etwa der Fläche von Mexiko entspricht. Für die Saison 2021/ 2022 prognostiziert das Landwirtschaftsministerium der Vereinigten Staaten (USDA) einen globalen Ertrag von über 1,2 Milliarden Tonnen Körnermais (https://apps.fas.usda.gov). Mais ist damit das ertragreichste Getreide weltweit, gefolgt von Weizen und Reis, für die in der gleichen Saison etwa halb so viel Ertrag vorhergesagt wird. Neben Weizen und Reis zählt Mais zu den Grundnahrungsmitteln in vielen Ländern und kommt außerdem als Viehfutter und Bio-Treibstoff zum Einsatz.

Bereits Mitte des 19. Jahrhunderts wurde Mais als Modellpflanze für genetische Analysen genutzt. Gregor Mendel konnte seine postulierten Vererbungsregeln anhand von Kreuzungsversuchen mit Mais bestätigen (Rhoades, 1984). Mehrere Eigenschaften machen Mais zu einem geeigneten Modellorganismus für genetische Studien: Zum einen kann die Pflanze ganzjährig in Klimakammern oder Gewächshäusern und im Sommer im Freiland kultiviert werden. Zum anderen ist Mais eine einhäusig getrenntgeschlechtliche (monözische) Pflanze, bei der die weiblichen und männlichen Blütenstände voneinander getrennt sind. Während sich die weiblichen Blüten an mehreren achselständigen Kolben des bis zu 3 m hohen, selten verzweigten Halms entwickeln, befindet sich die männliche Blüte (Fahne) an der Pflanzenspitze. Mais kann daher gezielt mit sich selbst bestäubt (geselbstet) oder ausgekreuzt werden und produziert eine hohe Anzahl an Nachkommen, die für genetische Analysen genutzt werden können (Candela und Hake, 2008).

Mit dem Ziel, die Gene von Mais funktionell zu charakterisieren, wurden bereits Anfang des 20. Jahrhunderts die ersten Maismutanten beschrieben (Emerson et al., 1935). In den letzten Jahrzehnten wurden viele weitere Mutationen durch Mutagenese induziert und die resultierenden Mutanten charakterisiert. Genmutationen können durch physikalische Mutagene, z. B. ionisierende Strahlung (Stadler und Sprague, 1936), Chemikalien, wie Ethylmethansulfonat (EMS, Neuffer, 1994), oder durch biologische Mutagene, z. B. Transposons (Walbot, 2000; Marcon et al., 2020), erzeugt werden. Ein Meilenstein in der Geschichte der Maisgenetik war schließlich die Genomsequenzierung der Inzuchtlinie B73 (Schnable et al., 2009; Jiao et al., 2017), die funktionelle Genomanalysen (Genomik) wesentlich vereinfacht hat. Die funktionelle Genomik zielt darauf ab Funktion, Regulation und Zusammenspiel von Genen und Proteinen auf Ebene des gesamten Genoms zu analysieren. Die Sequenzierung des Maisgenoms vereinfachte die Isolierung von Genen, die bestimmten Mutanten unterliegen, und verbesserte die Analyse von Transkriptom- und Proteomdaten. Die Inzuchtlinie B73 wurde als erste Referenzlinie ausgewählt, da sie in weiten Teilen der Vereinigten Staaten ertragreich kultiviert werden kann und für die Züchtung moderner Maislinien genutzt wird (Candela und Hake, 2008). Das Genom von B73 umfasst etwa 40 000 proteinkodierende Gene und hat eine Gesamtgröße von 2,3 Gigabasen (Schnable et al., 2009). Damit ist das Maisgenom etwa fünfmal größer als das Reisgenom. Das kann zumindest teilweise auf die hohe Anzahl mobiler DNA-Elemente (Transposons) zurückgeführt werden, die 85% des Maisgenoms (Schnable et al., 2009), aber nur etwa 35% des Reisgenoms ausmachen (International Rice Genome Sequencing Project, 2005).

Maisinzuchtlinien zeichnen sich durch eine hohe genetische Diversität aus. Das Genom von Inzuchtlinien kann z. B. in der Anzahl von Genkopien variieren oder für Gene kodieren, die in anderen Inzuchtlinien fehlen (Schnable et al., 2009; Springer et al., 2009). Gemessen an der Anzahl zusätzlicher Einzelnukleotid-Polymorphismen (SNPs, engl. *single nucleotide polymorphisms*) sind Maislinien untereinander ähnlich verschieden wie der Mensch vom Schimpansen (Buckler et al., 2006). Aufgrund der Körner-Phänotypen können Zahnmais (Dent Pool) und Hartmais (Flint Pool) unterschieden werden. Maiskörner aus dem Dent Pool besitzen ein weicheres Endosperm als die Körner des Flint Pools und weisen deshalb eine charakteristische Einkerbung auf. Maiskörner aus dem Flint Pool besitzen dagegen keine Einkerbung und sind durch eine harte, porzellanartige Außenschale gekennzeichnet (Unterseer et al., 2016). Während die US-amerikanische Inzuchtlinie B73 zum Dent Pool gehört, lassen sich europäische Linien wie Lacaune (F7, Frankreich), Gelber Badischer Landmais (DK105, Deutschland) und Lizargarate (EP1, Spanien) dem Flint Pool zuordnen. Die genetische Diversität zwischen dem Dent und dem Flint Pool wird durch die frühe geographische Trennung dieser Genotypen erklärt. Der Anbau von Flintlinien in nördlichen Breiten erforderte eine Selektion auf Kältetoleranz und auf die frühe Abreife des Saatguts

(Brown und Anderson, 1947). Vor dem Hintergrund, dass sich Inzuchtlinien durch eine hohe genetische Diversität auszeichnen, ist es interessant, die Genome unterschiedlicher Maislinien zu sequenzieren und zu mutagenisieren. Das Genom von B73 deckt bereits einen Großteil des sogenannten Pangenoms, also die Gesamtheit der Kern-Gene (engl. *core genes*) ab, die in allen Verwandtschaftsgruppen von Mais vorkommen. Durch die Sequenzierung weiterer Inzuchtlinien wird das Pangenom aber durch Gene ergänzt, die zwar nicht lebensnotwendig für die Maispflanze sind, jedoch einen genotypspezifischen Selektionsvorteil bieten könnten (Haberer et al., 2020).

9.1.2 Das Wurzelsystem von Mais

Wurzelsysteme erfüllen mit der Verankerung der Pflanze im Boden und der Wasser- sowie Nährstoffaufnahme drei wichtige Funktionen. Das Wurzelsystem monokotyledoner Modellpflanzen wie dem Mais zeichnet sich durch embryonal und postembryonal angelegte Wurzeln aus (Feldman, 1994; Hochholdinger, 2009; Marcon et al., 2013a; Hochholdinger et al., 2018a). Diese werden zu unterschiedlichen Zeitpunkten in der Pflanzenentwicklung gebildet. Die Primärwurzel und eine unterschiedliche Anzahl an Seminalwurzeln (Abbildung 1A) sind endogene Bestandteile des Embryos. Bereits wenige Tage nach der Keimung tritt die Primärwurzel aus dem basalen Meristem und die endogen angelegten Seminalwurzeln aus dem Skutellarknoten des Embryos hervor. In der postembryonalen Entwicklungsphase treten die sprossbürtigen Kron- und Stützwurzeln an den fortlaufenden Knoten der Sprossachse auf. Während sich die Kronwurzeln an den Knoten unterhalb der Erdoberfläche bereits etwa 10 Tage nach der Keimung entwickeln, wachsen die oberirdischen Stützwurzeln erst etwa sechs Wochen nach der Keimung. Eine Gemeinsamkeit aller Wurzeltypen ist die Fähigkeit postembryonale Wurzelhaare und Lateralwurzeln auszubilden (Abbildung 1A). Das trifft auch auf die Stützwurzeln zu, sobald diese ins Erdreich hineinwachsen.

In longitudinaler Richtung ist die terminale Spitze der Maiswurzel durch die Wurzelkappe gekennzeichnet, an die sich die meristematische Zone anschließt (Abbildung 1B). Die meristematische Zone besteht aus jungen, nicht differenzierten Zellen, die sich in der angrenzenden Elongationszone zunächst strecken und in der darauffolgenden Zone differenzieren (Ishikawa und Evans, 1995). In dieser sogenannten Differenzierungszone gibt es funktionell unterschiedliche Zelltypen, wie z. B. Wurzelhaare (Abbildung 1B). Wurzelhaare tragen als röhrenförmige Verlängerungen spezialisierter Epidermiszellen (Abbildung 1C) wesentlich zur Vergrößerung der resorbierenden Oberfläche der Maiswurzel und damit zur Wasser- und Nährstoffaufnahme bei (Marzec et al., 2015).

In radialer Orientierung lässt sich die Maiswurzel in den Zentralzylinder (Stele) und das die Stele umgebende Cortex-Gewebe unterteilen (Hochholdinger, 2009). Die Stele besteht aus

den Leitbündeln (Phloem und Xylem) und dem Mark, welches durch die Perizykelzellschicht vom Cortex-Gewebe abgegrenzt wird. Das Cortex-Gewebe ist aus mehreren Zellschichten aufgebaut. Von innen nach außen betrachtet besteht es aus einem Ring Endodermiszellen, gefolgt vom mehrzelllagigen kortikalen Parenchym. Das kortikale Parenchym wird durch jeweils einen Ring Exodermis- und Epidermiszellen abgeschlossen (Abbildung 1C). Die Epidermis verbindet die Wurzel mit der umgebenden Rhizosphäre.

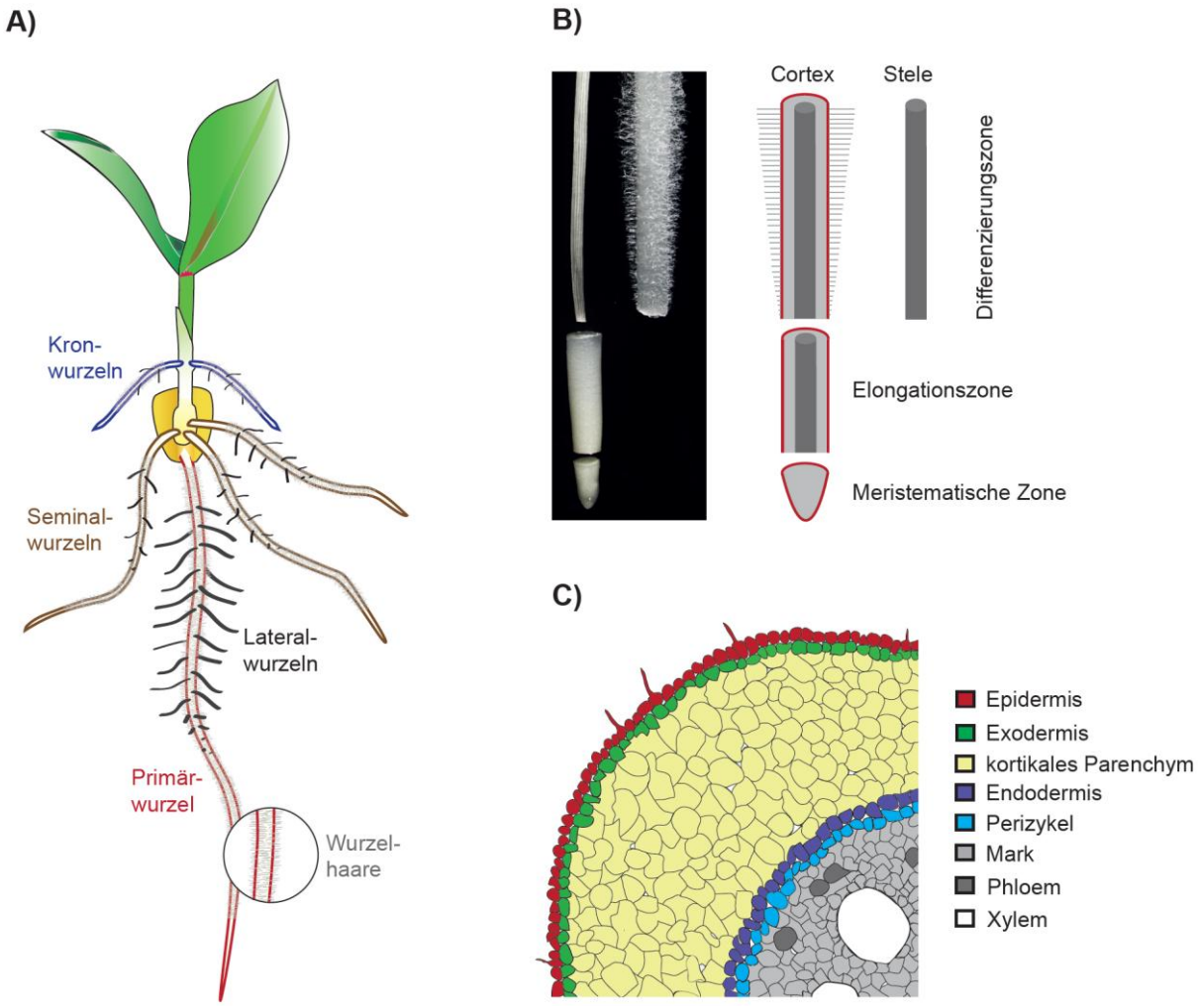


Abbildung 1. Keimlingswurzeln und Primärwurzelgewebe bei Mais. **A)** Das junge Wurzelsystem eines 10 Tage alten Maiskeimlings besteht aus embryonalen Primär- und Seminalwurzeln, sowie postembryonalen sprossbürtigen Kronwurzeln, die jeweils Lateralwurzeln und Wurzelhaare ausbilden (verändert nach Hochholdinger et al., 2018b). **B)** Die Gewebe einer 2-4 cm langen Primärwurzel lassen sich manuell in die meristematische Zone, die Elongationszone, sowie in Cortex und Stele der Differenzierungszone zerlegen (verändert nach Marcon et al., 2015). **C)** Querschnitt durch die Primärwurzel. Der innenliegende Zentralzylinder besteht aus Mark, Phloem, Xylem und der abschließenden Perizykel-Zellschicht. Das kortikale Parenchym des außenliegenden Cortex-Gewebes ist nach innen durch die Endodermis begrenzt und wird durch Exodermis sowie die Epidermis abgeschlossen.

Obwohl alle Maiswurzeltypen anatomisch ähnlich aufgebaut sind, können Unterschiede in der Zellgröße, sowie eine variable Anzahl Leitbündelelemente und kortikaler Zellschichten zu signifikanten Änderungen des Wurzeldurchmessers führen. So besitzen die embryonalen Seminalwurzeln im Vergleich zu Primär- und Kronwurzeln einen geringeren Durchmesser, was

auf eine reduzierte Anzahl an Leitbündeln in der Stele und weniger Zelllagen des kortikalen Parenchyms zurückgeführt werden kann (Tai et al., 2016; Yu et al., 2016).

9.2 Neue Studien zur funktionellen Genomik des jungen Maiswurzelsystems

In den nachfolgenden Kapiteln stelle ich meine nach der Promotion geleisteten Beiträge zur Maiswurzelgenomik in den wissenschaftlichen Kontext. Diese Arbeiten beinhalten neben molekulargenetischen Analysen zur Wurzelentwicklung, Studien zur Transkriptom- und Proteom-Plastizität der Primärwurzel, sowie die Etablierung einer Maismutanten-Ressource für funktionelle Genanalysen.

9.2.1 Molekulargenetische Analysen der Wurzelentwicklung

Die Identifizierung von Genen, die an der Wurzelentwicklung beteiligt sind, erfolgte bislang zumeist durch die funktionelle Charakterisierung von Mutanten monogen vererbter Merkmale und anschließender Klonierung des kausalen Gens. Diese Methode ist unter dem Begriff der klassischen Genetik oder Vorwärts-Genetik bekannt. Mutanten der Wurzelentwicklung können dabei durch Mutagenese induziert werden. Die Transposonmutagenese, bei der Mutanten durch die ungerichtete Insertion mobiler DNA-Elemente in Gensequenzen erzeugt werden, ist ein wesentlicher Bestandteil der vorliegenden Arbeit, der in Kapitel 9.2.4 erläutert wird. Mutagene wirken sich häufig nur auf die Funktion einzelner Gene aus und können zum Verlust spezifischer Wurzeltypen führen (zusammengefasst in Hochholdinger et al., 2018a). Die Phänotypisierung mutagenisierter Maispopulationen im großen Stil war in der Vergangenheit im Hinblick auf die Wurzelmerkmale herausfordernd: Die unterirdisch wachsende Wurzel ist nicht direkt für die Phänotypisierung zugänglich. Außerdem besitzt das Maiswurzelsystem eine beachtliche Größe, ist komplex aufgebaut und kann aufgrund verschiedener Umweltbedingungen stark variieren. Um die Identifizierung von Mutanten der Wurzelentwicklung zu ermöglichen, wurde daher eine standardisierte Methode etabliert. Dabei werden mutagenisierte Maiskörner segregierender F_2-Generationen in einem Papierrollensystem unter definierten Licht- und Temperaturbedingungen angezogen und nach etwa 10 Tagen mit bloßem Auge auf Wurzelphänotypen überprüft (Hetz et al., 1996). Molekulargenetische Analysen der so isolierten Wurzelmutanten konnten in den vergangenen Jahrzehnten bereits Schlüsselgene identifizieren, die für die Funktion von Maiswurzeln und die Regulation des Wachstums notwendig sind (zusammengefasst in Hochholdinger et al., 2018a).

9.2.1.1 Klonierung und Charakterisierung der *rth5* und *rth6* Gene

Die Durchmusterung junger, mutagenisierter Maiskeimlinge führte u. a. zur Isolierung von Mutanten, die in der Initiation oder Elongation von Wurzelhaaren beeinträchtigt sind (zusammengefasst in Hochholdinger et al., 2018a). Die Elongation der Wurzelhaare wird durch lokalisiertes Spitzenwachstum ermöglicht. Die dazu nötigen Zellwandbestandteile, z. B. Zellulose, gelangen dabei über Vesikeltransport und polare Exozytose in den extrazellulären Raum. Mit der Klonierung des *roothairless1* (*rth1*) Gens konnte in der Vergangenheit ein Gen beschrieben werden, das bei der Exozytose eine Rolle spielt (Wen et al., 2005). Das *rth1* Gen kodiert für eine SEC3-Untereinheit des Exozyst-Komplexes. Diese Untereinheit ist in der Plasmamembran lokalisiert und dort an der Fusion von Exozytosevesikeln beteiligt, die das Spitzenwachstum ermöglichen. Die Mutante *rth1* zeigt zwar eine normale Wurzelhaarinitiation, jedoch ist die Elongation dieser Zellen stark reduziert. Eine weitere Wurzelhaarmutante, *rth3*, ist ebenfalls in der Elongation der Wurzelhaare beeinträchtigt (Hochholdinger et al., 2008). Das *rth3* Gen kodiert für ein COBRA-like Protein, welches an der Organisation der synthetisierten Zellulose beteiligt ist und somit die Zellverlängerung von Wurzelhaaren unterstützt.

Im Rahmen der vorliegenden Arbeit war ich an der Klonierung und Charakterisierung der *rth5* und *rth6* Gene beteiligt, die neuartigen Wurzelhaarmutanten unterliegen (Nestler et al., 2014; Li et al., 2016).

Die Maismutante *rth5*, die durch EMS-Mutagenese erzeugt wurde, weist verkürzte und eine reduzierte Anzahl von Wurzelhaaren an allen Wurzeltypen auf (Nestler et al., 2014). Letzteres deutete darauf hin, dass das RTH5 Protein eine Rolle in der Wurzelhaarinitiation spielen könnte. Markergestützte Kartierungsexperimente, basierend auf einer *Bulked Segregant* Analyse (BSA; Michelmore et al., 1991), konnten das *rth5* Gen auf dem langen Arm von Chromosom 3 lokalisieren. Für die verwendete BSA-Methode wurde anhand der Phänotypen zunächst ein Wildtyp-Pool und ein Mutanten-Pool hergestellt und die DNA aus dem jeweiligen Pool isoliert. Anschließend konnten mittels DNA-Sequenzierung im Hochdurchsatz genetische Marker (hier: SNPs und Insertions-Deletions-Polymorphismen) identifiziert werden, die mit dem Mutanten-Pool, nicht jedoch mit dem Wildtyp-Pool, assoziiert sind. Mit Hilfe der genetischen Marker, die mit der Mutation verlinkt sind, konnte das Kartierungsintervall auf eine 280 kb große Region eingeengt werden, auf der fünf proteinkodierende Gene liegen. Eines dieser Gene, mit der Identifzierungsnummer Zm00001eb148780, enthielt am 3‘-Ende einen durch EMS induzierten Basenaustausch von Guanin zu Adenin. Diese Mutation führte zu einem Aminosäureaustausch, der vermutlich für ein nicht-funktionales RTH5 Protein verantwortlich ist. Unabhängige Mutantenallele, die jeweils eine Transposoninsertion in *rth5* aufwiesen, zeigten einen ähnlichen Wurzelhaarphänotyp wie das Referenzallel *rth5* und

bestätigten das Kandidatengen. Das *rth5* Gen kodiert für eine NADPH Oxidase, die in Wurzelhaaren stark exprimiert wird (Nestler et al., 2014). NADPH Oxidasen produzieren apoplastisches Superoxid, eine reaktive Sauerstoffspezies (ROS, engl. *reactive oxygen species*), die durch den Elektronentransfer von NADPH an molekularen Sauerstoff gebildet wird. Superoxid und seine reaktiven Permutationsprodukte Wasserstoffperoxid und insbesondere Hydroxylradikale verursachen eine Lockerung der Zellwand, durch die Spaltung der darin enthaltenden Polysaccharide wie z. B. Zellulose. Dieser Prozess ermöglicht das Zellwachstum. In Lokalisierungsexperimenten konnte die Bildung von ROS in den Spitzen wachsender Wurzelhaare von Wildtypkeimlingen gezeigt werden. In Wurzelhaarspitzen von *rth5* Mutanten konnten ROS dagegen seltener nachgewiesen werden, was auf das mutierte RTH5 Protein zurückgeführt werden kann.

Die zweite Wurzelhaarmutante, *rth6,* wurde bei der Durchmusterung einer F_2-Maispopulation entdeckt, die durch Transposonmutagenese hergestellt wurde. *Rth6* Mutanten sind in der Elongation der Wurzelhaarzellen inhibiert (Li et al., 2016). Die Klonierung des *rth6* Gens erfolgte durch eine Kombination einer *Bulked Segregant* RNA-Sequenzierung (BSR-Seq; Liu et al., 2012) und anschließendem *Seq-Walking*. BSR-Seq ist eine Modifizierung der BSA-Methode, bei der polymorphe Marker (z. B. SNPs) in RNA-Seq-Daten beim Vergleich von Mutanten- mit Wildtyp-Pools identifiziert werden, um ein Gen in einer Population zu kartieren. Polymorphe Marker, die mit der betreffenden Mutation im *rth6* Gen verlinkt sind, dominierten im *rth6* Mutanten-Pool, nicht jedoch im Wildtyp-Pool. Mit Hilfe des BSR-Seq-Verfahrens konnte das *rth6* Gen zunächst in der Zentromerregion von Chromosom 1 lokalisiert werden. Die anschließende *Seq-Walking*-Methode ermöglichte die Amplifizierung und Sequenzierung der Transposon-flankierenden DNA-Abschnitte im *rth6* Mutanten-Pool. Insgesamt konnten 29 Transposoninsertionen im Mutanten-Pool identifiziert werden, die innerhalb des BSR-Seq-Kartierungsintervalls liegen, aber nicht im Referenzgenom der Inzuchtlinie B73 vorhanden sind. Besonders viele der im Hochdurchsatz sequenzierten kurzen DNA-Abschnitte (engl. *reads*) konnten für eine Insertionsstelle innerhalb des Gens Zm00001eb025100 nachgewiesen werden. Dass es sich bei Zm00001eb025100 um *rth6* handelt, wurde schließlich durch die Sequenzierung unabhängiger Mutantenallele bestätigt (Li et al., 2016). *Rth6* kodiert für eine CELLULOSE SYNTHASE-LIKE D5 (CSLD5), die aus acht Transmembrandomänen besteht, welche in Rosettenform an der Plasmamembran lokalisiert sind.

Die Funktion von RTH6 als Zellulosesynthase steht in direktem Zusammenhang mit RTH5 und RTH3. Nach Auflockerung der Zellwand durch die RTH5 bedingte Bildung von ROS katalysiert RTH6 die Neusynthese von Zellulose an der Plasmamembran. Die Zellulose wird auf Seite der Zellwand freigesetzt und schließlich durch RTH3 in der wachsenden Wurzelhaarspitze zellverlängernd angeordnet. Unabhängig davon bildet die SEC3-Untereinheit RTH1 in der Plasmamembran die Kontaktstelle für den Exozyst-Komplex und unterstützt damit das

exozytose-gesteuerte Spitzenwachstum von Wurzelhaaren (zusammengefasst in Hochholdinger et al., 2018a).

9.2.1.2 Eigenschaften der homeologen Gene *rum1* und *rul1*

Die Mutante *rootless with undetectable meristems 1* (*rum1*) wurde in einer F_2-Population identifiziert, die durch Transposoninsertion mutagenisiert wurde. Die *rum1* Mutante zeichnet sich durch das Fehlen der Seminalwurzeln und der Lateralwurzeln an der Primärwurzel aus (Woll et al., 2005). Das *rum1* Gen kodiert für ein Aux/IAA Protein, welches für die Initiation beider Wurzeltypen notwendig ist (von Behrens et al., 2011). Aux/IAA Proteine interagieren im Zellkern mit *Auxin Response* Faktoren (ARFs), die als Transkriptionsfaktoren die Aktivität von *Auxin Response* Genen regulieren. Das RUM1 Protein interagiert dabei mit ARF25 und ARF34. Entsprechend einem vorgeschlagenen Modell fungiert RUM1 als Transkriptionsrepressor in Zellen des Perizykels, aus dem die Lateralwurzeln hervorgehen: Bei niedriger Auxinkonzentration im Perizykel bindet RUM1 an ARF25 und ARF34, die ihre Zielgene dann nicht ansteuern können. In der Folge werden keine Lateralwurzeln gebildet. Bei hoher Auxinkonzentration geht RUM1 eine Interaktion mit dem SCF^{TIR} Komplex ein und wird dem Abbau durch das 26S Proteasom zugeführt. Gestützt wird dieses Modell durch Experimente, die eine kurze Halbwertszeit von RUM1 bei hohem Auxinlevel nachweisen konnten (von Behrens et al., 2011). Als Folge des Abbaus von RUM1 können die ARF Proteine nun die Expression entsprechender *Auxin Response* Gene aktivieren und damit die Lateralwurzelbildung initiieren.

Das Maisgenom beherbergt auch ein homeologes Gen von *rum1*, das *rum1-like1* (*rul1*) genannt wird. Homeologe Gene sind durch Chromosomenduplikation entstanden und zeichnen sich durch hohe Sequenzübereinstimmung aus. Die Aminosäuresequenz von RUM1 ist zu 92% identisch mit dem homeologen Protein RUM1-LIKE1 (RUL1; von Behrens et al., 2011). Beide Proteine besitzen vier konservierte Domänen, die für Aux/IAA Proteine typisch sind. Die Domäne I ist an der transkriptionellen Repression beteiligt. Domäne II ist für die Interaktion mit dem SCF^{TIR} Komplex notwendig und trägt mit Hilfe der sogenannten Degronsequenz zur Instabilität des Aux/IAA Proteins bei. Über die beiden anderen Domänen III und IV gehen RUM1 und RUL1 Interaktionen mit anderen Aux/IAA Proteinen oder ARF Proteinen ein. Im mutierten *rum1* Gen liegt eine Transposoninsertion vor, die zu einem alternativen Spleißprodukt führt. Das mutierte rum1 Protein zeichnet sich durch eine fast vollständige Deletion der Domäne II, inklusive ihrer Degronsequenz, aus. Das führt zu einer Stabilisierung des mutierten Proteins. Es wird vermutet, dass durch Proteinstabilisierung von rum1, und damit Stabilisierung des rum1/ARF Komplexes, keine Aktivierung der nachgeschalteten *Auxin Response* Gene erfolgen kann und so die Lateralwurzelbildung in

rum1 Mutanten unterdrückt wird (von Behrens et al., 2011; zusammengefasst in Hochholdinger et al., 2018a).

Im Rahmen eines Projekts war ich an zwei Studien beteiligt, in denen die molekulargenetischen Analysen der *rum1* Mutante fortgeführt wurden. In einer Studie stand die Transkriptomanalyse von *rum1* im Vergleich zum Wildtyp im Fokus (Zhang et al., 2014). Ergebnisse dieser Untersuchung werden im Kapitel 9.2.2.2 erläutert. In der zweiten Studie wurden konservierte und spezifische Funktionen der beiden homeologen Aux/IAA Proteine RUM1 und RUL1 charakterisiert (Zhang et al., 2016), die im Folgenden beschrieben werden.

Bislang sind in Mais nur wenige homeologe Genpaare, die durch Genomduplikation entstanden sind, funktionell untersucht worden. Darunter sind die Mutanten *dwarf plant 8* (*d8*) und *dwarf plant 9* (*d9*), die jeweils unabhängig einen Zwergwuchs verursachen (Winkler und Freeling, 1994). Beide Proteine, D8 und D9, kodieren für DELLA Proteine, die als negative Regulatoren des Gibberellin-Signalwegs bekannt sind (Lawit et al., 2010). D8 und D9 zeigen unterschiedliche Expressionsprofile. Während D8 bevorzugt in meristematischem Gewebe exprimiert wird, konnte D9 vor allem in ausdifferenzierten Zellen detektiert werden (Lawit et al., 2010).

Da die Homelogen RUM1 und RUL1 jeweils ein Kernlokalisierungssignal besitzen, wurden im ersten Schritt der Analyse GFP-Fusionsproteine generiert und die Lokalisierung der beiden Proteine im Zellkern bestätigt (von Behrens et al., 2011; Zhang et al., 2016). Eine weitere gemeinsame Eigenschaft von RUM1 und RUL1 ist ihre relativ kurze Halbwertszeit von etwa 23 Minuten, nach Zugabe von Auxin. Beide Proteine können jedoch durch spezifische Mutationen in ihrer Degronsequenz der Domäne II stabilisiert werden. Um die Stabilisierung des rul1-GFP-Fusionsproteins nachzuweisen, wurde die Degronsequenz von *rul1* mutiert und Fluoreszenzmessungen *in vivo* durchgeführt.

Erste Unterschiede zwischen den homeologen Genen wurden in Expressionsanalysen festgestellt. Mittels qRT-PCR (engl. *quantitative real-time*) wurde die Expression von *rum1* und *rul1* in 12 verschiedenen Geweben und unterschiedlichen Entwicklungsstadien gemessen und miteinander verglichen. Dabei zeigte sich, dass *rul1* in der Primärwurzel und anderen Geweben etwa 84-mal stärker als *rum1* exprimiert war (Zhang et al., 2016). Darüber hinaus konnten mit Hilfe bimolekularer Fluoreszenzkomplementation (biFC) und dem Hefe-Zwei-Hybrid-System Interaktionspartner von RUL1 identifiziert werden. Obwohl RUL1 und RUM1 miteinander Homodimere und jeweils Heterodimere mit ARF25 und ARF34 bilden können, interagieren sie auch mit unterschiedlichen Proteinen. Als ein Beispiel sei das RUM1-ASSOCIATED PROTEIN 1 (RAP1) genannt, das spezifisch mit RUM1, aber nicht mit RUL1 interagiert. Diese Ergebnisse deuten abschließend darauf hin, dass RUM1 und RUL1 trotz

ihrer hoch konservierten Sequenz und ihrer Funktion als Aux/IAA Proteine zumindest teilweise unterschiedliche Funktionen haben könnten.

9.2.2 Transkriptomanalysen der Wurzelentwicklung von Mais

Die RNA-Sequenzierung (RNA-Seq) ermöglicht weitreichende Einblicke in die Expression sämtlicher Gene eines Genoms. Diese Technik hat durch präzise quantitative Ermittlung globaler Expressionsmuster die Transkriptomanalysen in Mais revolutioniert (zusammengefasst in Yu et al., 2018) und ermöglicht die Identifizierung aller exprimierten Transkripte in einer Probe und dadurch einen hohen Anteil differentiell exprimierter Gene verschiedener Proben.

In diesem Kapitel werden eigene Beiträge zu Transkriptomstudien erläutert, die erstens Aufschluss über die Funktion der komplexen Maiswurzel-Architektur geben, zweitens Aspekte des Wurzelsystems unter Wasserdefizit und nach Nitratbehandlung untersucht haben und drittens die Entwicklung von Maishybriden beleuchten. Eine Übersicht der Transkriptomstudien, an denen ich nach meiner Promotion beteiligt war, ist in Tabelle 1 gezeigt.

Tabelle 1. Überblick über Transkriptomstudien an denen ich nach der Promotion beteiligt war. Sie stehen im Zusammenhang mit verschiedenen Wurzelmutanten, Wurzeltypen, Nitratbehandlung, Wasserdefizitstress und Heterosis.

Untersuchte Wurzel/ Wurzelgewebe	**Faktor**	**Erhöhte Expression**	**Referenzen**
Primär-, Seminal-, Kronwurzel	Wurzeltyp	stress-responsiver Gene und Transkriptionsregulatoren in Seminalwurzeln	Tai et al., 2016
Perizykelzellen der Primär-, Seminal-, Kron-, Stützwurzel	Nitratbehandlung	Nitrat-responsiver Gene in Verbindung mit Zellzyklus-relevanten Genen in Stützwurzeln	Yu et al., 2016
Primärwurzel	Wurzelmutante *rum1* versus WT	Auxin-induzierter Gene im WT von Lignin-Biosynthese-Genen in *rum1*	Zhang et al., 2016
Primärwurzel/ Meristematische Zone, Elongationszone, Cortex und Stele	Wasser-defizitstress	reaktiver Sauerstoffspezies von Zellwand-Biosynthesegenen in der meristematischen Zone	Opitz et al., 2014; Opitz et al., 2016
	Heterosis	von nicht-syntänischen und SPE-Genen	Paschold et al., 2014; Baldauf et al., 2016; Marcon et al., 2016; Marcon et al., 2017; Baldauf et al., 2018

9.2.2.1 Transkriptomprofile verschiedener Wurzeltypen und zellspezifische RNA-Seq-Analysen

Die Bildung embryonaler und postembryonaler Wurzeltypen wird genetisch kontrolliert (Kapitel 9.2.1). Um spezifische Genexpressionsmuster in der frühen Entwicklungsphase von Primär-, Seminal- und Kronwurzeln (Abbildung 1A) der Inzuchtlinie B73 zu identifizieren, wurden diese Wurzeltypen in einer RNA-Seq-Studie untersucht (Tai et al., 2016). Bei dieser Analyse stellte sich heraus, dass die Genexpressionsmuster in der embryonalen Primärwurzel mit denen in postembryonalen Kronwurzeln vergleichbar sind. Eine funktionelle Kategorisierung der exprimierten Gene zeigte, dass insbesondere die Zellwandbiosynthese sowie zelluläre Umbauprozesse in diesen beiden Wurzeltypen reguliert sind. Im Gegensatz dazu waren stress-responsive Gene und Transkriptionsregulatoren in Seminalwurzeln überrepräsentiert (Tabelle 1). Die funktionale Spezialisierung der Seminalwurzeln spiegelte sich auch in der anatomischen Struktur dieses Wurzeltyps wider. Im Vergleich zu den anatomisch ähnlichen Primär- und Kronwurzeln hatte die Seminalwurzel einen geringeren Durchmesser, der auf eine reduzierte Zahl von Zellschichten im Cortexgewebe zurückgeführt werden konnte (Tai et al., 2016; Yu et al., 2016).

Vor etwa 9000 Jahren begann die Züchtung moderner Maissorten aus dem Wildgras Teosinte (*Zea mays* ssp. *parviglumis*; Matsuoka et al., 2002). Während heutige Maissorten durchschnittlich vier Seminalwurzeln aufweisen, entwickelt ein Großteil der Teosintesorten diese Wurzeln nicht (Burton et al., 2013; Tai et al., 2016). In ähnlicher Weise trifft das auch auf domestizierte Gerste- und Weizensorten zu, die jeweils mehr Seminalwurzeln als ihre Vorfahren ausbilden (Robertson et al., 1979; Grando und Ceccarelli, 1995). Seminalwurzeln könnten sich im Laufe der Domestizierung von Mais durch Selektion durchgesetzt haben, um die Pflanze z. B. an veränderte Umweltbedingungen anzupassen. In diesem Zusammenhang steht eine Studie, die zeigen konnte, dass eine hohe Anzahl von Seminalwurzeln mit Trockentoleranz bei Maiskeimlingen korreliert (Qayyum et al., 2012).

Neben den Analysen ganzer Wurzeltypen ermöglicht die *Laser Capture Microdissection* (LCM; Ludwig und Hochholdinger, 2014) in Kombination mit RNA-Seq zellspezifische Transkriptomprofile zu erstellen. Diese Methode wurde genutzt um die Lateralwurzelinitiation in Perizykelzellen von Primär-, Seminal-, Kron- und Stützwurzeln in Abhängigkeit von Nitrat zu untersuchen (Yu et al., 2016). Für diese Studie wurden die Maispflanzen vor der Analyse für 24 h in einem *split-root*-System mit niedriger bzw. hoher Nitratlösung behandelt. Lateralwurzeln besitzen die Fähigkeit sich an die ungleich verteilten Ressourcen im Boden, wie z. B. Stickstoff-Vorkommen, anzupassen. So kann Nitrat im Boden als Botenstoff die Bildung von Lateralwurzeln in Mais initiieren und die Elongation fördern (Yu et al., 2015). Nach der Nitratbehandlung wiesen die vier Wurzeltypen Unterschiede in der Länge und Dichte der Lateralwurzeln auf (Yu et al., 2016). Vor allem die Stützwurzeln bildeten bei hoher

Nitratkonzentration lange und stark verzweigte Lateralwurzeln aus. Die Lateralwurzeldichte der drei anderen Wurzeltypen blieb dagegen bei hoher Nitratkonzentration unverändert.

Für die RNA-Seq-Studie wurden schließlich etwa 2000 Perizykelzellen aus jeweils drei biologischen Replikaten via LCM isoliert und die extrahierte RNA mittels Hochdurchsatzsequenzierung (HTS) sequenziert. Statistische Auswertungen zeigten, dass sich nach lokaler Nitratbehandlung besonders die Genexpression in Perizykelzellen der Stützwurzeln signifikant änderte. Von den insgesamt 3640 Genen, die bei hoher Nitratkonzentration differentiell in den Perizykelzellen aller Wurzeltypen exprimiert waren, konnte die Mehrzahl von 3046 Genen ausschließlich den Stützwurzeln zugeordnet werden (Yu et al., 2016). Unter diesen nitrat-regulierten und stützwurzelspezifischen Genen ließen sich mehrere 100 Gene funktionell der Zellzyklus-Kontrolle, Zellorganisation, DNA-Synthese und Chromatinstruktur zuordnen. Die morphologische Besonderheit der Stützwurzeln als Reaktion auf eine hohe Nitratkonzentration spiegelte sich demnach auch im Perizykeltranskriptom dieses Wurzeltyps wider. Die einzigartige Genexpression in Perizykelzellen der Stützwurzeln kann die nitrat-abhängige Anpassungsfähigkeit der Lateralwurzeln beeinflussen und so zumindest teilweise die Unterschiede zu den anderen drei Wurzeltypen erklären.

9.2.2.2 Differentielle Genexpression in der Wurzelmutante *rum1* versus WT

Transkriptomanalysen können auch genutzt werden, um die differentielle Expression von Genen in Mutanten im Vergleich zum Wildtyp zu untersuchen. Um RUM1-abhängige Expressionsmuster in der Differenzierungszone von Primärwurzeln zu untersuchen, wurde eine vergleichende RNA-Seq-Studie von *rum1* und Wildtyp durchgeführt (Zhang et al., 2014). Statistische Auswertungen dieser Experimente, die in vier biologischen Replikaten durchgeführt wurden, identifizierten 2801 differentiell exprimierte Gene. Davon waren 1741 Gene in der Mutante *rum1* herunterreguliert und 1060 hochreguliert. Unter den häufigsten funktionellen Klassen befanden sich Auxin regulierte Gene wie *like auxin 1 (lax1)* und *lax2*, die für Auxininfluxtransporter kodieren. Diese waren wie *rum1* und drei *plethora* (*plt*) Gene in der *rum1* Mutante herunterreguliert. Desweiteren waren die Gene *arf8* und *arf37* in *rum1* herunterreguliert. Homologe dieser *arf* Gene aus Arabidopsis sind an der Leitgewebebildung beteiligt (Hardtke und Berleth, 1998). In diesem Zusammenhang konnten anschließend histologische Untersuchungen Defekte in der Xylemdifferenzierung von *rum1* nachweisen (Zhang et al., 2014). Dieses Ergebnis korrelierte mit der RNA-Seq-Analyse, die zeigen konnte, dass die Expression von drei Enzymen der Ligninbiosynthese in *rum1* hochreguliert war (Tabelle 1). Die erhöhte Akkumulierung von Lignin in Primärwurzeln von *rum1* wurde durch die Färbung der histologischen Wurzelquerschnitte mit Phloroglucinol-HCl nachgewiesen. Zusammengefasst konnte diese Studie Gene identifizieren, die indirekt oder direkt von RUM1 reguliert werden. Die Kombination von RNA-Seq-Experimenten und histologischen

Untersuchungen zeigte neue Funktionen von RUM1 in der Lignifizierung und in der Entwicklung der Leitgewebe auf, die sonst unentdeckt geblieben wären.

9.2.2.3 Reaktion des Wurzel-Transkriptoms auf Wasserdefizitstress

Maispflanzen reagieren empfindlich auf Trockenheit. Dürreperioden verursachen jährlich erhebliche Ertragsverluste. Aufgrund der globalen Klimaerwärmung werden Dürreperioden zukünftig häufiger auftreten (Dai, 2013) und erfordern daher die Züchtung trockentoleranter Maissorten. Toleranter gegen Trockenstress sind z. B. Maispflanzen mit einem tiefgehenden Wurzelsystem (Ribaut et al., 2009). Kontinuierliches Längenwachstum der Primärwurzel bei niedrigem Wasserpotential ist für Mais notwendig, um tiefere Wasserschichten im Boden zu erreichen. Dabei beschreibt das Wasserpotential die Wasserverfügbarkeit im System von Wurzel und Boden. Bei Trockenheit wird das wenige Wasser im Erdreich verstärkt an Bodenpartikel gebunden und das Wasserpotential nimmt ab. Da der Transport von Wasser in die Wurzel durch einen Gradienten des Wasserpotentials angetrieben wird, ist die Wurzel bestrebt ihr Wasserpotential weiter zu verringern, damit sie Wasser vom Ort höheren Potentials aufnehmen kann. Dieser Prozess wird auch durch die Verdunstung von Wasser über die Blätter reguliert. Sinkt das Wasserpotential im Boden jedoch unter die Marke von -1,5 MPa, gelingt es der Wurzel meist nicht mehr das stark an Bodenpartikel gebundene Wasser zu extrahieren (O´Geen, 2012).

Auf physiologischer Ebene wird das Wurzelwachstum unter Wasserdefizitstress durch das Pflanzenhormon Abscisinsäure gesteuert, das den Transport von Prolin in die Wurzelspitze fördert. Die Aminosäure Prolin verursacht dort eine Erhöhung des osmotischen Drucks, der den zellulären Turgor schützt. Als Folge wird das Wasserpotential verringert, sodass der Influx von Wasser aus dem umgebenden Medium weiter möglich ist (Yamaguchi und Sharp, 2010). Im Gegensatz zu den physiologischen Prozessen, die durch Trockenstress hervorgerufen werden können, ist über Genexpressionsänderungen in der Maiswurzel infolge von Wasserdefizitstress wenig bekannt.

Unter Laborbedingungen kann Wasserdefizitstress bei Pflanzen durch die Behandlung mit osmotisch wirkendem Polyethylenglykol (PEG) 8000 simuliert werden. Um die Transkriptom-Antwort im Hinblick auf Intensität und Dauer des Wasserdefizits zu untersuchen, wurden Primärwurzeln von Mais jeweils für sechs und 24 Stunden moderatem (-0,2 Mpa) und starkem (-0,8 Mpa) PEG-vermittelten Stress ausgesetzt (Opitz et al., 2014). Änderungen in der Genexpression, d. h. differentiell exprimierte Gene wurden durch den Vergleich der Daten mit einer Kontrollgruppe von Keimlingen, die keinem Wasserdefizitstress ausgesetzt waren, ermittelt. Dabei wurde festgestellt, dass die Anzahl differentiell exprimierter Gene mit Intensität und Dauer der Wasserdefizitbehandlung zunimmt. Nach sechs Stunden leichtem Stress (-0,2

Mpa) waren lediglich etwa 250 Gene reguliert. Im Gegensatz dazu waren nach 24 Stunden starkem Stress (-0,8 Mpa) mehr als 12 800 Gene differentiell exprimiert. Der überwiegende Teil dieser Gene wurde durch die Stressbehandlung induziert. Eine funktionelle Kategorisierung der differentiell exprimierten Gene ermöglichte es, die durch Wasserdefizitstress beeinflussten Stoffwechselwege zu identifizieren. Damit konnten die differentiell exprimierten Gene vor allem mit dem Stoffwechsel der reaktiven Sauerstoffspezies (ROS) in Verbindung gebracht werden (Tabelle 1). ROS üben wichtige Funktionen als Signalmoleküle aus, können jedoch in hoher Konzentration Membranen, Makromoleküle und Nukleinsäuren schädigen (zusammengefasst in Mittler et al., 2011).

In einer Folgestudie wurde die Transkriptom-Reaktion einzelner Primärwurzelgewebe (Abbildung 1B) nach kurzer (sechs Stunden) und intensiver (-0,8 Mpa) Wasserdefizitbehandlung der Keimlinge untersucht (Opitz et al., 2016). Dabei zeigte sich, dass die Genexpression infolge der Stressbehandlung zwischen Kontrolle und den Geweben der meristematischen Zone, der Elongationszone, dem Cortex und der Stele variabel reguliert wird. So waren mehr als 75% der wasserdefizit-responsiven Gene (1501 von insgesamt 1915 regulierten Genen) gewebespezifisch reguliert. Im Gegensatz dazu waren nur etwa 3% der stress-responsiven Gene (50 Gene) einheitlich in allen vier Geweben reguliert. Erneut war der überwiegende Anteil aller differentiell exprimierten Gene durch die Stressbehandlung induziert.

Die größte Anzahl der stress-responsiven Gene entfiel auf das Cortexgewebe und die Elongationszone (Abbildung 1B). Diese stellt die Region der Wurzel dar, die durch Zellelongation Wachstum steuert und dieses sogar unter Trockenstressbedingungen aufrechterhalten kann. In der Elongationszone und im Cortex waren, wie bereits die Pilotstudie anhand der gesamten Primärwurzel zeigen konnte, besonders die Gene induziert, die eine Rolle in den Stoffwechselwegen der ROS spielen. Um schädigende ROS, u. a. Superoxid- und Hydroxyl-Radikale zu minimieren, werden enzymatisch und nicht-enzymatisch wirkende Antioxidantien gebildet (zusammengefasst in Cruz de Carvalho, 2008). Gene, die für ROS-detoxifizierende Enzyme, wie z. B. Katalasen, Peroxidasen oder Superoxiddismutasen, kodieren, waren im Cortex und in der Elongationszone nach Stressbehandlung überrepräsentiert. Weitere stress-responsive Gene, die in allen Wurzelgeweben reguliert waren, konnten den funktionellen Kategorien der Transkriptionsregulation und des Hormonstoffwechsels zugeordnet werden. Daneben waren in allen Geweben, außer der Stele, Gene zur Prolinsynthese induziert. Gene, die Enzyme zum Abbau von Prolin kodieren, waren dagegen reprimiert. Zusätzlich waren viele der differentiell exprimierten Gene in der meristematischen Wurzelspitze mit der Reorganisation der Zellwand assoziiert. Zusammenfassend machen die Ergebnisse deutlich, dass es in Primärwurzelgeweben zu

einer umfassenden Umprogrammierung des zellulären Stoffwechsels kommt, um eine Anpassung an Trockenstress zu gewährleisten.

9.2.2.4 Heterosis-assoziierte Transkriptomanalysen der Keimlingswurzeln von Maishybriden

Der Heterosiseffekt beschreibt die überlegene Leistungsfähigkeit heterozygoter Maishybride im Vergleich zu ihren parentalen homozygoten Inzuchtlinien. Das bereits im Jahr 1908 beschriebene Phänomen der Heterosis (East, 1908; Shull, 1908) dient als Grundlage der modernen Maiszüchtung, was sich darin zeigt, dass heute in Industrieländern ausschließlich Maishybridsorten angebaut werden. Hybride zeichnen sich u. a. durch hohe Erträge, schnelle Wachstumsgeschwindigkeit und Toleranz gegenüber nachteiligen Umwelteinflüssen aus.

Der Vorteil von Maishybriden zeigt sich unter dem Mikroskop bereits im jungen Embryo (Meyer et al., 2007) und setzt sich nach der Keimung fort. Embryonen der Hybride sind wenige Wochen nach der Bestäubung signifikant größer und weisen sichtbare Seminalwurzel-primordien auf, die in den Embryonen der Elternlinien noch nicht ausgeprägt sind (Marcon et al., 2010). Etwa vier Tage nach der Keimung wird der Wachstumsvorsprung in Form einer längeren Primärwurzel mit bloßem Auge erkennbar (Abbildung 2; zusammengefasst in Paschold et al., 2010; Marcon et al., 2016).

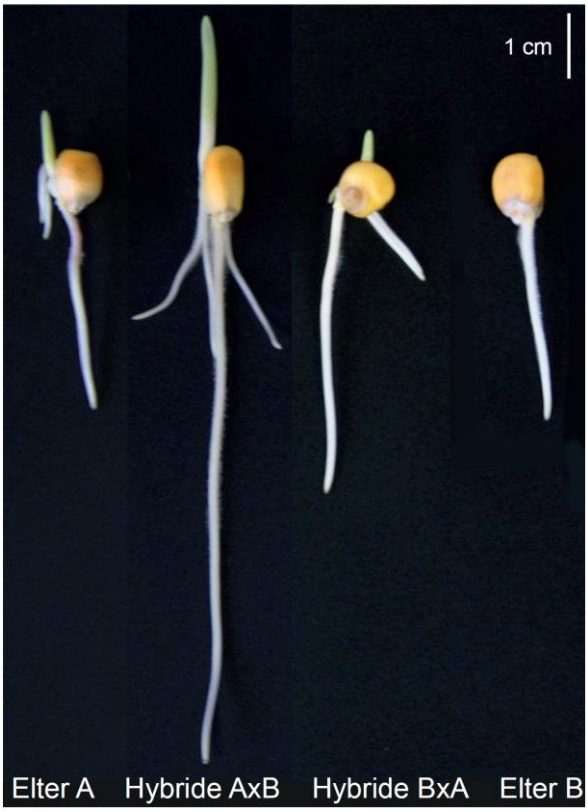


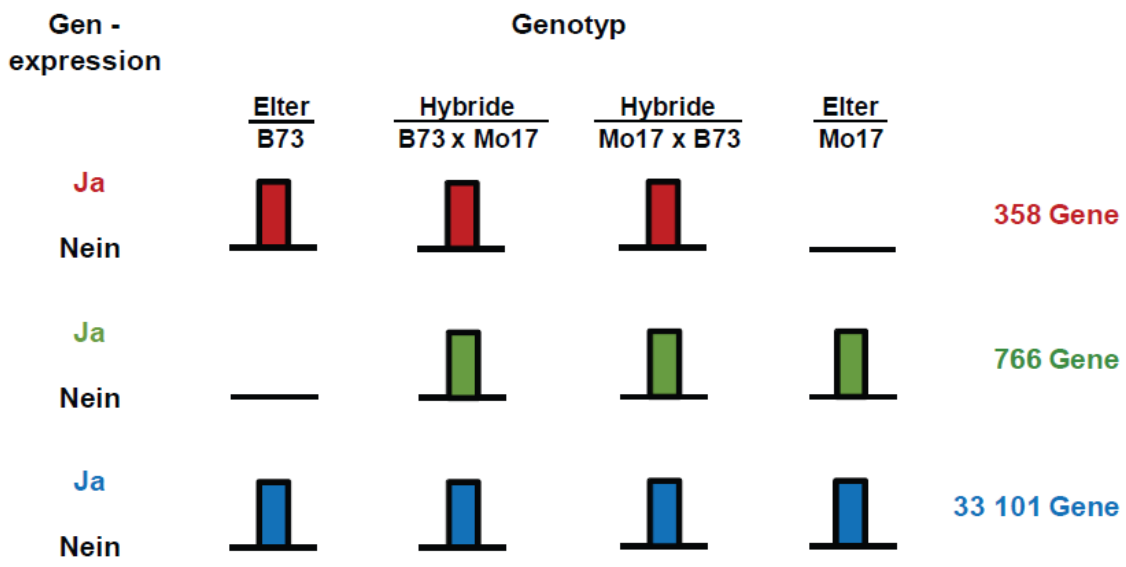
Abbildung 2. Der Heterosiseffekt gezeigt am Wachstumsvorsprungs junger Maishybride (Mitte) im Vergleich zu den jeweiligen Eltern A und B, vier Tage nach Keimung.

Da die Primärwurzel einfach strukturiert ist (Abbildung 1A, B und Abbildung 2), eignet sie sich besonders gut als Modell um die molekularen Grundlagen, die zur Überlegenheit der Hybride führen, zu entschlüsseln. Zu den ersten genetischen Erklärungsversuchen der Heterosis gehört die aus dem Jahr 1917 stammende Dominanz- oder Komplementationshypothese. Sie postuliert, dass vorteilhafte Allele in Hybridpflanzen dominant sind und das unterlegene Allel einer der Elternlinien komplementieren können (Jones, 1917). In diesem Zusammenhang steht auch die hohe genetische Diversität vieler Maisinzuchtlinien. Diese zeichnet sich z. B. durch eine unterschiedliche Anzahl an Genen oder Genkopien im Genom aus (Springer et al., 2009). Dadurch entstehen genetisch polymorphe Inzuchtlinien, die phänotypisch sehr unterschiedlich sein können. In Hybriden können folglich solche Gene, die in einer der parentalen Inzuchtlinie fehlen, durch die andere Inzuchtlinie komplementiert werden. Um zu ermitteln, ob Variationen im Genom auch zu Unterschieden in der quantitativen Genexpression von Hybriden führen, wurden Heterosis-assoziierte Wurzel-Transkriptomanalysen erstens mit unterschiedlichen Inzuchtlinien (Paschold et al., 2012; Baldauf et al., 2016; Marcon et al., 2016; Baldauf et al., 2018), zweitens mit verschiedenen Wurzelgeweben (Paschold et al., 2014) und drittens mit Wurzeln nach Wasserdefizitbehandlung (Marcon et al., 2017) durchgeführt (Tabelle 1).

9.2.2.5 Spezielle Expressionsmuster formen das Transkriptom reziproker Hybride

Eine Pilotstudie, an der ich als Doktorandin in der Arbeitsgruppe von Prof. Dr. Hochholdinger (Universität Bonn) beteiligt war, konnte einen Spezialfall der Komplementationshypothese in jungen Primärwurzeln auf Genexpressionsebene nachweisen (Paschold et al., 2012; zusammengefasst in Hochholdinger et al., 2013; Marcon et al., 2016). Ein RNA-Seq-Experiment konnte dabei zeigen, dass es in den Elternlinien B73 und Mo17 mehrere hundert Gene gibt, die nur in einer der beiden parentalen Inzuchtlinien, aber in beiden reziproken Hybriden aktiv sind (Abbildung 3A, B). Dieses Phänomen wird als *single-parent expression* (SPE) bezeichnet und beschreibt eine spezielle Form der Komplementation auf Ebene der Genexpression, die dazu führt, dass Hybride insgesamt mehr Gene als ihre Eltern exprimieren. SPE-Gene könnten somit zur überlegenen Leistungsfähigkeit der Maishybride beitragen. Zwar leistet jedes einzelne SPE-Gen vermutlich nur einen geringen Beitrag zum Heterosiseffekt, dennoch könnte die Summe an zusätzlich aktiven Genen mitverantwortlich dafür sein, dass Hybride leistungsfähiger sind als ihre Elternlinien.

A)



B)

Genotyp	B73	B73 x Mo17	Mo17 x B73	Mo17
Anzahl aktiver Gene	33 460	34 225	34 226	33 874

Abbildung 3. Komplementation der Genexpression durch SPE-Gene. **A)** Die meisten Gene (33 101) sind in allen vier untersuchten Genotypen exprimiert. Dennoch gibt es hunderte Gene (358 und 766), die in beiden Hybriden (Mitte), aber nur in einer der beiden Elternlinien aktiviert sind und folglich Komplementation auf Ebene der Genexpression zeigen. **B)** Da mehrere hundert Gene nur in einer der parentalen Inzuchtlinie, aber in beiden Hybriden exprimiert werden, sind in den Hybriden insgesamt mehr Gene aktiv als in den Elternlinien (Paschold et al., 2012; Hochholdinger et al., 2013; Marcon et al., 2016).

Neben den SPE-Genen konnten wir im Jahr 2012 weitere differentielle, sowie nicht-additive und allelische Genexpressionsmuster nachweisen, die das Transkriptom reziproker Hybride maßgeblich beeinflussen. Paarweise Vergleiche der Genexpression zwischen Hybriden und Elternlinien zeigten, dass zwischen den genetisch sehr unterschiedlichen Eltern B73 und Mo17 die höchste Zahl differentiell exprimierter Gene auftrat. Etwa 70% aller exprimierten Gene zeigten demnach zwischen den beiden Inzuchtlinien differentielle Expression. Dagegen waren zwischen den reziproken Hybriden und den einzelnen Inzucht-Hybrid-Kombinationen nur 15-55% aller Gene differentiell exprimiert. Dieses Ergebnis ließ sich dadurch erklären, dass beide Hybride 50% des jeweiligen Eltern-Genoms enthalten.

Während sich differentiell exprimierte Gene signifikant von mindestens einer der elterlichen Inzuchtlinie unterscheiden, weichen nicht-additiv exprimierte Gene signifikant vom Mittelwert der Eltern ab (Hochholdinger und Hoecker, 2007). Insgesamt waren etwa 10% aller untersuchten Gene in der Pilotstudie nicht-additiv in den reziproken Hybriden B73 x Mo17 und Mo17 x B73 exprimiert (Paschold et al., 2012).

Um allelspezifische Expressionsmuster in den Hybriden nachzuweisen, wurden die Expressionsverhältnisse der elterlichen Allele für alle identifizierten Gene ermittelt. Insgesamt zeigten 4100 Gene eine allelspezifische Expression in den reziproken Hybriden, die vom elterlichen Verhältnis der Allele abwich. Zusätzlich konnten etwa hundert Gene in den Hybriden detektiert werden, die nicht nur eine unerwartete Expression der Allele, sondern gleichzeitig eine nicht-additive Expression aufwiesen. Dieses Ergebnis verdeutlichte, dass eine veränderte Expression der Allele in Hybriden zumindest teilweise zur Nicht-Additivität führen kann.

9.2.2.6 *Single-parent expression* als genereller Mechanismus der Komplementation in verschiedenen Wurzelgeweben, unter Wasserdefizitstress und in diversen Maishybriden

Die Entdeckung der SPE-Gene und anderer extremer Expressionsklassen in Maishybriden (Paschold et al., 2012) führte zu einer Reihe von Folgestudien, die im weiteren Verlauf der vorliegenden Arbeit zusammengefasst werden. Das Hauptziel dieser Untersuchungen war es, die speziellen Expressionsmuster der SPE-Gene als generellen Komplementationsmechanismus in Maishybriden zu entschlüsseln. In einem breiten Spektrum von Heterosis-assoziierten RNA-Seq-Analysen wurden extreme Expressionsklassen untersucht: erstens in verschiedenen Primärwurzelgeweben (Paschold et al., 2014; Baldauf et al., 2016; zusammengefasst in Marcon et al., 2016), zweitens nach Wasserdefizitbehandlung (Marcon et al., 2017) und drittens in verschiedenen Entwicklungsstadien der Primärwurzel (Baldauf et al., 2018).

9.2.2.7 Gewebespezifische Dynamik der Genexpression in Hybriden

Um zu prüfen, ob das Phänomen der *single-parent expression* in einzelnen Wurzelgeweben beobachtet werden kann, wurden in einer Folgestudie Transkriptomprofile der vier Primärwurzelzonen erstellt (Abbildung 1B; Paschold et al., 2014). In bidirektionalen Vergleichen der vier Genotypen (B73, Mo17, B73 x Mo17, Mo17 x B73) mit den vier Geweben konnten hunderte gewebespezifische SPE-Gene identifiziert werden. Die meisten SPE-Gene kamen in der Stele (1287 SPE-Gene) der reziproken Hybride vor, die Wenigsten in der Elongationszone (1037 SPE-Gene). Insgesamt 1096 dieser Gene zeigten SPE_B-Expressionsmuster, d. h. sie waren in beiden Hybriden und in der Elternlinie B73 exprimiert.

SPE_M-Expression, d. h. Genexpression, die ausschließlich in der Elternlinie Mo17, aber in beiden Hybriden detektiert wurde, konnte für 1245 Gene nachgewiesen werden. Als Folge dieser extremen Komplementation auf Expressionsebene waren mehr als 440 zusätzliche Gene in den vier Geweben beider Hybride, verglichen zu den Elternlinien, aktiv. Lediglich 20% der SPE-Gene, zeigten dieses Muster in allen vier untersuchten Geweben. Dagegen war ein beträchtlicher Anteil der identifizierten SPE-Muster (80%; 1877 der 2341 SPE-Gene) gewebespezifisch. Im Cortexgewebe junger Primärwurzeln wurden beispielsweise 184 SPE_B- und 210 SPE_M-Gene identifiziert, die in den drei anderen Geweben nicht als SPE-Gene ermittelt wurden (Paschold et al., 2014). Zusammenfassend konnte festgestellt werden, dass die SPE-Gene einem dynamischen gewebespezifischen Expressionsmuster entsprechen. Viele dieser Gene folgten zwar einem SPE-Muster in bestimmten Geweben, zeigten jedoch in anderen Geweben ein anderes, nicht-SPE-spezifisches Expressionsmuster. Das lässt die Schlussfolgerung zu, dass die Expression von SPE-Genen in einzelnen Wurzelgeweben einer komplexen Regulation unterliegt.

Neben gewebespezifischen SPE-Genen wurden weitere Expressionsänderungen in den reziproken Hybriden im Vergleich zu den Elternlinien B73 und Mo17 untersucht (Baldauf et al., 2016). Dabei stellte sich heraus, dass die Genexpressionsmuster die Verwandtschafts-beziehung zwischen den einzelnen Wurzelgeweben und Genotypen widerspiegeln. Im Hinblick auf die Verteilung der differentiell exprimierten Gene, die zwischen den vier Genotypen und den vier Wurzelgeweben bestimmt wurden, zeigte sich dieses Phänomen besonders deutlich: Zwischen den beiden genetisch verschiedenen Inzuchtlinien B73 und Mo17 wurde jeweils die größte Anzahl an differentiell exprimierten Genen beobachtet. Dagegen waren zwischen den beiden reziproken Hybriden, die aus jeweils 50% des elterlichen Genoms bestehen, keine Gene differentiell exprimiert. In den jeweiligen Vergleichen von Hybriden und Elternlinien lag die Anzahl differentiell exprimierter Gene zwischen diesen Werten. Zusätzlich wurden hunderte nicht-additiv exprimierte Gene und allelspezifische Expressionsmuster in den Hybriden identifiziert, die eine erhebliche gewebespezifische Dynamik zeigten. Insgesamt deuteten die Ergebnisse darauf hin, dass die analysierten Gewebe unterschiedliche funktionelle Beiträge zum Heterosis-Effekt in Keimlingswurzeln leisten könnten.

9.2.2.8 Stabilität der SPE-Gene bei Wasserdefizitstress

In einer weiteren Untersuchung wurden Transkriptomprofile von Primärwurzeln erstellt, die einer Wasserdefizitbehandlung ausgesetzt waren (Marcon et al., 2017). Ähnlich wie in Kapitel 9.2.2.3 erläutert wurden junge Keimlinge der Inzuchtlinien B73 und Mo17 sowie die Hybriden B73 x Mo17 und Mo17 x B73 kurzem (sechs Stunden), aber intensivem Wasserdefizitstress (PEG 8000 Lösung, -0,8 Mpa) ausgesetzt oder verblieben unter Kontrollbedingungen in

Wasser. Anschließend erfolgte die RNA-Seq-Analyse von 2-4 cm langen Primärwurzeln der jeweiligen Kombination aus Genotyp und Behandlung (Kontrolle/ Wasserdefizitstress) in vier biologischen Replikaten. Im Mittelpunkt der nachfolgenden Untersuchungen stand die Frage, ob und welche extremen Expressionsmuster unter schwankenden Umweltbedingungen konserviert sind.

Zunächst stand die Analyse der Gene im Vordergrund, die durch die PEG-Behandlung im Vergleich zur Kontrollbehandlung in den einzelnen Geweben induziert oder reprimiert wurden. Die geringste Anzahl solcher stress-responsiven Gene (229 Gene) wurde in der Inzuchtlinie Mo17 festgestellt. Hier waren 140 Gene hoch- und 89 Gene herunterreguliert. Die höchste Anzahl stress-regulierter Gene (902 Gene) wurde in der Hybride B73 x Mo17 identifiziert. Viele der durch den Stress regulierten Gene unterlagen dabei einer genotypspezifischen Regulation. So waren mehr als die Hälfte dieser Gene ausschließlich in den reziproken Hybriden reguliert, nicht jedoch in den Elternlinien (Marcon et al., 2017). Ein beträchtlicher Anteil der stress-responsiven Gene (mehr als 60%) war in den jeweiligen Genotypen hochreguliert. Dieses Ergebnis impliziert, dass eine reduzierte Wasserverfügbarkeit zur Aktivierung der Genexpression führen könnte. Ähnliche Tendenzen konnten wir bereits in Primärwurzeln und in Primärwurzelgeweben von B73 feststellen, in denen bis zu 88% der stress-responsiven Gene induziert waren (Opitz et al., 2014; Opitz et al., 2016).

Darüber hinaus stellten die Experimente ähnlich wie vorangegangene Heterosis-assoziierte Transkriptomstudien differentielle und nicht-additive Genexpression sowie SPE-Muster fest (Paschold et al., 2012; Paschold et al., 2014; Baldauf et al., 2016). Um differentiell exprimierte Gene zwischen den einzelnen Genotypen zu identifizieren, wurden jeweils sechs paarweise Vergleiche der Kontrollgruppe und der Gruppe nach Wasserdefizitstress angestellt. Die meisten differentiell exprimierten Gene wurden unabhängig von der Behandlung zwischen den genetisch entfernt verwandten Genotypen B73 und Mo17 beobachtet. Die geringste Anzahl solcher Gene gab es dagegen in den reziproken Hybriden. Diese genotyp-assoziierten Tendenzen waren damit vergleichbar mit anderen Heterosis-assoziierten Transkriptomstudien (Paschold et al., 2012; Paschold et al., 2014; Baldauf et al., 2018). In allen paarweisen Vergleichen zeigte sich außerdem deutlich, dass bis zu 86% der differentiellen Genexpressionsmuster unabhängig der Wasserdefizitbehandlung konserviert waren (Marcon et al., 2017). Es gab jedoch auch hunderte von Genen, die nur unter Kontrollbedingung differentiell, aber unter Stressbedingung stabil exprimiert, oder umgekehrt nur unter Stressbedingung differentiell, aber unter Kontrollbedingung stabil waren.

Für tausende der identifizierten Gene konnte ein nicht-additives Expressionsmuster in Hybriden unter Kontrollbedingung und nach Behandlung mit PEG festgestellt werden. Bemerkenswert war, dass fast 50% der nicht-additiv exprimierten Gene dieses Muster sowohl unter Kontroll- als auch unter Stressbedingung zeigte. Dieses Ergebnis macht deutlich, dass

ein beträchtlicher Anteil nicht-additiv exprimierter Gene in Hybriden robust gegenüber Wasserdefizitstress ist. Diese Gene könnten daher bei der frühen Ausprägung von Heterosis eine Rolle spielen.

In dieser Untersuchung wurden auch mehrere hundert SPE-Gene identifiziert, die eine extreme Form der differentiellen Genexpression darstellen. Bemerkenswert war, dass unter allen genotypspezifischen Expressionsmustern nur die SPE-Muster eine hohe Stabilität ungeachtet mangelnder Wasserverfügbarkeit aufwiesen. So zeigten 75% aller SPE-Gene, die unter Kontrollbedingungen identifiziert wurden, dieses Expressionsmuster auch nach der PEG-Behandlung (Marcon et al., 2017). Im Gegensatz dazu blieben nur 17% der übrigen genotypspezifischen Expressionsmuster nach der Behandlung mit PEG gleich. Die überwiegende Mehrheit dieser genotypspezifisch exprimierten Gene (83%) änderte infolge der PEG-Behandlung ihr Expressionsmuster. Aufgrund der SPE-Gene waren in den reziproken Hybriden unter Kontroll- und Wasserdefizitbedingungen mehr Gene aktiv als in den beiden Elternlinien, was sich vorteilhaft auf das Wachstum der Hybridpflanzen auswirken könnte.

9.2.2.9 Genexpressionsprofile genetisch unterschiedlicher Inzuchtlinien und deren Hybride während der Primärwurzelentwicklung

In einer weiteren Studie wurden Genexpressionsänderungen in sieben genetisch unterschiedlichen Inzuchtlinien und deren Hybride während drei verschiedener Entwicklungsstadien des Primärwurzelwachstums analysiert (Baldauf et al., 2018). Ziel war es zu prüfen, ob die extreme SPE-Komplementation ein genereller Mechanismus ist, der in allen Genotypen und Entwicklungsstadien beobachtet werden kann.

Unter Berücksichtigung phylogenetischer Analysen von US-amerikanischen Inzuchtlinien (Lorenz und Hoegemeyer, 2013) wurden folgende Elternlinien aus sechs verschiedenen Verwandtschaftsgruppen ausgewählt, um eine möglichst hohe genetische Diversität zu erreichen: B73, A554, H84, H99, Mo17, Oh43 und W64A. Die Inzuchtlinie B73 diente bei der Herstellung der Hybride B73 x A554, B73 x H84, B73 x H99, B73 x Mo17, B73 x Oh43 und B73 x W64A jeweils als Mutterpflanze. Die Primärwurzeln aller Elternlinien und ihrer jeweiligen Hybride wurden in drei Entwicklungsstadien entsprechend ihrer Länge (2-4 cm; 6-8 cm; 10-12 cm) untersucht. Jede Kombination aus Genotyp und Entwicklungsstadium wurde in mindestens vier biologischen Replikaten einer RNA-Seq-Analyse unterzogen, was einer Probengröße von 180 entsprach (Baldauf et al., 2018).

Durchschnittlich konnten etwa 1000 SPE-Gene in allen Hybridkombinationen und über alle Entwicklungsstadien hinweg beobachtet werden. Dieses Ergebnis bestätigte die Hypothese, dass die Komplementation der Expression durch SPE-Gene ein genereller Mechanismus ist, der über verschiedene Genotypen hinweg konserviert ist. Entsprechend waren in den

untersuchten Hybriden etwa 600 Gene mehr aktiv als in den jeweiligen Elternlinien. Darüber hinaus wurde beobachtet, dass bis zu 66% aller SPE-Muster über die drei untersuchten Entwicklungsstadien eines Genotyps konserviert waren. Im Gegensatz dazu zeigten nur etwa 6% der Gene in den unterschiedlichen Genotyp-Kombinationen ein robustes SPE-Muster. Dieses Ergebnis lässt den Schluss zu, dass SPE-Gene zwar genotypspezifisch sind, jedoch über verschiedene Entwicklungsstadien der Primärwurzel stabil als SPE-Muster nachgewiesen werden können.

9.2.2.10 Komplementation der Genexpression in evolutionär jungen nicht-syntänischen Genen

Mais ist innerhalb der Süßgräser evolutionär nah mit der Mohrenhirse (*Sorghum bicolor* L.) verwandt (Abbildung 4A). Vor etwa 5-12 Millionen Jahren haben sich die beiden Arten getrennt und das Maisgenom hat sich dupliziert (Swigonová et al., 2004). Diese Genomduplikation hatte zur Folge, dass zunächst alle Maisgene im Vergleich zu den homologen Genen in der Mohrenhirse in doppelter Ausführung vorlagen (Schnable et al., 2011). Diese homologen Gene zwischen Mais und Mohrenhirse werden als syntänisch bezeichnet. Im Verlauf der Evolution hat das Maisgenom viele dieser homologen Genkopien wieder verloren. Gleichzeitig konnten jedoch durch erneute Duplikation einzelner Gene auch zusätzliche Maisgene entstehen (Woodhouse et al., 2010). Für diese neu entstandenen Gene gibt es keine Homologe im Genom der Mohrenhirse. Solche evolutionär jungen Gene werden als nicht-syntänisch bezeichnet und machen etwa 50% des heutigen Maisgenoms aus (zusammengefasst in Marcon et al., 2016).

Unser Interesse bestand nun darin den evolutionären Ursprung der Gene zu untersuchen, die in Hybriden spezielle Expressionsmuster aufweisen. Im Fokus standen vor allem SPE-Gene und Gene mit nicht-additivem und differentiellem Expressionsmuster, die in unseren Transkriptomstudien beschrieben wurden (Paschold et al., 2014; Baldauf et al., 2016; Marcon et al., 2017; Baldauf et al., 2018). Um einen Einblick in den evolutionären Ursprung der Gene dieser Expressionsklassen zu gewinnen, wurde zunächst für alle in der Primärwurzel exprimierten Gene ermittelt, ob sie syntänischen oder nicht-syntänischen Ursprungs sind. Bei diesem Vergleich zeigte sich, dass etwa ein Drittel aller exprimierten Gene nicht-syntänischen Ursprungs sind. Innerhalb der extremen Expressionsklassen (SPE, nicht-additiv, differentiell) lag der Anteil nicht-syntänischer Gene bei etwa zwei Drittel (Abbildung 4B).

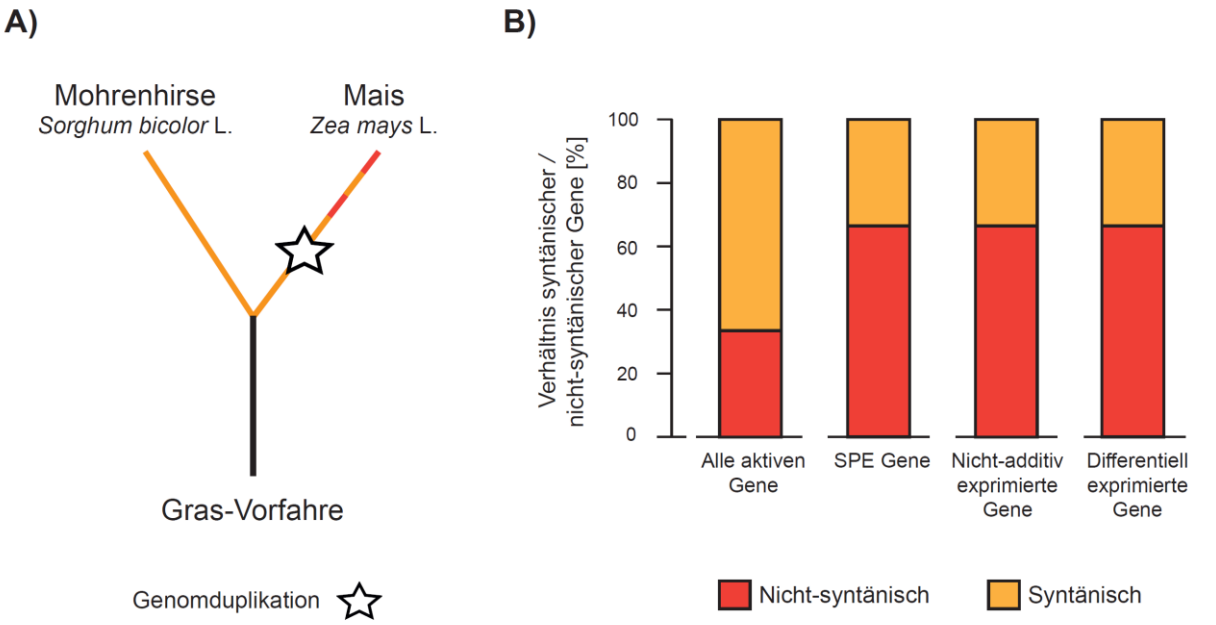


Abbildung 4. Evolutionäre Herkunft der Gene unterschiedlicher Expressionsklassen. **A)** Gene, die aus der letzten Duplikation des Maisgenoms hervorgegangen sind, sind mit den Genen der nah verwandten Mohrenhirse syntänisch (gelb). Im späteren Verlauf der Evolution entstanden durch Duplikation einzelner Gene weitere Maisgene, die nicht-syntänisch zu Genen der Mohrenhirse sind (rot). **B)** Im Vergleich zu allen aktiven Genen in der Primärwurzel von Mais sind nicht-syntänische Gene in den extremen Expressionsklassen deutlich angereichert (verändert nach Marcon et al., 2016).

Zusammenfassend zeigten unsere Heterosis-assoziierten Transkriptomanalysen übereinstimmend, dass spezielle Genexpressionsmuster in Hybriden deutlich von nicht-syntänischen, also evolutionär jungen Genen geprägt sind. Solche Gene wurden durch natürliche Selektion geformt und kodieren vermutlich für Proteine mit redundanter Funktion (Paschold et al., 2014; Schnable, 2015), die sich jedoch vorteilhaft auf das Wachstum auswirken können, z. B. bei nachteiligen Umwelteinflüssen. In diesem Zusammenhang wurden z. B. in nicht-syntänischen Regionen des Genoms von Ziegengras (*Aegilops tauschii*) einige Krankheitsresistenzgene identifiziert (Dong et al., 2016). Ebenso spricht die hohe Stabilität der nicht-snytänischen SPE-Gene nach einer Wasserdefizitbehandlung für einen positiven Einfluss solcher Gene auf die Anpassungsfähigkeit der Hybride (Marcon et al., 2017). Evolutionär ältere, syntänische Gene im Maisgenom haben dem Selektionsdruck und damit der Entfernung nachteiliger Allele über mehrere Millionen Jahre standgehalten. Sie kodieren für Proteine, die wichtige Funktionen ausüben (Birchler et al., 2005). Unter den syntänischen Genen sind auch die sogenannten klassischen Maisgene, deren Funktionen bereits beschrieben sind (Schnable und Freeling, 2011). Mutationen in diesen Genen können erhebliche Entwicklungsstörungen bei Mais verursachen. Erwartungsgemäß sind klassische Maisgene unter den SPE-Gene signifikant unterrepräsentiert, da diese vermehrt nicht-syntänischen Ursprungs sind (Marcon et al., 2017; Baldauf et al., 2018).

9.2.3 Proteom-Atlas der Primärwurzel von Mais

Als Proteom wird die Gesamtheit aller Proteine bezeichnet, die in einem untersuchten Gewebe zu einem bestimmten Entwicklungszeitpunkt akkumulieren. Studien der Proteomik befassen sich mit der Identifizierung der einzelnen Proteine aus einer Probe. Frühere Proteomik-Studien von Mais basierten auf der zweidimensionalen (2-D) Polyacrylamid Gelelektrophorese und nachfolgender massenspektrometrischer Analyse der Proteine (z. B. Hoecker et al., 2008; Marcon et al., 2010). Bei dieser Methode wurden lösliche Proteine nach Extraktion in der ersten Dimension nach ihrem isoelektrischen Punkt und in der zweiten Dimension entsprechend ihres Molekulargewichts aufgetrennt. Nach Färbung der Proteine durch Coomassieblau konnten in eigenen Vorarbeiten während meiner Doktorarbeit maximal 599 Proteine aus den 2-D-Proteingelen extrahiert werden (Marcon et al., 2010). Nach proteolytischer Spaltung dieser Proteine in Peptide erfolgte die massenspektrometrische Analyse.

Inzwischen werden komplexe Proteingemische mit oder ohne stabile Markierung direkt über hochauflösende Tandem-Massenspektrometrie (MS/MS) analysiert, d. h. ohne vorherige Auftrennung durch ein 2-D-Proteingel. Diese sogenannte *Shotgun*-Analyse ermöglichte es bereits im Rahmen meiner Doktorarbeit insgesamt 1918 unterschiedliche Proteine aus 2-4 cm langen Seminalwurzeln zu identifizieren (Marcon et al., 2013b). Bei der *Shotgun*-Analyse werden die Peptide zunächst chromatographisch aufgetrennt, mittels Elektrospray-Ionisierung (ESI) oder Matrix-unterstützter Laserdesorption/Ionisierung (MALDI) ionisiert und in die Gasphase überführt. Anschließend werden die Peptide im Massenspektrometer entsprechend ihrem Masse-Ladungs-Quotienten (m/z) aufgetrennt und an einem Detektor nachgewiesen (zusammengefasst in Zhu et al., 2010). Die aufgenommenen MS-Spektren sind für jedes Peptid spezifisch. Sie werden mit Proteindatenbanken abgeglichen und so den betreffenden Peptidsequenzen zugeordnet.

Proteomanalysen unterliegen im Vergleich zu Transkriptomanalysen mehreren technischen Herausforderungen: Zunächst ist es schwierig bis unmöglich alle Proteine einer Probe zu extrahieren, weil die Bandbreite der Proteine von stark hydrophil bis hochgradig hydrophob reicht. Außerdem können membranständige Proteine nur schwer in Lösung gebracht werden. Weiterhin können Proteine, im Gegensatz zu Nukleinsäuren, nicht amplifiziert werden. Insbesondere niedrig akkumulierte Proteine werden häufig nicht erfasst, weil die Messkapazität der Massenspektrometer begrenzt ist (Kuster et al., 2005). Trotz dieser technischen Einschränkungen wurde in den letzten Jahren die Reproduzierbarkeit der chromatographischen Peptidtrennung und die Sensitivität und Geschwindigkeit der Massenspektrometrie verbessert. Daher ist es mittlerweile möglich mehrere tausend Proteine

aus einem komplexen Proteingemisch zu detektieren und zu analysieren (z. B. Marcon et al., 2015).

In vorangegangenen Studien wurde das Maiswurzelproteom während der frühen Wurzelentwicklung (Hochholdinger et al., 2005) und in unterschiedlichen Genotypen untersucht (z. B. Marcon et al., 2013b). Dabei wurden jeweils ganze Maiswurzeln analysiert, was den Nachteil hat, dass die daraus isolierten Proteine nicht den einzelnen Wurzelgeweben zugeordnet werden konnten. Erste gewebespezifische Proteomik-Analysen wurden dadurch ermöglicht, dass sich die Maiswurzel durch einfache mechanische Handgriffe in die Wurzelgewebe Cortex und Stele separieren lässt (Saleem et al., 2010; Abbildung 1B).

Mit dem Ziel, einen gewebespezifischen Proteom-Atlas der Primärwurzel von B73 herzustellen, wurden lösliche Proteine aus der meristematischen Zone, der Elongationszone, dem Cortex und der Stele isoliert und mittels *Shotgun*-Proteomik analysiert (Marcon et al., 2015). Die Proteine wurden in vier biologischen Replikaten je Gewebe aus 2-4 cm langen Primärwurzeln extrahiert. Nach enzymatischer Spaltung der Proteine durch die Peptidase Trypsin wurden die Peptide direkt in einem Hochleistungsflüssig-Chromatograph (HPLC) aufgetrennt und mittels MS/MS analysiert. Ergänzend wurden phosphorylierte Peptide aus dem jeweiligen Peptidgemisch mittels Affinitätschromatographie angereichert, um einen Einblick in das Phosphoproteom und damit auf den Phosphorylierungsstatus der Proteine zu gewinnen. Die Phosphorylierung von Proteinen ist eine bedeutende posttranslationale Modifikation, die sich auf die Proteinfunktion auswirken kann. Die plasmamembranständige H^+-ATPase wird z. B. über Phosphorylierung reguliert (z. B. Ottmann et al., 2007; Speth et al., 2010). Als Protonenpumpe baut die H^+-ATPase unter Hydrolyse von ATP einen Protonengradienten über der Plasmamembran auf, der die treibende Kraft für den sekundären Transport von Nährstoffen in die Zelle darstellt.

Insgesamt wurden acht Millionen Massenspektren zur Quantifizierung verwendet. Damit konnten 11 552 nicht-modifizierte Proteine und 2852 phosphorylierte Proteine identifiziert werden (Marcon et al., 2015). Unter den 11 552 Proteinen waren 61% (7012 von 11 552) in allen vier Geweben der Primärwurzel akkumuliert. Weitere 28% (3208 von 11 552) der Proteine wurden in zwei oder drei Geweben identifiziert, die restlichen 11% (1332 von 11 552) waren gewebespezifisch. Dieses Ergebnis war vergleichbar mit einer Proteomstudie, in der 10% der in Maiskörnern akkumulierten Proteine eine gewebespezifische Akkumulierung zeigten (Walley et al., 2013). Im Gegensatz zu nicht-modifizierten Proteinen waren nur 30% der Phosphoproteine (844 von 2852) in allen Geweben aktiv. Ein größerer Anteil von 39% (1118 von 2852) war dagegen in zwei oder drei Geweben akkumuliert. Die übrigen 31% der phosphorylierten Proteine (890 von 2852) waren gewebespezifisch. Eine ähnliche Dynamik des Phosphoproteoms gegenüber dem nicht-modifizierten Proteom wurde bei einer vergleichbaren Analyse unterschiedlicher Blattgewebe festgestellt (Facette et al., 2013).

In Eukaryoten werden vornehmlich Serin- und Threoninreste phosphoryliert, die Aminosäure Tyrosin dagegen seltener (de la Fuente van Bentem und Hirt, 2009). Übereinstimmend damit stellte sich bei der Analyse der phosphorylierten Aminosäuren in den Proteinen heraus, dass zumeist Serinreste (84%), gefolgt von Threonin- und Tyrosinresten (15% und 1%) modifiziert waren (Marcon et al., 2015).

Eine funktionelle Charakterisierung der nicht-modifizierten und phosphorylierten Proteine, die ausschließlich in der meristematischen Zone akkumulierten, zeigte, dass dort vor allem die Kategorien DNA, RNA und Protein angereichert sind (Abbildung 5A). Insbesondere Proteine, die eine Rolle im Zellzyklus spielen, waren überrepräsentiert, was die zellteilungsaktive Rolle der meristematischen Zone unterstreicht. Die angereicherten Proteinklassen der meristematischen Zone ließen sich in den drei anderen Zonen graduell weniger nachweisen (Abbildung 5A). Vergleichbar dazu waren auch phosphorylierte Proteine der Kategorien DNA und RNA in der meristematischen Zone überrepräsentiert (Abbildung 5B). In den differenzierten Geweben der Stele und dem Cortex befanden sich vornehmlich nicht-modifizierte Proteine, die eine Rolle im Zellwand-, Fett-, Transport- und Sekundärstoffwechsel spielen (Abbildung 5A). Weiterhin waren in diesen beiden Geweben phosphorylierte Proteine des Kohlenhydratstoffwechsels, der Signaltransduktion und Transportproteine (Abbildung 5B), wie z. B. ABC-Transporter (ABC, engl. *ATP binding cassette*), angereichert. ABC-Transporter sind Transmembranproteine, die mittels ATP-Hydrolyse an ihrer ATP-bindenden Kassette Hormone und andere Moleküle transportieren können (Higgins, 1992).

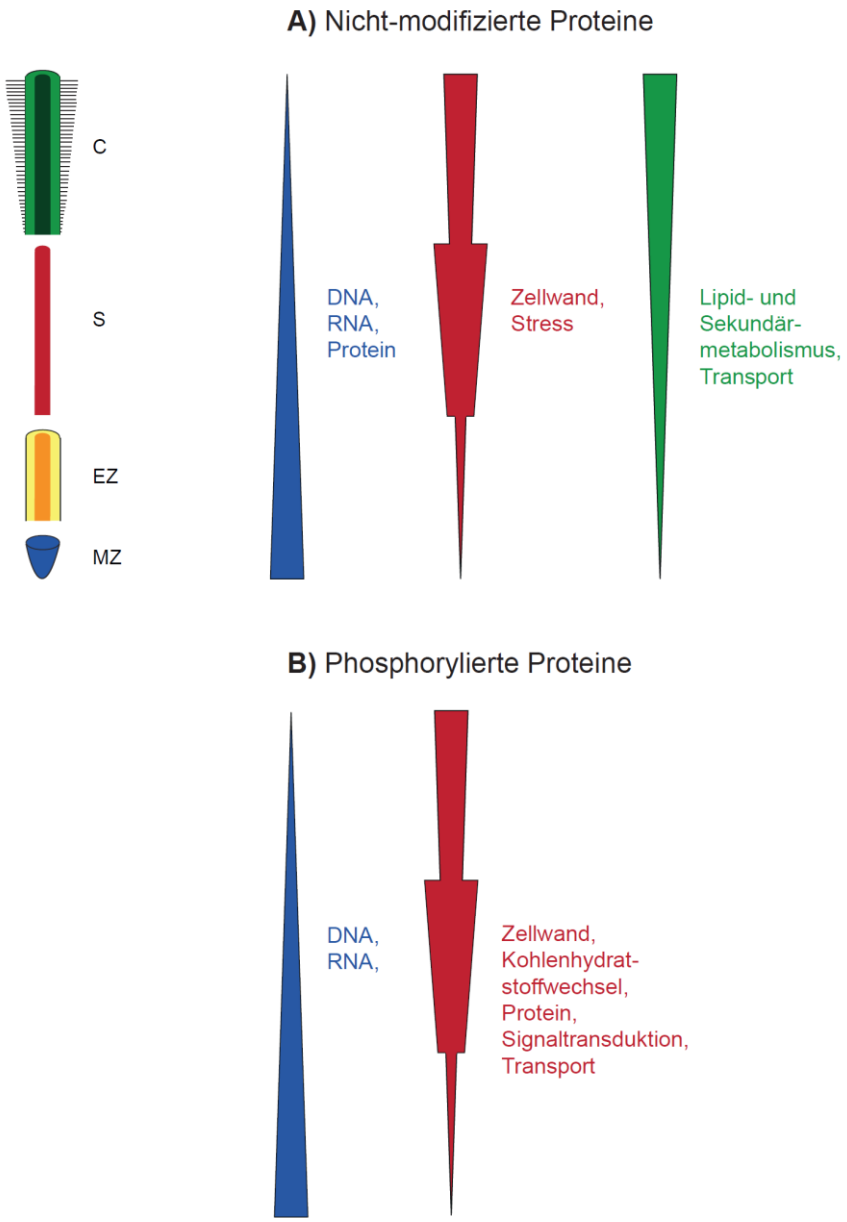


Abbildung 5. Gewebespezifische Akkumulierungstendenz funktioneller Proteinklassen (Marcon et al., 2015; C, Cortex; EZ, Elongationszone; MZ, meristematische Zone; S, Stele). **A)** Akkumulierungstendenz nicht-modifizierter Proteine. **B)** Akkumulierungstendenz phosphorylierter Proteine.

Neben der funktionellen Analyse der identifizierten Proteine interessierte uns auch die Korrelation von Transkriptlevel und Proteinakkumulation. Zur Berechnung der mRNA/Protein-Korrelationskoeffizienten in den einzelnen Geweben wurde auf einen unserer RNA-Seq-Datensätze zurückgegriffen (Paschold et al., 2014). Insgesamt wurden zwischen 5542 (Stele) und 7092 (meristematische Zone) annotierte Gene sowohl im Transkriptom- als auch im Proteomdatensatz identifiziert und standen für die Korrelationsanalyse zur Verfügung. Für jedes der vier Gewebe wurde ein moderater mRNA/Protein-Korrelationskoeffizient zwischen 0,38 und 0,56 berechnet. Obwohl in der Vergangenheit davon ausgegangen wurde, dass das Proteinlevel proportional zum Transkriptlevel ist (zusammengefasst in Vogel und Marcotte,

2012), weisen inzwischen viele Studien auf eine geringe bis mäßige mRNA/Protein-Korrelation hin (Schwanhäusser et al., 2011; Walley et al., 2013; Ponnala et al., 2014). Teilweise kann Letztere durch posttranskriptionale Regulationsmechanismen erklärt werden, die zu unterschiedlichen Translationsraten führen (Schwanhäusser et al., 2011). Zusätzlich können Variationen in der mRNA/Protein-Korrelation durch unterschiedliche Transkript- versus Proteinstabilität oder dem Transport von Proteinen zwischen einzelnen Geweben erklärt werden. Neben diesen biologischen Erklärungen gibt es auch technische Ursachen, die eine geringe Korrelation von Proteinen und korrespondierenden Transkripten begünstigen. In diesem Zusammenhang sind vor allem die unterschiedlichen Quantifizierungsmethoden zu nennen, die bei Transkriptom- und Proteomstudien eingesetzt werden.

Zusammenfassend repräsentiert unser hochauflösender Proteom-Atlas der Maiswurzel eine wertvolle Ressource, die funktionelle Gradienten in benachbarten Geweben sichtbar macht.

9.2.4 Insertionsmutagenese zur Erzeugung neuer Mutationen in Mais

Die genomweite Insertionsmutagenese ist ein Werkzeug, um Funktionsverlustmutanten für möglichst viele Gene eines Genoms zu generieren. Dabei werden Mutationen auf natürliche Weise durch zufällige Integration von mobilen DNA-Fragmenten innerhalb eines Genoms erzeugt (Feschotte et al., 2002). Die durch mobile DNA-Fragmente getroffenen Gene werden häufig in ihrer Funktion ausgeschaltet.

In Mais werden häufig Transposons als endogene Insertionsmutagene verwendet (Transposonmutagenese). Sie wurden erstmalig im Jahr 1951 als mobile DNA-Sequenzen entdeckt (McClintock, 1951). Transposons oder transponierende Elemente (TEs) können ihre Position im Genom ändern und somit auch Mutationen in genkodierenden Sequenzen verursachen. Diese Mutationen können dann erhebliche Auswirkungen auf den Phänotyp der Maispflanzen haben (siehe Kapitel 9.2.1).

Es gibt zwei Transposonklassen, die sich hauptsächlich in ihrer strukturellen Organisation und in ihrem Transpositionsmechanismus unterscheiden. Retrotransposons der Klasse 1 transponieren über den sogenannten *copy-and-paste*-Mechanismus, in dem ihre DNA zunächst in ein RNA-Zwischenprodukt transkribiert wird. Die RNA-Kopie des Elements wird dann mittels reverser Transkriptase in DNA umgewandelt und an einem neuen Ort im Genom integriert. Zu den Retrotransposons gehören z. B. LTR Transposons (engl. *long terminal repeats*), die terminale DNA-Wiederholungen besitzen (zusammengefasst in Feschotte et al., 2002). Im Gegensatz zu den Retrotransposons transponieren die Transposons der Klasse 2 ausschließlich auf Ebene der DNA. Dazu nutzen sie den *cut-and-paste*-Mechanismus, mit dessen Hilfe sie an einem Ort im Genom ausgeschnitten und an anderer Stelle wieder reintegriert werden (Wicker et al., 2007). Wenn die Transposition dieser TEs während der

Replikation der Chromosomen von einer bereits replizierten an eine noch nicht replizierte Stelle passiert, bleibt die Kopie des TEs am Ursprungsort erhalten (Greenblatt und Brink, 1962).

Beide Transposonklassen bestehen aus autonomen und nicht-autonomen Elementen. Autonome TEs kodieren u. a. für eine Transposase, die die selbstständige Transposition im Genom ermöglicht. Nicht-autonome Transposons benötigen für diesen Prozess ein autonomes Element. Während Transposons der Klasse 1 vornehmlich in nicht-kodierende DNA-Bereiche springen und wesentlich zur Genomgröße beitragen können (Bennetzen, 2000), transponieren TEs der Klasse 2 vor allem in genkodierende Regionen (Fernandes et al., 2004; Settles et al., 2004). Für eine genomweite Insertionsmutagenese, mit dem Ziel möglichst viele Gene durch Transposoninsertionen auszuschalten, sind daher DNA-Transposons der Klasse 2 am besten geeignet. Zu solchen Transposons gehört die Familie der *Mutator* (*Mu*)-Elemente, die sich durch umgekehrte DNA-Wiederholungen an den terminalen Enden (engl. *terminal inverted repeats*, TIRs) auszeichnen (Lisch, 2002). Das autonome Element *MuDR*, benannt nach dem Entdecker Donald S. Robertson (Robertson, 1978), kontrolliert seine eigene Transposition sowie die Reintegration von mindestens 16 Klassen nicht-autonomer *Mu*-Elemente (Lisch, 2015). Zum Zweck einer ungerichteten Transposonmutagenese besitzen *Mu*-Elemente einige Vorteile gegenüber anderen Familien der Klasse II Transposons, wie z. B. der *Ac*/*Ds*-Elemente (engl. *Activator/Dissociation*). *Mu*-Elemente zeichnen sich zum einen durch eine hohe Transpositionsrate und Kopienzahl aus. Zum anderen transponieren *Mu*-Transposons ungerichtet in Gene des gesamten Maisgenoms und korrelieren mit der Gendichte (Lisch, 2002; Schnable et al., 2009). *Ac*/*Ds*-Transposons weisen dagegen eine Präferenz für eine regionale Mutagenese auf, d. h. sie transponieren bevorzugt in benachbarte Genbereiche (Brutnell, 2002). Vor dem Hintergrund, dass *Mu*-Elemente konservierte TIR-Sequenzen an beiden Enden besitzen, können *Mu*-flankierende DNA-Sequenzen mit Hilfe von Adapter- und *Mu*TIR-vermittelter PCR amplifiziert und mittels HTS sequenziert werden. Dieses Verfahren kam in jüngerer Vergangenheit bei der Analyse sequenzindizierter Kollektionen von *Mu*-induzierten Maismutanten bereits zum Einsatz (McCarty et al., 2013; Liang et al., 2019; Marcon et al., 2020).

9.2.4.1 Etablierung der europäischen Maismutanten-Ressource *BonnMu*

Im Fokus meiner Arbeit stand in den vergangenen fünf Jahren die Etablierung der europäischen Maismutanten-Kollektion *BonnMu* (Marcon et al., 2020), mit dem Ziel neuartige Funktionsverlustmutanten mit Hilfe funktioneller Genanalysen zu beschreiben. Dabei kam die genomweite Transposonmutagenese zum Einsatz, um sequenzindizierte Mutationen in möglichst vielen Genen eines Genoms zu erzeugen.

Für die *BonnMu*-Pilotstudie wurden 1152 *Mu*-induzierte Maisfamilien im genetischen Hintergrund der Inzuchtlinie B73 hergestellt, um revers- und vorwärtsgenetische Analysen durchzuführen. Zunächst wurde eine *Mu*-aktive Linie (Robertson, 1983) in B73 gekreuzt und damit die Transposition endogener *Mu*-Elemente ausgelöst. Die Selbstung der heterozygoten F_1-Generation stellte die F_2-Population her, die für rezessive Transposon-induzierte Mutationen segregierte. Für die reversgenetische Untersuchung wurden diese F_2-Familien dann mit einer effizienten HTS-Methode namens *Mutant-Seq* (*Mu-Seq*; McCarty et al., 2013) sequenziert. Damit konnten dann zuerst neue Insertionsstellen/ Mutationen identifiziert werden, bevor die durch Insertionen bedingten Phänotypen analysiert wurden (reverse Genetik).

Zur Herstellung einer *Mu-Seq*-Bibliothek wurde im ersten Schritt Blattmaterial von 576 *Mu*-induzierten *BonnMu*-Familien entsprechend einem 24 x 24 Gittersystem gepoolt (Abbildung 6; (Marcon et al., 2020; Stöcker et al., 2021). Blattmaterial von mindestens drei Keimlingen je *BonnMu*-Familie wurde hierfür in einen der 24 Reihenpools und einen der 24 Spaltenpools geerntet. Konkret wurde eine Blattspitze je Keimling in einem definierten Reihenpool und die Blattspitze eines wechselständigen Blattes desselben Keimlings in einem definierten Spaltenpool gesammelt. Folglich repräsentierte jeder Knotenpunkt des Gittersystems eine *BonnMu*-Familie. Die Verwendung wechselständiger Blätter je Keimling, also Blattmaterial unterschiedlicher Zelllinien, ermöglichte in der späteren bioinformatischen Analyse vererbbare Keimbahninsertionen von somatischen Insertionen zu unterscheiden.

Nach Isolierung und Fragmentierung der DNA aus den 48 Blattproben (je 24 Reihen- und Spaltenpools) wurde ein universeller Adapter an die Fragmentenden ligiert und die *Mu*TIR-flankierende DNA in einer ersten PCR angereichert. In zwei weiteren Amplifikationsschritten wurden Sequenzierungs-Adapter und ein Barcode an die amplifizierten Fragmente angehängt (Abbildung 6). Der Barcode erfüllte den Zweck die einzelnen Pools zu markieren und ein Multiplexing der 48 Proben in einer einzigen Sequenzierungsbibliothek zu ermöglichen. Im Anschluss an die Sequenzierung der Bibliothek im Hochdurchsatzverfahren erfolgte die bioinformatische Analyse der *Mu-Seq-reads*. Diese beinhaltete u. a. das Alignment der *reads* auf das im Jahr 2020 aktuelle Referenzgenom von B73v4.0 (Version 4.0), das insgesamt 44 117 protein-kodierende Gene enthält. Die vorangegangene Beprobung unterschiedlicher somatischer Zelllinien je *BonnMu*-Familie ermöglichte dabei die Identifizierung von vererbbaren Keimbahninsertionen. Diese waren ausschließlich durch *reads* repräsentiert, die entsprechend dem 24 x 24 Gittersystem übereinstimmende Genomkoordinaten in einem Reihen- und einem Spaltenpool aufwiesen (Abbildung 6).

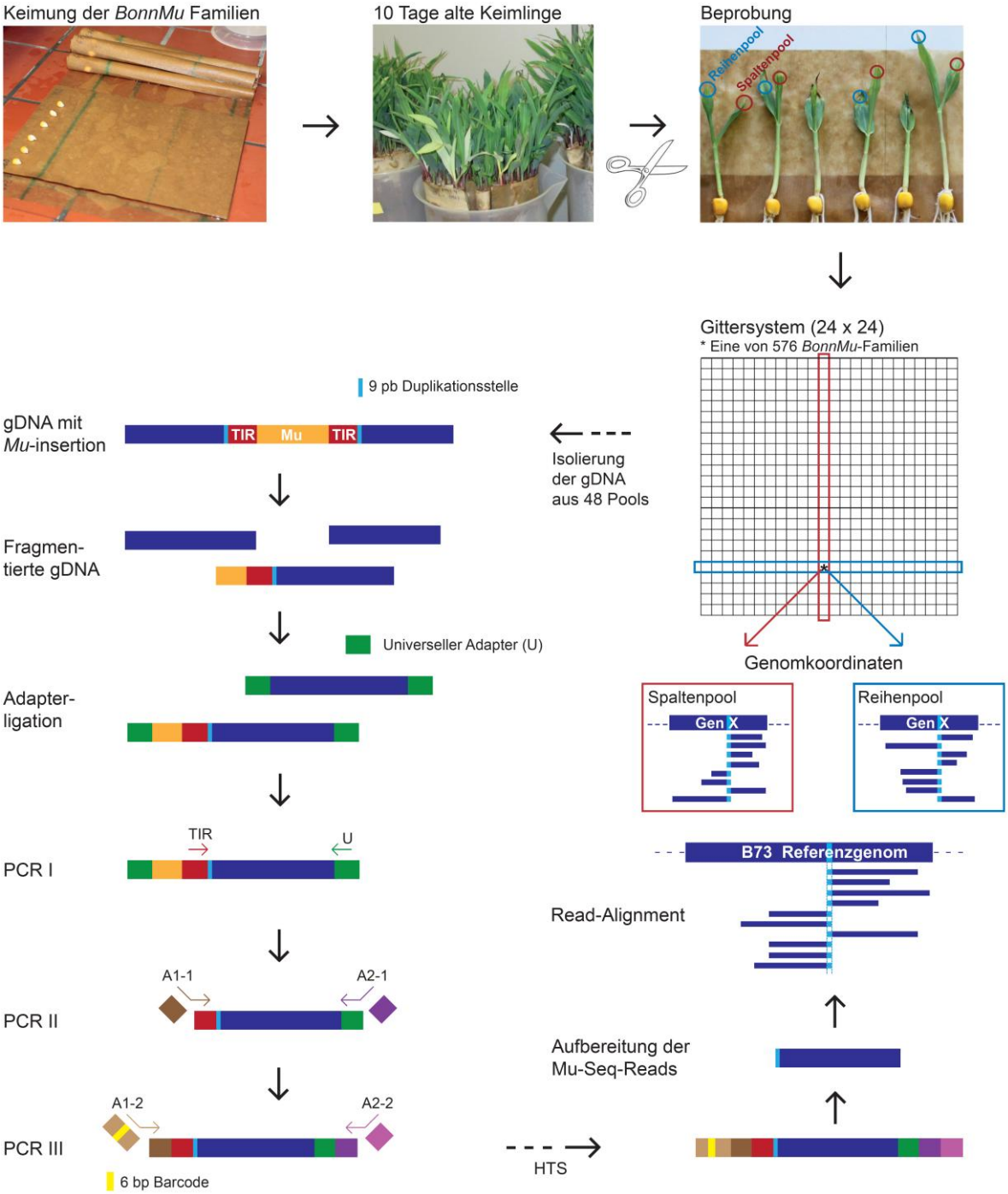


Abbildung 6. Herstellung einer sequenzindizierten *Mu-Seq*-Bibliothek. Nach der Keimung von 576 *BonnMu*-Familien (F_2-Population) wird Blattmaterial unterschiedlicher somatischer Zelllinien (hier: gegenüberliegende Blätter) geerntet. Die Blattbeprobung folgt einem 24 x 24 Gittersystem, wonach jede *BonnMu*-Familie in jeweils einem Reihenpool (blau) und Spaltenpool (rot) vertreten ist (Details siehe Haupttext). Anschließend wird die genomische DNA (gDNA) der 48 Proben isoliert und fragmentiert. Zur Anreicherung *Mu*-flankierender DNA-Fragmente und Inkorporation notwendiger Sequenzierungsadapter erfolgen drei nacheinander geschaltete Amplifikationsschritte (PCRI – PCRIII). Nach der Sequenzierung der *Mu*-flankierenden DNA-Fragmente im Hochdurchsatzverfahren erfolgt die bioinformatische Analyse der *Mu-Seq-reads*. Diese Analyse beinhaltet u. a. das Alignment der *reads* auf das Referenzgenom und die Zuordnung der *reads* in die jeweiligen *BonnMu*-Familien. Stimmen die Genomkoordinaten der 9 bp Duplikationsstelle mehrerer *reads* an der Schnittstelle von Reihen- und Spaltenpool überein, handelt es sich um eine Keimbahninsertion.

Insgesamt wurden 66 750 vererbbare *Mu*-Insertionen nachgewiesen, von denen 38% (25 664 von 66 750 Insertionen) auf nicht-kodierende DNA und 62% (41 086 von 66 750) auf kodierende Abschnitte des Genoms, also auf Gene, entfielen. Diese 41 086 Insertionen betrafen 16 392 (37%) der zu diesem Zeitpunkt annotierten Maisgene (Marcon et al., 2020; Stöcker et al., 2021). Weltweit existierten neben der initialen *BonnMu*-Ressource bereits die *UniformMu*- (McCarty et al., 2013) und die *ChinaMu*-Sammlung transposonmarkierter Gene (Liang et al., 2019). Während die *UniformMu*-Kollektion zu diesem Zeitpunkt 16 090 transposonmarkierte Gene im genetischen Hintergrund der Inzuchtlinie W22 umfasste, enthielt die *ChinaMu*-Bibliothek 20 693 solcher Gene in B73. Zusammengenommen konnten die drei Ressourcen in etwa 57% (25 140 von 44 117 Genen) aller Gene des Referenzgenoms *Mu*-Insertionen nachweisen. Ein beträchtlicher Anteil von 42% (10 623 von 25 140) der betroffenen Gene überschnitt sich dabei in allen drei Mutantenkollektionen. Da die Anzahl sich überschneidender Gene signifikant größer war als der zufällige Erwartungswert von 2724, implizierten die Daten, dass einige Gene bevorzugt von *Mu*-Insertionen getroffen werden. Fast 2000 Gene waren ausschließlich in der *BonnMu*-Ressource durch Insertion betroffen, nicht jedoch in den anderen Kollektionen. Damit stellt die europäische *BonnMu*-Datenbank eine wertvolle Ergänzung zu den beiden anderen Ressourcen *UniformMu* und *ChinaMu* dar.

In weiteren Untersuchungen prüften wir, ob und welche Genabschnitte vornehmlich von transponierten *Mu*-Elementen betroffen sind. Dazu wurde jeweils die genkodierende Sequenz aller transposonmarkierter Gene in 10 gleich große Abschnitte, sogenannte Dezile, eingeteilt. Auch die 5‘ und 3‘ untranslatierte Region (UTR) dieser Gene wurde berücksichtigt. Dann erfolgte die Sortierung der Transposoninsertionen in den jeweiligen Abschnitt. Das Ergebnis machte deutlich, dass *Mu*-Elemente bevorzugt in die 5‘ UTR und das erste Dezil der Gene springen (Abbildung 7A). Unter Berücksichtigung der jeweiligen Genlänge konnte dieses Phänomen noch präziser dargestellt werden: Lange Gene zeigten dabei einen eindeutigen Insertionsschwerpunkt im 5‘ UTR und im ersten Dezil, sowie einen kleineren Peak im 3‘ UTR (Abbildung 7B). Dagegen waren die *Mu*-Insertionen kürzerer Gene (Dezile 1% - 10% und 11% - 20%) gleichmäßig über die Gensequenz verteilt (Abbildung 7B). Dieses Phänomen kann aber zumindest teilweise dadurch erklärt werden, dass die absolute Entfernung von Insertionen zum 5‘ UTR selbst dann sehr kurz ist, wenn die Insertion am Genende liegt. Ähnlich verhielt es sich mit den *Mu*-Insertionen der *UniformMu*-Bibliothek (Marcon et al., 2020). Darüber hinaus zeigte die *BonnMu*-Studie, dass die Mehrzahl (75%) der *Mu*-Insertionsstellen näher am Transkriptionsstartpunkt, dem Beginn der 5‘ UTR, als am Startcodon der betroffenen Gene liegt. Diese Erkenntnis war insofern interessant, als dass aktive Gene im Bereich des Transkriptionsstartpunkts häufig durch offenes Chromatin gekennzeichnet sind (Springer et al., 2018). Als Chromatin wird der Komplex aus DNA und Protein bezeichnet, aus dem die Chromosomen bestehen. Aufgelockertes Chromatin könnte nicht nur leicht für die

Transkriptionsmaschinerie, sondern auch leicht für *Mu*-Transposons zugänglich sein. Diese These wird durch die in der *BonnMu*-Studie identifizierte hohe Anzahl von *Mu*-Insertionsstellen am Transkriptionsstart unterstützt, bedarf aber noch weiterer Untersuchungen.

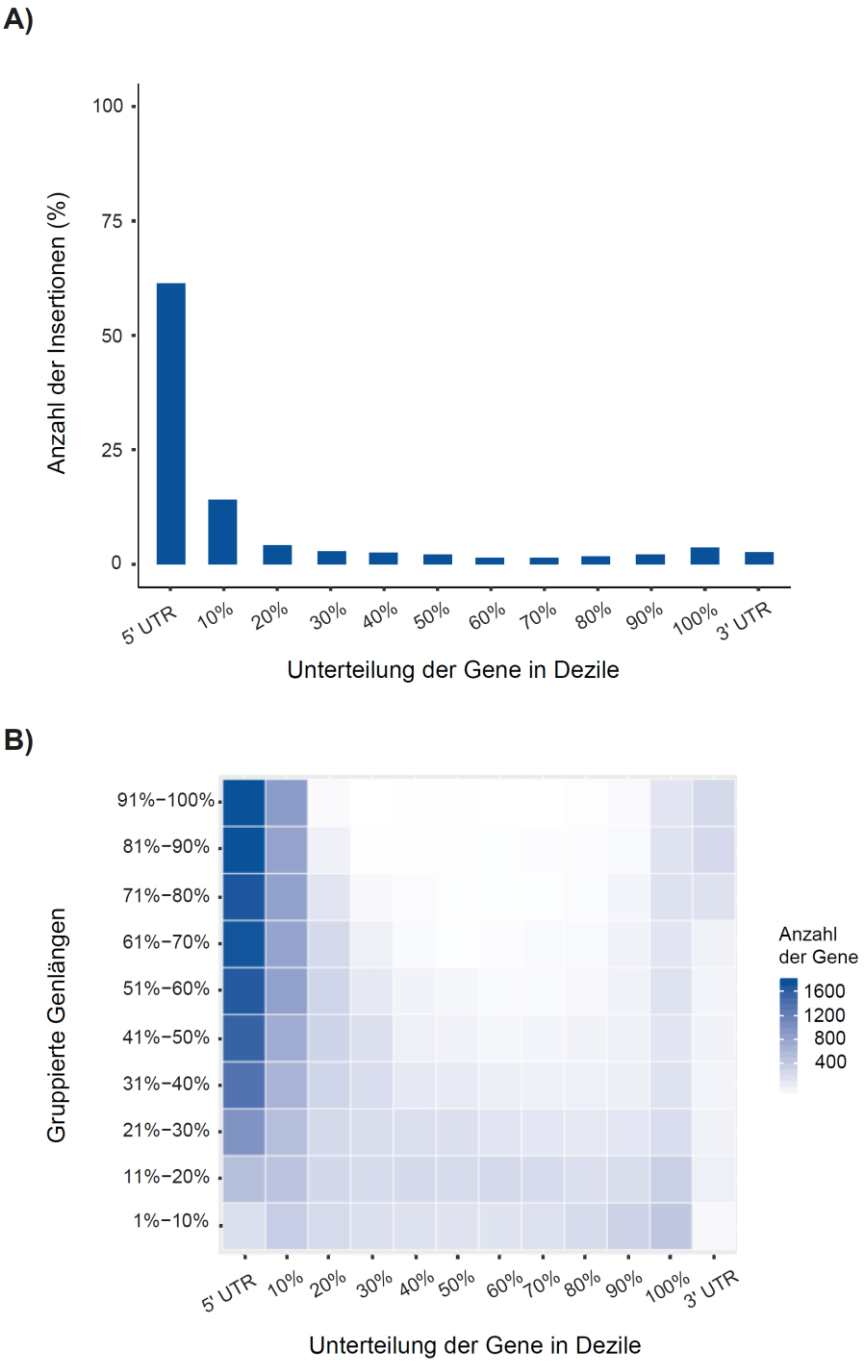


Abbildung 7. Verteilung der *Mu*-Insertionen in proteinkodierenden Genen. **A)** Prozentualer Anteil von *Mu*-Insertionen, die auf 5‘ UTR, 3‘ UTR und auf die in Zehntel-Abschnitte eingeteilte kodierende Sequenz aller transposonmarkierten Gene entfällt. **B)** Verteilung der Insertionen in 5‘ UTR, 3‘ UTR und in Dezile der kodierenden Gensequenz, nach Längeneinteilung aller transposonmarkierten Gene. Jede Gruppe enthält dieselbe Anzahl von Genen. Die Gruppe 1% - 10% repräsentiert die kürzesten Gene, die Gruppe 91% - 100% enthält die längsten Gene (verändert nach Marcon et al., 2020).

9.2.4.2 Automatisierte Identifizierung von Transposoninsertionen mit der *Mutant-Seq Workflow Utility* (*MuWU*)

Vor dem Hintergrund der expandierenden *BonnMu*-Ressource wurde mittlerweile ein Workflow entwickelt, der die komplexen bioinformatischen Schritte der *Mu-Seq-read*-Analyse automatisiert durchführt (Stöcker et al., 2021). Die sogenannte *Mutant-Seq Workflow Utility* (*MuWU*) kann dabei in einer einzigen Prozessierungs-Pipeline transposonmarkierte Gene identifizieren, zwischen somatischen und vererbbaren Insertionen unterscheiden und Insertionen direkt den entsprechenden *BonnMu*-Familien einer *Mu-Seq*-Bibliothek zuordnen. Das ermöglicht die schnelle bioinformatische Auswertung einer *Mu-Seq*-Bibliothek in weniger als einer Stunde. Die Parameter der *MuWU*-Pipeline sind dabei nicht auf die Auswertung von *Mu-Seq-reads* begrenzt, sondern können auch auf andere Spezies/TEs ausgeweitet werden (Stöcker et al., 2021). Zu diesem Zweck basiert der *MuWU*-Algorithmus auf einem Workflow-Management-System, namens Snakemake (Köster und Rahmann, 2012). Snakemake zielt darauf ab, komplexe Arbeitsabläufe mit Hilfe von einzelnen Modulen zu vereinfachen, die ähnlich einem Baukastensystem angepasst oder erweitert werden können.

Die ersten Schritte des *MuWU*-Algorithmus beinhalten eine Qualitätskontrolle und das Sequenzalignment von *Mu-Seq-reads* auf das Referenzgenom. Die Datenaufbereitung umfasst dabei z. B. die Entfernung von TIR- und Adaptersequenzen an beiden Enden der *reads* (Abbildung 6). Nachfolgend werden Insertionsstellen mit Hilfe der *Mu*-spezifischen 9 bp Duplikationsstelle, die beim Transpositionsprozess an jeder *Mu*-flankierenden Sequenz entsteht, detektiert. Anschließend erfolgt die Identifizierung der Keimbahninsertionen, die übereinstimmende Genomkoordinaten in jeweils nur einem Reihen- und einem Spaltenpool aufweisen dürfen. Die nachfolgende *MuWU*-Annotation zielt darauf ab, Insertionsstellen innerhalb von Genmodellen oder in den nahen 5‘ stromaufwärts-, bzw. 3‘ stromabwärts gelegenen Regionen von Genen in einer Ausgabe-Tabelle aufzulisten. Zu diesem Zweck werden alle Kombinationen aus Reihen- und Spaltenpool hinlänglich vererbbarer Insertionen gefiltert und den entsprechenden *BonnMu*-Familien der analysierten *Mu-Seq*-Bibliothek automatisch zugeordnet. Die finale Ausgabe-Tabelle, die als Nachschlagewerk für *BonnMu*-Insertionen dient, gibt Auskunft über die Genomkoordinaten der Keimbahninsertionen, die Anzahl der die Insertionen unterstützenden *reads* und die entsprechenden *BonnMu*-Familien, in der die Insertionen detektiert wurden. Darüber hinaus führt die Tabelle die Genidentifizierungsnummern der transposonmarkierten Gene und deren Länge in bp auf. Die Ausgabe-Tabelle ist Ausgangspunkt für reversgenetische Analysen.

9.2.4.3 Nutzung der *BonnMu*-Ressource für funktionelle Genanalysen

Die *BonnMu*-Insertionsstellen werden bei Verfügbarkeit neuer Mutanten regelmäßig in der *Maize Genetics and Genomics* Datenbank hinterlegt (*MaizeGDB*; Portwood et al., 2019) und so der Öffentlichkeit zugänglich gemacht.

Mittlerweile sind alle *BonnMu*-Insertionsstellen auf die entsprechenden Genomkoordinaten der Genomversion 5 von B73 (B73v5.0; Hufford et al., 2021) aktualisiert. Im Genombrowser von *MaizeGDB* (https://jbrowse.maizegdb.org) können die Insertionsstellen in den einzelnen Genen und Keimlings-Fotos der mutagenisierten Maisfamilien visualisiert werden (Abbildung 8; Marcon et al., 2020). Die fortlaufende Durchmusterung von jungen Keimlingen der segregierenden *BonnMu*-Familien ist dabei zusätzlich Bestandteil eines vorwärtsgenetischen Ansatzes zur Identifizierung von Mutanten. Die anschließende Identifizierung von Genen, die bestimmten Phänotypen unterliegen, wird durch die Kombination von BSR-Seq und der Verfügbarkeit der sequenzindizierten *Mu-Seq*-Bibliothek wesentlich vereinfacht. Die *Mu-Seq*-Bibliothek liefert eine begrenzte Anzahl an Kandidatengenen für jedes BSR-Seq-Intervall. Unter den analysierten Maisfamilien der *BonnMu*-Pilotstudie (Marcon et al., 2020) befanden sich neben Spross- und Blattmutanten (z. B. Albinos, Abbildung 8) auch Wurzelmutanten, die in derzeitig laufenden Projekten detailliert untersucht werden.

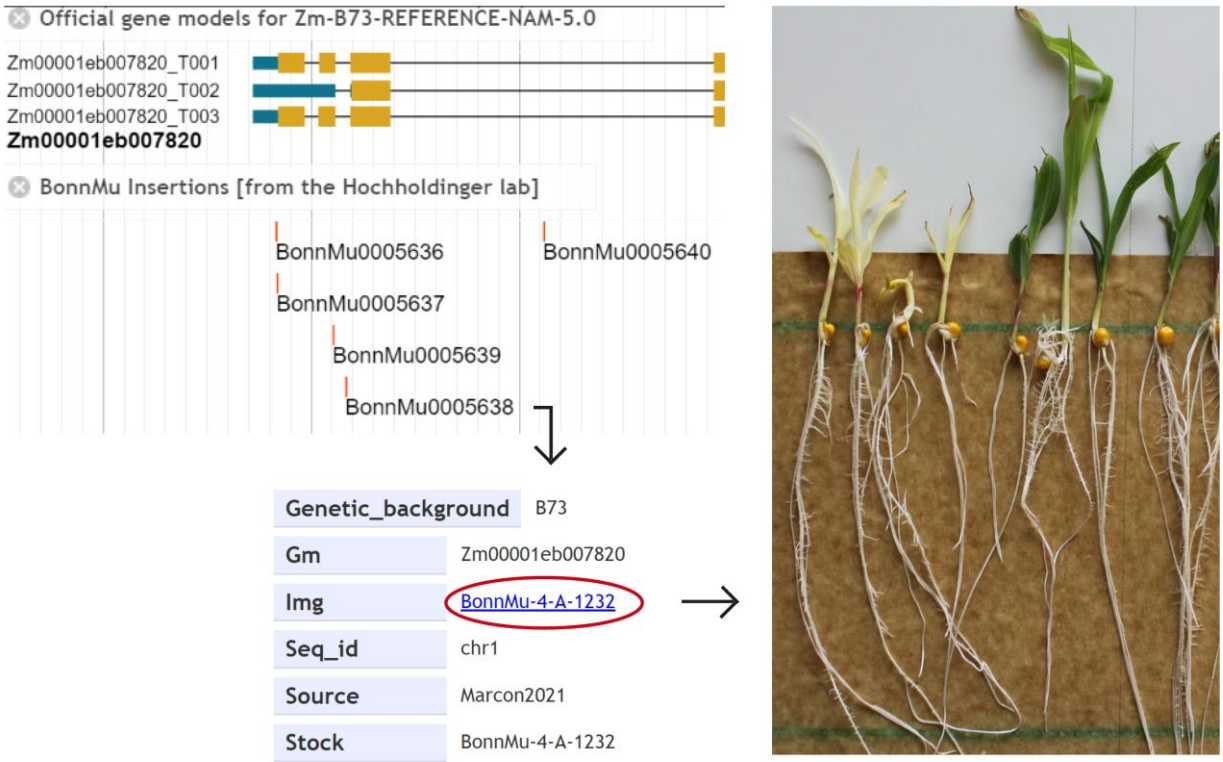


Abbildung 8. Insertionen der *BonnMu*-Bibliothek und Keimlings-Fotos der mutagenisierten Maisfamilien sind in der Datenbank *MaizeGDB* hinterlegt (https://jbrowse.maizegdb.org). Beispielhaft sind fünf *BonnMu*-Insertionen im Gen mit der Identifizierungsnummer Zm00001eb007820 gezeigt (links oben). Die Insertion mit der Endung 0005638 wurde in der Familie *BonnMu*-4-A-1232 identifiziert (links unten), die für einen Albino-Phänotyp segregiert (rechts).

Mutagenisierte F_2-Familien der *BonnMu*-Ressource werden bereits weltweit für reverse-genetische Analysen eingesetzt. Zu diesem Zweck wurden bislang über 1100 *BonnMu*-Familien an etwa 50 Forschungsgruppen verschickt.

9.2.4.4 Ausweitung der *BonnMu*-Ressource

Durch die fortlaufende Sequenzierung mutagenisierter Maisfamilien im *BonnMu*-Projekt konnten nach aktuellem Stand bereits über 250 000 Keimbahninsertionen identifiziert werden, die 27 896 (70%) der 39 756 Gene des Maisgenoms (B73v5) repräsentieren. Dafür wurden bislang 12 *Mu-Seq*-Bibliotheken, also 6912 *Mu*-induzierte F_2-Familien, sequenziert.

Neben Transposon-induzierten Maisfamilien in B73 wurden der *BonnMu*-Kollektion inzwischen auch Inzuchtlinien im genetischen Hintergrund von Colorado125 (Co125, Dent Pool) und die europäischen Linien F7, DK105 und EP1 hinzugefügt. Die drei Inzuchtlinien F7, DK105 und EP1 aus dem Flint Pool stellen wichtige Ressourcen für die europäische Maisforschung dar: Zum einen zeigen diese Linien in den gemäßigten Breiten Europas stabiles Wachstum, und zum anderen stehen die entsprechenden Genomsequenzen seit Kurzem zur Verfügung (Haberer et al., 2020). Zukünftig kann die Mutagenisierung der europäischen Inzuchtlinien dazu genutzt werden genotypspezifische Mutationen/ Insertionen zu identifizieren und daraus hervorgehende Mutanten zu beschreiben.

9.3 Zusammenfassung und Ausblick

Mais ist das ertragreichste Getreide weltweit. Die Entschlüsselung molekulargenetischer Grundlagen der Wurzelentwicklung liefert Antworten auf viele entwicklungsbiologische Fragen. Darüber hinaus ist die Grundlagenforschung am Beispiel des Maiswurzelsystems von großem agronomischen Interesse. Die vorliegende Arbeit konnte durch molekulargenetische und funktionelle Genomanalysen einen wesentlichen Beitrag zum besseren Verständnis der Wurzelentwicklung leisten. Zunächst konnten mit Hilfe einzelner Mutanten Einblicke in die molekularen Netzwerke, die an der Wurzelhaarbildung von Mais beteiligt sind, gewonnen werden. Zusätzlich lieferte die Charakterisierung der Aux/IAA Proteine RUM1 und RUL1 Erkenntnisse über konservierte und spezifische Funktionen beider Homeologe. Fortgeführt wurde die Analyse der Lateralwurzelmutante *rum1* mit Hilfe einer Transkriptomstudie, die in Kombination mit histologischen Untersuchungen neue Funktionen von RUM1 in der Lignifizierung und Entwicklung der Leitgewebe aufdecken konnte.

Weitere umfangreiche RNA-Seq-Analysen spielten eine Schlüsselrolle in der vorliegenden Arbeit und gaben Aufschluss über die Funktion des komplexen Maiswurzelsystems. In diesem Zusammenhang konnten spezifische Genexpressionsmuster in der frühen Entwicklungsphase von Primär-, Seminal- und Kronwurzeln eine funktionale Spezialisierung der Seminalwurzeln

aufzeigen. Weiterhin wurden zellspezifische Transkriptomprofile erstellt, um die Initiation von Lateralwurzeln in Perizykelzellen von Primär-, Seminal-, Kron- und Stützwurzeln in Abhängigkeit von Nitrat zu untersuchen. Ausschließlich Stützwurzeln entwickelten bei hoher Nitratkonzentration lange und stark verzweigte Lateralwurzeln. Dieser morphologische Unterschied zu den anderen Wurzeltypen spiegelte sich auch im Perizykeltranskriptom der Stützwurzeln wider, denn die Mehrzahl der nitrat-regulierten Gene konnte spezifisch den Perizykelzellen dieses Wurzeltyps zugeordnet werden. Aus agronomischer Sicht war es darüber hinaus interessant die Reaktion des Wurzel-Transkriptoms auf abiotischen Wasserdefizitstress zu untersuchen. Dabei wurde festgestellt, dass die Anzahl differentiell exprimierter Gene in der Primärwurzel mit Intensität und Dauer der Stressbehandlung zunahm. Die Mehrzahl der differentiell exprimierten Gene wurde durch die Behandlung induziert und war gewebespezifisch reguliert. Gene, die für die Detoxifizierung reaktiver Sauerstoffspezies wichtig sind, waren nach Stressbehandlung im Cortexgewebe und in der Elongationszone überrepräsentiert und könnten demnach zu einer Anpassung der Wurzel an Trockenstress beitragen.

Desweiteren lieferten Heterosis-assoziierte RNA-Seq-Studien neue Erkenntnisse über einzigartige Expressionsmuster in Maishybriden. Insbesondere das Phänomen der *single-parent expression* (SPE) konnte als konservierter Komplementationsmechanismus in Hybriden identifiziert werden. Das spezielle Expressionsmuster der SPE-Gene zeigte sich dabei nicht nur in verschiedenen Wurzelgeweben. Es war auch unter Wasserdefizitbedingungen und in genetisch diversen Maishybriden hoch konserviert. Zusammen mit anderen Genen, die nicht-additive und differentielle Expressionsmuster in Hybriden aufwiesen, konnten SPE-Gene den evolutionär jungen Genen des Maisgenoms zugeordnet werden. Proteine, die durch evolutionär junge Gene kodiert werden, könnten sich vorteilhaft auf das Wachstum und die Anpassungsfähigkeit der Hybriden auswirken und so den Heterosis-Effekts zumindest teilweise erklären.

Vergleichend zur RNA-Seq-Analyse einzelner Wurzelgewebe repräsentierte eine umfassende *Shotgun*-Proteomik-Studie einen Atlas nicht-modifizierter und phosphorylierter Proteine, der funktionelle Gradienten in benachbarten Geweben aufdeckte. So waren insbesondere Proteine, die eine Funktion im Zellzyklus ausüben, in der zellteilungsaktiven meristematischen Zone überrepräsentiert, konnten jedoch in den drei anderen Geweben graduell weniger nachgewiesen werden.

Um zukünftig neuartige Funktionsverlustmutanten mit Hilfe funktioneller Genanalysen zu charakterisieren, stand die Etablierung der Maismutanten-Ressource *BonnMu* im Mittelpunkt dieser Arbeit. Die Sequenzierung mutagenisierter Maisfamilien im genetischen Hintergrund der Inzuchtlinie B73 zeigte, dass *Mutator* Transposons bevorzugt in die Nähe der Startcodons von Genen springen, was zu Funktionsverlustmutationen führen kann. Inzwischen wurden

bereits mehrere hunderttausend Keimbahnmutationen identifiziert, die in reversgenetischen Analysen untersucht werden können. Außerdem wird in der Zukunft die Mutagenisierung europäischer Inzuchtlinien dazu beitragen genotypspezifische Mutationen zu identifizieren und daraus hervorgehende Mutanten, u. a. in Transkriptom- und Proteomstudien, zu beschreiben. Im Fokus zukünftiger Arbeiten wird darüber hinaus die funktionelle Charakterisierung neuartiger Maiswurzelmutanten stehen, die bereits bei der Phänotypisierung der mutagenisierten *BonnMu*-Populationen entdeckt wurden.

9.4 Literaturverzeichnis

Baldauf, J., **Marcon, C.**, Paschold, A., and Hochholdinger, F. (2016). Nonsyntenic genes drive tissue-specific dynamics of differential, nonadditive and allelic expression patterns in maize hybrids. Plant Physiol 171, 1144-1155.

Baldauf, J., **Marcon, C.**, Lithio, A., Vedder, L., Altrogge, L., Piepho, H., Schoof, H., Nettleton, D., and Hochholdinger, F. (2018). Single-parent expression is a general mechanism driving extensive complementation of non-syntenic genes in maize hybrids. Curr Biol 28, 431-437.

Bennetzen, J.L. (2000). Transposable element contributions to plant gene and genome evolution. Plant Mol Biol 42, 251-269.

Birchler, J.A., Riddle, N.C., Auger, D.L., and Veitia, R.A. (2005). Dosage balance in gene regulation: Biological implications. Trends in Genetics 21, 219-226.

Brown, W.L., and Anderson, E. (1947). The northern flint corn. Ann Mo Bot Gard 34, 1-28.

Brutnell, T.P. (2002). Transposon tagging in maize. Funct Integr Genomics 2, 4-12.

Buckler, E.S., Gaut, B.S., and McMullen, M.D. (2006). Molecular and functional diversity of maize. Curr Opin Plant Biol 9, 172-176.

Burton, A.L., Brown, K.M., and Lynch, J.P. (2013). Phenotypic diversity of root anatomical and architectural traits in *Zea* species. Crop Science 53, 1042-1055.

Candela, H., and Hake, S. (2008). The art and design of genetic screens: Maize. Nat Rev Genet 9, 192-203.

Cruz de Carvalho, M.H. (2008). Drought stress and reactive oxygen species: Production, scavenging and signaling. Plant Signal Behav 3, 156-165.

Dai, A. (2013). Increasing drought under global warming in observations and models. Nature Clim Change 3, 52-58.

de la Fuente van Bentem, S., and Hirt, H. (2009). Protein tyrosine phosphorylation in plants: More abundant than expected? Trends Plant Sci 14, 71-76.

Dong, L., Huo, N., Wang, Y., Deal, K., Wang, D., Hu, T., Dvorak, J., Anderson, O.D., Luo, M.C., and Gu, Y.Q. (2016). Rapid evolutionary dynamics in a 2.8-Mb chromosomal region containing multiple prolamin and resistance gene families in Aegilops tauschii. Plant J 87, 495-506.

East, E.M. (1908). Inbreeding in corn. Conn Agric Exp Stn Rep 1907, 419-428.

Emerson, R.A., Beadle, G.W., and Fraser, A.C. (1935). A summary of linkage studies in maize. Cornell Univ Agric Exp Sta Mem 180, 1-83.

Facette, M.R., Shen, Z., Björnsdóttir, F.R., Briggs, S.P., and Smith, L.G. (2013). Parallel proteomic and phosphoproteomic analyses of successive stages of maize leaf development. Plant Cell 25, 2798-2812.

Feldman, L. (1994). The maize root. In M. Freelings, V. Walbot eds. The maize handbook (New York: Springer), 29-37.

Fernandes, J., Dong, Q., Schneider, B., Morrow, D.J., Nan, G.L., Brendel, V., and Walbot, V. (2004). Genome-wide mutagenesis of *Zea mays* L. using *RescueMu* transposons. Genome Biol 5, R82.

Feschotte, C., Jiang, N., and Wessler, S.R. (2002). Plant transposable elements: Where genetics meets genomics. Nat Rev Genet 3, 329-341.

Grando, S., and Ceccarelli, S. (1995). Seminal root morphology and coleoptile length in wild (*Hordeum vulgare* ssp. *spontaneum*) and cultivated (*Hordeum vulgare* ssp. *vulgare*) barley. Euphytica 86, 73-80.

Greenblatt, I.M., and Brink, R.A. (1962). Twin mutations in medium variegated pericarp maize. Genetics 47, 489-501.

Haberer, G., Kamal, N., Bauer, E., Gundlach, H., Fischer, I., Seidel, M.A., Spannagl, M., **Marcon, C.**, Ruban, A., Urbany, C., Nemri, A., Hochholdinger, F., Ouzunova, M., Houben, A., Schon, C.C., and Mayer, K.F.X. (2020). European maize genomes highlight intraspecies variation in repeat and gene content. Nat Genet 52, 950-957.

Hardtke, C.S., and Berleth, T. (1998). The Arabidopsis gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development. EMBO J 17, 1405-1411.

Hetz, W., Hochholdinger, F., Schwall, M., and Feix, G. (1996). Isolation and characterization of *rtcs*, a maize mutant deficient in the formation of nodal roots. Plant J 10, 845-857.

Higgins, C.F. (1992). ABC transporters: From microorganisms to man. Annu Rev Cell Biol 8, 67-113.

Hochholdinger, F. (2009). The maize root system: Morphology, anatomy, and genetics. In: Handbook of maize: Its biology, J.L. Bennetzen and S.C. Hake, eds (Springer), 145-160.

Hochholdinger, F., and Hoecker, N. (2007). Towards the molecular basis of heterosis. Trends Plant Sci 12, 427-432.

Hochholdinger, F., Paschold, A., and **Marcon, C.** (2013). Heterosis: Maishybride exprimieren mehr Gene als ihre Eltern. BiuZ, 11-12.

Hochholdinger, F., Yu, P., and **Marcon, C.** (2018a). Genetic control of root system development in maize. Trends Plant Sci 23, 79-88.

Hochholdinger, F., Woll, K., Guo, L., and Schnable, P.S. (2005). The accumulation of abundant soluble proteins changes early in the development of the primary roots of maize (*Zea mays* L.). Proteomics 5, 4885-4893.

Hochholdinger, F., **Marcon, C.**, Baldauf, J.A., Frey, F., and Yu, P. (2018b). Proteomics of maize root development. Front Plant Sci 9, 143.

Hochholdinger, F., Wen, T.J., Zimmermann, R., Chimot-Marolle, P., Silva, O.D.E., Bruce, W., Lamkey, K.R., Wienand, U., and Schnable, P.S. (2008). The maize (*Zea mays* L.) *roothairless3* gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield. Plant J 54, 888-898.

Hoecker, N., Keller, B., Muthreich, N., Chollet, D., Descombes, P., Piepho, H.P., and Hochholdinger, F. (2008). Comparison of maize (*Zea mays* L.) F_1-hybrid and parental inbred line primary root transcriptomes suggests organ-specific patterns of nonadditive gene expression and conserved expression trends. Genetics 179, 1275-1283.

Hufford, M.B., Seetharam, A.S., Woodhouse, M.R., Chougule, K.M., Ou, S., Liu, J., Ricci, W.A., Guo, T., Olson, A., Qiu, Y., Della Coletta, R., Tittes, S., Hudson, A.I., Marand, A.P., Wei, S., Lu, Z., Wang, B., Tello-Ruiz, M.K., Piri, R.D., Wang, N., Kim, D.W., Zeng, Y., O'Connor, C.H., Li, X., Gilbert, A.M., Baggs, E., Krasileva, K.V., Portwood, J.L., 2nd,

Cannon, E.K.S., Andorf, C.M., Manchanda, N., Snodgrass, S.J., Hufnagel, D.E., Jiang, Q., Pedersen, S., Syring, M.L., Kudrna, D.A., Llaca, V., Fengler, K., Schmitz, R.J., Ross-Ibarra, J., Yu, J., Gent, J.I., Hirsch, C.N., Ware, D., and Dawe, R.K. (2021). De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. Science 373, 655-662.

International Rice Genome Sequencing Project. (2005). The map-based sequence of the rice genome. Nature 436, 793-800.

Ishikawa, H., and Evans, M.I. (1995). Specialized zones of development in roots. Plant Physiol 109, 725-727.

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M.C., Wang, B., Campbell, M.S., Stein, J.C., Wei, X., Chin, C.S., Guill, K., Regulski, M., Kumari, S., Olson, A., Gent, J., Schneider, K.L., Wolfgruber, T.K., May, M.R., Springer, N.M., Antoniou, E., McCombie, W.R., Presting, G.G., McMullen, M., Ross-Ibarra, J., Dawe, R.K., Hastie, A., Rank, D.R., and Ware, D. (2017). Improved maize reference genome with single-molecule technologies. Nature 546, 524-527.

Jones, D.F. (1917). Dominance of linked factors as a means of accounting for heterosis. Genetics 2, 466-479.

Köster, J., and Rahmann, S. (2012). Snakemake - a scalable bioinformatics workflow engine. Bioinformatics 28, 2520-2522.

Kuster, B., Schirle, M., Mallick, P., and Aebersold, R. (2005). Scoring proteomes with proteotypic peptide probes. Nat Rev Mol Cell Biol 6, 577-583.

Lawit, S.J., Wych, H.M., Xu, D., Kundu, S., and Tomes, D.T. (2010). Maize DELLA proteins dwarf plant8 and dwarf plant9 as modulators of plant development. Plant Cell Physiol 51, 1854-1868.

Li, L., Hey, S., Liu, S., Liu, Q., McNinch, C., Hu, H.C., Wen, T.J., **Marcon, C.**, Paschold, A., Bruce, W., Schnable, P.S., and Hochholdinger, F. (2016). Characterization of maize *roothairless6* which encodes a D-type cellulose synthase and controls the switch from bulge formation to tip growth. Sci Rep 6, 34395.

Liang, L., Zhou, L., Tang, Y., Li, N., Song, T., Shao, W., Zhang, Z., Cai, P., Feng, F., Ma, Y., Yao, D., Feng, Y., Ma, Z., Zhao, H., and Song, R. (2019). A sequence-indexed mutator insertional library for maize functional genomics study. Plant Physiol 181, 1404-1414.

Lisch, D. (2002). Mutator transposons. Trends Plant Sci 7, 498-504.

Lisch, D. (2015). Mutator and MULE Transposons. Microbiol Spectr 3, MDNA3-0032-2014.

Liu, S., Yeh, C.T., Tang, H.M., Nettleton, D., and Schnable, P.S. (2012). Gene mapping via bulked segregant RNA-Seq (BSR-Seq). PLoS One 7, e36406.

Lorenz, A., and Hoegemeyer, T. (2013). The phylogenetic relationships of US maize germplasm. Nat Genet 45, 844-845.

Ludwig, Y., and Hochholdinger, F. (2014). Laser microdissection of plant cells. Methods Mol Biol 1080, 249-258.

Marcon, C., Paschold, A., and Hochholdinger, F. (2013a). Genetic control of root organogenesis in cereals. In: De Smet I. (eds) Plant Organogenesis. Methods Mol Biol 959, 69-81.

Marcon, C., Baldauf, J.A., and Hochholdinger, F. (2016). Komplementation von Genexpressionsmustern in Maishybriden. Biospektrum 22, 603-605.

Marcon, C., Schützenmeister, A., Schütz, W., Madlung, J., Piepho, H.-P., and Hochholdinger, F. (2010). Non-additive protein accumulation patterns in maize *(Zea mays* L.) hybrids during embryo development. J Proteome Res 9, 6511-6522.

Marcon, C., Lamkemeyer, T., Malik, W.A., Ungrue, D., Piepho, H.P., and Hochholdinger, F. (2013b). Heterosis-associated proteome analyses of maize (*Zea mays* L.) seminal roots by quantitative label-free LC-MS. J Proteomics 93, 295-302.

Marcon, C., Malik, W.A., Walley, J.W., Shen, Z., Paschold, A., Smith, L.G., Piepho, H.P., Briggs, S.P., and Hochholdinger, F. (2015). A high-resolution tissue-specific proteome and phosphoproteome atlas of maize primary roots reveals functional gradients along the root axes. Plant Physiol 168, 233-246.

Marcon, C., Paschold, A., Malik, W.A., Lithio, A., Baldauf, J.A., Altrogge, L., Opitz, N., Lanz, C., Schoof, H., Nettleton, D., Piepho, H.P., and Hochholdinger, F. (2017). Stability of single-parent gene expression complementation in maize hybrids upon water deficit stress. Plant Physiol 173, 1247-1257.

Marcon, C., Altrogge, L., Win, Y., Stöcker, T., Gardiner, J., Portwood, J., 2nd, Opitz, N., Kortz, A., Baldauf, J.A., Hunter, C., McCarty, D., Koch, K., Schoof, H., and Hochholdinger, F. (2020). *BonnMu*: A sequence-indexed resource of transposon-induced maize mutations for functional genomics studies. Plant Physiol 184, 620-631.

Marzec, M., Melzer, M., and Szarejko, I. (2015). Root hair development in the grasses: What we already know and what we still need to know. Plant Physiol 168, 407-414.

Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez, G.J., Buckler, E., and Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci U S A 99, 6080-6084.

McCarty, D.R., Latshaw, S., Wu, S., Suzuki, M., Hunter, C.T., Avigne, W.T., and Koch, K.E. (2013). Mu-Seq: Sequence-based mapping and identification of transposon induced mutations. PLoS One 8, e77172.

McClintock, B. (1951). Mutable loci in maize. In: Annual report of the director of the department of genetics, Carnegie institution of Washington year book No. 50, 1950-1951. Cold Spring Harbor, New York, 174-181.

Meyer, S., Pospisil, H., and Scholten, S. (2007). Heterosis associated gene expression in maize embryos 6 days after fertilization exhibits additive, dominant and overdominant pattern. Plant Mol Biol 63, 381-391.

Michelmore, R.W., Paran, I., and Kesseli, R.V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci U S A 88, 9828-9832.

Mittler, R., Vanderauwera, S., Suzuki, N., Miller, G., Tognetti, V.B., Vandepoele, K., Gollery, M., Shulaev, V., and Van Breusegem, F. (2011). ROS signaling: The new wave? Trends Plant Sci 16, 300-309.

Nestler, J., Liu, S., Wen, T.J., Paschold, A., **Marcon, C.**, Tang, H.M., Li, D., Li, L., Meeley, R.B., Sakai, H., Bruce, W., Schnable, P.S., and Hochholdinger, F. (2014). *Roothairless5*, which functions in maize (*Zea mays* L.) root hair initiation and elongation encodes a monocot-specific NADPH oxidase. Plant J 79, 729-740.

O´Geen, A. (2012). Soil water dynamics. Nature Education Knowledge 3, 12.

Opitz, N., Paschold, A., **Marcon, C.**, Malik, W.A., Lanz, C., Piepho, H.P., and Hochholdinger, F. (2014). Transcriptomic complexity in young maize primary roots in response to low water potentials. BMC Genomics 15, 741.

Opitz, N., **Marcon, C.**, Paschold, A., Malik, W.A., Lithio, A., Brandt, R., Piepho, H.P., Nettleton, D., and Hochholdinger, F. (2016). Extensive tissue-specific transcriptomic plasticity in maize primary roots upon water deficit. J Exp Bot 67, 1095-1107.

Ottmann, C., Marco, S., Jaspert, N., **Marcon, C.**, Schauer, N., Weyand, M., Vandermeeren, C., Duby, G., Boutry, M., Wittinghofer, A., Rigaud, J.L., and Oecking, C. (2007).

Structure of a 14-3-3 coordinated hexamer of the plant plasma membrane H^+ -ATPase by combining X-ray crystallography and electron cryomicroscopy. Mol Cell 25, 427-440.

Paschold, A., **Marcon, C.**, Hoecker, N., and Hochholdinger, F. (2010). Molecular dissection of heterosis manifestation during early maize root development. Theor Appl Genet 120, 383-388.

Paschold, A., Larson, N.B., **Marcon, C.**, Schnable, J.C., Yeh, C.T., Lanz, C., Nettleton, D., Piepho, H.P., Schnable, P.S., and Hochholdinger, F. (2014). Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. Plant Cell 26, 3939-3948.

Paschold, A., Jia, Y., **Marcon, C.**, Lund, S., Larson, N.B., Yeh, C.T., Ossowski, S., Lanz, C., Nettleton, D., Schnable, P.S., and Hochholdinger, F. (2012). Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents. Genome Res 22, 2445-2454.

Piperno, D.R., Ranere, A.J., Holst, I., Iriarte, J., and Dickau, R. (2009). Starch grain and phytolith evidence for early ninth millennium B.P. maize from the central balsas river valley, Mexico. Proc Natl Acad Sci U S A 106, 5019-5024.

Ponnala, L., Wang, Y., Sun, Q., and van Wijk, K.J. (2014). Correlation of mRNA and protein abundance in the developing maize leaf. Plant J 78, 424-440.

Qayyum, A., Ahmad, S., Liaqat, S., Malik, W., Noor, E., Saeed, H.M., and Hanif, M. (2012). Screening for drought tolerance in maize (*Zea mays* L.) hybrids at an early seedling stage. Afr J Agric Res 7, 3594-3604.

Rhoades, M.M. (1984). The early years of maize genetics. Annu Rev Genet 18, 1-29.

Ribaut, J.M., Betran, J., Monneveux, P., and Setter, T. (2009). Drought tolerance in maize. In Handbook of Maize: Its biology, J.L. Bennetzen and S.C. Hake, eds (Springer), 311-344.

Robertson, B., Waines, J., and Gill, B. (1979). Genetic variability for seedling root numbers in wild and domesticated wheats. Crop Science 19, 843-847.

Robertson, D.S. (1978). Characterization of a *Mutator* system in maize. Mutat Res 51, 21-28.

Robertson, D.S. (1983). A possible dose-dependent inactivation of *Mutator* (*Mu*) in maize. Mol Gen Genet 191, 86-90.

Saleem, M., Lamkemeyer, T., Schützenmeister, A., Madlung, J., Sakai, H., Piepho, H.P., Nordheim, A., and Hochholdinger, F. (2010). Specification of cortical parenchyma and stele of maize primary roots by asymmetric levels of auxin, cytokinin, and cytokinin-regulated proteins. Plant Physiol 152, 4-18.

Schnable, J.C., and Freeling, M. (2011). Genes identified by visible mutant phenotypes show increased bias towards one of two subgenomes of maize. PLOS One 6, e17855.

Schnable, J.C. (2015). Genome evolution in maize: From genomes back to genes. Annu Rev Plant Biol 66, 329-343.

Schnable, J.C., Springer, N.M., and Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci U S A 108, 4069-4074.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F.S., Pasternak, S., Liang, C.Z., Zhang, J.W., Fulton, L., Graves, T.A., Minx, P., Reily, A.D., Courtney, L., Kruchowski, S.S., Tomlinson, C., Strong, C., Delehaunty, K., Fronick, C., Courtney, B., Rock, S.M., Belter, E., Du, F.Y., Kim, K., Abbott, R.M., Cotton, M., Levy, A., Marchetto, P., Ochoa, K., Jackson, S.M., Gillam, B., Chen, W.Z., Yan, L., Higginbotham, J., Cardenas, M., Waligorski, J., Applebaum, E., Phelps, L., Falcone, J., Kanchi, K., Thane, T., Scimone, A., Thane, N., Henke, J., Wang, T., Ruppert, J., Shah, N., Rotter, K., Hodges, J., Ingenthron, E., Cordes, M., Kohlberg, S., Sgro, J., Delgado, B., Mead, K., Chinwalla,

A., Leonard, S., Crouse, K., Collura, K., Kudrna, D., Currie, J., He, R.F., Angelova, A., Rajasekar, S., Mueller, T., Lomeli, R., Scara, G., Ko, A., Delaney, K., Wissotski, M., Lopez, G., Campos, D., Braidotti, M., Ashley, E., Golser, W., Kim, H., Lee, S., Lin, J.K., Dujmic, Z., Kim, W., Talag, J., Zuccolo, A., Fan, C., Sebastian, A., Kramer, M., Spiegel, L., Nascimento, L., Zutavern, T., Miller, B., Ambroise, C., Muller, S., Spooner, W., Narechania, A., Ren, L.Y., Wei, S., Kumari, S., Faga, B., Levy, M.J., McMahan, L., Van Buren, P., Vaughn, M.W., Ying, K., Yeh, C.T., Emrich, S.J., Jia, Y., Kalyanaraman, A., Hsia, A.P., Barbazuk, W.B., Baucom, R.S., Brutnell, T.P., Carpita, N.C., Chaparro, C., Chia, J.M., Deragon, J.M., Estill, J.C., Fu, Y., Jeddeloh, J.A., Han, Y.J., Lee, H., Li, P.H., Lisch, D.R., Liu, S.Z., Liu, Z.J., Nagel, D.H., McCann, M.C., SanMiguel, P., Myers, A.M., Nettleton, D., Nguyen, J., Penning, B.W., Ponnala, L., Schneider, K.L., Schwartz, D.C., Sharma, A., Soderlund, C., Springer, N.M., Sun, Q., Wang, H., Waterman, M., Westerman, R., Wolfgruber, T.K., Yang, L.X., Yu, Y., Zhang, L.F., Zhou, S.G., Zhu, Q., Bennetzen, J.L., Dawe, R.K., Jiang, J.M., Jiang, N., Presting, G.G., Wessler, S.R., Aluru, S., Martienssen, R.A., Clifton, S.W., McCombie, W.R., Wing, R.A., and Wilson, R.K. (2009). The B73 maize genome: complexity, diversity, and dynamics. Science 326, 1112-1115.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature 473, 337-342.

Settles, A.M., Latshaw, S., and McCarty, D.R. (2004). Molecular analysis of high-copy insertion sites in maize. Nucleic Acids Res 32, e54.

Shull, G.H. (1908). The composition of a field of maize. J Hered 4, 296-301.

Speth, C., Jaspert, N., **Marcon, C.**, and Oecking, C. (2010). Regulation of the plant plasma membrane H^+-ATPase by its C-terminal domain: What do we know for sure? Eur J Cell Biol 89, 145-151.

Springer, N.M., Ying, K., Fu, Y., Ji, T.M., Yeh, C.T., Jia, Y., Wu, W., Richmond, T., Kitzman, J., Rosenbaum, H., Iniguez, A.L., Barbazuk, W.B., Jeddeloh, J.A., Nettleton, D., and Schnable, P.S. (2009). Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLOS Genet 5, e1000734.

Springer, N.M., Anderson, S.N., Andorf, C.M., Ahern, K.R., Bai, F., Barad, O., Barbazuk, W.B., Bass, H.W., Baruch, K., Ben-Zvi, G., Buckler, E.S., Bukowski, R., Campbell, M.S., Cannon, E.K.S., Chomet, P., Dawe, R.K., Davenport, R., Dooner, H.K., Du, L.H., Du, C., Easterling, K.A., Gault, C., Guan, J.C., Hunter, C.T., Jander, G., Jiao, Y., Koch, K.E., Kol, G., Kollner, T.G., Kudo, T., Li, Q., Lu, F., Mayfield-Jones, D., Mei, W., McCarty, D.R., Noshay, J.M., Portwood, J.L., 2nd, Ronen, G., Settles, A.M., Shem-Tov, D., Shi, J., Soifer, I., Stein, J.C., Stitzer, M.C., Suzuki, M., Vera, D.L., Vollbrecht, E., Vrebalov, J.T., Ware, D., Wei, S., Wimalanathan, K., Woodhouse, M.R., Xiong, W., and Brutnell, T.P. (2018). The maize W22 genome provides a foundation for functional genomics and transposon biology. Nat Genet 50, 1282-1288.

Stadler, J.L., and Sprague, G.F. (1936). Genetic effects of ultra-violet radiation in maize. I. Unfiltered radiation. Proc Natl Acad Sci U S A 22, 572-578.

Stöcker, T., Altrogge, L., **Marcon, C.**, Win, Y.N., Hochholdinger, F., and Schoof, H. (2021). *MuWU*: Mutant-Seq library analysis and annotation. Bioinformatics btab679.

Swigonová, Z., Lai, J.S., Ma, J.X., Ramakrishna, W., Llaca, V., Bennetzen, J.L., and Messing, J. (2004). Close split of sorghum and maize genome progenitors. Genome Res 14, 1916-1923.

Tai, H., Lu, X., Opitz, N., **Marcon, C.**, Paschold, A., Lithio, A., Nettleton, D., and Hochholdinger, F. (2016). Transcriptomic and anatomical complexity of primary, seminal, and crown roots highlight root type-specific functional diversity in maize (*Zea mays* L.). J Exp Bot 67, 1123-1135.

Unterseer, S., Pophaly, S.D., Peis, R., Westermeier, P., Mayer, M., Seidel, M.A., Haberer, G., Mayer, K.F., Ordas, B., Pausch, H., Tellier, A., Bauer, E., and Schon, C.C. (2016). A comprehensive study of the genomic differentiation between temperate Dent and Flint maize. Genome Biol 17, 137.

Vogel, C., and Marcotte, E.M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet 13, 227-232.

von Behrens, I., Komatsu, M., Zhang, Y.X., Berendzen, K.W., Niu, X.M., Sakai, H., Taramino, G., and Hochholdinger, F. (2011). *Rootless with undetectable meristem 1* encodes a monocot-specific AUX/IAA protein that controls embryonic seminal and post-embryonic lateral root initiation in maize. Plant J 66, 341-353.

Walbot, V. (2000). Saturation mutagenesis using maize transposons. Curr Opin Plant Biol 3, 103-107.

Walley, J.W., Shen, Z., Sartor, R., Wu, K.J., Osborn, J., Smith, L.G., and Briggs, S.P. (2013). Reconstruction of protein networks from an atlas of maize seed proteotypes. Proc Natl Acad Sci U S A 110, E4808-4817.

Wen, T.J., Hochholdinger, F., Sauer, M., Bruce, W., and Schnable, P.S. (2005). The *roothairless1* gene of maize encodes a homolog of *sec3*, which is involved in polar exocytosis. Plant Physiol 138, 1637-1643.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., and Schulman, A.H. (2007). A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8, 973-982.

Winkler, R.G., and Freeling, M. (1994). Physiological genetics of the dominant gibberellin-nonresponsive maize dwarfs, *dwarf8* and *dwarf9*. Planta 193, 341-348.

Woll, K., Borsuk, L.A., Stransky, H., Nettleton, D., Schnable, P.S., and Hochholdinger, F. (2005). Isolation, characterization, and pericycle-specific transcriptome analyses of the novel maize lateral and seminal root initiation mutant *rum1*. Plant Physiol 139, 1255-1267.

Woodhouse, M.R., Schnable, J.C., Pedersen, B.S., Lyons, E., Lisch, D., Subramaniam, S., and Freeling, M. (2010). Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. PLOS Biol 8, e1000409.

Yamaguchi, M., and Sharp, R.E. (2010). Complexity and coordination of root growth at low water potentials: recent advances from transcriptomic and proteomic analyses. Plant Cell Environ 33, 590-603.

Yu, P., Eggert, K., von Wiren, N., Li, C., and Hochholdinger, F. (2015). Cell type-specific gene expression analyses by RNA sequencing reveal local high nitrate-triggered lateral root initiation in shoot-borne roots of maize by modulating auxin-related cell cycle regulation. Plant Physiol 169, 690-704.

Yu, P., **Marcon, C.**, Baldauf, J.A., Frey, F., Baer, M., and Hochholdinger, F. (2018). Transcriptomic dissection of maize root system development. The *Zea mays* Genome. 1st ed. (eds. J. Bennetzen, S. Flint-Garcia, C. Hirsch and R. Tuberosa), 247-257.

Yu, P., Baldauf, J.A., Lithio, A., **Marcon, C.**, Nettleton, D., Li, C., and Hochholdinger, F. (2016). Root type-specific reprogramming of maize pericycle transcriptomes by local high nitrate results in disparate lateral root branching patterns. Plant Physiol 170, 1783-1798.

Zhang, Y., **Marcon, C.**, Tai, H., von Behrens, I., Ludwig, Y., Hey, S., Berendzen, K.W., and Hochholdinger, F. (2016). Conserved and unique features of the homeologous maize Aux/IAA proteins ROOTLESS WITH UNDETECTABLE MERISTEM 1 and RUM1-like 1. J Exp Bot 67, 1137-1147.

Zhang, Y., Paschold, A., **Marcon, C.**, Liu, S., Tai, H., Nestler, J., Yeh, C.T., Opitz, N., Lanz, C., Schnable, P.S., and Hochholdinger, F. (2014). The Aux/IAA gene *rum1* involved in seminal and lateral root formation controls vascular patterning in maize (*Zea mays* L.) primary roots. J Exp Bot 65, 4919-4930.

Zhu, W., Smith, J.W., and Huang, C.M. (2010). Mass spectrometry-based label-free quantitative proteomics. J Biomed Biotechnol 2010, 840518.

10 Anhang

Die in dieser Arbeit vorgestellten Forschungsergebnisse zur funktionellen Genomik der Keimlingswurzeln von Mais sind in Form von Orginal-Publikationen angehängt. Die Publikationen sind chronologisch nach ihrem Erscheinungsdatum geordnet.

Bioinformatics, 2021, 1–2
doi: 10.1093/bioinformatics/btab679
Advance Access Publication Date: 29 September 2021
Applications Note

Downloaded from https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btab679/6377773 by ulbbonn user on 21 December 2021

Genome analysis

MuWU: Mutant-seq library analysis and annotation

Tyll Stöcker [1,*], Lena Altrogge [1], Caroline Marcon [2], Yan Naing Win [2], Frank Hochholdinger [2] and Heiko Schoof [1,*]

[1]Crop Bioinformatics, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, 53115 Bonn, Germany and [2]Crop Functional Genomics, Institute of Crop Science and Resource Conservation (INRES), University of Bonn, 53115 Bonn, Germany

*To whom correspondence should be addressed.

Associate Editor: Tobias Marschall

Received on April 13, 2021; revised on July 22, 2021; editorial decision on September 22, 2021; accepted on September 23, 2021

Abstract

Motivation: Insertional mutagenesis allows for the creation of loss-of-function mutations on a genome-wide scale. In theory, every gene can be 'knocked out' via the insertion of an additional DNA sequence. Resources of sequence-indexed mutants of plant and animal model organisms are instrumental for functional genomics studies. Such repositories significantly speed up the acquisition of interesting genotypes and allow for the validation of hypotheses regarding phenotypic consequences in reverse genetics. To create such resources, comprehensive sequencing of flanking sequence tags using protocols such as Mutant-seq requires various downstream computational tasks, and these need to be performed in an efficient and reproducible manner.

Results: Here, we present MuWU, an automated **Mu**tant-seq **w**orkflow **u**tility initially created for the identification of *Mutator* insertion sites of the *BonnMu* resource, representing a reverse genetics mutant collection for functional genetics in maize (*Zea mays*). MuWU functions as a fast, one-stop downstream processing pipeline of Mutant-seq reads. It takes care of all complex bioinformatic tasks, such as identifying tagged genes and differentiating between germinal and somatic mutations/insertions. Furthermore, MuWU automatically assigns insertions to the corresponding mutated seed stocks. We discuss the implementation and how parameters can easily be adapted to use MuWU for other species/transposable elements.

Availability and implementation: MuWU is a Snakemake-based workflow and freely available at https://github.com/tgstoecker/MuWU.

Contact: tyll.stoecker@uni-bonn.de or schoof@uni-bonn.de

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Both in forward and reverse genetic studies, sequence-indexed insertional libraries constitute essential resources for researchers around the globe. Notable examples are transgenic T-DNA insertion lines in Arabidopsis (Alonso *et al.*, 2003) or several mutant collections based on the use of transposable elements (TEs) in rice (Hirochika *et al.*, 2004) and maize (Liang *et al.*, 2019; Marcon *et al.*, 2020; McCarty *et al.*, 2005). For efficiency, when screening large mutant collections, targeted sequencing approaches such as Mutant-seq (Mu-seq; McCarty *et al.*, 2013; Supplementary Fig. S1) are used.

Mutagenized families are pooled according to a grid design (Supplementary Fig. S2; McCarty *et al.*, 2013; Urbański *et al.*, 2012). Mu-seq can take this one step further by distinguishing between somatic and germinal (heritable) mutations. If row and column pools are taken from independent somatic cell lineages, heritable insertions have to appear in both axes of the grid and can thus be singled out for further analyses. To our knowledge, MuWU is the first openly available tool for the analysis of insertional libraries generated using a Mu-seq approach. MuWU efficiently combines the necessary bioinformatics analyses to enable the benefits of the Mu-seq approach: filtering of pre-existing insertion sites as well as somatic insertions. In contrast to other approaches, MuWU does not require paired-end sequencing.

MuWU was created as a solution to the recurring processing of *Mu* insertional mutagenesis sequencing libraries in maize as part of the newly created and expanding *BonnMu* resource (Marcon *et al.*, 2020), the first European mutant resource of its kind. *BonnMu* insertions are continuously being integrated into the MaizeGDB.org genome browser (https://www.maizegdb.org; Portwood *et al.*, 2019).

2 Software description

2.1 Input files and data preparation

All software and dependencies are either installed at runtime or run inside a singularity container allowing for completely automated detection and annotation of insertion sites. The grid design of a

© The Author(s) 2021. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Mu-seq experiment results in $2n$ samples, with the size n in our experiments usually being 24 pools of 24 F2 families each. This results in the sequencing of 576 families in 48 sequencing libraries (24 row and 24 column pools). In addition to sequencing reads, MuWU requires genome sequence FASTA file as well as a suitable annotation file. Also, a library-specific stock matrix table should be supplied indicating the position of each mutagenized maize family in the row x column grid design to infer germinal insertions (Supplementary Fig. S2).

The first steps of MuWU deal with quality control and alignment of the reads to the reference genome sequence. Most notable are the removal of transposon terminal inverted repeat (TIR) sequences and adapters from both ends of the raw Mu-seq reads. A summary of all statistics is generated in one HTML report.

2.2 Annotation procedure

Insertion sites are identified with our Python tool (insertions.py) which takes advantage of the TE-specific target site duplication (TSD) at the insertion flanking region. At this first step, at least two sequencing reads at both sides of the TSD are required to support the insertion. In a second step, germinal insertion sites of one row and one column pool of the grid layout are filtered to allow assignment of a distinct mutagenized family. Thus, heritable mutations are identified by keeping insertions with shared genomic coordinates in only one row and one column sample each. Parallel implementation of both steps of the algorithm allows us to circumvent this initial computational bottleneck - with a 1 thread per sample granularity reducing runtime significantly.

In MuWU's subsequent annotation, we analyze insertion sites inside or near gene models. By identifying the specific combination of row and column pool for each germinal insertion event, we can perform a matrix lookup in the library-specific seed stock table and add stock information to the output. The final outputs are a table of germinal insertions, their read coverage and corresponding stock, as well as id, gene lengths and coordinates of the genes the insertions were assigned to, and a similar table for all insertions. The complete workflow is shown in Supplementary Figure S3.

3 Conclusion

MuWU is an efficient workflow solution reducing the bioinformatics steps of the *BonnMu* database resource to a one-command job finished in <1 h per library (Intel(R) Xeon(R) CPU E5-2690 v4@ 2.60 GHz; 24 cores). It implements the bioinformatics part of Mu-seq, an improved strategy to distinguish between germinal and somatic insertions, combining the advantages of an experimental layout in grid design with the analysis of TSDs (Liang *et al.*, 2019; McCarty *et al.*, 2013). Automation of library annotation is essential for reproducibility and consistency as we continue to expand our database effort with upcoming libraries.

While MuWU is not the only bioinformatics tool for TE insertion detection, it is to our knowledge the only openly available tool suitable for Mu-seq data and, in contrast to other tools for TE insertion detection, works with single-end sequencing as it does not rely on the analysis of discordant read pairs. MuWU is not specific for *Mutator* insertions or maize but can detect any insertion event that causes TSDs. The user can configure TE-specific and adapter sequences, TSD length and read support threshold. We have implemented a secondary mode ('GENERIC') which does not require an experimental design that allows inferring germinal insertion events and is thus more widely applicable. As the workflow is built on Snakemake (Köster and Rahmann, 2012), the addition of further analyses to the automatic handling of Mu-seq libraries allows for modular expansion of its current state.

Acknowledgement

The authors thank Alexa Brox (University of Bonn) for technical assistance.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft [MA 8427/1-1 to C.M.].

Conflict of Interest: none declared.

References

Alonso,J.M. *et al.* (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. *Science*, **301**, 653–657.

Hirochika,H. *et al.* (2004) Rice mutant resources for gene discovery. *Plant Mol. Biol.*, **54**, 325–334.

Köster,J. and Rahmann,S. (2012) Snakemake-a scalable bioinformatics workflow engine. *Bioinformatics*, **28**, 2520–2522.

Liang,L. *et al.* (2019) A sequence-indexed mutator insertional library for maize functional genomics study. *Plant Physiol.*, **181**, 1404–1414.

Marcon,C. *et al.* (2020) *BonnMu*: a sequence-indexed resource of transposon-induced maize mutations for functional genomics studies. *Plant Physiol.*, **184**, 620–631.

McCarty,D.R. *et al.* (2005) Steady-state transposon mutagenesis in inbred maize. *Plant J.*, **44**, 52–61.

McCarty,D.R. *et al.* (2013) Mu-seq: sequence-based mapping and identification of transposon induced mutations. *PLoS One*, **8**, e77172.

Portwood,J.L. *et al.* (2019) Maizegdb 2018: the maize multi-genome genetics and genomics database. *Nucleic Acids Res.*, **47**, D1146–D1154.

Urbański,D.F. *et al.* (2012) "Genome-wide LORE1 retrotransposon mutagenesis and high-throughput insertion detection in Lotus japonicus". *Plant J.*, **69**, 731–741.

Downloaded from https://academic.oup.com/bioinformatics/advance-article/doi/10.1093/bioinformatics/btab679/6377773 by ulbbonn user on 21 December 2021

ARTICLES
https://doi.org/10.1038/s41588-020-0671-9
nature genetics

Check for updates

OPEN

European maize genomes highlight intraspecies variation in repeat and gene content

Georg Haberer[1], Nadia Kamal[1], Eva Bauer[2], Heidrun Gundlach[1], Iris Fischer[1], Michael A. Seidel[1], Manuel Spannagl[1], Caroline Marcon[3], Alevtina Ruban[4,5], Claude Urbany[5], Adnane Nemri[5], Frank Hochholdinger[3], Milena Ouzunova[5], Andreas Houben[4], Chris-Carolin Schön[2]✉ and Klaus F. X. Mayer[1,6]✉

The diversity of maize (*Zea mays*) is the backbone of modern heterotic patterns and hybrid breeding. Historically, US farmers exploited this variability to establish today's highly productive Corn Belt inbred lines from blends of dent and flint germplasm pools. Here, we report de novo genome sequences of four European flint lines assembled to pseudomolecules with scaffold N50 ranging from 6.1 to 10.4 Mb. Comparative analyses with two US Corn Belt lines explains the pronounced differences between both germplasms. While overall syntenic order and consolidated gene annotations reveal only moderate pangenomic differences, whole-genome alignments delineating the core and dispensable genome, and the analysis of heterochromatic knobs and orthologous long terminal repeat retrotransposons unveil the dynamics of the maize genome. The high-quality genome sequences of the flint pool complement the maize pangenome and provide an important tool to study maize improvement at a genome scale and to enhance modern hybrid breeding.

Since its domestication ~10,000 years ago by Native Americans, maize (*Zea mays* ssp. *mays*) has become one of the most important sources for human nutrition and animal feeding. Extensive variation in landraces and breeding germplasm such as dent or flint corns underpins the enormous phenotypic and genetic diversity of maize[1–3]. Today, US hybrids produced from inbred lines of different heterotic groups; for example, stiff stalk and nonstiff stalk[4] are highly productive and agriculturally important worldwide. These US Corn Belt dents resulted from crosses in the nineteenth century between Southern dent lines introduced to the United States from Mexico and Northern flints (NF) which were the predominant germplasm in the pre-Columbian era grown by the Native Americans[5,6]. Historical records and genetic and molecular data strongly indicate that NFs originated from Native American populations of Southwestern America and the Great Plains[2,6,7]. Improvement by early farmers and adaptation to cooler climate and different photoperiods extended NF growth as far north as Southern Canada. After the colonization of the New World, maize was spread to Europe both from the Caribbean islands and Northeast United States[8]. The early maturing and cold-tolerant flints were key to a successful maize cultivation in temperate regions of Europe. Consequently, NF germplasm still makes a major contribution to modern European maize breeding material while Corn Belt dent genomes contain on average one-quarter from their blending with NFs[2,6].

The complete genome sequence of the dent line B73 and its continuous updates provide a high-quality reference for maize breeding, genetic and genomic research[9,10]. However, very high diversity at both the sequence and genic level has been reported between maize inbred lines for targeted regions and large-scale comparisons[3,11–14]. Hence, the B73 reference sequence captures only a portion of the maize pangenome. To overcome these limitations, a draft genome of the Iodent line PH207 (ref. [15]) and—more recently—reference sequences of SK, Mo17 and W22, a tropical and two US dent lines, have been released[16–18]. Worldwide, many hybrid breeding programs focus on dent germplasm, whereas breeding programs in cooler regions of Central Europe exploit heterotic effects between dent and flint lines. Several studies have shown a clear differentiation of the North American and Northern European flint germplasm from the rest of the world[19]. While reference-quality sequences exist for several dent inbred lines, the flint pool is still underexploited. To date, only one fragmented draft assembly with an N50 ~13.9 kilobases (kb) of one flint line, F2, is available covering ~65% of the estimated genome size[20]. To better understand the extent and quality of genomic differences between the flint and dent germplasm pools we generated de novo high-quality sequences for four flint inbred lines representing important ancestors of maize hybrid breeding in Central Europe[21] and contrasted in this study their gene and repeat contents to two US Corn Belt dents.

Results

Whole-genome sequencing. To access the genomic landscape of European flint maize, we assembled four flint lines to reference genome quality. Three lines are important founders of European breeding programs and represent distinct flint germplasm sources such as populations Lacaune (F7, Southern France), Lizargarate (EP1, Northern Spain) and Gelber Badischer Landmais (DK105, Southern Germany). The fourth sample, PE0075, is a doubled-haploid line derived from the Petkuser Ferdinand Rot population, a landrace from Northern Germany. We generated Illumina paired-end and mate-pair sequences equivalent to ~220–320× coverage (Supplementary Table 1) and assembled the reads to

[1]Plant Genome and Systems Biology, Helmholtz Center Munich, Neuherberg, Germany. [2]Plant Breeding, School of Life Sciences, Technical University Munich, Munich, Germany. [3]Crop Functional Genomics, Institute for Crop Science and Resource Conservation, University of Bonn, Bonn, Germany. [4]Leibniz Institute of Plant Genetics and Crop Plant Research, Seeland, Germany. [5]KWS SAAT SE, Einbeck, Germany. [6]School of Life Sciences, Technical University Munich, Munich, Germany. ✉e-mail: chris.schoen@tum.de; k.mayer@helmholtz-muenchen.de

Table 1 | Genome statistics of maize dent and flint lines used in this study

	Zea mays 'flint' lines				*Zea mays* 'dent' lines	
Line	**EP1**	**F7**	**DK105**	**PE0075**	**B73 (v.4)**	**PH207**
Assembly size (Mb)	2,455.3	2,392.8	2,288.2	2,198.5	2,135.1	2,156.2
Chromosome (Mb)	2,321.0	2,255.5	2,176.1	2,140.7	2,106.3	2,060.3
Scaffolds (*n*)	60,567	62,610	797	972	267	4,3291
Scaffold N50 (Mb)	6.13	9.48	10.39	8.64	ND	ND
Genes (*n*)	47,174	48,068	46,697	46,742	48,003	46,207
Orthologs	45,630	45,932	45,697	45,824	45,928	43,777
Class I repeats (% of assembly)	79.1	78.6	78.3	77.8	77.2	74.9
Class II repeats (% of assembly)	2.03	2.07	2.11	2.16	2.28	2.29

The table summarizes total assembly size, total size of assembled chromosomes, number of pseudochromosomes including unanchored scaffolds, N50 of scaffolds before pseudochromosome generation, total size of undefined sequence in the genome, number of genes (including low-confidence genes; see Main text and Supplementary Information), orthologs (defined as bidirectional best Blast hits) and total percentages of class I and class II repeats. ND, not determined.

scaffolds and pseudomolecules using the DeNovoMagic pipeline[15,16,22]. Total assembly sizes of the pseudochromosomes range from 2.14 to 2.32 megabases (Mb) and amount to 92.9–100.5% of the genome sizes estimated by *K*-mer analysis (Table 1 and Supplementary Fig. 1). Most (94.3–97.3%) contig and scaffold sequences were integrated into ten pseudochromosomes (Table 1, Supplementary Table 2 and Extended Data Fig. 1a). A genetic map generated from an F_2 mapping population of an EP1 × PH207 cross demonstrated a high consensus between genetic and physical map corroborating quality and contiguity of the maize assemblies (Extended Data Fig. 2). Sequence accuracy was high with less than one erroneous base per 100 kb (Supplementary Table 3) and remapping genomic reads using SQUAT[23] reported <2% poorly mapped reads. Complete BUSCO[24] genes totaled >95%, strongly supporting high coverage of the gene space while only ~1.5% and 0.8–2.4% of the BUSCO Liliopsidae set were absent or fragmented in each of the four assemblies on average, respectively (Supplementary Table 2). In addition, the long terminal repeat (LTR) assembly index[25], a measure for the correct reconstruction of the repetitive transposon space, supports the reference quality of our four flint assemblies (Supplementary Table 4).

Pan-gene variation among flint and dent is moderate. To assess genic presence–absence variations (PAVs) in the six maize lines we used a three-layer gene prediction pipeline. For F7 and EP1 we predicted protein-coding genes as consensus models using protein homologies of known monocotyledonous protein sequences and a broad spectrum of transcriptome evidences from F7 and EP1 RNA-seq data (Supplementary Table 5). Subsequently, cross-mapping of the EP1 and F7 models as well as annotations of B73 (v.4) and PH207 (v.1.1) complemented the gene sets of all six maize lines including PE0075 and DK105. In total, we identified ~46,200–48,000 consolidated gene models per line (Table 1). Cross-mapping and consolidation substantially improved the number of pairwise reciprocal best blast orthologs (Fig. 1a and Supplementary Table 6) as well as the completeness of orthologous clusters (Fig. 1b and Supplementary Fig. 2).

Overall, 43,700–45,900 genes (94.7–98% of all genes per line) have an orthologous counterpart or reciprocal best blast match in at least one of the other lines (Fig. 1b). Comparison of the original and consolidated annotations of B73 and PH207 suggests that part of this gene complement displays functional signatures such as homology to angiosperm proteins or expression in the respective line (Fig. 1a and Supplementary Fig. 3). The cross-mapped annotation significantly improved the completeness of syntelog clusters with 34,352 clusters that have syntelogs for all six lines, and 94.6% of these contain exactly one gene copy per line while 1,861 (5.4%) contain additional co-orthologous members (Fig. 1b). Genic PAVs strongly support flint and dent as distinctive germplasm (Extended Data Fig. 3). Overall, core genes present in all six maize lines show high expression levels (mean and median 25.8 and 5.9 transcripts per million, TPM) while PAV genes show significantly lower expression levels (median 0.02–0.06 TPM) and have lower homology support in a comparison to 27 angiosperms (Fig. 1c,d). Note that genes found in only one line in general show more pronounced support for protein-coding sequences compared to genes of cluster sizes 2–4. Genes with no corresponding maize ortholog (cluster size 1) are enriched for tandem duplications (Fisher exact test with *P* values between $<1\times10^{-8}$ and 1×10^{-70}). The very low median expression of syntenic dispensable and singleton genes suggests no, only minor or highly specialized line-specific functions for most of these genes.

Consistent with the genome analysis of the maize inbred line Mo17 (ref. [17]), our cross-consolidation revealed a sizeable number of predicted genes with large effect mutations. Between 2,487 (B73) and up to 4,264 (DK105) gene models had no contiguous ORF due to internal stop codons or frameshifts. Genes with disrupted ORFs had highly similar expression levels (Pearson correlation $0.93 < r < 0.95$) to their syntenic partner with contiguous reading frame (Extended Data Fig. 4). Close inspection revealed a series of reasons including pseudogenization, missing splice variants or skipped exons in the respective line.

Dynamics of the full-length retrotransposon landscape. Approximately 80% of the maize genome is constituted by repeats of different types. In addition to the six lines above, the recently published genomes Mo17 and W22 were analyzed for their repeat content[16,17]. No pronounced differences in transposable element (TE) composition among the eight lines analyzed was detected (Extended Data Fig. 1 and Supplementary Table 7). We detected almost 15,000 high-quality full-length LTRs (fl-LTRs) per line, matching the expected genome size to fl-LTR ratio[22] (Supplementary Table 8a). The lower number observed for PH207 (6,838 fl-LTRs) can be attributed the overall lower quality assembly compared to the other five lines. Correspondingly, the PH207 assembly was excluded from evaluation depending on the analysis criteria. For the remaining five lines, the number and age distribution of fl-LTRs confirm the high quality of the respective assemblies (Extended Data Fig. 5).

To quantify transposon dynamics between the different maize lines we identified still-shared syntenic fl-LTRs by clustering TE junctions with high stringency. Only 3% of all fl-LTR locations were found to be shared between six lines. While the percentage of shared elements among the different lines shows a marked decrease, the

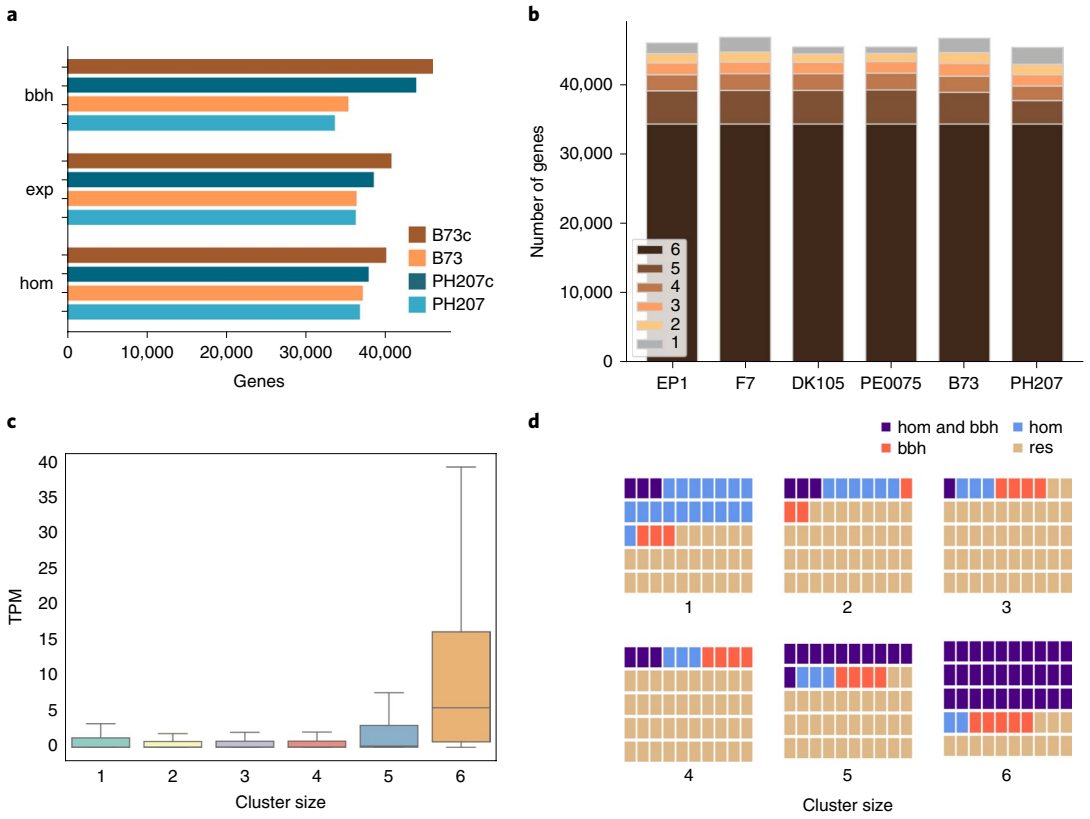


Fig. 1 | Gene consolidation and characteristics. a, Comparison of reciprocal best blast orthologs (bbh), genes with a TPM ≥0.1 (exp) and angiosperm proteins (hom) between the consolidated (B73c and PH207c) and initial B73 and PH207 gene models. **b**, The number of genes per line that are either singletons (label 1) or syntenic between two and six lines (labels 2–6). **c**, Expression levels (TPM) of EP1 and F7 genes graded by size of their orthologous clusters. Genes either are singletons (cluster size 1) or have orthologs in one to five maize lines (cluster sizes 2–6). Syntelogs of orthologous clusters comprising 2–4 lines show very low median expression levels (TPM ~0.06). **d**, Illustration per cluster size of the proportion of orthologous groups with support by bbh and homology (purple), only bbh (red), only homology (hom, blue) and residual clusters (res, tan) matching neither the homology nor the bbh criterion. To record homology, a minimum alignment length of 80% of both maize and angiosperm protein was required in at least three species. The high conservation observed in cluster size 1 reflects, in part, enrichment of tandemly repeated genes duplicated from orthologs of clusters present in five or six lines.

pangenome of the fl-LTR space almost doubles the number of elements accordingly. This high dynamic is in stark contrast to genes where 76.1% are retained at their syntenic position (Fig. 2a). Pairwise cross-comparison among the different genomes detected between 18 and 32% of still syntenic fl-LTRs to be present in the corresponding line. Differences in pairwise shared numbers match the phylogenies from gene-derived phylogenetic relationships and reveal a clear distinction between flint and dent lines (Fig. 2b). Line-specific fl-LTRs are younger and depleted in the pericentromeric regions. Along with increased sharing, a continuous enrichment towards the central low recombining compartments is observed and associated with an increase in age (Fig. 2c). Our findings illustrate the rapid turnover of the intergenic space most likely driven by elimination through illegitimate recombination.

Heterochromatic knob islands in the maize genome. Knob regions are heterochromatic regions in the genome that have been demonstrated to affect local recombination[26,27]. They belong to the group of satellite tandem repeats which are well represented in our flint assemblies (Extended Data Fig. 6a, Supplementary Fig. 4 and Supplementary Table 8b). Maize knobs are composed of 180 bp and closely related 202-bp tandemly repeated sequence units (Extended Data Fig. 6c). Variations in position and extent of knob regions in maize have been documented[26,28]. Intensity and position of knob regions detected by FISH for the different karyotypes clearly separate dent from flint lines (Fig. 3a and Extended Data Fig. 6b). In particular, extensive knob regions detected on chromosomes 7 and 8 of the dent lines are absent in flint lines. For other positions on the genome, a clear flint/dent separation among lines cannot be observed. However, the observed variations reveal the pervasive dynamics of knob regions in the genome. FISH requires a minimum number of adjacent tandem units for their detection. High-quality genome assemblies resolving a larger proportion of the repetitive space allow analyzing the so-far-hidden, cryptic, knob sites which we annotated by homology to the 180 and 202 bp monomers. Besides the large knob regions detected by FISH, numerous additional positions with fewer tandem repeat units are found in the sequence assembly on all chromosomes (Fig. 3b and Extended Data Fig. 7). A strikingly pronounced syntenic conservation of minor knob sites common to all analyzed flint and dent lines was observed. Given the observations of rapid diversification based on repeat units detailed above, this might indicate numerous potential (shadowed) knob regions that can rapidly expand or shrink by, for example, illegitimate recombination and can cause recombinational isolation of the affected regions.

In fission yeast, knob structures have been shown to implicate significant downregulation of surrounding genes[29]. To check for similar effects in maize we analyzed the distance of genes to the nearest knob signature in the assembly and related it to expression data[30] generated for the lines B73, EP1 and F7. Even more pronounced than observed for other types of repeat elements, we find the close

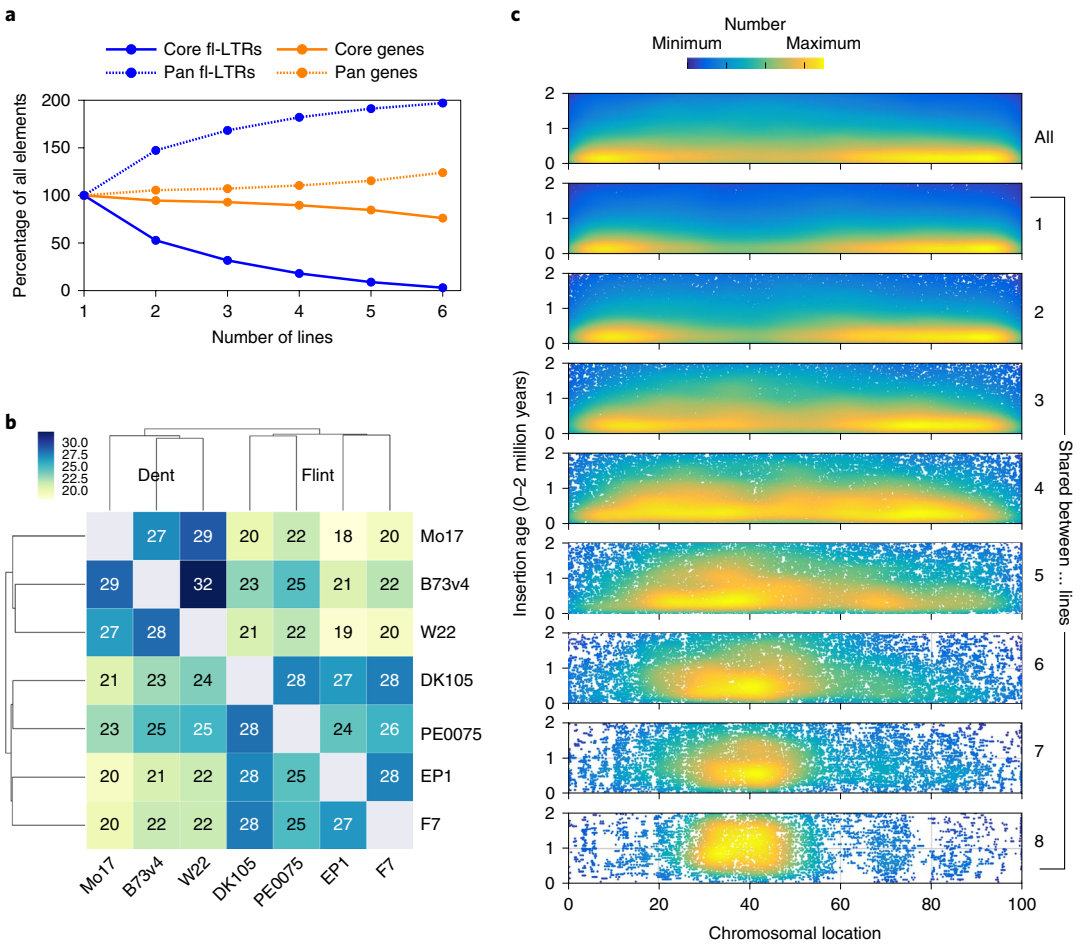


Fig. 2 | Pan and core characteristics of fl-LTR. a, Proportion of the pan and core sets for fl-LTRs and genes with increasing line numbers. **b**, Percentage of pairwise shared and still-intact fl-LTRs at syntenic positions for seven maize lines. Reading direction is column to row; for example, EP1 shares 27% of its fl-LTRs with F7 and F7 shares 28% with EP1. On the basis of the similarity matrix, the seven lines cluster into a relationship context that separates flint and dent. Within flint and dent, around 30% of the locations are shared; between flint and dent, most of the values are reduced to about 20%, except for PE0075 (25%). The most pronounced overlap of intact and shared fl-LTR locations is found between W22 and B73 (32%); the least pronounced is found between EP1 and Mo17 (18%). A corresponding evalution for genes gives pairwise shared numbers between 82 and 91%. **c**, Insertion age (*y* axis) and chromosomal distribution of all fl-LTRs (top row), line unique (label 1) and cluster constellations of increased sharing ranging from two to all eight lines. The chromosomal location is collapsed for all ten chromosomes and given in percentage of the respective chromosome length. The line-specific or shared among fewer lines fl-LTRs contain a higher proportion of younger elements and are less frequently found in the central, low recombining regions. There is a continuous shift towards a more pericentromeric location and towards older elements with the increase of lines sharing corresponding elements.

surrounding of knob repeat units to be devoid of genes (Fig. 3c). For genes surrounding knob regions, we observe a trend for lower expression values (Extended Data Fig. 7d).

Whole-genome alignment. Numerous large structural variations (SV) in maize have been reported[11,12,14]. To gain insight into the extent of SVs in the maize flint versus dent genomes we generated pairwise whole-genome alignments (WGAs) for the six lines and postprocessed the resulting alignments to single alignment blocks (SAB) that represent the highest scoring one-to-one relationships between each genome pair. To approximate overall contiguity between two genomes, SABs were further concatenated to merged alignment blocks (MABs) if they followed a strict and unambiguous order in both genome sequences (Supplementary Fig. 5). On average, ~50% of the genome sequence aligned in each of the pairwise comparisons by SAB scoring, while WGAs (MABs) associated 80–90% (1.7–2 Gb) of the genome sequences (Supplementary Tables 9 and 10). Mean sizes of SABs and MABs were ~10 kb and ~100 kb, respectively (Supplementary Fig. 6). Evaluation of pairwise WGA regions strongly supported orthology of segments in parallel orientations (Supplementary Fig. 7) while an assessment of most rearrangements (inversions and transpositions) was inconclusive due to insufficient or ambiguous information at the breakpoints.

To delineate regions that do not align with any other of the five lines (unaligned), align solely with flint or dent maize lines (group-specific) or are aligned in all pairwise alignments (core), we combined pairwise WGAs of each line and projected them onto its genome sequence (Supplementary Fig. 8). We classified a genomic region as core type if it aligned to four out of the five possible lines due to the high amount of gap sequences in PH207 that likely lead to an underestimation of the true core genome. The core genome defined by SABs comprised on average ~850 Mb (40%) of the total genome while 71 Mb (3.3%) were group-specific and ~460 Mb (22%) were unaligned. This indicates a large fraction of uniquely inserted or deleted sequences in each line. Positional analysis of genomic bins revealed significant positive correlations (Kendall's $\tau \geq 0.38$; $P < 10^{-36}$) between the core and repeat density while gene densities were negatively correlated ($-0.4 < \tau < -0.33$; $P < 10^{-26}$)

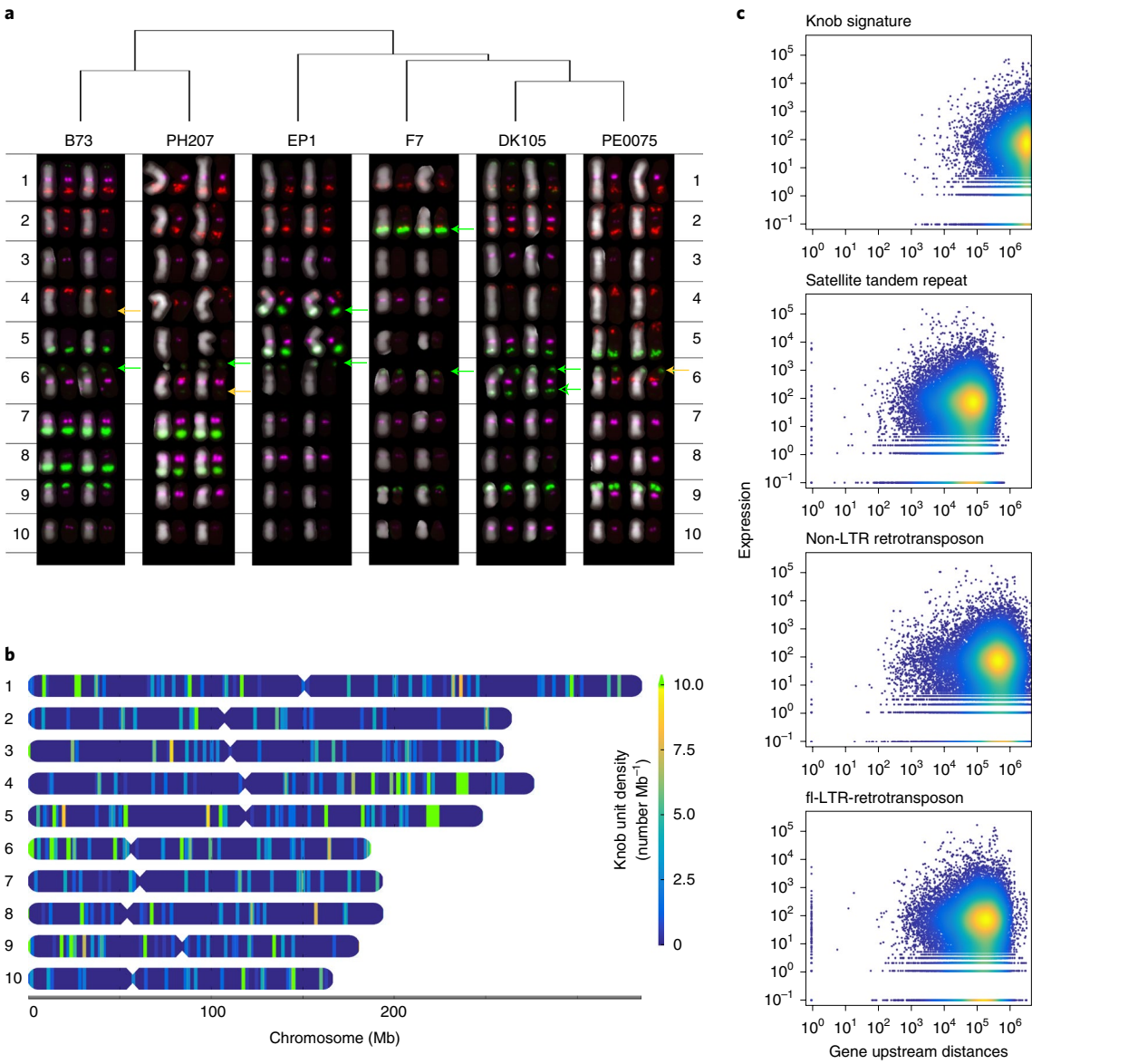


Fig. 3 | Diversity of knob locations in six maize lines. a, Karyotyping of all six maize lines by FISH. Colors show: red, (ACT)10; green, Knob-2 (180-bp knob repeat) and MR68-3 (chromosome 6 clone MR68, ID AF020265.1); magenta, CentC69-1. Green arrows indicate co-occurrences of Knob-2 and MR68-3; orange arrows indicate a sole occurrence of MR68-3. **b**, Chromosomal locations of all knob sequences in EP1. Besides the three major Mb scaled knob regions identified by FISH on chromosome 4L, 5L and 6S, many smaller knob sequences (<10 kb; below the FISH detection limit) are found scattered along all chromosomes. Interestingly, many of the minor knob locations are shared between the lines (Extended Data Fig. 7). The polymorphisms seen by FISH seem to be based on large size variations of the mostly syntenic minor sites (7). **c**, Relation between gene neighborhood and gene expression. For each gene the maximal expression of seven different conditions is plotted against its upstream distance (bp) to the next neighboring element of a specific type. Both axes are logarithmic. It is noteworthy that knob sequences in particular seem to be a 'bad' neighborhood for genes. Here, in contrast to other elements a 1,000-bp perimeter was found to be devoid of genes.

(Fig. 4a and Extended Data Fig. 8). Notably, for all six lines highest core densities located at regions adjacent to the centromere and correspond to above-average SAB sizes and below-average recombination rates (Fig. 4b). These findings are reminiscent of observations in the Triticeae gene space[22] mirroring the recombinogenic properties of centromeric and telomeric regions.

Genetic mosaicism in US dents and European flints. The core genomic regions served as positional anchors to generate multiple sequence alignments. One-to-one unambiguous alignments for the six lines total 287 Mb (~15%) of each genome and comprise 6.25×10^6 orthologous SNPs with positional information for each of the six genomic coordinate systems. The detected SNPs showed an excellent agreement of 99.3% (4,827 mismatches out of 653,398 total scored calls) to a published set of SNP calls using the Affymetrix Axiom Maize Genotyping Array with 600,000 variants[31]. On the basis of sequences of identical SNP pairs, we determined significant haplotypes between all 15 line combinations (Fig. 4c). Mean and total genomic spans covered by these pairwise near-identical haplotypes range from 40 to 68 kb and 394 to 664 Mb, respectively. The proportion of one-fifth to one-third shared genomic regions between flint and dent lines is in close agreement with previous studies of NF content in stiff stalk and nonstiff stalk dents[2,6]. Number, mean and total sizes were highly consistent with their phylogenetic relationship

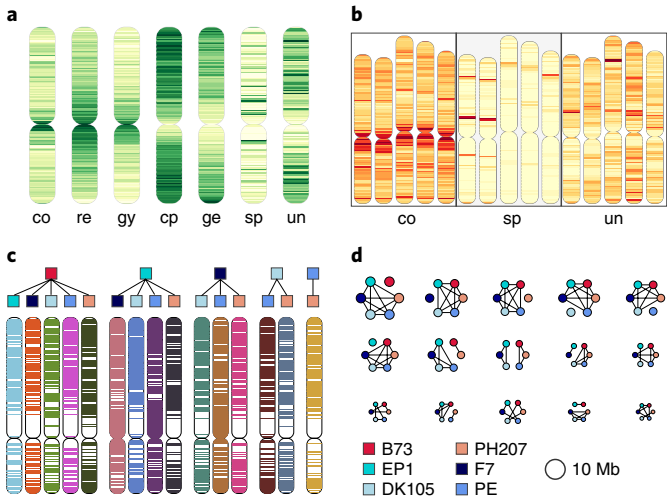


Fig. 4 | Comparative genome characteristics for the six maize lines. a, The density distribution of the WGA (co, core; sp, germplasm-specific; un, unaligned regions) and functional genomic elements (re, all repeats; gy and cp, gypsy and copia LTR elements; ge, genes). The core WGAs are significantly positively correlated to the repeat and gypsy densities, similar to densities of copia elements and genes. For the unaligned and germplasm-specific densities, we detected no highly significant correlations with one of the other functional elements. **b**, Densities of the core, germplasm-specific and unaligned WGA regions showed similar distributions for all maize lines, exemplified for chromosome 1 of B73, PH207, EP1, F7 and PE0075. Core regions are enriched at (peri-)centromeric regions, germplasm-specific obviously cluster in a group-specific manner and unaligned parts are random within all six lines. **c**, Runs-of-identity for SNPs from core regions for chromosome 7 and all 15 pairwise combinations. Color code above the chromosome pictograms follows legend in **d**. **d**, Sizes of higher order haplotypes of the six lines. Pictograms are proportional to their total genomic span, the legend provides the color coding of each of the six lines. Black lines indicate identity. Interestingly, the top six higher order haplotypes each group one maize variety as outlier and five lines that are identical by their SNP runs for these genomic segments. Of the more complex groupings, the most prominent haplotype separates the two germplasm, flint and dent.

deduced from high-confidence orthologs (Extended Data Fig. 9). However, each line of one group (flint or dent) shared substantial genomic portions with the other germplasm group (394 to 480 Mb for B73 and EP1 or PE0075, respectively). This is in line with the history of US Corn Belt dents that originated from crosses of Southern dents and flints of North America and Canada[6]. The latter established many founder lines of modern European flint breeding and were introduced to Europe as early as the discovery of the New World[8]. Alternatively, these haplotypes could be introgressions of US germplasm into European flint material that started in the 1950s to broaden genetic diversity in European breeding programs. Notably, the DH line of the European landrace Petkuser showed the largest genome-wide similarity of all flints towards B73.

Next, we surveyed higher order haplotypes considering all 15 pairwise similarities simultaneously as combined binary patterns. Limiting these haplotypes to a minimum run size of 40 SNPs, we identified 31 fitting this size criterion (Fig. 4d and Supplementary Table 11). Since the identified regions exceed random expectation by far and show near-complete SNP identity across particular line subsets/combinations, these regions are likely identical-by-descent and share a common ancestor. In total, the 31 distinct higher order haplotypes span ~288 Mb and comprise $>1.1 \times 10^6$ orthologous SNPs. Note that ordering by the number of SNPs as well as the genomic region covered, the top five/six haplotypes encode combinatorial groupings with five out of six lines being identical. The haplotype distinguishing between the studied dent and European flint lines ranked only as the sixth (by size) or seventh (by SNP number) most frequent (Fig. 4d and Supplementary Table 11). The distinct higher order haplotypes distribute evenly along the ten chromosomes but also comprised several striking clusters. Although not the most prominent identical-by-descent type, three regions were enriched for the haplotype distinctive between flint and dent lines: two at the distal sites of chromosomes 4 and 7 and a large portion spanning ~12 Mb in the proximal part of chromosome 8. The latter strongly coincides with group-specific WGAs on chromosome 8 (Extended Data Fig. 8) and encloses the major flowering time quantititive trait locus *vgt1* in maize[32,33].

Haplotype-informed differential gene expression analysis. To explore putative functional consequences of the haplotype differentiating our dent and European flint lines, we analyzed its effect on the maize transcriptome. From the maize association panel we selected a subset of 40 out of 282 lines, which were highly similar to either B73 or F7/EP1 within regions defined by this haplotype (Supplementary Fig. 9)[34,35]. We analyzed genome-wide expression levels between the two contrasting groups underlying three different genome references (B73, EP1 and F7) and using expression data from seven different conditions and tissues[30]. In total, 4,761 genes orthologous between B73, EP1 and F7 (out of 35,389 triplets) were differentially expressed (DEGs) in all three references and in at least one of the tissues and conditions (Fig. 5).

Comparing identified DEGs with curated maize genes revealed a striking pattern for genes involved in the establishment of the kernel texture. We identified a plethora of genes involved in starch metabolism: sugar transporters including several proteins of the SWEET family, *hex4* and *hex7*, sugar transporter *stp1*, the rate-limiting gene for starch biosynthesis (*shrunken2*), *starch synthase1* (*ss1*), *starch phosphorylase*, the $\alpha(1\rightarrow6)$-glucosylhydrolase *sugary*[36] and starch branching enzyme *sbe1* catalyzing the formation of $\alpha(1\rightarrow6)$-glycosidic bonds in amylopectin (Fig. 5c). We also detected a number of genes that establish and regulate the prolamin content and structure of the maize endosperm (Fig. 5b) including *opaque endosperm o7* and *o10*, *empty pericarp 16*, *floury endosperm fl3* and *naked endosperm nkd1*[37–40].

Discussion

The comparative analysis of the gene and repeat content together with the analysis of WGAs and SNPs confirms the European flints as highly differentiated germplasm of cultivated maize at a whole-genome scale. For all these analyses, the four flint and the two dent lines form two distinct groups, with each line sharing larger numbers of syntenic genes, repeats, aligned regions and haplotypes with members of the same group than with any member of the other group. Nevertheless, we also observed significant overall conservation of chromosome structure and gene content. Aligned blocks ordered by consecutive syntenic order span ~85% of the entire genome in all lines. Cross-consolidation of gene models revealed a moderate number of nonsyntenic genes, suggesting that genic PAVs are less pronounced in maize than previously reported. Consistently, significant lower support by expression and homology indicates that less conserved orthologous clusters are likely enriched for nonfunctional and poor gene models. Nevertheless, the latter set also contains many genes expressed at moderate to high levels, which might contribute to line-specific adaptations and are of interest for maize improvement and breeding. Additionally, some of these genes may have highly confined expression patterns or conditions that were missed in our transcriptome analysis. The cross-consolidation approach also had several limitations including the potential transfer of poor gene models and the identification

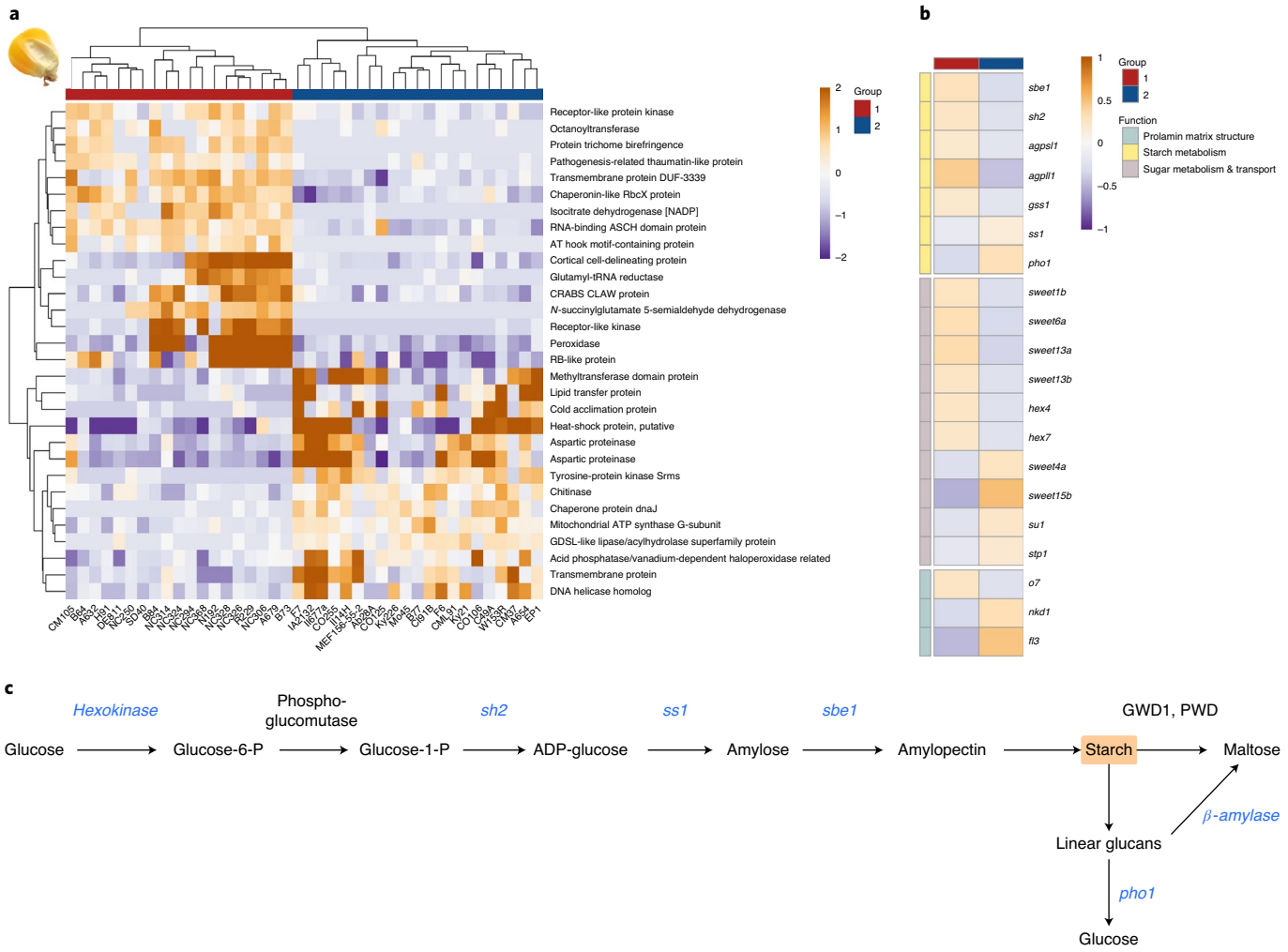


Fig. 5 | Haploblock-informed DEGs. a, Heatmap of top 30 DEGs (log_2-fold change above 2 or below −2, adjusted $P \leq 0.05$, variance stabilized) between the two haploblock-informed groups in kernels. Group 1, lines with haploblock-similarity to B73 (dent-alike); group 2, lines with haploblock-similarity to EP1/F7 (flint-alike). **b**, DEGs between the two groups that encode for components of the starch biosynthesis pathway, for proteins involved in sugar metabolism and transport into the kernel, and for proteins organizing the prolamin matrix structure of the kernel. *sbe1*, starch branching enzyme 1; *agpsl1*, ADP-glucose pyrophosphorylase small subunit leaf1; *sh2*, *shrunken2*; *ss1*, starch synthase 1; *pho1*, starch phosphorylase 1; sweet, sugar transporter; *su1*: *sugary1*; *stp1*, sugar transport protein 1; hex, hexokinase; *o7*, *opaque7*; *nkd1*, *naked endosperm*; *fl3*, *floury endosperm 3*; *PWD*, phosphoglucan, water dikinase; *GWD1*, glucan, water dikinase 1. **c**, Main reactions of starch biosynthesis and catalyzing enzymes, genes differentially expressed between the two groups of maize lines highlighted in blue.

of genes that contain large effect mutations and disrupted ORFs. Inspection of the latter set revealed several causes including likely true pseudogenizations or truncations, probable differences in the exon–intron structure between the applied informant and target gene model as well as mapping errors.

In contrast to the overall syntenic conservation, a large amount of unaligned, nonorthologous sequences for WGAs and the analysis of the repeat space demonstrated enormous diversity and dynamics of the maize genome. Only 2% of shared orthologous fl-LTRs are conserved in eight lines. Consistent with previous reports, large FISH visible knobs display substantial variation in position among the maize lines studied[28,41]. Remarkably, both shared sets of orthologous LTRs and larger knob positions perfectly reproduced the relationship of the six lines based on coding sequences, indicating smooth transitions between these features during breeding and admixture. In contrast to previous studies reporting low knob numbers in NF or even knob-less NF lines, the European flints contained similar knob numbers to the US dents[41]. It is unclear whether this is a characteristic of the NF ancestral lines migrated to Europe, reflects the breeding history in European flints or represents a technical advantage given the progress in methodology to detect weaker signals. Consistent with fiber-FISH studies, we also detected numerous locations of the 180-bp knob repeat unit throughout the chromosome assemblies[42]. The potential contribution of these widespread knob repeats on gene density, expression levels and rare deleterious alleles requires future studies and adds further complexity to known effects of large-scale knobs in maize on recombination rates, meiotic drive and phenotypic traits such as flowering time[26,43,44].

WGAs enabled us to identify long runs of pairwise and higher order combinations of genomic regions with identical haplotypes. Size and distribution of these haplotype blocks reflect complexity of historic recombination, intercrossing and breeding history over the last centuries[3,5,7,19,45–47]. Similar to the gene and repeat comparisons, the SNP runs highlight the separation of flints and dents of this study but also illustrate and delineate candidate regions of common ancestry in the US Corn Belt and European flint lines. We identified a set of regions that differentiated the studied Corn Belt dents from the European germplasm. Intriguingly, genes

differentially expressed between two groups of maize varieties that have been selected by their genotypic similarity to these regions comprised a nearly complete starch biosynthesis pathway and several genes involved in the establishment and organization of the endosperm texture. Hence, besides the detection of selective sweeps and other genomic signatures, it might be equally informative to survey *trans*-effects of such genomic regions.

Given the high dynamics and tessellation of the maize genome, we estimate that many high-quality genome sequences will be required for improving breeding efficacy and understanding maize biology. The four European flint lines of this study close an important gap in our knowledge and complement the maize pangenome.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-020-0671-9.

Received: 21 June 2019; Accepted: 25 June 2020;
Published online: 27 July 2020

References

1. Doebley, J. F., Goodman, O. M. & Stuber, C. W. Exceptional genetic divergence of northern flint corn. *Am. J. Bot.* **73**, 64–69 (1986).
2. Liu, K. et al. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics* **165**, 2117–2128 (2003).
3. Chia, J. M. et al. Maize HapMap2 identifies extant variation from a genome in flux. *Nat. Genet* **44**, 803–807 (2012).
4. Duvick, D. N. The contribution of breeding to yield advances in maize (*Zea mays* L.). *Adv. Agron.* **86**, 83–145 (2005).
5. Anderson, E. & Brown, W. L. *The History of the Common Maize Varieties of the United States Corn Belt* (Univ. of California Press, 1952).
6. Doebley, J. et al. The origin of cornbelt maize: the isozyme evidence. *Econ. Bot.* **42**, 120–131 (1988).
7. van Heerwaarden, J., Hufford, M. B. & Ross-Ibarra, J. Historical genomics of North American maize. *Proc. Natl Acad. Sci. USA* **109**, 12420–12425 (2012).
8. Tenaillon, M. I. & Charcosset, A. A European perspective on maize history. *C. R. Biol.* **334**, 221–228 (2011).
9. Jiao, Y. et al. Improved maize reference genome with single-molecule technologies. *Nature* **546**, 524 (2017).
10. Schnable, P. S. et al. The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112–1115 (2009).
11. Lai, J. et al. Gene loss and movement in the maize genome. *Genome Res.* **14**, 1924–1931 (2004).
12. Springer, N. M. et al. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet.* **5**, e1000734 (2009).
13. Swanson-Wagner, R. A. et al. Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. *Genome Res.* **20**, 1689–1699 (2010).
14. Wang, Q. & Dooner, H. K. Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus. *Proc. Natl Acad. Sci. USA* **103**, 17644–17649 (2006).
15. Hirsch, C. N. et al. Draft assembly of elite inbred line PH207 provides insights into genomic and transcriptome diversity in maize. *Plant Cell* **28**, 2700–2714 (2016).
16. Springer, N. M. et al. The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nat. Genet.* **50**, 1282–1288 (2018).
17. Sun, S. et al. Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nat. Genet.* **50**, 1289–1295 (2018).
18. Yang, N. et al. Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nat. Genet.* **51**, 1052–1059 (2019).
19. Unterseer, S. et al. A comprehensive study of the genomic differentiation between temperate Dent and Flint maize. *Genome Biol.* **17**, 137 (2016).
20. Darracq, A. et al. Sequence analysis of European maize inbred line F2 provides new insights into molecular and chromosomal characteristics of presence/absence variants. *BMC Genomics* **19**, 119 (2018).
21. Messmer, M. M. et al. RFLP analyses of early-maturing European maize germ plasm: I. Genetic diversity among flint and dent inbreds. *Theor. Appl. Genet.* **83**, 1003–1012 (1992).
22. International Wheat Genome Sequencing Consortium et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**, 7191 (2018).
23. Yang, L. A. et al. SQUAT: a Sequencing Quality Assessment Tool for data quality assessments of genome assemblies. *BMC Genomics* **19**, 238 (2019).
24. Simão, F. A. et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
25. Ou, S., Chen, J. & Jiang, N. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* **46**, e126 (2018).
26. Ghaffari, R. et al. Maize chromosomal knobs are located in gene-dense areas and suppress local recombination. *Chromosoma* **122**, 67–75 (2013).
27. Stack, S. M. et al. Meiotic crossing over in maize knob heterochromatin. *Genetics* **205**, 1101–1112 (2017).
28. Ananiev, E. V., Phillips, R. L. & Rines, H. W. Complex structure of knobs and centromeric regions in maize chromosomes. *Tsitol. Genet.* **34**, 11–15 (2000).
29. Matsuda, A. et al. Highly condensed chromatins are formed adjacent to subtelomeric and decondensed silent chromatin in fission yeast. *Nat. Commun.* **6**, 7753 (2015).
30. Kremling, K. A. G. et al. Dysregulation of expression correlates with rare-allele burden and fitness loss in maize. *Nature* **555**, 520 (2018).
31. Unterseer, S. et al. A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array. *BMC Genomics* **15**, 823 (2014).
32. Salvi, S. et al. Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc. Natl Acad. Sci. USA* **104**, 11376–11381 (2007).
33. Ducrocq, S. et al. Key impact of *Vgt1* on flowering time adaptation in maize: evidence from association mapping and ecogeographical information. *Genetics* **178**, 2433–2437 (2008).
34. Bukowski, R. et al. Construction of the third-generation *Zea mays* haplotype map. *Gigascience* **7**, 1–12 (2018).
35. Cook, J. P. et al. Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiol.* **158**, 824–834 (2012).
36. James, M. G., Robertson, D. S. & Myers, A. M. Characterization of the maize gene sugary1, a determinant of starch composition in kernels. *Plant Cell* **7**, 417–429 (1995).
37. Yang, J., Ji, C. & Wu, Y. Divergent transactivation of maize storage protein zein genes by the transcription factors Opaque2 and OHPs. *Genetics* **204**, 581–591 (2016).
38. Yao, D. et al. Maize *opaque10* encodes a cereal-specific protein that is essential for the proper distribution of zeins in endosperm protein bodies. *PLoS Genet.* **12**, e1006270 (2016).
39. Yi, G. et al. The naked endosperm genes encode duplicate INDETERMINATE domain transcription factors required for maize endosperm cell patterning and differentiation. *Plant Physiol.* **167**, 443–456 (2015).
40. Zhang, S., Zhan, J. & Yadegari, R. Maize opaque mutants are no longer so opaque. *Plant Reprod.* **31**, 319–326 (2018).
41. Brown, W. L. Numbers and distribution of chromosome knobs in United States maize. *Genetics* **34**, 524–536 (1949).
42. Adawy, S. S., Stupar, R. M. & Jiang, J. Fluorescence in situ hybridization analysis reveals multiple loci of knob-associated DNA elements in one-knob and knobless maize lines. *J. Histochem. Cytochem.* **52**, 1113–1116 (2004).
43. Jian, Y. et al. Maize (*Zea mays* L.) genome size indicated by 180-bp knob abundance is associated with flowering time. *Sci. Rep.* **7**, 5954 (2017).
44. Kanizay, L. B. et al. Intragenomic conflict between the two major knob repeats of maize. *Genetics* **194**, 81–89 (2013).
45. Hufford, M. B. et al. Comparative population genomics of maize domestication and improvement. *Nat. Genet.* **44**, 808 (2012).
46. Lai, J. et al. Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat. Genet.* **42**, 1027 (2010).
47. Reif, J. C. et al. Trends in genetic diversity among European maize cultivars and their parental components during the past 50 years. *Theor. Appl. Genet.* **111**, 838–845 (2005).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020

Methods

Plant material and genome assembly. The four flint inbred lines were chosen to represent landraces from different European ancestry. While PE0075 (a doubled-haploid line derived from the landrace Petkuser Ferdinand Rot) and DK105 (derived from Gelber Badischer Landmais) can be classified as a representatives of the NF, EP1 and F7 represent Pyrenean–Galician ancestry and were derived from populations Lizargarate and Lacaune, respectively[48]. Contigs, scaffolds and pseudochromosomes were assembled de novo using the DeNovoMAGIC 2.0 technology for lines EP1 and F7, and DeNovoMAGIC 3.0 for lines DK105 and PE0075 (Energin R Technologies 2009), respectively (Supplementary Table 2) using the B73 (v.4) reference genome to assist anchoring and orienting scaffolds and contigs to pseudochromosomes. The application of this toolset has been described previously for several plant genome assemblies[15,16,49]. Briefly, bulked leaf tissue from 15–24 seedlings was harvested, immediately ground in liquid nitrogen and genomic DNA was then isolated using a modified CTAB protocol[50]. Illumina sequencing of paired-end and mate-pair libraries was performed as specified in Supplementary Table 1. Remappings of the paired-end and mate-pair libraries were used to estimate gap sizes in the assemblies (Supplementary Fig. 10).

A cross of parental lines PH207 × EP1 was used to establish a high-density genetic map from an F_2 mapping population. Parents and 192 F_2 plants were analyzed using the Affymetrix Axiom Maize Genotyping Array and data were processed as described in ref. [31], which resulted in 174,616 markers. The markers were clustered into groups showing identical segregation patterns using POPSEQ[51]. A first genetic map was calculated with 9,404 binmap markers with the R package ASMap v.1.0-2 (ref. [52]) using the function mstmap with the following parameters: pop.type="RIL2", dist.fun="kosambi", objective.fun="COUNT", p.value=1e-22, noMap.dist=15, noMap.size=2, miss.thresh=0.00. Linkage groups were assigned to the ten maize chromosomes on the basis of previously mapped markers. Unlinked small groups with only a few markers were discarded. Ten F_2 plants exhibited very high numbers of crossovers and were excluded from further analyses. In multiple rounds of mapping using the same parameters as stated above, markers with highly distorted segregation or which led to double-crossovers were identified using the function statMark in the ASMap R package and discarded before final map construction. The final genetic linkage map contained 8,869 markers. All markers from the initial dataset which had a Hamming distance of 0 with one of the mapped binmap markers were inserted into the map, resulting in a genetic map with 174,071 markers.

Gene annotation and consolidation. We predicted protein-coding structures for F7 and EP1 as consensus models using an approach as previously described[53,54]. Briefly, consensus gene models are based on transcriptome evidences from F7 and EP1 RNA-seq data generated within this study (see RNA preparation and Expression analysis) and protein homologies of known monocotyledonous protein sequences (maize B73 and PH207, *Sorghum bicolor* v.3.1., *Oryza sativa* v.7 MSU and *Brachypodium distachyon* v.3.1)[53,55,56]. Evidences were mapped onto the genome assembly applying GenomeThreader[57] using a minimum alignment coverage of 50% and seed sizes 7–10 for protein and 18 for nucleic acid matches. Before mapping, transcript data were assembled using Trinity with default parameters and Bridger with *K*-mer sizes 25 and 29 bp (refs. [58,59]). Subsequently, these sequences were combined by the evidential gene pipeline (http://arthropods.eugenes.org/EvidentialGene) to obtain the final EP1 and F7 transcriptome assemblies. Initial consensus gene models were subsequently consolidated as described in the Supplementary Notes to derive final gene sets for the four flints and two dent lines.

FISH and karyotyping. Maize chromosomes were prepared from root meristems of 2-day-old seedlings. Roots were cut and treated with 2 mM of 8-hydroxyquinoline solution for 3.5 h. Fixation was performed overnight at room temperature in 3:1 (ethanol:acetic acid) fixative. Slides were prepared according to ref. [60] and in situ hybridization was performed as described in ref. [61]. Oligonucleotide probes for karyotyping and identification of knob repeats were chosen from ref. [62].

Images were taken using an Olympus BX61 microscope equipped with an ORCA-ER CCD camera (Hamamatsu). All images were acquired in gray scale and pseudocoloured with Adobe Photoshop CS5 (Adobe Systems). Karyotyping was done according to ref. [62] using the maize line B73 as a reference.

Repeat analysis. To obtain equivalent transposon and tandem repeat data for comparative analyses we annotated all six lines with the same annotation workflows. An homology search against the Panicoidae section of the PGSB transposon library[63] resulted in a basal transposon annotation. The REdat_9.8_Panicoideae used contains publicly available Panicoideae transposons templates as well as de novo detected full-length LTR-retrotranspons from maize (12,510 elements from B73 (v.2)) and sorghum (3,368 elements). The program vmatch (http://www.vmatch.de) was used as a fast and efficient matching tool well suited for large and highly repetitive genomes under the following parameter setup: identity ≥70%, minimal hit length 75 bp, seedlength 12 bp (exact commandline: -d -p -l 75 -identity 70 -seedlength 12 -exdrop 5). The vmatch output was filtered for redundant hits via a priority-based approach, which assigns higher scoring matches first and either shortens (<90% coverage and ≥50-bp rest length) or removes lower scoring overlaps. The resulting annotation is free of overlaps. Elements that have been interrupted by other transposon insertions (nesting) are not defragmented into a higher order instance (such as exons belonging to one gene).

Full-length LTR-retrotransposons (fl-LTR) where identified with LTRharvest[64] using the following parameters: overlaps best -seed 30 -minlenltr 100 -maxlenltr 2000 -mindistltr 3000 -maxdistltr 25000 -similar 85 -mintsd 4 -maxtsd 20 -motif tgca -motifmis 1 -vic 60 -xdrop 5 -mat 2 -mis -2 -ins -3 -del -3. All candidates from the LTRharvest output were subsequently annotated with PfamA domains using hmmer3 (http://hmmer.org) and stringently filtered for false positives by several criteria, the main ones being the presence of at least one typical retrotransposon domain (for example, reverse transcriptase (RT), RNase H (RH), integrase (INT), protease (PR) and so on) and a tandem repeat content <25%. The inner domain order served as a criterion for the classification into the Gypsy (RT–RH–INT) or Copia (INT–RT–RH) superfamily abbreviated as RLG and RLC. Elements missing either INT or RT were classified as RLX. The insertion age of each full-length LTR-retrotransposon was estimated on the basis of the accumulated divergence between its 5′ and 3′ LTRs and a random mutation rate of 1.3×10^{-8} (ref. [65]).

Tandem repeats were identified with the TandemRepeatFinder under default parameters[66] and subjected to an overlap removal as described above, prioritizing longer and higher scoring elements. *K*-mer frequencies were calculated with Tallymer[67].

Syntentic fl-LTRs where identified by sequence clustering (vmatch dbcluster, 98% identity and 98% coverage) of TE junctions from the de novo annotated fl-LTR locations from each line. The junctions consisted of 2 × 100-bp sequence signatures spanning the upstream and downstream insertion sites with each 50 bp inside and 50 bp outside of the TE element. The LTR assembly index was calculated using the LTR_retriever program suite[68] as suggested on the Github pages of LTR_retriever (github.com/oushujun/LTR_retriever). LTR_FINDER (ref. [69]) was run via the wrapper LTR_FINDER_parallel[70] using the following parameters: -w 2 -C -D 15000 -d 1000 -L 7000 -l 100 -p 20 -M 0.85 (20 threads and 5-Mb batches). LTRHarvest[64] was run with: -minlenltr 100 -maxlenltr 7000 -mintsd 4 -maxtsd 6 -motif TGCA -motifmis 1 -similar 85 -vic 10 -seed 20.

Whole-genome alignments. Block alignments of high identity were generated for all pairwise combinations of the six maize lines using the MUMMER v.3 suite[71]. Initial alignments were computed using nucmer with a minimal cluster size of 250 bp and a seed size of 20 bp. Results were piped through the delta-filter tool selecting the best one-to-one blocks (global option -1) with a minimal size of 500 bp (-l 500). Despite this filter step, many (~41,000–81,000) putatively translocated blocks aligned small regions between different chromosomes or unanchored scaffolds. These blocks frequently overlapped 'regular' blocks linking same orthologous chromosomes and showed significantly lower sequence identities in comparison to their respective regular blocks with which they overlapped. Hence, they likely represented paralogous alignments triggered by tandem repeats or the absence of the truly orthologous sequences in one of the lines. To derive the final set of SAB, we only included candidate interchromosomal translocations with a minimal sequence identity of 99% and a maximum overlap of 10 bp to its adjacent alignments. To gain an overview of the contiguity between the six maize genomes, SABs were connected to MABs if SABs were directly adjacent in both genomes with a consistent orientation (Supplementary Fig. 11).

To identify and delineate genomic regions for each line that aligned (1) to none of the other lines (designated as 'unaligned'), (2) to all lines of the same group (either flint or dent maize) and to none of the other group (aka 'group-specific) or (3) to all five other lines ('core6'), we superimposed SABs and MABs of one line on its genome coordinates by recording from which group and how many times a single genomic base participated in an alignment. Thereby, we classified each base and concatenated adjacent bases of the same type (1–3) to derive the unaligned, group-specific and core genomic part. Reversing this approach, identified core regions of EP1 were reprojected onto the pairwise alignments to decode the sequence and coordinates of matching core elements in the other five lines. On the basis of these coordinates, sequences for each core block were extracted and aligned applying the Fast Sequence Aligner[72] with default parameters.

Determination of paired and higher order haplotypes. SNPs were directly determined from multiple sequence alignments of the WGA core blocks (omitting insertions/deletions). Their position in each of the six genomes was derived from the position in the alignment and the known genomic start- and end-positions of the respective core block. Given the observed SNP frequencies, and applying randomization studies and run-of-head statistics, the expected maximal run lengths of identical SNPs between two lines range from 21 to 30 consecutive SNPs. We selected genomic windows of ≥40 identical SNPs as seeds to identify shared haplotypes of likely common ancestry and used those as seeds for a greedy extension to delineate genomic segments with ≥98% sequence identity between all 15 pairwise line combinations. To deduce higher order haplotypes, we transformed the 15 pairwise runs-of-identities to binary SNP patterns and surveyed seeds with ≥40 consecutive identical binary configurations. An iterative greedy algorithm linking the seed with adjacent upstream and downstream runs extended such seeds if the run exhibited a haplotype identical to the seed and less than three

nonmatching SNPs were observed between the current end of the extension and the candidate run.

RNA preparation. Different tissues of the European maize (*Z. mays*) inbred lines EP1 and F7 were sampled according to refs. [73,74] and subsequently subjected to transcriptome sequencing (Supplementary Table 5). Samples 1–3 were collected from seeds imbibed (whole seed, sample 1) or germinated (primary root, sample 2; coleoptile, sample 3) in paper rolls[75] in a 16 h light (28 °C) and 8 h dark (21 °C) regime of a growth cabinet (Conviron CMP6010, http://www.conviron.com). Tissue samples 4–24 were taken from plants grown in a climate chamber either in small pots (14 cm top diameter, 8 cm height, 0.25 l volume for samples 4–17) or in big pots (28 cm top diameter, 21 cm height, 10 l volume for samples 18–24) containing soil substrate type ED 73 (https://www.meyer-shop.com). Growth conditions in the climate chamber were 16 h light (28 °C) and 8 h dark (21 °C). Two individual plants were collected for each of the 24 tissues. Harvested plant material was immediately frozen in liquid nitrogen and stored at −80 °C until RNA extraction. For total RNA extraction, the 24 samples per genotype were separately ground in liquid nitrogen. Subsequently, one spatula of each of the 24 samples per genotype was mixed to generate one pool per inbred line EP1 and F7, respectively. Total RNA was extracted from each of the two pools using the Qiagen RNeasy mini kit according to the manufacturer's protocol (Qiagen, https://www.qiagen.com), including on-column DNA digestion. RNA quality was determined by agarose gel electrophoresis and by a Bioanalyzer using an Agilent RNA 6000 Nano Chip (Agilent Technologies, https://www.agilent.com). Both RNA samples were of excellent quality with RNA integrity number values[76] between 9.8 and 10. In total, 9 µg of total RNA per pool were used for RNA-seq at Novogene (Novogene). After mRNA enrichment using oligo(dT) beads, sequencing libraries were constructed with the NEBNext Ultra RNA Library Prep Kit for Illumina (New England Biolabs) and Illumina PE (150 bp × 2) sequencing was performed on a HiSeq4000 by Novogene. The number of raw reads and quality-filtered reads are shown in Supplementary Table 12. Reads containing adapters, reads containing >10% undefined bases and reads containing >50% low quality bases (Qscore 5) were removed.

Expression analysis and analysis of DEGs. Genomic variants overlapping the flint/dent specific haplotype identified in this study were extracted from the Maize Hapmap v.3.2.1 panel[34] for lines of the 282 inbred maize association panel[77]. Selected variants were subjected to phylogenetic analysis using FastTree (v.2.1.5 SSE3)[78] with default parameters and the resulting phylogenetic tree was visualized using iTOL[79] (Supplementary Fig. 18). Two nonoverlapping subtrees were selected such that three of the lines presented in this study (B73, EP1 and F7) are contained, together with other topologically close lines. All lines in the subtree containing B73 were assigned to group one; all those in the subtree containing both EP1 and F7 were assigned to group two.

For the resulting subset of 41 lines (Supplementary Fig. 9) samples from seven different tissues (germinating shoot and root, the tip and base of leaves as well as leaf samples under light and dark conditions and kernels) were downloaded from SRA (PRJNA383416)[30]. Read preprocessing was performed as described in the Lexogen QuantSeq user guide (https://www.lexogen.com/wp-content/uploads/2018/10/015UG108V0201-QuantSeq-Data-Analysis-Pipeline.pdf). Trimmed reads were mapped to the maize reference sequences B73 (v.4), EP1 and F7 using STAR aligner v.2.5.1a (ref. [80]) applying parameters outFilterMultimapNmax=10, outFilterMismatchNoverLmax=0.04,outFilterIntronMotifs=RemoveNoncanonicalUnannotated To obtain nonunique gene-level counts from the mapping files, HTSeq (v.0.11.2) with the 'nonunique all'-method was used[81]. Normalization of read counts was performed by library sequence depth using the R package DESeq2 (v.1.23.3)[82]. For differential gene expression analysis with DESeq2, lines classified into either the flint- or dent-like group were treated as biological replicates. In this study, we report only genes for which all three orthologs of RNA-seq read mappings to B73, EP1 and F7 were significant at Benjamini–Hochberg adjusted $P \leq 0.05$.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The de novo assembled genomes and raw reads were released in collaboration with the Maize Genetics and Genomics Database MaizeGDB[18] and NCBI BioProjects PRJNA360923 (F7, DK105 and PE0075) and PRJNA360920 (EP1). Additionally, RNA-seq raw data were also deposited under these NCBI accessions.

References

48. Mayer, M. et al. Is there an optimum level of diversity in utilization of genetic resources? *Theor. Appl. Genet.* **130**, 2283–2295 (2017).
49. Avni, R. et al. Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* **357**, 93–97 (2017).
50. Saghai-Maroof, M. A. et al. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl Acad. Sci. USA* **81**, 8014–8018 (1984).
51. Mascher, M. et al. Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J.* **76**, 718–727 (2013).
52. Taylor, J. & Butler, D. R package ASMap: efficient genetic linkage map construction and diagnosis. *J. Stat. Softw.* **79**, 1–29 (2017).
53. Paterson, A. H. et al. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**, 551–556 (2009).
54. Wang, W. et al. The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. *Nat. Commun.* **5**, 3311 (2014).
55. Kawahara, Y. et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**, 4 (2013).
56. The International Brachypodium Initiative Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**, 763–768 (2010).
57. Gremme, G. et al. Engineering a software tool for gene structure prediction in higher organisms. *Inf. Softw. Technol.* **47**, 965–978 (2005).
58. Chang, Z. et al. Bridger: a new framework for de novo transcriptome assembly using RNA-seq data. *Genome Biol.* **16**, 30 (2015).
59. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
60. Aliyeva-Schnorr, L., Ma, L. & Houben, A. A fast air-dry dropping chromosome preparation method suitable for fish in plants. *J. Vis. Exp.* **106**, 53470 (2015).
61. Badaeva, E. D. et al. in *Fluorescence In Situ Hybridization (FISH): Application Guide* (ed. Liehr, T.) 477–494 (Springer, 2017).
62. Zhu, M. et al. A simple and efficient nondenaturing FISH method for maize chromosome differentiation using single-strand oligonucleotide probes. *Genome* **60**, 657–664 (2017).
63. Spannagl, M. et al. PGSB PlantsDB: updates to the database framework for comparative plant genome research. *Nucleic Acids Res.* **44**, D1141–D1147 (2016).
64. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18 (2008).
65. SanMiguel, P. et al. The paleontology of intergene retrotransposons of maize. *Nat. Genet.* **20**, 43–45 (1998).
66. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
67. Kurtz, S. et al. A new method to compute K-mer frequencies and its application to annotate large repetitive plant genomes. *BMC Genomics* **9**, 517 (2008).
68. Ou, S. & Jiang, N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* **176**, 1410–1422 (2018).
69. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
70. Ou, S. & Jiang, N. LTR_FINDER_parallel: parallelization of LTR_FINDER enabling rapid identification of long terminal repeat retrotransposons. *Mob. DNA* **10**, 48 (2019).
71. Kurtz, S. et al. Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
72. Bradley, R. K. et al. Fast statistical alignment. *PLoS Comput. Biol.* **5**, e1000392 (2009).
73. Downs, G. S. et al. A developmental transcriptional network for maize defines coexpression modules. *Plant Physiol.* **161**, 1830–1843 (2013).
74. Sekhon, R. S. et al. Genome-wide atlas of transcription during maize development. *Plant J.* **66**, 553–563 (2011).
75. Hetz, W. et al. Isolation and characterization of rtcs, a maize mutant deficient in the formation of nodal roots. *Plant J.* **10**, 845–857 (1996).
76. Schroeder, A. et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.* **7**, 3 (2006).
77. Flint-Garcia, S. A. et al. Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J.* **44**, 1054–1064 (2005).
78. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
79. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
80. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
81. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
82. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

Acknowledgements

This study was funded by the Bavarian State Ministry of the Environment and Consumer Protection (grant no. TGC01GCUFuE69741; Project BayKlimaFit, http://www.bayklimafit.de/) and the German Federal Ministry of Education and Research (BMBF) within the funding initiative 'Plant Breeding Research for the Bioeconomy' (grant no. 031B0195A; Project MAZE, http://www.europeanmaize.net). In addition K.F.X.M.

acknowledges support from the BMBF within the project de.NBI grant no. 031A536B. We thank N. Opitz (University of Bonn) for helping to collect maize tissue for RNA-seq.

Author contributions

C.-C.S., K.F.X.M., M.O. and G.H. designed the study. K.F.X.M. and G.H. wrote the manuscript. E.B. and C.-C.S. supervised the genomic sequencing. E.B., C.U. and A.N. generated the F_2 population and constructed the genetic map. G.H. analyzed genes and whole-genome alignments. H.G. performed the repeat analysis. A.H. and A.R. contributed the FISH analysis. C.M. and F.H. prepared and sequenced the RNA-seq samples. N.K., M.A.S. and I.F. conducted the DEG expression analysis and the phylogenetic analysis. A.S. and M.A.S. provided bioinformatics support. All authors critically read and edited the manuscript.

Competing interests

The authors declare no competing interests.

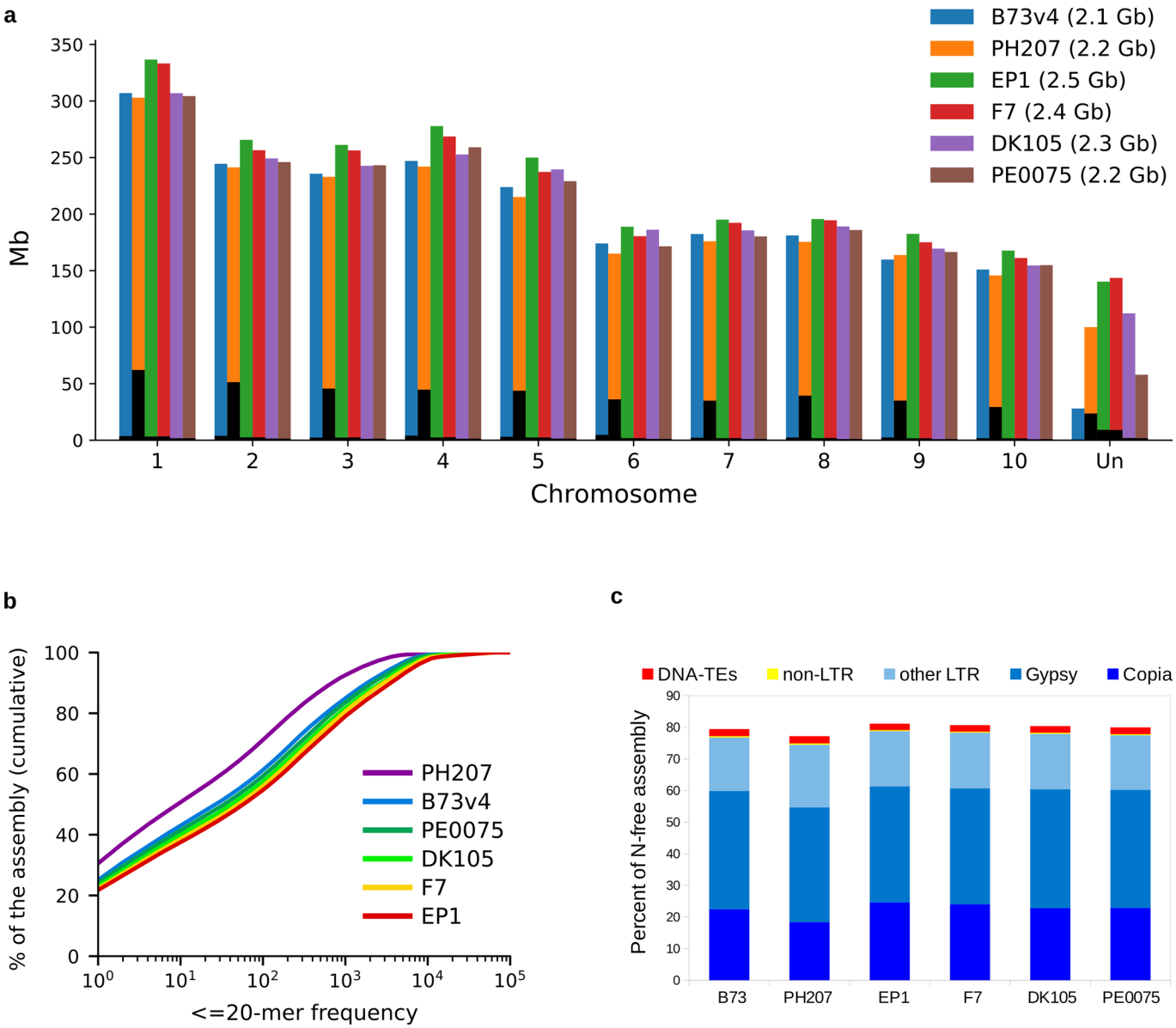
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-020-0671-9.

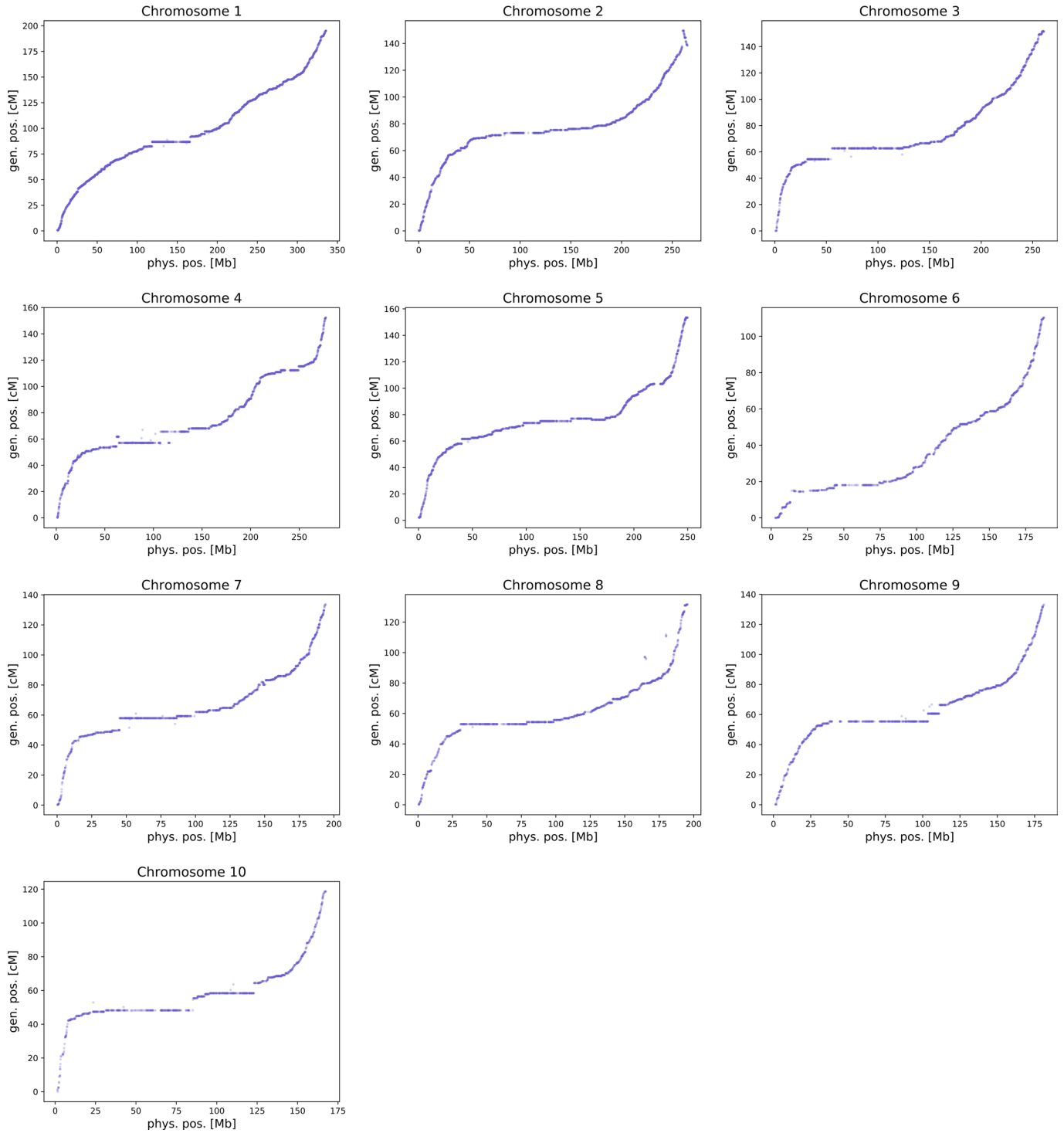
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-020-0671-9.

Correspondence and requests for materials should be addressed to C.-C.S. or K.F.X.M.

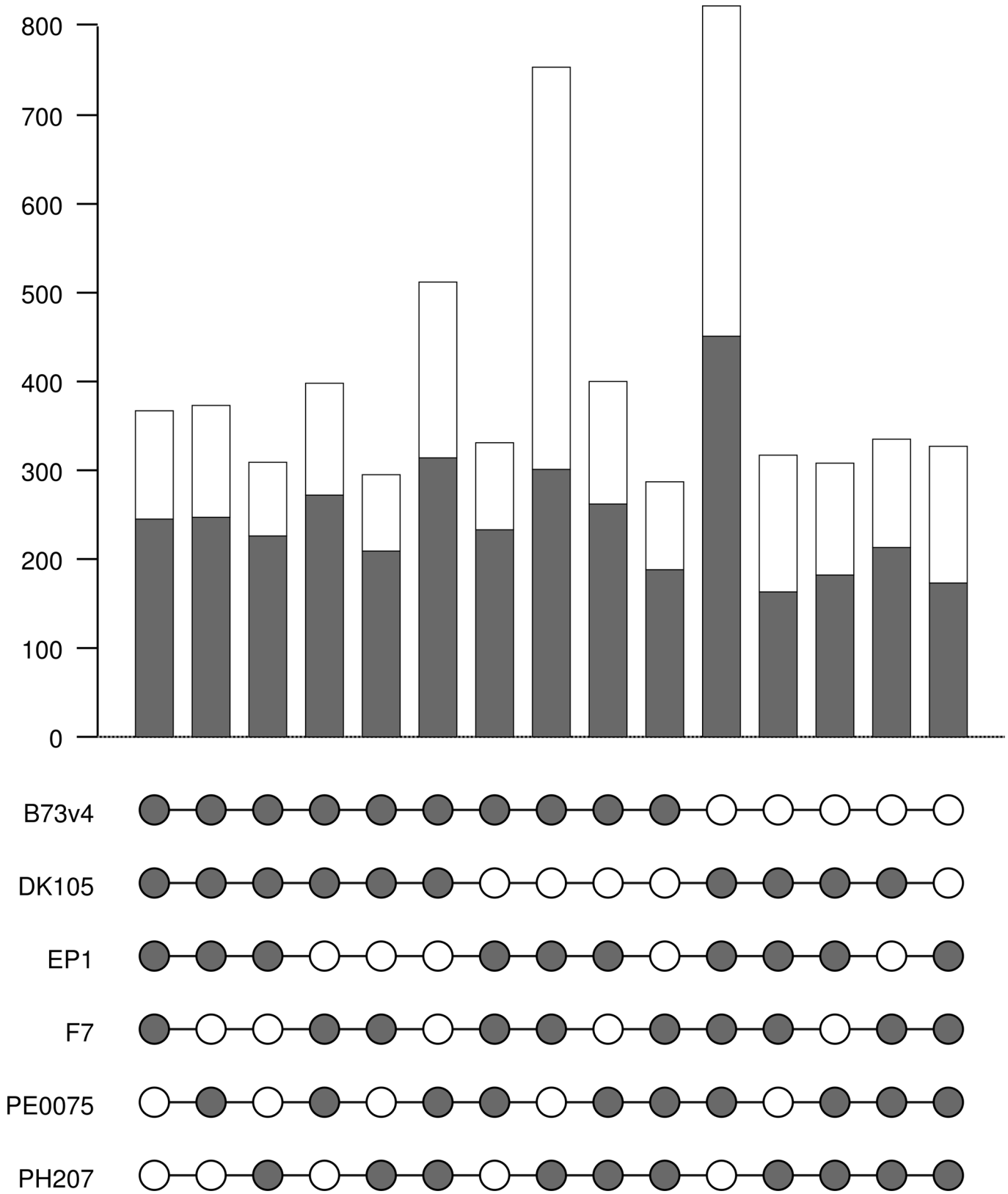
Reprints and permissions information is available at www.nature.com/reprints.



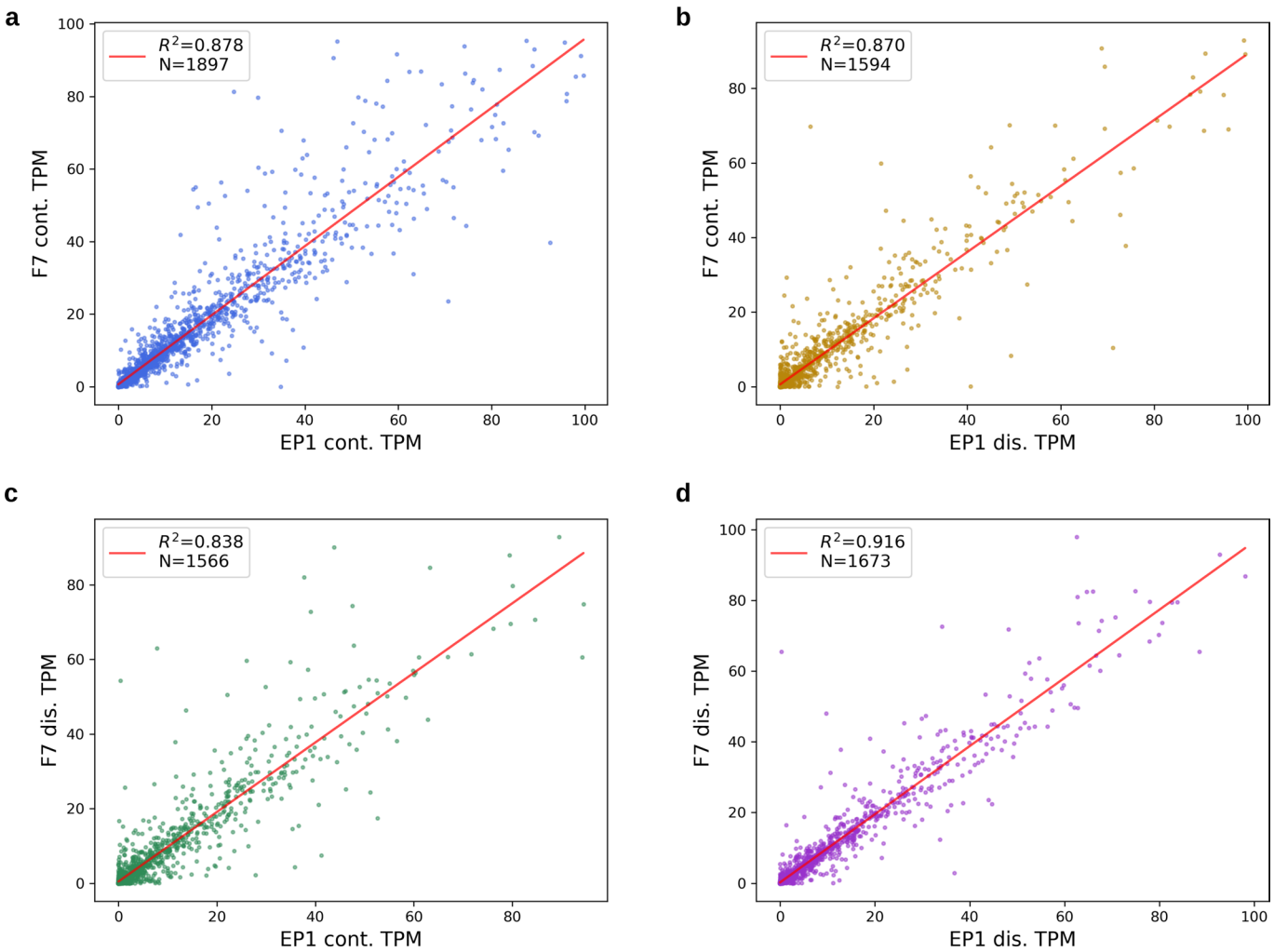
Extended Data Fig. 1 | Assembly sizes and overall repeat content of 8 maize lines. a, Chromosome sizes and N-content (black bars), **b**, Basic repetitivity in form of 20mer frequencies. Cumulative plot, lower curves represent higher repetitivity e.g. 50% of the PH207 assembly consists of 20mers occurring <=10 times, for the other lines this value is only around 40%. **c**, Transposon content.



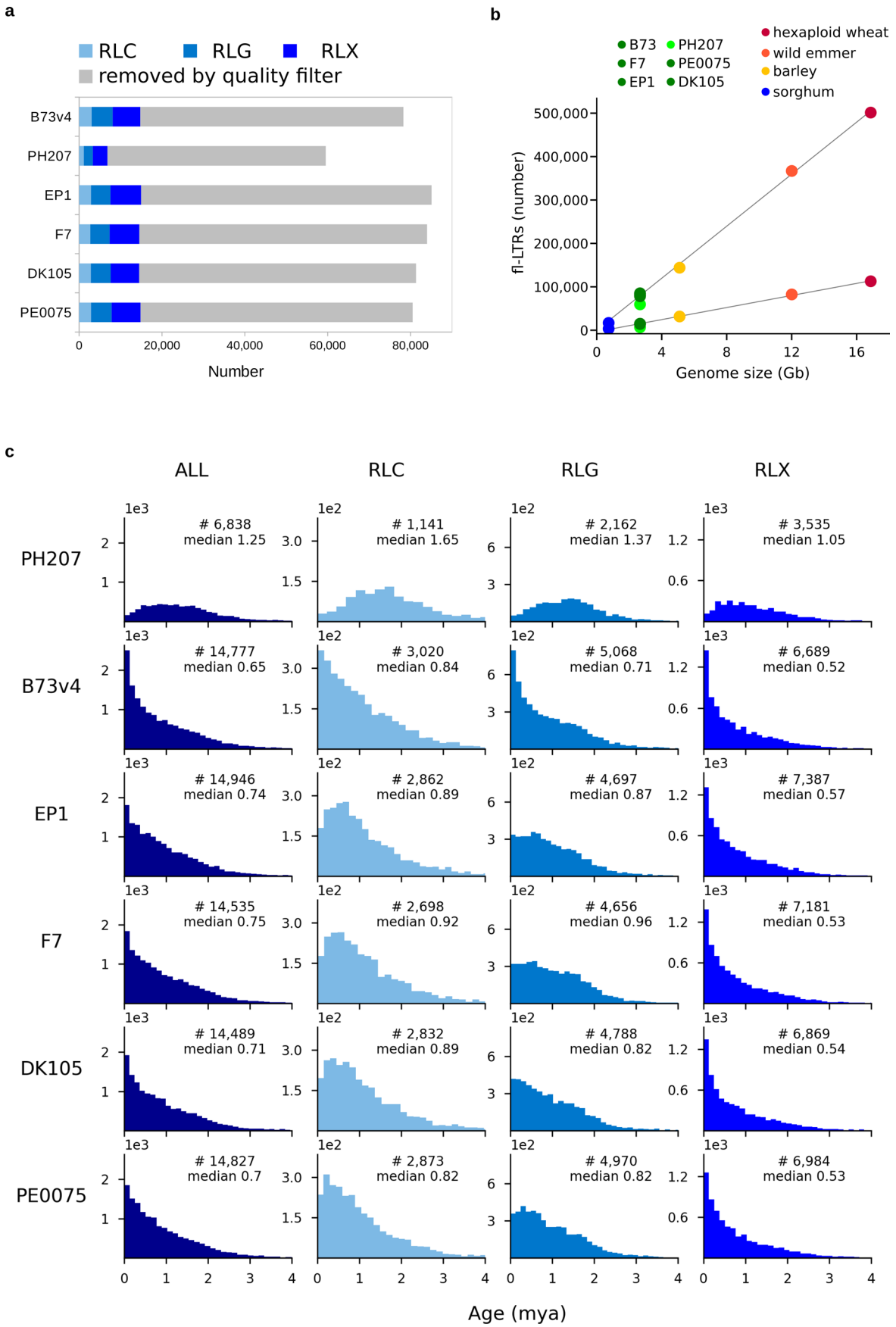
Extended Data Fig. 2 | Genetic versus physical distances for all ten chromosomes of the EP1xPH207 cross. X-axis: physical position on the respective EP1 chromosome in [Mb], Y-axis: genetic position in centiMorgan [cM].



Extended Data Fig. 3 | Group-specific syntelogs for all (4,2)-combinations of the six lines and their syntelog set. We divided the six maize lines of this study into two groups of 4 (gray circles) and 2 (white circles) lines, respectively. For all 15 combinations, the number of group-specific syntelogs for each group was determined under the prior of the identified orthologous gene set. The germplasm-specific separation of flint (EP1, F7, DK105 and PE0075) versus dent (B73 version 4 and PH207) genotypes showed the largest distinction between the two groups.

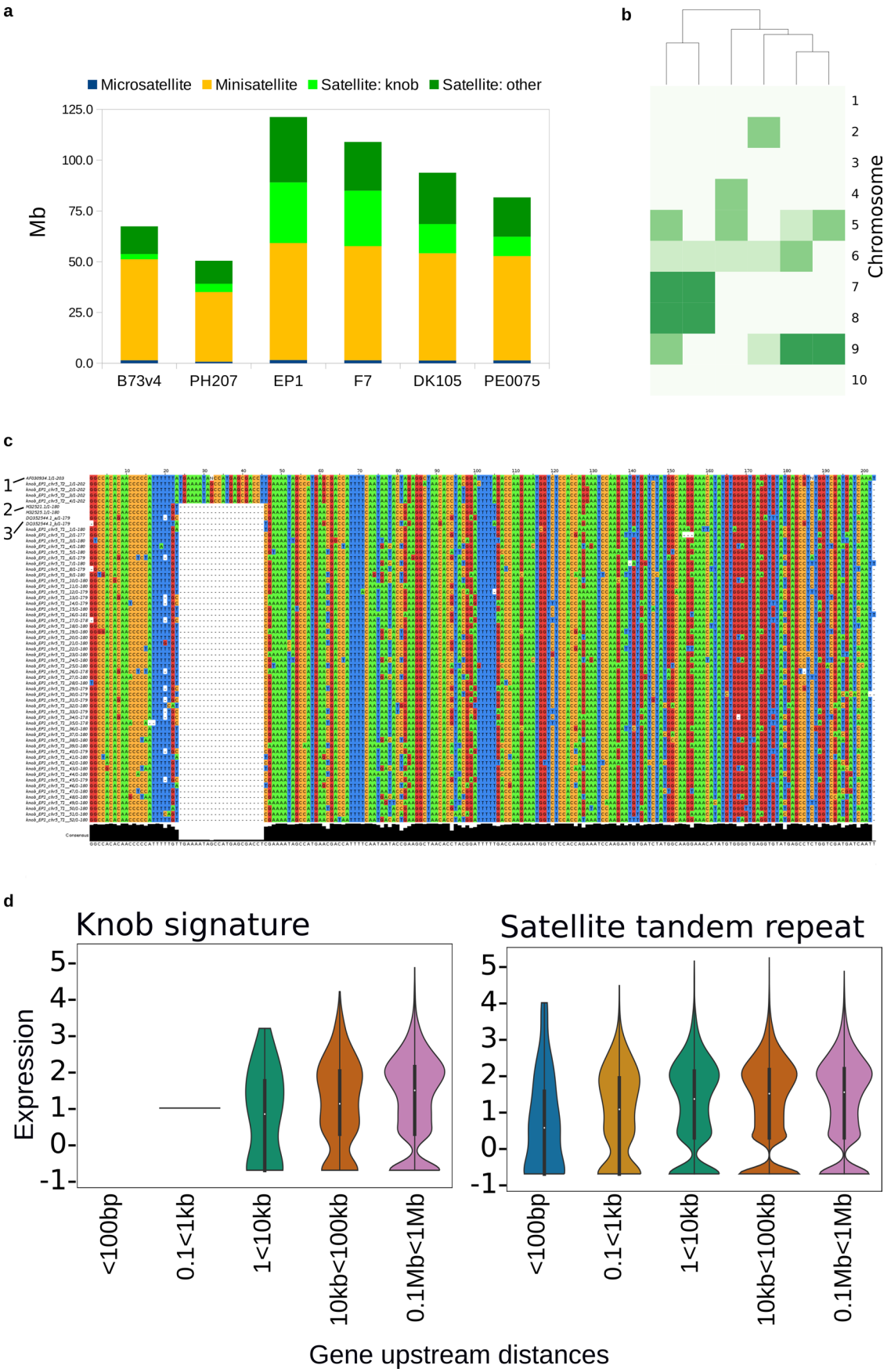


Extended Data Fig. 4 | Expression levels of orthologs with contiguous and disrupted open reading frame (ORF). Consolidated mappings discovered several thousand genes for which the entire cross-mapping of the top-scoring consolidation model had no contiguous but a disrupted ORF. These genes exhibit highly similar expression levels to their nondisrupted orthologs illustrated here for lines EP1 and F7. **a**, shows the control correlation and linear regression (red line) of a random sample of F7 and EP1 orthologous pairs for which both genes have a contiguous ORF. To obtain comparable sample sizes and expression ranges, random sampling was restricted to gene pairs with a maximal TPM ≤ 100. **b**, and **c**, display expression correlations between gene pairs with exactly one disrupted and one contiguous ORF, while in subfigure **d**, both orthologous partners contain a disrupted ORF. N: number of pairs, r2: coefficient of determination.



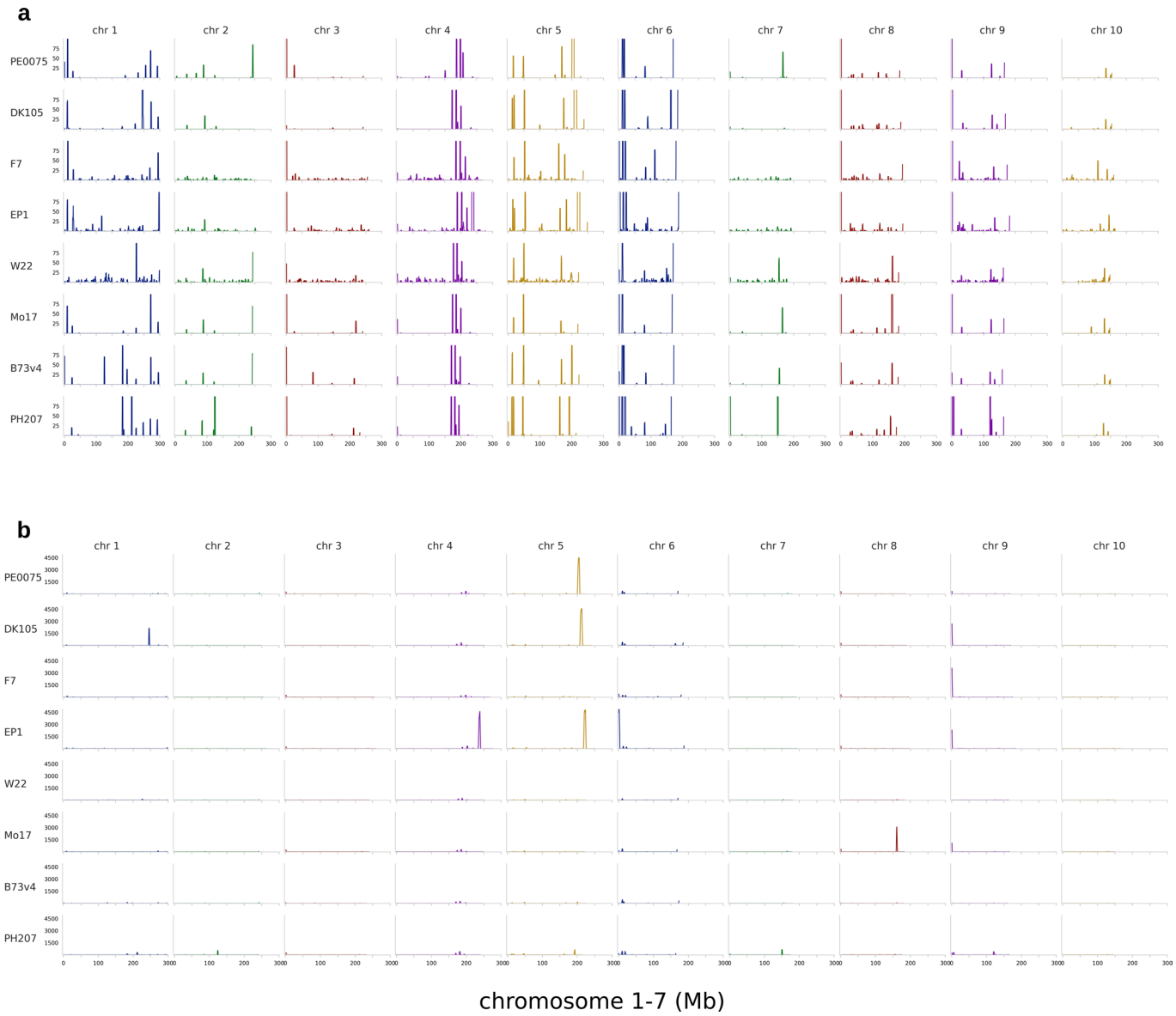
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Full-length LTR-retrotransposon detection in different maize lines. a, *De novo* detection and stringent quality filtering of fl-LTR elements. **b**, Retrieved fl-LTR numbers as proxy for assembly quality. Upper data points give candidate fl-LTRs, lower data points give filtered, high-quality fl-LTR numbers. Five of the six assemblies contain the expected number of fl-LTRs. The by 50% reduced numbers of high-quality elements in PH207 reflect an older not optimized assembly approach. **c**, Insertion age distribution of fl-LTRs. The four NRGene assemblies have very similar age distribution patterns. The B73 assembly has a higher proportion of very young Copia and Gypsy elements, which can probably be attributed to technical assembly differences. In PH207, the young elements are completely missing in the assembly and thus explain the reduced numbers. Due to their (almost) identical long terminal repeats the correct structure of young elements is more difficult to resolve.

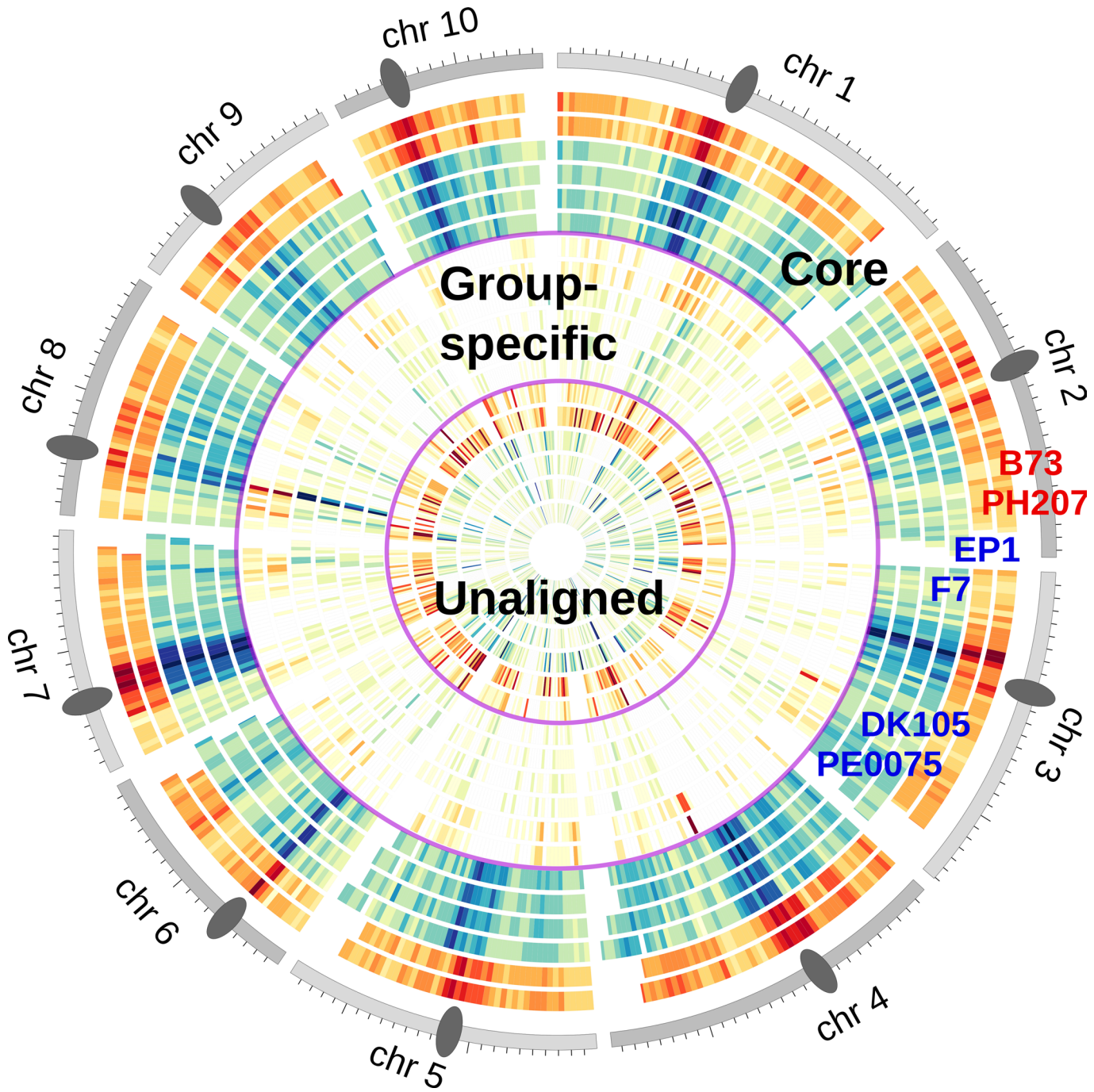


Extended Data Fig. 6 | See next page for caption.

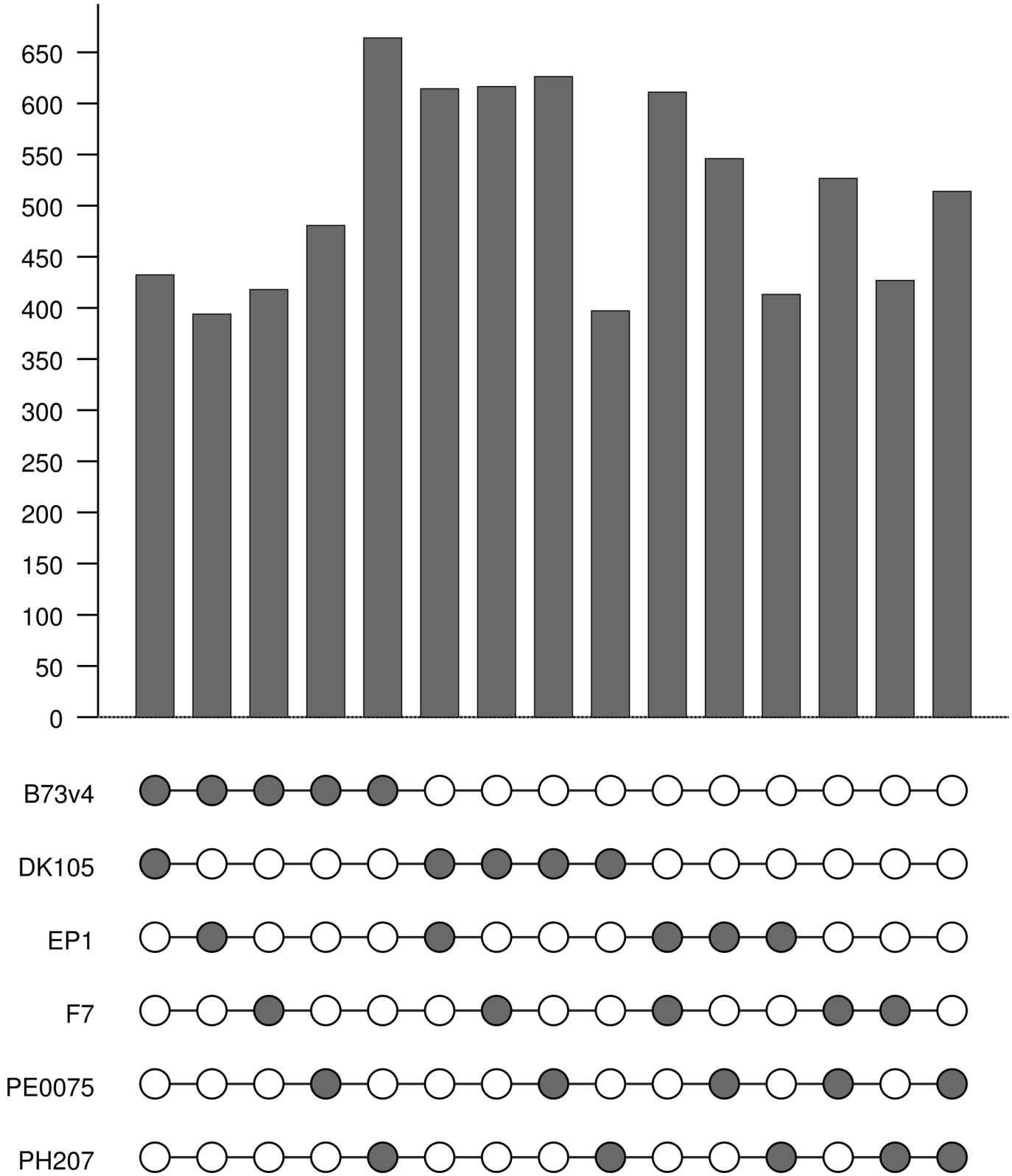
Extended Data Fig. 6 | Tandem repeat content and composition in the genome assemblies. a, The assemblies of EP1 and F7 captured the highest amounts of tandem repeats and knob sequences. The in spite of high fish intensities much lower contents in the B73 and PH207 assemblies reflect the assembly difficulties of these highly repetitive genome regions. **b**, Relationship between the six lines derived from knob location similarities. The input matrix for clustering contained three relative intensities of fish knob locations per chromosome. The knob patterns of the two Dents B73 and PH207 are very similar, as well as the patterns of DK105 and PE0075. **c**, Multiple sequence alignment of selected knob monomers from EP1 together with known knob monomers. The knob sequences in the EP1 assembly consist of 180 and 202 bp monomers with a surplus of the 180 bp monomer by a factor of 6.8. Both monomers are highly similar to previously reported knob monomers from maize with the following Genbank IDs marked as 1, 2, 3 in the figure, 1: AF030934.1, 2: M32521.1 and M32525.1, 3: DQ352544.1_a and DQ352544.1_b. Consensus sequences of the monomers where used to identify all major and minor knob locations in the assemblies. **d**, Gene expression (maximal expression of 7 different conditions per Gene, log10) in relation to the nearest upstream knob signature (left) and tandem repeat (right). Both axes are logarithmic.



Extended Data Fig. 7 | Syntenic knob signatures. x-axis: number of knob units per 2 Mb windows, y axis: chromosome length in Mb. **a**, **Minor knob locations** in the assemblies (y max=100). The majority of the minor knob locations are syntenic between the 8 lines, especially for chromosomes 4,5,and 6, and display no clear distinction between flint and dent. This is in contrast to the knob intensities and extent detected by FISH hybridizations (Fig. 3a). We hypothesize that these minor locations act as 'anchor points' for knob movement or expansion, which seem to occur within very short evolutionary time-spans since the separation of different breeding lines. Larger knob stretches could impact the expression of surrounding genes and influence local recombination rates. **b**, **Major knob locations** in the assemblies (Y max=4000).



Extended Data Fig. 8 | Densities of the core, group-specific and unaligned regions for the six maize lines studies. Intensities are normalized within one track, dark red/blue values representing a higher fraction of aligned regions per 5 Mb bins. Outer chromosome ideograms show the largest of all six lines, dark gray ellipses indicate the approximate centromere position derived from the repeat analysis. Group of six circular tracks show from outer to inner the core, germplasm-specific and unaligned densities. Each group consists of six tracks, displaying from outer to inner densities for B73, PH207, EP1, F7, DK105 and PE0075.



Extended Data Fig. 9 | Genomic sizes in [Mb] of SNP runs-of-identity (RoI) for all 15 pairwise comparisons of the six maize lines. Comparisons within one germplasm (flint: EP1, F7, DK105 and PE0075; dent: B73 version 4 and PH207) share larger genomic fractions than inter-germplasm RoI totals. Y-axis: total sum of RoIs in [Mb], X-axis: line pairs, gray circles indicate maize line used in the respective comparison.

nature research

Corresponding author(s): Klaus FX Mayer
Chris Carolin Schoen

Last updated by author(s): May 29, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Illumina HiSeq 2500 and 4000 produced genomic and RNAseq paired-end reads, respectively;
Data analysis	Mummer v3.23; blat v34; ncbi-blast+ 2.2.28+; exonerate 2.2.0-x86 64 bit version; fast sequence aligner fsa-1.15.9; GenomeThreader v1.6.4; bwa 0.7.17; STAR v2.5.1a; vmatch 2.3.0; samtools v1.3; bcftools v1.3; vcftools v0.1.15; HTseq 0.11.2; R packages ASMap v. 1.0-2, DESeq2 v1.23.3; Trinity 2.4.0; Bridger version r2014-12-01 ; TandemRepeatFinder 4.09; genometools (including LTRharvest, LTRdigest & Tallymer) gt-1.5.3; hmmer3; LTR retriever (LTR Finder within retriever suite); fasttree v2.1.5 SSE3; bbmap v38.31; custom python and R scripts/code will be made available upon request

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data has been released as NCBI bioproject PRJNA360923 and PRJNA360920. Genome sequences and gene models (GFFs) are also available at maizeGDB (https://www.maizegdb.org)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sequence read depths followed recommendations of NRGene's DeNovoMagick assembly pipeline
Data exclusions	No data excluded
Replication	No replication required/meaningful for study
Randomization	Not applicable
Blinding	Not applicable

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Maize (Zea mays ssp. mays) lines EP1, F7, DK105, PE0075 (doubled haploid line from Petkuser population)
Wild animals	not applicable
Field-collected samples	not applicable
Ethics oversight	no ethical guidance required

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Check for updates

BonnMu: A Sequence-Indexed Resource of Transposon-Induced Maize Mutations for Functional Genomics Studies[1][OPEN]

Caroline Marcon,[a,2,3] Lena Altrogge,[b,2] Yan Naing Win,[a] Tyll Stöcker,[b] Jack M. Gardiner,[c] John L. Portwood II,[d] Nina Opitz,[a,4] Annika Kortz,[a] Jutta A. Baldauf,[a] Charles T. Hunter,[e,5] Donald R. McCarty,[e] Karen E. Koch,[e] Heiko Schoof,[b,3] and Frank Hochholdinger[a,3,6]

[a]Institute of Crop Science and Resource Conservation, Crop Functional Genomics, University of Bonn, 53113 Bonn, Germany
[b]Institute of Crop Science and Resource Conservation, Crop Bioinformatics, University of Bonn, 53115 Bonn, Germany
[c]Division of Animal Sciences, University of Missouri, Columbia, Missouri 65211
[d]USDA-ARS Corn Insects and Crop Genetics Research Unit, Ames, Iowa 50011
[e]Horticultural Sciences Department, Plant Molecular and Cellular Biology Program, Genetics Institute, University of Florida, Gainesville, Florida 32611

ORCID IDs: 0000-0001-6699-9441 (C.M.); 0000-0002-6871-235X (L.A.); 0000-0002-4043-5194 (Y.N.W.); 0000-0001-7184-9472 (T.S.); 0000-0002-1804-0110 (J.M.G.); 0000-0002-6339-0881 (J.L.P.); 0000-0003-0776-4890 (N.O.); 0000-0002-1825-0660 (A.K.); 0000-0002-7286-0558 (J.A.B.); 0000-0002-2652-9485 (C.T.H.); 0000-0001-8694-5117 (D.R.M.); 0000-0003-1469-8799 (K.E.K.); 0000-0002-1527-3752 (H.S.); 0000-0002-5155-0884 (F.H.)

Sequence-indexed insertional libraries in maize (*Zea mays*) are fundamental resources for functional genetics studies. Here, we constructed a *Mutator* (*Mu*) insertional library in the B73 inbred background designated *BonnMu*. A total of 1,152 *Mu*-tagged F_2-families were sequenced using the *Mu*-seq approach. We detected 225,936 genomic *Mu* insertion sites and 41,086 high quality germinal *Mu* insertions covering 16,392 of the annotated maize genes (37% of the B73v4 genome). On average, each F_2-family of the *BonnMu* libraries captured 37 germinal *Mu* insertions in genes of the Filtered Gene Set (FGS). All *BonnMu* insertions and phenotypic seedling photographs of *Mu*-tagged F_2-families can be accessed via MaizeGDB.org. Downstream examination of 137,410 somatic and germinal insertion sites revealed that 50% of the tagged genes have a single hotspot, targeted by *Mu*. By comparing our *BonnMu* (B73) data to the *UniformMu* (W22) library, we identified conserved insertion hotspots between different genetic backgrounds. Finally, the vast majority of *BonnMu* and *UniformMu* transposons was inserted near the transcription start site of genes. Remarkably, 75% of all *BonnMu* insertions were in closer proximity to the transcription start site (distance: 542 bp) than to the start codon (distance: 704 bp), which corresponds to open chromatin, especially in the 5′ region of genes. Our European sequence-indexed library of *Mu* insertions provides an important resource for functional genetics studies of maize.

The genome of the maize (*Zea mays*) inbred line B73v4 is predicted to contain 44,117 genes, including 39,179 high confidence protein encoding gene models and 4,938 genes encoding for large intergenic non-coding RNAs, transfer RNAs, and microRNAs (Jiao et al., 2017). To date only a few hundred of these genes have been functionally characterized (Schnable and Freeling, 2011). The identification of specific mutants in forward genetic screens and the subsequent cloning of the underlying genes has been the method of choice to determine gene functions for several decades (Candela and Hake, 2008). Until now, reverse genetics using specific gene sequences as a starting point for the identification of the corresponding mutants (Candela and Hake, 2008) has mainly been used to independently validate candidate genes identified by forward genetics experiments (e.g. Hochholdinger et al., 2008; Nestler et al., 2014). Nevertheless, reverse genetics has also been successfully applied to identify developmental mutants of genes without prior knowledge of their phenotype (Xu et al., 2015). Genome-wide insertional mutagenesis is a tool to create loss-of-function mutations for virtually all genes in a genome by the insertion of DNA into the gene of interest. Sequence-indexed collections of such mutants have been generated for Arabidopsis (*Arabidopsis thaliana*; Alonso et al., 2003), rice (*Oryza sativa*; Hirochika et al., 2004), and maize (Fernandes et al., 2004; McCarty et al., 2005; Liang et al., 2019).

Endogenous transposons or transposable elements (TEs) are mobile DNA sequences initially discovered in maize by McClintock (1951). Once activated, a transposon can move from one location in the genome to another, thereby disrupting genes. The two major classes of transposons are class I retrotransposons and class II DNA transposons. The main difference between these classes is their mode of transposition. Class I retrotransposons first require transcription of their DNA to an RNA intermediate, which is then reversely transcribed into DNA, before it is reinserted into a new

 © 2020 American Society of Plant Biologists. All Rights Reserved.

Downloaded from https://academic.oup.com/plphys/article/184/2/620/6117865 by ulbbonn user on 21 December 2021

genomic location (for review, see Feschotte et al., 2002). This mechanism is catalyzed by a reverse transcriptase. In contrast, the TE itself is the transposition intermediate for class II DNA transposons. Both classes have in common that they consist of autonomous and nonautonomous TEs. Autonomous TEs can transpose by themselves, because they encode a transposase, whereas nonautonomous TEs require another autonomous transposon to move. Class I transposons are predominantly intergenic, contributing to plant genome size (Bennetzen, 2000). In contrast, class II TEs insert preferentially in and around genes (Dietrich et al., 2002; Fernandes et al., 2004; Settles et al., 2004), which is a prerequisite for genome-wide insertional mutagenesis screens.

The most active class II transposon family in maize is the family of *Mutator* (*Mu*) elements (Lisch, 2002). Discovered by Robertson (1978), *MuDR,* an autonomous TE, controls the transposition of itself and 16 classes of nonautonomous *Mu* elements (Lisch, 2015). All *Mu* transposons share highly conserved 215-bp terminal inverted repeats (TIRs) at both ends of the element and upon insertion they generate 9-bp target site duplications directly flanking the *Mu* transposon sequences. For untargeted insertional mutagenesis, *Mu* exhibits a number of advantages over other maize class II transposon families, such as Activator/Dissociation (*Ac*/*Ds*) or Enhancer/Suppressor-mutator (*En*/*Spm*), including its high transposition rate and high-copy number (Chandler and Hardeman, 1992). While *Ac*/*Ds* transposons exhibit a preference for regional mutagenesis (i.e. transposition into closely linked genes; Brutnell, 2002), *Mu* elements randomly target genes throughout the maize genome (Lisch, 2002). As such, *Mu* insertion site frequencies strongly correlate with gene density (Liu et al., 2009; Schnable et al., 2009; Springer et al., 2018).

In the past, adapter-mediated (Frey et al., 1998) and thermal asymmetric interlaced-PCR (Liu and Whittier, 1995) have been used to amplify *Mu*-flanking fragments for sequencing. The presence of highly conserved TIR sequences at both ends of the transposons is ideally suited to design *Mu*-specific primers used in thermal asymmetric interlaced-PCR. This PCR technique combines nested TIR-specific primers and degenerative arbitrary primers to amplify sequences neighboring known sequences by using high and low annealing temperature cycles (Settles et al., 2004). More recently, high-throughput next generation sequencing enables efficient, reproducible, and sequence-based identification and mapping of transposon insertion sites in maize (McCarty et al., 2013; Liu et al., 2016; Liang et al., 2019).

Thus far, several transposon-tagged maize populations have been generated using the *Mu* transposon system (Fernandes et al., 2004; Stern et al., 2004; McCarty et al., 2005; McCarty et al., 2013; Liang et al., 2019). Nevertheless, to date only about 52% of the annotated maize gene models (Liang et al., 2019) have mapped *Mu* insertions. Among these collections, the insertion-tagged *UniformMu* population (McCarty et al., 2005) has been developed by backcrossing an active *MuDR* element eight times into the genetically uniform inbred line W22. A genetically uniform genetic inbred background is one key feature of the *UniformMu* resource, because it allows to distinguish between parental and newly created mutations. Therefore, a uniform background aids to identification of insertions related to new phenotypes (McCarty et al., 2005). A large set of more than 14,000 *UniformMu* lines are publicly available (Liu et al., 2016). Another *Mu* insertional library, called *ChinaMu*, has been generated by crossing a *Mu*-starter line to the inbred line B73 (Liang et al., 2019). The *Mu*-tagged sequences of 2,581 F_2-*Mu* lines were isolated by a *Mu* tag enrichment approach and sequenced by high-throughput sequencing.

Here we used the *Mutator*-seq approach (*Mu*-seq; McCarty et al., 2013; Liu et al., 2016) to construct multiplexed sequencing libraries in the inbred line B73 that guarantees unbiased coverage of mutations in the maize genome. The *Mu*-seq reverse genetics approach enables the identification of F_2-families carrying transposon insertions in genes of interest in a sequence-indexed transposon tagged population. These newly identified transposon induced mutations can be used for subsequent molecular and genetic analyses. Furthermore, this European database (*BonnMu*) of sequence-indexed *Mu* transposon insertion sites in B73 complements the *UniformMu* and *ChinaMu* resources and allows for forward and reverse genetics experiments. Besides the sequence-indexed *Mu* insertional libraries from North America (*UniformMu*) and Asia (*ChinaMu*), our resource of transposon induced

[1]This work was supported by the Deutsche Forschungsgemeinschaft (grant no. MA 8427/1–1 to C.M.), the Bundesministerium für Bildung und Forschung (grant no. 031 B195C to F.H.), and the United States Department Agriculture, Agricultural Research Service to J.L.P. and C.T.H.

[2]These authors contributed equally to this article.

[3]Senior authors.

[4]Present address: Federal Ministry of Food and Agriculture, 53123 Bonn, Germany

[5]Present address: USDA-ARS Chemistry Research Unit, Gainesville, FL 32608

[6]Author for contact: hochholdinger@uni-bonn.de.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Caroline Marcon (marcon@uni-bonn.de).

C.M. and L.A. interpreted the data and drafted the article; C.M. and F.H. conceived the research; C.M., Y.N.W., N.O., and A.K. generated the *Mu*-seq libraries; C.M. and Y.N.W. phenotyped the *Mu*-seq families and prepared seedling photos for public access via MaizeGDB.org; L.A., T.S., and H.S. conceived and carried out the bioinformatics analyses; J.M.G. and J.L.P. curated insertions and phenotypes of the *Mu*-tagged families at MaizeGDB.org; C.T.H., D.R.M., and K.E.K. provided W22 datasets and participated in data interpretation; J.A.B. was involved in statistical analysis; H.S. and F.H. participated in data interpretation, coordinated the study, and helped to draft the article; all authors approved the final draft of the article.

[OPEN]Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.20.00478

Downloaded from https://academic.oup.com/plphys/article/184/2/620/6117865 by ulbbonn user on 21 December 2021

Table 1. Mu-*seq library alignment statistics*

Statistical Analysis	*Mu*-seq 1	*Mu*-seq 2	Combined
Raw read pairs	89,833,184	55,248,154	–
Read pairs after trimming[a]	65,773,604	43,651,352	–
Average alignment rate	89%	83%	–
Remaining read pairs after removing duplicate reads[b]	23,251,540	18,785,728	–
Union (somatic and germinal insertions)[c]			
Insertion sites	135,108	112,205	225,936
Insertion sites in FGS genes	81,698	68,413	137,410
Insertion sites in exons of FGS genes	66,270	55,221	110,730
Insertion sites in cds of FGS genes	22,743	19,743	40,140
Insertion sites in 5′ UTR of FGS genes	46,362	37,945	75,201
Insertion sites in 3′ UTR of FGS genes	4,085	3,361	7,125
Insertion sites in introns of FGS genes	14,586	12,404	25,203
Insertion sites upstream of FGS genes (≤ 500 bp)	23,268	19,411	38,609
Intergenic insertion sites	53,410	43,792	88,562
Mu-tagged FGS genes	19,162	17,923	22,362
Intersection (germinal insertions)[d]			
Insertion sites	43,225	24,234	66,750
Insertion sites in noncoding DNA	16,462	9,323	25,664
Insertions sites in FGS genes	26,762	14,912	41,086
Mu-tagged FGS genes	13,590	9,538	16,392
Mu-tagged FGS genes, affected in their cds	5,605	3,487	7,582

[a]*Mu*-seq reads were trimmed and aligned to the B73v4 genome as described in the "Materials and Methods" section. [b]Duplicate reads were removed as described in the "Materials and Methods" section. [c]All insertion sites identified in any of the samples. Insertion sites were identified as described in the "Materials and Methods" section. [d]Insertion sites identified in only one row and one column sample, respectively.

mutations will offer easier access for European researchers but also for maize geneticists around the globe.

RESULTS

Mu-seq Library Alignment Statistics of *BonnMu*

To generate a sequence-indexed collection of transposon tagged maize mutants, the *Mu*-seq protocol (McCarty et al., 2013; Liu et al., 2016) was applied to generate two libraries each containing 576 Mu active B73 F_2-families (*BonnMu*) of maize. These 1,152 mutagenized F_2-families were pooled according to a 24 × 24 grid system per library (see "Materials and Methods"). Sequencing of the two libraries yielded 55 and 90 million raw read pairs (Table 1; Supplemental Dataset S1). Each *Mu*-seq library was parsed according to the individual 6-bp barcodes of the 48 multiplexed pools (Supplemental Table S1), resulting in an averaged output of 1,151,003 and 1,871,525 read pairs per pool (Supplemental Dataset S1). Downstream statistical analysis of the two *Mu*-seq libraries are summarized in Table 1. After U-adapter and TIR sequences were trimmed, 83% and 89% (Table 1) of the remaining reads (43 and 65 million) were aligned to the B73 Filtered Gene Set (FGS) AGPv4.36, containing 44,117 high confidence gene models. After duplicate read removal, 18 and 23 million read pairs remained and were used to identify genomic insertion sites, in (1) any of the 48 pools (somatic and germinal insertions, named union in Table 1) and (2) at intersections of one row and one column pool (germinal insertions, named intersection in Table 1). For insertion site identification, reads were counted in each of the 48 pools. Only insertion sites supported by at least four reads were used for further analyses. In total, 225,936 genomic insertion sites were detected in the two *Mu*-seq libraries (Table 1; Supplemental Dataset S2). Among those insertion sites, 61% (137,410 of 225,936 insertions) mapped to FGS genes, tagging in total 22,362 genes (Table 1). Thus, there were on average six somatic or germinal insertions per tagged FGS gene. Further parsing of insertion sites identified 110,730 insertion sites in exons of FGS genes, 40,140 of such sites in the coding sequence, 75,201 sites in 5′ untranslated region (UTR), 7,125 insertion sites in 3′ UTR, and 25,203 insertion sites in introns of FGS genes. In addition, 38,609 insertion sites ≤500 bp upstream of FGS genes and 88,562 intergenic insertions were detected (Table 1). Based on the genome annotation of AGPv4.36, insertion sites can be assigned to more than one feature (e.g. both the 5′ UTR and exonic region), within a gene. To calculate the number of insertion sites in introns, the union of insertion sites located in exons/coding sequence (cds) and UTRs was subtracted from the number of insertion sites in FGS genes. Therefore, the number of insertion sites located in introns cannot be recalculated from Table 1.

Finally, to only consider heritable (germinal) insertions the information from the different pools was combined to track down *Mu*-tagged genes in specific F_2-families. Each F_2-family is represented by the intersection of one column and one row pool (according to the 24 × 24 grid system; Supplemental Table S1). By that criterion, we identified 66,750 germinal insertion sites in the two *Mu*-seq libraries, of which 38% (25,664

Downloaded from https://academic.oup.com/plphys/article/184/2/620/6117865 by ulbbonn user on 21 December 2021

of 66,750) mapped to noncoding DNA (Table 1). The remaining 62% (41,086 of 66,750) of the insertion sites were located in genes of the FGS (Table 1; Supplemental Dataset S3). On average, each F_2-family harbored 37 germinal insertions in FGS genes (Supplemental Tables S2 and S3; Supplemental Dataset S3). Furthermore, 46% of mutated FGS genes (7,582 of 16,392 FGS genes) contained insertions in coding sequences.

European-Based *BonnMu* Complements the North American *UniformMu* and the Asian *ChinaMu* Resource

Based on the analysis of two *Mu*-seq libraries in B73 (*BonnMu*), *Mu* insertions affected 16,392 of 44,117 (37%; Fig. 1A; Table 1) high-confidence gene models of maize. Of the 16,392 genes, 41% (6,763) were identified in F_2-families of both *Mu*-seq libraries, whereas 59% (9,656) of the genes were uniquely detected in one of the *Mu*-seq libraries (Fig. 1A). We observed a significantly higher overlap of *Mu*-tagged genes in both *Mu*-seq libraries than expected (2,938; Supplemental Table S4). This finding supports the notion that genes are not randomly targeted by *Mu* transposons, but that there is a preference for *Mu* tagging of specific genes.

The 16,392 genes of the *BonnMu* collection in the B73 background was compared with 16,090 *Mu*-tagged genes of the *UniformMu* resource in the W22 background (McCarty et al., 2013) and to 20,396 genes tagged in B73 in the *ChinaMu* dataset (Liang et al., 2019) and personal communication with R. Song). The three collections cover 57% (25,140 of 44,117) of the genes present in the B73 reference genome AGPv4.36 (Fig. 1B). A considerable proportion of *Mu*-tagged genes, that is, 42% (10,623 of 25,140) were hit in all three collections. This number of overlapping genes was significantly higher than expected by pure chance (2,724; Supplemental Table S4), suggesting that at least some genes are preferentially targeted by *Mu* insertions.

BonnMu insertions can be browsed at the MaizeGDB Web site (https://www.maizegdb.org; Portwood et al., 2019) as previously described for the *UniformMu* database (Liu et al., 2016). A unique feature of *BonnMu* is that photos of seedling phenotypes of segregating F_2-families are linked to the database entries of each transposon insertion. To access and visualize B73 insertions in maizegdb.org, a gene model identifier is needed for the genome browser view. When "select track" and "Bonn*Mu*" are clicked, insertion sites become visible in the locus. Clicking on the insertion site links to the *Mu* insertion identifiers (e.g. *BonnMu*0000812) and the respective F_2-*Mu*-seq family (e.g. *BonnMu*-2-A-0596) and its respective phenotype 10 d after germination. Among the analyzed F_2-families various leaf color mutants were identified such as albino (e.g. *BonnMu*-2-A-0784) or pale green (e.g. *BonnMu*-2-A-0784). The mutation rate, detected on the basis of albino and pale green leaf phenotype, was 13% (149 of 1,152 F_2-families). Furthermore, shoot gravitropism mutants (e.g. *BonnMu*-2-A-0596, *BonnMu*-2-A-0598) were identified.

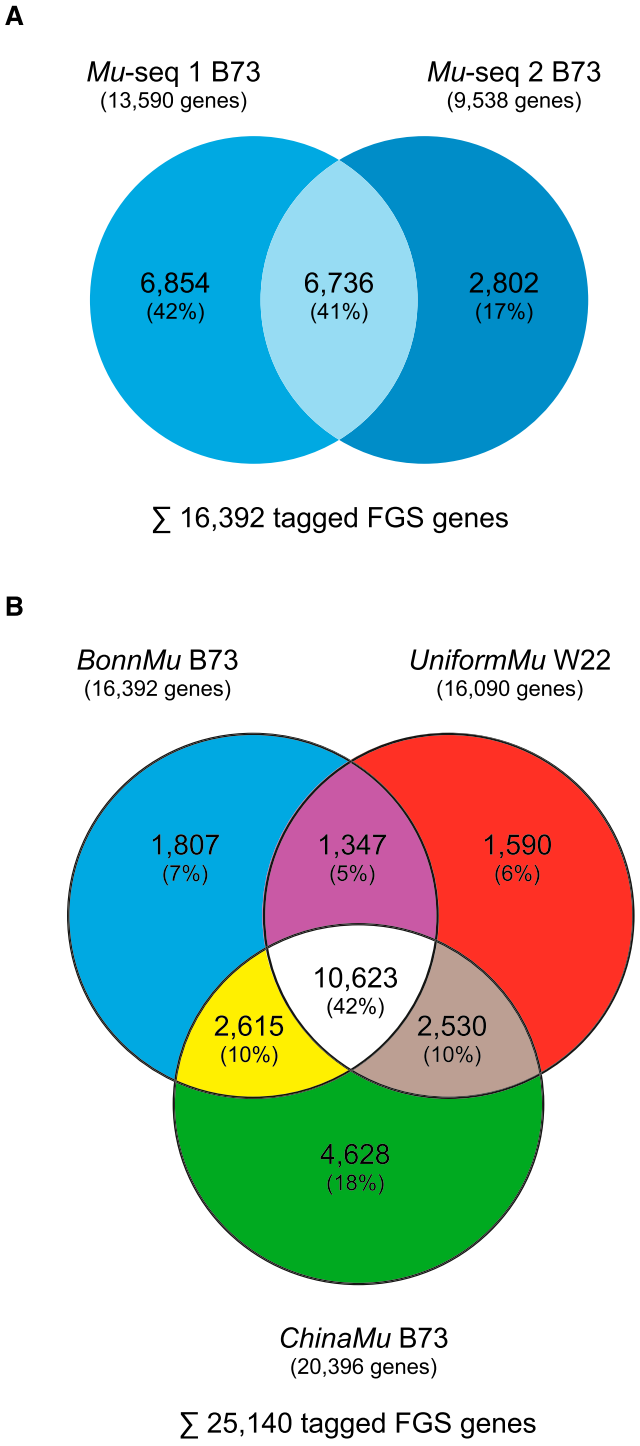


Figure 1. Overlap of FGS genes tagged by *Mu* transposon insertions. A, Tagged FGS genes among the two *Mu*-seq libraries in the B73 background. B, FGS genes with *Mu* insertions in the three available *Mu*-tagged resources in the B73 (blue, *BonnMu*; green, *ChinaMu*) and W22 inbred backgrounds (red, *UniformMu*).

Downloaded from https://academic.oup.com/plphys/article/184/2/620/6117865 by ulbbonn user on 21 December 2021

The Number of *Mu* Insertions Correlates with Gene Length in a Size-Dependent Manner

The present dataset contains 41,086 *Mu* insertions in 16,392 B73 high confidence gene models (Table 1; Supplemental Dataset S3). More specifically, 39% (6,455 of 16,392) of the high-confidence gene models were tagged by a unique insertion (Fig. 2A), whereas 61% (9,937 of 16,392) contained at least two *Mu* insertions. A similar distribution was observed when the 7,582 FGS genes harboring insertions in their cds (Table 1; Supplemental Dataset S3) were analyzed. Among those, 33% (2,535 of 7,582) had a unique insertion, whereas 67% (5,047 of 7,582) were tagged by at least two *Mu* insertions (Fig. 2A). An extreme example is a gene encoding an adaptor protein 4 (AP-4) complex subunit μ (Zm00001d002925), which harbored 33 independent insertions in its cds (Supplemental Dataset S3).

The length of the affected genes ranged between 72 bp and 195 kb (Fig. 2B; Supplemental Dataset S3). The three shortest genes (72 bp) encode transfer RNAs (Zm00001d006818, Zm00001d007104, and Zm00001d034123), whereas the largest gene of 195 kb (Zm00001d001010) represents a large intergenic noncoding RNA. We observed 882 tagged genes (5% of all 16,392 *Mu*-tagged FGS genes) in a size range of 72 to 999 bp (Fig. 2B). Then, the size and the number of tagged genes increased, with a peak of 2,548 affected genes (16%) having a size between 3 and 3.999 kb. After this peak the number of affected genes decreased, while their length increased. Overall, 12,510 tagged genes (76%) had a size between 72 bp and 6.999 kb, whereas only 3,882 genes (24%) were >7 kb. In contrast with the observed distribution of the 16,392 tagged genes according to their gene length, the size distribution of all 44,117 B73v4 high-confidence gene models differed. Although 47% of all B73v4 gene models had a size of <2 kb, only 18% of all *Mu*-tagged genes of the present dataset ranged in that size (Supplemental Fig. S1). In contrast, 53% of all B73v4 gene models had a length between 2 and 194.5 kb, whereas among all *Mu*-tagged genes 82% ranged between these lengths.

Next, we tested the hypothesis that the number of *Mu* insertions increased with the length of the affected genes. Although a strong linear correlation of insertion number with gene size was detected in the gene size range of 72 bp to 3.999 kb, there was no correlation with gene size for genes >4 kb (Fig. 2B). Hence, the calculated Pearson correlation over all groups of gene sizes revealed no correlation (r = 0.06; Fig. 2C). This abrupt transition in gene-size dependence at 4 kb was also detected in the W22 dataset (Supplemental Fig. S2; Supplemental Dataset S4). Due to the fact that a large subset of 48% affected genes (7,838 B73 and 7,790 W22 genes) have a size of <4 kb, the gene-size-dependent correlation is important to consider.

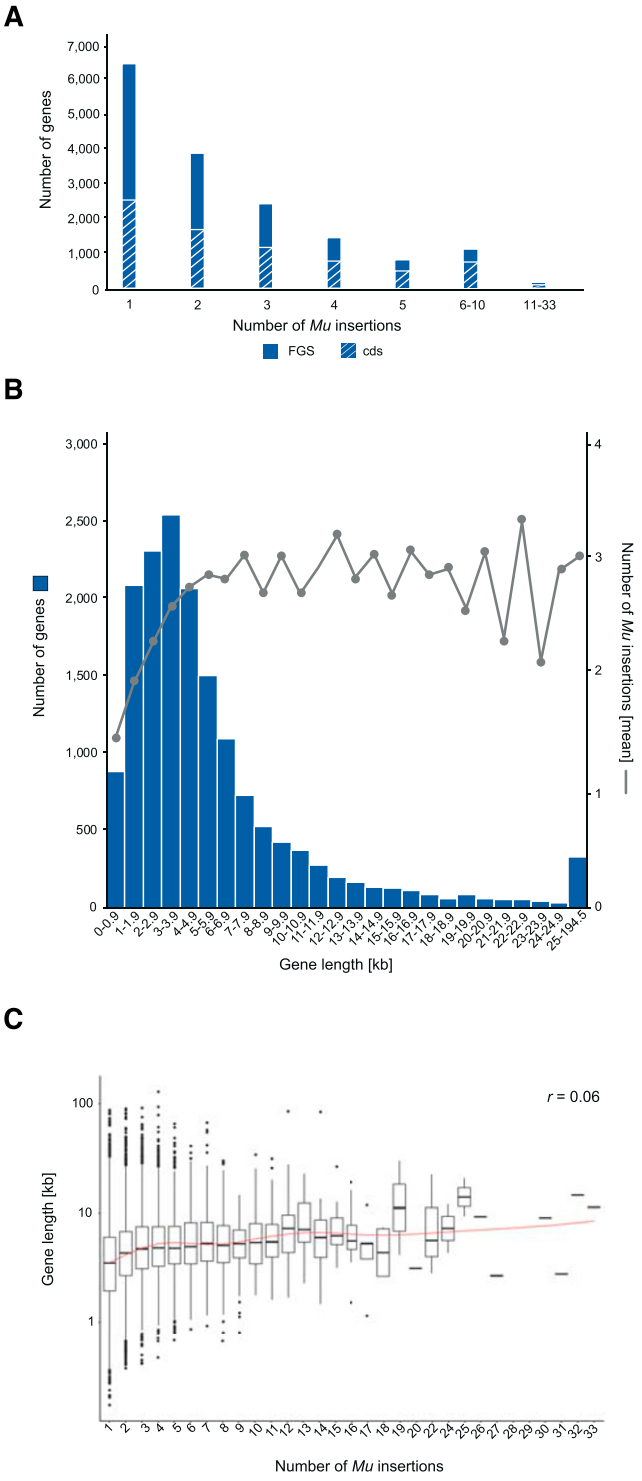


Figure 2. Number of *Mu* insertions and its correlation to the length of affected FGS genes of B73. A, Distribution and number of *Mu* insertions among 16,392 affected FGS genes and the subset of 7,582 genes harboring insertions in the cds of FGS genes. B, Number of tagged genes and associated mean number of *Mu* insertions plotted against the corresponding gene size. C, Length distribution of affected genes plotted against the individual number of *Mu* insertions (Pearson correlation coefficient, r = 0.06).

Downloaded from https://academic.oup.com/plphys/article/184/2/620/6117865 by ulbbonn user on 21 December 2021

Table 2. *Number of expected and observed unique and overlapping bins (i.e. 200-bp sized regions of genes) targeted by* Mu *transposon insertions in B73 and W22*

Genotype	B73 YES	B73 NO	Total	Observed/Expected
W22 YES	293[a]	138	431	Observed
	50	381	–	Expected[b]
W22 NO	1,406	12,899	14,305	Observed
	1,649	12,656	–	Expected[b]
Total	1,699	13,037	14,736	–

[a] $P < 2.2 \times e^{-16}$ (obtained from χ^2-testing). [b] Expected value of bins being hit in B73 and W22 was calculated as total number of bins identified in W22 (431)* total number of bins identified in B73 (1,699) / total number of bins (14,736).

Mu Insertion Hotspots Are Conserved Among Affected Genes of B73 and W22

Next, we investigated if *Mu* transposons show preferences regarding their insertion sites in the genes. Therefore, we examined all insertion sites in any of the samples, including somatic and germinal insertions of the present *BonnMu* datasest (Table 1; Supplemental Dataset S2). The majority of these insertion sites (61%; 137,410 of 225,936) was located in FGS genes. Each of the 22,362 *Mu*-tagged genes was divided into bins of 200 bp in size. Then, all insertion sites were sorted into the corresponding bins. Each bin with at least one insertion site was counted as a hotspot. In total, 38% (8,502 of 22,362 genes) of the *Mu*-tagged genes had one hotspot, which was predominantly targeted by the transposons (Supplemental Fig. S3A). The number of hotspots per affected gene ranged from 1 to 15, with a median number of two hotspots, independent of gene length (Supplemental Fig. S3B).

To investigate if insertion hotspots are conserved among affected genes of our *Mu*-seq repository in B73 and the *UniformMu* repository in the W22 genetic background, one exemplary W22 *Mu*-seq dataset (SRA accession SRP028545) was reanalyzed. A total of 558 identified insertion sites in W22 supported by at least two reads could be assigned to 423 FGS genes.

Among these 423 *Mu*-tagged genes in W22, 413 (98%) were also affected in B73. For the overlapping 413 genes a hotspot analysis was performed. Because we were interested in only conserved hotspots of new *Mu* insertion events, we first corrected the affected genes for endogenous *Mu* insertions that are present in the genomes of B73 and W22. About 40 canonical *Mu* elements exist in the two genomes (Springer et al., 2018; D. R. McCarty, personal communication; Supplemental Dataset S5). Due to the fact that such parental *Mu* elements are fixed in every B73/W22 plant, such elements account for a large proportion of reads. Hence, the endogenous elements could contribute to an apparent hotspot overlap between B73 and W22. After correcting for endogenous *Mu* elements, 22,348 B73 genes and 420 W22 genes remained. Among those, 412 genes overlapped between the two datasets and were divided into bins of 200 bp in length. Then, *Mu* transposon insertion sites identified in B73 and W22 were sorted into the corresponding bin. Out of 14,736 bins, 293 were targeted in both B73 and W22, 1,406 only in B73, 138 only in W22, and 12,899 in neither B73 nor W22 (Table 2). A χ^2-test was performed to test the hypothesis that there is an association between insertion sites in B73 and W22. The observed value of overlapping bins in both B73 and W22 (293) significantly exceeded the expected value of 50 bins, demonstrating that there are conserved insertion site hotspots in both genomes. This observation is exemplarily shown in Figure 3 for gene Zm00001d002371. Here, *Mu* transposon insertions were concentrated around two hotspots within that gene in B73 and one overlapping hotspot in W22.

Mu Transposons Preferentially Insert near the Transcription Start Site of Genes in B73 and W22

To investigate if *Mu* transposons show preferences regarding their insertion sites in genes, the union of all B73 *Mu* transposon insertion sites located in FGS genes was analyzed. A total of 110,730 of the insertion sites were located in exons of FGS genes (Table 1). Whereas 40,140 of the insertion sites were assigned to the cds of FGS genes, the majority of insertion sites (75,201) were located in the 5′ UTR of FGS genes.

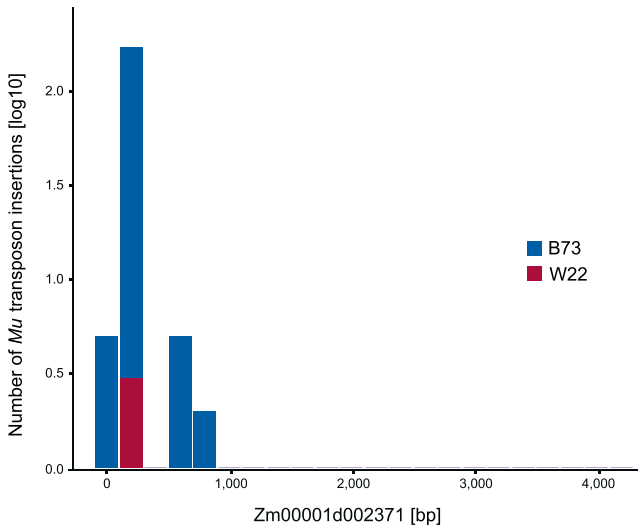


Figure 3. A *Mu* insertion hotspot in the representative gene Zm00001d002371 affected in the B73 and W22 genomes. The number of *Mu* insertions is plotted against location in Zm00001d002371.

Downloaded from https://academic.oup.com/plphys/article/184/2/620/6117865 by ulbbonn user on 21 December 2021

To examine *Mu* transposon insertion site preferences in B73 compared with W22, for each affected FGS gene with annotated cds, the cds was divided into deciles. Subsequently, all insertion sites located in the cds of FGS genes were sorted into the corresponding cds decile. Insertion sites located up- or downstream of the cds were designated as insertions in the 5′ or 3′ UTR, respectively. The majority of *Mu* insertion sites was detected in the 5′ UTR and the first decile of the cds in both B73 (76%) and W22 (62%), with another smaller peak at the end of the cds (Fig. 4A). To account for the length of the FGS genes, all affected genes were sorted into gene sets based on their length (Supplemental Fig. S4A). Each of the sets (deciles) contained 2,152 affected B73 and either 42 or 43 affected W22 genes. Whereas in shorter genes of B73 (deciles 1% to 10%, and 11% to 20%) *Mu* insertions were evenly distributed among the gene, longer genes showed an insertion peak at the start region (i.e. the 5′ UTR and first 10% of the genes' cds; Fig. 4B). Another small peak of insertions was detected at the end of the cds (100%) of B73 genes.

Although the W22 dataset (SRP028545) only contained 423 genes, a similar preference of *Mu* insertions to target the 5′ UTR of the genes was observed (Fig. 4B). The observation that *Mu* insertions were evenly distributed among the gene of shorter genes, especially in B73, can be explained by the very short 5′ UTR of those genes. By comparing the length of the 5′ UTRs of each of the ten gene sets it became clear that the first two gene sets (1% to 10%, 135–1214 bp; 11% to 20%, 1215–1872 bp) had very short or even no 5′ UTRs (Supplemental Fig. S4B). Therefore, the absolute distance of insertions to the 5′ UTRs in those genes was small, even if the insertion was close to the 3′ end of the gene.

To verify if the insertion sites were located closer to the 5′ UTR start or closer to the start codon (ATG) of the affected genes, the distance (bp) from insertions inside or upstream of FGS genes to both the 5′ UTR start and cds start codon was calculated for the B73 dataset. The density plot (Fig. 5) represents only insertion sites located 500 bp upstream of the 5′ UTR to 2,000 bp downstream of the genes' start codon. Irrespective of this cutoff, the vast majority of 93% of all insertion sites are in the range of the above-mentioned distance to the 5′ UTR and 90% of all insertion sites are in such a distance to the cds. This result suggested that the majority of *Mu* insertion sites are close to the 5′ UTR and the cds start of the affected genes. This is also visible in the density plot (Fig. 5) where the majority of *Mu* insertion sites peaked in close proximity to the 5′ UTR and cds start. The slope of the curve revealed that the *Mu* transposon insertions are closer to the 5′ UTR than to the genes' start codon. Finally, 25% of all *Mu* insertion sites were located in close proximity to the 5′ UTR (46 bp) and to the cds start (111 bp; Fig. 5). When 50% of all insertion sites were considered, the distance was on average 135 bp to the 5′ UTR and 271 bp to the cds. Remarkably, 75% of all *Mu* insertion sites were located near the 5′ UTR (542 bp) and close to the start codon

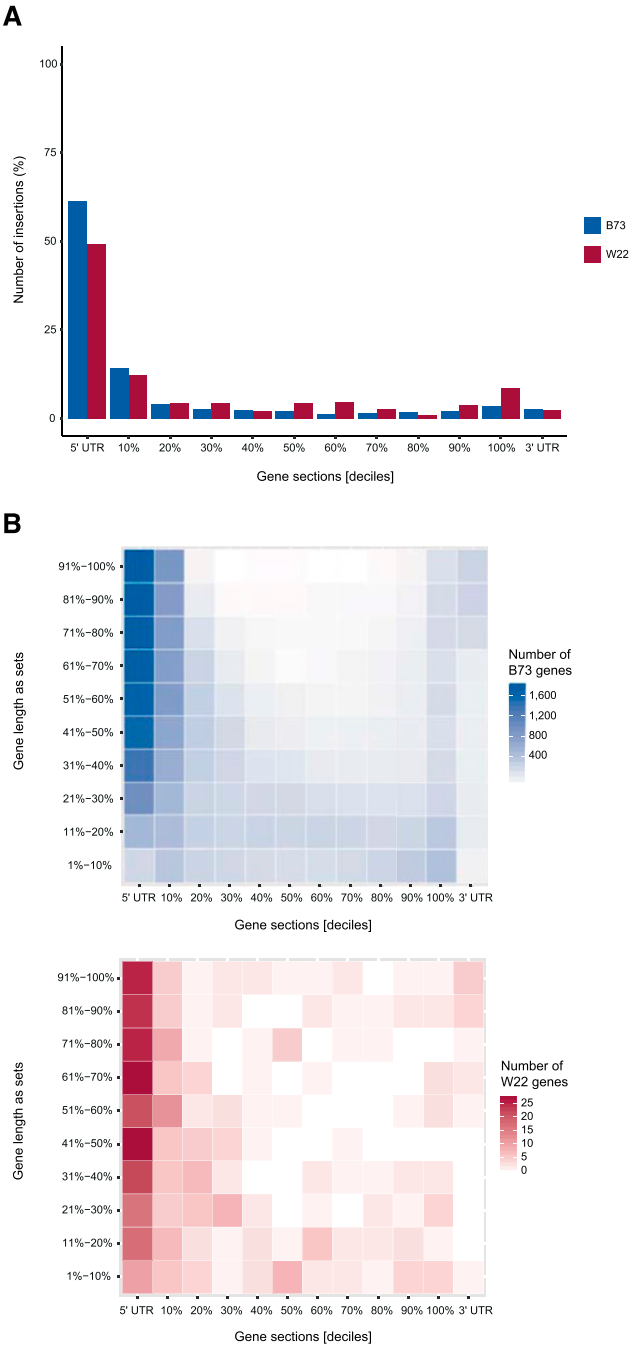


Figure 4. Distribution of *Mu* transposon insertion sites in FGS genes. A, Proportion of *Mu* insertions located in each decile of the cds, 5′ UTR, and 3′ UTR (B73, blue; W22, red). B, All FGS genes targeted by *Mu* transposons in B73 and W22 (SRP028545) were divided into gene sets based on their length (1% to 10% represent the shortest set of genes and 91% to 100% represent the longest genes) and the number of genes with insertions in each cds decile, 5′ UTR and 3′ UTR (B73, blue; W22, red).

(704 bp) of the affected genes. In summary, this data revealed that the majority of *Mu* insertion sites are located closer to the 5′ UTR TSS than to the cds start of affected genes.

Downloaded from https://academic.oup.com/plphys/article/184/2/620/6117865 by ulbbonn user on 21 December 2021

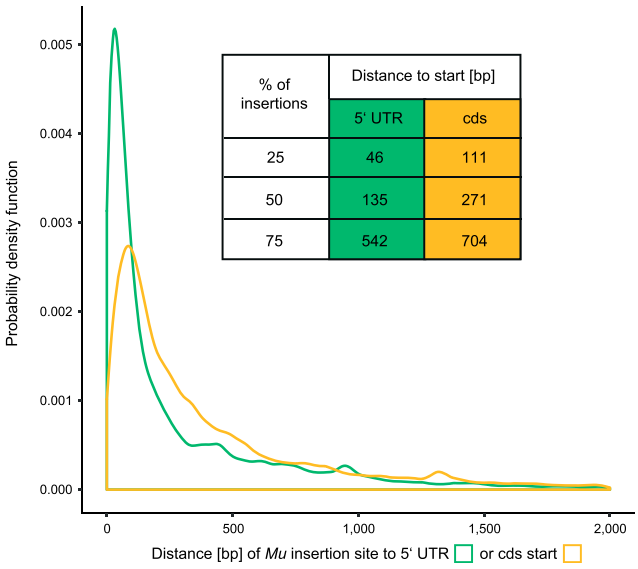


Figure 5. Distance of *Mu* insertion sites to the start of the 5′ UTR (transcription start site [TSS]) and the cds of affected B73 genes. Density plot illustrating the distances of the *Mu* insertion sites to the 5′ UTR start (green) and the start of the cds (yellow). Only insertion sites located between 500 bp upstream of the 5′ UTR and 2,000 bp downstream of the ATG were considered. Calculated distances of *Mu* insertion sites to the 5′ UTR and the start of affected genes are given in the table.

DISCUSSION

Construction of the European sequence-indexed insertional library (*BonnMu*) in maize discovered 41,086 high quality germinal *Mu* insertions in 16,392 high confidence gene models (Fig. 1), tagging 37% of the B73 genome. This number is comparable with the North American *UniformMu* resource (McCarty et al., 2013; Liu et al., 2016), covering 16,090 genes, which were tagged by 43,943 germinal *Mu* insertions. Recently, an Asian insertional library (*ChinaMu*; Liang et al., 2019) identified 66,565 high-quality germinal insertions representing 20,396 annotated maize genes. In comparison with *UniformMu* and *ChinaMu*, 1,807 genes were newly tagged by *BonnMu*. Hence, with our new European sequence-indexed resource, the percentage of *Mu*-tagged maize genes in public repositories increased from 52% to 57%. Although, this number is still less than that of Arabidopsis (74% of the genes covered; Alonso et al., 2003), it is comparable with rice (about 60% tagged genes; Wang et al., 2013). However it is worth noting that different techniques, such as transfer-DNA insertional mutagenesis (Alonso et al., 2003; Toki et al., 2006), two-component transposon system Ac/Ds based mutagenesis (van Enckevort et al., 2005), or Tos17 retrotransposon mutagenesis (Miyao et al., 2003) were applied to construct insertional libraries in Arabidopsis and rice. Nevertheless, because adequate *Mu*-tagged F_2-families are available in the three global resources (Liang et al., 2019) expanding the datasets is already being implemented to increase the coverage of tagged genes. As described by Liang et al. (2019) we also detected an increase in the number of insertional alleles per gene reducing the efficiency of tagging new genes. A reduction of efficient new gene tagging was also observed in insertional libraries of rice, where more than 246,000 insertions only covered 60% of the genes (Wang et al., 2013). Therefore, our data and previous studies (Kolesnik et al., 2004; Hsing et al., 2007; Zhang, 2007; Liang et al., 2019) suggest that it is difficult, if not impossible, to achieve whole-genome saturation using insertional mutagenesis.

Maize is an excellent model for transposon genetics and mutagenesis studies (Settles et al., 2004). Several endogenous DNA transposon systems, such as *Ac/Ds* (McClintock, 1950) and *Mu* transposons (Robertson, 1978), are highly active in maize and can be easily adopted for large-scale mutagenesis. Genic transposon insertions are usually large enough to cause substantial disruptions of gene function, which are genetically stable (McCarty and Meeley, 2009). Moreover, transposon insertions are relatively easy to identify using molecular genetics and high-throughput sequencing. Nevertheless, early constructed *Mu*-based mutation libraries in maize (Bensen et al., 1995; May et al., 2003; Fernandes et al., 2004) had to deal with (1) inefficient recovery and identification of newly transposed insertions and (2) the challenge to discriminate between germinal and the high background of somatic insertions. Recent studies have overcome both of these challenges. The *Mu*-seq method (McCarty et al., 2013; this study) applies high-throughput NextGen sequencing of *Mu*-tag enriched replicated samples to identify germinal *Mu* insertions. More recently, Liang et al. (2019) achieved *Mu*-tag enrichment utilizing probe hybridization, coupled with high-throughput sequencing to efficiently recover *Mu* insertions in F_2-families of maize. To distinguish germinal from somatic insertions Liang et al. (2019) used a practical criterion (i.e. normalized *Mu*-flanking sequence tag read counts), in addition to the replicated samples as were used for the *UniformMu* analysis (McCarty et al., 2013).

Since the present European *BonnMu* and the North American *UniformMu* insertional library was constructed using the *Mu*-seq approach, *Mu* insertion sites of *BonnMu* were analyzed and compared in more detail with the *UniformMu* library. Due to multiple-round backcrossing between the mutagenic parents and the uniform W22 inbred line, the *UniformMu* lines captured on average five new germinal insertions (McCarty et al., 2005), already excluding about 40 ancestral insertions in the W22 genome (D. R. McCarty, personal communication; Supplemental Dataset S5). However, the *BonnMu* libraries presented here captured 37 germinal insertions in genes of the FGS per F_2-family on average (Supplemental Tables S2 and S3; Supplemental Dataset S3).

For about 50% of the *BonnMu* and *UniformMu*-tagged genes, we observed a gene-size– dependent number of insertions (Fig. 2B; Supplemental Fig. S2). On average, 2.5 insertions were detected per *Mu*-tagged gene, and we observed an association between insertion sites in

Downloaded from https://academic.oup.com/plphys/article/184/2/620/6117865 by ulbbonn user on 21 December 2021

B73 and W22 (i.e. hotspots of insertions), which are conserved among tagged genes of both genomes (Fig. 3). The availability of the W22v2 assembly (Springer et al., 2018) allows for comparisons with the B73v4 reference genome at multiple scales, such as *Mu* transposon composition. In total, 75% to 80% of the genes are present at syntenic locations in both maize inbred lines (Springer et al., 2018), making a comparison suitable. To provide additional support for the notion that insertion hotspots are conserved among tagged genes in B73 and W22, it will be essential to analyze and compare a larger W22-*Mu*-seq dataset.

Previous studies revealed that *Mu* insertions concentrate in genomic regions with epigenetic marks of open chromatin near the TSS (Dietrich et al., 2002; Liu et al., 2009; Springer et al., 2018). Indeed, *Mu* elements do exhibit a strong preference for 5′ UTRs of genes (Fig. 4; Dietrich et al., 2002; Liang et al., 2019). Even more precisely, we demonstrated that 50% of all *BonnMu* insertions are located within 135 bp of the TSS (Fig. 5). Pericentromeric regions, which are rich in heterochromatin, display low frequencies of *Mu* insertions and meiotic recombinations (Liu et al., 2009), suggesting an association of site selection for *Mu* insertions with chromatin structure. Indeed, locations for novel transposon insertions can be sensitive to chromatin structure (Naito et al., 2009; Sultana et al., 2017; Springer et al., 2018). For instance, DNA methylation patterns were analyzed by whole-genome bisulfite sequencing and chromatin accessibility throughout the W22 genome using a Micrococcal Nuclease assay (Springer et al., 2018). The enrichment of open chromatin, based on epigenetic marks, such as Cytosin methylation (CHH, where H = A, T, or C) and Micrococcal Nuclease treatment at the TSS and the 3′ UTR of genes (Regulski et al., 2013; Springer et al., 2018) corresponds with the majority of *BonnMu* insertions detected near the 5′ UTR region of genes (Figs. 4 and 5) and a smaller peak near the 3′ UTR (Fig. 4A). Hence, it is likely that chromatin structure plays a key role in site selection for *Mu* insertions. Further indirect support of the hypothesis that open chromatin drives *Mu* insertion sites preferences is shown in Figure 1. The high number of observed *Mu*-tagged genes overlapping among the libraries suggests that the genes are not randomly selected by *Mu*. To further test the hypothesis that open chromatic marks drives *Mu* insertion site preferences, chemical treatment to promote chromatin opening (Baubec et al., 2009) could be applied, thereby facilitating *Mu* insertions in heterochromatic (i.e. inaccessible regions). Another indirect strategy to support this hypothesis is to test if the subset of genes that is not yet *Mu* tagged in any of the published libraries belongs to transcriptionally silenced genes with densely methylated chromatin.

CONCLUSION

In summary, our publicly available European *BonnMu* insertional library provides a starting point for forward and reverse genetics studies in maize. To facilitate functional genomics study in maize, specific mutants and respective phenotypes of segregating F_2-families can be identified based on gene sequences of interest through MaizeGDB.org as described in Liu et al. (2016). Detected on the basis of albino and pale green leaf phenotypes, the mutation rate was ~13%, which is consistent with previously published mutation rates (Robertson, 1983), suggesting a high transposon activity in these stocks. The European *BonnMu* resource is still expanding to archive additional genes tagged by germinal *Mu* insertions. To accelerate the identification of *Mu* insertion sites of future *Mu*-seq libraries, we created a *Mu*-seq workflow utility (*Mu*WU), which is publicly available at https://github.com/Crop-Bioinformatics-Bonn/MuWU. Our *Mu*WU makes use of bioconda (Grüning et al., 2018) and snakemake (Köster and Rahmann, 2012) to provide an easy to install and completely automated framework for the detection and annotation of *Mu* insertion sites.

MATERIAL AND METHODS

Plant Material

The *Mu*-tagged maize (*Zea mays*) F_1-population was generated in the summer season of 2014 by crossing an active *Mu* line (Mu^4 per se; Robertson, 1983) into B73 at the experimental field station Campus Klein Altendorf (University of Bonn). The F_1-generation, which is heterozygous for the new *Mu* insertions, was self-pollinated during winter nursery 2014 to 2015 in Chile and in the summer season 2015 at Campus Klein Altendorf to produce segregating F_2-families.

Mu-seq Library Production

For construction of two multiplexed *Mu*-seq libraries we used 1,152 (2×576) F_2-families in the B73 background and followed the previously described *Mu*-seq approach (McCarty et al., 2013). Briefly, each of the two *Mu*-seq libraries contained 576 F_2-families, which were pooled according to a two-dimensional 24×24 grid design (Supplemental Table S1). Hence, the two *Mu*-seq libraries consisted of 1,152 ($2 \times 24 \times 24$) F_2-families. Construction of the *Mu*-seq libraries started with germination of 2×576 maize F_2-families (6 seeds per F_2-family) in paper rolls (Hetz et al., 1996). Sampling of leaf-tissue was carried out 10 to 12 d after germination, according to the 24×24 grid design, where each family contributed to one distinct column and one distinct row pool. In total, there were 24 row and 24 column pools, which created 48 multiplexed leaf samples. Depending on the germination rate, 3 to 6 seedlings were sampled per F_2-family. If less than 3 seedlings germinated, seeds of these F_2-families were regerminated to complement the number of seedlings. A probability was calculated of 99% (Supplemental Table S5) to obtain at least one mutant allele per tagged gene among the 3 to 6 germinated plants per F_2-family. To ensure the specific identification of heritable germinal insertions at intersections of rows and columns in the grid, leaf samples of row and column pools were taken from independent somatic cell lineages, that is, from alternate leaves of each seedling. By utilizing this approach, nonheritable, somatic insertions appeared only in a single axis of the grid and they were excluded from further analyses. Subsequently, gDNA was isolated from these 48 pools (Nalini et al., 2003) and randomly fragmented to a size of about 1 kb using 1 to 3 cycles of 3 min sonication using an Ultrasonication unit followed by 3 min rigorous vortexing. The single stranded overhang of the randomly sheared gDNA fragments was filled in by an enzyme mix to create blunt ends that enabled ligation of a double-stranded universal (U) adapter. Afterward, *Mu*-flanking amplicons were enriched by ligation-mediated PCR using a *Mu*-TIR-specific and a specific primer for the ligated U adapter (PCR I). Two subsequent rounds of PCR (PCR II and III) used nested TIR primers in combination with Illumina sequencing adapter primers to successively incorporate the adapters, necessary for sequencing of the fragments. To reduce the number of very short *Mu*-flanking fragments, PCR II

Downloaded from https://academic.oup.com/plphys/article/184/2/620/6117865 by ulbbonn user on 21 December 2021

products were purified to a size range of 300 to 800 bp on 1.2% (w/v) agarose gels. To avoid sample contamination, each of the 48 PCR II products was loaded on a separate gel. The final PCR III completed the required sequencing adapters and incorporated a 6-bp barcode allowing multiplexing of the 48 pools. Finally, each *Mu*-seq library was quantified using a Bioanalyzer, DNA 7500 chip (Agilent Technologies). The final concentrations of the two *Mu*-seq libraries were 73 and 102 nM. Paired-end sequencing (100 bp) of the multiplexed *Mu*-seq libraries was performed in two separate lanes of a HiSeq 2500 (Illumina) sequencer.

Identification of *Mu* Insertion Sites in B73

Adapters and TIR sequences were cut from both ends of the raw *Mu*-seq reads with cutadapt v2.3 (-e 0.2; Martin, 2011). Low-quality bases were filtered with Trimmomatic v0.36 (SLIDINGWINDOW:4:15; MINLEN:12; Bolger et al., 2014) and mapped to the maize reference genome AGPv4.36 (Jiao et al., 2017) with Bowtie2 v2.5.0 (Langmead and Salzberg, 2012). Duplicate reads were removed with the Picard MarkDuplicates tool v2.5.0 (Picard Toolkit, 2019: https://github.com/broadinstitute/picard). *Mu* insertion sites were identified using a customized Python script based on the characteristic 9 bp overlap of *Mu* insertion flanking sequences (McCarty et al., 2013). An insertion site was selected if the 9-bp overlap was supported by at least two reads on each side. To detect insertions that can be assigned to a specific maize F_2-family, all insertion sites with matching genomic coordinates in only one row and one column sample were selected. To identify *Mu* insertion sites inside or upstream of genes of the B73 Filtered Gene Set v4.36 (FGS; Jiao et al., 2017), the Bioconductor R packages IRanges v2.16 and GenomicRanges v1.36 (Lawrence et al., 2013), as well as ChIPpeakAnno v3.16.1 (Zhu et al., 2010) were subsequently used. Only *Mu* insertion sites located inside or upstream (≤500 bp) of FGS genes were considered. The Pearson correlation coefficient was calculated in R v3.5.2 to test the hypothesis that the number of *Mu* insertions increased with the length of the affected genes.

To determine the expected number of FGS genes being *Mu* tagged among *BonnMu* libraries and global sequence-indexed libraries (Fig. 1; Supplemental Table S4), a generalized linear model with a log-link and Poisson distribution was fitted under the assumption of independence between the three available *Mu*-seq resources using *proc genmod* in SAS 9.4 (SAS Institute). For each comparison, the model under independence comprised main effects for each of the compared resources. The expected and observed numbers of *Mu*-tagged FGS genes were compared using a χ^2-test for the two- or three-way interaction effect of the compared resources, respectively. Significant differences were determined at a significance level of $\alpha = 0.05$. Because it was not known which genome versions were used to align the reads of the *UniformMu* or the *ChinaMu* datasets, the reference genome AGPv4.36, used in this study, was applied. Hence, only 16,066 *UniformMu* genes and 20,133 *ChinaMu* genes (Fig. 1B) were considered for the calculation of expected values.

Identification of *Mu* Insertion Sites in W22

Single-end raw *Mu*-seq reads in W22 background (SRA accession SRP028545) were trimmed with cutadapt v2.3 (-e 0.2; Martin, 2011). The barcode and the TIR sequence were cut from the beginning of each read. Reads without TIR sequence were discarded and bases were cut from the end of the remaining reads with Trimmomatic v0.36 (Bolger et al., 2014) if they fell below a quality score of 20 (TRAILING:20; MINLEN:12). Reads were mapped to the maize. reference genome W22 v2.0 (Springer et al., 2018) with Bowtie2 v2.5.0 (Langmead and Salzberg, 2012), and duplicate reads were subsequently removed using the Picard MarkDuplicates tool v2.5.0 (Picard Toolkit, 2019: https://github.com/broadinstitute/picard). Reads that mapped to more than one position were excluded from the deduplicated SAM files and *Mu* insertion sites were identified using a customized Python script. Genomic coordinates were tagged as *Mu* transposon insertion sites if the alignment start of at least two reads matched at the same position. Insertion sites located inside W22 v2.0 genes (Springer et al., 2018) were identified with the R Bioconductor packages IRanges v2.16, GenomicRanges v1.36 (Lawrence et al., 2013), and ChIPpeakAnno v3.16.1 (Zhu et al., 2010), and all insertion sites located inside W22 genes were selected.

Analysis of *Mu* Insertion Sites in B73 and W22

To analyze hotspots of *Mu* insertion sites in B73 and W22, all insertion sites were sorted into corresponding bins. To this end, the length of each gene was divided into bins of 200 bp in size and the number of bins (i.e. hotspots) containing *Mu* transposon insertions was counted for each gene. To compare hotspots between B73 and W22, *Mu* flanking sequences in W22 were extracted (50 bp upstream – 50 bp downstream of the insertion site) using BEDTools v2.25.0 (Quinlan and Hall, 2010) and aligned to the maize. reference genome AGPv4.36 (Jiao et al., 2017) with Bowtie2 v2.5.0 (Langmead and Salzberg, 2012). Multireads (i.e. reads aligning to multiple genomic locations) were excluded and B73 coordinates of W22 reads were further analyzed to identify W22 *Mu* insertion sites within genes of the B73 FGS (Jiao et al., 2017) as described above. Identified insertion sites in W22 and B73 were filtered for endogenous *Mu* elements. To this end, matches between sequences of endogenous *Mu* elements (Springer et al., 2018) and *Mu* insertion flanking sequences (200 bp upstream – 200 bp downstream of the insertion site) were identified with BLAST v2.3.0 (Camacho et al., 2009). Matches that exceeded the defined threshold (pident ≥ 95; bitscore ≥ 50) were excluded from further analyses. Remaining *Mu*-tagged genes that were hit in both B73 and W22 were divided into 200-bp bins. The filtered *Mu* insertions were sorted into the corresponding bins. A χ^2-test ($\alpha \leq 0.01$) was performed on the confusion matrix given in Table 2 in R v3.5.2 to test the hypothesis that there was an association between hotspots in B73 and W22.

To analyze insertion site preferences of *Mu* transposons, the longest transcript for each affected gene in either B73 or W22 was selected and the cds was divided into deciles based on the B73 FGS (Jiao et al., 2017). Genes without annotated cds were excluded from further analyses. All identified insertion sites located in the cds, including endogenous *Mu* elements, were sorted into the corresponding cds decile. Insertion sites located within FGS genes, but up- or downstream of the cds, were counted as an insertion site in the 5′ UTR and 3′ UTR, respectively. To determine the distance from each *Mu* insertion site to the TSS, the longest transcript of each FGS gene was selected. For each *Mu* insertion site identified in B73 inside or upstream (≤500 bp) of B73 FGS genes, the absolute distance to both the cds start and, if available, the 5′ UTR start of the associated gene was calculated.

Accession Numbers

Raw sequencing data are stored at the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) with the accession number PRJNA608624.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Size distribution of 16,392 affected FGS genes in comparison to the size distribution of all 44,117 B73v4 genes.

Supplemental Figure S2. Number of affected W22 genes and associated number of *Mu* insertions plotted against the corresponding gene size.

Supplemental Figure S3. Number of regions (200 bp bins) in FGS genes preferentially targeted by *Mu* transposons.

Supplemental Figure S4. Affected genes sorted into gene sets according to their length (deciles).

Supplemental Table S1. Grid design for 24 x 24 F_2-families, i.e. 576 F_2-families per *Mu*-seq library.

Supplemental Table S2. Number of germinal *Mu* insertions in FGS genes per 576 F_2-families (*Mu*-seq 1).

Supplemental Table S3. Number of germinal *Mu* insertions in FGS genes per 576 F_2-families (*Mu*-seq 2).

Supplemental Table S4. Number of observed and expected *Mu*-tagged genes among sequence-indexed *Mu* insertional libraries (related to Fig. 1).

Supplemental Table S5. Calculated probability to obtain at least one mutant allele per tagged gene among the 3-6 germinated plants per F_2-family.

Supplemental Dataset S1. Number of raw read pairs per row and column sample sequenced per *Mu*-seq library with HiSeq2500 (Illumina) sequencer.

Supplemental Dataset S2. A total of 225,936 genomic insertion sites detected in the two *Mu*-seq libraries.

Downloaded from https://academic.oup.com/plphys/article/184/2/620/6117865 by ulbbonn user on 21 December 2021

Supplemental Dataset S3. A total of 41,086 *Mu* insertion sites detected in 16,392 genes of the FGS.

Supplemental Dataset S4. List of 43,943 *UniformMu* insertions within 16,090 genes in B73.

Supplemental Dataset S5. Canonical *Mu* elements in W22 and B73.

ACKNOWLEDGMENTS

We thank Hans-Peter Piepho (University of Hohenheim) for statistical assistance and Patrick S. Schnable (Iowa State University) for providing seeds of *Mu*[4] per se. We also appreciate the help of Peng Yu (University of Bonn) in this project. We thank Karl-Josef Wiesel (Campus Klein Altendorf, University of Bonn), Christa Schulz, and Helmut Rehkopf (University of Bonn) for technical assistance.

Received April 20, 2020; accepted July 27, 2020; published August 7, 2020.

LITERATURE CITED

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, et al (2003) Genome-wide insertional mutagenesis of *Arabidopsis thaliana*. Science **301:** 653–657

Baubec T, Pecinka A, Rozhon W, Mittelsten Scheid O (2009) Effective, homogeneous and transient interference with cytosine methylation in plant genomic DNA by zebularine. Plant J **57:** 542–554

Bennetzen JL (2000) Transposable element contributions to plant gene and genome evolution. Plant Mol Biol **42:** 251–269

Bensen RJ, Johal GS, Crane VC, Tossberg JT, Schnable PS, Meeley RB, Briggs SP (1995) Cloning and characterization of the maize *An1* gene. Plant Cell **7:** 75–84

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics **30:** 2114–2120

Brutnell TP (2002) Transposon tagging in maize. Funct Integr Genomics **2:** 4–12

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: Architecture and applications. BMC Bioinformatics **10:** 421

Candela H, Hake S (2008) The art and design of genetic screens: maize. Nat Rev Genet **9:** 192–203

Chandler VL, Hardeman KJ (1992) The *Mu* elements of *Zea mays*. Adv Genet **30:** 77–122

Dietrich CR, Cui F, Packila ML, Li J, Ashlock DA, Nikolau BJ, Schnable PS (2002) Maize *Mu* transposons are targeted to the 5′ untranslated region of the *gl8* gene and sequences flanking *Mu* target-site duplications exhibit nonrandom nucleotide composition throughout the genome. Genetics **160:** 697–716

Fernandes J, Dong Q, Schneider B, Morrow DJ, Nan GL, Brendel V, Walbot V (2004) Genome-wide mutagenesis of *Zea mays* L. using RescueMu transposons. Genome Biol **5:** R82

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: Where genetics meets genomics. Nat Rev Genet **3:** 329–341

Frey M, Stettner C, Gierl A (1998) A general method for gene isolation in tagging approaches: Amplification of insertion mutagenised sites (AIMS). Plant J **13:** 717–721

Grüning B, Dale R, Sjödin A, Chapman BA, Rowe J, Tomkins-Tinch CH, Valieris R, Köster J; Bioconda Team (2018) Bioconda: Sustainable and comprehensive software distribution for the life sciences. Nat Methods **15:** 475–476

Hetz W, Hochholdinger F, Schwall M, Feix G (1996) Isolation and characterization of *rtcs*, a maize mutant deficient in the formation of nodal roots. Plant J **10:** 845–857

Hirochika H, Guiderdoni E, An G, Hsing YI, Eun MY, Han CD, Upadhyaya N, Ramachandran S, Zhang Q, Pereira A, et al (2004) Rice mutant resources for gene discovery. Plant Mol Biol **54:** 325–334

Hochholdinger F, Wen TJ, Zimmermann R, Chimot-Marolle P, da Costa e Silva O, Bruce W, Lamkey KR, Wienand U, Schnable PS (2008) The maize (*Zea mays* L.) *roothairless3* gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield. Plant J **54:** 888–898

Hsing YI, Chern CG, Fan MJ, Lu PC, Chen KT, Lo SF, Sun PK, Ho SL, Lee KW, Wang YC, et al (2007) A rice gene activation/knockout mutant resource for high throughput functional genomics. Plant Mol Biol **63:** 351–364

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS, et al (2017) Improved maize reference genome with single-molecule technologies. Nature **546:** 524–527

Kolesnik T, Szeverenyi I, Bachmann D, Kumar CS, Jiang S, Ramamoorthy R, Cai M, Ma ZG, Sundaresan V, Ramachandran S (2004) Establishing an efficient Ac/Ds tagging system in rice: Large-scale analysis of Ds flanking sequences. Plant J **37:** 301–314

Köster J, Rahmann S (2012) Snakemake--a scalable bioinformatics workflow engine. Bioinformatics **28:** 2520–2522

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods **9:** 357–359

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ (2013) Software for computing and annotating genomic ranges. PLOS Comput Biol **9:** e1003118

Liang L, Zhou L, Tang Y, Li N, Song T, Shao W, Zhang Z, Cai P, Feng F, Ma Y, et al (2019) A sequence-indexed *Mutator* insertional library for maize functional genomics study. Plant Physiol **181:** 1404–1414

Lisch D (2002) *Mutator* transposons. Trends Plant Sci **7:** 498–504

Lisch D (2015) Mutator and MULE transposons. Microbiol Spectr **3:** MDNA3-0032-2014

Liu P, McCarty DR, Koch KE (2016) Transposon mutagenesis and analysis of mutants in *UniformMu* maize (*Zea mays*). Curr Protoc Plant Biol **1:** 451–465

Liu S, Yeh CT, Ji T, Ying K, Wu H, Tang HM, Fu Y, Nettleton D, Schnable PS (2009) *Mu* transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. PLoS Genet **5:** e1000733

Liu YG, Whittier RF (1995) Thermal asymmetric interlaced PCR: Automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics **25:** 674–681

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J **17:** 10–12

May BP, Liu H, Vollbrecht E, Senior L, Rabinowicz PD, Roh D, Pan X, Stein L, Freeling M, Alexander D, et al (2003) Maize-targeted mutagenesis: A knockout resource for maize. Proc Natl Acad Sci USA **100:** 11541–11546

McCarty DR, Latshaw S, Wu S, Suzuki M, Hunter CT, Avigne WT, Koch KE (2013) Mu-seq: Sequence-based mapping and identification of transposon induced mutations. PLoS One **8:** e77172

McCarty DR, Meeley RB (2009) Transposon resources for forward and reverse genetics in maize. In J.L. Bennetzen, and S. Hake, eds, Handbook of Maize. Springer, New York, pp 561–584

McCarty DR, Settles AM, Suzuki M, Tan BC, Latshaw S, Porch T, Robin K, Baier J, Avigne W, Lai J, et al (2005) Steady-state transposon mutagenesis in inbred maize. Plant J **44:** 52–61

McClintock B (1950) The origin and behavior of mutable loci in maize. Proc Natl Acad Sci USA **36:** 344–355

McClintock B (1951) Mutable loci in maize. Carnegie Institution of Washington Yearbook **50:** 174–181

Miyao A, Tanaka K, Murata K, Sawaki H, Takeda S, Abe K, Shinozuka Y, Onosato K, Hirochika H (2003) Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. Plant Cell **15:** 1771–1780

Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, Okumoto Y, Tanisaka T, Wessler SR (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature **461:** 1130–1134

Nalini E, Bhagwat SG, Jawali N (2003) A simple method for isolation of DNA from plants suitable for long term storage and DNA marker analysis. BARC Newsletter **249:** 208–214

Nestler J, Liu S, Wen TJ, Paschold A, Marcon C, Tang HM, Li D, Li L, Meeley RB, Sakai H, et al (2014) *Roothairless5*, which functions in maize (*Zea mays* L.) root hair initiation and elongation encodes a monocot-specific NADPH oxidase. Plant J **79:** 729–740

Portwood JL II, Woodhouse MR, Cannon EK, Gardiner JM, Harper LC, Schaeffer ML, Walsh JR, Sen TZ, Cho KT, Schott DA, et al (2019) MaizeGDB 2018: The maize multi-genome genetics and genomics database. Nucleic Acids Res **47**(D1): D1146–D1154

Downloaded from https://academic.oup.com/plphys/article/184/2/620/6117865 by ulbbonn user on 21 December 2021

Quinlan AR, Hall IM (2010) BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics **26:** 841–842

Regulski M, Lu Z, Kendall J, Donoghue MTA, Reinders J, Llaca V, Deschamps S, Smith A, Levy D, McCombie WR, et al (2013) The maize methylome influences mRNA splice sites and reveals widespread paramutation-like switches guided by small RNA. Genome Res **23:** 1651–1662

Robertson DS (1978) Characterization of a *mutator* system in maize. Mutat Res **51:** 21–28

Robertson DS (1983) A possible dose-dependent inactivation of *Mutator* (*Mu*) in maize. Mol Gen Genet **191:** 86–90

Schnable JC, Freeling M (2011) Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS One **6:** e17855

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science **326:** 1112–1115

Settles AM, Latshaw S, McCarty DR (2004) Molecular analysis of high-copy insertion sites in maize. Nucleic Acids Res **32:** e54

Springer NM, Anderson SN, Andorf CM, Ahern KR, Bai F, Barad O, Barbazuk WB, Bass HW, Baruch K, Ben-Zvi G, et al (2018) The maize W22 genome provides a foundation for functional genomics and transposon biology. Nat Genet **50:** 1282–1288

Stern DB, Hanson MR, Barkan A (2004) Genetics and genomics of chloroplast biogenesis: Maize as a model system. Trends Plant Sci **9:** 293–301

Sultana T, Zamborlini A, Cristofari G, Lesage P (2017) Integration site selection by retroviruses and transposable elements in eukaryotes. Nat Rev Genet **18:** 292–308

Toki S, Hara N, Ono K, Onodera H, Tagiri A, Oka S, Tanaka H (2006) Early infection of scutellum tissue with Agrobacterium allows high-speed transformation of rice. Plant J **47:** 969–976

van Enckevort LJ, Droc G, Piffanelli P, Greco R, Gagneur C, Weber C, González VM, Cabot P, Fornara F, Berri S, et al (2005) EU-OSTID: A collection of transposon insertional mutants for functional genomics in rice. Plant Mol Biol **59:** 99–110

Wang N, Long T, Yao W, Xiong L, Zhang Q, Wu C (2013) Mutant resources for the functional analysis of the rice genome. Mol Plant **6:** 596–604

Xu C, Tai H, Saleem M, Ludwig Y, Majer C, Berendzen KW, Nagel KA, Wojciechowski T, Meeley RB, Taramino G, et al (2015) Cooperative action of the paralogous maize lateral organ boundaries (LOB) domain proteins RTCS and RTCL in shoot-borne root formation. New Phytol **207:** 1123–1133

Zhang Q (2007) Strategies for developing green super rice. Proc Natl Acad Sci USA **104:** 16402–16409

Zhu LJ, Gazin C, Lawson ND, Pagès H, Lin SM, Lapointe DS, Green MR (2010) ChIPpeakAnno: A bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics **11:** 237

Downloaded from https://academic.oup.com/plphys/article/184/2/620/6117865 by ulbbonn user on 21 December 2021

Check for updates

Experimental platforms for the investigation of spatiotemporal patterns in the rhizosphere—Laboratory and field scale

Doris Vetterlein[1,2]*, Eva Lippold[1], Susanne Schreiter[1], Maxime Phalempin[1], Toni Fahrenkampf[1], Frank Hochholdinger[3], Caroline Marcon[3], Mika Tarkka[4,5], Eva Oburger[6], Mutez Ahmed[7], Mathieu Javaux[8,9], and **Steffen Schlüter[1]**

[1] Department of Soil System Science, Helmholtz Centre for Environmental Research–UFZ, Theodor-Lieser-Str. 4, 06120 Halle/Saale, Germany
[2] Soil Science, Martin-Luther-University Halle-Wittenberg, Von-Seckendorff-Platz 3, 06120 Halle/Saale, Germany
[3] Institute for Crop Science and Resource Conservation, Crop Functional Genomics, University of Bonn, Friedrich-Ebert-Allee 144, 53113 Bonn, Germany
[4] Department of Soil Ecology, Helmholtz Centre for Environmental Research–UFZ, Theodor-Lieser-Str. 4, 06120 Halle/Saale, Germany
[5] German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5, 04103 Leipzig, Germany
[6] Institute of Soil Research, Department of Forest and Soil Sciences, BOKU–University of Natural Resources and Life Sciences, Konrad-Lorenz Strasse 24, 3430 Tulln, Austria
[7] Chair of Soil Physics, Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Universitätsstr. 30, 95447 Bayreuth, Germany
[8] Université catholique de Louvain, Earth and Life Institute, Croix du Sud L7.05.02, 1348 Louvain-la-Neuve, Belgium
[9] Agrosphere (IBG3), Forschungszentrum Juelich GmbH, 52425 Juelich, Germany

Abstract

The numerous feedback loops between roots, microorganisms, soil chemical and physical properties, and environmental variables result in spatial parameter patterns which are highly dynamic in time. In order to improve our understanding of the related rhizosphere processes and their relevance at the soil–plant system scale, experimental platforms are required. Those platforms should enable (1) to relate small scale observations (nm to dm) to system behaviour, (2) the integration of physical, chemical and biological sampling approaches within the same experiment, and (3) sampling at different time points during the life cycle of the system in question. Here we describe what requirements have to be met and to what extent this has been achieved in practice by the experimental platforms which were set up within the framework of DFG priority programme 2089 "Rhizosphere Spatiotemporal Organisation—a key to rhizosphere functions". It is discussed to what extent theoretical considerations could be accommodated, in particular for the comparison across scales, *i.e.*, from laboratory to field scale. The latter scale is of utmost importance to overcome the trade-off between fraction of life cycle covered and the avoidance of unrealistic root length densities.

Key words: destructive sampling / rhizosphere / root / undisturbed sampling / X-ray CT / *Zea mays*

Accepted July 17, 2020

Supporting Information available online

1 Introduction

The processes at the soil–root interface are key to unraveling plant resource acquisition, soil carbon storage and soil structure formation as well as the communication of plants with the soil microbiome which is thought to be the basis for plant health (*Hinsinger* et al., 2009; *Bardgett* and *van der Putten*, 2014; *Keiluweit* et al., 2015; *Roose* et al., 2016; *York* et al., 2016).

The numerous feedback loops between roots, microorganisms, soil chemical and physical properties, and environmental variables result in spatial parameter patterns which are highly dynamic in time. It was recently hypothesized that such patterns are the result of self-organization and that they are highly relevant for understanding the properties of soil–plant systems such as stability under disturbances and change (*Vetterlein* et al., 2020). A detailed discussion of current knowledge gaps in rhizosphere research and how applying the concept of self-organization as a system approach can contribute to advance rhizosphere research is found in *Vetterlein* et al. (2020). Here we focus on the experimental prerequisites required to follow such a system approach. The experimental platforms should enable to relate small scale observations (nm to dm) to system behavior, the integration of physical, chemical and biological sampling approaches within the same experiment and, ideally, sampling at different time points and at selected sites during the life cycle of the system in question (*Oburger* and *Schmidt*, 2016; *Roose* et al., 2016).

In the following, we describe what requirements have to be met by joined experimental platforms, and provide the experi-

* Correspondence: D. Vetterlein; e-mail: doris.vetterlein@ufz.de

© 2020 The Authors. *Journal of Plant Nutrition and Soil Science* published by Wiley-VCH GmbH www.plant-soil.com
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

mental details of a laboratory and a field platform that were established within the framework of DFG priority programme 2089 "Rhizosphere Spatiotemporal Organisation—a key to rhizosphere functions". In retrospect, we discuss the outcomes of these approaches and sum up in how far theoretical considerations could be accommodated.

Specifically, the following points will be addressed and respective requirements derived:

- Gradients of interest and associated sampling strategies
- Relevance for system behavior
- Integration of physical, biological and chemical sampling strategies within the same experiment
- Suitable drivers for challenging spatiotemporal rhizosphere patterns
- Boundary conditions *suitable for rhizosphere studies—avoid luxury supply*
- Homogeneous initial conditions—*work load versus statistical requirements*

1.1 Gradients of interest and associated sampling strategies

Gradients can be observed along roots, *i.e.*, with changing root ontogeny and topology or with increasing distance from the root surface (*Hinsinger* et al., 2009). These gradients constitute the rhizosphere, which can be defined in the most general terms as the zone of soil affected by roots. However, it is not straightforward to sample the rhizosphere. There are operationally defined ways of sampling the rhizosphere, which are applied if the methods require a certain amount of homogenized soil material for analyses. Here, the rhizosphere soil is the amount of soil adhering to the roots after extracting them from bulk soil. It is thereafter removed from the root surface by brushing or washing. For these sampling methods the rhizosphere soil sample is a mixed sample across the whole root system. It does not refer to a certain distance from the root surface, but rather to the binding strength of soil to the root surface. The properties of such "rhizosphere samples" are conventionally compared to the properties of the bulk soil samples, *i.e.*, the soil remaining after the roots with adhering soil were removed.

Alternatively, *in situ* approaches can be applied, for which the spatial context between root and soil is not altered by sampling and the distance between sampling location and root surface is known. This is the case for all imaging approaches, whether they are 2D, requiring a transparent access window, or 3D, requiring a specific geometry. For the latter, we can distinguish between completely non-invasive, like intact roots within soil columns being scanned during growth, and undisturbed soil columns with intact root and soil structure extracted from the field. All *in situ* approaches provide information on spatial gradients without predefining the extent of the rhizosphere. 2D approaches come with the risk of confounding diffusion in soil with diffusion in the contact medium between the soil and the measurement device. Moreover, roots may be altered by the growth along a solid plate. 2D approaches can however easily address root topology (root branching pattern), whereas 3D approaches mostly only provide root architecture.

A third category of sampling approaches aims at obtaining spatial information from destructive samples by carefully dissecting volume samples to identify a root and physically sampling at increasing distance from the root channel (*Barej* et al., 2014). Similarly, spatial resolution for destructive sampling can be obtained by extracting such samples along roots after removing the transparent access window.

Finally, there is a range of sampling and monitoring devices (*e.g.*, micro suction cups, tensiometers, TDR moisture sensors) which can be installed in the soil at a known position from the root surface or from a compartment within which root growth is constrained (*Vetterlein* and *Jahn*, 2004). The results obtained can be then interpreted in relation to the distance from the root surface or to the compartment.

Requirement 1: Allow access for a range of non-invasive and destructive sampling approaches.

1.2 Relevance for system behaviour

Frequently, physiological properties are investigated under controlled conditions in the laboratory to exclude uncertainties introduced by fluctuating environmental parameters (*e.g.*, temperature, light, water availability, pests). As a result, most of our current knowledge on rhizosphere processes is derived from young plants two to six weeks of age. Growth limitation within small pots is recognized as a potential artifact (*Passioura*, 2006). Most laboratory experiments show high root length densities and hence are not comparable to the field situation in terms of potential pay-off of resource acquisition strategies.

Requirement 2: Run laboratory (short term) and field (growing season) experiment with similar treatments.

1.3 Integration of physical, biological and chemical sampling strategies within the same experiment

For many research questions, it is indispensable to work with the same sample, or at least to sample the same system in order to advance our understanding. This is true whether one wants to relate the composition of the rhizosphere microbiome to the release of root exudates, the extent of P depletion zone to root age, or the root water uptake rate to aquaporin expression, to name just a few examples. If sampling techniques are mutually exclusive, as is the case for destructive sampling of rhizosphere microbiome and sampling of root exudates requiring an intact root system, at least parallel setups should be used.

Partly, subsampling strategies can be used to overcome such difficulties; this is particularly promising for combining non-invasive methods (X-ray computed tomography—X-ray CT, Magnetic resonance imaging—MRI) with destructive sampling approaches (plant gene expression, microbiome analy-

© 2020 The Authors. *Journal of Plant Nutrition and Soil Science* published by Wiley-VCH GmbH

15222624, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jpln.202000079 by Universitats und Landesbibliothek B, Wiley Online Library on [19/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

sis) or *in situ* sample collection (soil solution sampling). Subsampling can also be used to combine analyses at different scales like detailed chemical mapping at the nanometer to the micrometer scale with the whole plant nutrient uptake.

Requirement 3: Standardized experimental protocols. Non-invasive and in-situ measurements as reference points for destructive sampling.

1.4 Suitable drivers for challenging spatiotemporal rhizosphere patterns

In order to motivate scientists of different disciplines to find a common specific experimental platform, drivers expected to fundamentally alter physical, chemical and biological parameter spatial patterns in the rhizosphere have to be identified.

The most prominent soil property affecting water and nutrient fluxes and hence parameter spatial patterns is texture (*Jarvis*, 2007). Indeed, textural fractions shape the pore network as well as the number of sorption sites relevant for chemical gradients to develop (*Tinker* and *Nye*, 2000).

For root traits, the spatial extent of patterns, such as depletion zones for a certain chemical element or the zone impacted by root exudation, is expected to vary with the presence of root hairs. Fortunately, there are root hair mutants available for some plant species that facilitate questioning their importance for the establishment of rhizosphere gradients.

Requirement 4: Drivers of general interest, one related to soil property, one related to plant property.

1.5 Boundary conditions suitable for rhizosphere studies – avoid luxury supply

Rhizosphere processes are in most cases associated with an investment of carbohydrates (root growth, specific root traits, release of substances, symbioses, expression of genes) in order to improve resource acquisition. Such investment will only be advantageous if resource availability is limiting growth. Hence, the system boundary conditions have to be selected accordingly, *i.e.*, avoiding over-supply and luxury consumption of nutrients, in particular P.

Requirement 5: Marginal nutrient supply.

1.6 Homogenous initial conditions – work load *versus* statistical requirements

Ideally, chemical, physical and biological parameters should initially be homogeneous at all scales, as unwanted heterogeneity would modify rooting patterns at the profile scale but would also be detrimental to the interpretation of the observed gradients at the nano or micrometer scale. Practically, such homogeneous conditions are unrealistic. One way of coping with heterogeneity is to increase the amount of replicates (independent samples, biological replicates), and/or number of sampling points within one replicate (technical replicates, pooled subsamples). The limit to these workarounds is not just the work load but likewise the large soil area and volume consumption.

Requirement 6: Homogenous initial conditions.

In the following, we describe the experimental platforms in the laboratory and the field of the DFG priority programme 2089 "Rhizosphere Spatiotemporal Organisation—a key to rhizosphere functions". For the design of these platforms, we tried to fulfill the requirements 1–6. In this review, we provide the design, initial chemical and physical properties of the substrates as well as climatic conditions. We further discuss as to how far requirements 1–6 could be fulfilled for laboratory *versus* field scale and provide a first comparison of plant growth at laboratory and field scales for the common growth stage (BBCH 14).

2 Material and methods

2.1 Experimental design

The laboratory based experiments will be denoted as soil column experiments (SCE) throughout this review, the field based experiment as soil plot experiment (SPE). For both scales, the basic design is a two factorial, randomized block design with six replicates (Fig. 1). The term replication here refers to six individual soil columns (SCE) or six individual field plots (SPE). Factor one is substrate with two levels [loam (L), sand (S)]. Factor two is the *Zea mays* genotype with two levels comprising wild type (WT) and a root hair defective mutant (*rth3*).

2.2 Genotypes

For the experiments, the *Zea mays* root hair defective mutant *rth3* and the corresponding wild-type siblings were selected (*Wen* and *Schnable*, 1994).

The monogenic mutant *rth3* is transposon induced and shows normal root hair initiation but disturbed elongation. The mutant does not display an aberrant shoot phenotype, but yield is reduced by 20 to 40% compared to the wildtype (*Hochholdinger* et al., 2008). The mutated gene encodes a GPI-anchored COBRA-like cell wall protein RTH3 that is involved in the organization of the synthesized cellulose (*Hochholdinger* et al., 2018). The *rth3* mutants used in these experiments are genetically highly homozygous because they have been backcrossed to the inbred line B73 for > 8 generations. Seeds were multiplied at the experimental station Endenich of the Faculty of Agriculture of the University of Bonn.

2.3 Substrate

The substrate loam (L) was obtained by excavating 700 t of a haplic Phaeozem soil (from 0 to 50 cm depth) in Schladebach, Germany (51°18′31.41′′ N; 12°6′16.31′′ E). The haplic Phaeozem has been under agricultural use until excavation. Before the excavation, the area was planted with oilseed rape. The excavated material was sieved at a gravimetric water content of 14% utilizing a heavy duty double deck

© 2020 The Authors. *Journal of Plant Nutrition and Soil Science* published by Wiley-VCH GmbH

15222624, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jpln.202000079 by Universitats und Landesbibliothek B, Wiley Online Library on [19/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

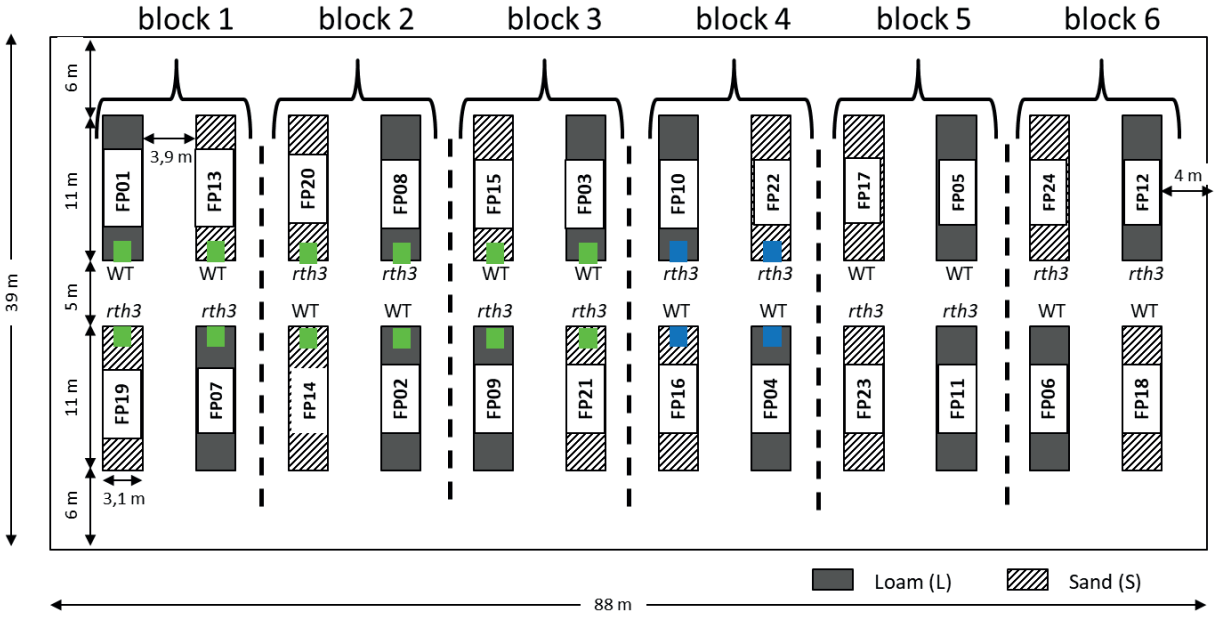


Figure 1: Soil plot experiment established as randomized block design with 6 replicates per treatment. Factor one: substrate with two levels [Loam (L), Sand (S)]; factor two: plant genotype with two levels [*Zea mays* B73 wild type (WT), respective hairless mutant (*rth3*)]. Position of root windows and monitoring devices for water relations are indicated in green and blue, respectively.

vibrating screen (Keestrack Combo; 3.2 m^2 area per deck). An upper level was mounted with a 40 mm mesh; a lower level had a 20 mm mesh. The throughput was in the range of 15 tons substrate per hour. Out of the three fractions obtained, the coarse one (> 40 mm) was discarded (about 5% of the total), and the medium one was added back to the ongoing sieving process. The whole amount was sieved within one week and stored by wheel loaders under a carport sheltered from rain. Setting up the field plots started right after the sieving operation in October 2018.

The substrate sand (S) was obtained by repeated mixing and sieving of (L) with quartz sand (550 t, WF 33, Quarzwerke Weferlingen, Germany). This procedure was chosen as no suitable sandy substrate (no carbonate, dominated by medium sand fraction, clay in the range of 5%) was available within a distance < 80 km (transport costs) from the research station. Initial mixing was conducted with wheel loaders based on the volume of their buckets (about 9:1). For homogenization of the pre-mixed material a drum screening plant (20 mm mesh) was used. Within one week, about 600 t of substrate (S) were produced and stored protected from rain, in the same fashion as for the loam substrate.

Stocks of initial material loam and quartz sand were kept for the SCE. These were stored in plastic containers (610 L, Kiga Kunststofftechnik GmbH) in a shelter protected from rain and sun.

2.4 Field plot and soil column design

Field plot experiment (SPE) was established in the research station Bad Lauchstädt, Germany (51°22′0′′ N, 11°49′60′′E). Individual field plots (11 × 3.1 m) were excavated to a depth of 1 m (Fig. 2a). Vertical side walls were covered with a root barrier (420 Xavan® 420 g/m^2, RootBarrier BV, Netherlands). The bottom of the plots was filled with a 25 cm gravel layer (0/32) and a drainage textile was then placed on top of the gravel (Fig. 2b). Substrates were then filled up to the original soil surface with a thickness of 75 cm (Fig. 2a). For weed growth prevention, the plots were covered thereafter with a nonwoven fabric made of 100% continuous polypropylene filaments (Plantex®Gold 125 g/m^2, DuPont, Luxembourg). The prevention canvas is permeable to air and water. Plots were established in October and November 2018 and remained covered until April 2019.

Individual soil columns for the SCE consist of an acrylic glass tube (25 cm height, 7 cm inner diameter). A nylon mesh (30 μm mesh size) is placed at the bottom of the column in order to retain the soil. The columns were then filled up to 23 cm height with the investigated substrates. With such a set-up, the volume available for plant growth is 885 cm^3.

2.5 Sieving and packing

For field plots treatment (L), loam was gradually placed in layers of 15 cm by a wheel loader, evened out with wheel loader bucket and compacted with a vibrating plate (weight 70 kg, support pressure 8 N/cm^2, reduced speed). This procedure proved to be suitable in a pilot experiment, as X-ray CT scans of extracted undisturbed soil cylinders showed no layering (Fig. 3). For treatment (S), sand was packed similarly; however, no vibrating plate was used.

For SCE, the loam with a moisture content of 10% is sieved step by step to < 4 mm, < 2 mm, and finally < 1 mm. The substrate sand corresponds to a mixture of 16.7% loam and 83.3% quartz sand, and it is established for each individual experiment (Tab. 2).

© 2020 The Authors. *Journal of Plant Nutrition and Soil Science* published by Wiley-VCH GmbH

15222624, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jpln.202000079 by Universitats und Landesbibliothek B, Wiley Online Library on [19/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

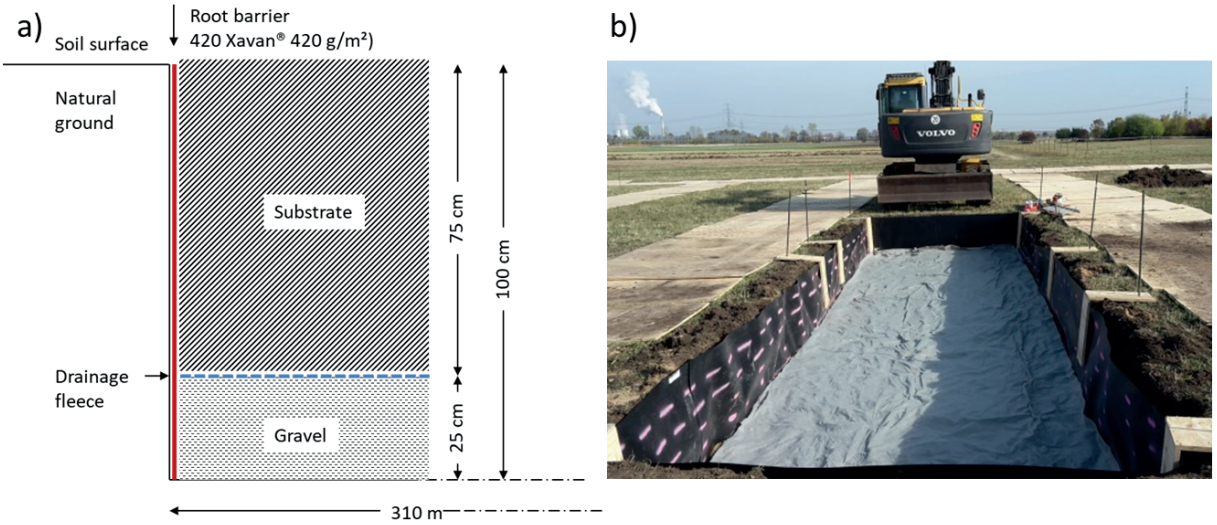


Figure 2: (a) Field plot cross section with root barrier and drainage layer, (b) field plot during establishment, drainage fleece is still visible.

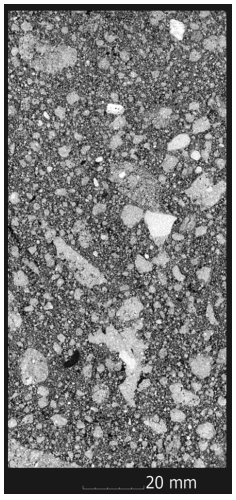


Figure 3: X-ray CT scan of an undisturbed soil column (20 cm height, 10 cm diameter) extracted from the field in a pilot study assessing different packing options. The substrate loam was packed with a vibrating plate in layers of 15 cm. Boundary between two subsequent layers would be expected halfway up the image.

Columns are packed carefully avoiding particle sorting by introducing a coarse sieve (4 mm) during filling which is continuously moved laterally. Compaction is achieved by "stamping" the entire column on a flat surface. For details regarding the packing of the columns, see https://www.ufz.de/spp-rhizosphere/index.php?en=43229.

2.6 Fertilization

In order for investment into resource acquisition to pay-off, the aim was to achieve a nutrient level which is in the range between slightly nutrient deficient to adequate nutrition for the WT genotype. Substrate as a driver was chosen primarily to introduce differences in water and nutrient transport; hence, differences in nutrient availability should ideally be compensated by fertilization.

Pre-trials were run with different doses of NPK application to achieve similar shoot growth rates for the WT on both substrates in column experiments, without visible symptoms of nutrient deficiency. These trials were associated with tissue analyses and the determination of plant available P and K in the substrates.

As a result N, P, K, and Mg were added at a dose twice as high in (S) compared to (L), and Ca as well as micronutrients were only applied to (S) (Tab. 1). This was the case for SCE and likewise for SPE.

The type of fertilizer differs between SCE and SPE as for column experiment laboratory grade chemicals of high purity are used, while for the field conventional fertilizers were applied (Tab. 1). For SCE, fertilizers were mixed with the soil prior to filling the columns, but for SPE, fertilizers were surface applied, 50% prior to seeding (24/04/2019), and the remaining 50% after first sampling (14/06/2019).

2.7 Agronomic measures – growth conditions

For field plots maize was sown (24/04/2019) to a depth of 5 cm by hand using a V-shaped planting device for single grain seeding. Distance within row was 20 cm, between rows 45 cm. This resulted in six rows with 54 plants each and a planting density of 9.5 plants m^{-2}, corresponding to a soil volume per plant of 78.935 dm^3 down to a depth of 75 cm.

During the first year, no pesticides were applied and weeding was done by hand. No heavy machinery was allowed to pass over the plots to avoid modification of soil structure.

The climatic conditions at the research station Bad Lauchstädt (average for the years 2008–2018) are depicted in Fig. 4 (Meteorological data of Bad Lauchstädt, Helmholtz Centre for Environmental Research–UFZ, Department of Soil System Science).

© 2020 The Authors. *Journal of Plant Nutrition and Soil Science* published by Wiley-VCH GmbH

15222624, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jpln.202000079 by Universitats und Landesbibliothek B, Wiley Online Library on [19/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1: Substrate specific fertilization in soil column experiments (SCE) and soil plot experiments (SPE). Data for SPE are for the first year after establishment of plots. 50% of the fertilizer was applied prior to seeding (24/04/2019), 50% after first sampling (14/06/2019).

(a) SCE

Substrate	Nutrient	Application rate (mg nutrient kg^{-1})	Type
Loam	N	50	NH_4NO_3
	P	40	$CaHPO_4$
	K	50	K_2SO_4
	Mg	25	$MgCl_2 \times 6\ H_2O$
Sand	N	100	NH_4NO_3
	P	80	$CaHPO_4$
	K	100	K_2SO_4
	Mg	50	$MgCl_2 \times 6\ H_2O$
	Ca	100	$CaSO_4 \times 2\ H_2O$
	Mn	3.25	$MnSO_4 \times H_2O$
	Zn	0.79	$Zn(NO3)_2 \times 4\ H_2O$
	Cu	0.5	$CuSO_4 \times 5\ H_2O$
	B	0.17	H_3BO_3
	Fe	3.25	Fe-EDTA

(b) SPE

Substrate	Nutrient	Application rate (kg nutrient ha^{-1})	Type
Loam	N	50	Calcium ammonium nitrate
	P	12	Triple superphosphate
	K	50	60s corn potash
	Mg	18	Epsom salt
	Ca	27	(applied with other fertilisers)
Sand	N	100	Calcium ammonium nitrate
	P	24	Triple superphosphate
	K	100	60s corn potash
	Mg	33	Epsom salt
	Ca	52	(applied with other fertilisers)
	Micronutrients	100	Excello 331 Spezial

Weather data are available from the DWD weather station which is located at the research station Bad Lauchstädt (ID 02878), https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/).

For column experiments, the maize seeds are surface sterilized (10% H_2O_2) and sown to a depth of 1 cm. Surface is covered with inert quartz gravel to reduce evaporation. Columns are carefully watered from top and bottom to an average volumetric water content of 22% for loam and 18% for sand. Fluctuation of water content is low as watering intervals are shortened as plant transpiration increases. Plants are grown at 22°C during the day and 18°C at night with 12 h light-period, 350 μM m^{-2} s^{-1} photosynthetically active radiation in a climate chamber. Relative humidity is kept constant at 65%. Growth duration in SCE is 21 d, *i.e.*, harvest is conducted on day 22 after seeding.

2.8 Definition of growth stages

As a common definition of growth stages for the field, BBCH-scale (*Bleiholder* et al., 2001) is used for defining sampling points in time. BBCH 14 (four leaves unfolded) was selected

© 2020 The Authors. *Journal of Plant Nutrition and Soil Science* published by Wiley-VCH GmbH

15222624, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jpln.202000079 by Universitats und Landesbibliothek B, Wiley Online Library on [19/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

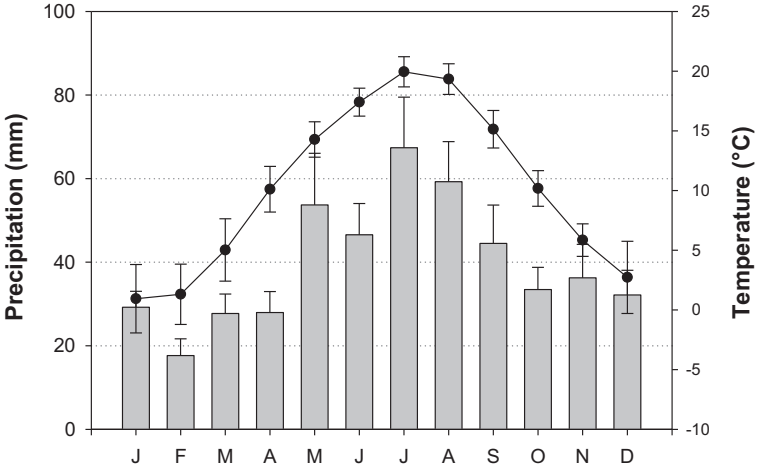


Figure 4: Mean temperature (symbols) and precipitation (bars) for the field station Bad Lauchstädt for 2008–2018. Source: Meteorological data of Bad Lauchstädt, Helmholtz Centre for Environmental Research–UFZ, Department of Soil System Science.

as the first sampling point as this corresponds to the developmental stage achieved after 21 d in SCE. BBCH 19 (nine or more leaves unfolded) was selected as second time point representing exponential growth. BBCH59 (end of tassel emergence) was selected as time point during the transition from vegetative to generative growth and BBCH 83 (early dough) as a growth stage representing ripening phase.

2.9 Installations for sampling and monitoring

2.9.1 Soil and plant water status monitoring (Fig. 5)

Sensors for the measurement of soil temperature, water-content (TEROS 10; Meter Group AG) and matric potential (T5 and TEROS 21; Meter Group AG) were installed at a depth of 10, 20, 40, and 60 cm at three different positions in relation to the planted maize rows. T5 is used to acquire accurate measurements of soil matric potential in the wet range, whereas TEROS 21 provides reliable measurements in the dry range. In addition, four plants per field were equipped with leaf psychrometers (PSY1; ICT International, Armidale, NSW, Australia) and sap flow sensors (Dynamax Inc., Houston, TX) to monitor respectively plant water status (as ψ_{leaf}) and transpiration. Measured data are logged every 10 min and transmitted to University of Bayreuth, which are responsible for the monitoring network. Due to high costs of installation only one out of the six blocks of SPE (Fig. 1) is equipped with these sensors.

2.9.2 Root windows (Fig. 6)

Inspired by the installation of root observation windows in forest sites (*Marschner* et al., 1991) Oburger et al. (pers. comm., manuscript with construction details in preparation) developed new root windows providing higher flexibility for obtaining partial access to roots growing along the transparent plane. These windows with a width of 60 cm and a depth of 60 cm were installed in three out of six field blocks (Fig. 1).

2.9.3 Isotope labelling (Fig. 7)

Stable isotope probing is used increasingly in different disciplines to trace uptake or release of substances and their subsequent distribution in functional networks (*Heinrich* et al., 2015). To enable such investigations in the soil plot experiments, an area of 70 × 120 cm was reserved within each plot for the three growth stages BBCH 14, 19, and 59. To trace nitrogen uptake, ^{15}N was applied to the soil surface as KNO_3 dissolved in water 60 h prior to sampling. The amount of water was calculated to allow infiltration to at least 0–10 cm. Application rates of 0.1, 0.2, 0.4 g m^{-2} were chosen at BBCH 14, 19, 59, respectively (99% ^{15}N KNO_3, Euriso-Top GmbH, Germany).

To trace the fate of assimilated carbon in the rhizosphere, ^{13}C pulse labeling was conducted on the same subplots 24 h prior to sampling. Gas tight chambers (70 × 120 cm, covering nine plants) with increasing height were set up and ^{13}C-CO_2 (99% ^{13}C Na_2CO_3, Euriso-Top GmbH, Germany) was released from Na_2CO_3 by adding sulfuric acid to the initial solution following a protocol adapted from *Heinrich* et al. (2015). Two short term pulses were applied over a 4-h period from 9 am to 1 pm. 8, 16, 32 g of $Na_2{}^{13}CO_3$ tracer were applied within the 4-h period to the nine plants in the chambers for BBCH 14, 19, and 59, respectively. Soil surface was sealed during ^{13}C labeling with a plastic foil to avoid direct gas exchange (Fig. 7).

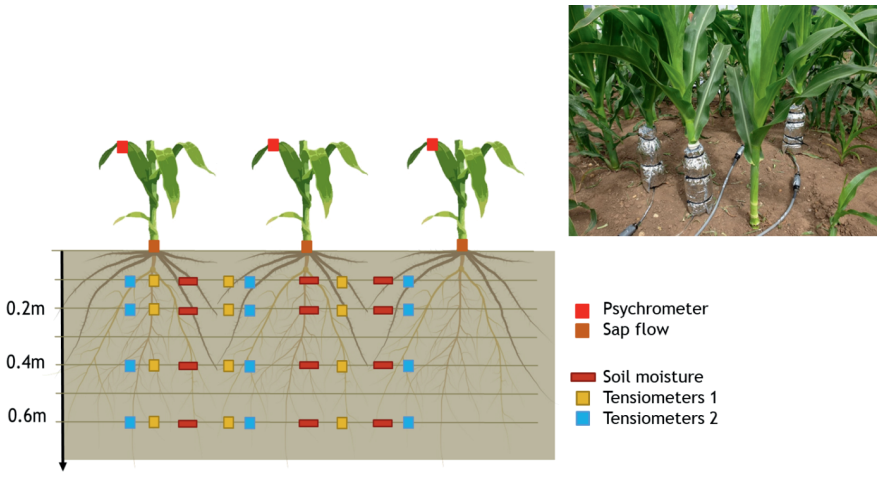


Figure 5: Sensor installation for the measurement of soil temperature, water content (TEROS 10), and matric potential (T5 and TEROS 21) at different depths and different positions in relation to the planted maize rows. In addition, plant water status is monitored with psychrometers (PSY1) and sap flow sensors (Dynamax TX).

© 2020 The Authors. *Journal of Plant Nutrition and Soil Science* published by Wiley-VCH GmbH

15222624, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jpln.202000079 by Universitats und Landesbibliothek B, Wiley Online Library on [19/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Figure 6: Root windows allowing visual monitoring, as well as repeated and non-destructive access to field-grown maize roots during the entire vegetation period. Idea and design: Eva Oburger, BOKU—University of Natural Resources and Life Sciences; Technical construction: Gottfried Wieshammer (gottfried.wieshammer@aon.at), Technisches Büro für Bodenkultur, Austria.

2.10 Characterization of initial soil conditions

Initial chemical characteristics were determined for SPE after establishment of plots in November 2018. For each plot, a bulk sample from 5 to 10 cm depth was obtained by mixing material collected at nine points per plot. Samples were air dried, homogenized and sieved to < 2 mm prior to chemical analyses.

2.10.1 Chemical analyses (Tab. 2, Tab. S1)

Soil pH was measured using a 0.01 M $CaCl_2$ solution (soil:solution ratio of 1:2.5). Total contents of elements were determined by X-ray fluorescence analysis (Spectro XEPOS HE, SPECTRO Analytical Instruments and S4 PIONEER, Bruker-AXS). Cation exchange capacity was determined with ammonium acetate method at pH 7 according to Lavkulich (*Gregorich* and *Carter*, 2007). Carbonate was below detection limit, hence, organic carbon (C_{org}) contents and total nitrogen (N_t) were determined using a CNS analyzer (Vario Max CN Element Analyzer, Hanau, Germany) without further pretreatment. Plant available P and K were determined by CAL-method (*Schüller*, 1969). Iron-oxides were characterized by oxalate (Fe_o) and dithionate (Fe_d) extraction (*Mehra* and *Jackson*, 1960; *Schwertmann*, 1964).

Mineral nitrogen (N_{min}) was analyzed for the field plots in April 2019 (NO_3-N and NH_4-N according to VDLUFA-Methods).

2.10.2 Physical analyses

2.10.2.1 Soil texture

Texture analysis was carried out according to International Standards (*ISO 11277*, 1998; *ISO 11277*, 2002). In brief, samples were air-dried and sieved down to 2 mm. Soil organic matter was oxidized with hydrogen peroxide solution (30 Vol%). A solution of tetra-sodiumdiphosphate decahydrate was added as a dispersing agent and samples were put in an overhead shaker for 18 h. The suspension was sieved with mesh sizes of 630 μm, 200 μm, or 63 μm. The residues on the sieves were oven-dried at 105°C to determine the fractions of coarse, medium, and fine sand, respectively. The remaining suspension contained particles below 63 μm in diameter and was shaken for 1 min prior to sedimentation. Aliquots of 10 mL were taken at defined time steps, now containing only particles smaller than 20, 6.3, or 2 μm. These aliquots were transferred into glass beakers and dried at 105°C. All fractions were weighed out and their proportion to the dry mineral matter of the sample was calculated, with a correction for the weight of the dispersing agent being applied for the sedimented fractions.

2.10.2.2 Water retention curves—bulk density

In order to characterize the water retention curve (WRC) of the substrates, we used a simplified evaporation method coupled with the HYPROP apparatus (UMS GmbH Munich). For the SCE, the soil was sieved to < 1 mm and then packed to a bulk density of 1.26 g and 1.50 g, respectively, for the loam and sand. Five replicates were packed for each substrate. For SPE, one sample was extracted per plot at a depth of –15 cm prior to seeding (03/04/2019). All 250 cm^3 samples were then saturated overnight by raising a water table gradually before being placed in the HYPROP device. After the measurements, the samples were oven-dried overnight at 105°C to determine bulk density. The obtained WRC are shown in Fig. 11.

The saturated hydraulic conductivity (Ks) of the investigated soils was characterized by using a falling head method with the KSAT device (UMS GmbH Munich) for the SCE substrates. To do so, another set of five samples per soil type was prepared using the same packing procedure as for

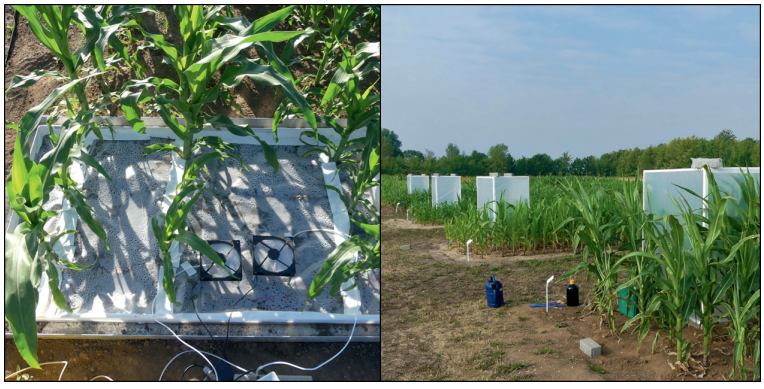


Figure 7: ^{13}C Isotope labeling in the soil plot experiment at BBCH 19. On the right, chambers in the field during pulse labeling; on the left, sealing of soil surface with a plastic foil to avoid direct diffusion into the soil.

© 2020 The Authors. *Journal of Plant Nutrition and Soil Science* published by Wiley-VCH GmbH

15222624, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jpln.202000079 by Universitats und Landesbibliothek B, Wiley Online Library on [19/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2: Selected characteristics of the substrates 'loam', 'sand' after establishment of soil plot experiment in the field. Treatments were established in 12 replicates for each substrate. Initial values for loam, quartz sand and the mixture 'sand' (16.7% loam, 83.3% quartz sand) are provided for comparison. The 'loam' was obtained by sieving and homogenizing 700 t of a haplic Phaeozem excavated from 0–50 cm depth at Schladebach (51°18′31.41″ N; 12°6′16.31″ E). The quartz sand is WF33 from Quarzwerke GmbH. Standard errors of mean are given in brackets for the field plots (n = 12). *calculated based on results of loam and quartz sand and the mixing ratio 16.7:83.3.

	Depth	Bulk density	pH ($CaCl_2$)	Carbonate	Sand	Silt	Clay	CEC	C_{org}	N_t	P plant available	K plant available	Fe_o
	(cm)	(g cm^{-3})		(g kg^{-1})	(%)	(%)	(%)	(mmol$_c$ kg^{-1})	(%)	(%)	(mg kg^{-1})	(mg kg^{-1})	(g kg^{-1})
SPE 'loam'	5–10	1.39 (*0.01*)	6.37	< 1 (*0. 0*)	32.5 (*0.36*)	47.9 (*0.17*)	19.5 (*0.26*)	98.6 (*4.7*)	0.85 (*0.01*)	0.083 (*0.001*)	32.69 (*0.40*)	28.51 (*0.72*)	1.32 (*0.01*)
SPE 'sand'	5–10	1.50 (*0.01*)	6.29	< 1 (*0. 0*)	91.8 (*0.51*)	5.6 (*0.35*)	2.6 (*0.17*)	33.1 (*2.6*)	0.15 (*0.01*)	0.017 (*0.001*)	8.29 (*0.37*)	7.84 (*0.61*)	0.25 (*0.01*)
SCE 'loam'		1.26	6.21	< 1	33.2	47.7	19.1	76.1	0.84 (*0.01*)	0.084	33.41	26.67	1.32
SCE 'sand'*		1.47	6.25	< 1	88.6	8.1	3.3	13.0	0.14	0.014	5.67 (*0.01*)	5.44	0.22
Quartz sand WF33			5.80	< 1	99.8	0.1	0.1	0.33	0.00	0.000	0.11	1.19	0.00

the WRC determination. For the SPE substrates, we used tension disk infiltrometers and performed two measurements per plot next to each other at a depth of –15 cm on the 15/04/2019. During the infiltration, a tension of –0.5 cm and –2 cm was imposed on the disk and the Ks was then derived by solving the equation proposed by *Reynolds* and *Elrick* (1991).

In order to fit the experimentally acquired data of both the SCE and SPE, we used the software HYPROP-fit (UMS GmbH Munich). The experimental data was fitted to the bimodal Mualem–van Genuchten model (*Durner*, 1994) and by allowing all parameters of the model to be fitted, except the Ks. The resulting fitted curves and parameter sets are respectively shown on Fig. 11 and in Tab. 3.

Equation of *Durner* (1994) for multimodal pore size distribution:

$$\frac{\theta(h)-\theta_r}{\theta_s-\theta_r}=\sum_{i=1}^{k} w_i\left[\frac{1}{1+(\alpha_i|h|)^{n_i}}\right]^{m_i} \quad (1)$$

where h is the capillary pressure, θ_r and θ_s are respectively the residual and saturated VWC, k is an integer number equal to the number of subcurves considered (k = 2), w_i are the weighting factors between the subcurves, α_i is a parameter related to the inverse of the air entry capillary pressure for the subcurve i, n_i and m_i are empirical parameters for the subcurve i (note that we consider $m_i = 1–1/n_i$).

2.11 Sampling strategies and documentation

Combining monitoring approaches and invasive sampling with different strategies over time requires rigorous documentation of activities. This is done by tagging any activity in the field and recording the position by a GPS device (Leica Viva GNSS-GS08 plus) with a precision of ≤ 2 cm in x, y and z direction. As a coordinate system EPSG 25832 (ETRS89/UTM zone 32N) was used and positions are all kept in specific GIS files. These are used to avoid unintended overlap of destructive/invasive sampling activities on the one hand and to relate data sampled from close proximity on the other hand.

To accommodate a multitude of activities despite the limited size of plots, individual investigators are encouraged to share samples or access to samples whenever possible. At the same time this improves the possibility for data integration and joint interpretation.

For soil column experiments, the basic approach is the same, *i.e.*, as far as possible different groups are encouraged to share samples from the same SCE. Subsampling schemes are developed to spatially register data from different groups to each other and to integrate results obtained with different approaches (see sect. 3.3).

© 2020 The Authors. *Journal of Plant Nutrition and Soil Science* published by Wiley-VCH GmbH

15222624, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jpln.202000079 by Universitats und Landesbibliothek B, Wiley Online Library on [19/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3: Fitted parameters for the bimodal Mualem–van Genuchten model for substrates loam (L) and sand (S) for soil plot experiment (SPE) and soil column experiment (SCE). Equation (1) for establishing water retention curve according to *Durner* (1994).

Parameters	Units	Loam		Sand	
		SPE	SCE	SPE	SCE
θ_s	cm^3/cm^3	0.411	0.482	0.346	0.414
θ_r	cm^3/cm^3	0	0.004	0	0
n_1	–	3.103	3.609	9.996	7.706
α_1	1/ cm	0.0332	0.0256	0.0338	0.0315
n_2	–	1.108	1.334	1.060	1.505
α_2	1/ cm	0.1670	0.0137	0.5	0.0221
w_2	–	0.833	0.732	0.327	0.328
K_s	cm/d	441	245	1174	1864
τ	–	–1.089	1.309	–0.125	1.939

3 Results and discussion

3.1 Selected drivers and their implementation

Seeds of the two plant genotypes (WT, *rth3*) had been produced in sufficient quantity in the summer season 2018 at the experimental station of the University of Bonn to supply all SCE and the SPE with seeds from the same lot. Hence, differences in seed quality can be excluded as explaining variable.

For the substrates, providing equally homogenous quality was much more difficult. Comparing two substrates under the same environmental conditions in the field offers unique possibilities, but meant that substrates had to be homogenized as in soil column experiments and soil profiles had to be established artificially. This resulted in tremendous amounts of substrate required (700 t of loam, 550 t of quartz sand) and hence tools (wheel loader, heavy duty vibrating screen, vibrating plate) not commonly used in rhizosphere research. Substrate selection could not only be based on scientific criteria (textures sand and loam, free of carbonate, low content of organic matter, small differences in pH) but had to account for local availability in large quantity at a given time point, transport cost and handling. The selected substrates (Tab. 2) show, as intended, large differences in texture with small variability between field plots, and likewise, between field plots and soil column experiments. Along with texture, there are strong differences in cation exchange capacity and concentrations of plant available P and K. Values in (L) are by a factor of 3–4 higher than those in (S). The levels of plant available P and K are low (B) and very low (A) according to the rating of the agricultural classification system (*LLFG*, 2008). These differences were accounted for by the selection of fertilization dose, being twice as high in (S) compared to (L) for the macronutrients. Resulting tissue concentrations for column experiments are in the range of 2.1–5.4 mg g^{-1} for P and 26–45 mg g^{-1} for K for (L) and (S), respectively. This classifies them as low to adequate (*Bergmann*, 1986) and, hence, the goal to avoid luxury supply for these nutrients was achieved. N_{min} was likewise higher in (L) (1.1 mg kg^{-1}) compared to (S) (0.4 mg kg^{-1}) and in general very low. Soil pH is in the neutral range for both substrates showing low variability (Fig. 8).

C_{org} and N_t concentrations for (L) are in a range typical for a loess derived substrate under the given climatic conditions (*LLFG*, 2008) and differ by a factor of 5–6 between substrates. C_{org} (being zero in quartz sand) values were used to calculate the actual mixing ratio achieved for the soil plot experiment (Fig. 8). The median derived for the twelve field plots with substrate (S) is used as the mixing ratio for soil column experiments.

Bulk density values differ between substrates in both SPE and SCE. While similar bulk densities at laboratory and field scale could be achieved for (S), this was not possible for (L). The field bulk density of 1.39 g cm^{-3} for loam cannot be established under laboratory conditions without using special devices.

A major difference between SPE and SCE is found for soil structure (Fig. 9) as mixing and sieving of the total batch of substrate required for the field could not be done in a similar fashion as it is done for SCE. Utilizing the substrates as they were packed in the field plots for SCE without further sieving would have been an option; however, this would have introduced a large heterogeneity in SCE which might hinder the detection of chemical and physical gradients at nanometer to the millimeter scale.

A column experiment in which the two structures (Fig. 9 b, d) were compared for (S) showed no impact of the structural differences on shoot or root growth, at least not at the whole column scale (Fig. 10).

Water retention curves differ, as expected, fundamentally between (L) and (S) (Fig. 11). The differences between laboratory and field scale are a result of differences in structure and bulk density as discussed above. The substrate specific volu-

© 2020 The Authors. *Journal of Plant Nutrition and Soil Science* published by Wiley-VCH GmbH

15222624, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jpln.202000079 by Universitats und Landesbibliothek B, Wiley Online Library on [19/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

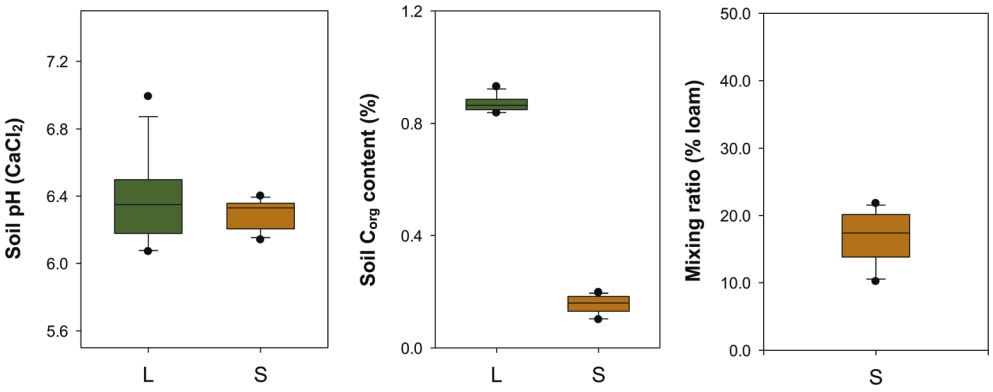


Figure 8: Soil pH and soil C_{org} content for 5–10 cm in the soil plot experiment (SPE) after establishment (November 2018). The mixing ratio for substrate sand (S), which was created by mixing substrate loam (L) with quartz sand (Q), was calculated based on C_{org} values. The median of 16.7% L is used for establishment of respective treatments in soil column experiments (SCE).

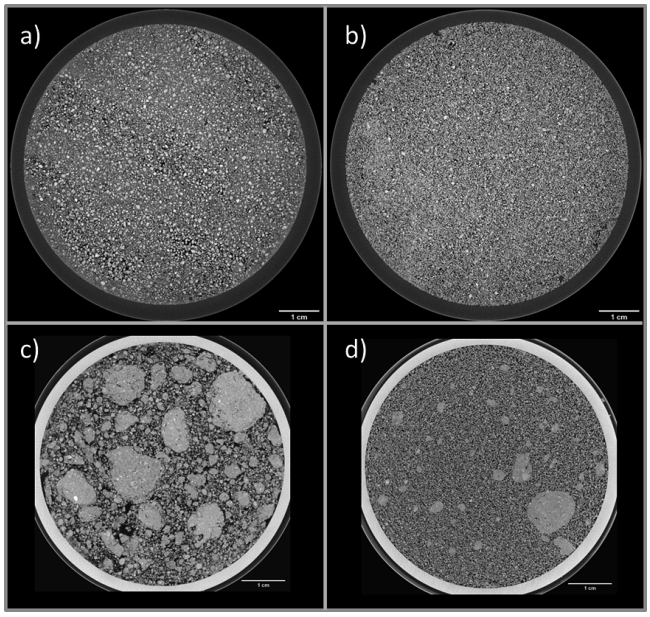


Figure 9: Soil structure in soil column experiment (SCE) and the soil plot experiment (SPE) after establishment (November 2018). Structure was visualized by X-ray CT scanning of whole columns (SCE; 7cm diameter) and extracted undisturbed soil cores (SPE, 5 cm diameter), respectively. (a) SCE–loam, (b) SCE–sand, (c) SPE–loam, (d) SPE–sand.

metric water content of 22 and 18% for (L) and (S) in SCE takes into account the steeper water potential gradient in (S) across the column height induced by lower capillarity, and the associated risk of water logging at the bottom of the sand-columns, as well as the higher share of water content below the permanent wilting point for (L). With the chosen watering levels, plant available water content averaged across the column is similar between (L) and (S) for SCE.

For SPE being exposed to natural rainfall, soil hydraulic conditions are different. The lower ability of (S) to store water is modified by the fact that a drainage layer was introduced at the lower boundary. The gravel layer acts as a capillary barrier and therefore reduces groundwater recharge and retains more water in the sand layer above.

The decision for limiting the soil profile to 75 cm and establishing a drainage layer at that depth was governed by practical considerations. The total amount of substrate to be handled had to be limited and a depth larger than 75 cm would have required construction of concrete side walls. This would have incurred substantially higher costs and time requirement.

3.2 Extrapolation from laboratory to field scale – relevance of rhizosphere processes for system behaviour

Addressing both the laboratory and field scale for rhizosphere research was motivated by two main issues: (1) the strong discrepancy in soil volume available for root growth and the potential consequences for the obtained results (*Hess* and *De Kroon*, 2007; *Poorter* et al., 2012; *Ray* and *Sinclair*, 1998), and (2) the need to judge the relevance of small scale rhizosphere processes for system behavior and the problem of extrapolating from experiments with very young plants to the whole plant life cycle.

3.2.1 Soil volume

A fundamental problem for any type of pot experiment is that available soil volume is by order of magnitude smaller in pot than in field trials. For the present example soil volume for one plant in SCE is 0.885 dm^3 while it is 78.935 dm^3 for SPE. This discrepancy remains even if we assume that for the early growth stage (BBCH 14) roots are not found below 20 cm (= 21.049 dm^3). The resulting effect for studies of root growth or rhizosphere processes is reflected in the root length density (RLD) data shown in Fig. 12. While in the field values are in the range of 0.1 cm cm^{-3}, they are 100 times higher in the SCE even at this very early growth stage (BBCH 14). Taking the classical simple concept of half mean distance (*Gardner*, 1960), assuming that all roots within a given cube are aligned in parallel [Eq. (2)], this corresponds to half mean distance

© 2020 The Authors. *Journal of Plant Nutrition and Soil Science* published by Wiley-VCH GmbH

15222624, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jpln.202000079 by Universitats und Landesbibliothek B, Wiley Online Library on [19/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

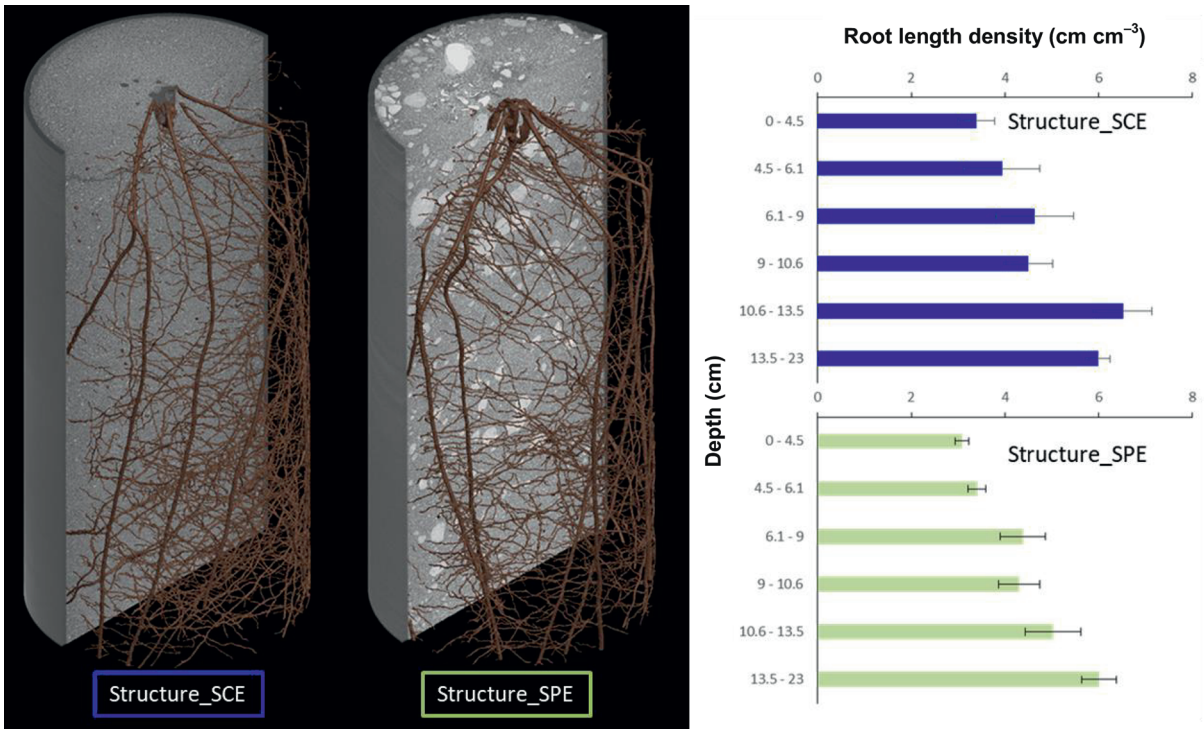


Figure 10: Root growth of *Zea mays* (B73 wild-type) in a column experiment (diameter 7cm) in substrate sand (S) for soil structure as it was established in soil plot experiments (Structure_SPE) *versus* soil column experiments (Structure_SCE). Root system architecture segmented from X-ray CT scans (left) and root length distribution with depth (right) derived from destructive root sampling. One-factorial ANOVA, $n = 3$ for individual depths, bars refer to standard error, comparison of means by Tukey-test, no significant differences were observed in any of the depths.

(HMD) of 1.78 mm for the SCE compared to 10 mm for the SPE. In other words, for a RLD of 10 cm cm^{-3} potential radial gradients around roots would start to overlap if they extent further than 1.78 mm.

$$HMD = \frac{1}{\sqrt{\pi\, RLD}}. \quad (2)$$

Alternatives are shortening the growth duration or enlarging pots. Both options have their drawbacks. Shortening the growth period means restricting the investigations to the very early growth stage in which interactions with the microbiome just start and part of the nutrients are still derived from the seed (*Julia* et al., 2018). Enlarging pots results in loss of resolution for all non-invasive imaging techniques (X-ray CT, MRI, neutron tomography). However, high resolution is required in order to detect fine roots. For the genotypes selected here depending on the substrate and age, up to 40% of the roots are smaller than 100 μm in diameter. Nominal resolution obtained with the geometry of SCE (7 cm diameter) is at best 45 μm.

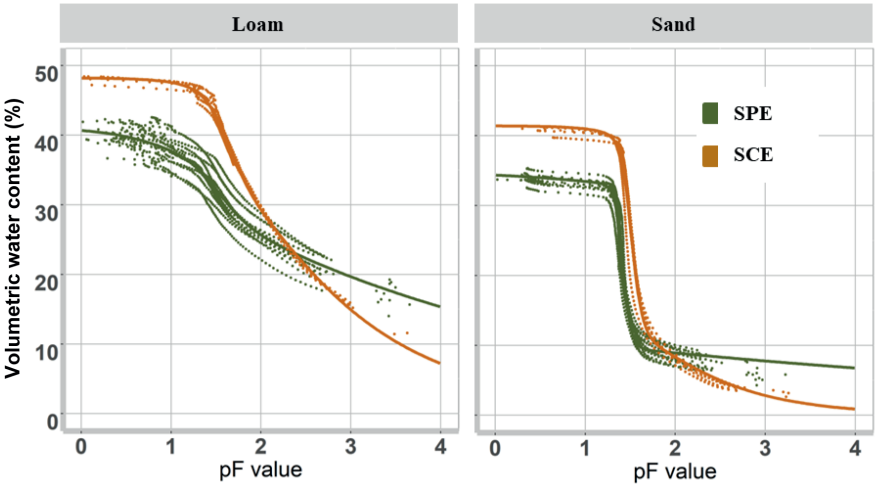


Figure 11: Measured (symbols) and fitted (solid lines) water retention curves for substrates loam (L) and sand (S) for soil plot experiment (SPE) and soil column experiment (SCE). Fitted parameters for the bimodal Mualem–van Genuchten model are presented in Tab. 3.

3.2.2 Life cycle

Accepting limitation of pot experiments to early developmental stages leads to the necessity to investigate later stages in the field. As it is well known that root function and hence associated rhizosphere processes are changing with root ontogeny and with plant age (*Vetterlein* and *Doussan*, 2016), extrapolation from the early growth stage to the whole life cycle is not justified. In addition, some of the prominent plant

© 2020 The Authors. *Journal of Plant Nutrition and Soil Science* published by Wiley-VCH GmbH

15222624, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jpln.202000079 by Universitats und Landesbibliothek B, Wiley Online Library on [19/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

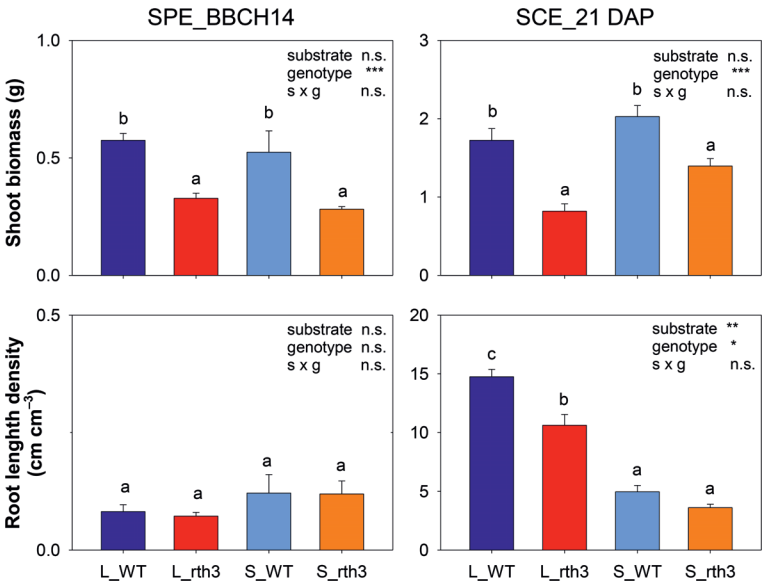


Figure 12: Impact of substrate (L, S) and genotype (WT, *rth3*) on shoot biomass and root length density in soil plot experiment (SPE) and soil column experiment (SCE) for comparable growth stage BBCH 14, which is obtained in SCE 21 d after planting (DAP). Two-factorial ANOVA, $n = 6$, bars refer to standard error, comparison of means by Tukey-test, significant differences are indicated by different letters.

microbiome interactions, like mycorrhizal symbioses, need two to several weeks for establishment, depending on, *e.g.*, soil temperature, nutrient levels, and plant host (*Jakobsen* and *Nielsen*, 1983; *Abbott* and *Robson*, 1991).

Under field conditions, plants are typically confronted with soil characteristics changing with depth and this is shaping root distribution and hence the rhizosphere (*Lynch*, 2013). This is not the case in SPE at the initial state except for soil water distribution driven by gravity and capillarity. However, starting with these homogenous conditions will allow studying changes induced by plant growth in general and more specifically by root activity. It is expected that depth gradients in soil structure, nutrient availability, carbon input and microbial activity and composition will develop over time. SPE is expected to run for six consecutive years.

A further difference between laboratory and field scale is the fact that SCE investigates the individual plants while in the field (SPE) the plant stock is investigated. It cannot be ruled out, that differences arise from interactions of roots from the same individual (SCE) *versus* roots from neighboring plants (SPE), even though neighboring plants share the same genotype (*Hess* and *De Kroon*, 2007; *Poorter* et al., 2012).

3.3 Integration of sampling strategies across disciplines

The easiest compromise for joint interpretation of data across different disciplines and between different working groups is using the same drivers (here substrate and genotype) for independent experiments. A higher degree of integration is achieved by using whenever possible the same experimental platforms (here SPE, SCE). The optimal situation for integrative research is to sample the same experiment at the same point in time and, when applicable, at the same location. This final step poses a particular challenge. In order to obtain accurate information of rhizosphere activities, two further aspects of sampling are taken most seriously: short duration and accurate timing of harvests. Particular challenge for sampling in the field is that increasing sampling time lowers the quality of enzyme activity or metabolite or gene expression level analyses. For instance, inconsistencies in RNA quality causes widespread changes in RNA sequencing based gene expression levels (*Romero* et al., 2014). Sample collection timing is another important factor for the analysis of activities, which follow diurnal rhythms, such as root gene expression (*Covington* et al., 2008) and rhizosphere microbiome structure and function (*Hubbard* et al., 2018), albeit not significant for others, such as root exudation (*Badri* et al., 2010).

3.3.1 Non-invasive approaches for data integration and guiding of sampling

For soil column experiments, a workflow was developed to provide temporally resolved information on root architecture based on X-ray CT (*Schlüter* et al., 2018). This workflow can be used to describe the spatial context of *in situ* sampling over time, such as soil solution sampling (*Gao* et al., 2019; *Lohse* et al., 2020), but they can also be used as a guide and provide the spatial context for destructive sampling and subsampling (Fig. 13). To obtain information from different depths, 1.6 cm thick layers are cut with a sharp steel blade without disturbing soil structure. Such layers have been used for the harvest of rhizospheres for microbiome analyses and sampling of roots for RNA sequencing (*Ganther* et al., 2020). The layers are further divided into undisturbed subsamples (diameter 1.6 cm) which, after sample fixation, high resolution 3D X-ray scanning, and resin embedding, are used for a range of 2D chemical mapping techniques working at the nm and μm resolution. This will enable visualization, quantification and merging of chemical and physical gradients around individual roots of known age. These data can be embedded into information on the gene expression and microbiome composition data derived from the same layer from the material surrounding the subsamples and the root age and root spatial distribution data from CT analysis. The data set can be complemented by information describing the whole system like overall root and shoot growth, plant nutrient and water uptake or root exudation. The list is not exhaustive and it is not always possible to perform all measurements simultaneously in the same system. In particular, measurement of root exudation requires intact root systems (*Oburger* and *Jones*, 2018). For biological parameters, potential artifacts like irradiation damage caused by X-ray CT scanning has to be quantified or ruled out prior the establishment of the workflow (*Schmidt* et al., 2015; *Blaser* et al., 2018; *Ganther* et al., 2020).

For the field, X-ray CT cannot fulfill the same purpose in terms of integration of data. It can still be used for investigation of

© 2020 The Authors. *Journal of Plant Nutrition and Soil Science* published by Wiley-VCH GmbH

15222624, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jpln.202000079 by Universitats und Landesbibliothek B, Wiley Online Library on [19/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

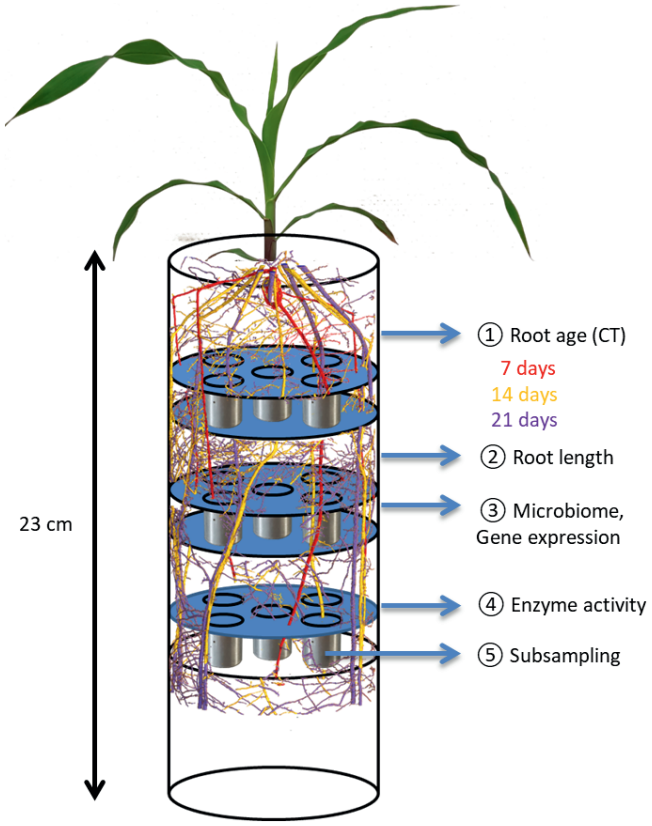


Figure 13: Illustration of a workflow for integrated sampling in soil column experiments (SCE) guided by X-ray CT scanning.

root architecture and for the work-flow associated with subsampling, but root age cannot be derived since *in vivo* application is not possible. Transparent interfaces in a soil profile, root windows, are used to describe root growth and architecture and to merge different techniques as it has been shown for rhizoboxes (*Neumann* et al., 2009; *Spohn* et al., 2015; *Kreuzeder* et al., 2018).

Another issue arises in the field experiment, *i.e.*, a given variability within a plot that needs to be captured by an appropriate sample size or number of technical replicates. An increased number of technical replicates is in stark conflict with the objective of minimal field area consumption and the idea of keeping working time and costs reasonable. Here, the compromise is oriented towards keeping both space consumption and workload as low as possible by sharing the sample material among scientists. Further benefit of this approach is that using the same sample facilitates the joint interpretation of the data. Similarly to the column experiment, detailed multidisciplinary foci on rhizosphere processes are complemented by information describing the whole plant–soil system, like overall root and shoot growth, plant nutrient and water uptake, soil carbon and nitrogen and microbial biomass levels.

First results on root and shoot growth from SCE and SPE (Fig. 12) show that, among them, the relative differences between treatments are similar regarding shoot growth at the early growth stage (BBCH 14). In both experimental platforms, the maize WT plants grow better than the *rth3* plants. This observation is in line with the expectation that under marginal P supply, implemented in both systems, root hairs contribute to P uptake. That the substrate showed no impact on shoot growth, is in part explained with our choice of fertilization level for the two substrates aiming at similar shoot size for the WT. Surprisingly, root growth at this early stage shows contrasting trends for laboratory and field scale. While at the laboratory scale root growth is higher in loam compared to sand, we observed an opposite trend in the field. For root length density, differences between genotypes are absent in the field and small in the laboratory compared to the impact of substrate.

4 Conclusion and outlook

With the establishment of the experimental platforms, a basis for interdisciplinary cooperation was laid. We consider this interdisciplinarity as indispensable for a fundamental gain of knowledge in rhizosphere research. Carrying out joint sampling and exchanging samples from the same system not only facilitates interpretation of the complex interactions and feedbacks between soil, root system and microorganisms, but also raises awareness of the needs of the different discipline-specific methods. The combination of laboratory and field scale is a great challenge but indispensable (see also *Darrah* et al., 2006; *York* et al., 2016). In principle, it is expected that although the absolute values differ, the relative differences between treatments point in the same direction in both experimental platforms. Whenever this is not the case, theories must be questioned, which can lead to the conception of new experiments and advancing the field to new areas. For the field scale, SPE provides the unique possibility to follow long-term changes of soil properties induced by a rhizosphere trait specific difference which is related to a mutation of a single gene and affects only root hair formation.

Acknowledgments

Umweltschutz-Mitte GmbH, Kieswerke Schladebach, Marcel Barnikol-Veit (excavator operator), as well as *Benjamin Pietzner, Detlef Bunge, Sven Freitag, Konrad Kirsch, Ines Merbach* (UFZ) supported the establishment of field plots.

For UFZ we thank *Sebastian Häusler* for maintaining the soil plot experiment and conducting root sampling; *Sylvia Schmögner* for washing and sorting of root samples; *Luise Ohmann* and *Bernd Apelt* for support of isotope labeling; *Maximilian John Köhne* and *Sebastian Blaser* for support with X-ray CT scanning and field sampling; *Ralf Gründling* for establishing GIS maps for all sampling events; UFZ Workshop for construction of columns, labeling chambers. Further we thank *Christa Schulz* and *Helmut Rehkopf* (University of Bonn) for propagation of maize seeds. *Ulla Rosskopf* (University of Kassel) for texture analysis; *Sandra Spielvogel* (University of Kiel) for Fe_o-analysis; *Christine Krenkewitz, Klaus Kaiser* (University of Halle) for ICP-OES measurements of soil extracts; *Minh Ganther* (UFZ) and *Desiree Bienert* (IPK) for nutrient analyses of plant samples, Landesanstalt für Landwirtschaft und Gartenbau Sachsen-Anhalt for N_{min}

© 2020 The Authors. *Journal of Plant Nutrition and Soil Science* published by Wiley-VCH GmbH

15222624, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jpln.202000079 by Universitats und Landesbibliothek B, Wiley Online Library on [19/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

analysis; *Johanna Pausch* (University of Bayreuth) for advice in setting up isotope labeling. Student assistants *Jakob Streuber, Eric Braatz, Tanja Knauerhase, Anne-Sophie Wachter, Julius Schäfer, Friedrich Jalowy*, and *Franziska Busch* provided their support whenever needed.

We would like to thank *Wolfgang Durner* (TU Braunschweig) for providing the T5 tensiometers. *Andrea Carminati* (University of Bayreuth) and *Jan Vanderborgth* (Research Centre Jülich) contributed in designing the measurements of soil and plant water status in the field. We thank *Patrick Duddek, Andreas Kolb, Daniel-Sebastian Moser, Faisal Hayat, Mohanned Abdalla* (University of Bayreuth), *Helena Jorda* (Research Centre Jülich), and *Sébastien François* (Université catholique de Louvain) for their help in installing the soil and plant sensors.

This project was carried out in the framework of the *priority programme 2089* "*Rhizosphere spatiotemporal organisation—a key to rhizosphere functions*" funded by *DFG, German Research Foundation* (*project number 403801423*).

Open access funding enabled and organized by Projekt DEAL.

References

Abbott, L. K., *Robson, A. D.* (1991): Factors influencing the occurrence of vesicular-arbuscular mycorrhizas. *Agric. Ecosys. Environ.* 35, 121–150.

Badri, D. V., *Loyola-Vargas, V. M.*, *Broeckling, C. D.*, *Vivanco, J. M.* (2010): Root secretion of phytochemicals in *Arabidopsis* is predominantly not influenced by diurnal rhythms. *Mol. Plant* 3, 491–498.

Bardgett, R. D., *van der Putten, W. H.* (2014): Belowground biodiversity and ecosystem functioning. *Nature* 515, 505–511.

Barej, J. A. M., *Pätzold, S.*, *Perkons, U.*, *Amelung, W.* (2014): Phosphorus fractions in bulk subsoil and its biopore systems. *Eur. J. Soil Sci.* 65, 553–561.

Bergmann, W. (1986): Ernährungsstörungen bei Kulturpflanzen. Fischer, Jena, Germany.

Blaser, S. R., *Schlüter, S.*, *Vetterlein, D.* (2018): How much is too much?—Influence of X-ray dose on root growth of faba bean (*Vicia faba*) and barley (*Hordeum vulgare*). *PloS One* 13. DOI: https://doi.org/10.1371/journal.pone.0193669.

Bleiholder, H., *Weber, E.*, *Lancashire, P. D.*, *Feller, C.*, *Buhr, L.*, *Hess, M.*, *Wicke, H.*, *Hack, H.*, *Meier, U.*, *Klose, R.*, *Stauss, R.*, *Van Den Boom, T.* (2001): Growth stages of mono-and dicotyledonous plants, BBCH monograph. Federal Biological Research Centre for Agriculture and Forestry, Berlin/Braunschweig, Germany.

Covington, M. F., *Maloof, J. N.*, *Straume, M.*, *Kay, S. A.*, *Harmer, S. L.* (2008): Global transcriptome analysis reveals circadian regulation of key pathways in plant growth and development. *Genome Biol.* 9. DOI: https://doi.org/10.1186/gb-2008-9-8-r130.

Darrah, P. R., *Jones, D. L.*, *Kirk, G. J. D.*, *Roose, T.* (2006): Modelling the rhizosphere: a review of methods for 'upscaling' to the whole-plant scale. *Eur. J. Soil Sci.* 57, 13–25.

Durner, W. (1994): Hydraulic conductivity estimation for soils with heterogeneous pore structure. *Water Resour. Res.* 30, 211–223.

Ganther, M., *Yim, B.*, *Ibrahim, Z.*, *Bienert, D.*, *Lippold, E.*, *Maccario, L.*, *Sorensen, S. J.*, *Bienert, P.*, *Vetterlein, D.*, *Heintz-Buschart, A.*, *Blagodatskaya, E.*, *Smalla, K.*, *Tarkka, M.* (2020): Compatibility of X-ray computed tomography with plant gene expression, soil bacterial communities and enzyme activities. *J. Exp. Bot.* DOI: https://doi.org/10.1093/jxb/eraa262.

Gao, W., *Blaser, S. R.*, *Schlüter, S.*, *Shen, J.*, *Vetterlein, D.* (2019): Effect of localised phosphorus application on root growth and soil nutrient dynamics *in situ*—comparison of maize (*Zea mays*) and faba bean (*Vicia faba*) at the seedling stage. *Plant Soil* 441, 469–483.

Gardner, W. R. (1960): Dynamic aspects of water availability to plants. *Soil Sci.* 89, 63–73.

Gregorich, E. G., *Carter, M. R.* (2007): Soil Sampling and Methods of Analysis. CRC Press, Boca Raton, FL, USA.

Heinrich, S., *Dippold, M. A.*, *Werner, C.*, *Wiesenberg, G. L.*, *Kuzyakov, Y.*, *Glaser, B.* (2015): Allocation of freshly assimilated carbon into primary and secondary metabolites after *in situ* ^{13}C pulse labelling of Norway spruce (*Picea abies*). *Tree Physiol.* 35, 1176–1191.

Hess, L., *De Kroon, H.* (2007): Effects of rooting volume and nutrient availability as an alternative explanation for root self/non-self discrimination. *J. Ecol.* 95, 241–251.

Hinsinger, P., *Bengough, A. G.*, *Vetterlein, D.*, *Young, I. M.* (2009): Rhizosphere: biophysics, biogeochemistry and ecological relevance. *Plant Soil* 321, 117–152.

Hochholdinger, F., *Wen, T.-J.*, *Zimmermann, R.*, *Chimot-Marolle, P.*, *da Costa e Silva, O.*, *Bruce, W.*, *Lamkey, K. R.*, *Wienand, U.*, *Schnable, P. S.* (2008): The maize (*Zea mays L.*) *roothairless3* gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield. *Plant J.* 54, 888–898.

Hochholdinger, F., *Yu, P.*, *Marcon, C.* (2018): Genetic control of root system development in maize. *Trend. Plant Sci.* 23, 79–88.

Hubbard, C. J., *Brock, M. T.*, *van Diepen, L. T.*, *Maignien, L.*, *Ewers, B. E.*, *Weinig, C.* (2018): The plant circadian clock influences rhizosphere community structure and function. *ISME J.* 12, 400–410.

ISO 11277 (1998): Soil Quality—Determination of Particle Size Distribution in Mineral Soil Material—Method by Sieving and Sedimentation. Beuth, Berlin, Germany.

ISO 11277 (2002): Corrigendum 1 for ISO11277:1998. Beuth, Berlin, Germany.

Jakobsen, T., *Nielsen, E.* (1983): Vesicular-arbuscular mycorrhiza in field-grown crops: I. Mycorrhizal infection in cereals and peas at various times and soil depths. *New Phytol.* 93, 401–413.

Jarvis, N. J. (2007): A review of non-equilibrium water flow and solute transport in soil macropores: Principles, controlling factors and consequences for water quality. *Eur. J. Soil Sci.* 58, 523–546.

Julia, C. C., *Rose, T. J.*, *Pariasca-Tanaka, J.*, *Jeong, K.*, *Matsuda, T.*, *Wissuwa, M.* (2018): Phosphorus uptake commences at the earliest stages of seedling development in rice. *J. Exp. Bot.* 69, 5233–5240.

Keiluweit, M., *Bougoure, J. J.*, *Nico, P. S.*, *Pett-Ridge, J.*, *Weber, P. K.*, *Kleber, M.* (2015): Mineral protection of soil carbon counteracted by root exudates. *Nature Clim. Chang.* 5, 588–595.

Kreuzeder, A., *Santner, J.*, *Scharsching, V.*, *Oburger, E.*, *Hoefer, C.*, *Hann, S.*, *Wenzel, W. W.* (2018): *In situ* observation of localized, sub-mm scale changes of phosphorus biogeochemistry in the rhizosphere. *Plant Soil* 424, 573–589.

LLFG (2008): Richtwerte für die Untersuchung und Beratung sowie zur fachlichen Umsetzung der Düngeverordnung (DüV).

© 2020 The Authors. *Journal of Plant Nutrition and Soil Science* published by Wiley-VCH GmbH

15222624, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jpln.202000079 by Universitats und Landesbibliothek B, Wiley Online Library on [19/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Gemeinsame Hinweise der Länder Brandenburg, Mecklenburg Vorpommern und Sachsen-Anhalt. Ed. Land Sachsen Anhalt – Landesanstalt für Landwirtschaft, Forsten und Gartenbau des Landes Sachsen-Anhalt, Bernburg, Germany.

Lohse, M., *Blaser, S. R. G. A.*, *Vetterlein, D.*, *Schlüter, S.*, *Oburger, E.*, *Reemtsma, T.*, *Lechtenfeld, O. J.* (2020): Online nano-solid phase extraction Fourier-transform ion cyclotron resonance mass spectrometry workflow to analyse small scale gradients of soil solution organic matter in the rhizosphere. *Anal. Chem.* 92, 10442–10449.

Lynch, J. P. (2013): Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems. *Ann. Bot.* 112, 347–357.

Marschner, H., *Häussling, M.*, *George, E.* (1991): Ammonium and nitrate uptake rates and rhizosphere pH in non-mycorrhizal roots of Norway spruce [*Picea abies* (L.) Karst.]. *Trees* 5, 14–21.

Mehra, O. P., *Jackson, M. L.* (1960): Iron oxide removal from soils and clays by a dithionite-citrate system buffered with sodium bicarbonate. Proceedings of the 7th National Conference on Clays and Clays Minerals, October 20–23, 1958, Washington, DC, USA.

Neumann, G., *George, T. S.*, *Plassard, C.* (2009): Strategies and methods for studying the rhizosphere—the plant science toolbox. *Plant Soil* 321, 431–456.

Oburger, E., *Schmidt, H.* (2016): New methods to unravel rhizosphere processes. *Trend. Plant Sci.* 21, 243–255.

Oburger, E., *Jones, D. L.* (2018): Sampling root exudates—mission impossible? *Rhizosphere* 6, 116–133.

Passioura, J. B. (2006): The perils of pot experiments. *Funct. Plant Biol.* 33, 1075–1079.

Poorter, H., *Bühler, J.*, *van Dusschoten, D.*, *Climent, J.*, *Postma, J. A.* (2012): Pot size matters: a meta-analysis of the effects of rooting volume on plant growth. *Funct. Plant Biol.* 39, 839–850.

Ray, J. D., *Sinclair, T. R.* (1998): The effect of pot size on growth and transpiration of maize and soybean during water deficit stress. *J. Exp. Bot.* 49, 1381–1386.

Reynolds, W. D., *Elrick, D. E.* (1991): Determination of hydraulic conductivity using a tension infiltrometer. *Soil Sci. Soc. Am. J.* 55, 633–639.

Romero, I. G., *Pai, A. A.*, *Tung, J.*, *Gilad, Y.* (2014): RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biol.* 12, 1–13.

Roose, T., *Keyes, S. D.*, *Daly, K. R.*, *Carminati, A.*, *Otten, W.*, *Vetterlein, D.*, *Peth, S.* (2016): Challenges in imaging and predictive modeling of rhizosphere processes. *Plant Soil* 407, 9–38.

Schlüter, S., *Blaser, S. R.*, *Weber, M.*, *Schmidt, V.*, *Vetterlein, D.* (2018): Quantification of root growth patterns from the soil perspective *via* root distance models. *Front. Plant Sci.* 9. DOI: https://doi.org/10.3389/fpls.2018.01084.

Schmidt, H., *Vetterlein, D.*, *Köhne, J. M.*, *Eickhorst, T.* (2015): Negligible effect of X-ray μ-CT scanning on archaea and bacteria in an agricultural soil. *Soil Biol. Biochem.* 84, 21–27.

Schüller, H. (1969): Die CAL-Methode, eine neue Methode zur Bestimmung des pflanzenverfügbaren Phosphates in Böden. *Z. Pflanz. Bodenkunde* 123, 48–63.

Schwertmann, U. (1964): Differenzierung der Eisenoxide des Bodens durch Extraktion mit Ammoniumoxalat-Lösung. *Z. Pflanz. Bodenkunde Düng.* 105, 194–202.

Spohn, M., *Treichel, N. S.*, *Cormann, M.*, *Schloter, M.*, *Fischer, D.* (2015): Distribution of phosphatase activity and various bacterial phyla in the rhizosphere of *Hordeum vulgare* L. depending on P availability. *Soil Biol. Biochem.* 89, 44–51.

Tinker, P. B., *Nye, P. H.* (2000): Solute movement in the rhizosphere. Oxford University Press, Oxford, UK.

Vetterlein, D., *Jahn, R.* (2004): Combination of micro suction cups and time-domain reflectometry to measure osmotic potential gradients between bulk soil and rhizosphere at high resolution in time and space. *Eur. J. Soil Sci.* 55, 497–504.

Vetterlein, D., *Doussan, C.* (2016): Root age distribution: how does it matter in plant processes? A focus on water uptake. *Plant Soil* 407, 145–160.

Vetterlein, D., *Carminati, A.*, *Kögel-Knabner, I.*, *Bienert, G. P.*, *Smalla, K.*, *Oburger, E.*, *Schnepf, A.*, *Banitz, T.*, *Tarkka, M.*, *Schlüter, S.* (2020): Rhizosphere spatiotemporal organisation – a key to rhizosphere functions. *Front. Agron.* 2020. DOI: https://doi.org/10.3389/fagro.2020.00008.

Wen, T.-J., *Schnable, P. S.* (1994): Analyses of mutants of three genes that influence root hair development in *Zea mays* (Gramineae) suggest that root hairs are dispensable. *Am. J. Bot.* 81, 833–842.

York, L. M., *Carminati, A.*, *Mooney, S. J.*, *Ritz, K.*, *Bennett, M. J.* (2016): The holistic rhizosphere: integrating zones, processes, and semantics in the soil influenced by roots. *J. Exp. Bot.* 67, 3629–3643.

© 2020 The Authors. *Journal of Plant Nutrition and Soil Science* published by Wiley-VCH GmbH

15222624, 2021, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/jpln.202000079 by Universitats und Landesbibliothek B, Wiley Online Library on [19/05/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

European maize genomes unveil pan-genomic dynamics of repeats and genes

Authors

Haberer, G.[1], Bauer E.[2], Kamal, N.[1], Gundlach, H.[1], Fischer, I.[1], Seidel, M.A.[1], Spannagl, M.[1], Marcon, C.[3], Ruban, A.[4,5], Urbany, C.[5], Nemri, A..[5], Hochholdinger, F.[3], Ouzunova, M.[5], Houben A.[4], Schön, C.-C.[2] , Mayer, K.F.X.[1,6]

Affiliations

[1] Plant Genome and Systems Biology, Helmholtz Center Munich, Neuherberg, Germany,
[2] Plant Breeding, TUM School of Life Sciences, Technical University Munich, Germany
[3] Crop Functional Genomics, Institute for Crop Science and Resource Conservation, University of Bonn, Germany
[4] Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Seeland, Germany
[5] KWS SAAT SE, Einbeck, Germany
[6] TUM School of Life Sciences, Technical University Munich, Germany

author for correspondance: Klaus F.X. Mayer (k.mayer@helmholtz-muenchen.de) & Chris-Carolin Schön (chris.schoen@tum.de)

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

The exceptional diversity of maize (*Zea mays*) is the backbone of modern heterotic patterns and hybrid breeding. Historically, US farmers exploited this variability to establish today's highly productive Corn Belt inbred lines from blends of dent and flint germplasm pools. Here, we report high quality *de novo* genome sequences of the four European flint lines EP1, F7, DK105 and PE0075 assembled to pseudomolecules with scaffold N50 ranging between 6.1 to 10.4 Mb. Comparative analyses with the two US Corn Belt genomes B73 and PH207 elucidates the pronounced differences between both germplasm groups. While overall syntenic order and consolidated gene annotations reveal only moderate pan-genomic differences, whole genome alignments delineating the core and dispensable genome, and the analysis of repeat structures, heterochromatic knobs and orthologous long terminal repeat retrotransposons (LTRs) unveil the extreme dynamics of the maize genome. Haplotypes derived from core genome SNPs demonstrate the tessellation of modern maize resulting from a complex breeding history. The high quality genome sequences of the flint pool are a crucial complement to the maize pan-genome and provide an important tool to study maize improvement at a genome scale and to enhance modern hybrid breeding.

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

Since its domestication ~10,000 years ago by Native Americans, maize (*Zea mays ssp. mays*) has become one of the most important sources for human nutrition and animal feeding. Extensive variation in landraces and breeding germplasm like dent or flint corns underpin the enormous phenotypic and genetic diversity of maize [1-3]. Today, US hybrids produced from inbed lines of different heterotic groups – eg. Stiff-Stalk and non Stiff-Stalk [4], are highly productive and agriculturally important worldwide. These US Corn Belt dents resulted from crosses in the nineteenth century between Southern dent lines introduced to the United States from Mexico, and Northern flints (NF) which were the predominant germplasm in the pre-Columbian era grown by the American Indians [5, 6]. Phenotypes and genotypes of NFs constitute a highly distinct maize germplasm and possess rounded hard kernels while dent lines contain a softer endosperm texture that caves in during maturation forming dentate kernels. Historical records and genetic and molecular data strongly indicate that NFs originated from American Indian populations of southwestern America and the Great Plains [2, 6, 7]. Improvement by early farmers and adaptation to cooler climate and different photoperiods extended NF growth as north as southern Canada. After the colonization of the New World, maize was spread to Europe from both the Caribbean Islands and northeast US [8]. The early maturing and cold tolerant flints were key to a successful maize cultivation in temperate regions of Europe. In consequence, Northern flint germplasm still has a major contribution to modern European maize breeding material while Corn Belt dent genomes contain on average one quarter from their blending with NFs [2, 6].

The complete genome sequence of the dent line B73 represented a milestone in maize breeding, genetic and genomic research [9]. Applying the rapid progress in sequencing technologies, the B73 genome and gene models have been updated several times to reference quality [10]. On the other hand, very high diversity at both the sequence and genic level has been reported between maize inbred lines for targeted regions and large-scale comparisons [3, 11-14]. Hence, the B73 reference sequence captures only a portion of the maize pan-genome. To overcome these limitations, a draft genome of the Iodent line PH207 [15], and - more recently- reference sequences of Mo17 and W22, two additional US dent lines, have been released [16, 17]. Genome-wide comparisons confirmed the high dynamics of the maize genome emphasizing the need for several high-quality genome sequences to characterize and exploit the pan-genomic variability in maize. Worldwide, many hybrid breeding programs focus on dent germplasm, whereas breeding programs in cooler regions of Central Europe exploit heterotic effects between dent and flint lines. Several studies have shown a clear differentiation of the Northern American and Northern European flint germplasm from the rest of the world [18]. While reference quality sequences exist for several dent inbred lines, the flint pool is still underexploited. To date, only one fragmented draft assembly with an N50 ~13.9 kb of one flint line, F2, is available covering ~65% of the estimated genome size [19]. We generated *de novo* reference sequences for four flint inbred lines representing predominant ancestors of maize hybrid breeding in Central Europe [20]. The four lines represent distinct flint germplasm sources such as populations Lacaune (F7, Southern France), Lizargarate (EP1, Northern Spain), Gelber Badischer Landmais (DK105, Southern Germany) and Petkuser Ferdinand Rot (PE0075, Northern Germany). Understanding the extent

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and quality of genomic differences between the flint and dent germplasm pools may contribute to a better understanding of the maize pan-genome and of the complementarity in heterotic patterns exploited in hybrid breeding. For a first assessment of the pan-genome, we contrasted gene and repeat contents of European flints to two US Corn Belt dents applying a consolidated gene annotation and whole genome alignments. The length and quality of the assemblies now allow aligning and comparing also repeat-dense intergenic regions and reveal the degree of repeat conservation and rapid turnover. Our analysis reveals the presence of widespread knob repeat "seed" regions and their influence on the surrounding gene landscape and expression characteristics. Finally, a haplotype-informed expression analysis surveys the effect of germplasm-specific regions on the maize transcriptome and reveals gene loci that are prime candidates for differences in the kernel phenotype among both gene pools. We conclude that for larger and complex plant genomes pan-genomic analysis will require a multidirectional angle and, beside gene-centric analyses, has to consider the structural and functional environment that might impose important *cis* regulatory effects.

Results

Whole genome sequencing

To access the genomic landscape of European flint maize, and to compare it with lines from the dent pool, we assembled four flint lines to reference genome quality. Three lines - EP1, F7 and DK105, are important founders of European breeding programs. The fourth line, PE0075, is a doubled haploid line derived from the Petkuser Ferdinand Rot population, a European landrace. We generated Illumina paired-end and mate pair sequences equivalent to ~ 220-320 x coverage (Supplementary Table 1), and assembled the reads to scaffolds and pseudo-molecules using the DeNovoMagic pipeline [15, 21, 22]. The large majority (94.3% up to 97.3%) of contig and scaffold sequences were integrated into 10 pseudochromosomes underpinning the high quality of the four new maize flint references (Table 1; Supplementary Table 2; Supplementary Figure 2A). We evaluated the completeness and sequence qualities of the four assemblies applying a BUSCO analysis [23]. Only ~2.5% of BUSCO genes were absent in each of the four assemblies on average, and another ~2% additional fragmented matches were observed. Complete BUSCO genes totaled >95% strongly supporting the high quality of the flint assemblies (Supplementary Table 2). Short read mapping of EP1 and F7 Illumina paired-end data supported a high sequence accuracy with less than one erroneous base per 100 kb (Supplementary Table 5). A genetic map generated from a F_2 mapping population of a EP1xPH207 cross demonstrated a high consensus between genetic and physical map further corroborating the high quality and contiguity of the maize assemblies (Supplementary Figure 1).

Pan-gene variation among flint and dent is moderate

To assess genic presence-absence variations (PAVs) in the six maize lines we employed a three-layer gene prediction pipeline. This ensured identification of gene variability among these lines and at the same time improved robustness of the structural gene annotation using comparative genomics components. For F7 and EP1 we predicted protein-coding genes as consensus models using protein homologies of known monocotyledonous protein sequences and a broad spectrum

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of transcriptome evidences from F7 and EP1 RNA-seq data (Supplementary Table 3). Initial gene sets comprised 39,352 and 41,387 protein coding genes for EP1 and F7, respectively. To exploit the potential of comparative genomics and to ensure a uniform and comparable gene annotation, we cross-mapped all gene models of the dent and flint maize lines and, eventually, complemented existing annotations. After filtering for transposon or transposon-derived genes this gene set (GS1) comprised ~45-47.5k protein coding gene calls per line (Table 1).

We evaluated and scored gene models by two metrics, orthologous relations and a multi-layer perceptron classifying genes into low (LC) and high confidence (HC) models, using the consistency of protein structures to known annotations. The eligibility of our classifier to identify problematic gene models is demonstrated by significantly lower expression levels of LC genes (median tpm_{LC} = 0.07 versus tpm_{HC} = 6.39), lower exon number per gene (2.9 exons per LC gene versus 5.3 exons per HC gene) and a higher fraction of single exon genes (29.1% versus 21.7%; Fisher test $p < 1^{-300}$). In addition, out of 3,917 initial genes containing InterPro signatures indicative for transposons, 93% were classified as LC suggesting that the classifier is well capable to identify transposon-derived genes. The majority (~85-90%) of novel models added to GS1 were LC genes. However, between 650 and 900 HC genes complemented the initial structural gene calls per line. The second round of cross-consolidation limited to HC genes and LC genes with at least one syntenic counterpart adjusted gene structures and finalized gene set #2 (GS2) to a total of 41 to 44.5k genes with a mean of 42.5k genes per line. Variability in gene numbers is mainly reflecting underlying variations in assembly quality of the different lines. For example, the lower gene count found in PH207 is likely attributable to comparably larger gaps in the genome assembly whereas the higher gene count in EP1 and F7 is due to a higher amount of genes located in unanchored scaffolds that presumably represent assembly duplications.

Overall, 37,700 to 39,500 genes (88% to 93% of all genes per line) have a syntenic counterpart in at least one of the other lines and 95% of the genes in each genome had an orthologous copy in at least one of the other maize lines. Notably, GS2 increased the number of syntelogs by 2,200 to 2,600 genes per line. The cross-mapped annotations also identified between 1,380 (B73) and 1,680 (PH207) genes that were out-of-syntenic order but detected as a top hit by a syntelog of another line. 91.3% of these top hits represented bi-directional best blast hits (bbhs) to at least one of their proposed syntelog clusters suggesting that these orthologous candidates translocated in the respective maize line or represent small-scale misassemblies [24]. The remaining 1,376 up to 3,801 singleton genes had no syntelog cluster assignment. In all six lines, singleton genes were highly enriched for tandemly repeated genes, and the majority (60% in B73, PH207, DK105 and PE0075) of singletons were tandem duplications of syntenic genes indicating tandem gene duplications as a major source for gene copy number variations.

In summary, approximately 95% of the genes in each genome had an orthologous copy in at least one other maize line. GS2 significantly improved the completeness of syntelog clusters with 30,109 clusters that have syntelogs for all six lines (25,316 found for GS1) (Figure 1A and B). The addition of out-of-synteny bbhs further increased the number to 32,295 complete clusters, similar to the number of syntenic genes reported in the pairwise comparisons of W22 and Mo17 to B73 [17, 22]. However, shared pairwise gene content of GS2 – including out-of-synteny bbhs - range between 36,466 (between EP1 and PH207) and 38,133 (between DK105 and PE0075) indicating considerably higher conservation between our maize lines than previously reported.

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Overall, core genes present in all maize lines exhibit high expression levels (mean and median 27.5 and 5.6 tpm) while PAV genes show significantly lower expression levels (median 0.07 tpm) and constitute the dispensable and highly variable gene fraction in maize (Figure 1C). The very low median expression of syntenic dispensable and singleton genes suggests no, only minor or highly specialized, line-specific functions for the majority of these genes.

Consistent with the genome analysis of the maize inbred line Mo17 [17], our cross-consolidation revealed a sizeable number of predicted genes with large effect mutations. Between 2,182 (B73) and up to 3,458 (PE0075) gene models had no contiguous open reading frame due to internal stop codons or frameshifts. Noteworthy, genes with disrupted ORFs had highly similar expression levels to their syntenic partner with contiguous reading frame. For example, pairwise comparison of F7 and EP1 syntelog expressions for the three classes – none, one or both syntelog(s) with a disrupted ORF – exhibited clear linear correlations with coefficients $0.8 < r < 0.95$ (Figure 1D). Thus, a large majority of disrupted genes are not different in their expression to their syntenic partners with uninterrupted ORFs, and their mean and median expression are undiscernible to syntelog clusters from five lines. Such expressed a-, hypo- or even antimorphic alleles might represent a rich source of heterosis in maize and could also be involved in recently described nonsense induced transcriptional compensation (NITC) [25, 26].

The full-length retrotransposon landscape illustrates the dynamics of transposable element (TE) generation and purging

Approximately 80% of the maize genome is constituted by repeats of different types. No pronounced differences in TE composition among the six lines analyzed was detected (Supplementary Table 6; Supplementary Figure 2). Full-length LTR retrotransposons (fl-LTR) have been shown to be very good indicators of the assembly quality and a linear relationship of the number of fl-LTRs and genome size has been observed [21]. For the maize flint lines sequenced and assembled in this study, we detected almost 15,000 high quality fl-LTRs per line, matching the expected genome size to fl-LTR ratio (Supplementary Table 7). The lower number observed for PH207 (6,838 fl-LTRs) can be attributed the overall lower quality assembly compared to the other five lines. Correspondingly, the PH207 assembly has to be treated with caution and depending on the analysis criteria was excluded from evaluation. For the remaining five lines, the number of fl-LTRs and their age distribution confirm the high quality of the respective assemblies (Supplementary Figure 3).

To quantify transposon dynamics between the different maize lines we identified still shared syntenic fl-LTRs by clustering TE junctions (50 bp outside and 50 bp inside the element) with high stringency. Only 8.9% of all fl-LTR locations where found to be shared between five lines. While the percentage of shared elements among the different lines shows a marked decrease, accordingly the pan-genome of the fl-LTR space almost doubles the number of elements. This high dynamic is in stark contrast to genes where 84.6% are retained at their syntenic position (Figure 2A). A pairwise cross comparison among the different genomes detected 23-31% of still syntenic fl-LTRs to be present in the corresponding line. The differences in pairwise shared numbers match the phylogenies from gene derived phylogenetic relationships (Figure 2B). Line unique fl-LTRs are younger and enriched in the more distal chromosome compartments

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

concordant with higher recombination rates. Increasing the number of lines for shared fl-LTR elements reveals their continuous enrichment towards the distal low recombining compartments accompanied by an increase in age (Figure 2C). Our findings illustrate the rapid turnover of the intergenic space most likely driven by elimination through illegitimate recombination and, potentially, also functional consequences selected for by breeders and farmers during the cultivation and breeding process over the last centuries. The observation that lineage-specific transposable elements are more enriched in gene-rich regions towards the distal ends of the chromosomes, a genome environment that is usually lower in (fl-) TEs, and the observation that TE presence in the surrounding of genes frequently interferes with expression and functionality [27] thus allows to speculate that in part the lineage-specific TEs located in gene-rich regions might underlie functional and trait differences. Our results complement a similar recent comparison on the mobile genome content of four maize lines [28] and highlight the enormous dynamics of TE generation and purging which are already visible within the -in evolutionary time dimensions - very narrow time window after domestication and during the breeding process of maize and the spread along the Americas and towards Europe.

Heterochromatic knob islands in the maize genome

Knob regions are heterochromatic regions in the genome that have been demonstrated to affect local recombination [29, 30]. They belong to the group of satellite tandem repeats which are well represented in our flint assemblies (Supplementary Figure 4 and 6A, Supplementary Table 6B). The Maize knobs are composed of 180 bp and closely related 202 bp tandemly repeated sequence units (Supplementary Figure 5C). Variations in position and extent of knob regions in maize have been documented [29, 31]. We compared knob positions and extent in the six lines analyzed using fluorescence in situ hybridization (FISH). A broad variety of position and strength of hybridization signals was observed (Figure 3A;). Intensity and position of knob regions detected for the different karyotypes clearly separate dent lines (B73 and PH207) from flint lines (EP1, F7, DK105, PE0075) (Figure 3A; Supplementary Figure 5B). In particular extensive knob regions detected on chromosomes 7 and 8 of the dent lines are not found in flint lines. For other positions on the genome, a clear flint/dent separation among lines cannot be observed. However, the variations observed among the lines reveals the pervasive dynamics of knob regions in the genome that also influence local recombination rates [29]. FISH only detects longer regions that carry the respective sequence signature while shorter regions containing fewer tandem units are not detectable. High quality genome assemblies resolving a larger proportion of the repetitive space now allow analyzing the so far hidden, cryptic, knob sites. We used the assembly of EP1, the best resolving assembly, to analyze for the characteristic sequence signatures. While all signals also detected by FISH (chromosomes 4L, 5L, 6S) have clearly detectable sequence counterparts in the assembly numerous additional positions with fewer tandem repeat units are found in the sequence assembly on all chromosomes (Figure 3B). Given the observations on the rapid diversification based on repeat units detailed above and the sequence and genome-based observations this might indicate numerous potential (shadowed) knob regions that can rapidly expand or shrink by e.g. illegitimate recombination and can cause recombinational isolation of the affected regions.

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In fission yeast, knob structures have been shown to implicate significant downregulation of surrounding genes [32]. To check for similar effects in maize we analyzed the distance of genes to the nearest knob signature in the assembly and related it to expression data [33] generated for the lines B73, EP1 and F7. We find the close surrounding of knob repeat units to be devoid of genes (Figure 3C). The observation is even more pronounced than observed for other types of repeat elements and could be attributable to the heterochromatic characteristics of knob regions. Also for the genes surrounding knob regions, we observe a trend for lower expression values (Supplementary Figure 5D). Thus, the widespread presence of knob-type tandem repeats, the pronounced differences between lines, their rapid expansion and/or shrinking and the marked impact on the expression of genes located in the neighborhood might contribute to the appearance and impact of rare alleles and rare deleterious variants associated with crop fitness [33].

Whole genome alignments

Numerous large structural variations (SV) in maize have been reported [11, 12, 14]. To gain insight into the extent of SVs in the maize flint versus dent genomes we generated whole-genome alignments (WGA) among all of the genomes used in this study. We thereby overcame limitations caused by commonly employed short read sequence read alignments (e.g. large extent of repetitive genome space with ambiguous mapping), and made use of the information gain provided by the full genome assemblies of all maize lines used in this study. Thus, contextual sequence and full genome information have been exploited and non-genic elements that make up for the overwhelming majority of the genomic space were considered for the WGAs.

We generated pairwise WGAs for the six lines and post-processed the resulting alignments to single alignment blocks (SAB) that represent the highest scoring one-to-one relationships between each genome pair. Single alignment blocks (SABs) were further concatenated to merged alignment blocks (MABs) if they followed a strict and unambiguous order in both genome sequences (Supplementary Figure 6). MABs provide an approximation of the overall contiguity between two genomes by bridging sequence gaps or regions of high sequence difference due to sequence insertions, deletions or inversions. On average, approximately 50% of the genome sequence aligned in each of the pairwise comparisons by SAB scoring(Supplementary Table 8).

WGAs (MABs) assigned 80-90% (1.7 - 2 Gbp) of the genome sequences into 17,000 to 21,000 syntenic segments (Supplementary Table 9). Mean sizes of SABs and MABs were ~10kb and ~100 kb, respectively (Supplementary Figure 7). Consistent with previous findings, the analysis of MABs revealed numerous genomic rearrangements including translocations (~500 up to 1,650 inter-chromosomal) and inversions (~2,500 to 3,700) among the different genotypes. Inversions and translocations span a total of 48-115 Mbp and 5.3-12.8 Mbp of contiguous genomic sequences, respectively. While translocations are in general small and have a mean size of 8.6kb, the length distribution of inversions showed both small and larger sizes including 427 inversions ≥500 kb and up to 4.8 Mb (Supplementary Figure 8). All of them have potential influence on local recombination rates and separate individual lines genetically and functionally. To delineate regions that do not align with any other of the five lines (‘unaligned’), align solely with flint or dent

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

maize lines ('group-specific') or are aligned in all pairwise alignments ('core'), we combined pairwise WGAs of each line and projected them onto its genome sequence (Supplementary Figure 9). We classified a genomic region as core type if it aligned to four out of the five possible lines due to the high amount of gap sequences in PH207 that likely result in an underestimate of the true core genome. The core genome defined by SABs comprised on average ~850 Mb (40%) of the total genome while 71 Mb (3.3%) were group-specific and ~460 Mb (22%) unaligned. This indicates a large fraction of uniquely inserted or deleted sequences in each line. Positional analysis of genomic bins revealed significant positive correlations ($\tau \geq 0.38$; $p<10^{-36}$) between the core and repeat density while gene densities were negatively correlated ($-0.4 < \tau < -0.33$; $p<10^{-26}$) (Figure 1A and B). Notably, for all six lines the highest core densities located at regions adjacent to the centromere corresponding to above average SAB sizes and below average recombination rates (Supplementary Figure 10). These observations are reminiscent of observations in the *Triticeae* gene space [21] mirroring the recombinogenic and functional properties of more central chromosomal regions vs. regions with higher recombination and genic innovation near the telomeric ends, characteristic for the dispensable genome. Recombination inertness of the central regions is also mirrored by the high repetitive content of these regions but also an overall better alignability compared to the more distal regions of the chromosomes.

Genetic Mosaicism and Haplotypes in US Corn Belt dents and European flints

The core genomic regions served as positional anchors to generate multiple sequence alignments. One-to-one unambiguous (i.e. no overlapping tandem duplications) alignments for the six lines total 287 Mb (~15%) of each genome and comprise $6.25x10^6$ orthologous SNPs with positional information for each of the six genomic coordinate systems. The detected SNPs showed an excellent agreement of 99.3% (4,827 mismatches out of 653,398 total scored calls) to a published set of SNP calls using the 600K Affymetrix™ array [34]. This congruence is well within the error margins of genotyping arrays [35], and demonstrates the orthology and high accuracy of the multiple WGAs for the six lines. Given the observed SNP frequencies, and applying randomization studies and run-of-head statistics, the expected maximal run length of identical SNPs between two lines range between 21 to 30 consecutive SNPs. We applied a conservative threshold of ≥40 identical SNPs as seeds to identify shared haplotypes of likely common ancestry, and selected genomic windows with a minimum number ≥40 identical SNPs as seeds for a greedy extension to delineate genomic segments with ≥98% sequence identity between all 15 pairwise line combinations (Figure 4C). Mean and total genomic spans covered by these pairwise near-identical haplotypes range between 40 to 68 kb and 394 up to 664 Mb suggesting between one fifth and up to one third of the genome shared by a recent common ancestor. The proportion of shared genomic regions between flint and dent lines is in close agreement with previous studies of Northern flint content in Stiff-Stalk and non-Stiff Stalk dents [2, 6]. Number, mean and total sizes were highly consistent with their phylogenetic relationship deduced from high confidence orthologs (Supplementary Figure 11). However, each line of one group (flint or dent) contained substantial shared genomic portions of the other germplasm group (394 Mb up to 480 Mb for B73 and EP1 or PE0075, respectively). This is in line with the history of US Corn Belt dents that originated from crosses of Southern Dents and Flints of Northern America and Canada [6]. The

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

latter established many founder lines of modern European flint breeding and were introduced to Europe starting as early as the discovery of the New World [8]. Alternatively, these haplotypes could be introgressions of US germplasm into European flint material that started in the 1950's to broaden genetic diversity in European breeding programs. Notably, the DH line of the European landrace Petkuser exhibited the largest genome wide similarity of all flints towards B73.

Next, to pairwise runs-of-identities, we surveyed higher order haplotypes considering all 15 pairwise similarities simultaneously as combined binary patterns. Limiting these haplotypes to a minimum run size of 40 SNPs (see Methods), we identified 31 fitting our minimal size criterion (Figure 4D; Supplementary Table 10). Since the identified regions exceed random expectation by far and exhibit near complete SNP identity across particular line subsets/combinations, these regions are likely identical by descent and share a common ancestor. In total, the 31 distinct higher order haplotypes span ~288 Mb and comprise >1.1x10^6 orthologous SNPs. Noteworthy, ordering by the number of SNPs as well as the genomic region covered, the top five/six haplotypes encode combinatorial groupings with five out of six lines being identical. The haplotype distinguishing between the studied dent (B73, PH207) and European flint lines ranked only as the sixth (by size) or seventh (by SNP number) most frequent (Figure 4D; Supplementary Table 10). Pairwise and higher order haplotypes depict the high genetic mosaicism of the pan-genome built from European flint and the two US Corn Belt dent lines. The distinct higher order haplotypes distribute quite evenly along the 10 chromosomes. We observed, however, also several striking clusters: for example a cluster on chromosome 3 with the three flints as one group and the 2 dents together with the Petkuser line form two distinct IBD regions. For two other regions on chromosome 2, five lines show identical SNP patterns while either F7 or B73 had distinct haplotypes. Though not the most prominent IBD type, three regions were enriched for the haplotype distinctive between flint and dent lines: two at the distal sites of chromosomes 4 and 7, and a large portion spanning a total around 12 Mbp in the proximal part of chromosome 8. The germplasm specific haplotypes differentiating our flint and dent lines strongly coincide with the group-specific WGAs on chromosome 8 (Supplementary Figure10) and enclose the major flowering time QTL *vgt1* in maize for which a *cis*-element 70kb upstream of the AP2 transcription factor gene rap2.7 has been proposed as causative locus [36, 37].

Expression differences among different maize lineages

To explore putative functional consequences of the haplotype clearly differentiating our dent and European flint lines, we analyzed its effect on the maize transcriptome. From the maize association panel comprising 282 lines we selected a subset of 40 lines, which were highly similar to either B73 or F7/EP1 within regions identified by the haplotype analysis (Supplementary Figure 13) [38, 39]. We then analyzed genome-wide expression levels between the two contrasting groups using expression data from seven different conditions and tissues including root, shoot, leaf base and tip, leaf under light and dark, and kernels [33]. In total, 7,684 genes were differentially expressed (DEGs) over all seven tissues and conditions (Figure 5). A major phenotypic difference between the two germplasm groups analyzed is the kernel type. Maize endosperm contains soft starch granules that are embedded in a vitreous protein matrix of zein prolamines. Dent corn has softer starch at the kernel top, which caves in during maturation

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

forming the characteristic tooth-like shape while flint kernels have a harder outer shell and remain round shaped.

Scanning the identified DEGs for curated maize genes with known functions and phenotypes revealed a striking pattern for genes involved in the establishment of the kernel texture. We identified a plethora of genes involved in starch metabolism: sugar transporters including several proteins of the SWEET family, *hexokinase 4* and *7*, sucrose synthase *shrunken1*, the rate-limiting gene for starch biosynthesis (*shrunken2*), *starch synthase 1* (SS1), *starch phosphorylase*, the $\alpha(1\rightarrow6)$-glucosylhydrolase *sugary* [40] and starch branching enzyme *sbe1* catalyzing the formation of $\alpha(1\rightarrow6)$-glycosidic bonds in amylopectin (Figure 5C). Interestingly, the rice ortholog of *rap2.7* is *rice starch regulator 1* (*rsr1*) for which it has been shown that a T-DNA mutant has altered ratios of amylose to amylopectin and different starch granularity [41]. Carbohydrate metabolism complexly correlates with flowering time and *rap2.7* may therefore have a dual role in maize [42, 43]. In summary, this set of differentially expressed genes widely covers synthesis and turnover steps of the starch pathway.

We also detected a number of genes that function in the establishment and regulation of prolamin content and structure of the maize endosperm (Figure 5B). These genes encode the direct transcriptional regulator *ohp1* of α and γ zeins and several genes regulating the packaging of zeins and starch granules: opaque endosperm *o7* and *o10*, *empty pericarp 16*, floury endosperm *fl1* and naked endosperm *nkd1* [44-47].

Discussion

Currently available maize reference genomes were limited to four US dent lines, B73, PH207, W22 and Mo17 [10, 15-17]. The full genome sequencing of the four different European flint genomes EP1, F7, DK105 and PE0075 to reference quality closes an important knowledge gap in maize research and complements the maize pan-genome. Gene and repeat content together with the analysis of WGAs and SNPs confirms the European flints as highly different germplasm of cultivated maize. For all these analyses, the four flint and the two dent lines, B73 and PH207, form two distinct groups, with each line sharing larger numbers of syntenic genes, repeats, aligned regions and SNP genotypes with members of the same group than with any member of the other group. Nevertheless, we also observed significant overall conservation of chromosome structure and gene content. Aligned blocks ordered by consecutive syntenic order span ~85% of the entire genome in all lines. A moderate number of non-syntenic genes also reflects this high degree of syntenic conservation. Our comparative cross-mapping strategy complemented in all lines a considerable number of previously missed genes. Consistent with reports in Mo17 [17], enforcing full-length mappings of top-scoring consolidation models revealed a substantial fraction (>8% of the annotations) of genes that contain large effect mutations and disrupted ORFs. These may contribute to widespread dominance effects and consequently heterosis reported in several studies in maize [48]. The consolidated annotation suggests on average 35-36,000 syntenic protein-coding genes and another 1,000-2,000 non-syntenic orthologues shared between two lines. Taking larger sequence gaps of PH207 into account, we estimate the core gene number in

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

maize from the cluster sizes comprising syntelogs of five or six lines to range between 35-36,000 genes. A considerable number of non-syntenic genes were tandemly repeated genes suggesting a significant contribution to genic CNVs. Moreover, a large number of syntelogs found in only a subset of lines were expressed very lowly or not at all in a dataset of 26 different tissues and conditions contesting functionality of these predictions. However, the latter set also contains genes expressed at moderate to very high levels, which might contribute to line-specific adaptations and are of high interest for maize improvement and breeding. Additionally, some of these genes may have highly confined expression patterns or conditions that were missed in our transcriptome analysis. Though genic PAVs were moderate in our maize lines, there are also clear pieces of evidence that genic copy number variations play an important role in the differentiation of maize germplasm. Assessing the abundance of all four combinations of the set of syntelogs shows the largest difference for the division into 4 flint versus 2 dent lines (Supplementary Figure 12) clearly demonstrating that genic PAVs adhere to the expected pedigree information and germplasm relationships.

In contrast to the overall syntenic conservation, WGAs, the analysis of fl-LTRs and knob repeat patterns demonstrated enormous diversity and dynamics of the maize genome. This is also basis for the high amount of unaligned, non-orthologous genomic sequences, which are significantly larger than germplasm-specific WGAs and approximately ~½ of the core genome. Even more pronounced, however, is the asymmetry for shared orthologous fl-LTRs. Less than one tenth of them are conserved in all six lines. Consistent with previous reports, and similar to LTRs, large heterochromatic knobs display substantial variation in position (and likely also extent) among the maize lines studied [31, 49]. Remarkably, both shared sets of orthologous LTRs and knob positions perfectly reproduced the relationship of the six lines based on coding sequences, indicating smooth transitions between these features during breeding and admixture. In contrast to previous studies reporting low knob numbers in NF or even knob-less NF lines, the European flints including DK105 and the Petkuser line, contained similar knob numbers like the US dents [49]. It is unclear whether this is a characteristic of the NF ancestral lines migrated to Europe, reflects the breeding history in European flints or represents a technical advantage given the progress in methodology to detect weaker signals. For the first time, larger or nearly complete knobs - particularly for EP1- assembled into pseudomolecule sequences. Consistent with fiber-FISH studies, we also detected numerous locations of the 180 bp knob repeat unit throughout the chromosome assemblies [50]. These widespread knob repeats affected gene density and expression levels of adjacent genes adding further complexity to known effects of large-scale knobs in maize on recombination rates, meiotic drive and phenotypic traits like flowering time [29, 51, 52].

WGAs enabled to identify long runs of pairwise and higher-order combinations of genomic regions with identical haplotypes. Size and distribution of these haplotype blocks reflect complexíty of historic recombination, intercrossing and breeding history over the last centuries [3, 5, 7, 18, 53-55]. Similar to the gene and repeat comparisons, the SNP runs highlight the separation of flints and dents of this study but also illustrate and delineate candidate regions of common ancestry in the US Corn Belt and European flint lines. We identified a set of regions that differentiated the studied Corn Belt dents from the European germplasm. Intriguingly, genes differentially

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

expressed between two groups of maize varieties that have been selected by their genotypic similarity to these regions comprised a nearly complete starch biosynthesis pathway and several genes involved in the establishment and organization of the endosperm texture. These loci are certainly good candidates for the genetic analysis of hard and soft kernel phenotype of flint and dents, respectively. However, within each group of our selected maize varieties, there are also lines classified as semi-dent or semi-flint. The complexity of the kernel/endosperm phenotype suggests that probably hundreds of loci contribute to it and that their impact will be on many regulatory levels, including transcriptional regulation but also other *trans*-active effects like protein-protein interactions or *cis*-effects from alleles modifying protein function. Hence, we speculate that altered expression of the identified genes represent a predisposition for the kernel phenotype rather than the sole causative factors. Moreover, the genetic mosacism observed within our six lines suggests that there are few if any such factors specific for one of the germplasm groups. Nevertheless, besides the detection of selective sweeps and other genomic signatures, it might be equally informative to survey *trans*-effects of such genomic regions. We demonstrate the value of full genome sequencing and assembly - rather than restriction to resequencing approaches - that allows analyzing genomic variations beyond the sole gene complement. While sequencing and assembly of complex grass genomes, such as maize, was for a long time complicated by technical and economic limitations, recent developments in assemblies of complex plant genome now appear to be reasonably economical and feasible. Haplotype characterization and contemporary methods to e.g. calculate breeding value and genome wide association mappings (GWAS) might profit from complementing gene-centric analysis by differences of knobs and LTRs in the circumference of genes associated to traits of interest.

References

1. Doebley, J.F., o.M. Goodman, and C.W. Stuber, *EXCEPTIONAL GENETIC DIVERGENCE OF NORTHERN FLINT CORN.* American Journal of Botany, 1986. **73**(1): p. 64-69.
2. Liu, K., et al., *Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites.* Genetics, 2003. **165**(4): p. 2117-28.
3. Chia, J.M., et al., *Maize HapMap2 identifies extant variation from a genome in flux.* Nat Genet, 2012. **44**(7): p. 803-7.
4. Duvick, D.N., *The contribution of breeding to yield advances in maize (Zea mays L.).* Advances in Agronomy, Volume 86, 2005. **86**: p. 83-145.
5. Anderson, E. and W.L. Brown, *The history of the common maize varieties of the United States corn belt*. 1952, Berkeley, Calif.: University of California Press. 8 pages.
6. Doebley, J., et al., *The origin of cornbelt maize: The isozyme evidence.* Economic Botany, 1988. **42**(1): p. 120-131.
7. van Heerwaarden, J., M.B. Hufford, and J. Ross-Ibarra, *Historical genomics of North American maize.* Proc Natl Acad Sci U S A, 2012. **109**(31): p. 12420-5.
8. Tenaillon, M.I. and A. Charcosset, *A European perspective on maize history.* C R Biol, 2011. **334**(3): p. 221-8.

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

9. Schnable, P.S., et al., *The B73 maize genome: complexity, diversity, and dynamics.* Science, 2009. **326**(5956): p. 1112-5.
10. Jiao, Y., et al., *Improved maize reference genome with single-molecule technologies.* Nature, 2017. **546**: p. 524.
11. Lai, J., et al., *Gene loss and movement in the maize genome.* Genome research, 2004. **14**(10A): p. 1924-1931.
12. Springer, N.M., et al., *Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content.* PLoS Genet, 2009. **5**(11): p. e1000734.
13. Swanson-Wagner, R.A., et al., *Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor.* Genome Res, 2010. **20**(12): p. 1689-99.
14. Wang, Q. and H.K. Dooner, *Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus.* Proc Natl Acad Sci U S A, 2006. **103**(47): p. 17644-9.
15. Hirsch, C.N., et al., *Draft Assembly of Elite Inbred Line PH207 Provides Insights into Genomic and Transcriptome Diversity in Maize.* Plant Cell, 2016. **28**(11): p. 2700-2714.
16. Springer, N.M., et al., *The maize W22 genome provides a foundation for functional genomics and transposon biology.* Nature Genetics, 2018. **50**(9): p. 1282-1288.
17. Sun, S., et al., *Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes.* Nature Genetics, 2018. **50**(9): p. 1289-1295.
18. Unterseer, S., et al., *A comprehensive study of the genomic differentiation between temperate Dent and Flint maize.* Genome Biol, 2016. **17**(1): p. 137.
19. Darracq, A., et al., *Sequence analysis of European maize inbred line F2 provides new insights into molecular and chromosomal characteristics of presence/absence variants.* BMC Genomics, 2018. **19**(1): p. 119.
20. Messmer, M.M., et al., *RFLP analyses of early-maturing European maize germ plasm : I. Genetic diversity among flint and dent inbreds.* Theor Appl Genet, 1992. **83**(8): p. 1003-12.
21. International Wheat Genome Sequencing, C., et al., *Shifting the limits in wheat research and breeding using a fully annotated reference genome.* Science, 2018. **361**(6403).
22. Springer, N.M., et al., *The maize W22 genome provides a foundation for functional genomics and transposon biology.* Nat Genet, 2018. **50**(9): p. 1282-1288.
23. Simão, F.A., et al., *BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.* Bioinformatics, 2015. **31**(19): p. 3210-3212.
24. Brohammer, A.B., et al., *Limited role of differential fractionation in genome content variation and function in maize (Zea mays L.) inbred lines.* bioRxiv, 2017: p. 179119.
25. El-Brolosy, M.A., et al., *Genetic compensation triggered by mutant mRNA degradation.* Nature, 2019. **568**(7751): p. 193-197.
26. Ma, Z., et al., *PTC-bearing mRNA elicits a genetic compensation response via Upf3a and COMPASS components.* Nature, 2019. **568**(7751): p. 259-263.
27. Makarevitch, I., et al., *Transposable elements contribute to activation of maize genes in response to abiotic stress.* PLoS Genet, 2015. **11**(1): p. e1004915.
28. Anderson, S.N., et al., *Transposable elements contribute to dynamic genome content in maize.* bioRxiv, 2019: p. 547398.
29. Ghaffari, R., et al., *Maize chromosomal knobs are located in gene-dense areas and suppress local recombination.* Chromosoma, 2013. **122**(1-2): p. 67-75.
30. Stack, S.M., et al., *Meiotic Crossing Over in Maize Knob Heterochromatin.* Genetics, 2017. **205**(3): p. 1101-1112.

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

31. Ananiev, E.V., R.L. Phillips, and H.W. Rines, *Complex structure of knobs and centromeric regions in maize chromosomes.* Tsitol Genet, 2000. **34**(2): p. 11-5.
32. Matsuda, A., et al., *Highly condensed chromatins are formed adjacent to subtelomeric and decondensed silent chromatin in fission yeast.* Nat Commun, 2015. **6**: p. 7753.
33. Kremling, K.A.G., et al., *Dysregulation of expression correlates with rare-allele burden and fitness loss in maize.* Nature, 2018. **555**: p. 520.
34. Unterseer, S., et al., *A powerful tool for genome analysis in maize: development and evaluation of the high density 600 k SNP genotyping array.* BMC Genomics, 2014. **15**: p. 823.
35. Saunders, I.W., J. Brohede, and G.N. Hannan, *Estimating genotyping error rates from Mendelian errors in SNP array genotypes and their impact on inference.* Genomics, 2007. **90**(3): p. 291-6.
36. Salvi, S., et al., *Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize.* Proceedings of the National Academy of Sciences, 2007. **104**(27): p. 11376-11381.
37. Ducrocq, S., et al., *Key Impact of Vgt1 on Flowering Time Adaptation in Maize: Evidence From Association Mapping and Ecogeographical Information.* Genetics, 2008. **178**(4): p. 2433-2437.
38. Bukowski, R., et al., *Construction of the third-generation Zea mays haplotype map.* Gigascience, 2018. **7**(4): p. 1-12.
39. Cook, J.P., et al., *Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels.* Plant Physiol, 2012. **158**(2): p. 824-34.
40. James, M.G., D.S. Robertson, and A.M. Myers, *Characterization of the maize gene sugary1, a determinant of starch composition in kernels.* Plant Cell, 1995. **7**(4): p. 417-29.
41. Fu, F.-F. and H.-W. Xue, *Coexpression Analysis Identifies Rice Starch Regulator1, a Rice AP2/EREBP Family Transcription Factor, as a Novel Rice Starch Biosynthesis Regulator.* Plant Physiology, 2010. **154**(2): p. 927-938.
42. El-Lithy, M.E., et al., *Relation among plant growth, carbohydrates and flowering time in the Arabidopsis Landsberg erecta x Kondara recombinant inbred line population.* Plant Cell Environ, 2010. **33**(8): p. 1369-82.
43. Wahl, V., et al., *Regulation of flowering by trehalose-6-phosphate signaling in Arabidopsis thaliana.* Science, 2013. **339**(6120): p. 704-7.
44. Yang, J., C. Ji, and Y. Wu, *Divergent Transactivation of Maize Storage Protein Zein Genes by the Transcription Factors Opaque2 and OHPs.* Genetics, 2016. **204**(2): p. 581-591.
45. Yao, D., et al., *Maize opaque10 Encodes a Cereal-Specific Protein That Is Essential for the Proper Distribution of Zeins in Endosperm Protein Bodies.* PLoS Genet, 2016. **12**(8): p. e1006270.
46. Yi, G., et al., *The naked endosperm genes encode duplicate INDETERMINATE domain transcription factors required for maize endosperm cell patterning and differentiation.* Plant Physiol, 2015. **167**(2): p. 443-56.
47. Zhang, S., J. Zhan, and R. Yadegari, *Maize opaque mutants are no longer so opaque.* Plant Reprod, 2018. **31**(3): p. 319-326.
48. Schnable, P.S. and N.M. Springer, *Progress Toward Understanding Heterosis in Crop Plants.* Annual Review of Plant Biology, 2013. **64**(1): p. 71-88.
49. Brown, W.L., *Numbers and Distribution of Chromosome Knobs in United States Maize.* Genetics, 1949. **34**(5): p. 524-536.

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

50. Adawy, S.S., R.M. Stupar, and J. Jiang, *Fluorescence in situ hybridization analysis reveals multiple loci of knob-associated DNA elements in one-knob and knobless maize lines.* J Histochem Cytochem, 2004. **52**(8): p. 1113-6.
51. Jian, Y., et al., *Maize (Zea mays L.) genome size indicated by 180-bp knob abundance is associated with flowering time.* Scientific Reports, 2017. **7**(1): p. 5954.
52. Kanizay, L.B., et al., *Intragenomic conflict between the two major knob repeats of maize.* Genetics, 2013. **194**(1): p. 81-89.
53. Hufford, M.B., et al., *Comparative population genomics of maize domestication and improvement.* Nature Genetics, 2012. **44**: p. 808.
54. Lai, J., et al., *Genome-wide patterns of genetic variation among elite maize inbred lines.* Nature Genetics, 2010. **42**: p. 1027.
55. Reif, J.C., et al., *Trends in genetic diversity among European maize cultivars and their parental components during the past 50 years.* Theor Appl Genet, 2005. **111**(5): p. 838-45.

Acknowledgements

We thank Nina Opitz (University of Bonn) for helping to collect maize tissue for RNA-seq.

Funding

This study was funded by the Bavarian State Ministry of the Environment and Consumer Protection (Grant ID TGC01GCUFuE69741; Project BayKlimaFit; http://www.bayklimafit.de/) and the German Federal Ministry of Education and Research (BMBF) within the funding initiative “Plant Breeding Research for the Bioeconomy” (Grant ID 031B0195A; Project MAZE; http://www.europeanmaize.net). In addition K.F.X Mayer acknowledges support from the BMBF within the project de.NBI Grant ID 031A536B.

Author contributions

S.C-C., M.KFX., O.M. and H.G. designed the study. M.KFX. and H.G. wrote the manuscript. B.E. and S.C-C. supervised the genomic sequencing. B.E., U.C. and N.A. generated the F2 population and constructed the genetic map. H.G. analyzed genes and whole genome alignments. G.H. performed the repeat analysis. H.A. and R.A. contributed the FISH analysis. M.C. and H.F. prepared and sequenced the RNAseq samples. K.N., S.MA. and F.I. conducted the DEG expression analysis and the phylogenetic analysis. S.A. and S.MA. provided bioinformatics support. All authors critically read and edited the manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Plant material and Genome assembly. The four flint inbred lines were chosen to represent landraces from different European ancestry. While PE0075, a doubled-haploid line derived from the landrace Petkuser Ferdinand Rot, and DK105, derived from Gelber Badischer Landmais, can be classified as a representatives of the Northern flints, EP1 and F7 represent Pyrenean-Galician ancestry and were derived from populations Lizargarate and Lacaune, repectively [56]. EP1, F7 and DK105 are important founder line of European breeding programs. All four lines were assembled using the DeNovoMAGIC pipeline of NRGENE (Energin .R Technologies 2009 Ltd.; Ness Ziona, Israel). The application of this toolset has been previously described for several plant genome assemblies [15, 16, 57]. Briefly, genomic DNA was prepared from bulked leaf tissue from 15-24 seedlings per line. Leaf tissue was harvested, immediately ground in liquid nitrogen and DNA was isolated using a modified CTAB protocol [58]. Illumina sequencing of paired-end (PE) and mate-pair (MP) libraries was performed as specified in Supplementary Table 1. Contigs, scaffolds and pseudo-chromosomes were assembled *de novo* using the DeNovoMAGIC 2.0 technology for lines EP1, F7, and DeNovoMAGIC 3.0 for lines DK105 and PE0075, respectively, using the B73v4 reference genome to assist anchoring and orienting scaffolds and contigs to pseudo-chromosomes (Supplementary Table 2).

A cross of parental lines PH207 x EP1 was used to establish a high-density genetic map from an F_2 mapping population. DNA from the parental lines and 192 F_2 plants were analysed using the Affymetrix Axiom™ Maize 600k Array and data were processed as described by [34]. Genotype calls from the quality category “Poly High Resolution” were filtered for markers polymorphic between the parental lines, allowing no missing values, which resulted in 174,616 markers. This dataset was processed through a script from the POPSEQ pipeline to cluster markers into groups showing identical segregation patterns [59]. A first genetic map was calculated with 9404 binmap markers with the R package ASMap v. 1.0-2 [60] using the function mstmap with the following parameters: pop.type="RIL2", dist.fun="kosambi", objective.fun="COUNT", p.value=1e-22, noMap.dist=15, noMap.size=2, miss.thresh=0.00. Linkage groups were assigned to the ten maize chromosomes based on previously mapped markers. Unlinked small groups with only a few markers were discarded. Ten F2 plants exhibited very high numbers of crossovers and were excluded from further analyses. In multiple rounds of mapping using the same parameters as stated above, markers with highly distorted segregation or which led to double-crossovers were identified using the function statMark in the ASMap R package and discarded before final map construction. The final genetic linkage map contained 8869 markers. All markers from the initial dataset which had a Hamming distance of 0 with one of the mapped binmap markers were inserted into the map, resulting in a genetic map with 174,071 markers.

Gene annotation, Consolidation and Synteny. We predicted protein-coding structures for F7 and EP1 as consensus models using an approach as previously described [61, 62]. Briefly, consensus gene models are based on transcriptome evidences from F7 and EP1 RNAseq data generated within this study (see RNA preparation and expression analysis) and protein homologies of known monocotyledonous protein sequences (maize B73 and PH207, *Sorghum bicolor* v3.1., *Oryza sativa* v7 MSU and *Brachypodium distachyon* v3.1) [61, 63, 64]. Evidences were mapped onto the genome assembly applying GenomeThreader [65] using a minimum alignment coverage of 50% and seed sizes between 7-10 for protein and 18 for nucleic acid

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

matches. Before mapping, transcript data were assembled using Trinity with default parameters and Bridger with kmer sizes 25 and 29 bp [66, 67]. Subsequently, these sequences were combined by the evidential gene pipeline (http://arthropods.eugenes.org/EvidentialGene) to obtain the final EP1 and F7 transcriptome assemblies. Additional candidate loci for the initial gene sets of EP1 and F7, and the original representative gene models of B73 and PH207 [68] were searched in all six genome sequences (including PE0075 and DK105) by blat alignments [69]. We denote gene models used for the mapping as 'informants', and those derived from informant mappings to another genome sequence as 'target' models. Homologous loci were further refined by exonerate surveys [70] spanning the genomic region detected by blat and 20 kb up- and downstream sequence. Blat and exonerate alignments showed a linear correlation allowing direct conversion and comparison of both scoring schemes ($r>0.999$). For each blat/exonerate target pair, the top scoring model was selected and added to the initial models if the target (i) did not overlap with existing gene models, (ii) had a contiguous ORF, and (iii) ≥95% of the informant sequence was covered by the target. The resulting gene set constituted gene set 1 (GS1).

Current gene annotations largely rely on automatic pipelines and still contain poor gene models like partial, merged or transposon-derived structures. Cross mapping bears the danger to transfer such models between each genome and thereby potentially enriches poor annotations. To estimate this effect and to evaluate gene-based statements of this study, we trained a multi-layer perceptron (MLP) classifying genes into low (LC) and high confidence (HC) models based on their consistency to known protein structures of other annotations. We applied the implementation of scikit-learn (https://scikit-learn.org) of an MLPclassifier using backpropagation learning with the stochastic gradient-based optimizer, and two hidden layers of 6 and 3 neurons, respectively. A training set was based on swissprot maize proteins and a set COG clusters [71] that were constructed from bi-directional best blast hits (bbhs) between 10 plant species: *Arabidopsis thaliana, Oryza sativa, Hordeum vulgare, Triticum aestivum, Sorghum bicolor, Sertaria italica, Brachypodium distachyon, Ananas comosus, Leersia oryzoides, Phyllostachus edulis*. At least three mutual bbhs with an alignment coverage ≥95% defined a COG. The high confidence (HC) training set ('positives') consisted of maize swissprot proteins entirely matching a maize fl-cDNA [72] or B73 proteins with an alignment coverage ≥95% to one of the COG clusters above. Maize genes with matches <50% coverage to any of the COG cluster genes formed the low confidence (LC) set ('negatives'). For each gene in both sets, we constructed an 18-tuple feature vector recording identity, query and hit coverage of its top blastp match against Arabidopsis, Rice, Sorghum, Brachypodium, emmer and all curated swissprot proteins. Positives and negatives were well separated by PCA, and 10x cross-validation showed a high accuracy >99% of the MLPclassifier. A set of 3,917 transposon genes identified from GS1 based on Interpro domains and description lines (https://github.com/groupschoof/AHRD) was excluded from training data and separately classified afterwards. It should be emphasized that LC genes cannot solely considered as biologically non-functional genes but comprise both, true genes displaying erroneous structures caused by problematic underlying evidences, sequences, gene calls etc., as well as potential over-predictions.

To derive gene set 2 (GS2), we first selected HC singletons and genes from syntelog clusters (see below) that comprised only syntelogs of maximal 5 lines and/or syntelogs differing from the CDS mean size of the respective cluster by > 5%. Each of these gene models were cross-mapped to the other five lines using the blat/exonerate alignment procedure described above. We

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

collected all matches between the target and the informant genome that were located at a neighboring (±5 Mb) syntenic position as defined by the WGA (see below). If the informant model was located outside of an alignment, we approximated its syntenic position using WGAs framing the informant and set the midpoint in the target genome proportional to the distances to both flanking alignments. Subsequently, for each syntelog cluster we determined the top scoring combination of matches over all lines to infer gene models that optimize the fit and score for all participating lines and matches of this cluster. First, GS1 models and all selected matches were clustered into groups with congruent CDS sizes (maximal difference ±5%). For each group, we selected the top scoring model per line and summed their scores to obtain a group score. The final annotation was then based on models of the top scoring group and replaced or complemented by the GS1 model in one or more lines of the syntelog cluster. This gene set was further cleaned for transposon derived genes by the removal of singleton genes and syntelog clusters for which more than half of the cluster genes showed Interpro domains matching transposon domains.

We applied i-adhore v3 [73] running the hybrid cluster mode, a tandem gene distance of 10 and a minimum of 5 anchors to generate higher order syntenic relations for all six maize lines. Input pairwise gene similarities were derived from all-against-all blastn searches against coding sequences (CDS). To address putative ambiguities in the syntelog assignment caused by tandem arrays and the WGD in maize, we constructed a syntelog graph G = (V,E) with gene IDs as vertices V and multiplicon pairs provided by i-adhore as edges E. Edge weights were proportional to the size of a multiplicon, ie. the number of syntelogs within one syntenic block. In case of multiple edges from one vertex/gene, we only kept the highest scoring edge(s). This rule was applied in two consecutive steps, first to intra- and then inter-line connectors. In a second post-processing step, remaining clusters with vertices having two or more links to one line were further split into maximal cliques and a score was assigned to each clique as the sum of its edge/multiplicon weights. Next, we disjoint the cluster graph into non-overlapping clique subgraphs with a top-down approach thereby keeping cliques with lower scores only if none of its nodes is contained in the already selected subgraphs. The underlying rationale to select for syntelog cliques is analogous to the bidirectional best blast hit and the COG schemes.

Fluorescent *in situ* hybridization (FISH) and karyotyping. Maize chromosomes were prepared from root meristems of two days old seedlings. Roots were cut and treated with 2mM 8-hydroxyquinoline solution for 3.5 h. Fixation was performed overnight at room temperature in 3: 1 (ethanol: acetic acid) fixative. Slides were prepared according to [74] and *in situ* hybridization was performed as described in [75]. Oligonucleotide probes for karyotyping and identification of knob repeats were chosen from [76]. Following sequences were synthetized and 5'-labeled by Eurofins Genomics (Ebersberg, Germany):

$(ACT)_{10}$: [TAM]ACTACTACTACTACTACTACTACTACT

Knob-2 (180 bp Knob repeat):

[FAM]GAAGGCTAACACCTACGGATTTTTGACCAAGAAATGGTCTCCACCAGAAATCCAAAAAT

MR68-3 (Chromosome 6 clone MR68, AF020265.1):

[FAM]CTCTCGTGCGAATACAATGCCCTCAATATCATAGAAACACTATTCCTTTAGTGTGAATA

CentC69-1 (Centromeric repeat clone CentC69, AY530284.1):

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[CY5]CCCAATCCACTACTTTAGGTCCAAAACTCATGTTTGGGGTGGTTTCGCGCAATTTCGT
T

Images were taken using an Olympus BX61 microscope equipped with an ORCA-ER CCD camera (Hamamatsu). All images were acquired in grey scale and pseudocoloured with Adobe Photoshop CS5 (Adobe Systems). Karyotyping was done according to [76] using the maize line B73 as a reference. Chromosomes were identified based on distribution of repetitive sequences, morphology and centromeric index.

Repeat analysis. To obtain equivalent transposon and tandem repeat data for comparative analyses we annotated all six lines with the same annotation workflows. A homology search against the Panicoidae section of the PGSB transposon library [77] resulted in a basal transposon annotation. The used REdat_9.8_Panicoideae contains publicly available Panicoideae transposons templates as well as *de novo* detected full length LTR-retrotranspons from Maize (12,510 elements from B73v2) and Sorghum (3.368 elements). The program vmatch (http://www.vmatch.de), was used as a fast and efficient matching tool well suited for large and highly repetitive genomes under the following parameter setup: identity>=70%, minimal hit length 75 bp, seedlength 12 bp (exact commandline: -d -p -l 75 -identity 70 -seedlength 12 -exdrop 5). The vmatch output was filtered for redundant hits via a priority based approach, which assigns higher scoring matches first and either shortens (<90% coverage and >=50bp rest length) or removes lower scoring overlaps. The resulting annotation is free of overlaps. Elements that have been interrupted by other transposon insertions (nesting) are not defragmented into a higher order instance (like exons belonging to one gene).

Full-length LTR retrotransposons (fl-LTR) where identified with LTRharvest [78] using the following parameters: "overlaps best -seed 30 -minlenltr 100 -maxlenltr 2000 -mindistltr 3000 -maxdistltr 25000 -similar 85 -mintsd 4 -maxtsd 20 -motif tgca -motifmis 1 -vic 60 -xdrop 5 -mat 2 -mis -2 -ins -3 -del -3". All candidates from the LTRharvest output were subsequently annotated for PfamA domains using hmmer3 (http://hmmer.org) and stringently filtered for false positives by several criteria, the main ones being the presence of at least one typical retrotransposon domain [e.g. reverse transcriptase (RT), RNase H (RH), integrase (INT), protease (PR), etc.)] and a tandem repeat content < 25%. The inner domain order served as a criterion for the classification into the Gypsy (RT-RH-INT) or Copia (INT-RT-RH) superfamily abbreviated as RLG and RLC. Elements missing either INT or RT were classified as RLX. The insertion age of each full-length LTR retrotransposon was estimated based on the accumulated divergence between its 5' and 3' long terminal repeats and a random mutation rate of 1.3×10-8 [79].

Tandem repeats where identified with the TandemRepeatFinder under default parameters [80] and subjected to an overlap removal as decribed above prioritizing longer and higher scoring elements. Kmer frequencies were calculated with Tallymer [81].

Syntentic fl-LTRs where identified by sequence clustering (vmatch dbcluster, 98% identity and 98% coverage) of TE-junctions which consisted of 2x100 pb sequence signatures spanning the upstream and downstream insertion sites with each 50bp inside and 50 bp outside of the TE element.

Whole genome alignments. Block alignments of high identity were generated for all pairwise combinations of the six maize lines using the MUMMER v3 suite [82]. Initial alignments were

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

computed using nucmer with a minimal cluster size of 250 bp (-c 250) and a seed size of 20 bp (-l 20). Results were piped through the delta-filter tool selecting the best one-to-one blocks (global option -1) with a minimal size of 500bp (-l 500). Despite this filter step, a large number (~41-81k) of putatively translocated blocks aligned small regions between different chromosomes or unanchored scaffolds. These blocks frequently overlapped 'regular' blocks linking same orthologous chromosomes and showed significantly lower sequence identities in comparison to their respective regular blocks with which they overlapped. Hence, they likely represented paralogous alignments triggered by tandem repeats or the absence of the truly orthologous sequences in one of the lines. To derive the final set of single alignment blocks (SAB), we only included candidate inter-chromosomal translocations with a minimal sequence identity of 99% and a maximum overlap of 10bp to its adjacent alignments. To gain an overview of the contiguity between the six maize genomes, SABs were connected to merged alignment blocks (MAB) if SABs were directly adjacent in both genomes and had a consistent orientation (Supplementary Figure 3). MABs therefore comprise genomic regions of conserved global order and skip genomic parts that were not aligned due to larger sequence differences or undefined sequences (stretches of 'N's).

To identify and delineate genomic regions for each line that align (i) to none of the other lines (designated as 'unaligned'); (ii) to all lines of the same group (either flint or dent maize) and to none of the other group (aka 'group-specific); (iii) to all five other lines ('core6'); we superimposed SABs and MABs of one line on its genome coordinates by recording from which group and how many times a single genomic base participated in an alignment. Thereby, we classified each base and concatenated adjacent bases of the same type (i-iii, see above) to derive the unaligned, group-specific and core genomic part. Reversing this approach, identified core regions of EP1 were re-projected onto the pairwise alignments to decode the sequence and coordinates of matching core elements in the other five lines. Based on these coordinates, sequences for each core block were extracted and aligned applying the Fast Sequence Aligner [83] with default parameters. SNPs were directly determined from these multiple sequence alignments (omitting insertions/deletions) and their position in each of the six genomes was derived from the position in the alignment and the known start and end genomic positions for the core block. To deduce higher order haplotypes, we used seeds of the 15 pairwise runs-of-identities that showed for 40 consecutive SNPs an identical haplotype for a particular combination of lines. An iterative greedy algorithm linking the seed with adjacent up- and downstream runs extended such seeds if the run exhibited a haplotype identical to the seed and less than 3 non-matching SNPs were observed between the current end of the extension and the candidate run.

RNA Preparation and Expression analysis. Different tissues of the European maize (*Zea mays*) inbred lines EP1 and F7 were sampled according to [84, 85] and subsequently subjected to transcriptome sequencing. Supplementary Table 3 provides a detailed description of the collected tissues. Samples 1-3 were collected from seeds imbibed (whole seed, sample 1) or germinated (primary root, sample 2; coleoptile, sample 3) in paper rolls [86] in a 16 h light (28 °C) and 8 h dark (21 °C) regime of a growth cabinet (Conviron CMP6010, http://www.conviron.com). Tissue samples 4-24 were taken from plants grown in a climate chamber either in small pots (14 cm top diameter, 8 cm height, 0,25 l volume, samples 4-17) or in big pots (28 cm top diameter, 21 cm height, 10 l volume, samples 18-24) containing soil substrate type ED 73 (https://www.meyer-shop.com). Growth conditions in the climate chamber were 16 h light (28 °C) and 8 h dark (21

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

°C). Two individual plants were collected for each of the 24 tissues. Harvested plant material was immediately frozen in liquid nitrogen and stored at -80 °C until RNA extraction. For total RNA extraction, the 24 samples per genotype were separately ground in liquid nitrogen. Subsequently, one spatula of each of the 24 samples per genotype was mixed to generate one pool per inbred line EP1 and F7, respectively. Total RNA was extracted from each of the two pools using the Qiagen RNeasy mini kit according to the manufacturer‘s protocol (Qiagen, https://www.qiagen.com), including on-column DNA digestion. RNA quality was determined by agarose gel electrophoresis and by a Bioanalyzer using an Agilent RNA 6000 Nano Chip (Agilent Technologies, https://www.agilent.com). Both RNA samples were of excellent quality with RNA integrity number values [87] between 9.8 and 10. In total, 9 µg total RNA per pool were used for RNA-seq at Novogene (Novogene Co. Ltd, Bejing, China). After mRNA enrichment using oligo(dT) beads, sequencing libraries were constructed with the NEBNext® Ultra™RNA Library Prep Kit for Illumina® (New England Biolabs, Ipswich, USA), and Illumina PE (150 bp x 2) sequencing was performed on a HiSeq4000 by Novogene. The number of raw reads and quality-filtered reads are shown in Supplementary Table 4. Reads containing adapters, reads containing more than 10% undefined bases and reads containing more than 50% low quality bases (Qscore 5) were removed.

Expression analysis and analysis of differentially expressed genes (DSG). We identified regions in SAB, which were fixed within either flint or dent but segregating between the two germplasm of the six lines under study. The corresponding variants were extracted from the Maize Hapmap v3.2.1 panel [38] for lines of the 282 inbred maize association panel [88]. Selected variants were subjected to phylogenetic analysis using FastTree (Version 2.1.5 SSE3) [89] with default parameters and the resulting phylogenetic tree was visualized using iTOL [90] (Supplementary Figure 13). Two non-overlapping subtrees were selected such that three of the lines presented in this study (B73, EP1 and F7) are contained together with other topologically close lines. All lines in the subtree containing B73 were assigned the group one, all those in the subtree containing both EP1 and F7 were assigned group two.

For the resulting subset of of 41 lines (Supplementary Figure 13) samples from seven different tissues (germinating shoot and root, the tip and base of leaves as well as leaf samples under light and dark conditions and kernels) were downloaded from SRA (PRJNA383416) [91]. Trimming of the first thirteen basepairs and cropping of bases with quality values below 20 at the 3’end of reads was performed using bbbduk from the bbmap toolkit [92] as described in the QuantSeq user guide from Lexogen (http://www.bluebee.com/wp-content/uploads/2017/08/QuantSeq-Data-Analysis-Pipeline-User-Guide.pdf).

FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc) was used before and after trimming to check the quality of the reads. Trimmed reads were mapped to the maize reference sequences B73_v4 using the splice aware STAR aligner (version 2.5.1a) [93]. A read was allowed to map in at most 10 locations (–outFilterMultimapNmax 10) with a maximum of 4% mismatches (–outFilterMismatchNoverLmax 0.04) and all non-canonical intron motifs were filtered out (–outFilterIntronMotifs RemoveNoncanonicalUnannotated). The SAM-format files were converted to BAM-format files and sorted using SAMtools [94]. To obtain non-unique gene-level counts from the mapping files, HTSeq (version 0.9.1) with the ‘nonunique all’-method was used [95]. Normalization of read counts was performed by library sequence depth using the R-package

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DESeq2 (version 1.23.3) [96]. For differential gene expression analysis with DESeq2 lines classified into either the flint or dent like group were treated as biological replicates. A gene was considered as differentially expressed when the Benjamini-Hochberg adjusted p-value was equal or below 0.05.

Data availability. The *de novo* assembled genomes and raw reads were released in collaboration with the Maize Genetics and Genomics Database MaizeGDB [18] and NCBI BioProjects PRJNA360923 (F7, DK105 and PE0075) and PRJNA360920 (EP1). Additionally, RNAseq raw data were also deposited under these NCBI accessions.

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

56. Mayer, M., et al., *Is there an optimum level of diversity in utilization of genetic resources?* Theoretical and Applied Genetics, 2017. **130**(11): p. 2283-2295.
57. Avni, R., et al., *Wild emmer genome architecture and diversity elucidate wheat evolution and domestication.* Science, 2017. **357**(6346): p. 93-97.
58. Saghai-Maroof, M.A., et al., *Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics.* Proceedings of the National Academy of Sciences, 1984. **81**(24): p. 8014-8018.
59. Mascher, M., et al., *Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ).* The Plant Journal, 2013. **76**(4): p. 718-727.
60. Taylor, J. and D. Butler, *R Package ASMap: Efficient Genetic Linkage Map Construction and Diagnosis.* 2017, 2017. **79**(6): p. 29.
61. Paterson, A.H., et al., *The Sorghum bicolor genome and the diversification of grasses.* Nature, 2009. **457**: p. 551.
62. Wang, W., et al., *The Spirodela polyrhiza genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle.* Nature Communications, 2014. **5**: p. 3311.
63. Kawahara, Y., et al., *Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data.* Rice, 2013. **6**(1): p. 4.
64. The International Brachypodium, I., et al., *Genome sequencing and analysis of the model grass Brachypodium distachyon.* Nature, 2010. **463**: p. 763.
65. Gremme, G., et al., *Engineering a software tool for gene structure prediction in higher organisms.* Information and Software Technology, 2005. **47**(15): p. 965-978.
66. Chang, Z., et al., *Bridger: a new framework for de novo transcriptome assembly using RNA-seq data.* Genome Biol, 2015. **16**: p. 30.
67. Grabherr, M.G., et al., *Full-length transcriptome assembly from RNA-Seq data without a reference genome.* Nat Biotechnol, 2011. **29**(7): p. 644-52.
68. Portwood, J.L., 2nd, et al., *MaizeGDB 2018: the maize multi-genome genetics and genomics database.* Nucleic Acids Res, 2019. **47**(D1): p. D1146-D1154.
69. Kent, W.J., *BLAT--the BLAST-like alignment tool.* Genome Res, 2002. **12**(4): p. 656-64.
70. Slater, G.S. and E. Birney, *Automated generation of heuristics for biological sequence comparison.* BMC Bioinformatics, 2005. **6**: p. 31.
71. Tatusov, R.L., et al., *The COG database: a tool for genome-scale analysis of protein functions and evolution.* Nucleic Acids Res, 2000. **28**(1): p. 33-6.
72. Soderlund, C., et al., *Sequencing, mapping, and analysis of 27,455 maize full-length cDNAs.* PLoS Genet, 2009. **5**(11): p. e1000740.
73. Proost, S., et al., *i-ADHoRe 3.0--fast and sensitive detection of genomic homology in extremely large data sets.* Nucleic Acids Res, 2012. **40**(2): p. e11.
74. Aliyeva-Schnorr, L., L. Ma, and A. Houben, *A Fast Air-dry Dropping Chromosome Preparation Method Suitable for FISH in Plants.* Journal of Visualized Experiments : JoVE, 2015(106): p. 53470.
75. Badaeva, E.D., et al., *In Situ Hybridization to Plant Chromosomes*, in *Fluorescence In Situ Hybridization (FISH): Application Guide*, T. Liehr, Editor. 2017, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 477-494.
76. Zhu, M., et al., *A simple and efficient non-denaturing FISH method for maize chromosome differentiation using single-strand oligonucleotide probes.* Genome, 2017. **60**(8): p. 657-664.

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

77. Spannagl, M., et al., *PGSB PlantsDB: updates to the database framework for comparative plant genome research.* Nucleic Acids Res, 2016. **44**(D1): p. D1141-7.
78. Ellinghaus, D., S. Kurtz, and U. Willhoeft, *LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons.* BMC Bioinformatics, 2008. **9**: p. 18.
79. SanMiguel, P., et al., *The paleontology of intergene retrotransposons of maize.* Nat Genet, 1998. **20**(1): p. 43-5.
80. Benson, G., *Tandem repeats finder: a program to analyze DNA sequences.* Nucleic Acids Res, 1999. **27**(2): p. 573-80.
81. Kurtz, S., et al., *A new method to compute K-mer frequencies and its application to annotate large repetitive plant genomes.* BMC Genomics, 2008. **9**: p. 517.
82. Kurtz, S., et al., *Versatile and open software for comparing large genomes.* Genome Biology, 2004. **5**(2): p. R12.
83. Bradley, R.K., et al., *Fast statistical alignment.* PLoS Comput Biol, 2009. **5**(5): p. e1000392.
84. Downs, G.S., et al., *A developmental transcriptional network for maize defines coexpression modules.* Plant Physiol, 2013. **161**(4): p. 1830-43.
85. Sekhon, R.S., et al., *Genome-wide atlas of transcription during maize development.* Plant J, 2011. **66**(4): p. 553-63.
86. Hetz, W., et al., *Isolation and characterization of rtcs, a maize mutant deficient in the formation of nodal roots.* The Plant Journal, 1996. **10**(5): p. 845-857.
87. Schroeder, A., et al., *The RIN: an RNA integrity number for assigning integrity values to RNA measurements.* BMC Molecular Biology, 2006. **7**(1): p. 3.
88. Flint-Garcia, S.A., et al., *Maize association population: a high-resolution platform for quantitative trait locus dissection.* Plant J, 2005. **44**(6): p. 1054-64.
89. Price, M.N., P.S. Dehal, and A.P. Arkin, *FastTree 2--approximately maximum-likelihood trees for large alignments.* PLoS One, 2010. **5**(3): p. e9490.
90. Letunic, I. and P. Bork, *Interactive Tree Of Life (iTOL) v4: recent updates and new developments.* Nucleic Acids Res, 2019.
91. Kremling, K.A.G., et al., *Dysregulation of expression correlates with rare-allele burden and fitness loss in maize.* Nature, 2018. **555**(7697): p. 520-523.
92. Bushnell, B., *BBMap: A Fast, Accurate, Splice-Aware Aligner*. Conference: 9th Annual Genomics of Energy & Environment Meeting, Walnut Creek, CA, March 17-20, 2014. 2014: ; Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States). Medium: ED.
93. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
94. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
95. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-9.
96. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Tables

	***Zea mays* "flint" lines**				***Zea mays* "dent" lines**	
Line	**EP1**	**F7**	**DK105**	**PE0075**	**B73v4**	**PH207**
assembly size [Mb]	2455.3	2392.8	2288.2	2198.5	2135.1	2156.2
chromosome [Mb] [a]	2321.0	2255.5	2176.1	2140.7	2106.3	2060.3
scaffolds [#]	60567	62610	797	972	267	43291
scaffold N50 [Mb]	6.13	9.48	10.39	8.64	n.a.	n.a.
Undefined/gaps [Mb]	25.48	25.62	14.33	14.67	30.73	442.21
Genes, GS1 [#] [b]	44,958	47,204	46,433	46,482	44,588	44,321
Syntelogs GS1	36,479	36,811	36,729	36,801	37,274	35,158
Genes, GS2 [#] [b]	43,353	44,413	42,683	42,608	42,509	40,782
Syntelogs GS2	38,847	39,126	39,058	39,208	39,526	37,726
Orthologs GS2	40,308	40,483	40,560	40,621	40,753	39,253
Class I repeats [% of assembly]	79,1	78,6	78,3	77,8	77,2	74,9
Class II repeats [% of assembly]	2,03	2,07	2,11	2,16	2,28	2,29

Table 1. Genome statistics of maize dent and flint lines used in this study.

Table summarizes total assembly size, total size of assembled chromosomes ([a] only 10 pseudo-chromosomes), number of pseudo-chromosomes including unanchored scaffolds, N50 of scaffolds before pseudo-chromosome generation, total size of undefined sequence in the genome, number of genes ([b] referring to the gene sets, see main text and supplement), syntelogs (referring to genes syntenic to at least one line), orthologs (i.e. syntelogs complemented by bidirectional best Blast hits for out-of-syntenic order genes) and total percentages of class I and class II repeats.

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figures

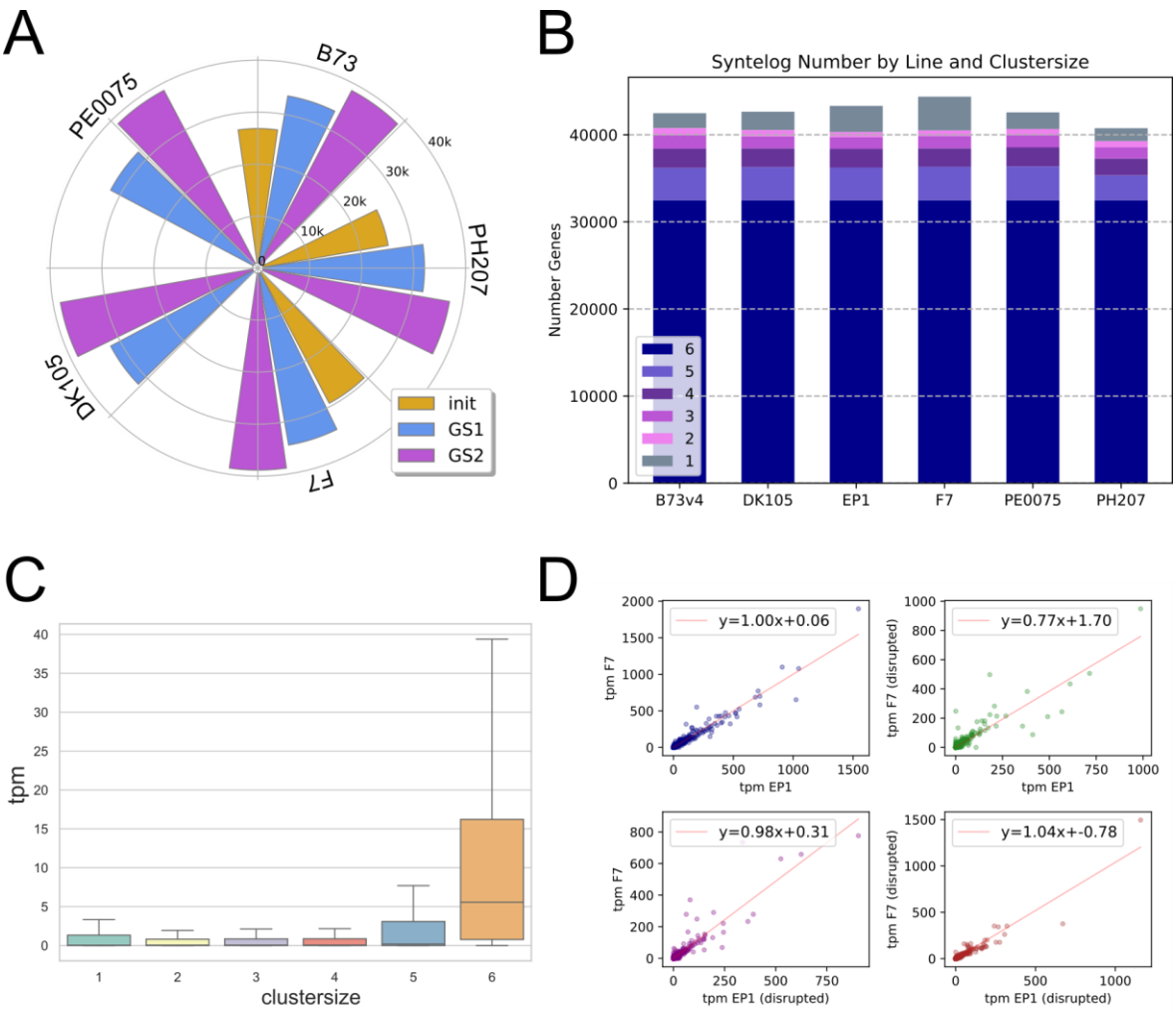


Figure 1. Gene consolidation and characteristics.
(A) illustrates the iterative increase of pairwise syntelogs between EP1 and the other five maize lines applying an annotation consolidation by cross-mapped gene models. Syntelog count increases along the initial prediction ('init') and two consecutive consolidation steps reporting gene set 1 and 2 (GS1, GS2) for all pairwise comparisons. **(B)** shows the number of genes per line that are either singletons (label 1) or are syntenic between 2 and up to 6 lines (labels 2 to 6). **(C)** grades expression levels (tpm) of EP1 and F7 genes by size of their orthologous clusters. Genes either are singletons (cluster size 1) or have orthologues in one and up to five maize lines (cluster sizes 2 to 6). Syntelogs of orthologous clusters comprising 2-4 lines exhibit very low median expression levels (tpm ~0.06). **(D)** Consolidated mappings discovered several thousand genes for which the entire cross mapping of the top-scoring consolidation model had no contiguous but a disrupted ORF. These genes exhibit highly similar expression levels to their non-disrupted syntelogs (top right and bottom left panel). Overall, disrupted syntelogs have an equal linear correlation like non-disrupted syntenic pairs (top left panel).

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

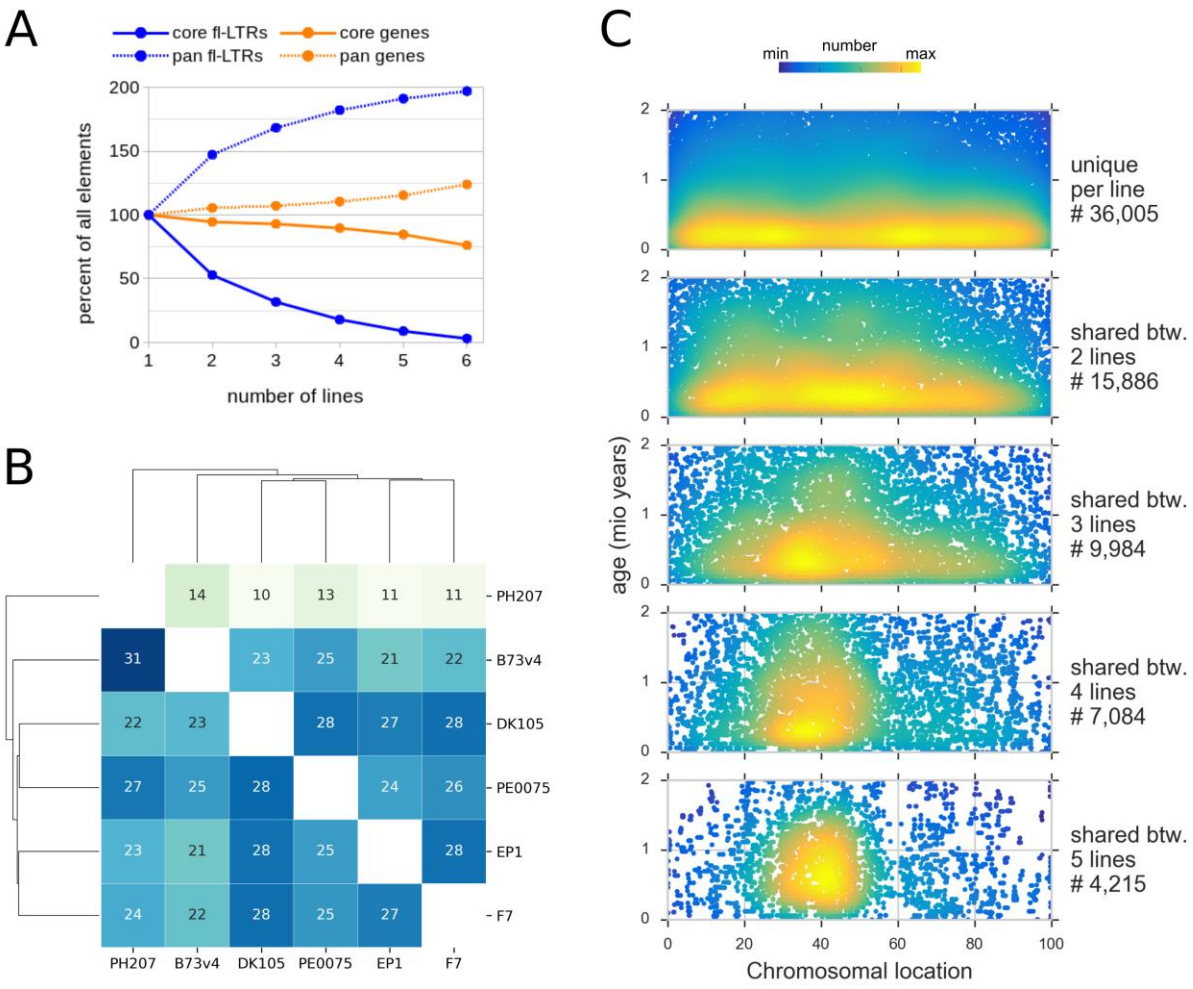


Figure 2. Pan and core characteristics of full length LTR-retrotransposons (fl-LTR).
(A) Proportion of the pan and core sets for fl-LTRs and gene with increasing line numbers. **(B)** Percent of pairwise shared still intact fl-LTRs at syntenic positions for six maize lines. Reading direction is column to row, e.g. EP1 shares 27% of its fl-LTRs with F7 and vice versa F7 28% with EP1. The PH207 row has lower values because of its lower overall fl-LTR amount. Similar reciprocal values indicate a comparable assembly quality level. Based on the similarity matrix the six lines cluster into a relationship context, which distinguishes between flint and dent and places EP1 next to F7 and DK105 close to PE0075. A corresponding evaluation for genes gives pairwise shared numbers between 82 and 91 %. **(C)** Insertion age and chromosomal distribution of line unique and shared fl-LTRs in the different cluster constellations. The line unique fl-LTRs contain a higher proportion of younger elements. There is a continuous shift towards a more pericentromeric location and towards older elements if more lines share the same syntenic fl-LTR locus.

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

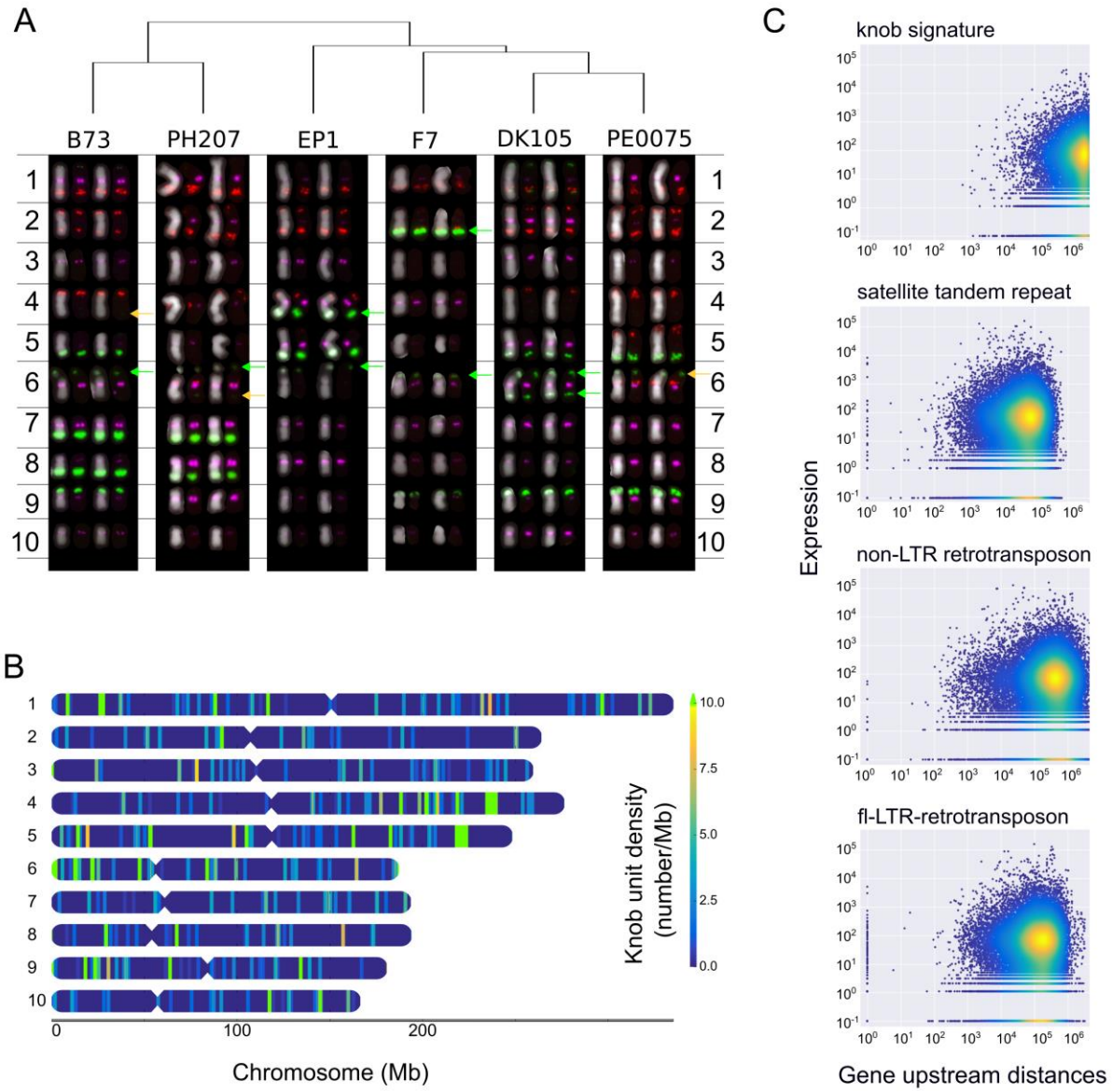


Figure 3: Diversity of knob locations in six maize lines.
(A) Karyotyping of all six maize lines by FISH. Red- (ACT)10, green- Knob-2 (180 bp Knob repeat) and MR68-3 (Chromosome 6 clone MR68, ID AF020265.1), magenta-CentC69-1. Green arrows indicate co-occurrences of Knob-2 and MR68-3, orange arrows a sole occurrence of MR68-3. **(B)** Chromosomal locations of all knob sequences in EP1. Besides the three major Mb scaled knob regions identified by FISH on chromosome 4L, 5L and 6S many smaller knob sequences (<10 kb) are found scattered along all chromosomes. **(C)** Relation between gene neighborhood and gene expression. For each gene, the maximal expression of 7 different conditions is plotted against its upstream distance [bp] to the next neighboring element of a specific type. Both axes are logarithmic. It is noteworthy that especially knob sequences seem to be a "bad" neighborhood for genes. Here, in contrast to other element a 1,000 bp perimeter was found to be devoid of genes.

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

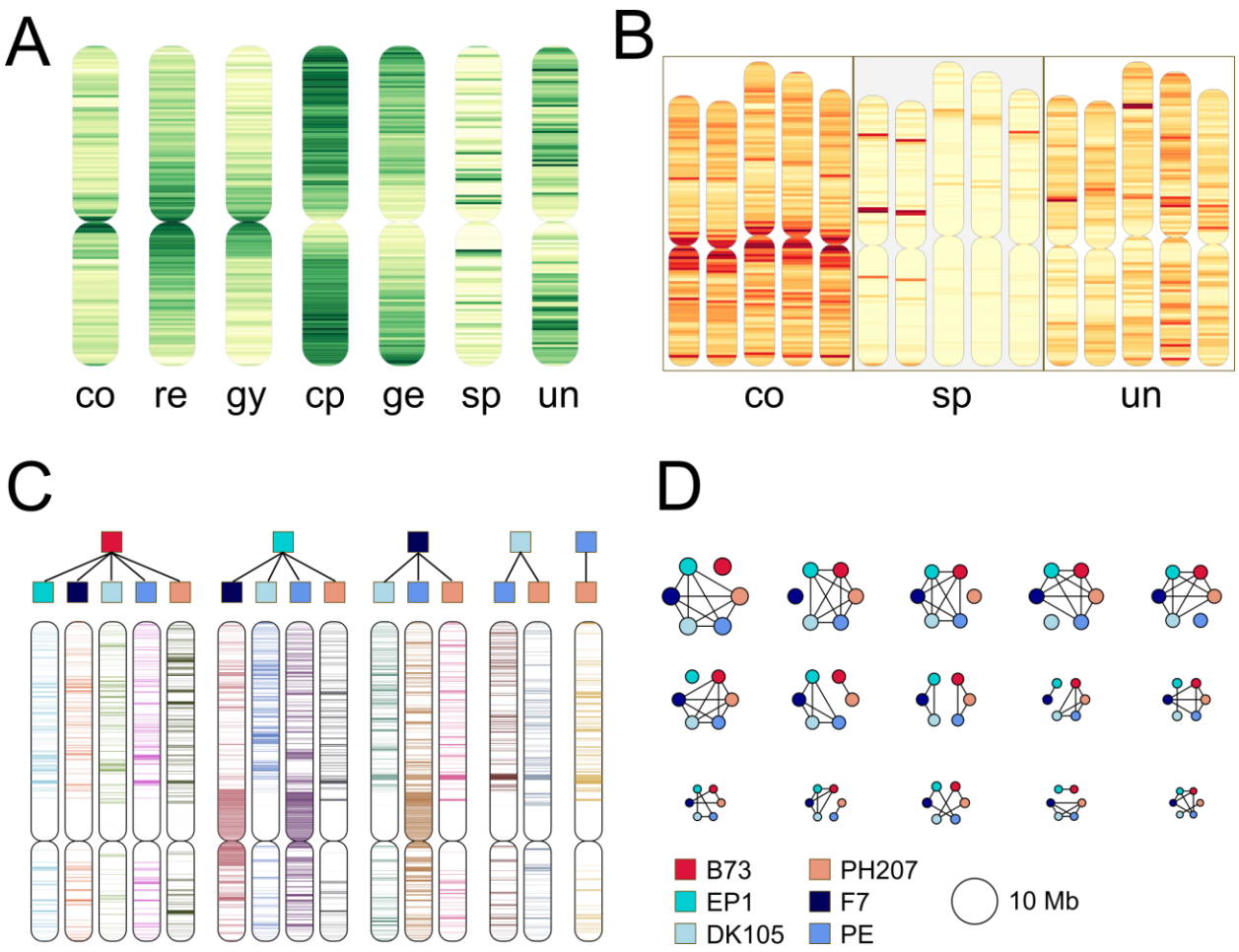


Figure 4. Comparative genome characteristics for the six maize lines.
(A) shows the density distribution of the WGA (co: core, sp: germplasm-specific and un: unaligned regions) and functional genomic elements (re: all repeats, gy and cp: gypsy and copia LTR elements, ge: genes). The core WGAs are significantly positively correlated to the repeat and gypsy densities, similar to densities of copia elements and genes. For the unaligned and germplasm-specific densities, we detected no highly significant correlations with one of the other functional element. **(B)** Densities of the core, germplasm-specific and unaligned WGA regions exhibited similar distributions for all maize lines, exemplified for chr1 of B73, PH207, EP1, F7 and PE0075. Core regions are enriched at (peri-)centromeric regions, germplasm-specific obviously cluster in a group-specific manner and unaligned parts are random within all six lines. **(C)** shows runs-of-identity for SNPs from core regions for chr7 and all 15 pairwise combinations. Color code above the chromosome pictograms follows legend in 4D. **(D)** Sizes of higher order haplotypes of the six lines. Pictograms are proportional to their total genomic span, the legend provides the color coding of each of the six. Black lines indicate identity. Interestingly, the top six higher order haplotypes each group one maize variety as outlier and five lines that are identical by their SNP runs for these genomic segments. Of the more complex groupings, the most prominent haplotype separates the two germplasm, flint and dent.

bioRxiv preprint doi: https://doi.org/10.1101/766444; this version posted September 12, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

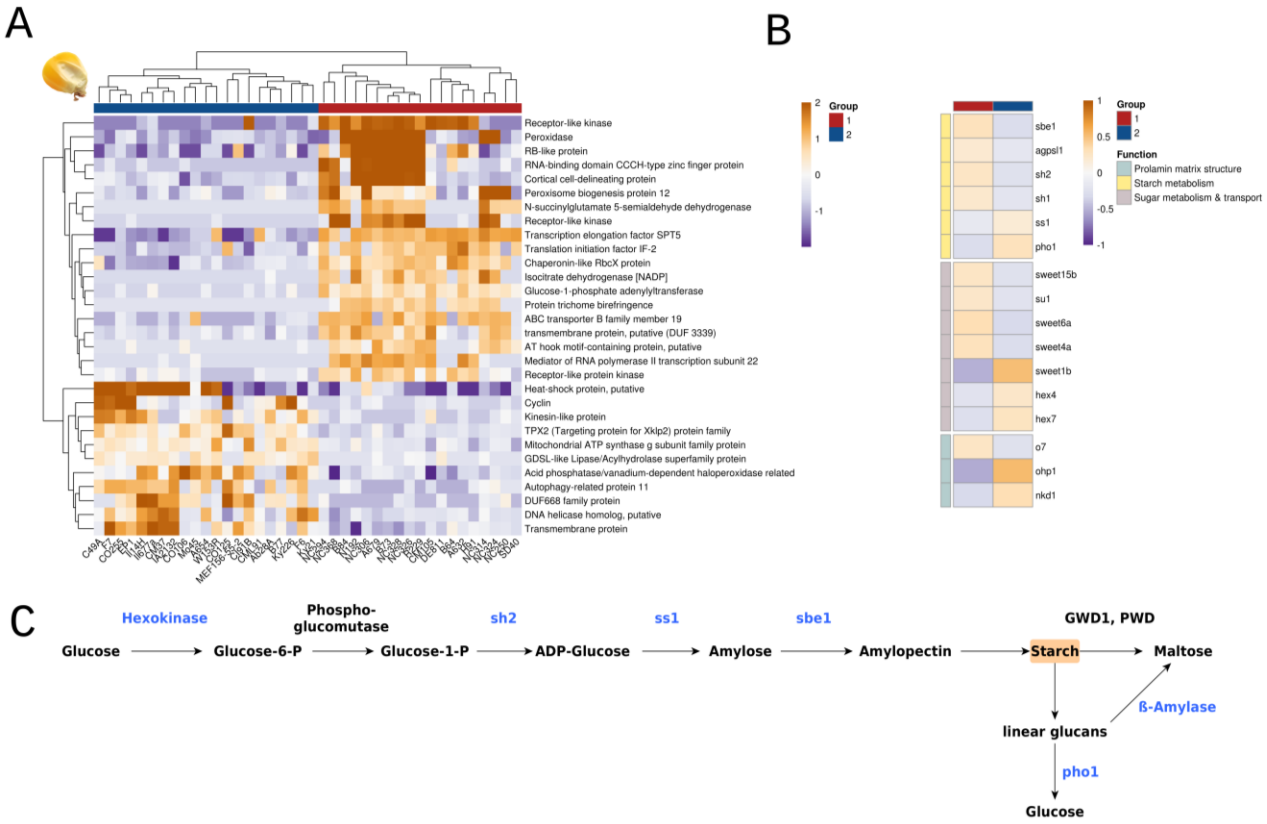


Figure 5. Haploblock-informed differentially expressed genes.

(A) Heat map of top 30 differentially expressed genes (log2 fold change above 2 or below -2, adjusted p-value <= 0.05, variance stabilized) between the two haploblock-informed groups in kernels. Group 1: lines with haploblock-similarity to B73 (dent-alike), group 2: lines with haploblock-similarity to EP1/F7 (flint-alike). **(B)** Differentially expressed genes between the two groups that encode for components of the starch biosynthesis pathway, for proteins involved in sugar metabolism and transport into the kernel and for proteins organizing the prolamin matrix structure of the kernel. (sbe1: starch branching enzyme 1; agpsl1: ADP-glucose pyrophosphorylase small subunit leaf1; sh2: shrunken 2; sh1: shrunken 1; pho1: starch phosphorylase 1; sweet: sugar transporter; su1: sugary 1; hex: hexokinase; o7: opaque7; ohp1: opaque2; nkd1: naked endosperm 1). **(C)** Main reactions of starch biosynthesis and catalyzing enzymes, genes differentially expressed between the two groups of maize lines are indicated in blue color.

Report

Current Biology

Single-Parent Expression Is a General Mechanism Driving Extensive Complementation of Non-syntenic Genes in Maize Hybrids

Highlights

- Gene expression complementation is a general mechanism in maize hybrids
- Maize hybrids express more genes than their parental inbred lines
- Evolutionary younger non-syntenic genes drive expression complementation
- Significant complementation of non-syntenic bZIP and bHLH transcription factors

Authors

Jutta A. Baldauf, Caroline Marcon, Andrew Lithio, ..., Heiko Schoof, Dan Nettleton, Frank Hochholdinger

Correspondence

hochholdinger@uni-bonn.de

In Brief

Baldauf et al. examine the plasticity of the transcriptomic landscape of a diverse panel of maize inbred lines and their F_1 hybrids during primary root development. They demonstrate that extreme expression complementation driven by non-syntenic genes is a general mechanism that extensively shapes the transcriptomic plasticity of hybrids.

Baldauf et al., 2018, Current Biology *28*, 431–437
February 5, 2018 © 2017 Elsevier Ltd.
https://doi.org/10.1016/j.cub.2017.12.027

CellPress

Current Biology
Report

Single-Parent Expression Is a General Mechanism Driving Extensive Complementation of Non-syntenic Genes in Maize Hybrids

Jutta A. Baldauf,[1] Caroline Marcon,[1] Andrew Lithio,[2] Lucia Vedder,[3] Lena Altrogge,[3] Hans-Peter Piepho,[4] Heiko Schoof,[3] Dan Nettleton,[2] and Frank Hochholdinger[1,5,*]
[1]Institute for Crop Science and Resource Conservation, Crop Functional Genomics, University of Bonn, Friedrich-Ebert-Allee 144, 53113 Bonn, Germany
[2]Department of Statistics, Iowa State University, 2438 Osborne Dr., Ames, IA 50011-1210, USA
[3]Institute for Crop Science and Resource Conservation, Crop Bioinformatics, University of Bonn, Katzenburgweg 2, 53115 Bonn, Germany
[4]Institute for Crop Science, Biostatistics Unit, University of Hohenheim, Fruwirthstraße 23, 70599 Stuttgart, Germany
[5]Lead Contact
*Correspondence: hochholdinger@uni-bonn.de
https://doi.org/10.1016/j.cub.2017.12.027

SUMMARY

Maize (*Zea mays* L.) displays an exceptional degree of structural genomic diversity [1, 2]. In addition, variation in gene expression further contributes to the extraordinary phenotypic diversity and plasticity of maize. This study provides a systematic investigation on how distantly related homozygous maize inbred lines affect the transcriptomic plasticity of their highly heterozygous F_1 hybrids. The classical dominance model of heterosis explains the superiority of hybrid plants by the complementation of deleterious parental alleles by superior alleles of the second parent at many loci [3]. Genes active in one inbred line but inactive in another represent an extreme instance of allelic diversity defined as single-parent expression [4]. We observed on average ~1,000 such genes in all inbred line combinations during primary root development. These genes consistently displayed expression complementation (i.e., activity) in their hybrid progeny. Consequently, extreme expression complementation is a general mechanism that results on average in ~600 additionally active genes and their encoded biological functions in hybrids. The modern maize genome is complemented by a set of non-syntenic genes, which emerged after the separation of the maize and sorghum lineages and lack syntenic orthologs in any other grass species [5]. We demonstrated that non-syntenic genes are the driving force of gene expression complementation in hybrids. Among those, the highly diversified families of bZIP and bHLH transcription factors [6] are systematically overrepresented. In summary, extreme gene expression complementation extensively shapes the transcriptomic plasticity of maize hybrids and might therefore be one factor controlling the developmental plasticity of hybrids.

RESULTS AND DISCUSSION

Transcriptome Profiling of a Diverse Panel of Maize Inbred Lines and Their F_1 Hybrids during Early Primary Root Development

Maize exhibits an exceptional degree of intra-specific genomic and phenotypic diversity compared to other crop species [2, 7, 8]. In the present study, RNA sequencing (RNA-seq) was applied to survey how the exceptional degree of genotypic diversity of maize is reflected in the transcriptomic landscape of F_1 hybrids and their parental inbred lines during early primary root development. To this end, we selected seven inbred lines representing six divergent clades of a phylogenetic tree of maize (Figure 1A) [9]. In this tree, the clade harboring the closely related inbred lines B73 and H84 contains stiff stalk lines, whereas all other clades and the genotypes selected from them are non-stiff stalk lines. The inbred line B73 was selected as a common female parent to generate F_1 hybrids in this study because it is the female parent in many hybrid breeding programs and the first maize reference genome sequence was determined in this genotype. The transcriptomes of the six paternal inbred lines and their F_1 hybrids were investigated at three developmental stages early in primary root development (Figure S1A). The three selected stages cover the developmental transition from root hair formation (stage I) via lateral root initiation (stage II) to lateral root formation (stage III). The common maternal inbred line B73 was subjected to RNA-seq in 12 biological replicates per developmental stage to maximize the number of direct comparisons between B73 and other genotypes, whereas all other genotypes were sequenced in four biological replicates per developmental stage (Figure S1B).

Kinship Relations among the Diverse Panel of Genotypes Are Reflected in the Highly Reproducible Transcriptome Data

Relationships among the diverse genotypes and developmental stages were explored in a multidimensional scaling (MDS) plot

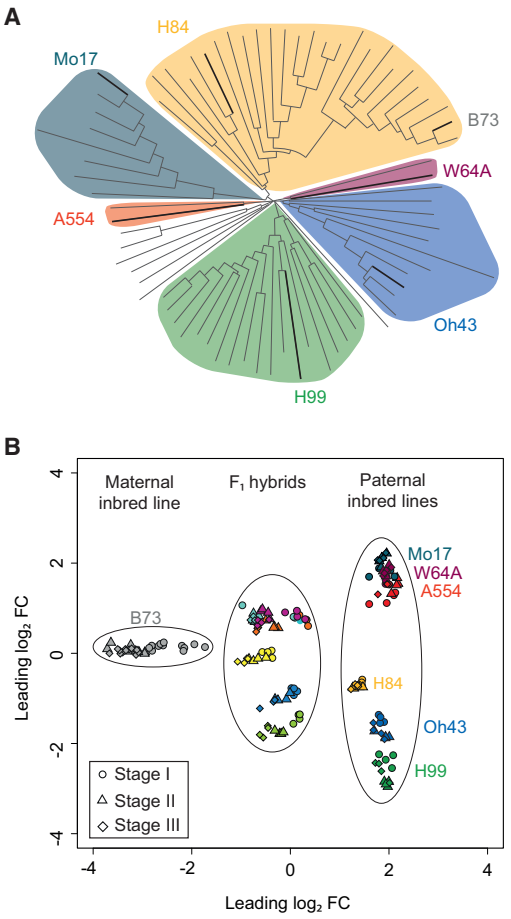


Figure 1. Phylogenetic and Transcriptomic Relationship of Maize Inbred Lines Used in This Study

(A) Selection of maize inbred lines based on their phylogenetic relationship (see also Figure S2). The selected inbred lines are highlighted by a thicker branch. Different clades of the phylogenetic tree are indicated by different colors.

(B) Multidimensional scaling (MDS) plot of all 180 RNA-seq samples. Each genotype is highlighted by a different color (see also Figure S1B). Inbred line clusters are labeled with the corresponding genotype tag. Hybrids are represented in lighter shade of the same color as the paternal inbred line. The three developmental stages are distinguished by the shapes of their features.

(Figure 1B). The biological replicates of specific genotypes and developmental stages form distinct clusters. Furthermore, the MDS plot also depicts the kinship relations among the surveyed genotypes: the clusters of the hybrids are located between the clusters of their two parental inbred lines with each of whom they share 50% of their genome. In addition, the closely related parental stiff stalk inbred lines B73 and H84, which were both selected from the same phylogenetic clade (Figure 1A), revealed a higher transcriptomic similarity than B73 and the remaining paternal inbred lines of the examined non-stiff stalk genotypes (Figure 1B). Hence, the close relationship of the replicated samples and the assembly of the different genotypes according to their genetic relationship highlight the high quality of this dataset. The transcriptomic relationship of the surveyed inbred lines is not only consistent with their phylogenetic relationship but also with their pedigree lineage (Figure S2) [10].

Single-Parent Expression Complementation Is Universal among Different Genotypes and along Development

Single-parent expression (SPE) is an extreme instance of complementation on the gene expression level. In SPE patterns, complementation is observed in hybrids because genes are expressed in only one of its two parental inbred lines [4]. We designate an observed pattern as SPE_B, if genes are expressed in the hybrid and in their maternal inbred line B73, but not in the corresponding paternal inbred line. In contrast, we refer to genes that are expressed in the hybrid and the corresponding paternal inbred line, but not in the maternal parent B73 as SPE_X (X = A554, H84, H99, Mo17, Oh43, and W64A). We determined for each developmental stage and within each parent-hybrid triplet the number of genes showing SPE patterns (Figure 2A; Data S1). Along the three developmental stages and among the different parent-hybrid triplets, between 916 and 1,183 genes display SPE_B patterns (Figure 2A). Among the different parent-hybrid triplets and along the three developmental stages, between 852 and 1,083 genes showed SPE_X patterns (Figure 2A). For all parent-hybrid triplets, fewer SPE_X than SPE_B gene expression patterns were observed (Figure 2A); because of the lack of genomic data, X-specific SPE_X genes could not be identified. Overall, the lowest number of SPE genes was identified between the parental inbred lines of B73 and H84 and their F_1 hybrid. This observation reflects the close kinship relation of these two parental stiff stalk inbred lines that map to the same phylogenetic clade [9] and is also consistent with their pedigree lineage (Figures 1A and S2) [10].

To obtain an estimate of the proportion of genes that display SPE patterns because of genes not present in one of the parental inbred lines, the number of such presence-absence variations (PAVs) between the available genomic sequences of the inbred lines B73 and Mo17 was determined. Along development, between 5% and 6% of the SPE_B patterns are due to genomic PAVs (Data S1). It was not possible to determine the fraction of PAVs present in Mo17 but absent in B73 among SPE_Mo17 expression patterns because all RNA-seq reads were mapped to the B73 reference genome. For the remaining parent-hybrid triplets, PAVs could not be determined because of the lack of genomic data. We observed similarly low proportions of PAVs among genes displaying SPE in the reciprocal hybrids of B73 and Mo17 in our previous studies [4, 11, 12]. These results suggest that most genes displaying SPE patterns are physically present in the genome but are only active in certain genotypes. This is in line with previous observations in the reciprocal hybrids of B73 and Mo17 [12]. The low frequency of genomic PAVs among the SPE_B genes suggests that genomic PAV is not a major driver of SPE.

F_1 Hybrids Express More Genes along Development than Their Parental Inbred Lines

Within each parent-hybrid triplet, the hybrids express more genes than their parental inbred lines (Figure 2B). Along the

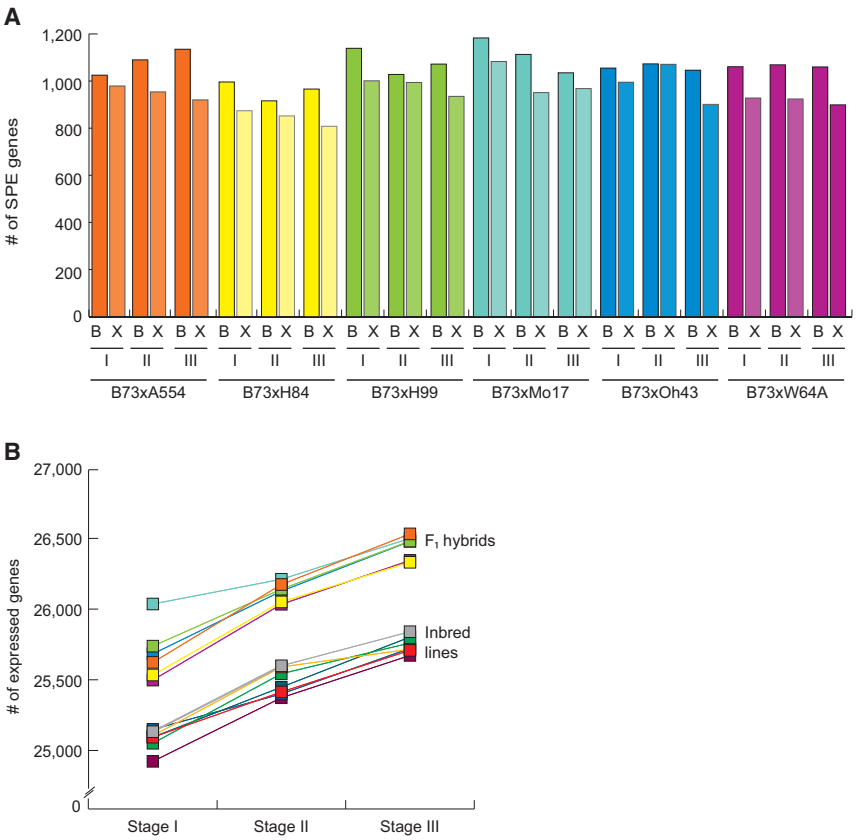


Figure 2. SPE Complementation Is Universal in Maize Hybrids

(A) Number of SPE_B (B) and SPE_X (X) genes in the six parent-hybrid triplets. I–III indicate the three developmental stages of the analyzed primary roots: I, 2–4 cm; II, 6–8 cm; and III, 10–12 cm (see also Figure S1A).

(B) Number of expressed genes in hybrids and their parental inbred lines at each of the three developmental stages (I–III).

SPE, single-parent expression.

three developmental stages, hybrids express on average 584 (stage I), 593 (stage II), and 660 (stage III) genes more than their parental inbred lines. Moreover, the total number of expressed genes per genotype increases during development (Figure 2B; Data S1). Of the 39,469 core protein-coding genes, 72% (28,593/39,469) were expressed among the different genotypes and developmental stages. The increasing number of active genes reflects the increase of morphological complexity in these roots starting from bare primary roots with root hairs to longer, more differentiated primary roots with emerging lateral roots [13].

In previous studies, we discovered SPE patterns in maize primary roots [4] and their tissues [12] in reciprocal hybrids that emerged from crosses of the parental inbred lines B73 and Mo17. By extending our gene activity studies to a diverse panel of genotypes and different developmental stages, we demonstrated that gene expression complementation in F_1 hybrids is a general mechanism to increase the total number of active genes in maize hybrids compared to their parental inbred lines by several hundred (Figures 2B and S4C).

SPE Patterns Are Highly Conserved along Development, but Not Between Different Genotypes

All SPE_B and SPE_X patterns were highly conserved along development (58% to 68%) (Figure S3A) and their overlap by far exceeded the statistically expected values under a null model of independence (Table S3). However, only between 3% and 6% of the SPE genes were conserved among all six parent-hybrid combinations (Figure S3B). In contrast, between 7% and 22% of SPE genes were specific for one parent-hybrid triplet (Figure S3B). In general, SPE_B genes displayed a higher genotype-specificity than their SPE_X counterparts, because SPE patterns of genes only present in X could not be captured.

The high developmental and the low genotypic conservation of SPE patterns is in line with their transcriptomic relationship (Figure 1B). Similar observations have been made in a comparison of different genotypes of maize during seminal root development [14]. In addition, SPE patterns displayed 75% conservation in maize roots subjected to water deficit treatment versus control conditions [11]. The observation of highly conserved SPE patterns during development and upon environmental changes supports the notion of a biological role of these SPE patterns.

Non-syntenic Genes Are Significantly Overrepresented among SPE Genes

Maize encountered a last whole genome duplication ~5 to 12 million years ago [15], shortly after the divergence of the maize and sorghum lineages [5]. Syntenic genes are evolutionarily old and can be traced back to a maize progenitor already before the separation of maize and sorghum. In most instances, these genes encode important functions and were therefore not eliminated by selection in the course of evolution. In contrast, non-syntenic genes evolved after the separation of maize and sorghum. To examine their evolutionary origin, the proportion of non-syntenic genes among the SPE genes was determined. In total, 38% (15,175/39,469) [16] of all protein-coding genes of the B73v3 reference genome are non-syntenic (Figure 3, dashed line). Non-syntenic genes are underrepresented among all expressed genes that are not SPE within each parent-hybrid triplet at all three developmental stages (Figure 3; Data S1). On average, only between 18% and 19% of expressed genes in inbred lines and hybrids, which are not SPE, are non-syntenic. Remarkably, at all three developmental stages non-syntenic genes were exceptionally and significantly overrepresented among all SPE_B and SPE_X patterns compared to their prevalence among all expressed genes (Figure 3; Data

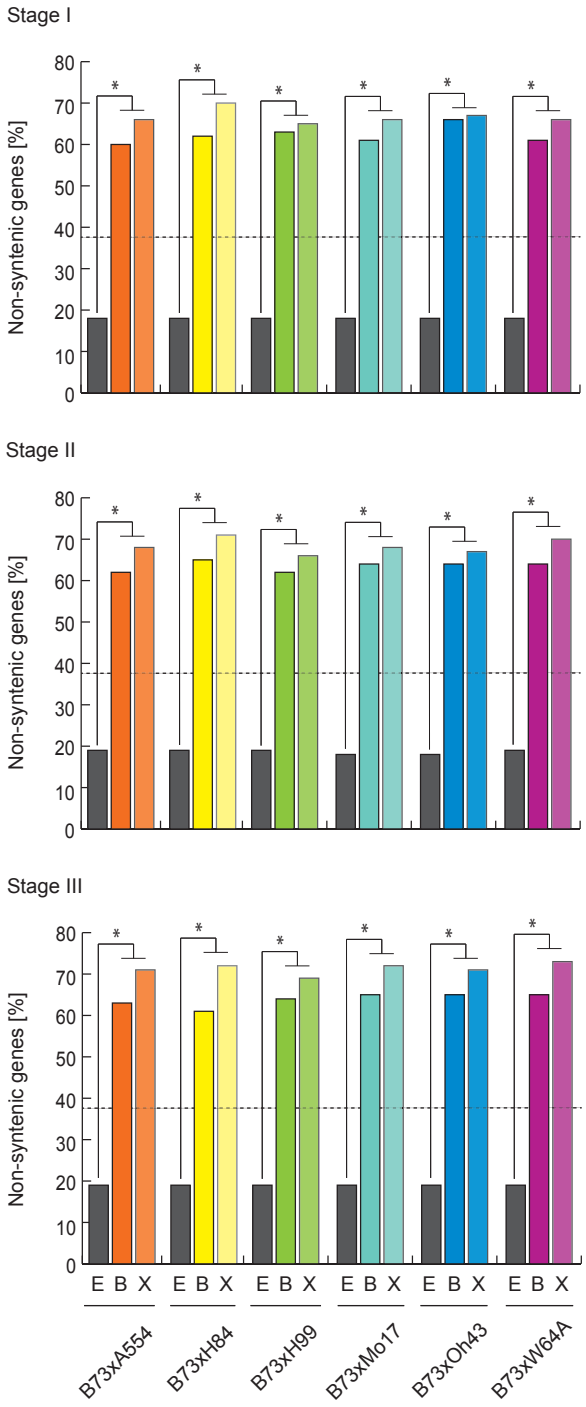


Figure 3. Overrepresentation of Non-syntenic Genes among SPE Genes

Proportion of non-syntenic genes among SPE_B (B, darker color) and SPE_X (X, lighter color) patterns and among the total number of expressed genes that are not SPE (E, gray) per parent-hybrid triplet during the three developmental stages I–III. The proportion of non-syntenic genes among all protein-coding genes is represented by the dashed line. Significant differences ($\alpha = 0.05$) are indicated by an asterisk.

See also Data S1 and Figure S4D.

S1). Along development, between 60% and 66% of the genes displaying SPE_B patterns were non-syntenic. Among the genes showing SPE_X patterns between 65% and 73% were non-syntenic along the three developmental stages (Figure 3). In all 18 genotype-by-stage combinations, SPE_X patterns displayed a higher prevalence of non-syntenic genes than SPE_B patterns.

Non-syntenic genes, which evolved more recently than their syntenic counterparts, are considered to encode genetically redundant functions and thereby are able to buffer the function of particular genes [17]. It was suggested that non-syntenic genes contribute to the capacity of plants to constantly adapt to fluctuating environments. In comparative studies, disease resistance genes were overrepresented among genes of non-syntenic regions in *Arabidopsis thaliana* [18], *Aegilops tauschii* [19] and different accessions of cultivated and wild rice [20]. Notably, the recently cloned wheat *Fusarium*-resistance gene *Fhb1* is non-syntenic [21]. In general, maize hybrids can cope better with environmental fluctuations than their parental inbred lines [22]. A possible function of non-syntenic genes, overrepresented among genes that display expression complementation, could be related to the adaptation of hybrid plants to environmental changes. The suggested role of SPE genes in environmental adaptation is also underscored by their extensive stability under water deficit conditions [11].

Non-syntenic bZIP and bHLH Transcription Factors Are Significantly Overrepresented among SPE Genes

In maize, 2,298 genes encoding transcription factors (TFs) were identified [6]. TFs are classified into 56 families. Among those, bZIP (216 members) and bHLH (308 members) are the two most prevalent TF families with adequate sample size to be examined. In the bZIP and bHLH TF families, non-syntenic genes were systematically overrepresented among SPE patterns (bZIP: 33% to 100%; bHLH: 17% to 40%) (Figure 4) compared to the prevalence of non-syntenic genes among expressed genes of these families that do not display SPE (bZIP: 2% to 7%; bHLH: 4% to 7%) (Figure 4; Table S4).

Transcription factors (TFs) modulate a multitude of regulatory processes by controlling downstream gene expression cascades. Mutations in a single gene encoding a TF such as the lob domain TF *rtcs*, which controls maize shoot-borne root development [23], can affect the regulation of thousands of downstream genes and molecular networks [24]. To maintain the balance of the regulatory genes, any variation in their expression, which could drastically reprogram the transcriptomic landscape of the affected plant, is largely selected against [25]. bZIP and bHLH control a multitude of physiological mechanisms [26–30], such as the adaptation to drought [31]. A total of 91% of bZIP and 88% of bHLH TFs are encoded by evolutionary old syntenic genes (Data S1). The remaining non-syntenic bZIP and bHLH TFs were highly overrepresented in SPE patterns compared to all expressed TFs of these families (Figure 4). These likely genetically redundant TFs could function in the stabilization of molecular networks by expression complementation.

SPE Complementation Is a General Mechanism in Maize

To further generalize the concept of SPE complementation in maize hybrids independently of the inbred line B73, we

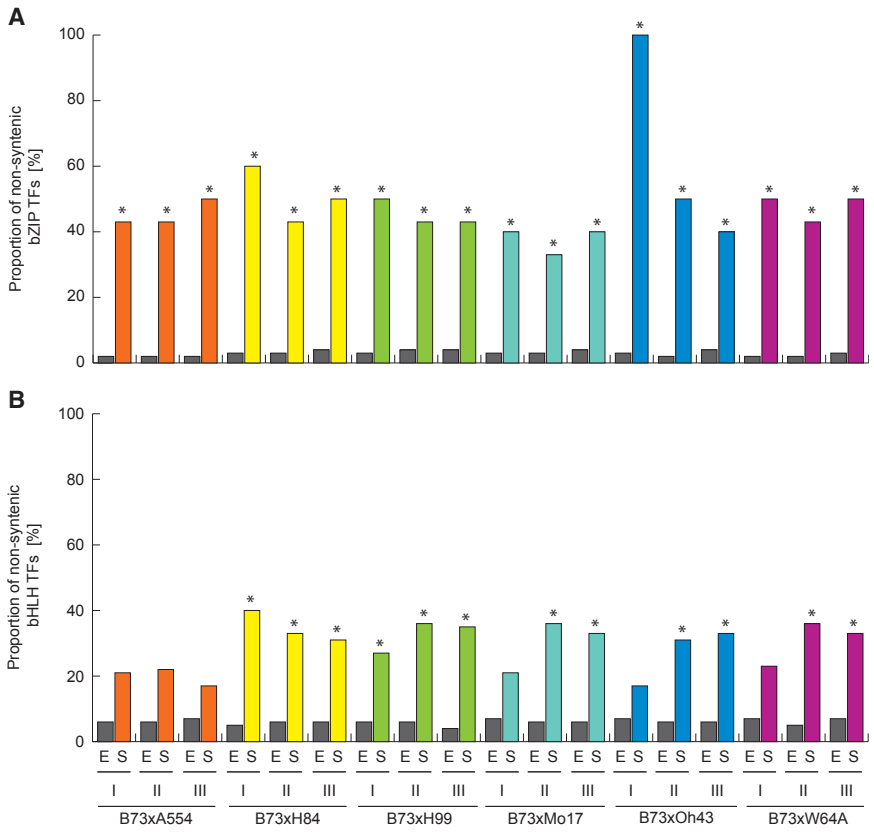


Figure 4. Overrepresentation of Non-syntenic bZIP and bHLH Transcription Factors TFs among All TFs Displaying SPE Patterns

(A and B) Proportion of non-syntenic genes among all expressed bZIP (A) and bHLH (B) TFs that do not display SPE (E, gray bars; sample size n = 106–113 bZIP, 135–145 bHLH) and of bZIP (A) and bHLH (B) TFs displaying SPE (S, colored bars; sample size n = 3–8 bZIP, 6–17 bHLH) for each of the six parent-hybrid triplets and developmental stages (I–III). The proportion of non-syntenic TFs displaying SPE refers to the sum of SPE_B and SPE_X genes per parent-hybrid triplet. Significant differences ($\alpha = 0.05$) are indicated by an asterisk.

See also Figure S4, Table S4, and Data S1.

surveyed the transcriptomes of the reciprocal F_1 hybrids Oh43xW64A and W64AxOh43 and their distantly related parental inbred lines (Figures 1A and 1B). All analyses with these B73-independent genotypes displayed the same results as the B73-dependent experiments including kinship relations of transcriptomic samples (Figure S4A), SPE patterns (Figure S4B), increased number of expressed genes in hybrids (Figure S4C), overrepresentation of non-syntenic genes among SPE genes (Figure S4D), and overrepresentation of non-syntenic bZIP TFs among all bZIP TFs displaying SPE patterns (Figure S4E; Table S4). On the other hand, non-syntenic bHLH TFs were not enriched (Figure S4F; Table S4).

The observation of substantial gene expression complementation in hybrids resulting in more active genes than in their parents is in accordance with the complementation hypothesis of heterosis, which was suggested a century ago by Donald F. Jones [3] as one possible mechanism to explain the increased vigor of hybrids. A role of gene expression complementation in heterosis is also consistent with the reduced phenotypic performance of hybrids observed after subsequent cycles of self-pollination, i.e., inbreeding depression. In each round of selfing the number of heterozygous loci is reduced by 50% and as a consequence the number of SPE patterns and thus the number of additionally active genes in these plants also consistently decrease. To obtain a maximum level of heterosis in F_1 hybrids, the maize germplasm is divided into heterotic groups representing genetically distinct inbred lines [32]. Inter-group hybrids show a greater heterosis effect than intra-group crosses [33–36]. This is in line with the higher number of SPE patterns in inter- versus intra-group crosses observed in the present study. Although several other mechanisms must be considered in heterosis manifestation, non-syntenic genes displaying SPE patterns might collectively contribute at least to a certain degree to the better performance and plasticity of hybrid maize. Nevertheless, the function of hundreds of highly genotype-specific genes with a minor individual influence on plant performance is very difficult to access genetically. However, the additional activity of hundreds of such genes in hybrids could make a difference. The identified expression profiles might be beneficial to support genomic prediction of hybrid performance in breeding cycles, which has been successfully demonstrated in combination with DNA marker information [37, 38].

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - RNA isolation and RNA-sequencing (RNA-seq)
  - Analysis of raw sequencing data: read mapping and raw read counting
  - Multidimensional scaling analysis for data quality check
  - Genomic presence-absence variations between B73 and Mo17
  - Identification of non-syntenic maize genes and transcription factors among SPE genes
- QUANTIFICATION AND STATISTICAL ANALYSES
  - Analysis of expression complementation
  - Fisher's exact test
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures, four tables, and two data files and can be found with this article online at https://doi.org/10.1016/j.cub.2017.12.027.

ACKNOWLEDGMENTS

We thank Helmut Rehkopf and Christa Schulz (University of Bonn) for their support in propagating the genetic material for this study. We would like to thank the staff of the experimental field station Campus Klein-Altendorf (University Bonn) for crop cultivation. Statistical method development for gene activity analysis was funded by the National Institute of General Medical Sciences (NIGMS) of the NIH and the joint National Science Foundation (NSF)/NIGMS Mathematical Biology Program under award number R01GM109458 to D.N. This work was funded by the Deutsche Forschungsgemeinschaft (DFG) grant HO2249/9-3 to F.H.

AUTHOR CONTRIBUTIONS

J.A.B. carried out the experiments, conducted the statistical analyses, interpreted the data, and drafted the manuscript. C.M. was involved in performing the wet lab experiments. A.L. and D.N. carried out the gene activity analyses. L.A., L.V., and H.S. were involved in the bioinformatics handling of the data. H.-P.P. provided the experimental design for the RNA-seq experiment and helped with the statistical analyses. F.H. conceived and coordinated the study and participated in data interpretation and drafting the manuscript. All authors edited the manuscript and approved the final draft.

DECLARATION OF INTERESTS

All authors declare no competing financial interests.

Received: November 8, 2017
Revised: December 8, 2017
Accepted: December 13, 2017
Published: January 18, 2018

REFERENCES

1. Springer, N.M., Ying, K., Fu, Y., Ji, T., Yeh, C.T., Jia, Y., Wu, W., Richmond, T., Kitzman, J., Rosenbaum, H., et al. (2009). Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet. *5*, e1000734.
2. Swanson-Wagner, R.A., Eichten, S.R., Kumari, S., Tiffin, P., Stein, J.C., Ware, D., and Springer, N.M. (2010). Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res. *20*, 1689–1699.
3. Jones, D.F. (1917). Dominance of linked factors as a means of accounting for heterosis. Genetics *2*, 466–479.
4. Paschold, A., Jia, Y., Marcon, C., Lund, S., Larson, N.B., Yeh, C.-T., Ossowski, S., Lanz, C., Nettleton, D., Schnable, P.S., and Hochholdinger, F. (2012). Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents. Genome Res. *22*, 2445–2454.
5. Schnable, J.C., Springer, N.M., and Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc. Natl. Acad. Sci. USA *108*, 4069–4074.
6. Jin, J., Tian, F., Yang, D.-C., Meng, Y.-Q., Kong, L., Luo, J., and Gao, G. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res. *45* (D1), D1040–D1045.
7. Marroni, F., Pinosio, S., and Morgante, M. (2014). Structural variation and genome complexity: is dispensable really dispensable? Curr. Opin. Plant Biol. *18*, 31–36.
8. Flint-Garcia, S.A., Thuillet, A.-C., Yu, J., Pressoir, G., Romero, S.M., Mitchell, S.E., Doebley, J., Kresovich, S., Goodman, M.M., and Buckler, E.S. (2005). Maize association population: a high-resolution platform for quantitative trait locus dissection. Plant J. *44*, 1054–1064.
9. Lorenz, A., and Hoegemeyer, T. (2013). The phylogenetic relationships of US maize germplasm. Nat. Genet. *45*, 844–845.
10. Troyer, F.A. (2001). Temperate corn - background, behaviour, and breeding. In Specialty Corns, A.R. Hallauer, ed. (CRC Press), pp. 393–466.
11. Marcon, C., Paschold, A., Malik, W.A., Lithio, A., Baldauf, J.A., Altrogge, L., Opitz, N., Lanz, C., Schoof, H., Nettleton, D., et al. (2017). Stability of single-parent gene expression complementation in maize hybrids upon water deficit stress. Plant Physiol. *173*, 1247–1257.
12. Paschold, A., Larson, N.B., Marcon, C., Schnable, J.C., Yeh, C.-T., Lanz, C., Nettleton, D., Piepho, H.-P., Schnable, P.S., and Hochholdinger, F. (2014). Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. Plant Cell *26*, 3939–3948.
13. Hochholdinger, F., Woll, K., Sauer, M., and Dembinsky, D. (2004). Genetic dissection of root formation in maize (*Zea mays*) reveals root-type specific developmental programmes. Ann. Bot. *93*, 359–368.
14. Tai, H., Opitz, N., Lithio, A., Lu, X., Nettleton, D., and Hochholdinger, F. (2017). Non-syntenic genes drive RTCS-dependent regulation of the embryo transcriptome during formation of seminal root primordia in maize (*Zea mays* L.). J. Exp. Bot. *68*, 403–414.
15. Swigonová, Z., Lai, J., Ma, J., Ramakrishna, W., Llaca, V., Bennetzen, J.L., and Messing, J. (2004). Close split of sorghum and maize genome progenitors. Genome Res. *14* (10A), 1916–1923.
16. Schnable, J.C., Zang, Y., and Ngu, D.W.C. (2016). Pan-grass syntenic gene set (sorghum referenced). https://figshare.com/articles/Pan_Grass_Syntenic_Gene_Set_sorghum_referenced_/3113488.
17. Schnable, J.C. (2015). Genome evolution in maize: from genomes back to genes. Annu. Rev. Plant Biol. *66*, 329–343.
18. Freeling, M., Lyons, E., Pedersen, B., Alam, M., Ming, R., and Lisch, D. (2008). Many or most genes in *Arabidopsis* transposed after the origin of the order Brassicales. Genome Res. *18*, 1924–1937.
19. Dong, L., Huo, N., Wang, Y., Deal, K., Wang, D., Hu, T., Dvorak, J., Anderson, O.D., Luo, M.-C., and Gu, Y.Q. (2016). Rapid evolutionary dynamics in a 2.8-Mb chromosomal region containing multiple prolamin and resistance gene families in *Aegilops tauschii*. Plant J. *87*, 495–506.
20. Xu, X., Liu, X., Ge, S., Jensen, J.D., Hu, F., Li, X., Dong, Y., Gutenkunst, R.N., Fang, L., Huang, L., et al. (2011). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat. Biotechnol. *30*, 105–111.
21. Rawat, N., Pumphrey, M.O., Liu, S., Zhang, X., Tiwari, V.K., Ando, K., Trick, H.N., Bockus, W.W., Akhunov, E., Anderson, J.A., and Gill, B.S. (2016). Wheat *Fhb1* encodes a chimeric lectin with agglutinin domains and a pore-forming toxin-like domain conferring resistance to Fusarium head blight. Nat. Genet. *48*, 1576–1580.
22. Betrán, F.J., Beck, D., Bänziger, M., and Edmeades, G.O. (2003). Secondary traits in parental inbreds and hybrids under stress and non-stress environments in tropical maize. Field Crops Res. *83*, 51–65.
23. Taramino, G., Sauer, M., Stauffer, J.L., Jr., Multani, D., Niu, X., Sakai, H., and Hochholdinger, F. (2007). The maize (*Zea mays* L.) RTCS gene encodes a LOB domain protein that is a key regulator of embryonic seminal and post-embryonic shoot-borne root initiation. Plant J. *50*, 649–659.
24. Muthreich, N., Majer, C., Beatty, M., Paschold, A., Schützenmeister, A., Fu, Y., Malik, W.A., Schnable, P.S., Piepho, H.-P., Sakai, H., and Hochholdinger, F. (2013). Comparative transcriptome profiling of maize coleoptilar nodes during shoot-borne root initiation. Plant Physiol. *163*, 419–430.
25. Birchler, J.A., and Veitia, R.A. (2010). The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. New Phytol. *186*, 54–62.

26. Kumar, I., Swaminathan, K., Hudson, K., and Hudson, M.E. (2016). Evolutionary divergence of phytochrome protein function in *Zea mays* PIF3 signaling. J. Exp. Bot. *67*, 4231–4240.
27. Muszynski, M.G., Dam, T., Li, B., Shirbroun, D.M., Hou, Z., Bruggemann, E., Archibald, R., Ananiev, E.V., and Danilevskaya, O.N. (2006). *delayed flowering1* encodes a basic leucine zipper protein that mediates floral inductive signals at the shoot apex in maize. Plant Physiol. *142*, 1523–1536.
28. Pautler, M., Eveland, A.L., LaRue, T., Yang, F., Weeks, R., Lunde, C., Je, B.I., Meeley, R., Komatsu, M., Vollbrecht, E., et al. (2015). *FASCIATED EAR4* encodes a bZIP transcription factor that regulates shoot meristem size in maize. Plant Cell *27*, 104–120.
29. Nan, G.-L., Zhai, J., Arikit, S., Morrow, D., Fernandes, J., Mai, L., Nguyen, N., Meyers, B.C., and Walbot, V. (2017). MS23, a master basic helix-loop-helix factor, regulates the specification and development of the tapetum in maize. Development *144*, 163–172.
30. Ludwig, S.R., Habera, L.F., Dellaporta, S.L., and Wessler, S.R. (1989). *Lc*, a member of the maize R gene family responsible for tissue-specific anthocyanin production, encodes a protein similar to transcriptional activators and contains the myc-homology region. Proc. Natl. Acad. Sci. USA *86*, 7092–7096.
31. Rabara, R.C., Tripathi, P., and Rushton, P.J. (2014). The potential of transcription factor-based genetic engineering in improving crop tolerance to drought. OMICS *18*, 601–614.
32. Melchinger, A.E., and Gumber, R.K. (1998). Overview of heterosis and heterotic groups in agronomic crops. In Concepts and Breeding of Heterosis in Crop Plants, K.R. Larnkey, and J.E. Staub, eds. (CSSA), pp. 29–44.
33. Adeyemo, O., Menkir, A., Melaku, G., and Omidiji, O. (2011). Genetic diversity assessment and relationship among tropical-yellow endosperm maize inbred lines using SSR markers. Maydica *56*, 17–23.
34. Barbosa, A.M.M., Garaldi, I.O., Benchimol, L.L., Garcia, A.A.F., Souza, C.L., and Souza, A.P. (2003). Relationship of intra- and interpopulations tropical maize single cross hybrid performance and genetic distances computed from AFLP and SSR markers. Euphytica *130*, 87–99.
35. Bruel, D.C., Carpentieri-Pípolo, V., Gerage, A.C., Fonseca, N.D.S., Prete, C.E.C., Ruas, C.D.F., Ruas, P.M., De Souza, S.G.H., and Garbuglio, D.D. (2006). Genetic distance estimated by RAPD markers and its relationship with hybrid performance in maize. Pesq. Agropec. Bras. *41*, 1491–1498.
36. Melchinger, A.E., Messmer, M.M., Lee, M., Woodman, W.L., and Lamkey, K.R. (1991). Diversity and relationships among U.S. maize inbreds revealed by restriction fragment length polymorhisms. Crop Sci. *31*, 669–678.
37. Westhues, M., Schrag, T.A., Heuer, C., Thaller, G., Utz, H.F., Schipprack, W., Thiemann, A., Seifert, F., Ehret, A., Schlereth, A., et al. (2017). Omics-based hybrid prediction in maize. Theor. Appl. Genet. *130*, 1927–1939.
38. Zenke-Philippi, C., Frisch, M., Thiemann, A., Seifert, F., Schrag, T., Melchinger, A.E., Scholten, S., and Herzog, E. (2017). Transcriptome-based prediction of hybrid performance with unbalanced data from a maize breeding programme. Plant Breeding *136*, 331–337.
39. Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol. Biol. *7*, 3.
40. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.
41. Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. Science *326*, 1112–1115.
42. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.
43. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.
44. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.
45. Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. *15*, R29.
46. Smyth, G.K. (2005). Limma: linear models for microarray data. In Bioinformatics and Computational Biology Solutions using R and Bioconductor, R. Gentleman, V. Carey, W. Huber, R. Irizarry, and S. Dudoit, eds. (Springer), pp. 397–420.
47. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinformatics *10*, 421.
48. Schnable, J.C., Freeling, M., and Lyons, E. (2012). Genome-wide analysis of syntenic gene deletion in the grasses. Genome Biol. Evol. *4*, 265–277.
49. Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.
50. Van De Wiel, M.A., Leday, G.G.R., Pardo, L., Rue, H., Van Der Vaart, A.W., and Van Wieringen, W.N. (2013). Bayesian analysis of RNA sequencing data by estimating multiple shrinkage priors. Biostatistics *14*, 113–128.
51. Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J. R. Stat. Soc. Series B Stat. Methodol. *71*, 319–392.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Zea mays L.; B73	Hochholdinger lab	Lab Ac 14-1137-12; 14-1145-5; 14-1150-7
Zea mays L.; Mo17	Hochholdinger lab	Lab Ac 14-1152-2; 14-1152-3; 14-1152-8
Zea mays L.; A554	Hochholdinger lab	Lab Ac 14-1156-5;14-1104-3;14-1106-5
Zea mays L.; H84	Hochholdinger lab	Lab Ac 14-1154-1; 14-1154-2; 14-1153-1
Zea mays L.; H99	Hochholdinger lab	Lab Ac 14-1157-8; 14-1157-9; 14-1158-9
Zea mays L.; Oh43	Hochholdinger lab	Lab Ac 14-1159-5; 14-1160-1; 14-1160-4 16-1008-1
Zea mays L.; W64A	Hochholdinger lab	Lab Ac 14-1120-1; 14-1161-1; 14-1161-8 16-1006-1
Zea mays L.; B73xMo17	Hochholdinger lab	Lab Ac 14-1101/1107-1; 14-1105/1107-3; 14-1109/1107-5
Zea mays L.; B73xA554	Hochholdinger lab	Lab Ac 14-1101/1104-4; 14-1101/1106-10; 14-1101/1106-11
Zea mays L.; B73xH84	Hochholdinger lab	Lab Ac 14-1105/1115-6; 14-1119/1115-3; 14-1119/1115-7
Zea mays L.; B73xH99	Hochholdinger lab	Lab Ac 14-1101/1114-3; 14-1101/1112-8; 14-1109/1114-1
Zea mays L.; B73xOh43	Hochholdinger lab	Lab Ac 14-1105/1160-2; 14-1105/1160-4; 14-1105/1160-10
Zea mays L.; B73xW64A	Hochholdinger lab	Lab Ac 14-1119/1122-1; 14-1119/1122-5; 14-1121/1122-1
Zea mays L.; Oh43xW64A	Hochholdinger lab	Lab Ac 16-1055-10/1054-1
Zea mays L.; W64AxOh43	Hochholdinger lab	Lab Ac 16-1054-7/1055-2
Chemicals, Peptides, and Recombinant Proteins		
RNeasy Plant Mini Kit	QIAGEN, Venlo, the Netherlands	Cat No./ID: 74104
Critical Commercial Assays		
Agilent RNA 6000 Nano Chip	Agilent Technologies, Santa Clara, CA, USA	N/A
Deposited Data		
B73 reference genome sequence version AGPv3.31	Ensembl Genomes [39]	ftp://ftp.ensemblgenomes.org/pub/plants/release-31/fasta/zea_mays/dna/
Maize protein-coding genes	Ensembl Genomes	ftp://ftp.ensemblgenomes.org/pub/plants/release-31/gtf/zea_mays/
Mo17 genome assembly version 1 (ZM-Mo17-REFERENCE-NRGENE-1.0)	Ensembl Genomes	http://ftp.maizegdb.org/MaizeGDB/FTP/Mo17/2013_XinEtAl_PlantCell/RefGen_Mo17_v1.fasta.gz
Pan-Grass Syntenic Gene Set (sorghum referenced)	[40]	http://www.schnablelab.org/data.html
Plant Transcription Factor Database version 4.0	[6]	http://planttfdb.cbi.pku.edu.cn/
Software and Algorithms		
Trimmomatic version 0.36	[41]	http://www.usadellab.org/cms/?page=trimmomatic
Bowtie2 version 2.2.9	[42]	https://sourceforge.net/projects/bowtie-bio/files/bowtie2/
TopHat2 version 2.1.1	[43]	https://ccb.jhu.edu/software/tophat/tutorial.shtml
MarkDuplicates version 2.5.0		http://broadinstitute.github.io/picard/command-line-overview.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HTSeq-count version 0.6.1p1	[44]	https://htseq.readthedocs.io/en/release_0.9.1/
NCBI blastn tool (BLAST Command Line Application, National Center for Biotechnology Information (US))	[45]	https://www.ncbi.nlm.nih.gov/books/NBK279670/
SAS 9.4		(SAS Institute, Cary, NC, USA, 2014)
R version 3.3.1 (2016-06-21)		

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Frank Hochholdinger (hochholdinger@uni-bonn.de).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

In a first experiment, kernels of the maize (*Zea mays* L.) parental inbred lines A554, B73, H84, H99, Mo17, Oh43 and W64A and the hybrids B73xA554, B73xH84, B73xH99, B73xMo17, B73xOh43 and B73xW64A were grown in filter paper rolls in a climate chamber with a 16 h light (28°C), 8 h dark (21°C) cycle in distilled water. For each genotype, 36 paper rolls each containing 10 seeds were prepared. For germination, all rolls were randomly arranged in 39 10-L buckets. Primary roots were harvested at three developmental stages (stage I, II and III). Each developmental stage was defined by the length of the respective roots. Stage I was represented by 2–4 cm long primary roots, stage II by 6–8 cm long roots and stage III by 10–12 cm long roots. Ten primary roots were pooled for each of the four biological replicates per genotype and developmental stage. For B73, which was the common parent of all hybrid crosses, we generated twelve biological replicates per developmental stage to maximize the number of direct comparisons between B73 and other genotypes in the RNA-seq experiment described below.

In a second experiment, kernels of the parental inbred lines Oh43 and W64A and their reciprocal F_1-hybrids, Oh43xW64A and W64AxOh43, were germinated in paper rolls under the same conditions as described above. For each of the four biological replicates per genotype, 10 primary roots were harvested at developmental stage I, i.e., at a length of 2–4 cm.

METHOD DETAILS

RNA isolation and RNA-sequencing (RNA-seq)

Pooled primary roots of both experimental setups were ground in liquid nitrogen and total RNA was isolated with the RNeasy Plant Mini Kit (QIAGEN, Venlo, the Netherlands). RNA quality was assessed via agarose gel electrophoresis and a Bioanalyzer (Agilent RNA 6000 Nano Chip; Agilent Technologies, Santa Clara, CA, USA). For all samples, a RIN (RNA integrity number [39]) $\geq$ 8.9 was detected. cDNA libraries for Illumina sequencing were constructed according to the manufacturer's protocol (TruSeq RNA sample preparation; Illumina, San Diego CA, USA). For sequencing, each library was indexed with an Illumina TruSeq Adaptor.

The 180 indexed libraries of the first experiment were loaded in nine lanes of two flow cells with 20 pooled libraries per lane. The experimental design was set up to obtain small standard errors for all pairwise comparisons involving genotypes from the same cross (parents and hybrids). Genotypes in the same lane can be compared directly, and such comparisons are most accurate. Each triplet consists of one hybrid, one paternal inbred line and the common parent B73. The combination of a hybrid with a paternal inbred line will be denoted as a "pair." In total we analyzed six such pairs (A554 and B73xA554, H84 and B73xH84, H99 and B73xH99, Mo17 and B73xMo17, Oh43 and B73xOh43, W64A and B73xW64A). To each lane, we assigned four pairs (i.e., the hybrid and their paternal inbred line) of two developmental stages (= 16 libraries). Moreover, two libraries of the parental inbred line B73 of both developmental stages were added per lane. Cluster preparation and paired-end sequencing were performed according to the manufacturer's protocol (HiSeq 4000, Illumina).

Analysis of raw sequencing data: read mapping and raw read counting

Downstream analyses of 100 bp long paired-end reads generated with an Illumina HiSeq 4000 were performed as follows: raw sequencing reads were quality trimmed by removing low quality reads and sequencing adapters using Trimmomatic [40] version 0.36 (http://www.usadellab.org/cms/?page=trimmomatic). In paired-end mode the following settings were used: ILLUMINACLIP:3:30:10, LEADING:3, TRAILING:3, MAXINFO:30:0.8 and MINLEN:40. The B73 reference genome sequence [41] version AGPv3.31 (B73v3, ftp://ftp.ensemblgenomes.org/pub/plants/release-31/fasta/zea_mays/dna/) was indexed using Bowtie2 [42] version 2.2.9 (https://sourceforge.net/projects/bowtie-bio/files/bowtie2/) and raw sequence reads were mapped to the reference genome by TopHat2 [43] version 2.1.1 (https://ccb.jhu.edu/software/tophat/tutorial.shtml). A slight bias was introduced by mapping all genotypes to the B73 genome, the only available high-quality reference genome. The minimum intron size was set to 5 and the maximum intron size was set to 60,000. For GTF guided mapping, a GTF file was employed containing information for core protein-coding

genes (ftp://ftp.ensemblgenomes.org/pub/plants/release-31/gtf/zea_mays/). All other parameters were set to default. In the following step, duplicates, i.e., read pairs that have identical 5′ coordinates and orientations, were marked and removed from the mapped reads except the read pair having the highest sum of base qualities. This step was performed using the picard tool MarkDuplicates version 2.5.0 (http://broadinstitute.github.io/picard/command-line-overview.html). The RNA-seq of the 180 samples resulted on average in 17 million 100 bp long paired-end reads (Table S1). After quality trimming, between 82% and 96% of all sequencing reads mapped to the B73v3 reference genome sequence (Table S1) [41]. In the second RNA-seq experiment, the 16 indexed libraries were pooled in a single lane of a flow cell. On average, 20 million 100 bp long paired-end reads were obtained from RNA-seq (Table S2). After quality trimming, between 87% and 89% of all sequencing reads mapped to the B73v3 reference genome sequence (Table S2) [41].

To obtain the raw read counts for each protein-coding gene, the aligned reads were sorted according to sequence names and separated into single- and paired-end reads. Read counts per protein-coding gene were generated using HTSeq-count [44] version 0.6.1p1 (https://htseq.readthedocs.io/en/release_0.9.1/) using default settings. During this step, reads mapping at multiple positions of the genome and reads which mapped to overlapping gene models were discarded. For further downstream analyses, the separated single- and paired-end reads were merged into a single data file.

Multidimensional scaling analysis for data quality check

The following analyses were applied to genes which were declared expressed or active based on the analysis of expression complementation. The raw sequencing reads of those genes were normalized by sequencing depth and $\log_2$-transformed to meet the assumptions of linear modeling. In addition, the mean-variance trend within the count data was estimated and a precision weight was assigned to each observation [45]. A multidimensional scaling (MDS) analysis was performed to visualize any sample relations using the plotMDS function of the Bioconductor package *limma* [46] in R (R version 3.3.1 (2016-06-21), limma_3.30.0). The distance between each pair of samples was estimated as the root-mean-square deviation for the top genes with the largest standard deviations across all samples.

Genomic presence-absence variations between B73 and Mo17

Genomic presence-absence variations (PAVs) are genes present in a certain genotype but absent in another. To determine the proportion of genomic PAVs among SPE_B expression patterns identified in the parent-hybrid triplet of B73 and Mo17, these genes were blasted against the Mo17 genome assembly version 1 (ZM-Mo17-REFERENCE-NRGENE-1.0; http://ftp.maizegdb.org/MaizeGDB/FTP/Mo17/2013_XinEtAl_PlantCell/RefGen_Mo17_v1.fasta.gz) using the command line blastn tool [47] of NCBI (BLAST® Command Line Application, National Center for Biotechnology Information (US); https://www.ncbi.nlm.nih.gov/books/NBK279670/). For the five remaining parental inbred lines, genomic PAVs could not be determined because no reference genome data is available for these lines.

SPE_B genes were defined as putative PAV if no BLAST hit with sequence identity > 90, a bit score > 50 and a query coverage > 90 was observed. These putative PAVs were compared to their expression levels of the Mo17 RNA-seq samples. Based on these criteria, genes with no reads in Mo17 (FPKM = 0) were determined as putative PAVs.

Identification of non-syntenic maize genes and transcription factors among SPE genes

First, the syntenic or non-syntenic evolutionary origin of SPE genes was determined by comparisons with the lists of these genes [48] (http://www.schnablelab.org/data.html). Second, for each parent-hybrid triplet, the distribution of non-syntenic transcription factors (TFs; according to the Plant Transcription Factor Database version 4.0 [6] (http://planttfdb.cbi.pku.edu.cn/) was determined among the genes showing SPE patterns in comparison to the total number of expressed TFs that are not SPE of a corresponding TF class.

QUANTIFICATION AND STATISTICAL ANALYSES

Analysis of expression complementation

For the data of the first experiment, the transcriptional activity status of all genes (active/inactive) in each of the thirteen genotypes at each of the three developmental stages was determined using hierarchical gene-specific generalized linear mixed models with negative binomial responses. In brief, within each developmental stage, the log of the mean for any given gene was assumed to be a linear combination of gene-specific effects for genotype, lanes and barcodes, plus sample and gene-specific normalization factors (described below). The log of the TMM normalization factor [49] was added to the linear predictor for normalization across samples, and a smooth function of gene length and GC content was used to normalize across genes.

For each genotype, the gene-specific genotype effects were assumed to be independent draws from a scaled *t*-distribution with three degrees of freedom, an unknown and unrestricted genotype-specific mean, and an unknown genotype-specific scaling parameter. Furthermore, the gene-specific genotype effects were modeled as independent across genotypes. The lane and barcode effects were modeled as mean-zero normal random variables with gamma-distributed gene-specific precisions. The parameters of the gamma distribution corresponding to barcode-effect precisions were specified to create a vague distribution, but the gamma distribution corresponding to lane-effect precisions was treated as having unknown shape and rate to be estimated from the data rather than specified. The logs of the gene-specific negative binomial dispersion parameters were assumed to be independent draws from a normal distribution with unknown mean and variance. An empirical Bayes procedure via the R package *ShrinkBayes* [50] was

used to estimate the unknown parameters. However, the unknown parameters were estimated initially by assuming normal distributions for gene-specific genotype effects. The aforementioned *t*-distributions for gene-specific genotype effects were subsequently specified using the initial genotype-specific mean estimates, and genotype-specific scale parameter estimates set to match the 10^{th} and 90^{th} percentiles of the initially estimated genotype-specific normal distributions. This allows for the use of heavier tailed *t*-distributions that better represent the distributions of gene-specific genotype effects, while still leveraging the empirical Bayes routine and the computational advantages of integrated nested Laplace approximation [51].

The activity status of each gene was determined by computing P_{ijk} (T), the posterior probability that the genotype effect for gene (i), genotype (j), and developmental stage (k) was larger than a given threshold T. A gene (i) was called active for genotype (j) and developmental stage (k) if P_{ijk} (T) > 0.5 and otherwise inactive. This method classifies genes as active or inactive based on the posterior distribution of fixed effects considering raw read count, sequencing differences from sample to sample, gene length, and GC content differences. The same approach was applied to the dataset of the second experiment. Because this dataset consists of only four genotypes at one developmental stage with all RNA samples sequenced together in a single lane, the model and estimation procedure described above was simplified accordingly for this second analysis.

To determine the expected number of SPE genes, which are developmentally conserved within each parent-hybrid triplet, a generalized linear model with a log-link and Poisson distribution was fitted under the assumption of independence between the three stages using *proc genmod* in SAS 9.4 (SAS Institute, Cary, NC, USA, 2014). The model under independence comprised main effects for the three stages. The expected and observed numbers of SPE genes overlapping between the three stages were compared using a χ^2-test for the three-way interaction effect of the three stages. Significant differences between expected and observed values are indicated by * at a significance level of $\alpha = 0.05$.

Fisher's exact test

Fisher's exact test (function *fisher.test* in R version 3.3.1 (2016-06-21)) was applied to evaluate the distribution of non-syntenic genes and non-syntenic bZIP and bHLH TFs among the SPE genes in comparison to all expressed genes that are not SPE and to all expressed TFs that are not SPE, respectively, within each parent-hybrid triplet. The tests were conducted at significance level 0.05.

DATA AND SOFTWARE AVAILABILITY

The accession numbers for the raw sequencing data reported in this paper are Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra): SRP110782 and SRP110862.

frontiers in Plant Science

MINI REVIEW
published: 05 March 2018
doi: 10.3389/fpls.2018.00143

Proteomics of Maize Root Development

Frank Hochholdinger, Caroline Marcon, Jutta A. Baldauf, Peng Yu and Felix P. Frey*

Crop Functional Genomics, Institute of Crop Science and Resource Conservation, University of Bonn, Bonn, Germany

Maize forms a complex root system with structurally and functionally diverse root types that are formed at different developmental stages to extract water and mineral nutrients from soil. In recent years proteomics has been intensively applied to identify proteins involved in shaping the three-dimensional architecture and regulating the function of the maize root system. With the help of developmental mutants, proteomic changes during the initiation and emergence of shoot-borne, lateral and seminal roots have been examined. Furthermore, root hairs were surveyed to understand the proteomic changes during the elongation of these single cell type structures. In addition, primary roots have been used to study developmental changes of the proteome but also to investigate the proteomes of distinct tissues such as the meristematic zone, the elongation zone as well as stele and cortex of the differentiation zone. Moreover, subcellular fractions of the primary root including cell walls, plasma membranes and secreted mucilage have been analyzed. Finally, the superior vigor of hybrid seedling roots compared to their parental inbred lines was studied on the proteome level. In summary, these studies provide novel insights into the complex proteomic interactions of the elaborate maize root system during development.

Keywords: heterosis, lateral root, maize, proteomics, primary root, root hairs, seminal root, shoot-borne root

OPEN ACCESS

Edited by:
Ive De Smet,
Flanders Institute for Biotechnology, Belgium

Reviewed by:
Raffaele Dello Ioio,
Sapienza Università di Roma, Italy
Jian Xu,
National University of Singapore, Singapore

****Correspondence:***
Frank Hochholdinger
hochholdinger@uni-bonn.de

Specialty section:
This article was submitted to Plant Proteomics, a section of the journal Frontiers in Plant Science

Received: *18 December 2017*
Accepted: *25 January 2018*
Published: *05 March 2018*

Citation:
Hochholdinger F, Marcon C, Baldauf JA, Yu P and Frey FP (2018) Proteomics of Maize Root Development. Front. Plant Sci. 9:143. doi: 10.3389/fpls.2018.00143

INTRODUCTION

Maize forms a complex root system to capture limited and unevenly distributed water and nutrient resources from soil and allocating them to the energy-delivering, aboveground parts of the plant. The basic blueprint of the elaborate three-dimensional structure of the maize root system is encoded in an intrinsic genetic program, while the responsiveness to environmental signals secures the adaptive plasticity of the maize root stock (Hochholdinger et al., 2004a,b, 2017). The maize root system is composed of primary and seminal roots formed during embryogenesis and of shoot-borne and lateral roots initiated post-embryonically after germination (**Figure 1**) (Hochholdinger et al., 2004a; Hochholdinger, 2009; Atkinson et al., 2014). The primary root and a variable number of seminal roots are already preformed during embryogenesis. After germination, these roots are important for early seedling vigor (Sanguineti et al., 1998). Whorls of shoot-borne roots are initiated at successive nodes of the shoot and dominate the rootstock of mature plants (Hochholdinger et al., 2004b). Finally, lateral roots are formed in all root types by the meristematic activity of phloem pole pericycle cells. Together with root hairs, which are unicellular extensions of epidermal cells, lateral roots substantially increase the absorbing surface of the maize root system (Yu et al., 2016).

Large-scale quantitative approaches can provide a comprehensive snapshot of biologically relevant molecules such as transcripts, metabolites or proteins in a biological sample. While the transcriptomic dissection of maize root development has been recently reviewed elsewhere (e.g., Yu et al., 2016; Hochholdinger et al., 2017), here we will give an update on the status of the proteomic dissection of maize root development. The proteome is defined as the protein complement of a given biological sample, while proteomics refers to the global identification of these proteins (Wilkins et al., 1995). Initial proteomic studies in maize (Chang et al., 2000; Porubleva et al., 2001) were based on the two-dimensional electrophoretic (2-DE) separation of protein extracts and the subsequent in-gel visualization of distinct protein spots by a variety of staining techniques (Hochholdinger et al., 2006). Proteins separated on 2-DE gels were then individually excised, proteolytically digested and subjected to mass spectrometry. More recently, protein samples are directly subjected to a shotgun mass spectrometric analysis without prior separation (e.g., Marcon et al., 2015). During mass spectrometry, peptides are ionized by electrospray ionization (ESI) or matrix-assisted laser desorption ionization (MALDI). Subsequently, these peptides are separated in the gas phase, according to their mass-to-charge (m/z) ratio in a mass analyzer and recorded with a detector. Finally, the peptides are identified by comparing the detected mass spectra with theoretical spectra generated from *in silico* digestions of protein databases. Details on the mass spectrometric analysis of proteins have been reviewed elsewhere (e.g., Chait, 2011).

The focus of this review is to highlight the status of the proteomic dissection of processes involved in the endogenous genetic control of maize root development. Endogenous factors that alter the proteome of maize roots such as the availability of mineral nutrients (e.g., Li et al., 2015) or the influence of abiotic stress (e.g., Ghosh and Xu, 2014; Ghatak et al., 2017) will not be discussed.

PROTEOMICS OF MAIZE ROOT SYSTEM ARCHITECTURE

Primary Root

The single primary root is preformed in the embryo and emerges from the basal pole of the seed 2–3 days after germination (Hochholdinger et al., 2004a). It has been demonstrated that the majority of the soluble proteome of maize primary roots of the inbred line B73 significantly changes between 5 and 9 days after germination. Only 28% of proteins identified in 5-day-old

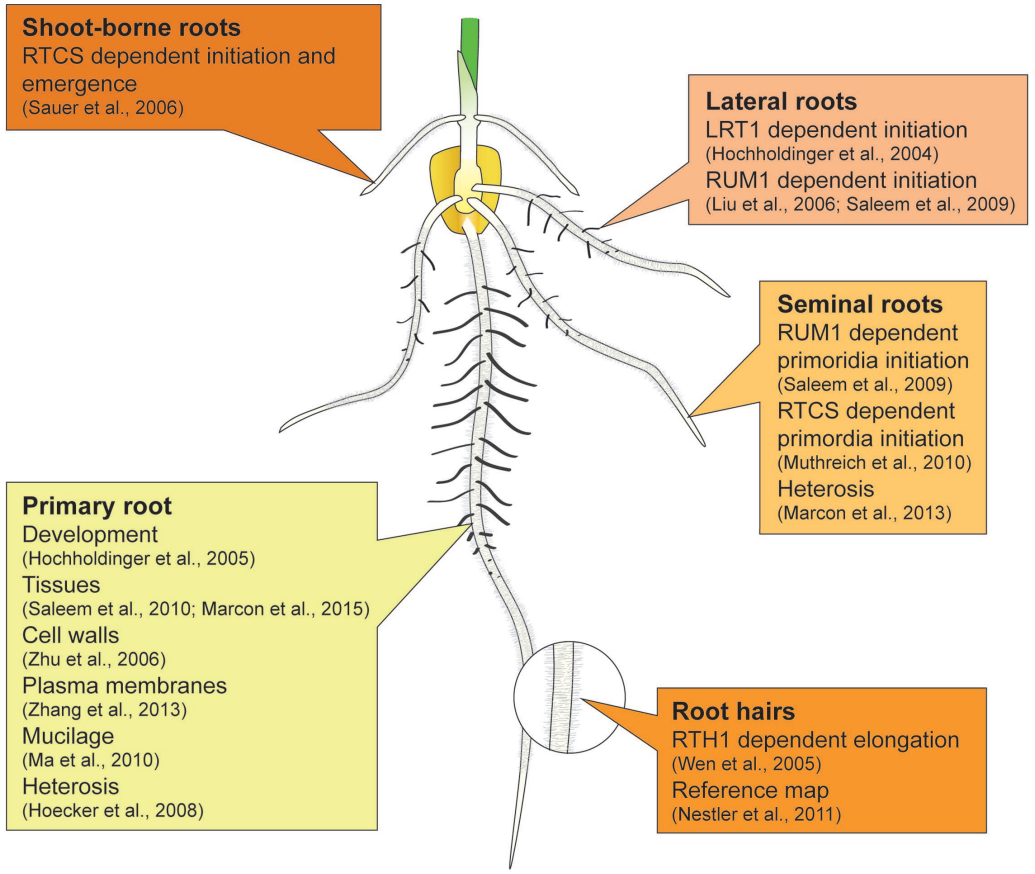


FIGURE 1 | Summary of proteome studies related to individual root types of maize.

primary roots were also observed in 9-day-old primary roots (Hochholdinger et al., 2005).

In longitudinal orientation, maize roots are organized in several partially overlapping zones of development comprising the meristematic, elongation and differentiation zones (Ishikawa and Evans, 1995). To study the molecular function of radial zones of development, the differentiation zone of 2.5-day-old primary roots was manually separated into cortical parenchyma and stele (Saleem et al., 2010). The stele comprises the central vascular cylinder and the pericycle, while the cortical parenchyma includes all tissues surrounding the central vascular cylinder. The cortical parenchyma comprises a single endodermal cell layer, several layers of cortex cells and a single epidermal cell file. The stele transports water, nutrients and photosynthates, while the cortical parenchyma represents the ground tissue of the root, which has metabolic functions but is also involved in the transport of material into the stele. Phytohormone profiling revealed enhanced levels of auxin in the stele and of cytokinin in the cortical parenchyma (Saleem et al., 2010). Several cytokinin-dependent proteins were upregulated in the cortical parenchyma. This localization is consistent with the preferential accumulation of cytokinin in this tissue. Some of these proteins are enzymes, including a β-glucosidase that hydrolyze the cytokinin *cis*-zeatin *O*-glucoside and four enzymes involved in ammonium assimilation (Saleem et al., 2010). The antagonistic levels of auxin and cytokinin in the stele and cortical parenchyma and the cortical parenchyma-specific accumulation of cytokinin-regulated proteins highlight a molecular framework that regulates the function of these distinct root tissues (Saleem et al., 2010). Recently, a high-resolution map of the proteome and phosphoproteome of different tissues of the maize primary root (Marcon et al., 2015) complemented this initial tissue specific proteomic dissection of the maize primary root. In this analysis cortical parenchyma, stele, meristematic zone, and elongation zone were separately subjected to high-performance liquid chromatography coupled with tandem mass spectrometry. As a result, 11,552 distinct non-modified and 2,852 phosphorylated proteins were identified in these four root tissues (Marcon et al., 2015). Along the longitudinal axis, the abundance of functional protein classes was highlighted by two gradients. Along these gradients, the functional classes "RNA", "DNA", and "protein" peaked in the meristematic zone, whereas the categories "cell wall", "lipid metabolism", "stress", "transport", and "secondary metabolism" displayed maximum accumulation in the differentiation zone (Marcon et al., 2015). Comparison of this primary root tissue data set with high-resolution datasets of maize seed (Walley et al., 2013) and leaf (Facette et al., 2013) proteomes revealed that 13% of the identified proteins were exclusively detected in the primary root. These root-specific proteins displayed a high degree of tissue specific functionalization and underscored the plasticity of tissue specific proteomes in maize primary roots (Marcon et al., 2015).

On the subcellular level, the maize primary root was also a model to study the proteomes associated with cell walls (Zhu et al., 2007) and plasma membranes (Zhang et al., 2013). Cell wall proteins have been studied in the elongation zone of maize primary roots (Zhu et al., 2007). Among the identified cell wall proteins, many were involved in cell elongation and cell wall metabolism. Maize forms type II cell walls (Carpita, 1996). Therefore, several of the cell wall proteins identified in the elongation zone of maize primary roots have not been identified in proteomic studies that have focused on type I walls (Zhu et al., 2007). Similarly, the distribution of plasma membrane proteins in the growth zone of the maize primary root was studied during development (Zhang et al., 2013). The plasma membrane is the interface between the plant cell and the extracellular space or the apoplast and thus involved in the exchange of cellular molecules and in signal integration. This study demonstrated that cellulose synthases became more abundant with increasing distance from the root apex. This is consistent with the expected localization of cell wall deposition (Zhang et al., 2013).

Among other compounds, maize root cap cells secrete proteins into the rhizosphere via an amorphous gel structure called mucilage. From the mucilage secreted by 3 to 4-day-old primary roots 2,848 distinct extracellular proteins were identified by nanoLC–MS/MS (Ma et al., 2010). Among those, metabolic proteins represented the largest class. A comparison with the mucilage proteins previously identified in several dicot species suggested a considerable overlap between monocot and dicot mucilage proteomes (Ma et al., 2010).

Seminal Roots

Seminal roots are the second embryonic root type of maize formed after the primary root. They emerge from the scutellar node and their number is variable between different genotypes but also within a given genotype (Hochholdinger et al., 2004a). The maize mutants *rtcs* (*rootless concerning crown and seminal roots*; Hetz et al., 1996) and *rum1* (*rootless with undetectable meristems 1*; Woll et al., 2005) do not initiate seminal roots. Both genes encode key components of auxin signal transduction. Auxin triggers transcriptional responses via the regulation of the activity of AUXIN RESPONSE FACTOR (ARF) proteins (reviewed in, Lavy and Estelle, 2016). At low auxin levels, Auxin/INDOLE-3-ACETIC ACID (Aux/IAA) transcriptional repressors interact with ARFs and repress their activity. In contrast, at high auxin levels TRANSPORT INHIBITOR RESPONSE 1/AUXIN SIGNALING F-BOX PROTEIN (TIR1/AFB) proteins bind to Aux/IAA transcriptional repressors and mediate their degradation via the proteasome. As a consequence, ARF transcription factors activate direct downstream target genes such as LOB domain transcription factors by binding to auxin response elements in their promoter (Xu et al., 2016). While *rtcs* encodes a LOB domain transcription factor (Taramino et al., 2007; Xu et al., 2015, 2016), *rum1* encodes an Aux/IAA transcriptional regulator (von Behrens et al., 2011). It was demonstrated that *rtcs* and *rum1* are not only critical for seminal root initiation but also for seminal root number because they likely underlie the two major QTLs controlling this trait (Salvi et al., 2016). Immature wild-type and *rtcs* embryos were subjected to a comparative proteome analysis 25 days after pollination (Muthreich et al., 2010). At this stage wild-type embryos started the initiation of seminal roots while mutant embryos did not. Among the differentially accumulated proteins, two phosphoglycerate kinases and a malate dehydrogenase,

which are crucial checkpoints of cellular energetics, were preferentially accumulated in wild-type versus mutant embryos (Muthreich et al., 2010). Similarly, the proteomes of wild-type and *rum1* embryos were compared 30 days after pollination (Saleem et al., 2009). In maize, GLOBULIN 1 and GLOBULIN 2 are the most prevalent storage proteins in embryos (Kriz, 1989; Belanger and Kriz, 1991). These proteins are exclusively expressed during embryo development (Belanger and Kriz, 1991; Kriz, 1999). In total, seven GLOBULIN 1 isoforms and four GLOBULIN 2 isoforms were differentially accumulated between wild-type and *rum1* embryos 30 days after pollination suggesting significant regulation of these proteins by RUM1 (Saleem et al., 2009). It was pointed out that the regulation of GLOBULIN1 by the phytohormone abscisic acid (ABA), which stimulates root elongation and branching, might link this class of proteins to root formation (Saleem et al., 2009). Only little overlap was observed between embryonic proteins differentially accumulated between wild-type and *rtcs* embryos 25 days after pollination (Muthreich et al., 2010) and wild-type and *rum1* embryos 30 days after pollination (Saleem et al., 2009). This highlights the distinct regulation of the maize embryo proteome by RTCS and RUM1, but this difference could also be in part due to the analysis of the distinct stages of embryo development.

Shoot-Borne Roots

After germination, whorls of shoot-borne roots are formed at successive nodal structures of the shoot (Hochholdinger et al., 2004b). The maize rootstock develops ∼70 shoot-borne roots in the course of development (Hoppe et al., 1986). Therefore, shoot-borne roots make up the major proportion of the adult maize root system. The previously described maize mutant *rtcs* (Hetz et al., 1996) is impaired in the initiation of all shoot-borne roots. In wild-type maize, the first node from which shoot-borne roots emerge is the coleoptilar node. In a proteome study, the soluble proteomes of wild-type versus mutant *rtcs* coleoptilar nodes were compared 5 and 10 days after germination (Sauer et al., 2006). These stages coincided with the initiation and emergence of shoot-borne roots in wild-type coleoptilar nodes, respectively. Several of the differentially accumulated proteins detected in this study are involved in the regulation of plant development (Sauer et al., 2006). The differentially accumulated proteins at the two developmental stages displayed only little overlap, indicating that distinct sets of proteins control initiation and emergence of lateral roots. RNA gel blot experiments of a subset of differentially accumulated proteins demonstrated that these expression differences are already manifested on the RNA level (Sauer et al., 2006).

Lateral Roots

After germination, lateral roots emerge by division of pericycle cells in the differentiation zone of all root types (Hochholdinger et al., 2004b). The maize mutants *lrt1* (*lateral rootless 1*; Hochholdinger and Feix, 1998) and *rum1* (Woll et al., 2005) are impaired in the initiation of lateral roots. Differentially accumulated proteins were determined in a proteome survey of 9-day-old wild-type primary roots with lateral roots and *lrt1* primary roots without lateral roots (Hochholdinger et al., 2004c). Remarkably, 10% of proteins identified in this study were predominantly accumulated in *lrt1* primary roots that lack lateral roots suggesting that the presence of lateral roots can significantly influence the proteome of the primary root (Hochholdinger et al., 2004c). To study the molecular processes prior to lateral root formation, 2.5-day-old wild-type and mutant *rum1* primary roots were subjected to a proteome analysis (Liu et al., 2006). At this early stage of development no morphological differences were observed between the primary roots of the two genotypes. Nevertheless, differentially accumulated proteins involved in defense, lignin biosynthesis and in the citrate cycle were identified (Liu et al., 2006). In a follow up study the differentiation zone of 2.5-day-old wild-type and mutant *rum1* primary roots were manually separated in cortical parenchyma and stele and subjected to a proteome analysis (Saleem et al., 2009). Differentially accumulated proteins between the mutant *rum1* and wild-type demonstrated that RUM1 regulates the proteome of cortical parenchyma and stele tissues. Among the biochemical pathways regulated by RUM1, enzymes related to glycolysis were differentially expressed between these tissues (Saleem et al., 2009). It has been demonstrated that key enzymes of glycolysis such as hexokinase (Halford and Paul, 2003) and pyruvate kinase (Moore et al., 2003) control root elongation via sugar sensing and signaling. It was therefore suggested that the reduced primary root length of the *rum1* mutant could be conferred by tissue specific imbalances of glycolytic enzymes during root development (Saleem et al., 2009). Finally, in a pioneering work, pericycle cells which give rise to lateral roots were captured via laser microdissection (Schnable et al., 2004) and subjected to a proteome analysis (Dembinsky et al., 2007). As a starting point for future cell-type specific analyses, this initial study identified the twenty most abundant soluble proteins of the maize pericycle cell proteome (Dembinsky et al., 2007).

Root Hairs

Root hairs are tubular extensions of trichoblasts in the epidermis, which significantly increase the absorbing surface of the root and thus the uptake of nutrients (Gilroy and Jones, 2000). In maize, root hair patterning is random and it is not predictable which epidermal cells form root hairs (Dolan, 1996). Several genes control root hair elongation in maize. The *root hair defective* genes *rth3* (Hochholdinger et al., 2008), *rth5* (Nestler et al., 2014), and *rth6* (Li et al., 2015) are functionally linked in the processes of cell wall loosening, cellulose synthesis, and organization of synthesized cellulose, respectively. In contrast, the gene *rth1* (Wen et al., 2005) encodes a SEC3 subunit of the exocyst complex (Hala et al., 2008). To gain a better insight into the root hair proteome, a reference set of 2,573 soluble proteins of maize root hairs of 4-day-old primary roots of the inbred line B73 was identified (Nestler et al., 2011). Root hairs are an ideal model for single cell proteome analyses in maize because they can be easily separated from the primary root after freezing the roots in liquid nitrogen. Among these root hair proteins, homologs of 252 proteins have been associated with root hair development in other species

(Nestler et al., 2011). A comparison of the maize root hair reference proteome with those of soybean revealed conserved, but also unique protein functions related to root hairs in these species (Nestler et al., 2011).

PROTEOMICS OF HETEROSIS MANIFESTATION IN MAIZE ROOTS

Heterozygous F_1-hybrids of maize typically perform better than their homozygous genetically distinct parental inbred lines (Hochholdinger and Hoecker, 2007). This phenomenon, known as heterosis can already be observed during early maize primary root development (Hoecker et al., 2006). It has been demonstrated that very young primary roots of hybrids display protein accumulation patterns different from the average of their parental values even before the manifestation of developmental differences in comparison to their parental inbred lines (Hoecker et al., 2008). Among these non-additively accumulated proteins, the functional classes "metabolism" and "disease/defense" were the most abundant (Hoecker et al., 2008). Similar patterns of non-additive protein accumulation were observed in maize embryos of reciprocal hybrids and their parental inbred lines 25 and 35 days after pollination (Marcon et al., 2010). In these embryos, the functional classes "development", "protein metabolism", "redox-regulation", "glycolysis", and "amino acid metabolism" were most prevalent among non-additively accumulated proteins. In 35-day-old embryos, enzymes related to "glucose metabolism" were significantly upregulated and their expression often even exceeded the expression level of the better parent (Marcon et al., 2010). Finally, the 970 most abundant soluble proteins of 2–4 cm long seminal roots were quantified by label-free LC–MS/MS in a shot-gun approach (Marcon et al., 2013). Consistent with the results of the comparative proteomic dissection of primary roots (Hoecker et al., 2008) and embryos (Marcon et al., 2010), the category "protein metabolism" was also the most prevalent functional class of non-additive proteins in seminal roots (Marcon et al., 2013) when comparing hybrids and their parental inbred lines. In this category, 16 of 17 non-additively accumulated ribosomal proteins displayed high or above better-parent expression in seminal roots (Marcon et al., 2013).

The major findings summarized in Sections "Proteomics of Maize Root System Architecture" and "Proteomics of Heterosis Manifestation in Maize Roots" have been highlighted in **Figure 1**.

CONCLUSION

In recent years, numerous studies have surveyed proteomic changes in different root types of maize to understand the complex molecular interactions during root development. Proteome studies are an important tool to complement transcriptome studies because proteins are the biologically active molecules in the cell and there is typically only a moderate correlation between the levels of a transcript and protein encoded by a same gene (e.g., Marcon et al., 2015). This discrepancy can be explained by differences in RNA versus protein stability but also by the fact that one gene can give rise to many proteins by alternative splicing or post-translational modifications. Proteome studies are technically more challenging than transcriptome studies because it is impossible to solubilize all proteins in a single experiment because each tissue contains proteins that range from highly hydrophobic to hydrophilic. In contrast to nucleic acids, proteins cannot be amplified and therefore lowly abundant proteins cannot be detected. Despite these technical limitations it was recently possible to survey tissue specific differences of the primary root proteome and phosphoproteome in high resolution detecting more than 11,000 proteins. With the advent of mass spectrometry, even higher resolution proteome analyses and thus more detailed accounts of proteomic changes during maize root development will become available. While transcriptome analyses are already performed on the level of individual cell-types, such proteome analyses are currently only available for very specific and exposed cell-types that can be easily separated from the rest of the root such as root hair cells (Nestler et al., 2011). With future technological advances of mass spectrometry, the detection of minute amounts of proteins isolated from complex tissues by techniques such as laser microdissection might become available.

AUTHOR CONTRIBUTIONS

All authors contributed to the writing of this minireview. JB drew the maize seedling displayed in **Figure 1**.

ACKNOWLEDGMENTS

Root research in FH's laboratory is supported by the DFG (German Research Foundation) and the BMBF (Federal Ministry of Education and Research).

REFERENCES

Atkinson, J. A., Rasmussen, A., Traini, R., Voß, U., Sturrock, C., Mooney, S. J., et al. (2014). Branching out in roots: uncovering form, function, and regulation. *Plant Physiol.* 166, 538–550. doi: 10.1104/pp.114.245423

Belanger, F. C., and Kriz, A. L. (1991). Molecular basis for allelic polymorphism of the maize *Globulin-1* gene. *Genetics* 129, 863–872.

Carpita, N. C. (1996). Structure and biogenesis of the cell walls of grasses. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47, 445–476. doi: 10.1146/annurev.arplant.47.1.445

Chait, B. T. (2011). Mass spectrometry in the postgenomic era. *Annu. Rev. Biochem.* 80, 239–246. doi: 10.1146/annurev-biochem-110810-095744

Chang, W. W., Huang, L., Shen, M., Webster, C., Burlingame, A. L., and Roberts, J. K. (2000). Patterns of protein synthesis and tolerance of anoxia in root tips of maize seedlings acclimated to a low-oxygen environment, and identification of proteins by mass spectrometry. *Plant Physiol.* 122, 295–318. doi: 10.1104/pp.122.2.295

Dembinsky, D., Woll, K., Saleem, M., Liu, Y., Fu, Y., Borsuk, L. A., et al. (2007). Transcriptomic and proteomic analyses of pericycle cells of the maize (*Zea mays* L.) primary root. *Plant Physiol.* 145, 575–588. doi: 10.1104/pp.107.106203

Dolan, L. (1996). Pattern in the root epidermis: an interplay of diffusible signals and cellular geometry. *Ann. Bot.* 77, 547–553. doi: 10.1093/aob/77.6.547

Facette, M. R., Shen, Z., Björnsdóttir, F. R., Briggs, S. P., and Smith, L. G. (2013). Parallel proteomic and phosphoproteomic analyses of successive stages of maize leaf development. *Plant Cell* 25, 2798–2812. doi: 10.1105/tpc.113.112227

Ghatak, A., Chaturvedi, P., and Weckwerth, W. (2017). Cereal crop proteomics: systemic analysis of crop drought stress responses towards marker-assisted selection breeding. *Front. Plant Sci.* 8:757. doi: 10.3389/fpls.2017.00757

Ghosh, D., and Xu, J. (2014). Abiotic stress responses in plant roots: a proteomics perspective. *Front. Plant Sci.* 5:6. doi: 10.3389/fpls.2014.00006

Gilroy, S., and Jones, D. L. (2000). Through form to function: root hair development and nutrient uptake. *Trends Plant Sci.* 5, 56–60. doi: 10.1016/S1360-1385(99)01551-4

Hala, M., Cole, R., Synek, L., Drdova, E., Pecenkova, T., Nordheim, A., et al. (2008). An exocyst complex functions in plant cell growth. *Plant Cell* 20, 1330–1345. doi: 10.1105/tpc.108.059105

Halford, N. G., and Paul, M. J. (2003). Carbon metabolite sensing and signalling. *Plant Biotechnol. J.* 1, 381–398. doi: 10.1046/j.1467-7652.2003.00046.x

Hetz, W., Hochholdinger, F., Schwall, M., and Feix, G. (1996). Isolation and characterization of *rtcs*, a maize mutant deficient in the formation of nodal roots. *Plant J.* 10, 845–857. doi: 10.1046/j.1365-313X.1996.10050845.x

Hochholdinger, F. (2009). "The maize root system: morphology, anatomy and genetics," in *The Handbook of Maize*, eds J. Bennetzen and S. Hake (New York, NY: Springer), 145–160. doi: 10.1007/978-0-387-79418-1

Hochholdinger, F., and Feix, G. (1998). Early post-embryonic root formation is specifically affected in the maize mutant *lrt1*. *Plant J.* 16, 247–255. doi: 10.1046/j.1365-313x.1998.00280.x

Hochholdinger, F., Guo, L., and Schnable, P. S. (2004a). Lateral roots affect the proteome of the primary root of maize (*Zea mays* L.). *Plant Mol. Biol.* 56, 397–412. doi: 10.1007/s11103-004-3476-9

Hochholdinger, F., and Hoecker, N. (2007). Towards the molecular basis of heterosis. *Trends Plant Sci.* 12, 427–432. doi: 10.1016/j.tplants.2007.08.005

Hochholdinger, F., Park, W. J., Sauer, M., and Woll, K. (2004b). From weeds to crops: genetic analysis of root development in cereals. *Trends Plant Sci.* 9, 42–48.

Hochholdinger, F., Sauer, M., Dembinsky, D., Hoecker, N., Muthreich, N., Saleem, M., et al. (2006). Proteomic dissection of plant development. *Proteomics* 6, 4076–4083. doi: 10.1002/pmic.200500851

Hochholdinger, F., Wen, T. J., Zimmermann, R., Chimot-Marolle, P., da Costa e Silva, O., Bruce, W., et al. (2008). The maize (*Zea mays* L.) *roothairless3* gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield. *Plant J.* 54, 888–898. doi: 10.1111/j.1365-313X.2008.03459.x

Hochholdinger, F., Woll, K., Guo, L., and Schnable, P. S. (2005). The accumulation of abundant soluble proteins changes early in the development of the primary roots of maize (*Zea mays* L.). *Proteomics* 5, 4885–4893. doi: 10.1002/pmic.200402034

Hochholdinger, F., Woll, K., Sauer, M., and Dembinsky, D. (2004c). Genetic dissection of root formation in maize (*Zea mays*) reveals root-type specific developmental programs. *Ann. Bot.* 93, 359–368.

Hochholdinger, F., Yu, P., and Marcon, C. (2017). Genetic control of root system development in maize. *Trends Plant Sci.* 23, 79–88. doi: 10.1016/j.tplants.2017.10.004

Hoecker, N., Keller, B., Piepho, H.-P., and Hochholdinger, F. (2006). Manifestation of heterosis during early maize (*Zea mays* L.) root development. *Theor. Appl. Genet.* 112, 421–429. doi: 10.1007/s00122-005-0139-4

Hoecker, N., Lamkemeyer, T., Sarholz, B., Paschold, A., Fladerer, C., Madlung, J., et al. (2008). Analysis of non-additive protein accumulation in young primary roots of a maize (*Zea mays* L.) F1-hybrid compared to its parental inbred lines. *Proteomics* 8, 3882–3894. doi: 10.1002/pmic.200800023

Hoppe, D. C., McCully, M. E., and Wenzel, C. L. (1986). The nodal roots of *Zea*: their development in relation to structural features of the stem. *Can. J. Bot.* 64, 2524–2537. doi: 10.1139/b86-335

Ishikawa, H., and Evans, M. L. (1995). Specialized zones of development in roots. *Plant Physiol.* 109, 725–727. doi: 10.1104/pp.109.3.725

Kriz, A. L. (1989). Characterization of embryo globulins encoded by the maize Glb genes. *Biochem. Genet.* 27, 239–251. doi: 10.1007/BF02401804

Kriz, A. L. (1999). "7S globulins of cereals," in *Seed Proteins*, eds P. R. Shewry and R. Casey (Dordrecht: Kluwer Academic Publishers), 477–498. doi: 10.1007/978-94-011-4431-5_20

Lavy, M., and Estelle, M. (2016). Mechanisms of auxin signaling. *Development* 143, 3226–3229. doi: 10.1242/dev.131870

Li, Z., Phillip, D., Neuhäuser, B., Schulze, W. X., and Ludewig, U. (2015). Protein dynamics in young maize root hairs in response to macro and micronutrient deprivation. *J. Proteome Res.* 14, 3362–3371. doi: 10.1021/acs.jproteome.5b00399

Liu, Y., Lamkemeyer, T., Jakob, A., Mi, G., Zhang, F., Nordheim, A., et al. (2006). Comparative proteome analyses of maize (*Zea mays* L.) primary roots prior to lateral root initiation reveal differential protein expression in the lateral root initiation mutant *rum1*. *Proteomics* 6, 4300–4308. doi: 10.1002/pmic.200600145

Ma, W., Muthreich, N., Liao, C., Franz-Wachtel, M., Schütz, W., Zhang, F., et al. (2010). The mucilage proteome of maize (*Zea mays* L.) primary roots. *J. Proteome Res.* 9, 2968–2976. doi: 10.1021/pr901168v

Marcon, C., Lamkemeyer, T., Malik, W. A., Ungrue, D., Piepho, H.-P., and Hochholdinger, F. (2013). Heterosis-associated proteome analyses of maize (*Zea mays* L.) seminal roots by quantitative label-free LC-MS. *J. Proteomics* 93, 295–302. doi: 10.1016/j.jprot.2013.04.015

Marcon, C., Malik, W. A., Walley, J. W., Shen, Z., Paschold, A., Smith, L. G., et al. (2015). A high resolution tissue-specific proteome and phosphoproteome atlas of maize primary roots reveals functional gradients along the root axis. *Plant Physiol.* 168, 233–246. doi: 10.1104/pp.15.00138

Marcon, C., Schützenmeister, A., Schütz, W., Madlung, J., Piepho, H. P., and Hochholdinger, F. (2010). Non-additive protein accumulation patterns in maize (*Zea mays* L.) hybrids during embryo development. *J. Proteome Res.* 9, 6511–6522. doi: 10.1021/pr100718d

Moore, B., Zhou, L., Rolland, F., Hall, Q., Cheng, W. S., Liu, Y.-X., et al. (2003). Role of the *Arabidopsis* glucose sensor HXK1 in nutrient, light, and hormonal signaling. *Science* 300, 332–336. doi: 10.1126/science.1080585

Muthreich, N., Schützenmeister, A., Schütz, W., Madlung, J., Krug, K., Nordheim, A., et al. (2010). Regulation of the maize (*Zea mays* L.) embryo proteome by RTCS which controls seminal root initiation. *Eur. J. Cell Biol.* 89, 242–249. doi: 10.1016/j.ejcb.2009.11.014

Nestler, J., Liu, S., Wen, T.-J., Paschold, A., Marcon, C., Tang, H. M., et al. (2014). *Roothairless5*, which functions in maize (*Zea mays* L.) root hair initiation and elongation encodes a monocot-specific NADPH oxidase. *Plant J.* 79, 729–740. doi: 10.1111/tpj.12578

Nestler, J., Schütz, W., and Hochholdinger, F. (2011). Conserved and unique features of the maize root hair proteome. *J. Proteome Res.* 10, 2525–2537. doi: 10.1021/pr200003k

Porubleva, L., Van der Velden, K., Kothari, S., Oliver, D. J., and Chitnis, P. R. (2001). The proteome of maize leaves: use of gene sequences and expressed sequence tag data for identification of proteins with peptide mass fingerprints. *Electrophoresis* 22, 1724–1738. doi: 10.1002/1522-2683(200105)22:9<1724::AID-ELPS1724>3.0.CO;2-2

Saleem, M., Lamkemeyer, T., Schuetzenmeister, A., Fladerer, C., Piepho, H.-P., Nordheim, A., et al. (2009). Tissue specific control of the maize (*Zea mays* L.) embryo, cortical parenchyma, and stele proteomes by RUM1 which regulates seminal and lateral root initiation. *J. Proteome Res.* 8, 2285–2297. doi: 10.1021/pr8009287

Saleem, M., Lamkemeyer, T., Schutzenmeister, A., Madlung, J., Sakai, H., Piepho, H.-P., et al. (2010). Specification of cortical parenchyma and stele of maize (*Zea mays* L.) primary roots by asymmetric levels of auxin, cytokinin and cytokinin-regulated proteins. *Plant Physiol.* 152, 4–18. doi: 10.1104/pp.109.150425

Salvi, S., Giuliani, S., Ricciolini, C., Carraro, N., Maccaferri, M., Presterl, T., et al. (2016). Two major quantitative trait loci controlling the number of seminal roots in maize co-map with the root developmental genes *rtcs* and *rum1*. *J. Exp. Bot.* 67, 1149–1159. doi: 10.1093/jxb/erw011

Sanguineti, M. C., Giuliani, M. M., Govi, G., Tuberosa, R., and Landi, P. (1998). Root and shoot traits of maize inbred lines grown in the field and in hydroponic culture and their relationships with root lodging. *Maydica* 43, 211–216.

Sauer, M., Jakob, A., Nordheim, A., and Hochholdinger, F. (2006). Proteomic analysis of shoot-borne root initiation in maize (*Zea mays* L.). *Proteomics* 6, 2530–2541. doi: 10.1002/pmic.200500564

Schnable, P. S., Hochholdinger, F., and Nakazono, M. (2004). Global expression profiling applied to plant development. *Curr. Opin. Plant Biol.* 7, 50–56. doi: 10.1016/j.pbi.2003.11.001

Taramino, G., Sauer, M., Stauffer, J., Multani, D., Niu, X., Sakai, H., et al. (2007). The *rtcs* gene in maize (*Zea mays* L.) encodes a lob domain protein that is required for postembryonic shoot-borne and embryonic seminal root initiation. *Plant J.* 50, 649–659. doi: 10.1111/j.1365-313X.2007.03075.x

von Behrens, I., Komatsu, M., Zhang, Y., Berendzen, K. W., Niu, X., Sakai, H., et al. (2011). *Rootless with undetectable meristem1* encodes a monocot-specific AUX/IAA protein that controls embryonic seminal and postembryonic lateral root initiation in maize. *Plant J.* 66, 341–353. doi: 10.1111/j.1365-313X.2011.04495.x

Walley, J. W., Shen, Z., Sartor, R., Wu, K. J., Osborn, J., Smith, L. G., et al. (2013). Reconstruction of protein networks from an atlas of maize seed proteotypes. *Proc. Natl. Acad. Sci. U.S.A.* 110, E4808–E4817. doi: 10.1073/pnas.1319113110

Wen, T. J., Hochholdinger, F., Sauer, M., Bruce, W., and Schnable, P. S. (2005). The *roothairless1* gene of maize (*Zea mays*) encodes a homolog of *sec3*, which is involved in polar exocytosis. *Plant Physiol.* 138, 1637–1643. doi: 10.1104/pp.105.062174

Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F., et al. (1995). Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.* 13, 19–50. doi: 10.1080/02648725.1996.10647923

Woll, K., Borsuk, L., Stransky, H., Nettleton, D., Schnable P. S., and Hochholdinger, F. (2005). Isolation, characterization, and pericycle-specific transcriptome analyses of the novel maize lateral and seminal root initiation mutant *rum1*. *Plant Physiol.* 139, 1255–1267. doi: 10.1104/pp.105.067330

Xu, C., Luo, F., and Hochholdinger, F. (2016). LOB domain proteins: beyond lateral organ boundaries. *Trends Plant Sci.* 21, 159–167. doi: 10.1016/j.tplants.2015.10.010

Xu, C., Tai, H., Saleem, M., Ludwig, Y., Majer, C., Berendzen, K., et al. (2015). Cooperative action of the paralogous maize LOB domain proteins RTCS and RTCL in shoot-borne root formation. *New Phytol.* 207, 1123–1133. doi: 10.1111/nph.13420

Yu, P., Gutjahr, C., Li, C., and Hochholdinger, F. (2016). Genetic control of lateral root formation in cereals. *Trends Plant Sci.* 21, 951–961. doi: 10.1016/j.tplants.2016.07.011

Zhang, Z., Voothuluru, P., Yamaguchi, M., Sharp, R. E., and Peck, S. C. (2013). Developmental distribution of the plasma membrane-enriched proteome in the maize primary root growth zone. *Front. Plant Sci.* 4:33. doi: 10.3389/fpls.2013.00033

Zhu, J., Alvarez, S., Marsh, E. L., Lenoble, M. E., Cho, I. J., Sivaguru, M., et al. (2007). Cell wall proteome in the maize primary root elongation zone. II. Region-specific changes in water soluble and lightly ionically bound proteins under water deficit. *Plant Physiol.* 145, 1533–1548. doi: 10.1104/pp.107.107250

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Hochholdinger, Marcon, Baldauf, Yu and Frey. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

CellPress
REVIEWS

Review

Genetic Control of Root System Development in Maize

Frank Hochholdinger,[1,*] Peng Yu,[1] and Caroline Marcon[1]

The maize root system comprises structurally and functionally different root types. Mutant analyses have revealed that root-type-specific genetic regulators intrinsically determine the maize root system architecture. Molecular cloning of these genes has demonstrated that key elements of auxin signal transduction, such as LOB domain (LBD) and Aux/IAA proteins, are instrumental for seminal, shoot-borne, and lateral root initiation. Moreover, genetic analyses have demonstrated that genes related to exocytotic vesicle docking, cell wall loosening, and cellulose synthesis and organization control root hair elongation. The identification of upstream regulators, protein interaction partners, and downstream targets of these genes together with cell-type-specific transcriptome analyses have provided novel insights into the regulatory networks controlling root development and architecture in maize.

Trends

The maize root system comprises different root types formed at different stages of development with distinct anatomic and functional properties.

Genetic analyses have demonstrated that genes involved in auxin signal transduction are key elements for the initiation of the seminal, lateral, and shoot-borne roots of maize.

High-resolution tissue- and cell-type-specific transcriptome studies have identified genes and molecular pathways associated with the initiation of seminal, lateral, and shoot-borne roots.

Genetic analyses have demonstrated that genes involved in exocytotic vesicle fusion and cellulose synthesis are instrumental for root hair elongation.

The Maize Root System as a Genetic Model for Cereals

Maize has a wider global geographic distribution than any other major crop species [1]. Moreover, maize is the crop species with the highest global annual yield. The International Grains Council predicted that global maize production would reach an all-time high in 2016/2017 with a forecasted yield of 1035 million tons per year (http://www.igc.int/en/downloads/grainsupdate/IGC_5year_projections2016.pdf). The unique root system architecture of maize is instrumental for its superior productivity because it efficiently extracts mineral nutrients and water from soil and allocates them to the energy-delivering, aboveground part of the plant. Nevertheless, the root system remains a largely untapped resource for future maize improvement [2–6].

The complex root system of maize comprises embryonic roots formed during embryogenesis and postembryonic roots initiated after germination [7–9]. The embryonic root system comprises a single primary root and a variable number of seminal roots. These roots are important for seedling vigor during early development [10]. Postembryonic shoot-borne roots dominate the root system of adult plants [11]. Finally, all root types form postembryonic lateral roots. They significantly increase the absorbing surface of the maize root system and are common to all main root types of maize [12].

In the longitudinal direction, all maize roots represent a developmental gradient with the youngest cells in the root apex and the oldest cells in the proximal part of the root. This developmental gradient comprises three partially overlapping zones [13]. In the root apex, undifferentiated cells divide in the meristematic zone. Subsequently, these cells increase in size in the elongation zone and reach their final size and function in the differentiation zone. A specific attribute of the differentiation zone is the formation of root hairs. Root hairs are tubular extensions of epidermal cells and significantly contribute to the increase of the absorbing surface of the root [14].

[1]INRES, Institute of Crop Science and Resource Conservation, Crop Functional Genomics, University of Bonn, 53113 Bonn, Germany

*Correspondence:
hochholdinger@uni-bonn.de
(F. Hochholdinger).
URL: http://www.cfg.uni-bonn.de.

© 2017 Elsevier Ltd. All rights reserved.

In the radial orientation, the maize root system comprises functionally distinct cell types and tissues [8]. The stele, which represents the central vascular cylinder, contains xylem vessels responsible for the transport of water and nutrients and phloem vessels for the transport of photosynthates. Pericycle cells in the outermost layer of the central cylinder and endodermal cells in the innermost layer of the ground tissue connect these tissues. A variable number of cortex cell files constitute the majority of the ground tissue in maize. The epidermis, which contains root-hair-forming trichoblasts and non-root-hair-forming atrichoblasts is the outermost cell layer of maize roots and connects the root system to the rhizosphere. The functional diversity of these zones and tissues of maize roots is illustrated by remarkable expression differences along the longitudinal and across the radial axis as demonstrated in the transcriptome [15–18].

All maize root types display a similar anatomical organization. Nevertheless, differences in cell size and the number of cortical cell files can lead to significant differences in the average diameter of the various root types [19]. Moreover, the organization and number of vascular elements can differ between different maize root types [19] or between different whorls of shoot-borne roots along their development [20,21].

Genetic Dissection of Maize Root Development

The ease of controlled genetic crosses as a result of having separate male and female reproductive organs made maize a favored model species for plant geneticists more than 100 years ago. Carl Correns had reported the earliest examples of Mendelian inheritance in maize by 1901 [22], just 1 year after de Vries, Tschermak, and Correns himself independently rediscovered Mendel's laws of inheritance. By 1935 an impressive collection of >350 phenotypic mutations had been described in maize [23]. Among those, *rootless* (*rt*), which displays a reduced number of shoot-borne roots, was the only mutant affected in root development [24].

The difficulty of finding maize mutants defective in root development has several reasons. Most importantly and obviously, roots are hidden in soil and are therefore, in contrast to the aboveground part of the plant, not directly accessible to genetic analyses. In line with this, the first root mutant *rt* was initially identified by a secondary effect because these plants toppled over when the aboveground parts became heavier than the limited root system [24]. Only subsequent analyses revealed that a significant reduction of the shoot-borne root system conditions the aboveground phenotype of this early mutant [24]. Another factor that complicates the identification of root mutants is the exceptional degree of plasticity necessary to adapt root architecture to changing conditions in their soil environment. Moreover, maize root systems are of considerable size and complexity because different root types emerge at different phases of development. The problems imposed by the plasticity and complexity of maize root systems were circumvented in the first systematic screens for monogenic maize root mutants in the 1990s [25,26]. In these screens mutagenized F_2 populations grew under standardized conditions in paper rolls and Mendelian segregation ratios of root traits were determined at the seedling stage within 2 weeks after germination. This system has several advantages. First, growth under standardized conditions and analysis at the seedling stage reduces the developmental plasticity of root stock development. Moreover, most root types of maize, including primary, seminal, shoot-borne, and lateral roots but also root hairs, develop within 2 weeks after germination. Furthermore, this growth system allows the analysis of intact root systems after opening of the paper rolls. By contrast, extraction of root systems from their natural soil environment typically damages the root system. Finally, the limited space needed by seedlings in the growth chamber or greenhouse allows large-scale genetic screens.

Molecular Cloning and Characterization of Maize Genes Controlling Root System Architecture and Root Hair Formation

In recent years several genes controlling major checkpoints of maize root development and root hair formation have been cloned and characterized (Table 1). These genes and their molecular functions, which have been identified based on their mutant phenotypes, are highlighted in the following paragraphs along with the different root types of maize.

Shoot-Borne Roots

Shoot-borne roots emerging from consecutive shoot nodes are unique to monocot species [8]. They provide mechanical support and constitute the major backbone of the adult root system. The mutant *rtcs* is completely devoid of shoot-borne roots [27] (Figure 1). The mutation in *rtcs* acts before the initiation of shoot-borne roots, as demonstrated by the absence of shoot-borne root primordia in the coleoptilar and consecutive shoot nodes [27]. In line with the lack of shoot-borne root primordia, four maize cyclins, which are molecular markers for meristematic activity, did not show any expression in shoot nodes of the mutant *rtcs* [28]. By contrast, the cyclin-dependent kinase *cdc2*, a marker for the competence of cell division, was expressed in the mutant [28]. Hence, the affected cells are still competent to divide but the mutation in *rtcs* suppresses their division and proliferation. That these cells are able to divide was later demonstrated by introducing the mutant *rtcs* allele into *gaspe flint*, a genotype that is characterized by extensive tillering from lower shoot nodes [29]. While *rtcs* still did not initiate any shoot-borne roots, tillers emerged from the nodal tissue of the mutant after *gaspe flint* introgression [29].

Map-based cloning demonstrated that *rtcs* encodes a member of the plant-specific family of LBD transcription factors [30]. LBD proteins are essential for lateral organ development in higher plants [31]. Phylogenetic analyses revealed that the maize genome contains a paralog of *rtcs*, which was designated *rtcs-like* (*rtcl*) [30]. The *rtcs–rtcl* gene pair emerged by the last whole-genome duplication of maize *ca* 5–12 Mya [32]. Functional analyses of the *rtcs* and *rtcl* genes demonstrated that they are both involved in auxin signal transduction [30] (Figure 2A). Expression of *rtcs* and *rtcl* is activated by binding of auxin response factor (ARF)34 to LBD elements (5′-GCGGCG-3′) in their promoter [33]. As expected for transcription factors, *rtcs* and *rtcl* are localized to the nucleus [34] and can bind to LBD domains in promoter elements of downstream genes [33,34]. The RTCS and RTCL proteins can homo- and heterointeract with each other [34]. Their interactions with other proteins are in part conserved but also in part unique to RTCS and RTCL [33]. Genetic analyses demonstrated that the *rtcl* mutant displays significantly shorter shoot-borne roots, demonstrating that this gene is involved in the elongation of this root type [33] (Figure 1). Hence, while *rtcs* controls shoot-borne root initiation, its

Table 1. Cloned Genes with Known Function in Maize Root Development

Gene name	Affected developmental process	Gene function	Gene family	Refs
rtcs	Shoot-borne and seminal root initiation	Auxin signaling	LOB domain	[30]
rtcl	Shoot-borne root elongation	Auxin signaling	LOB domain	[33]
rum1	Seminal and lateral root initiation	Auxin signaling	Aux/IAA	[43]
bige1	Number of seminal and shoot-borne roots	Multidrug and toxin extrusion transport	MATE	[39]
rth1	Root hair elongation	Exocytotic vesicle fusion	SEC3	[70]
rth3	Root hair elongation	Cellulose organization	COBRA-like	[69]
rth5	Root hair elongation	Cell wall loosening	NADPH oxidase	[66]
rth6	Root hair elongation	Cellulose synthesis	CSLD5	[67,68]

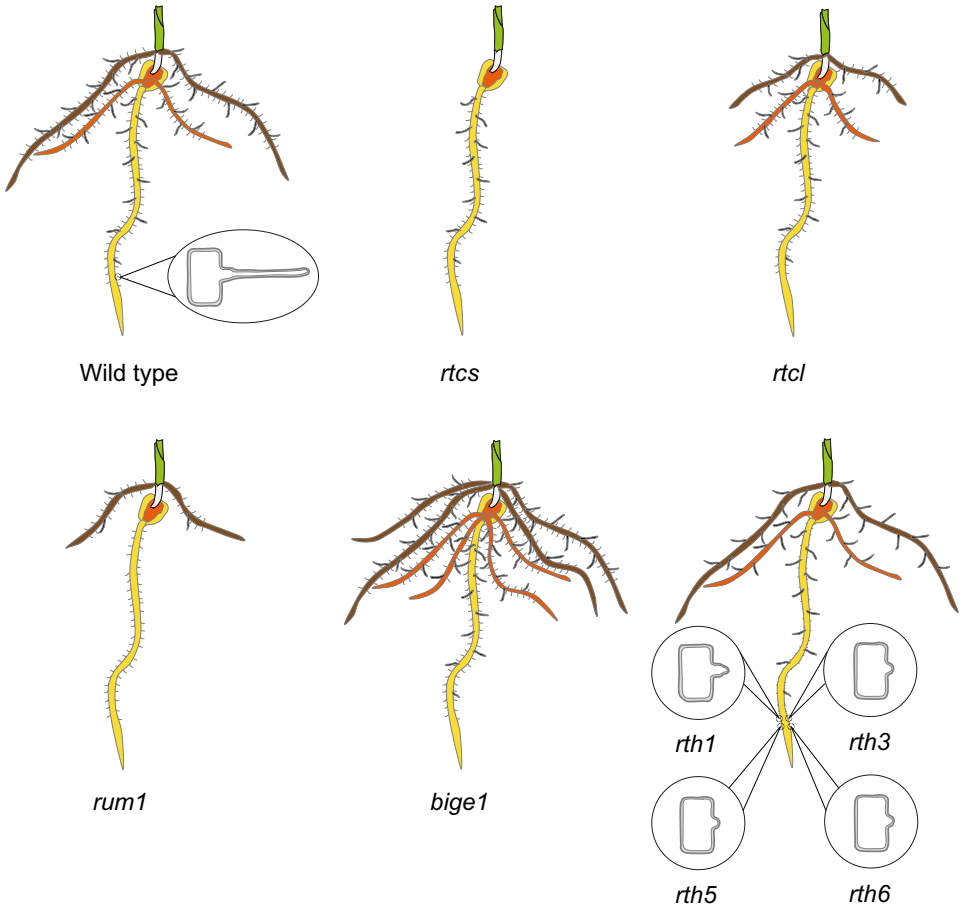
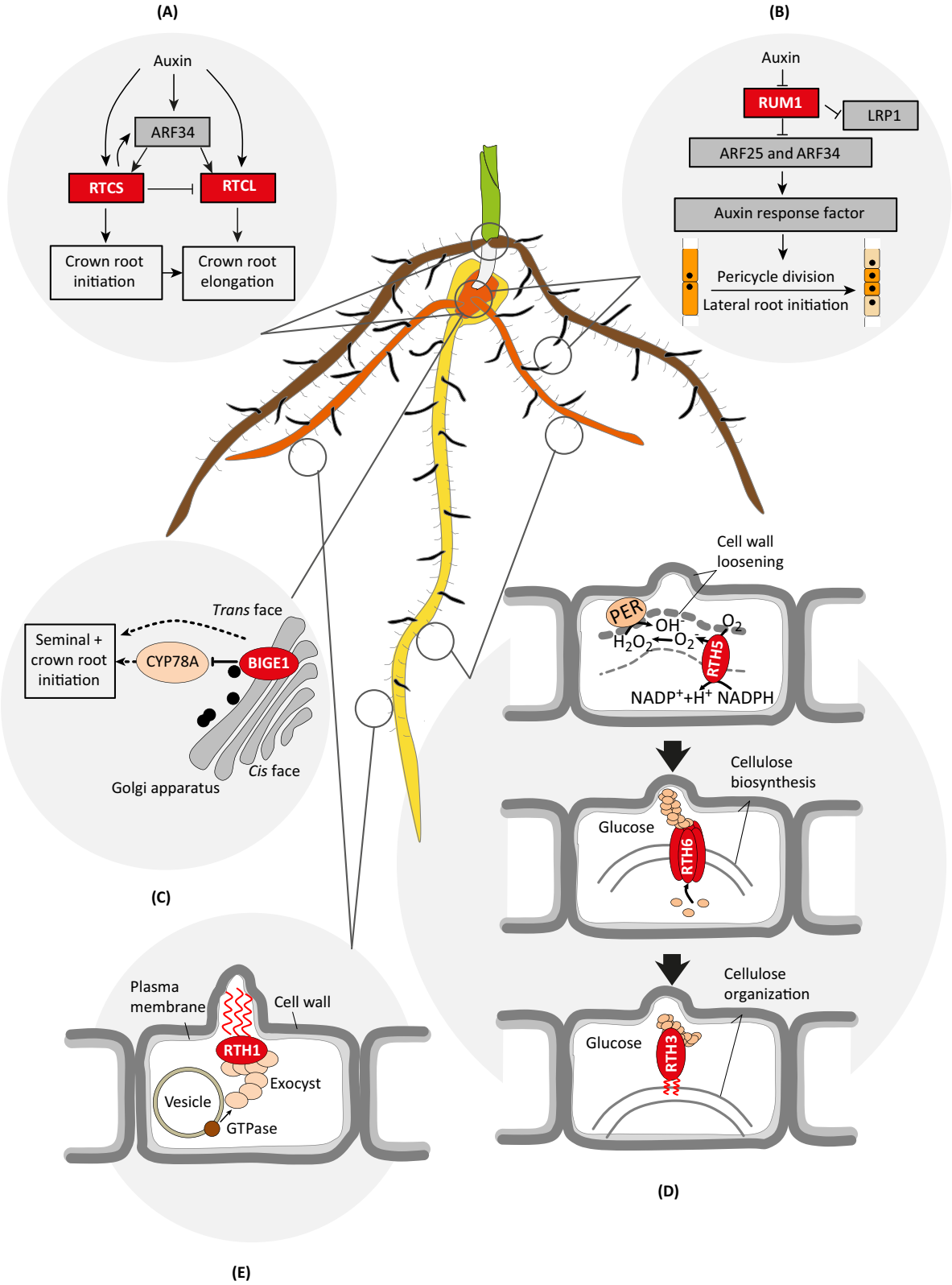


Figure 1. Seedling Root Phenotypes of Monogenic Mutants of Maize for Which the Affected Genes Have Been Cloned. For details on these phenotypes, refer to the main text and to Table 1. Color code for root structures: yellow, primary root; orange, seminal roots; brown, shoot-borne crown roots; black, lateral roots; grey, root hairs; orange, scutellar node from which seminal roots emerge; white, mesocotyl connecting the seed and the first shoot node from which shoot-borne roots emerge. Root hair length is illustrated by enlarged epidermal cells in enlarged circles. *rtcs*, *rootless concerning crown and seminal roots*; *rtcl*, *rootless concerning crown and seminal roots like*; *rum1*, *rootless with undetectable meristem 1*; *bige1*, *big embryo 1*; *rth*, *roothairless*.

paralog *rtcl* controls the distinct developmental process of shoot-borne root elongation. Recent physiological studies revealed that crown root initiation and elongation are coordinated under drought-stress conditions by the formation of fewer but longer crown roots [35]. Moreover, it has been demonstrated that *rtcs* is involved in the suppression of shoot-borne root formation in response to water deficit [36]. The previously demonstrated interaction of RTCS and RTCL proteins [34], which control shoot-borne root initiation and elongation, could be the link that coordinates these distinct developmental processes under drought conditions.

The allelic rice genes *ARL1* [37] and *CRL1* [38], which are homologs of maize *rtcs*, also control the initiation of shoot-borne roots. In contrast to the *rtcs* mutant, which is not affected in lateral root formation, mutations in the *crl1/arl1* gene of rice result in a 30% reduction of lateral roots [38]. Although *rtcs* and *rtcl* exhibit highly correlated and overlapping expression patterns in the primary root, no reduction in lateral root density was detectable in *rtcs–rtcl* double mutants [33], suggesting that lateral root formation might be controlled by other genes.



Trends in Plant Science

Figure 2. Molecular Function and Context of Proteins Associated with Maize Root Development. (A) RTCS and RTCL are paralogous LOB domain proteins involved in auxin signal transduction controlling shoot-borne and seminal root initiation and shoot-borne root elongation, respectively. (B) RUM1 is an Aux/IAA protein

(See figure legend on the bottom of the next page.)

Finally, it was recently demonstrated that loss-of-function mutations in the *Big embryo 1* (*Bige1*) gene encoding a multidrug and toxin extrusion (MATE) transporter cause an increased number of crown roots in maize [39].

Lateral Roots

Lateral roots are initiated postembryonically from phloem pole pericycle cells of all root types and therefore contribute significantly to the plasticity of the maize root system and the efficient uptake of water and nutrients [12]. It has only recently been demonstrated that pericycle founder cells in maize cease their proliferative activity after exiting the meristematic zone and are reactivated in phloem pole pericycle cells that are committed to initiate primordia [40]. The radial positioning of lateral roots of maize and thus the reactivation of these phloem pole pericycle cells is determined by auxin response maxima [41]. The maize mutant *rum1* is defective in lateral root initiation in the primary root [42] (Figure 1). Map-based cloning demonstrated that *rum1* encodes a canonical Aux/IAA protein that is a central regulator of auxin signaling [43]. Aux/IAA proteins regulate the activity of downstream auxin response genes by their interaction with ARF proteins in the nucleus [44]. RUM1 interacts with ARF25 and ARF34 [43] (Figure 2B). It was suggested that at low auxin levels in the pericycle, RUM1 acts as a transcriptional repressor of downstream target genes by interacting with ARF25 and ARF34. By contrast, at high auxin levels RUM1 interacts with the SCF^{TIR1} complex, which results in the degradation of RUM1 by the 26S proteasome. This notion is supported by the very short half-life of RUM1 proteins at high auxin levels [43]. The transcriptional activator LATERAL ROOT PRIMORDIA 1 (LRP1) is repressed by the binding of RUM1 to its promoter [45]. An ancient whole-genome duplication of a maize progenitor resulted in the emergence of the homeologous *rum1* and *rum1-like* (*rul1*) genes [43]. RUL1 and RUM1 exhibit sequence identity of 92% and share the canonical features of Aux/IAA proteins [46]. Although it was demonstrated that RUL1 and RUM1 can homodimerize and heterodimerize *in vivo*, they interact in part with different proteins, suggesting that these homeologs are at least in part interwoven into different molecular networks [46]. Early laser-capture microdissection studies of maize pericycle cells coupled with downstream 12k microarray analyses provided first insights into genes controlling the specification of this cell type [47] and the processes preceding lateral root initiation [42]. A comprehensive RNA-seq analysis of the differentiation zone of wild-type versus mutant *rum1* primary roots revealed that several genes involved in auxin signal transduction are controlled by *rum1* [48]. Based on the downregulation of *arf8* and *arf37*, it was discovered that *rum1* also controls vascular patterning in maize primary roots [48]. The 34 members of the maize Aux/IAA family are subject to root-type- and tissue-specific developmental expression regulation [49] and a plethora of molecular interactions [50]. It is therefore very likely that some of these genes control as-yet-unknown aspects of root formation. Recent cell-type-specific transcriptome analyses highlighted that the root-type-dependent transcriptome divergence of pericycle cells is correlated with exceptional architectural plasticity and functional adaptability during lateral root initiation [51–53].

Seminal Roots

In addition to the single primary root, maize forms a variable number of seminal roots. Seminal roots evolved in the maize lineage quite recently after divergence from the closely related species sorghum, which lacks this type of root [54]. The number of seminal roots in modern maize is highly variable within a genotype and between different genotypes [7]. While the

involved in the Aux/IAA–auxin response factor (ARF) auxin signal transduction module. (C) BIGE1 is a multidrug and toxin intrusion transport protein that controls the number of seminal and shoot-borne roots. (D) RTH3, RTH5, and RTH6 interact during cellulose synthesis-dependent root hair elongation. RTH5 is a NADPH oxidase involved in cell wall loosening, RTH6 is a CSLD5 cellulose synthase controlling cellulose biosynthesis, and RTH3 organizes the newly synthesized cellulose. (E) RTH1 is a SEC3 homolog defining the domain in the plasma membrane where the exocyst complex tethers exocytotic vesicles before their fusion during the exocytotic tip growth of root hairs. Proteins encoded by genes associated with a developmental phenotype upon mutation are highlighted by white font on red background.

emergence of seminal roots in the course of maize evolution remains obscure, an increase of seminal root number has been observed during the domestication of maize from teosinte [55]. It was suggested that beneficial root traits were inadvertently selected for during domestication because of their crucial role in anchorage and soil resource uptake [56]. The increased number of seminal roots might be such a beneficial trait because it was demonstrated that their number correlates with, for instance, drought tolerance at the seedling stage [57]. A recent comparative analysis of the three major root types of maize illustrated a metabolically cost-effective anatomic organization and a unique transcriptomic landscape of seminal roots compared with primary and crown roots [19].

Thus far two maize genes involved in seminal root initiation have been cloned based on their mutant phenotypes. The *rtcs* mutant is defective in embryonic seminal root initiation in addition to the previously described defects in postembryonic shoot-borne root initiation [30] (Figure 1). It was suggested that, during seminal root evolution, the *rtcs* gene has recruited evolutionarily younger (non-syntenic) genes particularly associated with auxin signaling, transcription factor action, and cell proliferation [58]. Non-syntenic genes are maize genes that emerged after the separation of the lineages that led to maize and sorghum [32]. The involvement of non-syntenic genes in the appearance of seminal roots would be intriguing because it was suggested that most non-syntenic genes have little or no biological function [59]. It has been pointed out that the observations of a role for non-syntenic genes in seminal root appearance could also contribute to the molecular interpretation of quantitative trait loci (QTL) [60]: while strong QTLs correspond most frequently with syntenic genes, weaker QTLs could be explained by non-syntenic genes.

Similarly, *rum1* does not initiate embryonic seminal roots in addition to the previously described defect in postembryonic lateral root formation [43] (Figure 1). Remarkably, it was demonstrated in a recent study that these two genes, *rtcs* and *rum1*, underlie the two major QTLs controlling seminal root number in maize [61].

Finally, loss-of-function mutations in the previously mentioned *Bige1* gene, which result in an increased number of shoot-borne roots, cause accelerated seminal root formation and an increased number of seminal roots (Figure 1) accompanied by enlargement of the embryo scutellum [39]. The BIGE1 protein is involved in the extrusion of endogenous compounds and xenobiotics from the cytosol into the Golgi lumen by feedback regulation of a CYP78A pathway (Figure 2C). This might suggest that mobile signals could play a role in embryonic seminal root initiation.

Root Hairs

Root hairs are tubular extensions of trichoblast cells of the epidermis. They are instrumental for nutrient uptake and optimal plant development [62]. Three types of epidermal patterning have been observed in plant roots [63]. In maize, root hair patterning is random (type I) and it cannot be predicted which cells of the root epidermis will form root hairs. In rice, root hairs develop from the smaller cell of an asymmetric division (type II). Finally, in arabidopsis (type III), only epidermis cells situated in a cleft between two underlying cortical cells (H position) give rise to root hairs [64]. The diversification of genes involved in root hair development among different species of vascular plants including maize has been studied recently and suggested a substantial divergence in the structure and expression of genes involved in root hair patterning in different species [65].

Thus far, four maize genes controlling root hair elongation (Figure 2D,E) have been cloned based on their mutant phenotypes (Figure 1). These genes were designated *roothairless* (*rth*). The proteins encoded by *rth3*, *rth5*, and *rth6* are functionally linked during maize root hair

formation (Figure 2D). RTH5 produces apoplastic superoxide [66], which results in cell wall loosening via hydroxyl radicals and also controls the expression of cellulose biosynthesis genes. After cell wall loosening, the transmembrane cell wall protein RTH6 [67], which corresponds to cellulose synthase-like D5 (CSLD5) [68], is arranged in rosettes and synthesizes cellulose at the plasma membrane that is extruded to the inner side of the cell wall in the root hair tip and thus reinforces the tubular shaft. Finally, the GPI-anchored COBRA-like cell wall protein RTH3 [69] is involved in the organization of the synthesized cellulose.

By contrast, the *rth1* gene encodes the SEC3 subunit of the exocyst complex, which controls exocytotic growth of the root hair tip [70]. It has been demonstrated that SEC3 defines the domain in the plasma membrane where the exocyst complex tethers exocytotic vesicles before their fusion [71] (Figure 2E).

A recent transcriptome analysis highlighted genes expressed exclusively in root hair cells compared with other non-root-hair cell types of maize [72], providing a valuable resource for further reverse genetic analyses. This publication also implemented a 'maize root view' of the maize eFP browser (http://bar.utoronto.ca/efp_maize/cgi-bin/efpWeb.cgi?dataSource=Maize_Root), which allows the study of the expression profiles of a diverse set of maize root transcriptomes and complements another recently published maize gene expression atlas to explore root development [17].

Concluding Remarks and Outlook

In recent years genetic analyses have identified key genes instrumental for the intrinsic regulation and function of the various root types of maize and their root hairs. Genomic analyses have contributed to our understanding of the molecular networks in which these genes are embedded. Nevertheless, a plethora of questions related to maize root development and function remain unanswered and need to be addressed by future research (see Outstanding Questions).

To exploit limited soil resources, the maize root system readjusts its 3D structure and its physiological, biochemical, and metabolic performance to changing environmental conditions in the course of development [35,73–76]. It is largely unclear how intrinsic regulators of root structure and function perceive environmental cues. In turn, it remains to be elucidated which pathways regulate the specific alterations of root system architecture mediated by different abiotic environmental signals such as drought, salinity, or nutrient deficiency [16,77–80]. In this context it will be interesting to understand how the different root types of maize communicate with each other to coordinate and integrate these abiotic processes.

Furthermore, it remains unclear which genes control the positioning and specification of certain cell types, such as phloem pole pericycle cells or trichoblasts giving rise to lateral roots and root hairs, respectively. Laser microdissection, which allows the isolation of specific cell types from complex tissues coupled with next-generation sequencing approaches, will be instrumental for the study of individual cell types.

To date molecular analysis of maize root development has mainly focused on the determination of genes controlling qualitative traits. In the future it will be feasible with the help of modern phenotyping technologies in combination with genome-wide association techniques to identify genes underlying subtle quantitative differences in root system architecture as reflected by natural variation [81–85].

Finally, the knowledge obtained in maize and other model cereals can be translated to the understanding and improvement of cereal species with less agronomic relevance that have

Outstanding Questions

Which intrinsic regulators of maize root structure and function are modulated by environmental cues and which pathways regulate in turn the specific alterations of root system architecture mediated by different environmental signals?

How do different root types communicate with each other and coordinate the overall architecture of the root system?

Which genes control quantitative differences in root system architecture as reflected by natural variation?

Which genes are involved in the positioning and specification of phloem pole pericycle cells and determine that not all of these cell populations initiate lateral roots?

Which genes regulate the unpredictable specification of trichoblasts in the epidermis and the subsequent initiation of root hairs?

How conserved are the molecular processes determining maize root development in other cereals?

gained less attention thus far. Such studies will also contribute to our understanding of conserved and diverse mechanisms of cereal root development.

Acknowledgments

Root research in F.H.'s laboratory is supported by the German Research Foundation (DFG) and the Federal Ministry of Education and Research (BMBF).

References

1. Hake, S. and Ross-Ibarra, J. (2015) Genetic, evolutionary and plant breeding insights from the domestication of maize. *Elife* 4, e05861
2. Gewin, V. (2010) An underground revolution. *Nature* 466, 552–553
3. Lynch, J.P. (2013) Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems. *Ann. Bot.* 112, 347–357
4. Meister, R. *et al.* (2014) Challenges of modifying root traits in crops for agriculture. *Trends Plant Sci.* 19, 779–788
5. Rogers, E.D. and Benfey, P.N. (2015) Regulation of plant root system architecture: implications for crop advancement. *Curr. Opin. Biotechnol.* 32, 93–98
6. Hochholdinger, F. (2016) Untapping root system architecture for crop improvement. *J. Exp. Bot.* 67, 4431–4433
7. Hochholdinger, F. *et al.* (2004) From weeds to crops: genetic analysis of root development in cereals. *Trends Plant Sci.* 9, 42–48
8. Hochholdinger, F. (2009) The maize root system: morphology, anatomy and genetics. In *The Handbook of Maize* (Bennetzen, J. and Hake, S., eds), pp. 145–160, Springer
9. Atkinson, J.A. *et al.* (2014) Branching out in roots: uncovering form, function, and regulation. *Plant Physiol.* 166, 538–550
10. Sanguineti, M. *et al.* (1998) Root and shoot traits of maize inbred lines grown in the field and in hydroponic culture and their relationships with root lodging. *Maydica* 43, 211–216
11. Hochholdinger, F. *et al.* (2004) Genetic dissection of root formation in maize (*Zea mays*) reveals root-type specific developmental programs. *Ann. Bot.* 93, 359–368
12. Yu, P. *et al.* (2016) Genetic control of lateral root formation in cereals. *Trends Plant Sci.* 21, 951–961
13. Ishikawa, H. and Evans, M.L. (1995) Specialized zones of development in roots. *Plant Physiol.* 109, 725–727
14. Marzec, M. *et al.* (2015) Root hair development in the grasses: what we already know and what we still need to know. *Plant Physiol.* 168, 407–414
15. Paschold, A. *et al.* (2014) Non-syntenic genes drive highly dynamic complementation of gene expression in maize hybrids. *Plant Cell* 26, 3939–3948
16. Opitz, N. *et al.* (2016) Extensive tissue-specific transcriptomic plasticity in maize primary roots upon water deficit. *J. Exp. Bot.* 67, 1095–1107
17. Stelpflug, S.C. *et al.* (2016) An expanded maize gene expression atlas based on RNA sequencing and its use to explore root development. *Plant Genome* 9, 1
18. Marcon, C. *et al.* (2015) A high resolution tissue-specific proteome and phosphoproteome atlas of maize primary roots reveals functional gradients along the root axis. *Plant Physiol.* 168, 233–246
19. Tai, H. *et al.* (2016) Transcriptomic and anatomic complexity of primary, seminal and crown roots highlight root type-specific functional diversity in maize (*Zea mays* L.). *J. Exp. Bot.* 67, 1123–1135
20. Hoppe, D.C. *et al.* (1986) The nodal roots of *Zea*: their development in relation to structural features of the stem. *Can. J. Bot.* 64, 2524–2537
21. Yu, P. *et al.* (2015) Root-type-specific plasticity in response to localized high nitrate supply in maize (*Zea mays*). *Ann. Bot.* 116, 751–762
22. Correns, C. (1901) Bastarde zwischen Maisrassen, mit besonderer Berücksichtigung der Xenien. *Bibl. Bot.* 53, 1–161 (in German)
23. Emerson, R.A. *et al.* (1935) A summary of linkage studies in maize. *Cornell Univ. Agric. Exp. Sta. Mem.* 180, 1–83
24. Jenkins, M.T. (1930) Heritable characters of maize. XXXIV – rootless. *J. Hered.* 21, 79–80
25. Wen, T.J. and Schnable, P.S. (1994) Analyses of mutants of three genes that influence root hair development in *Zea mays* (Gramineae) suggest that root hairs are dispensable. *Am. J. Bot.* 81, 833–842
26. Hochholdinger, F. and Feix, G. (1998) Early post-embryonic development is specifically affected in the maize mutant *lrt1*. *Plant J.* 16, 247–255
27. Hetz, W. *et al.* (1996) Isolation and characterisation of *rtcs*, a mutant deficient in the formation of nodal roots. *Plant J.* 10, 845–857
28. Hochholdinger, F. and Feix, G. (1998) Cyclin expression is completely suppressed at the site of crown root formation in the nodal region of the maize root mutant *rtcs*. *J. Plant Physiol.* 153, 425–429
29. Hochholdinger, F. and Feix, G. (1998) Tiller formation in *gaspe flint* is not affected by the *rtcs* locus. *Maize Genet. Coop. Newsl.* 72, 30–31
30. Taramino, G. *et al.* (2007) The *rtcs* gene in maize (*Zea mays* L.) encodes a LOB domain protein that is required for postembryonic shoot-borne and embryonic seminal root initiation. *Plant J.* 50, 649–659
31. Xu, C. *et al.* (2016) LOB domain proteins: beyond lateral organ boundaries. *Trends Plant Sci.* 21, 159–167
32. Schnable, J.C. *et al.* (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci. U. S. A.* 108, 4069–4074
33. Xu, C. *et al.* (2015) Cooperative action of the paralogous maize LOB domain proteins RTCS and RTCL in shoot-borne root formation. *New Phytol.* 207, 1123–1133
34. Majer, C. *et al.* (2012) Molecular interactions of RTCS, a LOB domain protein regulating shoot-borne root initiation in maize (*Zea mays* L.). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1542–1551
35. Gao, Y. and Lynch, J.P. (2016) Reduced crown root number improves water acquisition under water deficit stress in maize (*Zea mays* L.). *J. Exp. Bot.* 67, 4545–4557
36. Sebastian, J.C. *et al.* (2016) Grasses suppress shoot-borne roots to conserve water during drought. *Proc. Natl. Acad. Sci. U. S. A.* 113, 8861–8866
37. Liu, H. *et al.* (2005) ARL1, a LOB-domain protein required for adventitious root formation in rice. *Plant J.* 43, 47–56
38. Inukai, Y. *et al.* (2005) *crown rootless1*, which is essential for crown root formation in rice, is a target of an AUXIN RESPONSE FACTOR in auxin signaling. *Plant Cell* 17, 1387–1396
39. Suzuki, M. *et al.* (2015) Conserved functions of the MATE transporter BIG EMBRYO1 in regulation of lateral organ size and initiation rate. *Plant Cell* 27, 2288–2300
40. Alarcón, M.V. *et al.* (2016) The initiation of lateral roots in the primary roots of maize (*Zea mays* L.) implies a reactivation of cell proliferation in a group of founder pericycle cells. *J. Plant Physiol.* 192, 105–110
41. Jansen, L. *et al.* (2012) Phloem-associated auxin response maxima determine radial positioning of lateral roots in maize. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1525–1533
42. Woll, K. *et al.* (2005) Isolation, characterization and pericycle specific transcriptome analyses of the novel maize (*Zea mays*

L.) lateral and seminal root initiation mutant *rum1*. *Plant Physiol.* 139, 1255–1267

43. von Behrens, I. *et al.* (2011) *rootless with undetectable meristem1* encodes a monocot-specific AUX/IAA protein that controls embryonic seminal and postembryonic lateral root initiation in maize. *Plant J.* 66, 341–353
44. Ulmasov, T. *et al.* (1997) Aux/IAA proteins repress expression of reporter genes containing natural and highly active synthetic auxin response elements. *Plant Cell* 9, 1963–1971
45. Zhang, Y. *et al.* (2015) LATERAL ROOT PRIMORDIA 1 of maize acts as a transcriptional activator in auxin signaling downstream of the Aux/IAA gene *rootless with undetectable meristem1*. *J. Exp. Bot.* 66, 3855–3863
46. Zhang, Y. *et al.* (2016) Conserved and unique features of the homeologous maize Aux/IAA proteins ROOTLESS WITH UNDETECTABLE MERISTEM 1 and RUM1-like 1. *J. Exp. Bot.* 67, 1137–1147
47. Dembinsky, D. *et al.* (2007) Transcriptomic and proteomic analyses of pericycle cells of the maize (*Zea mays* L.) primary root. *Plant Physiol.* 145, 575–588
48. Zhang, Y. *et al.* (2014) The Aux/IAA gene *rum1* involved in seminal and lateral root formation controls vascular patterning in maize (*Zea mays* L.) primary roots. *J. Exp. Bot.* 65, 4919–4930
49. Ludwig, Y. *et al.* (2013) The maize (*Zea mays* L.) AUXIN/INDOLE-3-ACETIC ACID gene family: phylogeny, synteny, and unique root-type and tissue-specific expression patterns during development. *PLoS One* 8, e78859
50. Ludwig, Y. *et al.* (2014) Diversity of stability, localization, interaction and control of downstream gene activity in the maize Aux/IAA protein family. *PLoS One* 9, e107346
51. Jansen, L. *et al.* (2013) Comparative transcriptomics as a tool for the identification of root branching genes in maize. *Plant Biotechnol. J.* 11, 1092–1102
52. Yu, P. *et al.* (2015) Cell-type specific gene expression analyses by RNA-seq reveal local high nitrate triggered lateral root initiation in shoot-borne roots of maize by modulating auxin-related cell cycle-regulation. *Plant Physiol.* 169, 690–704
53. Yu, P. *et al.* (2016) Root type specific reprogramming of maize pericycle transcriptomes by local high nitrate results in disparate lateral root branching patterns. *Plant Physiol.* 170, 1783–1798
54. Singh, V. *et al.* (2010) Morphological and architectural development of root systems in sorghum and maize. *Plant Soil* 333, 287–299
55. Burton, A.L. *et al.* (2013) Phenotypic diversity of root anatomical and architectural traits in Zea species. *Crop Sci.* 53, 1042–1055
56. Lynch, J.P. and Brown, K. (2006) Whole plant adaptations to low phosphorus availability. In *Plant–Environment Interactions* (Huang, B., ed.), pp. 209–242, CRC Press
57. Qayyum, A. *et al.* (2012) Screening for drought tolerance in maize (*Zea mays* L.) hybrids at an early seedling stage. *Afr. J. Agric. Res.* 7, 3594–3604
58. Tai, H. *et al.* (2017) Non-syntenic genes drive RTCS-dependent regulation of the embryo transcriptome during formation of seminal root primordia in maize (*Zea mays* L.). *J. Exp. Bot.* 67, 1137–1147
59. Schnable, J.C. and Freeling, M. (2011) Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. *PLoS One* 6, e17855
60. Salvi, S. (2017) An evo–devo perspective on root genetic variation in cereals. *J. Exp. Bot.* 68, 351–354
61. Salvi, S. *et al.* (2016) Two major quantitative trait loci controlling the number of seminal roots in maize co-map with the root developmental genes *rtcs* and *rum1*. *J. Exp. Bot.* 67, 1149–1159
62. Gilroy, S. and Jones, D.L. (2000) Through form to function: root hair development and nutrient uptake. *Trends Plant Sci.* 5, 56–60
63. Dolan, L. (1996) Pattern in the root epidermis: an interplay of diffusible signals and cellular geometry. *Ann. Bot.* 77, 547–553
64. Salazar-Henao, J.E. *et al.* (2016) The regulation and plasticity of root hair patterning and morphogenesis. *Development* 143, 1848–1858
65. Huang, L. *et al.* (2017) Diversification of root hair development genes in vascular plants. *Plant Physiol.* 174, 1697–1712
66. Nestler, J. *et al.* (2014) *roothairless5*, which functions in maize (*Zea mays* L.) root hair initiation and elongation encodes a monocot-specific NADPH oxidase. *Plant J.* 79, 729–740
67. Li, L. *et al.* (2016) Characterization of maize *roothairless6* which encodes a D-type cellulose synthase and controls the switch from bulge formation to tip growth. *Sci. Rep.* 6, 34395
68. Penning, B.W. *et al.* (2009) Genetic resources for maize cell wall biology. *Plant Physiol.* 151, 1703–1728
69. Hochholdinger, F. *et al.* (2008) The maize (*Zea mays* L.) *roothairless3* gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield. *Plant J.* 54, 888–898
70. Wen, T.J. *et al.* (2005) The *roothairless1* gene of maize (*Zea mays*) encodes a homolog of *sec3*, which is involved in polar exocytosis. *Plant Physiol.* 138, 1637–1643
71. Hala, M. *et al.* (2008) An exocyst complex functions in plant cell growth. *Plant Cell* 20, 1330–1345
72. Hey, S. *et al.* (2017) Complexity and specificity of the maize (*Zea mays* L.) root hair transcriptome. *J. Exp. Bot.* 68, 2175–2185
73. Zhan, A. *et al.* (2015) Reduced lateral root branching density improves drought tolerance in maize. *Plant Physiol.* 168, 1603–1615
74. Zhan, A. and Lynch, J.P. (2015) Reduced frequency of lateral root branching improves N capture from low-N soils in maize. *J. Exp. Bot.* 66, 2055–2065
75. Chimungu, J.G. *et al.* (2014) Large root cortical cell size improves drought tolerance in maize. *Plant Physiol.* 166, 2166–2178
76. Chimungu, J.G. *et al.* (2014) Reduced root cortical cell file number improves drought tolerance in maize. *Plant Physiol.* 166, 1943–1955
77. Opitz, N. *et al.* (2014) Transcriptomic complexity in young maize primary roots in response to low water potentials. *BMC Genomics* 15, 741
78. Yu, P. *et al.* (2014) A novel morphological response of maize (*Zea mays*) adult roots to heterogeneous nitrate supply revealed by a split-root experiment. *Physiol. Plant.* 150, 133–144
79. Yu, P. *et al.* (2014) Phenotypic plasticity of the maize root system in response to heterogeneous nitrogen availability. *Planta* 240, 667–678
80. Zhang, M. *et al.* (2015) Comparative transcriptome profiling of the maize primary, crown and seminal root in response to salinity stress. *PLoS One* 10, e0121222
81. Iyer-Pascuzzi, A.S. *et al.* (2010) Imaging and analysis platform for automatic phenotyping and trait ranking of plant root systems. *Plant Physiol.* 152, 1148–1157
82. Fiorano, F. and Schurr, U. (2013) Future scenarios for plant phenotyping. *Annu. Rev. Plant Biol.* 64, 267–291
83. Le Marié, C. *et al.* (2014) Rhizoslides: paper-based growth system for non-destructive, high throughput phenotyping of root development by means of image analysis. *Plant Methods* 10, 13
84. Zurek, P.R. *et al.* (2015) Quantitative trait locus mapping reveals regions of the maize genome controlling root system architecture. *Plant Physiol.* 167, 1487–1496
85. Downie, H.F. *et al.* (2015) Challenges and opportunities for quantifying roots and rhizosphere interactions through imaging and image analysis. *Plant Cell Environ.* 38, 1213–1232

Transcriptomic Dissection of Maize Root System Development

15

Peng Yu, Caroline Marcon, Jutta A. Baldauf, Felix Frey, Marcel Baer and Frank Hochholdinger

Abstract

RNA-Seq (RNA-Sequencing) allows for precise quantitative determination of global gene expression patterns and has therefore revolutionized transcriptome analyses in maize. In recent years, genetic analyses have identified numerous genes that control maize root system architecture and root hair elongation. In addition, RNA-Seq has been applied to dissect structure and function of individual roots. In this chapter, we summarize the current state of the transcriptomic dissection of maize root development on the level of whole roots, tissues, and individual cells. Moreover, we highlight the current knowledge of transcriptome responses of maize roots to drought stress and nutrient availability. Finally, we outline novel findings related to gene expression plasticity in primary roots of maize hybrids during the early manifestation of heterosis.

P. Yu · C. Marcon · J. A. Baldauf · F. Frey
M. Baer · F. Hochholdinger (✉)
INRES, Institute of Crop Science and Resource Conservation, Crop Functional Genomics, University of Bonn, 53113 Bonn, Germany
e-mail: hochholdinger@uni-bonn.de
URL: https://www.cfg.uni-bonn.de

15.1 Introduction

Genome sequencing of the maize inbred line B73 was a major landmark in maize genetics (Schnable et al. 2009). The recent *de novo* assembly and annotation of the maize B73 genome sequence led to the identification of 39,324 protein-coding genes (https://www.maizegdb.org/assembly, v4). Nevertheless, the function of most of these genes remains elusive (Jiao et al. 2017). Functional genomic tools such as RNA-Seq (RNA-Sequencing) are critical to explore the transcriptomic landscape during plant development, tissue patterning, and cell organization. Maize displays a complex root system architecture (Fig. 15.1) as illustrated by the presence of different embryonic and post-embryonic root types (reviewed in Yu et al. 2016; Hochholdinger et al. 2018a). This chapter summarizes the current progress in understanding the transcriptomic landscape controlling the function of the maize root system and its responses to environmental stimuli. Moreover, we highlight the transcriptomic plasticity of maize hybrids compared to their parental inbred lines during the early manifestation of heterosis. A review summarizing the status of the proteomic dissection of the maize root system has been published recently (Hochholdinger et al. 2018b).

© Springer Nature Switzerland AG 2018

J. Bennetzen et al. (eds.), *The Maize Genome*, Compendium of Plant Genomes,
https://doi.org/10.1007/978-3-319-97427-9_15

15.2 Transcriptomic Dissection of Different Root Types, Tissues, and Cell Types

15.2.1 Root Type-Specific Transcriptomics

The rootstock of maize consists of an embryonically formed primary root and a variable number of embryonic seminal roots and post-embryonic shoot-borne crown and brace roots (Fig. 15.1; reviewed in Hochholdinger et al. 2004a). In the first weeks after germination, primary and seminal roots make up the major portion of the seedling rootstock (Hochholdinger et al. 2004b) and are thus vital for the early vigor of young maize seedlings (Peter et al. 2009). Shoot-borne roots formed from consecutive shoot nodes below the soil level are designated crown roots, while their counterparts initiated from

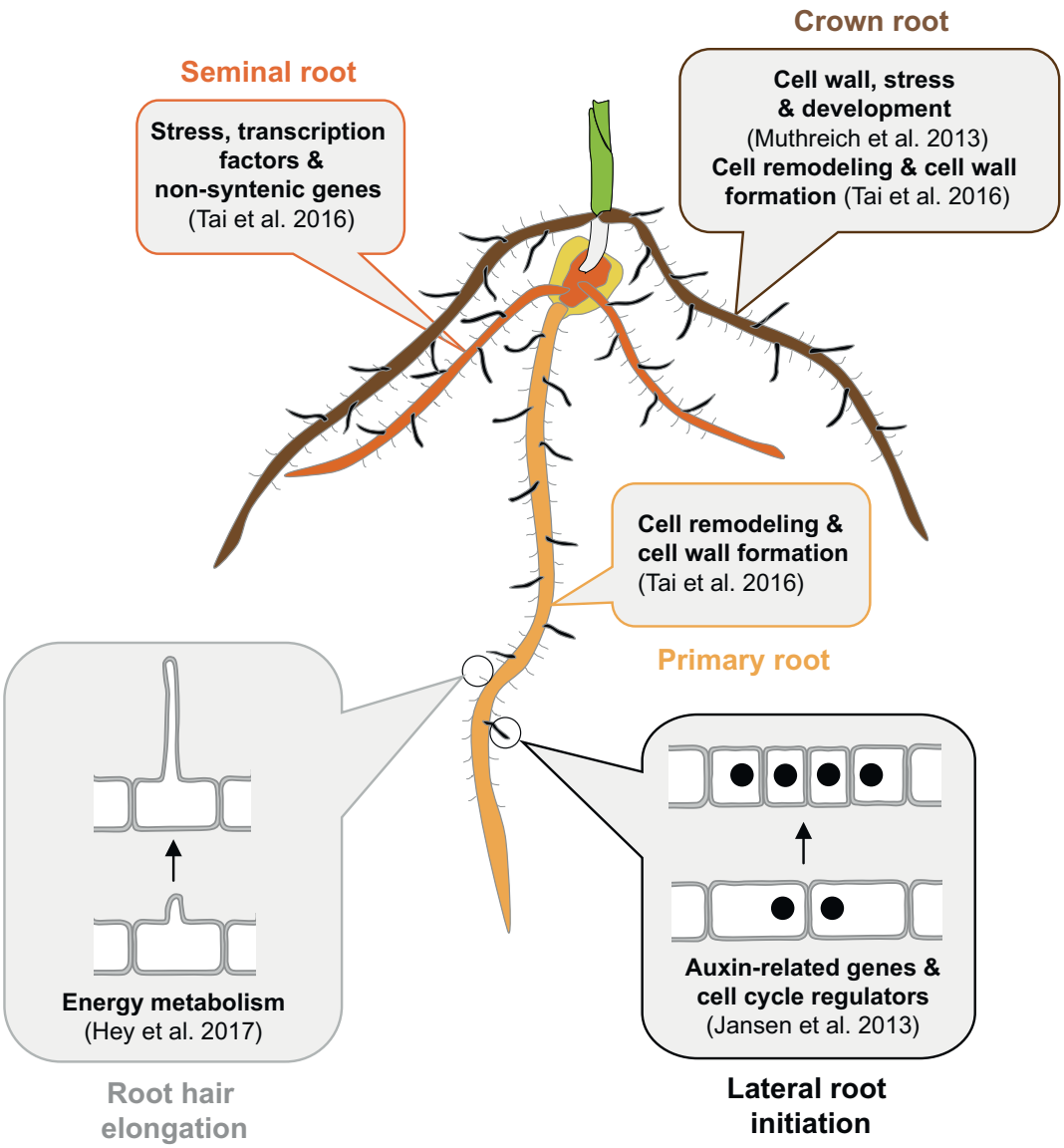


Fig. 15.1 Summary of transcriptome studies related to individual root types of maize

aboveground shoot nodes are brace roots. Shoot-borne roots take up most nutrients and are therefore critical for plant yield (Hochholdinger et al. 2004a). Genetic analyses demonstrated that root type-specific developmental programs control seminal, crown and lateral root formation in maize (Hochholdinger et al. 2004a).

The maize mutant *rtcs (rootless concerning crown and seminal roots)* does not initiate any shoot-borne roots (Hetz et al. 1996). A comparative microarray analysis of the first shoot-node of wild type and *rtcs* seedlings during three early stages of crown root initiation identified >800 transcripts that displayed RTCS-dependent expression. This suggests a direct or indirect role of these genes during shoot-borne root initiation (Muthreich et al. 2013). Functional classes of genes overrepresented among RTCS-dependent genes refer to cell wall, stress, and development-related processes (Muthreich et al. 2013).

A RNA-Seq study of primary, seminal and crown roots highlighted diverse transcriptomic patterns of these root types (Tai et al. 2016). Remarkably, embryonic primary and post-embryonic crown roots displayed similar anatomical features and transcriptomic profiles, while the transcriptome of seminal roots was distinct from these root types (Tai et al. 2016). In primary and crown roots, functions related to cell remodeling and cell wall formation were prominent. In contrast, in seminal roots stress-related genes and transcriptional regulators were overrepresented, which suggests a functional specialization of the different root types (Tai et al. 2016). Seminal roots also displayed unique anatomical features, which indicate a higher absorption efficiency in comparison with the other root types (Burton et al. 2013; Tai et al. 2016). In another RNA-Seq analysis of wild type and mutant *rtcs* embryos, which do not initiate seminal roots, it was demonstrated that evolutionary young non-syntenic genes were overrepresented among genes displaying RTCS-dependent expression during seminal root primordia formation (Tai et al. 2017). Non-syntenic genes evolved after the separation of the lineages leading to maize and sorghum (Schnable et al. 2011). It was suggested that these differentially expressed non-syntenic genes might have come under the transcriptional control of the syntenic gene *rtcs* during seminal root evolution (Tai et al. 2017).

A commonality of all embryonic and post-embryonic root types is their ability to initiate lateral roots and root hairs from pericycle and epidermis cells, respectively (Yu et al. 2016; Hochholdinger et al. 2018a). Continuous production of lateral roots and root hairs from the main root types substantially increases the root surface and thus facilitates the capturing of nutrient and water in maize (Rogers and Benfey 2015; Hochholdinger et al. 2018a). The maize mutant *lrt1 (lateralrootless 1)* is specifically affected in the formation of lateral roots on the embryonic primary root and seminal roots but not on the post-embryonic shoot-borne roots (Hochholdinger and Feix 1998). Similarly, the mutant *rum1 (rootless with undetectable meristem 1)* does not initiate lateral roots on the primary root but displays normal lateral root formation on crown and brace roots (Woll et al. 2005). These mutants demonstrate that root type-specific developmental programs control lateral root initiation in maize. The differential regulation of lateral root initiation in primary and crown roots of maize might be explained by members of root type-specific transcription factor families such as MYB (V-MYB AVIAN MYELOBLASTOSIS VIRAL ONCOGENE HOMOLOG), MYB-related, HOMEOBOX, and bHLH (basic helix-loop-helix) (Jansen et al. 2013).

15.2.2 Tissue Type-Specific Transcriptomics

Maize roots display highly differentiated tissue patterns along the longitudinal and radial axes. In longitudinal orientation, roots are structurally divided into the root cap at the terminal end, a subterminal meristematic zone, followed by the elongation and differentiation zone (Fig. 15.2; Ishikawa and Evans 1995; Hochholdinger et al. 2004a). These zones are functionally diversified

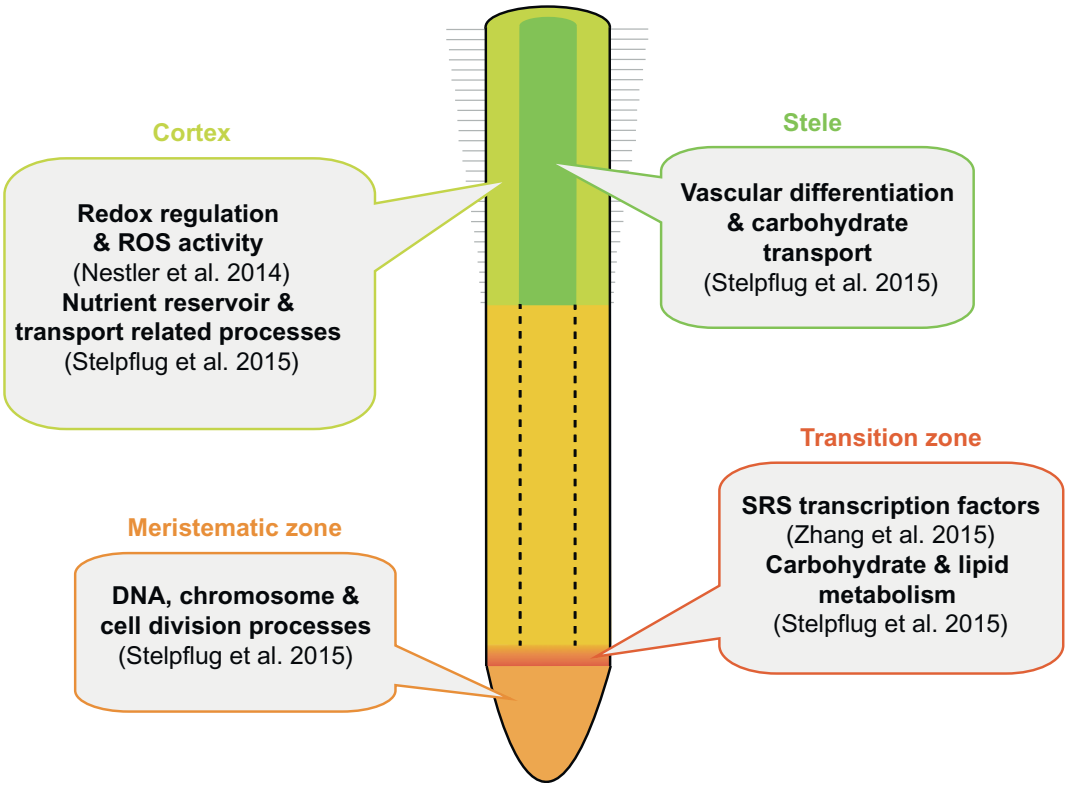


Fig. 15.2 Summary of transcriptome studies related to tissue patterning of the maize primary root. ROS, reactive oxygen species

by mitotic divisions in the meristematic zone, cell extension, and elongation in the elongation zone and highly differentiated cells with root hairs in the differentiation zone (Ishikawa and Evans 1995). In transverse orientation, the differentiation zone displays a number of functionally distinct cell types (Hochholdinger 2009; Yu et al. 2016). The stele, in the center of the root, contains differentiated xylem vessels and primary phloem elements. Xylem vessels are involved in water and nutrient transport, while primary phloem elements transport photosynthates. Both, xylem and phloem elements are embedded in the ground tissue of the central cylinder. The pericycle represents the outermost cell layer of the central cylinder. Phloem pole pericycle cells in the differentiation zone can divide and give rise to lateral roots. The pericycle is connected via a single layer of endodermis cells to the multilayered parenchyma. The outermost cell layer of the root is the epidermis that connects the root to the rhizosphere. The stele and the cell layers between endodermis and epidermis, hereafter referred to as cortical parenchyma, can be mechanically separated at the boundary between pericycle and endodermis without damaging these cell files, thus allowing subsequent analyses of these functionally diverse tissues (Saleem et al. 2009; Paschold et al. 2014; Yu et al. 2015). The functionally diverse cell types described above can be distinguished based on their position within the three-dimensional context of the root but also according to their anatomical characteristics and their staining ability by certain histological dyes (Hochholdinger 2009).

Maize root tissues are highly diversified in their biological functions along the longitudinal and transverse axes. A snapshot of the dynamics of maize root tissues was generated by co-expression analyses of 17 transcriptome

profiles generated from embryonic and post-embryonic root tissues (Stelpflug et al. 2015). Genes encoding enzymes associated with "translation", "ribosomal function and assembly", "protein metabolism", "DNA synthesis and replication", "transcriptional activation", "cell cycle regulation", "microtubule motor activity", "nucleosome assembly", and "cell wall organization" were highly enriched in the root tip. This is consistent with functions required by the rapidly dividing cells in the meristematic zone. Genes showing peak expression in the transition zone of the root, which is located between the meristematic and the elongation zone (Fig. 15.2), were enriched in "carbohydrate and lipid metabolic processes", "response to oxidative stress", "peroxidases", "lignin catabolism", and "cell wall organization" (Stelpflug et al. 2015). Moreover, SRS (SHI RELATED SEQUENCE) transcription factors , involved in auxin-mediated lateral root primordia initiation in maize were overrepresented in this region (Zhang et al. 2014). The differentiation zone featured predominant expression of genes encoding nutrient reservoir activity, transport, kinases, protein phosphorylation, regulation of transcription, and transcription factor activity (including enrichment for TIFY (TIFY DOMAIN), MYB, NAC (NAM, ATAF, and CUC) and WRKY (WRKY DOMAIN) families, monooxygenase activity, glutathione transferases, redox regulation, electron carrier activity, lipid metabolism, and biosynthesis of flavonoids (Stelpflug et al. 2015). Redox regulation and ROS (reactive oxygen species) activity are crucial for root hair initiation and elongation in maize, which occur in this zone (Nestler et al. 2014). In the cortical parenchyma, 61 GO (gene ontology) categories were enriched highlighting the prevalent biological functions in this complex tissue. In contrast, only six GO terms were enriched in the stele indicating the specialization of this tissue to processes such as vascular differentiation and carbohydrate transport (Stelpflug et al. 2015). Overall, GO enrichment analyses showed disparate patterns of gene expression between cortical parenchyma and stele, which is consistent with distinct biological functions of these tissues.

15.2.3 Cell Type-Specific Transcriptomics

Root hairs are tubular extensions of epidermis cells. Their easy accessibility made them a model for single cell-type research in plants. In maize, transcriptome profiling demonstrated that the root hair transcriptome is less complex than the transcriptome of primary roots without root hairs, which consists of multiple cell types (Hey et al. 2017). A substantial number of GO terms enriched in the transcriptome of root hairs are functionally related to "energy metabolism", which highlights the high-energy demand for the development of these cells and the maintenance of their function (Hey et al. 2017). LCM (laser capture microdissection) provides the opportunity to isolate specific cell types from inside complex tissues (Nakazono et al. 2003). RNA isolated from individual cells can be linearly amplified and subsequently subjected to downstream transcriptome analyses (Ludwig and Hochholdinger 2014). Early LCM studies in combination with microarray hybridization experiments revealed genes involved in the specification of pericycle cells (Dembinsky et al. 2007) and genes regulating the stages shortly before lateral root initiation (Woll et al. 2005). More recently, LCM-based pericycle-specific RNA-Seq analyses indicated that auxin-related genes and cell cycle regulators play a crucial role during lateral root initiation in maize (Yu et al. 2015). Notably, pericycle cells isolated from brace roots showed a unique transcriptomic landscape compared to other seedling root types, which is consistent with the distinct lateral root density in these root types (Yu et al. 2016).

15.3 Transcriptomic Responses to Different Supply of Mineral Nutrients in Maize Roots

The architectural modulation of the root system of modern maize in comparison to its teosinte counterpart can be explained in part by its adaptation to different local environmental conditions during domestication. Root morphology

Table 15.1 Summary of transcriptome studies related to heterosis, drought, hydropatterning, and nitrate. LBD, lateral organ boundary domain; NO, Nitric oxide; ROS, reactive oxygen species; SPE, single parent expression; PIN, pin-formed

Studied root type or zone	Factors	Enrichment of	References
Primary root	Heterosis	SPE and non-syntenic genes	Paschold et al. (2012, 2014), Baldauf et al. (2016, 2018), Marcon et al. (2017)
	Drought	Transcriptional regulation, ROS, and hormone metabolism	Opitz et al. (2014, 2016), Marcon et al. (2017)
Lateral root	Hydropatterning	Efflux and biosynthesis of auxin, LBD	Bao et al. (2014), Robbins and Dinneny (2018)
	Local nitrate	Auxin transport and PIN9	Liu et al. (2010), Yu et al. (2015)
Meristematic zone	Drought	Cell wall reorganization	Opitz et al. (2016)
Transition zone	Nitrate	Molecular NO signal	Trevisan et al. (2011), Manoli et al. (2014)

and architecture determine the potential to exploit unevenly distributed nutrient patches in the soil to maximize nutrient-use efficiency and yield (Yu et al. 2014). Physiological experiments suggest that maize plants optimize their root architecture by regulating lateral root formation based on the availability of soil nitrate. Sparsely spaced and long lateral roots are optimal for nitrate acquisition in maize (Lynch 2011, 2013). Recent transcriptome studies highlighted diverse transcriptomic changes in response to nitrogen-deficiency and nitrogen-enrichment conditions (Table 15.1).

15.3.1 Nitrate Deficiency Triggered Transcriptome Changes of Maize Root

Systemic nitrate deficiency in maize seedlings increases carbon partitioning to roots, accelerates root growth, and results in fewer, longer main roots with longer lateral roots (Gaudin et al. 2011; Trachsel et al. 2013). Recent RNA-Seq analyses revealed that several AP2-EREBP (APETALA2/ETHYLENE-RESPONSIVE ELEMENT BINDING PROTEIN) family members play important roles in the *rtcs*-dependent regulatory network related to both root development and nitrate-deficiency response in maize (He et al. 2015). Moreover, complex transcriptomic reprogramming occurs as an early response to nitrate supply and molecular NO (nitric oxide) in the transition zone between meristematic and elongation zone (Trevisan et al. 2011; Manoli et al. 2014). These results emphasize the role of the transition zone of maize roots in sensing and transducing nitrate signals (Trevisan et al. 2011, 2015).

15.3.2 Local Nitrate Enrichment Triggered Transcriptome Changes of Maize Roots

In resource-depleted environments, heterogeneously distributed nutrient availability directs lateral root growth preferentially into nutrient-rich patches (Forde 2014; Giehl and von Wirén 2014). The application of auxin and auxin transport inhibitors revealed the pivotal role of auxin shoot-to-root transport in lateral root growth in response to localized nitrate supply (Wang et al. 2005; Tian et al. 2008; Guo et al. 2005; Liu et al. 2010). Microarray analyses indicated that early responsive genes related to cell division and expansion, such as α-expansin, cellulose synthase, kinesin, plasma membrane and tonoplast aquaporins, are possibly involved in localized nitrate stimulation of lateral root development in maize (Wang et al. 2005; Liu

et al. 2010). Among the different root types of maize, brace roots displayed an exceptional plasticity for lateral root formation compared to other root types upon different nitrate concentrations. This is illustrated by an increased induction of lateral roots in brace roots by nitrate compared to other root types and is reflected by their unique transcriptional control of cell cycle and MYB-related genes (Yu et al. 2014, 2015, 2016). Notably, activity of the monocot-specific PIN9 (*PIN-FORMED 9*) gene in phloem pole cells modulates auxin efflux to pericycle cells and subsequent cell cycle activation (Yu et al. 2015). Taken together, these recent discoveries link transcriptional regulation and nutrient-triggered lateral root development in maize.

15.4 Transcriptomic Responses to Drought Stress in Maize Roots

In maize, a deeper root system enhances drought tolerance (Ribaut et al. 2009), while shallower seminal and crown roots increase phosphorus acquisition (Zhu et al. 2005). Recent studies indicated that seminal and crown roots are major targets for drought stress signaling in maize (Gao and Lynch 2016; Sebastian et al. 2016). To study the transcriptomic response to intensity and duration of drought stress simulated by PEG (polyethylene glycol) 8000 treatment, maize primary roots were subjected to mild (−0.2 MPa) and severe (−0.8 MPa) water deficit conditions for 6 h and 24 h (Opitz et al. 2014). In general, the number of drought-responsive genes increased with intensity and duration of water deficit. In this study, a set of 53 genes were drought-responsive independent of the type of water deficit treatment. Among the differentially expressed genes, the overrepresented GO categories "oxidoreductase activity" and "heme binding" connected water deficit response to ROS metabolism (Opitz et al. 2014). This study provided a first global insight into water deficit-responsive genes in young maize primary roots and provided candidate genes as a starting point for future genetic analyses. In a follow-up study, the transcriptomic dynamics of distinct tissues of maize primary roots in response to drought were surveyed by exposing seedlings to PEG 8000 treatment simulating a low water potential of −0.8 MPa for 6 h (Opitz et al. 2016). Comparison of gene expression between control and drought conditions in the meristematic zone, elongation zone and in cortex and stele of the differentiation zone revealed an remarkable transcriptomic plasticity of water deficit response in these tissues. The highest number of water deficit-responsive genes was detected in the cortex and the elongation zone. Furthermore, GO terms "transcriptional regulation" and "hormone metabolism" were most prominent among the differentially expressed genes in all tissues. This indicates a global reprogramming of cellular metabolism in adaptation to drought. Moreover, genes associated with cell wall reorganization were most abundant among differentially expressed genes in the root tip. These genes likely allow for continued root growth under water deficit conditions (Opitz et al. 2016).

It has been demonstrated that lateral root formation is repressed by transient water deficit in maize (Babé et al. 2012). Upon water deficit, pericycle founder cells are blocked irreversibly at the asymmetric division stage and at various later stages (Babé et al. 2012). Recently, it has been shown that roots of Arabidopsis, rice, and maize can sense heterogeneity in water availability by "hydropatterning" in transverse orientation at the sub-organ level (Bao et al. 2014). Hydropatterning is preceded by PIN-mediated auxin efflux and TAA1 (TRYPTOPHANE-PYRUVATE AMINOTRANSFERASE 1)-mediated auxin biosynthesis to determine pre-branch sites of lateral roots (Bao et al. 2014). The process of hydropatterning is independent of endogenous abscisic acid signaling, thus distinguishing it from a classical drought response (Bao et al. 2014). Transcriptomic characterization of the local response of maize lateral root formation in contact with a moist surface revealed extensive regulation of signaling pathways, such as upregulated genes encoding LOB (LATERAL ORGAN BOUNDARIES) domain transcriptional factors, including RTCS (Taramino et al. 2007; Robbins and Dinneny 2018).

15.5 Transcriptomic Plasticity During Heterosis Manifestation in Maize Hybrids

Maize hybrids perform significantly better than their parental inbred lines, a phenomenon known as heterosis (reviewed in Hochholdinger and Hoecker 2007; Schnable and Springer 2013). Young maize roots are an ideal system to study the manifestation of heterosis, because developmental differences between inbred lines and their hybrids are established within a few days after germination thus allowing these plants to grow under very controlled conditions (Hoecker et al. 2006). An early transcriptome study based on 12 k maize microarray chips demonstrated genotype- and organ-specific non-additive gene expression patterns in hybrids (Hoecker et al. 2008). More recently, RNA-Seq-based transcriptome analyses highlighted the transcriptomic plasticity of global gene expression patterns in primary roots of maize inbred lines versus their hybrids. It was demonstrated that hundreds of genes display an extreme instance of gene expression complementation by being expressed in both hybrids but in only one of the parental inbred lines (Paschold et al. 2012). These expression patterns were designated SPE (single parent expression) complementation. Such complementation patterns are consistent with the dominance model of heterosis (Jones 1917). In subsequent studies, it was demonstrated that SPE complementation patterns are highly dynamic in different root tissues (Paschold et al. 2014) and stable under water deficit stress (Marcon et al. 2017). The concept of SPE complementation was recently generalized by demonstrating that a panel of diverse inbred lines and their hybrids consistently displayed hundreds of SPE patterns during three stages of primary root development (Baldauf et al. 2018). Remarkably, evolutionary younger non-syntenic genes were significantly overrepresented among genes displaying SPE patterns (Paschold et al. 2014; Baldauf et al. 2018). It was also demonstrated that non-syntenic genes drive non-additive, allele-specific, and differential gene expression patterns in distinct tissues of maize primary roots (Baldauf et al. 2016). Furthermore, it was indicated that non-syntenic genes contribute to the adaption of plants under changing environmental conditions (Marcon et al. 2017). A possible function of non-syntenic SPE genes could therefore be associated with the increased potential of adaptation of hybrids to environmental changes.

15.6 Future Prospects

Functional genomics studies and systems biology approaches increasingly contribute to the understanding of plant biology. Integration of functional genomics and high-throughput phenotyping will help to better understand the molecular processes during maize root development. Currently, most transcriptome analyses are executed at the level of entire plants, organs, or tissues. Although these experiments contribute to the understanding of the transcriptomic landscapes of plants, they do not resolve the cellular complexity of the analyzed samples. Techniques to isolate and subsequently analyze the biomolecules of single-cell types provide unparalleled opportunities for high-resolution systemic dissection of root traits and complex biological phenomena such as heterosis. Finally, plant roots intensively interact with their environment via the rhizosphere, which is the narrow proportion of soil that is influenced by root secretions and soil microorganisms. A grand challenge of root biology is to decipher the complex interplay between root systems and microbes in the rhizosphere of natural habitats.

Acknowledgements Root research in FH's laboratory supported by the DFG (German Research Foundation) and the BMBF (German Federal Ministry of Education and Research).

Author Contributions All authors contributed to the writing of this book chapter.

References

Babé A, Lavigne T, Severin J-P, Nagel KA, Walter A, Chaumont F, Batoko H, Beeckman T, Draye X (2012) Repression of early lateral root initiation events by transient water deficit in barley and maize. Philos Trans R Soc B Biol Sci 367:1534–1541

Baldauf JA, Marcon C, Lithio A, Vedder L, Altrogge L, Piepho H-P, Schoof H, Nettleton D, Hochholdinger F (2018) Single-parent expression is a general mechanism driving extensive complementation of non-syntenic genes in maize hybrids. Curr Biol 28:431–437

Baldauf JA, Marcon C, Paschold A, Hochholdinger F (2016) Nonsyntenic genes drive tissue-specific dynamics of differential, nonadditive and allelic expression patterns in maize hybrids. Plant Physiol 171:1144–1155

Bao Y, Aggarwal P, Robbins NE, Sturrock CJ, Thompson MC, Tan HQ, Tham C, Duan L, Rodriguez PL, Vernoux T et al (2014) Plant roots use a patterning mechanism to position lateral root branches toward available water. Proc Natl Acad Sci 111:9319–9324

Burton AL, Brown KM, Lynch JP (2013) Phenotypic diversity of root anatomical and architectural traits in *Zea* species. Crop Sci 53:1042–1055

Dembinsky D, Woll K, Saleem M, Liu Y, Fu Y, Borsuk LA, Lamkemeyer T, Fladerer C, Madlung J, Barbazuk B et al (2007) Transcriptomic and proteomic analyses of pericycle cells of the maize primary root. Plant Physiol 145: 575–588

Forde BG (2014) Nitrogen signalling pathways shaping root system architecture: an update. Curr Opin Plant Biol 21:30–36

Gao Y, Lynch JP (2016) Reduced crown root number improves water acquisition under water deficit stress in maize (*Zea mays* L.). J Exp Bot 67:4545–4557

Gaudin ACM, Mcclymont SA, Holmes BM, Lyons E, Raizada MN (2011) Novel temporal, fine-scale and growth variation phenotypes in roots of adult-stage maize (*Zea mays* L.) in response to low nitrogen stress. Plant Cell Environ 34:2122–2137

Giehl RFH, von Wiren N (2014) Root nutrient foraging. Plant Physiol 166:509–517

Guo Y, Chen F, Zhang F, Mi G (2005) Auxin transport from shoot to root is involved in the response of lateral root growth to localized supply of nitrate in maize. Plant Sci 169:894–900

He X, Qu B, Li W, Zhao X, Teng W, Ma W, Ren Y, Li B, Li Z, Tong Y (2015) The nitrate-inducible NAC transcription factor TaNAC2-5A controls nitrate response and increases wheat yield. Plant Physiol 169:1991–2005

Hetz W, Hochholdinger F, Schwall M, Feix G (1996) Isolation and characterization of *rtcs*, a maize mutant deficient in the formation of nodal roots. Plant J 10:845–857

Hey S, Baldauf JA, Opitz N, Lithio A, Pasha A, Provart N, Nettleton D, Hochholdinger F (2017) Complexity and specificity of the maize (*Zea mays* L.) root hair transcriptome. J Exp Bot 68:2175–2185

Hochholdinger F (2009) The maize root system: morphology, anatomy, and genetics. In: Bennetzen JL, Hake SC (eds) Handb. Maize, Springer, pp 145–160

Hochholdinger F, Feix G (1998) Early post-embryonic root formation is specifically affected in the maize mutant *lrt1*. Plant J 16:247–255

Hochholdinger F, Hoecker N (2007) Towards the molecular basis of heterosis. Trends Plant Sci 12:427–432

Hochholdinger F, Marcon C, Baldauf JA, Frey F, Yu P (2018b) Proteomics of maize root development. Front Plant Sci 9:143

Hochholdinger F, Park WJ, Sauer M, Woll K (2004a) From weeds to crops: genetic analysis of root development in cereals. Trends Plant Sci 9:42–48

Hochholdinger F, Woll K, Sauer M, Dembinsky D (2004b) Genetic dissection of root formation in maize (*Zea mays*) reveals root type specific developmental programmes. Ann Bot 93:359–368

Hochholdinger F, Yu P, Marcon C (2018a) Genetic control of root system development in maize. Trends Plant Sci 23:79–88

Hoecker N, Keller B, Muthreich N (2008) Comparison of maize (*Zea mays* L.) F_1-hybrid and parental inbred line primary root transcriptomes suggests organ-specific patterns of nonadditive gene expression and conserved expression trends. Genetics 179:1275–1283

Hoecker N, Keller B, Piepho HP, Hochholdinger F (2006) Manifestation of heterosis during early maize (*Zea mays* L.) root development. Theor Appl Genet 112:421–429

Ishikawa H, Evans ML (1995) Specialized zones of development in roots. Plant Physiol 109:725–727

Jansen L, Hollunder J, Roberts I, Forestan C, Fonteyne P, Van Quickenborne C, Zhen RG, Mckersie B, Parizot B, Beeckman T (2013) Comparative transcriptomics as a tool for the identification of root branching genes in maize. Plant Biotechnol J 11:1092–1102

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin C et al (2017) Improved maize reference genome with single-molecule technologies. Nature 546:524–527

Jones DF (1917) Dominance of linked factors as a means of accounting for heterosis. Genetics 2:466–479

Liu J, An X, Cheng L, Chen F, Bao J, Yuan L, Zhang F, Mi G (2010) Auxin transport in maize roots in response to localized nitrate supply. Ann Bot 106:1019–1026

Ludwig Y, Hochholdinger F (2014) Laser microdissection of plant cells. In: Žárský V, Cvrčková F (eds) Plant cell morphogenesis. Methods in molecular biology. Humana Press, Totowa, pp 249–258

Lynch JP (2011) Root phenes for enhanced soil exploration and phosphorus acquisition: tools for future crops. Plant Physiol 156:1041–1049

Lynch JP (2013) Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems. Ann Bot 112:347–357

Manoli A, Begheldo M, Genre A, Lanfranco L, Trevisan S, Quaggiotti S (2014) NO homeostasis is a key regulator of early nitrate perception and root elongation in maize. J Exp Bot 65:185–200

Marcon C, Paschold A, Malik WA, Lithio A, Baldauf JA, Altrogge L, Opitz N, Lanz C, Schoof H, Nettleton D et al (2017) Stability of single parent gene expression complementation in maize hybrids upon water deficit stress. Plant Physiol 173:1247–1257

Muthreich N, Majer C, Beatty M, Paschold A, Schutzenmeister A, Fu Y, Malik WA, Schnable PS, Piepho H-P, Sakai H et al (2013) Comparative transcriptome profiling of maize coleoptilar nodes during shoot-borne root initiation. Plant Physiol 163:419–430

Nakazono M, Qiu F, Borsuk LA, Schnable PS (2003) Laser-capture microdissection, a tool for the global analysis of gene expression in specific plant cell types: identification of genes expressed differentially in epidermal cells or vascular tissues of maize. Plant Cell 15:583–596

Nestler J, Liu S, Wen TJ, Paschold A, Marcon C, Tang HM, Li D, Li L, Meeley RB, Sakai H et al (2014) *Roothairless5*, which functions in maize (*Zea mays* L.) root hair initiation and elongation encodes a monocot-specific NADPH oxidase. Plant J 79:729–740

Opitz N, Marcon C, Paschold A, Malik WA, Lithio A, Brandt R, Piepho HP, Nettleton D, Hochholdinger F (2016) Extensive tissue-specific transcriptomic plasticity in maize primary roots upon water deficit. J Exp Bot 67:1095–1107

Opitz N, Paschold A, Marcon C, Malik WA, Lanz C, Piepho H-P, Hochholdinger F (2014) Transcriptomic complexity in young maize primary roots in response to low water potentials. BMC Genom 15:741

Paschold A, Jia Y, Marcon C, Lund S, Larson NB, Yeh C-T, Ossowski S, Lanz C, Nettleton D, Schnable PS et al (2012) Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents. Genome Res 22:2445–2454

Paschold A, Larson NB, Marcon C, Schnable JC, Yeh CT, Lanz C, Nettleton D, Piepho H-P, Schnable PS, Hochholdinger F (2014) Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. Plant Cell 26:3939–3948

Peter R, Eschholz TW, Stamp P, Liedgens M (2009) Early growth of flint maize landraces under cool conditions. Crop Sci 49:169–178

Ribaut JM, Betran J, Monneveux P, Setter T (2009) Drought tolerance in maize. In: Bennetzen JL, Hake SC (eds) Handb. Maize, Springer, pp 311–344

Robbins NE, Dinneny JR (2018) Growth is required for perception of water availability to pattern root branches in plants. Proc Natl Acad Sci 201710709

Rogers ED, Benfey PN (2015) Regulation of plant root system architecture: implications for crop advancement. Curr Opin Biotechnol 32:93–98

Saleem M, Lamkemeyer T, Schützenmeister A, Fladerer C, Piepho HP, Nordheim A, Hochholdinger F (2009) Tissue specific control of the maize (*Zea mays* L.) embryo, cortical parenchyma, and stele proteomes by RUM1 which regulates seminal and lateral root initiation. J Proteome Res 8:2285–2297

Schnable PS, Springer NM (2013) Progress toward understanding heterosis in crop plants. Annu Rev Plant Biol 64:13.1–13.18

Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Nat Acad Sci 108:4069–4074

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Sebastian J, Yee M-C, Goudinho Viana W, Rellán-Álvarez R, Feldman M, Priest HD, Trontin C, Lee T, Jiang H, Baxter I et al (2016) Grasses suppress shoot-borne roots to conserve water during drought. Proc Natl Acad Sci 113:8861–8866

Stelpflug SC, Rajandeep S, Vaillancourt B, Hirsch CN, Buell CR, De Leon N, Kaeppler SM (2015) An expanded maize gene expression atlas based on RNA-sequencing and its use to explore root development. Plant Genome 9:314–362

Tai H, Lu X, Opitz N, Marcon C, Paschold A, Lithio A, Nettleton D, Hochholdinger F (2016) Transcriptomic and anatomical complexity of primary, seminal, and crown roots highlight root type-specific functional diversity in maize (*Zea mays* L.). J Exp Bot 67:1123–1135

Tai H, Opitz N, Lithio A, Lu X, Nettleton D, Hochholdinger F (2017) Non-syntenic genes drive RTCS-dependent regulation of the embryo transcriptome during formation of seminal root primordia in maize (*Zea mays* L.). J Exp Bot 68:403–414

Taramino G, Sauer M, Stauffer JL, Multani D, Niu X, Sakai H, Hochholdinger F (2007) The maize (*Zea mays* L.) RTCS gene encodes a LOB domain protein that is a key regulator of embryonic seminal and post-embryonic shoot-borne root initiation. Plant J 50:649–659

Tian Q, Chen F, Liu J, Zhang F, Mi G (2008) Inhibition of maize root growth by high nitrate supply is correlated with reduced IAA levels in roots. J Plant Physiol 165:942–951

Trachsel S, Kaeppler SM, Brown KM, Lynch JP (2013) Maize root growth angles become steeper under low N conditions. Field Crop Res 140:18–31

Trevisan S, Manoli A, Begheldo M, Nonis A, Enna M, Vaccaro S, Caporale G, Ruperti B, Quaggiotti S (2011) Transcriptome analysis reveals coordinated spatiotemporal regulation of hemoglobin and nitrate reductase in response to nitrate in maize roots. New Phytol 192:338–352

Trevisan S, Manoli A, Ravazzolo L, Botton A, Pivato M, Masi A, Quaggiotti S (2015) Nitrate sensing by the maize root apex transition zone: a merged transcriptomic and proteomic survey. J Exp Bot 66:3699–3715

Wang Y, Mi G, Chen F, Zhang J, Zhang F (2005) Response of root morphology to nitrate supply and its

contribution to nitrogen accumulation in maize. J Plant Nutr 27:2189–2202

Woll K, Borsuk L a, Stransky H, Nettleton D, Schnable PS, Hochholdinger F (2005) Isolation, characterization, and pericycle-specific transcriptome analyses of the novel maize lateral and seminal root initiation mutant *rum1*. Plant Physiol 139:1255–1267

Yu P, Baldauf JA, Lithio A, Marcon C, Nettleton D, Li C, Hochholdinger F (2016) Root type-specific reprogramming of maize pericycle transcriptomes by local high nitrate results in disparate lateral root branching patterns. Plant Physiol 170:1783–1798

Yu P, Eggert K, von Wirén N, Li C, Hochholdinger F (2015) Cell type-specific gene expression analyses by RNA sequencing reveal local high nitrate-triggered lateral root initiation in shoot-borne roots of maize by modulating auxin-related cell cycle regulation. Plant Physiol 169:690–704

Yu P, White PJ, Hochholdinger F, Li C (2014) Phenotypic plasticity of the maize root system in response to heterogeneous nitrogen availability. Planta 240:667–678

Zhang Y, Paschold A, Marcon C, Liu S, Tai H, Nestler J, Yeh C, Opitz N, Lanz C, Schnable PS et al (2014) The *Aux/IAA* gene rum1 involved in seminal and lateral root formation controls vascular patterning in maize (*Zea mays* L.) primary roots. J Exp Bot 65:4919–4930

Zhu J, Kaeppler SM, Lynch JP (2005) Topsoil foraging and phosphorus acquisition efficiency in maize (*Zea mays*). Funct Plant Biol 32:749–762

bioRxiv preprint doi: https://doi.org/10.1101/103747; this version posted January 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

European Flint reference sequences complement the maize pan-genome

Unterseer, Sandra[1*]; Seidel, Michael A.[2*]; Bauer, Eva[1]; Haberer, Georg[2]; Hochholdinger, Frank[3]; Opitz, Nina[3]; Marcon, Caroline[3]; Baruch, Kobi[4]; Spannagl, Manuel[2]; Mayer, Klaus F.X.[2]; Schön, Chris-Carolin[1]

[1] Plant Breeding, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany, 85354

[2] Plant Genome and System Biology, Helmholtz Center Munich GmbH; Neuherberg, Germany, 85764

[3] Crop Functional Genomics, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany, 53113

[4] Energin .R Technologies 2009 Ltd (NRGENE), Ness Ziona, Israel, 7403648

* These authors contributed equally.

Abstract

The genomic diversity of maize is reflected by a large number of SNPs and substantial structural variation. Here, we report the *de novo* assembly of two European Flint maize lines to remedy the scarcity of sequence resources for the Flint pool. EP1 and F7 are important founder lines of European hybrid breeding programs. The lines were sequenced on an Illumina platform at 320X and 225X coverage. Using NRGene´s DeNovoMAGIC 2.0 technology, pseudochromosomes were assembled encompassing a total of 2,463 Mb for EP1 and 2,405 Mb for F7. Structural and functional annotation of the two genomes is currently in progress. The two high-quality *de novo* assemblies complement the existing maize pan-genome and will pave the way for future functional and comparative studies.

Keywords: short genome report, *Zea mays*, maize, Dent, Flint, assembly, genome

Introduction and organism information

Maize is an important source for food, livestock feed and industrial products. Two of the major germplasm pools exploited in breeding are Dent and Flint with their names referring to different kernel phenotypes [1]. Due to their historic geographical separation and adaptation to different environments, Dent and Flint are genetically divergent with respect to several traits differentiating the two pools, like cold tolerance, early vigour and flowering time. Worldwide, many hybrid breeding programs focus on Dent germplasm, especially the modern high yielding US Corn Belt Dent, whereas breeding programs in cooler regions of Central Europe exploit heterotic effects between Dent and Flint lines. Maize introduced into Europe after the discovery of the New World was substantially influenced by early maturing and cold tolerant North American Flint, but also by Caribbean germplasm [2]. Understanding the genomic differences between germplasm pools may contribute to a better understanding of the complementarity in heterotic patterns exploited in hybrid breeding and of mechanisms involved in the adaptation to different environments.

Maize has a flexible and dynamic genome with a tremendous amount of genetic diversity, which is reflected by a large number of SNPs and substantial structural variation [3-5]. The pan-genome concept presumes that the maize genome comprises genomic segments common to all lines and dispensable segments that can be line-specific or partially shared between lines [6]. High-quality

bioRxiv preprint doi: https://doi.org/10.1101/103747; this version posted January 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

genome sequences are essential for characterizing and exploiting the native genomic variation within maize. To date, a reference sequence of maize exists based on the Dent line B73 [3, 7] as well as a recently published *de novo* assembled genome of the Dent line PH207 [8]. To remedy the scarcity of sequence resources for the Flint pool, two reference sequences were generated *de novo* from inbred lines EP1 and F7. The two Flint lines are important founder lines of European hybrid breeding programs and trace back to the Spanish landrace Lizargarate and the French landrace Lacaune, respectively.

Genome sequencing information

For Illumina whole genome sequencing, DNA was extracted from leaf material frozen in liquid nitrogen following the protocol of [9]. DNA was extracted from 17 (EP1) or 15 (F7) single plants and then pooled for library construction. The libraries for both samples were prepared following manufacturer´s instructions. We obtained a raw sequencing data coverage of 320X for EP1 and of 225X for F7 based on libraries with five different insert sizes (470 bp, 800 bp, 3000 bp, 6000 bp, and 9000 bp). Contigs, scaffolds and pseudochromosomes were assembled using NRGene´s DeNovoMAGIC 2.0 technology. The resulting 137,249 and 130,426 contigs had an N50 length of 82,295 bp and 96,432 bp for EP1 and F7, respectively. The contig N50 count was 8,811 (EP1) and 7,368 (F7). The N50 length for 71,196 EP1 scaffolds was 6.134 Mb and the N50 length for 77,899 F7 scaffolds was 9.483 Mb. The scaffold N50 counts were 121 and 70, respectively. The final pseudochromosomes encompassed a total of 2,463 Mb (EP1) and 2,405 Mb (F7) including 2,438 Mb and 2,379 Mb ungapped sequence, respectively. Finally, we used BUSCO [10] to assess completeness of the two genome assemblies based on evolutionarily-informed expectations of gene content from 956 near-universal single-copy orthologs. The assemblies of EP1 and F7 revealed 97.3% and 96.1% of complete BUSCOs and 1.6% and 2.4% of fragmented BUSCOs, respectively.

Conclusions

High-quality *de novo* assembled reference sequences are fundamental for characterizing genomic and functional variation. The two reference sequences presented here will enable the integration of Flint diversity into the maize pan-genome and will pave the way for more detailed structural and functional analyses of Flint germplasm.

Availability of data

The *de novo* assembled genomes, Zm-EP1-REFERENCE-TUM-1.0 (Zm00010a) and Zm-F7-REFERENCE-TUM-1.0 (Zm00011a), were released in collaboration with the Maize Genetics and Genomics Database (MaizeGDB; http://www.maizegdb.org) and NCBI (BioProjects PRJNA360920 and PRJNA360923). The Whole Genome Shotgun projects have been deposited at GenBank under the accessions MTTA00000000 and MTTB00000000. The versions described here are versions MTTA01000000 and MTTB01000000. Functional annotation supported by RNA-seq data and structural analysis of the two genomes is currently in progress.

The Plant Breeding Group (Technical University of Munich) and the Plant Genome and Systems Biology Group (Helmholtz Center Munich) in collaboration with NCBI and MaizeGDB have released the two maize genomes prior to scientific publication in accordance with guidelines set forth by the Toronto Agreement for prepublication data sharing (Nature. 2009 461:168). The above groups reserve

bioRxiv preprint doi: https://doi.org/10.1101/103747; this version posted January 27, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

the first right to publish on the available EP1 and F7 data, including but not limited to whole-genome comparisons, genes, structural annotations, functional annotations, and genome-wide association studies, and to improve the sequence and its annotations.

By accessing these data users agree not to publish articles containing whole genome or chromosome scale analyses prior to publication by the Plant Breeding Group (Technical University of Munich) and the Plant Genome and Systems Biology Group (Helmholtz Center Munich). Any redistribution of these data should include the full text of the data use policy.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CCS, EB, KFXM, MS, GH and FH conceived the study; NO and CM performed sample preparation; KB assembled the genome; MAS conducted data analysis; SU and EB wrote the manuscript. All authors read and approved the final manuscript.

Funding acknowledgements

This study was funded by the Bavarian State Ministry of the Environment and Consumer Protection (Grant ID TGC01GCUFuE69741; Project BayKlimaFit; http://www.bayklimafit.de/) and the German Federal Ministry of Education and Research (BMBF) within the funding initiative "Plant Breeding Research for the Bioeconomy" (Grant ID 031B0195A; Project MAZE; http://www.europeanmaize.net).

References

1. Smith CW, Betrán J, Runge ECA: Corn : origin, history, technology, and production. Hoboken, N.J.: John Wiley; 2004.
2. Rebourg C, Chastanet M, Gouesnard B, Welcker C, Dubreuil P, Charcosset A: Maize introduction into Europe: the history reviewed in the light of molecular data. *Theor Appl Genet* 2003, 106:895-903.
3. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al: The B73 maize genome: complexity, diversity, and dynamics. *Science* 2009, 326:1112-1115.
4. Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J, et al: A first-generation haplotype map of maize. *Science* 2009, 326:1115-1117.
5. Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, et al: Maize HapMap2 identifies extant variation from a genome in flux. *Nat Genet* 2012, 44:803-807.
6. Morgante M, De Paoli E, Radovic S: Transposable elements and the plant pan-genomes. *Curr Opin Plant Biol* 2007, 10:149-155.
7. Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin C-S, et al: The complex sequence landscape of maize revealed by single molecule technologies. *bioRxiv* 2016.
8. Hirsch CN, Hirsch CD, Brohammer AB, Bowman MJ, Soifer I, Barad O, Shem-Tov D, Baruch K, Lu F, Hernandez AG, et al: Draft assembly of elite inbred line PH207 provides insights into genomic and transcriptome diversity in maize. *Plant Cell* 2016, 28:2700-2714.
9. Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW: Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences of the United States of America* 1984, 81:8014-8018.
10. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM: BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015, 31:3210-3212.

Stability of Single-Parent Gene Expression Complementation in Maize Hybrids upon Water Deficit Stress[1][OPEN]

Caroline Marcon[2], Anja Paschold[2,3], Waqas Ahmed Malik, Andrew Lithio, Jutta A. Baldauf, Lena Altrogge, Nina Opitz, Christa Lanz, Heiko Schoof, Dan Nettleton, Hans-Peter Piepho, and Frank Hochholdinger*

Institute of Crop Science and Resource Conservation, Crop Functional Genomics (C.M., A.P., J.A.B., N.O., F.H.) and Crop Bioinformatics (L.A., H.S.), University of Bonn, 53113 Bonn, Germany; Institute for Crop Science, Biostatistics Unit, University of Hohenheim, 70599 Stuttgart, Germany (W.A.M., H.-P.P.); Department of Statistics, Iowa State University, Ames, Iowa 50011-1210 (A.L., D.N.); and Department of Molecular Biology, Max Planck Institute for Developmental Biology, 72076 Tuebingen, Germany (C.L.)

ORCID IDs: 0000-0001-6699-9441 (C.M.); 0000-0002-7286-0558 (J.A.B.); 0000-0002-6871-235X (L.A.); 0000-0003-0776-4890 (N.O.); 0000-0002-1527-3752 (H.S.); 0000-0002-6045-1036 (D.N.); 0000-0001-7813-2992 (H.-P.P.); 0000-0002-5155-0884 (F.H.).

Heterosis is the superior performance of F1 hybrids compared with their homozygous, genetically distinct parents. In this study, we monitored the transcriptomic divergence of the maize (*Zea mays*) inbred lines B73 and Mo17 and their reciprocal F1 hybrid progeny in primary roots under control and water deficit conditions simulated by polyethylene glycol treatment. Single-parent expression (SPE) of genes is an extreme instance of gene expression complementation, in which genes are active in only one of two parents but are expressed in both reciprocal hybrids. In this study, 1,997 genes only expressed in B73 and 2,024 genes only expressed in Mo17 displayed SPE complementation under control and water deficit conditions. As a consequence, the number of active genes in hybrids exceeded the number of active genes in the parental inbred lines significantly independent of treatment. SPE patterns were substantially more stable to expression changes by water deficit treatment than other genotype-specific expression profiles. While, on average, 75% of all SPE patterns were not altered in response to polyethylene glycol treatment, only 17% of the remaining genotype-specific expression patterns were not changed by water deficit. Nonsyntenic genes that lack syntenic orthologs in other grass species, and thus evolved late in the grass lineage, were significantly overrepresented among SPE genes. Hence, the significant overrepresentation of nonsyntenic genes among SPE patterns and their stability under water limitation might suggest a function of these genes during the early developmental manifestation of heterosis under fluctuating environmental conditions in hybrid progeny of the inbred lines B73 and Mo17.

Maize (*Zea mays*) is a fundamental source of food, feed, and energy (Persson et al., 2009). Maize displays a remarkable degree of structural intraspecific genomic diversity, which is likely unparalleled among higher eukaryotes (Springer et al., 2009; Swanson-Wagner et al., 2010). The divergence between different maize genotypes is the consequence of copy number variation, presence/absence variation (PAV), single-nucleotide polymorphisms (SNPs), and insertion-deletion polymorphisms. Between the widely used inbred lines B73 and Mo17, thousands of structural variations (Springer et al., 2009; Swanson-Wagner et al., 2010), differences in total genome size (Laurie and Bennett, 1985), and content of repetitive DNA (Kato et al., 2004) have been detected, underscoring their high degree of intraspecific variation. Syntenic genes are defined by their preserved colocalization on chromosomes of different species. A syntenic comparison of the genomes of grass species demonstrated that the lineage leading to maize experienced a whole-genome duplication about 5 to 12 million years ago (Blanc and Wolfe, 2004; Swigonová et al., 2004). Comparisons of the duplicated genome of modern maize with orthologous regions of its unduplicated progenitor sorghum (*Sorghum bicolor*) revealed that many of the duplicated maize genes disappeared in the course of evolution (Haberer et al., 2005). Nevertheless, 3,000 to

[1] This work was supported by the Deutsche Forschungsgemeinschaft (grant no. HO2249-9/1 to F.H. and grant no. PI 377/12-1 to W.A.M. and H.-P.P.) and by the European Union's Seventh Framework Programme (FP7/2007-2013; grant no. 289300.27 to F.H.).

[2] These authors contributed equally to the article.

[3] Present address: Max Planck Institute for Plant Breeding Research, 50829 Cologne, Germany.

* Address correspondence to hochholdinger@uni-bonn.de.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Frank Hochholdinger (hochholdinger@uni-bonn.de).

C.M., A.P., and F.H. conceived the research; C.M., N.O., and C.L. carried out the RNA sequencing; W.A.M., A.L., J.A.B., L.A., H.S., D.N., and H.-P.P. carried out the statistical analyses; C.M., A.P., and F.H. interpreted the data; C.M. drafted the article; F.H. coordinated the study and helped to draft the article; all authors approved the final draft of the article.

[OPEN] Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.16.01045

 © 2017 American Society of Plant Biologists. All Rights Reserved.

Downloaded from https://academic.oup.com/plphys/article/173/2/1247/6115918 by ulbbonn user on 21 December 2021

5,000 pairs of paralogs are retained in modern maize (Schnable et al., 2011). The synteny to sorghum allows a subdivision of the maize genome into two distinct subgenomes. Each of these subgenomes originally contained a complete set of genomic information. Thus, the gene complement of modern maize can be divided into three categories: (1) pairs of duplicate genes that are shared by both subgenomes; (2) single-copy genes present in only one of the two subgenomes; and (3) genes that cannot be assigned to any subgenome. This third class of genes is characterized by a lack of syntenic orthologs in the genomes of other grass species. Most of these nonsyntenic genes emerged by single-gene duplication mechanisms, most likely after the last genome duplication of maize (Woodhouse et al., 2010).

Heterosis or hybrid vigor denotes the superior performance of heterozygous F1 plants compared with their genetically distinct, homozygous parents (for review, see Falconer and Mackay, 1996). Hybrid vigor was monitored for traits such as biomass and grain yield but also resistance to environmental cues, such as drought stress (Araus et al., 2010). The phenomenon of heterosis is often associated with transcriptional variation between inbred lines and hybrids (Hochholdinger and Hoecker, 2007). For instance, gene expression in hybrids that is significantly different from the average of the parental inbred lines (midparent value) is described as nonadditive (Birchler et al., 2003; Stupar and Springer, 2006). Moreover, in maize primary roots and their tissues, hundreds of genes were expressed exclusively in only one of the two parental inbred lines, whereas these genes were expressed in both reciprocal hybrids. This pattern, which is an extreme instance of complementation at the gene expression level, was termed single-parent expression (SPE; Paschold et al., 2012, 2014). As a consequence of SPE, hybrids display hundreds of more active genes than each of their parents.

In global agriculture, drought leads to more yield loss in crops such as maize than any other stress (Boyer, 1982). Global warming alters precipitation patterns, and the anticipated world population of nine billion by 2050 requires an estimated increase in crop production by 70% (Tester and Langridge, 2010). Water shortage severely influences many physiological and developmental aspects of plants (for review, see de Oliveira et al., 2013). For instance, continuous primary root elongation at low water potentials is an important adaptive trait in maize to access deeper water layers (Sharp et al., 1988). In rice (*Oryza sativa*), the recently characterized protein DEEPER ROOTING1 leads to enhanced downward directed root growth and avoidance of drought stress (Uga et al., 2013). The maintenance of root growth is mediated by the plant hormone abscisic acid. Abscisic acid prevents excessive ethylene production and growth inhibition by promoting the transport of Pro to the root apex (for review, see Yamaguchi and Sharp, 2010). Pro contributes to osmotic adjustment by protecting cellular turgor. Consequently, water influx into cells is enabled by a more negative water potential compared with the outside medium (for review, see Yamaguchi and Sharp, 2010). In contrast to these physiological processes, little is known about the metabolic and molecular networks involved in drought adjustment (Spollen et al., 2008; Krasensky and Jonak, 2012). Drought stress occurring in drying soil can be simulated by polyethylene glycol (PEG) treatment of plants (Zheng et al., 2004; Opitz et al., 2014, 2016). Simulation of moderate, naturally occurring drought conditions is of great agronomic importance, as such drought conditions are frequent and reduce yield (Sinclair, 2011; O'Geen, 2012). It was hypothesized that maize hybrids are better adapted to environmental stresses such as drought than their parental inbred lines, possibly as a result of improved water use efficiency (Araus et al., 2010).

To understand the differences in their transcriptomic response to drought, we subjected young primary roots of the maize hybrids B73xMo17 and Mo17xB73 and their parental inbred lines B73 and Mo17 to global gene expression experiments by RNA sequencing (RNA-seq). Our experiments revealed robust SPE patterns, which were significantly less susceptible to water deficit treatment than other genotype-specific gene activity patterns.

RESULTS

The Influence of Genotype and Water Deficit Treatment on Differential Gene Expression in Primary Root Transcriptomes of Maize Inbred Lines and Their Hybrid Progeny

The transcriptomic landscape of maize primary roots grown under control or water deficit conditions simulated by PEG treatment was determined in the inbred lines B73 and Mo17 and their reciprocal hybrid progeny B73xMo17 and Mo17xB73 by RNA-seq. Each genotype-by-treatment combination was surveyed in four biological replicates. After quality control and stacked read removal (see "Materials and Methods"), on average, 14 million reads of the 32 cDNA libraries (67%) mapped uniquely to the maize reference genome (Supplemental Table S1). On average, 77% of these sequences mapped to the 39,656 high-confidence gene models of the maize filtered gene set (FGSv2; Supplemental Table S1).

The relationship of the 32 sequenced transcriptomes is summarized in a multidimensional scaling (MDS) plot indicating the overall relationship of the different RNA-seq samples (Fig. 1A). Samples representing control and water deficit conditions for each of the four genotypes clustered closely together, illustrating their high transcriptomic correlation. The similarity between the reciprocal hybrids B73xMo17 and Mo17xB73, which contain identical nuclear genomes, was much higher than that between their parental inbred lines B73 and Mo17.

In the four genotypes, between 229 (140 up + 89 down; Mo17) and 902 (582 up + 320 down; B73xMo17) genes were regulated by water deficit stress (Fig. 1B). Between 61% (Mo17) and 67% (Mo17xB73) of these genes were

Downloaded from https://academic.oup.com/plphys/article/173/2/1247/6115918 by ulbbonn user on 21 December 2021

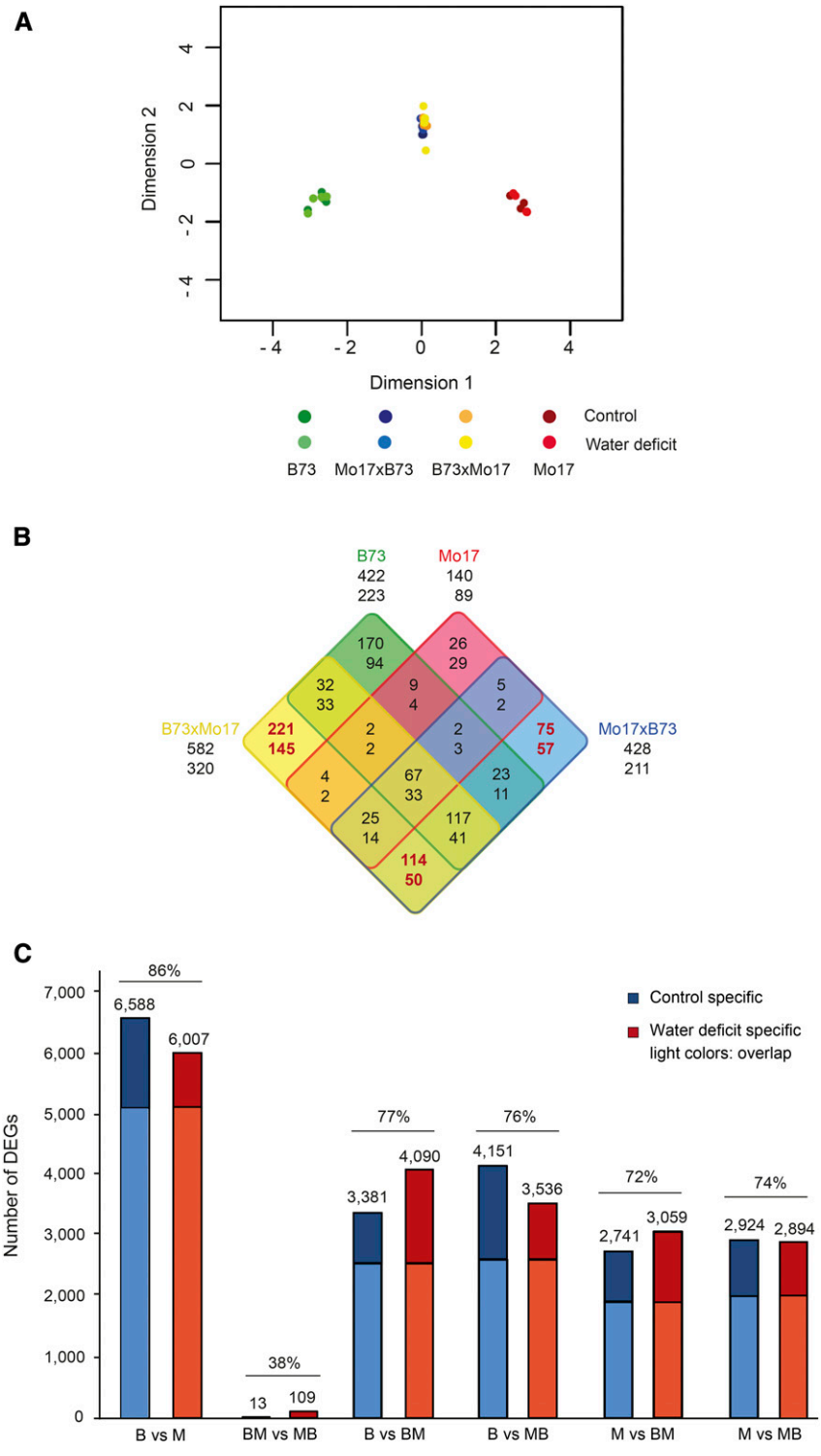


Figure 1. Sample relationship and differential gene expression. A, MDS plot illustrating the level of similarity of the transcriptomes of the four genotypes after control and water deficit treatments. B, Four-way Venn diagram of the number and distribution of genes differentially expressed between control water deficit conditions and overlap of these genes between the different genotypes. The four genotypes are color coded as indicated by the color of the genotype labels. To facilitate interpretation, the eight boxes per genotype are encircled by the corresponding genotype color. Hence, the central field, which represents the intersection of all four genotypes, comprises genes consistently up- or down-regulated by drought stress in all four genotypes. False discovery rate (FDR) $\leq$ 5%, $\log_2$ fold change $>$ 1; top number, up-regulated under water deficit; bottom number, down-regulated under water deficit. C, Number of differentially expressed genes (DEGs; FDR $\leq$ 5%, $\log_2$ fold change $>$ 1) between individual genotypes after control and water deficit conditions (percentages indicate overlapping DEGs). B, B73; BM, B73xMo17; M, Mo17; MB, Mo17xB73.

up-regulated upon water deficit treatment (top numbers below each genotype), while the remaining genes were down-regulated upon drought stress (bottom numbers below each genotype). Remarkably, a substantial number of drought-regulated genes displayed genotype-specific regulation. In particular, between 46% (Mo17xB73) and 59% (B73xMo17) of genes regulated in the reciprocal hybrids were only regulated in hybrids but not in their parental inbred lines (Fig. 1B, red numbers).

The number of differentially expressed genes between the four genotypes was determined in six pairwise comparisons that were performed for control and water deficit-treated samples separately (Fig. 1C). The highest number of differentially expressed genes for

Downloaded from https://academic.oup.com/plphys/article/173/2/1247/6115918 by ulbbonn user on 21 December 2021

both treatments was observed between the genetically distinct parental inbred lines B73 and Mo17, and the lowest number was detected between the reciprocal hybrids B73xMo17 and Mo17xB73 (Fig. 1C). In all six comparisons, differential gene expression patterns were highly conserved up to 86% (B73 versus Mo17) irrespective of the treatment (Fig. 1C, light colors on the base of the bars). Nevertheless, hundreds of genes were differentially expressed only under control conditions and stable under water deficit stress or vice versa.

Nonadditive Gene Expression Is Often Conserved between Reciprocal Hybrids

To determine gene expression patterns in hybrids relative to their parents, nine expression categories were defined (Table I). These expression classes were based on three pairwise Student's *t* tests to estimate genes differentially expressed between the hybrid and the lower (low-parent) and higher (high-parent) expressing inbred lines. According to this classification scheme, most genes were assigned to classes 1 to 4, in which hybrid expression is between the parental inbred lines. Nevertheless, hundreds of genes displayed extreme expression levels in hybrids above the high parent (classes 5 and 6) or below the low parent (classes 7 and 8). The majority of these genes (4,469) displayed below low-parent expression, whereas 1,968 genes revealed above high-parent expression in the reciprocal hybrids under control and water deficit conditions. The remaining genes, which were not covered by the criteria defined for classes 1 to 8, were summarized in class 9. Within the nine expression classes, a subset of genes exhibited nonadditive gene expression (i.e. expression levels in a hybrid that were significantly different from the average of the parental values [midparent value]). In summary, in the hybrid B73xMo17, 17% (4,549) of genes under control and 38% (10,144) of genes under water deficit conditions displayed nonadditive expression. Similarly, in the hybrid Mo17xB73, 27% (7,185) of genes were significantly different from the midparent value under control conditions, whereas 35% (9,230) of genes displayed nonadditive expression under water deficit. The overlap of 2,192 (48%; expected 317) nonadditively expressed genes in the hybrid B73xMo17 under both conditions and 3,344 (47%; expected 490) in Mo17xB73 significantly exceeded the expected numbers by 7-fold in both genotypes ($P \leq 0.01$; Table I), suggesting a biological relevance of these patterns.

Treatment-Dependent Overlap of Genes Exhibiting Allelic Expression Ratios in Hybrids Different from Their Parental Ratio

For genes expressed in all four genotypes (see "Materials and Methods"), the expression ratio of the two parental alleles for each gene was determined in control and water deficit-treated hybrids and compared with the allelic expression ratio of the two parental alleles in

Table I. *Determination of nonadditive expression classes in reciprocal maize hybrids under control and water deficit conditions*

Class[a]	Expression Pattern	Criteria			Total No. of Genes						No. of Nonadditive Genes[b]							
		LP ≠ H	HP ≠ H	LP ≠ HP	B73xMo17		Overlap[c]		Mo17xB73		Overlap[c]		B73xMo17		Overlap[c]		Mo17xB73		Overlap[c]	
					C	WD	E	O	C	WD	E	O	C	WD	E	O	C	WD	E	O
1	LP < H < HP	Y	Y	Y	3,783	3,109	440	2,398*	3,325	3,187	396	2,351*	1,179	1,196	53	829*	1,154	1,119	48	832*
2	LP = H < HP	Y	Y	Y	3,175	3,751	445	1,519*	4,400	3,724	613	2,110*	581	1,475	32	209*	1,556	1,273	74	508*
3	LP < H = HP	Y	N	Y	5,845	5,599	1,224	3,179*	5,856	5,276	1,156	3,468*	1,437	3,180	171	754*	3,254	2,207	269	1,309*
4	LP = HP = H	N	N	N	8,559	7,224	2,313	4,821*	6,960	8,555	2,227	4,839*	3	84	0	0	55	44	0	0
5	LP < HP < H	Y	Y	Y	47	267	0	9*	148	65	0	17*	47	267	0	9*	148	65	0	17*
6	LP = HP < H	Y	Y	N	199	641	5	81*	342	259	3	92*	199	641	5	81*	342	259	3	92*
7	H < LP < HP	Y	Y	Y	134	343	2	49*	260	265	3	72*	134	343	2	49*	260	265	3	72*
8	H < LP = HP	Y	Y	N	430	1,202	19	189*	1,022	813	31	322*	430	1,202	19	189*	1,022	813	31	322*
9	Ambiguous				4,565	4,601	786	1,158*	4,424	4,593	760	1,225*	539	1,756	35	72*	1,439	1,140	61	192*
Total					26,737	26,737	5,234	13,403*	26,737	26,737	5,189	14,496*	4,549	10,144	317	2,192*	9,230	7,185	490	3,344*

[a]The classification of the relative expression of hybrids (H) compared to their parental inbred lines (HP, high parent; LP, low parent) is based on the three pairwise tests listed in the Criteria columns (N, no; Y, yes). [b]Nonadditively expressed genes were determined relative to their midparent value. [c]Expected (E) values were calculated as (number of nonadditive genes in B73×Mo17 × number of nonadditive genes in Mo17×B73)/number of all expressed genes. The number of all expressed genes is the sum of classes 1 to 9. C, Control; O, observed; WD, water deficit. The numbers of expected and observed overlapping genes were compared using a χ^2 test. Significant values are indicated by asterisks: $P \leq 0.01$.

Downloaded from https://academic.oup.com/plphys/article/173/2/1247/6115918 by ulbbonn user on 21 December 2021

the inbred lines. In total, between 491 (B73xMo17 control) and 563 (B73xMo17 water deficit) genes displayed allelic ratios in hybrids significantly different from the expression ratio of the parental inbred lines (Fig. 2). Genes of the reciprocal hybrids displaying allelic expression different from their parental ratio in a defined treatment significantly overlapped and exceeded the expected values by 13-fold (water deficit; 437:32) to 15-fold (control; 418:27; Supplemental Table S2). Similarly, the number of genes with allelic ratios different from their parental values significantly exceeded the parental ratios by 6-fold (Mo17xB73; 188:29) to 7-fold (B73xMo17; 177:27) when comparing the same hybrids at different treatments. Between 25% (125; B73Mo17 control) and 60% (324; Mo17xB73 water deficit) of the genes, which displayed allelic ratios in hybrids different from the allelic ratios of their parental inbred lines, also showed nonadditive gene expression (Supplemental Table S2). Again, the observed overlap of nonadditively expressed genes between the reciprocal hybrids under the same condition and the same hybrids under water deficit versus control conditions significantly exceeded the expected values by 20- to 23-fold (163:8 and 93:4) and 11- to 13-fold (89:8 and 52:4), respectively.

The genes displaying allelic ratios in hybrids deviating from the allelic ratios of the parental inbred lines were compared between the reciprocal hybrids and both treatments. The results are summarized in a four-way Venn diagram (Fig. 2). This comparison indicated that 23% and 25% of the genes showing allelic ratios different from their parental lines were conserved in the two reciprocal hybrids after control and water deficit treatment, respectively. In contrast, only 0.5% and 0.6% of such genes overlapped exclusively between the same hybrids under different conditions, indicating a strong treatment effect rather than a reciprocal effect.

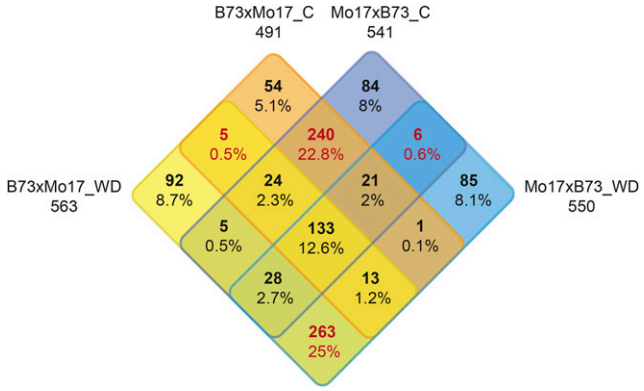


Figure 2. Distribution of genes exhibiting allelic expression ratios in hybrids different from their parental ratios under control and water deficit conditions.

Determination of the Activity Status of Genes

An extreme instance of differential gene expression is activity versus inactivity of a gene in different genotype-by-treatment combinations. The activity status (active/inactive) of individual genes between genotypes and treatments was determined by a generalized linear mixed model (see "Materials and Methods"). Water deficit treatment (bottom numbers under the genotypes) resulted in 53 (B73xMo17, 25,620 to 25,567) to 148 (B73, 24,803 to 24,655) more active genes in all four genotypes compared with control conditions (Fig. 3A). The vast majority (85%) of active genes were expressed in all four genotypes (B73, Mo17, B73xMo17, and Mo17xB73) under control (Fig. 3A, top numbers) or water deficit (Fig. 3A, bottom numbers) conditions. Among the genes not active in all genotypes, the majority displayed SPE (red numbers). SPE is an extreme instance of complementation in which genes are only active in one of the two parental inbred lines but in both reciprocal hybrid progeny. Under control conditions, 853 SPE_B genes active in B73 and both reciprocal hybrids but not in Mo17 and 1,144 SPE_M genes transcribed only in Mo17 and both hybrids but not in B73 were detected (Fig. 3A, top red numbers). In primary roots subjected to water deficit stress, 841 SPE_B and 1,183 SPE_M genes were identified (Fig. 3A, bottom red numbers). Among the 853 genes displaying SPE_B under control treatment, 643 (75%) genes also displayed SPE under water deficit conditions (Fig. 3A). Only a small fraction of 2.5% to 15% of 1,051 SPE_B patterns (see Venn diagram in Fig. 3A) are conditioned by genomic PAV (i.e. lack of these genes in the Mo17 genome; Supplemental Table S4). Similarly, among the 1,144 genes showing SPE_M under control conditions, 857 (75%) displayed the SPE pattern under water deficit stress as well. A detailed list indicating the genotype-specific activity status of each of the 29,986 genes active in at least one genotype is provided in Supplemental Data Set S1.

SPE Patterns Are Less Susceptible to Alterations by Water Deficit Than Other Genotype-Specific Expression Profiles

The stability of genotype-specific activity patterns upon water deficit treatment is reported in Supplemental Table S3. The diagonal green line highlights stable activity patterns unaltered under changing water availability (Supplemental Table S3), while the yellow fields indicate patterns that display different genotype-specific activity patterns upon water deficit. Subsequently, for each of the genotype-specific activity patterns, the ratio of stable (green) expression patterns versus dynamic (yellow) expression patterns was determined (Fig. 3B). As an example, in Supplemental Table S3, the dynamic expression patterns for the SPE genes are highlighted by a green frame while the stable expression patterns are highlighted by a dark green background. Significantly, among all genotype-specific expression patterns, only SPE patterns were highly

Downloaded from https://academic.oup.com/plphys/article/173/2/1247/6115918 by ulbbonn user on 21 December 2021

stable under changing water availability (Fig. 3B). While, on average, 75% of all SPE patterns were unaltered in response to PEG treatment, only 17% of the remaining genotype-specific expression patterns were not affected by water deficit. This highlights the superior robustness of SPE patterns compared with other genotype-specific expression patterns.

Nonsyntenic Genes Are Highly Overrepresented, While Classical Maize Genes Are Significantly Underrepresented among SPE Genes

To survey the evolutionary origin of SPE genes, the frequency of syntenic and nonsyntenic genes was determined. Syntenic genes are defined by their preserved colocalization on chromosomes of different species. In maize, these genes emerged after the last whole-genome duplication of a maize progenitor and display synteny to the unduplicated genome of sorghum, a close but unduplicated relative of maize. In contrast, genes that evolved after this whole-genome duplication are nonsyntenic to sorghum genes. Among the 39,656 genes of the FGSv2, 51% (20,291 of 39,656) were declared nonsyntenic (Schnable et al., 2011; Fig. 4A). In total, 38% (10,382 of 27,274) of all expressed genes, 39% (10,427 of 26,866) of the genes expressed under control conditions, and 40% (10,843 of 26,974) of the genes expressed after water deficit treatment were nonsyntenic. The relative number of expressed nonsyntenic genes under both conditions (39% and 40%) was significantly lower than the relative frequency of all nonsyntenic genes of FGSv2 (51%), suggesting that nonsyntenic genes are expressed less frequently in roots than expected by their abundance in the genome. By contrast, 74% (1,470 of 1,997) and 75% (1,511 of 2,024) of the SPE genes identified in control and water deficit-treated roots were nonsyntenic (Fig. 4A). This suggests that most of the SPE genes are evolutionarily younger and without synteny to sorghum genes. Moreover, 78%

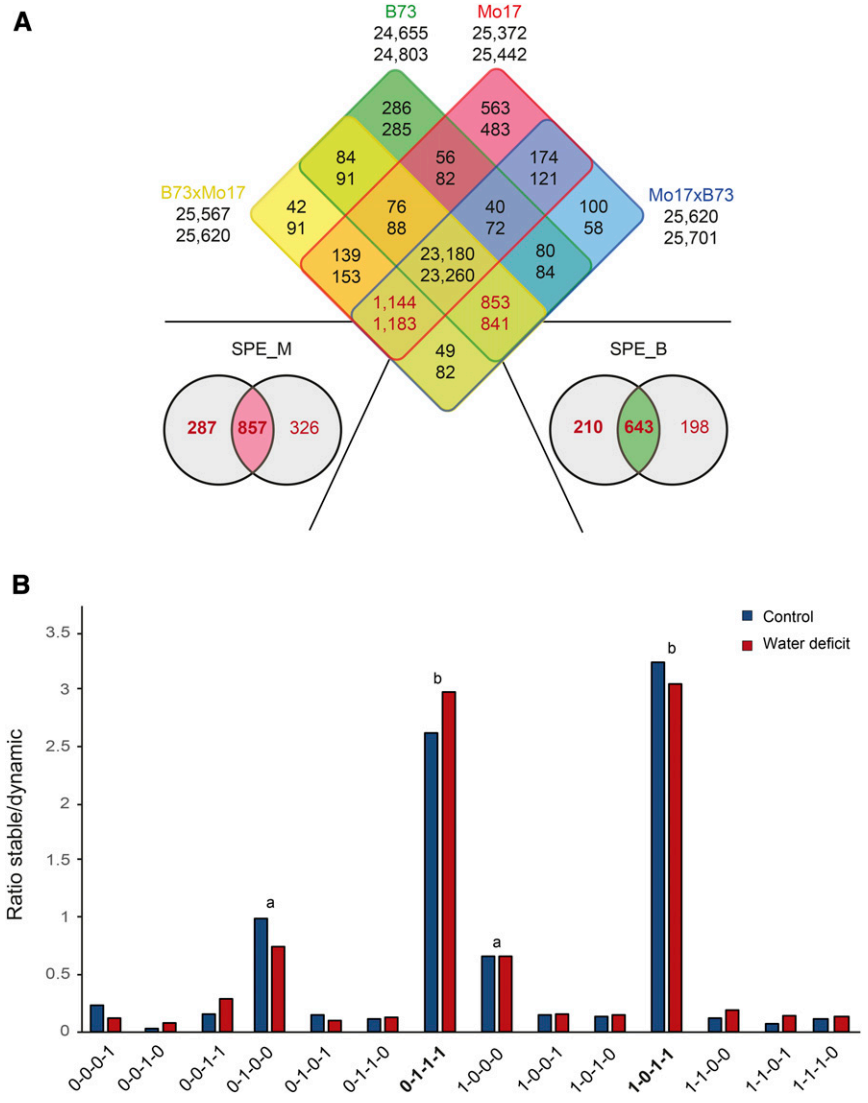


Figure 3. Distribution of SPE genes under control and water deficit conditions. A, Four-way Venn diagram summarizing gene activity (on/off) patterns in primary roots of the different genotypes and their overlap, after control (top numbers) and water deficit (bottom numbers) treatments. Genotype-specific color coding is as in Figure 1B. SPE genes are indicated in red. Overlap of SPE_M (expression in Mo17 and both hybrids) and SPE_B (expression in B73 and both hybrids) genes among control and water deficit-treated samples is shown in additional two-way Venn diagrams below the four-way Venn diagram. B, Ratio of stable genes (identical genotype-specific expression under control and water deficit treatments) to dynamic genes (variable genotype-specific expression pattern in control versus water deficit conditions). The binary code on the *x* axis explains the activity status of the four genotypes in the order B73-Mo17-B73xMo17-Mo17xB73. 0, Inactive; 1, active. Lowercase letters a and b illustrate significant differences to the ratios of all other genotype-specific gene expression profiles, according to an Edwards-Berry test at a family-wise type I error rate of $\alpha = 5\%$ (Piepho, 2012).

Downloaded from https://academic.oup.com/plphys/article/173/2/1247/6115918 by ulbbonn user on 21 December 2021

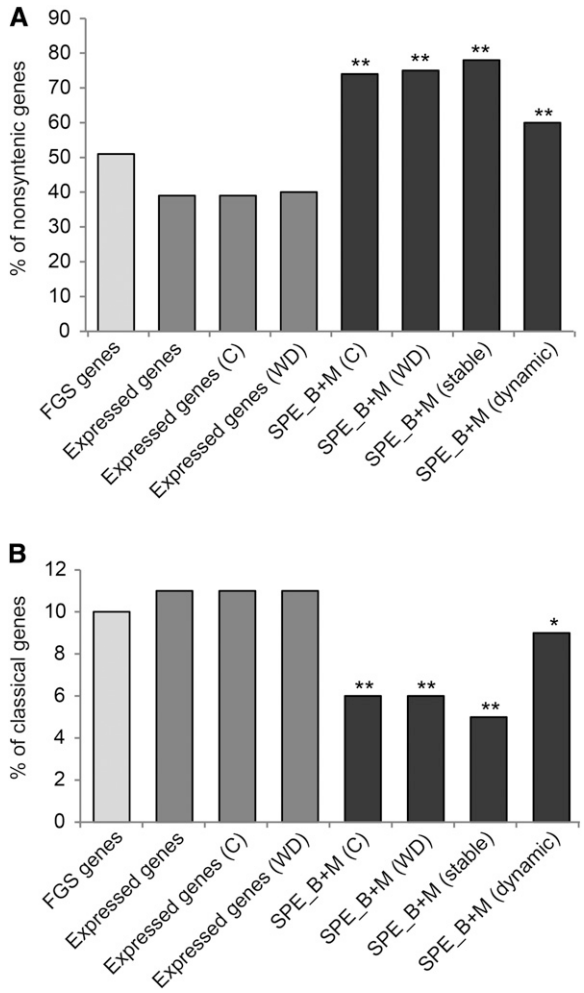


Figure 4. Distribution of nonsyntenic genes (A) and classical maize genes (B) among all expressed genes and genes displaying SPE. Significant differences ($\alpha \leq 5\%$) compared with expressed genes are indicated: **, $P \leq 0.001$. C, Control; WD, water deficit stress.

(1,170 of 1,500) of all stable SPE genes and 59% (604 of 1,016) of dynamic SPE genes were nonsyntenic.

In maize, a set of 3,946 hand-curated classical genes with experimentally determined functions has been defined (Schaeffer et al., 2011; Schnable and Freeling, 2011). Among the genes of the FGSv2, 10% (3,927 of 39,656) are classical maize genes. Similar proportions were observed among all genes expressed in control (11%; 3,008 of 26,866) and water deficit (11%; 3,037 of 26,974; Fig. 4B) conditions. In contrast, significantly fewer classical maize genes were identified among SPE genes observed under control (6%; 126 of 1,997) and drought stress (6%; 120 of 2,024) conditions. A similar number of classical maize genes was detected among the stable SPE genes (5%; 76 of 1,500), whereas 9% (93 of 1,016) of the dynamic SPE genes were assigned to classical genes. These findings underline the predominant classification of stable SPE genes to the set of nonsyntenic genes, representing classical genes to a lesser extent than syntenic genes.

DISCUSSION

Maize displays an exceptional degree of intraspecific genome diversity compared with other higher eukaryotes (Springer et al., 2009; Swanson-Wagner et al., 2010; Marroni et al., 2014). High-throughput sequencing technologies such as RNA-seq allow monitoring how this genotype-specific genomic diversity is reflected by the transcriptomic landscape of maize (Hansey et al., 2012; Hirsch et al., 2014). In this study, RNA-seq was applied to compare the primary root transcriptomes of B73, Mo17, and their reciprocal hybrids B73xMo17 and Mo17xB73 under control and water deficit conditions. Primary roots were selected as a model because they are the first line of defense against water deficit. Moreover, they display a simple organization, and heterosis is manifested in primary roots within 3 to 4 d after germination (Hoecker et al., 2006). An MDS plot illustrated that transcriptomic differences between genotypes were more pronounced than those between control and water deficit treatments in maize primary roots (Fig. 1A). Similarly, in a previous study, the transcriptomic disparity between four different maize root tissues was larger than the transcriptomic differences of the individual tissues under control and water deficit conditions (Opitz et al., 2016).

The sample relationships in the MDS plot (i.e. smaller distances between treatments than by genotypes) also were reflected by the smaller number of water deficit-regulated genes (Fig. 1B) compared with the large number of differentially expressed genes between the distantly related parental genotypes B73 and Mo17 (Fig. 1C).

In this study, the majority of genes differentially expressed in response to water deficit (Fig. 1B) was up-regulated (61%–67%), suggesting that limited water availability tends to activate gene expression. Previous studies displayed similar trends of preferentially up-regulated genes upon water deficit treatment, with 81% in whole B73 primary roots (Opitz et al., 2014) and between 68% and 88% in four different root tissues (Opitz et al., 2016). A higher number of drought-induced than drought-repressed genes also has been reported for the whole maize root system of the inbred line SRG-200 (Humbert et al., 2013) and maize seedling roots of the inbred line Han21 (Zheng et al., 2004).

Six pairwise comparisons of gene expression levels in the four genotypes revealed a large number of differentially expressed genes between the distantly related parental inbred lines B73 and Mo17 (Lorenz and Hoegemeyer, 2013), which was highly conserved (86%) in control and water deficit treatment (Fig. 1C). In contrast, only a small number of differentially expressed genes were detected between the genetically identical reciprocal hybrids (Fig. 1C). Independent of treatment, comparisons of gene expression in hybrids versus inbred lines, which share 50% genomic identity, revealed intermediate numbers of differentially expressed genes (Fig. 1C). Hence, the degree of genomic difference correlates with

Downloaded from https://academic.oup.com/plphys/article/173/2/1247/6115918 by ulbbonn user on 21 December 2021

differential gene expression (Paschold et al., 2012; Baldauf et al., 2016). Similar genotype-associated trends of differential expression have been reported previously in these four maize genotypes on the transcriptome level (Stupar and Springer, 2006; Paschold et al., 2012; Baldauf et al., 2016) and the proteome level (Marcon et al., 2013). Likewise, in Arabidopsis (*Arabidopsis thaliana*), the number of differentially expressed genes between three landraces significantly exceeded that of the corresponding reciprocal hybrids (Vuylsteke et al., 2005).

Nonadditively expressed genes display expression values in the hybrid that are significantly different from the average midparent value of the parent (Paschold et al., 2012). Such genes have been suggested to be associated with heterosis (Swanson-Wagner et al., 2006; Hoecker et al., 2008). In this study, thousands of genes displayed nonadditive expression under control (9,230) and water deficit (7,185) treatment. Notably, the overlap of nonadditively expressed genes between the two treatments (47%) by far exceeded the expected values (Table I). This implies that a major proportion of nonadditively expressed genes in hybrids is robust against water deficit stress and, therefore, might be of relevance for the early developmental manifestation of heterosis.

Allelic gene expression ratios in the hybrids deviating from the expression ratio of the corresponding parental alleles in the inbred lines were determined in this study (Fig. 2). In total, 614 and 676 genes (Supplemental Table S2) under control and water deficit conditions were identified for which the expression ratio of the parental inbred lines was different from that of the reciprocal hybrids. In whole maize primary roots (Paschold et al., 2012) and root tissues (Baldauf et al., 2016), 840 and 739 genes, respectively, were reported that show allelic expression ratios in hybrids deviating from the expression ratios of the parental alleles in the inbred lines. In a pioneering study for a set of 15 maize genes, allelic expression variation was observed in response to drought stress, indicating an unequal function of the parental alleles in hybrids (Guo et al., 2004). Such expression patterns could, at least in part, be the result of epigenetic regulation and be associated with heterosis. In Arabidopsis hybrids, it was shown that reduced levels of 24-nucleotide small RNAs in comparison with the parental inbred lines change the epigenetic landscape of hybrids, leading to the formation of heterotic phenotypes (Groszmann et al., 2011).

In a qualitative analysis of gene expression, transcriptionally active and inactive genes were determined for each genotype/treatment combination by a generalized linear mixed model approach. In this study, 27,274 genes were declared active in at least one genotype/treatment combination (Supplemental Data Set S1). Similar numbers of 27,438 active genes were detected in root tissues of B73 (Opitz et al., 2016) and 27,347 genes in maize seedling root tissues of the same genotypes as used in this study (Paschold et al., 2014). SPE patterns denote gene expression in only one parental inbred line but in both reciprocal hybrids (Paschold et al., 2012). For SPE_B patterns (i.e. for genes active only in B73 but not in Mo17), it has been determined that only 2.5% to 15% of these patterns are conditioned by PAV (i.e. the lack of these genes in the Mo17 genome; Supplemental Table S4). This is in line with previous observations where 18% of SPE_B genes in roots were considered PAVs (Paschold et al., 2012). On the level of gene expression, SPE represents a special instance of complementation, which results in several hundred more active genes in hybrids (Fig. 3A). In previous analyses, we also observed hundreds of more genes active in primary roots (Paschold et al., 2012) and in distinct primary root tissues (Paschold et al., 2014) of hybrids. It has been discussed that the activity of additional genes might be advantageous for hybrid plants (Paschold et al., 2012, 2014), because SPE genes encode rather redundant functions that could help to make biological processes more robust against unfavorable changes of environmental conditions (Paschold et al., 2014). This is substantiated by our finding here that, among all genotype- and treatment-specific expression profiles, SPE genes were exceptionally robust against changing water availability (Fig. 3B; Supplemental Table S3; i.e. that these expression patterns are not susceptible to alterations by water deficit conditions).

A whole-genome duplication ~5 to 12 million years ago (Blanc and Wolfe, 2004; Swigonová et al., 2004) separated maize from its close relative sorghum (Paterson et al., 2009). In modern maize, genes with synteny to sorghum resulted from this polyploidization event, while nonsyntenic genes evolved after this last whole-genome duplication (Woodhouse et al., 2010; Schnable et al., 2011), such as by transposition of existing genes (Freeling et al., 2008), by modular rearrangement of protein-encoding domains (Kersting et al., 2012), or by exon shuffling from transposons resulting in fusion genes (Barbaglia et al., 2012). In total, 51% of all maize genes are nonsyntenic (Fig. 4A). Among all genes expressed under control or water deficit conditions, ~40% were nonsyntenic (Fig. 4A). Thus, nonsyntenic genes were less frequently expressed compared with their genomic proportion. Similar observations were made in primary root tissues, in which only 36% of all expressed genes were of nonsyntenic descent (Paschold et al., 2014). However, nonsyntenic genes were significantly overrepresented among SPE genes (Fig. 4A) and made up 74% and 75% under control and water deficit conditions, respectively. A similar overrepresentation of SPE patterns also has been observed in a transcriptome study of root tissues (Paschold et al., 2014). An even higher proportion of genes with stable SPE expression patterns under control and water deficit conditions were nonsyntenic (78%) compared with SPE genes displaying treatment-dependent expression changes (59%). The overrepresentation and stability of nonsyntenic SPE genes might be an adaptation to cope with selective fluctuating environments such as water deficit.

As expected, classical maize genes, among which syntenic genes are enriched (Schnable and Freeling, 2011; Paschold et al., 2014), were significantly underrepresented among SPE genes in this study (Fig. 4B).

Downloaded from https://academic.oup.com/plphys/article/173/2/1247/6115918 by ulbbonn user on 21 December 2021

CONCLUSION

In summary, hundreds of nonadditive, allelic, and SPE patterns were observed under control and water deficit conditions in maize primary roots of the inbred lines B73 and Mo17 and their hybrid progeny. The majority of SPE genes were nonsyntenic and stable throughout changing water availability, suggesting a role of these SPE patterns in the developmental manifestation of heterosis in maize under fluctuating environmental conditions.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Seeds of the inbred maize (*Zea mays*) lines B73 and Mo17 and the two reciprocal hybrids B73xMo17 and Mo17xB73 were germinated in paper rolls in distilled water as described previously (Hoecker et al., 2006). Four to 5 d after germination in a 16-h-light (28°C) and 8-h-dark (21°C) regime, seedlings with a primary root length of 2 to 4 cm were transferred to new paper rolls and incubated in a PEG solution (PEG8000; M_r of 7,300–9,000; Roth) with a water potential of −0.8 MPa or in distilled water as a control. The water potential of the PEG8000 solution was determined as described previously (Opitz et al., 2014). The seedlings were incubated in aerated PEG solution or distilled water for 6 h. Subsequently, per biological replicate, 10 primary roots were harvested, immediately frozen in liquid nitrogen, and stored at −80°C until RNA isolation. Four biological replicates per control and PEG treatment were analyzed for each of the 32 samples.

RNA Isolation and Sequencing Library Construction

RNA was isolated from pools of primary roots pulverized in liquid nitrogen according to an established protocol (Winz and Baldwin, 2001). RNA quality was determined by agarose gel electrophoresis and by a Bioanalyzer using an Agilent RNA 6000 Nano Chip (Agilent Technologies). All RNA samples were of excellent quality with RNA integrity number (Schroeder et al., 2006) values of 9.1 or greater. cDNA libraries for Illumina sequencing were constructed according to the protocol of the manufacturer (TruSeq RNA Sample Preparation; Illumina). For sequencing, four libraries were pooled in one lane of a flow cell. One of the adapters AR001, AR008, AR010, or AR011 (Supplemental Table S1) indexed each library per lane. The indexed libraries were loaded onto a flow cell according to a split-plot design (https://www.vsni.co.uk/software/cycdesign/) assigning conditions (water deficit stress versus control) to lanes (main plots), with two adjacent lanes forming a complete block and assignments of genotypes to lanes regarded as subplots. Cluster preparation and single-read sequencing (100 bp) were performed according to the manufacturer's instructions (HiSeq 2000; Illumina).

Analysis of Sequencing Data and Read Mapping

Raw sequencing reads generated by the Illumina HiSeq 2000 system were processed and quality trimmed with SHORE (http://1001genomes.org/software/shore.html). Reads with more than two or five bases having quality scores of 3 or less in the first 12 or 25 bases, respectively, were rejected. Bases with quality scores of 5 or less at the 3′ end were trimmed until two succeeding bases had higher quality scores. Reads with two or more mismatches in adapter sequences were excluded. Only reads of 40 bp or greater were retained for subsequent analyses. Finally, adapter sequences were removed. Resulting reads had a length of 40 to 100 bp (60% or greater of all reads were 100 bp long) and quality scores of 26 or higher at all base positions.

Read mapping was performed with CLC Genomics Workbench (version 7.0; http://www.clcbio.com/products/clc-genomics-workbench/). All high-quality reads were mapped to the maize B73 reference genome sequence version 2 (RefGen_v2; http://ftp.maizesequence.org/current/; Schnable et al., 2009), allowing gaps of up to 50 kb to span introns. At least 75% of each read had to fit with 90% identity to the reference sequence to be included in the mapping. Stacked reads (i.e. redundant reads sharing the same start and end coordinates), sequencing direction, and sequence were merged into one. The remaining reads were projected to the filtered gene set (FGSv2; http://ftp.maizesequence.org/current/; release 5b.60) of the B73 reference genome, allowing a maximum of two mismatches for reads of 56 bp or less. Longer reads had to fit at least with 80% of their length and an identity of 90% or greater. Only those reads uniquely mapping to the reference data set were subsequently used for analyses.

An MDS plot was produced by the Bioconductor package limma (Smyth, 2005) in R (version 3.1.1 2014-07-10) using normalized expression values on the $\log_2$ scale, resulting from the function voom. The MDS plot illustrates and compares sample relationships.

Statistical Procedure to Determine Gene Activity or Inactivity

After excluding genes with an average of less than one read per replicate of each genotype-by-treatment combination, 29,986 genes remained for analysis (Supplemental Data Set S1). As described by Opitz et al. (2016), a generalized linear mixed model with a negative binomial response was used to model gene-wise read counts. The negative binomial dispersion parameter was assumed to be constant within each gene. Using a log link function, the linear predictor was taken to be the sum of a normalization factor, a linear function of random effects that account for experimental design, and a linear function of fixed effects specific to each combination of genotype and condition. Each normalization factor was calculated by adding the log of the trimmed mean of M values normalization factor (Robinson and Oshlack, 2010) for each sample to a term accounting for gene length. The data were considered as coming from a split-plot design. The whole-plot part is organized as a randomized complete block design with four blocks, with the sequencing lane acting as the whole-plot experimental unit and condition as the whole-plot factor. Finally, the genotype is the split-plot factor and the unique Illumina adapter used for multiplexing is an additional blocking factor. The principles of a Latin square were used to match each adapter with a genotype within each block. Correspondingly, the vector of random effects for each gene consists of a random effect for each lane, a random effect for each block, and a random effect for each adapter. The log of the negative binomial dispersion parameter and vectors of fixed and random effects combine to form a vector that was assumed to have a multivariate Gaussian distribution. Each fixed effect and the log of the negative binomial dispersion were assumed to be independent of other parameters and assigned a normal distribution with unknown mean and variance. The precisions of the random effects were assigned γ-distributions, and the parameters for the effects corresponding to block and adapter were specified to create vague distributions. The unknown hyperparameters in the distributions of the fixed effects, negative binomial dispersion, and precision of the lane effects were estimated using an empirical Bayes procedure via the R package ShrinkBayes (Van De Wiel et al., 2013). Integrated nested Laplace approximation (Rue et al., 2009) was used to approximate $P_{ijk}(T)$, the posterior probability that the linear combination of fixed effects for gene i, genotype j, and condition k was larger than a selected threshold represented by T. Gene i was called active for genotype j and condition k if $P_{ijk}(T) > 0.5$ and inactive otherwise. Using posterior distributions to classify genes as active or inactive has the advantage of accounting for sequencing differences from sample to sample, accounting for gene length differences, and borrowing information across genes via the empirical Bayes procedure. For these data, the selected threshold T resulted in calling a gene active if the expected number of reads per thousand bases of gene length was three or more for a sample with trimmed mean of M normalization factor equal to 1.

Statistical Procedures for Analyzing Differential Allele-Specific Expression and Differential and Nonadditive Gene Expression

The maize reference genome sequence (RefGen_v2; Schnable et al., 2009) was indexed using Bowtie2 (http://bowtie-bio.sourceforge.net/bowtie2/index.shtml, version 2.2.2), and raw sequence reads were mapped to the RefGen_v2 with TopHat2 (http://ccb.jhu.edu/software/tophat/manual.shtml, version 2.1.0). Transcript assembly and the quantification of aligned reads were performed with Cufflinks (http://cole-trapnell-lab.github.io/cufflinks/, version 2.2.1), and assembled transcripts uniquely projected onto the FGS_v2 (http://ftp.maizesequence.org/current/, release 5b.60) were extracted. For calculation of gene expression levels, the R package cummeRbund (http://compbio.mit.edu/cummeRbund/index.html, version 2.7.2) was used.

Aligned reads in Pileup format were generated from the TopHat2 alignments using samtools (http://samtools.sourceforge.net/, version 0.1.19). To estimate differences in allele-specific expression, a set of 106,335 SNPs identified between the B73 and Mo17 genotypes were extracted as described (Xin et al., 2013). B73-Mo17 SNPs were checked against the TopHat2 alignments using a customized

Downloaded from https://academic.oup.com/plphys/article/173/2/1247/6115918 by ulbbonn user on 21 December 2021

Perl script (http://cropbio.uni-bonn.de). Only SNPs with a coverage of at least five in at least one sample of each parent, B73 and Mo17, that further showed the expected allele in at least 80% of reads in all parental samples with a coverage of at least five, were used for subsequent analysis. SNPs overlapping with the FGSv2 were extracted using bedtools intersect (https://github.com/arq5x/bedtools2/, version 2.19.1). Allele counts per gene were summed across all SNPs within exons with a coverage of at least 10 reads using a customized Perl script (http://cropbio.uni-bonn.de; Supplemental Data Set S2).

To determine the number of genes in hybrids that exhibit allelic ratios deviating from their parental ratios, genes represented by a minimum of five mapped allele-specific reads in all four replicates of at least one genotype/treatment combination were selected. This filtering step was applied separately to the control and PEG-treated samples. The raw allele-specific read counts were RPKM normalized by sequencing depth and $\log_2$ transformed to meet the assumptions of linear models. Furthermore, the mean variance relationship within the count data were estimated, and precision weights for each observation were computed using limma in R (version 3.16.7; Law et al., 2014). To test the hypothesis that the allelic expression ratio of the parental alleles in the hybrids is equal to the allelic expression ratio in the parents, this test was performed separately for both reciprocal hybrids and each treatment. The resulting *P* values were FDR adjusted (Benjamini and Hochberg, 1995).

To identify differential and nonadditive gene expression, genes represented by a minimum of five mapped reads in all four replicates of at least one genotype/treatment combination were selected (Supplemental Data Set S3). An empirical Bayes approach was used in which the mean variance trend for $\log_2$ counts was estimated and then a weight was assigned to each observation based on its predicted variance. The weights were then used in the linear modeling process to adjust for heteroscedasticity (Law, 2013). To borrow strength across genes in the estimation of the residual error variance, an empirical Bayes approach (Smyth, 2004) implemented in the Bioconductor package limma in R were used. Applying a moderated *t* statistic (Smyth, 2004) makes use of all estimated sample variances, which are shrunk to a pooled estimate. Especially in cases where rather few replicates are available, inference is far more stable in comparison with the ordinary *t* statistic (Smyth, 2004). We fit the linear model

$$\log_2\left(y_{jkbl} + 0.5\right) = \alpha_{jk} + \beta_b + \gamma_{bl} + e_{jkbl}$$

for each feature, where y_{jkbl} is the RPKM normalized read count on the RNA level of the *j*th genotype and *k*th treatment in the *bl*th lane and *b*th block, α_{jk} is the fixed effect of the *j*th genotype for the *k*th treatment, β_b is the fixed effect of the *b*th block, γ_{bl} is the random effect of the *l*th lane in the *b*th block, assumed to be independently and identically distributed as $N(0,\sigma^2_l)$, and e_{jkbl} is the corresponding residual, assumed to be independently and identically distributed as $N(0,\sigma^2)$. Pairwise contrasts were calculated, and resulting *P* values of each contrast were corrected for multiplicity using an FDR approach (Benjamini and Yekutieli, 2001). This correction extends the method of Benjamini and Hochberg (1995) by controlling the FDR for general dependency structures.

Accession Numbers

Raw sequencing data are stored at the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) with the accession number SRP058750.

Supplemental Data

The following supplemental materials are available.

Supplemental Table S1. Overview of sample distribution within the flow cell, biological replication, RNA-seq output, and mapping results.

Supplemental Table S2. Genes whose allelic expression levels in hybrids differ from the expected values after control and water deficit treatments.

Supplemental Table S3. Genotype- and treatment-specific activities of expressed genes.

Supplemental Table S4. PAV in SPE genes could explain between 2.5% and 15% of SPE_B expression patterns.

Supplemental Data Set S1. List of active and inactive genes by genotype and treatment.

Supplemental Data Set S2. List of allele counts per gene.

Supplemental Data Set S3. Overview of the number of mapped reads of 26,737 expressed genes, represented by at least five reads per replicate in at least one genotype/treatment combination.

Received July 5, 2016; accepted December 18, 2016; published December 20, 2016.

LITERATURE CITED

Araus JL, Sánchez C, Cabrera-Bosquet L (2010) Is heterosis in maize mediated through better water use? New Phytol **187:** 392–406

Baldauf JA, Marcon C, Paschold A, Hochholdinger F (2016) Nonsyntenic genes drive tissue-specific dynamics of differential, nonadditive, and allelic expression patterns in maize hybrids. Plant Physiol **171:** 1144–1155

Barbaglia AM, Klusman KM, Higgins J, Shaw JR, Hannah LC, Lal SK (2012) Gene capture by Helitron transposons reshuffles the transcriptome of maize. Genetics **190:** 965–975

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol **57:** 289–300

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat **29:** 1165–1188

Birchler JA, Auger DL, Riddle NC (2003) In search of the molecular basis of heterosis. Plant Cell **15:** 2236–2239

Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell **16:** 1667–1678

Boyer JS (1982) Plant productivity and environment. Science **218:** 443–448

de Oliveira AB, Alencar NLM, Gomes-Filho E (2013) Comparison between the water and salt stress effects on plant growth and development. InTechOpen **4:** 67–94

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics, Ed 4. Longman Harlow, Essex, UK

Freeling M, Lyons E, Pedersen B, Alam M, Ming R, Lisch D (2008) Many or most genes in *Arabidopsis* transposed after the origin of the order Brassicales. Genome Res **18:** 1924–1937

Groszmann M, Greaves IK, Albertyn ZI, Scofield GN, Peacock WJ, Dennis ES (2011) Changes in 24-nt siRNA levels in *Arabidopsis* hybrids suggest an epigenetic contribution to hybrid vigor. Proc Natl Acad Sci USA **108:** 2617–2622

Guo M, Rupe MA, Zinselmeier C, Habben J, Bowen BA, Smith OS (2004) Allelic variation of gene expression in maize hybrids. Plant Cell **16:** 1707–1716

Haberer G, Young S, Bharti AK, Gundlach H, Raymond C, Fuks G, Butler E, Wing RA, Rounsley S, Birren B, et al (2005) Structure and architecture of the maize genome. Plant Physiol **139:** 1612–1624

Hansey CN, Vaillancourt B, Sekhon RS, de Leon N, Kaeppler SM, Buell CR (2012) Maize (*Zea mays* L.) genome diversity as revealed by RNA-sequencing. PLoS ONE **7:** e33071

Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, Peñagaricano F, Lindquist E, Pedraza MA, Barry K, et al (2014) Insights into the maize pan-genome and pan-transcriptome. Plant Cell **26:** 121–135

Hochholdinger F, Hoecker N (2007) Towards the molecular basis of heterosis. Trends Plant Sci **12:** 427–432

Hoecker N, Keller B, Muthreich N, Chollet D, Descombes P, Piepho HP, Hochholdinger F (2008) Comparison of maize (*Zea mays* L.) F_1-hybrid and parental inbred line primary root transcriptomes suggests organ-specific patterns of nonadditive gene expression and conserved expression trends. Genetics **179:** 1275–1283

Hoecker N, Keller B, Piepho HP, Hochholdinger F (2006) Manifestation of heterosis during early maize (*Zea mays* L.) root development. Theor Appl Genet **112:** 421–429

Humbert S, Subedi S, Cohn J, Zeng B, Bi YM, Chen X, Zhu T, McNicholas PD, Rothstein SJ (2013) Genome-wide expression profiling of maize in response to individual and combined water and nitrogen stresses. BMC Genomics **14:** 3

Kato A, Lamb JC, Birchler JA (2004) Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. Proc Natl Acad Sci USA **101:** 13554–13559

Kersting AR, Bornberg-Bauer E, Moore AD, Grath S (2012) Dynamics and adaptive benefits of protein domain emergence and arrangements during plant genome evolution. Genome Biol Evol **4:** 316–329

Downloaded from https://academic.oup.com/plphys/article/173/2/1247/6115918 by ulbbonn user on 21 December 2021

Krasensky J, Jonak C (2012) Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. J Exp Bot **63:** 1593–1608

Laurie DA, Bennett MD (1985) Nuclear-DNA content in the genera *Zea* and *Sorghum*: intergenetic, interspecific and intraspecific variation. Heredity **55:** 307–313

Law C (2013) Precision weights for gene expression analysis. PhD thesis. University of Melbourne, Melbourne, Australia

Law CW, Chen Y, Shi W, Smyth GK (2014) voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol **15:** R29

Lorenz A, Hoegemeyer T (2013) The phylogenetic relationships of US maize germplasm. Nat Genet **45:** 844–845

Marcon C, Lamkemeyer T, Malik WA, Ungrue D, Piepho HP, Hochholdinger F (2013) Heterosis-associated proteome analyses of maize (*Zea mays* L.) seminal roots by quantitative label-free LC-MS. J Proteomics **93:** 295–302

Marroni F, Pinosio S, Morgante M (2014) Structural variation and genome complexity: is dispensable really dispensable? Curr Opin Plant Biol **18:** 31–36

O'Geen A (2012) Soil water dynamics. Nature Education Knowledge **3:** 12

Opitz N, Marcon C, Paschold A, Malik WA, Lithio A, Brandt R, Piepho HP, Nettleton D, Hochholdinger F (2016) Extensive tissue-specific transcriptomic plasticity in maize primary roots upon water deficit. J Exp Bot **67:** 1095–1107

Opitz N, Paschold A, Marcon C, Malik WA, Lanz C, Piepho HP, Hochholdinger F (2014) Transcriptomic complexity in young maize primary roots in response to low water potentials. BMC Genomics **15:** 741

Paschold A, Jia Y, Marcon C, Lund S, Larson NB, Yeh CT, Ossowski S, Lanz C, Nettleton D, Schnable PS, et al (2012) Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents. Genome Res **22:** 2445–2454

Paschold A, Larson NB, Marcon C, Schnable JC, Yeh CT, Lanz C, Nettleton D, Piepho HP, Schnable PS, Hochholdinger F (2014) Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. Plant Cell **26:** 3939–3948

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, et al (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature **457:** 551–556

Persson T, Garcia y Garcia A, Paz J, Jones J, Hoogenboom G (2009) Maize ethanol feedstock production and net energy value as affected by climate variability and crop management practices. Agric Syst **100:** 11–21

Piepho HP (2012) A SAS macro for generating letter displays of pairwise mean comparisons. Communications in Biometry and Crop Science **7:** 4–13

Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol **11:** R25

Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. J R Stat Soc Ser B Stat Methodol **71:** 319–392

Schaeffer ML, Harper LC, Gardiner JM, Andorf CM, Campbell DA, Cannon EK, Sen TZ, Lawrence CJ (2011) MaizeGDB: curation and outreach go hand-in-hand. Database (Oxford) **2011:** bar022

Schnable JC, Freeling M (2011) Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS ONE **6:** e17855

Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci USA **108:** 4069–4074

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science **326:** 1112–1115

Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T (2006) The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol **7:** 3

Sharp RE, Silk WK, Hsiao TC (1988) Growth of the maize primary root at low water potentials. I. Spatial distribution of expansive growth. Plant Physiol **87:** 50–57

Sinclair TR (2011) Challenges in breeding for yield increase for drought. Trends Plant Sci **16:** 289–293

Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol **3:** Article3

Smyth GK (2005) Limma: linear models for microarray data. *In* W Gentleman, R Carey, V Dudoit, S Irizarry, R Huber, eds, Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer, New York, pp 397–420

Spollen WG, Tao W, Valliyodan B, Chen K, Hejlek LG, Kim JJ, Lenoble ME, Zhu J, Bohnert HJ, Henderson D, et al (2008) Spatial distribution of transcript changes in the maize primary root elongation zone at low water potential. BMC Plant Biol **8:** 32

Springer NM, Ying K, Fu Y, Ji T, Yeh CT, Jia Y, Wu W, Richmond T, Kitzman J, Rosenbaum H, et al (2009) Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet **5:** e1000734

Stupar RM, Springer NM (2006) *Cis*-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F_1 hybrid. Genetics **173:** 2199–2210

Swanson-Wagner RA, Eichten SR, Kumari S, Tiffin P, Stein JC, Ware D, Springer NM (2010) Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res **20:** 1689–1699

Swanson-Wagner RA, Jia Y, DeCook R, Borsuk LA, Nettleton D, Schnable PS (2006) All possible modes of gene action are observed in a global comparison of gene expression in a maize F_1 hybrid and its inbred parents. Proc Natl Acad Sci USA **103:** 6805–6810

Swigonová Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J (2004) Close split of sorghum and maize genome progenitors. Genome Res **14:** 1916–1923

Tester M, Langridge P (2010) Breeding technologies to increase crop production in a changing world. Science **327:** 818–822

Uga Y, Sugimoto K, Ogawa S, Rane J, Ishitani M, Hara N, Kitomi Y, Inukai Y, Ono K, Kanno N, et al (2013) Control of root system architecture by DEEPER ROOTING 1 increases rice yield under drought conditions. Nat Genet **45:** 1097–1102

Van De Wiel MA, Leday GG, Pardo L, Rue H, Van Der Vaart AW, Van Wieringen WN (2013) Bayesian analysis of RNA sequencing data by estimating multiple shrinkage priors. Biostatistics **14:** 113–128

Vuylsteke M, van Eeuwijk F, Van Hummelen P, Kuiper M, Zabeau M (2005) Genetic analysis of variation in gene expression in *Arabidopsis thaliana*. Genetics **171:** 1267–1275

Winz RA, Baldwin IT (2001) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. IV. Insect-induced ethylene reduces jasmonate-induced nicotine accumulation by regulating putrescine *N*-methyltransferase transcripts. Plant Physiol **125:** 2189–2202

Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. PLoS Biol **8:** e1000409

Xin M, Yang R, Li G, Chen H, Laurie J, Ma C, Wang D, Yao Y, Larkins BA, Sun Q, et al (2013) Dynamic expression of imprinted genes associates with maternally controlled nutrient allocation during maize endosperm development. Plant Cell **25:** 3212–3227

Yamaguchi M, Sharp RE (2010) Complexity and coordination of root growth at low water potentials: recent advances from transcriptomic and proteomic analyses. Plant Cell Environ **33:** 590–603

Zheng J, Zhao J, Tao Y, Wang J, Liu Y, Fu J, Jin Y, Gao P, Zhang J, Bai Y, et al (2004) Isolation and analysis of water stress induced genes in maize seedlings by subtractive PCR and cDNA macroarray. Plant Mol Biol **55:** 807–823

Downloaded from https://academic.oup.com/plphys/article/173/2/1247/6115918 by ulbbonn user on 21 December 2021

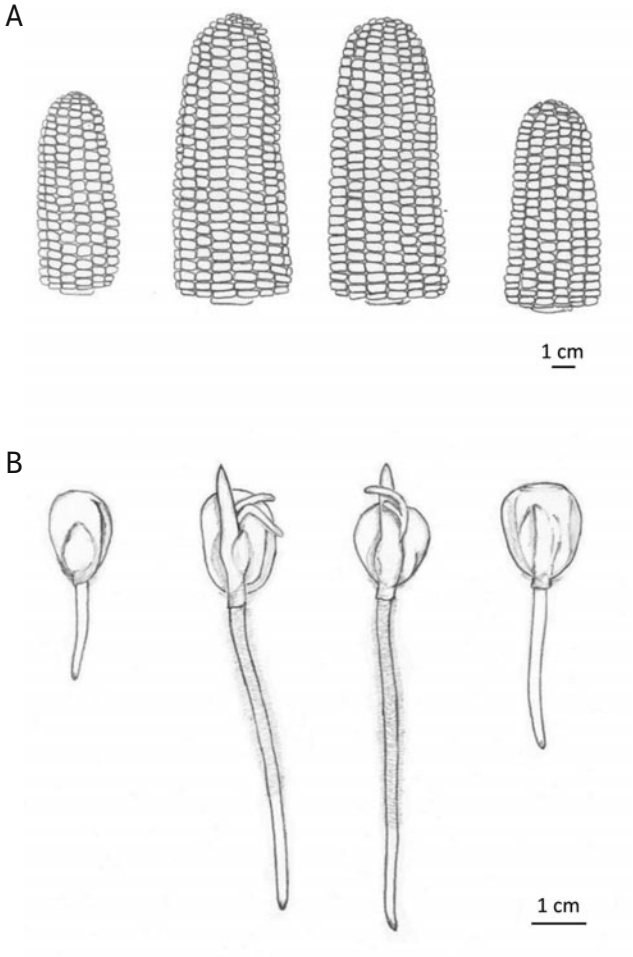
Heterosis

Komplementation von Genexpressionsmustern in Maishybriden

CAROLINE MARCON, JUTTA A. BALDAUF, FRANK HOCHHOLDINGER
INRES – INSTITUT FÜR NUTZPFLANZENWISSENSCHAFTEN UND RESSOURCENSCHUTZ, CROP FUNCTIONAL GENOMICS, UNIVERSITÄT BONN

Heterosis describes the superior performance of F_1-hybrids compared to their genetically distinct parents, a phenomenon which is extensively exploited in maize breeding. To better understand the molecular basis of heterosis in maize, we analyzed how the genetic diversity of parental inbred lines affects the root transcriptome of their F_1-progeny. We discovered complementation of activity in evolutionary younger genes and as a consequence more active genes in hybrids than in parental inbred lines.

DOI: 10.1007/s12268-016-0737-5
© Springer-Verlag 2016



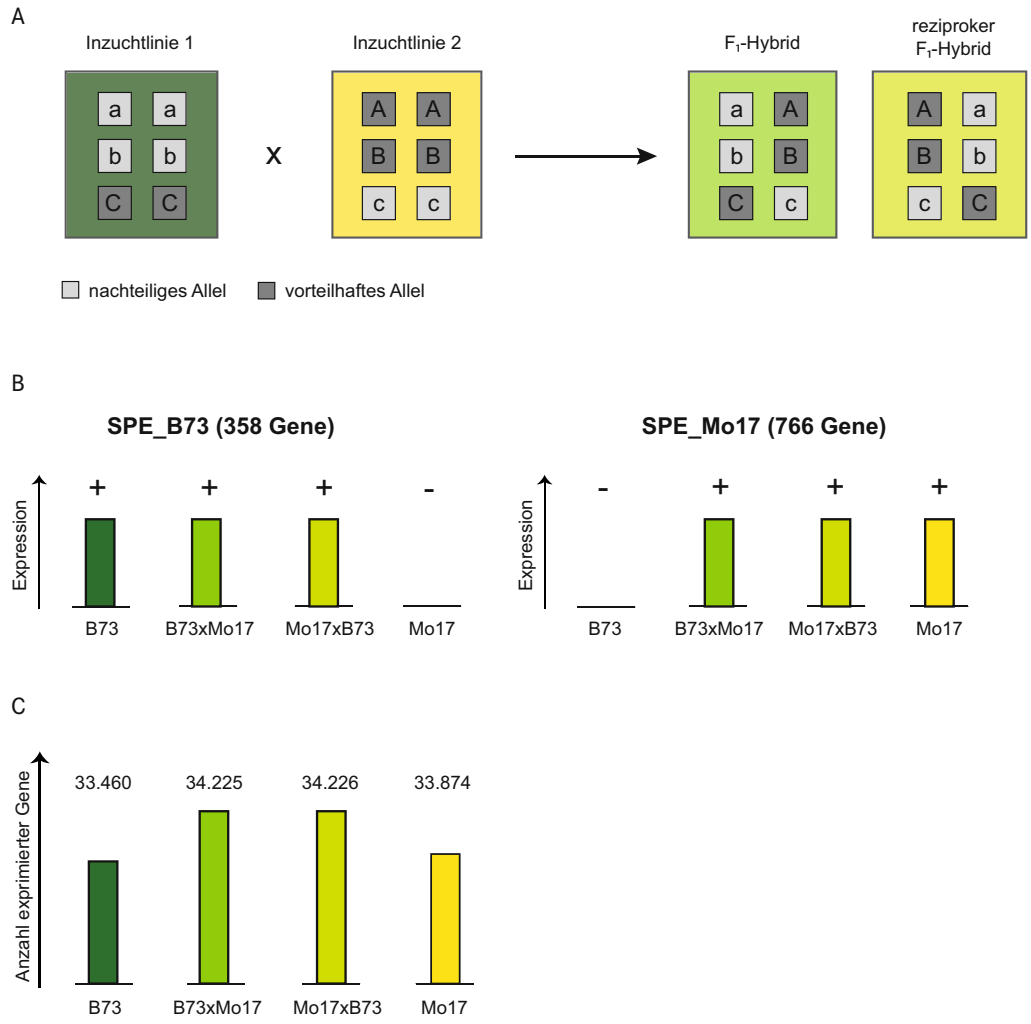
◀ **Abb. 1:** Heterosis wird zu verschiedenen Phasen der Pflanzenentwicklung in verschiedenen Merkmalen wie z. B. Kolbengröße (**A**) oder Primärwurzellänge (**B**) manifestiert. Heterosis bezeichnet das Phänomen, dass die Leistung der F_1-Nachkommen über dem Durchschnitt der elterlichen Werte eines Merkmals liegt (Zeichnungen: Jutta A. Baldauf).

■ Maishybride sind im Vergleich zum Durchschnitt ihrer genetisch unterschiedlichen und reinerbigen Eltern (Inzuchtlinien) deutlich leistungsfähiger, was sich z. B. in der Kolbengröße zeigt (**Abb. 1A**). Dieses als Heterosis bezeichnete Phänomen wurde bereits im Jahr 1908 von George H. Shull [1] und Edward M. East [2] beschrieben und dient als Grundlage der modernen Maiszüchtung. Der Heterosiseffekt ist von überragender wirtschaftlicher Bedeutung, was sich darin zeigt, dass heute ausschließlich Maishybridsorten angebaut werden. Diese sind im Vergleich zu offen bestäubten Sorten ertragreicher und toleranter gegenüber nachteiligen Umweltbedingungen. Die Erfolge der Hybridzüchtung haben in den letzten 50 Jahren maßgeblich zur Verfünffachung der weltweiten Maisproduktion auf heute etwa eine Milliarde Tonnen Körnermais pro Jahr beigetragen [3].

In jungen Maishybriden zeigt die Primärwurzel bereits vier Tage nach Keimung einen deutlichen Wachstumsvorsprung gegenüber den Elternlinien (**Abb. 1B**). Die einfach strukturierte Primärwurzel eignet sich daher besonders gut als Modell zur Erforschung des Heterosiseffekts in der frühen Entwicklung.

In Maishybriden sind mehr Gene aktiv als in ihren Elternlinien

Trotz ihrer überragenden wirtschaftlichen Bedeutung ist über die genetischen und molekularen Mechanismen, die zur Überlegenheit der Hybriden führen, wenig bekannt. Verschiedene Hypothesen postulieren, dass genetische Interaktionen für die Ausprägung von Heterosis verantwortlich sind. Nach der Dominanzhypothese komplementieren vorteilhafte Allele des einen Elters in Hybriden die Funktion des nachteiligen Allels des zweiten Elters und führen so zu einer Leistungsüberlegenheit der Hybriden (**Abb. 2A**, [4]). Phänotypische Merkmalsunterschiede verschiedener Genotypen, wie im Fall der Heterosis beobachtet, sind die Folge von Unterschieden in der Genexpression [5]. Eine spezielle Form der Komplementation auf Ebe-



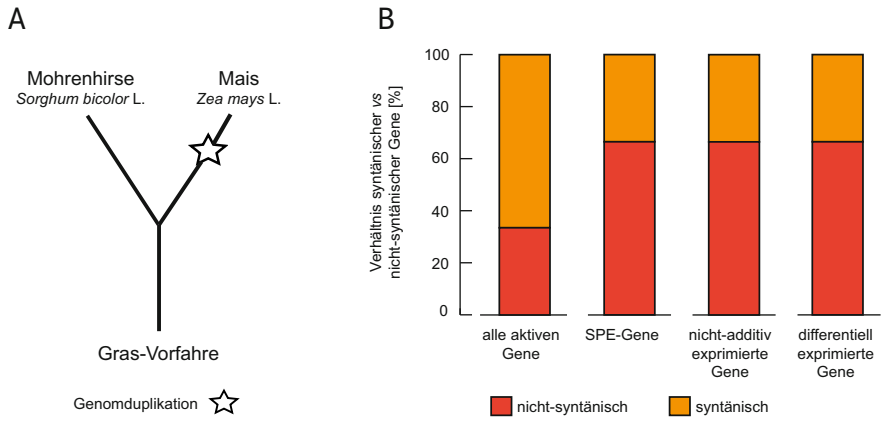
◀ **Abb. 2:** Komplementation in Maishybriden. **A**, Auf Ebene des Genoms werden nachteilige Allele aus den beiden homozygoten Inzuchtlinien in Hybride durch vorteilhafte Allele komplementiert, was exemplarisch an drei Genen (A bis C) dargestellt ist. Reziproke Hybriden enthalten dabei das identische Kerngenom, auch wenn die jeweils vor- und nachteiligen Allele von unterschiedlichen Eltern stammen. **B**, Im Transkriptom von Maiswurzeln zeigten mehrere Hundert Gene *single parent expression* (SPE). Diese Gene werden nur in einem Elter, aber in beiden Hybriden exprimiert. Dieses Muster ist ein extremes Beispiel für Komplementation auf Genexpressionsebene. **C**, Als Folge der SPE-Muster (B) sind in den reziproken Hybriden B73×Mo17 (34.225 aktive Gene) und Mo17×B73 (34.226) mehr Gene aktiv als in den Eltern B73 (33.460) und Mo17 (33.874).

ne der Genexpression liegt vor, wenn einzelne Gene jeweils nur in einem der beiden Eltern exprimiert werden, diese Gene aber in den durch die Kreuzung der Eltern entstandenen Hybriden aktiv sind. Solche Muster konnten wir im Transkriptom, also der Gesamtheit aller exprimierten Gene, von Primärwurzeln zweier genetisch unterschiedlicher Maisinzuchtlinien (B73 und Mo17) und deren reziproken Hybriden (B73×Mo17 und Mo17×B73) mittels RNA-Sequenzierung (RNA-Seq) für Hunderte von Genen nachweisen (**Abb. 2B**, [6]). Dieses Phänomen haben wir als *single parent expression* (SPE) bezeichnet. Eine Folge dieser Komplementation auf Genexpressionsebene ist, dass in Maishybriden Hunderte Gene mehr aktiv sind als in ihren Elternlinien (**Abb. 2C**). SPE-Gene könnten durch die Komplementation ihrer Expression zur überlegenen Leistungsfähigkeit der Hybriden beitragen. Zwar leisten die zusätzlich aktiven Gene im Einzelnen vermutlich nur einen geringen Beitrag zum Heterosiseffekt, könnten aber in ihrer Summe eine wichtige Rolle bei dessen Ausprägung spielen.

Komplementation von Genexpression wird vor allem in evolutionär jungen Genen beobachtet

Mais und Mohrenhirse (*Sorghum bicolor* L.) sind evolutionär nahe Verwandte innerhalb der Familie der Gräser (**Abb. 3A**). Ein Vergleich der Genome von Mohrenhirse und Mais zeigte, dass sich das Maisgenom nach Auftrennung dieser beiden Arten vor etwa zwölf bis fünf Millionen Jahren dupliziert hat [7] (**Abb. 3A**). Dadurch lagen direkt nach dieser Genomduplikation alle Maisgene im Vergleich zu den homologen Genen in Mohrenhirse in doppelter Ausführung vor [8]. Diese homologen Gene zwischen Mohrenhirse und Mais werden als syntänisch bezeichnet. Im Laufe der Evolution von Mais ist in vielen Fällen eine oder es sind auch beide Genkopien wieder verloren gegangen. In diesem Zeitraum sind andererseits durch die Duplikation einzelner Gene zusätzlich neue Maisgene entstanden [9]. Für die so entstandenen neuen Gene, finden sich keine Homologe im Genom der Mohrenhirse. Diese Gene werden daher als nicht-syntänisch bezeichnet. Das heutige Maisgenom besteht jeweils zu etwa 50 Prozent aus syntänischen und nicht-syntänischen Genen.

Um den evolutionären Ursprung von Genen zu studieren, die speziellen Expressionsmustern in Hybriden zuzuordnen sind, haben wir neben den SPE-Mustern auch Tausende differenziell und Hunderte nicht-additiv exprimierte Gene in Primärwurzeln von B73×Mo17 und Mo17×B73 untersucht [10, 11]. Als differenziell exprimierte Gene werden solche Gene bezeichnet, deren Expressionswert sich in Hybriden signifikant von mindestens einer der beiden Elternlinien unterscheidet. Nicht-additiv exprimierte Gene weisen dagegen einen Expressionswert auf, der signifikant von dem gemittelten Expressionswert beider Eltern abweicht [5]. Um einen Einblick in den evolutionären Ursprung der Gene dieser Expressionsklassen zu gewinnen, wurde



▲ **Abb. 3:** Evolutionärer Ursprung der Gene verschiedener Expressionsklassen. **A**, Vor zwölf bis fünf Millionen Jahren fand die letzte Duplikation des Maisgenoms statt. Aus dieser Duplikation hervorgegangene Gene sind mit Genen der nahe verwandten Mohrenhirse syntänisch. Danach entstanden durch Duplikationen einzelner Gene weitere Maisgene. Diese sind nicht-syntänisch zu Genen der Hirse. **B**, In verschiedenen Expressionsklassen von Maiswurzeln (Details siehe Text) sind jüngere, nicht-syntänische Gene (rot) im Vergleich zu ihrem Anteil an allen aktiven Genen deutlich überrepräsentiert [9, 10].

zunächst für alle in der Primärwurzel exprimierten Gene bestimmt, ob sie syntänischen oder nicht-syntänischen Ursprungs sind. Bei diesem Vergleich zeigte sich, dass ein Drittel aller aktiven Gene nicht-syntänisch sind [10, 11]. Innerhalb der Klassen der differenziellen, nicht-additiven und SPE-Gene ist der Anteil der nicht-syntänischen Gene signifikant höher und macht etwa zwei Drittel dieser Gene aus (**Abb. 3B**).

Die signifikante Überrepräsentation nicht-syntänischer Gene in den untersuchten Expressionsklassen und die Beobachtung, dass die Hybriden B73×Mo17 und Mo17×B73 mehr Gene als ihre Eltern exprimieren, könnten auf eine Rolle dieser Gene bei der frühen Ausprägung von Heterosis in Keimlingswurzeln hinweisen, die in weiteren Studien genauer untersucht werden soll. ■

Literatur

[1] Shull GH (1908) The composition of a field of maize. Rep Am Breed Assoc 5:51–59
[2] East EM (1908) Inbreeding in corn. Conn Agr Exp Sta Rpt 1907:419–428
[3] Food and Agriculture Organization of the United Nations, Statistic Division, http://faostat3.fao.org
[4] Jones DF (1917) Dominance of linked factors as a means of accounting for heterosis. Genetics 2:466–479
[5] Hochholdinger F, Hoecker N (2007) Towards the molecular basis of heterosis. Trends Plant Sci 12:427–432
[6] Paschold A, Jia Y, Marcon C et al. (2012) Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents. Genome Res 22:2445–2454
[7] Swigonová Z, Lai JS, Ma JX et al. (2004) Close split of sorghum and maize genome progenitors. Genome Res 14:1916–1923
[8] Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci USA 108:4069–4074
[9] Woodhouse MR, Schnable JC, Pedersen BS et al. (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. PLoS Biol 8:e1000409
[10] Paschold A, Larson NB, Marcon C et al. (2014) Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. Plant Cell 26:3939–3948
[11] Baldauf J, Marcon C, Paschold A et al. (2016) Nonsyntenic genes drive tissue-specific dynamics of differential, nonadditive and allelic expression patterns in maize hybrids. Plant Physiol 171:1144–1155

Korrespondenzadresse:
Prof. Dr. Frank Hochholdinger
INRES – Institut für Nutzpflanzenwissenschaften und Ressourcenschutz
Crop Functional Genomics
Universität Bonn
Friedrich-Ebert-Allee 144
D-53113 Bonn
Tel.: 0228-73-60334
Fax: 0228-73-60333
hochhold@uni-bonn.de
www.cfg.uni-bonn.de

AUTOREN

Caroline Marcon
2001–2007 Biologiestudium (Diplom) an der Universität Tübingen, dort **2012** Promotion. Seit **2012** wissenschaftliche Assistentin an der Universität Bonn.

Jutta A. Baldauf
2008–2011 Bachelorstudium in Agrarbiologie an der Universität Hohenheim, dort **2011–2014** Masterstudium in Crop Sciences. Seit **2014** Promotionsstudium an der Universität Bonn.

Frank Hochholdinger
1989–1995 Biologiestudium (Diplom) an der Universität Freiburg, dort **1999** Promotion. **1999–2001** Postdoktorand an der Iowa State University, USA. **2001–2010** Leiter einer unabhängigen Forschungsgruppe am Zentrum für Molekularbiologie der Pflanzen (ZMBP) der Universität Tübingen. Seit **2010** W3-Professur für funktionelle Genomik der Nutzpflanzen (*Crop Functional Genomics*) an der Universität Bonn.

Root Type-Specific Reprogramming of Maize Pericycle Transcriptomes by Local High Nitrate Results in Disparate Lateral Root Branching Patterns[1][OPEN]

Peng Yu, Jutta A. Baldauf, Andrew Lithio, Caroline Marcon, Dan Nettleton, Chunjian Li*, and Frank Hochholdinger*

Department of Plant Nutrition, China Agricultural University, 100193 Beijing, China (P.Y., C.L.); Institute of Crop Science and Resource Conservation, Crop Functional Genomics, University of Bonn, 53113 Bonn, Germany (P.Y., J.A.B., C.M., F.H.); and Department of Statistics, Iowa State University, Ames, Iowa 50011-1210 (A.L., D.N.)

ORCID IDs: 0000-0003-1670-8428 (P.Y.); 0000-0002-7286-0558 (J.A.B.); 0000-0001-6699-9441 (C.M.); 0000-0002-6045-1036 (D.N.); 0000-0001-7492-3949 (C.L.); 0000-0002-5155-0884 (F.H.).

Downloaded from https://academic.oup.com/plphys/article/170/3/1783/6114378 by ulbbonn user on 21 December 2021

The adaptability of root system architecture to unevenly distributed mineral nutrients in soil is a key determinant of plant performance. The molecular mechanisms underlying nitrate dependent plasticity of lateral root branching across the different root types of maize are only poorly understood. In this study, detailed morphological and anatomical analyses together with cell type-specific transcriptome profiling experiments combining laser capture microdissection with RNA-seq were performed to unravel the molecular signatures of lateral root formation in primary, seminal, crown, and brace roots of maize (*Zea mays*) upon local high nitrate stimulation. The four maize root types displayed divergent branching patterns of lateral roots upon local high nitrate stimulation. In particular, brace roots displayed an exceptional architectural plasticity compared to other root types. Transcriptome profiling revealed root type-specific transcriptomic reprogramming of pericycle cells upon local high nitrate stimulation. The alteration of the transcriptomic landscape of brace root pericycle cells in response to local high nitrate stimulation was most significant. Root type-specific transcriptome diversity in response to local high nitrate highlighted differences in the functional adaptability and systemic shoot nitrogen starvation response during development. Integration of morphological, anatomical, and transcriptomic data resulted in a framework underscoring similarity and diversity among root types grown in heterogeneous nitrate environments.

Sessile plants efficiently exploit nutrients and water by postembryonically formed branched lateral roots. In most plant species, lateral roots are initiated from pericycle cells of the central cylinder that undergo asymmetric divisions (Malamy and Benfey, 1997). Xylem and phloem pole pericycle cells differ in morphological (Beeckman et al., 2001) and ultrastructural (Parizot et al., 2008) characteristics and specific protein and gene expression patterns (Vanneste et al., 2005; Moreno-Risueno et al., 2010). Auxin gradients and their essential downstream molecular components have been identified as driving forces of lateral root development (Vanneste et al., 2005; De Smet et al., 2007, 2008; Dubrovsky et al., 2008; Laskowski et al., 2008). Moreover, a tight control of cell cycle progression in xylem-pole pericycle cells is instrumental for lateral root initiation in Arabidopsis (*Arabidopsis thaliana*; Himanen et al., 2002, 2004).

In contrast to Arabidopsis, the mechanisms underlying lateral root positioning in the economically important crop species maize (*Zea mays*) are largely unknown. Lateral root initiation is preceded by auxin response maxima in differentiating xylem cells and in cells surrounding the protophloem vessels in maize (Jansen et al., 2012; Yu et al., 2015a). The maize root system is an ideal model to study organ-specific similarity and diversity because of its spatio-temporal complexity and distinct genetic control of different root types (Hochholdinger and Zimmermann, 2008; Rogers

[1] This research was supported by grants from the National Natural Science Foundation of China (No. 31272232), the State Key Basic Research and Development Plan of China (No. 2013CB127402), the Innovative Group Grant (No.31421092), the Chinese Universities Scientific Fund (No. 2012YJ039), and the Post-graduate Study Abroad Program of China Scholarship Council (No. 201306350120) to P.Y. Root research in F.H.'s laboratory is supported by grants from the Deutsche Forschungsgemeinschaft.

* Address correspondence to lichj@cau.edu.cn or hochholdinger@uni-bonn.de.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Frank Hochholdinger (hochholdinger@uni-bonn.de).

P.Y. and C.L. conceived and designed research; P.Y. performed the experiments; P.Y. and J.A.B. analyzed data; A.L. and D.N. determined gene activity; C.M. helped edit the manuscript; F.H. coordinated the research, provided all the technical support, and participated in data interpretation; P.Y., C.L., and F.H. wrote the article; all authors participated in stimulating discussion and approved the final article.

[OPEN] Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.15.01885

 © 2016 American Society of Plant Biologists. All Rights Reserved.

and Benfey, 2015). The maize root system is formed by multiple root types that are of embryonic (primary and seminal roots) and postembryonic (crown and brace roots) origin (Hochholdinger et al., 2004a). These root types are initiated at different stages of development and all form lateral roots (Hochholdinger et al., 2004a). Understanding root system behavior at the scale of specific root types is therefore of great importance in rational breeding approaches aiming on improving crop stress tolerance, efficiency, and productivity (Rogers and Benfey, 2015).

Pericycle cells deeply embedded inside the root tissue can respond to environmental changes to guide lateral root development, thus shaping the highly branched root system in contact with soil (Malamy, 2005; Atkinson et al., 2014). A critical adaptive behavior of lateral roots is foraging of nonuniformly distributed soil resources in both ecological and agronomic contexts (Malamy, 2005; Ruffel et al., 2011; Guan et al., 2014; Yu et al., 2014a). Nitrogen, the most mobile element in soil, determines shoot biomass and yield (Marschner, 1995). Recent results highlighted the superior role of lateral root branching on nitrate and water capturing in maize (Postma et al., 2014; Zhan et al., 2015), suggesting considerable genetic potential to modify this trait. Most studies in Arabidopsis concluded that local nitrate application exerts either no (Zhang and Forde, 1998) or only a relatively small effect (Linkohr et al., 2002) on lateral root number or density. Recently, the auxin receptor AFB3 and the downstream transcription factors NAC4 and OBP4 were identified as key regulators of a nitrate-responsive gene network activating lateral root initiation by controlling cell cycle progression in the pericycle of Arabidopsis (Vidal et al., 2013).

Transcriptome analyses have provided novel insights in lateral root initiation in maize roots by identifying candidate genes associated with early cell divisions (Woll et al., 2005; Dembinsky et al., 2007; Yu et al., 2015a). To obtain an in-depth understanding of nitrate sensing during lateral root branching, this study surveyed the transcriptomic complexity of pericycle cells in the four major root types of maize in response to heterogeneous nitrate application using laser capture microdissection in combination with RNA-sequencing (LCM-RNAseq). We identified substantial differences between the transcriptomic landscapes of pericycle cells of different maize root types, among which functional enrichment was altered as a consequence of changed external nitrate levels. Intriguingly, pericycle cells displayed brace root specific responses to local nitrate stimuli that included unique development-related processes. Finally, by integrating morphological and anatomical observations with cell-specific transcriptome responses, we provided a comprehensive overview of branching mechanisms among maize root types and their response to heterogeneous nitrate during plant development.

RESULTS

Deep Profiling of Lateral Root Branching Response to Local High Nitrate

Emerged lateral roots in the branching zone were quantified to determine how heterogeneous nitrate availability contributes to root type-specific lateral root formation in maize. Twenty days after local high nitrate treatment, average lateral root length was significantly increased compared to control experiments in all root types (Fig. 1A). The local effect on the promotion of lateral root elongation in brace roots (600%) was more pronounced than that of the other three root types (50–75%; Fig. 1A). A significant increase in the density of emerged lateral roots was exclusively observed in brace roots (75%) but not in the other three root types (Fig. 1B).

Based on histological analyses, early divisions of pericycle cells were quantified to determine whether local high nitrate influences the stages of lateral root initiation at increasing distances from the root tip. It was previously demonstrated that anticlinal and periclinal divisions of pericycle cells were significantly induced in response to local high nitrate in aboveground shoot-borne brace roots (Yu et al., 2015a). In contrast, no significant induction of cell divisions in phloem pole pericycle cells were detected in this study in any region of the three seedling root types, primary, seminal, and crown roots of 80 mm length, by local high nitrate induction 24 h after treatment (Fig. 1C). Therefore, lateral root branching of brace root was clearly distinct from that of primary, seminal, and crown roots in response to heterogeneous nitrate supplies. This result supports the notion that brace roots are important for water and nutrient acquisition.

Anatomical Characteristics of the Four Maize Root Types

To further understand functional differences between the four root types, microscopic analyses were conducted to investigate anatomical traits in transverse sections of roots in the lateral root initiation zone grown in homogeneous low nitrate. Overall, in brace roots, the total transversal area, the stele transversal area, and the total meta-xylem area were much larger than those in the other three root types (Supplemental Table S1). Remarkably, primary and crown roots displayed similar traits in these transverse sections and differed significantly from those of seminal roots (Supplemental Fig. S1 and Supplemental Table S1). Apparently, brace roots had a higher number of cortical cell layers and meta-xylem elements than the other three root types (Supplemental Fig. S1 and Supplemental Table S1). Subsequently, the length and width of pericycle cells adjacent to phloem and xylem poles were quantified to explore the similarity of pericycle cell size among root types. Pericycle cells adjacent to the xylem or phloem poles in brace roots showed an equal size to those in crown roots but were significantly bigger than those of primary and seminal roots (Supplemental Table S1). Together, these data suggest anatomical diversity of the four maize root types.

Downloaded from https://academic.oup.com/plphys/article/170/3/1783/6114378 by ulbbonn user on 21 December 2021

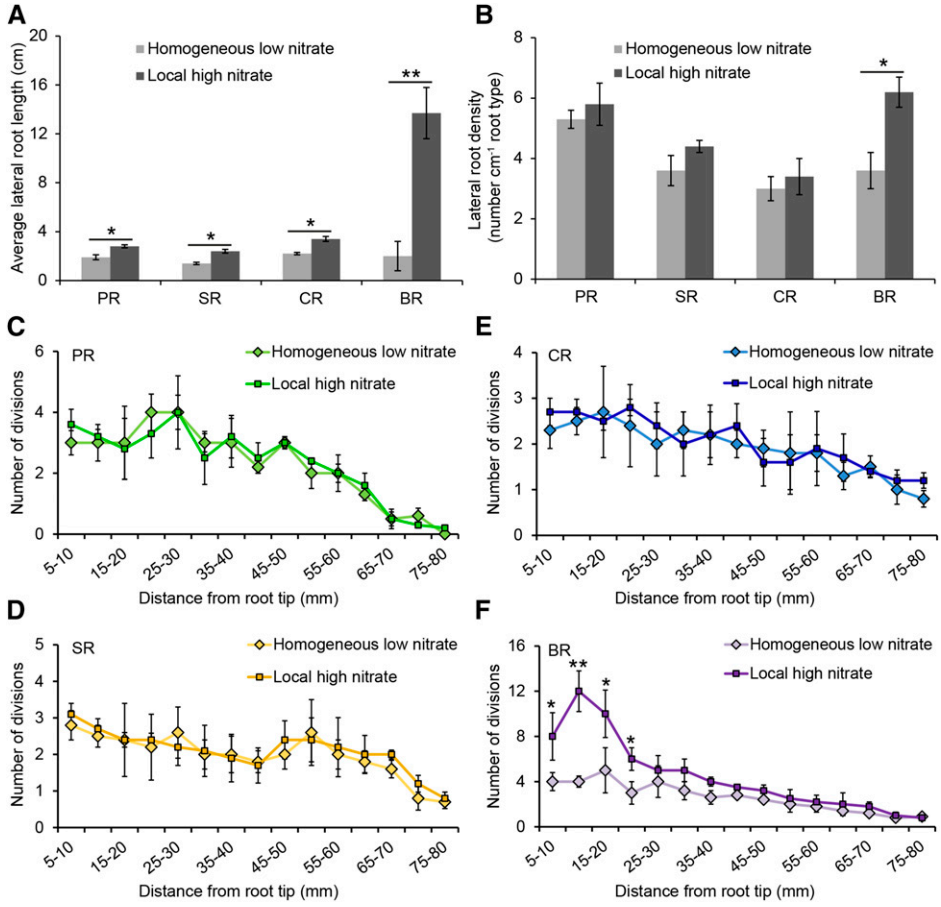


Figure 1. Effects of homogeneous low nitrate or local high nitrate supplies on lateral root development of four maize root types. Seedlings or adult plants were grown in homogeneous low nitrate nutrient solution before each root type was transferred to local high nitrate nutrient solution (see "Materials and Methods"). Roots were scanned and counted after 20 d of growth in these two nitrate conditions. A and B, Average lateral root length (A) and lateral root density (B) in the lateral root branching zone under homogenous low nitrate and local high nitrate conditions. Error bars indicate means ± SE; $n = 4$ (roots from four different maize plants). C to F, Early pericycle cell divisions in primary (C), seminal (D), crown roots (E), and brace roots (F) 24 h after local high nitrate supply. F is taken from Yu et al. (2015a) as a reference for comparisons with the other root types. Asymmetric pericycle cell divisions were tracked from 5 to 80 mm from root tip after Safranin O and Fast Green staining. Error bars indicate means ± SE; $n = 4$ (roots from four different) maize plants. Asterisks denote significant differences according to paired Student's t tests (*$P < 0.05$; **$P < 0.01$). PR, primary root; SR, seminal root; CR, crown root; BR, brace root. F was adapted from Yu et al. (2015a).

RNA Sequencing of Laser Captured Phloem Pole Pericycle Cells from Four Maize Root Types Grown under Different Nitrate Regimes

Laser capture microdissection was adopted to isolate phloem pole pericycle cells from four root types of the maize inbred line B73 upon local high nitrate and control treatment (Fig. 2A). Total RNA was extracted and amplified from three independent biological replicates of these captured pericycle cells per root type and treatment resulting in 24 samples (Fig. 2A). Subsequently, RNA populations were reverse-transcribed into cDNA, linearly amplified, and paired-end sequenced using an Illumina HiSeq 2000 platform (see "Materials and Methods"). On average, LCM-RNAseq experiments yielded approximately 23 million 100-bp paired-end reads per sample (Supplemental Table S2). After quality trimming and removal of stacked reads, approximately 68% of all sequences were mapped to the maize reference genome (ZmB73_RefGen_v2; Supplemental Table S2). Finally, approximately 52% of the remaining reads mapped uniquely to the filtered gene set of maize (Supplemental Table S2), which comprises 39,656 high-confidence gene models (FGSv2, release 5b.60).

The Transcriptomic Landscape of Pericycle Cells Differs between Root Types under Homogeneous Low or Local High Nitrate Conditions

Gene-wise transcriptional activity of pericycle cells for each root type/nitrate treatment combination was

Downloaded from https://academic.oup.com/plphys/article/170/3/1783/6114378 by ulbbonn user on 21 December 2021

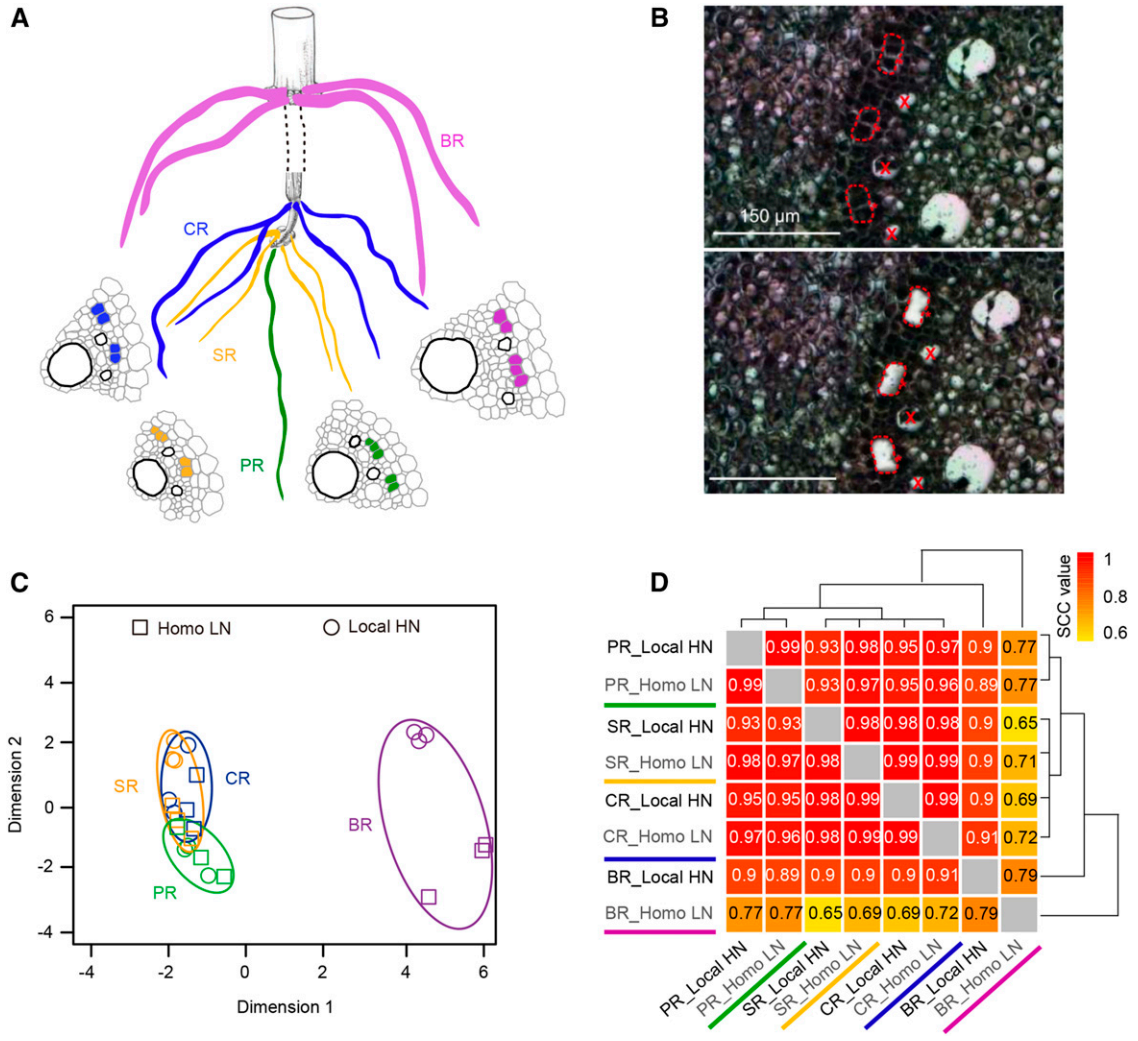


Figure 2. Maize root types, cellular organization, and laser captured pericycle cell transcriptomes under homogenous low nitrate or local high nitrate conditions. A, Graphic depiction of pericycle cells (in color) adjacent to phloem poles of the four maize root types primary root (PR), seminal root (SR), crown root (CR), and brace root (BR) subjected to RNA sequencing. B, Representative sections taken before and after laser capture microdissection of phloem pole pericycle cells. Asterisks, phloem poles. C, Multidimensional scaling plot visualizing overall similarity of RNA populations based on gene expression for all 24 root type-specific pericycle cells. Homo LN, homogeneous low nitrate; Local HN, local high nitrate. D, SCC analysis of the eight transcriptomes using $\log_2$-transformed FPKM values of all expressed genes. The hierarchical clustering dendrogram was inferred by applying 1-SCC as distance function.

computed. To decide if a gene was expressed ("active") or not ("inactive") in any root type/nitrate treatment combination, a generalized linear model analysis was conducted (see "Materials and Methods"). In total, 25,132 genes (63% of the filtered gene set; Figure 3A) were declared active in pericycle cells of at least one root type and nitrate condition. The specificity of pericycle transcriptomes in different root types/nitrate combinations was visualized in a Venn diagram. A substantial number of genes were constitutively active in all four tissues under homogeneous low (17,370; 71%) and local high nitrate (17,551; 73%) conditions (Fig. 3A). Notably, a significant number of genes were exclusively active in brace roots under both nitrate conditions (Fig. 3A). In contrast, a smaller number of genes were specifically active in the three seedling root types (Fig. 3A), suggesting root type specificity of genes expressed during early lateral root initiation (Fig. 3A). Both, a multidimensional scaling plot (Fig. 2C) and hierarchical clustering of Spearman correlation coefficient (SCC) analyses (Fig. 2D) revealed high correlation between pericycle transcriptomes of the three seedling root types and a distinct clustering of brace root transcriptomes. Gene-wise transcriptional activity was further computed to explore the functional specificity for each root type/nitrate combination. Remarkably, a substantially higher number of genes were specifically activated by local high nitrate

Downloaded from https://academic.oup.com/plphys/article/170/3/1783/6114378 by ulbbonn user on 21 December 2021

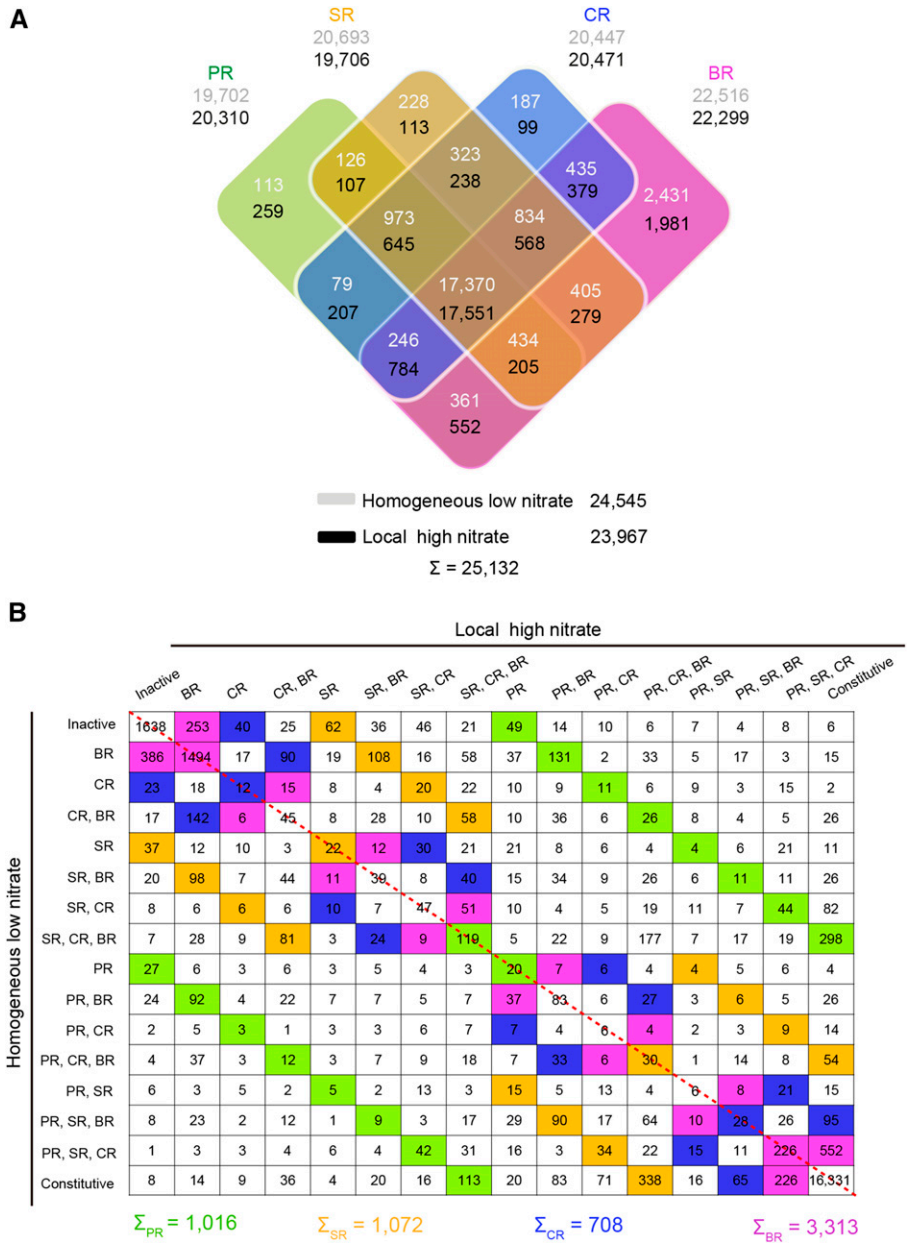


Figure 3. Root type-specific gene activity in pericycle cells of maize. A, Venn diagram illustrating the overlap of active genes in pericycle cells of the four root types under homogenous low nitrate or local high nitrate conditions. B, Overview of gene activity patterns and numbers of genes with these patterns in pericycle cells of the four maize root types grown under homogenous low nitrate or local high nitrate conditions. Each colored cell represents a root type-specific gene activity and states the number of active genes. Columns represent patterns under local high nitrate condition, while rows represent expression patterns under homogenous low nitrate condition. Increase of root type-specific gene activity patterns in pericycle cells after local high nitrate stimulus is indicated above the dashed diagonal red line, while increase of such patterns under homogenous low nitrate conditions is summarized below the dashed red line. PR, primary root; SR, seminal root; CR, crown root; BR, brace root.

in brace roots (3,313 genes) in comparison to the other root types (primary root, 1,016 genes; seminal root, 1,072 genes; crown root, 708 genes; Fig. 3B). Among the genes only active in brace roots, a majority of them (1,720; 52%) were constitutively active irrespective of the nitrate condition (Fig. 3B). This observation highlights the genetically determined specificity in brace roots, which is different from other root types (Fig. 3B). Hence, our observation might suggest that brace roots are distinct in some specific functions of pericycle cells typically not found in the other root types during lateral root initiation. These results are consistent with the morphological observations on lateral root branching in brace roots.

Dynamic Reprogramming of Pericycle Cell Transcriptomes after Changed External Nitrate

To understand the systemic transcriptome responses of different root types to heterogeneous nitrate distribution, we compared pericycle cell specific transcriptome profiles by K-means clustering. The relationship within clustering representation helped to illustrate the strong similarity and diversity between expression profiles among root types under different nitrate conditions. Thus, we identified 10 clusters among four root types covering all 18,985 expressed genes under homogeneously low nitrate conditions (Supplemental Fig. S2A) and 19,111 expressed genes under local high nitrate conditions (Supplemental

Downloaded from https://academic.oup.com/plphys/article/170/3/1783/6114378 by ulbbonn user on 21 December 2021

Fig. S2B). Among these, three main clusters (K1–K3) representing significant gene expression transitions from low in seedling root types to high in brace roots accounted for 20% of the expressed genes under homogeneous low nitrate and 13% of the expressed genes under local high nitrate (Supplemental Fig. S2A). We then adopted the MapMan functional annotation system to assign the genes in the six clusters (K1 to K3 homogenous low nitrate and K1 to K3 local high nitrate) to functional categories. In total, genes of these six specific clusters were significantly enriched in seven MapMan bins (Fig. 4B). Among those, genes that encode proteins for "signaling" were enriched in all six clusters (Fig. 4B). Moreover, genes that encode proteins related to "DNA" and "transport" were enriched in specific K-means clusters under changing external nitrate levels (Fig. 4B). Finally, genes associated with "RNA" and "protein"-related processes represented enriched functions related to nitrate induction and inhibition, respectively (Fig. 4B). Taken together, this dynamic response established the notion that nitrate availability may reprogram the transcriptome signature of pericycle cells at the root type level.

To further analyze the functions of the differentially expressed genes, the MapMan bin classification system was applied to compare the distribution of differentially expressed genes in the six pairwise comparisons of the four root types under homogenous low nitrate (Fig. 5A) and local high nitrate (Fig. 5B) conditions. For each comparison, the differentially expressed genes were split in two columns indicating the root type in which these genes were preferentially expressed (Fig. 5, A and B, upper panels). The ratios of these two columns for all six root type comparisons were defined as the expected values for further analyses. Subsequently, these ratios were determined for each of the six root type comparisons and both nitrate conditions for each of the 35 functional MapMan bins, and these were compared with the expected values by a two-tailed Fisher's exact test. For both nitrate conditions, these ratios differed significantly from the expected values in the same seven classes (Fig. 5, A and B). These classes were "RNA," "DNA," "protein," "signaling," "transport," "cell," and "development" (Fig. 5, A and B). Under homogeneously low nitrate (Fig. 5A), the functional classes "RNA," "DNA," and "protein" displayed a significant enrichment (T1) of genes preferentially expressed in pericycle cells of primary, seminal, and crown roots compared to brace roots (classes 3, 5, and 6) relative to the expected value. By contrast, comparisons of primary, seminal, and crown roots between each other (classes 1, 2, and 4) did not display any significant alterations of differentially expressed genes in these MapMan bins under homogenous low nitrate conditions compared to the reference values (Fig. 5A). These trends have been summarized in Figure 5C. In this panel, primary, seminal, and crown roots are depicted by the same dark color indicating no differences between these roots but an enrichment compared to the light color of brace roots. The opposite transcript

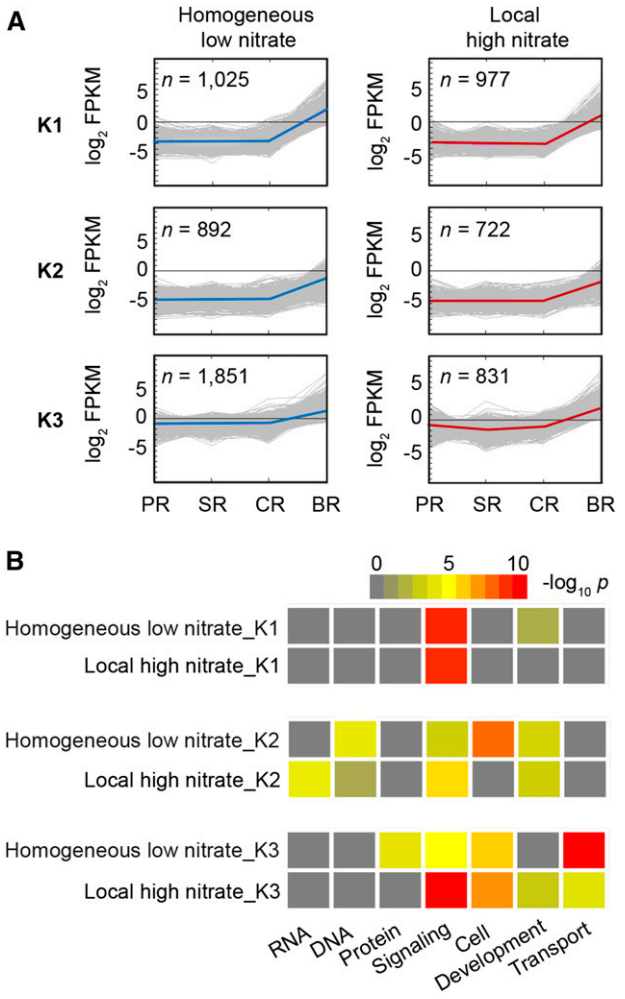


Figure 4. Dynamics and functional shift of pericycle cell transcriptomes of the four maize root types upon altered external nitrate levels. A, K-means clusters (K1–K3) showing the dynamics of pericycle transcriptomes among maize root types across homogenous low nitrate and local high nitrate conditions. Expressed genes that follow similar absolute values of Euclidean distance were indicated and clustered as an expression pattern. The K-means support module embedded in the MEV program (http://www.tm4.org/mev) was used to generate clusters. Each gene is plotted as a gray line, and the average expression pattern for each cluster is drawn as a blue or red line. *n*, the number of genes belonging to each cluster. B, Functional categories (extracted from MapMan bins) analyzed among the six induced K-means clusters depicted in A. Color code from gray to red represents significant enrichment of functional categories calculated by Fisher's exact test and normalized by $-\log_{10} P$.

accumulation trend (T2) was identified for the functional classes "signaling," "transport," "cell," and "development": brace root > primary root = seminal root = crown root, with higher accumulation in adult brace roots than the three seedling root types (Fig. 5A). This pattern is also illustrated by the color scheme summary in Figure 5C. The same analysis was repeated for the six comparisons between root types grown under local high nitrate conditions (Fig. 5B) and

Downloaded from https://academic.oup.com/plphys/article/170/3/1783/6114378 by ulbbonn user on 21 December 2021

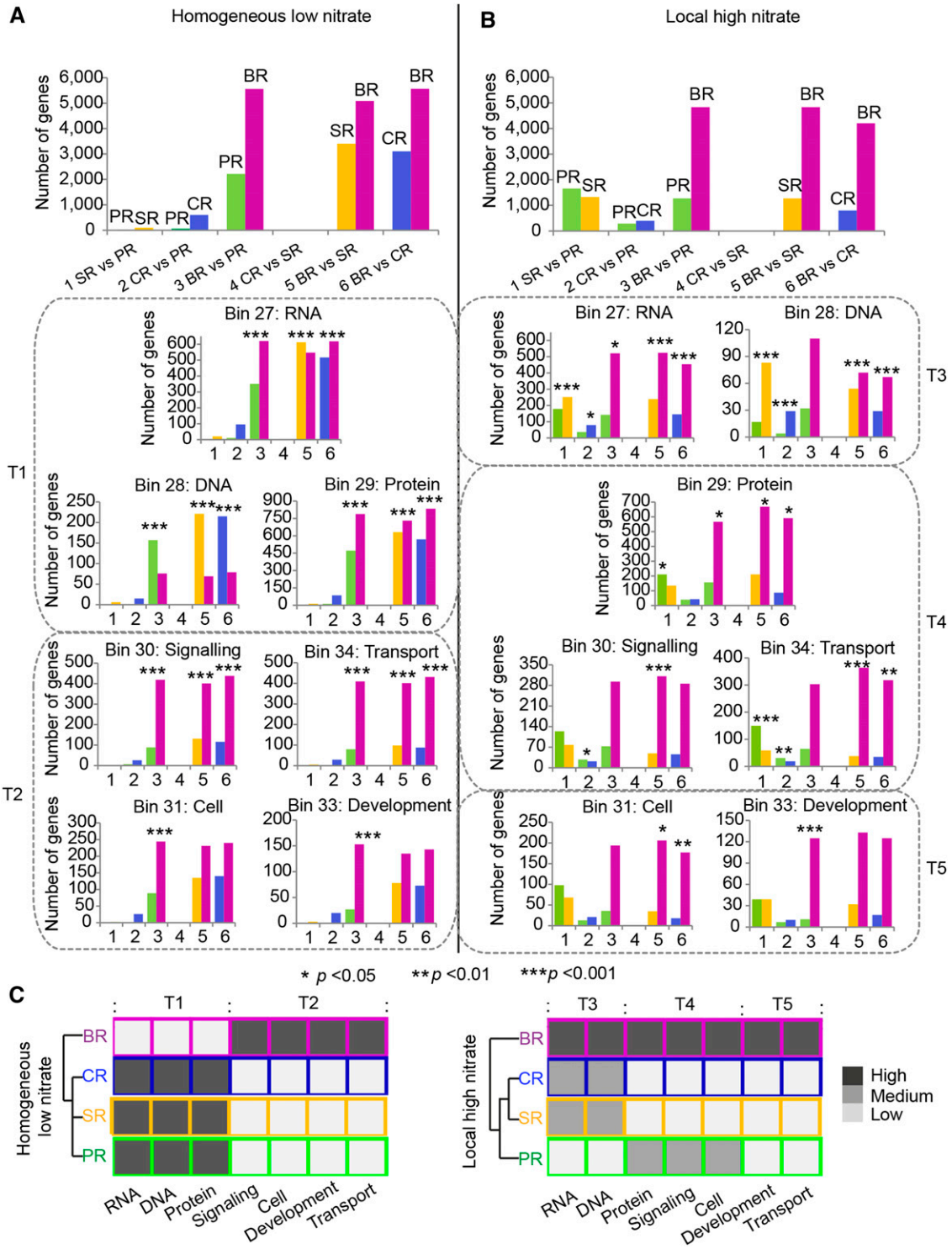


Figure 5. Preferentially expressed genes and functional distributions in pericycle cells among six pairwise root type comparisons under homogeneous low nitrate (A) and local high nitrate (B) conditions. All genes preferentially expressed in one root type in the six comparisons were mapped to the 35 MapMan bins. The different color-coded columns in each comparison indicate the number of genes preferentially expressed in the corresponding root type. For each pair of root type comparisons (1–6), functional classes in which the ratio of preferentially expressed genes significantly differed from the expected ratio calculated from all differentially expressed genes were combined based on similar trends in classes T1 to T5. $*P < 0.05$, $**P < 0.01$, and $***P < 0.001$. C, Root type-specific enrichment trends of MapMan classes according to the results of A and B. The hierarchical clustering dendrogram was inferred by applying PCC as distance function. Same fill colors of the squares indicate no relative enrichment of

Downloaded from https://academic.oup.com/plphys/article/170/3/1783/6114378 by ulbbonn user on 21 December 2021

summarized in Figure 5C. In this analysis another three accumulation trends based on altered transcript ratios were observed. For "RNA" and "DNA," the enrichment trend was brace root > seminal root = crown root > primary root (T3). For "protein," "signaling," and "transport," the trend was brace root > primary root > seminal root = crown root (T4), while for "cell" and "development," the trend we brace root > primary root = seminal root = crown root (T5). Taken together, this observation suggests a functional shift in the transcriptome profile of pericycle cells among root types upon change of the external nitrate stimulus.

Nitrate-Responsive Genes in Brace Roots of Maize

Contrasts were calculated to identify significant changes (false discovery rate [FDR] < 0.05; $|\log_2 Fc| \geq 1$) in response to local high nitrate for each root type using a log-transformed linear model. In total, 3,640 genes expressed in pericycle cells were differentially expressed across the four root types (Fig. 6A). Among those, a majority of 3,046 differentially expressed genes were shown in brace roots (Fig. 6A), while seminal roots contributed 24% of all differentially expressed genes (Fig. 6A). Remarkably, primary and crown roots showed almost no responsive genes when induced by local high nitrate (Fig. 6A). This suggests that upon local high nitrate stimulus, diverse root types showed significantly diverse transcriptome responses to heterogeneous nitrate supplies.

Subsequently, the 2,740 genes that displayed differential expression upon nitrate stimulation exclusively in brace roots were functionally classified according to Gene Ontology (GO) terms using agriGO. To identify overrepresented functional categories, a singular enrichment analysis (SEA) was performed. SEA compares each annotated gene to all annotated expressed genes. In total, 45 GO terms displayed overrepresentation (FDR < 0.001) (Supplemental Fig. S3). In parallel, differentially expressed genes were further assigned to MapMan functional categories to compare the distribution of overrepresented and underrepresented functional classes between nitrate treatments. Overall, 640 distinct genes were significantly overrepresented in MapMan bins "RNA," "DNA," and "cell," and 311 distinct genes were significantly underrepresented in MapMan bins "signaling" and "transport" (Fig. 6B). These functional analyses are consistent with those of the transcriptome expression patterns and suggest that brace root pericycle cells likely acts uniquely in comparison to the other root types. To gain insights into the biological role of the enriched functional classes, we investigated the complete set of 640 identified overrepresented genes for whether they were functionally predictive and associated them with the aid of the STRING algorithm. Coexpression analyses with high confidence of ≥ 0.7 were performed to construct a functionally connected network. Direct interactions of transcripts were clustered and corresponded to DNA synthesis and chromatin structure, DNA replication, cell cycle/division, and cell organization (Fig. 6C). This network supports a role of pericycle cells adjacent to phloem poles in cell cycle and cell division during lateral root initiation in brace roots.

Among the 2,740 nitrate-regulated brace root specific genes, 60 "classical" maize genes with experimentally confirmed functions (Schnable and Freeling, 2011) were determined (Supplemental Table S3). Twelve of them were related to DNA synthesis and cell cycle (Supplemental Table S3). Specifically, six "classical" maize genes were associated with histones (*his2a1*, *his2b1*, and *his2b2*) and cell cycle control (*cyc1*, *cyc3*, and *tub1*; Fig. 6C). Finally, we investigated transcription factors (TFs) that are key regulators of growth, development, and cell fate among the 2,740 nitrate-regulated brace root specific genes. Most differentially regulated TFs belonged to the NAC, bHLH, MYB, TALE, C2H2, and bZIP TF families (Fig. 7A; Supplemental Table S4). All these TFs were further assigned to STRING v10 to predict their functional connectivity. Only hits with high confidence of ≥ 0.7 were used to construct the network (Fig. 7B). Interactions comprised the MYB, MADS box, and CPP families, which were interconnected by response regulator (RR) ZmRR9 and maize GDB, curated "classical" genes ZmGATA34 and ZmGATA2 (Fig. 7B). The genes enriched in MapMan bins "DNA" and "cell" were used to develop a model to illustrate cell cycle control during lateral root initiation in brace roots (Fig. 8). The four histone families (H2A, H2B, H3, and H4) and linker histone H1, which constitute the chromatosome core are associated with DNA synthesis (MapMan bin classification) and were moderately induced by local high nitrate application (Fig. 8A). A heat map associated with the transition from metaphase to anaphase of mitosis showed upregulated genes encoding several spindle checkpoint proteins (MITOTIC ARREST DEFICIENT2 [MAD2]) and proteins involved in the ANAPHASE-PROMOTING COMPLEX (APC), SMC, and CELL DIVISION CYCLE (CDC) (Fig. 8B). At last, genes enriched in cell division and organization processes were classified into four groups: I, subunits of microtubules; II, structural constituents of

Figure 5. (*Continued.*)
differentially expressed genes in these MapMan bins between root types while darker colors indicate relative enrichment of differentially expressed genes in root types compared to root types with lighter fill colors. PR, primary root; SR, seminal root; CR, crown root; BR, brace root.

Downloaded from https://academic.oup.com/plphys/article/170/3/1783/6114378 by ulbbonn user on 21 December 2021

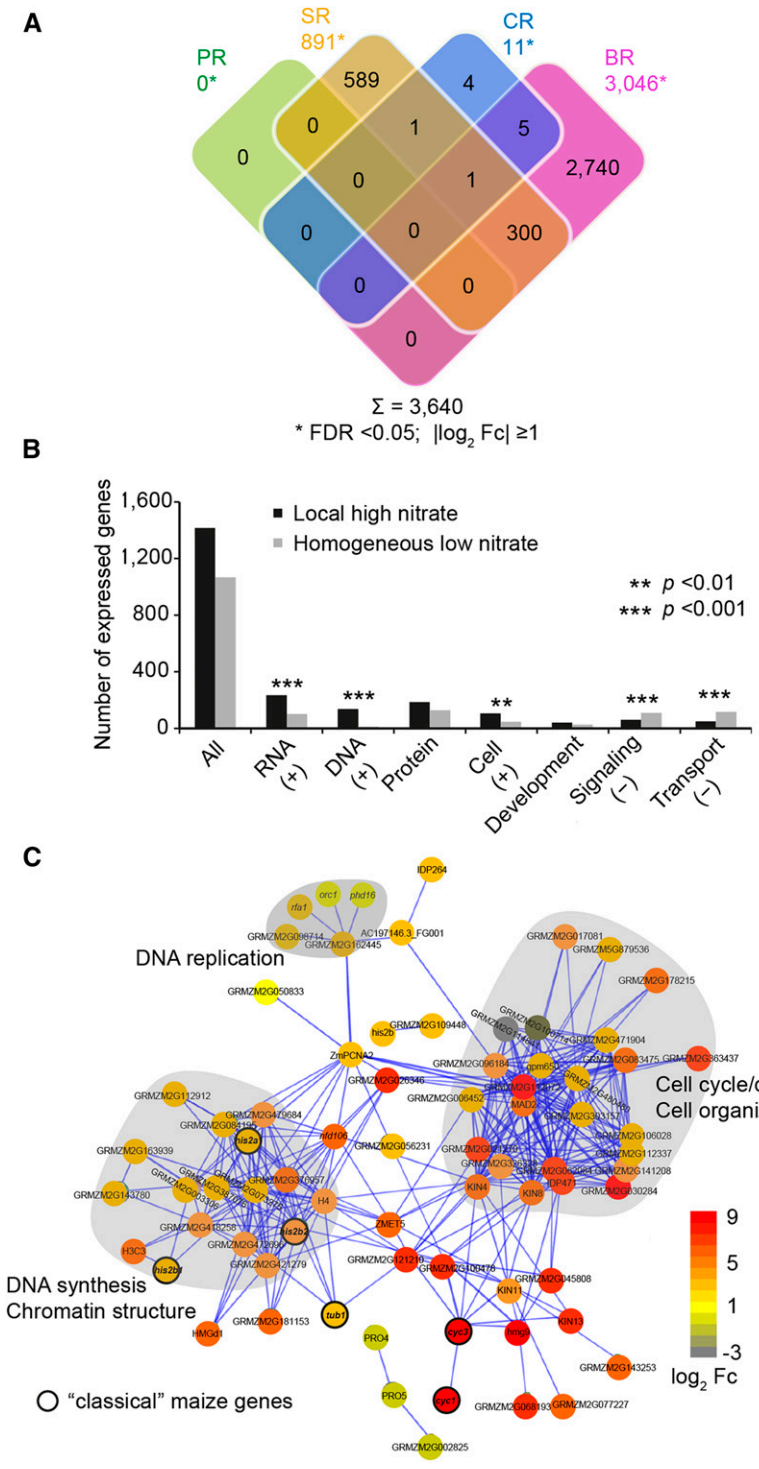


Figure 6. Functional specificity of the brace root specific transcriptome. A, Overlap of differentially expressed genes (FDR < 0.05, $|\log_2 Fc| \geq 1$) between pericycle transcriptomes of the four root types (PR, primary root; SR, seminal root; CR, crown root; BR, brace root) in response to local high nitrate supply versus homogenous low nitrate as control. B, Functional categories in which the ratio of preferentially expressed genes differed significantly from the ratio of all expressed genes. The expected number of genes for each functional category was calculated based on the distribution of functional categories among all expressed genes. To determine if significantly more overrepresented (+) or underrepresented (−) genes than expected were detected for each individual category, a χ^2 test for independence with Yates' continuity correction was performed. $^{**}P < 0.01$ and $^{***}P < 0.001$. C, Network view of genes with high coexpression scores of ≥0.7 generated using STRING v10 prediction algorithm. The visible genes associated with the categories "DNA" and "cell" were connected based on 640 genes extracted from overrepresented MapMan bins ("RNA," "DNA," and "cell" according to B). Three clusters of direct interaction partners are marked by gray background and correspond to DNA synthesis and chromatin structure, DNA replication, cell cycle/division, and cell organization, respectively. Thickened node outlines represent "classical" maize genes. Color codes from gray to red indicate expression level of nitrate-responsive transcripts normalized as $\log_2$ algorithm of fold changes ($\log_2$ Fc).

the cytoskeleton; III, microtubule binding and associated protein; IV, genes coding kinesin/like motor proteins (Fig. 8C). The complete list of genes associated with chromatin structure and core histones extracted from bin28 and genes associated with cell cycle, cell division, and cell organization extracted from bin31 of the MapMan analysis are shown in Supplemental Tables S5 and S6, respectively.

DISCUSSION

Lateral root branching is a key adaptive strategy in establishing a root system architecture that can optimally exploit heterogeneously distributed nutrients from soil (Ruffel et al., 2011; Guan et al., 2014; Postma et al., 2014; Zhan et al., 2015). Highly branched crop root systems are composed of multiple root types formed at different stages of development under the

Downloaded from https://academic.oup.com/plphys/article/170/3/1783/6114378 by ulbbonn user on 21 December 2021

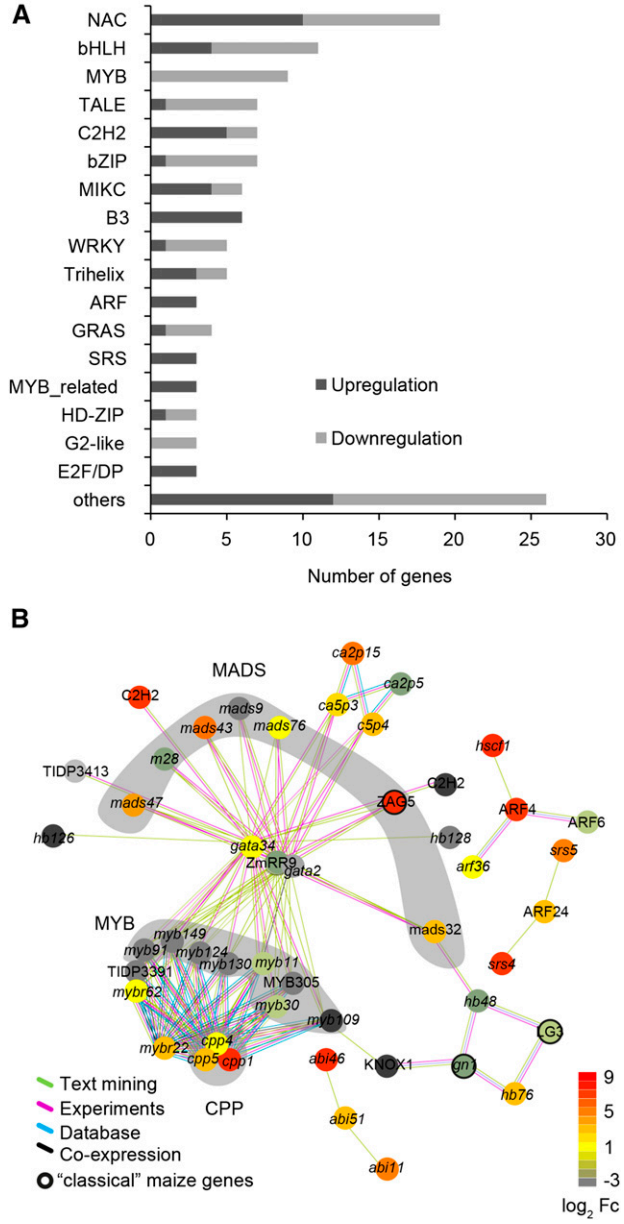


Figure 7. TF analyses among the brace root specific nitrate-responsive genes. A, Distribution of the 132 nitrate-responsive TFs according to families. TFs were classified according to PlantTFDB3.0 database annotation. B, Prediction of connectivity of 132 TFs identified from brace root specific nitrate-responsive genes using STRING v10. Color codes indicate expression values of transcripts normalized as $\log_2$ algorithm of fold changes. A green line indicates text mining evidence, a purple line indicates experimental evidence, a blue line indicates database evidence, and a black line indicates coexpression evidence. Clustered groups associated with "MADS box," "MYB," and "CPP" families were highlighted accordingly. Outlines of nodes representing "classical" maize transcription factors were thickened.

genetic control of distinct genes (Hochholdinger et al., 2004a). The formation of embryonic primary and seminal and postembryonic shoot-borne and lateral roots in monocot cereals is more divergent than root development of dicot root systems (Hochholdinger et al., 2004b). In this study, integration of anatomical, morphological, and a LCM-RNAseq-based specific transcriptome analyses in pericycle cells demonstrated that maize root types showed similar but also divergent responses to heterogeneous nitrate. In particular, brace roots displayed unique cell cycle control mechanisms controlling lateral root formation in response to local high nitrate compared to the other root types.

Local High Nitrate Supply Modulates the Transcriptome Response of Pericycle Cells of Different Root Types in Maize

Determination of specific gene expression in pericycle cells profiles prior to lateral root initiation (Figs. 2 and 3) in combination with morphological and histological analyses (Fig. 1; Supplemental Table S1) allow extrapolation on the plasticity of crop responses under heterogeneous nitrate environments. Cell type-specific RNA-seq experiments revealed that the transcriptomes of pericycle cells isolated from diverse root types represented distinct functions irrespective of external nitrate levels (Figs. 4 and 5). This is consistent with the observation that several maize mutants defective in lateral root formation showed root type-specific phenotypes but did not affect all major root types (Hochholdinger et al., 2004a). The mutants *rum1* (Woll et al., 2005) and *lrt1* (Hochholdinger and Feix, 1998) are both defective in the formation of lateral roots in primary roots but not affected in the shoot-borne root system. Consistent with this observation, profiling of pericycle transcriptomes revealed exceptional behavior of these cells in shoot-borne brace roots, which are very different from the seedling root types (Figs. 2, 3, and 4A). In rice (*Oryza sativa*), functionally distinct transcriptome profiles were identified between different root types in response to arbuscular mycorrhizal fungi (Gutjahr et al., 2015). Functional differences of pericycle transcriptomes between the four studied root types were reprogrammed in response to local high nitrate stimulation in maize (Fig. 5C). These results have important implications for understanding root adaptability in sensing environmental nitrate in plants. Remarkably, brace roots exhibited a high transcriptome activity in all seven enriched functional classes in response to local nitrate compared to seedling root types (Fig. 5C). In agricultural systems, adult brace roots contribute more to the branching response and participate in acquisition of nitrate than the seedling primary and seminal roots during the whole life cycle (Yu et al., 2014a, 2014b, 2015b). In particular, the shoot-borne root system, including brace roots, plays a crucial role in resource capturing, yield increase and lodging resistance via their branched roots throughout plant life compared to the seedling root system, which is eminent only during the early stages of plant development (Hochholdinger, 2009; Rogers and Benfey, 2015). Therefore, our findings suggest functional adaptability of the transcriptome signature

Downloaded from https://academic.oup.com/plphys/article/170/3/1783/6114378 by ulbbonn user on 21 December 2021

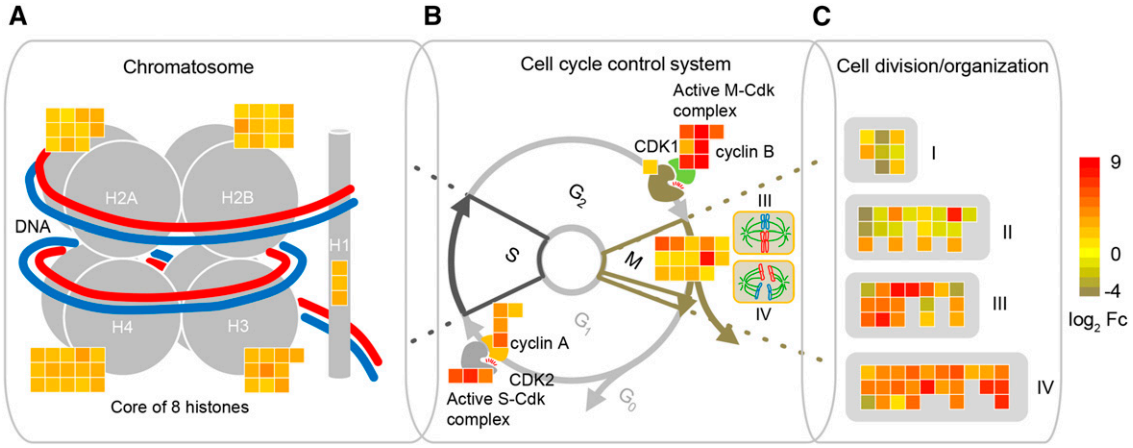


Figure 8. Nitrate-responsive regulatory genes and key regulators involved in the cell cycle control system regulating G1-S and G_2-M transitions during pericycle cell division. A, Schematic illustration of a chromatosome consisting of a histone octamer. Histones H2A, H2B, H3, and H4, core histones; Histone H1, linker histone. B, S-phase and mitosis-specific cell division (III, metaphase; IV, anaphase) regulation by cell cycle control system. Two activated cyclin and cyclin-dependent protein kinase (CDK) complexes progressing into S-phase and M-phase, respectively. A heat map associated with the transition from metaphase to anaphase showing upregulated genes on spindle checkpoint protein (MAD2), APC, SMC, and CDC protein. C, Transcripts enriched in cell division and organization processes were divided into four groups (I, subunits of microtubule; II, structural constituents of cytoskeleton; III, microtubule binding and associated protein; IV, genes encoding kinesin/like motor proteins). Color code from blue to yellow indicates expression level of nitrate-responsive cell cycle genes normalized as $\log_2$ algorithm of fold changes ($\log_2$ Fc).

of diverse root types. This highlights the notion of unique characteristics of the complex molecular networks in monocot cereals compared to the dicot model species Arabidopsis.

Heterogeneous Nitrate Supply Promotes Lateral Root Initiation by Triggering the Pericycle Cell Cycle Control System

Cell cycle control is a fundamental mechanism to regulate lateral root initiation by allowing specific cells to divide or not (Himanen et al., 2002, 2004; Vidal et al., 2013). Our study revealed that cell cycle control is integrated in brace roots by highly connective nitrate-responsive genes unique to this root type (Figs. 6 and 8). Functional analyses indicated enriched coregulated genes involved in nucleosome/chromatin assembly and microtubule motor activity/movement during mitosis (Supplemental Fig. S3; McIntosh et al., 2002). This is consistent with the enrichment of genes encoding proteins involved in DNA binding and microtubule movement upon auxin-induced lateral root initiation in maize (Jansen et al., 2013). The increased transcription of these genes involved at the G1-S phase boundary might be required to provide enough histone mRNA to keep pace with the need for new chromatin synthesis during S phase (Osley, 1991). Moreover, enrichment of genes required for the transition from metaphase to anaphase during mitosis was coordinated with enrichment of genes encoding microtubules and its associated motor proteins (Fig. 8C). This result is in line with the recent finding of active pericycle cell division in the stele of brace roots in the lateral root initiation zone in response to local high nitrate supply (Yu et al., 2015b). TFs play a key role in regulatory networks shaping growth and development. Therefore, identification of nitrate-responsive TFs enabled a better understanding of nitrate during lateral root initiation (Berckmans and De Veylder, 2009; Canales et al., 2014). We identified three TFs (ZmGATA34, ZmGATA2, and ZmRR9) that connected MYB and MADS box families as inferred from network predictions (Fig. 7B). In Arabidopsis, GATA23 plays a key role in specifying pericycle cells to become lateral root founder cells prior to lateral root initiation (De Rybel et al., 2010). Moreover, MADS box transcription factors are implicated in nitrate-induced changes of root architecture (Zhang and Forde, 1998). The observation of MYB transcription factors supports their role in cell cycle regulation and lateral root formation both in Arabidopsis (Ito et al., 2001; Shin et al., 2007; Mu et al., 2009) and maize (Jansen et al., 2013). Moreover, CPP-like TFs play an important role in development of reproductive tissues and control of cell division in plants (Yang et al., 2008), which is correlated to the distinct transcriptomic response in adult brace root. These results revealed strong transcriptional control in nitrate-responsive gene networks during lateral root initiation in brace roots.

By integrating MapMan outputs and network interactions, a model was devised to highlight how nitrate triggers pericycle cell cycle progression during lateral root initiation in brace roots of maize (Fig. 8). Accordingly, G_1/S phase-specific genes, such as *histone H4*, were induced together with the *CYCD3;1* cyclin and DNA replication genes by auxin during lateral root

Downloaded from https://academic.oup.com/plphys/article/170/3/1783/6114378 by ulbbonn user on 21 December 2021

initiation in Arabidopsis (Himanen et al., 2002, 2004). In addition, active S-CDK (cyclin-dependent protein kinases) and M-CDK complexes were progressing into the S-phase and M-phase, respectively (Fig. 8B; Polyn et al., 2015). Moreover, a link between MAD2 and the cell cycle regulatory proteins APC and CDC proteins has provided support for the notion that anaphase is initiated (Elledge, 1998; Yu et al., 1999). These results are consistent with genes identified in this study associated with the transition from metaphase to anaphase of mitosis (Fig. 8B; Supplemental Table S6). Dynamic instability of microtubule subunits (TUBLIN A and TUBLIN B) and associated actin-binding proteins (ADF, VILLIN, and PROFILIN4) suggest a driving force for movements of microtubules during cell division (Smith, 2001). Furthermore, microtubule-based molecular motors are functionally important in organelle transport in the cytoplasm and in cell division (Asada and Collings, 1997). Taken together, root type-specific responses induced by local high nitrate supply were demonstrated to be involved in cell cycle control during lateral root initiation, supporting a unique lateral root branching strategy for maize brace roots.

Lateral Root Foraging Strategies Depend on the Developmental Phases and Internal Nitrogen Status

The present study underscores the plasticity of different root types during lateral root initiation at different stages of maize development. The model in Figure 9 supports the notion of highly divergent developmental complexity of maize root system development compared to Arabidopsis. Shoot-borne roots initiate from the elongation stage onward and contribute to the vast majority of total nitrogen uptake in maize plants (Shane and McCully, 1999; Peng et al., 2010). Notably, primary and crown roots showed no changes of their pericycle transcriptomes in response to local nitrate stimulus (Fig. 6A). Their root type-specific reaction to nitrate may suggest that pericycle cells of maize respond differently to environmental nitrate stimulation. A microarray analysis of pericycle cells suggested common mechanisms for lateral root initiation in maize primary and crown roots (Jansen et al., 2013). Remarkably, divergent responses of seminal versus primary and crown roots might be related to their evolutionary origin because cereal species such as rice are lacking seminal roots entirely (Rogers and Benfey, 2015). Thus, we hypothesize that nitrate-induced functional shifts between root types reflect diverse responses on lateral root foraging (Fig. 5C), as illustrated by brace root-specific distribution of functional classes and constitutively active genes in response to local high nitrate stimulus (Fig. 3B).

Systemic morphological and histological analyses revealed that all seedling root types and early shoot-borne roots specifically exploit heterogeneous nitrate by promoting lateral root elongation, whereas shoot-borne roots initiated at late vegetative growth stage (60 d after germination) display more plastic responses showing both increased length and density of lateral roots (Yu et al., 2014a, 2014b, 2015a, 2015b). Such directed lateral root development depends on regulatory networks that integrate both local and systemic signals to coordinate them with the overall plant internal nitrogen status (Ruffel et al., 2011; Guan et al., 2014). Economizing the costs for root development is pivotal for a resource-efficient strategy in nutrient acquisition, as illustrated by the impact of the internal nitrogen status-dependent regulatory module CLAVATA3/EMBRYO-SURROUNDING REGION-related peptides-CLAVATA1 leucine-rich repeat receptor-like kinase (Araya et al., 2014; Tabata et al., 2014). In particular, hormones such as cytokinin, which serves as a long-distance systematic messenger mediated by nitrate, can signal the nitrogen demand of the whole plant (Ruffel et al., 2011). The nitrogen concentration of the shoot shows decreasing progression from seedling to flowering stage, which indicates systemic nitrogen starvation responses intensify gradually (Yu et al., 2014a, 2015b). It was suggested that root growth and plasticity depend on the systemic nutrient status and shoot growth potential (Forde, 2002; Gojon et al., 2009). This notion can explain

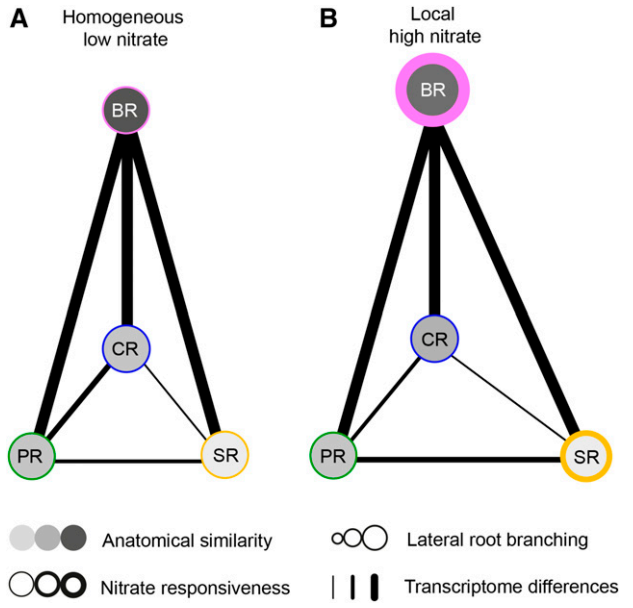


Figure 9. Nitrate-modulated similarities and differences of lateral root branching in maize root types. Network view of morphology, anatomy, and transcriptome profiling in relation to the four maize root types under homogenous low nitrate (A) and local high nitrate (B) conditions. Grey shades of nodes correspond to the similarity of anatomical traits among the four root types (according to Supplemental Fig. S1 and Supplemental Table S1). Node size corresponds to the degree of lateral root branching response to local nitrate stimulus (according to Fig. 1). Thickened node outlines indicate more nitrate-responsive genes induced by local high nitrate stimulation (according to Figs. 3B and 6A). Line thickness between two nodes is proportional to the number of significantly changing genes (according to Fig. 5).

Downloaded from https://academic.oup.com/plphys/article/170/3/1783/6114378 by ulbbonn user on 21 December 2021

the more substantial morphological and transcriptome responses to local high nitrate application in brace roots compared to the other root types.

Therefore, this work extends the current knowledge of root type-specific functions of pericycle cells in maize under changing nitrate availability. Mechanistic interactions between plant development and root plasticity highlight specific root foraging responses that depend on the root developmental age and endogenous nitrogen status of the plant (Araya et al., 2014; Tabata et al., 2014). Understanding the reprogramming of these transcriptomic networks under limited nitrate conditions in brace roots in the context of yield acquisition and adult fitness can be relevant for rational breeding approaches. To this end, while cell cycle control mechanisms are likely evolutionary conserved during lateral root initiation across angiosperms, root type-specific transcriptomic networks of pericycle cells are functionally diverse during lateral root initiation in maize.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Seeds of the maize (*Zea mays*) inbred line B73 were surface sterilized and germinated in paper rolls in distilled water as previously described (Hetz et al., 1996). Seven days after germination, the endosperm of each seedling was excised and uniform seedlings with two visible leaves and four embryonic roots were transferred into low nitrate (0.5 mM NO_3^-) nutrient solution and grown until a root length of 7 cm per root type. One primary, seminal, or crown root was separately treated for 24 h with homogeneous low nitrate (0.5 mM NO_3^-) or local high nitrate (4 mM NO_3^-) nutrient solution in a two-compartment split-root system (Yu et al., 2014a). For brace root treatment, maize plants were cultured in a 20-liter container until brace roots were initiated. Two neighboring brace roots were kept in a small plastic bag with low (0.5 mM NO_3^-) or high (4 mM NO_3^-) nitrate nutrient solution for 24 h. These split-root systems were aerated continuously to maintain the oxygen content. The experiment was conducted in a growth chamber in a 16-h-light, 25°C, and 8-h-dark, 21°C cycle. Setup of split nitrate supply to four root types was described in detail by Yu et al. (2015b).

Morphological, Histological, and Anatomical Analyses

To comprehensively determine the morphological and histological responses of lateral root branching, local nitrate-treated primary, seminal, crown, and brace roots were harvested after 20 d of treatment. The average length and density of lateral roots and early pericycle cell divisions for seedling root types were examined according to Yu et al. (2015a).

To determine the anatomical structure of the four maize root types, transverse sections collected at 3 mm from the root tip of 7-d-old primary roots, 9-d-old seminal roots, 15-d-old crown roots, and 60-d-old brace roots were prepared with a VT1200 Vibratome (Leica). Areas of total transverse, stele, meta-xylem vessels, and size of pericycle cells adjacent to phloem and xylem poles were measured under an AxioCam MRc microscope (Carl Zeiss Microimaging) and then documented with Axio-Imager software (Carl Zeiss Microimaging).

Laser Capture Microdissection of Pericycle Cells from Four Root Types

Root segments of 3 to 5 mm from the root tip were dissected and fixed immediately (Yu et al., 2015a). The strength of infiltration was adjusted to 400 MPa for seedling root types and 300 MPa for brace roots to maintain the morphology of the transversal sections and maximize the effects of the fixative and the cryoprotective during sample preparation. The fixative was replaced and vacuum infiltration/swirl steps were repeated twice. The fixative was then replaced by fresh Suc (Sigma-Aldrich) solution as cryoprotective and infiltrated as described previously to minimize the formation of ice crystals (Yu et al., 2015a). In addition, concentrations of Suc were adjusted from 10% and 15% steps for primary, seminal, and crown root to 15% and 40% steps for brace root. The root segments of the three individual seedling root types and brace roots from the adult root system were embedded in $7 \times 7 \times 5$-mm disposable molds (Polysciences Europe) with PolyFreeze-clear tissue freezing medium (Polysciences Europe). About 2,000 phloem pole pericycle cells were collected per root type and nitrate condition in three biological replicates by laser microdissection according to Yu et al. (2015a). The fine-scale capturing of phloem pole pericycle cells is shown in Figure 2B.

RNA Isolation and Amplification

RNA extraction, purification, and quality determination were performed according to Yu et al. (2015a). Only samples with a RIN value ≥6.5 were used for downstream amplification (Supplemental Fig. S4). For each of the 24 samples, 20 ng of high-quality RNA was used for cDNA synthesis and amplification. cDNA synthesis with oligo(dT) and random primers and cDNA amplifications were performed using the SMARTer PCR cDNA synthesis kit (Clontech Laboratories) following the manufacturer's protocol. The quality and profile of the RNA and amplified cDNA samples were checked on an Agilent 2100 Bioanalyzer (Agilent Technologies) using an Agilent RNA 6000 Pico Kit (Agilent Technologies) and an Agilent High Sensitivity DNA Kit (Agilent Technologies), respectively.

cDNA Library Construction and Illumina Sequencing

The cDNA libraries for Illumina sequencing were constructed according to the manufacturer (TruSeq DNA Sample Prep Kit V3; Illumina). An Agilent 2100 Bioanalyzer (Agilent DNA 1000 LabChip) and ABI StepOnePlus real-time PCR system (Applied Biosystems) were used for quantification and qualification of the sample libraries. For sequencing, 24 libraries were arranged randomly on four flow cell lanes. Cluster preparation and paired-end read sequencing were performed according to the manufacturer's guidelines (HiSeq 2000; Illumina). After the raw sequencing reads were generated, adapter sequences were trimmed using the Trimmomatic program (Bolger et al., 2014). Quality of the trimmed reads was checked using the FastQC program (Andrews, 2010).

Processing and Mapping of Illumina Sequencing Reads

Raw sequencing reads were processed and subsequently mapped with CLC Genomics Workbench software (version 8.0.1). Reads with more than one mismatch in the adapter sequence were excluded. By quality trimming, low-quality and ambiguous nucleotides of sequence ends and adapter contaminations were removed. Only ≥40-bp reads were retained for further analyses. These were initially mapped to the maize B73 reference genome sequence (RefGen_v2; ftp://ftp.gramene.org/pub/gramene/maizesequence.org/release-5b/assembly/), allowing large gaps of up to 50 kb to span introns. To be mapped, at least 90% of each read had to fit with 90% similarity to the reference. Stacked reads, i.e. redundant reads sharing the same start and end coordinate, sequencing direction, and sequence, were merged into one. The remaining reads were further projected to the "filtered gene set" (FGS v2; release 5b) of the B73 reference genome. Thereby, reads had to fit at least with 80% of their length comprising 90% similarity to the reference. Only reads uniquely mapping to the reference genome were further analyzed.

Statistical Procedures for Determining Active and Inactive Genes

We used a generalized linear model with a negative binomial response to model gene-wise transcriptional activity for each root type and nitrate treatment combination (Opitz et al., 2015). The log of the mean read count was assumed to be the sum of a normalization factor and a linear function of the fixed effects for each combination of root type and nitrate treatment. Normalization factors were calculated separately for each experimental unit by fitting a smoothing spline with response log(count+1) and with GC content and log gene length as explanatory variables, then using the fitted values of each estimated function at each gene's GC content and length.

The vector of fixed effects for each gene was assumed to be a draw from a multivariate normal distribution with an unknown and unrestricted mean and an unknown diagonal variance-covariance matrix. The log of the negative binomial dispersion parameter was assumed to be constant within a gene and a draw from a normal distribution with unknown mean and variance. The unknown

Downloaded from https://academic.oup.com/plphys/article/170/3/1783/6114378 by ulbbonn user on 21 December 2021

parameters in the distributions of the fixed effects and negative binomial dispersion were estimated using an empirical Bayes procedure via the R package ShrinkBayes (Van De Wiel et al., 2013), which uses integrated nested Laplace approximation (Rue et al., 2009) to approximate the posterior distribution for the fixed effect associated with gene g, root type r, and nitrate condition c.

For a given threshold T, the posterior distributions for gene g were used to find P_{grc} (T), the posterior probability that the fixed effect for gene g, root type r, and condition c was larger than T. Gene g was called "active" for root type r and condition c if P_{grc} (T) > 0.5 and "inactive" otherwise. Because fixed effects were estimated while accounting for sequencing differences from sample to sample, gene length, and GC content differences, classifying genes as active or inactive based on the posterior distribution of fixed effects is more meaningful than attempting to do so based on a single raw read count threshold applied to all genes. In terms of counts, the selected threshold T resulted in calling a gene of average length and average GC content active if the expected number of reads per million mapped reads was approximately one or more. The read count threshold was, in effect, adjusted up or down for other genes depending on gene length, GC content, and the observed empirical distribution between these variables and read count.

Statistical Procedures for Analyzing Differential Gene Expression

For the following analyses, only genes were considered that were represented by a minimum of five mapped reads in all three replicates of at least one sample. These genes were declared as "expressed." The raw sequencing reads were normalized by sequencing depth and were $\log_2$-transformed to meet the assumptions of linear models. Furthermore, the mean variance relationship within the count data were estimated and precision weights for each observation were computed (Law et al., 2014). Prior to further data analyses, sample relations were analyzed based on multidimensional scaling using the plotMDS function of the Bioconductor package *limma* (Smyth, 2005) in R (R version 3.1.1 2014-07-10, limma_3.20.9). The distance between each pair of samples was estimated as the root mean square deviation for the top genes with the largest standard deviations between samples. The data were further analyzed according to a completely randomized design including a fixed effect for treatment and a random error term. After model fit, an empirical Bayes approach was applied to shrink the sample variances toward a common value (Smyth, 2004). Hypotheses tests were performed using the contrasts.fit function of the Bioconductor package *limma* (Smyth, 2005). The resulting *P*-values of the performed pairwise *t* tests were used to determine the total number of differentially expressed genes for each comparison by controlling the FDR < 0.05 to adjust for multiple testing (Benjamini and Hochberg, 1995).

GO and Metabolic Pathway Analyses

The Web-based agriGO software was used to assign expressed or differentially expressed genes to GO functional categories (Du et al., 2010). With SEA, overrepresented categories were computed by comparing the lists of differentially expressed and all expressed genes. Multiple testing was corrected by FDR (Benjamini and Yekutieli, 2001), and a cutoff was introduced at 0.05, 0.01, and 0.001. The MapMan software (Thimm et al., 2004) assigned differentially expressed genes to metabolic pathways and subsequently visualized them based on the functional annotation file ZmB73_5b_FGS_cds_2012. Fisher's exact test was used to determine if the observed number of genes in each of the 35 major MapMan bins significantly deviated from the expected distribution of all expressed genes.

Correlation Analysis on Global Transcriptomes

SCC analysis was used to quantify the relationship among the transcriptomes under a given root type. SCCs were calculated from $\log_2$-transformed FPKM (fragments per kilobase of exon per million fragments mapped) values of expressed genes using R. The hierarchical clustering dendrogram was inferred by applying 1-SCC as distance function.

Expression Profile Correlation and Clustering among Root Types

K-means clustering was performed with the K-means support module embedded in the MEV program (http://www.tm4.org/mev). A scalar quantity, called the figure of merit (Yeung et al., 2001), was used to assess the quality of the clustering algorithm and to determine the optimal number of clusters. The K-means analysis was performed six times with each run generating 10 clusters using Euclidean distance by controlling FDR < 0.05 among the expressed genes of four root types under homogeneous low and local high nitrate conditions.

Known and Predicted Protein-Protein Interactions and Network Analyses

Association networks of the identified genes significantly enriched in MapMan bins ("RNA," "DNA," "signaling," "cell," and "transport") were constructed with the aid of the online analysis tool STRING v10 (Szklarczyk et al., 2015). Coexpression or similar functional connections were determined at high confidence of ≥ 0.7. All predicted connections available in STRING v10 were also computed to predict the connectivity among all the 132 transcription factors (according to the Plant Transcription Factor Database v3.0; http://planttfdb.cbi.pku.edu.cn/) detected from the 2,740 genes that were specifically nitrate regulated in brace roots.

Accession Numbers

Raw sequencing data are stored at the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under accession number SRP062897.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Transverse sections from the tip of maize root types.

Supplemental Figure S2. Expression patterns of the pericycle transcriptomes of maize root types.

Supplemental Figure S3. Simplified hierarchical tree graphs of overrepresented GO terms.

Supplemental Figure S4. Quality assessment of the 24 LCM-dissected cell samples.

Supplemental Table S1. Anatomical characteristics of the four maize root types.

Supplemental Table S2. Overview of LCM-RNAseq output, mapping results and alignments to the reference genome.

Supplemental Table S3. "Classical" maize genes detected among nitrate-responsive genes.

Supplemental Table S4. List of 132 transcription factor genes.

Supplemental Table S5. List of genes associated with DNA synthesis/chromatin structure.

Supplemental Table S6. List of genes associated with cell cycle, cell division and cell organization.

Supplemental Data Set S1. Complete list of 20,179 expressed genes.

Supplemental Data Set S2. Complete list of 26,770 active/inactive genes.

Supplemental Data Set S3. Complete list of clustered (K1-K3) genes by K-means.

ACKNOWLEDGMENTS

We thank Mikio Nakazono (Nagoya University) for stimulating discussions on the optimization of laser capture microdissection and Nina Opitz (University of Bonn), Stefan Hey (University of Bonn), and Huanhuan Tai (University of Bonn) for helpful discussions on the data analyses.

Received December 2, 2015; accepted January 24, 2016; published January 25, 2016.

LITERATURE CITED

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. Available at www.bioinformatics.babraham.ac.uk/projects/fastqc/

Downloaded from https://academic.oup.com/plphys/article/170/3/1783/6114378 by ulbbonn user on 21 December 2021

Araya T, Miyamoto M, Wibowo J, Suzuki A, Kojima S, Tsuchiya YN, Sawa S, Fukuda H, von Wirén N, Takahashi H (2014) CLE-CLAVATA1 peptide-receptor signaling module regulates the expansion of plant root systems in a nitrogen-dependent manner. Proc Natl Acad Sci USA **111:** 2029–2034

Asada T, Collings D (1997) Molecular motors in higher plants. Trends Plant Sci **2:** 29–37

Atkinson JA, Rasmussen A, Traini R, Voß U, Sturrock C, Mooney SJ, Wells DM, Bennett MJ (2014) Branching out in roots: uncovering form, function, and regulation. Plant Physiol **166:** 538–550

Beeckman T, Burssens S, Inzé D (2001) The peri-cell-cycle in *Arabidopsis*. J Exp Bot **52:** 403–411

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc B **57:** 289–300

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat **29:** 1165–1188

Berckmans B, De Veylder L (2009) Transcriptional control of the cell cycle. Curr Opin Plant Biol **12:** 599–605

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics **30:** 2114–2120

Canales J, Moyano TC, Villarroel E, Gutiérrez RA (2014) Systems analysis of transcriptome data provides new hypotheses about *Arabidopsis* root response to nitrate treatments. Front Plant Sci **5:** 22

Dembinsky D, Woll K, Saleem M, Liu Y, Fu Y, Borsuk LA, Lamkemeyer T, Fladerer C, Madlung J, Barbazuk B, et al (2007) Transcriptomic and proteomic analyses of pericycle cells of the maize primary root. Plant Physiol **145:** 575–588

De Rybel B, Vassileva V, Parizot B, Demeulenaere M, Grunewald W, Audenaert D, Van Campenhout J, Overvoorde P, Jansen L, Vanneste S, et al (2010) A novel aux/IAA28 signaling cascade activates GATA23-dependent specification of lateral root founder cell identity. Curr Biol **20:** 1697–1706

De Smet I, Tetsumura T, De Rybel B, Frei dit Frey N, Laplaze L, Casimiro I, Swarup R, Naudts M, Vanneste S, Audenaert D, et al (2007) Auxin-dependent regulation of lateral root positioning in the basal meristem of *Arabidopsis*. Development **134:** 681–690

De Smet I, Vassileva V, De Rybel B, Levesque MP, Grunewald W, Van Damme D, Van Noorden G, Naudts M, Van Isterdael G, De Clercq R, et al (2008) Receptor-like kinase ACR4 restricts formative cell divisions in the *Arabidopsis* root. Science **322:** 594–597

Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res **38:** W64–W70

Dubrovsky JG, Sauer M, Napsucialy-Mendivil S, Ivanchenko MG, Friml J, Shishkova S, Celenza J, Benková E (2008) Auxin acts as a local morphogenetic trigger to specify lateral root founder cells. Proc Natl Acad Sci USA **105:** 8790–8794

Elledge SJ (1998) Mitotic arrest: Mad2 prevents sleepy from waking up the APC. Science **279:** 999–1000

Forde BG (2002) Local and long-range signaling pathways regulating plant responses to nitrate. Annu Rev Plant Biol **53:** 203–224

Gojon A, Nacry P, Davidian JC (2009) Root uptake regulation: a central process for NPS homeostasis in plants. Curr Opin Plant Biol **12:** 328–338

Guan P, Wang R, Nacry P, Breton G, Kay SA, Pruneda-Paz JL, Davani A, Crawford NM (2014) Nitrate foraging by *Arabidopsis* roots is mediated by the transcription factor TCP20 through the systemic signaling pathway. Proc Natl Acad Sci USA **111:** 15267–15272

Gutjahr C, Sawers RJH, Marti G, Andrés-Hernández L, Yang SY, Casieri L, Angliker H, Oakeley EJ, Wolfender JL, Abreu-Goodger C, Paszkowski U (2015) Transcriptome diversity among rice root types during asymbiosis and interaction with arbuscular mycorrhizal fungi. Proc Natl Acad Sci USA **112:** 6754–6759

Hetz W, Hochholdinger F, Schwall M, Feix G (1996) Isolation and characterization of *rtcs*, a maize mutant deficient in the formation of nodal roots. Plant J **10:** 845–857

Himanen K, Boucheron E, Vanneste S, de Almeida Engler J, Inzé D, Beeckman T (2002) Auxin-mediated cell cycle activation during early lateral root initiation. Plant Cell **14:** 2339–2351

Himanen K, Vuylsteke M, Vanneste S, Vercruysse S, Boucheron E, Alard P, Chriqui D, Van Montagu M, Inzé D, Beeckman T (2004) Transcript profiling of early lateral root initiation. Proc Natl Acad Sci USA **101:** 5146–5151

Hochholdinger F (2009) The maize root system: morphology, anatomy, and genetics. *In* JL Bennetzen, Hake SC, eds, Handbook of Maize: Its Biology. Springer, New York pp 145–160

Hochholdinger F, Feix G (1998) Early post-embryonic root formation is specifically affected in the maize mutant *lrt1*. Plant J **16:** 247–255

Hochholdinger F, Park WJ, Sauer M, Woll K (2004b) From weeds to crops: genetic analysis of root development in cereals. Trends Plant Sci **9:** 42–48

Hochholdinger F, Woll K, Sauer M, Dembinsky D (2004a) Genetic dissection of root formation in maize (*Zea mays*) reveals root-type specific developmental programmes. Ann Bot (Lond) **93:** 359–368

Hochholdinger F, Zimmermann R (2008) Conserved and diverse mechanisms in root development. Curr Opin Plant Biol **11:** 70–74

Ito M, Araki S, Matsunaga S, Itoh T, Nishihama R, Machida Y, Doonan JH, Watanabe A (2001) G2/M-phase-specific transcription during the plant cell cycle is mediated by c-Myb-like transcription factors. Plant Cell **13:** 1891–1905

Jansen L, Hollunder J, Roberts I, Forestan C, Fonteyne P, Van Quickenborne C, Zhen RG, McKersie B, Parizot B, Beeckman T (2013) Comparative transcriptomics as a tool for the identification of root branching genes in maize. Plant Biotechnol J **11:** 1092–1102

Jansen L, Roberts I, De Rycke R, Beeckman T (2012) Phloem-associated auxin response maxima determine radial positioning of lateral roots in maize. Philos Trans R Soc Lond B Biol Sci **367:** 1525–1533

Laskowski M, Grieneisen VA, Hofhuis H, Hove CA, Hogeweg P, Marée AF, Scheres B (2008) Root system architecture from coupling cell shape to auxin transport. PLoS Biol **6:** e307

Law CW, Chen Y, Shi W, Smyth GK (2014) voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol **15:** R29

Linkohr BI, Williamson LC, Fitter AH, Leyser HM (2002) Nitrate and phosphate availability and distribution have different effects on root system architecture of *Arabidopsis*. Plant J **29:** 751–760

Malamy JE (2005) Intrinsic and environmental response pathways that regulate root system architecture. Plant Cell Environ **28:** 67–77

Malamy JE, Benfey PN (1997) Organization and cell differentiation in lateral roots of *Arabidopsis thaliana*. Development **124:** 33–44

Marschner H (1995) Mineral Nutrition of Higher Plants. Academic Press, London

McIntosh JR, Grishchuk EL, West RR (2002) Chromosome-microtubule interactions during mitosis. Annu Rev Cell Dev Biol **18:** 193–219

Moreno-Risueno MA, Van Norman JM, Moreno A, Zhang J, Ahnert SE, Benfey PN (2010) Oscillating gene expression determines competence for periodic *Arabidopsis* root branching. Science **329:** 1306–1311

Mu RL, Cao YR, Liu YF, Lei G, Zou HF, Liao Y, Wang HW, Zhang WK, Ma B, Du JZ, et al (2009) An R2R3-type transcription factor gene *AtMYB59* regulates root growth and cell cycle progression in *Arabidopsis*. Cell Res **19:** 1291–1304

Opitz N, Marcon C, Paschold A, Malik WA, Lithio A, Brandt R, Piepho H-P, Nettleton D, Hochholdinger F (2015) Extensive tissue-specific transcriptomic plasticity in maize primary roots upon water deficit. J Exp Bot **pii:** erv453

Osley MA (1991) The regulation of histone synthesis in the cell cycle. Annu Rev Biochem **60:** 827–861

Parizot B, Laplaze L, Ricaud L, Boucheron-Dubuisson E, Bayle V, Bonke M, De Smet I, Poethig SR, Helariutta Y, Haseloff J, et al (2008) Diarch symmetry of the vascular bundle in Arabidopsis root encompasses the pericycle and is reflected in distich lateral root initiation. Plant Physiol **146:** 140–148

Peng Y, Niu J, Peng Z, Zhang F, Li C (2010) Shoot growth potential drives N uptake in maize plants and correlates with root growth in the soil. Field Crops Res **115:** 85–93

Polyn S, Willems A, De Veylder L (2015) Cell cycle entry, maintenance, and exit during plant development. Curr Opin Plant Biol **23:** 1–7

Postma JA, Dathe A, Lynch JP (2014) The optimal lateral root branching density for maize depends on nitrogen and phosphorus availability. Plant Physiol **166:** 590–602

Rogers ED, Benfey PN (2015) Regulation of plant root system architecture: implications for crop advancement. Curr Opin Biotechnol **32:** 93–98

Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc B **71:** 319–392

Ruffel S, Krouk G, Ristova D, Shasha D, Birnbaum KD, Coruzzi GM (2011) Nitrogen economics of root foraging: transitive closure of the

Downloaded from https://academic.oup.com/plphys/article/170/3/1783/6114378 by ulbbonn user on 21 December 2021

nitrate-cytokinin relay and distinct systemic signaling for N supply vs. demand. Proc Natl Acad Sci USA **108:** 18524–18529

Schnable JC, Freeling M (2011) Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS One **6:** e17855

Shane MW, McCully ME (1999) Root xylem embolisms: implications for water flow to the shoot in single-rooted maize plants. Aust J Plant Physiol **26:** 107–114

Shin R, Burch AY, Huppert KA, Tiwari SB, Murphy AS, Guilfoyle TJ, Schachtman DP (2007) The *Arabidopsis* transcription factor MYB77 modulates auxin signal transduction. Plant Cell **19:** 2440–2453

Smith LG (2001) Plant cell division: building walls in the right places. Nat Rev Mol Cell Biol **2:** 33–39

Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol **3:** 1–25

Smyth GK (2005) Limma: linear models for microarray data. *In* RC Gentleman, VJ Carey, W Huber, RA Irizarry, S Dudoit, eds, Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer, New York, pp 397–420

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, et al (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res **43:** D447–D452

Tabata R, Sumida K, Yoshii T, Ohyama K, Shinohara H, Matsubayashi Y (2014) Perception of root-derived peptides by shoot LRR-RKs mediates systemic N-demand signaling. Science **346:** 343–346

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J **37:** 914–939

Van De Wiel MA, Leday GGR, Pardo L, Rue H, Van Der Vaart AW, Van Wieringen WN (2013) Bayesian analysis of RNA sequencing data by estimating multiple shrinkage priors. Biostatistics **14:** 113–128

Vanneste S, De Rybel B, Beemster GTS, Ljung K, De Smet I, Van Isterdael G, Naudts M, Iida R, Gruissem W, Tasaka M, et al (2005) Cell cycle progression in the pericycle is not sufficient for SOLITARY ROOT/IAA14-mediated lateral root initiation in *Arabidopsis thaliana.* Plant Cell **17:** 3035–3050

Vidal EA, Moyano TC, Riveras E, Contreras-López O, Gutiérrez RA (2013) Systems approaches map regulatory networks downstream of the auxin receptor AFB3 in the nitrate response of *Arabidopsis thaliana* roots. Proc Natl Acad Sci USA **110:** 12840–12845

Woll K, Borsuk LA, Stransky H, Nettleton D, Schnable PS, Hochholdinger F (2005) Isolation, characterization, and pericycle-specific transcriptome analyses of the novel maize lateral and seminal root initiation mutant *rum1.* Plant Physiol **139:** 1255–1267

Yang Z, Gu S, Wang X, Li W, Tang Z, Xu C (2008) Molecular evolution of the CPP-like gene family in plants: insights from comparative genomics of *Arabidopsis* and rice. J Mol Evol **67:** 266–277

Yeung KY, Fraley C, Murua A, Raftery AE, Ruzzo WL (2001) Model-based clustering and data transformations for gene expression data. Bioinformatics **17:** 977–987

Yu HG, Muszynski MG, Kelly Dawe R (1999) The maize homologue of the cell cycle checkpoint protein MAD2 reveals kinetochore substructure and contrasting mitotic and meiotic localization patterns. J Cell Biol **145:** 425–435

Yu P, Eggert K, von Wirén N, Li C, Hochholdinger F (2015a) Cell-type specific gene expression analyses by RNA-Seq reveal local high nitrate triggered lateral root initiation in shoot-borne roots of maize by modulating auxin-related cell cycle-regulation. Plant Physiol **169:** 690–704

Yu P, Hochholdinger F, Li C (2015b) Root-type-specific plasticity in response to localized high nitrate supply in maize (*Zea mays*). Ann Bot (Lond) **116:** 751–762

Yu P, Li X, Yuan L, Li C (2014a) A novel morphological response of maize (*Zea mays*) adult roots to heterogeneous nitrate supply revealed by a split-root experiment. Physiol Plant **150:** 133–144

Yu P, White PJ, Hochholdinger F, Li C (2014b) Phenotypic plasticity of the maize root system in response to heterogeneous nitrogen availability. Planta **240:** 667–678

Zhan A, Schneider H, Lynch JP (2015) Reduced lateral root branching density improves drought tolerance in maize. Plant Physiol **168:** 1603–1615

Zhang H, Forde BG (1998) An *Arabidopsis* MADS box gene that controls nutrient-induced changes in root architecture. Science **279:** 407–409

Downloaded from https://academic.oup.com/plphys/article/170/3/1783/6114378 by ulbbonn user on 21 December 2021

Nonsyntenic Genes Drive Tissue-Specific Dynamics of Differential, Nonadditive, and Allelic Expression Patterns in Maize Hybrids[1][OPEN]

Jutta A. Baldauf, Caroline Marcon, Anja Paschold[2], and Frank Hochholdinger*

Institute of Crop Science and Resource Conservation, Crop Functional Genomics, University of Bonn, 53113 Bonn, Germany

ORCID IDs: 0000-0002-7286-0558 (J.A.B.); 0000-0001-6699-9441 (C.M.); 0000-0001-6465-2882 (A.P.); 0000-0002-5155-0884 (F.H.).

Distantly related maize (*Zea mays*) inbred lines display an exceptional degree of genomic diversity. F_1 progeny of such inbred lines are often more vigorous than their parents, a phenomenon known as heterosis. In this study, we investigated how the genetic divergence of the maize inbred lines B73 and Mo17 and their F_1 hybrid progeny is reflected in differential, nonadditive, and allelic expression patterns in primary root tissues. In pairwise comparisons of the four genotypes, the number of differentially expressed genes between the two parental inbred lines significantly exceeded those of parent versus hybrid comparisons in all four tissues under analysis. No differentially expressed genes were detected between reciprocal hybrids, which share the same nuclear genome. Moreover, hundreds of nonadditive and allelic expression ratios that were different from the expression ratios of the parents were observed in the reciprocal hybrids. The overlap of both nonadditive and allelic expression patterns in the reciprocal hybrids significantly exceeded the expected values. For all studied types of expression - differential, nonadditive, and allelic - substantial tissue-specific plasticity was observed. Significantly, nonsyntenic genes that evolved after the last whole genome duplication of a maize progenitor from genes with synteny to sorghum (*Sorghum bicolor*) were highly overrepresented among differential, nonadditive, and allelic expression patterns compared with the fraction of these genes among all expressed genes. This observation underscores the role of nonsyntenic genes in shaping the transcriptomic landscape of maize hybrids during the early developmental manifestation of heterosis in root tissues of maize hybrids.

Maize (*Zea mays*) is one of the most prolific crops with a global production of 1.02 billion tons in 2014. In the past 50 years, global maize production has increased almost 5-fold (http://faostat3.fao.org/). Between 50% to 60% of this increase has been attributed to genetic improvements in hybrid breeding (Duvick, 2005).

Heterosis, or hybrid vigor, describes the superior performance of heterozygous F_1 hybrid progeny relative to the average of their homozygous parental inbred lines (midparent heterosis) or the better performing parent (best parent heterosis; Falconer and Mackay, 1996). The phenomenon of heterosis was first scientifically reported by Charles Darwin after he compared the performance of self-pollinated maize progeny to cross-pollinated maize plants (Darwin, 1876). In the early twentieth century, heterosis was rediscovered by Edward M. East (East, 1908) and George H. Shull (Shull, 1908). The first commercial hybrid maize was introduced in the 1930s (Duvick, 2005). Since then, heterosis has been widely exploited in plant breeding and agriculture. Approximately 95% of U.S. maize and 55% of the rice (*Oryza sativa*) acreage in China are planted with hybrid seeds (Ding et al., 2014).

The phenotypic effect of heterosis is most evident in adult plants and can be monitored for instance as increased plant height, biomass, yield, fertility, and improved resistance to abiotic or biotic stress (Falconer and Mackay, 1996). Nevertheless, it has been demonstrated that heterosis can already be observed a few days after germination for seedling root traits such as lateral root density or primary root length (Hoecker et al., 2006). Several genetic hypotheses have been proposed that make the combination of a remarkable number of genes responsible for heterosis. These genetic hypotheses attribute heterosis to the complementation of superior dominant alleles (dominance hypothesis), allelic interactions at one or multiple loci (overdominance hypothesis), or epistatic interactions between nonallelic genes (for review, see Birchler et al., 2006; Hochholdinger and Hoecker, 2007).

On the level of gene expression, the prevalence of expression patterns in hybrids different from the

[1] This work was supported by the Deutsche Forschungsgemeinschaft (Grant HO2249-9/3 to F.H.).

[2] Present address: Max-Planck-Institute for Plant Breeding Research, 50829 Cologne, Germany.

* Address correspondence to hochholdinger@uni-bonn.de.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Frank Hochholdinger (hochholdinger@uni-bonn.de).

J.A.B., A.P., and F.H. conceived the research; J.A.B. carried out the statistical analyses, interpreted the data, and drafted the article; C.M. and F.H. participated in data interpretation; F.H. coordinated the study and helped to draft the article; all authors approved the final draft of the article.

[OPEN] Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.16.00262

 © 2016 American Society of Plant Biologists. All Rights Reserved.

average of their parental values, i.e. nonadditive expression, widely varied between different species, tissues, and genotypes (Guo et al., 2006; Stupar and Springer, 2006; Swanson-Wagner et al., 2006; Uzarowska et al., 2007; Stupar et al., 2008; Hoecker et al., 2008b; Paschold et al., 2012). Nonadditive gene expression is the result of allelic interactions that modify regulatory networks that lead to gene activity patterns different from average parental values (Stupar and Springer, 2006). Furthermore, unexpected allelic expression patterns were observed in hybrids that deviated from the allelic ratios of their parents (Guo et al., 2004; Stupar and Springer, 2006; Springer and Stupar, 2007; Paschold et al., 2012; Song et al., 2013). Although nonadditive expression patterns and unexpected allelic ratios have been observed in hybrids of different species, none of these expression types have yet been associated with the evolutionary origin of these genes.

The evolutionary history of flowering plants included several rounds of hybridization and polyploidization events followed by fractionation and rediploidization (for review, see Doyle et al., 2008). Within the lineage leading to maize, a last whole genome duplication occurred between ~4.8 and ~11.9 million years ago by the hybridization of two progenitors, which led to a tetraploid genome (Swigonová et al., 2004). Comparisons of the duplicated regions of the genome of modern maize with orthologous regions of the unduplicated genomes of rice (*Oryza sativa*) and sorghum (*Sorghum bicolor*) indicated that the maize genome had lost many of its duplicated genes (Haberer et al., 2005). Based on the synteny to sorghum, the genome of maize can be subdivided into two subgenomes, which contain pairs of duplicated genes shared by both subgenomes but also single-copy genes present in only one of the two subgenomes (Schnable et al., 2011). In total, 19,365 genes, representing 49% of the 39,656 maize genes of the filtered gene set (FGSv2), were assigned to these two subgenomes (Schnable et al., 2011). Beside these syntenic maize genes, the genome of modern maize is complemented by a set of nonsyntenic genes that lack syntenic orthologs in other grass species (Schnable et al., 2011). These 20,291 (51%) maize genes (FGSv2) likely evolved by single gene duplication after the last whole genome duplication event (Woodhouse et al., 2010).

Along the longitudinal axis, maize roots are structurally divided into a subterminal meristematic zone at the terminal end, followed by the elongation and differentiation zones (Hochholdinger, 2009). In the meristematic zone, new root cells are formed by cell division and are dislocated proximally into the elongation zone, where they elongate. Finally, cells are translocated into the differentiation zone. Cells in the differentiation zone display diverse functions. In transverse orientation, the stele with the pericycle as its outermost cell layer contains xylem and primary phloem elements functioning in the transport of water, nutrients, and photosynthates. The surrounding cortical parenchyma consists of the endodermis, multiple layers of cortex tissue, and the epidermis that connects the root to the rhizosphere. Hence, roots represent gradients of development, with very young undifferentiated cells at the distal end near the root cap and fully differentiated cells toward the proximal end of the root (Hochholdinger, 2009).

A previous study of our research group revealed dynamic root tissue-specific patterns of single-parent expression (SPE), an extreme instance of expression complementation in which genes are expressed in both reciprocal hybrids but only in one parental inbred line (Paschold et al., 2014). In this study, this dataset was surveyed to determine tissue-specific nonadditive, differential, and allelic gene expression patterns and the relation of these expression patterns to the phylogenetic origin of these genes. We demonstrated that nonsyntenic genes were highly overrepresented among nonadditive and allelic expression patterns in hybrids, suggesting a possible role of these genes in the early developmental manifestation of heterosis in maize roots.

RESULTS

Transcriptome Relationships and Tissue-Specific Expression of Four Primary Root Tissues of the Maize Inbred Lines B73 and Mo17 and Their Reciprocal Hybrids

Relationships of RNA-sequencing (RNA-seq) samples of the meristematic zone, the elongation zone, and the cortex and stele of the differentiation zone of maize primary roots (Fig. 1A) of the inbred lines B73 and Mo17 and their reciprocal F1 hybrids B73xMo17 and Mo17xB73 were explored in a hierarchical cluster analysis (Fig. 1B) and in a multidimensional scaling plot (Fig. 1C). Each of the sixteen genotype/tissue combinations was represented by four biological replicates. The replicated samples of each root tissue marked with the same color code formed a cluster irrespective of their genotype. This indicates that the four distinct tissues of a single genotype display a higher degree of transcriptomic dissimilarity than the same tissues of different genotypes. Furthermore, among the four tissues, the meristematic and the elongation zone displayed the highest similarity (Fig. 1, B and C). On the genotype level, the transcriptomes of the two reciprocal hybrids that contain an identical nuclear genome were closely related in all four tissues, while the transcriptomes of the parental inbred lines were more distantly related (Fig. 1B).

Determination of Differentially Expressed Genes between Different Tissues

To determine the number of differentially expressed genes between the four different primary root tissues, six pairwise comparisons were performed per genotype (Fig. 2). For each comparison two bars are displayed in Figure 2 indicating the number of genes preferentially expressed in the color-coded tissue. In all four genotypes, the relative number of differentially expressed

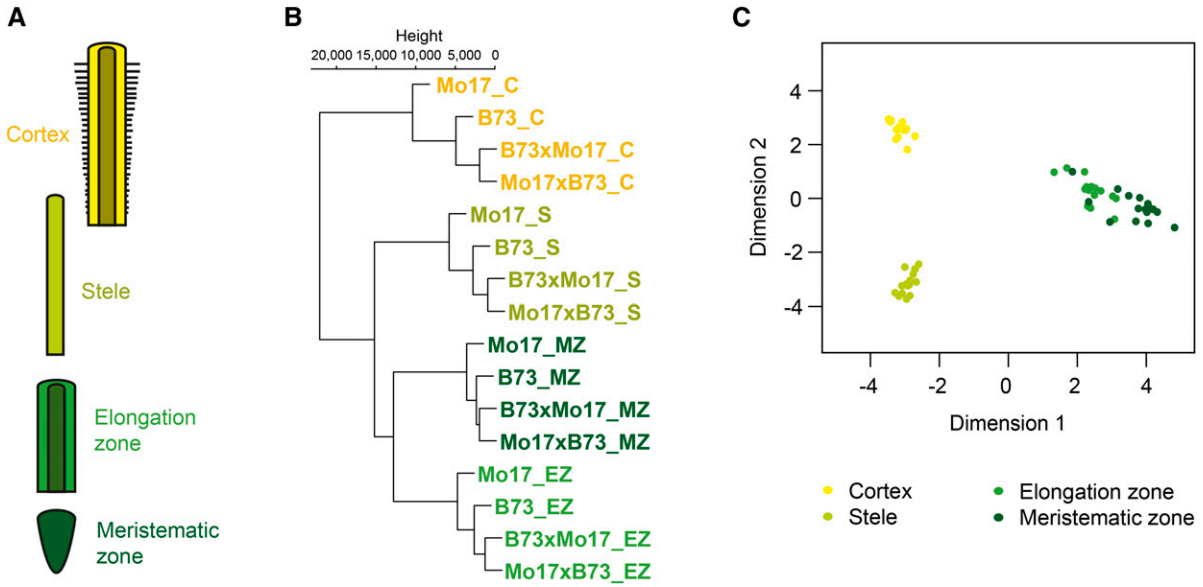


Figure 1. Overview of RNA-seq sample relationship. A, Color-coded schematic of the four mechanically separated maize primary root tissues: yellow, cortex (C); light green, stele (S); green, elongation zone (EZ); dark green, meristematic zone (MZ). B, Hierarchical cluster analysis of normalized read counts among all four replicates of each RNA-seq sample. C, Multidimensional scaling plot of all 64 RNA-seq samples.

genes in the six comparisons was similar. The comparison between cortex and stele, which are both complementary tissues of the differentiation zone and between the elongation and the meristematic zone, revealed the smallest numbers of differentially expressed genes (Fig. 2; Supplemental Data File S1). In contrast, the highest numbers of differentially expressed genes were observed in the comparisons of the meristematic zone with cortex and with stele. These tissues are located at opposite poles of the root, and the high number of differentially expressed genes underscores the dissimilarity between these primary root tissues. In all three pairwise comparisons of the stele with other tissues, a higher number of genes was preferentially expressed in the stele compared with the other tissues (Fig. 2). In pairwise comparisons of the remaining three tissues, a higher number of genes was predominantly expressed in the meristematic zone compared with cortex or elongation zone, while in the comparison of cortex versus elongation zone a higher number of differentially expressed genes was preferentially expressed in the cortex. These tendencies were observed for all comparisons and were statistically significant for 22 of 24 comparisons (Fig. 2). Cross-comparison of the differentially expressed genes between the different root tissues among the four maize genotypes B73, Mo17, and their reciprocal hybrids showed that between 44% (cortex *versus* stele) and 68% (cortex *versus* meristematic zone) of differentially expressed genes were conserved among the four genotypes (Supplemental Fig. S1). Among the genes that were differentially expressed in only one genotype, the parental inbred lines displayed between 2 to 5 times more such genes than the reciprocal hybrids (Supplemental Fig. S1).

Determination of Differentially Expressed Genes between Different Genotypes

Differential gene expression between the four different genotypes was determined for each primary root tissue (Fig. 3). For all four tissues, the pairwise comparison between the two parental inbred lines B73 and Mo17 revealed the highest number of differentially expressed genes. In contrast, no differential gene expression was observed between the two reciprocal hybrids for any of the four tissues. For all four possible parent-hybrid comparisons, in each tissue between 1,104 (B73 *versus* B73xMo17 in elongation zone) and 1,582 (Mo17 *versus* B73xMo17 in stele) differentially expressed genes were estimated by controlling the false discovery rate (FDR) at 5% and an absolute $\log_2$ fold change (FC) > 1 (Fig. 3; Supplemental Data File S1). Hence, the number of differentially expressed genes in parent-hybrid comparisons was substantially below parent-parent comparisons but higher than in hybrid-hybrid comparisons. These numbers reflect the genetic relationship of the analyzed genotypes. While the reciprocal hybrids, which contain an identical nuclear genome, did not display any differentially expressed genes, the genetically distinct parents displayed a high degree of differential gene expression. Intermediate numbers of differentially expressed genes were observed in comparisons of hybrids with one of the parental inbred lines. Hybrids and parental inbred lines share 50% of their nuclear genomes. In all inbred *versus* hybrid comparisons, a significantly higher number of differentially expressed genes was preferentially expressed in the hybrids. This conserved trend was

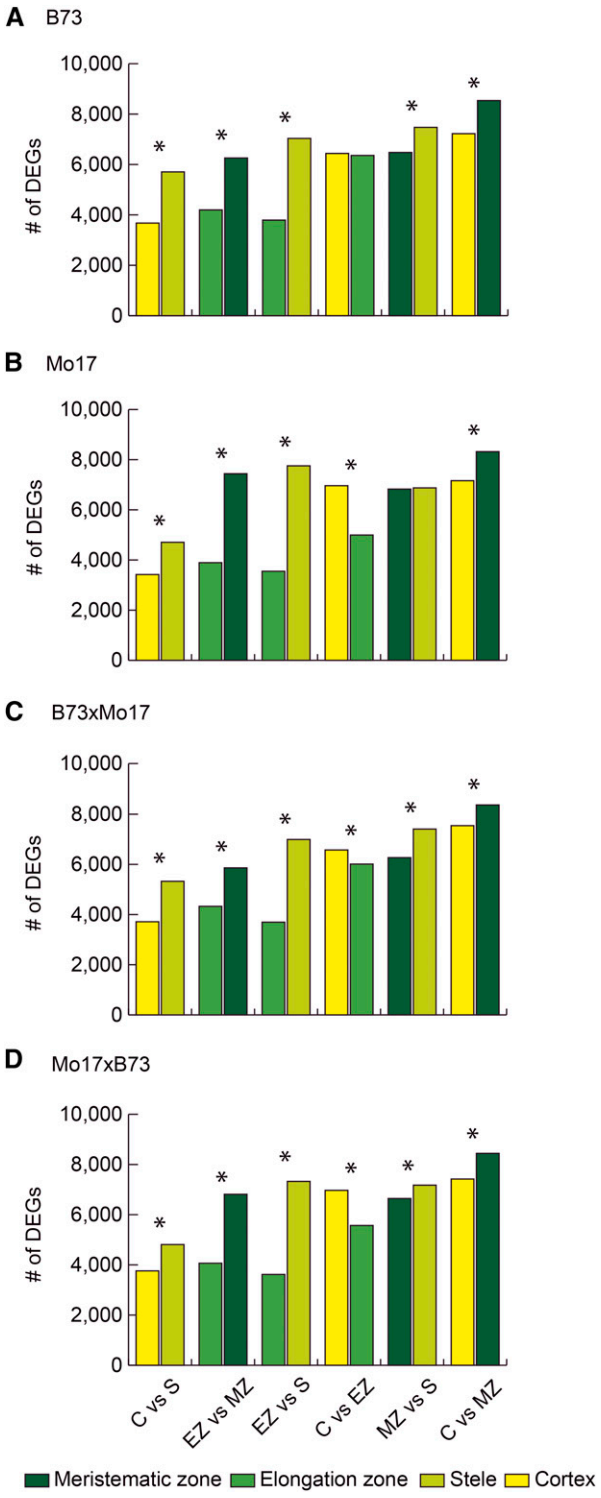


Figure 2. Differential expression between four different maize primary root tissues. Differentially expressed genes (DEGs) were identified on the basis of six pairwise comparisons between the four analyzed root tissues: cortex (C, yellow), stele (S, light green), elongation zone (EZ, green), and meristematic zone (MZ, dark green) among the four maize genotypes B73 (A), Mo17 (B), and their reciprocal hybrids B73xMo17 (C) and Mo17xB73 (D). The number of differentially expressed genes while controlling the FDR at 5% and with a cutoff of $|\log_2(FC)| > 1$ are shown. For each pairwise comparison, two color-coded bars illustrate the tissue preference of the differentially expressed genes. Significant differences ($\alpha \leq 5\%$) between the two bars per comparison are indicated by an asterisk.

statistically significant for 15 of 16 comparisons (Fig. 3). Irrespective of the compared genotypes, the overlap of differentially expressed genes showed that fewer genes (between 7% and 9%) were differentially expressed in all four primary root tissues, and a substantially higher number of genes (between 28% and 47%) was exclusively expressed in one tissue (Supplemental Fig. S2).

Genes Exhibiting Unexpected Allelic Expression Ratios in Hybrids Are Tissue-Specific

For each tissue, the allelic ratio was determined in the hybrids and compared with the allelic expression ratio of the two parental inbred lines. The number of genes for which these ratios were significantly different (unexpected allelic ratios) is summarized in Figure 4A. In total, between 87 (Mo17xB73 in cortex) and 228 (B73xMo17 in stele) genes displayed unexpected allelic ratios (Supplemental Table S1, column Observed). Overall, between 39% (41/106 in elongation zone) and 54% (47/87 in cortex) of unexpected allelic ratios were conserved between the reciprocal hybrids (Fig. 3A, dark shaded bars), which was significantly higher than expected (Supplemental Table S1, column Expected) as calculated by a χ^2 test. For each hybrid, the genes displaying unexpected allelic expression ratios were compared between the four primary root tissues. For B73xMo17, this comparison indicated that 74% to 84% of the genes with unexpected allelic ratios were exclusively expressed in one primary root tissue and only 1.3% of these genes were common in all four root tissues (Fig. 4B). Similarly, for Mo17xB73, between 76% and 89% of the genes with differential allelic expression were tissue-specific and only 1.4% of these genes were expressed in all four tissues (Fig. 4C). In total, 739 unique genes displaying unexpected allelic expression ratios were identified in this study (Supplemental Data File S1).

Nonadditive Gene Expression Patterns Are Conserved between Reciprocal Hybrids in All Four Primary Root Tissues

For each tissue, gene expression levels in the reciprocal hybrids relative to their parental inbred lines were characterized based on nine gene expression classes, which were derived from the classes previously defined by Hoecker et al. (2008a). Subsequently, the subset of genes showing expression levels that deviate significantly from the midparent value was determined in these nine classes. The number of genes in each expression class was estimated by combining three pairwise *t* tests that compared the expression level of the

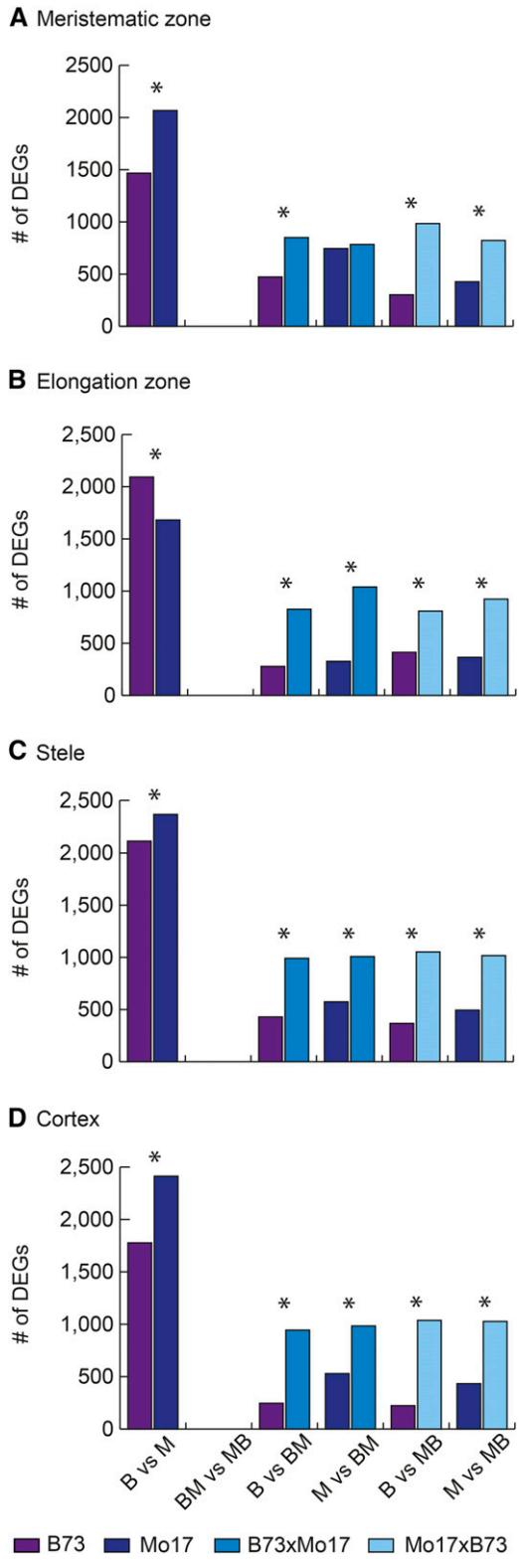


Figure 3. Differential expression between four different maize genotypes. Differentially expressed genes (DEGs) were identified on the basis of six pairwise comparisons between the four analyzed maize genotypes: B73 (B, purple), Mo17 (M, dark blue), B73xMo17 (BM, blue), and Mo17xB73 (MB, light blue) for each of the four primary root tissues: meristematic zone (A), elongation zone (B), stele (C), and cortex (D). The number of differentially expressed genes while controlling the FDR at 5% and with a cutoff of $|\log_2(FC)| > 1$ are displayed. For each pairwise comparison, two color-coded bars illustrate the genotype preference of the differentially expressed genes. Significant differences ($\alpha \leq 5\%$) between the two bars per comparison are indicated by an asterisk.

hybrid with each parent and between both parents (Supplemental Table S2). According to this classification scheme most of the expressed genes in the hybrids revealed no significant difference in their expression relative to their parents (class 4). Genes that show expression levels between the parental values or exhibited low- or high-parent expression were abundant and were assigned to expression classes 1, 2, and 3, respectively. Genes that exhibited either above high-parent (36 genes in B73xMo17; 17 genes in Mo17xB73) or below low-parent (147 genes in B73xMo17; 36 genes in Mo17xB73) expression in the hybrids were assigned to the expression classes 5 to 8. Genes with ambiguous gene expression levels in the hybrids were assigned to class 9. In total, 1,056 and 824 nonadditively expressed genes were exclusively identified in the hybrids B73xMo17 and Mo17xB73, respectively. Among those, 439 nonadditive genes overlapped between the reciprocal hybrids, which substantially exceeded the expected number of 35 overlapping genes as demonstrated by a χ^2 test (Supplemental Table S2).

Each expression class contained a subset of genes that exhibited nonadditive gene expression, i.e. gene expression in these hybrids was significantly different from the mean of expression of the two parental inbred lines. The highest number of nonadditively expressed genes was observed in gene expression class 1 (between 205 in cortex and 432 in meristematic zone) and class 3 (between 201 in meristematic zone and 339 in elongation zone; Supplemental Table S2). Within each class, the overlap of the nonadditively expressed genes between the two hybrids was significantly higher than expected. Lower numbers of nonadditively expressed genes were observed in gene expression class 2 and the classes to which genes with extreme expression patterns were assigned to (classes 5–8). For each tissue, the observed number of common genes in both hybrids significantly exceeded the number of genes expected purely by chance. Consistent with the definition of the gene expression pattern of class 4, no nonadditively expressed genes were observed in this class.

Among the genes with unexpected allelic ratios, between 26 (Mo17xB73, elongation zone) and 58 (B73xMo17, meristematic zone) genes showed nonadditive gene expression patterns (Supplemental Table S1). The comparison of the nonadditively expressed genes with unexpected allelic ratios between the two reciprocal hybrids B73xMo17 and Mo17xB73 within each primary root tissue revealed that a significantly higher number of these genes is common in both hybrids than expected.

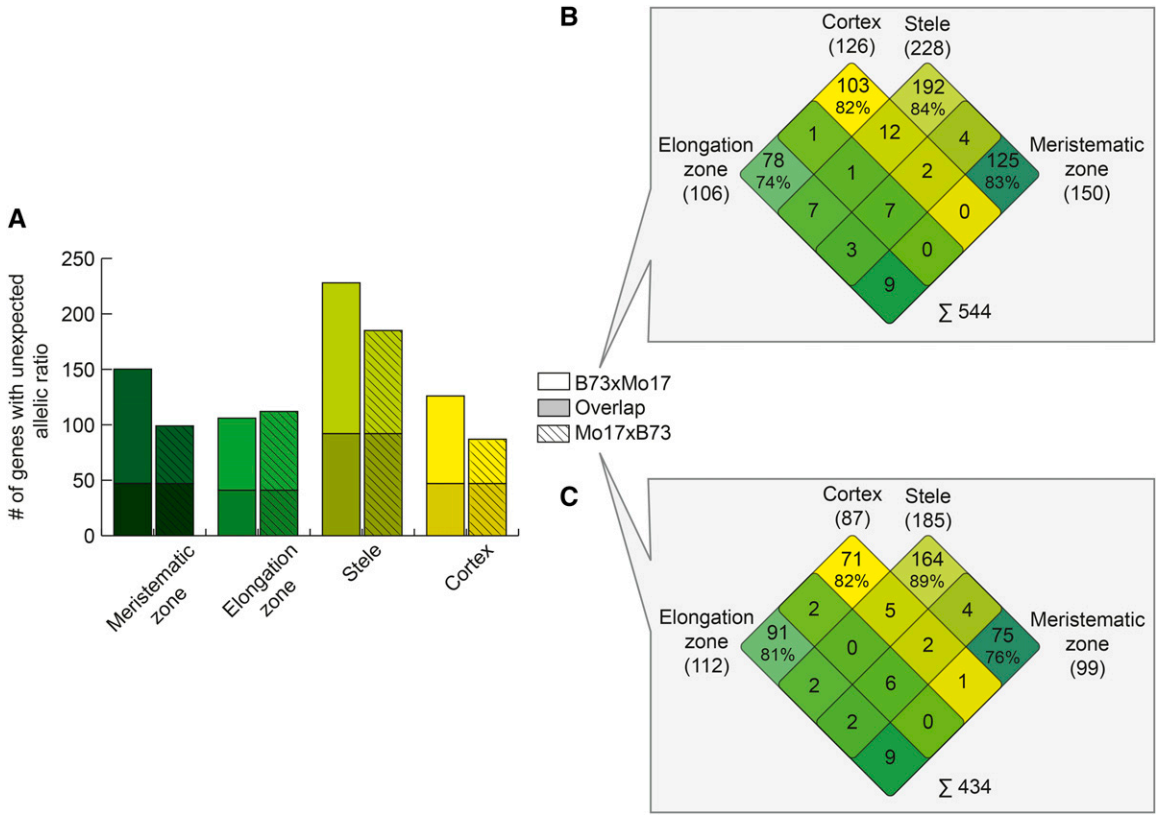


Figure 4. Genes with unexpected allelic expression in the hybrids. A, Number of genes showing unexpected allelic ratios in both hybrids (plain: B73xMo17; shaded: Mo17xB73) for each primary root tissue. The overlapping genes between the two hybrids within each tissue are highlighted by the darker color. B and C, Overlap of genes with unexpected allelic ratio between the four different primary root tissues in B73xMo17 (B) and in Mo17xB73 (C). The numbers in parentheses indicate the total number of genes with unexpected allelic ratio in the corresponding root tissue.

Nonadditively, Allelic, and Differentially Expressed Genes Are Overrepresented among Nonsyntenic Genes

To study their evolutionary origin, differentially and nonadditively expressed genes, and genes exhibiting unexpected allelic ratios, were compared with the set of nonsyntenic genes that constitute 51% of all genes in the filtered gene set (FGSv2) of the maize genome (Schnable et al., 2011). Among the genes expressed in each of the four tissues, only between 31% (6,653/21,276 in cortex) and 33% (6,860/20,899 in meristematic zone) were nonsyntenic (Fig. 5A; Supplemental Data File S1). Hence, nonsyntenic genes were substantially underrepresented among all expressed genes relative to their fraction of all genes present in FGSv2. Remarkably, in the elongation zone, 68% (501/740) of the nonadditively expressed genes were nonsyntenic and thus significantly overrepresented compared with their prevalence among all expressed genes (Fig. 5A). Similar values were observed for nonadditively expressed genes in the remaining three primary root tissues investigated: cortex 61% (407/666), meristematic zone 60% (667/1,115), and stele 63% (435/686). Similarly, nonsyntenic genes were significantly overrepresented among the genes with unexpected allelic expression ratios relative to all expressed genes with single nucleotide polymorphisms (SNPs) in each of the four root tissues (Fig. 5B). Among the expressed genes with SNPs, only between 28% (3,596/12,963 in cortex) and 30% (3,809/12,903 in meristematic zone) were nonsyntenic. In comparison, 47% (78/166) of genes with unexpected allelic ratios were nonsyntenic in cortex, 41% (73/177) in the elongation zone, 58% (118/202) in the meristematic zone, and 39% (124/321) in the stele (Fig. 5B).

Furthermore, nonsyntenic genes were significantly overrepresented among genes differentially expressed between the four genotypes compared with all expressed genes (Fig. 5C; Supplemental Data File S1). Their fraction ranged from 41% (1,704/4,190 of B73 *versus* Mo17 in cortex) to 62% (818/1,323 of B73 *versus* B73xMo17 in meristematic zone; Fig. 5C). However, among all expressed genes in the different tissues, only between 30% (cortex) and 33% (meristematic zone) were nonsyntenic.

Functional Classification of Nonsyntenic Nonadditively Expressed Genes

To identify overrepresented biological and molecular functions among the nonsyntenic genes with nonadditive and allelic expression patterns, a Gene Ontology

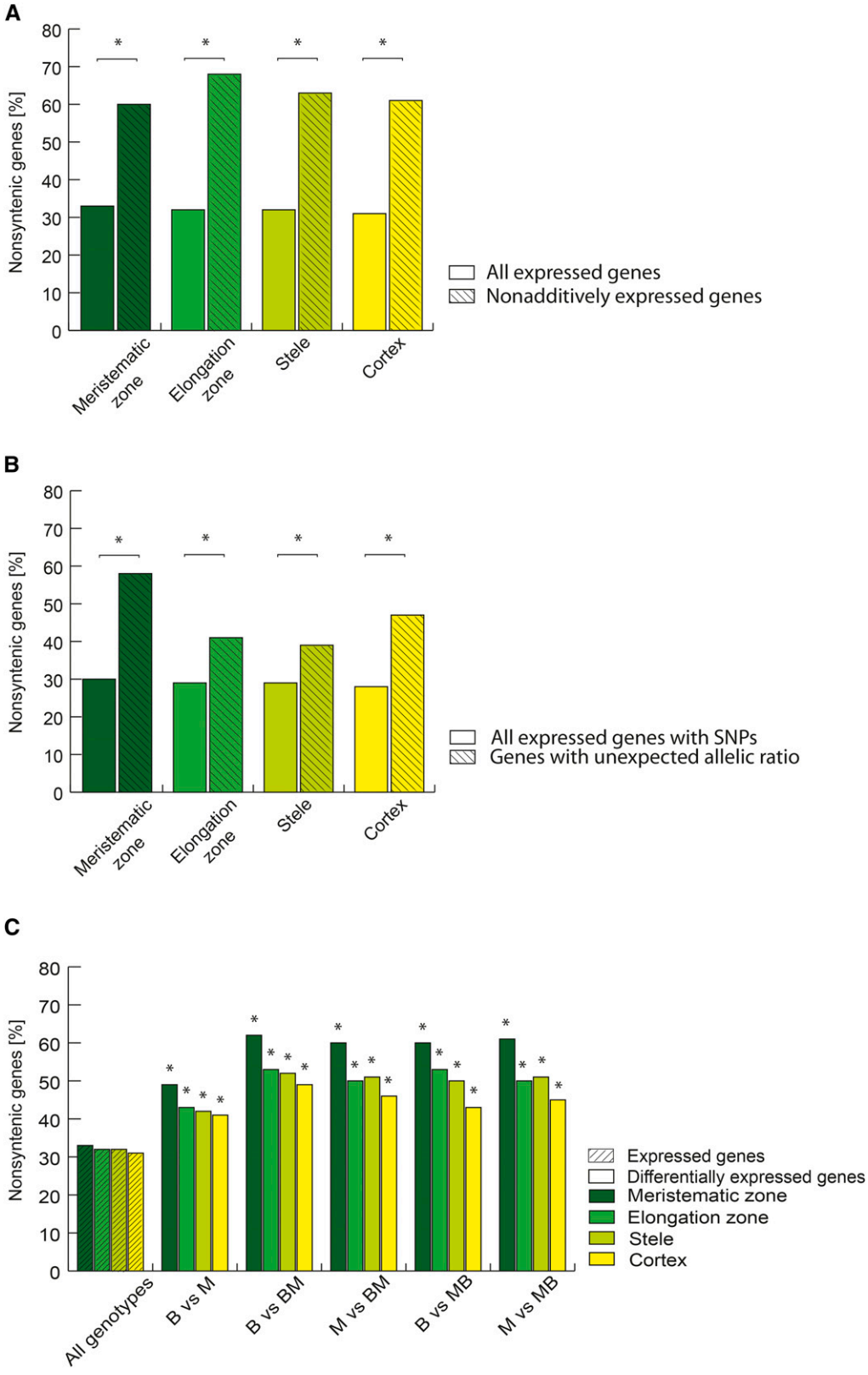


Figure 5. Distribution of nonsyntenic genes among different classes of genes. A, Proportion of nonsyntenic genes among non-additively expressed genes (shaded) in the hybrids relative to all expressed genes (not shaded) in each of the four primary root tissues. B, Proportion of nonsyntenic genes among genes showing unexpected allelic ratios (shaded) in the hybrids relative to all

(GO) analysis was performed. Only for the nonsyntenic nonadditively expressed genes of the cortex and the elongation zone, significantly enriched GO terms were identified (Supplemental Table S3). Most prominently, the biological processes "death" and "cell death" with the subcategories "programmed cell death" and "apoptosis" were overrepresented in the cortex samples of the reciprocal hybrids. Moreover, in B73xMo17, the molecular processes "iron ion binding" and "tetrapyrrole binding" with its subcategory "heme binding" were significantly enriched. Furthermore, in the reciprocal hybrid Mo17xB73, the molecular process "monooxygenase activity" was overrepresented. Finally, in the elongation zone, the molecular process "aspartic-type peptidase activity" and its subcategory "aspartic-type endopeptidase activity" were overrepresented among the nonsyntenic nonadditively expressed genes in Mo17xB73.

DISCUSSION

In this study, differential, nonadditive, and allelic gene expression patterns of the maize inbred lines B73 and Mo17 and their reciprocal hybrid progeny were analyzed in four maize primary root tissues (Fig. 1A), based on a RNA-seq dataset from our laboratory (Paschold et al., 2014).

Hierarchical clustering and multidimensional scaling plots of all samples revealed high correlation of gene expression in the adjacent meristematic and elongation zones and only low correlation of these zones with cortex and stele tissue of the differentiation zone (Fig. 1, B and C). This correlation reflects the developmental gradient of roots along the longitudinal axis with the youngest undifferentiated cells in the meristematic zone of the root tip (Ishikawa and Evans, 1995). Moreover, the low correlation of the cortex and stele can be explained by the functional differences of the disparate cell types present in these tissues (Saleem et al., 2010). As illustrated by these correlation analyses, overall gene expression is more divergent between different tissues of a genotype than between different genotypes within a tissue. This observation is supported by similar distributions of differentially expressed genes identified between the four root tissues in the four genotypes (Fig. 2). Less differentially expressed genes were estimated between adjacent tissues, whereas high numbers of differentially expressed genes were observed in comparisons of the differentiation zone with the meristematic or elongation zone. On the proteome level, a similar distribution of differentially accumulated proteins between the same four primary root tissues of B73 seedlings at the same developmental stage was observed (Marcon et al., 2015).

Pairwise comparison of gene expression levels revealed a high number of differentially expressed genes between the two parental inbred lines B73 and Mo17 but no differentially expressed genes between the genetically identical reciprocal hybrids of a specific tissue (Fig. 3). Phylogenetic analyses demonstrated that the inbred lines B73 and Mo17 are distantly related (Jiao et al., 2012; Lorenz and Hoegemeyer, 2013). Similarly, mapping of Mo17 genomic sequences to the B73 reference genome revealed that only 63% of the sequences were identical in both genotypes (Wei and Wang, 2013). This finding supports the notion that transcriptome diversity in the roots of young maize hybrids is conditioned by genomic differences of the parental inbred lines. A similar distribution of the differentially expressed genes between the two maize inbred lines and their reciprocal hybrids was previously reported for aboveground tissues of 11-d-old maize seedlings (Stupar et al., 2008) and for whole primary roots of the same developmental stage as described here (Paschold et al., 2012). As suggested by the dominance hypothesis, the combination of two diverse genomes in a hybrid might lead to their integration or the complementation of deleterious alleles and form a superior plant (Jones, 1917).

Remarkably, when comparing gene expression levels between hybrids and one parent in the four primary root tissues, in all instances more differentially expressed genes were preferentially expressed in hybrids than in the parental inbred line (Fig. 3). This result is consistent with the previous findings obtained in maize top ear shoots (Qin et al., 2013), in which approximately two-thirds of the differentially expressed genes were up-regulated in the F_1 hybrid relative to the parental inbred lines. Similar results were observed in early maize ear inflorescences (Ding et al., 2014).

Unexpected allelic gene expression describes the deviation of the allelic expression ratios in the hybrids from the expression ratios of their parental inbred lines (Hochholdinger and Hoecker, 2007). In whole maize primary roots, 840 genes were reported for which the expression ratio of the two parental alleles in the reciprocal hybrids significantly deviated from the expected allelic ratio of the inbred lines (Paschold et al., 2012). In this study, a total of 739 genes were identified that show unexpected allelic expression ratios (Fig. 4). Unexpected allelic expression patterns might be predetermined by the epigenetic status of the parental genomes and the activation of transposable elements or

Figure 5. (*Continued.*)
expressed genes containing an SNP (not shaded) in each of the four primary root tissues. C, Proportion of nonsyntenic genes among differentially expressed genes [FDR $\leq$ 5%; $|\log_2(FC)| > 1$] (not shaded) between the four different genotypes B73 (B), Mo17 (M), B73xMo17 (BM), and Mo17xB73 (MB) in cortex, stele, elongation, and meristematic zone relative to all expressed genes in all genotypes (shaded). Significant differences ($\alpha \leq 5\%$) compared with all expressed genes are indicated with an asterisk.

uniparental expression of noncoding RNAs (for review, see Groszmann et al., 2013). In this study, between 15% (stele) and 35% (meristematic zone) of genes with unexpected allelic expression ratios in hybrids revealed also nonadditive expression patterns, i.e. expression levels in hybrids that deviated significantly from the midparent value. This observation supports the notion that altered allelic activity in hybrids could contribute to nonadditive gene expression (Paschold et al., 2012). In this study, approximately 8% of all expressed genes were nonadditively expressed, which is similar to results previously observed in whole primary maize roots of the same developmental stage (Paschold et al., 2012). Another remarkable observation was the high degree of conservation of both the nonadditively expressed genes and those showing unexpected allelic expression ratios between the two reciprocal hybrids (Fig. 4; Supplemental Tables S1 and S2). This underscores the importance of nuclear genome composition in reciprocal hybrids irrespective of the origin of the two parental alleles. Such a high degree of expression conservation between reciprocal hybrids was already observed in whole primary maize roots (Paschold et al., 2012).

The genome of an ancient maize progenitor underwent several duplications, including that of a paleopolyploid ancestor (Paterson et al., 2004) and an additional whole genome duplication ~5 to 12 million years ago (Blanc and Wolfe, 2004a; Swigonová et al., 2004). This resulted in the divergence of modern maize from its close and unduplicated relative sorghum (Paterson et al., 2009). Comparative analyses of the maize and sorghum genomes identified syntenic paralogs between these species and nonsyntenic maize genes that developed after this genome duplication event (Woodhouse et al., 2010; Schnable et al., 2011). In a comparative genomic study in Arabidopsis (*Arabidopsis thaliana*), it was shown that disease resistance genes were significantly overrepresented among genes found in nonsyntenic regions (Freeling et al., 2008). This indicates that nonsyntenic genes are instrumental to cope with the selective fluctuating environment, which requires constant adaptations to new conditions. In contrast, syntenic genes may encode functions that are under continuous selection. We determined if the evolutionary origin of genes correlates with the intraspecific transcriptome diversity between parental inbred lines B73 and Mo17 and their reciprocal hybrids. To this end, all genes displaying differential, nonadditive, or allelic expression patterns were surveyed concerning their synteny with sorghum, hence their evolutionary origin. Syntenic genes were shaped by more than 5 million years of natural selection (Blanc and Wolfe, 2004b). Therefore, they are likely instrumental for maize development and thus highly conserved between different maize inbred lines. In general, nonsyntenic genes were expressed significantly less frequently in this study compared with their relative genomic proportion in the maize FGSv2. Nevertheless, nonsyntenic genes were significantly overrepresented among the nonadditively expressed genes and genes with unexpected allelic expression ratios in the hybrids (Fig. 5, A and B). In addition, nonsyntenic genes were significantly overrepresented among differentially expressed genes (Fig. 5C). This finding suggests that genotype-specific gene expression patterns in the inbred lines B73 and Mo17 and their hybrid progeny are determined to a disproportionally high level by evolutionarily younger genes. Although the maize tetraploidy with subsequent genome fractionation occurred already between 5 and 12 million years ago, there is evidence that biased gene loss and expression continues today (Woodhouse et al., 2010; Schnable et al., 2011). This ongoing change in genome structure might at least in part explain the remarkable genetic diversity found among different maize lines and the overrepresentation of nonsyntenic differentially expressed genes in this study. A similar observation was made in the four maize primary root tissues, where nonsyntenic genes were significantly overrepresented among genes displaying SPE patterns (Paschold et al., 2014). SPE is an extreme instance of expression complementation in which genes are expressed in both reciprocal hybrids but only in one of the two parental inbred lines (Paschold et al., 2012). Hence, this pattern represents a special instance of the dominance model of complementation on the gene expression level.

Among the nonsyntenic nonadditively expressed genes, the GO terms death and cell death were significantly enriched in the cortex in the reciprocal hybrids. Those genes are involved in several developmental processes in roots, e.g. the development of root cortical aerenchyma. Root cortical aerenchyma is formed by enlarged gas spaces in the root cortex that are the result of cell death or the separation of cells (Evans, 2003). It has been demonstrated that maize genotypes with higher root cortical aerenchyma formation have growth advantages over genotypes with lower root cortical aerenchyma formation under certain conditions (Postma and Lynch, 2011). To date no systematic comparisons of related inbred hybrid combinations are available with respect to root cortical aerenchyma formation. It will therefore be interesting to see in future experiments if the overrepresentation of cell death and apoptosis genes in the cortex is a general feature of maize hybrids and if this observation is related to an increase in root cortical aerenchyma in hybrids.

CONCLUSION

In summary, nonsyntenic genes were significantly overrepresented among nonadditive, differential, and unexpected allelic expression patterns relative to their prevalence among all expressed genes. These observations underscore the role of these genes in shaping the transcriptomic landscape of the reciprocal maize hybrids B73xMo17 and Mo17xB73, which might be associated with the developmental manifestation of heterosis.

MATERIALS AND METHODS

RNA-Seq and Allelic Read Calling

The RNA-seq dataset analyzed in this study was generated in our research group and described in detail by Paschold et al. (2014). In brief, primary roots of a length of 2 to 4 cm of the inbred lines B73, Mo17, and their reciprocal hybrids B73xMo17 and Mo17xB73 were separated into meristematic zone, elongation zone, cortex, and stele. The 64 cDNA libraries were loaded on two flow cells according to a split plot design (supplemental table S1 in Paschold et al., 2014). Subsequently, RNA-seq (100 bp, single read) of high-quality RNA (RNA integrity number $\geq$ 7.2; Schroeder et al., 2006) was performed with a GAIIx genome analyzer equipped with on-instrument sequencing control software SCS Version 2.8 and real-time analysis RTA1.8.7 in four biological replicates per tissue and genotype combination. Base calling and run statistics were performed with the genome analyzer data analysis pipeline OLB Version 1.8.0.

High-quality reads of all genotypes were aligned to the B73 reference genome using the short-read nucleotide alignment program GSNAP (http://research-pub.gene.com/gmap/ version 2012-01-11; Schnable et al., 2009), which was indexed by applying GMAP_BUILD (k-mer size of 15 and step size 3). For subsequent analyses only uniquely mapping reads with a maximum of two mismatches out of 36 bp and with ideally 5 bp tails for every 76 bp were used. Furthermore, stacked reads that share the same start and end coordinate, sequencing direction, and sequence were discarded from the set of uniquely mapping reads. The remaining set of reads was aligned to the filtered gene set (http://ftp.maizesequence.org/release-5b/filtered-set/ release 5b.60) of the B73 reference genome derived from the Maize Genome Sequencing Project using a Perl script. Mapping Mo17 reads to the B73 genome introduces a slight bias because Mo17 reads that contain polymorphisms beyond the parameters defined above are at disadvantage of being aligned to the B73 genome sequence. However, in this dataset this bias is very small. A total of 88.1% of B73 reads mapped to the B73 genomic sequence, while only a slightly smaller fraction of Mo17 reads (87.4% of reads) mapped to the B73 genome sequence (see supplemental table S2 in Paschold et al., 2014).The set of allelic reads was determined with the help of a Perl script by extracting the uniquely mapping reads to any of the 4,034,683 B73-Mo17 SNP positions, which were previously identified between B73 and Mo17 by the 123SNP software (http://schnablelab.plantgenomics.iastate.edu/software). Only uniquely mapping reads were considered, as this mapping procedure did not allow for quantifying sequence reads that map to repeats throughout the genome. Therefore, transcript isoforms of a given gene could not be distinguished.

Analysis of Differential Gene Expression and Allelic Expression

The following analyses were applied to genes that were declared expressed if represented by a minimum of five mapped reads in all four replicates of at least one sample. The raw sequencing reads were normalized by sequencing depth and were $\log_2$-transformed to meet the assumptions of linear models. Furthermore, the mean-variance relationship within the count data were estimated and precision weights for each observation were computed (Law et al., 2014). Prior to further data analyses, sample relations were analyzed based on multidimensional scaling using the plotMDS function of the Bioconductor package limma (Smyth, 2005) in R (R version 3.1.1 2014-07-10, limma_3.20.9). The distance between each pair of samples was estimated as the root-mean-square deviation for the top 500 genes with the largest standard deviations between samples. Additionally, a hierarchical cluster analysis was performed using the Euclidean distance of the average values of the four tissues and genotypes. To analyze the differences in expression between the primary root tissues and between the four genotypes, two different linear mixed models were fitted based on the complete dataset within the Bioconductor package limma (Smyth, 2005) in R. According to the experimental design used for sequencing, the model for the analysis of the differentially expressed genes between the different tissues included fixed effects for tissue, genotype, the interaction of both treatment factors, and for block. In addition, normally distributed random effects for main- and subplot were included. Both the main- and subplot represent in a split plot design a randomization unit and need therefore be represented by a random error term each. For the estimation of the differences in expression between the four genotypes within a specific primary root tissue, the statistical model was in such a way modified that it corresponds to a complete block design. Therefore, the model included a fixed effect for genotype and block and a normally distributed random error term. After model fit, an empirical Bayes approach was applied to shrink the sample variances toward a common value (Smyth, 2004). Hypothesis tests were performed using the contrasts.fit function of the Bioconductor package limma (Smyth, 2005). The resulting p values of the performed pairwise t tests were used to determine the total number of differentially expressed genes for each comparison by controlling the FDR $\leq$ 5% to adjust for multiple testing (Benjamini and Hochberg, 1995). Significant differences in the number of up- and down-regulated genes for each of the tissue and genotype pairwise comparisons were determined based on a McNemar test (McNemar, 1947) at a significance level of alpha $\leq$ 5%.

To estimate the differences in allelic gene expression between the parental inbred lines and the two reciprocal hybrids within a specific primary root tissue, a linear mixed model was fitted in R. The model is composed of a treatment model explaining the genotypic and allelic effects and of an interaction term of both factors. Additionally, a fixed effect for block and a normally distributed random error term were included in the analysis. Using the allelic read counts, the hypothesis test that the ratio of the expression of the parental alleles in both inbred lines equals to the ratio of the expression of the parental alleles in a specific hybrid was performed by applying the contrasts.fit function in limma (Smyth, 2005). The hypothesis test was performed for both reciprocal hybrids separately. The resulting p values were FDR-adjusted (Benjamini and Hochberg, 1995) and used to determine the number of genes showing unexpected allelic ratios.

Analysis and Classification of Gene Expression Patterns in the Hybrids

Gene expression levels in the hybrids were classified according to nine gene expression classes defined previously (Paschold et al., 2012). The expression classes describe the level of gene expression in a hybrid relative to the gene expression level of both parents. The contrasts.fit function of the limma package estimates coefficients and standard errors for a given set of contrasts (Smyth, 2005). The estimated coefficients that resulted from each hybrid-parent comparison within a specific root tissue directly supply the estimated expression level for each gene separately. The pairwise comparisons between two genotypes resulting in an adjusted $p \leq 0.05$ were used for the classification of the gene expression in the hybrid. The parental inbred line with the larger estimated expression level within a given gene was termed high parent (HP), whereas the other parent was labeled as low parent (LP).

In a second step, gene expression levels of the hybrid were compared with midparent values (MPV) to determine the number of nonadditively expressed genes. For each hybrid, a contrast was fitted within the linear mixed model framework of limma to compare the estimated log-expression value of the hybrid to the mean of the log-expression values of both parents. The determination of nonadditively expressed genes was performed for each of the nine gene expression classes separately.

Determination of Syntenic and Nonsyntenic Maize Genes and Singular Enrichment Analysis

The lists comprising genes with differential, nonadditive, and allelic expression patterns were compared with the genes assigned to the maize 1 and maize 2 subgenomes (Schnable et al., 2011; http://www.skraelingmountain.com/datasets.php).

The Web-based agriGO platform (http://bioinfo.cau.edu.cn/agriGO/analysis.php) was used to assign GO functional categories to nonsyntenic genes with nonadditive and allelic expression patterns. Singular enrichment analysis computed overrepresented categories in these two gene sets by comparing them with GO terms in the set of all expressed genes using Fisher's exact test (Du et al., 2010).

Accession Numbers

Raw sequencing data are stored at the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under accession number SRP029742.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Overlap of differentially expressed genes of all tissue comparisons.

Supplemental Figure S2. Overlap of differentially expressed genes of all genotype comparisons.

Supplemental Table S1. Genes showing unexpected allelic expression patterns in each tissue.

Supplemental Table S2. Gene expression classification in the reciprocal hybrids within each tissue.

Supplemental Table S3. Functional categorization of nonsyntenic nonadditively expressed genes in the reciprocal hybrids.

Supplemental Data File S1. Gene lists for all analyzed expression patterns in the four tissues of the primary maize roots.

ACKNOWLEDGMENTS

We thank Hans-Peter Piepho (Institute for Crop Sciences, Biostatistics Unit, University of Hohenheim, Stuttgart, Germany) for advice in statistical questions.

Received February 17, 2016; accepted April 18, 2016; published April 19, 2016.

LITERATURE CITED

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Methodol **57:** 289–300

Birchler JA, Yao H, Chudalayandi S (2006) Unraveling the genetic basis of hybrid vigor. Proc Natl Acad Sci USA **103:** 12957–12958

Blanc G, Wolfe KH (2004a) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell **16:** 1667–1678

Blanc G, Wolfe KH (2004b) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell **16:** 1679–1691

Darwin C (1876) The effects of cross- and self-fertilization in the vegetable kingdom, Ed 1. John Murray

Ding H, Qin C, Luo X, Li L, Chen Z, Liu H, Gao J, Lin H, Shen Y, Zhao M, et al (2014) Heterosis in early maize ear inflorescence development: a genome-wide transcription analysis for two maize inbred lines and their hybrid. Int J Mol Sci **15:** 13892–13915

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF (2008) Evolutionary genetics of genome merger and doubling in plants. Annu Rev Genet **42:** 443–461

Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res **38:** W64–W70

Duvick DN (2005) Genetic progress in yield of United States maize (*Zea mays* L.). Maydica **50:** 193–202

East EM (1908) Inbreeding in corn. Conn Agric Exp Sta Rpt **1907:** 419–428

Evans DE (2003) Aerenchyma formation. New Phytol **161:** 35–49

Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics, Ed 4. Longman, Harlow, UK

Freeling M, Lyons E, Pedersen B, Alam M, Ming R, Lisch D (2008) Many or most genes in *Arabidopsis* transposed after the origin of the order Brassicales. Genome Res **18:** 1924–1937

Groszmann M, Greaves IK, Fujimoto R, Peacock WJ, Dennis ES (2013) The role of epigenetics in hybrid vigour. Trends Genet **29:** 684–690

Guo M, Rupe MA, Yang X, Crasta O, Zinselmeier C, Smith OS, Bowen B (2006) Genome-wide transcript analysis of maize hybrids: allelic additive gene expression and yield heterosis. Theor Appl Genet **113:** 831–845

Guo M, Rupe MA, Zinselmeier C, Habben J, Bowen BA, Smith OS (2004) Allelic variation of gene expression in maize hybrids. Plant Cell **16:** 1707–1716

Haberer G, Young S, Bharti AK, Gundlach H, Raymond C, Fuks G, Butler E, Wing RA, Rounsley S, Birren B, et al (2005) Structure and architecture of the maize genome. Plant Physiol **139:** 1612–1624

Hochholdinger F (2009) The maize root system: morphology, anatomy and genetics. *In* J Bennetzen, S Hake, eds, Handbook of Maize. Springer, New York, pp 145–160

Hochholdinger F, Hoecker N (2007) Towards the molecular basis of heterosis. Trends Plant Sci **12:** 427–432

Hoecker N, Keller B, Muthreich N, Chollet D, Descombes P, Piepho HP, Hochholdinger F (2008a) Comparison of maize (*Zea mays* L.) F_1-hybrid and parental inbred line primary root transcriptomes suggests organ-specific patterns of nonadditive gene expression and conserved expression trends. Genetics **179:** 1275–1283

Hoecker N, Keller B, Piepho HP, Hochholdinger F (2006) Manifestation of heterosis during early maize (*Zea mays* L.) root development. Theor Appl Genet **112:** 421–429

Hoecker N, Lamkemeyer T, Sarholz B, Paschold A, Fladerer C, Madlung J, Wurster K, Stahl M, Piepho HP, Nordheim A, et al (2008b) Analysis of nonadditive protein accumulation in young primary roots of a maize (*Zea mays* L.) F_1-hybrid compared to its parental inbred lines. Proteomics **8:** 3882–3894

Ishikawa H, Evans ML (1995) Specialized zones of development in roots. Plant Physiol **109:** 725–727

Jiao Y, Zhao H, Ren L, Song W, Zeng B, Guo J, Wang B, Liu Z, Chen J, Li W, et al (2012) Genome-wide genetic changes during modern breeding of maize. Nat Genet **44:** 812–815

Jones DF (1917) Dominance of linked factors as a means of accounting for heterosis. Genetics **2:** 466–479

Law CW, Chen Y, Shi W, Smyth GK (2014) voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol **15:** R29

Lorenz A, Hoegemeyer T (2013) The phylogenetic relationships of US maize germplasm. Nat Genet **45:** 844–845

Marcon C, Malik WA, Walley JW, Shen Z, Paschold A, Smith LG, Piepho HP, Briggs SP, Hochholdinger F (2015) A high-resolution tissue-specific proteome and phosphoproteome atlas of maize primary roots reveals functional gradients along the root axes. Plant Physiol **168:** 233–246

McNemar Q (1947) Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika **12:** 153–157

Paschold A, Jia Y, Marcon C, Lund S, Larson NB, Yeh CT, Ossowski S, Lanz C, Nettleton D, Schnable PS, et al (2012) Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents. Genome Res **22:** 2445–2454

Paschold A, Larson NB, Marcon C, Schnable JC, Yeh CT, Lanz C, Nettleton D, Piepho HP, Schnable PS, Hochholdinger F (2014) Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. Plant Cell **26:** 3939–3948

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, et al (2009) The *Sorghum* bicolor genome and the diversification of grasses. Nature **457:** 551–556

Paterson AH, Bowers JE, Chapman BA (2004) Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA **101:** 9903–9908

Postma JA, Lynch JP (2011) Root cortical aerenchyma enhances the growth of maize on soils with suboptimal availability of nitrogen, phosphorus, and potassium. Plant Physiol **156:** 1190–1201

Qin J, Scheuring CF, Wei G, Zhi H, Zhang M, Huang JJ, Zhou X, Galbraith DW, Zhang HB (2013) Identification and characterization of a repertoire of genes differentially expressed in developing top ear shoots between a superior hybrid and its parental inbreds in *Zea mays* L. Mol Genet Genomics **288:** 691–705

Saleem M, Lamkemeyer T, Schützenmeister A, Madlung J, Sakai H, Piepho HP, Nordheim A, Hochholdinger F (2010) Specification of cortical parenchyma and stele of maize primary roots by asymmetric levels of auxin, cytokinin, and cytokinin-regulated proteins. Plant Physiol **152:** 4–18

Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci USA **108:** 4069–4074

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science **326:** 1112–1115

Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T (2006) The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol **7:** 3

Shull G (1908) The composition of a field of maize. J Hered **4:** 296–301

Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol **3:** e3

Smyth GK (2005) Limma: linear models for microarray data. *In* R Gentleman, V Carey, S Dudoit, R Irizarry, W Huber, eds, Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer, New York, pp 397–420

Song G, Guo Z, Liu Z, Cheng Q, Qu X, Chen R, Jiang D, Liu C, Wang W, Sun Y, et al (2013) Global RNA sequencing reveals that genotype-dependent

allele-specific expression contributes to differential expression in rice F_1 hybrids. BMC Plant Biol **13:** 221–233

Springer NM, Stupar RM (2007) Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. Plant Cell **19:** 2391–2402

Stupar RM, Gardiner JM, Oldre AG, Haun WJ, Chandler VL, Springer NM (2008) Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis. BMC Plant Biol **8:** 33

Stupar RM, Springer NM (2006) *Cis*-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F_1 hybrid. Genetics **173:** 2199–2210

Swanson-Wagner RA, Jia Y, DeCook R, Borsuk LA, Nettleton D, Schnable PS (2006) All possible modes of gene action are observed in a global comparison of gene expression in a maize F_1 hybrid and its inbred parents. Proc Natl Acad Sci USA **103:** 6805–6810

Swigonová Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J (2004) Close split of sorghum and maize genome progenitors. Genome Res **14:** 1916–1923

Uzarowska A, Keller B, Piepho HP, Schwarz G, Ingvardsen C, Wenzel G, Lübberstedt T (2007) Comparative expression profiling in meristems of inbred-hybrid triplets of maize based on morphological investigations of heterosis for plant height. Plant Mol Biol **63:** 21–34

Wei X, Wang X (2013) A computational workflow to identify allele-specific expression and epigenetic modification in maize. Genomics Proteomics Bioinformatics **11:** 247–252

Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. PLoS Biol **8:** e1000409

SCIENTIFIC REPORTS

OPEN

Received: 11 May 2016
Accepted: 13 September 2016
Published: 06 October 2016

Characterization of maize *roothairless6* which encodes a D-type cellulose synthase and controls the switch from bulge formation to tip growth

Li Li[1,*,†], Stefan Hey[2,*], Sanzhen Liu[1,‡], Qiang Liu[1,3], Colton McNinch[1], Heng-Cheng Hu[1], Tsui-Jung Wen[4], Caroline Marcon[2], Anja Paschold[2,#], Wesley Bruce[5,$], Patrick S. Schnable[1,3,6,7] & Frank Hochholdinger[2]

Root hairs are tubular extensions of the epidermis. Root hairs of the monogenic recessive maize mutant *roothairless 6* (*rth6*) are arrested after bulge formation during the transition to tip growth and display a rough cell surface. BSR-Seq in combination with Seq-walking and subsequent analyses of four independently generated mutant alleles established that *rth6* encodes CSLD5 a plasma membrane localized 129 kD D-type cellulose synthase with eight transmembrane domains. Cellulose synthases are required for the biosynthesis of cellulose, the most abundant biopolymer of plant cell walls. Phylogenetic analyses revealed that RTH6 is part of a monocot specific clade of D-type cellulose synthases. D-type cellulose synthases are highly conserved in the plant kingdom with five gene family members in maize and homologs even among early land plants such as the moss *Physcomitrella patens* or the clubmoss *Selaginella moellendorffii*. Expression profiling demonstrated that *rth6* transcripts are highly enriched in root hairs as compared to all other root tissues. Moreover, in addition to the strong knock down of *rth6* expression in young primary roots of the mutant *rth6*, the gene is also significantly down-regulated in *rth3* and *rth5* mutants, while it is up-regulated in *rth2* mutants, suggesting that these genes interact in cell wall biosynthesis.

The epidermis of plant roots comprises of two types of cells: trichoblasts and atrichoblasts. During development trichoblasts give rise to tubular extensions which develop into root hairs. Root hair formation is a three step process that is initiated by the swelling of a bulge, followed by tip growth initiation and elongation through polarized exocytosis[1]. Root hairs of cereals significantly increase the root surface and are therefore instrumental for nutrient uptake and optimal development[2–6].

Different modes of epidermis differentiation into trichoblasts and atrichoblasts have been observed in plants. Trichoblast formation in Arabidopsis is position-dependent. Only epidermis cells that are situated in a cleft between two underlying cortical cells become trichoblasts and form root hairs. In Arabidopsis, the molecular network involved in epidermis specification and thus the differentiation into trichoblasts and atrichoblasts is

[1]Department of Agronomy, Iowa State University, Ames, IA 50011-3650, USA. [2]INRES, Institute of Crop Science and Resource Conservation, Crop Functional Genomics, University of Bonn, 53113 Bonn, Germany. [3]Department of Plant Genetics & Breeding, China Agricultural University, Beijing 100193, China. [4]Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, IA 50011-3650, USA. [5]Pioneer Hi-Bred International, Inc. - A DuPont Company, Johnston 50131-0184, Iowa, USA. [6]Center for Plant Genomics, Iowa State University, Ames, IA 50011-3650, USA. [7]Plant Sciences Institute, Iowa State University, Ames, IA 50011-3650, USA. [†]Present address: Department of Plant Genetics and Breeding, China Agricultural University, Beijing 100193, P.R. China. [‡]Present address: Department of Plant Pathology, Kansas State University, Manhattan KS 66506, USA. [#]Present address: Max-Planck-Institute for Plant Breeding Research, Carl-von-Linné-Weg 10, 50829 Köln, Germany. [$]Present address: BASF, Research Triangle Park, NC 27709, USA. [*]These authors contributed equally to this work. Correspondence and requests for materials should be addressed to P.S.S. (email: schnable@iastate.edu) or F.H. (email: hochholdinger@uni-bonn.de)

well understood (reviewed in: Hochholdinger and Nestler[7]). In contrast, maize epidermis cells can differentiate into trichoblasts irrespective of their position[8,9]. In maize, *roothairless5* (*rth5*) is the only gene known to date that controls root hair density and thus the differentiation of epidermis cells into trichoblasts[10].

In maize, four genes that control root hair elongation have been identified. The *roothairless 1* (*rth1*) gene encodes a SEC3 subunit[11] of the exocyst complex which tethers exocytotic vesicles prior to their fusion thus mediating exocytotic tip growth of root hairs[12]. The *roothairless3* (*rth3*) gene is translated into a monocot-specific COBRA like cell wall protein[13]. A phylogenetic study on gene families related to cell wall formation revealed that mutant alleles of *ZmCslD5* exhibit a root hair defect[14]. Most recently, the *roothairless5* (*rth5*) gene was shown to give rise to a monocot-specific NADPH oxidase, which is involved in the tip growth of root hairs[10].

Plant cell walls provide structure and determine the shape of plant cells[15]. In general, two types of cell walls can be distinguished. All dicots and most monocots develop type I cell walls. Only species of the monophyletic clade commelinids which includes the poaceae family, which comprise cereals such as maize, rice and wheat develop type II cell walls[16]. Both, type I and type II cell walls contain mostly cellulose, synthesized by cellulose synthases of the CESA gene family[14]. Cell expansion integrates the loosening of existing cell wall components with the deposition of new components. In elongating cells, cellulose microfibrils are separated and reoriented by turgor forces and subsequently interlaced by matrix constituents. In class II cell walls in species such as maize, microfibrils are crosslinked by glucuronoarabinoxylans (GAXs) while pectins contribute to only a small fraction of the matrix polymers[17]. Moreover, networks of phenylpropanoids are deposited in type II cell walls. In contrast to other commelinids, the order of poales which includes the cereals synthesize a mixed-linkage $(1 \rightarrow 3)$, $(1 \rightarrow 4)$-β-glucan during cell expansion that is hydrolyzed when growth stops[18]. Root hairs develop two layers of cell walls. The thin α-layer at the root hair tip can be expanded and formed by turgor pressure, while the thick β-layer at the non-growing tubular parts of the root hair provides stability and strength[19]. Hence, the synthesis of cell wall components in the α-layer of root hairs is instrumental for their elongation as illustrated by mutants defective in this process[13,20–22].

In the present study we report the cloning and functional characterization of the *rth6* gene of maize which encodes CSLD5, a monocot and root hair-specific D-type cellulose synthase required for the transition from bulge formation to tip growth in root hair formation.

Results

The *roothairless6* gene controls root hair length in all root-types of maize. A mutant specifically affected in root hair elongation but otherwise normal was identified in a phenotypic screen of an F_2-population resulting from *Mutator* transposon mutagenesis. Genetic crosses established that this mutant phenotype is controlled by a single recessive allele. Moreover, genetic crosses of homozygous mutants of this new mutant allele with the previously isolated mutants *rth1* to *rth5* resulted in complementation thus demonstrating that these mutants are not allelic. Therefore, this new mutant was designated *roothairless6* (*rth6*). Subsequently, additional *rth6* alleles *rth6-2* to *rth6-6* were generated (see methods).

Stereo microscopy demonstrated that *rth6-1* mutants display significantly shorter root hairs than their wild-type siblings in all major root-types including primary-, seminal- and crown roots (Fig. 1A). Quantification of root hair length established that on average *rth6-1* root hairs developed only 4–5% of the length of wild-type root hairs in primary, seminal and crown roots (Fig. 1B). Root hair morphology of wild-type and *rth6* was analyzed via cSEM (cryo scanning electron microscopy). High-resolution cSEM pictures revealed that *rth6* forms only root hair bulges (Fig. 1F–H) which did not elongate compared to wild-type root hairs (Fig. 1C–E). The basal bulges of *rth6* were characteristically swollen (Fig. 1G) like the basal region of wild-type root hairs (Fig. 1D). In contrast to wild-type root hairs, the bulges of *rth6* were arrested at this developmental stage and unable to elongate. High-resolution pictures of root hair tips illustrated that wild-type root hairs had a smooth surface (Fig. 1E) while the surface of mutant *rth6* root hairs was rough as a consequence of defects during root hair formation (Fig. 1H). In addition, mutant *rth6-1* root hair tips (Fig. 1H) displayed small outgrowths with a very rough surface at the tip. These are likely the initiation sites which represent the switch from bulge formation to root hair elongation which is defective in *rth6-1*.

Cloning of *rth6* by a combination of BSR-Seq and Seq-walking. The *rth6* gene was genetically mapped to the short arm of chromosome 1 by B-A translocation stocks. Subsequently, 123 mutant and 123 wild-type individuals of a F_2-mapping population segregating for *rth6* mutant phenotypes were subjected to BSR-Seq. This analysis demonstrated that the causative gene maps to the centromeric region of chromosome 1 (Fig. 2A). Chromosome 1 was then scanned using a window containing 100 SNPs and a step size of 20 SNPs. Within each window, the median linkage probability obtained from a Bayesian BSA analysis across all 100 SNPs was determined and plotted against the middle physical position of the window. This process mapped the *rth6* gene to a 15.7 cM interval flanked by IDP (insertion deletion polymorphism) markers IDP6931 (130.8 cM) and IDP525 (146.5 cM) of ISU IBM Map7 1 (Fig. 2B) which corresponded to ~50 Mb on the physical map of chromosome 1 (ZmB73_AGPv1 release 4a53; http://ftp.maizesequence.org/release-4a.53/). Correlation of physical and genetic distances on chromosome 1 calculated via molecular markers from an IBM genetic map[23] indicated that the centromeric region denoted by a vertical line exhibited low rates of recombination per Mb (Fig. 2C).

Genomic DNA extracted from the inbred line B73 and a pool of seedlings derived from a self of a plant with the genotype *rth6-1*/*rth6-2* was used to construct Seq-walking libraries, which were then sequenced on an Ion Proton instrument (see methods). In total, 16,680,521 raw sequence reads were obtained, 7,175,209 from the *rth6* library and 9,505,312 from the B73 library. Reads were sorted and decoded according to their barcode sequences. Subsequently, barcodes, adapters and *Mu*-related sequences were removed. The remaining *Mu* flanking sequences that were longer than 90 bp (3,054,970 and 2,703,957, respectively) were aligned to the B73 reference genome.

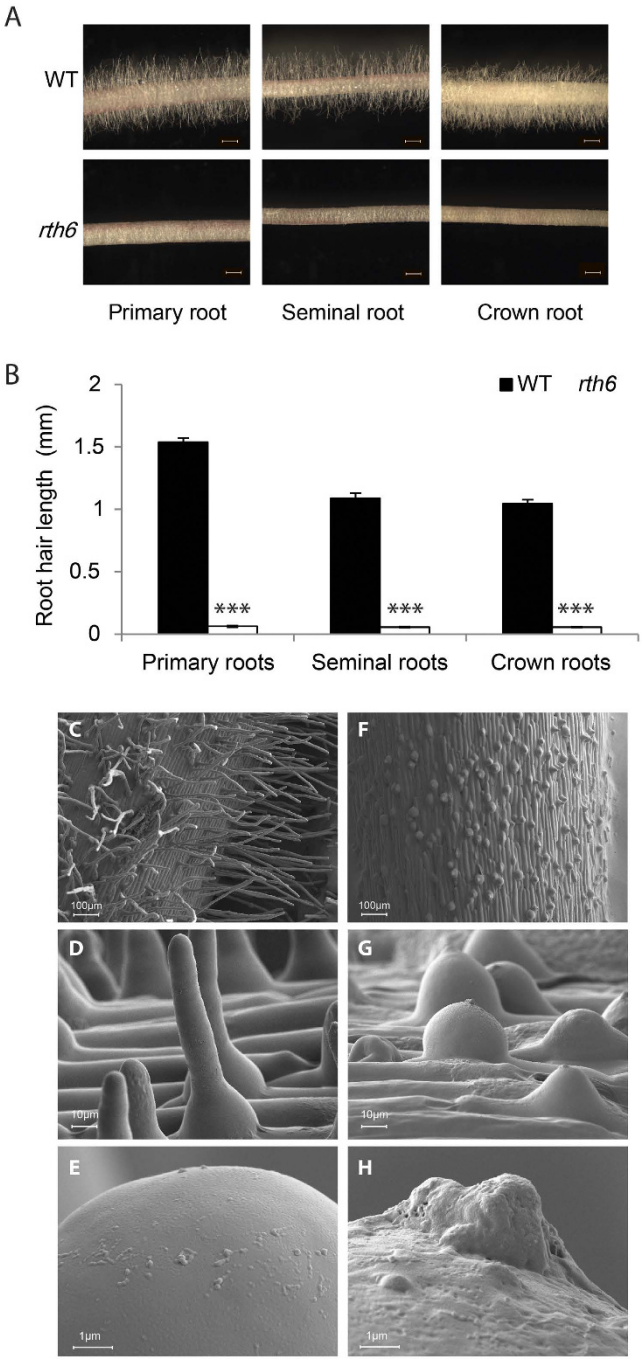


Figure 1. Phenotype of *rth6* mutants. (**A**) Microscopic images of 3–4 cm long primary roots, seminal roots and crown roots of wild-type (WT) (upper panels) and *rth6* mutants (lower panels). (Scale bar: 1 mm; n = 25, error bars indicate SD; Student's t-test ***$p \leq 0.001$). (**B**) Root hair length of 3–4 cm long primary roots, seminal roots and crown roots of WT and *rth6* mutants. (**C**–**H**) Cryo scanning electron microscopic (cSEM) images of WT primary root surface (**C**–**E**) and *rth6* primary root surface (**F**–**H**) at 100x magnification (**C**,**F**), 1,000x magnification (**D**,**G**) and tips of root hairs at 10,000x magnification (**E**,**H**).

Mu insertions recovered from the B73 Seq-Walking library were assumed not to be responsible for an *rth6* mutation. This subtractive process resulted in the identification of 29 *Mu* insertion sites within the *rth6* mapping interval (Chr1: from 102, 196, 700 bp to 144, 956, 100 bp) that had been defined by the BSR-Seq experiment (Fig. 2B). Many more reads (N = 22,945) were obtained for a single insertion site (Chr1: 104, 610, 900 bp) within the GRMZM2G436299 gene than for any other of the 28 sites within the mapping interval (Table S1). Confirmative PCR experiments on genomic DNA of the *rth6-1* and *rth6-2* alleles demonstrated that the observed *Mu* insertion derived from allele *rth6-2*, whereas the reference allele *rth6-1* did not contain a detectable *Mu* transposon insertion. Based on the *Mu* insertion in exon 3 of *rth6-2*, we declared GRMZM2G436299 an *rth6* candidate gene.

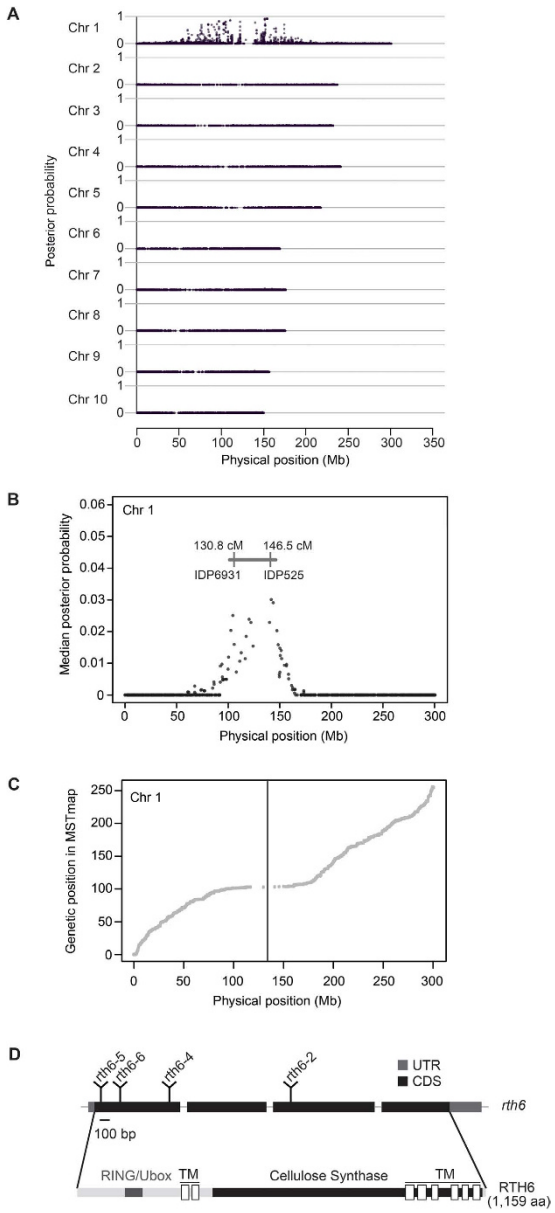


Figure 2. BSR-Seq mapping of *rth6*. (**A**) BSR-Seq analyses confines *rth6* to the centromeric region of chromosome 1. (**B**) BSR-Seq scanning defines a 15.7 cM *rth6* interval which corresponds to ~50 Mb. (**C**) Correlation of physical and genetic map positions illustrate the low recombination rates in the centromeric region of chromosome 1. The vertical line indicates the position of the *rth6* gene on the physical map. (**D**) Gene structure of *rth6*. Exons: black boxes, introns: thin lines between black boxes, UTR: untranslated region, CDS: coding sequence. The positions of transposon insertions in for novel alleles, *rth6-2*, *rth6-4*, *rth6-5* and *rth6-6* are indicated. Functional domains of the RTH6 protein are indicated. RING/Ubox: RING/Ubox like Zinc-finger domain, TM: transmembrane domain.

Confirmation of *rth6* identity via the generation of independent alleles. To confirm that GRMZM2G436299 indeed represents the *rth6* gene, two independent *Mu*-induced mutant alleles (*rth6-3* and *rth6-4*) and two *Ac*/*Ds* insertion alleles (*rth6-5* and *rth6-6*) obtained via a reverse genetic screen (see methods) were analyzed. All four of these novel alleles displayed the roothairless phenotype and transposon insertion sites were mapped by PCR and subsequent sequencing (Table 1). These experiments confirmed that GRMZM2G436299 is the *rth6* gene by independent mutant alleles which contain transposon insertions at different positions of exon 1 (*rth6-4*, *rth6-5*, *rth6-6*) and exon 3 (*rth6-2*) of the candidate gene. These results were further substantiated by crosses of the transposon induced alleles to the reference allele *rth6-1*, each of which failed to complement.

Structure and functional domains of RTH6. The *rth6* gene contains four exons and three introns encoding a 3,866 bp open reading frame which translates into a 1,159 aa protein (Fig. 2D) with a predicted molecular weight of 129 kD and an isoelectric point of 7.9 (http://web.expasy.org). RTH6 is predicted to be a membrane

Allele name	Allele ID	Type of mutation	Mutation site in bp on chr. 1 (AGPv2)
rth6-1	*rth6-ref*	n. d.[a]	n. d.
rth6-2	*rth6-Mu 10B-668*	*Mu1*	104,610,915
rth6-3	*rth6-Mu 11B-451*	n. d.	n. d.
rth6-4	*rth6-Mu 11B-453*	*Mu1*	104,609,772
rth6-5	*rth6-Ac*.mon00102	*Ac*	104,609,104
rth6-6	*rth6-Ds* I.S07.1244A	*Ds*	104,609,449

Table 1. Mutant alleles of *rth6*. [a]n.d.: not determined.

protein with 8 transmembrane domains that contains an N-terminal RING/Ubox like Zinc-finger domain, and a C-terminal cellulose synthase-like protein domain (http://www.ebi.ac.uk/Tools/hmmer).

Phylogenetic reconstructions reveal the presence of RTH6 homologs in early land plants and strict separation of mono and dicot clades. Sequence similarity searches and phylogenetic analyses revealed that *rth6* encodes a CELLULOSE SYNTHASE-LIKE D (CSLD) protein, which belongs to the D-type subfamily of the cellulose synthase superfamily. The CSLD proteins which are illustrated in Fig. 3 cluster into monocot (Fig. 3, red), dicot (Fig. 3, blue) and non-seed-plant (Fig. 3, green) subclades which are strictly separated. Each of the analyzed monocot species maize, rice, sorghum and brachypodium encodes five *CslD* genes. The five maize *CslD* genes have been previously designated *CslD1* to *CslD5*[14]. The *rth6* gene corresponds to *CslD5*. Proteins encoded by monocot *CslD* genes form five groups in which each of the four species is represented by one protein. Typically, clusters of monocot CSLD proteins group with clusters of dicot CSLD proteins. Even evolutionary distantly related plant species such as the moss *Physcomitrella patens* or the clubmoss *Selaginella moellendorffii* encode for CSLD proteins with a unexpectedly high degree of sequence identity of 67–68% with RTH6 of maize. No homologs were identified in green algae such as *Chlamydomonas reinhardii*.

Root-type and tissue specific expression of *rth6*. Expression of *rth6* was surveyed in a wide range of tissues and root types by qRT-PCR (Fig. 4). In primary roots of different lengths *rth6* displayed the highest expression in young primary roots of 1–2 cm and 2–4 cm, while significantly less expression was detected in older primary roots of 4–8 cm and 10–14 cm length (Fig. 4A). Lateral roots displayed similarly high expression levels as young primary roots, whereas seminal and crown roots showed lower expression levels similar to older primary roots (Fig. 4A). Expression of *rth6* in leaves as an example for non-root tissues, was significantly lower than in most root tissues (Fig. 4A).

Primary roots of 2–4 cm length were then dissected longitudinally into the meristematic zone (MZ), the elongation zone (EZ) and the differentiation zone. The differentiation zone was mechanically dissected into the cortical parenchyma (C) comprising all cell types between epidermis and endodermis and the stele (S). Finally, root hairs (RH) of the differentiation zone were collected separately. Root hairs displayed significantly higher *rth6* expression than all other tissues. Expression in elongation zone and cortical parenchyma could be attributed to trichoblasts present in these tissues (Fig. 4B).

Finally, *rth6* expression was monitored in primary roots of 2–4 cm length in wild-type and mutant *rth1*, *rth2*, *rth3*, *rth5*, and *rth6* seedlings (Fig. 4C). Expression of *rth6* in mutant primary roots was on average reduced to 6% of the expression levels in wild-type primary roots. Moreover, *rth6* expression was reduced in roots of the mutants *rth3* and *rth5* and significantly up-regulated in the mutant *rth2*. Expression of *rth6* was unaffected in the mutant *rth1*. Among the five closely related *CslD* genes of maize only *CslD1*, *CslD2* and *CslD5* (*rth6*) were expressed in roots (Fig. 4A). While, for instance *CslD2* displayed similar expression in older primary roots, seminal roots and crown roots as did *rth6* (Fig. 4A), *rth6* was the only *CslD* gene that displayed highly specific expression in root hairs (Fig. 4B).

Discussion

Root hair length in maize *rth6* mutants is reduced to 4–5% of wild-type seedlings in all root types including embryonic primary and seminal and postembryonic shoot-borne roots. Similarly, the maize mutants *rth1*[11], *rth2*[24], *rth3*[13] and *rth5*[10] also constitutively display defects in root hair formation in all root types. In contrast, other aspects of maize root development such as lateral root formation are controlled by root-type specific genetic programs. This is illustrated by the mutants *rum1*[25,26] and *lrt1*[27]. In both mutants only embryonic primary and seminal roots are defective in lateral root initiation while this process is not impaired in postembryonic shoot-borne roots.

The maize roothairless mutants identified thus far are affected at different stages of the three-step process of root hair development[1]. Both, the *rth3* mutant[13] and the *rth6* mutant characterized in the present study can form a bulge but are unable to initiate root hair elongation. In contrast, the *rth5* mutant is impaired in cell wall loosening at the root hair tip for subsequent turgor driven tip growth[10]. Therefore, *rth5* forms a bulge and initiates tip growth but ceases elongation soon after. Finally, the mutants *rth1*[11] and *rth2*[24] display the longest root hairs of all identified maize roothairless mutants. In both mutants root hairs initiate polar tip growth and elongate to a certain extent, but their overall final length lags significantly behind their wild-type siblings. This diversity of root hair mutant phenotypes illustrates the subtle multi step genetic regulation of this developmental process in maize.

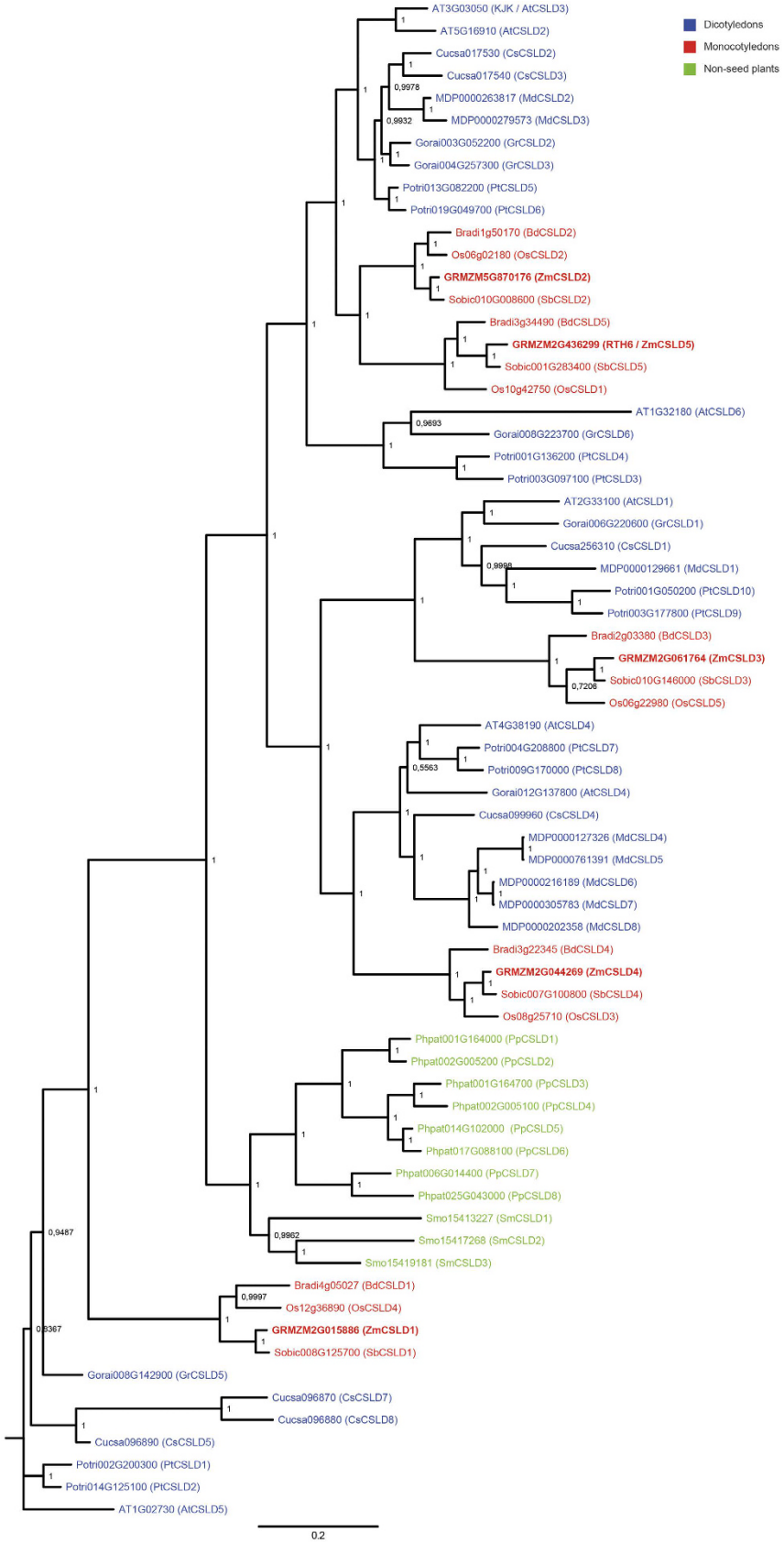


Figure 3. Phylogenetic reconstruction of the CSLD subfamily of Cellulose Synthases. Dicot species: blue, monocot species: red, non-seed plants: green. Posterior probabilities are indicated at the branching points.

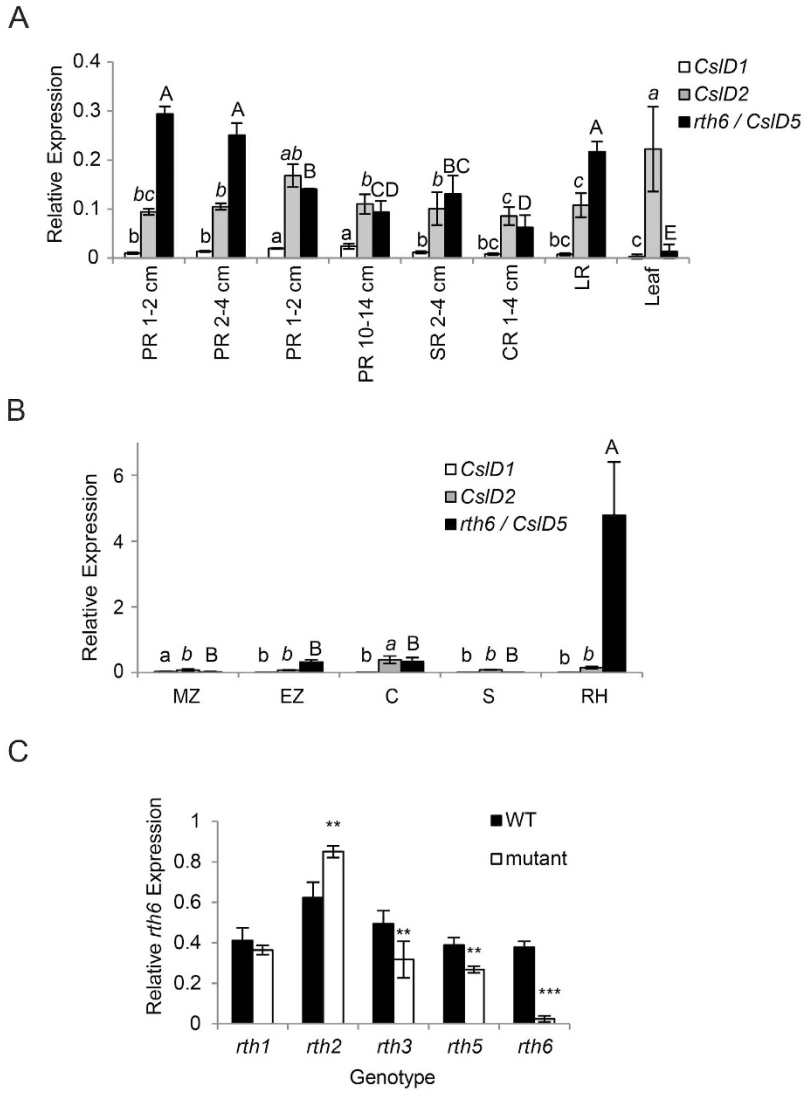


Figure 4. Expression of *rth6*. Expression of *CslD1* (white bars), *CslD2* (grey bars) and *CslD5*/*rth6* (black bars) in (**A**) primary roots (PR) at different developmental stages, seminal roots (SR), crown roots (CR), lateral roots (LR) and leaves, (**B**) in the meristematic zone (MZ), elongation zone (EZ), cortex (**C**), stele (S) and root hairs (RH) of 2–4 cm primary roots. (**C**) Expression of *rth6* in 2–4 cm primary roots of the mutants *rth1*, *rth2*, *rth3*, *rth5* and *rth6* and their wild-type (WT) siblings. Bars indicate mean relative expression ±SD (n = 4, ten plants per biological replicate). Different letters in (**A**) and (**B**) indicate significance levels at $p \leq 0.05$ after ANOVA. Small normal letters: *CslD1*; small italics letters: *CslD2*; capital letters: *rth6*/*CslD5*. No differential expression was found for *CslD2* in (**A**). (**C**) Student's t-test (**$p \leq 0.01$; ***$p \leq 0.001$).

Molecular cloning revealed that *rth6* encodes CSLD5 of the Cellulose synthase like D (CSLD) protein family[14]. Mutants of several members of this gene family including *maize csld5*[14], rice *csld1*[22] and *Arabidopsis csld3/kjk*[20,21], and *csld2*[28] have been demonstrated to be defective in root hair elongation. RTH6 is predicted to contain a C-terminal cellulose synthase domain and eight transmembrane domains. Based on shared sequence identity, all members of the cellulose synthase superfamily are predicted to be membrane bound processive glycosyltransferases that synthesize b-linked glycan polymers such as cellulose or hemicellulose backbones found in cell walls[29]. Glucan chains are synthesized at the plasma membrane by cellulose syntheses complexes to be extruded and deposited at the inside of the cell wall[15,30]. Evidence that CSLD proteins indeed act as cellulose synthases was provided by the complementation of the Arabidopsis *csld3* mutant by a chimeric CSLD3 protein containing a CESA domain known to catalyze cellulose synthesis[31]. The N-terminal RING/Ubox Zinc-finger-like domain predicted in RTH6 is specific for the CSLD and CESA subfamilies of Cellulose Synthases[29]. This domain was suggested to mediate protein-protein interactions[32,33].

Several members of the CSLD gene family have been functionally characterized and it has been demonstrated that these genes are involved in the tip growing processes. For example, *AtCslD1* and *AtCslD4*[28], have been shown to be involved in pollen tube growth. Moreover, *AtCslD2* and *AtCslD3/kjk* have divergent and redundant

functions in in root hair development and female gametophyte development[34]. Finally, ZmCSLD1 is involved in the establishment of new cross walls during cell division[35].

Despite the high degree of sequence conservation on the amino acid level, mutations in different members of the CSLD family result in morphologically different root hair phenotypes. For instance, the maize mutant *rth6* (*csld5*) and Arabidopsis *csld3*/*kjk* display very short root hairs which are arrested shortly after initiation and are unable to elongate after bulge formation. In contrast, root hairs in the rice mutant *csld1* are initiated normally but their elongation is impaired relatively late in development and root hairs display kinks and swellings along their length[22]. Hence, despite the fundamental structural differences in type I (Arabidopsis) and type II (maize and rice) cell wall compositions[16] these closely related CSLD proteins have similar functions in root hair formation.

CSLD proteins belong to one of ten subfamilies of the highly conserved cellulose synthase superfamily[14]. The high degree of evolutionary conservation of this subfamily is supported by the observation that early land plants such as the clubmoss Selaginella and the moss Physcomitrella contain *CslD* genes (Fig. 3). Selaginella belongs to the oldest living vascular plant division Lycopodiophyta which emerged ~410 million years ago. Phylogenetic reconstruction revealed that monocot, dicot, and non-seed-plant subclades are strictly separated which might illustrate that diversification of the gene family mainly occurred along the boundary of monocot, dicot, and non-seed plant species (Fig. 3). Therefore, a one-to-one correlation of monocot and dicot CSLD proteins is difficult although it has been demonstrated that monocot and dicot members of closely related clades can have similar functions as illustrated above for the monocot genes *rth6* (*ZmCsld5*) and *OsCSLD1*[22] and the dicot genes *AtCSLD2*[28] and *AtCSLD3/KJK*[20,21] in root hair development.

Root hair specific expression of *rth6* and the visible mutant phenotype implies that *rth6* has only limited functional redundancy with other members of the maize *CslD* family. Similar root hair specific expression and visible root hair defective mutant phenotypes have been observed for the rice *OsCslD1*[22] and maize *rth5*, which encodes a NADPH oxidase[10], also implying limited functional redundancy with other members of these gene families. For the Arabidopsis AtCSLD2 and AtCSLD3 proteins, partial redundancy has been observed because double mutants of these genes display even shorter root hairs than the single mutants[28].

Based on their low (*CslD2*) to absent (*CslD1*, *CslD3*, *CslD4*) expression in root hairs, the other four members of the maize *Cellulose synthase-like D* subfamily likely have other functions than controlling root hair formation. To date other than *rth6* (Zm*CslD5*), only Zm*CslD1* has been characterized by a mutant which is defective in cell division and expansion and therefore displays a significantly reduced width of several organs including leaves[14,35].

A comparative expression survey of *rth6* transcripts in wild-type versus known maize roothairless mutants revealed that the *rth2*, *rth3* and *rth5* genes are differentially expressed in wild-type versus mutant *rth6* primary roots. While the *rth2* gene has not yet been cloned, functional links between *rth3*, *rth5* and *rth6* can be established. In a comparative RNA-Seq analysis of wild-type versus mutant *rth5* roots only two GO terms were enriched[10]. In addition to "oxidation/reduction" which included some peroxidases which likely act downstream of RTH5 and facilitate the loosening of the elongating root hairs, the GO subgroup "cellulose biosynthesis" was overrepresented. Cellulose biosynthesis follows peroxidase mediated cell wall loosening[10]. Similarly, *rth3*[13] and *rth6* are functionally linked by secondary cell wall formation. Secondary cell wall formation is a two-step process of synthesis and organization[36,37]. The *rth3* gene encodes a COBRA-like cell wall protein[13]. COBRA (COB) has been identified as a putative regulator of cellulose synthesis[38]. Recently, it has been suggested in Arabidopsis that COBRA and the cellulose synthase complex reside in close proximity on the plasma membrane and that COBRA facilitates cellulose crystallization from the emerging β1-4-glucan chains by acting as a "polysaccharide chaperone"[39]. Consistent with these suggested functions, RTH6 synthesizes cellulose at the plasma membrane of the root hair tips and thus reinforces the tubular shaft, while RTH3 is involved in the organization of the synthesized cellulose. Hence, defects in either of both processes lead to non-functional cell walls which explains the highly similar phenotypes of *rth3* and *rth6* mutants. The functional relationship of RTH3, RTH5 and RTH6 has been summarized in the model in Fig. 5.

Methods

Isolation of the reference allele *rth6-1*. The mutant *rth6* was initially identified by screening segregating F_2-families derived from *Mu* active lines of the collection at Pioneer-Hybrid. The reference allele was designated *rth6-1* (Schnable Lab Ac #: 1351) and backcrossed into the inbred line B73 >8 times.

cSEM imaging. Wild-type and *rth6* mutant seeds for cSEM (cryo scanning electron microscopy) were germinated for 3–4 days. Seedling roots were cut into 1–2 cm pieces with a razor blade and mounted on a specimen holder with a mixture of Tissue-Tek® O.C.T.™Compound (Sakura Finetek Europe B.V., Alphen aan den Rijn, Netherlands) and colloidal graphite (Agar Scientific, Stansted, UK) and were immediately frozen in a nitrogen slush. Specimen were then transferred into a Quorum PP3010T cryo preparation chamber (Quorum Technologies, Laughton, UK) at −140 °C the specimen holder was heated to −80 °C for 40 min were water was sublimated. Subsequently, specimen were platinum sputtered at 10 mA for 60 sec and imaged in a Zeiss SIGMA VP cryo scanning electron microscope (Zeiss, Oberkochen, Germany). Images were taken at a magnification of 100x, 1,600x and 16,000x.

BSR-Seq mapping. A line carrying the *rth6-1* allele backcrossed for five generations into the inbred line B73 was crossed with the inbred line Mo17 and subsequently self-pollinated to generate multiple F_2-families. Kernels from one of the resulting F_2-families that were segregating for the *rth6-1* allele were rolled up in water-soaked germination paper (Anchor paper, Roseville, MN, USA) and placed in a 10 L bucket filled with ca. 2 L of distilled water and incubated at 25 °C in dark. Subsequently, 3 cm primary roots were harvested from both the *rth6* mutants and their wild-type siblings. In total, 123 mutant and 123 wild-type individuals were collected and separately pooled according to their phenotypes. RNA was extracted separately from each pool with the RNeasy

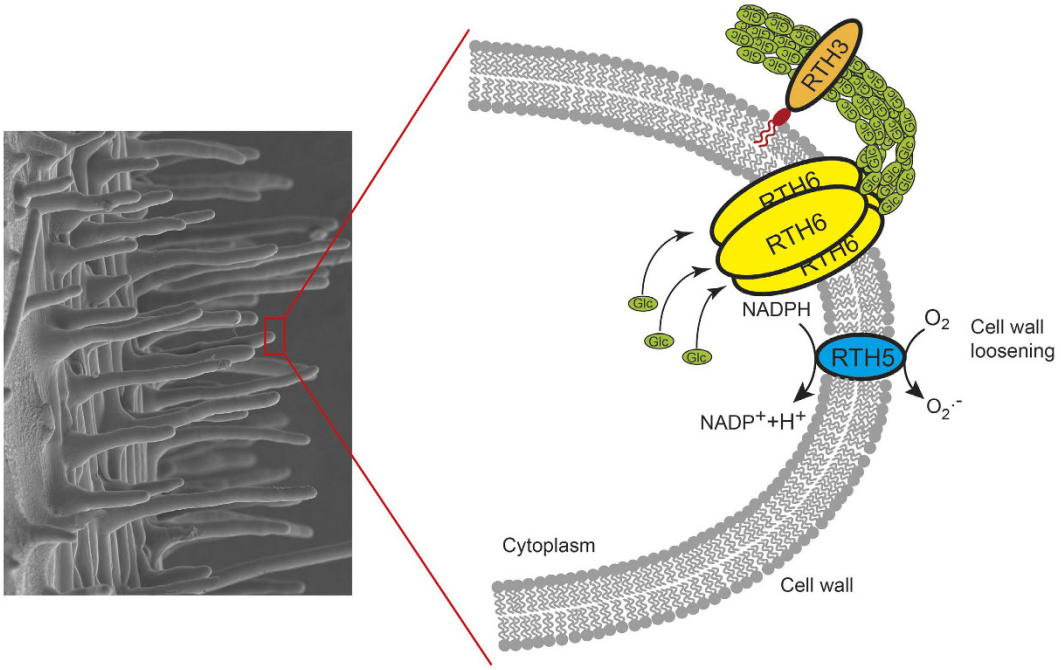


Figure 5. Model of the functional link of RTH3, RTH5 and RTH6 in root hair formation in maize. RTH5 produces apoplastic superoxide, which results in cell-wall loosening via hydroxyl radicals and also controls the expression of cellulose biosynthesis genes[10]. After cell wall loosening, the transmembrane protein RTH6 is arranged in rosettes and synthesizes cellulose at the plasma membrane[47] which is extruded to the inner side of the cell wall in the root hair tip and thus reinforces the tubular shaft, while the GPI-anchored COBRA-like cell wall protein RTH3[13] is involved in the organization of the synthesized cellulose. Glc: Glucose.

mini kit (Qiagen, Hilden, Germany) and subsequently subjected to a DNaseI treatment as per the manufacturer's protocol. RNA quality (RIN > 8) was checked on a Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA) using a RNA 6000 Nano chip. RNA-Seq libraries were constructed using the Illumina Truseq RNA-Seq sample preparation kit according to the manufacturer's protocol. Two different Illumina TruSeq barcodes were used for the two libraries[24]. The libraries were pooled and sequenced on an Illumina HiSeq2000 (Illumina, San Diego, CA, USA) sequencer, generating 99 bp single-end reads, which were analyzed as previously described[24]. Briefly, the two alleles of a given SNP site should be detected in approximately equal numbers of RNA-seq reads when considering both pools of RNA-seq reads. In contrast, only one allele of a SNP that is completely linked to the causal gene should be present in the RNA-seq reads from the mutant pool. Hence, linkage probabilities of each SNP with the causal gene were calculated and plotted versus the physical position of the gene. Subsequently, chromosome 1 was scanned by using a chromosome window containing 100 SNPs and the median linkage probability was plotted versus the middle physical position of the window. The window was slid with a step size of 20 SNPs. Sequence reads from the BSR-Seq experiment have been deposited in the SRA (sequence read archive) under AC: SRP044758.

Seq-Walking. A Seq-Walking library was generated from genomic DNA extracted from 24 mutant seedlings obtained via the self-pollination of a mutant plant with the genotype *rth6-1/rth6-2*. The isolated DNA was sheared by a BioRuptor-UCD-200 (Diagenode, Denville, NJ, USA) sonication system with 15 s/30 s on/off cycles at low speed. The library was prepared as described previously[24] and sequenced on an Ion Proton system (Life Technologies, Grand Island, NY, USA). As a control, Seq-Walking was also conducted on DNA isolated from the inbred line B73. The Seq-Walking library was prepared using the itp-Mu8 (GTCGAT) barcode, while the itp-Mu31 (CTGCTA) barcode was used for B73 library generation. Other primers used for Seq-Walking library preparation have been published previously[23].

Confirmation of the *rth6* candidate gene by independent transposon insertion alleles. Novel *Mutator* insertions in the *rth6* gene were generated by a direct transposon-tagging experiment in which plants homozygous for the *rth6-1* allele were crossed as males with *Mu*-active stocks. Three novel alleles *rth6-2* (*rth6-Mu 10B-668*), *rth6-3* (*rth6-Mu 11B-451*) and *rth6-4* (*rth6-Mu 11B-453*) were identified in a forward genetic screen for roothairless phenotypes among ~94,000 seeds of the resulting progeny. Mutant progeny were self-pollinated and in parallel crossed with plants homozygous for *rth6-1* to validate the new mutant alleles. Moreover, two *Ac*/*Ds* insertion alleles *rth6-5* (Plant GDB Ac#: *Ac*.mon00102; Schnable Lab Ac#: 5972) and *rth6-6* (Plant GDB Ac#: *Ds* I.S07.1244A; Schnable Lab Ac#: 5971) were obtained from Plant GDB (http://www.plantgdb.org/prj/AcDsTagging/v2/genes.php) by reverse genetic analysis. All *Mutator* and *Ac/Ds* insertion sites were mapped by

PCR and subsequent sequencing using transposon-specific oligonucleotide primers in combination with the *rth6* specific oligonucleotide primers listed in Table S2.

qRT-PCR. For quantitative real time PCR, seedlings of the maize inbred line B73 plants were grown for three to ten days in germination paper rolls (Anchor Paper) as previously described[40] under a 16 h light/8 h dark photoperiod at 28 °C and 24 °C, respectively. Subsequently, roots of ten plants were collected for each of the four biological replicates per tissue or developmental stage and immediately frozen in liquid nitrogen. Furthermore, four different tissues of young primary roots were collected. The meristematic zone comprising the first two mm of the root tip and the proximal elongation zone were cut with a razor blade under a stereo microscope (Zeiss Stemi 2000, Zeiss). The differentiation zone which was distinguished from the neighboring elongation zone by the presence of root hairs was separated into cortical parenchyma and stele tissues as previously described[41]. For each root hair sample, 100–150 primary roots of three-day-old seedlings were dipped into liquid nitrogen. Subsequently, frozen root hairs were broken off the primary root using a pre-cooled spatula and collected in a mortar where they were pulverized in liquid nitrogen using a pestle. Total RNA was extracted with the RNeasy kit (Qiagen). RNA integrity was measured using an Agilent 2100 Bioanalyzer and a RNA 6000 Nano chip (Agilent Technologies Inc., Santa Clara, USA). RIN values for all samples were >9.5. cDNA was synthesized from 1 μg of total RNA using the Quanta qScript™ cDNA SuperMix (Quanta, Gaithersburg, MD, USA). The cDNA was 1:2 diluted with water and a dilution series was prepared up to 1:128. Each biological replicate was measured in a BioRad CFX 384 Real-Time System (Biorad, Hercules, CA, USA) in three technical replicates using the Quanta PerfeCTa® SYBR® Green SuperMix (Quanta). Primer efficiencies were calculated using the following formula: PCR amplification efficiency $= 10^{-1/\text{slope}} - 1$[42]. Primer efficiencies were between 85% and 105% and R^2 was >0.995. Expression levels were calculated relative to a homolog of a *myosin* heavy-chain gene (GenBank accession AI941656) previously used as a reference for expression in maize roots[43].

Phylogenetic analysis. The predicted amino acid sequence of RTH6 was compared via tblastn against the translated nucleotide databases of *C. reinhardii*, *P. patens*, *S. moellendorffii*, *A. thaliana*, *G. raimondii*, *M. domestica*, *P. trichocarpa*, *C. sativus*, *S. bicolor*, *B. distachyon*, *Z. mays* and *O. sativa* from the Phytozome 10.1 plant genomics portal (http://phytozome.jgi.doe.gov/pz/portal.html). Homologous sequences were downloaded and aligned using ClustalW with a gap opening penalty of 10 and gap extension penalty of 0.1 using MEGA 6[44]. Alignments were exported into a NEXUS file and trees were generated by MrBayes[45,46]. The tree was calculated using a Markov Chain Monte Carlo approach and three hot chains and 5 million generations until the standard deviation of split frequencies dropped below 0.01. The phylogenetic tree was built using Fig Tree software (http://tree.bio.ed.ac.uk/software/figtree/).

References

1. Dolan, L. Pattern in the root epidermis: an interplay of diffusible signals and cellular geometry. *Ann. Bot.* **77,** 547–553 (1996).
2. Gilroy, S. & Jones, D. L. Through form to function: root hair development and nutrient uptake. *Trends Plant Sci.* **5,** 56–60 (2000).
3. Bates, T. R. & Lynch, J. P. The efficiency of *Arabidopsis thaliana* (Brassicaceae) root hairs in phosphorus acquisition. *Am. J. Bot.* **87,** 964–970, doi: 10.2307/2656995 (2000).
4. Gahoonia, T. S. & Nielsen, N. E. Phosphorus (P) uptake and growth of a root hairless barley mutant (*bald root barley, brb*) and wild type in low- and high-P soils. *Plant, Cell Environ.* **26,** 1759–1766, doi: 10.1046/j.1365-3040.2003.01093.x (2003).
5. Tanaka, N. *et al.* Characteristics of a root hair-less line of *Arabidopsis thaliana* under physiological stresses. *J. Exp. Bot.* doi: 10.1093/jxb/eru014 (2014).
6. Miguel, M. A., Postma, J. A. & Lynch, J. P. Phene Synergism between Root Hair Length and Basal Root Growth Angle for Phosphorus Acquisition. *Plant Physiol.* **167,** 1430–1439, doi: 10.1104/pp.15.00145 (2015).
7. Hochholdinger, F. & Nestler, J. Genetics and Genomics Of Plant Root Development. *Brenner's Encyclopedia of Genetics*, *2nd edition* Volume 5 (2012).
8. Row, H. C. & Reeder, J. R. Root-hair development as evidence of relationships among genera of Gramineae. *Am. J. Bot.* 596–601 (1957).
9. Clowes, F. Pattern in root meristem development in angiosperms. *New Phytol.* **146,** 83–94 (2000).
10. Nestler, J. *et al. Roothairless5*, which functions in maize (*Zea mays* L.) root hair initiation and elongation encodes a monocot-specific NADPH oxidase. *Plant J.* **79,** 729–740, doi: 10.1111/tpj.12578 (2014).
11. Wen, T.-J., Hochholdinger, F., Sauer, M., Bruce, W. & Schnable, P. S. The *roothairless1* Gene of Maize Encodes a Homolog of *sec3*, Which Is Involved in Polar Exocytosis. *Plant Physiol.* **138,** 1637–1643, doi: 10.1104/pp.105.062174 (2005).
12. Hála, M. *et al.* An exocyst complex functions in plant cell growth in Arabidopsis and tobacco. *Plant Cell* **20,** 1330–1345 (2008).
13. Hochholdinger, F. *et al.* The maize (*Zea mays* L.) *roothairless3* gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield. *Plant J.* **54,** 888–898, doi: 10.1111/j.1365-313X.2008.03459.x (2008).
14. Penning, B. W. *et al.* Genetic Resources for Maize Cell Wall Biology. *Plant Physiol.* **151,** 1703–1728, doi: 10.1104/pp.109.136804 (2009).
15. Taylor, N. G. Cellulose biosynthesis and deposition in higher plants. *New Phytol.* **178,** 239–252, doi: 10.1111/j.1469-8137.2008.02385.x (2008).
16. Carpita, N. C. & Gibeaut, D. M. Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J.* **3,** 1–30 (1993).
17. Carpita, N. C. Structure and Biogenesis of the Cell Walls of Grasses. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47,** 445–476, doi: 10.1146/annurev.arplant.47.1.445 (1996).
18. Buckeridge, M. S., Rayon, C., Urbanowicz, B., Tiné, M. A. S. & Carpita, N. C. Mixed Linkage (1→3),(1→4)-β-d-Glucans of Grasses. *Cereal Chem*. **81,** 115–127, doi: 10.1094/cchem.2004.81.1.115 (2004).
19. Galway, M. Root hair cell walls: filling in the framework This review is one of a selection of papers published in the Special Issue on Plant Cell Biology. *Botany* **84,** 613–621 (2006).
20. Favery, B. *et al. KOJAK* encodes a cellulose synthase-like protein required for root hair cell morphogenesis in Arabidopsis. *Genes Dev.* **15,** 79–89 (2001).
21. Wang, X. *et al. AtCSLD3*, a cellulose synthase-like gene important for root hair growth in Arabidopsis. *Plant Physiol.* **126,** 575–586 (2001).

22. Kim, C. M. *et al. OsCSLD1*, a Cellulose Synthase-Like D1 Gene, Is Required for Root Hair Morphogenesis in Rice. *Plant Physiol.* **143,** 1220–1230, doi: 10.1104/pp.106.091546 (2007).
23. Li, L. *et al.* The Maize *glossy13* Gene, Cloned via BSR-Seq and Seq-Walking Encodes a Putative ABC Transporter Required for the Normal Accumulation of Epicuticular Waxes. *PLoS ONE* **8,** e82333, doi: 10.1371/journal.pone.0082333 (2013).
24. Wen, T. J. & Schnable, P. S. Analyses of mutants of 3 genes that influence root hair development in *Zea mays* (gramineae) suggest that root hairs are dispensable. *Am. J. Bot.* **81,** 833–842, doi: 10.2307/2445764 (1994).
25. Woll, K. *et al.* Isolation, characterization, and pericycle-specific transcriptome analyses of the novel maize lateral and seminal root initiation mutant *rum1*. *Plant Physiol.* **139,** 1255–1267 (2005).
26. von Behrens, I. *et al. Rootless with undetectable meristem 1* encodes a monocot-specific AUX/IAA protein that controls embryonic seminal and post-embryonic lateral root initiation in maize. *Plant J.* **66,** 341–353 (2011).
27. Hochholdinger, F. & Feix, G. Early post-embryonic root formation is specifically affected in the maize mutant *lrt1*. *Plant J.* **16,** 247–255 (1998).
28. Bernal, A. *et al.* Functional Analysis of the Cellulose Synthase-Like Genes *CSLD1*, *CSLD2*, *and CSLD4* in Tip-Growing Arabidopsis Cells. *Plant Physiol.* **148,** 1238–1253, doi: 10.1104/pp.108.121939 (2008).
29. Richmond, T. A. & Somerville, C. R. The Cellulose Synthase Superfamily. *Plant Physiol.* **124,** 495–498, doi: 10.1104/pp.124.2.495 (2000).
30. Carpita, N. C. Update on mechanisms of plant cell wall biosynthesis: how plants make cellulose and other (1→ 4)-β-D-glycans. *Plant Physiol.* **155,** 171–184 (2011).
31. Park, S., Szumlanski, A. L., Gu, F., Guo, F. & Nielsen, E. A role for CSLD3 during cell-wall synthesis in apical plasma membranes of tip-growing root-hair cells. *Nat. Cell Biol.* **13,** 973–980, http://www.nature.com/ncb/journal/v13/n8/abs/ncb2294.html#supplementary-information (2011).
32. Kurek, I., Kawagoe, Y., Jacob-Wilk, D., Doblin, M. & Delmer, D. Dimerization of cotton fiber cellulose synthase catalytic subunits occurs via oxidation of the zinc-binding domains. *Proc. Natl. Acad. Sci. USA* **99,** 11109–11114, doi: 10.1073/pnas.162077099 (2002).
33. Gamsjaeger, R., Liew, C. K., Loughlin, F. E., Crossley, M. & Mackay, J. P. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem. Sci.* **32,** 63–70, http://dx.doi.org/10.1016/j.tibs.2006.12.007 (2007).
34. Yoo, C.-M., Quan, L. & Blancaflor, E. B. Divergence and redundancy in CSLD2 and CSLD3 function during *Arabidopsis thaliana* root hair and female gametophyte development. *Frontiers in Plant Science* **3,** doi: 10.3389/fpls.2012.00111 (2012).
35. Hunter, C. T. *et al.* Cellulose Synthase-Like D1 Is Integral to Normal Cell Division, Expansion, and Leaf Development in Maize. *Plant Physiol.* **158,** 708–724, doi: 10.1104/pp.111.188466 (2012).
36. Roudier, F. *et al.* COBRA, an Arabidopsis Extracellular Glycosyl-Phosphatidyl Inositol-Anchored Protein, Specifically Controls Highly Anisotropic Expansion through Its Involvement in Cellulose Microfibril Orientation. *Plant Cell* **17,** 1749–1763, doi: 10.1105/tpc.105.031732 (2005).
37. Ben-Tov, D. *et al.* COBRA-LIKE2, a Member of the Glycosylphosphatidylinositol-Anchored COBRA-LIKE Family, Plays a Role in Cellulose Deposition in Arabidopsis Seed Coat Mucilage Secretory Cells. *Plant Physiol.* **167,** 711–724, doi: 10.1104/pp.114.240671 (2015).
38. Schindelman, G. *et al.* COBRA encodes a putative GPI-anchored protein, which is polarly localized and necessary for oriented cell expansion in Arabidopsis. *Genes Dev.* **15,** 1115–1127, doi: 10.1101/gad.879101 (2001).
39. Sorek, N. *et al.* The Arabidopsis COBRA Protein Facilitates Cellulose Crystallization at the Plasma Membrane. *J. Biol. Chem.* **289,** 34911–34920, doi: 10.1074/jbc.M114.607192 (2014).
40. Hetz, W., Hochholdinger, F., Schwall, M. & Feix, G. Isolation and characterization of *rtcs*, a maize mutant deficient in the formation of nodal roots. *Plant J.* **10,** 845–857 (1996).
41. Saleem, M. *et al.* Tissue specific control of the maize (*Zea mays* L.) embryo, cortical parenchyma, and stele proteomes by RUM1 which regulates seminal and lateral root initiation. *J. Proteome Res.* **8,** 2285–2297 (2009).
42. Bustin, S. A. *et al.* The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clin. Chem.* **55,** 611–622, doi: 10.1373/clinchem.2008.112797 (2009).
43. Hoecker, N. *et al.* Comparison of maize (*Zea mays* L.) F_1-hybrid and parental inbred line primary root transcriptomes suggests organ-specific patterns of nonadditive gene expression and conserved expression trends. *Genetics* **179,** 1275–1283 (2008).
44. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **30,** 2725–2729, doi: 10.1093/molbev/mst197 (2013).
45. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17,** 754–755 (2001).
46. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19,** 1572–1574, doi: 10.1093/bioinformatics/btg180 (2003).
47. Olek, A. T. *et al.* The Structure of the Catalytic Domain of a Plant Cellulose Synthase and Its Assembly into Dimers. *Plant Cell* **26,** 2996–3009, doi: 10.1105/tpc.114.126862 (2014).

Acknowledgements

We thank Drs. An-Ping Hsia and Wei Wu and Mr. Cheng-Ting "Eddy" Yeh and for technical support and helpful discussions and Ms. Lisa Coffey for the generation and maintenance of genetic stocks used in this study (all of ISU). We thank Sabine Adler, Tanja Rollnik and Dominik Begerow of the Ruhr University of Bochum, Germany for technical support with cyro scanning electron microscopy. Root hair research in F.H.´s laboratory is supported by the DFG (Deutsche Forschungsgemeinschaft) grant HO2249/8-2.

Author Contributions

W.B. identified the reference allele of the mutant. L.L., S.H., P.S.S. and F.H. conceived and designed research; L.L., S.L., Q.L., C.M.N., H.C.H., A.P., C.M., T.-J.W. and P.S.S. performed the experiments and analyzed data for the mapping and cloning of the *rth6* gene; S.H. and F.H. performed and analyzed data for the functional characterization; L.L., S.H., P.S. and F.H. wrote the paper. All authors participated in stimulating discussion and approved the final article.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Li, L. *et al.* Characterization of maize *roothairless6* which encodes a D-type cellulose synthase and controls the switch from bulge formation to tip growth. *Sci. Rep.* **6**, 34395; doi: 10.1038/srep34395 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2016

Journal of Experimental Botany, Vol. 67, No. 4 pp. 1137–1147, 2016
doi:10.1093/jxb/erv519 Advance Access publication 15 December 2015
This paper is available online free of all access charges (see http://jxb.oxfordjournals.org/open_access.html for further details)

RESEARCH PAPER

Conserved and unique features of the homeologous maize Aux/IAA proteins ROOTLESS WITH UNDETECTABLE MERISTEM 1 and RUM1-like 1

Yanxiang Zhang[1,2], Caroline Marcon[2], Huanhuan Tai[2], Inga von Behrens[3], Yvonne Ludwig[2], Stefan Hey[2], Kenneth W. Berendzen[4] and Frank Hochholdinger[2,*]

[1] Center for Molecular Cell and Systems Biology, College of Life Science, Fujian Agriculture & Forestry University, 350002 Fuzhou, China
[2] Crop Functional Genomics, Institute of Crop Science and Resource Conservation, University of Bonn, 53113 Bonn, Germany
[3] ZMBP, Center for Plant Molecular Biology, General Genetics, University of Tuebingen, 72076 Tuebingen, Germany
[4] ZMBP, Center for Plant Molecular Biology, Central Facilities, University of Tuebingen, 72076 Tuebingen, Germany

* Correspondence: hochholdinger@uni-bonn.de

Received 16 September 2015; Accepted 13 November 2015

Editor: Adam Price, University of Aberdeen

Abstract

The maize (*Zea mays* L.) Aux/IAA protein RUM1 (ROOTLESS WITH UNDETECTABLE MERISTEM 1) is a key regulator of lateral and seminal root formation. An ancient maize genome duplication resulted in the emergence of its homeolog *rum1-like1* (*rul1*), which displays 92% amino acid sequence identity with RUM1. Both, RUL1 and RUM1 exhibit the canonical four domain structure of Aux/IAA proteins. Moreover, both are localized to the nucleus, are instable and have similar short half-lives of ~23 min. Moreover, RUL1 and RUM1 can be stabilized by specific mutations in the five amino acid degron sequence of domain II. In addition, proteins encoded by both genes interact *in vivo* with auxin response factors (ARFs) such as ZmARF25 and ZmARF34 in protoplasts. Although it was demonstrated that RUL1 and RUM1 can homo and heterodimerize *in vivo*, *rul1* expression is independent of *rum1*. Moreover, on average *rul1* expression is ~84-fold higher than *rum1* in the 12 tested tissues and developmental stages, although the relative expression levels in different root tissues are very similar. While RUM1 and RUL1 display conserved biochemical properties, yeast-two-hybrid in combination with BiFC experiments identified a RUM1-associated protein 1 (RAP1) that specifically interacts with RUM1 but not with RUL1. This suggests that RUM1 and RUL1 are at least in part interwoven into different molecular networks.

Key words: Aux/IAA, maize, protein interaction, root, RAP1, RUM1, RUL1.

Introduction

Maize (*Zea mays* L.) plays an important role as fodder, human food and a source of bioethanol. The maize root system is formed by embryonic primary and seminal roots and post-embryonic shoot-borne crown and brace roots, which are instrumental for water and nutrient uptake and for anchorage of plants in soil (Hochholdinger and Tuberosa, 2009). All these root-types form post-embryonic lateral roots, which are initiated from pericycle and endodermis cells and make up the major backbone of the plant (Hochholdinger *et al.*, 2004).

© The Author 2015. Published by Oxford University Press on behalf of the Society for Experimental Biology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

The phytohormone auxin is a key regulator of almost all developmental processes including root formation (Hochholdinger and Zimmermann, 2008; Peret *et al.*, 2009; Jansen *et al.*, 2012). Application of exogenous auxin or auxin transport inhibitors to roots of *Arabidopsis thaliana* and maize suggest that polar auxin transport is required for lateral root initiation (Reed *et al.*, 1998; Casimiro *et al.*, 2001; Jansen *et al.*, 2012). The semi-dominant maize mutant *rum1* (*rootless with undetectable meristem 1*) is affected in the initiation of embryonic seminal and postembryonic lateral roots of the primary root (Woll *et al.*, 2005). The mutant *rum1* showed an 83% reduction of polar auxin transport and delayed gravitropic response in the primary root (Woll *et al.*, 2005). The *rum1* gene encodes an Aux/IAA (auxin/indole-3-acetic acid) protein which is a key regulator of auxin signal transduction (von Behrens *et al.*, 2011). Canonical Aux/IAA proteins have four functional domains. Domain I is responsible for transcriptional repression (Tiwari *et al.*, 2004). In Arabidopsis, domain I is also predicted to be a protein-protein interaction domain. For instance, domain I of BD/IAA12 is responsible for interaction with TOPLESS which is required as a co-repressor for IAA12 (Szemenyei *et al.*, 2008). Domain II of Aux/IAA proteins is related to their instability of Aux/IAA proteins (Worley *et al.*, 2000; Dreher *et al.*, 2006). The degron motif GWPPV of domain II binds to the E3 ubiquitin ligase complex SCF^{TIR} (Tan *et al.*, 2007) leading to Aux/IAA protein ubiquitination and subsequent proteasomal degradation (Gray *et al.*, 2001). The maize *rum1* mutant and several Arabidopsis *Aux/IAA* gain-of-function mutants were identified due to mutations in the core region of domain II (Benjamins and Scheres, 2008; von Behrens *et al.*, 2011) resulting in increased stability of Aux/IAA proteins (Worley *et al.*, 2000; Dreher *et al.*, 2006; von Behrens *et al.*, 2011). Aux/IAA proteins can form homo- or heterodimers by their dimerization domains III and IV with Aux/IAA or ARF (auxin response factor) proteins (Kim *et al.*, 1997; Woodward and Bartel, 2005; Ludwig *et al.*, 2014). The Aux/IAA-ARF complex binds to AuxREs (auxin responsive elements) repressing the transcription of early/primary auxin-responsive genes such as *Aux/IAA*s, *SAUR*s and *GH3*s at low intracellular auxin concentrations (Woodward and Bartel, 2005). These genes often contain a conserved 5′ TGTCTC 3′ or 5′ TGTC 3′ motif in their upstream regulatory sequence (Ulmasov *et al.* 1997, 1999; Lau *et al.*, 2011). However, at high cellular auxin levels, ARF proteins are released from the Aux/IAA-ARF complex to promote transcription of auxin-responsive genes, whereas Aux/IAA proteins are degraded by the proteasome (Abel, 2007; von Behrens *et al.*, 2011).

The genome of a maize progenitor was subjected to a whole genome duplication 5–12 million years ago. In the course of evolution in many instances one copy of the duplicated genes was lost, a process called partial fractionation (Schnable and Freeling, 2011). In modern maize, ~50% of all syntenic genes are pairs of homeologs while the remaining 50% of genes are single copy genes (Schnable and Freeling, 2011). Thus far only a small number of homeologous maize gene pairs have been characterized in more detail. Among those, *dwarf plant8* (*d8*) and *dwarf plant9* (*d9*) encode DELLA proteins, which are negative regulators of gibberellin signaling (Lawit *et al.*, 2010). Moreover, *colored aleurone1* (*c1*) and *purple plant1* (*pl1*) encode MYB transcription factors (Cone *et al.*, 1993), while *colored 1* (*r1*) and *colored plant* (*b1*) encode bHLH transcriptional regulators (Chandler *et al.*, 1989), which all control the biosynthesis of anthocyanin. Finally, *discordia1* (*dcd1*) and its close relative *alternative discordia1* (*add1*) encode a putative B″ regulatory subunit of the PP2A phosphatase complex that is involved in preprophase band formation during cytokinesis (Wright *et al.*, 2009). In the present study, we characterized the unique and conserved features of the homeologous maize genes *rum1* and *rul1*, which encode Aux/IAA proteins.

Materials and methods

Plant material and growth conditions

Seeds of the maize inbred line B73, the mutant *rum1-R* (*rum1-Reference*) and its wild-type siblings were sterilized with 6% hypochlorite for 5 min under vacuum at 500 mPa and then rinsed five times with distilled water. Subsequently, seeds were germinated in paper rolls in a plant growth chamber in 16 h light at 28°C and 8 h dark at 21°C as previously described (Ludwig *et al.*, 2013). Five-day-old seedlings of the maize inbred line B73 were treated with the auxin analog 1-Naphthaleneacetic Acid (1-NAA, working solution 5 µM) for 3 h. Primary roots were harvested after 0, 1, 2 and 3 h of 1-NAA exposure, then immediately frozen in liquid nitrogen and stored at −80 °C until RNA isolation (Zhang *et al.*, 2015).

qRT-PCR expression analyses

For qRT-PCR, total RNA was isolated from distinct maize samples with the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany). All RNA samples were treated with RNase-free DNaseI (Fermentas, St. Leon-Roth, Germany) and were subsequently tested for contamination with genomic DNA by PCR as previously described (Zhang *et al.*, 2015). cDNA was synthesized from 500 ng total RNA using the qScript cDNA SuperMix (Quanta Biosciences, Gaithersburg, USA). qPCR was performed as previously described (Zhang *et al.*, 2014). Each genotype or treatment was assayed in four biological replicates by qRT-PCR. Each biological replicate was subjected to three qRT-PCR reactions (technical replications). An internal control gene (Genbank AC: 486090G09.x1; primers: 486090G09.x1-5′; 486090G09.x1-3′) was used in the qPCR as previously described for maize primary roots (Hoecker *et al.*, 2006). The oligonucleotide primers *rum1*-fw and *rum1*-rv, *rul1*-fw and *rul1*-rv, *rap1*-fw and *rap1*-rv were used for *rum1*, *rul1*, and *rap1* gene expression studies, respectively (Supplementary Table S1 available at *JXB* online). Differential gene expression was determined by Student's *t*-test (*, $P \leq 0.05$; **, $P \leq 0.01$; ***, $P \leq 0.001$; $n=4$). Correlation of expression values was calculated based on Student's *t*-distribution (degree of freedom, $n-2$; tails, 2).

Subcellular localization

To construct the vector pucHA-GFP for transient transformations, pUC-SPYCE was double digested by the restriction enzymes *Sma*I and *Sac*I, then ligated with the HA-GFP fragment replacing the SPYCE fragment (Lab AC: 765). The HA-GFP open reading frame was amplified by PCR from vector pCF203-GFP using the oligonucleotide primers HA-GFP-*Sma*I-fw and HA-GFP-*Sac*I-rv (Supplementary Table S1). For the *rul1*-GFP fusion construct, the open reading frame of *rul1* which was previously cloned into the pENTR/D-TOPO vector (Invitrogen, Darmstadt, Germany) was amplified using the oligonucleotide primers *rul1-Kpn*I-fw and *rul1-Bam*HI-rv (Supplementary Table S1) introducing 5′ *Kpn*I and 3′ *Bam*HI restriction sites and deleting the stop codon of the *rul1* cDNA. Subsequently, this PCR

product was introduced into the *Kpn*I and *Bam*HI restriction sites of the pucHA-GFP vector yielding a construct containing a constitutive cauliflower mosaic virus (CaMV) 35S promoter at the 5′ end of the coding sequence of *rul1* and a 3′ in-frame GFP sequence followed by the nopaline synthase (NOS) terminator (Lab AC: 818). Site-directed mutagenesis was used to change the proline (P) amino acid residues at positions 121 and 122 of RUL1 to lysine (L) according to the manufacturer's instructions (Stratagene). The oligonucleotide primers *rul1-P121L*-fw, *rul1-P121L*-rv and *rul1-P122L*-fw, *rul1-P122L*-rv were used for site-directed mutagenesis of *rul1-P121L* and *rul1-P122L* (Lab ACs: 820 and 821), respectively (Supplementary Table S1). All of the nucleotide sequence insertions were confirmed by sequencing.

Subcellular localization experiments were performed by transiently transforming the plasmids 35S::GFP, 35S::*rul1*-GFP, 35S::*rul1-P121L*-GFP and 35S::*rul1-P122L*-GFP into Arabidopsis Col-0 protoplasts. Cell cultures were incubated at 26°C in MSCol medium overnight in dark (Liu *et al.*, 2003). Protoplasts were generated as described in Negrutiu *et al.* (1987). Transformation was performed according to a PEG protocol (Merkle *et al.*, 1996). After overnight incubation, the transformed protoplasts were examined with a HCX PL APO 63×/1.2W CORR water immersion objective (Leica Microsystems, Wetzlar, Germany) of a TCS SP2 AOBS confocal microscope (Leica Microsystems). GFP was excited with an argon laser at 488 nm and the emitted fluorescence was detected with an argon-krypton laser at 509 nm. Image processing was performed with Leica Confocal Software (Leica Microsystems). Epifluorescence images were taken from the same protoplasts that were analysed for green fluorescence localization.

Stability of RUL1, rul1-P121L and rul1-P122L

The GFP constructs 35S::GFP, 35S::*rul1*-GFP, 35S::*rul1-P121L*-GFP and 35S::*rul1-P122L*-GFP that were generated for subcellular localization experiments were transformed into Arabidopsis protoplasts as described above. Sixteen hours after protoplast transformation, the auxin analog 1-NAA (working solution 10 µM) and the eukaryotic protein synthesis inhibitor cycloheximide (working solution 100 µg/ml) were added and samples were taken after 0, 10, 30, 60 and 120 min. The analyses with a Modular Flow (Beckman Coulter, Brea, CA) cytometer were performed as previously described (von Behrens *et al.*, 2011). Data acquisition and analysis were performed with MoFlo Summit 4.3 software. Three biological replicates were used for each experiment.

Protein expression of the RUL1-GFP, rul1-P121L-GFP and rul1-P122L-GFP fusion protein in protoplasts was analysed by precipitating protoplasts with a buffer containing 0.5 M mannitol, 15 mM $MgCl_2$, and 0.1% MES. Proteins were extracted using an extraction buffer containing 50 mM Tris pH 7.5, 100 mM NaCl, 0.1% Triton X-100 and protease inhibitor cocktail. Sixteen hours after incubation protoplasts were treated with cycloheximide and 1-NAA as described above. Per experiment, 1.5 ml of the transfected protoplasts were harvested after 0, 30 and 120 min. Proteins were separated on a 12% SDS-PAGE gel. Western blot analyses were performed as previously described (Saleem *et al.*, 2010) using a primary anti-GFP antibody (Roche, Germany) and a secondary anti-mouse IgG antibody (Sigma Aldrich, Germany). The secondary antibody was detected with NBT/BCIP (Roche, Germany) as previously described (Saleem *et al.*, 2010).

Yeast two-hybrid assay

For yeast two-hybrid experiments the bait construct pGBKT7-BD-*rum1* (Lab AC: 554) was used and a cDNA expression library was generated from mRNA of 2.5-day-old maize primary roots as previously described (von Behrens *et al.*, 2011). Screening for interaction of AD-prey proteins with BD-RUM1, preparation of yeast competent cells and transformation were performed according to the manufacturer's instructions (Clontech Laboratories, Paris, France). cDNA insertions of positive interaction partners of BD-RUM1 were sequenced and identified using http://ensembl.gramene.org.

BiFC and flow cytometric analysis

Generation of fusion proteins of full length *rum1* and the Aux/IAA domains of ZmARF25 and ZmARF34 with YFPC (Walter *et al.*, 2004) or $YFPN^{152}$ (Li *et al.*, 2010) has been previously described (von Behrens *et al.*, 2011). Moreover, subdomains of *rum1* were amplified with the oligonucleotide primers *DI-rum1-BamH*I-fw and *DI-rum1-Kpn*I-rv (domain I), *DII-rum1-BamH*I-fw and *DII-rum1-Kpn*I-rv (domain II) and *DIII–IV-rum1-BamH*I-fw and *DIII-IV-rum1-Kpn*I-rv (domain III–IV) (Supplementary Table S1) using TaKaRa La Taq Polymerase (Lonza, Basel, Switzerland) from the pENTR-TOPO-*rum1* vector (Lab AC: 473), which was constructed as previously described (von Behrens *et al.*, 2011). Subsequently, *Bam*HI and *Kpn*I fragments of domain I, domain II and domain III–IV of *rum1* were introduced into pUC-SPYCE and the modified pUC-SPYNE-152, respectively (Lab ACs: 767 and 769, 778 and 780, 768 and 770). Similarly, full length *rul1* was amplified with the oligonucleotide primers *rul1-Bam*HI-fw and *rul1-Kpn*I-rv (Supplementary Table S1) from the pENTR-TOPO-*rul1* (Lab AC: 475) vector, which was generated as described above. Subsequently, *Bam*HI and *Kpn*I fragments were introduced into pUC-SPYCE and the modified pUC-SPYNE-152 (Lab ACs: 529 and 530). Furthermore, the oligonucleotide primers *rum1-R-Xba*I-fw and *rum1-R-Sma*I-rv (Supplementary Table S1) were used to amplify the full length sequence of mutated *rum1-R* cDNA using pENTR-TOPO-*rum1-R* as a template (Lab AC: 474) (von Behrens *et al.*, 2011). This PCR product was introduced into the restriction sites *Xba*I and *Sma*I of pUC-SPYCE and the modified pUC-SPYNE-152 vectors (Lab ACs: 748 and 749). Finally, full length *rap1* was generated by PCR amplification with *rap1*-fw and *rap1*-rv oligonucleotide primers (Supplementary Table S1). Subsequently, this template was reamplified with the oligonucleotide primers *rap1-Xba*I-fw and *rap1-Sma*I-rv (Supplementary Table S1) to introduce *Xba*I and *Sma*I restriction sites, which were then introduced into the restriction sites of the BiFC vector pUC-SPYCE and the modified vector pUC-SPYNE-152 (Lab ACs: 710 and 711). The inserts were confirmed by sequencing and transformed into Arabidopsis Col-0 protoplasts for BiFC analyses.

Flow cytometry was performed as previously described (von Behrens *et al.*, 2011). Briefly, 300 µl of the transfected protoplasts per experiment were filtered through a 40 µM sieve and fluorescence signal intensity was analysed with a Modular Flow (Beckman Coulter, Brea, CA, USA) cytometer. YFP fluorescence was excited with a 488 nm (50 mW) argon laser and its principal emission was captured in FL1 (530/40) and plotted against autofluorescence in FL2 (580/30). After gating out cellular debris detected in the FSC/SSC plot, BiFC expressing cells were identified as those whose fluorescence was increased in the FL1 channel compared to the negative controls (NUO-YFPCE and NUO-YFPNE-152, Lab ACs: 705 and 706) as previously described (von Behrens *et al.*, 2011). Data acquisition and analysis was performed with MoFlo Summit 4.3 software.

Protein expression of each fusion protein described above was analysed in Arabidopsis protoplasts. Western blot analyses were performed as previously described (Saleem *et al.*, 2010) using a primary anti-HA antibody (Roche, Germany) or anti-MYC antibody (Cell Signaling Technology, USA) and a secondary anti-mouse IgG antibody (Sigma Aldrich, Germany). The secondary antibody was detected with NBT/BCIP (Roche, Germany) as described before (Saleem *et al.*, 2010).

Results

RUL1 is the homeolog of the Aux/IAA protein RUM1 and localizes to the nucleus

Based on their microsynteny it has been demonstrated that *rum1*, which is located on chromosome 3, and *rul1* (*rum1-like1*; GRMZM2G163848), which maps to chromosome 8, are homeologs (von Behrens *et al.*, 2011). A maize progenitor

underwent a whole genome duplication in ancient times. Therefore, many genomic regions of modern maize can be attributed to either of two subgenomes 1 and 2. While the *rum1* gene belongs to subgenome 1 its homeolog *rul1* was attributed to subgenome 2. RUL1 is predicted to contain the canonical four domain structure of Aux/IAA proteins and a bipartite nuclear localization signal (NLS; Fig. 1A). To determine the subcellular localization of RUL1, a RUL1-GFP fusion protein was transiently expressed in Arabidopsis Col-0 protoplasts. This experiment localized RUL1-GFP to the nucleus (Fig. 1B). Similarly, the fusion proteins rul1-P121L-GFP and rul1-P122L-GFP containing point mutations in the degron sequence of RUL1 were also localized to the nucleus. By contrast, the GFP control protein displayed a constitutive localization in the nucleus and the cytoplasm.

RUL1 is unstable

Aux/IAA proteins are unstable proteins that are rapidly degraded at increased cellular auxin levels. Therefore, the stability of the RUL1 wild-type protein was compared with two mutated isoforms of the protein, rul1-P121L and rul1-P122L in which the proline (P) at amino acid positions 121 and 122 was changed into lysine (L), respectively. Relative GFP-fluorescence intensities of RUL1-GFP, rul1-P121L-GFP and rul1-P122L-GFP fusion proteins were measured between 0 and 120 min with a flow cytometer (Fig. 1C). As a control, relative GFP-fluorescence intensities of constitutively expressed GFP were determined (Fig. 1C). This experiment demonstrated the instability of RUL1 with an average half-life of ~23 min, compared to the stable GFP protein (Fig. 1C). As predicted, the P-to-L amino acid exchange within the degron sequence of RUL1 at position 122 stabilized

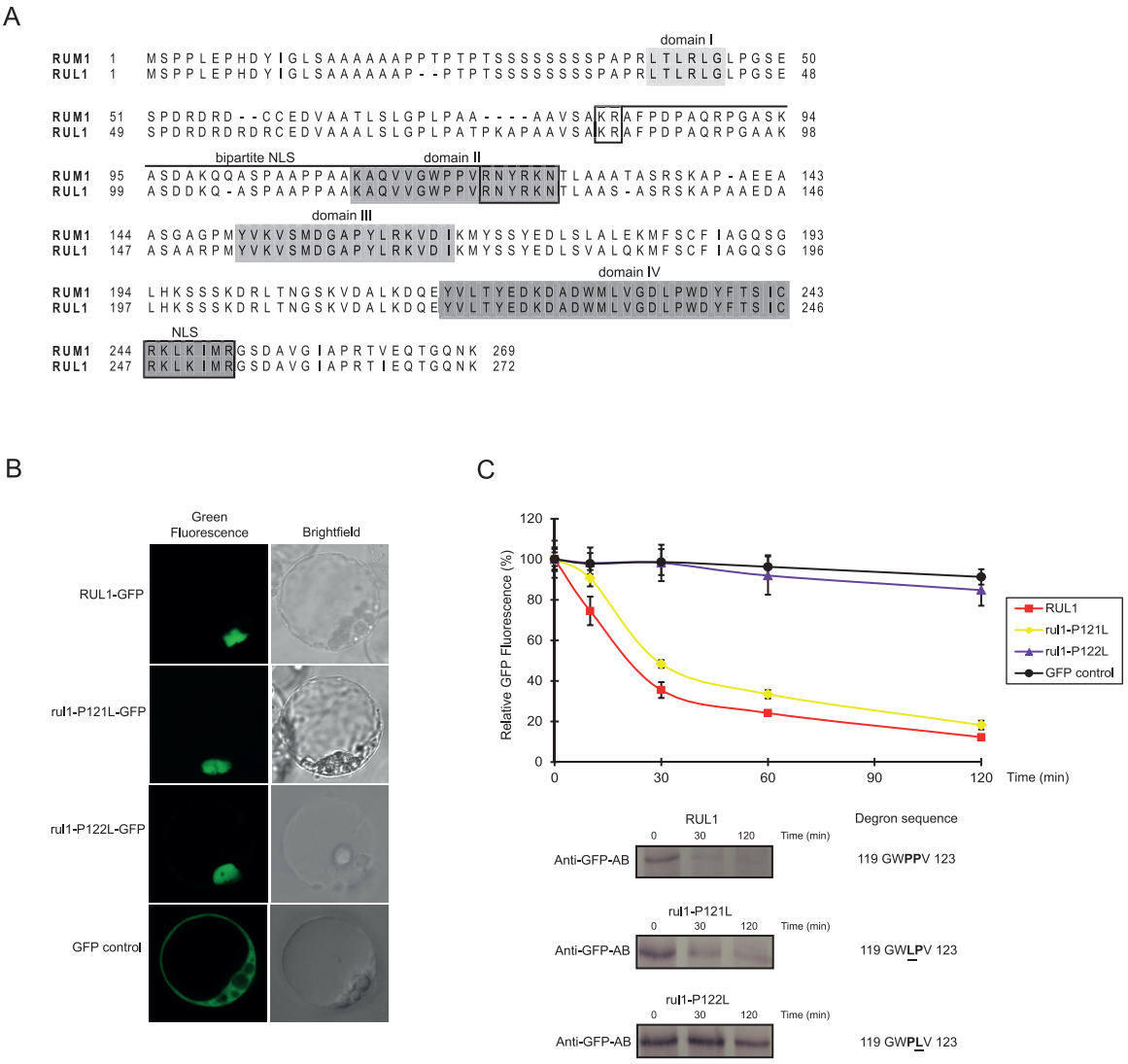


Fig. 1. Characteristics of *rul1*. (A) Alignment of the amino acid sequences encoded by the homeologous genes *rul1* and *rum1* revealed ~92% identity. The domain structures and the nuclear localization signal (NLS) are indicated. (B) Subcellular localization of RUL1, rul1-P121L-GFP and rul1-P122L-GFP and a GFP control. (C) Protein stability assay of RUL1, rul1-P121L, rul1-P122L and the GFP control. Relative fluorescence of these proteins was measured in Arabidopsis Col-0 protoplasts between 0 and 120 min after 1-NAA and cycloheximide treatment by flow cytometry. Protein stability of RUL1-GFP, rul1-P121L-GFP and rul1-P122L-GFP was confirmed at time points 0, 30, 120 min by Western blot analyses using an anti-GFP antibody.

the protein significantly. While the relative GFP fluorescence of wild-type RUL1 was reduced to ~10% within 120 min, GFP fluorescence of rul1-P122L was still at ~85% which is close to the GFP control. By contrast, the P-to-L amino acid exchange within the degron sequence of RUL1 at position 121 did not stabilize the protein. The rul1-P121L protein displayed a similar half-life of ~28 min as the wild-type RUL1 protein and after 120 min only ~20% of the GFP fluorescence was remaining (Fig. 1C). Western blot experiments with total protein extracts from protoplasts overexpressing RUL1-GFP, rul1-P121L-GFP and rul1-P122L-GFP fusion proteins confirmed these results by showing half-lives similar in value to those observed in flow cytometry (Fig. 1C).

Overall higher expression in rul1 *compared to* rum1

To compare the temporal and spatial expression patterns of *rum1* and *rul1*, 12 distinct tissues and developmental stages of the maize inbred line B73 were analysed by qRT-PCR (Fig. 2A). In all comparisons, *rul1* was significantly higher expressed than *rum1*. On average, *rul1* displayed a ~84-fold higher expression than *rum1*. Both, *rum1* and *rul1* displayed the highest expression in the cortex and stele of 3-day-old primary roots (Fig. 2A). Among all tested tissues, *rum1* and *rul1* displayed the lowest expression in the elongation zone of 3-day-old primary roots. Furthermore, both genes were expressed at higher levels in seminal and crown roots than in all other tested stages of primary roots. Despite the significant differences in overall expression, across all tissues *rum1* and *rul1* expression was significantly correlated (Fig. 2B). For all analysed tissues a Pearson correlation coefficient of $R=0.8$ ($P \leq 0.01$) was calculated. Pairwise *t*-tests comparing expression of the two genes in all tested tissues are summarized in Supplementary Table S2. Thus far two types of auxin response promoter elements (AuxRE), 5′ TGTCTC 3′ and 5′ TGTC 3′, have been described. Promoter analysis of 1 kb upstream of the ATG start codon of *rul1* revealed six 5′ TGTC 3′ elements. Auxin-inducibility of *rul1* was tested by qRT-PCR in 5-day-old B73 primary roots after 5 µM 1-NAA treatment (Fig. 2C). The experiment demonstrated that *rul1* is auxin inducible (FC=3.7) within 3 h. Only a relatively small expression difference of *rul1* at a moderate significance was observed in wild-type vs. *rum-R* mutant primary roots (Fig. 2D), suggesting that *rul1* expression might not be controlled by *rum1*.

RUL1 interacts with ZmARF25 and ZmARF34, RUM1 and itself

Aux/IAA proteins are characterized by their capability of interacting with ARF proteins. Interactions of RUL1 with ZmARF25 and ZmARF34, which have been previously demonstrated to interact with RUM1 (von Behrens *et al.*, 2011), were surveyed in BiFC (Bimolecular Fluorescence Complementation) experiments. This *in vivo* technique is based on the detection of a YFP (Yellow Fluorescent Protein) signal which is emitted when N (YFPN) and C (YFPC) terminal YFP parts come in close proximity by the interactions of fusion proteins coupled to these YFP subunits.

Quantification of the YFP fluorescence by flow cytometry demonstrated significant interactions of RUL1 with ZmARF25 and ZmARF34 in Arabidopsis protoplasts compared to the control experiments (Fig. 3A). Moreover, significant homo-interaction was observed for RUM1-RUM1 and RUL1-RUL1. Furthermore, hetero-interaction of RUM1-RUL1 in both orientations in comparison to the corresponding negative controls was demonstrated (Fig. 3B). Each experiment was performed in three biological replicates. The fusion proteins NUO-YFPC and NUO-YFPN, which were used as a negative control (von Behrens *et al.*, 2011), did not display any interaction with RUL1, ARF25, and ARF34. Expression of fusion proteins in BiFC assays according to Fig. 3A, B were confirmed by Western blot experiments (Supplementary Fig. S1).

RAP1specifically interacts with RUM1 but not with RUL1

To identify novel interaction partners for RUM1 and RUL1, a yeast two-hybrid assay was performed using RUM1 as bait and a cDNA expression library generated from mRNA of 2.5-day-old maize primary roots as prey (methods). This experiment revealed known interaction partners of RUM1 and RUL1 such as RUM1 and ARF25 (Supplementary Table S3). To identify interaction partners of RUM1 and RUL1 involved in auxin signal transduction, 1 kb promoter sequences upstream of the ATG start codon of yeast two-hybrid candidate genes were screened for auxin response elements (AuxREs). The promoter of a gene designated *rap1* contained seven AuxREs of the type 5′ TGTCTC 3′ (once) and 5′ TGTC 3′ (six times) (Fig. 4A). Auxin induction of *rap1* in 5-day-old wild-type primary roots was demonstrated after a 3 h treatment with 5 µM 1-NAA (Fig. 4B).

Subseqeunetly, BiFC experiments were performed to analyse the interaction of RAP1 (RUM1 ASSOCIATED PROTEIN 1) with RUM1, RUL1 and rum1-R, a mutated form of RUM1 lacking 24 amino acids in domain II. These experiments confirmed the interaction of RUM1 and rum1-R with RAP1 in Arabidopsis protoplasts. Remarkably, interaction of the mutant protein rum1-R with RAP1 in Arabidopsis protoplasts was significantly stronger than RUM1-RAP1 interaction (Fig. 4C), possibly due to the increased stability of the mutated proteins. In contrast, no interaction was observed between RUL1 and RAP1 compared to the corresponding negative controls in BiFC analyses (Fig. 4C). Western blot analyses were performed with Anti-HA (haemagglutinin) antibodies against the YFPC and anti-c-Myc antibodies against the YFPN tag, which demonstrated the expression of the corresponding proteins in Arabidopsis protoplasts (Supplementary Fig. S1).

BiFC assays with domain I, II, and III-IV of RUM1 with RAP1 were performed to resolve which domain of the RUM1 protein interacts with RAP1. These experiments demonstrated that the interaction of domain I of RUM1 with RAP1 was significantly stronger than the interaction of domain III-IV of RUM1 with RAP1 (Fig. 4D). In contrast, no interaction was detected between domain II of RUM1 and RAP1 in BiFC experiments relative to the negative controls (Fig. 4D). These results suggested that domain I of RUM1 is mainly responsible for interaction with RAP1 although the truncated domain III-IV in RUM1 also interacted with RAP1. Expression of all

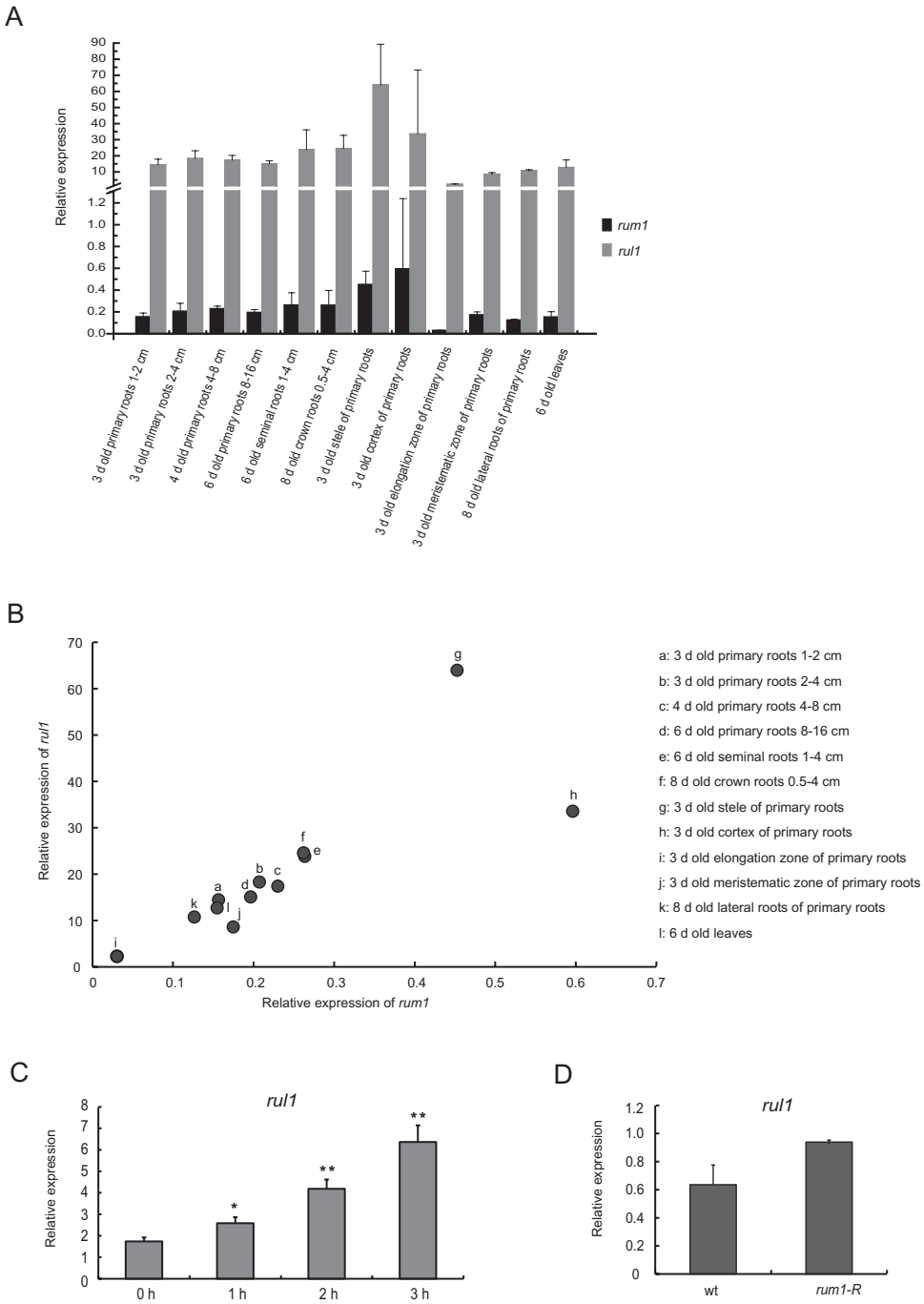


Fig. 2. Expression of *rum1* and *rul1*. (A) qRT-PCR analyses of *rul1* versus *rum1* in 12 different tissues at different developmental stages. (B) Correlation of *rum1* and *rul1* gene expression patterns. (C) Auxin induction of *rul1* in primary roots of wild-type seedlings assayed by qRT-PCR after treatment with 5 μM 1-NAA at 0, 1, 2 and 3 h, respectively (*, $P \leq 0.05$; **, $P \leq 0.01$; $n=4$). (D) Expression of *rul1* in wild-type versus *rum1-R* primary roots assayed by qRT-PCR experiments (Fc=1.5; $P \leq 0.1$).

experimental fusion-proteins in Arabidopsis protoplasts were confirmed by Western blot assays (Supplementary Fig. S1).

The RAP1 family in maize

Homology searches using the maize RAP1 protein sequence as query revealed six additional proteins of this family ZmRAP1-like1 (ZmRAL1) to ZmRAP1-like6 (ZmRAL6). Four of seven maize genes were assigned to maize subgenome 1, while the remaining three members were not assigned to a subgenome. All four maize genes assigned to subgenome 1 (ZmRAL1, ZmRAL3, ZmRAL5, ZmRAL6) have an ortholog in rice and sorghum (Supplementary Table S4).

Discussion

RUM1 and RUL1 display characteristics of canonical Aux/IAA proteins

About 5–12 million years ago a maize progenitor had undergone a whole genome duplication which led to the emergence

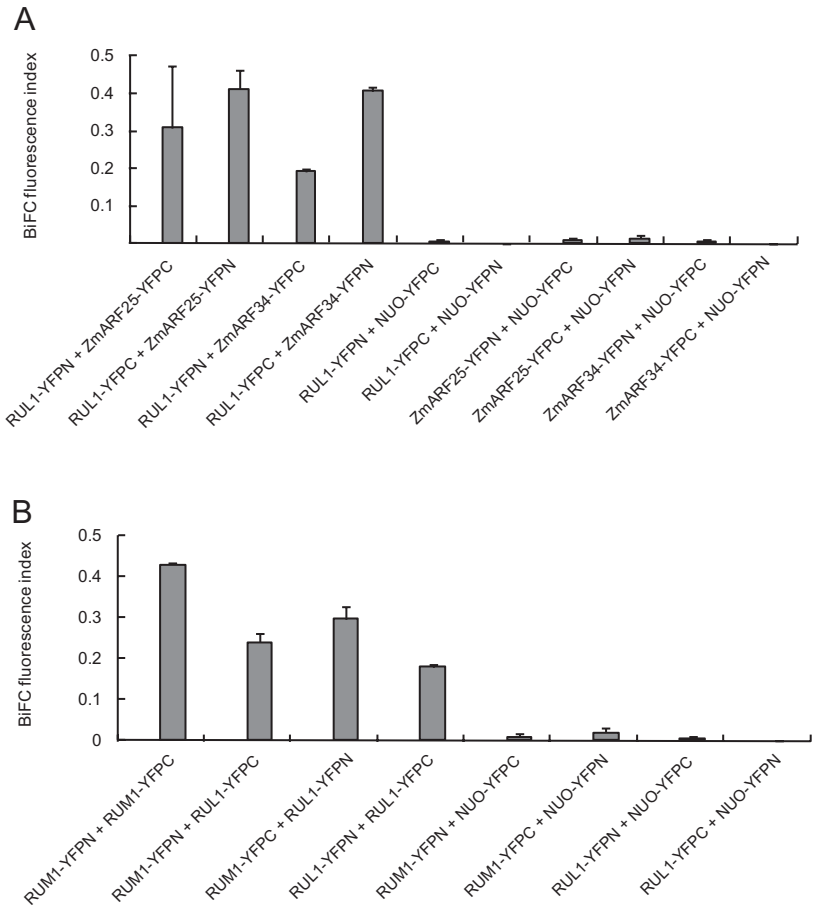


Fig. 3. Quantification of RUL1 protein interactions with ZmARF25, ZmARF34 and RUM1 by the BiFC system in Arabidopsis Col-0 protoplasts. (A) RUL1 interaction with ZmARF25 and ZmARF34. (B) Homo- and hetero-interaction of RUM1 and RUL1.

of two subgenomes (Schnable *et al.*, 2011). During evolution, subgenome 2 experienced more gene loss than subgenome 1. Moreover, mutant phenotypes identified in forward genetic screens are often the result of a mutation in genes of maize subgenome 1 (Schnable and Freeling, 2011). This observation is explained by the hypothesis that subgenome 1 genes have predominantly retained the ancestral function while subgenome 2 genes potentially adopted new, or less essential functions (Schnable and Freeling, 2011). In line with this, several genes controlling maize root development such as *rtcs* (Taramino *et al.*, 2007), *rth3* (Hochholdinger *et al.*, 2008) and *rum1* (von Behrens *et al.*, 2011) belong to subgenome 1. Their homeologs in maize subgenome 2 are designated *rtcl* (Taramino *et al.*, 2007), *rtl3* (Hochholdinger *et al.*, 2008) and *rul1* (von Behrens *et al.*, 2011).

The proteins encoded by *rum1* and *rul1* display the canonical four domain architecture defining Aux/IAA proteins (Liscum and Reed, 2002). Aux/IAA proteins are involved in the transcriptional regulation of auxin responsive genes (Quint and Gray, 2006) and are therefore localized in the nucleus. In the present study, nuclear localization of RUL1 was demonstrated by RUL1-GFP localization in Arabidopsis protoplasts (Fig. 1B). This localization pattern was also observed for RUM1 (von Behrens *et al.*, 2011) and several other Aux/IAA proteins (Ludwig *et al.*, 2014).

Aux/IAA proteins are unstable with short half-lives of between six and 80 minutes due to their interaction with the SCF^{TIR} complex via domain II and subsequent proteasomal degradation (Abel *et al.*, 1994; Gray *et al.*, 2001; Ouellet *et al.*, 2001). The instability of Aux/IAA proteins is conferred by interaction with the SCF^{TIR1} complex at the conserved degron sequence GWPPV (Dreher *et al.*, 2006). Point mutations in this short amino acid stretch are often sufficient to prohibit the interaction and thus enhance the stability of Aux/IAA proteins (Tian *et al.*, 2003). In the present study, wild-type RUL1-GFP displayed a half-life of ~23 min (Fig. 1C) which was similar to RUM1 (~22 min; von Behrens *et al.*, 2011). Remarkably, while the point mutation that led to a P-to-L exchange in position 122 was sufficient to stabilize the rul1-P122L protein, the same amino exchange at position 121 did not stabilize the protein (Fig. 1C) suggesting that SCF^{TIR1} interaction and subsequent proteasomal degradation is still possible in this mutated protein to a considerable degree. In Arabidopsis several mutants with a developmental phenotype have been identified with P-to-L exchanges that correspond to position 121 (Rogg *et al.*, 2001; Tatematsu *et al.*, 2004) and 122 (Rouse *et al.*, 1998; Knox *et al.*, 2003; Tatematsu *et al.*, 2004) in the degron sequence. These aberrant phenotypes are likely conditioned by stabilized mutated Aux/IAA proteins. In maize, as suggested by rul1-P121L-GFP

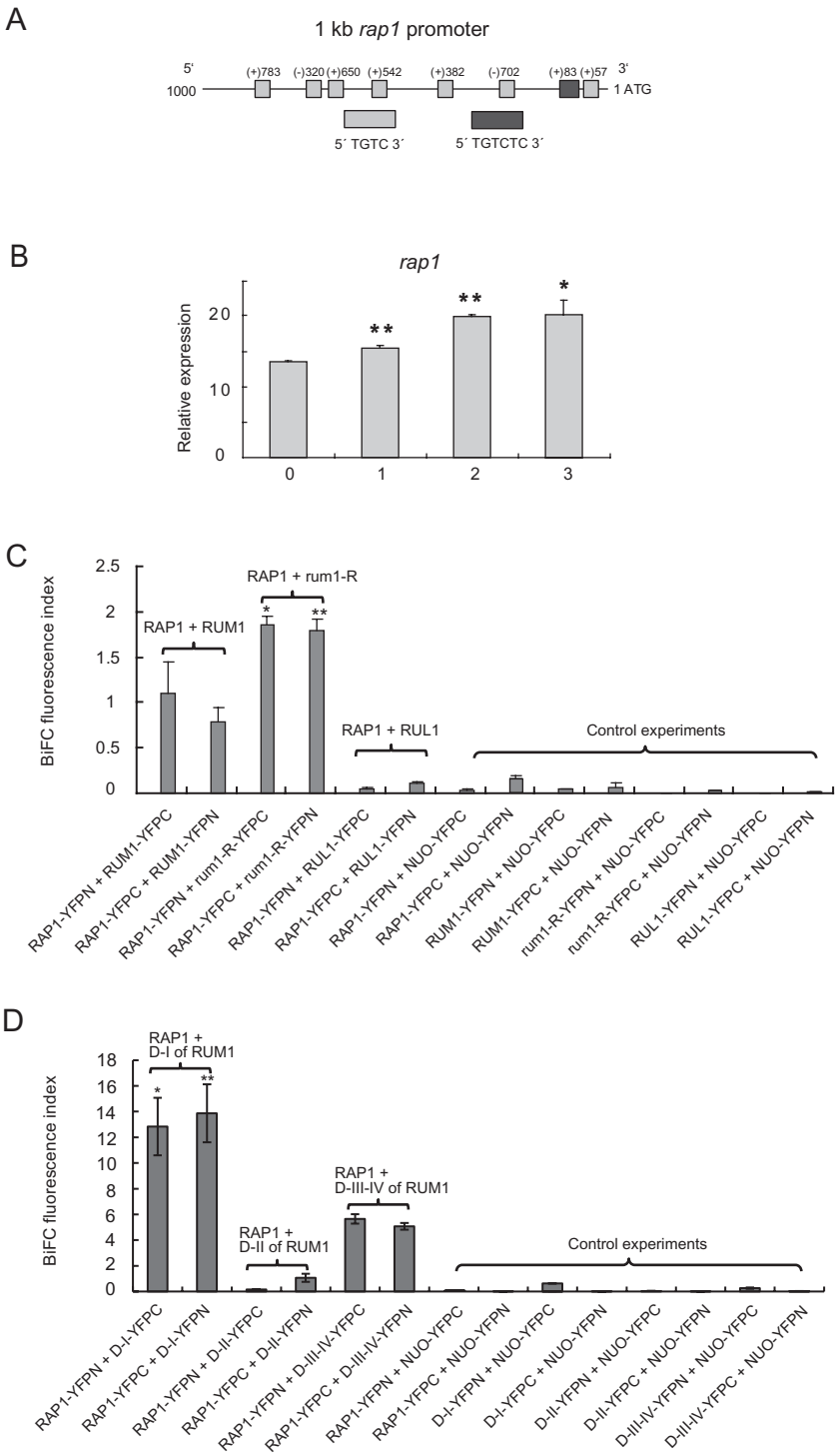


Fig. 4. Auxin related characteristics of *rap1* and specific interaction of RAP1 with RUM1 but not RUL1. (A) AuxRE analysis of 1-kb *rap1* promoter. 5′ TGTCTC 3′ is represented by a dark grey box. 5′ TGTC 3′ is denoted by a light grey box. (B) Expression of *rap1* gene in primary roots of wild-type seedlings after treatment with 5 μM 1-NAA at 0, 1, 2 and 3 h, respectively, assayed by qRT-PCR (*, $P \leq 0.05$; **, $P \leq 0.01$; n=3). (C) RAP1 interaction with RUM1, rum1-R and RUL1. (D) RAP1 interaction with domains I, II and III–IV of RUM1. In all experiments the NUO protein was used as a negative control.

and rul1-P122L-GFP, distinct positions in the degron motif might contribute differently to protein interaction with the SCF^{TIR} complex.

Aux/IAA proteins repress the transcription of early auxin-responsive genes by their interaction with ARF proteins (Woodward and Bartel, 2005). In the present study, we demonstrated interaction of RUL1 with ZmARF25 and ZmARF34 (Fig. 3A) as previously demonstrated for RUM1 (von Behrens *et al.*, 2011). It was suggested that this interaction blocks lateral root formation in non-precursor pericycle cells (von Behrens *et al.*, 2011). In Arabidopsis, multiple models of Aux/IAA-ARF-dependent auxin response signalling involved in lateral root development were proposed. First, the IAA28-ARF-dependent model

mediates the specification of lateral root founder cell identity (De Rybel *et al.*, 2010). Second, the SLR/IAA14-ARF7-ARF19 module controls the division of early founder cells of lateral roots, and subsequently the BDL/IAA12-ARF5 module regulates lateral root initiation and organogenesis (De Smet *et al.*, 2010). Hence, RUM1 and RUL1 might also be involved in different pathways involved in lateral root formation.

In addition to the canonical features of RUL1, it was demonstrated in the present survey that RUM1 and RUL1 can form homo- and heterodimers *in vivo* (Fig. 3B). It was previously demonstrated that the Arabidopsis proteins IAA1, IAA2 and the pea protein IAA4 can form homodimers *in vitro* (Kim *et al.*, 1997). Nevertheless, the function of Aux/IAA interactions still remains elusive and it has been suggested that it might allow interaction with downstream genes without the formation of Aux/IAA-ARF complexes (Paciorek and Friml, 2006). Similarly, direct binding of RUM1 to the promoter of *lrp1* (*lateral root primordia 1*) has also been suggested (Zhang *et al.*, 2015).

In general rul1 *displays higher expression than* rum1

It was observed that on average genes of maize subgenome 1 were expressed at higher levels than their homeologous genes in subgenome 2 (Schnable *et al.*, 2011). In contrast to this trend, *rul1* (subgenome 2) displayed on average a ~84-fold higher expression than *rum1* (subgenome 1) in the 12 tissues surveyed in the present study (Fig. 2A). Despite highly correlated expression patterns (Fig. 2B), the significantly differential expression intensities of *rul1* and *rum1* might suggest distinct functions of these two homeologous proteins in root development.

Promoter analysis of the sequence 1 kb upstream of the ATG start codon revealed 13 putative AuxREs in *rum1* and six AuxRE in *rul1*. Despite the different number and position of AuxRE elements it was demonstrated that *rul1* is auxin inducible as previously demonstrated for *rum1* (von Behrens *et al.*, 2011). Moreover, by comparing *rul1* expression in wild-type and *rum1-R* mutant primary roots it was demonstrated that the expression of *rul1* was not regulated by *rum1*. This result supports the notion that these genes might act in different molecular pathways and might therefore have different functions in root development.

For the recessive loss-of-function mutants *rtcs* and *rth3*, different functions compared to their homeologous genes were demonstrated since the homeologs were not able to complement the mutant phenotypes (Taramino *et al.*, 2007; Hochholdinger *et al.*, 2008). However for the semi-dominant mutation *rum1-R* the situation is different. The *rum1-R* mutant phenotype is conferred by a stabilization of the rum1-R/ARF complex, which inhibits the expression of downstream gene expression (von Behrens *et al.*, 2011). Hence, RUL1 cannot complement the mutant phenotype because downstream gene expression is already blocked by the gene product of the gain-of-function allele rum1-R which may not allow redundancy in this process.

RAP1 specifically interacts with RUM1 but not with RUL1

AtSPR1 is a plant-specific small protein. Homology searches revealed similarity of AtSPR1 with a nitrilase-associated protein (GenBank AC: Z96936) in Arabidopsis (Nakajima *et al.*, 2004). Yeast two-hybrid experiments demonstrated interaction of RUM1 with a novel protein which is a homolog of AtSPR1 and which we designated RAP1. Homology searches (ensembl.gramene.org) identified a total of seven homeologous maize genes *rap1* and *rap1-like1* (*ral1*) to *ral6* (Supplementary Table S4). Four of seven *rap1-like* gene family members (57%) were assigned to maize subgenome 1 (Supplementary Table S4). This tendency was also observed for the *lrp1-like* gene family where five of nine (56%) assigned to subgenome 1. The remaining three (43%) *rap1-like* genes likely emerged by single copy duplications after the last whole genome duplication because they did not map to any of the subgenomes. Similarly, seven *Aux/IAA* genes were not associated with a subgenome, suggesting that they emerged after the ancient genome duplication of maize (Ludwig *et al.*, 2013).

Quantification of the interaction of RUM1, rum1-R and RUL1 with RAP1 (Fig. 4C) demonstrated interaction of RUM1 with RAP1, and an even stronger interaction of rum1-R with RAP1 *in vivo*. This is most likely a consequence of the observation that rum1-R is more stable than RUM1 (von Behrens *et al.*, 2011). The homeologous maize proteins RUM1 and RUL1 share 92% identity on the protein level (von Behrens *et al.*, 2011). However, no interaction was detected between RUL1 and RAP1. A domain interaction analysis revealed that the interaction of domain I of RUM1 with RAP1 is significantly stronger than the interaction of domains III–IV, while no interaction was detected with domain II in BiFC experiments (Fig. 4D). It has been demonstrated that domain I of Aux/IAA is also responsible for protein-protein heterodimerization. For example, domain I of BDL/IAA12 was sufficient to interact with TOPLESS which is a co-repressor regulating embryogenesis in Arabidopsis (Szemenyei *et al.*, 2008).

In summary, we demonstrated that both RUM1 and RUL1 display all characteristics of functional Aux/IAA proteins. Distinct functions of the two proteins are suggested by a different promoter architecture and overall differences in gene expression levels. Moreover, it was demonstrated that *rul1* is not regulated by RUM1 suggesting their activity in independent pathways. Finally, RUM1 specific interaction with RAP1 suggests that these homeologous genes, despite their role as Aux/IAA proteins, have at least in part diverse interaction partners and might thus be functioning in distinct molecular networks.

Supplementary data

Supplementary data are available at *JXB* online.

Fig. S1. Expression of fusion proteins in Arabidopsis Col-0 protoplasts detected by Western blot experiments.

Table S1. Sequences of oligonucleotide primers used in this study.

Table S2. Pairwise comparison of *rum1* (upper table) and *rul1* (lower table) expression in different tissues and at different developmental stages according to Fig. 2A.

Table S3. RUM1 interaction partners identified via yeast two-hybrid experiments.

Table S4. Characteristics of members of the maize *rap1-like* gene family in maize.

Acknowledgements

We would like to thank Claudia Oecking (ZMBP, University of Tuebingen, Germany) for the modified pUC-SPYNE-152 vector and Caterina Brancato (ZMBP, Transformation Unit, University of Tuebingen, Germany) for the preparation of protoplasts and excellent technical support. This project was supported by a DFG (Deutsche Forschungsgemeinschaft) grant to FH and a CSC (China Scholarship Council) fellowship to YZ.

References

Abel S. 2007. Auxin is surfacing. ACS Chemical Biology **2,** 380–384.

Abel S, Oeller PW, Theologis A. 1994. Early auxin-induced genes encode short-lived nuclear proteins. Proceedings of the National Academy of Sciences, USA **91,** 326–330.

Benjamins R, Scheres B. 2008. Auxin: the looping star in plant development. Annual Reviews in Plant Biology **59,** 443–465.

Casimiro I, Marchant A, Bhalerao RP, *et al*. 2001. Auxin transport promotes Arabidopsis lateral root initiation. The Plant Cell **13,** 843–852.

Chandler VL, Radicella JP, Robbins TP, Chen J, Turks D. 1989. Two regulatory genes of the maize anthocyanin pathway are homologous: isolation of B utilizing R genomic sequences. The Plant Cell **1,** 1175–1183.

Cone KC, Cocciolone SM, Burr FA, Burr B. 1993. Maize anthocyanin regulatory gene *pl* is a duplicate of *c1* that functions in the plant. The Plant Cell **5,** 1795–1805.

De Rybel B, Vassileva V, Parizot B, *et al*. 2010. A novel aux/IAA28 signaling cascade activates GATA23-dependent specification of lateral root founder cell identity. Current Biology **20,** 1697–1706.

De Smet I, Lau S, Voss U, *et al*. 2010. Bimodular auxin response controls organogenesis in Arabidopsis. Proceedings of the National Academy of Sciences USA **107,** 2705–2710.

Dreher KA, Brown J, Saw RE, Callis J. 2006. The Arabidopsis Aux/IAA protein family has diversified in degradation and auxin responsiveness. The Plant Cell **18,** 699–714.

Gray WM, Kepinski S, Rouse D, Leyser O, Estelle M. 2001. Auxin regulates SCF(TIR1)-dependent degradation of AUX/IAA proteins. Nature **414,** 271–276.

Hochholdinger F, Park WJ, Sauer M, Woll K. 2004. From weeds to crops: genetic analysis of root development in cereals. Trends in Plant Science **9,** 42–48.

Hochholdinger F, Tuberosa R. 2009. Genetic and genomic dissection of maize root development and architecture. Current Opinion in Plant Biology **12,** 172–177.

Hochholdinger F, Wen TJ, Zimmermann R, Chimot-Marolle P, da Costa e Silva O, Bruce W, Lamkey KR, Wienand U, Schnable PS. 2008. The maize (*Zea mays* L.) *roothairless 3* gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield. The Plant Journal **54,** 888–898.

Hochholdinger F, Zimmermann R. 2008. Conserved and diverse mechanisms in root development. Current Opinion in Plant Biology **11,** 70–74.

Hoecker N, Keller B, Piepho HP, Hochholdinger F. 2006. Manifestation of heterosis during early maize (*Zea mays* L.) root development. Theoretical and Applied Genetics **112,** 421–429.

Jansen L, Roberts I, De Rycke R, Beeckman T. 2012. Phloem-associated auxin response maxima determine radial positioning of lateral roots in maize. Philosophical Transactions of the Royal Society London B: Biological Sciences **367,** 1525–1533.

Kim J, Harter K, Theologis A. 1997. Protein-protein interactions among the Aux/IAA proteins. Proceedings of the National Academy of Sciences, USA **94,** 11786–11791.

Knox K, Grierson CS, Leyser O. 2003. AXR3 and SHY2 interact to regulate root hair development. Development **130,** 5769–5777.

Lawit SJ, Wych HM, Xu D, Kundu S, Tomes DT. 2010. Maize DELLA proteins dwarf plant8 and dwarf plant9 as modulators of plant development. Plant and Cell Physiology **51,** 1854–1868.

Li M, Doll J, Weckermann K, Oecking C, Berendzen K, Schöffl F. 2010. Detection of in vivo interactions between Arabidopsis class A-HSFs, using a novel BiFC fragment, and identification of novel class B-HSF interacting proteins. European Journal of Cell Biology **89,** 126–132.

Liscum E, Reed JW. 2002. Genetics of Aux/IAA and ARF action in plant growth and development. Plant Molecular Biology **49,** 387–400.

Liu LH, Ludewig U, Frommer WB, von Wiren N. 2003. *AtDUR3* encodes a new type of high-affinity urea/H+ symporter in Arabidopsis. The Plant Cell **15,** 790–800.

Ludwig Y, Berendzen KW, Xu C, Piepho HP, Hochholdinger F. 2014. Diversity of stability, localization, interaction and control of downstream gene activity in the Maize Aux/IAA protein family. PLoS One **9,** e107346.

Ludwig Y, Zhang Y, Hochholdinger F. 2013. The maize (*Zea mays* L.) *AUXIN/INDOLE-3-ACETIC ACID* gene family: phylogeny, synteny, and unique root-type and tissue-specific expression patterns during development. PLoS One **8,** e78859.

Merkle T, Leclerc D, Marshallsay C, Nagy F. 1996. A plant in vitro system for the nuclear import of proteins. The Plant Journal **10,** 1177–1186.

Nakajima K, Furutani I, Tachimoto H, Matsubara H, Hashimoto T. 2004. *SPIRAL1* encodes a plant-specific microtubule-localized protein required for directional control of rapidly expanding Arabidopsis cells. The Plant Cell **16,** 1178–1190.

Negrutiu I, Shillito R, Potrykus I, Biasini G, Sala F. 1987. Hybrid Genes in the Analysis of Transformation Conditions. Plant Molecular Biology **8,** 363–373.

Ouellet F, Overvoorde PJ, Theologis A. 2001. IAA17/AXR3: biochemical insight into an auxin mutant phenotype. The Plant Cell **13,** 829–841.

Paciorek T, Friml J. 2006. Auxin signaling. Journal of Cell Science **119,** 1199–1202.

Peret B, De Rybel B, Casimiro I, Benkova E, Swarup R, Laplaze L, Beeckman T, Bennett MJ. 2009. Arabidopsis lateral root development: an emerging story. Trends in Plant Science **14,** 399–408.

Quint M, Gray WM. 2006. Auxin signaling. Current Opinion in Plant Biology **9,** 448–453.

Reed RC, Brady SR, Muday GK. 1998. Inhibition of auxin movement from the shoot into the root inhibits lateral root development in Arabidopsis. Plant Physiology **118,** 1369–1378.

Rogg LE, Lasswell J, Bartel B. 2001. A gain-of-function mutation in IAA28 suppresses lateral root development. The Plant Cell **13,** 465–480.

Rouse D, Mackay P, Stirnberg P, Estelle M, Leyser O. 1998. Changes in auxin response from mutations in an *AUX/IAA* gene. Science **279,** 1371–1373.

Saleem M, Lamkemeyer T, Schützenmeister A, Madlung J, Sakai H, Piepho HP, Nordheim A, Hochholdinger F. 2010. Specification of cortical parenchyma and stele of maize primary roots by asymmetric levels of auxin, cytokinin, and cytokinin-regulated proteins. Plant Physiology **152,** 4–18.

Schnable JC, Freeling M. 2011. Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS One **6,** e17855.

Schnable JC, Springer NM, Freeling M. 2011. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proceedings of the National Academy of Sciences, USA **108,** 4069–4074.

Szemenyei H, Hannon M, Long JA. 2008. TOPLESS mediates auxin-dependent transcriptional repression during Arabidopsis embryogenesis. Science **319,** 1384–1386.

Tan X, Calderon-Villalobos LI, Sharon M, Zheng C, Robinson CV, Estelle M, Zheng N. 2007. Mechanism of auxin perception by the TIR1 ubiquitin ligase. Nature **446,** 640–645.

Taramino G, Sauer M, Stauffer JL, Jr., Multani D, Niu X, Sakai H, Hochholdinger F. 2007. The maize (*Zea mays* L.) RTCS gene encodes a LOB domain protein that is a key regulator of embryonic seminal and post-embryonic shoot-borne root initiation. The Plant Journal **50,** 649–659.

Tatematsu K, Kumagai S, Muto H, Sato A, Watahiki MK, Harper RM, Liscum E, Yamamoto KT. 2004. *MASSUGU2* encodes Aux/IAA19, an auxin-regulated protein that functions together with the transcriptional activator NPH4/ARF7 to regulate differential growth responses of hypocotyl and formation of lateral roots in *Arabidopsis thaliana*. The Plant Cell **16,** 379–393.

Tian Q, Nagpal P, Reed JW. 2003. Regulation of Arabidopsis SHY2/IAA3 protein turnover. The Plant Journal **36,** 643–651.

Tiwari SB, Hagen G, Guilfoyle TJ. 2004. Aux/IAA proteins contain a potent transcriptional repression domain. The Plant Cell **16,** 533–543.

Ulmasov T, Hagen G, Guilfoyle TJ. 1997. ARF1, a transcription factor that binds to auxin response elements. Science **276,** 1865–1868.

von Behrens I, Komatsu M, Zhang YX, Berendzen KW, Niu XM, Sakai H, Taramino G, Hochholdinger F. 2011. *Rootless with undetectable meristem 1* encodes a monocot-specific AUX/IAA protein that controls embryonic seminal and post-embryonic lateral root initiation in maize. The Plant Journal **66,** 341–353.

Walter M, Chaban C, Schutze K, *et al*. 2004. Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. The Plant Journal **40,** 428–438.

Woll K, Borsuk LA, Stransky H, Nettleton D, Schnable PS, Hochholdinger F. 2005. Isolation, characterization, and pericycle-specific transcriptome analyses of the novel maize lateral and seminal root initiation mutant *rum1*. Plant Physiology **139,** 1255–1267.

Woodward AW, Bartel B. 2005. A receptor for auxin. The Plant Cell **17,** 2425–2429.

Worley CK, Zenser N, Ramos J, Rouse D, Leyser O, Theologis A, Callis J. 2000. Degradation of Aux/IAA proteins is essential for normal auxin signalling. The Plant Journal **21,** 553–562.

Wright AJ, Gallagher K, Smith LG. 2009. *discordia1* and *alternative discordia1* function redundantly at the cortical division site to promote preprophase band formation and orient division planes in maize. The Plant Cell **21,** 234–247.

Zhang Y, Behrens I, Zimmermann R, Ludwig Y, Hey S, Hochholdinger F. 2015. LATERAL ROOT PRIMORDIA 1 of maize acts as a transcriptional activator in auxin signalling downstream of the *Aux/IAA* gene *rootless with undetectable meristem 1*. Journal of Experimental Botany **66,** 3855–3863.

Zhang Y, Paschold A, Marcon C, *et al*. 2014. The *Aux/IAA* gene *rum1* involved in seminal and lateral root formation controls vascular patterning in maize (*Zea mays* L.) primary roots. Journal of Experimental Botany **65,** 4919–4930.

Journal of Experimental Botany, Vol. 67, No. 4 pp. 1123–1135, 2016
doi:10.1093/jxb/erv513 Advance Access publication 30 November 2015
This paper is available online free of all access charges (see http://jxb.oxfordjournals.org/open_access.html for further details)

RESEARCH PAPER

Transcriptomic and anatomical complexity of primary, seminal, and crown roots highlight root type-specific functional diversity in maize (*Zea mays* L.)

Huanhuan Tai[1], Xin Lu[2], Nina Opitz[1], Caroline Marcon[1], Anja Paschold[1,*], Andrew Lithio[3], Dan Nettleton[3] and Frank Hochholdinger[1,†]

[1] Institute of Crop Science and Resource Conservation, Crop Functional Genomics, University of Bonn, D-53113 Bonn, Germany
[2] Experimental Medicine and Therapy Research, University of Regensburg, D-93053 Regensburg, Germany
[3] Department of Statistics, Iowa State University, Ames, IA 50011-1210, USA

* Present address: Max-Planck-Institute for Plant Breeding Research, D-50829 Cologne, Germany.
† Correspondence: hochholdinger@uni-bonn.de

Received 9 October 2015; Accepted 6 November 2015

Editor: Adam Price, University of Aberdeen

Abstract

Maize develops a complex root system composed of embryonic and post-embryonic roots. Spatio-temporal differences in the formation of these root types imply specific functions during maize development. A comparative transcriptomic study of embryonic primary and seminal, and post-embryonic crown roots of the maize inbred line B73 by RNA sequencing along with anatomical studies were conducted early in development. Seminal roots displayed unique anatomical features, whereas the organization of primary and crown roots was similar. For instance, seminal roots displayed fewer cortical cell files and their stele contained more meta-xylem vessels. Global expression profiling revealed diverse patterns of gene activity across all root types and highlighted the unique transcriptome of seminal roots. While functions in cell remodeling and cell wall formation were prominent in primary and crown roots, stress-related genes and transcriptional regulators were over-represented in seminal roots, suggesting functional specialization of the different root types. Dynamic expression of lignin biosynthesis genes and histochemical staining suggested diversification of cell wall lignification among the three root types. Our findings highlight a cost-efficient anatomical structure and a unique expression profile of seminal roots of the maize inbred line B73 different from primary and crown roots.

Key words: Crown root, primary root, RNA-seq, root type, seminal root, transcriptome, *Zea mays*.

Introduction

Roots play a vital role in plant development and fitness because they provide mechanical support, mediate water and nutrient uptake from soil, and interact with microbial communities in the rhizosphere (reviewed in Hawkes *et al.*, 2007; Zhu *et al.*, 2011; Villordon *et al.*, 2014). Sophisticated root system architecture is a prerequisite for optimal capturing of soil resources (Giehl and von Wirén, 2014). For instance, in maize, a deeper root system enhances drought tolerance (Ribaut *et al.*, 2009), while shallower root angles of seminal and crown roots increase phosphorus acquisition (Zhu *et al.*, 2005).

© The Author 2015. Published by Oxford University Press on behalf of the Society for Experimental Biology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

The root stock of maize consists of an embryonically formed primary root and a variable number of embryonic seminal roots and post-embryonic shoot-borne crown and brace roots (reviewed in Feldman, 1994; Hochholdinger *et al.*, 2004a, b). These root types are initiated from distinct tissues during successive stages of development. Primordia of primary roots are formed at the basal pole of the embryo and become visible as distinct structures within the embryo ~10–15 d after pollination (Tillich, 1977). After germination, the primary root emerges within 2–3 d. Seminal root primordia are formed in the embryo 22–40 d after pollination (Erdelska and Vidovencova, 1993). The number of seminal roots is largely determined by their genetic background (Sass, 1977; Erdelska and Vidovencova, 1993). Seminal roots are initiated at the scutellar node ~1 week after germination. Up to 12 seminal roots are formed per seedling, which thus dominate the seedling root stock (Abbe and Stein, 1954). In the first weeks after germination, primary and seminal roots make up the major portion of the seedling root stock (Hochholdinger *et al.*, 2004b) and are thus vital for the early vigor of young maize seedlings (Peter *et al.*, 2009). In contrast, the post-embryonic shoot-borne root system shapes the adult root stock (Martin and Harris, 1976). Shoot-borne roots formed from consecutive shoot nodes below the soil level are designated crown roots, while brace roots are initiated from above-ground shoot nodes (Hochholdinger *et al.*, 2004a). The first whorl of shoot-borne crown roots is formed ~10–14 days after germination (dag) (Hochholdinger *et al.*, 2004b). An adult maize root stock develops ~70 shoot-borne roots, which are organized on average in six whorls of underground crown roots and 2–3 whorls of above-ground brace roots (Hoppe *et al.*, 1986). A commonality of all major embryonic and post-embryonic root types is their ability to initiate post-embryonic lateral roots from pericycle and endodermis cells (Fahn, 1990).

A number of maize developmental mutants with specific defects in their root system architecture have been identified. These mutants are affected in shoot-borne root formation (Hetz *et al.*, 1996), lateral root formation (Hochholdinger and Feix, 1998; Hochholdinger *et al.*, 2001; Woll *et al.*, 2005), or root hair formation (Wen and Schnable, 1994; Wen *et al.*, 2005; Hochholdinger *et al.*, 2008; Nestler *et al.*, 2014). Among those, seminal and shoot-borne root formation are controlled by *rtcs* (*rootless concerning crown and seminal roots*) (Hetz *et al.*, 1996) which encodes an auxin-inducible LOB (lateral organ boundaries) domain transcription factor (TF) (Taramino *et al.*, 2007) acting downstream of the auxin response factor ARF34 (Majer *et al.*, 2012; Xu *et al.*, 2015). The *rum1* (*rootless with undetectable meristems 1*) gene encodes a monocot-specific Aux/IAA protein (von Behrens *et al.*, 2011) that controls the formation of seminal roots and lateral roots at the primary root (Woll *et al.*, 2005). These genetic studies supported the notion of root type-specific control of maize root system formation (Hochholdinger *et al.*, 2004a). Gene expression profiles of different maize root tissues generated by microarray analyses (Woll *et al.*, 2005; Jiang *et al.*, 2006; Dembinsky *et al.*, 2007; Muthreich *et al.*, 2013) and RNA sequencing (RNA-seq) studies (Paschold et al., 2012, 2014; Zhang *et al.*, 2014) provided initial cues of molecular networks involved in maize root development. In this study, we applied RNA-seq to profile global expression patterns of genes expressed in primary, seminal, and crown roots of the maize inbred line B73 during early development. In parallel, a detailed comparative investigation of anatomical and histochemical features across the root types was conducted. The goal of this study was to survey anatomical and transcriptomic differences to reveal functional differences between the different root types.

Materials and methods

Plant material and growth conditions

Seeds of the maize inbred line B73 were surface sterilized with 6% hypochlorite for 10 min, and then germinated in paper rolls imbibed in distilled water for 2 d (Hetz *et al.*, 1996). Subsequently, these paper rolls were transferred to Hoagland's solution (Zhao *et al.*, 2012) in a growth chamber (16 h light, 28 °C, and 8 h dark, 22 °C; 60% humidity) until the roots were harvested. The culture solution was replaced every second day.

RNA sequencing and read mapping

For RNA-seq, pools of 10 roots (length: 20–30 mm) per root type were sampled per biological replicate. For each root type, four independent biological replicates were analyzed. The sampled plant material was immediately frozen in liquid nitrogen and stored at –80 °C until RNA isolation.

Total RNA was extracted by using the RNeasy Plant Mini Kit (Qiagen, Venlo, The Netherlands). RNA integrity and quality were assayed by agarose gel electrophoresis and a Bioanalyzer using an Agilent RNA 6000 Nano Chip (Agilent Technologies, Santa Clara, CA, USA). High quality samples with RIN (RNA integrity number) values >8 were used for RNA-seq. cDNA libraries were prepared according to the manufacturer´s protocol (Illumina, San Diego, CA, USA) and subsequently sequenced on an Illumina HiSeq 2000 platform to generate 90 bp paired-end reads. Alignment of the processed raw reads to the maize reference genome sequence version 2 (B73 RefGen_v2) was performed with the CLC Genomics Workbench (Version 7.0.1; http://www.clcbio.com/products/clc-genomics-workbench/). Stacked reads that shared the same starting point, sequence, and orientation were merged into one read. Maximum gaps of 50 kb were allowed to span introns. A read was considered only if at least 75% of its sequence fitted with 90% similarity to the reference sequence. Mapped reads were further projected to the filtered gene set (FGS v2; Release 5b, ftp://ftp.gramene.org/pub/gramene/maizesequence.org/release-5b/filtered-set/). Only reads that mapped uniquely to the filtered gene set with 80% of their length displaying 90% similarity were considered in subsequent analyses. RNA-seq data have been deposited in the NCBI sequencing read archive (SRA; http://www.ncbi.nlm.nih.gov/sra; Accession: SRP052697).

Statistical analyses to determine active and differentially expressed genes

The activity status of genes (on/off) was determined by fitting a hierarchical model to read count data. Read counts were modeled as draws from negative binomial distributions where the variance is a quadratic function of the mean. The negative binomial distributions were parameterized with a mean μ, dispersion parameter φ, and variance $\mu+\varphi\mu 2$. To account for the experimental design, randomly distributed lane effects were included in the model. Conditional on random effects, each gene was considered as a draw from a negative binomial distribution with a gene-specific dispersion parameter and a mean determined by the sum of a normalization factor and

a linear combination of gene-specific fixed and random effects. The normalization factor for each sample and gene was taken as the fitted value of a sample-specific smoother using the log of the raw read count plus one as the response, and GC content and log of gene length as covariates. The log of the dispersion parameter and each fixed effect were considered as independent draws from separate normal distributions, and the precisions for the random effects were modeled as draws from gamma distributions. The parameters of each normal distribution and the distribution corresponding to the lane random effects were estimated through an empirical Bayes procedure. Activity or inactivity of a gene was determined by computing the posterior probability that the expected number of reads is less than a given threshold. A gene was declared as 'inactive' in a specific root type if the resulting posterior probability was >0.5. The R-package ShrinkBayes, which is based on the ideas of Van de Wiel *et al.* (2012), was used to obtain empirical Bayes estimates of hyperparameters and compute the desired posterior probabilities. A key idea of ShrinkBayes is to use Integrated Nested Laplace Approximations (INLAs) to obtain very fast, but accurate, approximations of the marginal posteriors.

For differential expression analysis, gene expression was normalized as FPKM (fragments per kilobase of exon model per million mapped reads) values. The R-package limma based on linear models (Ritchie *et al.*, 2015) was used for determination of differentially expressed genes in pairwise comparisons. Euclidean algorithm based K-means clustering [false discovery rate (FDR) <5%] was performed, to generate 12 differential expression patterns among the three root types.

PCA and cluster analysis

A PCA (principal component analysis) was performed by using the prcomp function in R with default settings. Hierarchical clustering of samples was performed based on Pearson correlations in the CLC Genomics Workbench. Log_2-transformed FPKM values were used for both PCA and hierarchical clustering.

Heat maps of expression levels of differentially expressed TFs and lignin biosynthesis genes were based on log_2-transformed FPKM values and were generated via combined tools of Gene Cluster 3.0 based on Euclidean distances (de Hoon *et al.*, 2004) and the Java Tree View (Saldanha, 2004).

Functional annotation of differently expressed genes

Gene Ontology (GO) analyses were conducted using singular enrichment analysis (SEA) with an online agriGO platform (http://bioinfo.cau.edu.cn/agriGO/analysis.php; Du *et al.*, 2010). SEA compares input list of genes (the number of root type exclusively expressed genes or differentially expressed genes in this study) with reference genes (24 687 expressed genes) to identify enriched GO terms. Resulting *P*-values were further adjusted by introducing FDR <5% (Benjamini and Yekutieli, 2001).

Significantly over- and under-represented categories were based on the MapMan annotation (http://mapman.gabipd.org/web/guest/mapman; Thimm *et al.*, 2004). Declaration of a significantly changed category was based on comparison of detected genes with expected number in the category (χ^2 tests with Yate's continuity correction). The expected number of genes per category was calculated based on the distribution of all expressed genes in the MapMan categories.

Anatomy and histochemistry

Fresh hand cross-sections of 30, 60, and 90 mm long roots were prepared for each root type in 10 mm increments for anatomical analyses. Segments of samples were fixed in 4% paraformaldehyde in 1× phosphate buffer (containing 10 mM NaH_2PO_4 and 10 mM Na_2HPO_4, pH 7) and then sectioned by hand. Transverse sections were stained with 0.1% toluidine blue for 30 s, washed, and mounted in 50% xylene, and then photographed under a bright field microscope (PixCell IIe System, Zeiss, Jena, Germany). ImageJ2 software (version 2.0.0-rc-28; http://fiji.sc/Downloads) was used for the quantification of anatomical traits. Cell wall lignification was visualized by berberine–aniline blue staining according to Brundrett *et al.* (1988). After staining, fluorescence was observed under a PixCell IIe microscope (Zeiss) using the excitation filter G365, the chromatic splitter FT395, and the barrier filter LP 420.

Results

Morphological and anatomical characterization of maize primary, seminal, and crown roots early in development

In maize, different root types are formed consecutively in early development. In the inbred line B73, the primary root emerges 2–3 dag, seminal roots become visible 5–6 dag, and shoot-borne crown roots are formed between 10 and 11 dag (Fig. 1A). The number of seminal roots is variable between different individuals of a genotype. In the inbred line B73, the number of seminal roots varies between two and four (Supplementary Table S1 available at *JXB* online). Moreover, the number of seminal roots varies widely between different genotypes. In a panel of 30 different maize inbred lines, the inbred line UH005 did not form any seminal roots while the inbred line CML322 formed on average 5.4 seminal roots (Supplementary Table S1). Individuals of the inbred line Ki11 formed up to 10 seminal roots (Supplementary Table S1).

To study the anatomical structure of B73 primary, seminal, and crown roots, serial transverse sections were compared in a developmental time course experiment in roots of 30, 60, and 90 mm length in 10 mm increments (Supplementary Fig. S1 at *JXB* online). Within each of the three root types, no difference in their anatomical organization was observed along the longitudinal axes and between the developmental stages (Supplementary Fig. S1). Subsequently, transverse sections of the proximal parts of B73 primary, seminal, and crown roots of 20–30 mm length were subjected to a detailed comparative anatomical analysis (Fig. 1B). These microscopic analyses revealed that seminal roots of the inbred line B73 have a smaller diameter and a different organization of the vasculature compared with primary and crown roots. Quantification of anatomical parameters (Fig. 1C) in transverse orientation demonstrated that embryonic seminal roots displayed a significantly smaller total area, a smaller total stele area, and a reduced number of cortical cell files compared with primary and crown roots (Fig. 1C). Crown roots displayed the highest number of meta- and protoxylem elements, followed by primary and seminal roots (Fig. 1C). Finally, the proportion of stele area relative to the total root area was not different in primary and seminal roots, while it was slightly but significantly higher in crown roots. In contrast, the proportion of xylem vessels relative to the total stele area was significantly higher in seminal roots than in primary and crown roots (Fig. 1C).

RNA-seq of maize primary, seminal, and crown root transcriptomes

To survey how their anatomical differences are reflected in global gene expression profiles, RNA-seq of B73 primary,

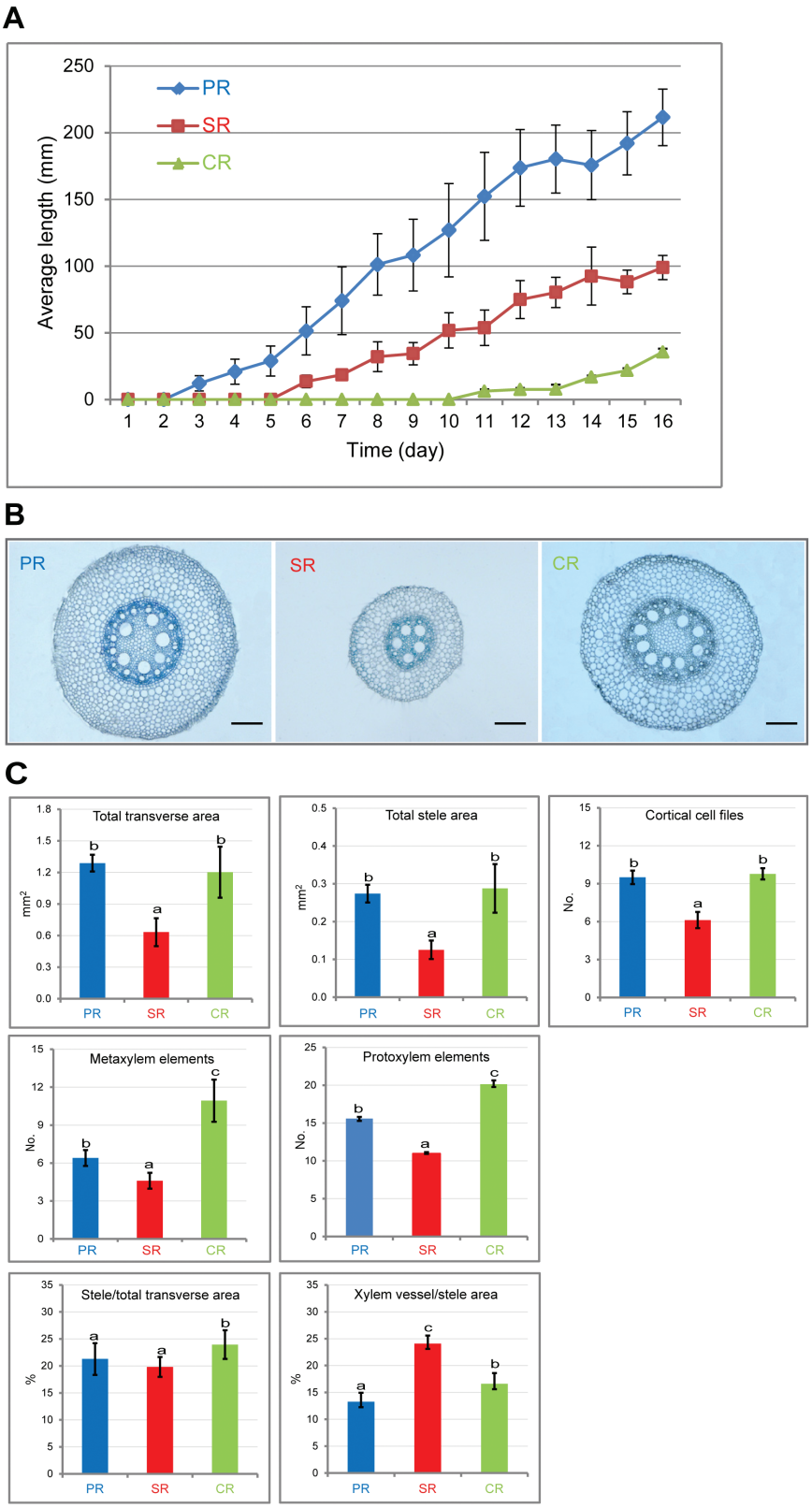


Fig. 1. Morphological and anatomical analyses of maize primary, seminal, and crown roots. (A) Growth of the three root types within 16 d after germination (n=30 per time point; error bars indicate ±SD). (B) Transverse sections of the proximal parts of 20 mm long roots of each type stained with toluidine blue. Scale bars=200 µm. (C) Quantification of root anatomical traits in transverse sections of the three root types (P<0.05, n=15; error bars indicate ±SD). PR, primary roots; SR, seminal roots; CR, crown roots.

seminal, and crown roots of 20–30 mm length was performed in four biological replicates per root type. The RNA-seq experiment resulted in 22–24 million high quality paired-end 90 bp reads per biological replicate (Supplementary Table S2 at *JXB* online). Among all reads, 82–87% mapped uniquely to the maize reference genome sequence (ZmB73_RefGen_v2;

Supplementary Table S2). After removal of redundant reads sharing the same start and end co-ordinate, sequencing direction, and sequence ('stacked reads'), between 81% and 87% of the remaining reads mapped uniquely to the 'filtered gene set', a set of 39 656 gene models predicted with high confidence (ZmB73 FGS_5B_FGSv2; Supplementary Table S2).

To assess the reproducibility of the data set, relationships between the 12 samples were determined by a PCA (Fig. 2A) and hierarchical clustering (Fig. 2B). Both analyses showed high correlation among biological replicates of each root type. Furthermore, the transcriptomes of primary and crown roots clustered in close proximity, while a distinct cluster comprising the replicate samples of seminal roots was observed (Fig. 2A, B).

Determination of the activity status of genes and their root type-specific expression

A hierarchical model with a negative binomial response (see the Materials and methods) was applied to determine the activity status (on/off) of genes in the three root types. In total, 24 687 of 39 656 (62%) high confidence gene models (ZmB73 FGS_5B_FGSv2) were expressed in at least one root type (Fig. 2C; Supplementary Table S3 at *JXB* online). Among those, 23 156 genes were expressed in primary roots, 23 736 in seminal roots, and 23 501 in crown roots (Fig. 2C). In total, 96% (22 206/23 156) of genes expressed in primary roots, 94% (22 206/23 736) of genes active in seminal roots, and 95% (22 206/23 501) of genes transcribed in crown roots were also active in the other two root types. In contrast, 1187 genes were exclusively expressed in only one of the three root types, comprising 583 seminal root-, 309 crown root-, and 295 primary root-specific genes (Fig. 2C). A GO analysis was conducted to identify enriched biological functions among these root type-specific genes. Enriched GO terms were calculated by SEA with agriGO (see the Materials and methods). Among seminal root-specific genes, only the GO term 'transcription factor activity' (GO: 0003 700) was enriched. Among crown root-specific genes, only the GO term 'photosynthesis' (GO: 0015 979) was over-represented. Among the primary root-specific genes, no GO term was enriched.

Genes differentially expressed between distinct root types

Pairwise contrasts between the three root types were determined to identify differentially expressed genes. The distribution of fold change (Fc) versus mean relative expression revealed that differential expression was observed for genes with both low and high expression (Supplementary Fig. S2A at *JXB* online). In pairwise comparisons of the three root types, more differentially expressed genes (FDR <5%, $|\log_2 Fc| \geq 1$) were observed in comparisons involving seminal roots compared with contrasts determined from primary and crown roots (Fig. 2D; Supplementary Fig. S2B). In comparisons of seminal versus primary and crown roots, substantially more differentially expressed genes were preferentially expressed in seminal roots. SEA with agriGO (see the Materials and methods) revealed that molecular functions such as antioxidant activity, catalytic activity, and electron carrier activity

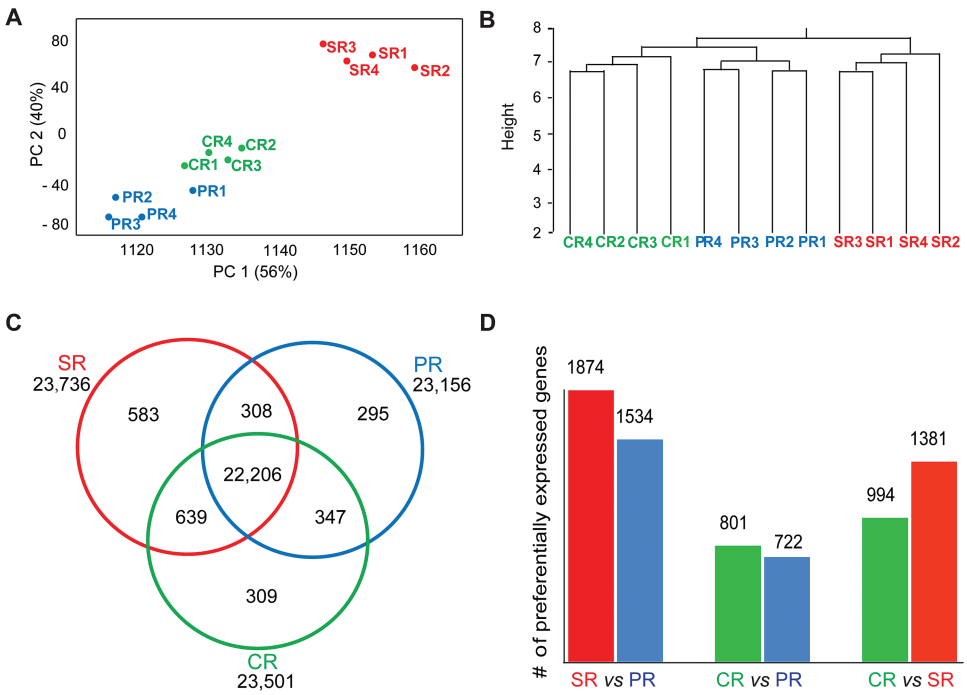


Fig. 2. Relationship of root transcriptome samples, root type-specific gene expression, and differential gene expression. (A) Principal component analysis (PCA) of the 12 RNA-seq samples of the three root types. The first and second principal components collectively explain 96% of variance. (B) Hierarchical clustering of the 12 RNA-seq samples of the three root types based on Euclidian distance. The *y*-axis indicates the degree of the variance. (C) Number of genes with conserved and root type-specific expression in the three root types. (D) Pairwise comparisons of differentially expressed (FDR <5%, $|\log_2 Fc| \geq 1$) genes among the three root types. PR, primary roots; SR, seminal roots; CR, crown roots; PC 1 and PC2, principal component one and two.

were enriched among the differentially expressed genes in all three pairwise comparisons (Supplementary Fig. S3A–C). In addition, regulation of biological process, biological regulation, and transcriptional regulator activity were enriched in the comparisons of primary versus seminal roots and crown versus seminal roots (Supplementary Fig. S3A, C). Photosynthesis was enriched in the comparisons of primary versus crown roots and crown versus seminal roots (Supplementary Fig. S3B, C).

Exploration of global gene expression patterns in the three root types

To explore differential expression across the three root types, K-means clustering (see the Materials ad methods) identified 12 clusters representing all possible patterns of differential expression (Fig. 3A; Supplementary Table S4 at *JXB* online). In total, 12 207 of 24 687 (49%) genes were classified into the 12 clusters, while the remaining genes (51%) were constitutively expressed. The 12 clusters were further classified into six patterns according to peak expression in one or two root types (Fig. 3B). Patterns K1–K3 displayed peak expression in primary roots, while patterns K4–K6 peaked in seminal roots and K7–K9 in crown roots (see Fig. 3A). In contrast, patterns K10, K11, and K12 displayed peak expression in two of three root types (see Fig. 3A). According to this classification, genes of pattern K4–K6 (3869/12 207, 32%) that displayed peak expression in seminal roots comprised the highest number of genes of the six patterns.

Over- and under-representation of MapMan functional categories was determined in the six patterns of differentially expressed genes compared with all expressed genes (Fig. 3B; see the Materials and methods). The analysis revealed that enriched or under-represented categories among genes preferentially expressed in primary (K1+K2+K3) or crown roots (K7+K8+K9) were relatively similar, whereas genes preferentially expressed in seminal roots (K4+K5+K6) displayed distinct enrichment patterns. For instance, functional categories such as cell organization, cell wall biosynthesis, secondary metabolism, signaling pathways, and transport processes were over-represented among genes preferentially expressed in primary and crown roots, while they were under-represented or not enriched among genes peaking in seminal roots. Genes involved in auxin metabolism such as *aux/iaa33* and *arf34* were enriched in primary roots. In contrast, genes that are required for minor CHO metabolism (such as sugar and sugar derivate metabolism), metal handling, abscisic acid (ABA) metabolism (including the key ABA biosynthesis genes *vp14* and *zep2* and numerous ABA-induced genes), and RNA metabolism were highly over-represented among genes with peak expression in seminal roots. These categories were under-represented or not significantly changed among genes preferentially expressed in primary and crown roots. Moreover, stress-related genes were highly over-represented

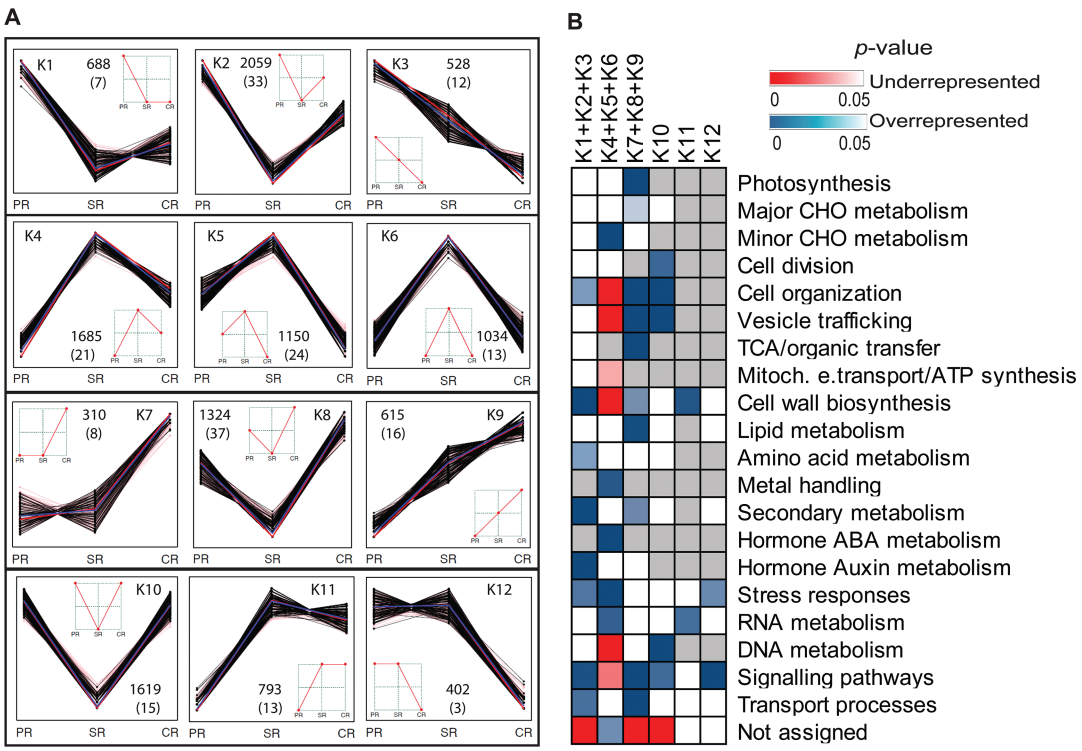


Fig. 3. Global expression profiling of genes among the three root types. (A) K-means clustering showing all 12 possible expression patterns of 12 207 differentially expressed genes (202 classical maize genes in parentheses). Each black line represents a gene; the red lines represent median values; the blue lines represent the mean value. (B) Over- or under-representation of MapMan functional categories among six patterns representing peak expression in one or two root types. Clusters K1–K3 (primary roots), K4–K6 (seminal roots), and K7– K9 (crown roots) were combined in this analysis because they displayed peak expression in one of the three root types (χ^2 tests, $P<0.05$; red, significantly under-represented; blue, significantly over-represented; white, not significant; gray, not detected). PR, primary roots; SR, seminal roots; CR, crown roots.

in the group of genes peaking in seminal roots, and moderately enriched among genes peaking in primary roots. Genes required for cell wall biosynthesis and RNA metabolism were enriched in cluster K11, the group of genes preferentially expressed in both crown and seminal roots. Genes preferentially expressed in both primary and seminal roots (K12) were enriched in the functional categories stress and signaling.

Preferential expression of classical maize genes in different root types

In maize, a set of 468 hand-curated 'classical genes' with experimentally confirmed function is available (Schnable and Freeling, 2011). In total, 348 of these classical genes were detected in this study. Among those, 202 were differentially expressed among the three root types and assigned to the 12 expression clusters defined in Fig. 3A. Novel root type-specific functions were revealed for classical maize genes displaying peak expression in one of the three root types (Supplementary Table S5 at *JXB* online) including 52 in primary roots (K1+K2+K3), 58 in seminal roots (K4+K5+K6), and 61 in crown roots (K7+K8+K9). For instance, among the group of 52 classical genes with peak expression in the primary root, seven of eight cellulose synthesis genes of the classical gene set were observed. Moreover, six of nine tubulin synthesis genes of the classical gene set were detected among the 61 genes displaying peak expression in crown roots. Finally, among 58 classic genes peaking in seminal roots, seven of 10 genes encoding aquaporin proteins and all four genes of the classical gene set translating into heat shock proteins were present. Furthermore, genes encoding stress-related proteins such as a WOUND INDUCED PROTEIN 1 (WIP1), an NaCl STRESS PROTEIN (NCA1), and a MAIZE INSECT RESISTANCE PROTEIN (MIR1) were highly expressed in seminal roots. The 31 classical genes assigned to clusters K10–K12 displayed diverse functions (Supplementary Table S5).

Expression dynamics of transcriptional regulators in the three root types

Expression dynamics of genes involved in transcriptional regulation were surveyed to investigate their role in the developmental control of specific root types. In total, 2442 genes related to transcriptional regulation were identified based on their MapMan annotation. Among those, 1252 genes were differentially expressed between the three root types (Supplementary Table S6 at *JXB* online). Hierarchical clustering classified these genes into two clusters (Fig. 4A). Cluster 1, containing 684 (55%) of the differentially expressed genes, displayed high expression in seminal roots. The remaining 568 (45%) genes that were grouped into cluster 2 showed high expression in primary and crown roots (Fig. 4A). A comparison of these 1252 genes with the maize TF database PlantTFDB v3.0 (http://planttfdb.cbi.pku.edu.cn/; Jin *et al.*, 2013) identified 718 differentially expressed genes in the two clusters which encode TFs (457 in cluster 1 and 261 in cluster 2). Furthermore, family-specific expression trends of each cluster were determined by Fisher's exact test (Fig. 4B). Six gene families encoding TFs [BASIC/HELIX–LOOP–HELIX (bHLH), DNA BINDING WITH ONE FINGER (DOF), ETHYLENE RESPONSE FACTOR (ERF), LATERAL ORGAN BOUNDARY DOMAIN (LBD), GOLDEN2 (G2)-like, and WRKY] were over-represented among genes with peak expression in seminal roots (cluster 1; Fig. 4B). In contrast, genes encoding the AUXIN RESPONSE FACTORs (ARFs) and the HOMEODOMAIN-LEUCINE ZIPPERs (HD-ZIPs) were over-represented in primary and crown roots (cluster 2; Fig. 4B). These results suggest distinct roles for transcriptional regulators in seminal root development compared with primary and crown roots.

Root type diversification of cell wall lignification

Cell proliferation and enlargement accompanied by active cell wall reorganization are the main processes that drive root growth during early development. Lignin is an important component of plant cell walls. To explore the developmental status of cell walls in the three root types, we focused on the expression of genes encoding members of the eight major families involved in lignin biosynthesis (Fig. 5A). Based on MapMan annotations, 53 genes involved in lignin biosynthesis were identified in this study, and 31 displayed differential expression across the three root types (Fig. 5B; Supplementary Table S7 at *JXB* online). These 31 genes represented seven of eight major gene families involved in lignin biosynthesis (Fig. 5A, B). Hierarchical clustering classified genes active in primary roots into a distinct group including 12 genes with higher expression in this root type (Fig. 5B). This suggested that a different set of lignin biosynthesis genes was active in primary roots compared with seminal and crown roots. Moreover, expression diversification among homologs of genes encoding enzymes such as phenylalanine ammonia lyase (PAL), 4-coumaroyl-CoA synthase (4CL), cinnamyl-alcohol dehydrogenase (CAD), and caffeoyl CoA *O*-methyltransferase (CCoAOMT) was observed in the three root types (Fig. 5B). For example, two of four genes encoding PAL which control the first step of lignin biosynthesis showed significantly higher expression in primary roots, while the other two genes were preferentially expressed in seminal and crown roots (Supplementary Table S7). Only one gene (GRMZM2G068917) encoding cinnamoyl-CoA reductase (CCR) was identified in this study which was constitutively expressed in all three root types.

Subsequently, cell wall lignification was visualized in transverse sections of the three root types through berberine–aniline blue staining (see the Materials and methods). A strong staining intensity of the endodermal Casparian strip and lignified cell walls of xylem vessels was detected in the primary root, while only a very weak staining of the Casparian strip was observed in crown roots. In seminal root sections, the Casparian strip was not detected by staining. Staining intensities of cell walls of xylem vessels in crown and seminal roots were similar but substantially less than in primary roots (Fig. 5C).

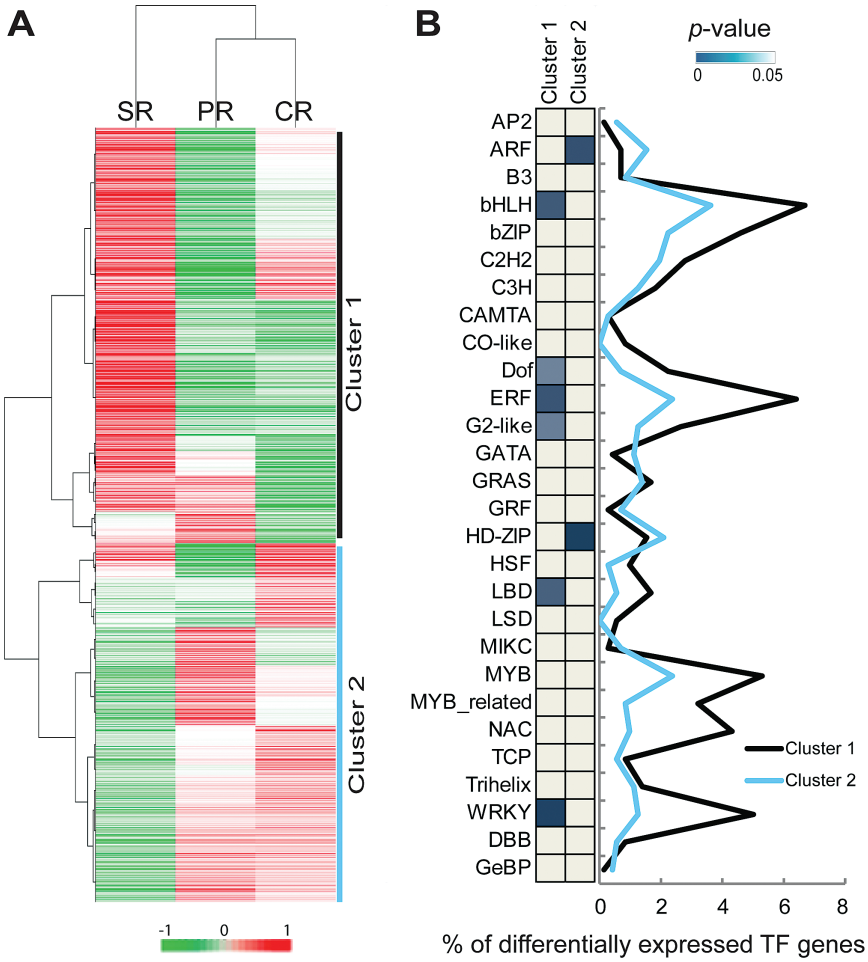


Fig. 4. Dynamics of the expression profiles of transcriptional regulators. (A) Hierarchical clustering analysis of 1252 differentially expressed genes involved in transcriptional regulation of the three root types based on average expression values. (B) Classification of genes in 28 major transcription factor families of the two clusters from (A) as a proportion (in %) of all differentially expressed transcription factor genes. Significantly enriched families in each cluster were determined by Fisher's exact tests (P<0.05) and indicated by blue blocks. Gray, not significant. PR, primary roots; SR, seminal roots; CR, crown roots.

Discussion

Morphology: grass species-specific development and variability of seminal roots

The maize root stock is composed of embryonic primary and seminal roots and of post-embryonic shoot-borne roots (reviewed in Feldman, 1994; Hochholdinger *et al.*, 2004a, b). Primary and shoot-borne roots are formed in all cereal species, while seminal roots are present in maize (Robertson *et al.*, 1979), wheat (Erdelska and Vidovencova, 1993), and barley (Luxová, 1986), but not in rice (Morita and Abe, 1994), sorghum (Singh *et al.*, 2010), and the grass model species brachypodium (Watt *et al.*, 2009). Remarkably, the formation of seminal roots is not related to the phylogenetic relationship of these species. For instance, despite their close phylogenetic relationship (Schnable and Freeling, 2011), maize does form seminal roots while sorghum does not. Seminal roots might have therefore evolved independently in these species as functional adaptation to changing environmental challenges.

The number of seminal roots in modern maize is highly variable within a genotype and between different genotypes. While the emergence of seminal roots in the course of grass evolution remains obscure, an increase of seminal root number has been observed during the domestication of maize, barley, and wheat. Modern maize (*Zea mays* ssp. *mays*) was domesticated ~9000 years ago from its ancestor teosinte (*Z. mays* ssp. *Parviglumis*; Matsuoka *et al.*, 2002; Piperno *et al.*, 2009) and has since then been subjected to intensive improvement from maize landraces (Hufford *et al.*, 2012). A study by Burton *et al.* (2013) revealed that the majority of teosinte accessions (62%) did not form seminal roots, while the remaining accessions only formed a maximum of three seminal roots. In contrast, maize landraces formed on average 3.9 seminal roots, ranging between one and 11 per plant (Burton *et al.*, 2013). This is in the range of the 30 modern maize inbred lines that were surveyed in this study. They formed on average 3.7 seminal roots and between 0 and 10 per plant (Supplementary Table S1 at *JXB* online). Consistently, barley landraces and their modern cultivars formed significantly more seminal roots than their ancestors (Grando and Ceccarelli, 1995). Similarly, increased numbers of seminal roots were observed in domesticated wheat genotypes (Robertson *et al.*, 1979). It was suggested that root traits were probably inadvertently selected for during domestication because of their crucial role in anchorage and soil resource uptake (Lynch and Brown,

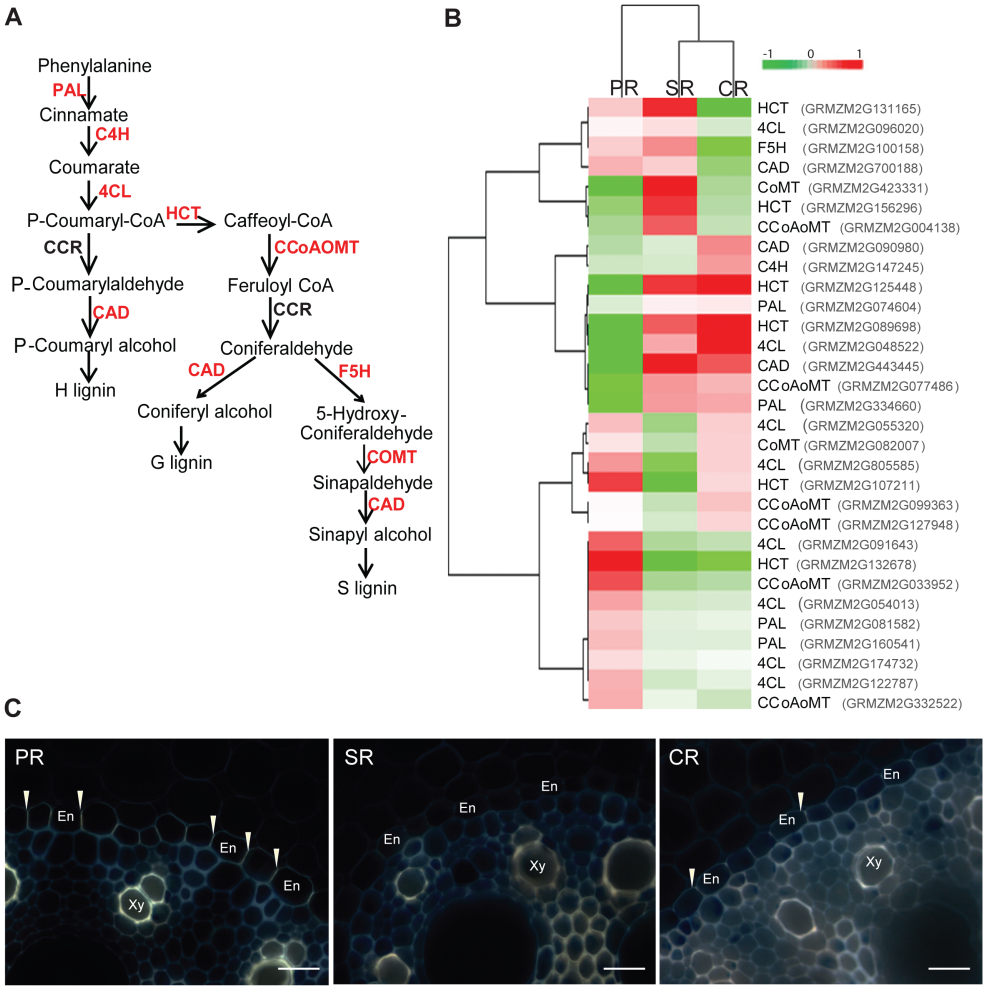


Fig. 5. Dynamics of lignin biosynthesis among the three root types. (A) The lignin biosynthesis pathway based on MapMan. PHENYLALANINE AMMONIA LYASES (PAL), CINNAMATE-4-HYDROXYLASE (C4H), 4-COUMAROYL-COA SYNTHASE (4CL), HYDROXYCINNAMOYL TRANSFERASE (HCT), CAFFEOYL COA *O*-METHYLTRANSFERASE (CCoAOMT), CINNAMOYL-COA REDUCTASE (CCR), FERULATE 5-HYDROXYLASE (F5H), CAFFEIC ACID/5-HYDROXYFERULIC ACID 3/5-*O*-METHYLTRANSFERASE (COMT), CINNAMYL-ALCOHOL DEHYDROGENASE (CAD). Differentially expressed genes are highlighted in red. (B) Hierarchical clustering analysis of 31 differentially expressed genes predicted to be involved in lignin biosynthesis. Differential expression of genes is indicated as $\log_2$ of the FPKM value. (C) Cell wall lignification of transverse sections in the proximal parts of the three root types of 20 mm length detected by berberine–aniline blue staining. Scale bars=25 μm; Xy, xylem; En, endodermis; PR, primary roots; SR, seminal roots; C, crown roots.

2006). An increased number of seminal roots is probably a trait that underwent selection. The notion that maize plants would benefit from an increased number of seminal roots is supported by the observation that the number of seminal roots in maize correlates with drought tolerance at the seedling stage (Qayyum *et al.*, 2012). Similarly, it was demonstrated that the number and length of seminal roots were positively yet weakly correlated with shoot biomass in the field under low phosphorus (Zhu *et al.*, 2006). Taken together, these results imply that inadvertent selection of genotypes with increased seminal root number during maize domestication probably has contributed to the agronomic success of modern maize.

Anatomy: cost-efficient anatomical structure of seminal roots in the maize inbred line B73

The early establishment of the maize seedling root system depends on the availability of carbohydrates stored in the starchy endosperm of the seed (Lopes and Larkins, 1993). These carbohydrates provide the primary energy for the emergence of embryonic primary roots from the basal pole of the seed and seminal roots from the scutellar node in the first days after germination (Hochholdinger *et al.*, 2004a). Nutrient acquisition depends more on the root surface than on root volume. Formation of a single primary root and several seminal roots might be an efficient way of allocating seed-stored carbohydrates for maize seedling root system establishment and thus enable an efficient acquisition of soil resources and early anchorage of the seedling.

Quantification of anatomical traits in the inbred line B73 revealed thinner seminal roots with significantly fewer cortical cell files relative to primary and crown roots (Fig. 1C). A reduced number of cortical cell files was associated with increased drought tolerance in maize (Chimungu *et al.*, 2014). Moreover, the ratio of meta-xylem vessels to the whole stele area was significantly increased in B73 seminal roots compared with the primary root, while the ratio of stele to transverse area showed no difference (Fig. 1C). The ratio of the

stele to transverse area probably influences the root system cost and function (Burton *et al.*, 2013). Meta-xylem vessels in the stele are the largest vascular elements and mainly responsible for water and nutrient transport from the root to the shoot (Hochholdinger, 2009). A larger stele is likely to accommodate more xylem vessels for increased water and nutrient transport, whereas initial carbon investment for establishing a larger stele is greater on a per root basis (Burton *et al.*, 2013). The observation that seminal roots of the modern maize inbred line B73 displayed a greater ratio of meta-xylem vessels to the whole stele area but a similar ratio of stele to transverse area compared with primary root suggests they might be better equipped for efficient water and nutrient transport without additional costs for stele establishment.

Comparative transcriptome analyses highlight the uniqueness of seminal roots

RNA-seq of different root types of the maize reference inbred line B73 revealed a distinct transcriptomic landscape of seminal versus primary and crown roots. First, hierarchical clustering and PCA of all expressed genes revealed that seminal roots were only distantly related to primary and crown roots which clustered closely together (Fig. 2A, B). This correlates with the unique anatomy of seminal roots of the inbred line B73 that is distinct from the more similar primary and crown roots (Fig. 1B, C). Secondly, substantially more genes were exclusively expressed in seminal roots than in the other root types (Fig. 2C). Moreover, more genes were differentially expressed in seminal roots and primary and crown roots than between primary and crown roots (Fig. 2D; Supplementary S2B at *JXB* online). Similarly, observations that anatomical or physiological differences correlate with transcription profiles have been made in maize (Sekhon *et al.*, 2011; Downs *et al.*, 2013), rice (Wang *et al.*, 2010), Arabidopsis (Ma *et al.*, 2005), and tobacco (Edwards *et al.*, 2010). This suggests that tissue or organ identity is a primary factor that explains transcriptome variation (Downs *et al.*, 2013; Raherison *et al.*, 2015), while changing environmental conditions affect a much smaller set of genes, as demonstrated for salt stress in maize (Zhang *et al.*, 2015) and a plethora of stresses in Arabidopsis (Aceituno *et al.*, 2008). Functional annotation of differentially expressed genes further revealed root type-specific developmental functions (Fig. 3B). Remarkably, seminal roots displayed distinct enriched functional categories compared with primary and crown roots which were relatively similar (Fig. 3B), suggesting functional divergence of seminal roots. Some of these differences will be discussed in the following sections.

Distinct roles of transcriptional regulators in seminal roots

Hierarchical clustering of transcriptional regulators suggested specific transcriptional regulation in seminal roots compared with similar regulation in primary and crown roots (Fig. 4A). For instance, members of the ARF and HD-ZIP families were enriched in primary and crown roots (Fig. 4B). Both families are involved in growth, development, and cell division, and respond to environmental stimuli (Van Ha *et al.*, 2013; Chen *et al.*, 2014). In contrast, members of the LBD, G2-like, bHLH, ERF, DOF, and WRKY families were over-represented in seminal roots (Fig. 4B). Among those, the LBD protein RTCS of maize is instrumental in seminal root initiation (Taramino *et al.*, 2007). Moreover, G2-like, bHLH, ERF, DOF, and WRKY TFs play different roles in root growth and development including xylem or phloem cell differentiation, cell elongation, root hair development, stress responses, ethylene responses, and ABA signaling (Zhao *et al.*, 2005; Nakano *et al.*, 2006; Feller *et al.*, 2011; Rushton *et al.*, 2012; Guo and Qiu, 2013). Coincidently, genes involved in ABA metabolism including many ABA-induced genes were significantly enriched in seminal roots (Fig. 3B). ABA is produced under osmotic stress conditions such as drought and salinity, and plays an important role in stress response and tolerance of plants (Jakab *et al.*, 2005). Specific enrichment of members of these TF families related to stress response in seminal roots suggests that stress activates different signaling pathways in seminal roots of the inbred line B73.

Diversification of cell wall lignification among the root types

The secondary metabolite lignin is an important component in secondary cell walls of developing maize roots (Zeier *et al.*, 1999). Lignin is deposited in the secondary cell walls of endodermis cells and xylem vessels to strengthen the root and enhance plant anchorage (Vermerris *et al.*, 2010). Compared with Arabidopsis, the number of lignin biosynthesis genes has substantially increased in maize (Penning *et al.*, 2009). Dynamic expression patterns of lignin biosynthesis genes across the three root types underpin the diversity and complexity of transcriptional regulation in maize root development (Fig. 5A, B). Remarkably, while seminal roots displayed distinct features for most anatomical and transcriptomic features from primary and crown roots, hierarchical clustering revealed similar expression patterns for lignin biosynthesis genes in seminal and crown roots which were distinct from expression profiles in primary roots. Organ-specific expression patterns of lignin genes underscoring the diversification of the lignin pathway in maize have also been observed for other vegetative organs (Sekhon *et al.*, 2011). Histochemical staining further supported the differences observed in the transcriptomic profiles by revealing differences in cell wall lignification of the endodermis and xylem vessels between the three root types (Fig. 5C). Increased cell wall lignification coupled with diversification of the lignin pathway in primary roots might indicate relatively active formation and regulation of secondary cell walls in primary root development (Fig. 5B, C). Similarly, the *brown midrib2* (*bm2*) mutant of maize which is defective in cell wall biosynthesis displayed tissue-specific patterns of lignification (Vermerris and Boon, 2001). Excessive lignin deposition was observed in the mutant *rum1* of maize which is defective in vasculature formation (Zhang *et al.*, 2014).

Root type-specific expression of genes of the auxin signal transduction pathway

Genes involved in auxin signal transduction are important for various aspects of maize root development. The *Aux/IAA* gene *rum1* (Woll *et al.*, 2005) controls seminal and lateral root initiation in primary roots and depends on the activity of ZmARF34 (von Behrens *et al.*, 2011). This is in line with preferential expression of *ZmARF34* in primary roots in the present study. Interaction of RUM1 with ZmIAA33 has been revealed by Ludwig *et al.* (2014). The *ZmIAA33* gene that emerged after the whole-genome duplication of maize (Ludwig *et al.*, 2014) displayed peak expression in seminal roots in this study. Consistently, *ZmIAA33* displayed substantially higher expression in seminal roots relative to primary and crown roots (Ludwig *et al.*, 2014). Sorghum (Singh *et al.*, 2010) and rice (Morita and Abe, 1994), which both do not form seminal roots, do not have homologs of *ZmIAA33* (Ludwig *et al.*, 2014). These findings support functions of auxin-related genes with root type-specific expression profiles in maize root development.

In summary, seminal roots of the maize inbred line B73 display a distinct anatomical structure and transcriptomic landscape compared with primary and crown roots related to the different functions of these root types during development. The formation of seminal roots in modern maize might have allowed maize to explore new habitats which were inaccessible for maize progenitors without seminal roots.

Supplementary Data

Supplementary data are available at *JXB* online.

Figure S1. Serial transverse sections of 30, 60, and 90 mm long primary (PR), seminal (SR), and crown roots (CR) of maize in 10 mm increments.

Figure S2. MA and Volcano plots displaying the correlation of gene expression changes versus mean expression in the pairwise comparisons.

Figure S3. Singular enrichment analyses (SEA) with AgriGO revealed significantly enriched GO terms for differentially expressed genes in the three pairwise comparisons.

Table S1. Summary of seminal root counts of 30 modern maize inbred lines.

Table S2. RNA-seq output and mapping results.

Table S3. List of 24 687 genes expressed in at least one of the three root types in four biological replicates.

Table S4. Classification of the 12 207 differentially expressed genes among the three root types into 12 K-means clusters.

Table S5. List of 202 classical maize genes differentially expressed in the three root types.

Table S6. List of 1252 differentially expressed genes involved in transcriptional regulation.

Table S7. List of 31 differentially expressed genes involved in lignin biosynthesis.

Acknowledgements

HT was supported by a China Scholarship Council fellowship (number 201206350004). Root research in FH's laboratory is supported by a Deutsche Forschungsgemeinschaft (DFG) grant (2249/9-3) and the University of Bonn.

References

Abbe E, Stein O. 1954. The origin of the shoot apex in maize: embryogeny. American Journal of Botany **41,** 285–293.

Aceituno FF, Moseyko N, Rhee SY, Gutiérrez RA. 2008. The rules of gene expression in plants: organ identity and gene body methylation are key factors for regulation of gene expression in *Arabidopsis thaliana*. BMC Genomics **9,** 438.

Benjamini Y, Yekutieli D. 2001. The control of the false discovery rate in multiple testing under dependency. Annals of Statistics **29,** 1165–1188.

Brundrett MC, Enstone DE, Peterson CA. 1988. A berberine–aniline blue fluorescent staining procedure for suberin, lignin, and callose in plant tissue. Protoplasma **146,** 133–142.

Burton AL, Brown KM, Lynch JP. 2013. Phenotypic diversity of root anatomical and architectural traits in *Zea* species. Crop Science **53,** 1042–1055.

Chen X, Chen Z, Zhao H, Zhao Y, Cheng B, Xiang Y. 2014. Genome-wide analysis of soybean HD-Zip gene family and expression profiling under salinity and drought treatments. PLoS One **9,** e0087156.

Chimungu JG, Brown KM, Lynch JP. 2014. Reduced root cortical cell file number improves drought tolerance in maize. Plant Physiology **166,** 1943–1955.

de Hoon MJ, Imoto S, Nolan J, Miyano S. 2004. Open source clustering software. Bioinformatics **20,** 1453–1454.

Dembinsky D, Woll K, Saleem M, *et al*. 2007. Transcriptomic and proteomic analyses of pericycle cells of the maize (*Zea mays* L.) primary root. Plant Physiology **145,** 575–588.

Downs GS, Bi Y-M, Colasanti J, Wu W, Chen X, Zhu T, Rothstein SJ, Lukens L. 2013. A developmental transcriptional network for *Zea mays* defines coexpression modules. Plant Physiology **161,** 1830–1843.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z. 2010. agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Research **38,** 64–70.

Edwards KD, Bombarely A, Story GW, Allen F, Mueller LA, Coates SA, Jones L. 2010. TobEA: an atlas of tobacco gene expression from seed to senescence. BMC Genomics **11,** 142.

Erdelska O, Vidovencova Z. 1993. Development of adventitious seminal root primordia of maize during embryogenesis. Biologia Plantarum **48,** 85–88.

Fahn A. 1990. Plant anatomy, 4th edn. Oxford: Pergamon Press

Feldman L. 1994. The maize root. In: Freeling M, Walbot V, eds. The maize handbook. New York: Springer, 29–37.

Feller A, Machemer K, Braun EL, Grotewold E. 2011. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. The Plant Journal **66,** 94–116.

Giehl RF, von Wirén N. 2014. Root nutrient foraging. Plant Physiology **166,** 509–517.

Grando S, Ceccarelli S. 1995. Seminal root morphology and coleoptile length in wild (*Hordeum vulgare* ssp. *spontaneum*) and cultivated (*Hordeum vulgare* ssp. *vulgare*) barley. Euphytica **86,** 73–80.

Guo Y, Qiu LJ. 2013. Genome-wide analysis of the Dof transcription factor gene family reveals soybean specific duplicable and functional characteristics. PLoS One **8,** e76809.

Hawkes CV, De Angelis KM, Firestone MK. 2007. Root interactions with soil microbial communities and processes. In: Cardon ZG, Whitbeck JL, eds. The rhizosphere, an ecological perspective. New York: Elsevier Academic Press, 1–29.

Hetz W, Hochholdinger F, Schwall M, Feix G. 1996. Isolation and characterization of *rtcs*, a maize mutant deficient in the formation of nodal roots. The Plant Journal **10,** 845–857.

Hochholdinger F. 2009. The maize root system: morphology, anatomy, and genetics. In: Bennetzen J, Hake S eds. Handbook of maize: its biology. New York: Springer, 145–160.

Hochholdinger F, Feix G. 1998. Early post-embryonic root formation is specifically affected in the maize mutant *lrt1*. The Plant Journal **16,** 247–255.

Hochholdinger F, Park WJ, Feix GH. 2001. Cooperative action of SLR1 and SLR2 is required for lateral root-specific cell elongation in maize. Plant Physiology **125,** 1529–1539.

Hochholdinger F, Park WJ, Sauer M, Woll K. 2004*a*. From weeds to crops: genetic analysis of root development in cereals. Trends in Plant Science **9,** 42–48.

Hochholdinger F, Wen TJ, Zimmermann R, Chimot-Marolle P, da Costa e Silva O, Bruce W, Lamkey KR, Wienand U, Schnable PS. 2008. The maize (*Zea mays* L.) *roothairless3* gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield. The Plant Journal **54,** 888–898.

Hochholdinger F, Woll K, Sauer M, Dembinsky D. 2004*b*. Genetic dissection of root formation in maize (*Zea mays*) reveals root type-specific developmental programmes. Annals of Botany **93,** 359–368.

Hoppe D, McCully M, Wenzel C. 1986. The nodal roots of *Zea*: their development in relation to structural features of the stem. Canadian Journal of Botany **64,** 2524–2537.

Hufford MB, Xu X, Van Heerwaarden J, Pyhäjärvi T, Chia J-M, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM. 2012. Comparative population genomics of maize domestication and improvement. Nature Genetics **44,** 808–811.

Jakab G, Ton J, Flors V, Zimmerli L, Métraux J-P, Mauch-Mani B. 2005. Enhancing Arabidopsis salt and drought stress tolerance by chemical priming for its abscisic acid responses. Plant Physiology **139,** 267–274.

Jiang K, Zhang S, Lee S, Tsai G, Kim K, Huang H, Chilcott C, Zhu T, Feldman LJ. 2006. Transcription profile analyses identify genes and pathways central to root cap functions in maize. Plant Molecular Biology **60,** 343–363.

Jin J, Zhang H, Kong L, Gao G, Luo J. 2013. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Research **42,** 1182–1187.

Lopes MA, Larkins BA. 1993. Endosperm origin, development, and function. The Plant Cell **5,** 1383–1391.

Ludwig Y, Berendzen KW, Xu C, Piepho H-P, Hochholdinger F. 2014. Diversity of stability, localization, interaction and control of downstream gene activity in the maize Aux/IAA protein family. PLoS One **9,** e107346.

Luxová M. 1986. The seminal root primordia in barley and the participation of their non-meristematic cells in root construction. Biologia Plantarum **28**, 161–167.

Lynch J, Brown K. 2006. Whole plant adaptations to low phosphorus availability. In: Huang B, ed. Plant–environment interactions. Boca Raton, FL: CRC Press, 209–242.

Ma L, Sun N, Liu X, Jiao Y, Zhao H, Deng XW. 2005. Organ-specific expression of Arabidopsis genome during development. Plant Physiology **138,** 80–91.

Majer C, Xu C, Berendzen KW, Hochholdinger F. 2012. Molecular interactions of ROOTLESS CONCERNING CROWN AND SEMINAL ROOTS, a LOB domain protein regulating shoot-borne root initiation in maize (*Zea mays* L.). Philosophical Transactions of the Royal Society B: Biological Sciences **367,** 1542–1551.

Martin EM, Harris WM. 1976. Adventitious root development from the coleoptilar node in *Zea mays* L. American Journal of Botany **63,** 890–897.

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez J, Buckler E, Doebley J. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. Proceedings of the National Academy of Sciences, USA **99,** 6080–6084.

Morita S, Abe J. 1994. Modeling root system morphology in rice. In: Davis TD, Haissig BE, eds. Biology of adventitious root formation. New York: Springer, 191–202.

Muthreich N, Majer C, Beatty M, Paschold A, Schützenmeister A, Fu Y, Malik WA, Schnable PS, Piepho H-P, Sakai H. 2013. Comparative transcriptome profiling of maize coleoptilar nodes during shoot-borne root initiation. Plant Physiology **163,** 419–430.

Nakano T, Suzuki K, Fujimura T, Shinshi H. 2006. Genome-wide analysis of the ERF gene family in Arabidopsis and rice. Plant Physiology **140,** 411–432.

Nestler J, Liu S, Wen TJ, Paschold A, Marcon C, Tang HM, Li D, Li L, Meeley RB, Sakai H. 2014. *Roothairless 5*, which functions in maize (*Zea mays* L.) root hair initiation and elongation encodes a monocot-specific NADPH oxidase. The Plant Journal **79,** 729–740.

Paschold A, Jia Y, Marcon C, Lund S, Larson NB, Yeh CT, Ossowski S, Lanz C, Nettleton D, Schnable PS, Hochholdinger F. 2012. Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents. Genome Research **22,** 2445–2454.

Paschold A, Larson NB, Marcon C, Schnable JC, Yeh CT, Lanz C, Nettleton D, Piepho HP, Schnable PS, Hochholdinger F. 2014. Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. The Plant Cell **26,** 3939–3948.

Penning BW, Hunter CT, Tayengwa R, Eveland AL, Dugard CK, Olek AT, Vermerris W, Koch KE, McCarty DR, Davis MF. 2009. Genetic resources for maize cell wall biology. Plant Physiology **151,** 1703–1728.

Peter R, Eschholz TW, Stamp P, Liedgens M. 2009. Early growth of flint maize landraces under cool conditions. Crop Science **49,** 169–178.

Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R. 2009. Starch grain and phytolith evidence for early ninth millennium BP maize from the Central Balsas River Valley, Mexico. Proceedings of the National Academy of Sciences, USA **106,** 5019–5024.

Qayyum A, Ahmad S, Liaqat S, Malik W, Noor E, Muhammad H. 2012. Screening for drought tolerance in maize (*Zea mays* L.) hybrids at an early seedling stage. African Journal of Agricultural Research **7,** 3594–3604.

Raherison ES, Giguère I, Caron S, Lamara M, MacKay JJ. 2015. Modular organization of the white spruce (*Picea glauca*) transcriptome reveals functional organization and evolutionary signatures. New Phytologist **207,** 172–187.

Ribaut JM, Betran J, Monneveux P, Setter T. 2009. Drought tolerance in maize. In: Bennetzen J, Hake S, eds. Handbook of maize: its biology. Berlin: Springer, 311–344.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research **43,** e47.

Robertson B, Waines J, Gill B. 1979. Genetic variability for seedling root numbers in wild and domesticated wheats. Crop Science **19,** 843–847.

Rushton DL, Tripathi P, Rabara RC, Lin J, Ringler P, Boken AK, Langum TJ, Smidt L, Boomsma DD, Emme NJ. 2012. WRKY transcription factors: key components in abscisic acid signalling. Plant Biotechnology Journal **10,** 2–11.

Saldanha AJ. 2004. Java Treeview—extensible visualization of microarray data. Bioinformatics **20,** 3246–3248.

Sass J. 1977. Morphology. In: Sprague GF, ed. Corn and corn improvement. Madison, WI: American Society of Agronomy Publishers, 89–110.

Schnable JC, Freeling M. 2011. Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS One **6,** e17855.

Sekhon RS, Lin H, Childs KL, Hansey CN, Buell CR, de Leon N, Kaeppler SM. 2011. Genome-wide atlas of transcription during maize development. The Plant Journal **66,** 553–563.

Singh V, van Oosterom EJ, Jordan DR, Messina CD, Cooper M, Hammer GL. 2010. Morphological and architectural development of root systems in sorghum and maize. Plant and Soil **333,** 287–299.

Taramino G, Sauer M, Stauffer J, Multani D, Niu X, Sakai H, Hochholdinger F. 2007. The rtcs gene in maize (*Zea mays* L.) encodes a lob domain protein that is required for postembryonic shoot-borne and embryonic seminal root initiation. The Plant Journal **50,** 649–659.

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M. 2004. MapMan: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. The Plant Journal **37,** 914–939.

Tillich H. 1977. Vergleichend-morphologische Untersuchungen zur Identitaet der Gramineen-Primaerwurzel. (Comparative morphological investigations on the identity of the primary root of the Gramineae.) Flora **166,** 415–421.

Van De Wiel MA, Leday GG, Pardo L, Rue H, Van Der Vaart AW, Van Wieringen WN. 2012. Bayesian analysis of RNA sequencing data by estimating multiple shrinkage priors. Biostatistics **14,** 113–128.

Van Ha C, Le DT, Nishiyama R, Watanabe Y, Sulieman S, Tran UT, Mochida K, Van Dong N, Yamaguchi-Shinozaki K, Shinozaki K. 2013. The auxin response factor transcription factor family in soybean:

genome-wide identification and expression analyses during development and water stress. DNA Research **20,** 511–524.

Vermerris W, Boon JJ. 2001. Tissue-specific patterns of lignification are disturbed in the brown *midrib2* mutant of maize (*Zea mays* L.). Journal of Agricultural and Food Chemistry **49,** 721–728.

Vermerris W, Sherman DM, McIntyre LM. 2010. Phenotypic plasticity in cell walls of maize brown midrib mutants is limited by lignin composition. Journal of Experimental Botany **61,** 2479–2490.

Villordon AQ, Ginzberg I, Firon N. 2014. Root architecture and root and tuber crop productivity. Trends in Plant Science **19,** 419–425.

von Behrens I, Komatsu M, Zhang Y, Berendzen KW, Niu X, Sakai H, Taramino G, Hochholdinger F. 2011. *Rootless with undetectable meristem 1* encodes a monocot-specific AUX/IAA protein that controls embryonic seminal and postembryonic lateral root initiation in maize. The Plant Journal **66,** 341–353.

Wang L, Xie W, Chen Y, Tang W, Yang J, Ye R, Liu L, Lin Y, Xu C, Xiao J. 2010. A dynamic gene expression atlas covering the entire life cycle of rice. The Plant Journal **61,** 752–766.

Watt M, Schneebeli K, Dong P, Wilson IW. 2009. The shoot and root growth of Brachypodium and its potential as a model for wheat and other cereal crops. Functional Plant Biology **36,** 960–969.

Wen TJ, Hochholdinger F, Sauer M, Bruce W, Schnable PS. 2005. The *roothairless1* gene of maize encodes a homolog of *sec3*, which is involved in polar exocytosis. Plant Physiology **138,** 1637–1643.

Wen TJ, Schnable PS. 1994. Analyses of mutants of three genes that influence root hair development in *Zea mays* (Gramineae) suggest that root hairs are dispensable. American Journal of Botany **81,** 833–842.

Woll K, Borsuk LA, Stransky H, Nettleton D, Schnable PS, Hochholdinger F. 2005. Isolation, characterization, and pericycle-specific transcriptome analyses of the novel maize lateral and seminal root initiation mutant *rum1*. Plant Physiology **139,** 1255–1267.

Xu C, Tai H, Saleem M, Ludwig Y, Majer C, Berendzen KW, Nagel KA, Wojciechowski T, Meeley RB, Taramino G. 2015. Cooperative action of the paralogous maize lateral organ boundaries (LOB) domain proteins RTCS and RTCL in shoot borne root formation. New Phytologist **207,** 1123–1133.

Zeier J, Ruel K, Ryser U, Schreiber L. 1999. Chemical analysis and immunolocalisation of lignin and suberin in endodermal and hypodermal/rhizodermal cell walls of developing maize (*Zea mays* L.) primary roots. Planta **209,** 1–12.

Zhang M, Kong X, Xu X, Li C, Tian H, Ding Z. 2015. Comparative transcriptome profiling of the maize primary, crown and seminal root in response to salinity stress. PLoS One **10,** e0121222.

Zhang Y, Paschold A, Marcon C, Liu S, Tai H, Nestler J, Yeh C-T, Opitz N, Lanz C, Schnable PS. 2014. The Aux/IAA gene *rum1* involved in seminal and lateral root formation controls vascular patterning in maize (*Zea mays* L.) primary roots. Journal of Experimental Botany **65,** 4919–4930.

Zhao C, Craig JC, Petzold HE, Dickerman AW, Beers EP. 2005. The xylem and phloem transcriptomes from secondary tissues of the Arabidopsis root-hypocotyl. Plant Physiology **138,** 803–818.

Zhao M, Tai H, Sun S, Zhang F, Xu Y, Li W-X. 2012. Cloning and characterization of maize miRNAs involved in responses to nitrogen deficiency. PLoS One **9,** e0029669.

Zhu J, Ingram PA, Benfey PN, Elich T. 2011. From lab to field, new approaches to phenotyping root system architecture. Current Opinion in Plant Biology **14,** 310–317.

Zhu J, Kaeppler SM, Lynch JP. 2005. Topsoil foraging and phosphorus acquisition efficiency in maize (*Zea mays*). Functional Plant Biology **32,** 749–762.

Zhu J, Mickelson SM, Kaeppler SM, Lynch JP. 2006. Detection of quantitative trait loci for seminal root traits in maize (*Zea mays* L.) seedlings grown under differential phosphorus levels. Theoretical and Applied Genetics **113,** 1–10.

Journal of Experimental Botany, Vol. 67, No. 4 pp. 1095–1107, 2016
doi:10.1093/jxb/erv453 Advance Access publication 13 October 2015
This paper is available online free of all access charges (see http://jxb.oxfordjournals.org/open_access.html for further details)

RESEARCH PAPER

Extensive tissue-specific transcriptomic plasticity in maize primary roots upon water deficit

Nina Opitz[1], Caroline Marcon[1], Anja Paschold[1,*], Waqas Ahmed Malik[2], Andrew Lithio[3], Ronny Brandt[4], Hans-Peter Piepho[2], Dan Nettleton[3] and Frank Hochholdinger[1,†]

[1] Institute of Crop Science and Resource Conservation, Crop Functional Genomics, University of Bonn, D-53113 Bonn, Germany
[2] Institute for Crop Science, Biostatistics Unit, University of Hohenheim, D-70599 Stuttgart, Germany
[3] Department of Statistics, Iowa State University, Ames, IA 50011-1210, USA
[4] Leibniz Institute of Plant Genetics and Crop Plant Research, D-06466 Gatersleben, Germany

[†] Correspondence: hochholdinger@uni-bonn.de
[*] Present address: Max-Planck-Institute for Plant Breeding Research, D-50829 Cologne, Germany.

Received 23 June 2015; Accepted 25 September 2015

Editor: Adam Price

Abstract

Water deficit is the most important environmental constraint severely limiting global crop growth and productivity. This study investigated early transcriptome changes in maize (*Zea mays* L.) primary root tissues in response to moderate water deficit conditions by RNA-Sequencing. Differential gene expression analyses revealed a high degree of plasticity of the water deficit response. The activity status of genes (active/inactive) was determined by a Bayesian hierarchical model. In total, 70% of expressed genes were constitutively active in all tissues. In contrast, <3% (50 genes) of water deficit-responsive genes (1915) were consistently regulated in all tissues, while >75% (1501 genes) were specifically regulated in a single root tissue. Water deficit-responsive genes were most numerous in the cortex of the mature root zone and in the elongation zone. The most prominent functional categories among differentially expressed genes in all tissues were 'transcriptional regulation' and 'hormone metabolism', indicating global reprogramming of cellular metabolism as an adaptation to water deficit. Additionally, the most significant transcriptomic changes in the root tip were associated with cell wall reorganization, leading to continued root growth despite water deficit conditions. This study provides insight into tissue-specific water deficit responses and will be a resource for future genetic analyses and breeding strategies to develop more drought-tolerant maize cultivars.

Key words: Drought, expression pattern, low water potential, maize, primary root, RNA-Sequencing (RNA-Seq), tissue specificity, transcriptome, water deficit.

Introduction

In agriculture, drought is responsible for more yield losses than any other abiotic stress (Boyer, 1982). Besides widespread poor soil moisture, variations in water availability within fields result in uneven crop stands and cause yield losses (Nafziger *et al.*, 1991). As a consequence of global warming, land areas subjected to drought conditions increase each year. Thus, new insights into the molecular mechanisms of drought response and adaptation are required to advance

© The Author 2015. Published by Oxford University Press on behalf of the Society for Experimental Biology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

breeding and enable an estimated 70% increase in crop production until 2050 to feed the growing world population (Tester and Langridge, 2010).

When plants experience drought conditions, their growth is inhibited. Root development is less sensitive to water deficit than shoot development or leaf growth (Westgate and Boyer, 1985; Sharp *et al.*, 1988). Root growth maintenance is an important adaptive trait ensuring that plants can access deep water and nutrient resources to survive (Rodrigues *et al.*, 1995). It relies predominantly on the activity of the plant hormone abscisic acid (ABA). ABA accumulation prevents excessive ethylene production and growth inhibition. Furthermore, ABA promotes transport of proline to the root apex and thereby contributes to osmotic adjustment. Hence, cellular turgor is protected and a favourable water potential gradient towards the environment is maintained; that is, a more negative water potential than in the outside medium, enabling water influx. Maintenance of root elongation further depends on an increase of longitudinal cell wall extensibility (reviewed in Yamaguchi and Sharp, 2010). Other drought stress responses limit cellular damage by sustaining protein structure, maintaining membrane integrity, and removing toxic compounds such as reactive oxygen species (ROS) to restore redox balance (reviewed in Krasensky and Jonak, 2012). In spite of the well-known physiological processes, relatively little is known about the underlying gene regulatory networks that translate environmental changes to metabolic alterations needed to gain stress tolerance (Spollen *et al.*, 2008; Krasensky and Jonak, 2012).

Maize is one of the most important economic crops worldwide and an important source for food, fodder, and energy. Due to climate change, drought is a growing threat to most maize cultivation areas (IPCC, 2014). However, breeding approaches have not yet decreased maize drought sensitivity (Lobell *et al.*, 2014), making it important to understand the molecular mechanisms underlying the drought response.

Global gene expression profiles in response to water deficit have been monitored in different parts of maize including roots, leaves, and kernels by microarray chip hybridization experiments (Zheng *et al.*, 2004; Andjelkovic and Thompson, 2006; Jia *et al.*, 2006; Zhuang *et al.*, 2007; Yue *et al.*, 2008; Hayano-Kanashiro *et al.*, 2009; Marino *et al.*, 2009; Luo *et al.*, 2010; Downs *et al.*, 2013; Humbert *et al.*, 2013) and more recently by RNA-Sequencing (RNA-Seq) (Kakumanu *et al.*, 2012; Opitz *et al.*, 2014). RNA-Seq allows for fully quantitative gene expression analyses with absolute values and the capture of even subtle expression changes. These studies provided global insight into the maize water deficit response; however, the composite nature of the studied organs limited their spatial resolution. Plant organs are made up of different cell and tissue types each featuring a unique transcriptome (Brady *et al.*, 2007). Hence, transcriptome analyses of entire organs yield average gene expression profiles integrated over all cell types, thus potentially masking genes of interest (Schnable *et al.*, 2004). To date, there are only few transcriptome studies addressing this problem in maize roots focusing on different parts of the elongation zone after drought stress treatment (Bassani *et al.*, 2004; Sharp *et al.*, 2004; Poroyko *et al.*, 2007; Spollen *et al.*, 2008). Plant roots can be divided into specialized zones of development, including from the terminal end the root cap, the apical meristem, the cell elongation zone, and the maturation zone (Ishikawa and Evans, 1995). The root cap covers and thus protects the root tip, secretes mucilage to facilitate root movement, and acts as a sensor for gravity, light, temperature, and moisture gradients. The root apical meristem forms new cells that start to elongate and move proximally into the elongation zone where elongation reaches a maximum. The elongation zone is critical for the reaction to exogenous signals including the root growth response to water deficit. Eventually, cells dislocate into the maturation or differentiation zone and acquire their distinct functions (Ishikawa and Evans, 1995). The differentiation zone depicts in radial orientation several functionally diverse cell types. As an outermost layer, an epidermis that consists of root hair-forming trichoblasts and non-root hair-forming atrichoblasts encloses the root. The epidermis takes up water and nutrients, which are then either transported into aboveground plant organs or metabolized within the root. The epidermis is followed by multiple layers of cortex tissue and a single layer of endodermis. The adjacent pericycle represents the outermost layer of the central cylinder or stele which contains alternating poles of differentiated xylem vessels functioning in water and nutrient transport, and primary phloem elements instrumental in transport of photosynthates. The vasculature elements are embedded by parenchymal pith tissue that forms the centre of the root (Hochholdinger, 2009).

Polyethylene glycol (PEG) treatment was demonstrated to simulate the occurrence of drought stress in drying soil (Zheng *et al.*, 2004). In the present study, maize seedlings were exposed for 6 h to PEG8000 solution with a water potential of –0.8 MPa. Thus, water availability was in the mid range of naturally occurring, plant-usable soil water potentials, representing mild water deficit conditions (O'Geen, 2012). From an agronomic viewpoint, such moderate drought conditions are of great importance as they are frequent and reduce yield (Sinclair, 2011). During long-term (>24 h) exposure to such a low water potential, maize primary root growth is reduced by 30–50% (Saab *et al.*, 1992; Sharp *et al.*, 2004; Opitz *et al.*, 2014). However, short-term treatment did not impair root growth (Opitz *et al.*, 2014) and can therefore be used to monitor gene expression changes that precede the physiological changes. After treatment, the meristematic zone, the elongation zone, and the cortex and stele of the differentiation zone were mechanically separated and their transcriptomes analysed by RNA-Seq. The main goal of this study was to explore the tissue-specific plasticity of the transcriptomic landscape in response to water deficit.

Materials and methods

Plant material and water deficit treatment

Surface-sterilized seeds of the maize inbred line B73 were germinated in paper rolls as previously described (Ludwig *et al.*, 2013). After 4 d, seedlings with a primary root length of 2–4 cm were transferred to new paper rolls saturated with PEG8000 (M_r 7300–9000 Da; Roth, Karlsruhe, Germany) solution with a water

potential of –0.8 MPa for water deficit treatment, or distilled water for control experiments, as previously described (Opitz *et al.*, 2014). After 6 h of incubation in aerated PEG solution or distilled water, primary roots were dissected into four distinct tissues. First, the apical 2 mm of the root were collected, containing the root cap and meristematic zone. Secondly, the proximal zone adjacent to the root tip up to the part of the root where the root hairs became visible was sampled, which corresponds to the elongation zone. Finally, the distal differentiation zone, from the root hair zone to the coleorhiza, was mechanically separated into cortex (including epidermis, cortical parenchyma, and endodermis) and stele (including pericycle, phloem, xylem, and pith parenchyma) according to Saleem *et al.* (2010). Briefly, after cutting the apical tissues off, the root was incised with a razor blade close to the coleorhiza. The cortex was then pulled off the stele by gently bending the root while the stele remained attached to the remainder of the seedling. Collected plant material was immediately frozen in liquid nitrogen and stored at –80 °C until RNA isolation. Experiments were performed in four biological replicates each consisting of 10 pooled root tissues from the same 10 seedlings.

RNA isolation and sequencing library preparation

Pooled primary root tissues were ground in liquid nitrogen, and total RNA was isolated with the RNeasy Plus Universal Mini Kit (Qiagen, Venlo, The Netherlands). RNA quality was assessed via agarose gel electrophoresis and a Bioanalyzer (Agilent RNA 6000 Nano Chip; Agilent Technologies, Santa Clara, CA, USA). For all samples, a RIN (RNA integrity number) ≥8.5 was detected. cDNA libraries for Illumina sequencing were constructed according to the protocol of the manufacturer (TruSeq RNA Sample Preparation; Illumina, San Diego CA, USA). For sequencing, each library was indexed by an Illumina TruSeq Adapter in which no adaptor was used more than twice (Supplementary Table S1 available at *JXB* online). Indexed libraries were loaded onto a flow cell following an incomplete block design (generated with CycDesigN 5.1) with four pooled libraries per lane, which was considered as an incomplete block (Supplementary Table S1). Cluster preparation and single read sequencing were performed according to the manufacturer's instructions (HiSeq 2000, Illumina).

Processing and mapping of sequencing reads

Raw sequencing reads were processed and subsequently mapped with CLC Genomics Workbench (Version 7.0; http://www.clcbio.com/products/clc-genomics-workbench/). Reads with more than one mismatch in the adapter sequence were excluded. Quality trimming removed low quality and ambiguous nucleotides of sequence ends and adapter contamination. Only reads ≥40 bp were retained for further analyses. These were initially mapped to the maize B73 reference genome (RefGen_v2; ftp://ftp.gramene.org/pub/gramene/maizesequence.org/release-5b/assembly/; Schnable *et al.*, 2009) allowing large gaps of up to 50 kb to span introns. To be mapped, at least 90% of each read had to fit with 90% similarity to the reference. Stacked reads (i.e. redundant reads sharing the same start and end co-ordinate), sequencing direction, and sequence were merged into one. The remaining reads were further projected to the 'filtered gene set' (FGS_v2; ftp://ftp.gramene.org/pub/gramene/maizesequence.org/release-5b/filtered-set/; Release 5b) of the B73 reference genome. Thereby, reads had to fit at least with 80% of their length comprising 90% similarity to the reference. Only reads uniquely mapping to the reference were further analysed.

Sample relationships were displayed and compared in a hierarchical clustering, and a multidimensional scaling plot was produced with the Bioconductor package 'edgeR' in R. The multidimensional scaling plot displays the distances between the samples as the leading-$\log_2$ fold change corresponding to the estimated root-mean-square deviation of the top genes with the largest standard deviations between samples.

Statistical procedures to determine active and inactive genes

A generalized linear mixed model with a negative binomial response was used to model gene-wise transcriptional activity for each tissue–treatment combination. Conditional on random effects for sequencing lanes and biological replicates, the log of the mean read count was assumed to be the sum of a normalization factor and a linear function of the random and fixed effects for each combination of tissue and treatment. Each normalization factor was calculated by adding the log of the TMM normalization factor (Robinson and Oshlack, 2010) for each sample to the fitted value of a smoother on the log of the mean of the raw counts using GC content and gene length as covariates.

The vector of fixed effects for each gene was assumed to be a draw from a multivariate normal distribution with an unknown and unrestricted mean and an unknown diagonal variance–covariance matrix. The log of the negative binomial dispersion parameter was assumed to be constant within a gene, and a draw from a normal distribution with unknown mean and variance. The precisions of the random effects were assumed to follow gamma distributions, where the parameters for the lane effects were specified to create a vague distribution. The unknown parameters in the distributions of the fixed effects, negative binomial dispersion, and precision of the biological replicate effects were estimated using an empirical Bayes procedure via the R package 'ShrinkBayes' (Van De Wiel *et al.*, 2012). Integrated nested Laplace approximation (Rue *et al.*, 2009) was used to approximate the posterior distribution for the fixed effect associated with gene, tissue, and condition.

For a given threshold T, the posterior distribution for gene g, tissue t, and condition c was used to find $P_{gtc}(T)$, which is the posterior probability that the fixed effect for gene g, tissue t, and condition c was larger than T. Gene g was called 'active' for tissue t and condition c if $P_{gtc}(T) > 0.5$ and 'inactive' otherwise. Because fixed effects were estimated while accounting for sequencing differences from sample to sample, gene length, and GC content differences, classifying genes as active or inactive based on the posterior distribution of fixed effects is more meaningful than attempting to do so based on a single raw read count threshold applied to all genes. In terms of counts, the selected threshold T resulted in calling a gene of average length and average GC content active if the expected number of reads per million mapped reads was approximately ≥2. The read count threshold was, in effect, adjusted up or down for other genes depending on gene length, GC content, and the observed empirical distribution between these variables and read count.

Statistical procedures for analysing differential gene expression

For differential expression analyses, read counts (+0.5) were log-transformed and mean variance trends estimated. Only genes with a minimum of five mapped reads in all four replicates of at least one tissue sample were considered. Based on the predicted variance, weights were assigned to each observation to adjust for heteroscedasticity in the linear modelling process (Law *et al.*, 2014). To borrow strength across genes in the estimation of the residual error variance, the empirical Bayes approach implemented in the Bioconductor package 'limma' (Smyth, 2005) was used in R. The data were further analysed by fitting a linear mixed model that included a lane effect, where lanes corresponded to the incomplete blocks of the experimental design. The lane effect was considered as a random effect, thus allowing the recovery of the interblock information. Four pairwise comparisons were computed, comparing water deficit with control treatment for each tissue. Resulting *P*-values of contrasts were corrected for multiplicity using the false discovery rate (FDR) approach of Benjamini and Hochberg (1995).

Gene Ontology (GO) and metabolic pathway analyses

The web-based agriGO software (http://bioinfo.cau.edu.cn/agriGO/index.php) was used to assign GO functional categories to differentially expressed genes. Singular enrichment analysis computed

over-represented categories in the sets of differentially expressed genes by comparing them with GO terms in the set of all expressed genes using Fisher's exact test (Du *et al.*, 2010). Multiple testing was corrected by FDR (Benjamini and Yekutieli, 2001), and a cut-off was introduced at 5%.

The Mapman software (http://mapman.gabipd.org/web/guest/mapman; Thimm *et al.*, 2004) assigned differentially expressed genes to metabolic pathways and subsequently visualized them based on the functional annotation file ZmB73_5b_FGS_cds_2011. A χ^2 test was used to determine if more genes than expected from the distribution of all expressed genes were assigned to each of the 32 major Mapman categories. The same test was used to evaluate the distribution of differentially expressed transcription factors (TFs) into 55 TF families (according to the Plant Transcription Factor Database v3.0; http://planttfdb.cbi.pku.edu.cn/; Jin *et al.*, 2013).

Results

RNA-Sequencing and mapping of root tissue transcriptomes

To survey the transcriptomic dynamics of maize root tissues in response to water deficit, seedlings were subjected to a low water potential of –0.8 MPa and control conditions for 6 h (see the Materials and methods). Subsequently, four root tissues were sampled, namely the meristematic zone comprising the apical meristem and root cap (Mz), the elongation zone (Ez), and both the cortex (Co) and stele (St) of the differentiation zone (Fig. 1A–C). For transcriptome sequencing, RNA was extracted in four biological replicates per tissue and treatment and converted into cDNA libraries for sequencing procedures. RNA-Seq yielded on average ~49 million 100 bp reads per sample. The sequencing data have been deposited in the NCBI sequencing read archive (SRA; http://www.ncbi.nlm.nih.gov/sra; accession no. SRP052697). On average, 85% of high-quality reads mapped uniquely to the maize reference genome (Supplementary Table S1 at *JXB* online). After removal of duplicated reads, 74% of the remaining reads mapped to unique positions in the 'filtered gene set' (FGS; Supplementary Table S1), a set of 39 656 high confidence gene models.

Gene activity under control and water deficit conditions

Gene-wise transcriptional activity for each tissue–treatment combination was computed. To decide if a gene was expressed ('active') in any tissue–treatment combination or not ('inactive') (Supplementary Table S2; Supplementary Fig. S1 at *JXB* online), a generalized linear mixed model was applied (see the Materials and methods). In total, 27 438 genes (69% of the FGS; Fig. 1E) were declared active. In the surveyed tissues, between 235 and 752 more genes were active after water deficit treatment (Fig. 1D). A substantial number of genes was constitutively active in all four tissues under control (19 463: 72%) and water deficit (20 220: 75%) conditions. In contrast, only 2300 (8%) and 1914 (7%) genes were specifically active in a single tissue (Fig. 1D). The majority of active genes (19 163: 70%) were constitutively active in all tissues under both conditions (Fig. 1E). The activity of the remaining genes was classified either as 'stable' or as 'dynamic'. Stable genes were detected in the same tissues under both conditions and thus did not change their activity status; that is, they were neither turned on nor turned off in any tissue following treatment (4329 genes). A total of 31% of stable genes (1336/4329) were active in a single tissue, while 69% (2993/4329) were active in multiple tissues (Fig. 1E; Supplementary Fig. S1). On the other hand, 3946 genes were classified as dynamic genes. These genes changed their activity status in at least one tissue following water deficit treatment. Functionally, genes with stable and dynamic expression were related to similar metabolic pathways. However, genes involved in stress response processes were more numerous among the dynamic genes. Genes with dynamic activity were further categorized according to the type of change: the activity of 2607 genes spatially expanded and included more tissues after treatment, while the activity of 1157 genes was restricted to fewer tissues. A minority of 182 genes displayed an activity change in both directions (i.e. included different tissues after water deficit; Fig. 1E; Supplementary Fig. S1). Spatial activity changes were quantified; for example, a single change indicated a gene that was active in one tissue under control conditions and in the same tissue plus in an additional tissue after treatment. The maximum of four changes indicated a gene that was inactive under control conditions but active in all tissues after treatment (and vice versa). Ten genes displayed this kind of pattern, encoding HSF (heat shock factor) TFs 8, 17, and 20, TCP TF 36, a putative ABA signalling protein, two putative LEA (late embryogenesis abundant) proteins, and three unknown proteins. Overall, single changes were most common, while multiple changes were less frequent the more changes they included (Fig. 1E).

Tissue-specific gene regulation in response to water deficit

Hierarchical clustering of sample data revealed that the transcriptomes of each root tissue clustered together under control and water deficit conditions (Fig. 2A). Therefore, differences between tissues were more severe than those between treatments. Additionally, the young, apical tissues of the meristematic and elongation zone and the mature tissues of the differentiation zone stele and cortex formed separate clusters (Fig. 2A). In a multidimensional scaling plot, the samples of the meristematic and elongation zone clustered closely together, while cortex and stele samples were more distantly related (Fig. 2B). To identify genes differentially regulated in response to water deficit treatment, pairwise contrasts between control and water deficit conditions were computed for each tissue, yielding four sets of differentially expressed genes (Fig. 2C; for a complete gene list including normalized expression values, $\log_2$ fold changes (Fc), and *q*-values, see Supplementary Table S3 at *JXB* online). The pattern of responsive gene quantities was cortex >elongation zone >meristematic zone >stele (Fig. 2C). Thereby, 2878 genes were differentially regulated in the cortex, 2313 in the elongation zone, 1348 in the meristematic zone, and 846 in the stele (FDR <1%; Fig. 2C). Among differentially expressed genes, more than half displayed a small absolute

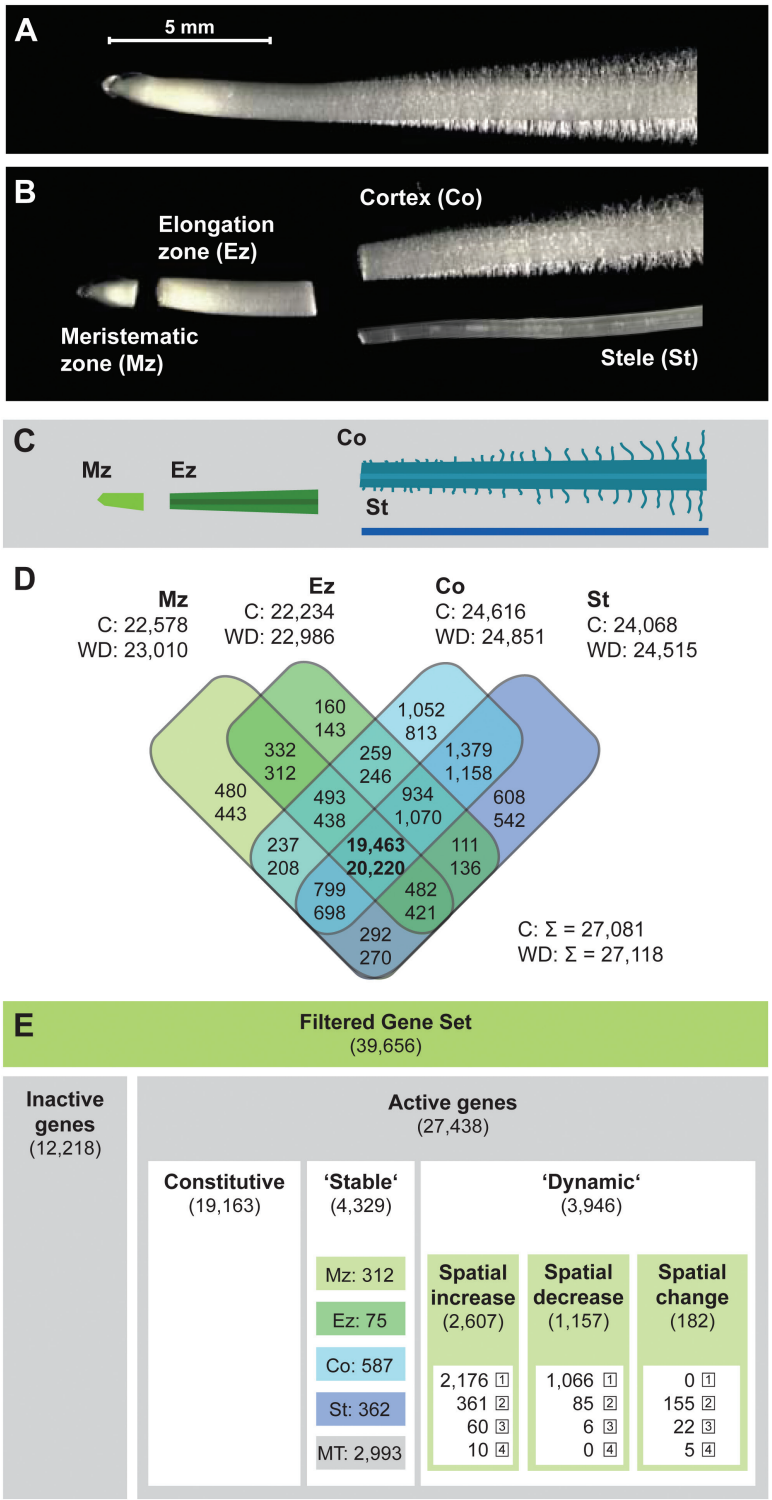


Fig. 1. Maize primary root tissues and tissue-specific gene activity. (A) Maize primary root prior to tissue sampling. (B) Manually dissected root with separated meristematic zone, elongation zone, cortex, and stele. (C) Colour-coded schematic of dissected root. (D) Overlap of active genes under control (upper numbers) or water deficit conditions (lower numbers) in each tissue. (E) Subdivision of genes according to their activity status (see the text). Stable genes were active in the same tissue(s) under both conditions (activity in single tissues: Mz, Ez, Co, St; activity in multiple tissues: summed in MT). Dynamic genes were active in additional tissues (spatial increase), fewer tissues (spatial decrease), or different tissues (spatial change) upon water deficit. Numbers in boxes depict the number of activity status changes. Mz, meristematic zone; Ez, elongation zone; Co, cortex; St, stele; C, control; WD, water deficit treatment.

$\log_2$Fc of ≤1 (Fig. 2D; Supplementary Fig. S2). In subsequent analyses, we focused on genes with a $|\log_2 Fc| \geq 1$. This arbitrary cut-off resulted in 466 differentially expressed genes in the meristematic zone, 995 in the elongation zone, 708 in the cortex, and 345 in the stele (Fig. 2D). In the elongation zone, the highest portion (43%) of differentially expressed genes

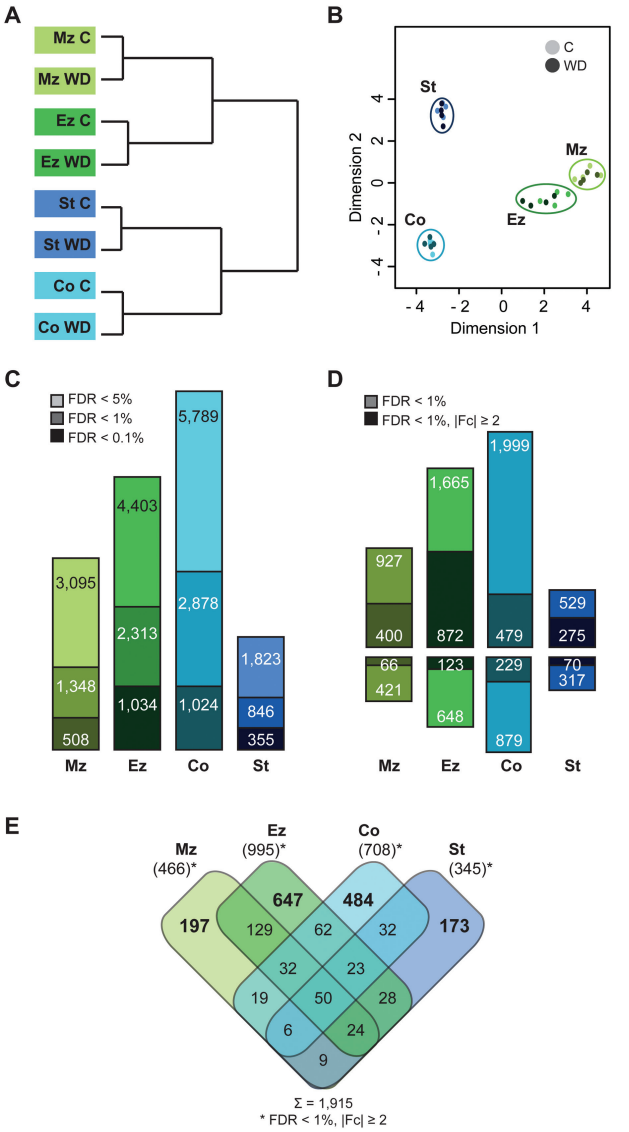


Fig. 2. Sample relationship and differential gene expression. (A) Hierarchical clustering and (B) multidimensional scaling plot of root tissue transcriptomes. (C) Numbers of differentially expressed genes in the four root tissues at different significance levels (light colour, FDR <5%; medium dark colour, FDR <1%; dark colour, FDR <0.1%). (D) Numbers of up- and down-regulated genes (FDR <1%); dark bars represent strongly regulated genes ($|\log_2 Fc| \geq 1$). (E) Overlap between differentially expressed genes (FDR <1%; $|\log_2 Fc| \geq 1$) in the four root tissues. Mz, meristematic zone; Ez, elongation zone; Co, cortex; St, stele; C, control; WD, water deficit treatment.

exceeded the Fc cut-off, while only a quarter of regulated genes in the cortex displayed such a high Fc (Supplementary Fig. S2). Similarly, the maximal absolute Fc was highest in the elongation zone and lowest in the cortex [Ez (340-fold) >Mz (116-fold) >St (93-fold) >Co (50-fold)].

Most differentially expressed genes (68–88%; Fig. 2D) were up-regulated. The direction of regulation was conserved among 407 (98%) of the 414 water deficit-responsive genes that were differentially expressed in more than one tissue.

Of a total of 915 water deficit-responsive genes, only 50 were consistently regulated in all tissues (48 persistently up-regulated, two persistently down-regulated), while the majority of 1501 genes were specifically regulated in a single tissue, illustrating a high degree of tissue-specific transcriptomic plasticity in response to water deficit (Fig. 2E).

Metabolic pathway responses to water deficit

An overview of metabolic processes regulated in response to water deficit in each of the four tissues was generated with Mapman (Supplementary Fig. S3 at *JXB* online). Differentially expressed genes (FDR <1%; $|\log_2 Fc| \geq 1$) were categorized into 32 major functional Mapman categories. A variety of different biochemical pathways were affected by water deficit treatment in all tissues, including major and minor cartbohydratre (CHO) metabolism (i.e. sucrose, starch, and sugar derivate metabolism), energy generation, and lipid metabolism. A category was declared enriched if significantly more differentially expressed genes were assigned to it as expected from the distribution of all expressed genes. For all tissues, such enrichment was detected for the categories 'hormone metabolism', 'stress', and—in all but the meristematic zone—'transport' (Fig. 3B). In the class 'hormone metabolism', genes were involved in biosynthesis, degradation, and signalling processes of all plant hormones, but most in auxin, ABA, and ethylene pathways. Tissue-specific regulation was prominent, as different genes—albeit with similar functions—were affected in different tissues. Exceptions were ABA-related processes: eight genes were commonly regulated including up-regulation of the ABA biosynthesis genes *zep1*, *zep2*, and *vp14*. Most genes annotated in the category 'stress' belong to the subgroup 'abiotic stress' and encode, for example, many putative heat shock proteins (HSPs) and HSFs. In the category 'transport', diverse processes including transport of sugars, peptides, and amino acids were altered by water deficit. The transcriptomes of the meristematic and elongation zones further displayed an enrichment of the category 'cell wall metabolism'. 'Amino acid metabolism' was significantly enriched in the meristematic zone and in the cortex. Altered pathways included, for example, down-regulation of proline degradation and up-regulation of cysteine, lysine, and tryptophan biosynthesis. In total, the cortex displayed the largest number of enriched categories, including in addition 'secondary metabolism' with terpenoid and phenylpropanoid metabolic processes.

Gene Ontology categorization of water deficit-responsive genes

Further functional categorization was performed according to GO terms with agriGO. Over-represented GO categories were computed by singular enrichment analyses. In total, 33 different GO terms were over-represented, 29 in the ontology 'biological process', three in 'molecular function', and one term in 'cellular component' (Fig. 3A). Two categories were conserved among differentially regulated genes (FDR <1%; $|\log_2 Fc| \geq 1$) in all tissues: 'transcription factor activity' and 'transcription, DNA-templated'. A comparison with the maize transcription factor database identified 199 (of 1652 expressed TFs: 10%) TFs among water deficit-responsive genes. Figure 3C displays the percentage of differentially expressed TFs within 27 major TF families. As among all expressed genes, most differentially regulated TFs

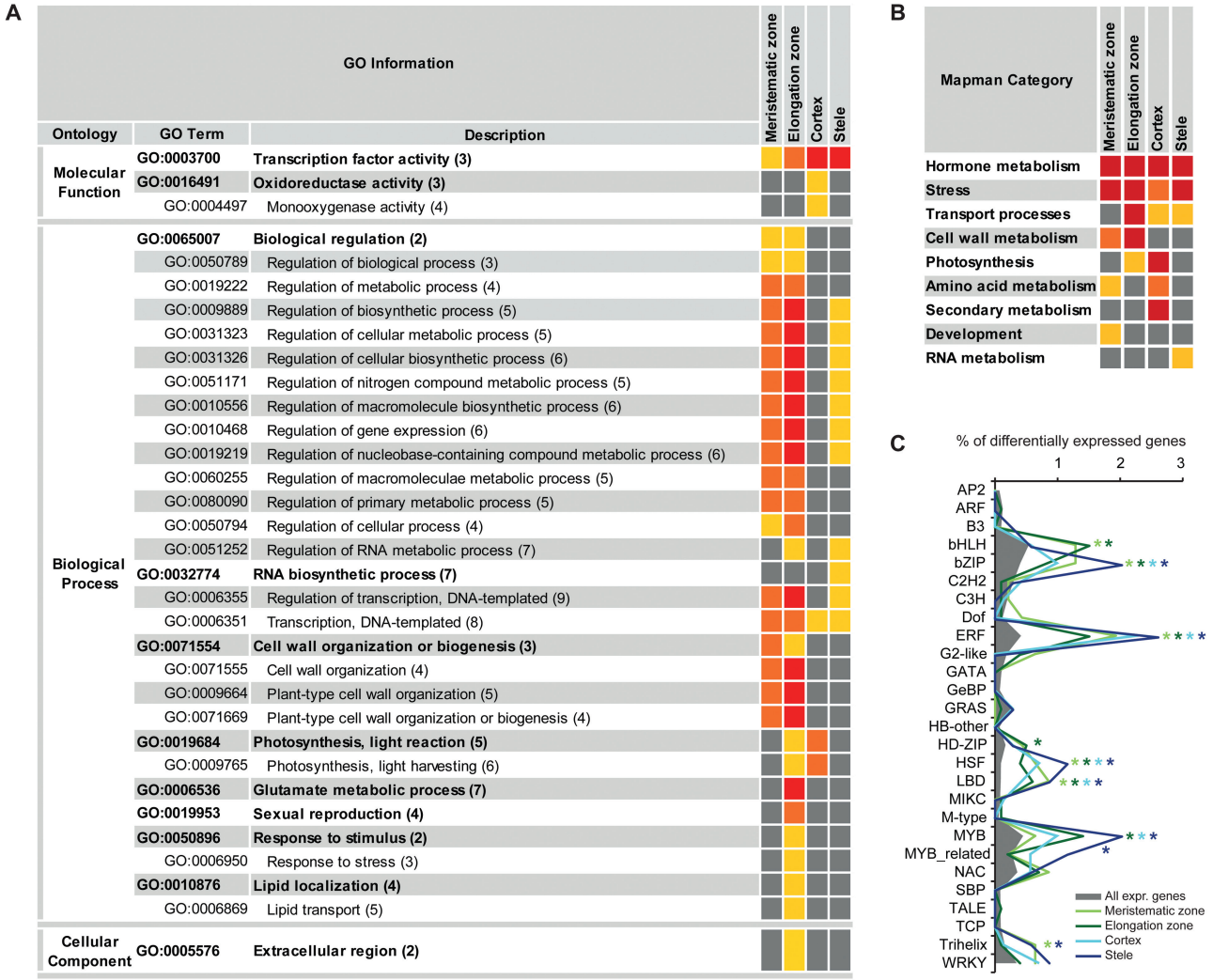


Fig. 3. Functional categorization of differentially expressed genes. (A) Cross-comparison of enriched GO terms and (B) Mapman categories among differentially expressed genes (FDR <1%; $|\log_2 Fc| \geq 1$) in the four root tissues in response to water deficit treatment. [Numbers in parentheses indicate the GO level; different colours in the right-hand columns represent different significance levels (*q*-values in A; *P*-values in B) of the over-representation; yellow, <0.05; orange, <0.01; red, <0.001.] (C) Distribution of differentially expressed genes (FDR <1%; $|\log_2 Fc| \geq 1$) into TF families as a percentage of all differentially expressed genes (grey colour represents percentages of TFs per family among all expressed genes; only 27 TF families with >15 expressed members are shown). Asterisks indicate TF families with more differentially expressed genes than expected, χ^2 test, $P<0.05$.

belonged to bHLH, bZIP, ERF, MYB, and NAC TF families. However, different genes were regulated in the distinct tissues, with 144 (of 199) TFs being specifically affected in a single tissue. Compared with all expressed genes, bHLH, bZIP, ERF, HD-ZIP, HSF, LBD, MYB, MYB-related, and Trihelix TFs were significantly over-represented among water deficit-responsive genes in one or more tissues (Fig. 3C). In addition to gene expression regulation by TFs, further GO categories and subcategories related to regulation of transcription such as 'RNA biosynthetic process' were enriched among water deficit-responsive genes, in the meristematic zone, elongation zone, and stele (Fig. 3A). Additionally, many groups related to regulation of general metabolic processes were enriched in these three tissues. In the stele, no other category was over-represented. In the meristematic and elongation zone, cell wall metabolism-related processes were commonly over-represented, as already indicated by the Mapman analyses. Predominantly, up-regulated genes in this group were annotated as expansins or xyloglucan endotransglycosylases/hydrolases (XTHs). The cortex and elongation zone shared two over-represented terms, 'photosynthesis, light reaction' and 'photosynthesis, light harvesting', that included mostly genes annotated as encoding chlorophyll-binding proteins. The molecular functions 'oxidoreductase activity' and 'monooxygenase activity' corresponding to redox regulation were specifically over-represented in the cortex. The highest number of enriched GO terms was found among the responsive genes identified in the elongation zone. These included 'glutamate metabolic process', 'sexual reproduction', 'response to stimulus/stress', 'lipid localization', and the cellular component 'extracellular region'. The last-mentioned term, however, includes many expansins that are also annotated in cell wall metabolism. This holds true for genes included in 'sexual reproduction' as these might be pollen allergens and/or expansins and are also annotated in both 'cell wall organization' and 'extracellular region'.

Water deficit-responsive expression changes of 'classical maize genes'

Among the 39 656 high confidence gene models of the maize FGS, 3599 are currently hand curated with experimentally confirmed functions ('classical maize genes'; Schaeffer *et al.*, 2011; Schnable and Freeling, 2011). Of these, 311 genes were water deficit responsive (FDR <1%; $|\log_2 Fc| \geq 1$; Supplementary Table S4 at *JXB* online). A majority of 236 (76%) were tissue specifically regulated in a single tissue, a similar proportion as among all expressed genes (Fig. 4; compare Fig. 2E). More than half (>54%) of differentially expressed, classical genes encode TFs (48/82 in Mz, 97/158 in Ez, 63/116 in Co, 46/70 in St), which is a slightly higher proportion than in all expressed classical genes (1489/2989). Other water deficit-responsive, classical genes had diverse functions; many were related to stress responses, transport, or hormone metabolism. Overall, the identified water deficit-responsive classical genes reflect the functions observed for all differentially expressed genes. For instance, the aquaporin genes *pip1b*, *pip1e*, and *pip2d*, the cell division-related gene *zyp1*, and the cell elongation-related gene *expa1* were specifically regulated in the meristematic zone. Also involved in cell wall modification is *xth1* that was specifically regulated in the elongation zone upon water deficit. Furthermore, several genes involved in gibberellin (GA) metabolism (*ga16,17ox1*, *ga2ox1*, *ga2ox2*, and *ga2ox3*) were regulated in this tissue. In the cortex, the ABA-related gene *aasr6* was specifically differentially expressed, as were some genes involved in secondary metabolism (*ccd7*, *ggps2*, *hyd4*, and *ks2*) and transport of potassium (*kch1* and *kch4*). Genes solely regulated in the stele included the ion homeostasis regulator *pmpm4* (Fig. 4).

Discussion

The first plant organ that detects a limitation of the water supply is the emerging root system. Roots represent a gradient of development, with undifferentiated cells near the tip and fully differentiated cells at the basal end. In the present study, transcriptome changes in four root tissues were surveyed to gain insight into tissue-specific water deficit responses. The meristematic zone was collected as the apical 2 mm of the root and included the region in which cell division rates peak (Sacks *et al.*, 1997). Under severe water deficit conditions, fewer cells are produced and the primary root meristem is smaller than under optimal conditions (Fraser *et al.*, 1990). Sacks *et al.* (1997) detected the end of the cell division zone at 2.2 mm from the root tip under control conditions and at 1.8 mm under low water potentials as applied in this study. As newly formed cells start to elongate within the meristematic zone, no definite discrimination from the elongation zone can be made (Ishikawa and Evans, 1995). However, in the apical 2 mm, no differences in elongation rate between control and water deficit conditions were observed (Sharp *et al.*, 1988; Fraser *et al.*, 1990; Saab *et al.*, 1992; Sacks *et al.*, 1997). Thus, the apical 2 mm seems a good sample size to represent the apical meristem. Secondly, the elongation zone was sampled up to the part of the root where root hairs emerge. Root hair development is a common indicator for the region of mature cells. From the remaining part of the root, the outer cortex (containing epidermis, cortical parenchyma, and endodermis) and inner stele (including pericycle, phloem, xylem, and pith parenchyma) were sampled, corresponding to two sets of distinct cell types with very disparate physiological functions.

Gene activity expands to more tissues upon water deficit

Per gene, the transcriptional activity (i.e. presence or absence of gene expression) was determined for each tissue–treatment combination. In contrast to analysing relative expression differences, the conversion into binary values (activity or inactivity) allows the analysis of fluctuation of expression at the level of detection (Birnbaum *et al.*, 2003). In total, 27 438 genes were active in at least one tissue–treatment combination. A similar number of 27 347 genes had been detected in a study analysing the same primary root tissues of untreated maize seedlings (Paschold *et al.*, 2014). Only 8% of the genes were specifically active in a single root tissue, which is a similar proportion to that detected by Paschold *et al.* (2014) and

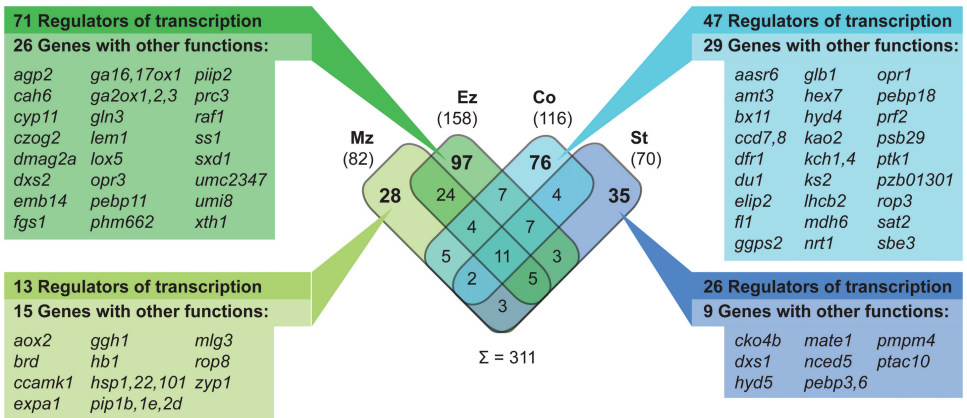


Fig. 4. Classical maize genes among water deficit-responsive genes (FDR <1%, $|\log_2 Fc| \geq 1$) and their overlap between tissues. For tissue-specific genes, the numbers of genes with functions in regulation of gene expression (e.g. TFs) are indicated as well as genes with other functions. (For the complete list of water deficit-responsive classical genes see Supplementary Table S4 atr *JXB* online.) Mz, meristematic zone; Ez, elongation zone; Co, cortex; St, stele.

in maize leaves (Li *et al.*, 2010). The vast majority of genes, however, were constitutively active under both conditions in all four tissues. The remaining genes displayed specific activity patterns. They were either active in the same tissue(s) under both conditions (stable genes) or gene activity changed in response to water deficit (dynamic genes). Of the latter, many genes were active in more tissues upon water deficit than under control conditions. Furthermore, 10 genes that were inactive under control conditions but active in all tissues after water deficit treatment were identified. These genes included known stress-responsive genes such as three HSF TFs, a putative ABA signalling protein, and two putative LEA proteins. In general, functions of stable and dynamic genes were similar. However, a higher number of genes involved in stress response processes were observed among unstable genes, reflecting a highly dynamic response to water deficit.

Cortex and elongation zone transcriptomes are most strongly affected by water deficit

Differentially expressed genes between control and water deficit treatment were identified in each tissue by pairwise comparisons. The largest effects of water deficit on gene regulation were detected in the cortex and elongation zone. The cortex displayed not only the highest number of water deficit-responsive genes but also the highest number of active genes *per se*. Similar results were obtained in analyses of *Arabidopsis* and *Brachypodium* leaves where more genes were drought responsive in mature and expansion zones (Skirycz *et al.*, 2010; Verelst *et al.*, 2012). These results indicate that differentiated tissues in general require a greater variety of gene expression levels than immature, dividing cells. The highest number of differentially expressed genes exceeding the Fc cut-off ($|\log_2 Fc| \geq 1$) was detected in the elongation zone. This strong impact of water deficit on elongating tissue reflects the uniqueness of this region as it mediates root growth responses and maintains root elongation during drought stress (see below; Sharp *et al.*, 2004).

Water deficit responses are tissue specific and mediated by up-regulation of gene expression

Only 50 of 1915 genes were differentially expressed in all four tissues in response to water deficit. Hence, while 70% (19 163/27 438) of the active genes were constitutively active in all tissues under both conditions, <3% of water deficit-responsive genes were consistently regulated in all tissues. This indicates extensive tissue-specific dynamics of the water deficit response. In total, 78% of differentially expressed genes were specific for a single tissue. Tissue specificity of drought stress responses was also detected in individual parts of the root elongation zone (Sharp *et al.*, 2004; Zhu *et al.*, 2007; Spollen *et al.*, 2008) and distinct leaf regions of *Arabidopsis* (Skirycz *et al.*, 2011) and *Brachypodium* (Verelst *et al.*, 2012). Genes commonly regulated in more than one tissue displayed the same direction of change in the present study and in a proteome study of the maize primary root elongation zone (Zhu *et al.*, 2007). Overall, the predominant direction of expression changes was induction. This was similarly detected in other maize root zones and whole primary roots (Spollen *et al.*, 2008; Opitz *et al.*, 2014), cotton and soybean roots (Bowman *et al.*, 2013; Fan *et al.*, 2013), *Sorghum* root/shoot samples (Dugas *et al.*, 2011), maize leaf meristems (Kakumanu *et al.*, 2012), and *Arabidopsis* leaves (Skirycz *et al.*, 2010). Furthermore, gene activity analyses in this study revealed that gene activity more frequently expanded to additional tissues than decreased to fewer tissues upon water deficit. Thus, water deficit might generally trigger induction of gene expression rather than inhibition, and an increase rather than a decrease in spatial gene activity.

Transcription factors and hormones mediate early water deficit responses in all primary root tissues

Alterations of environmental conditions trigger multiple signalling pathways that activate gene transcription and the downstream physiological adaptation. TFs play a key role in this process as main regulators of gene expression. However, expression of TFs is likewise influenced by exogenous signals. The maize FGS comprises 2231 annotated TFs. As in other plants, most of these belong to a few large multigene families such as bHLH, bZIP, ERF, and MYB (Jiang *et al.*, 2012). In this study, an over-representation of GO terms related to regulation of transcription was detected among water deficit-responsive genes in all primary root tissues. In total, ~10% of differentially expressed genes encoded TFs. About two-thirds of these were regulated in a single tissue, reflecting the high tissue-specific dynamics. Nevertheless, regulated TFs belonged to the same families such as bHLH, bZIP, ERF, and HSF that were enriched among the differentially expressed genes. Differentially expressed bHLH TFs were over-represented in the meristematic and elongation zone transcriptomes. bHLH TFs are generally up-regulated in response to drought stress and co-ordinate several processes including stomatal development, root hair formation, and hormone metabolism (reviewed in Castilhos *et al.*, 2014). Several bHLH TFs are involved in root growth control in *Arabidopsis*, for example by restricting root meristem size (Makkena and Lamb, 2013). Additionally, bHLH TFs regulate cell elongation under control of brassinosteroids (Zhang *et al.*, 2009) and therefore might be involved in growth adaptation in the root tip during water deficit. Further abiotic stress responses are mediated by bZIP TFs. Induced by ABA, they activate downstream responses such as expression of chaperones (Fujita *et al.*, 2005). In this data set, bZIP TFs were over-represented among differentially expressed genes in all tissues. However, individual genes were tissue specifically regulated. This is in agreement with the suggestion by Tang *et al.* (2012) that diverse bZIP TFs might function in distinct processes leading to stress tolerance. While bZIP TFs function via an ABA-dependent pathway, ERF TFs can integrate drought stress responses independently of ABA (Mizoi *et al.*, 2012). However, ERF TFs have also been reported to engage in ABA-mediated gene expression and to interact with TFs of other families and hormones (Xu *et al.*, 2011). In this study, ERF TFs were over-represented among regulated genes in

all tissues and were the most numerous, reflecting their central role in the acquisition of stress tolerance. In comparison with the other TFs, HSF TFs have a limited influence on gene expression as they primarily regulate expression of HSPs. Most HSPs act as molecular chaperones and prevent protein unfolding and aggregation, thereby maintaining cellular protein homeostasis and cellular structures (reviewed in Wang *et al.*, 2004). Expression of HSF TFs is regulated by different stresses, including drought, and is tissue specific (Lin *et al.*, 2011; Yang *et al.*, 2014). This was also observed in the present study, where different HSF TFs were over-represented among regulated genes in all tissues. Consequently, several HSPs were up-regulated, for example the classical genes *hsp1*, *hsp22*, and *hsp101* in the meristematic zone.

Transcriptional regulation of TFs and downstream genes is co-ordinated by hormones that act as central integrators of stress adaptation signalling cascades. Thereby the balance between the different hormones determines the appropriate response to the experienced stress (Xia *et al.*, 2015). The Mapman category 'hormone metabolism' was over-represented among differentially expressed genes in all root tissues. Genes involved in ABA metabolism were commonly regulated in all tissues, while for all other hormone-related genes tissue specificity was pronounced. ABA is a universal stress hormone that regulates nearly 10% of the protein-coding genes under diverse abiotic stress conditions (Sreenivasulu *et al.*, 2012). The ABA biosynthesis genes *zep1*, *zep2*, and *vp14* were up-regulated in all tissues in this study. Furthermore, the *vp14* homologue *nced5* was strongly induced in the stele. The *vp14* gene had the highest Fc in the meristematic and elongation zones, reflecting the fact that ABA accumulates towards the root apex where it is involved in maintenance of elongation through prevention of excess ethylene production. ABA accumulation might also be involved in osmotic adjustment, cell wall extensibility, and ion homeostasis (reviewed in Yamaguchi and Sharp, 2010). However, root elongation is only maintained in the apical region of the elongation zone, while more basal cells cease to elongate (Sharp *et al.*, 1988). This process is co-ordinated by GA which stimulates growth of most organs during plant development. However, during stress, active GA is reduced via up-regulation of *GA2ox* genes, whose products deactivate GA (Colebrook *et al.*, 2014). Consequently, *ga2ox1*, *ga2ox2*, and *ga2ox3* were specifically up-regulated in the elongation zone in the present study, reflecting the molecular processes needed to limit cell elongation to the apical part of the elongation zone.

Water deficit signalling induces cell protection processes in all primary root tissues

The complex interplay of different hormones and multiple TFs co-ordinates cellular responses to water deficit that include adjustments of the membrane system, modifications of the cell wall architecture, and changes in cell division. Metabolism is altered in various ways, including production of compatible solutes that are able to stabilize proteins and cellular structures and/or to maintain cell turgor by osmotic adjustment, and redox metabolism to remove excess levels of ROS and re-establish the cellular redox balance (Krasensky and Jonak, 2012). The effect of water deficit on cell membrane properties is mitigated by membrane lipid remodelling. Furthermore, lipids can act as signalling molecules during the stress response (reviewed in Okazaki and Saito, 2014). This was observed mainly in the elongation zone of the root, where the GO term 'lipid localization/transport' was over-represented among water deficit-responsive genes. In addition to a suitable structural barrier to the environment, ion homeostasis is crucial for physiological processes in living cells. Especially in roots, K^+ leakage mediated by potassium channels is a common phenomenon leading to irreversible K^+ loss under water deficit stress (Demidchik *et al.*, 2014). To prevent this, genes encoding potassium channels are down-regulated (Shabala and Pottosin, 2014); for example, *kch1* in the root cortex. Furthermore, up-regulation of *pmpm4* was specifically detected in the stele. The protein PMPM4 is involved in cation uptake and regulates the membrane potential, to maintain intracellular ion homeostasis (Fu *et al.*, 2012). Ion transport is regulated by ROS that are formed under stress (Demidchik *et al.*, 2014). ROS play a central role as signalling molecules, and spatial and temporal accumulation has a strong impact on hormone synthesis, transport, localization, and signalling (reviewed in Mittler *et al.*, 2011). However, if they reach a certain level, ROS, such as singlet oxygen, superoxide radicals, hydrogen peroxide, and hydroxyl radicals, become extremely deleterious and damage membranes, macromolecules, and nucleic acids. ROS are therefore kept under control by a scavenging system including enzymatic and non-enzymatic antioxidants (reviewed in Cruz de Carvalho, 2008). Genes encoding enzymes detoxifying ROS such as catalases, superoxide dismutases, and peroxidases were particularly over-represented in the elongation zone (GO category 'response to stimulus/stress') and cortex (GO category 'oxidoreductase activity/monooxygenase activity'). Molecular chaperones that stabilize proteins and membranes and assist in protein refolding under stress conditions, especially genes encoding HSPs and LEA-type proteins, were up-regulated in all tissues and over-represented, combined in the Mapman category 'stress'. Molecular chaperone function has also been identified for the amino acid proline. Proline can protect the integrity of different enzymes including those that act as ROS scavengers, and is further an important osmolyte. During water stress, proline biosynthesis is activated and its catabolism repressed (reviewed in Szabados and Savouré, 2010). Proline dehydrogenases that catalyse the first step of proline degradation were down-regulated in all tissues except in the stele, while transcription of several enzymes involved in proline biosynthesis was up-regulated. Furthermore, genes of the glutamate biosynthesis process were up-regulated, providing a source for proline production. In the elongation zone, this was reflected by over-representation of the GO term 'glutamate metabolic process' among differentially expressed genes and the up-regulation of the classic genes *fgs1* and *gln3*. In the Mapman analysis, the category 'amino acid metabolism' was further over-represented among regulated genes in the meristematic zone and cortex. In maize roots experiencing water deficit, proline is mainly transported to the root tip, whereby the basal region of the elongation

zone might serve as a source for transport. In the elongation zone, proline might also be used for synthesis of extensins that are involved in cell wall modification during drought adaptation (Yamaguchi and Sharp, 2010).

Cell wall loosening is induced to maintain root growth in apical root zones

Roots have the ability to elongate continuously under drought conditions. Thereby, cell elongation rates near the apex are maintained while more basal cells cease to elongate. This results in a shortened elongation zone (Sharp *et al.*, 1988). Maintenance of elongation despite reduced turgor pressure is achieved by cell wall-loosening proteins that render cell walls more extensible. Proteins with wall-loosening properties include mainly expansins and XTHs whose expression and activity increase in the root apex during drought conditions (reviewed in Wu and Cosgrove, 2000). Cell wall loosening can further be non-enzymatically enhanced by hydroxyl radicals that cause polysaccharide scission. Hydroxyl radicals are generated from hydrogen peroxide that increases in the apical elongation zone upon water deficit stress (Voothuluru and Sharp, 2012). The basal part of the elongation zone, on the other hand, ceases to elongate, and cell wall extension properties are inhibited possibly due to accumulation of lignin and cell wall-bound phenolics (Wu *et al.*, 1996; Bassani *et al.*, 2004; Fan *et al.* 2006; Spollen *et al.*, 2008). Under the experimental conditions in the present study, root growth is not visibly affected (Opitz *et al.*, 2014). Nevertheless, water deficit treatment leads to reduced turgor and cells adjust by increased expression of expansin-encoding genes to maintain local elongation rates (Yamaguchi and Sharp, 2010). Among differentially expressed genes in the apical root regions, the functional categories 'cell wall metabolism' and 'cell wall biosynthesis/organization' were over-represented, and the classical genes *exp1* and *xth1* were exclusively up-regulated in the meristematic and elongation zone, respectively. Additionally, ROS metabolism-related genes were over-represented among differentially expressed genes in the elongation zone (GO category 'response to stimulus/stress'). Besides cell wall loosening, cell expansion requires the co-ordinated uptake of water. The transcellular water transport is dependent on the amount and activity of water channels—aquaporins—in plasma membranes (reviewed in Chaumont and Tyerman, 2014; Li *et al.*, 2014). In the present study, the 11 classical *ZmPip* genes were expressed in all root tissues, and their expression was strongest in the differentiated tissues, as similarly described by Hachez *et al.* (2006). However, following water deficit treatment, up-regulation of *pip1b*, *pip1e*, and *pip2d* was detected only in the meristematic zone, and that of *pip1f* and *pip2a* in the meristematic and elongation zones. Similar up-regulation of several *pip* genes in maize has been found in whole seedling roots (Hachez *et al.*, 2012). Such enhanced aquaporin accumulation in root tips might allow maintenance of water uptake and, together with increased cell wall loosening, ensure continuous cell division and elongation despite water deficit so that plants might reach deeper water resources.

In summary, short-term mild water deficit that does not phenotypically affect maize plants had a considerable effect on root tissue transcriptomes. Regulation of gene expression was highly tissue specific and dominated by induction. The strongest changes were detected in the elongation zone, the root region important for growth adaptation. The water deficit response was mediated by TFs and changes in hormone metabolism in all root tissues, however by distinct sets of genes. This data set provides information for future analyses needed to understand the mechanisms underlying root adaptation to water scarcity and enable targeted breeding strategies to develop more water stress-tolerant varieties.

Supplementary data

Supplementary data are available at *JXB* online.

Figure S1. Overview of gene activity patterns and numbers of genes within these patterns.

Figure S2. Fc distribution of differentially expressed genes (FDR <1%).

Figure S3. Distribution of differentially expressed genes (FDR <1%) into major metabolic processes as visualized by Mapman.

Table S1. Overview of the sample distribution within the flow cell, biological replication, RNA-Seq output, and mapping results.

Table S2. List of activity patterns for 29 506 genes considered in a presence/absence analysis of gene expression.

Table S3. Comprehensive list of 27 386 expressed genes including normalized expression values, Fc, and *q*-values.

Table S4. Comprehensive list of expressed classical maize genes including Fc and *q*-values.

Acknowledgements

We thank Jutta Baldauf and Huanhuan Tai for technical assistance. FH received funding from the European Union's Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 289300.27. WAM and H-PP are funded by the Deutsche Forschungsgemeischaft (grant no. PI 377/17-1).

References

Andjelkovic V, Thompson R. 2006. Changes in gene expression in maize kernel in response to water and salt stress. Plant Cell Reports **25,** 71–79.

Bassani M, Neumann PM, Gepstein S. 2004. Differential expression profiles of growth-related genes in the elongation zone of maize primary roots. Plant Molecular Biology **56,** 367–380.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society **57,** 289–300.

Benjamini Y, Yekutieli D. 2001. The control of the false discovery rate in multiple testing under dependency. Annals of Statistics **29,** 1165–1188.

Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, Galbraith DW, Benfey PN. 2003. A gene expression map of the *Arabidopsis* root. Science **302,** 1956–1960.

Bowman MJ, Park W, Bauer PJ, Udall JA, Page JT, Raney J, Scheffler BE, Jones DC, Campbell BT. 2013. RNA-Seq transcriptome profiling of upland cotton (*Gossypium hirsutum* L.) root tissue under water-deficit stress. PLoS One **8,** 12.

Boyer JS. 1982. Plant productivity and environment. Science **218,** 443–448.

Brady SM, Orlando DA, Lee J-Y, Wang JY, Koch J, Dinneny JR, Mace D, Ohler U, Benfey PN. 2007. A high-resolution root spatiotemporal map reveals dominant expression patterns. Science **318,** 801–806.

Castilhos G, Lazzarotto F, Spagnolo-Fonini L, Bodanese-Zanettini MH, Margis-Pinheiro M. 2014. Possible roles of basic helix–loop–helix transcription factors in adaptation to drought. Plant Science **223,** 1–7.

Chaumont F, Tyerman SD. 2014. Aquaporins: highly regulated channels controlling plant water relations. Plant Physiology **164,** 1600–1618.

Colebrook EH, Thomas SG, Phillips AL, Hedden P. 2014. The role of gibberellin signalling in plant responses to abiotic stress. Journal of Experimental Biology **217,** 67–75.

Cruz de Carvalho MH. 2008. Drought stress and reactive oxygen species. Plant Signaling and Behavior **3,** 156–165.

Demidchik V, Straltsova D, Medvedev SS, Pozhvanov GA, Sokolik A, Yurin V. 2014. Stress-induced electrolyte leakage: the role of K^+-permeable channels and involvement in programmed cell death and metabolic adjustment. Journal of Experimental Botany **65,** 1259–1270.

Downs GS, Bi Y-M, Colasanti J, Wu W, Chen X, Zhu T, Rothstein SJ, Lukens L. 2013. A developmental transcriptional network for *Zea mays* defines coexpression modules. Plant Physiology **161,** 1830–1843.

Du Z, Zhou X, Ling Y, Zhang Z, Su Z. 2010. agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Research **38,** W64–W70.

Dugas DV, Monaco MK, Olson A, Klein RR, Kumari S, Ware D, Klein PE. 2011. Functional annotation of the transcriptome of *Sorghum bicolor* in response to osmotic stress and abscisic acid. BMC Genomics **12,** 514.

Fan L, Linker R, Gepstein S, Tanimoto E, Yamamoto R, Neumann PM. 2006. Progressive inhibition by water deficit of cell wall extensibility and growth along the elongation zone of maize roots is related to increased lignin metabolism and progressive stelar accumulation of wall phenolics. Plant Physiology **140,** 603–612

Fan X-D, Wang J-Q, Yang N, *et al*. 2013. Gene expression profiling of soybean leaves and roots under salt, saline–alkali and drought stress by high-throughput Illumina sequencing. Gene **512,** 392–402.

Fraser TE, Silk WK, Rost TL. 1990. Effects of low water potential on cortical cell length in growing regions of maize roost. Plant Physiology **93,** 648–651

Fu J, Zhang D-F, Liu Y-H, Ying S, Shi Y-S, Song Y-C, Li Y, Wang T-Y. 2012. Isolation and characterization of maize *PMP3* genes involved in salt stress tolerance. PLoS One **7,** 2.

Fujita Y, Fujita M, Satoh R, Maruyama K, Parvez MM, Seki M, Hiratsu K, Ohme-Takagi M, Shinozaki K, Yamaguchi-Shinozaki K. 2005. AREB1 is a transcription activator of novel ABRE-dependent ABA signaling that enhances drought stress tolerance in *Arabidopsis*. The Plant Cell **17,** 3470–3488.

Hachez C, Moshelion M, Zelazny E, Cavez D, Chaumont F. 2006. Localization and quantification of plasma membrane aquaporin expression in maize primary root: a clue to understanding their role as cellular plumbers. Plant Molecular Biology **62,** 305–323.

Hachez C, Veselov D, Ye Q, Reinhardt H, Knipfer T, Fricke W, Chaumont F. 2012. Short-term control of maize cell and root water permeability through plasma membrane aquaporin isoforms. Plant, Cell and Environment **35,** 185–198.

Hayano-Kanashiro C, Calderón-Vázquez C, Ibarra-Laclette E, Herrera-Estrella L, Simpson J. 2009. Analysis of gene expression and physiological responses in three Mexican maize landraces under drought stress and recovery irrigation. PLoS One **4,** 10.

Hochholdinger F. 2009. The maize root system: morphology, anatomy and genetics. In: Bennetzen JL, Hake SC, eds. The handbook of maize: its biology . Berlin: Springer, 145–160.

Humbert S, Subedi S, Cohn J, Zeng B, Bi Y-M, Chen X, Zhu T, McNicholas P, Rothstein S. 2013. Genome-wide expression profiling of maize in response to individual and combined water and nitrogen stresses. BMC Genomics **14,** 3.

IPCC. 2014. Summary for policymakers. In: Field CB, Barros VR, Dokken DJ, ***et al.***, eds. Climate Change 2014: impacts, adaptation, and vulnerability. Part A: global and sectoral aspects . Cambridge and New York: Cambridge University Press, 1–32.

Ishikawa H, Evans ML. 1995. Specialized zones of development in roots. Plant Physiology **109,** 725–727.

Jia J, Fu J, Zheng J, *et al.* 2006. Annotation and expression profile analysis of 2073 full-length cDNAs from stress-induced maize (*Zea mays* L.) seedlings. The Plant Journal **48,** 710–727.

Jiang Y, Zeng B, Zhao H, Zhang M, Xie S, Lai J. 2012. Genome-wide transcription factor gene prediction and their expressional tissue-specificities in maize. Journal of Integrative Plant Biology **54,** 616–630.

Jin J, Zhang H, Kong L, Gao G, Luo J. 2013. PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. Nucleic Acids Research **42,** D1182–D1187.

Kakumanu A, Ambavaram MMR, Klumas C, Krishnan A, Batlang U, Myers E, Grene R, Pereira A. 2012. Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-seq. Plant Physiology **160,** 846–867.

Krasensky J, Jonak C. 2012. Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. Journal of Experimental Botany **63,** 1593–1608.

Law CW, Chen Y, Shi W, Smyth GK. 2014. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biology **15,** R29.

Li G, Santoni V, Maurel C. 2014. Plant aquaporins: roles in plant physiology. Biochimica et Biophysica Acta **1840,** 1574–1582.

Li P, Ponnala L, Gandotra N, *et al*. 2010. The developmental dynamics of the maize leaf transcriptome. Nature Genetics **42,** 1060–1067.

Lin Y-X, Jiang H-Y, Chu Z-X, Tang X-L, Zhu S-W, Cheng B-J. 2011. Genome-wide identification, classification and analysis of heat shock transcription factor family in maize. BMC Genomics **12,** 76.

Lobell DB, Roberts MJ, Schlenker W, Braun N, Little BB, Rejesus RM, Hammer GL. 2014. Greater sensitivity to drought accompanies maize yield increase in the U.S. midwest. Science **344,** 516–519.

Ludwig Y, Zhang Y, Hochholdinger F. 2013. The maize (*Zea mays* L.) *AUXIN/INDOLE-3-ACETIC ACID* gene family: phylogeny, synteny, and unique root-type and tissue-specific expression patterns during development. PLoS One **8,** 11.

Luo M, Liu J, Lee RD, Scully BT, Guo B. 2010. Monitoring the expression of maize genes in developing kernels under drought stress using oligo-microarray. Journal of Integrative Plant Biology **52,** 1059–1074.

Makkena S, Lamb R. 2013. The bHLH transcription factor SPATULA regulates root growth by controlling the size of the root meristem. BMC Plant Biology **13,** 1.

Marino R, Ponnaiah M, Krajewski P, Frova C, Gianfranceschi L, Pè ME, Sari-Gorla M. 2009. Addressing drought tolerance in maize by transcriptional profiling and mapping. Molecular Genetics and Genomics **281,** 163–179.

Mittler R, Vanderauwera S, Suzuki N, Miller G, Tognetti VB, Vandepoele K, Gollery M, Shulaev V, Van Breusegem F. 2011. ROS signaling: the new wave? Trends in Plant Science **16,** 300–309.

Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. 2012. AP2/ERF family transcription factors in plant abiotic stress responses. Biochimica et Biophysica Acta **1819,** 86–96.

Nafziger ED, Carter PR, Graham EE. 1991. Response of corn to uneven emergence. Crop Science **31,** 811–815.

O'Geen A. 2012. Soil water dynamics. Nature Education Knowledge **3,** 12.

Okazaki Y, Saito K. 2014. Roles of lipids as signaling molecules and mitigators during stress response in plants. The Plant Journal **79,** 584–596.

Opitz N, Paschold A, Marcon C, Malik WA, Lanz C, Piepho H-P, Hochholdinger F. 2014. Transcriptomic complexity in young maize primary roots in response to low water potentials. BMC Genomics **15,** 741.

Paschold A, Larson NB, Marcon C, Schnable JC, Yeh C-T, Lanz C, Nettleton D, Piepho H-P, Schnable PS, Hochholdinger F. 2014. Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. The Plant Cell **26,** 3939–3948.

Poroyko V, Spollen W, Hejlek L, Hernandez A, LeNoble M, Davis G, Nguyen H, Springer G, Sharp R, Bohnert H. 2007. Comparing regional transcript profiles from maize primary roots under well-watered and low water potential conditions. Journal of Experimental Botany **58,** 279–289.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology **11,** R25.

Rodrigues ML, Pacheco CMA, Chaves MM. 1995. Soil–plant water relations, root distribution and biomass partitioning in *Lupinus albus* L. under drought conditions. Journal of Experimental Botany **46,** 947–956.

Rue H, Martino S, Chopin N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society **71,** 319–392.

Saab IN, Sharp RE, Pritchard J. 1992. Effect of inhibition of abscisic acid accumulation on the spatial distribution of elongation in the primary root and mesocotyl of maize at low water potentials. Plant Physiology **99,** 26–33

Sacks MM, Silk WK, Burman P. 1997. Effect of water stress on cortical cell division rates within the apical meristem of primary roots of maize. Plant Physiology **114,** 519–527

Saleem M, Lamkemeyer T, Schützenmeister A, Madlung J, Sakai H, Piepho H-P, Nordheim A, Hochholdinger F. 2010. Specification of cortical parenchyma and stele of maize primary roots by asymmetric levels of auxin, cytokinin, and cytokinin-regulated proteins. Plant Physiology **152,** 4–18.

Schaeffer ML, Harper LC, Gardiner JM, Andorf CM, Campbell DA, Cannon EKS, Sen TZ, Lawrence CJ. 2011. MaizeGDB: curation and outreach go hand-in-hand. Database **2011**, Article ID: bar022.

Schnable JC, Freeling M. 2011. Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS One **6,** 3.

Schnable PS, Hochholdinger F, Nakazono M. 2004. Global expression profiling applied to plant development. Current Opinion in Plant Biology **7,** 50–56.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA. 2009. The B73 maize genome: complexity, diversity, and dynamics. Science **326,** 1112–1115.

Shabala S, Pottosin I. 2014. Regulation of potassium transport in plants under hostile conditions: implications for abiotic and biotic stress tolerance. Physiologia Plantarum **151,** 257–279.

Sharp RE, Poroyko V, Hejlek LG, Spollen WG, Springer GK, Bohnert HJ, Nguyen HT. 2004. Root growth maintenance during water deficits: physiology to functional genomics. Journal of Experimental Botany **55,** 2343–2351.

Sharp RE, Silk WK, Hsiao TC. 1988. Growth of the maize primary root at low water potentials: I. Spatial distribution of expansive growth. Plant Physiology **87,** 50–57.

Sinclair TR. 2011. Challenges in breeding for yield increase for drought. Trends in Plant Science **16,** 289–293.

Skirycz A, De Bodt S, Obata T, *et al*. 2010. Developmental stage specificity and the role of mitochondrial metabolism in the response of *Arabidopsis* leaves to prolonged mild osmotic stress. Plant Physiology **152,** 226–244.

Skirycz A, Vandenbroucke K, Clauw P, *et al*. 2011. Survival and growth of *Arabidopsis* plants given limited water are not equal. Nature Biotechnology **29,** 212–214.

Smyth GK. 2005. *limma*: linear models for microarray data. In: Gentleman R, Carey V, Huber W, Irizarry R, Dudoit S, eds. Bioinformatics and computational biology solutions using R and Bioconductor . New York: Springer, 397–420.

Spollen W, Tao W, Valliyodan B, *et al*. 2008. Spatial distribution of transcript changes in the maize primary root elongation zone at low water potential. BMC Plant Biology **8,** 32.

Sreenivasulu N, Harshavardhan VT, Govind G, Seiler C, Kohli A. 2012. Contrapuntal role of ABA: does it mediate stress tolerance or plant growth retardation under long-term drought stress? Gene **506,** 265–273.

Szabados L, Savouré A. 2010. Proline: a multifunctional amino acid. Trends in Plant Science **15,** 89–97.

Tang N, Zhang H, Li X, Xiao J, Xiong L. 2012. Constitutive activation of transcription factor OsbZIP46 improves drought tolerance in rice. Plant Physiology **158,** 1755–1768.

Tester M, Langridge P. 2010. Breeding technologies to increase crop production in a changing world. Science **327,** 818–822.

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M. 2004. Mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. The Plant Journal **37,** 914–939.

Van De Wiel MA, Leday GG, Pardo L, Rue H, Van Der Vaart AW, Van Wieringen WN. 2012. Bayesian analysis of RNA sequencing data by estimating multiple shrinkage priors. Biostatistics **14,** 113–128.

Verelst W, Bertolini E, De Bodt S, Vandepoele K, Demeulenaere M, Pè ME, Inzé D. 2012. Molecular and physiological analysis of growth-limiting drought stress in *Brachypodium distachyon* leaves. Molecular Plant **6,** 311–322.

Voothuluru P, Sharp RE. 2012. Apoplastic hydrogen peroxide in the growth zone of the maize primary root under water stress. I. Increased levels are specific to the apical region of growth maintenance. Journal of Experimental Botany **64,** 1223–1233

Wang W, Vinocur B, Shoseyov O, Altman A. 2004. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends in Plant Science **9,** 244–252.

Westgate ME, Boyer JS. 1985. Osmotic adjustment and the inhibition of leaf, root, stem and silk growth at low water potentials in maize. Planta **164,** 540–549.

Wu Y, Cosgrove DJ. 2000. Adaptation of roots to low water potentials by changes in cell wall extensibility and cell wall proteins. Journal of Experimental Botany **51,** 1543–1553.

Wu Y, Sharp RE, Durachko DM, Cosgrove DJ. 1996. Growth maintenance of the maize primary root at low water potentials involves increases in cell-wall extension properties, expansin activity, and wall susceptibility to expansins. Plant Physiology **111,** 765–772.

Xia X-J, Zhou Y-H, Shi K, Zhou J, Foyer CH, Yu J-Q. 2015. Interplay between reactive oxygen species and hormones in the control of plant development and stress tolerance. Journal of Experimental Botany **66,** 2839–2856.

Xu Z-S, Chen M, Li L-C, Ma Y-Z. 2011. Functions and application of the AP2/ERF transcription factor family in crop improvement. Journal of Integrative Plant Biology **53,** 570–585.

Yamaguchi M, Sharp RE. 2010. Complexity and coordination of root growth at low water potentials: recent advances from transcriptomic and proteomic analyses. Plant, Cell and Environment **33,** 590–603.

Yang Z, Wang Y, Gao Y, Zhou Y, Zhang E, Hu Y, Yuan Y, Liang G, Xu C. 2014. Adaptive evolution and divergent expression of heat stress transcription factors in grasses. BMC Evolutionary Biology **14,** 147.

Yue G, Zhuang Y, Li Z, Sun L, Zhang J. 2008. Differential gene expression analysis of maize leaf at heading stage in response to water-deficit stress. Bioscience Reports **28,** 125–134.

Zhang L-Y, Bai M-Y, Wu J, *et al*. 2009. Antagonistic HLH/bHLH transcription factors mediate brassinosteroid regulation of cell elongation and plant development in rice and *Arabidopsis*. The Plant Cell **21,** 3767–3780.

Zheng J, Zhao J, Tao Y, *et al*. 2004. Isolation and analysis of water stress-induced genes in maize seedlings by subtractive PCR and cDNA macroarray. Plant Molecular Biology **55,** 807–823.

Zhu J, Alvarez S, Marsh EL, *et al.* 2007. Cell wall proteome in the maize primary root elongation zone. II. Region-specific changes in water soluble and lightly ionically bound proteins under water deficit. Plant Physiology **145,** 1533–1548.

Zhuang Y, Ren G, Yue G, Li Z, Qu X, Hou G, Zhu Y, Zhang J. 2007. Effects of water-deficit stress on the transcriptomes of developing immature ear and tassel in maize. Plant Cell Reports **26,** 2137–2147.

A High-Resolution Tissue-Specific Proteome and Phosphoproteome Atlas of Maize Primary Roots Reveals Functional Gradients along the Root Axes[1][OPEN]

Caroline Marcon, Waqas Ahmed Malik, Justin W. Walley[2], Zhouxin Shen, Anja Paschold, Laurie G. Smith, Hans-Peter Piepho, Steven P. Briggs*, and Frank Hochholdinger*

Institute of Crop Science and Resource Conservation, Crop Functional Genomics, University of Bonn, 53113 Bonn, Germany (C.M., A.P., F.H.); Institute of Crop Science, Biostatistics Unit, University of Hohenheim, 70599 Stuttgart, Germany (W.A.M., H.-P.P.); and Section of Cell and Developmental Biology, University of California-San Diego, La Jolla, California 92093 (J.W.W., Z.S., L.G.S., S.P.B.)

ORCID IDs: 0000-0001-6699-9441 (C.M.); 0000-0001-7553-2237 (J.W.W.); 0000-0002-8294-1589 (L.G.S.); 0000-0001-7813-2992 (H.-P.P.); 0000-0002-7226-8618 (S.P.B.); 0000-0002-5155-0884 (F.H.).

Downloaded from https://academic.oup.com/plphys/article/168/1/233/6113647 by ulbbonn user on 21 December 2021

A high-resolution proteome and phosphoproteome atlas of four maize (*Zea mays*) primary root tissues, the cortex, stele, meristematic zone, and elongation zone, was generated. High-performance liquid chromatography coupled with tandem mass spectrometry identified 11,552 distinct nonmodified and 2,852 phosphorylated proteins across the four root tissues. Two gradients reflecting the abundance of functional protein classes along the longitudinal root axis were observed. While the classes RNA, DNA, and protein peaked in the meristematic zone, cell wall, lipid metabolism, stress, transport, and secondary metabolism culminated in the differentiation zone. Functional specialization of tissues is underscored by six of 10 cortex-specific proteins involved in flavonoid biosynthesis. Comparison of this data set with high-resolution seed and leaf proteome studies revealed 13% (1,504/11,552) root-specific proteins. While only 23% of the 1,504 root-specific proteins accumulated in all four root tissues, 61% of all 11,552 identified proteins accumulated in all four root tissues. This suggests a much higher degree of tissue-specific functionalization of root-specific proteins. In summary, these data illustrate the remarkable plasticity of the proteomic landscape of maize primary roots and thus provide a starting point for gaining a better understanding of their tissue-specific functions.

The maize (*Zea mays*) root system consists of embryonic primary and seminal roots and postembryonic lateral and shoot-borne roots (Hochholdinger, 2009). The primary root is the first organ that emerges after germination, providing water and nutrients for the growing seedling. The longitudinal structure of maize roots is characterized by a root cap at the terminal end, a subterminal meristematic zone, followed by zones in which newly formed cells elongate and differentiate (Ishikawa and Evans, 1995). The differentiation zone, in which functionally distinct cell types are formed, can be tracked by the presence of epidermal root hairs. Hence, roots represent a gradient of cell differentiation along the longitudinal axis: young and undifferentiated cells are located at the distal end near the root tip, whereas differentiated cells are located toward the proximal end of the root (Ishikawa and Evans, 1995).

Radially, maize roots can be divided into the stele and the surrounding cortical parenchyma. The stele comprises vascular and ground tissue, which is enclosed by the single-layered pericycle as the outermost boundary. The multilayered cortical parenchyma comprises, from the center to the periphery, a single ring of endodermis cells, a multilayered cortical parenchyma, and single files of exodermis and epidermis cells connecting the root to the rhizosphere (Hochholdinger, 2009). The epidermis of the differentiation zone is densely populated by tubular root hairs, which are instrumental for the uptake of nutrients that are either transported into the shoot or metabolized in the cortical parenchyma (Marschner, 2011).

[1] This work was supported by the Deutsche Forschungsgemeinschaft (grant no. HO2249/9–1 to F.H. and grant no. PI 377/12–1 to H.-P.P.) and by the National Science Foundation (grant no. 0924023 to S.P.B.).

[2] Present address: Department of Plant Pathology and Microbiology, 351 Bessey Hall, Iowa State University, Ames, IA 50011.

* Address correspondence to sbriggs@ucsd.edu and hochholdinger@uni-bonn.de.

C.M., F.H., and S.P.B. designed the research; C.M. performed the experiment, conducted downstream analysis, drafted the article, and designed the figures; C.M. and L.G.S. sampled the tissues; W.A.M. and H.-P.P. performed the statistical analysis; Z.S. performed the mass spectrometric analysis; J.W.W. organized the raw data; A.P. provided transcriptome data and was involved in the correlation analysis (mRNA versus proteins); J.W.W., L.G.S., A.P., and F.H. participated in data interpretation; F.H. and S.P.B. conceived and coordinated the study and drafted the article.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Frank Hochholdinger (hochholdinger@uni-bonn.de).

[OPEN] Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.15.00138

 © 2015 American Society of Plant Biologists. All Rights Reserved.

In recent years, soluble proteomes of whole maize roots were analyzed with respect to development (Hochholdinger et al., 2005), genotypic variation (Hochholdinger et al., 2004; Wen et al., 2005; Liu et al., 2006; Sauer et al., 2006; Hoecker et al., 2008), and environmental interactions (Chang et al., 2000; Li et al., 2007).

A major limitation of analyzing whole roots is their composite nature, being made up of distinct tissues. Each tissue provides unique protein accumulation patterns that have the potential to be masked when whole roots are analyzed. This constraint can be overcome by tissue-specific analyses. Thus far, only a few surveys have analyzed distinct tissues such as the primary root tip (Chang et al., 2000), the elongation zone (Zhu et al., 2006, 2007), and cortical parenchyma and stele (Saleem et al., 2010). The mechanical separation of maize root cortical parenchyma (hereafter referred to as cortex) and stele made the most abundant soluble proteins of these tissues accessible (Saleem et al., 2010). Combining two-dimensional gel electrophoresis and electrospray tandem mass spectrometry (MS/MS) identified 59 proteins preferentially expressed in the cortex and 11 proteins predominantly accumulated in the stele (Saleem et al., 2010). Among these proteins, a β-glucosidase functioning in the release of free bioactive cytokinin (Brzobohatý et al., 1993) was preferentially accumulated in the cortical parenchyma (Saleem et al., 2010). That study gave first insights into tissue-specific protein accumulation in the differentiation zone of maize roots. All these initial studies were based on the combination of two-dimensional SDS-PAGE with subsequent mass spectrometric analyses of selected proteins and thus were of limited resolution and depth.

More recently, gel-free shotgun proteomics further enhanced the depth of proteomic profiles by quantitative liquid chromatography-MS/MS analyses (Zhu et al., 2010). This technique allowed for the identification of more than 12,000 proteins in a maize leaf developmental time-course experiment (Facette et al., 2013) and a similar number of proteins isolated from developing maize seeds (Walley et al., 2013). These studies provided novel insights into protein accumulation patterns in these maize organs during development, which are instrumental for the better understanding of these biological systems.

In this study, we used gel- and label-free high-performance liquid chromatography (HPLC)-MS/MS to quantify the abundance of 11,552 nonmodified proteins and 2,852 phosphoproteins in four distinct tissues of the maize primary root. Proteins unique to certain root tissues or differentially accumulated between root tissues were functionally annotated and analyzed in detail. Moreover, the maize root proteome was compared with maize leaf and seed proteomes of similar size to identify conserved and organ-specific protein accumulation patterns. Furthermore, tissue-specific accumulation patterns of proteins encoded by classical maize genes were surveyed to reveal tissue-specific functions. Finally, proteome and RNA sequencing (RNA-Seq) data of primary root tissues were compared to determine the correlation of these classes of biomolecules.

RESULTS

Tissue-Specific Accumulation of 11,552 Nonmodified Proteins and 2,852 Phosphoproteins from Four Maize Primary Root Tissues

Maize primary roots of 2 to 4 cm in length were separated into four distinct tissues along the longitudinal axis of the primary root. These tissues were the meristematic zone, including the root cap; the elongation zone; and the differentiation zone, which was subdivided radially into the cortical parenchyma, including all cell types between the endodermis and epidermis, and the stele, comprising all cell types enclosed by the pericycle (Fig. 1A). The elongation zone was distinguished from the differentiation zone by root hairs present only in epidermis cells of the differentiation zone.

Proteins were extracted in four biological replicates per tissue and digested by trypsin. The resulting peptides were either directly analyzed by HPLC-MS/MS or first enriched for phosphopeptides. More than 8 million identified mass spectra were used to quantify protein (Supplemental Data Set S1) and phosphoprotein (Supplemental Data Set S2) abundance by spectral counting. The resulting spectra were identified using the protein database ZmB73 RefGen_v2 5a working gene set, downloaded from www.gramene.org. In total, 11,552 unique nonmodified proteins (protein groups) were identified from 73,938 distinct peptides (Supplemental Data Sets S1 and S3; see "Materials and Methods"). Approximately 96% of the identified proteins (11,042/11,552) mapped to the filtered gene set, a subset of the working gene set that consists of 39,656 high-confidence gene models. The remaining 510 proteins (4%) mapped exclusively to the working gene set and were annotated as pseudogenes, transposons, or other low-confidence gene models (Supplemental Data Sets S1 and S3).

Phosphorylated proteins were identified after phosphopeptide enrichment using cerium oxide and subsequent HPLC-MS/MS analyses. In total, 11,226 phosphorylated amino acids in 9,480 distinct peptides were detected, matching to a parsimonious set of 2,852 phosphorylated proteins (Supplemental Data Set S2). For 7,292 residues assigned to 6,080 phosphopeptides, the phosphorylated amino acid was localized (Supplemental Data Set S4). Specific Ser (84%; 6,150), Thr (15%; 1,070), and Tyr (1%; 72) phosphorylations were assigned using the variable modification localization (VML) score in the Spectrum Mill software version 3.03 (Agilent Technologies).

Pearson correlation coefficients of protein abundance were calculated between the distinct tissues by using averaged normalized spectral counts (SPCs) of each tissue. Correlation coefficients for the four tissues

Downloaded from https://academic.oup.com/plphys/article/168/1/233/6113647 by ulbbonn user on 21 December 2021

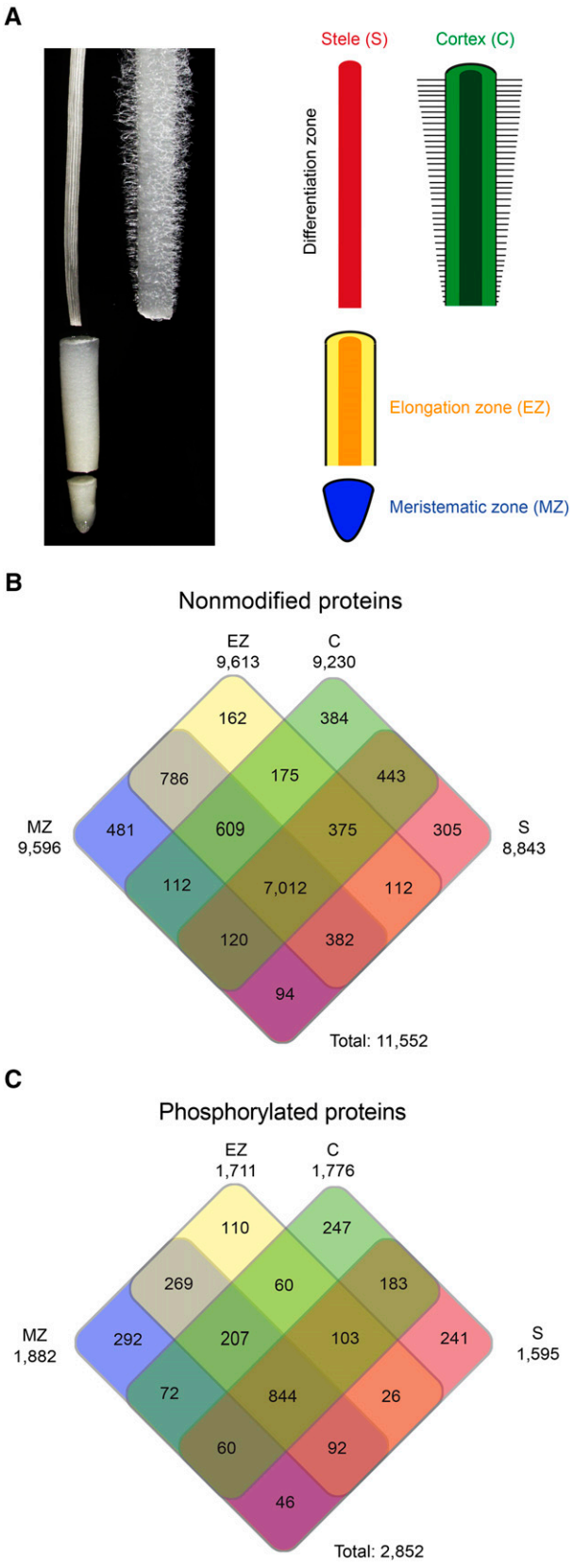


Figure 1. Tissues of young maize primary roots surveyed in this study and tissue-specific protein accumulation patterns. A, Maize primary root manually dissected into four tissues (left) and color-coded schematic of the four tissues (right): meristematic zone (MZ), elongation zone (EZ), cortex (C; covering all radial tissues between the endodermis and epidermis), and stele (S; covering all central cell types covered by the pericycle). B and C, Tissue-specific accumulation of nonmodified (B) and phosphorylated (C) proteins in the four maize primary root tissues.

were on average 0.94 for the nonmodified and 0.76 for the phosphorylated proteins (Table I). The abundance of nonmodified and phosphorylated proteins showed higher correlation between adjacent tissues (e.g. meristematic zone versus elongation zone; 0.96 for nonmodified proteome and 0.76 for phosphorylated proteome) than between distant root tissues (e.g. meristematic zone versus cortex; 0.84 for nonmodified proteome and 0.42 for phosphorylated proteome).

A comparison of biological replicate correlations using averaged normalized SPCs of the four biological replicates was conducted to assess reproducibility. On average, Pearson correlation coefficients of 0.94 and 0.65 were determined between the replicates of the nonmodified proteomes and the phosphorylated proteomes (Table I).

Analysis of the 11,552 nonmodified proteins revealed that 61% (7,012/11,552) of these proteins were expressed in all four tissues (Fig. 1B). The remaining 39% (4,540/11,552) of proteins were either accumulated in two or three tissues (28%; 3,208/11,552) or only in one tissue (11%; 1,332/11,552). Among the 1,332 proteins accumulated in a single tissue, 36% (481/1,332) were identified exclusively in the meristematic zone, 12% (162/1,332) in the elongation zone, 29% (384/1,332) in the cortex, and 23% (305/1,332) in the stele. In contrast, among 2,852 phosphorylated proteins, only 30% (844/2,852) were present in all four tissues, whereas 39% (1,118/2,852) were accumulated in two or three tissues and 31% (890/2,852) showed tissue-specific expression (Fig. 1B).

Functional Analysis of Tissue-Specific Nonmodified and Phosphorylated Proteins

Tissue-specific proteins of the nonmodified and phosphorylated data sets were functionally analyzed using agriGO (http://bioinfo.cau.edu.cn/agriGO/analysis.php). agriGO is based on Gene Ontology (GO) terms describing gene product species in three independent categories: biological process, molecular function, and cellular component (http://www.geneontology.org). To identify significantly overrepresented (false discovery rate [FDR] $< 5\%$) functional categories, a singular enrichment analysis (SEA) was performed. SEA compares each annotated protein class with all annotated expressed proteins in a tissue. Among 11,552 nonmodified proteins across all tissues, 8,017 (69%) were annotated by GO terms. Moreover, GO terms were available for 1,902 of 2,852 (67%) phosphorylated proteins identified in the different tissues. Overall, enrichment for 65 terms related to biological processes and 53 terms related to molecular functions were observed,

Downloaded from https://academic.oup.com/plphys/article/168/1/233/6113647 by ulbbonn user on 21 December 2021

Table I. *Pearson correlation coefficients between meristematic zone (MZ), elongation zone (EZ), cortex (C), and stele (S) and between the four replicates of each tissue (in boldface) of the nonmodified (first number in each column) and phosphorylated (second number in each column) protein data sets based on averaged normalized SPCs*

Tissue	MZ	EZ	C	S
MZ	**0.97/0.71**			
EZ	0.96/0.76	**0.96/0.66**		
C	0.84/0.42	0.91/0.61	**0.90/0.66**	
S	0.85/0.43	0.89/0.65	0.91/0.71	**0.95/0.60**

whereas only two terms reflecting cellular components were identified (Supplemental Table S1).

Among the 481 nonmodified proteins specific for the meristematic zone (Fig. 1B), 279 were annotated, and 35 GO terms displayed overrepresentation (Supplemental Table S1). The majority of these proteins are involved in biological processes, particularly in the regulation of cellular and metabolic processes. Among the 292 meristematic zone-specific phosphorylated proteins, 172 were annotated by agriGO, and only four GO terms, GO:0006139, GO:0006259, GO:0003677, and GO:0003676, which are involved in DNA binding and DNA metabolic processes, were overrepresented. These four terms were also overrepresented among the nonmodified meristematic zone-specific proteins.

Only 162 nonmodified proteins (101 annotated) and 110 phosphorylated proteins (73 annotated) displayed elongation zone-specific accumulation. SEA revealed no overrepresented GO term.

Among 384 nonmodified cortex-specific proteins (274 annotated), nine overrepresented GO terms were identified. Eight of the nine GO terms belonged to the category molecular function, predominantly reflecting proteins with catalytic activity (e.g. GO:0016747 and GO:0016491) or binding function (e.g. GO:0046906 and GO:0020037; Supplemental Table S1). No GO term showed overrepresentation among the 248 phosphorylated cortex-specific proteins (171 annotated).

In total, 18 GO terms were overrepresented among the 305 stele-specific nonmodified proteins (187 annotated). Four of these GO terms were predominantly involved in lignin metabolism (Supplemental Table S1), while two GO terms were assigned to the category cellular component (GO: 0048046 and GO:0005576). Almost 50% (59/120) of all overrepresented GO terms were detected among the 241 phosphorylated stele-specific proteins (101 annotated). These proteins show, among others, transferase and transporter activities. The GO term GO:0009055, referring to proteins with electron carrier activity, was overrepresented among nonmodified proteins of the cortex and stele. This indicates that there are at least some conserved functions in these two primary root tissues.

In summary, only one of the 120 overrepresented GO terms was detected in two different tissues, whereas 119 GO terms showed tissue-specific overrepresentation, indicating highly regulated tissue-specific functions of some proteins.

Differential Expression of Protein Accumulation between the Four Primary Root Tissues

Differentially accumulated nonmodified and phosphorylated proteins were identified based on the six possible pairwise combinations between the four tissues: (1) elongation zone versus meristematic zone, (2) cortex versus meristematic zone, (3) stele versus meristematic zone, (4) cortex versus elongation zone, (5) stele versus elongation zone, and (6) stele versus cortex. Among 11,552 nonmodified proteins, between 326 (meristematic zone versus elongation zone) and 3,530 (meristematic zone versus cortex) were differentially expressed between two tissues (Supplemental Table S2). After functional annotation, between 318 (meristematic zone versus elongation zone) and 3,460 (meristematic zone versus cortex) were differentially expressed (Fig. 2A). In all comparisons, the elongation zone preferentially accumulated more differential proteins compared with the other zones, followed by the meristematic zone, cortex, and stele, leading to following protein accumulation trend: elongation zone > meristematic zone > cortex > stele.

Among the 2,852 phosphorylated proteins, between 46 (meristematic zone versus elongation zone and cortex versus stele) and 919 (meristematic zone versus cortex) displayed differential accumulation (Supplemental Table S2), resulting in 44 (meristematic zone versus elongation zone) and 886 (meristematic zone versus cortex) differentially accumulated proteins after functional MapMan annotation (Supplemental Fig. S1A). Here, the following trend of protein abundance was observed: cortex > elongation zone > stele > meristematic zone. In comparisons 1 (meristematic zone versus elongation zone) and 6 (cortex versus stele), only a small number of differentially accumulated proteins were detected in both data sets, which might be the consequence of the close relation of the meristematic zone with the elongation zone and the cortex with the stele.

MapMan Annotation of Differentially Accumulated Root Tissue Proteins

The MapMan bin classification system was used to further analyze the function of the differentially accumulated nonmodified and phosphorylated proteins (Fig. 2A; Supplemental Fig. S1A) and to compare the distribution of up- and down-regulated proteins between tissues. For each comparison, the distribution of up- and down-regulated proteins across all MapMan bins was used as the expected value for further analyses. In a second step, the distribution of up- and down-regulated proteins in each of the 35 functional classes (MapMan bins) was compared with the expected value using Fisher's exact test (two tailed). In eight classes, the distribution of up- and down-regulated proteins differed significantly from the expected value for at least four of six tissue comparisons (Fig. 2A), including cell wall, DNA, lipid metabolism, protein, RNA, secondary metabolism, stress, and

Downloaded from https://academic.oup.com/plphys/article/168/1/233/6113647 by ulbbonn user on 21 December 2021

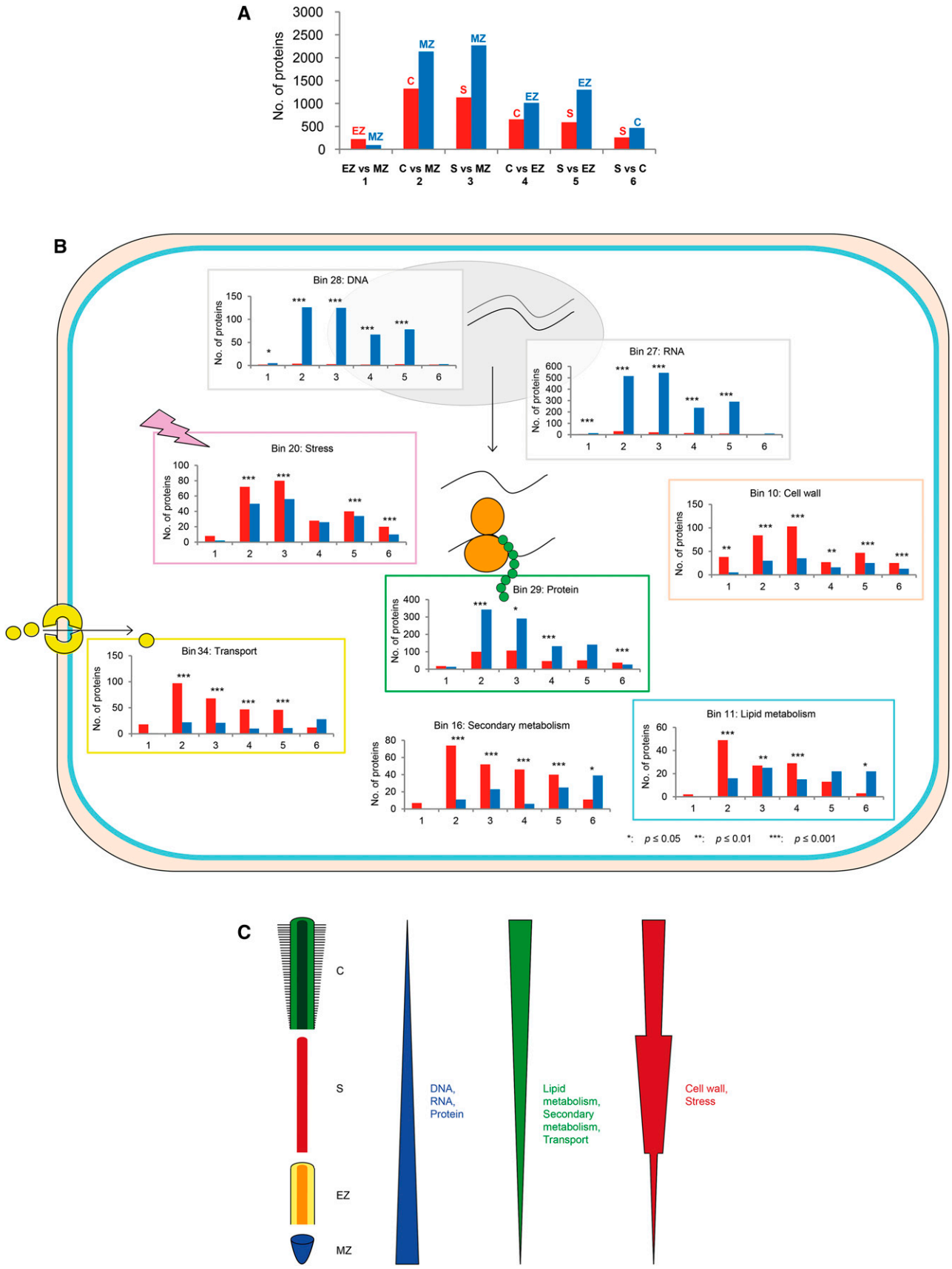


Figure 2. Number of up- and down-regulated nonmodified proteins in six pairwise tissue comparisons. A, Total number of up- and down-regulated proteins mapped to the 35 MapMan bins in the six comparisons. Blue and red bars in each comparison indicate the number of genes preferentially expressed in the tissue indicated on top of the bars. B, Functional classes in which

Downloaded from https://academic.oup.com/plphys/article/168/1/233/6113647 by ulbbonn user on 21 December 2021

transport (Fig. 2B; Supplemental Fig. S2). In the functional classes DNA, RNA, and protein, very similar distributions of protein up- and down-regulation were observed, resulting in the following trend of protein abundance: meristematic zone > elongation zone > stele > cortex, with the highest accumulation in the meristematic zone and gradually lower abundance in the elongation zone, the stele, and the cortex (Fig. 2C). Exactly the opposite protein accumulation trend was identified for the functional classes lipid metabolism, secondary metabolism, and transport: cortex > stele > elongation zone > meristematic zone, with the highest abundance in the cortex and the lowest abundance in the meristematic zone. Up- and down-regulated proteins among the six pairwise comparisons involved in cell wall and stress showed an expression trend of stele > cortex > elongation zone > meristematic zone (Fig. 2C).

The same analysis was conducted for the six sets of differentially accumulated phosphorylated proteins (Supplemental Figs. S1A and 2). Again, the assignment of all of these proteins to 35 functional classes was used as the expected distribution of up- and down-regulated proteins for subsequent analyses. Afterward, the bin-specific distribution of phosphorylated up- and down-regulated proteins was compared with the expected value. This comparison revealed numerical changes in the ratio of up- and down-regulated proteins in seven MapMan bins for at least one of six tissue comparisons (Supplemental Fig. S1B). For the differentially accumulated phosphorylated proteins in the classes DNA and RNA, the same expression trend as for nonmodified proteins was detected: meristematic zone > elongation zone > stele > cortex (Supplemental Fig. S1C). In contrast, proteins of the classes cell wall, major carbohydrate metabolism, protein, signaling, and transport showed a different expression trend: stele > cortex > elongation zone > meristematic zone (Supplemental Fig. S1C).

Classical Maize Genes among the Nonmodified Proteins

Classical maize genes are genes that were initially identified by visible mutant phenotypes before being cloned and characterized (Schnable and Freeling, 2011). To further characterize the tissue-specific nonmodified proteins, they were compared with the list of 468 classical maize genes (https://genomevolution.org/wiki/index.php/Classical_Maize_Genes; Schnable and Freeling, 2011). Among the nonmodified proteins, 242 proteins encoded by classical maize genes were detected (Fig. 3A; Supplemental Table S3). Similar numbers of classical maize proteins, ranging from 206 (stele) to 218 (meristematic zone), were identified in the distinct tissues. Among the nonmodified proteins, between two (elongation zone and stele) and 12 (meristematic zone) proteins encoded by genes of the classical maize gene list were present in only one tissue (Fig. 3A). Among the 10 cortex-specific proteins encoded by classical maize genes, six enzymes catalyze the initial steps of flavonoid biosynthesis, the synthesis of reduced glutathione (GSH)-anthocyanidin-glucoside from phenylpropanoids to GSH-anthocyanidin-glucoside (Fig. 3B).

Correlation of Protein and Transcript Abundance

Global correlations between mRNA and protein levels in the four tissues were surveyed by comparing the nonmodified proteome data set with RNA-Seq transcript profiling data from our laboratory (Paschold et al., 2014). For the meristematic zone, 7,092 accessions present in the transcriptome and proteome data set were identified, whereas 6,436 accessions for the elongation zone, 5,850 accessions for the cortex, and 5,542 accessions for the stele were determined (Supplemental Data Set S5). Pearson correlations revealed mRNA-protein correlation coefficients of 0.5 in the meristematic zone, 0.43 in the elongation zone, 0.38 in the cortex, and 0.38 in the stele (Table II). Spearman rank correlation coefficients were slightly higher for the meristematic zone (0.56), elongation zone (0.45), and 0.41 (stele) and similar for the cortex (0.38). In summary, Pearson and Spearman rank correlation coefficients displayed moderate positive correlation of RNA and protein abundance in the distinct primary root tissues. Tissue-specific comparison of transcripts and proteins demonstrated that, in all four tissues, a transcript was detected in 93% of proteins. Remarkably, for 7% of proteins, no transcript was found in the corresponding tissue (Supplemental Fig. S3).

Comparison of Nonmodified Proteins between Roots, Seeds, and Leaves

Proteome data sets of seed and leaf tissues of the maize inbred line B73 were generated previously by the same analysis pipeline, revealing 12,453 and 12,032 nonmodified proteins from seeds (Walley et al., 2013) and leaves (Facette et al., 2013), respectively. Comparison of these protein sets with the 11,552 nonmodified proteins identified in primary root tissues in this study allowed for the identification of common and tissue-specific proteins. In total, 7,902 proteins,

Figure 2. (*Continued.*)
the ratio of up- and down-regulated proteins differed significantly from the expected value (compare with A) in at least four of six pairwise tissue comparisons. $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$. C, Tissue-specific protein accumulation trends based on the results displayed in B. C, Cortex; EZ, elongation zone; MZ, meristematic zone; S, stele.

Downloaded from https://academic.oup.com/plphys/article/168/1/233/6113647 by ulbbonn user on 21 December 2021

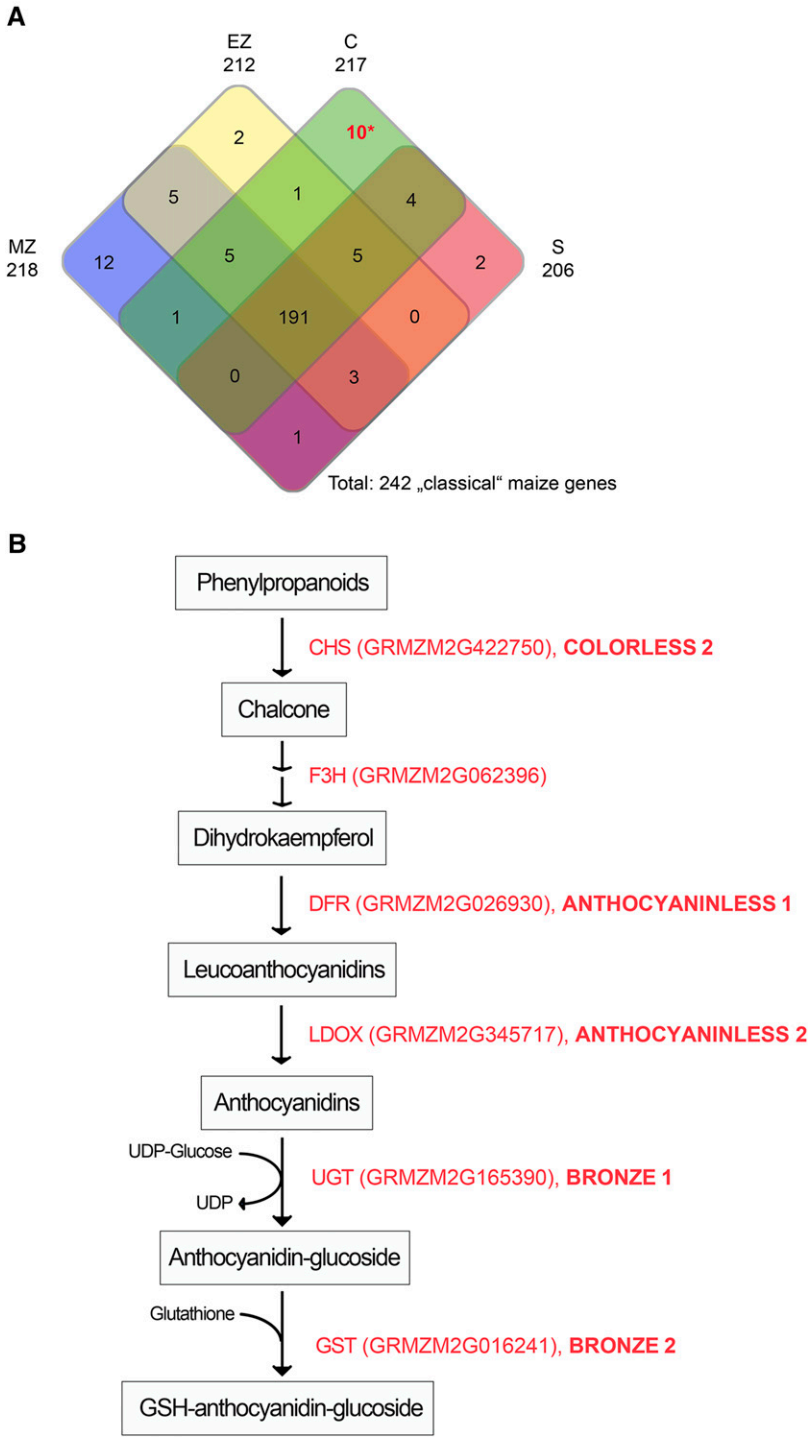


Figure 3. Maize proteins encoded by classical maize genes (i.e. genes that were first identified based on their mutant phenotypes). A, Tissue-specific accumulation of classical maize proteins identified among nonmodified proteins. *, Six of 10 cortex-specific proteins highlighted in red catalyze the initial steps of flavonoid biosynthesis. C, Cortex; EZ, elongation zone; MZ, meristematic zone; S, stele. B, Schematic of the initial steps of flavonoid biosynthesis including six of the 10 cortex-specific (asterisk in A) classical maize proteins (red): CHS, CHALCONE SYNTHASE; F3H, FLAVONONE 3-HYDROXYLASE; DFR, DIHYDROFLAVONOL-4-REDUCTASE; LDOX, LEUCOANTHOCYANIDIN DIOXYGENASE; UGT, UDP-GLUCOSYLTRANFERASE; GST, GLUTATHIONE *S*-TRANSFERASE. Classical gene names are highlighted in boldface.

Downloaded from https://academic.oup.com/plphys/article/168/1/233/6113647 by ulbbonn user on 21 December 2021

which correspond to 68% of all root proteins identified in this study, accumulated in all three maize organs (Fig. 4A). Organ-specific accumulation of nonmodified proteins was detected for 1,504 accessions in roots (13% of root proteins), 1,611 in leaf (13% of leaf proteins), and 2,110 in seed (20% of seed proteins). In summary, 16,680 different nonmodified proteins representing unique accessions were identified in roots, seeds, and leaves. Among 1,504 root-specific proteins, 342 proteins (23%) were detected in all four analyzed tissues (Fig. 4B), while 618 proteins (41%) accumulated in two or three tissues. Tissue specificity was observed for 544 proteins (36%), including 156 proteins in the meristematic zone, 68 in the elongation zone, 171 in the cortex, and 149 in the stele. agriGO SEA of the organ-specific proteins revealed 41 overrepresented GO

Table II. *Correlation between mRNA and protein levels within the four tissues: meristematic zone (MZ), elongation zone (EZ), cortex (C), and stele (S)*

***$P \leq 0.0001$.

Tissue	No. of Genes	Pearson Correlation	Spearman Rank Correlation
MZ	7,092	0.50***	0.56***
EZ	6,436	0.43***	0.45***
C	5,850	0.38***	0.38***
S	5,542	0.38***	0.41***

terms among 873 annotated proteins identified exclusively in roots. These GO terms are mainly involved in response to chemical stimulus, oxidative stress, and lignin metabolic processes (FDR < 5%; Supplemental Table S4). In the seed-specific data set, 1,201 GO terms were annotated and 19 were overrepresented, reflecting, among others, defense proteins and biosynthesis-regulating proteins. Among 36 overrepresented GO terms (out of 892 annotated GO terms) of the leaf-specific data sets, mainly proteins involved in photosynthesis were detected, which is in line with their role in converting light energy in carbohydrates (Supplemental Table S4).

DISCUSSION

Analysis of four maize primary root tissues revealed that, among 11,552 identified nonmodified proteins, 61% were expressed in all four tissues (Fig. 1B), compared with 28% that were accumulated in two or three tissues and 11% active in only one tissue. Similarly, 10% of the accumulated proteins in developing maize seeds exhibited accumulation in only one tissue (Walley et al., 2013), whereas in different regions of the maize leaf, only 4% of nonmodified proteins showed zone-specific accumulation (Facette et al., 2013). In contrast to the nonmodified proteins, only 30% of phosphorylated proteins were detected in all four tissues (844/2,852 phosphoproteins), whereas 39% were present in two or three tissues and 31% were present in only one tissue. Similar results were reported from differentiating leaf zones (Facette et al., 2013) and differentiating human embryonic stem cells (Rigbolt et al., 2011), where the phosphoproteome displayed more dynamic changes than in the nonmodified proteome.

The identification of peptide phosphorylation sites in this study revealed the relative abundance of Ser (84%), Thr (15%), and Tyr (1%) phosphorylation. Recent studies showed similar distributions of Ser (86%), Thr (13%), and Tyr (1%) in maize leaves and seeds (Facette et al., 2013; Walley et al., 2013). This allowed for the reconstruction of a kinases-substrate regulatory network in maize seeds (Walley et al., 2013) and identified previously unknown phosphorylation sites in cell wall-related proteins of developing maize leaves (Facette et al., 2013). In general, in eukaryotes, Ser and Thr residues are predominantly phosphorylated, whereas phosphorylation of Tyr residues is less common (de la Fuente van Bentem and Hirt, 2009). This was observed not only in maize but also in Arabidopsis (*Arabidopsis thaliana*; Ser [85%], Thr [11%], and Tyr [4%]; Sugiyama et al., 2008) and human embryonic stem cells (Ser [80%], Thr [17%], and Tyr [3%]; Rigbolt et al., 2011).

Functional analysis of nonmodified proteins and phosphoproteins accumulated in only one tissue was based on the GO analysis toolkit (Du et al., 2010). In parallel, differentially accumulated proteins between six pairwise comparisons were functionally characterized using the MapMan bin classification system (Thimm et al., 2004). Among the nonmodified and phosphorylated proteins accumulated exclusively in

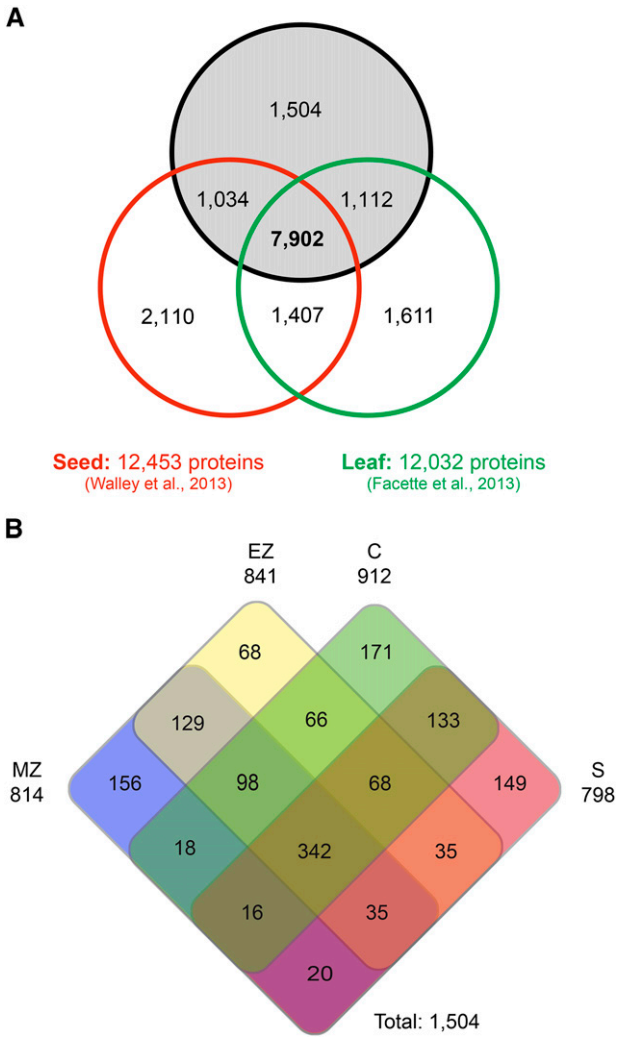


Figure 4. A, Organ-specific accumulation of nonmodified proteins detected in primary root, seed, and leaf. B, Protein accumulation pattern of the 1,504 primary root-specific proteins in distinct primary root tissues. C, Cortex; EZ, elongation zone; MZ, meristematic zone; S, stele.

Downloaded from https://academic.oup.com/plphys/article/168/1/233/6113647 by ulbbonn user on 21 December 2021

the meristematic zone, the GO terms DNA binding and DNA metabolic processes were overrepresented. In line with that, proteins belonging to the MapMan classes DNA, RNA, and protein were up-regulated in the meristematic zone and decreased gradually in more differentiated tissues, underscoring the high mitotic activity of the meristematic zone (Ishikawa and Evans, 1995; Verbelen et al., 2006). The increased rate of mitosis in the meristematic zone was underscored by overrepresented GTP-binding proteins (G proteins; bin 30.5), cell division (bin 31.2), and cell cycle proteins (bin 31.3; Supplemental Fig. S4). G proteins are regulators for diverse signaling pathways in plants (e.g. flower, leaf, and fruit development and seed germination; for review, see Perfus-Barbeoch et al., 2004). Several studies showed that G proteins also modulate cell division in Arabidopsis roots (Ullah et al., 2003; Chen et al., 2006) and rice (*Oryza sativa*; Ashikari et al., 1999). Therefore, an overrepresentation of G proteins might lead to an increase of cell cycle proteins, resulting in higher cell division rates in the meristematic zone.

In contrast, differentially accumulated nonmodified proteins related to cell wall, lipid metabolism, secondary metabolism, stress, and transport were higher expressed in the differentiated cortex and stele versus the undifferentiated meristematic and elongation zones (Fig. 2C). In line with the differentiation status of the stele, several GO terms involved in lignin biosynthesis (related to the bins cell wall and secondary metabolism) were overrepresented in this tissue (Supplemental Table S1). Lignin is deposited in the xylem elements of the stele to provide structural support (Vermerris et al., 2010; Zhang et al., 2014). In line with the overrepresentation of lignin biosynthesis-related GO terms, six laccases (GRMZM2G146152_P01, GRMZM2G164467_P01, GRMZM5G814718_P01, GRMZM2G367668_P01, GRMZM2G388587_P01, and GRMZM2G305526_P01) were accumulated exclusively in the stele (Supplemental Data Sets S1 and S3). Laccases are multicopper-containing glykoproteins that are involved in lignin polymerization in plants (for review, see O'Malley et al., 1993). In a variety of plant species, laccases are expressed in lignified cells, as for instance in the vasculature of Arabidopsis (Zhao et al., 2013), the xylem of loblolly pine (*Pinus taeda*; Bao et al., 1993), and maize roots (Liang et al., 2006). Furthermore, stele-specific phosphoproteins related to the GO term transmembrane transporter activity were enriched, suggesting that transporters are more extensively regulated by phosphorylation. Accordingly, the MapMan analysis revealed an up-regulation of phosphorylated transport proteins (bin 34) in cortex and stele (Supplemental Fig. S1). Among such transport proteins, 50% of all ATP-binding cassette (ABC) transporters (11 of 22) identified in the phosphorylated data set preferentially accumulated in cortex or stele (Supplemental Fig. S5). ABC transporters are transmembrane proteins enabling the transport of hormones, lipids, metals, and other molecules (for review, see Higgins, 1992) via ATP hydrolysis at the ABC, representing a nucleotide-binding domain (Higgins and Linton, 2004). ABC transporters, at least in humans, are regulated by phosphorylation via protein kinases, which was recently reviewed (Stolarczyk et al., 2011). Among the stele-specific phosphoproteins, GO terms related to protein kinases were also enriched (Supplemental Table S1), which could be in line with the high ABC transporter activity and its activation by kinase-dependent phosphorylation in this tissue (Stolarczyk et al., 2011).

In a previous study based on two-dimensional gel electrophoresis, 70 differentially accumulated protein spots, representing 56 different GenBank accessions, were identified between cortex and stele of 2.5-d-old maize primary roots (Saleem et al., 2010). Only 18 of the 56 unique proteins (Supplemental Data Set S3) were also present in the B73 RefGen_v2 5a working gene set (www.gramene.org), while the remaining 38 proteins were identified based on homology with proteins of other plant species. Among these 18 cortex-specific maize proteins (Saleem et al., 2010), 15 (83%) were preferentially accumulated in the cortex versus the stele in the present data set. In the study by Saleem et al. (2010), preferential accumulation of a β-glucosidase (GRMZM2G016890; GenBank accession no. P49235) enzyme in the cortex was reported. This enzyme functions in the release of free bioactive cytokinin (Brzobohatý et al., 1993). This finding was underscored and extended by this study, as predominant accumulation of not only one but all 12 β-glucosidase enzyme accessions identified in this study were preferentially accumulated in the cortex (Supplemental Data Set S3). Furthermore, four enzymes involved in ammonium assimilation showed cortex-specific expression in the study by Saleem et al. (2010). Those authors suggested that the maize primary root cortex is the site of ammonium assimilation in young seedlings before exogenous nitrogen is assimilated from the soil. Similarly, in this study, most enzymes related to ammonium assimilation showed preferential accumulation in the cortex (Supplemental Data Set S3): two Glu dehydrogenases showed cortex-specific expression, among five Gln synthetases four revealed this trend, three of five Asp aminotransferases showed preferential accumulation in cortex compared with stele, and finally, among four Asn synthetases two were cortex specific.

To gain further functional insights, the tissue-specific nonmodified proteins were compared with the list of classical maize genes that were initially identified by their mutant phenotypes (Schnable and Freeling, 2011). This survey identified six cortex-specific proteins that are involved in the initial steps of flavonoid biosynthesis by catalyzing the synthesis of GSH-anthocyanidin-glucoside (Fig. 3B). Flavonoids belong to a large family of polyphenolic secondary metabolites, which are involved in a wide range of biological processes such as signaling in plant-microbe interactions and protection against ultraviolet light (Koes et al., 1994). Flavonoids accumulate in most

Downloaded from https://academic.oup.com/plphys/article/168/1/233/6113647 by ulbbonn user on 21 December 2021

plant tissues, often in the epidermis of plant organs, to protect subjacent tissues from radiation damage (Saslowsky and Winkel-Shirley, 2001). In contrast to aboveground tissue, where flavonoid accumulation can easily be traced by the coloring of flowers or seeds (Koes et al., 1994), there is little knowledge about flavonoid function in maize root tissues. Accumulation of flavonoids has been demonstrated in the endodermis, cortex, and epidermis of white clover (*Trifolium repens*; Hassan and Mathesius, 2012), which is in line with the accumulation of enzymes of flavonoid biosynthesis, as demonstrated here. ABC transporters, discussed above, are able to transport flavonoid glycosides such as GSH-anthocyanidin-glucoside to the vacuole (Goodman et al., 2004) or to export them through the plasma membrane for long-distance transport (Buer et al., 2007).

The analysis of three different maize organs, including root tissues surveyed in this study, seeds (Walley et al., 2013), and leaves (Facette et al., 2013), revealed between 1,504 and 2,110 organ-specific proteins (Fig. 4A) and nearly 8,000 proteins accumulated in diverse organs. SEA identified more general overrepresented GO terms, such as protein folding or cellular amine metabolic process, among the overlapping proteins (data not shown), whereas more specific overrepresented GO terms were identified among the organ-specific proteins (Supplemental Table S4). In line with their function, several GO terms related to lignin and cell wall biosynthesis were overrepresented among the root-specific proteins. In the endodermis and around xylem vessels of roots, the cell wall polymer lignin is deposited to a high extent to provide mechanical support for the plant body (Vermerris et al., 2010). In young developing roots, cell wall biosynthesis and, therefore, also lignification might be increased (e.g. because of rapid cell division). Among the leaf-specific proteins, many GO terms related to photosynthesis were overrepresented. This is related to the major function of leaves, which is the photosynthetic conversion of light energy into chemical energy in the form of carbohydrates. Compared with the relatively low number (11%) of tissue-specific nonmodified proteins identified among all 11,552 proteins identified in our data set (Fig. 1B), more than three times as many proteins (36%) were detected to be tissue specific among the 1,504 root-specific proteins (Fig. 4B). This suggests that root-specific proteins have more tissue-specific functions compared with general proteins, which are also expressed in nonroot tissues.

Traditionally, mRNA abundance measurements were used to infer protein abundance (for review, see Vogel and Marcotte, 2012). However, a growing body of research has found only moderate correlation between transcript and protein abundance (Gygi et al., 1999; Schwanhäusser et al., 2011; Petricka et al., 2012; Vogel and Marcotte 2012; Walley et al., 2013; Ponnala et al., 2014). In this study, Pearson and Spearman rank correlation coefficients revealed moderate positive relationships between transcript and protein abundance in primary root tissues (Table II). Both coefficients were used in previous studies to calculate mRNA-protein relationships (for review, see Maier et al., 2009; Vogel and Marcotte, 2012). In general, the relationship between mRNA and protein abundance is considered positive in a wide range of organisms but varies greatly between the studies (yeast [Lee et al., 2011], bacteria [Maier et al., 2011], maize endosperm and embryos [Walley et al., 2013], and maize leaf sections [Ponnala et al., 2014]). Variations in mRNA-protein correlation might be explained by many factors, such as bias in RNA and protein stability and movement, experimental measurement, and quantification methods (Ponnala et al., 2014). In addition, one transcript must not necessarily encode for only one protein but can give rise to many different proteins by alternative splicing or a wide range of protein modifications. These variants might have different half-lives; therefore, the observed protein stabilities reflect average stabilities of multiple isoforms rather than one-to-one protein-transcript relationships. Furthermore, the correlation of transcript and protein might vary as a result of translational mechanisms regulating and fine-tuning the translation of transcripts, such as by changes in the relative synthesis rate (Kristensen et al., 2013) or adjusting ribosome density (Liu et al., 2013). In this study, between 5,543 (stele) and 7,092 (meristematic zone) gene accessions were detected in both the transcriptome and proteome analyses (Table II). In all four tissues, 93% of all identified proteins were also found in the RNA-Seq data set. Remarkably, for 7% of proteins, no corresponding transcript was identified.

A similar observation was made previously in a correlation analysis of maize seeds, where many abundant proteins were not supported by a corresponding mRNA (Walley et al., 2013). The lack of transcripts for detected proteins can have different reasons. For instance, while proteins can be very stable, their corresponding transcript levels can oscillate because of circadian regulation of gene expression (Khan et al., 2010; Walley et al., 2013). In our data set, we identified three proteins where the corresponding genes are known to be regulated by the circadian rhythm (Khan et al., 2010): GRMZM2G055795 (elongation zone), GRMZM2G083841 (elongation zone and stele), and GRMZM2G097457 (cortical parenchyma). As a consequence, we only identified proteins but no transcripts of these accessions. Similarly, in maize endosperm and embryo, circadian-regulated proteins displayed a higher abundance than their transcripts (Walley et al., 2013). Another reason for the lack of transcripts is that proteins move into another tissue after translation. There are several examples in our data set that support this notion. For instance, a heat shock protein (GRMZM5G833699) was present in all tissues, whereas its transcript was detected only in the elongation zone, cortical parenchyma, and stele but not in the meristematic zone (Supplemental Data Set S3; Paschold et al., 2014). Similarly, a SUCROSE SYNTHASE2 (GRMZM2G318780)

Downloaded from https://academic.oup.com/plphys/article/168/1/233/6113647 by ulbbonn user on 21 December 2021

protein was detected in all four tissues, whereas the corresponding mRNA was lacking in the elongation zone. Similarly, a short-chain dehydrogenase/reductase (GRMZM2G066840) protein was present in all four tissues, whereas transcripts were detected only in the meristematic zone, elongation zone, and cortical parenchyma. In a recent study comparing endosperm versus whole seed proteomes and transcriptomes of maize, 15 proteins were detected in both tissues, while the RNA was detected in only one (Walley et al., 2013).

In summary, our data set represents a high-resolution proteome map of four distinct maize primary root tissues, revealing the functional diversity of the distinct root zones. Therefore, this data set will be a helpful resource and starting point for future in-depth analyses of metabolic functions of root tissues.

MATERIALS AND METHODS

Plant Material

Seeds of the maize (*Zea mays*) inbred line B73 were germinated in paper rolls as described previously (Hetz et al., 1996). Seedlings were grown for 5 d in the greenhouse without supplemental lighting. Then, primary roots of 2 to 4 cm in length were dissected into four distinct tissues (Fig. 1). The first 2 mm of the root tip, containing the root cap and the subterminal meristematic zone, was collected as the meristematic zone. Moreover, the proximal zone adjacent to the root tip up to the region where the root hairs emerge was sampled as the elongation zone. Finally, the differentiation zone, starting with the root hair zone up to the coleorhiza, was mechanically separated into cortex and stele according to Saleem et al. (2010). For each tissue, 200 roots were pooled for each of the four biological replicates. All samples were immediately frozen in liquid nitrogen and stored at −80°C until protein extraction.

Protein Extraction and Phosphopeptide Enrichment

Proteins were isolated from 2 g of frozen tissue per biological replicate. After grinding the tissue in frozen nitrogen, proteins were precipitated in methanol containing 0.2 mM Na_3VO_4 and washed three times in methanol followed by three washing steps in acetone. Precipitation and washing of proteins were performed at −20°C. Protein pellets were dried in a vacuum concentrator at 4°C. Subsequently, protein pellets were resuspended in 1 mL of extraction buffer (0.1% [w/v] SDS, 1 mM EDTA, and 50 mM HEPES buffer, pH 7). Cys residues were reduced using 1 mM Tris(2-carboxyethyl)phosphine at 95°C for 5 min and subsequently alkylated with 2.5 mM iodoacetamide at 37°C in the dark for 15 min. Proteins were digested overnight using trypsin (Roche; enzyme:substrate = 1:100, w/w). A second tryptic digestion (enzyme:substrate = 1:100, w/w) was performed the next day for 4 h. Digested peptides were purified on a Waters Oasis MCX cartridge to remove SDS. Peptides were eluted from the MCX column with 1 mL of 50% (v/v) isopropyl alcohol and 400 mM NH_4HCO_3 (pH 9.5) and then dried in a vacuum concentrator at 4°C. For proteome profiling of nonmodified proteins, peptides were resuspended in 1% (v/v) formic acid to a final pH of 3 and used for mass spectrometric analyses. For phosphoproteome profiling, peptides were resuspended in 3% (v/v) trifluoroacetic acid to a final pH of 1 and then subjected to phosphopeptide enrichment.

For phosphopeptide enrichment, 1% (w/v) colloidal CeO_2 was added to the acidified peptide solution at 1:10 (w/w). CeO_2 with captured phosphopeptides was briefly vortexed and then centrifuged at 1,000*g* for 1 min. After removing the supernatant, the CeO_2 pellet was washed with 1 mL of 1% (v/v) trifluoroacetic acid. By adding eluting buffer [200 mM $(NH_4)_2HPO_4$, 2 M $NH_3 \cdot H_2O$, and 10 mM EDTA, pH 9.5; identical volume as the added 1% [w/v] colloidal CeO_2], phosphopeptides were eluted.

Mass Spectrometry and Peptide Identification

Mass spectrometry (MS) and peptide identification were performed as described previously (Facette et al., 2013; Walley et al., 2013). In brief, an 1100 HPLC system (Agilent Technologies) delivered a flow rate of 600 nL min^{-1} to a three-phase capillary chromatography column through a splitter. A custom pressure cell was used to pack 5 μm of Zorbax SB-C18 (Agilent Technologies) into fused silica capillary tubing (250-μm i.d., 360-μm o.d., 30 cm long) to form the first dimension reverse-phase (RP1) column. A 5-cm-long strong cation-exchange (SCX) column packed with 5 μm of polysulfoethyl (PolyLC) was connected to the RP1 column using a zero dead volume 1-μm filter (Upchurch; M548) attached to the exit of the RP1 column. As an analytical column, the fused silica capillary (200-μm i.d., 360-μm o.d., 20 cm in length) packed with 5 μm of Zorbax SB-C18 (Agilent Technologies) was connected to the SCX column. The electrospray tip of the fused silica tubing was pulled to a sharp tip with the i.d. smaller than 1 μm using a P-2000 (Sutter Instrument) laser puller. Using the custom pressure cell, the peptide mixtures were loaded onto the RP1 column. For each liquid chromatography-MS/MS analysis, a new set of columns was used. First, the peptides were eluted from the RP1 column to the SCX column using a 0% to 80% acetonitrile gradient for 150 min. The peptides were then fractionated by the SCX column using a series of salt gradients (from 10 mM to 1 M ammonium acetate for 20 min each), followed by high-resolution reverse-phase separation using an acetonitrile gradient of 0% to 80% for 120 min.

Spectra were analyzed on the LTQ Velos linear ion trap tandem mass spectrometer (Thermo Electron), which employed automated, data-dependent acquisition. The mass spectrometer was operated in positive ion mode with a source temperature of 250°C. As a final fractionation step, gas phase separation in the ion trap was employed to separate the peptides into three mass classes prior to scanning (300–800, 800–1,100, and 1,100–2,000 D). Each MS scan was followed by five MS/MS scans of the most intense ions from the parent MS scan. A dynamic exclusion of 1 min was used to improve the duty cycle.

Database Search

Spectrum Mill version 3.03 (Agilent Technologies) was used to extract and search the raw data. MS/MS spectra with a sequence tag length of 1 or less were discarded. The remaining MS/MS spectra were searched against the maize B73 RefGen_v2 5a working gene set downloaded from www.gramene.org. The enzyme parameter was limited to full tryptic peptides, allowing a maximum miscleavage of 1. All other search parameters were set to Spectrum Mill's default settings (carbamidomethylation of Cys, ±2.5 D for precursor ions and ±0.7 D for fragment ions; and a minimum matched peak intensity of 50%). For the phosphoproteome data, oxidation of Met, N-terminal pyro-Gln, and phosphorylation on Ser, Thr, or Tyr were defined as variable modifications. A maximum of two modifications per peptide were used. A 1:1 concatenated forward:reverse database was constructed to calculate the FDR. The tryptic peptides in the reverse database were compared with the forward database and were shuffled if they matched to any tryptic peptides from the forward database. Peptide cutoff scores were dynamically assigned to each data set to maintain the FDR less than 0.1% at the peptide level. Maintaining a data set level FDR of 0.1% resulted in FDRs for the nonmodified proteome of 0.57 at the protein level (66 decoy hits among 11,618 protein group leaders) and 2.1 at the phosphoprotein level (61 decoy hits among 2,913 phosphoprotein group leaders). Phosphorylation sites were localized to a particular amino acid within a peptide using the VML score in Agilent's Spectrum Mill software (Chalkley and Clauser, 2012). A sequence VML score of greater than 0.5 defined a particular amino acid as phosphorylated, whereas a VML score of 0.05 or less defined an ambiguous residue. Proteins that share common peptides were grouped using principles of parsimony to address the protein database redundancy. Thus, proteins within the same group shared the same set or subset of unique peptides. A specific protein of each protein group was assigned as a group leader if it was presented by the highest number of unique peptide identifications or showed the highest value of protein coverage. Protein abundance and phosphorylation levels were quantified by spectral counting. Spectral counts for each protein represent the total number of peptide spectral matches to that protein (Liu et al., 2004; Huttlin et al., 2010; Walley et al., 2013). Biological replicates were normalized as described previously (Marcon et al., 2010, 2013). Briefly, raw SPCs of each biological replicate were normalized by assigning 500,000 SPCs to the sum of all proteins detected in the distinct run. The resulting normalization factor (assigned sum of SPCs divided by the sum of SPCs of each run) was then multiplied with each protein SPC of each run to allocate the assigned sum of 500,000 SPCs. Expressed proteins were defined by at least one SPC in one of the four biological replicates.

Downloaded from https://academic.oup.com/plphys/article/168/1/233/6113647 by ulbbonn user on 21 December 2021

GO Analyses

GO functional categories were assigned to tissue specifically accumulated proteins using the Web-based agriGO software (http://bioinfo.cau.edu.cn/agriGO/analysis.php). An SEA was used to compute enriched categories by comparing a list of tissue-specific proteins or tissue-specific phosphoproteins with all expressed proteins (11,552) or phosphoproteins (2,852), as described previously (Du et al., 2010). According to Benjamini and Yekutieli (2001), multiple comparison correction was performed and FDR was controlled at 5%. The SEACOMPARE tool allowed for combined cross comparisons of SEA results.

Functional Analysis of Differentially Accumulated Proteins

Raw data were normalized by the sum of the proteins detected in a sample and then converted to counts per million. A small value of 0.5 was added to avoid zero counts. Count data typically show nontrivial mean variance relationships. Raw counts show increasing variance with increasing count size, while log counts typically show a decreasing mean variance trend. We estimated the mean variance trend for log counts and then assigned a weight to each observation based on its predicted variance. The weights were then used in the linear modeling process to adjust for heteroscedasticity (Law et al., 2014).

To borrow strength across proteins in the estimation of the residual error variance, we used the empirical Bayes approach of Smyth (2004) implemented in the R package *limma*. Using Smyth's moderated *t* statistic makes use of all estimated sample variances, which are shrunk to a pooled estimate. Especially in cases where few replicates are available, inference is far more stable in comparison with the ordinary *t* statistic. We fitted the following linear model for each gene:

$$\log_2(y_{ij}) = \alpha_i + \beta_j + e_{ij}$$

where $\log_2(y_{ij})$ is the $\log_2$-transformed value of the normalized spectral read counts per million on protein level of the *i*th tissue in the *j*th sample. α_i is the fixed effect of the tissue ($i = 1, 2,\ldots,4$), β_j is the random sample effect ($j = 1, 2,\ldots,4$), and e_{ij} is the corresponding residual, assumed to be independently and identically distributed as $N(0,\sigma^2)$. *P* values over genes were then corrected for multiplicity using the FDR (Benjamini and Hochberg, 1995).

The MapMan bin classification system (Thimm et al., 2004) was used for functional annotations of proteins displaying differential accumulation in the following tissue comparisons: meristematic zone versus elongation zone, meristematic zone versus cortex, meristematic zone versus stele, elongation zone versus cortex, elongation zone versus stele, and cortex versus stele. The MapMan functional annotation file Zm_B73_5b_FGS_cds_2012 (http://mapman.gabipd.org) was used for the analysis.

Correlation Analysis

To assess the accuracy of the biological replicates, Pearson correlation coefficients of averaged normalized SPCs of the four biological replicates per tissue were calculated as described previously (Walley et al., 2013). To determine the relationship of mRNA and protein, RNA-Seq data from our laboratory (Paschold et al., 2014) were compared with the nonmodified protein data set of this study. The RNA-Seq data set was generated using the same tissues, in the same genotype (B73) at the same developmental stage, and mapped to the ZmB73 FGS_5B_filtered gene set version 2 (www.gramene.org). For global correlation analyses of transcript and protein abundance, the R package was used to calculate Pearson correlation and Spearman rank correlation of normalized $\log_2$-transformed data. To be included in the correlation analysis, we required that both the mRNA and protein were detected and that the mRNA and protein were present in at least three biological replicates. Pearson correlation measures a linear correlation between two variables, which is more appropriate if the variables have a jointly normal distribution, whereas the Spearman rank correlation quantifies monotonic correlation between two variables, which is less restrictive than Pearson correlation (Ponnala et al., 2014). Such a monotonic correlation could be the more appropriate assumption for our correlation analysis, since mRNA-protein abundance relationships may be nonlinear (de Sousa Abreu et al., 2009; Maier et al., 2009).

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Number of up- and down-regulated phosphorylated proteins resulting from six pairwise comparisons of the four maize primary root tissues.

Supplemental Figure S2. Distribution of differentially accumulated nonmodified and phosphorylated proteins between six tissue pairwise comparisons into 35 distinct MapMan bins.

Supplemental Figure S3. Number of gene accessions identified among nonmodified proteins and transcripts.

Supplemental Figure S4. Distribution of differentially accumulated nonmodified proteins in six pairwise comparisons of the four primary root tissues into distinct MapMan bins.

Supplemental Figure S5. Preferential accumulation of 11 ABC transporters in the cortex and stele of primary roots.

Supplemental Table S1. Enriched GO terms of tissue-specific proteins of nonmodified and phosphorylated protein data sets using SEA.

Supplemental Table S2. Number of differentially accumulated proteins (FDR = 5%) in the nonmodified and phosphorylated protein data set.

Supplemental Table S3. Classical maize protein identifications and their tissue-specific accumulation identified among nonmodified proteins.

Supplemental Table S4. A SEA of proteins being organ-specifically accumulated among nonmodified proteins of maize.

Supplemental Data Set S1. Raw spectral counts of nonmodified proteins.

Supplemental Data Set S2. Raw spectral counts of phosphorylated proteins.

Supplemental Data Set S3. Normalized spectral counts of nonmodified proteins.

Supplemental Data Set S4. Normalized spectral counts of phosphorylated proteins.

Supplemental Data Set S5. Raw and normalized reads and spectral counts for each gene which was observed in at least three biological replicates of the mRNA and nonmodified protein data set.

ACKNOWLEDGMENTS

We thank Michelle Facette (University of California, San Diego) for helpful discussions and Philipp Zimmermann (University of Bonn) for helping to organize parts of the "Supplemental Data."

Received January 29, 2015; accepted March 14, 2015; published March 16, 2015.

LITERATURE CITED

Ashikari M, Wu J, Yano M, Sasaki T, Yoshimura A (1999) Rice gibberellin-insensitive dwarf mutant gene *Dwarf 1* encodes the alpha-subunit of GTP-binding protein. Proc Natl Acad Sci USA **96:** 10284–10289

Bao W, O'Malley DM, Whetten R, Sederoff RR (1993) A laccase associated with lignification in loblolly pine xylem. Science **260:** 672–674

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B **57:** 289–300

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat **29:** 1165–1188

Brzobohatý B, Moore I, Kristoffersen P, Bako L, Campos N, Schell J, Palme K (1993) Release of active cytokinin by a beta-glucosidase localized to the maize root meristem. Science **262:** 1051–1054

Buer CS, Muday GK, Djordjevic MA (2007) Flavonoids are differentially taken up and transported long distances in Arabidopsis. Plant Physiol **145:** 478–490

Chalkley RJ, Clauser KR (2012) Modification site localization scoring: strategies and performance. Mol Cell Proteomics **11:** 3–14

Downloaded from https://academic.oup.com/plphys/article/168/1/233/6113647 by ulbbonn user on 21 December 2021

Chang WW, Huang L, Shen M, Webster C, Burlingame AL, Roberts JK (2000) Patterns of protein synthesis and tolerance of anoxia in root tips of maize seedlings acclimated to a low-oxygen environment, and identification of proteins by mass spectrometry. Plant Physiol **122:** 295–318

Chen JG, Gao Y, Jones AM (2006) Differential roles of Arabidopsis heterotrimeric G-protein subunits in modulating cell division in roots. Plant Physiol **141:** 887–897

de la Fuente van Bentem S, Hirt H (2009) Protein tyrosine phosphorylation in plants: more abundant than expected? Trends Plant Sci **14:** 71–76

de Sousa Abreu R, Penalva LO, Marcotte EM, Vogel C (2009) Global signatures of protein and mRNA expression levels. Mol Biosyst **5:** 1512–1526

Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res **38:** W64–W70

Facette MR, Shen Z, Björnsdóttir FR, Briggs SP, Smith LG (2013) Parallel proteomic and phosphoproteomic analyses of successive stages of maize leaf development. Plant Cell **25:** 2798–2812

Goodman CD, Casati P, Walbot V (2004) A multidrug resistance-associated protein involved in anthocyanin transport in *Zea mays*. Plant Cell **16:** 1812–1826

Gygi SP, Rochon Y, Franza BR, Aebersold R (1999) Correlation between protein and mRNA abundance in yeast. Mol Cell Biol **19:** 1720–1730

Hassan S, Mathesius U (2012) The role of flavonoids in root-rhizosphere signalling: opportunities and challenges for improving plant-microbe interactions. J Exp Bot **63:** 3429–3444

Hetz W, Hochholdinger F, Schwall M, Feix G (1996) Isolation and characterization of *rtcs* a mutant deficient in the formation of nodal roots. Plant J **10:** 845–857

Higgins CF (1992) ABC transporters: from microorganisms to man. Annu Rev Cell Biol **8:** 67–113

Higgins CF, Linton KJ (2004) The ATP switch model for ABC transporters. Nat Struct Mol Biol **11:** 918–926

Hochholdinger F (2009) The maize root system: morphology, anatomy and genetics. *In* J Bennetzen, S Hake, eds, Handbook of Maize. Springer, New York, pp 145–160

Hochholdinger F, Guo L, Schnable PS (2004) Lateral roots affect the proteome of the primary root of maize (*Zea mays* L.). Plant Mol Biol **56:** 397–412

Hochholdinger F, Woll K, Guo L, Schnable PS (2005) The accumulation of abundant soluble proteins changes early in the development of the primary roots of maize (*Zea mays* L.). Proteomics **5:** 4885–4893

Hoecker N, Keller B, Muthreich N, Chollet D, Descombes P, Piepho HP, Hochholdinger F (2008) Comparison of maize (*Zea mays* L.) F_1-hybrid and parental inbred line primary root transcriptomes suggests organ-specific patterns of nonadditive gene expression and conserved expression trends. Genetics **179:** 1275–1283

Huttlin EL, Jedrychowski MP, Elias JE, Goswami T, Rad R, Beausoleil SA, Villén J, Haas W, Sowa ME, Gygi SP (2010) A tissue-specific atlas of mouse protein phosphorylation and expression. Cell **143:** 1174–1189

Ishikawa H, Evans ML (1995) Specialized zones of development in roots. Plant Physiol **109:** 725–727

Khan S, Rowe SC, Harmon FG (2010) Coordination of the maize transcriptome by a conserved circadian clock. BMC Plant Biol **10:** 126

Koes RE, Quattrocchio F, Mol JNM (1994) The flavonoid biosynthetic pathway in plants: function and evolution. BioEssays **16:** 123–132

Kristensen AR, Gsponer J, Foster LJ (2013) Protein synthesis rate is the predominant regulator of protein expression during differentiation. Mol Syst Biol **9:** 689

Law CW, Chen Y, Shi W, Smyth GK (2014) voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol **15:** R29

Lee MV, Topper SE, Hubler SL, Hose J, Wenger CD, Coon JJ, Gasch AP (2011) A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. Mol Syst Biol **7:** 514

Li K, Xu C, Zhang K, Yang A, Zhang J (2007) Proteomic analysis of roots growth and metabolic changes under phosphorus deficit in maize (*Zea mays* L.) plants. Proteomics **7:** 1501–1512

Liang M, Haroldsen V, Cai X, Wu Y (2006) Expression of a putative laccase gene, ZmLAC1, in maize primary roots under stress. Plant Cell Environ **29:** 746–753

Liu H, Sadygov RG, Yates JR III (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem **76:** 4193–4201

Liu MJ, Wu SH, Wu JF, Lin WD, Wu YC, Tsai TY, Tsai HL, Wu SH (2013) Translational landscape of photomorphogenic *Arabidopsis*. Plant Cell **25:** 3699–3710

Liu Y, Lamkemeyer T, Jakob A, Mi G, Zhang F, Nordheim A, Hochholdinger F (2006) Comparative proteome analyses of maize (*Zea mays* L.) primary roots prior to lateral root initiation reveal differential protein expression in the lateral root initiation mutant *rum1*. Proteomics **6:** 4300–4308

Maier T, Güell M, Serrano L (2009) Correlation of mRNA and protein in complex biological samples. FEBS Lett **583:** 3966–3973

Maier T, Schmidt A, Güell M, Kühner S, Gavin AC, Aebersold R, Serrano L (2011) Quantification of mRNA and protein and integration with protein turnover in a bacterium. Mol Syst Biol **7:** 511

Marcon C, Lamkemeyer T, Malik WA, Ungrue D, Piepho HP, Hochholdinger F (2013) Heterosis-associated proteome analyses of maize (*Zea mays* L.) seminal roots by quantitative label-free LC-MS. J Proteomics **93:** 295–302

Marcon C, Schützenmeister A, Schütz W, Madlung J, Piepho HP, Hochholdinger F (2010) Nonadditive protein accumulation patterns in maize (*Zea mays* L.) hybrids during embryo development. J Proteome Res **9:** 6511–6522

Marschner P, editor (2011) Mineral Nutrition of Higher Plants. Academic Press, London

O'Malley DM, Whetten R, Bao W, Chen CL, Sederoff R (1993) The role of laccase in lignification. Plant J **4:** 751–757

Paschold A, Larson NB, Marcon C, Schnable JC, Yeh CT, Lanz C, Nettleton D, Piepho HP, Schnable PS, Hochholdinger F (2014) Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. Plant Cell **26:** 3939–3948

Perfus-Barbeoch L, Jones AM, Assmann SM (2004) Plant heterotrimeric G protein function: insights from Arabidopsis and rice mutants. Curr Opin Plant Biol **7:** 719–731

Petricka JJ, Schauer MA, Megraw M, Breakfield NW, Thompson JW, Georgiev S, Soderblom EJ, Ohler U, Moseley MA, Grossniklaus U, et al (2012) The protein expression landscape of the Arabidopsis root. Proc Natl Acad Sci USA **109:** 6811–6818

Ponnala L, Wang Y, Sun Q, van Wijk KJ (2014) Correlation of mRNA and protein abundance in the developing maize leaf. Plant J **78:** 424–440

Rigbolt KT, Prokhorova TA, Akimov V, Henningsen J, Johansen PT, Kratchmarova I, Kassem M, Mann M, Olsen JV, Blagoev B (2011) System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. Sci Signal **4:** rs3

Saleem M, Lamkemeyer T, Schützenmeister A, Madlung J, Sakai H, Piepho HP, Nordheim A, Hochholdinger F (2010) Specification of cortical parenchyma and stele of maize primary roots by asymmetric levels of auxin, cytokinin, and cytokinin-regulated proteins. Plant Physiol **152:** 4–18

Saslowsky D, Winkel-Shirley B (2001) Localization of flavonoid enzymes in Arabidopsis roots. Plant J **27:** 37–48

Sauer M, Jakob A, Nordheim A, Hochholdinger F (2006) Proteomic analysis of shoot-borne root initiation in maize (*Zea mays* L.). Proteomics **6:** 2530–2541

Schnable JC, Freeling M (2011) Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS ONE **6:** e17855

Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian gene expression control. Nature **473:** 337–342

Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol **3:** Article 3

Stolarczyk EI, Reiling CJ, Paumi CM (2011) Regulation of ABC transporter function via phosphorylation by protein kinases. Curr Pharm Biotechnol **12:** 621–635

Sugiyama N, Nakagami H, Mochida K, Daudi A, Tomita M, Shirasu K, Ishihama Y (2008) Large-scale phosphorylation mapping reveals the extent of tyrosine phosphorylation in Arabidopsis. Mol Syst Biol **4:** 193

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J **37:** 914–939

Ullah H, Chen JG, Temple B, Boyes DC, Alonso JM, Davis KR, Ecker JR, Jones AM (2003) The β-subunit of the *Arabidopsis* G protein negatively regulates auxin-induced cell division and affects multiple developmental processes. Plant Cell **15:** 393–409

Downloaded from https://academic.oup.com/plphys/article/168/1/233/6113647 by ulbbonn user on 21 December 2021

Verbelen JP, De Cnodder T, Le J, Vissenberg K, Baluska F (2006) The root apex of *Arabidopsis thaliana* consists of four distinct zones of growth activities: meristematic zone, transition zone, fast elongation zone and growth terminating zone. Plant Signal Behav **1:** 296–304

Vermerris W, Sherman DM, McIntyre LM (2010) Phenotypic plasticity in cell walls of maize brown midrib mutants is limited by lignin composition. J Exp Bot **61:** 2479–2490

Vogel C, Marcotte EM (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet **13:** 227–232

Walley JW, Shen Z, Sartor R, Wu KJ, Osborn J, Smith LG, Briggs SP (2013) Reconstruction of protein networks from an atlas of maize seed proteotypes. Proc Natl Acad Sci USA **110:** E4808–E4817

Wen TJ, Hochholdinger F, Sauer M, Bruce W, Schnable PS (2005) The *roothairless1* gene of maize encodes a homolog of *sec3*, which is involved in polar exocytosis. Plant Physiol **138:** 1637–1643

Zhang Y, Paschold A, Marcon C, Liu S, Tai H, Nestler J, Yeh CT, Opitz N, Lanz C, Schnable PS, et al (2014) The Aux/IAA gene *rum1* involved in seminal and lateral root formation controls vascular patterning in maize (*Zea mays* L.) primary roots. J Exp Bot **65:** 4919–4930

Zhao Q, Nakashima J, Chen F, Yin Y, Fu C, Yun J, Shao H, Wang X, Wang ZY, Dixon RA (2013) Laccase is necessary and nonredundant with peroxidase for lignin polymerization during vascular development in *Arabidopsis*. Plant Cell **25:** 3976–3987

Zhu J, Alvarez S, Marsh EL, Lenoble ME, Cho IJ, Sivaguru M, Chen S, Nguyen HT, Wu Y, Schachtman DP, et al (2007) Cell wall proteome in the maize primary root elongation zone: II. Region-specific changes in water soluble and lightly ionically bound proteins under water deficit. Plant Physiol **145:** 1533–1548

Zhu J, Chen S, Alvarez S, Asirvatham VS, Schachtman DP, Wu Y, Sharp RE (2006) Cell wall proteome in the maize primary root elongation zone. I. Extraction and identification of water-soluble and lightly ionically bound proteins. Plant Physiol **140:** 311–325

Zhu W, Smith JW, Huang CM (2010) Mass spectrometry-based label-free quantitative proteomics. J Biomed Biotechnol **2010:** 840518

Downloaded from https://academic.oup.com/plphys/article/168/1/233/6113647 by ulbbonn user on 21 December 2021

The Plant Cell, Vol. 26: 3939–3948, October 2014, www.plantcell.org © 2014 American Society of Plant Biologists. All rights reserved.

Nonsyntenic Genes Drive Highly Dynamic Complementation of Gene Expression in Maize Hybrids[W]

Anja Paschold,[a] Nick B. Larson,[b,1] Caroline Marcon,[a] James C. Schnable,[c] Cheng-Ting Yeh,[d] Christa Lanz,[e] Dan Nettleton,[b] Hans-Peter Piepho,[f] Patrick S. Schnable,[d] and Frank Hochholdinger[a,2]

[a] Institute of Crop Science and Resource Conservation, Crop Functional Genomics, University of Bonn, 53113 Bonn, Germany
[b] Department of Statistics, Iowa State University, Ames, Iowa 50011-1210
[c] Department of Agronomy and Horticulture, University of Nebraska, Lincoln, Nebraska 68588
[d] Department of Agronomy and Center for Plant Genomics, Iowa State University, Ames, Iowa 50011-3650
[e] Department of Molecular Biology, Max-Planck-Institute for Developmental Biology, 72076 Tuebingen, Germany
[f] Institute for Crop Science, Biostatistics Unit, University of Hohenheim, 70599 Stuttgart, Germany

ORCID IDs: 0000-0001-6465-2882 (A.P.); 0000-0002-5155-0884 (F.H.)

Maize (*Zea mays*) displays an exceptional level of structural genomic diversity, which is likely unique among higher eukaryotes. In this study, we surveyed how the genetic divergence of two maize inbred lines affects the transcriptomic landscape in four different primary root tissues of their F1-hybrid progeny. An extreme instance of complementation was frequently observed: genes that were expressed in only one parent but in both reciprocal hybrids. This single-parent expression (SPE) pattern was detected for 2341 genes with up to 1287 SPE patterns per tissue. As a consequence, the number of active genes in hybrids exceeded that of their parents in each tissue by >400. SPE patterns are highly dynamic, as illustrated by their excessive degree of tissue specificity (80%). The biological significance of this type of complementation is underpinned by the observation that a disproportionally high number of SPE genes (75 to 82%) is nonsyntenic, as opposed to all expressed genes (36%). These genes likely evolved after the last whole-genome duplication and are therefore younger than the syntenic genes. In summary, SPE genes shape the remarkable gene expression plasticity between root tissues and complementation in maize hybrids, resulting in a tissue-specific increase of active genes in F1-hybrids compared with their inbred parents.

INTRODUCTION

Maize (*Zea mays*) is a highly polymorphic species with an extraordinary level of intraspecific genomic diversity compared with other eukaryotes (Springer et al., 2009; Swanson-Wagner et al., 2010). This diversity is mediated by single nucleotide polymorphisms (SNPs), insertion-deletion polymorphisms (INDELs), and structural variations. Major genotyping efforts determined exceptionally high frequencies of SNPs and INDELs (~1 per 80 bp and 1 per 300 bp genomic sequence, respectively) (Bi et al., 2006; Barbazuk et al., 2007) between the commonly used inbred lines B73 and Mo17. While these polymorphisms represent small genomic alterations, structural variations such as copy number variations, presence-absence variations (PAVs), and chromosomal rearrangements refer to segmental alterations of DNA >1 kb (Lai et al., 2010; Olsen and Wendel, 2013). Between the inbred lines B73 and Mo17, thousands of structural variants have been reported (Springer et al., 2009; Swanson-Wagner et al., 2010). In addition, inbred line-specific differences in content of repetitive DNA (Kato et al., 2004) and of total genome size (Laurie and Bennett, 1985) have been detected further emphasizing the high degree of intraspecific genomic divergence. To determine how genomic variation translates to quantitative gene expression differences, transcriptome surveys of various maize inbred lines have been conducted (Hansey et al., 2012). This study reported not only extensive sequence variation in annotated gene models between 21 inbred lines but also defined the variable part of the maize transcriptome as opposed to the conserved section. Recent studies support the concept of the pan-transcriptome consisting of a core and a dispensable subtranscriptome in maize (Hirsch et al., 2014; Marroni et al., 2014), which further illustrates the extensive intraspecific variation.

As demonstrated by syntenic comparison to the genomes of other grass species, the lineage leading to maize experienced a whole genome duplication 5 to 12 million years ago (Blanc and Wolfe, 2004; Swigonová et al., 2004), which created two distinct subgenomes each of which originally contained a complete set of genes and regulatory sequences (Schnable et al., 2011). In many cases, one of the two gene copies was lost from the genome in a process known as fractionation; however 3000 to 5000 gene pairs are retained in the modern maize genome (Schnable et al., 2011). Therefore, the gene complement of maize can be divided into three categories: pairs of duplicate genes shared by both subgenomes, single-copy genes present in only one subgenome, and genes that cannot be assigned to a subgenome. The third category is characterized by a lack of syntenic orthologs in the genomes of other grass species, and most of these nonsyntenic genes are created by single gene duplication mechanisms, likely after the maize whole-genome duplication (Woodhouse et al., 2010).

[1] Current address: Mayo Clinic, Department of Health Sciences Research, Division of Biomedical Statistics and Informatics, Rochester 55905, MN.
[2] Address correspondence to hochholdinger@uni-bonn.de.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: Frank Hochholdinger (hochholdinger@uni-bonn.de).
[W] Online version contains Web-only data.
www.plantcell.org/cgi/doi/10.1105/tpc.114.130948

Downloaded from https://academic.oup.com/plcell/article/26/10/3939/6101583 by ULB Bonn user on 21 December 2021

The structural genomic diversity of the maize inbred lines B73 and Mo17 is also reflected on the transcriptome level. In maize roots, hundreds of genes were identified that were exclusively expressed in one parental inbred line but not in the second parent, whereas all these genes were expressed in both reciprocal hybrids, a phenomenon termed single-parent expression (SPE; Paschold et al., 2012). As a result, hybrids have hundreds more active genes than their parental inbred lines (Paschold et al., 2012).

In longitudinal orientation, roots are structurally divided into the root cap at the terminal end, a subterminal meristematic zone, followed by the elongation and differentiation zones (Figure 1; Ishikawa and Evans, 1995). Cells formed by mitotic divisions in the meristematic zone are dislocated proximally into the elongation zone, where they start to elongate. The elongation zone partly overlaps with the meristematic zone because cell elongation begins in the meristematic zone (Ishikawa and Evans, 1995). Finally, root cells transit into the differentiation zone where the various cell types acquire their final functions. Hence, the cells along the longitudinal axis of a root represent a gradient of cell differentiation, with very young undifferentiated cells at the distal end near the root tip and differentiated cells toward the proximal end of the root. The differentiation zone can be spotted by the presence of epidermal root hairs (Ishikawa and Evans, 1995).

In transverse orientation, the differentiation zone displays a number of functionally diverse cell types that can be distinguished by their anatomical features. The stele with the pericycle as its outermost cell layer is connected to the multilayered cortical parenchyma via a ring of endodermis cells. In maize, the stele contains differentiated xylem vessels that function in water and nutrient transport and primary phloem elements that function in the transport of photosynthates. These elements of the vasculature are embedded in parenchymal pith tissue (Hochholdinger, 2009). The cortical parenchyma that surrounds the stele consists of the endodermis, multiple layers of cortex tissue, and a single epidermal layer that connects the root to the rhizosphere (Hochholdinger, 2009). The epidermis, which is densely populated by tubular root hairs, takes up nutrients, which are then either carried via the cortex and xylem into the shoot or metabolized in the cortex of the root (Marschner, 2011). Stele and cortical parenchyma (hereafter referred to as cortex) can be separated mechanically without damaging pericycle and endodermis cells allowing subsequent analyses of these functionally diverse tissues (Saleem et al., 2009).

In this study, we globally surveyed the transcriptomes of four functionally distinct tissues of young maize primary roots in the inbred lines B73 and Mo17 and their reciprocal F1-hybrid progeny via RNA-seq. These experiments revealed dynamic tissue-specific patterns of SPE, which preferentially affected nonsyntenic genes.

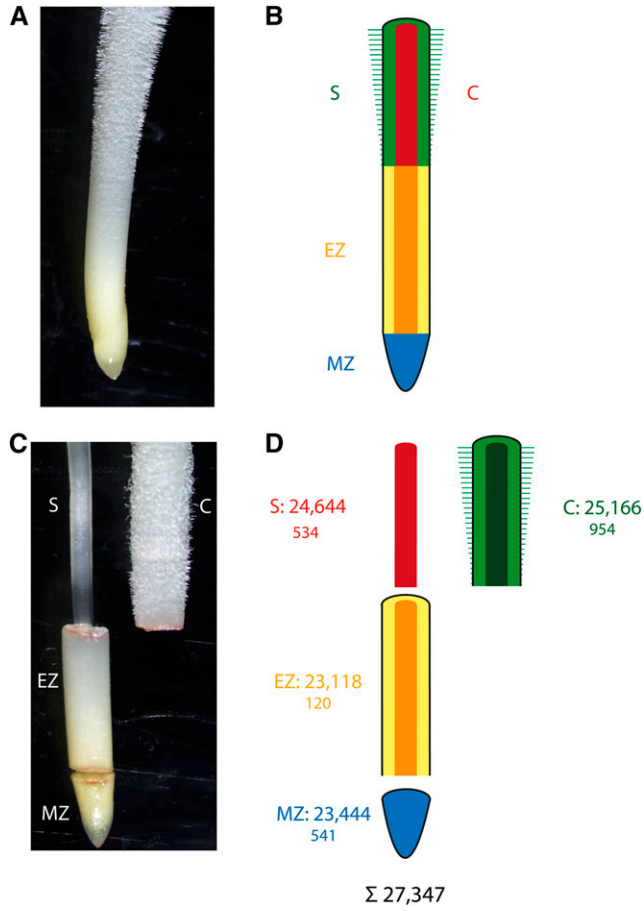


Figure 1. Tissues of Young Maize Primary Roots Surveyed in This Study.

(A) Primary root prior to tissue sampling.
(B) Color-coded schematic of the different tissues of a maize primary root.
(C) Manually dissected primary root with separated meristematic zone, elongation zone, cortex and stele.
(D) Schematic of separated primary root tissues. Upper numbers indicate the total number of genes expressed in a tissue, while the smaller lower numbers indicate the number of genes that were exclusively expressed in a tissue (in any of the analyzed genotypes). In total, 27,347 genes were expressed in at least one tissue. Red (S), stele; green (C), cortex; yellow/orange (EZ), elongation zone; blue (MZ), meristematic zone.

RESULTS

RNA-seq Profiling of Four Root Tissues of Maize Inbred Lines and Their Hybrid Offspring

In longitudinal orientation, maize primary roots display specialized zones of development, including the meristematic zone, the elongation zone and, as parts of the differentiation zone, stele and cortex (Figures 1A and 1B), which can be mechanically separated (Figures 1C and 1D). The transcriptomes of these four distinct tissues were surveyed in 2- to 4-cm primary roots of the maize inbred lines B73 and Mo17 and their reciprocal hybrids B73xMo17 and Mo17xB73. Each tissue per genotype was analyzed in four biological replicates (Supplemental Figure 1). RNA-seq of those 64 samples resulted on average in 9 million 100-bp reads per sample (Supplemental Table 1). After quality trimming and removal of duplicate reads, 79 to 81% of all sequences mapped to the maize reference genome (ZmB73_RefGen_v2; Supplemental Table 2). For each of the four genotypes, 96% of the reads mapped to the filtered gene set of maize (ZmB73 FGS_5B_FGSv2; Supplemental Table 2), which comprises 39,656 high-confidence gene models.

Downloaded from https://academic.oup.com/plcell/article/26/10/3939/6101583 by ULB Bonn user on 21 December 2021

In total, 27,347 of the 39,656 high-confidence genes (69%) were expressed in at least one tissue of at least one genotype (Figure 1D). Among those, 23,444 genes (86%) were expressed in the meristematic zone, 23,118 (85%) in the elongation zone, 25,166 (92%) in the cortex, and 24,644 genes (90%) in the stele (Figure 1D). Moreover, 2149 genes were exclusively expressed in only one of the four analyzed tissues (Figure 1D). While 120 genes representing 0.5% of all expressed genes in that tissue (120/23,118) were only expressed in the elongation zone, several hundred genes were exclusively expressed in cortex (954/25,166 ≙ 3.8%), stele (534/24,644 ≙ 2.2%) and meristematic zone (541/23,444 ≙ 2.3%).

Tissue-Specific Expression Complementation by SPE

SPE is an extreme instance of expression complementation in which genes are expressed in only one parent but in both reciprocal hybrid progeny. To assay for SPE in the four root tissues, activity of all genes in all four genotypes [B73 (B), Mo17 (M), B73xMo17 (BxM), Mo17xB73 (MxB)] was analyzed by a Bayesian data-augmented Markov chain Monte Carlo approach (see Methods). The results were summarized in four-way Venn diagrams (Figure 2). In each of the four root tissues, the vast majority of genes (94 to 95%) were expressed in all four genotypes (B_M_BxM_MxB). However, most of the genes that were not expressed in all four genotypes displayed SPE (Figure 2, square fields encircled by a bold black line). SPE_B genes (B_BxM_MxB) are expressed in B73 and in the reciprocal hybrids, but not in Mo17, whereas SPE_M genes (M_BxM_MxB) are expressed in Mo17 and the reciprocal hybrids, but not in B73. SPE was detected in each of the four tissues and the number of SPE genes was similar between tissues and ranged between 1037 (elongation zone) and 1287 (stele; Figure 2). Four genes switched their SPE pattern between tissues (Supplemental Table 3) and were not considered for further calculations because it is difficult to determine if they have escaped our error rate correction or if they indeed represent rare examples where SPE has switched between different tissues. This resulted in 1096 genes showing SPE_B and 1245 genes showing SPE_M in at least one tissue. As a consequence of complementation, both hybrids displayed in all four tissues >440 active genes more than their inbred parents (Figure 2).

To determine how many SPE_B genes are the result of PAV, i.e., these genes are not expressed in Mo17 because they are absent from that genotype, SPE_B genes were compared with a set of genes present in the B73 and absent in the Mo17 genome (Paschold et al., 2012). Overall, 4.4% (48/1096) of the SPE_B genes represent genomic PAVs (Supplemental Data Set 1).

Tissue-Specific Dynamics of SPE

To determine their tissue-specific dynamics, SPE patterns were summarized in four Venn diagrams (Figure 3). This analysis revealed that 20% (218/1096) of the SPE_B and 20% (250/1245) of the SPE_M genes display this pattern constitutively in all four analyzed tissues (Figures 3A and 3B). By contrast, 80% (1877/2341) of all SPE patterns were tissue specific, i.e., not observed in all four tissues. Among these, 67% (1225/1877) were detected in only one of the four analyzed tissues (Figures 3A and 3B, groups C, MZ, EZ, and S). For instance, 184 SPE_B and 210 SPE_M genes showed this particular expression pattern only in the cortex of young primary roots, while in the elongation zone, meristematic zone and stele these genes did not display SPE (Figures 3A and 3B, left corners). To determine if tissue-specific SPE is the result of exclusive expression of these genes in these tissues, the number of SPE genes that are only expressed in the tissues where they display SPE was determined (Figure 3, smaller numbers in the Venn diagrams). Together with the 468 SPE genes that displayed SPE in all four tissues, a total of 55% (1294/2341) of all SPE genes displayed SPE in all tissues in which they were expressed. However, the remaining 45% (1051/2341) of SPE genes displayed non-SPE expression patterns in some or all of the remaining tissues. In summary, while 55% of all SPE genes exhibited SPE in all tissues in which they were expressed, 45% of all SPE genes displayed complex tissue-specific regulation of SPE and non-SPE patterns.

Allele-Specific Contribution to SPE

To better understand the parental contribution to the SPE patterns, the allele-specific expression of SPE genes was determined in hybrids. The aim of this analysis was to determine if parental alleles that are not expressed in the inbred line are "reactivated" in the hybrid. SPE_B genes, which are known to be PAVs (Supplemental Table 3), were excluded from this analysis. Hence, for the remaining genes, both parental alleles are present in the hybrid genome. For this analysis >4 million SNPs previously called between B73 and Mo17 were used (Paschold et al., 2012). These SNPs allowed distinction between the expressed parental alleles of 2049 SPE genes. Allele-specific read mapping resulted in up to 197 tissue-specific SPE genes in which the "silent" allele was reactivated in the hybrid (Table 1). In a more conservative approach, an arbitrary threshold of 10 allele-specific reads per gene across the four replicates was introduced. Applying this cutoff reduced the number of SPE genes that expressed both alleles in the hybrids to between 0 and 4 (Table 1). This result indicates that most "silent" alleles remain inactive in hybrids if putative mapping artifacts are neglected.

Evolutionary Implications of Tissue-Specific SPE Patterns

To survey the evolutionary origin of the genes that display tissue-specific SPE, their assignment to maize subgenomes 1 or 2 was determined. For each of these analyses, comparisons with all 39,656 genes of the filtered gene set and with all 27,347 genes expressed in this analysis were performed. In total, 19,365 of all 39,656 high confidence FGS genes (49%) can be assigned to one of the two maize subgenomes (www.skraelingmountain.com/datasets/maize_indexed_by_subgenome.csv), which are the result of an ancient genome duplication (Table 2). Among all expressed genes 64% (17,402/ 27,347) were assigned to one of the two subgenomes pointing to their ancient origin. The extent to which this percentage exceeds the percentage of all FGS genes that can be assigned to a subgenome (19,365/39,656 [49%]; Fisher's exact test, $P < 0.001$) suggests that evolutionary older (syntenic) genes are expressed more frequently than nonsyntenic genes, which likely emerged after the last whole-genome duplication of maize. By contrast, only 18 to 25% of the SPE genes identified in the individual tissues were assigned to a subgenome, which is 2.6 to 3.6 times less than expected (Table 2). This pattern was statistically

Downloaded from https://academic.oup.com/plcell/article/26/10/3939/6101583 by ULB Bonn user on 21 December 2021

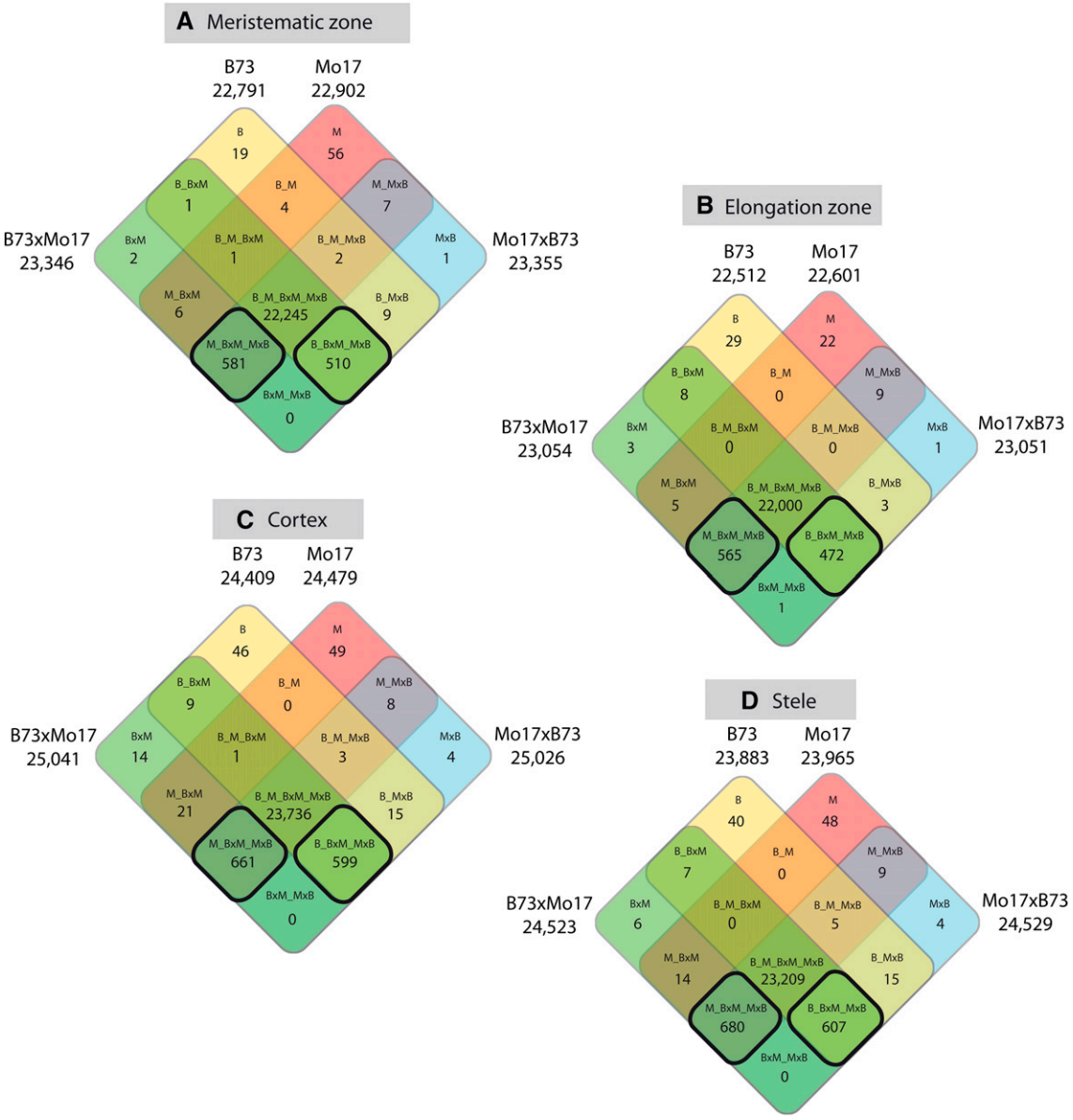


Figure 2. Genotype-Specific Expression Patterns in Four Different Maize Primary Root Tissues.

Transcriptomic expression complementation by SPE is dominant among the genes that are not expressed in all four genotypes (encircled by a bold black line). SPE genes are expressed in only one parental line but both hybrids. Meristematic zone **(A)**, elongation zone **(B)**, cortex **(C)**, and stele **(D)**. The different genotypes and number of genes of the various expression classes are indicated in each square field of the four-way Venn diagrams. The total number of expressed genes per genotype and tissue are indicated at the border of each Venn diagram. Note that in each tissue, hybrids express more genes than their parental inbred lines.

significant (Fisher's exact test, $P < 0.001$) for the SPE genes of each of the four tissues. Among the SPE genes that were assigned to subgenomes, the ratio of genes assigned to subgenome 1 (62%) and 2 (38%) was nearly identical to the ratio among all filtered gene set genes assigned to a subgenome and to the ratio among all expressed genes (Table 2).

In a previous analysis, 1124 genes that displayed a SPE pattern were identified from whole primary roots (WPRs) of the same genotypes (Paschold et al., 2012). Reanalysis of this data set using the improved algorithm introduced in this study (see Methods) resulted in 28,113 expressed genes and 946 SPE genes, 17,996 (64%) and 131 (14%) of which were assigned to subgenomes, respectively (Table 2). Again, significantly fewer SPE-WPR genes than expected were assigned to the subgenomes, indicating that most SPE-WPR genes are not conserved at syntenic loci in other species (Table 2; Fisher's exact test, $P < 0.001$).

Downloaded from https://academic.oup.com/plcell/article/26/10/3939/6101583 by ULB Bonn user on 21 December 2021

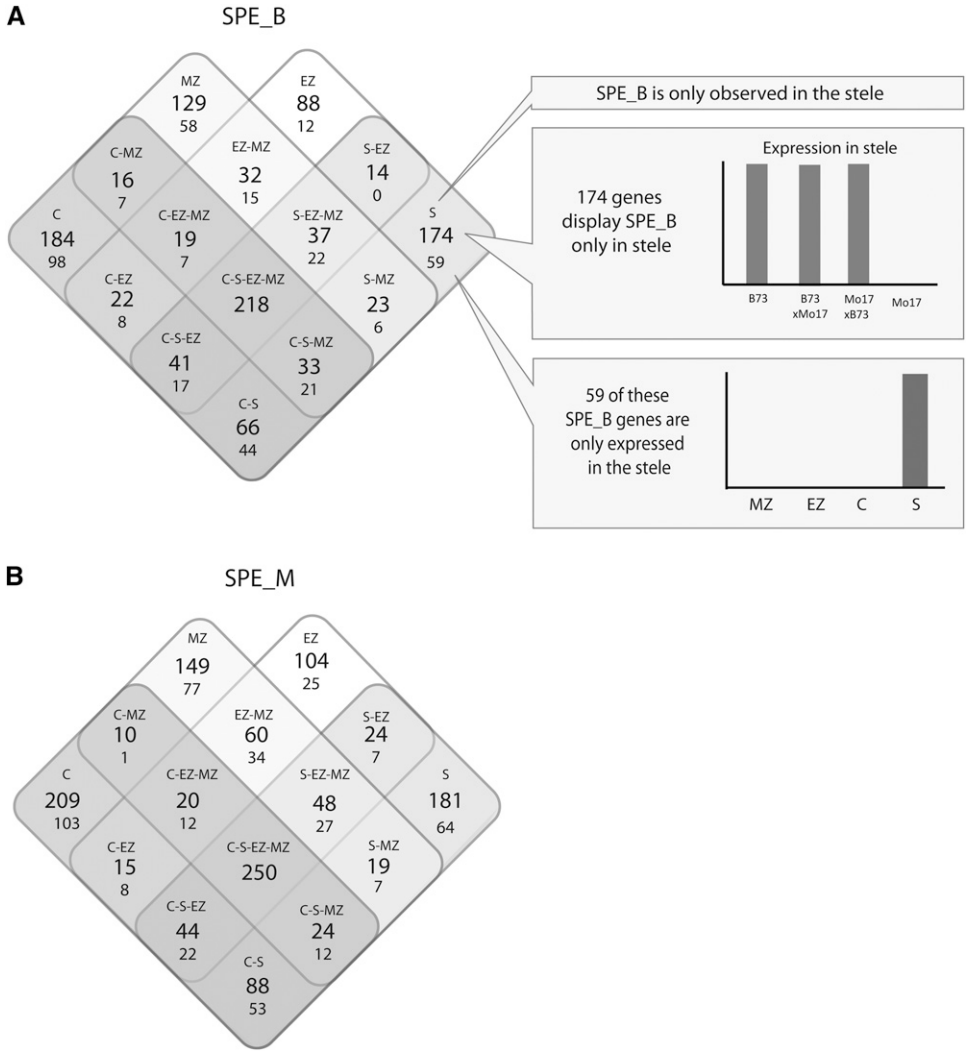


Figure 3. Summary of SPE Genes Conserved between and Unique for Different Tissues in Four-Way Venn Diagrams.

(A) SPE_B genes expressed only in B73 and the two reciprocal hybrids.
(B) SPE_M genes expressed only in Mo17 and the two reciprocal hybrids. The larger numbers in each square field indicate all genes that display the SPE pattern, and the smaller numbers below indicate how many of the SPE genes are expressed only in the tissue(s) in which they display SPE.

Functional Classification of SPE Genes

To identify overrepresented biological functions among the four groups of tissue-specific SPE genes a Gene Ontology (GO) analysis was conducted. Only the biological processes "death" and its subcategories "cell death," "programmed cell death," and "apoptosis" were overrepresented in each of the four tissue-specific SPE data sets. Among the four tissue-specific sets of SPE genes, between 11 and 18 genes were assigned to these GO terms. These genes overlap between tissues, which results in a total of 29 genes assigned to the GO term apoptosis and the parent terms (Supplemental Table 3). Eighteen of the 29 genes display SPE_B, while 11/29 genes show SPE_M expression patterns, indicating that both B73 and Mo17 are deficient in the expression of cell death-related genes. The majority (24/29) of these genes cannot be assigned to subgenomes, suggesting that they emerged after the last whole-genome duplication. Many of them are suggested to encode enzymes with nucleoside triphosphatase activity, which plays a role in the degradation of nucleic acids (Supplemental Table 3).

DISCUSSION

The high degree of intraspecific genome diversity is one remarkable feature of maize that has been studied in great detail (Springer et al., 2009; Swanson-Wagner et al., 2010; Chia et al., 2012; Jiao et al., 2012; Marroni et al., 2014). Recent studies investigated how genomic diversity translates to intraspecific transcriptome diversity (Stupar and Springer, 2006; Hansey et al., 2012; Hirsch et al., 2014), which is fostered by the advent of high-throughput sequencing technologies. Using RNA-seq as one of these approaches, this study surveyed four primary root tissues of maize for gene activity in the inbred lines B73 and Mo17 and their reciprocal hybrids.

Downloaded from https://academic.oup.com/plcell/article/26/10/3939/6101583 by ULB Bonn user on 21 December 2021

Table 1. Biallelic Expression of SPE Genes in the Reciprocal Hybrids B73xMo17 and Mo17xB73

Group of SPE Genes[a]	No. of SPE Genes	No. of SPE Genes with SNPs	No. of SPE Genes with SNPs, PAVs Excluded[b]	Subset of SPE Genes Expressing Both Parental Alleles			
				B73xMo17		Mo17xB73	
				≥1 Read	≥10 Reads	≥1 Read	≥10 Reads
SPE_B-MZ	510	399	374	73	0	74	0
SPE_M-MZ	581	525	–	116	2	132	4
SPE_B-EZ	472	381	356	70	0	62	0
SPE_M-EZ	565	502	–	134	2	124	2
SPE_B-C	599	475	451	95	0	97	0
SPE_M-C	661	611	–	176	3	197	2
SPE_B-S	607	488	457	107	1	117	1
SPE_M-S	680	611	–	183	2	169	3

An allele was considered to be expressed with either one read or 10 reads minimum. PAVs were excluded from all SPE_B genes.
[a]SPE_B, gene expressed in B73 and hybrids but not in Mo17; SPE_M, gene expressed in Mo17 and hybrids but not in B73. B, B73; M, Mo17; MZ, meristematic zone; EZ, elongation zone; C, cortex; S, stele.
[b]PAVs (genes present in the B73 genome but absent from the Mo17 genome).

In general, between 94 and 95% of all genes active in the four tissues were expressed in all four genotypes (Figure 2). This number is similar to a recently published data set of whole primary roots of the same developmental stage in which 97% of all active genes were expressed in each of the four genotypes (Paschold et al., 2012). Remarkably, as in whole roots, the major proportion of the genes not expressed in every genotype displayed SPE, indicating that a gene is only expressed in one parental inbred line but is expressed in both reciprocal hybrids (Paschold et al., 2012). SPE represents a special instance of complementation on the level of gene expression. On the transcriptomic level, the concept of complementation implies that the combination of two distinct genomes leads to the activity of many genes in hybrids that are only active in one parental inbred line. Support for this hypothesis has been reported for whole roots where the hybrid with the lower number of active genes expressed 352 more genes than the inbred line with higher number of active genes (Paschold et al., 2012). This study went one step ahead and increased the resolution to the level of individual root tissues. The bidirectional comparison of four genotypes and four primary root tissues revealed not only hundreds of tissue-specifically expressed genes but also hundreds of SPE genes in each tissue. Similar to what has been reported in whole roots (Paschold et al., 2012) both hybrids expressed more genes in each individual tissue than either parent. Comparing the hybrid showing fewer expressed genes with the parental line displaying more expressed genes this increase ranged between 444 (meristematic zone) and 558 (stele) additional active genes.

To allow for comparison of the SPE genes of this tissue-specific data set with the results of the previously published whole root data set, we reanalyzed the whole root data (Paschold et al., 2012) with the improved SPE annotation algorithm used in this study. This approach reduced the number of whole root SPE genes from 1124 to 946 and increased the difference to the 2341 genes displaying SPE in at least one root tissue. Because both data sets were generated by the same laboratories, following the same experimental design, data analysis pipelines, and statistical procedures, the discrepancy between whole roots and individual tissues is likely the result of biological differences, not of technical artifacts or variations. Among all tissue-specific SPE genes, 55% display SPE in all tissues in which they are expressed while 45% displayed SPE in at least one tissue and general expression (i.e., the gene is expressed in each of the four genotypes) in at least one other tissue.

Table 2. SPE Genes of the Individual Root Tissues and Their Assignment to Maize Subgenomes

	No. of Genes	No. of Genes Assigned to Subgenomes	%	Subgenome 1	%	Subgenome 2	%
FGS genes[a]	39,656	19,365	49	12,190	63	7,175	37
Expressed genes PRT	27,347	17,402	64	10,962	63	6,440	37
Expressed genes WPR	28,113	17,996	64	11,343	63	6,653	37
SPE class							
SPE total	2,341	696***	30	429	62	267	38
SPE_C	1,260	310***	25	197	64	113	36
SPE_S	1,287	294***	23	184	63	110	37
SPE_EZ	1,037	226***	22	156	68	71	31
SPE_MZ	1,091	199***	18	113	57	86	43
SPE WPR	946	131***	14	87	66	44	34

In addition to our tissue-specific analysis (PRT), a published WPR data set (Paschold et al., 2012) was reanalyzed using the same method as in this study. Fisher's exact test was used to determine if more or less SPE genes than expected are assigned to subgenomes (***$P < 0.001$).
[a]Filtered gene set of maize B73.

Downloaded from https://academic.oup.com/plcell/article/26/10/3939/6101583 by ULB Bonn user on 21 December 2021

SPE genes with such dynamic tissue-specific expression patterns are unlikely to be identified as SPE genes from complete organs because the SPE patterns are masked by the general expression of these genes in other tissues. This notion is supported by the observation that between 70% (SPE_B) and 52% (SPE_M) of all whole-root SPE patterns overlap with tissue-specific SPE patterns that are exclusively expressed in the tissues where they display SPE. This indicates that a large portion of the SPE genes identified from whole roots lack tissue-dependent gene expression dynamics. Alternatively, some of the SPE genes identified from whole roots might be expressed in each tissue, but the expression level of one particular tissue exceeds that of the others. Hence, the plasticity of SPE patterns is conditioned by the tissue-specific regulation of gene expression, indicating that a silenced allele of a SPE gene in an inbred line is not per se silenced in all tissues. This should not be confused with general tissue specificity of gene expression (Brady et al., 2007) as it refers to activity/ inactivity of genes in particular tissues. SPE, on the other hand, refers to genotype-specific gene activity and different alleles of a gene in two inbred lines.

Genotype-specific activity was detected for hundreds of expressed genes and most of these differences were complemented in the reciprocal hybrids. Only a minor proportion of these transcript variations were caused by differences in the genomic sequence (i.e., by genomic PAVs). This suggests that most of the genes are physically present in the genome but that they are inactivated. This inactivity can be caused by changes in the genomic sequence of the considered gene (*cis*-modification) or by alterations of proteins such as transcription factors regulating the activity of these genes (*trans*-modification) (Wittkopp et al., 2004). Most of the inbred line-specific differences in gene activity are complemented in the hybrids and the analysis of allele-specific gene expression levels revealed, by applying a conservative cutoff threshold, only a minor degree of reactivation of inactive alleles in the hybrids. This implies predominant *cis*-regulation of the SPE genes and genomic differences leading to transcriptional variation between inbred lines. Omission of any cutoff threshold resulted in biallelic expression of dozens of SPE genes in the hybrids. This observation needs to be treated with caution because the majority of these genes express the second allele at levels below 10 reads per transcript. When mapping such small read numbers to SNP positions, sequencing errors are more likely to result in mapping errors ultimately leading to false positives (Macmanes and Eisen, 2013). To account for this uncertainty, we set an arbitrary threshold cutoff of 10 reads

Phylogenetic analyses of the SPE genes identified in different root tissues were conducted to elucidate their evolutionary origin. The maize genome was duplicated several times during evolution including a whole-genome duplication (autopolyploidization) ~5 to 12 million years ago (Blanc and Wolfe, 2004; Swigonová et al., 2004), which separated maize from its close relative, *Sorghum bicolor* (Paterson et al., 2009). In modern maize, the result of this most recent whole-genome duplication can be surveyed by comparing the duplicated maize and the unduplicated *Sorghum* genomes. Such comparisons enable the identification of recognizable homoeologs and the assignment of a considerable proportion of all current maize genes to one of the two subgenomes that resulted from this genome duplication (Schnable et al., 2011). To determine if the phylogenetic origin of a gene conditioned its contribution to intraspecific transcriptome diversity, all genes that displayed SPE patterns were analyzed for their assignment to subgenomes. In total, 49% (19,365) of all maize genes were assigned to subgenomes 1 or 2 (Schnable et al., 2011). The remaining 51% likely emerged more recently, e.g., by modular rearrangement of protein encoding domains (Kersting et al., 2012), by the transposition of existing genes (Freeling et al., 2008), or by exon shuffling from helitrons and other transposons leading to fusion genes (Barbaglia et al., 2012). Among the 27,347 genes expressed in this study, 17,402 (64%) genes were assigned to one of the two subgenomes. This implies that in general ancient genes are more frequently expressed than nonsyntenic ones and is consistent with the observation that this same population of old genes is highly enriched in genes with visible mutant phenotypes (Schnable and Freeling, 2011). Since these genes have survived more than 5 million years of natural selection they likely have crucial functions in maize development. While ancient genes are overrepresented among all expressed genes they are significantly underrepresented among the SPE genes. Only between 18 and 25% of all SPE genes in the four tissues can be assigned to subgenomes 1 or 2. The number of nonsyntenic genes (i.e., genes not assigned to any subgenome) might be inflated by false positives caused by ancestral genes present in one or both maize subgenomes but lost from the genome of *Sorghum*, by genes apparently present at nonsyntenic locations only as a result of assembly errors in either maize or *Sorghum*, or by gene prediction errors. However, these false positives, if any, are expected to be equally common among all expressed genes including both SPE and non-SPE genes, so any bias introduced would be toward proportions of nonsyntenic genes among these two categories. In fact, we observed the opposite pattern. These results imply that a disproportionate number of SPE genes emerged after the separation of maize from *Sorghum* and the subsequent duplication of the maize genome.

Recently emerged genes might be less likely to play essential roles during the early stages of seedling development. Maize was domesticated from its ancestor teosinte only ~10,000 years ago (Doebley et al., 2006), resulting in the so-called landraces, which in turn have been improved during recent decades to generate elite inbred lines. However, SPE genes are likely unrelated to domestication and improvement as they were neither under nor overrepresented (data not shown) among genes found to be under selection during these processes (Hufford et al., 2012).

The observation that F1-hybrids complement gene expression patterns of contrasting parental inbred lines and express more genes could have implications for the manifestation of heterosis, which describes the increased vigor of hybrid plants compared with the average of their parental inbred lines (Shull, 1908) and is of enormous economic relevance in maize breeding (Duvick, 2005). There is consensus that heterosis requires genetic variation between the parents leading to highly heterozygotic hybrids and that heterosis is conditioned by the action of many loci (reviewed in Schnable and Springer, 2013).

It has been suggested that genotypic differences, such as polymorphisms, the number and distribution of repetitive sequences, gene copy number variation, or PAV contribute to heterosis (Zhang et al., 2008; Lai et al., 2010; Swanson-Wagner et al., 2010). Moreover, it has been demonstrated that despite the consensus that all organs have the same complement of the nuclear genome,

Downloaded from https://academic.oup.com/plcell/article/26/10/3939/6101583 by ULB Bonn user on 21 December 2021

the degree of heterosis largely varies for different tissues or phenotypic traits (Flint-Garcia et al., 2009). This observation suggests that heterosis is associated with tissue-specific gene activity, which in turn is conditioned by genetic variation. Tissue-specific SPE patterns as they are reported from this study support this hypothesis; however, it is likely that several other processes contribute to heterosis (Schnable and Springer, 2013).

In line with this concept, one could hypothesize that the ability to express ~2% more genes than their parental inbred lines provides a phenotypic advantage to hybrid plants. Their evolutionary young age argues against SPE genes encoding essential proteins and rather suggests that these genes encode proteins having redundant functions thus making the respective biological processes more robust, e.g., in stress situations.

The finding that apoptosis-related genes, which are involved in several developmental processes in roots (Fukuda, 1996; Buckner et al., 1998; Drew et al., 2000), are overrepresented among the SPE genes supports this hypothesis. Hybrids might benefit from the complementation of these deficiencies, e.g., by improved regulation of programmed cell death.

In summary, SPE is observed for hundreds of genes in very dynamic tissue-specific patterns. Many genes display SPE only in certain tissues while they show general expression in other tissues, which underpins the subtle regulation of these expression patterns and tightly controlled transcriptomic plasticity in root tissues. Finally, most of the SPE genes are not conserved at syntenic loci in the genomes of other grasses such as *Sorghum*, suggesting that in the maize genome, the average SPE gene is evolutionarily younger than the average expressed gene. These findings imply a role for SPE genes in the manifestation of heterosis in maize, which needs to be further analyzed in the future.

METHODS

Plant Material and Growth Conditions

Seeds of the maize (*Zea mays*) inbred lines B73 and Mo17 and the two reciprocal hybrids B73xMo17 and Mo17xB73 were germinated in paper rolls in distilled water as previously described (Hoecker et al., 2006). For each genotype, 18 paper rolls each containing 10 kernels were prepared. All material was germinated in a single 10-liter bucket in which the 72 paper rolls were randomly arranged. Primary roots of 2 to 4 cm length were harvested 4 d after germination. Even at this very early developmental stage, many hybrid primary roots were longer than those of their parental inbred lines. To avoid differential gene expression patterns based simply on root length, pools of roots having the same lengths (2 to 4 cm) were selected for all genotypes and replicates. From each root, four different tissues were sampled (Figure 1). First, 3 mm of the root tip was collected, which corresponded to the terminal root cap and the subterminal meristematic zone (Dembinsky et al., 2007). Second, the zone adjacent to the root tip up to the part of the root where root hairs became visible was separated, which basically corresponds to the elongation zone. Finally, we separated the beginning of the root hair zone up to the coleorhiza, which is connected to the differentiation zone. In the differentiation zone, cortex and stele were mechanically separated without damaging pericycle and endodermis cells (Saleem et al., 2009). For each tissue, 14 roots were pooled for each of the four biological replicates per genotype.

RNA Isolation and Sequencing Library Construction

The pooled primary root tissues were ground under liquid nitrogen and RNA was isolated as previously described (Winz and Baldwin, 2001). RNA quality was assayed via agarose gel electrophoresis and a Bioanalyzer (Agilent Technologies). Only samples with an RNA integrity number (Schroeder et al., 2006) $\geq$ 7.2 were used for downstream analyses. The cDNA libraries for sequencing-by-synthesis were constructed according to the protocol of the manufacturer (Illumina). The 64 samples were loaded on two flow cells in a predefined order (Supplemental Table 1). The experiment was randomized according to a split plot design. Four contiguous lanes of a flow cell were regarded as a complete block, so there were four complete blocks (biological replicates), distributed across two flow cells with eight lanes each. The four tissues were randomized within each block (set of four lanes). A lane was considered as a main plot. For the same tissue, the four genotypes were allocated to each lane, so the genotypes correspond to subplots of the design. The different tissues for the same genotype in a block originated from the same plants. Cluster generation and sequencing (100 bp, single read) was performed as per the instructions of the manufacturer (Genome Analyzer IIx; Illumina).

Analysis of Raw Sequencing Data, Read Mapping, and Allele-Specific Read Calling

Sequencing was performed on the GAIIx genome analyzer equipped with on-instrument sequencing control software SCS Version 2.8 and real-time analysis RTA1.8.7. The genome analyzer data analysis pipeline OLB version 1.8.0 was used for base calling and run statistics on all 16 lanes. On average, a raw cluster density of 387,145 clusters per tile was achieved, yielding between ~3.25 and ~3.76 Gb per lane after sample post filtering (see Supplemental Figure 1 and Supplemental Table 1 for sample allocation and sequencing performance per lane).

The maize B73 reference genome (Schnable et al., 2009) was indexed using GMAP_BUILD (k-mer size of 15 and step size 3) and all high-quality reads were aligned to the whole reference genome using the short reads aligner GSNAP (http://research-pub.gene.com/gmap/, version 2012-01-11). Only reads uniquely mapping to the B73 reference genome (Schnable et al., 2009) with a maximum of two mismatches out of 36 bp and with at most 5-bp tails for every 76 bp were extracted and subsequently analyzed. In addition, redundant reads mapping at the same starting coordinate and mapping orientation (stacked reads) were removed from the set of uniquely mapping RNA-seq reads. The remaining reads of all samples were projected to the filtered gene set (http://ftp.maizesequence.org/release-5b/filtered-set/, release 5b.60) of the B73 reference genome derived from maize genome sequencing project using a Perl script. Allele-specific gene expression was analyzed with a set of 4,034,683 SNPs (identified between the B73 and Mo17 genotypes via 123SNP software, which is available for download from http://schnablelab.plantgenomics.iastate.edu/software. B73-Mo17 SNPs were called by aligning B73 RefGen_v2 sequences and Mo17 sequences retrieved from GenBank Sequence Read Archive records (accession numbers SRA009756, SRA049859, SRA048055, SRA053592, and SRA050451) and maizegdb (http://ftp.maizegdb.org/MaizeGDB/FTP/Mo17/).

The allele-specific reads were extracted from the reads uniquely projected onto the FGSv2 and associated with any B73-Mo17 SNP using a customized Perl script. Since the mapping procedure presented here did not allow for quantifying sequence reads mapping to repeats throughout the genome, only uniquely mapping genes were considered. In addition, transcript isoforms of a given gene locus could not be distinguished.

Detection of Expression Complementation and Functional Analysis

Classification of the gene-wise transcriptional activity (presence/absence) for each genotype was determined by a hierarchical Bayesian latent class model using Markov Chain Monte Carlo methods, similar to Paschold et al. (2012). We again modeled read counts via the mean-dispersion parameterization of the negative binomial distribution, using the median read count multiplied by the exon length (in kilobases) of the specific transcript as an offset value for normalization. Expression was modeled

Downloaded from https://academic.oup.com/plcell/article/26/10/3939/6101583 by ULB Bonn user on 21 December 2021

log-linearly, with overall gene and gene-specific genotype effects, which we assumed were both normally distributed. To account for the split-plot design, random normally distributed plot effects were also included in the model. Each gene was assumed to have its own unique dispersion parameter, which was treated as random from a common log-normal distribution. To account for falsely mapped reads, we included a fixed background error rate estimated from the data. The model was defined using MCMC software JAGS (Plummer, 2003) and fit in R using the R2jags library package (Su and Yajima, 2011). Due to the size of such a model under the full data, we initially trained the model on a representative subset (4000 total) of the genes, each selected at random. We ran two Markov chains at a total of 15,000 iterations each, treating the first 10,000 as a burn-in and then thinning the remaining iterations by saving every 5th iteration, resulting in posterior sample parameter traces of length 1000. We analyzed the posterior sample to verify that the model had converged by visually inspecting the mixture of the chains and using posterior checking procedures from the coda output diagnostics package (Plummer et al., 2006) in R. We then applied Monte Carlo integration under the learned posterior distributions of the hyperparameters to obtain gene-wise posterior probabilities of each possible state vector over the entire set of genes, conditional on the observed read counts. Classification of each gene was determined by the maximum posterior predictive probability of a given state vector. Separate model fits were run for each analysis. PAVs were called as previously described (Paschold et al., 2012).

The GO of specific genes and the overrepresentation of GO terms were conducted using the single enrichment analysis of the AgriGo platform (http://bioinfo.cau.edu.cn/agriGO/analysis.php).

Accession Numbers

Raw sequencing data are stored at the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) under accession number SRP029742.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Distribution of 64 Samples across 16 Lanes on Two Flow Cells for RNA-seq.

Supplemental Table 1. Summary of the Sequencing Output of the 64 Primary Root Tissue Samples.

Supplemental Table 2. Summary of the Sequencing Output after Trimming of the Reads, Mapping Them to the Maize Genomic Sequence, Removal of Redundant Reads, and Determination of Reads in the Gene Space of the Maize Genome.

Supplemental Table 3. Summary of the SPE Genes Assigned to the GO Term Apoptosis.

Supplemental Data Set 1. Summary of All Expressed Genes and Their Distribution across Tissues, Presence-Absence Variations, and Their Assignment to Subgenomes.

ACKNOWLEDGMENTS

This article is dedicated to Guenter Feix on the occasion of his 80th birthday. We thank Nahal Ahmadinejad (INRES, Crop Bioinformatics, University of Bonn) for assistance with the allele-specific expression analyses and Jens Riexinger (Max Planck Institute for Developmental Biology, Tuebingen, Germany) for technical assistance during library production and sequencing. This work was funded by Deutsche Forschungsgemeinschaft Grants HO2249-9/1 to F.H. and PI 377/12-1 to H.-P.P.

AUTHOR CONTRIBUTIONS

A.P. carried out the experiment and tissue sampling, conducted downstream analyses, drafted the article, and designed figures. N.B.L. carried out the complementation analysis. C.M. isolated RNAs, constructed RNA-seq libraries, and participated in writing the article. J.C.S. participated in the subgenome analysis and in article writing. C.-T.Y. carried out the mapping and analysis of raw sequencing data. C.L. carried out the RNA sequencing and participated in library construction. D.N. and H.-P.P. provided the sequencing design and performed the statistical analysis. P.S.S. participated in data interpretation and helped to draft the article. F.H. conceived and coordinated the study and participated in data interpretation and drafting the article. All authors approved the final article.

Received August 12, 2014; revised September 5, 2014; accepted September 24, 2014; published October 14, 2014.

REFERENCES

Barbaglia, A.M., Klusman, K.M., Higgins, J., Shaw, J.R., Hannah, L.C., and Lal, S.K. (2012). Gene capture by Helitron transposons reshuffles the transcriptome of maize. Genetics **190:** 965–975.

Barbazuk, W.B., Emrich, S.J., Chen, H.D., Li, L., and Schnable, P.S. (2007). SNP discovery via 454 transcriptome sequencing. Plant J. **51:** 910–918.

Bi, I.V., McMullen, M.D., Sanchez-Villeda, H., Schroeder, S., Gardiner, J., Polacco, M., Soderlund, C., Wing, R., Fang, Z., and Coe, E.H. (2006). Single nucleotide polymorphisms and insertion-deletions for genetic markers and anchoring the maize fingerprint contig physical map. Crop Sci. **46:** 12–21.

Blanc, G., and Wolfe, K.H. (2004). Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. Plant Cell **16:** 1667–1678.

Brady, S.M., Orlando, D.A., Lee, J.Y., Wang, J.Y., Koch, J., Dinneny, J.R., Mace, D., Ohler, U., and Benfey, P.N. (2007). A high-resolution root spatiotemporal map reveals dominant expression patterns. Science **318:** 801–806.

Buckner, B., Janick-Buckner, D., and Gray, J. (1998). Cell-death mechanisms in maize. Trends Plant Sci. **3:** 218–223.

Chia, J.M., et al. (2012). Maize HapMap2 identifies extant variation from a genome in flux. Nat. Genet. **44:** 803–807.

Dembinsky, D., et al. (2007). Transcriptomic and proteomic analyses of pericycle cells of the maize primary root. Plant Physiol. **145:** 575–588.

Doebley, J.F., Gaut, B.S., and Smith, B.D. (2006). The molecular genetics of crop domestication. Cell **127:** 1309–1321.

Drew, M.C., He, C.J., and Morgan, P.W. (2000). Programmed cell death and aerenchyma formation in roots. Trends Plant Sci. **5:** 123–127.

Duvick, D.N. (2005). The contribution of breeding to yield advances in maize (*Zea mays* L.). Adv. Agron. **86:** 83–145.

Flint-Garcia, S.A., Buckler, E.S., Tiffin, P., Ersoz, E., and Springer, N.M. (2009). Heterosis is prevalent for multiple traits in diverse maize germplasm. PLoS ONE **4:** e7433.

Freeling, M., Lyons, E., Pedersen, B., Alam, M., Ming, R., and Lisch, D. (2008). Many or most genes in Arabidopsis transposed after the origin of the order Brassicales. Genome Res. **18:** 1924–1937.

Fukuda, H. (1996). Xylogenesis: Initiation, progression, and cell death. Annu. Rev. Plant Physiol. Plant Mol. Biol. **47:** 299–325.

Hansey, C.N., Vaillancourt, B., Sekhon, R.S., de Leon, N., Kaeppler, S.M., and Buell, C.R. (2012). Maize (*Zea mays* L.) genome diversity as revealed by RNA-sequencing. PLoS ONE **7:** e33071.

Hirsch, C.N., et al. (2014). Insights into the maize pan-genome and pan-transcriptome. Plant Cell **26:** 121–135.

Downloaded from https://academic.oup.com/plcell/article/26/10/3939/6101583 by ULB Bonn user on 21 December 2021

Hochholdinger, F. (2009). The maize root system: morphology, anatomy, and genetics. In Handbook of Maize: Its Biology, J.L. Bennetzen and S.C. Hake, eds (New York: Springer), pp. 145–160.

Hoecker, N., Keller, B., Piepho, H.P., and Hochholdinger, F. (2006). Manifestation of heterosis during early maize (*Zea mays* L.) root development. Theor. Appl. Genet. **112:** 421–429.

Hufford, M.B., et al. (2012). Comparative population genomics of maize domestication and improvement. Nat. Genet. **44:** 808–811.

Ishikawa, H., and Evans, M.L. (1995). Specialized zones of development in roots. Plant Physiol. **109:** 725–727.

Jiao, Y., et al. (2012). Genome-wide genetic changes during modern breeding of maize. Nat. Genet. **44:** 812–815.

Kato, A., Lamb, J.C., and Birchler, J.A. (2004). Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. Proc. Natl. Acad. Sci. USA **101:** 13554–13559.

Kersting, A.R., Bornberg-Bauer, E., Moore, A.D., and Grath, S. (2012). Dynamics and adaptive benefits of protein domain emergence and arrangements during plant genome evolution. Genome Biol. Evol. **4:** 316–329.

Lai, J., et al. (2010). Genome-wide patterns of genetic variation among elite maize inbred lines. Nat. Genet. **42:** 1027–1030.

Laurie, D.A., and Bennett, M.D. (1985). Nuclear-DNA content in the genera *Zea* and *Sorghum* - intergeneric, interspecific and intraspecific variation. Heredity **55:** 307–313.

Macmanes, M.D., and Eisen, M.B. (2013). Improving transcriptome assembly through error correction of high-throughput sequence reads. PeerJ **1:** e113.

Marroni, F., Pinosio, S., and Morgante, M. (2014). Structural variation and genome complexity: is dispensable really dispensable? Curr. Opin. Plant Biol. **18:** 31–36.

Marschner, P. (2011). Mineral Nutrition of Higher Plants. (London: Academic Press).

Olsen, K.M., and Wendel, J.F. (2013). A bountiful harvest: genomic insights into crop domestication phenotypes. Annu. Rev. Plant Biol. **64:** 47–70.

Paschold, A., Jia, Y., Marcon, C., Lund, S., Larson, N.B., Yeh, C.T., Ossowski, S., Lanz, C., Nettleton, D., Schnable, P.S., and Hochholdinger, F. (2012). Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents. Genome Res. **22:** 2445–2454.

Paterson, A.H., et al. (2009). The *Sorghum bicolor* genome and the diversification of grasses. Nature **457:** 551–556.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In 3rd International Workshop on Distributed Statistical Computing, K. Hornik, F. Leisch, and A. Zeileis, eds (Vienna, Austria: Technical University Vienna).

Plummer, M., Best, N., Cowles, K., Vines, K. (2006). Convergence diagnosis and output analysis for MCMC. R News **6:** 7–11.

Saleem, M., Lamkemeyer, T., Schützenmeister, A., Fladerer, C., Piepho, H.P., Nordheim, A., and Hochholdinger, F. (2009). Tissue specific control of the maize (*Zea mays* L.) embryo, cortical parenchyma, and stele proteomes by RUM1 which regulates seminal and lateral root initiation. J. Proteome Res. **8:** 2285–2297.

Schnable, J.C., and Freeling, M. (2011). Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS ONE **6:** e17855.

Schnable, J.C., Springer, N.M., and Freeling, M. (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc. Natl. Acad. Sci. USA **108:** 4069–4074.

Schnable, P.S., and Springer, N.M. (2013). Progress toward understanding heterosis in crop plants. Annu. Rev. Plant Biol. **64:** 71–88.

Schnable, P.S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. Science **326:** 1112–1115.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol. Biol. **7:** 3.

Shull, G.F. (1908). The composition of a field of maize. American Breeders' Association. Annual Report **4:** 296–301.

Springer, N.M., et al. (2009). Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. PLoS Genet. **5:** e1000734.

Stupar, R.M., and Springer, N.M. (2006). *Cis*-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F_1 hybrid. Genetics **173:** 2199–2210.

Su, Y., and Yajima, M. (2011). R2jags: A package for running JAGS from R. In R Package Version 0.02-17, http://cran.r-project.org/web/packages/R2jags/.

Swanson-Wagner, R.A., Eichten, S.R., Kumari, S., Tiffin, P., Stein, J.C., Ware, D., and Springer, N.M. (2010). Pervasive gene content variation and copy number variation in maize and its undomesticated progenitor. Genome Res. **20:** 1689–1699.

Swigonová, Z., Lai, J., Ma, J., Ramakrishna, W., Llaca, V., Bennetzen, J.L., and Messing, J. (2004). Close split of sorghum and maize genome progenitors. Genome Res. **14:** 1916–1923.

Winz, R.A., and Baldwin, I.T. (2001). Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. IV. Insect-Induced ethylene reduces jasmonate-induced nicotine accumulation by regulating putrescine N-methyltransferase transcripts. Plant Physiol. **125:** 2189–2202.

Wittkopp, P.J., Haerum, B.K., and Clark, A.G. (2004). Evolutionary changes in *cis* and *trans* gene regulation. Nature **430:** 85–88.

Woodhouse, M.R., Schnable, J.C., Pedersen, B.S., Lyons, E., Lisch, D., Subramaniam, S., and Freeling, M. (2010). Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. PLoS Biol. **8:** e1000409.

Zhang, H.Y., He, H., Chen, L.B., Li, L., Liang, M.Z., Wang, X.F., Liu, X.G., He, G.M., Chen, R.S., Ma, L.G., and Deng, X.W. (2008). A genome-wide transcription analysis reveals a close correlation of promoter INDEL polymorphism and heterotic gene expression in rice hybrids. Mol. Plant **1:** 720–731.

Downloaded from https://academic.oup.com/plcell/article/26/10/3939/6101583 by ULB Bonn user on 21 December 2021

Opitz et al. BMC Genomics 2014, 15:741
http://www.biomedcentral.com/1471-2164/15/741

BMC Genomics

RESEARCH ARTICLE Open Access

Transcriptomic complexity in young maize primary roots in response to low water potentials

Nina Opitz[1], Anja Paschold[1], Caroline Marcon[1], Waqas Ahmed Malik[2], Christa Lanz[3], Hans-Peter Piepho[2] and Frank Hochholdinger[1*]

Abstract

Background: Widespread and more frequently occurring drought conditions are a consequence of global warming and increase the demand for tolerant crop varieties to feed the growing world population. A better understanding of the molecular mechanisms underlying the water deficit response of crops will enable targeted breeding strategies to develop robust cultivars.

Results: In the present study, the transcriptional response of maize (*Zea mays* L.) primary roots to low water potentials was monitored by RNA sequencing (RNA-Seq) experiments. After 6 h and 24 h of mild (-0.2 MPa) and severe (-0.8 MPa) water deficit conditions, the primary root transcriptomes of seedlings grown under water deficit and control conditions were compared. The number of responsive genes was dependent on and increased with intensification of water deficit treatment. After short-term mild and severe water deficit 249 and 3,000 genes were differentially expressed, respectively. After a 24 h treatment the number of affected genes increased to 7,267 and 12,838 for mild and severe water deficit, respectively, including more than 80% of the short-term responsive genes. About half of the differentially expressed genes were up-regulated and maximal fold-changes increased with treatment intensity to more than 300-fold. A consensus set of 53 genes was differentially regulated independently of the nature of deficit treatment. Characterization revealed an overrepresentation of the Gene Ontology (GO) categories "oxidoreductase activity" and "heme binding" among regulated genes connecting the water deficit response to ROS metabolism.

Conclusion: This study gives a comprehensive insight in water deficit responsive genes in young maize primary roots and provides a set of candidate genes that merit further genetic analyses in the future.

Keywords: Drought, Low water potential, Maize, RNA-Seq, Root, Transcriptome, Water deficit

Background

Population growth and global warming are major challenges for global food security. It is estimated that the demand for agricultural products will increase by ~50% until 2030 [1]. This requires historically unprecedented annual production growth rates [2]. Worldwide, about 70% of the food production is provided by cereals [3]. Maize (*Zea mays* L.) outcompeted all other cereals with an estimated global yield of 863 million tons in 2012/2013 [4]. While cereal production needs to be significantly increased, climate change adversely affects global maize production with an estimated loss of ~4% relative to what could have been achieved without the climate trends [5]. Poor soil moisture is widespread among arable land and as a consequence of global warming more areas are affected by drought conditions each year [6]. Since water availability is the most critical environmental factor for plant growth [7], drought can limit crop productivity more than any other abiotic stress. Furthermore, variations in water availability within fields can result in uneven crop stands that cause yield losses [8].

Under drought conditions, when water loss through transpiration is high, it is essential that roots maintain the capacity to acquire soil water and nutrients. This is reflected by the ability of roots to continue elongation even under severe water deficit conditions albeit at a slower rate [9]. From a physiological perspective, root growth maintenance is predominantly regulated by the plant hormone abscisic acid (ABA). Accumulation of ABA suppresses excessive ethylene production and thereby prevents growth

* Correspondence: hochholdinger@uni-bonn.de
[1]Institute of Crop Science and Resource Conservation (INRES), Crop Functional Genomics, University of Bonn, 53113 Bonn, Germany
Full list of author information is available at the end of the article

BioMed Central

© 2014 Opitz et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

inhibition. ABA is further involved in the processes leading to osmotic adjustment as it promotes the transport of proline to the root apex. At the more basal regions of the root, hexoses are the predominant solutes providing osmotic adjustment and maintaining turgor pressure [reviewed in 10]. From a cellular viewpoint, the processes related to the water deficit response begin with stress perception, followed by signal transduction, and a change in gene expression that finally confers the complex metabolic and physiological alterations necessary to gain stress tolerance [11,12]. On the molecular level, genes regulated by water deficit can be grouped into two categories. The first group of genes encodes proteins providing direct stress tolerance such as chaperones, transporters, osmolytic and detoxifying proteins, and repair-enzymes [13]. The second category includes proteins involved in stress response by regulating signal transduction and gene expression for instance transcription factors, protein kinases and phosphatases, and other signaling molecules [13]. The high quantity of genes regulated upon water deficit reflects the complexity of the stress response [14]. Nevertheless, details of the translation of environmental changes to metabolic responses i.e. the adjustment of transcriptional and post-transcriptional modifications of metabolic enzymes still remains unclear [12].

In the past, microarray chip hybridization experiments monitored gene expression profiles of maize leaves and roots to elucidate the transcriptional changes upon water deficit [14-19]. Recently developed next-generation sequencing approaches such as RNA sequencing (RNA-Seq) allow fully quantitative gene expression analyses [20] of all 39,656 (FGSv2; [21], release 5b.60) high-confidence maize gene models currently annotated [22]. The digital nature of the method enables the detection of a large dynamic range of expression levels with absolute values and the capture of even subtle gene expression changes [23,24].

In the present study, a tightly controlled, reproducible experimental setup was applied to expose maize seedlings to water deficit conditions. Young seedlings were grown in paper rolls soaked with polyethylene glycol (PEG) solutions. Previously, it was demonstrated that PEG treatment is an effective way to simulate drought stress conditions occurring in drying soil [14]. Water potentials of -0.2 MPa and -0.8 MPa were used for mild and severe water deficit treatments, respectively. At -0.2 MPa root growth is not or only slightly affected while shoot growth is reduced by half [25]. The lower water potential of -0.8 MPa completely inhibits elongation of maize shoots and leaves while roots continue to elongate consistently [9,25,26]. This maintenance of root growth in drying soil is beneficial to plants as they can reach deeper water resources. As top soil layers are prone to drying, it is particularly important for seedlings to adapt to low water potentials [9]. Although maize is most susceptible to drought stress during the flowering period [27], drought conditions during the seedling stage can negatively affect its yield [28].

To gain a better understanding of the early molecular responses to water deficit we utilized an RNA-Seq approach and compared the root transcriptomes of stressed and control maize seedlings. Seedlings were subjected to mild and severe water deficit conditions for 6 h and 24 h. The overall goal of this study was to identify a set of genes involved in initial water deficit responses in maize primary roots. Furthermore, application of mild and severe water deficit conditions at two time points aimed at detecting specifically and commonly regulated genes across treatment intensities and time. This data set will be a resource for future genetic analyses of candidate genes involved in water deficit response in young maize primary roots.

Results

Characterization of water deficit treatment and phenotypic responses

Kernels of the maize inbred line B73 were germinated in paper rolls soaked with distilled water until seedlings had a primary root length of 2 to 4 cm (Figure 1A). For mild and severe water deficit conditions, seedlings were transferred to PEG8000 solution with water potentials of -0.2 MPa and -0.8 MPa, respectively (see Methods). Water deficit treatment was applied for 6 h and 24 h. Each treatment was performed in four biological replicates each consisting of 10 roots.

To analyze phenotypic stress responses, primary root length was measured before and after treatment (Figure 1B). Short-term water deficit did not affect root elongation significantly. On average roots elongated 0.75 cm in 6 h. Similarly, after 24 h of water deficit treatment no differences of root growth between control and mildly stressed seedlings were observed (2.5 cm). However, severe water deficit significantly reduced root elongation (1.8 cm in 24 h) by almost 30% (p <0.001).

RNA sequencing and mapping of maize primary root transcriptomes

To identify genes responsive to water deficit in young maize primary roots, global gene expression was surveyed by Illumina RNA-Seq. On average, the RNA-Seq experiments yielded between 20 and 34 million reads per sample. The raw sequencing data has been deposited at the NCBI sequencing read archive (SRA, [29], Accession: SRP032921). Among all reads, 83 to 90% mapped to unique positions in the maize reference genome (ZmB73_RefGen_v2; Additional file 1). After removal of redundant reads sharing the same start and end coordinate, sequencing direction, and sequence ("stacked reads"), 76 to 78% of the remaining reads mapped uniquely to the "filtered gene set" (FGSv2, release 5b.60; Additional file 1), a set of 39,656 high confidence gene models predicted by a combination of

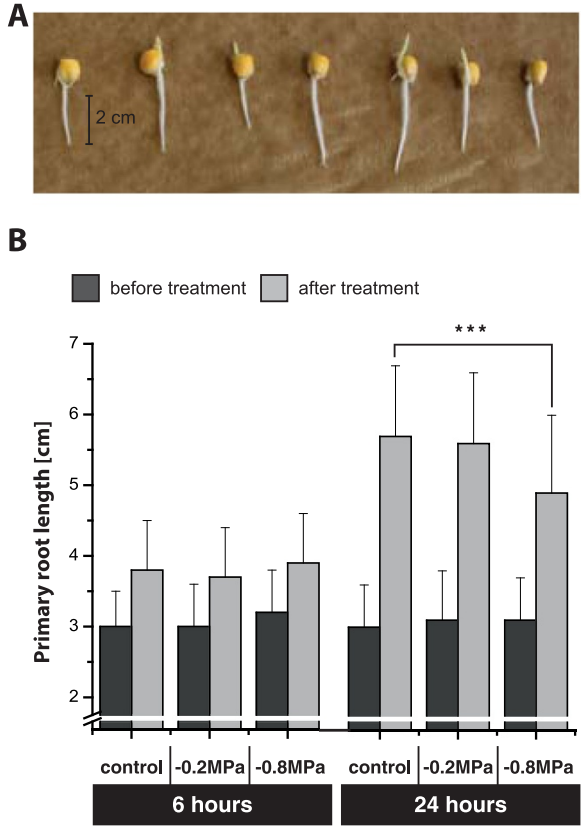


Figure 1 Root length measurement. A, Maize (B73) seedlings before treatment. **B**, Primary root length of seedlings before treatment (black bars) and after 6 h and 24 h of control or water deficit treatment (gray bars; water potential of -0.2 MPa: mild deficit, water potential of -0.8 MPa: severe deficit); n = 40, error bars: standard deviation; ***p-value <0.001.

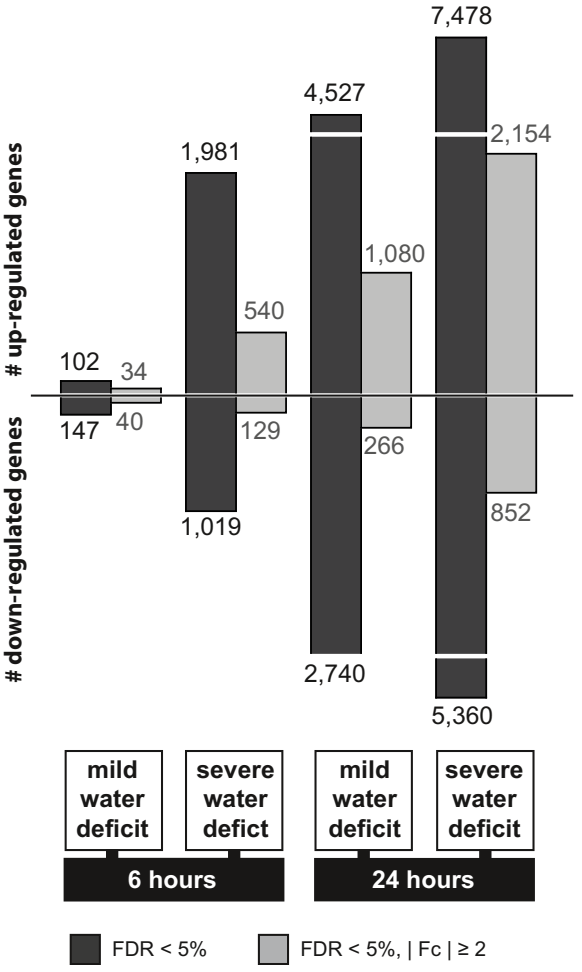


Figure 2 Number of differentially expressed genes. Bars represent up and down-regulated genes in the four pairwise comparisons of control groups and water deficit treatments.

evidence-based and *ab initio* approaches followed by stringent TE filtering [21,22]. A gene was declared expressed if a minimum of five reads mapped in all four replicates of a sample. As a result, 25,570 genes (64%) of the FGSv2 were expressed in at least one of the experimental conditions. A complete list of expressed genes with normalized expression values is provided in Additional file 2.

Exploration of differentially expressed genes in response to water deficit

To determine genes differentially expressed between control and water deficit conditions four pairwise comparisons of control groups versus the different water deficit treatments (mild deficit, 6 h (1), severe deficit, 6 h (2), mild deficit, 24 h (3), and severe deficit 24 h (4)) were performed. When controlling false discovery rate (FDR) at 5%, 249 and 3,000 genes were differentially expressed in response to 6 h mild and severe water deficit, respectively. After 24 h, the number of differentially expressed genes increased to 7,267 and 12,838 genes for mild and severe water deficit, respectively (Figure 2, black bars; Additional file 2). Small fold-changes (Fc) dominated among differentially expressed genes as about three-quarters of all affected genes display a |Fc| ≤2 (Additional file 3). To specifically focus on genes with a strong response to water deficit only genes with a |Fc| ≥2 were considered in subsequent analyses. This arbitrary cutoff reduced the numbers of differentially expressed genes to 74 (vs. 249) and 669 (vs. 3,000) for 6 h mild and severe water deficit and 1,346 (vs. 7,267) and 3,006 (vs. 12,838) for 24 h mild and severe deficit, respectively (Figure 2, gray bars; Additional file 2).

In summary, the results demonstrated that the duration and intensity of water deficit conditions significantly influence the number of responsive genes and the intensity of the response. In line with that observation, the maximal absolute Fc detected increased from 10-fold after short-term mild water deficit to 304-fold after 24 h severe water deficit. However, the average |Fc| was similar under all conditions and ranged between 2.8 and 3.6-fold.

The gene expression patterns in response to short-term mild water deficit differed from the patterns in response to the more intense or longer stress treatments. After 6 h of mild stress 46% of differentially expressed genes were up-regulated. In contrast, in the other three comparisons 72 to 82% of the water deficit responsive genes were up-regulated (Figure 2).

The overlap between the gene sets of the four comparisons is visualized in Figure 3A. Cross-comparison of responsive genes showed that 83% of genes responding to short-term mild water deficit and 81% of genes responding to short-term severe water deficit are also responsive after 24 h of water deficit. Similarly, 96% and 90% of genes responding to 6 h and 24 h mild water deficit were also responsive to severe water deficit. A set of 53 genes was differentially expressed (|Fc| ≥2) independently of water deficit level and treatment period. Among those, 30 were down and 23 up-regulated in all treatments. Such conservation of regulation direction was observed for most (99%) of the differentially expressed genes overlapping between two or more treatments. All 53 consistently responding genes were included in the set of 74 genes responsive to 6 h mild stress. Only 3 of 74 genes that were differentially regulated after 6 h of mild water deficit were specific for this treatment. After short-term severe water deficit and long-term mild water deficit 175 and 177 genes were uniquely affected, respectively. The highest number of uniquely affected genes was detected for 24 h severe water deficit (1,750, 58% of all responsive genes).

Besides the pairwise comparisons a two-way analysis of variance was performed to determine treatment main effects, time main effects, and treatment by time interactions. This analysis yields genes differentially regulated

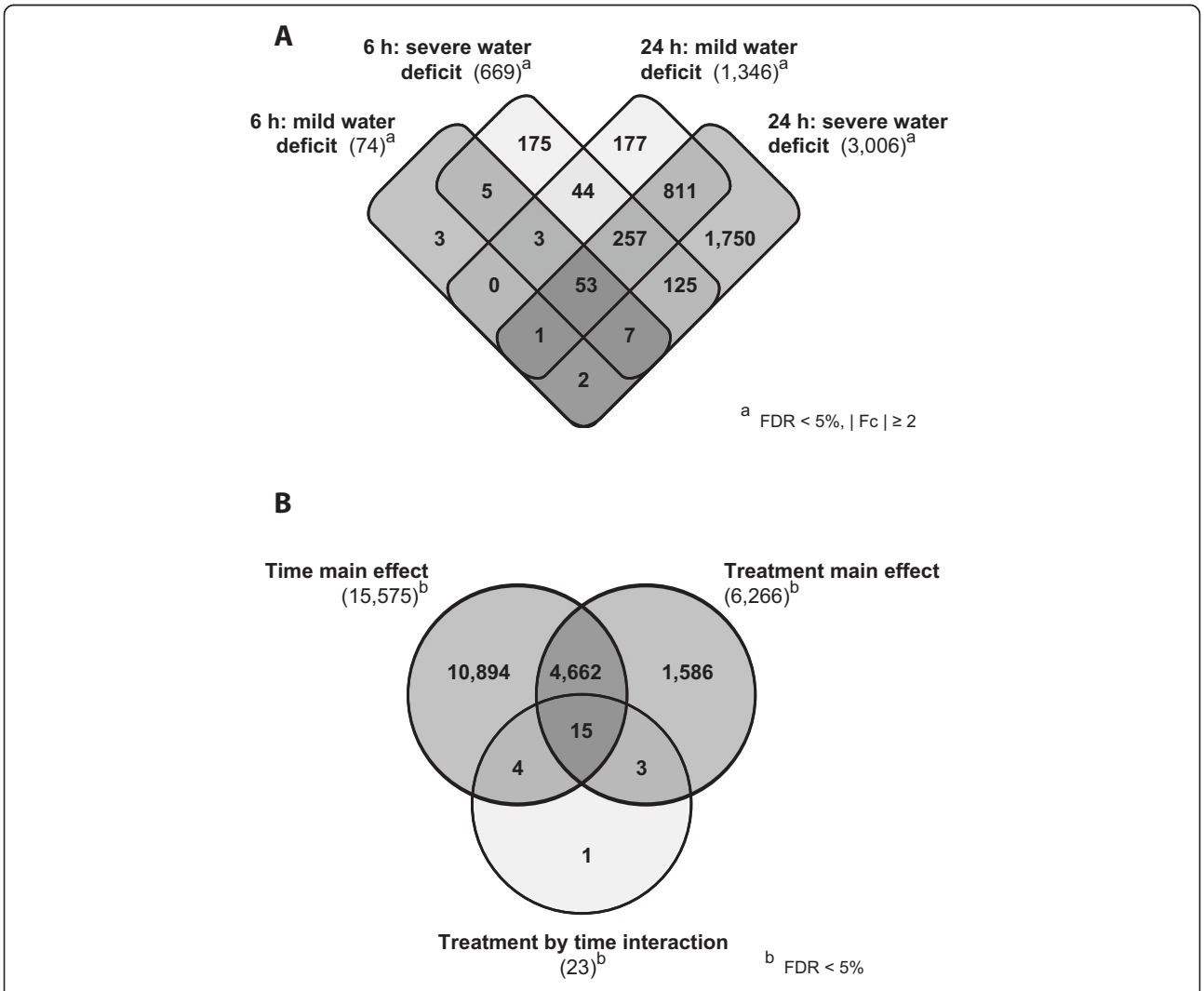


Figure 3 Overlap between different sets of water deficit responsive genes. A, Overlap between differentially expressed genes responsive to 6 h mild water deficit, 6 h severe water deficit, 24 h mild water deficit, and 24 h severe water deficit. (The total number of affected genes is given in brackets). **B**, Overlap between differentially expressed genes identified as time main effect, treatment main effect, and treatment by time interaction.

in response to both water deficit levels, both treatment periods, and to the combination of treatment intensity and duration, respectively. The latter gene set includes genes whose expression does not change initially but changes over time and treatment intensity [30]. As a result of this survey, 6,266 and 15,575 genes displayed treatment main effects and time main effects, respectively. Moreover, 23 genes displayed treatment by time interactions (Additional file 4). Between the three gene sets some overlap was identified: 15 genes were present in all data sets representing 65% (15/23) of the genes that displayed treatment by time interactions. Furthermore, 75% of the genes that showed treatment main effects were among the genes displaying time main effects (Figure 3B).

Functional categorization of stress responsive genes

An overview of the metabolic processes regulated by water deficit was generated with Mapman [31] (Figure 4). Only few of the 6 h mild water deficit responsive genes (FDR <5%, |Fc| ≥2) were included in the metabolic pathways overview, mainly in minor CHO metabolism (sugar and sugar derivate metabolism) and amino acid metabolism (Figure 4A). In response to more intense water deficit treatment, genes involved in major CHO metabolism (biosynthesis and degradation of starch and sucrose), cell wall metabolism and secondary metabolism were differentially regulated (Figure 4B). 24 h of treatment increased the number of responsive genes in all pathway categories with many differentially regulated genes annotated in major

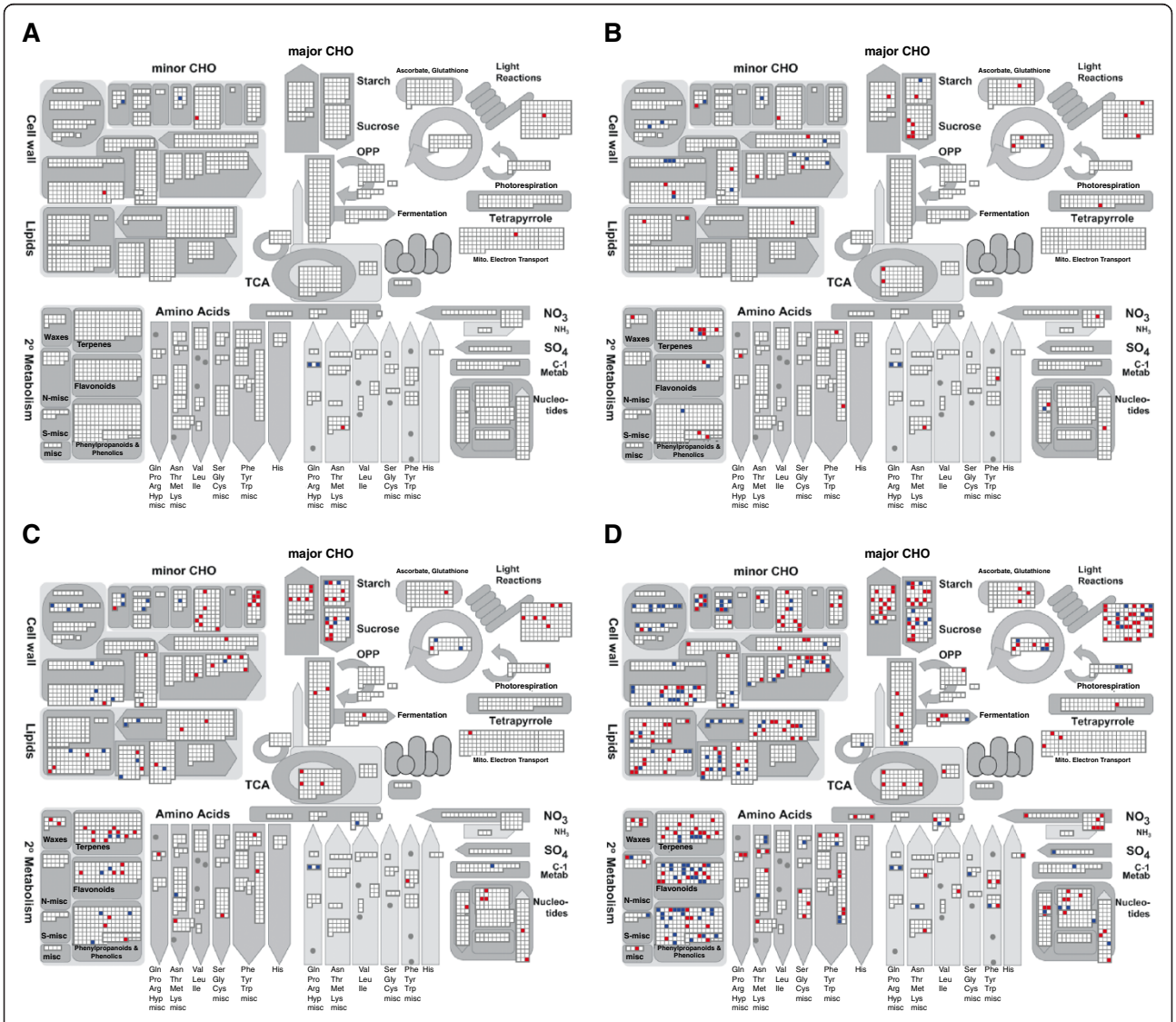


Figure 4 Overview of metabolic responses to water deficit. Genes differentially regulated (FDR <5%, |Fc| ≥2) in response to **A**, 6 h mild water deficit, **B**, 6 h severe water deficit, **C**, 24 h mild water deficit, and **D**, 24 h severe water deficit as visualized by Mapman. 74, 669, 1,346, and 3,006 differentially expressed genes corresponded to 80, 715, 1,431, and 3,210 unique transcript identifiers imported in Mapman [31]. Among those, 9, 66, 131, and 380 entities are visible in metabolic overviews. Transcripts which were up and down-regulated are represented in red and blue, respectively.

and minor CHO metabolisms, cell wall metabolism, lipid metabolism, and secondary metabolism (Figure 4C, D). An additional biochemical pathway analysis with Plant MetGenMAP [32] classified water deficit responsive genes into analogical pathways. In total, responsive genes were involved in 187 different pathways (Additional file 5). Among those, 126 describe biosynthetic processes, primarily biosynthesis of amino acids, carbohydrates, hormones, lipids, and cell wall compounds. The remaining pathways include degradation/assimilation (39; mainly carbohydrates and amino acids), energy generation (17; e.g. TCA cycle, glycolysis), detoxification (1), or a combination of these (10). Of the only four common pathways between the data sets, three describe down-regulation of proline degradation and one up-regulation of methylglyoxal degradation (Additional file 5).

Differentially expressed genes (FDR <5%, |Fc| ≥2) were further functionally classified according to Gene Ontology (GO) terms using agriGO [33]. For 65% (48/74) of the genes responding to short-term mild water deficit and 52% (351/669) of the genes responsive to severe water deficit, at least one GO term was identified. In the 24 h water deficit treatments, GO terms were identified for 46% (622/1,346) of mild deficit and 52% of (1,551/3,000) severe deficit responsive genes (Additional file 2). A singular enrichment analysis (SEA) was performed to discover significantly overrepresented functional categories (FDR <5%) by comparing each annotated gene set to all expressed genes (for 13,618 of the 25,570 expressed genes GO terms were available). The analyses revealed enrichment of terms related to biological processes and molecular function but not of cellular components. In the two 6 h treatment groups one and seven terms were enriched in the molecular function category (Additional file 6). After 24 h of mild and severe water deficit, 15 and 23 GO terms were overrepresented, respectively (Additional file 6). All commonly and uniquely enriched GOs are summarized in Figure 5. The GO term "oxidoreductase activity" (GO:0016491) was overrepresented in all data sets and was the only significant term among 6 h mild deficit responsive genes. Other commonly overrepresented GO terms included the molecular functions "binding" ("iron ion binding" GO:0005506, "heme binding" GO:0020037, "tetrapyrrole binding" GO:0046906), "monooxygenase activity" (GO:0004497), and "electron carrier activity" (GO:0009055). These covered almost all additionally enriched categories of the short-term severe water deficit regulated genes. Most overrepresented GO terms among the genes responding to 24 h mild water deficit were identical to those responding to 24 h severe water deficit including two terms related to hydrolase activity, "peroxidase activity" (GO:0004601), and "oxidoreductase activity" (GO:0016684). After 24 h of severe water deficit treatment 10 additional GO terms were enriched among responsive genes; three categories related to stimulus responses ("response to stimulus" GO:0050896, "response to stress" GO:0006950, "response to chemical stimulus" GO:0042221), four terms referring to C compound metabolism ("carbohydrate metabolic process" GO:0005975, "disaccharide metabolic process" GO:0005984, and "glycoside metabolic process" GO:0016137, "photosynthesis, light reaction" GO:0019684), two categories describing transcriptional regulation ("transcription factor activity" GO:0003700, "transcription regulator activity" GO:0030528), as well as the molecular function "antioxidant activity" (GO:0016209).

GO enrichment analyses were also performed for treatment and time main effects and treatment by time interaction specific genes. For 8,609 (55%) of 15,575 time main effect, 3,826 (58%) of 6,266 treatment main effect genes, and 14 (61%) of 23 treatment by time interaction genes at least one GO term was assigned (Additional file 4). In contrast to the analysis of pairwise comparisons, comparison of these gene sets with the set of all expressed genes by SEA yielded many overrepresented cellular component categories. Among the treatment by time interaction genes no enriched term was found. Among treatment main effect and time main effect genes 10 and 16 categories were overrepresented, respectively (Additional file 7). Each of the terms enriched among treatment main effect genes was included in the enriched time main effect terms. Four terms were connected to protein synthesis ("translation" GO:0006412, "structural constituent of ribosome" GO:0003735, "ribonucleoprotein complex" GO:0030529, and "ribosome" GO:0005840). Further categories were "structural molecule activity" (GO:0005198), the cellular components organelles ("intracellular non-membrane-bounded organelle" GO:0043232, "non-membrane-bounded organelle" GO:0043228), cytosol ("cytoplasmic part" GO:0044444, "cytoplasm" GO:0005737), and "macromolecular complex" (GO:0032991). Furthermore, the time main effect terms included five transport-related terms ("hydrogen ion transmembrane transporter activity" GO:0015078, "monovalent inorganic cation transmembrane transporter activity" GO:0015077, "inorganic cation transmembrane transporter activity" GO:0022890, "proton-transporting ATP synthase complex" GO:0045259, "proton-transporting two-sector ATPase complex GO:0016469) and the nucleosome (GO:0000786) as additional cellular compartment.

Validation of differentially expressed genes by qRT-PCR

To independently confirm the RNA-Seq results a subset of nine genes differentially regulated in response to all treatments was selected for quantitative real-time PCR (qRT-PCR) analyses (Additional file 8). As a result, 30 of 36 RNA-Seq expression patterns were confirmed i.e. trends of

Figure 5 Cross-comparison of enriched GO terms among differentially expressed genes in response to 6 h mild water deficit, 6 h severe water deficit, 24 h mild water deficit, and 24 h severe water deficit by single enrichment analysis (SEA). Different colors in the right columns represent the different significance levels of the overrepresentation; yellow: FDR <0.05, orange: FDR <0.01, red: FDR <0.001. Superscript letters indicate higher-ranking GO terms (GO level 2); [a]Catalytic activity (GO:0003824); [b]Binding (GO:0005488); [c]Antioxidant activity (GO:0016209); [d]Metabolic process (GO:0008152); [e]Response to stimulus (GO:0050896).

gene expression were identical in the sequencing and qRT-PCR approach. For nine of the 30 data points the analysis did not have enough statistical power to detect significant differences while for 21/30 data points statistically significant differences (Student's *t*-test: *p*-value <0.05) were measured.

Discussion

To promote plant survival, roots are often able to continue growth at low water potentials that completely inhibit shoot elongation [9,25,26]. As a result, roots might escape dry soil layers and reach deeper water resources. When water availability is so low to be life threatening, yield will be very low even if plants survive. Thus, from an agronomic viewpoint, mere survival of grain crops is irrelevant [34]. Therefore rather moderate water deficits were chosen for the present study. Water potentials of -0.2 MPa and -0.8 MPa were used to simulate mild and severe water deficit conditions, respectively. Short-term (6 h) water deficit treatment did not affect root elongation in the present study. This was expected as marked differences in steady maize root elongation rates between control conditions and low water potentials were observed only after 10 h of treatment [9]. Similarly, 24 h of mild water deficit did not

impair primary root growth, while severe water deficit significantly reduced primary root elongation. These findings are comparable to results by Sharp and co-workers [9,25] whereas Westgate and Boyer [26] measured considerable growth reduction only at water potentials lower than -0.8 MPa. This discrepancy might at least in part be explained by the different maize genotypes that were analyzed in these studies. Since 6 h of water deficit treatment did not affect root growth this experimental period was used to monitor gene expression changes at the initial phase of the water deficit response. Long-term treatment with mild water deficit had likewise no primary root growth effect suggesting that the root is able to tolerate this mild drought. In contrast, severe water deficit decelerated primary root elongation implying that the root had adapted by metabolic alterations.

RNA-Seq technology was utilized to compare the primary root transcriptomes of seedlings treated with low water potentials to those grown under control conditions. The seedlings used for the experiment were selected by their root length rather than by age minimizing variability of gene expression due to developmental differences [35]. In total, 25,570 genes were expressed under at least one of the time/treatment combinations. In the present study, 74 and 669 genes were identified that responded with expression changes to 6 h of mild and severe water deficit, respectively ($|Fc| \geq 2$). After 24 h, 1,981 and 3,006 genes were differentially expressed ($|Fc| \geq 2$) in response to mild and severe water deficit, respectively, reflecting complex acclimatization processes. A similar pattern of an increased number of responsive genes following more intense stress treatment has been reported for maize (Han21) seedling roots by Zheng, et al. [14], who considered 190 uniquely expressed sequence tags, and also for the whole maize (SRG-200) root system [15]. Comparative analyses in this study revealed that most (≥80%) of the short-term water deficit responsive genes were a subset of the long-term water deficit responsive genes. Similarly, a large overlap (≥90%) between mild and severe water deficit responsive genes was detected. An analogous pattern was observed for drought-stressed maize (SRG-200) and barley root systems although stress treatment was more intense in these studies [15,36]. However, neither maize nor barley leaves showed this pattern [15,19,36]. This indicates a root-specific response that is already established after the short period and low intensity of water deficit used in this study. Besides, there is further indication that the drought stress response is organ-specific. A microarray analysis of maize root, leaf, and shoot gene expression changes in response to drought stress revealed only a rather limited overlap between stress regulated genes [15]. In line with that observation the functional categories affected by the transcriptional changes were very distinct across tissues [15]. Similarly, a comparison of differentially expressed genes identified in the present study to genes identified by Kakumanu, et al. [37] via Illumina sequencing of RNAs isolated from basal leaf meristem tissues of maize plants grown under drought stress yielded very little overlap both on gene and functional category level. However, plant developmental stages (seedlings vs. adult plants, reproductive stage) and treatment period (24 h vs. 4 d) were quite different.

Among the water deficit responsive genes identified in this study several GO categories were overrepresented in comparison to all expressed genes. Cross-comparison of these categories resulted in one commonly enriched term, “oxidoreductase activity”. Its child term “monooxygenase activity” was also enriched among three of the four gene sets. Several previous studies detected differential expression of oxidoreductase and monooxygenase genes in response to water deficit in maize, barley, and soybean [14,18,36,38-40]. However, the complexity of these studies was limited to a few hundred or thousand entities and only individual genes or proteins of the oxidoreductase category were identified along with other differentially expressed genes. The reactions catalyzed by oxidoreductases can result in scavenging as well as in the generation of reactive oxygen species (ROS). Differential expression, including both up and down-regulation, of ROS scavengers in response to drought stress has also been shown in Arabidopsis [reviewed in 41]. While ROS can cause cellular damage [42], they also play an important role as signaling molecules [reviewed by 43] and are associated with growth and development [reviewed in 44]. For instance, the expression and protein abundance of an oxalate oxidase that produces hydrogen peroxide increased in stressed maize roots [18,40] and is probably involved in root growth regulation upon water deficit [45]. Other commonly enriched GO terms among water deficit responsive genes in the present study were “heme binding”, its ancestor term “tetrapyrrole binding” and the functionally related category “iron ion binding”. Likewise, genes annotated in the Mapman category “metal handling” were commonly up-regulated in response to water deficit (data not shown). In wheat roots an increase in total iron content during drought stress was detected [46]. Besides, the level of bound iron in soybean roots was substantially higher under water stress due to sequestration by ferritin proteins [10]. The expression of these and other metal-chelating proteins was also up-regulated in stressed maize roots [18]. Sequestration of metal ions is beneficial as it prevents the formation of the highly toxic hydroxyl radicals via the metal-dependent Haber-Weiss reaction or the Fenton reaction [41]. Additional studies are needed to fully elucidate the complex interaction of ROS metabolism and regulation of gene expression upon water deficit.

After 24 h of both mild and severe water deficit treatment, further enrichment of GO terms related to sugar

and general carbohydrate metabolic processes ("carbohydrate metabolic process", its child terms "polysaccharide metabolic process", "polysaccharide catabolic process", and "disaccharide metabolic process" as well as "glycoside metabolic process" and its child term "hydrolase activity") was detected. Accordingly, many of the long-term water deficit responsive genes were annotated in pathways of C compound metabolisms. Accumulation of carbohydrates in response to water deficit has been reported in several species and plant organs [reviewed in 47]. In maize primary roots, soluble carbohydrates together with proline account for osmotic adjustment which has an essential role in maintenance of root elongation at low water potentials [48,49]. It has been reported that both increases in proline synthesis and decreases in proline oxidation occur in response to low water potentials to increase proline concentrations [49]. Accordingly, three of four biochemical pathways conserved among water deficit responsive genes to all treatments included down-regulated proline oxidases. Besides functioning in osmotic adjustment, C compounds also participate in signaling and transcriptional and post-translational regulatory processes in metabolic and developmental programs [47] and the regulation of their synthesis/degradation is therefore tight and complex in response to environmental stresses.

An analysis of variance yielded differentially expressed genes in treatment main effect, time main effect, and treatment by time interactions. The highest number of genes (15,575) appeared in time main effect including more than half of all expressed genes. This suggests strong differences between root transcriptomes of seedlings differing in age as seedlings after long-term experiments were 18 h (about 20%) older and on average 1.6 cm (about 42%) longer than seedlings after short-term experiments. This is in accordance with the report of differences in protein abundance between two early stages (5 and 9 DAG) of maize root development in the inbred line B73 [50] which was also studied in the present survey. As most of the treatment-specific genes (4,662 of 6,266) are included in time effect genes, the pairwise comparisons between treatments within each time period might be more accurate. The test for treatment by time interaction could identify genes that do not differ across water deficit conditions initially but develop differences across conditions with longer treatment periods [30]. A comparable low number of 23 genes were assigned to this category. Hence, only a minor fraction of all water deficit responsive genes changed their expression from one time point to the other. This is in line with the large overlap observed between mild and severe deficit as well as between short and long-term treatment responsive genes in the pairwise comparisons.

A GO analysis revealed largely the same overrepresented categories among genes that display time main effects and treatment main effects. About one quarter and one half of the enriched GO terms, respectively, are related to protein biosynthesis including the biological process translation and the ribosome as cellular component. This points to a developmental function of these genes specific for time and for treatment effects as ribosomes are generally considered as housekeeping components of the cell and are involved in growth processes. Genes encoding ribosomal proteins are highly expressed in proliferating, elongating, and differentiating cells [51]. However, it was demonstrated for Arabidopsis that expression of these genes changes in response to abiotic stress [52,53]. Furthermore, a study in maize indicated that modifications of the translational machinery emerge in response to hypoxic stress [54]. In the present study the same genes functionally related to ribosomes were identified for time main effects and treatment main effects. This supports the concept of these genes playing a double role in normal development/growth and in water deficit response.

Conclusions

In the present study, hundreds of genes were identified that are differentially expressed in response to water deficits in maize seedling primary roots. The number of responsive genes was dependent on water deficit condition and duration and increased with intensification of treatment. The products of the differentially expressed genes are known to be involved in perception and signal transduction or confer adaptation and tolerance to low water potentials. Analyses of associated GO categories and underlying biochemical pathways connect the water deficit response to ROS and carbohydrate metabolisms and signaling, revealing their complex transcriptional regulation. Further functional analyses of these genes will promote our understanding of the molecular mechanisms underlying root adaptation to water deficit and enable targeted breeding strategies. Thus, more tolerant crop varieties, that can be cultivated with stable yield in dry regions as well as on fields with uneven soil moisture, might be generated.

Methods

Plant material and water deficit stress treatment

Seeds of the maize inbred line B73 were surface sterilized and germinated in paper rolls as described in Ludwig, et al. [55]. Four to 5 days (16 h light, 28°C; 8 h dark, 21°C) after germination seedlings with a primary root length of 2 to 4 cm were transferred to new paper rolls soaked with polyethylene glycol (PEG8000 M_r 7,300-9,000; Roth, Karlsruhe, Germany) solutions with a water potential of -0.2 MPa and -0.8 MPa for mild and severe water deficit, respectively, or distilled water for control experiments. Because manufactured PEG8000 includes a range of molar masses water potentials could not be calculated

directly. Osmolyte concentrations ($mol*l^{-1}$) of PEG solutions with defined mass concentrations ($g*l^{-1}$) were measured with an osmometer (OSMOMAT 030-D, Gonotec GmbH, Berlin, Germany). This data was used to estimate water potentials (water potential [MPa] = concentration [$mol*l^{-1}$] * gas constant [8.314 $Pa*l*mol^{-1}*K^{-1}$] * temperature [298.15 K]). Paper rolls with seedlings were incubated in aerated PEG solutions or distilled water for 6 h and 24 h (24°C, 24 h included 8 h darkness, 18°C). Before and after treatment, seedlings were photographed and primary root length was measured using WinRHIZO (http://www.regent.qc.ca/). Whole primary roots were harvested, immediately frozen in liquid nitrogen and stored at -80°C until RNA isolation. Experiments were performed in four biological replicates each consisting of 10 pooled roots.

RNA isolation and sequencing library preparation

Pooled primary roots were ground in liquid nitrogen, and RNA was extracted as previously described [56]. RNA quality was assessed via agarose gel electrophoresis and a Bioanalyzer (Agilent RNA 6000 Nano Chip, Agilent Technologies, Santa Clara CA, USA). Only samples with an RIN [RNA integrity number; 57] ≥9.1 were used for downstream analyses. The cDNA libraries for Illumina sequencing were constructed in accordance with the protocol of the manufacturer (TruSeq RNA Sample Preparation, Illumina, San Diego CA, USA). For sequencing, four libraries were pooled in one lane of a flow cell. Each library per lane was indexed by one of the adapters AR001, AR008, AR010, or AR011 (Additional file 1). The indexed libraries were loaded onto a flow cell according to an incomplete block design (generated with CycDesigN, http://www.vsni.co.uk/software/cycdesign; Additional file 1). Cluster preparation and single read sequencing were performed according to the manufacturer's instructions (HiSeq 2000, Illumina, San Diego CA, USA).

Processing and mapping of Illumina sequencing reads

Raw sequencing reads generated by the Illumina HiSeq 2000 system were initially processed and quality trimmed with SHORE (http://1001genomes.org/software/shore.html). Reads with more than 2 or 5 bases having quality scores ≤3 in the first 12 or 25 bases, respectively, were rejected [58]. Bases with quality scores ≤5 at the 3' end were trimmed until two succeeding bases with higher quality scores. Reads with ≥2 mismatches in adapter sequences were excluded. Only reads ≥40 bp were retained for subsequent analyses. Finally, adapter sequences were removed. Resulting reads had a length of 40 to 100 bp (≥60% of all reads were 100 bp long) and quality scores of 26 or higher at all base positions. Read mapping was performed with CLC Genomics Workbench (http://www.clcbio.com/products/clc-genomics-workbench/). All high quality reads were mapped to the maize B73 reference genome [RefGen_v2; 21,22] allowing large gaps of up to 50 kb to span introns. At least 75% of each read had to fit with 90% similarity to the reference to be mapped. Stacked reads i.e. redundant reads sharing the same start and end coordinate, sequencing direction, and sequence were merged into one. The remaining reads were projected to the filtered gene set (FGSv2; [21], release 5b.60) of the B73 reference genome derived from the maize genome sequencing project (MGSP) allowing a maximum of two mismatches for reads ≤56 bp. Longer reads had to fit at least with 80% of their length thereby comprising 90% similarity. Only those reads uniquely mapping to the reference data set were subsequently used for analyses.

Statistical procedures for analyzing differential gene expression

The read counts were normalized to RPKM (reads per kilobase of exon model per million mapped reads) values [59] and log2 transformed to meet the assumptions of linear models. Further statistical analysis followed the empirical Bayes approach of Smyth [60]. The mean-variance trend for log-counts was estimated and a weight assigned to each observation based on its predicted variance. The weights were then used in the linear modeling process to adjust for heteroscedasticity [61]. To borrow strength across genes in the estimation of the residual error variance, the empirical Bayes approach of Smyth [60] implemented in the R-package *limma* was used. A linear mixed model with lane effect was applied accounting for the lanes as incomplete blocks. The lane effect is considered as random effect, thus allowing the recovery of the inter-block information. The experimental setup allowed four comparisons of control groups against different water deficit levels: mild deficit, 6 h (1), severe deficit, 6 h (2), mild deficit, 24 h (3), and severe deficit 24 h (4). After computing these contrasts, resulting *p*-values of each contrast were corrected for multiplicity using the FDR-approach of Benjamini and Yekutieli [62]. Additionally, a two-way analysis of variance (ANOVA) was performed with treatment main effect, time main effect and treatment by time interaction according to the empirical Bayes approach of Smyth [60] implemented in the R-package *limma*. Computed *p*-values were corrected for multiplicity using the positive false discovery rate [pFDR; 63].

Gene Ontology (GO) and metabolic pathway analyses

GO functional categories were assigned to differentially expressed genes using the web-based agriGO software [33]. Enriched categories were computed using singular enrichment analysis (SEA) by comparing the list of differentially expressed genes to all expressed genes as described in Du, et al. [64]. Multiple comparison correction [62] was

performed and FDR controlled at 5%. Through the SEACOMPARE tool analysis results were combined for cross-comparisons. Differentially expressed genes were assigned to metabolic pathways and subsequently visualized using the Mapman software [31,65] based on the functional annotation file ZmB73_5b_FGS_cds_2011 [65]. Additionally, genes were assigned to biochemical pathways with the web-based Plant MetGenMAP [32,66] software according to the MaizeCyc database [version 2.1; 67].

Quantitative real-time PCR (qRT-PCR) analysis

To confirm gene expression levels detected by RNA-Seq, quantitative real-time PCR was performed in a Bio-Rad CFX 384TM Real-Time System (Bio-Rad, Munich, Germany) using gene-specific oligonucleotides (Additional file 9). The cDNA for qRT-PCR analyses was synthesized from 1 μg total RNA with the qScript cDNA SuperMix (Quanta Biosciences, Gaithersburg, MD, USA) using the same RNA samples as for the cDNA library construction. Each PCR reaction contained 4 μl MESA Blue qPCR™ Mastermix Plus for SYBR Assay no ROX (Eurogentec, Cologne, Germany), 1 μl cDNA sample and 100 nM gene-specific oligonucleotide primers to a final volume of 8 μl. The primer efficiency of each oligonucleotide was calculated using the following dilution series: 1, 1/2, 1/4, 1/8, 1/16, 1/32, 1/64, and 1/128. The relative expression levels of the transcripts were calculated with reference to the housekeeping gene *myosin* (Genbank AC: 486090G09.x1). Significant differences in gene expression levels were determined by a two-sided Student's *t*-test.

Additional files

Additional file 1: Overview of the sample distribution within the flow cell, biological replication, RNA-Seq output and mapping results.

Additional file 2: Comprehensive list of the 25,570 expressed genes, normalized expression values, fold-changes, *q*-values, and GO annotation.

Additional file 3: Fold-change distribution of differentially expressed genes (FDR <5%) between water deficit treatment and control groups.

Additional file 4: Comprehensive list of genes identified as time main effect, treatment main effect, and treatment by time interaction specific genes.

Additional file 5: List of pathways containing genes differentially regulated in response to water deficit.

Additional file 6: Comprehensive results of the singular enrichment analyses (SEA) among differentially expressed genes in pairwise comparisons between water deficit treatment and control groups.

Additional file 7: Comprehensive results of the singular enrichment analyses (SEA) of differentially expressed genes in time and treatment main effects and in treatment by time interaction.

Additional file 8: Comparison of RNA-Seq data and qRT-PCR results.

Additional file 9: List of oligonucleotide primers used for qRT-PCR experiments.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

NO performed the experiments, analyzed the data, and drafted the manuscript. AP and CM participated in experiments and data analysis. CL carried out the RNA sequencing and participated in library construction. WAM and HPP contributed to the statistical analysis. FH conceived the study, participated in its coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Jörg Hagmann (Max-Planck-Institute for Developmental Biology, Tübingen) for initial data analysis and Heiner Goldbach and Sven Berkau (University of Bonn, Bonn) for support with measuring water potentials. This project received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 289300.27.

Author details

[1]Institute of Crop Science and Resource Conservation (INRES), Crop Functional Genomics, University of Bonn, 53113 Bonn, Germany. [2]Institute for Crop Science, Bioinformatics Unit, University of Hohenheim, 70599 Stuttgart, Germany. [3]Department of Molecular Biology, Max-Planck-Institute for Developmental Biology, 72076 Tübingen, Germany.

Received: 14 May 2014 Accepted: 22 August 2014
Published: 29 August 2014

References

1. Wheeler T, von Braun J: **Climate change impacts on global food security.** *Science* 2013, **341**(6145):508–513.
2. Tester M, Langridge P: **Breeding technologies to increase crop production in a changing world.** *Science* 2010, **327**(5967):818–822.
3. Chandler VL, Brendel V: **The maize genome sequencing project.** *Plant Physiol* 2002, **130**(4):1594–1597.
4. **International Grains Council.** [www.igc.int]
5. Lobell DB, Schlenker W, Costa-Roberts J: **Climate trends and global crop production since 1980.** *Science* 2011, **333**(6042):616–620.
6. IPCC: **Summary for Policymakers**. In *The Physical Science Basis Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Edited by Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press; 2007.
7. Boyer JS: **Plant productivity and environment.** *Science* 1982, **218**(4571):443–448.
8. Nafziger ED, Carter PR, Graham EE: **Response of corn to uneven emergence.** *Crop Sci* 1991, **31**(3):811–815.
9. Sharp RE, Silk WK, Hsiao TC: **Growth of the maize primary root at low water potentials: I. Spatial distribution of expansive growth.** *Plant Physiol* 1988, **87**(1):50–57.
10. Yamaguchi M, Sharp RE: **Complexity and coordination of root growth at low water potentials: recent advances from transcriptomic and proteomic analyses.** *Plant Cell Environ* 2010, **33**(4):590–603.
11. Bray EA: **Plant responses to water deficit.** *Trends Plant Sci* 1997, **2**(2):48–54.
12. Krasensky J, Jonak C: **Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks.** *J Exp Bot* 2012, **63**(4):1593–1608.
13. Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Aklyama K, Taji T, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K: **Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray.** *Plant J* 2002, **31**(3):279–292.
14. Zheng J, Zhao J, Tao Y, Wang J, Liu Y, Fu J, Jin Y, Gao P, Zhang J, Bai Y, Wang G: **Isolation and analysis of water stress-induced genes in maize seedlings by subtractive PCR and cDNA macroarray.** *Plant Mol Biol* 2004, **55**(6):807–823.
15. Humbert S, Subedi S, Cohn J, Zeng B, Bi Y-M, Chen X, Zhu T, McNicholas P, Rothstein S: **Genome-wide expression profiling of maize in response to individual and combined water and nitrogen stresses.** *BMC Genomics* 2013, **14**(1):3.

16. Jia J, Fu J, Zheng J, Zhou X, Huai J, Wang J, Wang M, Zhang Y, Chen X, Zhang J, Zhao J, Su Z, Lv Y, Wang G: **Annotation and expression profile analysis of 2073 full-length cDNAs from stress-induced maize (*Zea mays* L.) seedlings.** *Plant J* 2006, **48**(5):710–727.
17. Poroyko V, Spollen W, Hejlek L, Hernandez A, LeNoble M, Davis G, Nguyen H, Springer G, Sharp R, Bohnert H: **Comparing regional transcript profiles from maize primary roots under well-watered and low water potential conditions.** *J Exp Bot* 2007, **58**(2):279–289.
18. Spollen W, Tao W, Valliyodan B, Chen K, Hejlek L, Kim J-J, LeNoble M, Zhu J, Bohnert H, Henderson D, Schachtman DP, Davis GE, Springer GK, Sharp RE, Nguyen HT: **Spatial distribution of transcript changes in the maize primary root elongation zone at low water potential.** *BMC Plant Biol* 2008, **8**(1):32.
19. Yue G, Zhuang Y, Li Z, Sun L, Zhang J: **Differential gene expression analysis of maize leaf at heading stage in response to water-deficit stress.** *Biosience Rep* 2008, **028**(3):125–134.
20. Metzker ML: **Sequencing technologies - the next generation.** *Nat Rev Genet* 2010, **11**(1):31–46.
21. **MaizeSequence.org FTP site.** [ftp://ftp.gramene.org/pub/gramene/maizesequence.org/]
22. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B: **The B73 maize genome: complexity, diversity, and dynamics.** *Science* 2009, **326**(5956):1112–1115.
23. Agarwal A, Koppstein D, Rozowsky J, Sboner A, Habegger L, Hillier L, Sasidharan R, Reinke V, Waterston R, Gerstein M: **Comparison and calibration of transcriptome data from RNA-seq and tiling arrays.** *BMC Genomics* 2010, **11**(1):383.
24. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y: **RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays.** *Genome Res* 2008, **18**(9):1509–1517.
25. Sharp RE, Poroyko V, Hejlek LG, Spollen WG, Springer GK, Bohnert HJ, Nguyen HT: **Root growth maintenance during water deficits: physiology to functional genomics.** *J Exp Bot* 2004, **55**(407):2343–2351.
26. Westgate ME, Boyer JS: **Osmotic adjustment and the inhibition of leaf, root, stem and silk growth at low water potentials in maize.** *Planta* 1985, **164**(4):540–549.
27. Zinselmeier C, Sun Y, Helentjaris T, Beatty M, Yang S, Smith H, Habben J: **The use of gene expression profiling to dissect the stress sensitivity of reproductive development in maize.** *Field Crops Res* 2002, **75**(2–3):111–121.
28. Bruce WB, Edmeades GO, Barker TC: **Molecular and physiological approaches to maize improvement for drought tolerance.** *J Exp Bot* 2002, **53**(366):13–25.
29. **NCBI Sequence Read Archive (SRA).** [www.ncbi.nlm.nih.gov/sra]
30. Nettleton D: **A discussion of statistical methods for design and analysis of microarray experiments for plant scientists.** *Plant Cell* 2006, **18**(9):2112–2121.
31. Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M: **Mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes.** *Plant J* 2004, **37**(6):914–939.
32. **Plant MetGenMAP.** [http://bioinfo.bti.cornell.edu/cgi-bin/MetGenMAP/home.cgi]
33. agriGO: **GO analysis toolkit and database for the agricultural community.** [http://bioinfo.cau.edu.cn/agriGO/analysis.php]
34. Sinclair TR: **Challenges in breeding for yield increase for drought.** *Trends Plant Sci* 2011, **16**(6):289–293.
35. Paschold A, Jia Y, Marcon C, Lund S, Larson NB, Yeh C-T, Ossowski S, Lanz C, Nettleton D, Schnable PS, Hochholdinger F: **Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents.** *Genome Res* 2012, **22**(12):2445–2454.
36. Ozturk ZN, Talamé V, Deyholos M, Michalowski C, Galbraith D, Gozukirmizi N, Tuberosa R, Bohnert H: **Monitoring large-scale changes in transcript abundance in drought- and salt-stressed barley.** *Plant Mol Biol* 2002, **48**(5–6):551–573.
37. Kakumanu A, Ambavaram MMR, Klumas C, Krishnan A, Batlang U, Myers E, Grene R, Pereira A: **Effects of drought on gene expression in maize reproductive and leaf meristem tissue revealed by RNA-seq.** *Plant Physiol* 2012, **160**(2):846–867.
38. Andjelkovic V, Thompson R: **Changes in gene expression in maize kernel in response to water and salt stress.** *Plant Cell Rep* 2006, **25**(1):71–79.
39. Fan X-D, Wang J-Q, Yang N, Dong Y-Y, Liu L, Wang F-W, Wang N, Chen H, Liu W-C, Sun Y-P, Wu J-Y, Li H-Y: **Gene expression profiling of soybean leaves and roots under salt, saline–alkali and drought stress by high-throughput Illumina sequencing.** *Gene* 2013, **512**(2):392–402.
40. Zhu J, Alvarez S, Marsh EL, LeNoble ME, Cho I-J, Sivaguru M, Chen S, Nguyen HT, Wu Y, Schachtman DP, Sharp RE: **Cell wall proteome in the maize primary root elongation zone. II Region-specific changes in water soluble and lightly ionically bound proteins under water deficit.** *Plant Physiol* 2007, **145**(4):1533–1548.
41. Mittler R, Vanderauwera S, Gollery M, Van Breusegem F: **Reactive oxygen gene network of plants.** *Trends Plant Sci* 2004, **9**(10):490–498.
42. Asada K, Takahashi M: **Production and scavenging of active oxygen in photosynthesis**. In *Photoinhibition*. Edited by Kyle DJ, Osborne CB, Arntzen CJ. Amsterdam: Elsevier; 1987:227–287.
43. Mittler R, Vanderauwera S, Suzuki N, Miller G, Tognetti VB, Vandepoele K, Gollery M, Shulaev V, Van Breusegem F: **ROS signaling: the new wave?** *Trends Plant Sci* 2011, **16**(6):300–309.
44. Kar RK: **Plant responses to water stress: role of reactive oxygen species.** *Plant Signal Behav* 2011, **6**(11):1741–1745.
45. Voothuluru P, Sharp RE: **Apoplastic hydrogen peroxide in the growth zone of the maize primary root under water stress. I. Increased levels are specific to the apical region of growth maintenance.** *J Exp Bot* 2013, **64**(5):1223–1233.
46. Price AH, Hendry GAF: **Iron-catalysed oxygen radical formation and its possible contribution to drought damage in nine native grasses and three cereals.** *Plant Cell Environ* 1991, **14**(5):477–484.
47. Muller B, Pantin F, Génard M, Turc O, Freixes S, Piques M, Gibon Y: **Water deficits uncouple growth from photosynthesis, increase C content, and modify the relationships between C and growth in sink organs.** *J Exp Bot* 2011, **62**(6):1715–1729.
48. Sharp RE, Hsiao TC, Silk WK: **Growth of the maize primary root at Low water potentials: II. Role of growth and deposition of hexose and potassium in osmotic adjustment.** *Plant Physiol* 1990, **93**(4):1337–1346.
49. Voetberg GS, Sharp RE: **Growth of the maize primary root at Low water potentials: III. Role of increased proline deposition in osmotic adjustment.** *Plant Physiol* 1991, **96**(4):1125–1130.
50. Hochholdinger F, Woll K, Guo L, Schnable PS: **The accumulation of abundant soluble proteins changes early in the development of the primary roots of maize (*Zea mays* L.).** *Proteomics* 2005, **5**(18):4885–4893.
51. Byrne ME: **A role for the ribosome in development.** *Trends Plant Sci* 2009, **14**(9):512–519.
52. Sormani R, Masclaux-Daubresse C, Daniele-Vedele F, Chardon F: **Transcriptional regulation of ribosome components are determined by stress according to cellular compartments in *Arabidopsis thaliana*.** *PLoS ONE* 2011, **6**(12):e28070.
53. Wang J, Lan P, Gao H, Zheng L, Li W, Schmidt W: **Expression changes of ribosomal proteins in phosphate- and iron-deficient *Arabidopsis* roots predict stress-specific alterations in ribosome composition.** *BMC Genomics* 2013, **14**(1):783.
54. Bailey-Serres J, Freeling M: **Hypoxic stress-induced changes in ribosomes of maize seedling roots.** *Plant Physiol* 1990, **94**(3):1237–1243.
55. Ludwig Y, Zhang Y, Hochholdinger F: **The maize (*Zea Mays* L.) *AUXIN/INDOLE-3-ACETIC ACID* gene family: phylogeny, synteny, and unique root-type and tissue-specific expression patterns during development.** *PLoS ONE* 2013, **8**(11):e78859.
56. Winz RA, Baldwin IT: **Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. IV. Insect-induced ethylene reduces jasmonate-induced nicotine accumulation by regulating Putrescine N-Methyltransferase transcripts.** *Plant Physiol* 2001, **125**(4):2189–2202.
57. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, Lightfoot S, Menzel W, Granzow M, Ragg T: **The RIN: An RNA integrity number for assigning integrity values to RNA measurements.** *BMC Mol Biol* 2006, **7**:3.
58. Ewing B, Green P: **Base-calling of automated sequencer traces using *Phred*. II. Error probabilities.** *Genome Res* 1998, **8**(3):186–194.
59. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-seq.** *Nat Methods* 2008, **5**(7):621–628.

60. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3**(1):Article3.
61. Law C: *Precision Weights for Gene Expression Analysis.* University of Melbourne: PhD thesis; 2013.
62. Benjamini Y, Yekutieli D: **The control of the false discovery rate in multiple testing under dependency.** *Ann Stat* 2001, **29**(4):1165–1188.
63. Storey JD: **A direct approach to false discovery rates.** *J Roy Stat Soc Ser B (Stat Method)* 2002, **64**(3):479–498.
64. Du Z, Zhou X, Ling Y, Zhang Z, Su Z: **agriGO: a GO analysis toolkit for the agricultural community.** *Nucleic Acids Res* 2010, **38**:W64–W70.
65. **MapMan.** [http://mapman.gabipd.org/web/guest/mapman]
66. Joung J-G, Corbett AM, Fellman SM, Tieman DM, Klee HJ, Giovannoni JJ, Fei Z: **Plant MetGenMAP: an integrative analysis system for plant systems biology.** *Plant Physiol* 2009, **151**(4):1758–1768.
67. **Metabolic pathways in maize: MaizeCyc.** [http://pathway.gramene.org/gramene/maizecyc.shtml]

doi:10.1186/1471-2164-15-741
Cite this article as: Opitz *et al.*: Transcriptomic complexity in young maize primary roots in response to low water potentials. *BMC Genomics* 2014 **15**:741.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central

the plant journal

SEB Society for Experimental Biology

The Plant Journal (2014) **79,** 729–740 doi: 10.1111/tpj.12578

Roothairless5, which functions in maize (*Zea mays* L.) root hair initiation and elongation encodes a monocot-specific NADPH oxidase

Josefine Nestler[1,†], Sanzhen Liu[2,3], Tsui-Jung Wen[4,‡], Anja Paschold[1], Caroline Marcon[1], Ho Man Tang[4,§], Delin Li[5], Li Li[2,6], Robert B. Meeley[7], Hajime Sakai[8], Wesley Bruce[7,¶], Patrick S. Schnable[2,5,9,*] and Frank Hochholdinger[1,*]

[1]*Institute of Crop Science and Resource Conservation, Crop Functional Genomics, University of Bonn, 53113 Bonn, Germany,*
[2]*Department of Agronomy, Iowa State University, Ames, IA 50011-3650, USA,*
[3]*Department of Plant Pathology, Kansas State University, Manhattan, KS 66506-5502, USA,*
[4]*Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, IA 50011-3650, USA,*
[5]*Department of Plant Genetics & Breeding, China Agricultural University, Beijing 100193, China,*
[6]*College of Agronomy, Northwest A&F University, Yangling 712100, China,*
[7]*Pioneer Hi-Bred International, Inc. – A DuPont Company, Johnston, IA 50131-0184, USA,*
[8]*DuPont Crop Genetics Research, Experimental Station, PO Box 80353, Wilmington, DE 19880-0353, USA, and*
[9]*Center for Plant Genomics, Iowa State University, Ames, IA 50011-3650, USA,*

Received 20 March 2014; revised 26 May 2014; accepted 28 May 2014; published online 5 June 2014.
*For correspondence (e-mails hochholdinger@uni-bonn.de, schnable@iastate.edu).
J.N. and S.L. contributed equally to this study.
†Present address: Crop, Environment and Livestock Division, Japan International Center for Agricultural Sciences (JIRCAS), Tsukuba, Ibaraki 305-8686 Japan.
‡Present address: Redmond, WA 98052 USA.
§Present address: Center for Cell Dynamics, Department of Biological Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD 21205 USA.
¶Present address: BASF, Research Triangle Park, NC 27709 USA.

SUMMARY

Root hairs are instrumental for nutrient uptake in monocot cereals. The maize (*Zea mays* L.) *roothairless5* (*rth5*) mutant displays defects in root hair initiation and elongation manifested by a reduced density and length of root hairs. Map-based cloning revealed that the *rth5* gene encodes a monocot-specific NADPH oxidase. RNA-Seq, *in situ* hybridization and qRT-PCR experiments demonstrated that the *rth5* gene displays preferential expression in root hairs but also accumulates to low levels in other tissues. Immunolocalization detected RTH5 proteins in the epidermis of the elongation and differentiation zone of primary roots. Because superoxide and hydrogen peroxide levels are reduced in the tips of growing *rth5* mutant root hairs as compared with wild-type, and Reactive oxygen species (ROS) is known to be involved in tip growth, we hypothesize that the RTH5 protein is responsible for establishing the high levels of ROS in the tips of growing root hairs required for elongation. Consistent with this hypothesis, a comparative RNA-Seq analysis of 6-day-old *rth5* versus wild-type primary roots revealed significant over-representation of only two gene ontology (GO) classes related to the biological functions (i.e. oxidation/reduction and carbohydrate metabolism) among 893 differentially expressed genes (FDR <5%). Within these two classes the subgroups 'response to oxidative stress' and 'cellulose biosynthesis' were most prominently represented.

Keywords: maize, root hairs, *rth5*, NADPH oxidase, RNA-Seq.

INTRODUCTION

Root hairs are tubular extensions of epidermis cells and comprise up to 77% of the root surface of cereals (Parker *et al.*, 2000). They are, therefore, considered to be instrumental for nutrient uptake and to support plant growth and development (Gilroy and Jones, 2000), a view that is supported by replicated field trials involving near-isogenic lines of the maize *roothairless 3* mutant, which conferred significant losses in grain yield (Hochholdinger *et al.*, 2008).

Generally, root hair bearing epidermis cells (trichoblasts) are short and contain denser cytoplasm, whereas non-root hair forming cells (atrichoblasts) are long and contain large vacuoles (Cormack, 1947). Several types of root hair patterning have been described. In dicot species such as

© 2014 The Authors
The Plant Journal © 2014 John Wiley & Sons Ltd

Arabidopsis thaliana, position-dependent root hair initiation leads to alternating longitudinal files of trichoblasts and atrichoblasts (Schiefelbein and Somerville, 1990). In contrast, some monocot species such as rice form trichoblasts via asymmetric cell divisions. Finally, random, unpredictable initiation without development of cytologically detectable distinct atrichoblasts and trichoblasts is observed in some monocot and dicot species such as maize and soybean (Clowes, 2000).

In Arabidopsis, epidermis cells overlaying two cortical cells (H cell position) develop into trichoblasts which give rise to root hairs. Cells overlaying only one cortical cell (N cell position) differentiate into hairless atrichoblasts (Schiefelbein and Somerville, 1990). In Arabidopsis, the molecular network involved in epidermis specification and thus root hair initiation is well understood (Dolan, 2001; Ishida *et al.*, 2008; Hochholdinger and Nestler, 2012). The Arabidopsis homeodomain transcription factor GLABRA2 (GL2) which is specifically expressed in atrichoblasts inhibits the expression of genes responsible for root hair formation (Masucci *et al.*, 1996). GL2 expression is activated by WEREWOLF (WER), a MYB-like transcription factor (Lee and Schiefelbein, 1999). The homologous MYB-like transcription factor CAPRICE (CPC) replaces WER in trichoblasts. CPC does not contain an activation domain and can therefore not activate *Gl2* expression (Wada *et al.*, 2002).

Root hair growth is divided into three phases: first, defined swelling to form a bulge, second, transition to tip growth and finally tip growth by oriented exocytosis (Dolan *et al.*, 1994). Recently, root hair formation from initiation to elongation in Arabidopsis was characterized via genome-wide transcriptome analyses of wild-type and 17 mutant lines leading to a complex model of transcriptional regulation (Bruex *et al.*, 2012).

Reactive oxygen species (ROS) and ROS-related proteins play an important role in root hair tip growth. Superoxide is produced by NADPH oxidases (NOXs) using oxygen as electron donor (Gapper and Dolan, 2006). Plant NADPH oxidases were identified as homologs of the mammalian $gp91^{phox}$, the catalytic subunit of the phagocyte NADPH oxidase, and named respiratory burst oxidase homologs (RBOHs) (Torres *et al.*, 1998). To date, the most thoroughly investigated *rboh* gene is *AtrbohC* or *ROOTHAIR DEFECTIVE 2* (*RHD2*) of Arabidopsis, which is involved in establishing hydrogen peroxide and superoxide peaks in growing root hair tips (Foreman *et al.*, 2003). The *rhd2* mutant lacks these high hydrogen peroxide and superoxide concentrations and the associated cytoplasmic Ca^{2+} influx (Wymer *et al.*, 1997). The membrane localized RHD2 protein is activated by Ca^{2+} (Takeda *et al.*, 2008). Although root hair bulges form in *rhd2* mutants they do not elongate (Schiefelbein and Somerville, 1990). Recently, it was demonstrated that *rhd2* mutants form macrotubules leading to a disturbance in mitosis in the root tip (Livanos *et al.*, 2012). In addition, the potential functions in root and root hair growth, several putative functions in shoot tissues and plant immune response have been assigned to *rboh* genes (Kwak *et al.*, 2003; Wong *et al.*, 2007; Galletti *et al.*, 2008; Lin *et al.*, 2009; Yi *et al.*, 2010; Yamauchi *et al.*, 2011; Wi *et al.*, 2012).

In maize, only a few mutants defective in root hair formation have been isolated. Two pleiotropic mutants, *disorganized aleurone layer 1* and *2* (*dil1* & *dil2*) display in addition to other developmental defects fewer root hairs and aberrant root hair morphology (Lid *et al.*, 2004). Three *roothairless* mutants (*rth1* to *rth3)* are defective at different stages of root hair formation (Wen and Schnable, 1994). Thus far only two maize genes involved in root hair elongation have been cloned. Mutants of the r*oothairless1* (*rth1*) gene condition short root hairs and a dwarf plant phenotype (Wen and Schnable, 1994) due to a defect in the encoded SEC homolog involved in polar exocytosis (Wen *et al.*, 2005). *Roothairless3* (*rth3*) encodes a monocot-specific COBRA-like protein that participates in cell-wall expansion and biosynthesis (Hochholdinger *et al.*, 2008); mutants in this gene result in very short root hairs defective in bulge formation (Wen and Schnable, 1994).

Here we describe the identification, cloning and characterization of the maize *roothairless5* gene controlling root hair initiation and outgrowth which encodes a monocot-specific NADPH oxidase involved in superoxide production.

RESULTS

The *roothairless5* gene controls both root hair length and density

An EMS mutagenesis screen yielded a recessive mutant (Schnable Laboratory Accession # 1350) that affects root hair development but has no other obvious effects on plant growth and development. Genetic crosses demonstrated that this mutant is not allelic to the previously described mutants *rth1*, *rth2* and *rth3* (Wen and Schnable, 1994). The affected gene and reference allele were designated *rth5* and *rth5-1*, respectively. Compared with wild-type (Figure 1a,c), the *rth5* mutant displays significantly shorter root hairs (Figure 1b,d). Root hair length and density of wild-type and *rth5* primary roots were analyzed via eSEM (environmental scanning electron microscopy). The length of root hairs on primary roots of 4-day-old (4 day) *rth5* mutants was decreased to 4% of wild-type length (Figure 1e). Moreover, root hair density of mutants was reduced to 64% of wild-type density (Figure 1f). In 10-day-old seedlings it is possible to observe root hairs on seminal and shoot-borne roots and the length and number of root hairs on these roots are also reduced in the mutant (Figure S1).

© 2014 The Authors

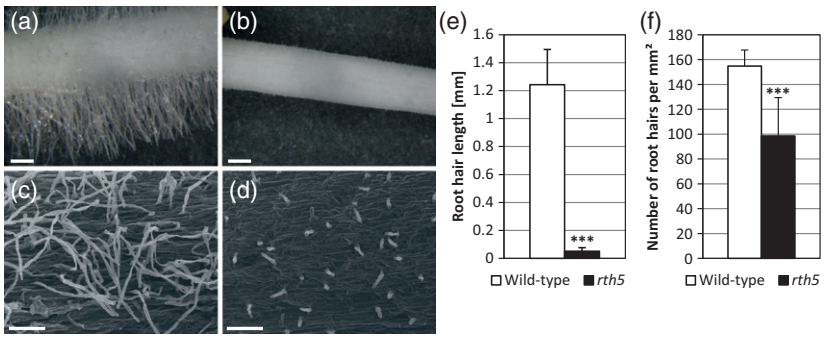


Figure 1. Phenotype of *roothairless5.*
(a–d) Binocular and eSEM pictures of 4-day-old wild-type (a, c) and *rth5* mutant (b, d) primary roots.
(e) Root hair length (n = 12).
(f) Root hair density per mm^2 (n = 3).
P-values were obtained via Student's *t*-test, *P*-value ***: ≤0.001. Error bars indicate standard deviation. Scale bars represent (a, b) 500 μm, (c, d) 100 μm.

Map-based cloning of *rth5* identified an NADPH oxidase

Using Sequenom-based bulked segregant analysis (BSA) (Liu *et al.*, 2010) the *rth5* locus was mapped to the long arm of chromosome 3. By analyzing F_2 and F_1BC populations (Experimental Procedures) the *rth5* gene was mapped to the interval 180.1–180.4 Mb of chromosome 3 (assembly version: ZmB73_AGPv1 release 4a53; http://ftp.maizesequence.org/release-4a.53/), flanked by insertion deletion polymorphism (IDP) markers IDP4064 and C3.184743 (Figure 2a). This interval harbors only five gene models, including GRMZM2G426953, encoding an NADPH oxidase (NOX), which contains a G-to-A transition at position 2462, resulting in a cysteine (C) to tyrosine (Y) conversion at amino acid position 821 close to the C-terminus of the protein (Figure 2b). This amino acid exchange is the result of a G-to-A transition relative to the B73 allele, which is characteristic of EMS-induced mutations (Greene *et al.*, 2003).

The affected cysteine residue is also conserved among divergent NADPH oxidases ranging from plants to yeast and human (Figure 2c). Hence, an alteration of this conserved residue in the predicted NAD substrate binding region might lead to a functional deficiency in the *rth5* mutant.

Confirmation of *rth5* identity via the generation of independent alleles

To confirm that the mutation in candidate gene GRMZM2G426953 indeed confers the phenotype of the *rth5* mutant, several independent putative *Mutator* alleles were isolated from Pioneer's trait utility system of corn (TUSC, Bensen *et al.*, 1995). Among these putative *Mu* insertion alleles, three displayed the roothairless phenotype and each of these alleles contained a *Mutator* insertion in the candidate gene as demonstrated by sequencing. Allele *rth5-3* contains a *Mu1* insertion in exon 5 (Figure 2b) while *rth5-2* and *rth5-4* contain *MuDR* insertions in exons 3 and 5, respectively (Figure 2b).

The gene structure

The B73 allele of the *rth5* gene consists of 14 exons and 13 introns (Figure 2b) encoding a 3792 bp mRNA (including the 5′ and 3′ UTR). This mRNA encodes an 852

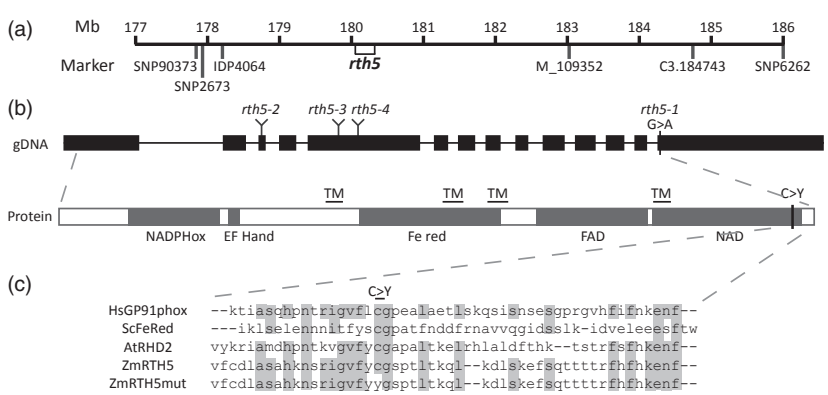


Figure 2. Mapping of *rth5.*
(a) Map-based cloning experiments positioned the *rth5* gene within a 280-kb interval of chromosome 3.
(b) A G-to-A (G→A) substitution in the last exon of the candidate gene GRMZM2G426953 results in a cysteine by tyrosine (C→Y) substitution in the protein sequence of the *rth5-1* allele at position 821. The positions of *Mu* transposon insertions in three additional alleles, *rth5-2*, *rth5-3*, and *rth5-4* are indicated.
(c) Alignment of the C-terminal 50 amino acids of human, yeast and Arabidopsis NADPH oxidases and wild-type and mutant RTH5.
At, *Arabidopsis thaliana*; EF hand, Ca^{2+} binding site; Fe red, Ferric reduction domain; FAD/NAD, FAD/NAD binding site; Hs, *Homo sapiens*; NADPHox, NADPH oxidase domain; Sc, *Saccharomyces cerevisiae*; TM, transmembrane; Zm, *Zea mays.*

© 2014 The Authors

amino acid protein with a predicted molecular weight of ~96 kDa.

The RTH5 protein is predicted to contain four transmembrane (TM) domains, two EF hand motifs, FAD and NAD co-factor binding sites, a ferric reductase domain, and the NADPH oxidase domain characteristic for NADPH oxidases (Figure 2b).

The NADPH oxidase family in maize

Based on sequence similarity to the RTH5 protein, 17 respiratory burst oxidase homologs (*rboh*) genes were identified in the maize filtered gene set (FGS, release: 4a53; http://ftp.maizesequence.org/release-4a.53/filtered-set/), a set of high confidence gene models. Phylogenetic reconstructions based on the full-length protein sequences of all known members of the maize, Arabidopsis, soybean and rice NADPH oxidase gene families (Figure 3) were performed. Two major groups of NADPH oxidases were observed. All NADPH oxidases exhibit diverse sequences in their N-terminus. The smaller group II is characterized by less conservation or the absence of the NADPH oxidase domain and a different conserved sequence in the NAD co-factor binding region (Figure S2). Interestingly, no soybean and only one rice and maize protein were included in group II, whereas eight Arabidopsis RBOH proteins belong to that subgroup. Along with two other maize and two rice RBOHs RTH5 is a member of a monocot-specific sub-clade of group I (Figure 3). No homolog of RTH5 resulting from an ancient genome duplication was found in the maize genome. *In silico* searches for syntenic homologs (genomeevolution.org/CoGe) identified the rice gene Os01g61880, which was also the closest homolog of RTH5 in the phylogenetic tree (Figure 3). RTH5 and its two closest maize homologs RBOH11 and RBOH13 are located in a subgroup of the phylogenetic tree characterized by the presence of a few additional amino acids at the end of the FAD co-factor binding site (Figure S2).

The rth5 gene is preferentially expressed in root hairs

Tissue-specific expression of *rth5* and all 17 *rboh* family members was examined via qRT-PCR in six different seedling tissues including the cap of the primary root, elongation zone, differentiation zone without root hairs, isolated root hairs, the coleoptilar node and the first leaf. *Rth5* transcripts accumulated in all tested tissues, but the significantly highest level was detected in root hairs (Figure 4b). Expression data of the remaining *rboh* family members in these tissues are summarized in Figure S3. Moreover, the qTeller tool (www.qteller.com) was used to compare the accumulation of *rth5* transcripts in RNA-Seq data from a wide range of tissues and organs. The highest expression values were observed in seedling roots (Figure S4). The relative expression levels of all 17 *rboh* genes in root hairs were determined via quantitative real-time PCR (qRT-PCR).

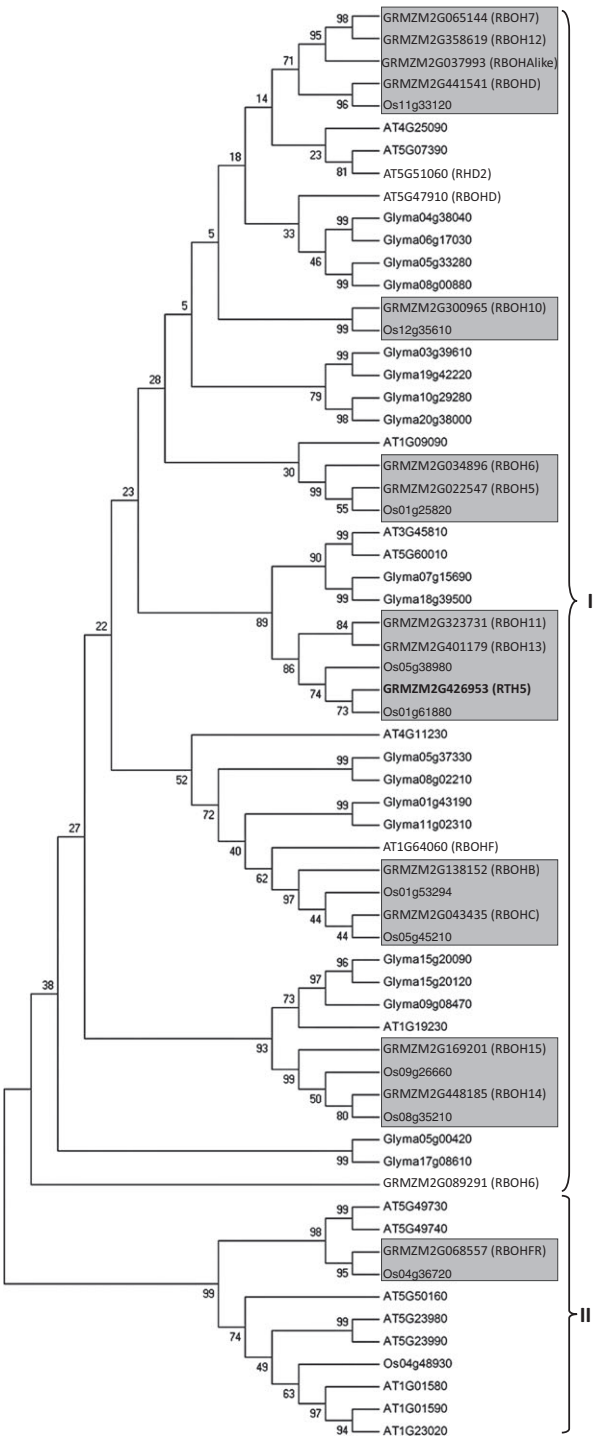


Figure 3. Phylogenetic tree of NADPH oxidases of maize, rice, Arabidopsis and soybean.
Two subfamilies can be distinguished by the presence or absence of several conserved domains in groups I and II (Figure S2). Monocot-specific clades/subclades are highlighted with gray boxes.

The *rth5* gene displayed the highest expression, accounting for 52% of the total *rboh* expression in maize root hairs (Figure 4a). The genes *rboh11* and *rboh13*, the closest

© 2014 The Authors

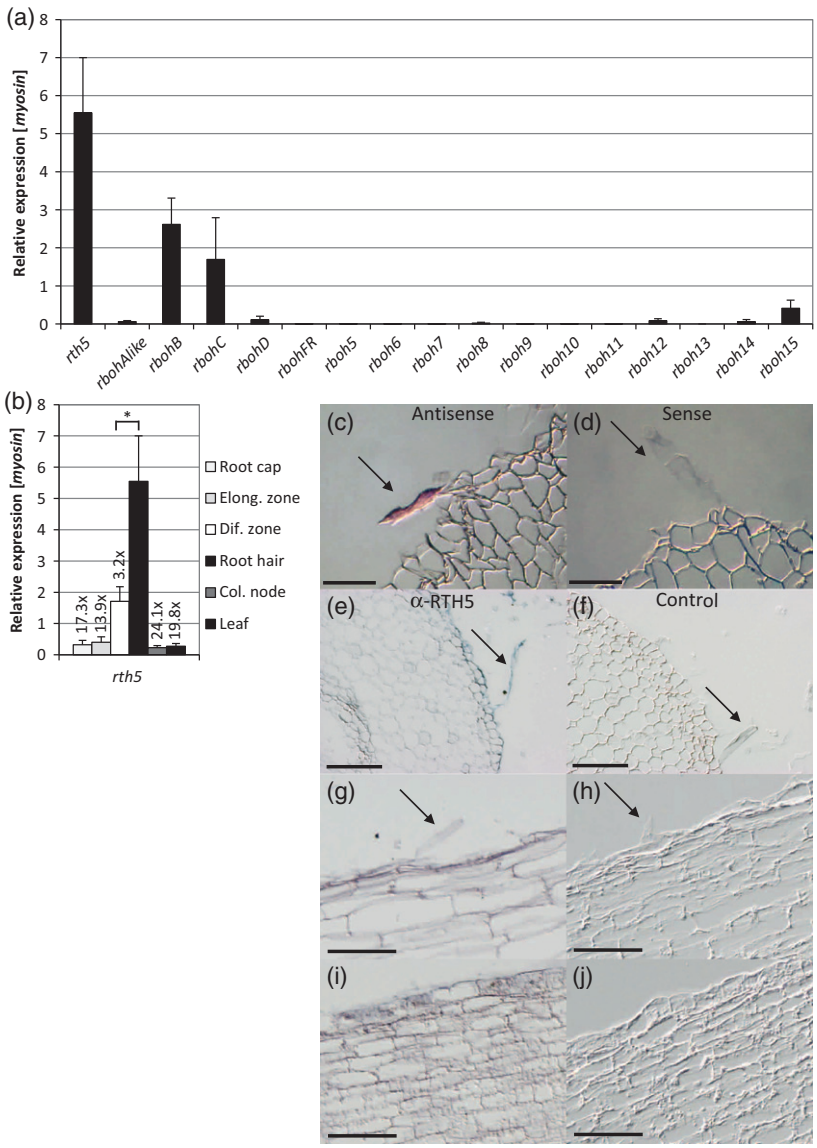


Figure 4. Tissue-specific expression of *rth5* transcripts and accumulation of RTH5 proteins.
(a) Expression of the *rth5* gene family members in root hairs by qPCR.
(b) *rth5* expression in different maize tissues by qPCR. Error bars indicate standard deviations; fold changes are indicated relative to root hairs. *P*-value obtained via Student's *t*-test, *P*-value *0.05, $n = 3$.
(c, d) RNA *in situ* hybridization of *rth5* antisense (c) and sense (d) probes on primary root cross-sections.
(e–j) Immunohistochemistry was performed on cross-sections (e, f) and longitudinal sections of the differentiation (g, h) or elongation zone (i, j). (e, g, i) Sections incubated with anti-RTH5 antibody. (f, h, j) Control sections incubated with buffer. Arrows indicate root hairs. Scale bars in (c, d) 50 μm, (e–j) 100 μm. For all experiments in (c–j), 3-day-old B73 seedling were used.

homologs of *rth5* were only very weakly expressed in root hairs.

RNA *in situ* hybridization experiments were used to study root tissue-specific expression patterns. Cross-sections hybridized with an *in vitro* transcribed *rth5* RNA antisense probe resulted in a signal in root hairs whereas the sense probe yielded no signal (Figure 4c,d).

The RTH5 protein is localized in epidermal cells of the primary root

Immunohistochemical experiments were performed to localize the RTH5 protein *in situ* (Figure 4e–j). In cross-sections of the differentiation zone, RTH5 antibody signals were detected in all epidermis cells and in root hairs (Figure 4e), whereas the control sections displayed no signal (Figure 4f).

In longitudinal sections an RTH5 signal was detected in root hairs and all epidermis cells of the root hair zone (Figure 4g). Interestingly, an RTH5 signal was also detected in longitudinal sections of the elongation zone, which does not form root hairs (Figure 4i). The control sections, showed no detectable color reaction (Figure 4h,j).

Both cross-sections and longitudinal sections displayed no difference in signal intensity or tissue organization between *rth5* and wild-type apart from the root hair defect (Figure S5).

The *rth5* mutant accumulates less ROS during root hair tip growth

Superoxide is a very short-lived radical that is rapidly converted into hydrogen peroxide (Figure 5a). Staining for both molecules was performed in wild-type and *rth5*

© 2014 The Authors

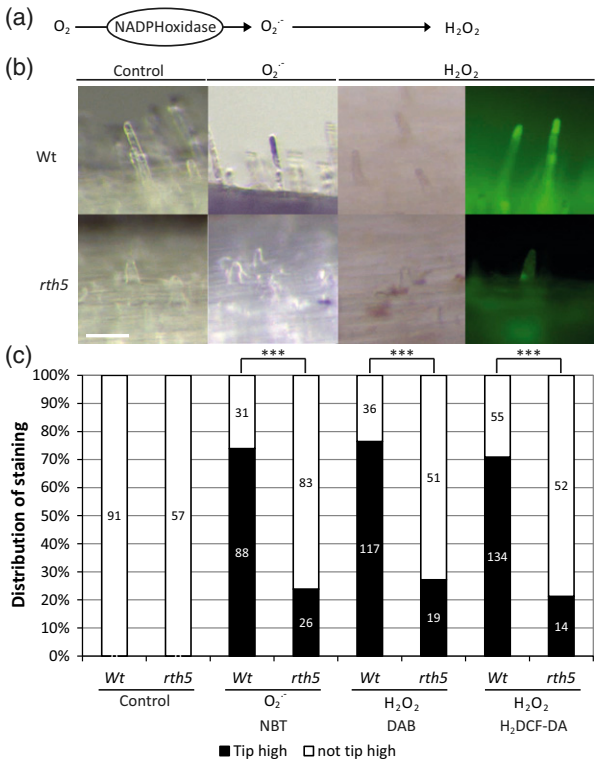


Figure 5. Reactive oxygen species (ROS) deficiency in *rth5* mutants.
(a) NADPH oxidases catalyze the formation of superoxide which is permutated into hydrogen peroxide.
(b) ROS staining of wild-type (*Wt*, upper lane) and *rth5* (lower lane) root hairs (from left to right): water control, NBT for superoxide, DAB, and H_2DCF-DA for hydrogen peroxide detection. A GFP filter was used for H_2DCF-DA detection. Scale bar 50 μm.
(c) Ratio of tip-high signals after staining in growing root hairs. Number of analyzed root hairs is indicated in the bars. *P*-values were obtained via Fisher's exact test, *P*-value ***: ≤0.001.

mutant seedlings (Figure 5b) by NBT for superoxide and DAB and H_2DCF-DA for hydrogen peroxide. NBT and DAB exclusively stained root hair cells, while H2DCF-DA not only exhibited the highest intensity of fluorescence in root hairs but also slightly stained other epidermis cells upon longer incubation. Subsequently, presence or absence of high ROS signals in root hair tips were quantified (Figure 5c). While 74% of wild-type root hairs displayed the superoxide (NBT) signal at the tip, the frequency of a tip-high signal was significantly reduced to 24% in *rth5* root hairs. Similarly, 77 and 71% of wild-type root hairs exhibited a high hydrogen peroxide signal in the tips of root hairs as detected by DAB or H_2DCF-DA, respectively, while only 27 and 21% of *rth5* root hairs displayed this signal specifically in root tips (Figure 5c). Fisher's exact test showed that these differences were significant at a *P*-value ≤0.001.

RNA-Seq of *rth5* mutants versus wild-type

To understand the influence of *rth5* on global patterns of transcription, RNA-Seq was performed on 6-day-old roots from the *rth5* mutant and wild-type seedlings (Experimental Procedures). Two biological replicates were analyzed for each genotype. Approximately 98% of raw reads from each sample passed the quality check and trimming procedure. In total, 3.7–4.7 million reads per replicate (~88% of the post-trimmed reads) were uniquely and confidently mapped to the B73 reference genome (Table S2) (Experimental Procedures). Among 21 007 genes that have a total of ≥20 RNA-Seq reads summed across all four biological replicates, 893 differentially expressed genes were identified at a false discovery rate (FDR) of 5%. Among those, 444 genes were preferentially expressed in wild-type primary roots, including 100 genes displaying a fold change (Fc) ≥2. Moreover, 449 genes were preferentially expressed in *rth5* primary roots including 195 genes displaying a Fc ≥2 (Table S3 and Figure S6). A gene ontology (GO) enrichment analysis of the 893 differentially expressed genes revealed statistically significant over-representation of three GO terms related to 'oxidation/reduction' and six GO terms related to 'carbohydrate metabolism' (Table S4). The 'oxidation/reduction' category contained four differentially expressed *rboh* genes, including *rth5.* While the *rth5* gene (GRMZM2G426953) was preferentially expressed in wild-type versus mutant primary roots (Fc: 1.4), the genes *rboh7* (GRMZM2G065144), *rboh6* (GRMZM2G034896) and *rbohAlike* (GRMZM2G037993) were preferentially expressed in mutant *rth5* roots. The relation of the 'oxidation/reduction' and 'carbohydrate metabolism' subgroups are summarized in Tables S5 and S6, respectively. Within these two classes, the subgroups 'response to oxidative stress' and 'cellulose biosynthesis' were most prominently represented. Overall, these RNA-Seq results are consistent with the view that *rth5/rbohA* functions in producing ROS ('response to oxidative stress') required for elongation ('cellulose biosynthesis') of root hairs.

DISCUSSION

The maize *rth5* mutant displays significantly shortened root hairs and reduced root hair density while aboveground development remains unaffected. Hence, *rth5* specifically controls both the elongation of root hairs and the specification of epidermis cells or the initiation of root hairs. More specifically, root hair elongation in the mutant *rth5* is affected during the transition from bulge to tip growth. In contrast to *rth5*, root hairs of the *rth3* mutant are characterized by disrupted bulges while *rth1* and *rth2* root hairs fail to elongate after transition to tip growth (Wen and Schnable, 1994). The *rth5* mutant resembles the phenotype of *rth2* and *rth3* (Wen and Schnable, 1994), which are specifically affected in root hair formation. In contrast, the maize mutants *rth1* (Wen and Schnable, 1994), *dil1* and *dil2* (Lid *et al.*, 2004) which are defective in root hair formation also display pleiotropic effects during development. The *rth1* mutant displays general growth abnormalities during development (Wen and Schnable,

© 2014 The Authors

1994) while *dil1* and *dil2* control cell division orientation in the aleurone, leaf and root epidermis. The maize mutant *rth5* forms fewer root hairs as compared to wild-type primary roots suggesting that the *rth5* gene is involved not only in root hair elongation but also in epidermis specification and/or root hair initiation.

In *rth5*, all root types including primary, seminal, lateral, and shoot-borne roots display defects in root hair formation, suggesting that the RTH5 protein is critical for root hair formation in all root types. A general defect in root hair formation affecting all root types has also been observed for the *rth1*, *rth2* and *rth3* mutants (Wen and Schnable, 1994; Hochholdinger *et al.*, 2008). In contrast to this general mechanism of root hair formation in all root types, the control of lateral root formation is root type-specific. For instance *rum1* (Woll *et al.*, 2005) and *lrt1* (Hochholdinger and Feix, 1998) display root type-specific defects which are confined to embryonic roots. In these mutants post-embryonic shoot-borne roots do not display any lateral root defects.

The RTH5 protein is characterized by several functional domains (Figure 2b) including four trans-membrane domains, responsible for the membrane embedding and two EF hand motifs, which are required for Ca^{2+} regulation. Moreover, the FAD and NAD domains confer binding to the co-factor FAD and the substrate NADPH, respectively. Regulation of the RBOHA/RTH5 NADPH oxidase by a mitogen-activated protein kinase cascade controlled by abscisic acid signalling was demonstrated in maize leaves (Lin *et al.*, 2009).

Based on the identification of the co-factor binding site and the functional assignment, the *rth5* gene is predicted to encode a NADPH oxidase (NOX). The NOX protein family is present in most eukaryotic species (Bedard *et al.*, 2007). The *rth5* mutant was induced by EMS, which resulted in a guanine to adenine substitution in the genomic sequence changing the cysteine (C) residue in amino acid position 821 to tyrosine (Y) (Figure 2b). Cysteine residues often form intra- or inter protein disulfide bridges to enable secondary structures. As this cysteine residue is highly conserved among a very diverse set of species (Figure 2c) it is likely that the C821Y mutation is disabling a functionally important protein domain. A point mutation of this conserved C in the human NADPH oxidase 2 (*NOX2*) gene results in the genetic disorder chronic granulomatous disease (Kawahara *et al.*, 2007). This disease affects cells of the immune system which have difficulties in forming reactive oxygen compounds, including superoxide used to kill certain ingested pathogens.

Phylogenetic reconstruction of RBOH (RESPIRATORY BURST OXIDASE HOMOLOG) proteins of rice, maize, soybean, and Arabidopsis revealed the existence of two subgroups. These subgroups are distinguished based on differences in the sequences of the NAD binding region, which are conserved within each group (Figure S2). Ten Arabidopsis RBOH proteins are classified to group I, which is consistent with previous annotations of superoxide producing NOX proteins (Kawahara *et al.*, 2007). Interestingly, all soybean proteins are assigned to group I, while eight of 18 members of the Arabidopsis RBOH protein family are classified in group II. According to their annotation, group II proteins are often connected to iron deficiency and might function as ferric reductases which are NOX proteins that do not produce superoxide (Ivanov *et al.*, 2012).

The nine members of the group I subclade that includes RTH5 are characterized by nine additional amino acids in the FAD co-factor domain as compared to the remaining group I proteins. The five monocot-specific proteins in this subclade lack 15 amino acids in their NADPH binding region. Both, the additional amino acids in the FAD, and the missing amino acids in the NAD region might confer altered binding to cofactors or substrates leading to specialized functions.

While several types of NOX proteins are found in mammals ranging from ancestral types to peroxidase containing DUOXs, in plants only homologs of EF-hand-containing NOX5-types have been identified (Bedard *et al.*, 2007). Human NOX proteins consist of two trans-membrane heterodimers ($gp91^{phox}$ and $p22^{phox}$) and four regulatory subunits ($p40^{phox}$, $p47^{phox}$, $p67^{phox}$, and Rac2) in the cytoplasm (Lam *et al.*, 2010). Thus far, no homologs for the regulatory subunits $p47^{phox}$, $p67^{phox}$, or $p22^{phox++}$ have been found in plants (Bedard *et al.*, 2007).

The human NOX2 has been shown to play a role in the production of superoxide as a first defense response to microorganisms invading neutrophils and macrophages (Nauseef, 2008; Lam *et al.*, 2010). Plant NOX proteins have been shown to function in plant immunity and several developmental processes. Pepper and tobacco NOX proteins were demonstrated to function in plant immunity (Yi *et al.*, 2010; Wi *et al.*, 2012). Aerchenyma formation after waterlogging was shown to be conferred by the maize NOX protein GRMZM2G300965 (Yamauchi *et al.*, 2011), which showed specific expression in the root cap and the elongation zone (Figure S4; RBOH10). Finally, Arabidopsis RBOHD and RBOHF are involved in ABA-dependent stomata closure (Kwak *et al.*, 2003) and lead to an oxidative burst following infection (Galletti *et al.*, 2008).

The Arabidopsis RBOH protein RBOHC (RHD2), which is distantly related to RTH5, also controls root hair growth. Arabidopsis plants defective in *RHD2* gene function form very short root hairs that do initiate bulges which do not elongate (Schiefelbein and Somerville, 1990). Moreover, such plants display stunted root growth (Foreman *et al.*, 2003), an effect that was not observed for *rth5* mutant roots. *RHD2* is expressed in root epidermal cells, but also in the cortical cells of the elongation- and differentiation zone (Foreman *et al.*, 2003).

Consistent with its mutation in the C-terminus of the protein, the expression of the *rth5-1* mutant reference

© 2014 The Authors

allele was only slightly reduced as compared with wild-type primary roots (Figure S5). Similarly, protein levels detected by immunohistochemistry and western blot analyses did not show any obvious change between wild-type and *rth5-1* mutants (see Figures 4 and S5). Although expression and protein abundance are not dramatically affected, this mutation significantly affects the function of the protein as illustrated by the mutant phenotype.

Expression of *rth5* was detected in several shoot and root tissues at low levels with expression maxima in root hairs (Figures 4 and S4). Similarly, the rice *rboh* genes are also expressed in multiple plant organs and tissues (Wong *et al.*, 2007). In maize root hairs *the rth5* gene accounts for 52% of the expression of all 17 *rboh* genes and hence the majority of NADPH oxidase transcripts in root hairs. The RTH5 protein was detected preferentially in root hairs and epidermis cells but weaker signals were also detected in root cortex cells (Figure 4). In two studies mainly focusing on the influence of hormones on NADPH oxidase activity in maize leaves it was demonstrated that expression of *rth5* (*rbohA*) is induced after abscisic acid and brassinosteroid treatment as well as hydrogen peroxide application (Lin *et al.*,2009; Zhang *et al.*, 2010). These experiments suggest that RTH5 may play a role in hormone-mediated oxidative defense in leaves. Further analyses will be required to determine whether RTH5 plays a similar role in roots, in addition to its known roles in root hair initiation and elongation.

Arabidopsis root hairs display high Ca^{2+} and superoxide concentrations in their growing tips (Foreman *et al.*, 2003). The tips of root hairs of the mutant *rth5* displayed significantly less superoxide staining and accumulate less hydrogen peroxide than wild-type plants (Figure 5). These results suggest that as a consequence of a functional defect of the RTH5 protein the *rth5* mutant is unable to establish high levels of ROS in the tips of root hairs. Not all growing tips of wild-type root hairs displayed staining for superoxide and hydrogen peroxide. Some root hairs may have died during the staining procedure; some may not have taken up the indicator chemicals. Consistent with expression data the presence of RTH5 homologous RBOH proteins may account for the residual ROS content detected in the mutant (Figure S3).

A model for the function of RTH5 in root hair elongation is summarized in Figure 6(a). Superoxide molecules in the root hair tips produced by RTH5 are transformed rapidly into hydrogen peroxide, which in turn is converted to hydroxyl radicals by apoplastic peroxidases (Liszkay *et al.*, 2003). These hydroxyl radicals are cleaving polysaccharides (Fry, 1998) and therefore cause cell-wall loosening, thereby enabling polarized cell growth (Bibikova *et al.*, 1998; Liszkay *et al.*, 2004).

In Arabidopsis the molecular interactions resulting in position-dependent root hair initiation are well characterized (reviewed in: Hochholdinger and Nestler, 2012). In contrast, the molecular network underlying random root hair patterning in maize remains enigmatic (Clowes, 2000). The reduced root hair density observed in mutant seedlings suggests that RTH5 is involved not only in root hair tip growth, but also in root hair initiation or epidermis specification. The presence of RTH5 in all epidermal cells is consistent with the hypothesis that superoxide is only produced upon activation of RTH5, which thus acts as a signal converter, similar to the human NOX2 (Lam *et al.*, 2010), which subsequently promotes root hair initiation. This is supported by the finding of selective root hair cell superoxide staining in the maize differentiation zone. In contrast, all epidermis cells display superoxide staining in the elongation zone (Liszkay *et al.*, 2004). To act in root hair initiation, we hypothesize that a yet unknown root hair initiation signal molecule activates RTH5 mediated superoxide production, leading to both the establishing of root hairs and their subsequent elongation. Residual root hair initiation in the *rth5* mutant might be controlled by RTH5-independent pathways. Due to the lack of functional RTH5 proteins these root hairs fail the transition to tip growth.

The human small GTPase RAC2 is necessary for the assembly of the functional NOX2 protein (Lam *et al.*, 2010). In Arabidopsis the ROP (RHO of plants) GTPase ROP2 is required for ROS formation in root hairs (Jones *et al.*, 2007). Another possible activation/regulation pathway involves Ca^{2+}. In Arabidopsis, it was demonstrated that NADPH oxidase activity leads to higher amounts of cytoplasmic Ca^{2+}, which in turn has a positive effect on their activity mediated through EF hands (Kimura *et al.*, 2012). Similarly, a trichoblast-specific signal might activate the RTH5 protein through a small GTPase or calcium-dependent protein kinase (CDPK) to facilitate trichoblast differentiation (Figure 6b). It is likely that RTH5 is also regulated by and/or participates in feedback loops involving these factors.

As a first step towards the elucidation of the RTH5-dependent molecular networks in young maize roots, a comparative RNA-Seq analysis was performed on 6-day-old wild-type versus *rth5* primary roots. Consistent with the suggested function of RTH5 only two major biological functions were overrepresented among the differentially expressed genes in a GO enrichment analysis. First, the GO term 'oxidation/reduction' was enriched. Genes in this class included four members of the *rboh* gene family, including *rth5* as discussed above. Moreover, this class also contains many genes related to oxidative stress including several peroxidases. As discussed above these peroxidases act downstream of RTH5 and facilitate the loosening of the elongating root hairs. This function connects the 'oxidation/reduction' GO class with the GO subgroup 'cellulose biosynthesis.' Cellulose is the most abundant plant polysaccharide (Carpita and McCann, 2000) and biopolymer *per se.* Cellulose biosynthesis is a crucial step in root hair elongation that follows peroxidase

© 2014 The Authors

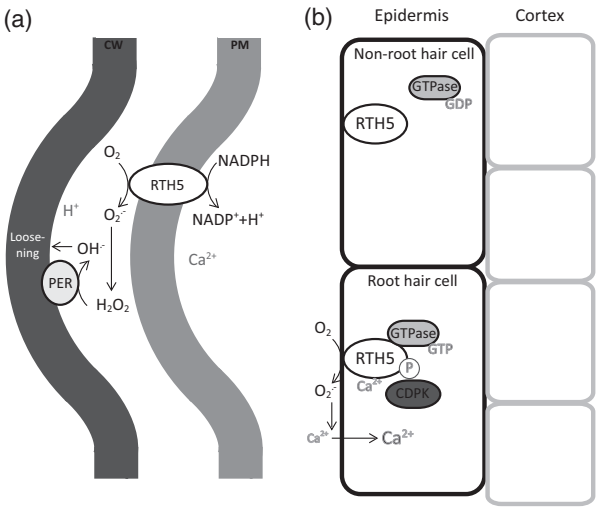


Figure 6. Models for the predicted function of RTH5 in root hair initiation and growth.
(a) Illustration of the putative role of RTH5 in tip growth by producing apoplastic superoxide, which is rapidly converted into hydrogen peroxide. Apoplastic peroxidases (PER) generate hydroxyl radicals which are cleaving celluloses and hemicelluloses leading to cell-wall loosening. The acidic cell wall (CW) allows oriented growth at the softened site. PM, plasma membrane.
(b) Suggested function of RTH5 in root hair initiation. A trichoblast-specific signal might activate the RTH5 protein through a small GTPase or calcium-dependent protein kinase (CDPK) to facilitate trichoblast differentiation.

mediated cell-wall loosening (Figure 6a). In Arabidopsis it has been demonstrated that mutations in cellulose biosynthesis genes can result in defective tip growth and aberrant elongation of root hairs (Favery *et al.*, 2001; Wang *et al.*, 2001; Park *et al.*, 2011). In summary, both overrepresented GO terms 'oxidation/reduction' and 'carbohydrate biosynthesis' are directly connected to maize root hair formation. Future reverse genetic analyses of candidate genes identified in this RNA-Seq analysis may therefore help to better define the genetic control of maize root hair formation.

EXPERIMENTAL PROCEDURES

Plant material and growth conditions

The mutation was initially induced by ethyl methanesulfonate (EMS) mutagenesis of the Pioneer inbred line 114748/AD. Mutants used in the present study were progeny of plants backcrossed 10 times to the inbred line B73 (Schnable laboratory stock 00g-1542-1).

Kernels were surface sterilized as described previously (Nestler *et al.*, 2011) and germinated in moist paper rolls at a 16 h light, 8 h dark cycle at 26°C in distilled water.

Binocular imaging and environmental scanning electron microscopy

A Zeiss Stemi SV8 Binocular (www.zeiss.com) coupled with a Powershot G2 camera (Canon) (www.canon.de) was used to document root hairs at a ×20 magnification. Root hair length was determined via WinRhizo Software (Regent Instruments Inc.) (www.regent.qc.ca). A Philips XL30 FEI environmental scanning electron microscope (eSEM; www.philips.de) was used for surface illustration of fresh 4-day-old primary roots without fixation or tissue drying.

Bulked segregant analysis

Sequenom-based BSA (Liu *et al.*, 2010) was used to map the *rth5* gene. In total, 89 individuals from an F_2-family were phenotypically divided into two bulks: mutant and non-mutant. Equal quantities of root tissue were pooled from each individual in each bulk. DNA was extracted from the two tissue bulks and then subjected to Sequenom-based SNP typing using ~1000 SNP markers.

Fine-scale mapping of *rth5*

Multiple markers located on the long arm of chromosome 3 were used to genotype 80 *rth5* mutant individuals from an F_2-population (Table S1). An *rth5* mutant individual in a B73 genetic background was crossed with the Mo17 inbred line and the resulting F_1-seeds were backcrossed to Mo17 to create an F_1BC population. Two single nucleotide polymorphism (SNP) markers, SNP2673 and SNP6262, were used to genotype 512 F_1BC individuals. In total, 65 recombinants were identified between SNP2673 and SNP6262, each of which was self-pollinated. The root hair phenotypes of self-pollinated progeny were scored to infer whether or not specific F_1BC_1 recombinants carried the *rth5* mutant allele. In parallel, the recombinants were genotyped with a series of molecular markers (Figure 2a). Based on an analysis of the resulting genotyping and phenotyping data, the *rth5* mutant was mapped to the 179.9–182.3 Mb interval of chromosome 3 in AGPv1. To further narrow down the *rth5* interval, a larger F_1BC population (n = ~5000) was genotyped with SNP markers SNP90373 and M_109352. In total, 440 recombinants between these two marker loci were identified and genotyped for *rth5* as described above. After genotyping all available recombinants (65 + 440) with 35 SNP markers (15 of which are co-dominant in this population) the *rth5* gene was mapped to the interval 180.1–180.4 Mb.

Identification of independent *rth5* mutant alleles

Three independent alleles, *rth5-2*, *rth5-3*, *and rth5-4*, were identified by a reverse genetic screen of *Mutator* stocks at Pioneer Hi-Bred. Confirmation of the candidate seedlings displaying a roothairless phenotype was performed by PCR mapping of the *Mutator* insertion with a general *Mutator* (*Mu*) oligonucleotide primer (Dietrich *et al.*, 2002) and an *rth5*-specific oligonucleotide primer (5′-GCACATCTCCCGGATAAATTG-3′). The amplification product contained 39 bp beyond the Mu TIR sequence sufficient to identify the *Mutator* element (Dietrich *et al.*, 2002). *Mu* insertions were found in positions 1300, 1931, and 1968 bp of the genomic *rth5* sequence counting from the ATG start codon in the alleles *rth5-2*, *rth5-3*, and *rth5-4*, respectively.

RTH5 protein domain prediction

Several bioinformatic tools were used to predict protein domains: InterProScan (EMBL-EBI) (www.ebi.ac.uk/Tools/pfa/iprscan), MyHits (SIB) (myhits.isb-sib.ch), SMART (University of Heidelberg) (smart.embl-heidelberg.de) and TMHMM (CBS of DTU, Technical University of Denmark) (www.cbs.dtu.dk/services/TMHMM). Domains predicted by at least two search tools are summarized in the protein domain structure (Figure 2b).

Phylogenetic analyses

The protein sequences of RTH5 homologs were obtained by blasting the RTH5 sequence against species-specific databases at

© 2014 The Authors
The Plant Journal © 2014 John Wiley & Sons Ltd,

maizesequence.org, rice.plantbiology.msu.edu, soybean.org, and TAIR.org. The DNA Baser software (Heracle BioSoft) (dnabaser.com) was used to combine all sequences into a Fasta file which was used for alignment and calculation of a phylogenetic tree by MEGA4.0 (Tamura *et al.*, 2007) using the Neighbor joining algorithm (Bootstrap, 1000 replications).

RNA *in situ* hybridization

Primary roots of 3-day-old seedlings were fixed and embedded in paraffin as described previously (Hochholdinger *et al.*, 2008). The RNA probe for hybridization was generated by amplification of 282 bp from the 3′-end of the *rth5* gene with the oligonucleotide primers forward 5′-GTGTACCCGAAGATCCGATG-3′, and reverse 5′-GACAGCTCGGGCAGAAAGAC-3′. The amplicon was cloned into the pGEM T-easy vector (Promega, www.promega.com) in sense and antisense directions. *In vitro* transcription of the probes and digoxigenin-labeling was performed using the SP6 polymerase (NEB, www.neb.com/nebecomm/default.asp) according to the manufacturer's protocol. RNA *in situ* hybridization was performed according to (Jackson, 1992). A Zeiss Axioplan 2 microscope (www.zeiss.com) in combination with a Photometrics Cool Snapcamera (Roper Scientific GmbH, www.roperscientific.com) was used for image acquisition.

Quantitative RT-PCR

For qRT-PCR, RNA was extracted from approximately 100 mg of different maize tissues with the RNeasy Plant Mini Kit (Qiagen, www.qiagen.com) according to the manufacturer's protocol, including on-column DNase I digestion (Fermentas, www.fermentas.de). Isolation and harvesting of root hairs was performed as described previously (Nestler *et al.*, 2011). cDNA was prepared from 500 ng total RNA by using the qScript cDNA Super Mix (Quanta BioSciences, www.quantabio.com). qRT-PCR experiments were performed using MESA Green or Blue qPCR Mastermix Plus for SYBR Assay no ROX kit (Eurogentec, www.eurogentec.com). All experiments were conducted in three biological replicates and three technical replicates per biological replicate in a CFX384 Real-Time PCR Detection System (Bio-Rad, www.bio-rad.com). The efficiency for every oligonucleotide primer pair (summarized in Table S7) was determined by a dilution series. Expression was calculated relative to the reference gene *myosin* (GenBank AC486090G09.x1), which has previously been used as expression standard in maize root assays (Dembinsky *et al.*, 2007).

Antibody preparation and immune histochemistry

A polyclonal antibody was produced by incubating rabbits with a specific RTH5 peptide with the amino acid sequence VAGMRPGRMTRMQSSAQM.

Sample preparation was slightly modified from the RNA *in situ* hybridization protocol. Fixation was performed using 4% formamide solution. After sectioning, the samples were deparaffined using RotiClear (Roth, www.carl-roth.de) and rehydrated using decreasing ethanol concentrations (90, 50, and 25%) in Microtubules stabilizing buffer (MtSB) (Albertini *et al.*, 1984). Unspecific binding was blocked by 3% bovine serum albumin (BSA) in 1 × MtSB, followed by incubation with the RTH5 antibody (1:100 in 1 × MtSB) overnight at 4°C. After six washing steps in 1 × MtSB the samples were incubated for 2 h with the secondary antibody, ZytoChem Plus AP Polymer anti-rabbit (ZytoMed, www.zytomed-systems.de) according to the manufacturer's protocol. After washing with a buffer containing 100 mM Tris buffer, 100 mM NaCl, and 50 mM $MgCl_2$, the signal was detected using the Roche NBT/BCIP stock solution (www.roche.de) by incubation in the dark for 5–10 min. The reaction was stopped by three water incubations. Images were obtained as described for RNA *in situ* hybridization experiments.

Detection of superoxide and hydrogen peroxide

Superoxide was detected by incubating root samples in 0.5 mM Nitro blue tetrazolium chloride (NBT) in 0.1 M KCl/0.1 M NaCl solution which forms an insoluble blue formazan precipitate upon reaction with superoxide (Bielski *et al.*, 1980). To detect hydrogen peroxide two dyes were used: DAB (3,3-diaminobenzidine) and 2,7-dichlorodihydrofluorescein diacetate (H_2DCF-DA). DAB forms a brown precipitate when oxidized by peroxidase activity (Thordal-Christensen *et al.*, 1997). Seedlings were incubated overnight in 1 mg ml^{-1} DAB dissolved in water. H_2DCF-DA emits a green fluorescence that corresponds to cytoplasmic H_2O_2 levels (Keston and Brandt, 1965). The stock solution was prepared by dissolving 1 mg H_2DCF-DA in 1 ml DMSO, and mixing with 1 ml H_2O. To obtain the working solution a dilution in 200 ml H_2O was prepared.

Three-day-old seedlings were incubated for 45 min in the dark in Petri dishes, each of which contained a detection solution. The signal was detected by mounting primary roots in water on a microscopic slide covered by a 60-mm cover slip. Excitation of H_2DCF-DA was performed at 488 nm. Emission was detected at 525 nm using a Zeiss Axioplan 2 microscope (www.zeiss.com).

RNA-Seq experiment

Seeds from an F_2-family segregating for *rth5* after nine generations of backcrossing to the inbred B73 were grown as described above. For this step, 3-cm root tips were collected from 6-day-old *rth5* mutants and wild-type siblings. Two mutant and two wild-type pools were collected. Each pool consisted of root tissues from six individuals. From each pool RNA was extracted using RNeasy Mini Kits (Qiagen) with DNase I treatment following the manufacturer's protocol. RNA quality was analyzed using a Bioanalyzer 2100 RNA Nanochip. RNA-Seq libraries were constructed using an Illumina RNA-Seq sample preparation kit following the manufacturer's protocol. The four libraries were sequenced on an Illumina HiSeq2000 at the Iowa State University DNA facility, generating 99-bp single-end reads and deposited in the GenBank database (accession no. SRP020528).

Trimming and mapping of RNA-Seq reads

Raw reads were subjected to quality checking and trimming to remove low quality bases using a custom trimming pipeline (Liu *et al.*, 2012). Trimmed reads were aligned to the B73 reference genome (ZmB73_RefGen_v2) using GSNAP (Wu and Nacu, 2010), allowing ≤2 mismatches every 50 and 2 bp tails per 50 bp. Only uniquely mapped reads were used for subsequent analyses. The read depth of each gene in the filtered gene set (ZmB73_FGSv2; http://ftp.maizesequence.org/current/filtered-set/, Release 5b.60) was computed based on the coordinates of mapped reads and the annotated locations of genes in the B73 reference genome.

Differential expression analysis of RNA-Seq

Genes with ≥20 RNA-Seq reads across all four biological replicates were tested for differential expression between the *rth5* mutant and wild-type using the R statistical package *QuasiSeq* (Iowa State University) (http://cran.r-project.org/web/packages/QuasiSeq). The

© 2014 The Authors

negative binomial *QLSpline* method implemented in the *QuasiSeq* package was used to compute a *P*-value for each gene. The 0.75 quantile of reads from each sample was used as the normalization factor (Bullard *et al.*, 2010). An approach for controlling for multiple testing (Benjamini and Hochberg, 1995) was used to convert *P*-values to *Q*-values. To approximately control the false discovery rate (FDR) at 5%, only genes with *Q*-values <0.05 were declared to be differentially expressed.

Gene ontology enrichment analysis of significantly differentially expressed genes

The GO of specific genes and the over-representation of GO terms were conducted using the single enrichment analysis of the AgriGo platform (http://bioinfo.cau.edu.cn/agriGO/analysis.php).

ACKNOWLEDGEMENTS

We thank Lisa Coffey (Iowa State University) for maintaining genetic stocks and conducting genetic crosses and Knut Wichterich (University of Bonn) for his support in generating the eSEM images of primary roots. We thank Mitzi J. Wilkening at the Genomic Technologies Facility at Iowa State University for conducting the Sequenom-based Bulk Segregation Analysis (BSA) experiment. This project was funded by the Deutsche Forschungsgemeinschaft (DFG) project HO2249/9 to F.H.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Phenotype of *roothairless 5* in different root types.

Figure S2. Alignment of RTH5 and its homologs from maize, soybean, rice and Arabidopsis.

Figure S3. Expression analysis of maize *rboh* genes.

Figure S4. Expression of *rth5* in various tissues.

Figure S5. *Rth5* transcript and RTH5 protein abundance in wild-type and *rth5* mutants.

Figure S6. RNA-Seq analysis of *rth5* versus wild-type primary roots.

Table S1. Summary of markers used in the fine-mapping experiment.

Table S2. Summary of RNA-Seq mapping.

Table S3. Summary of reads per gene and replicate and statistical analysis of differential expression between *rth5* mutants and wild-type roots.

Table S4. Gene ontology terms related to biological functions that are overrepresented among the differentially expressed genes.

Table S5. Relation of genes included in the GO term oxidation/reduction and its subgroups.

Table S6. Relation of genes included in the GO term carbohydrate metabolic process and its subgroups.

Table S7. Oligonucleotide primers used for qRT-PCR experiments.

REFERENCES

Albertini, D.F., Herman, B. and Sherline, P. (1984) *In vivo* and *in vitro* studies on the role of HMW-MAPs in taxol-induced microtubule bundling. *Eur. J. Cell Biol.* **33**, 134–143.

Bedard, K., Lardy, B. and Krause, K.H. (2007) NOX family NADPH oxidases: not just in mammals. *Biochimie*, **89**, 1107–1112.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B*, **57**, 289–300.

Bensen, R.J., Johal, G.S., Crane, V.C., Tossberg, J.T., Schnable, P.S., Meeley, R.B. and Briggs, S.P. (1995) Cloning and characterization of the maize *An1* gene. *Plant Cell*, **7**, 75–84.

Bibikova, T.N., Jacob, T., Dahse, I. and Gilroy, S. (1998) Localized changes in apoplastic and cytoplasmic pH are associated with root hair development in *Arabidopsis thaliana*. *Development*, **125**, 2925–2934.

Bielski, B.H.J., Shiue, G.G. and Bajuk, S. (1980) Reduction of nitro blue tetrazolium by CO_2^- and O_2^- Radicals. *J. Phys. Chem.* **84**, 830–833.

Bruex, A., Kainkaryam, R.M., Wieckowski, Y. *et al.* (2012) A gene regulatory network for root epidermis cell differentiation in *Arabidopsis*. *PLoS Genet.* **8**, e1002446.

Bullard, J.H., Purdom, E., Hansen, K.D. and Dudoit, S. (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*, **11**, 94.

Carpita, N. and McCann, M. (2000) The cell wall. In *Biochemistry and Molecular Biology of Plants* (Buchanan, B., Gruissem, W. and Jones, R., eds). Rockville, MD: American Society of Plant Physiologists, pp. 52–108.

Clowes, F.A.L. (2000) Pattern in root meristem development in angiosperms. *New Phytol.* **146**, 83–94.

Cormack, R.G.H. (1947) A comparative study of developing epidermal cells in white mustard and tomato roots. *Am. J. Bot.* **34**, 310–314.

Dembinsky, D., Woll, K., Saleem, M. *et al.* (2007) Transcriptomic and proteomic analyses of pericycle cells of the maize primary root. *Plant Physiol.* **145**, 575–588.

Dietrich, C.R., Cui, F., Packila, M.L., Li, J., Ashlock, D.A., Nikolau, B.J. and Schnable, P.S. (2002) Maize *Mu* transposons are targeted to the 5′ untranslated region of the *gl8* gene and sequences flanking *Mu* target-site duplications exhibit nonrandom nucleotide composition throughout the genome. *Genetics*, **160**, 697–716.

Dolan, L. (2001) How and where to build a root hair. *Curr. Opin. Plant Biol.* **4**, 550–554.

Dolan, L., Duckett, C.M., Grierson, C., Linstead, P., Schneider, K., Lawson, E., Dean, C., Poethig, S. and Roberts, K. (1994) Clonal relationships and cell patterning in the root epidermis of *Arabidopsis*. *Development*, **120**, 2465–2474.

Favery, B., Ryan, E., Foreman, J., Linstead, P., Boudonck, K., Steer, M., Shaw, P. and Dolan, L. (2001) KOJAK encodes a cellulose synthase-like protein required for root hair cell morphogenesis in *Arabidopsis*. *Genes Dev.* **15**, 79–89.

Foreman, J., Demidchik, V., Bothwell, J.H.F. *et al.* (2003) Reactive oxygen species produced by NADPH oxidase regulate plant cell growth. *Nature*, **422**, 442–446.

Fry, S.C. (1998) Oxidative scission of plant cell wall polysaccharides by ascorbate-induced hydroxyl radicals. *Biochem. J.* **332**(Pt 2), 507–515.

Galletti, R., Denoux, C., Gambetta, S., Dewdney, J., Ausubel, F.M., De Lorenzo, G. and Ferrari, S. (2008) The AtrbohD-mediated oxidative burst elicited by oligogalacturonides in *Arabidopsis* is dispensable for the activation of defense responses effective against *Botrytis cinerea*. *Plant Physiol.* **148**, 1695–1706.

Gapper, C. and Dolan, L. (2006) Control of plant development by reactive oxygen species. *Plant Physiol.* **141**, 341–345.

Gilroy, S. and Jones, D.L. (2000) Through form to function: root hair development and nutrient uptake. *Trends Plant Sci.* **5**, 56–60.

Greene, E.A., Codomo, C.A., Taylor, N.E. *et al.* (2003) Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics*, **164**, 731–740.

Hochholdinger, F. and Feix, G. (1998) Early post-embryonic root formation is specifically affected in the maize mutant *lrt1*. *Plant J.* **16**, 247–255.

Hochholdinger, F. and Nestler, J. (2012) Genetics and genomics of plant root development. *Brenner's Encyclopedia of Genet*, **5**, 349–352.

Hochholdinger, F., Wen, T.J., Zimmermann, R., Chimot-Marolle, P., Silva, O.D.E., Bruce, W., Lamkey, K.R., Wienand, U. and Schnable, P.S. (2008) The maize (*Zea mays* L.) *roothairless3* gene encodes a putative GPI-anchored, monocot-specific, COBRA-like protein that significantly affects grain yield. *Plant J.* **54**, 888–898.

Ishida, T., Kurata, T., Okada, K. and Wada, T. (2008) A genetic regulatory network in the development of trichomes and root hairs. *Annu. Rev. Plant Biol.* **59**, 365–386.

Ivanov, R., Brumbarova, T. and Bauer, P. (2012) Fitting into the harsh reality: regulation of iron-deficiency responses in dicotyledonous plants. *Mol. Plant*, **5**, 27–42.

© 2014 The Authors
The Plant Journal © 2014 John Wiley & Sons Ltd,

Jackson, D. (1992) *In situ* hybridization in plants. In *Molecular Plant Pathology: A Practical Approach* (Bowles, D.J., Gurr, S.J. and McPhereson, M., eds). Oxford: Oxford University Press, pp. 163–174.

Jones, M.A., Raymond, M.J., Yang, Z. and Smirnoff, N. (2007) NADPH oxidase-dependent reactive oxygen species formation required for root hair growth depends on ROP GTPase. *J. Exp. Bot.* **58**, 1261–1270.

Kawahara, T., Quinn, M.T. and Lambeth, J.D. (2007) Molecular evolution of the reactive oxygen-generating NADPH oxidase (Nox/Duox) family of enzymes. *BMC Evol. Biol.* **7**, 109.

Keston, A.S. and Brandt, R. (1965) Fluorometric analysis of ultramicro quantities of hydrogen peroxide. *Anal. Biochem.* **11**, 1–5.

Kimura, S., Kaya, H., Kawarazaki, T., Hiraoka, G., Senzaki, E., Michikawa, M. and Kuchitsu, K. (2012) Protein phosphorylation is a prerequisite for the Ca^{2+}-dependent activation of *Arabidopsis* NADPH oxidases and may function as a trigger for the positive feedback regulation of Ca^{2+} and reactive oxygen species. *Biochim. Biophys. Acta*, **1823**, 398–405.

Kwak, J.M., Mori, I.C., Pei, Z.M., Leonhardt, N., Torres, M.A., Dangl, J.L., Bloom, R.E., Bodde, S., Jones, J.D. and Schroeder, J.I. (2003) NADPH oxidase *AtrbohD* and *AtrbohF* genes function in ROS-dependent ABA signaling in *Arabidopsis*. *EMBO J.* **22**, 2623–2633.

Lam, G.Y., Huang, J. and Brumell, J.H. (2010) The many roles of NOX2 NADPH oxidase-derived ROS in immunity. *Semin. Immunopathol.* **32**, 415–430.

Lee, M.M. and Schiefelbein, J. (1999) WEREWOLF, a MYB-related protein in *Arabidopsis*, is a position-dependent regulator of epidermal cell patterning. *Cell*, **99**, 473–483.

Lid, S.E., Al, R.H., Krekling, T., Meeley, R.B., Ranch, J., Opsahl-Ferstad, H.G. and Olsen, O.A. (2004) The maize *disorganized aleurone layer 1* and *2* (*dil1, dil2*) mutants lack control of the mitotic division plane in the aleurone layer of developing endosperm. *Planta*, **218**, 370–378.

Lin, F., Ding, H., Wang, J., Zhang, H., Zhang, A., Zhang, Y., Tan, M., Dong, W. and Jiang, M. (2009) Positive feedback regulation of maize NADPH oxidase by mitogen-activated protein kinase cascade in abscisic acid signalling. *J. Exp. Bot.* **60**, 3221–3238.

Liszkay, A., Kenk, B. and Schopfer, P. (2003) Evidence for the involvement of cell wall peroxidase in the generation of hydroxyl radicals mediating extension growth. *Planta*, **217**, 658–667.

Liszkay, A., van der Zalm, E. and Schopfer, P. (2004) Production of reactive oxygen intermediates $O^{-}_{2}{}^{\cdot}$, H_2O_2, and $OH^{-\cdot}$ by maize roots and their role in wall loosening and elongation growth. *Plant Physiol.* **136**, 3114–3123.

Liu, S., Chen, H.D., Makarevitch, I., Shirmer, R., Emrich, S.J., Dietrich, C.R., Barbazuk, W.B., Springer, N.M. and Schnable, P.S. (2010) High-throughput genetic mapping of mutants via quantitative single nucleotide polymorphism typing. *Genetics*, **184**, 19–26.

Liu, S., Yeh, C.-T., Tang, H.M., Nettleton, D. and Schnable, P.S. (2012) Gene mapping via bulked segregant RNA-Seq (BSR-Seq). *PLoS ONE*, **7**, e36406.

Livanos, P., Galatis, B., Quader, H. and Apostolakos, P. (2012) Disturbance of reactive oxygen species homeostasis induces atypical tubulin polymer formation and affects mitosis in root-tip cells of *Triticum turgidum* and *Arabidopsis thaliana*. *Cytoskeleton (Hoboken)*, **69**, 1–21.

Masucci, J.D., Rerie, W.G., Foreman, D.R., Zhang, M., Galway, M.E., Marks, M.D. and Schiefelbein, J.W. (1996) The homeobox gene *GLABRA 2* is required for position-dependent cell differentiation in the root epidermis of *Arabidopsis thaliana*. *Development*, **122**, 1253–1260.

Nauseef, W.M. (2008) Biological roles for the NOX family NADPH oxidases. *J. Biol. Chem.* **283**, 16961–16965.

Nestler, J., Schutz, W. and Hochholdinger, F. (2011) Conserved and unique features of the maize (*Zea mays* L.) root hair proteome. *J. Proteome Res.* **10**, 2525–2537.

Park, S., Szumlanski, A.L., Gu, F., Guo, F. and Nielsen, E. (2011) A role for CSLD3 during cell-wall synthesis in apical plasma membranes of tip-growing root-hair cells. *Nat. Cell Biol.* **13**, 973–980.

Parker, J.S., Cavell, A.C., Dolan, L., Roberts, K. and Grierson, C.S. (2000) Genetic interactions during root hair morphogenesis in *Arabidopsis*. *Plant Cell*, **12**, 1961–1974.

Schiefelbein, J.W. and Somerville, C. (1990) Genetic control of root hair development in *Arabidopsis thaliana*. *Plant Cell*, **2**, 235–243.

Takeda, S., Gapper, C., Kaya, H., Bell, E., Kuchitsu, K. and Dolan, L. (2008) Local positive feedback regulation determines cell shape in root hair cells. *Science*, **319**, 1241–1244.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599.

Thordal-Christensen, H., Zhang, Z., Wei, Y. and Collinge, D.B. (1997) Subcellular localization of H_2O_2 in plants. H_2O_2 accumulated in papillae and hypersensitive response during barley-powdery mildew interaction. *Plant J.* **11**, 1187–1194.

Torres, M.A., Onouchi, H., Hamada, S., Machida, C., Hammond-Kosack, K.E. and Jones, J.D. (1998) Six *Arabidopsis thaliana* homologues of the human respiratory burst oxidase ($gp91^{phox}$). *Plant J.* **14**, 365–370.

Wada, T., Kurata, T., Tominaga, R., Koshino-Kimura, Y., Tachibana, T., Goto, K., Marks, M.D., Shimura, Y. and Okada, K. (2002) Role of a positive regulator of root hair development, CAPRICE, in *Arabidopsis* root epidermal cell differentiation. *Development*, **129**, 5409–5419.

Wang, X., Cnops, G., Vanderhaeghen, R., De Block, S., Van Montagu, M. and Van Lijsebettens, M. (2001) AtCSLD3, a cellulose synthase-like gene important for root hair growth in arabidopsis. *Plant Physiol.* **126**, 575–586.

Wen, T.J. and Schnable, P.S. (1994) Analyses of mutants of 3 genes that influence root hair development in *Zea mays* (Gramineae) suggest that root hairs are dispensable. *Am. J. Bot.* **81**, 833–842.

Wen, T.J., Hochholdinger, F., Sauer, M., Bruce, W. and Schnable, P.S. (2005) The *roothairless1* gene of maize encodes a homolog of *sec3,* which is involved in polar exocytosis. *Plant Physiol.* **138**, 1637–1643.

Wi, S.J., Ji, N.R. and Park, K.Y. (2012) Synergistic biosynthesis of biphasic ethylene and reactive oxygen species in response to hemibiotrophic *Phytophthora parasitica* in tobacco plants. *Plant Physiol.* **159**, 251–265.

Woll, K., Borsuk, L.A., Stransky, H., Nettleton, D., Schnable, P.S. and Hochholdinger, F. (2005) Isolation, characterization, and pericycle-specific transcriptome analyses of the novel maize lateral and seminal root initiation mutant *rum1*. *Plant Physiol.* **139**, 1255–1267.

Wong, H.L., Pinontoan, R., Hayashi, K. *et al.* (2007) Regulation of rice NADPH oxidase by binding of Rac GTPase to its N-terminal extension. *Plant Cell*, **19**, 4022–4034.

Wu, T.D. and Nacu, S. (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, **26**, 873–881.

Wymer, C.L., Bibikova, T.N. and Gilroy, S. (1997) Cytoplasmic free calcium distributions during the development of root hairs of *Arabidopsis thaliana*. *Plant J.* **12**, 427–439.

Yamauchi, T., Rajhi, I. and Nakazono, M. (2011) Lysigenous aerenchyma formation in maize root is confined to cortical cells by regulation of genes related to generation and scavenging of reactive oxygen species. *Plant Signal. Behav.* **6**, 759–761.

Yi, S.Y., Lee, D.J., Yeom, S.I., Yoon, J., Kim, Y.H., Kwon, S.Y. and Choi, D. (2010) A novel pepper (*Capsicum annuum*) receptor-like kinase functions as a negative regulator of plant cell death via accumulation of superoxide anions. *New Phytol.* **185**, 701–715.

Zhang, A., Zhang, J., Ye, N., Cao, J., Tan, M. and Jiang, M. (2010) ZmMPK5 is required for the NADPH oxidase-mediated self-propagation of apoplastic H_2O_2 in brassinosteroid-induced antioxidant defence in leaves of maize. *J. Exp. Bot.* **61**, 4399–4411.

© 2014 The Authors

Journal of Experimental Botany, Vol. 65, No. 17, pp. 4919–4930, 2014
doi:10.1093/jxb/eru249 Advance Access publication 13 June, 2014
This paper is available online free of all access charges (see http://jxb.oxfordjournals.org/open_access.html for further details)

Journal of Experimental Botany
www.jxb.oxfordjournals.org

RESEARCH PAPER

The *Aux/IAA* gene *rum1* involved in seminal and lateral root formation controls vascular patterning in maize (*Zea mays* L.) primary roots

Yanxiang Zhang[1], Anja Paschold[1], Caroline Marcon[1], Sanzhen Liu[2], Huanhuan Tai[1], Josefine Nestler[1], Cheng-Ting Yeh[3], Nina Opitz[1], Christa Lanz[4], Patrick S. Schnable[2,3] and Frank Hochholdinger[1,*]

[1] INRES, Institute of Crop Science and Resource Conservation, Crop Functional Genomics, University of Bonn, Friedrich-Ebert-Allee 144, 53113 Bonn, Germany
[2] Department of Agronomy, Iowa State University, Ames 50011-3650, Iowa, USA
[3] Center for Plant Genomics, Iowa State University, Ames 50011-3650, Iowa, USA
[4] Department of Molecular Biology, Max-Planck-Institute for Developmental Biology, 72076 Tuebingen, Germany

* To whom correspondence should be addressed. E-mail: hochholdinger@uni-bonn.de

Received 24 March 2014; Revised 3 May 2014; Accepted

Abstract

The maize (*Zea mays* L.) Aux/IAA protein RUM1 (ROOTLESS WITH UNDETECTABLE MERISTEMS 1) controls seminal and lateral root initiation. To identify RUM1-dependent gene expression patterns, RNA-Seq of the differentiation zone of primary roots of *rum1* mutants and the wild type was performed in four biological replicates. In total, 2 801 high-confidence maize genes displayed differential gene expression with Fc ≥2 and FDR ≤1%. The auxin signalling-related genes *rum1*, *like-auxin1* (*lax1*), *lax2*, (*nam ataf cuc 1 nac1*), the *plethora* genes *plt1* (*plethora 1*), *bbm1* (*baby boom 1*), and *hscf1* (*heat shock complementing factor 1*) and the *auxin response factors arf8* and *arf37* were down-regulated in the mutant *rum1*. All of these genes except *nac1* were auxin-inducible. The maize *arf8* and *arf37* genes are orthologues of *Arabidopsis MP/ARF5* (*MONOPTEROS/ARF5*), which controls the differentiation of vascular cells. Histological analyses of mutant *rum1* roots revealed defects in xylem organization and the differentiation of pith cells around the xylem. Moreover, histochemical staining of enlarged pith cells surrounding late metaxylem elements demonstrated that their thickened cell walls displayed excessive lignin deposition. In line with this phenotype, *rum1*-dependent mis-expression of several lignin biosynthesis genes was observed. In summary, RNA-Seq of RUM1-dependent gene expression in maize primary roots, in combination with histological and histochemical analyses, revealed the specific regulation of auxin signal transduction components by RUM1 and novel functions of RUM1 in vascular development.

Key words: Auxin, lateral roots, lignification, RNA-Seq, RUM1, vasculature, xylem.

Introduction

Maize (*Zea mays* L.) plays an important agronomic role as feed, food, and source of bioethanol. The complex root system of maize facilitates water and nutrient uptake and anchorage of the plant (Aiken and Smucker, 1996). The embryonic root system of maize consists of a primary root and a variable number of seminal roots. The post-embryonic root system comprises lateral and shoot-borne roots. In all maize root types lateral roots are initiated from phloem pole pericycle and endodermis cells (Hochholdinger *et al.*, 2004a).

The plant hormone auxin is a key regulator of many aspects of plant development. In roots, auxin controls lateral root development (Jansen *et al.*, 2012; Peret *et al.*, 2009a, b), root architecture (Overvoorde *et al.*, 2010) and vascular development (Wilson *et al.*, 1991; Uggla *et al.*, 1996; Mattsson *et al.*, 1999). It was suggested that polar auxin transport controls

© The Author 2014. Published by Oxford University Press on behalf of the Society for Experimental Biology.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

lateral root initiation in maize (Jansen *et al.*, 2012). Moreover, an auxin maximum was observed from late metaxylem to protoxylem in cells surrounding protophloem cells in the transgenic maize marker line DR5-RFP (Jansen *et al.*, 2012). This auxin maximum is likely to be essential for lateral root positioning and initiation. By contrast, inhibition of polar transport prevents an auxin response maximum from xylem to phloem pole cells in an anticlinal orientation and results in disorganized vascular tissues (Mattsson *et al.*, 1999; Jansen *et al.*, 2012).

Central regulators of auxin signalling include the *T*ransport *I*nhibitor *R*esponse *1* (TIR1) protein, members of the TIR1-like *A*uxin *F-B*ox (AFBs) family, *Aux*in/*I*ndole *A*cetic *A*cid (Aux/IAA) proteins, and *A*uxin *R*esponse *F*actor (ARF) proteins (Mockaitis and Estelle, 2008). At low intracellular auxin concentrations, Aux/IAA proteins act as transcriptional repressors that interact with ARF proteins via their domains III and IV. The ARF proteins of these complexes interact with *Aux*in *R*esponsive *E*lements (AuxREs) in the promoters of downstream genes, thereby repressing their transcription. By contrast, high auxin levels stabilize interactions between Aux/IAA proteins and SCF^{TIR1} E3 ubiquitin–ligase complexes, resulting in the degradation of Aux/IAA proteins by the 26S proteasome (Gray *et al.*, 2001; Tian *et al.*, 2002; Woodward and Bartel, 2005; Tan *et al.*, 2007). As a consequence, ARF proteins released from the Aux/IAA interactions can promote auxin-responsive target gene transcription.

The semi-dominant maize *rum1* (*rootless with undetectable meristem 1*) mutant is blocked in the initiation of embryonic seminal roots and post-embryonic lateral roots of the primary root (Woll *et al.*, 2005). The mutant *rum1* displays several auxin-related defects in root development. While exogenous auxin (αNAA) application induced additional lateral roots in wild-type primary roots, αNAA did not initiate any lateral roots in *rum1* (Woll *et al.*, 2005). Moreover, polar auxin transport in *rum1* primary roots was reduced by 83% compared with wild-type primary roots while it was not affected in the coleoptile of the mutant *rum1* (Woll *et al.*, 2005). The *rum1* gene (Genbank AC: GRMZM2G037368) encodes ZmIAA10, a member of the Aux/IAA protein family (Wang *et al.*, 2010). RUM1 interacts with the transcriptional activators ARF25 and ARF34 (von Behrens *et al.*, 2011). The mutated rum1 protein lacks 24 amino acids including the degron motif 'GWPPV' in domain II of rum1. Therefore, rum1 cannot interact with the SCF^{TIR1} E3 ubiquitin–ligase complexes which prevents its ubiquitin-mediated proteasomal degradation and resulting in constitutive repression of downstream gene expression (von Behrens *et al.*, 2011). In the present study, genes differentially expressed between wild-type and *rum1* primary roots were identified via RNA-Seq suggesting direct or indirect regulation of these genes by RUM1. In combination with histological and histochemical analyses, a RUM1-dependent gene network was identified and novel functions of RUM1 in vascular development were revealed.

Materials and methods

Plant material and growth conditions

Seeds of the F_8-generation of the maize mutant *rum1* and its homozygous wild type obtained by seven cycles of selfing of heterozygous plants were used in these experiments. Seeds were sterilized with 6% hypochlorite under vacuum at 500 mPa for 5 min, rinsed five times in distilled water, and germinated in paper rolls in a plant growth chamber at 28 °C with a 16/8 h light/dark regime at 21 °C (Woll *et al.*, 2005). For the RNA-Seq experiment, primary roots, 2 cm in length, of the mutant *rum1* and its homozygous wild type were harvested. Subsequently, 5 mm of the root tip, including the meristematic and elongation zones, were removed. Hence, only the differentiation zone of these roots was analysed by RNA-Seq. For auxin induction experiments, 5-d-old seedlings of the maize inbred line B73 that were grown under the same conditions were treated with the auxin analogue αNAA (α-naphthyl acetic acid; working solution 5 µM) for 3 h. The differentiation zone of the ~5 cm primary roots was harvested after 0, 1, 2, and 3 h of αNAA exposure (Taramino *et al.*, 2007) for subsequent analyses. For histological experiments, wild-type and *rum1* seedlings were grown in αNAA (0.1 µM) or the auxin transport inhibitor NPA (1-*N*-naphthylphthalamic acid, 10 µM).

RNA extraction and RNA sequencing

Total RNA was extracted via the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) and 3–4 roots were pooled per replicate, subsequently treated with RNase-free DNaseI (Fermentas, St Leon-Roth, Germany). RNA quality of all samples was assayed via a Bioanalyzer (Agilent Technologies, Boeblingen, Germany). As suggested by Agilent only samples with an RIN (RNA integrity number ≥7 were used for downstream analyses. Per genotype (wild type versus *rum1*) or treatment (auxin induction) four biological replicates were surveyed. The cDNA libraries for RNA-Seq were constructed using the TruSeq RNA sample preparation kit (Illumina Inc., San Diego, CA, USA). Each of the four replicates of each genotype was bar coded with one of the Illumina indices AR001, AR003, AR008, and AR009. Subsequently, the four wild-type replicates were pooled and loaded onto lane three of a flow cell, while the four *rum1* replicates were pooled on lane four of a flow cell. These lanes were then sequenced using a Genome Analyzer II (Illumina Inc., San Diego, CA, USA) according to the manufacturer's instructions resulting in 146 bp single-end reads.

RNA-Seq mapping and statistical analysis

RNA-Seq reads were trimmed according to Liu *et al.* (2012) and subsequently mapped to the B73 reference genome (ZmB73_RefGen_v2) (Schnable *et al.*, 2009) using GSNAP (Wu and Nacu, 2010). For subsequent analyses, only reads mapping uniquely (≤2 mismatches every 36 bp and fewer than five bases for every 75 bp as tails) to the reference genome were used. Genes with an average of at least five uniquely mapped reads across libraries and at least two samples with positive read counts (n=22 833) were tested for differential expression using the R package *QuasiSeq* (http://cran.r-project.org/web/packages/QuasiSeq). The 0.75 quartile of reads from each library was used as the normalization factor (Bullard *et al.*, 2010). The negative binomial *QLSpline* method implemented in the *QuasiSeq* package was used to compute a p-value for each gene (Lund *et al.*, 2012). A multiple test controlling approach (Nettleton *et al.*, 2006) was used to convert the p-values to q-values (Storey, 2002). Genes with FDR ≤1% and Fc ≥2 were declared to be differentially expressed (Li *et al.*, 2011). Subsequently, genes were classified into functional categories via MapMan (Thimm *et al.*, 2004). To determine if specific functional groups are overrepresented among the differentially expressed genes with reference to all expressed genes, the expected number of genes for each functional category was calculated based on the distribution of functional categories among all expressed genes. To determine if significantly more or less genes than expected were detected for each individual category a χ^2 test for independence with Yates' continuity correction was performed.

qRT-PCR expression analyses

cDNA for qRT-PCR was synthesized from 500 ng total RNA using the qScript cDNA SuperMix (Quanta Biosciences, Gaithersburg, MD, USA). qPCR was performed in a CFX384

Real-Time PCR detection system (Bio-Rad, Munich, Germany) for each of the four biological replicates in three technical replications in a total reaction volume of 8 µl using the MESA Green qPCR Mastermix Plus for SYBR Assay no ROX kit (Eurogentech, Cologne, Germany). Primers with an efficiency between 0.9 and 1.1 were used for qPCR, which was tested in a dilution series (1, 1/2, 1/4, 1/8, 1/16, 1/32, 1/64, 1/128). Gene expression for each genotype and each time point of auxin induction was assayed relative to *myosin* (Genbank AC: 486090G09.x1; primers: 486090G09.x1-5′; 486090G09.x1-3′) which has been previously used as a qPCR standard for expression analyses in maize roots (von Behrens *et al.*, 2011). The oligonucleotide primers *rum1*-fw and *rum1*-rv (*rum1*, GRMZM2G037368), *arf8*-fw and *arf8*-rv (*arf8*, GRMZM2G034840), *arf37*-fw and *arf37*-rv (*arf37*, GRMZM2G086949), *nac1*-fw and *nac1*-rv (*nac1*, GRMZM2G081930), *lax1*-fw and *lax1*-rv (*lax1*, GRMZM2G129413), *lax2*-fw and *lax2*-rv (*lax2*, GRMZM2G149481), *plt1*-fw and *plt1*-rv (*plt1*, GRMZM2G141638), *bbm1*-fw and *bbm1*-rv (*bbm1*, GRMZM2G366434), *hscf1*-fw and *hscf1*-rv (*hscf1*, GRMZM2G139082), *cad*-fw and *cad*-rv (*cad*, GRMZM2G443445), and *f5h*-fw and *f5h*-rv (*f5h*, GRMZM2G100158) were used for surveying expression of these genes (see Supplementary Table S5 at *JXB* online). Differential gene expression was determined by Student´s *t* tests (**p* ≤0.05; ***p* ≤0.01; ****p* ≤0.001; *n*=4).

Histology and histochemistry

Feulgen staining of whole roots was performed as previously described by Woll *et al.* (2005). For histological analyses and subsequent lignin staining, root fragments of the differentiation zone of the primary root were fixed in 4% paraformaldehyde in 1× phosphate buffer (containing 0.01 M Na_2HPO_4 and 0.01 M NaH_2PO_4, pH 7) for 2 h under vacuum at 100 mPa. Subsequently, the root fragments were sectioned by hand, mounted with water and then transferred into 65% glycerol. Transverse sections were analysed under bright field using a dissection microscope (PixCell IIe System, Zeiss). For lignin staining, transverse sections were incubated in 1% phloroglucinol in 12% HCl for 10 min, then transferred into acid solution (75% glycerol in 10% H_2SO_4) and observed under bright field conditions using a dissection microscope (PixCell IIe System, Zeiss).

Results

Comparative RNA-Seq analysis of rum1 *and wild-type primary roots*

The maize *rum1* mutant is impaired in lateral root initiation in the primary root. To understand the molecular network regulated by the Aux/IAA protein RUM1, ~2 cm long primary roots of the mutant *rum1* and homozygous wild-type siblings were subjected to an RNA-Seq analysis (Fig. 1A). At this early developmental stage no lateral root primordia were detectable in primary roots of wild-type seedlings by Feulgen staining (data not shown). Hence, at this stage, primary roots of wild-type and *rum1* were morphologically indistinguishable. Lateral roots in maize are formed in the differentiation zone of roots. Therefore, only the differentiation zone of primary roots, i.e. primary roots, after the removal of 5 mm of the root tip which comprises the meristematic zone and most of the elongation zone, were used for subsequent transcriptome profiling. A flow chart of the RNA-Seq experiment is provided in Fig. 1B. Sequencing of four biological replicates of wild-type and *rum1* libraries yielded, on average, ~15 million 146 bp single-end reads per sample (see Supplementary Table S1 at *JXB* online). After removal of bar-code tags from *rum1* and wild-type reads, on average, 76% and 73% of the trimmed reads mapped to unique positions of the maize B73 reference genome (ZmB73_RefGen_v2; see Supplementary Table S1 at *JXB* online), respectively. Of these reads, 94–95% were associated with high-confidence gene models of the 'filtered gene set' (4a.53_v2; maizesequence.org).

Identification of differentially expressed genes

The RNA-Seq reads from the differentiation zone of 2-cm-long maize primary roots aligned to 22 833 of 39 656 (58%) genes in the high confidence 'filtered gene set' (see Supplementary Table S2 at *JXB* online). The distribution of fold-changes versus the mean of gene expression in the differentiation zone of primary roots between *rum1* and the wild type demonstrated that differential gene expression was observed for genes with both low and high expression levels (Fig. 1C). In total, 8 309 (FDR ≤5%), 5 701 (FDR ≤1%), and 3 502 (FDR ≤0.1%) genes displayed differential expression at different significance levels irrespective of their fold-changes (Fig. 1D). Hence, these classes represented 36%, 25%, and 15% of all genes for which expression was detected. At FDR ≤1%, 2 801 (12% of all) genes displayed a Fc (fold change) ≥2 (see Supplementary Table S2 at *JXB* online), including 1 741 genes that were down-regulated (Fig. 1E, red dots) and 1 060 genes that were up-regulated in the *rum1* mutant (Fig. 1E, blue dots). Of those genes, 1 723 (62%) were functionally annotated using the MapMan software (see Supplementary Table S3 at *JXB* online), including 205 genes that were assigned to more than one functional class. To identify over- and underrepresented functional classes among the differentially expressed genes, the expected number of genes for each functional class was calculated based on all expressed genes and subsequently compared with the detected number of genes in this class. This analysis revealed that significantly more genes than expected ($p \leq 0.001$) were assigned to the classes *photosynthesis*, *cell wall*, *metal handling*, *secondary metabolism, miscellaneous*, and *DNA* (see Supplementary Table S4 at *JXB* online).

qRT-PCR confirmation of differentially expressed genes

To confirm the RNA-Seq results independently, expression levels of a subset of differentially expressed genes related to auxin signalling were analysed by qRT-PCR. Consistent with prior results (von Behrens *et al.*, 2011), the *Aux/IAA* gene *rum1* (GRMZM2G037368), used as a positive control, was down-regulated in the *rum1* mutant compared with the wild type (Fig. 2A). Moreover, three *plethora* (*plt*) genes *plt1* (GRMZM2G141638), *bbm1* (GRMZM2G366434), and *hscf1* (GRMZM2G139082) were down-regulated in *rum1* (Fig. 2A). Furthermore, two auxin response factors *arf8* (GRMZM2G034840) and *arf37* (GRMZM2G086949), two auxin influx transporters *like-aux1* (*lax1*, GRMZM2G129413) and *lax2* (GRMZM2G149481), and a *nac1* (GRMZM2G081930) gene were down-regulated in *rum1* compared with wild-type primary roots.

Auxin inducibility of differentially expressed genes involved in auxin signal transduction

To test the auxin inducibility of these auxin signal transduction genes, time-course experiments were performed in

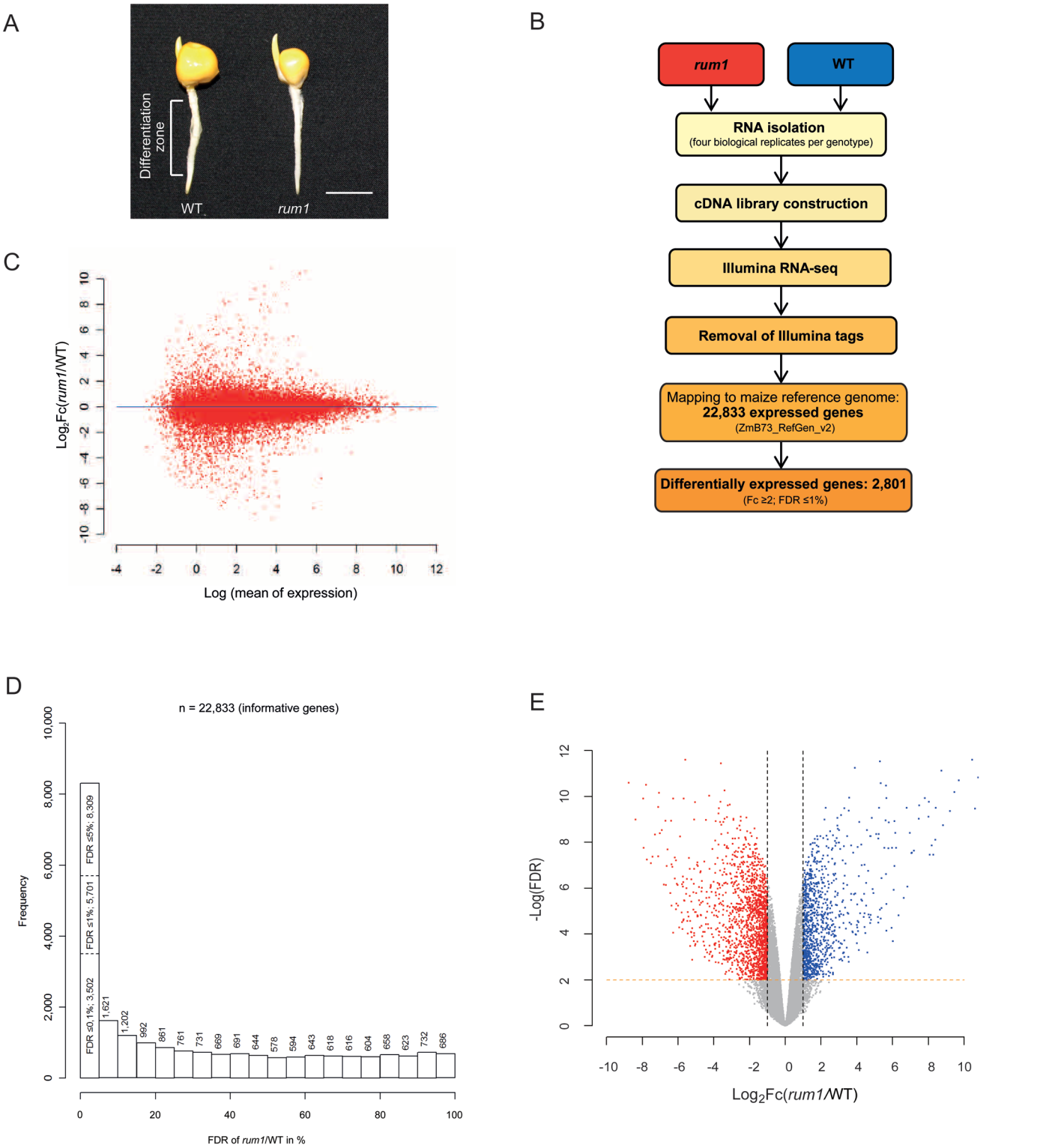


Fig. 1. (A) Phenotype of 2 cm primary roots of homozygous wild type (left) and *rum1* (right). The differentiation zone of the primary roots was subjected to RNA-Seq experiments. Lateral root primordia do not develop at this stage (Woll *et al.*, 2005). Scale bar: 1 cm. (B) Flow chart of the Illumina RNA-Seq experimental design. (C) MA-similar plot providing an overview of the correlation of gene expression changes versus expression intensity by plotting the $\log_2$ of the fold change of *rum1*/wild type against the $\log_2$ of the mean of gene expression in *rum1* and the wild type from RNA-Seq data. Each dot represents an expressed gene. (D) Histogram of FDR-values resulting from the comparison of gene expression between the *rum1* mutant and wild-type primary roots. (E) Volcano plot displaying differential gene expression patterns between *rum1* and the wild type. The *Y*-axis denotes negative $\log_{10}$ FDR of each expressed gene. The *X*-axis denotes $\log_2$-fold changes of *rum1*/wild type. Each dot represents an expressed gene. Red dots denote down-regulated genes in *rum1*, blue dots denote up-regulated genes in *rum1* with cut-off thresholds of Fc ≥2; FDR ≤1%. (This figure is available in colour at *JXB* online.)

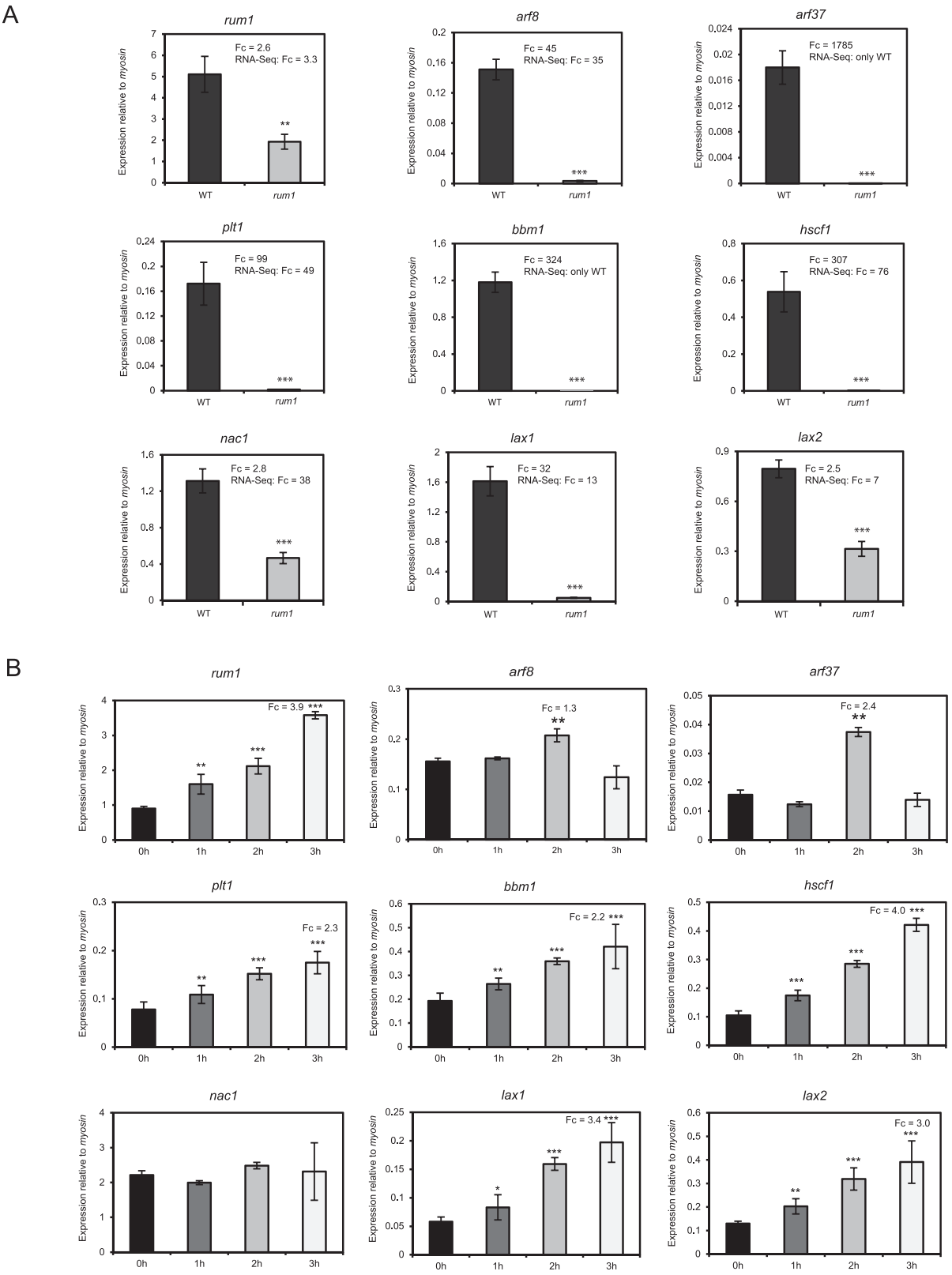


Fig. 2. (A) Confirmation of differential gene expression of nine genes involved in auxin signal transduction using qRT-PCR. Each of these genes was differentially expressed between *rum1* and the wild type in the RNA-Seq experiment. In each graph fold-changes (Fc; *rum1*/wild type) from the qRT-PCR and the RNA-Seq experiment are displayed. (B) Auxin inducibility was tested for the nine auxin-related genes by a qRT-PCR time-course experiment with reference to time point 0 h. Wild-type primary roots (5-d-old) were treated with αNAA and harvested at 0, 1, 2, and 3 h of exposure. Statistical analyses were performed with a two-sided Student's *t* test (*$p \leq 0.05$; **$p \leq 0.01$; ***$p \leq 0.001$).

5-d-old primary roots treated with 5 µM αNAA (Fig. 2B). Expression was measured relative to the housekeeping gene *myosin* (Genbank AC: 486090G09.x1) for each time point. The genes *rum1*, *lax1*, and *lax2* and all three *plt* genes (*plt1*, *bbm1*, and *hscf1*) were induced within 3 h of αNAA exposure. Remarkably, expression of *arf8* and *arf37* was induced within 2 h of αNAA exposure but returned to control levels 3 h after auxin treatment. The transcription of *nac1* was not affected by αNAA exposure.

RUM1 regulates vascular development

The homoeologues *arf8* (GRMZM2G034840) and *arf37* (GRMZM2G086949), which were significantly down-regulated in *rum1* (Fig. 2A) encode putative transcriptional activators. Their closest homologue in Arabidopsis is *MP*/*ARF5,* which regulates vascular development (Przemeck *et al.*, 1996; Hardtke and Berleth, 1998). To determine whether *rum1* mutants are impaired in vascular development, a comparative histological analysis of wild-type and *rum1* primary roots was performed at different developmental stages. In both genotypes, roots of 20, 40, 60, and 80 mm length were dissected by serial transverse hand sections in 10 mm increments (Fig. 3). In wild-type primary roots, at all four developmental stages (20, 40, 60, and 80 mm), protoxylem elements, early metaxylem elements, and parenchymous pith cells surrounding the late metaxylem elements were well developed and properly arranged throughout development (Fig. 3A–T). By contrast, in primary roots of the mutant *rum1*, xylem elements and the pith cells were largely in disarray. Only the oldest part of the mutant roots at all four developmental stages displayed properly arranged xylem elements (20 mm: Fig. 3A′; 0 mm: Fig. 3C′; 60 mm: Fig. 3G′; 80 mm: Fig. 3M′). Moreover, in 60 mm and 80 mm *rum1* roots, only the oldest parts of the

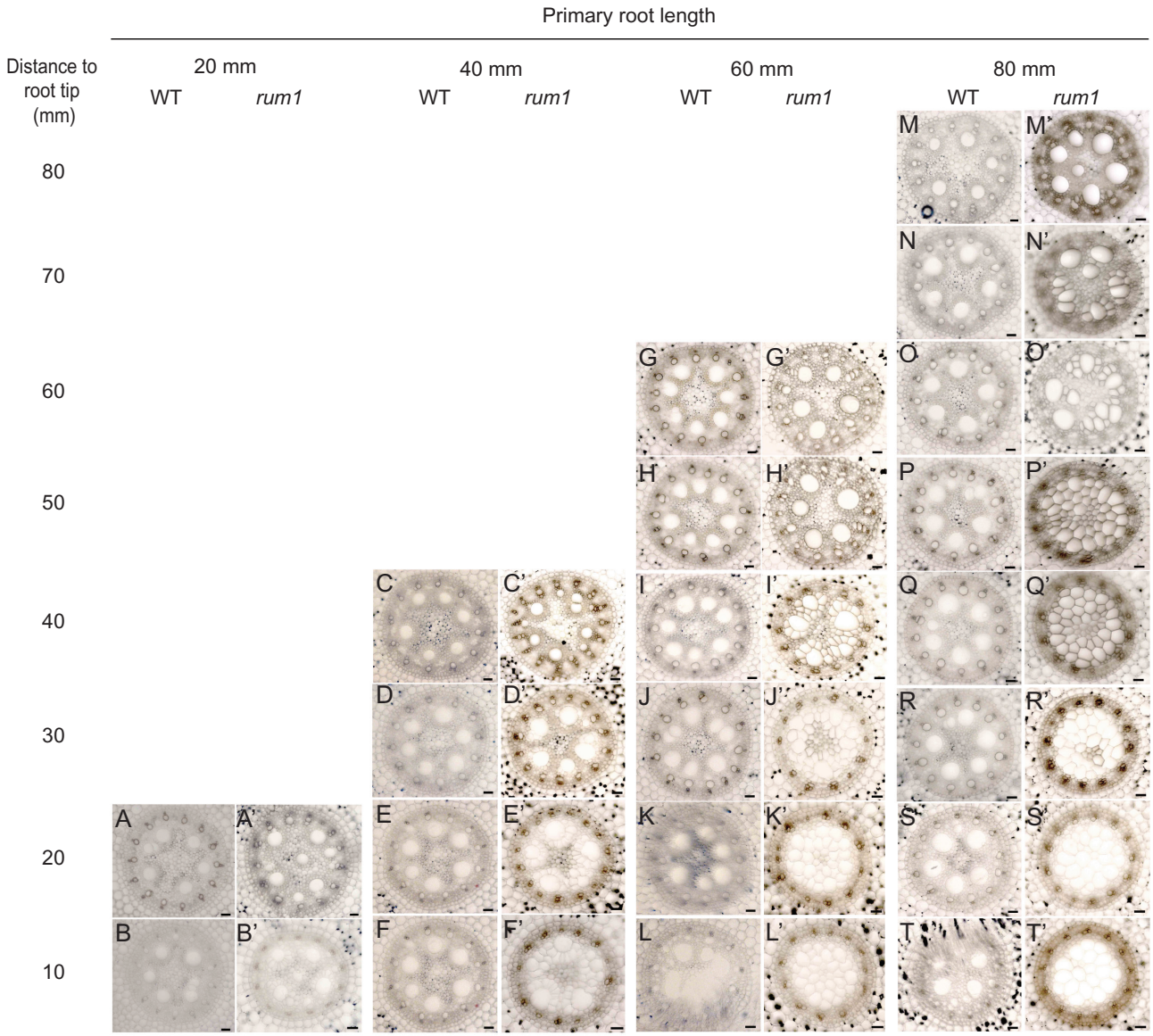


Fig. 3. Series of transverse sections displaying the central cylinder of primary roots at different developmental stages: 20 mm wild-type (A, B) and *rum1* (A′, B′) primary roots, 40 mm wild-type (C–F) and *rum1* (C′–F′) primary roots, 60 mm wild-type (G–L) and *rum1* (G′–L′) primary roots, 80 mm wild-type (M–T) and *rum1* (M′–T′) primary roots. Sections were taken every 10 mm and represent a single, representative primary root per genotype and developmental stage. The distances are indicated with reference to the root tip. Scale bar: 50 µm. (This figure is available in colour at *JXB* online.)

primary root (60 mm: Fig. 3G′–J′; 80 mm: Fig. 3M′–O′) displayed well differentiated xylem elements, while the younger parts of these roots (60 mm: Fig. 3K′–L′; 80 mm: Fig 3P′–T′) did not show any differentiated xylem elements. Hence, the magnitude of alterations in the vasculature of *rum1* correlated with the developmental stage at which these parts of the root were formed. In all analysed mutant roots, vascular differentiation was not affected in cells that were released from the meristematic zone soon after germination (Fig. 3A′, C′, G′, M′) whereas the defects became more severe the later these cells were released from the meristem.

The number of xylem elements and the number and size of pith cells was determined in representative cross-sections of 80 mm *rum1* roots at a distance of 70 mm from the root tip. In these sections, the number of early and late metaxylem elements was not altered, whereas the number of pith cells surrounding the late metaxylem was significantly reduced while their size in a radial direction was significantly enlarged compared with cells in the corresponding region of wild-type primary roots (Table 1).

Auxin control of vascular differentiation

In a previous study, it was demonstrated that polar auxin transport was decreased by 83% in 3-d-old *rum1* primary roots (Woll *et al.*, 2005). To test the effect of auxin on vascular development in maize primary roots, wild-type and *rum1* seedlings were germinated either in the presence of the auxin transport inhibitor NPA (10 µM), the synthetic auxin αNAA (0.1 µM) or water. Primary roots of 60 mm length were sectioned in 10 mm increments (Fig. 4). In the mutant *rum1*, disorganized vascular patterns and enlarged pith cells were observed irrespective of water, NPA or αNAA treatment (Fig. 4A′–R′). Hence, treatment with αNAA, which can passively diffuse into the cells, did not recover the wild-type vascular phenotype in *rum1* primary roots (Fig. 4M′–R′). By contrast, while αNAA treatment of wild-type primary roots did not alter the vascular system (Fig. 4M–R), NPA treatment led to a similar disorganization in vascular patterning and enlarged pith cells surrounding the late metaxylem (Fig. 4J–L) as observed in the mutant *rum1*. This suggests that these defects are a consequence of reduced polar auxin transport.

The rum1 *mutant displays enhanced lignification of aberrant pith cells*

Mapman analyses of genes differentially expressed between wild-type and *rum1* primary roots revealed that three genes which encode key enzymes of lignin biosynthesis (4-COUMARATE-CoA LIGASE, 4CL; GRMZM2G165844, FERULATE 5-HYDROXYLASE, F5H; GRMZM2G100158 and CINNAMYLALCOHOLDEHYDROGENASE, CAD; GRMZM2G443445) were strongly induced in 2 cm *rum1* primary roots (see Supplementary Table S3 at *JXB* online).

Secondary cell walls accumulate lignin or other secondary metabolites. Lignin deposition in the vasculature of wild-type (Fig. 5A) and *rum1* (Fig. 5B) primary roots was surveyed by phloroglucinol–HCl staining and identified by red staining of protoxylem and early metaxylem, and endodermis cells forming the Casparian strip. In contrast to wild-type primary roots (Fig. 5A), strong staining was detected in the larger and thickened pith cells surrounding the late metaxylem in *rum1* (Fig. 5B). Pith cell walls were significantly thicker in *rum1* than in wild-type primary roots (Fig. 5C), while cell wall thickness was not affected in *rum1* endodermis cells (Fig. 5D), which also deposit lignin. Hence, enhanced lignin deposition in mutant primary roots is probably not directly controlled by *rum1* but rather is a result of the defects during the differentiation of vascular cells.

Discussion

In the present RNA-Seq study, 22 833 genes were expressed in the differentiation zone of 2 cm primary roots of maize. A similar number of genes was expressed in a transcriptome analysis of maize primary roots by SAGE where it was extrapolated from 14 850 expressed genes that ~22 000 genes are active in the root tips of the cultivar FR697 (Poroyko *et al.*, 2005). These results suggest that at least ~55% of all high confidence gene models (ZmB73_RefGen_v2) are expressed in the differentiation zone and in root tips of young maize primary roots.

RUM1 controls expression of genes associated with auxin signal transduction

Aux/IAA proteins such as RUM1 (von Behrens *et al.*, 2011) are auxin signal transduction regulators that control diverse aspects of plant development. In the present RNA-Seq study, *rum1* expression was significantly reduced in the differentiation zone of *rum1* primary roots compared with the wild type, confirming previous qRT-PCR experiments (von Behrens *et al.*, 2011). Several genes known to act downstream of *Aux/*

Table 1. *Quantification of xylem cell number and radial size in 80 mm WT versus rum1 roots at a distance of 70 mm from the root tip*

	Number of metaxylem cells		Pith cells around late metaxylem	
	Early	Late	Number	Size (µm)
WT	13±1	6±1	110±7	9.4±2.6
rum1	13±2	6±1	76±18[a]	20.6±12.5[b]

[a] Number of WT vs *rum1* pith cells: $p \leq 0.05$.
[b] Size of WT vs *rum1* pith cells: $p \leq 0.01$.

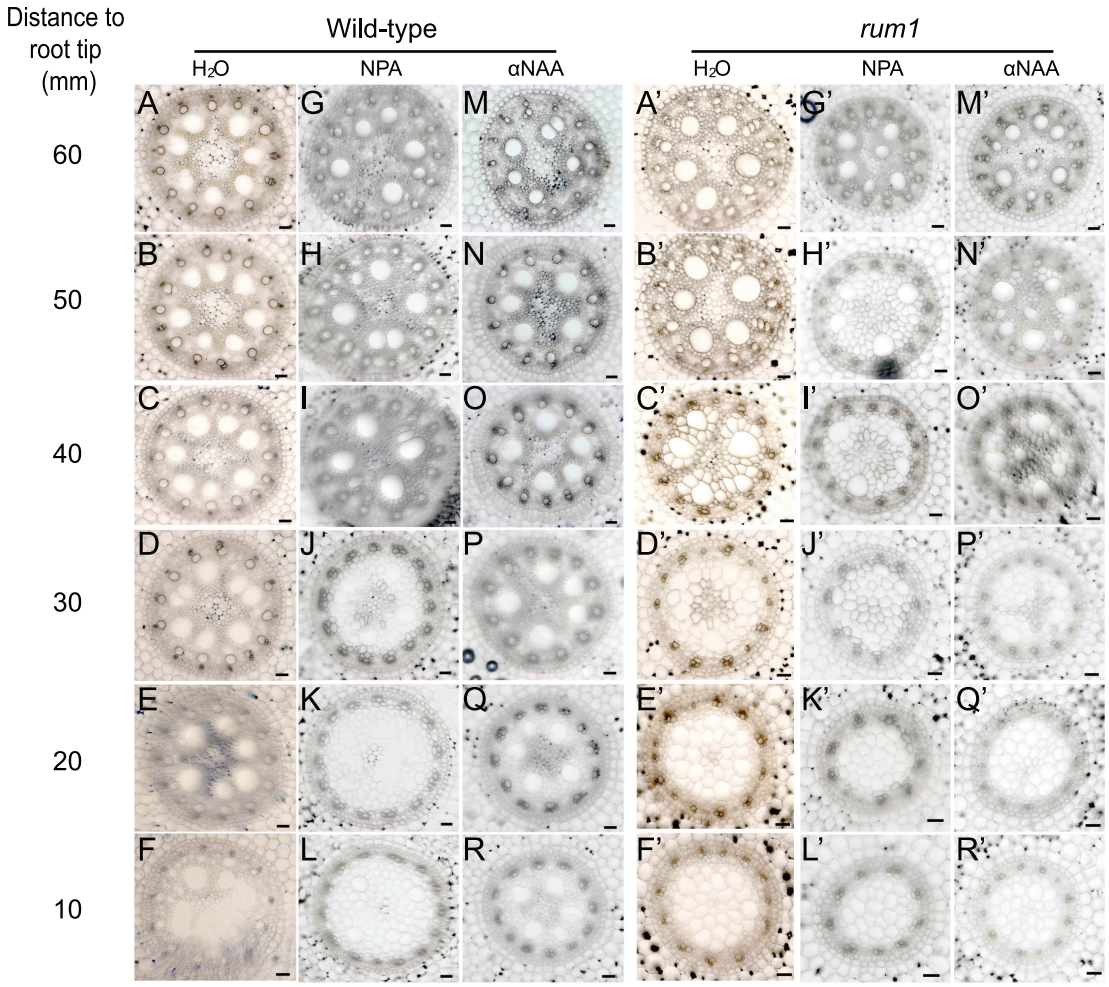


Fig. 4. Effect of NPA and αNAA application on the formation of the vascular system of 60 mm wild-type and *rum1* primary roots. Series of transverse sections displaying the central cylinder of wild-type (A–R) and *rum1* (A′–R′) primary roots grown in water (wild type: A–F; *rum1*: A′–F′), 10 μM NPA (wild type: G–L; *rum1*: G′–L′) and 0.1 μM αNAA (wild type: M–R; *rum1*: M′–R′). Sections were taken every 10 mm and show a single, representative primary root per genotype and growth condition. The distances are indicated with reference to the root tip. Scale bar: 50 μm. (This figure is available in colour at *JXB* online.)

IAA genes displayed RUM1-dependent expression and were auxin-inducible. A putative role of these genes in maize root development is suggested by the function of their orthologues in root formation in other plant species.

First, *nac1* (GRMZM2G081930), which encodes the NAC domain-containing protein 21/22 is down-regulated in the differentiation zone of *rum1* primary roots. The closest *Arabidopsis* homologue of this gene encodes the transcription factor NAC1 which controls lateral root initiation via an Aux/IAA-ARF-dependent auxin signalling module (Xie *et al.*, 2000, 2002).

Three maize *plethora* (*plt*) genes were down-regulated in *rum1* including *plt1*, (GRMZM2G141638), *bbm1* (GRMZM2G366434), and *hscf1* (GRMZM2G139082). These genes encode members of the AP2/EREBP (APETALA2/ETHYLENE-RESPONSIVE ELEMENT BINDING PROTEIN) class transcription factor family. In *Arabidopsis*, several *PLT* genes are down-regulated in the mutant *slr/iaa14* (Aida *et al.*, 2004) which is blocked in lateral root formation. The maize *plt1* and *bbm1* genes are orthologues of *AtBBM,* which controls adventitious root formation (Srinivasan *et al.*, 2007). Moreover, maize *hscf1* is the closest relative of *Arabidopsis PLT3* and *PLT7* (see Supplementary Fig. S1 at *JXB* online), which are involved in the regulation of the auxin efflux transporter PIN1 (Prasad *et al.*, 2011). Although overall *pin1* levels were not altered in the *rum1* mutant (Woll *et al.*, 2005) cell-type specific alterations of *pin1* expression might explain the reduction in polar auxin transport by 83% in the *rum1* primary root.

Furthermore, in the present study the maize *like-aux1* (*lax1*, GRMZM2G129413) and *lax2* (GRMZM2G149481) which encode auxin influx carriers, were down-regulated in *rum1*. In *Arabidopsis*, the *AUX/LAX* gene family members *AUX1* and *LAX3* act concomitantly in lateral root initiation (Marchant *et al.*, 2002) and emergence (Swarup *et al.*, 2008).

Finally, maize *lbd24* (GRMZM2G075499) was down-regulated more than 4-fold in *rum1* primary roots. The closest homologue of maize *lbd24* is *Arabidopsis LBD16* (Majer and Hochholdinger, 2011). AtLBD16 controls asymmetric division of lateral root founder cells in *Arabidopsis* (Goh *et al.*, 2012), which is directly regulated by the SLR/IAA14-ARF7/ARF19-dependent auxin signalling module which controls *Arabidopsis* lateral root initiation (Okushima *et al.*, 2007).

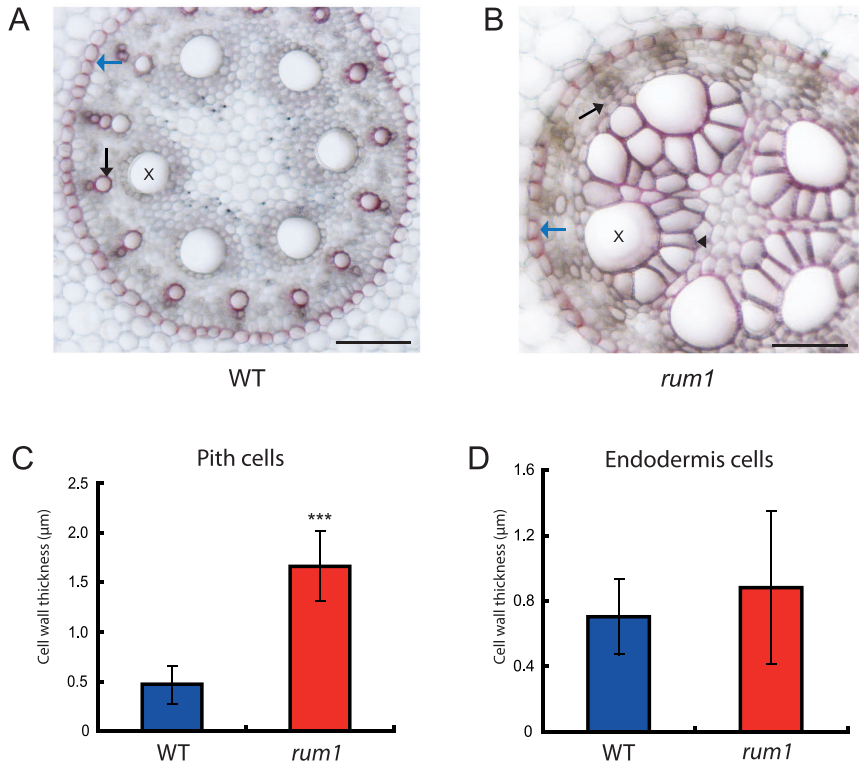


Fig. 5. Analyses of lignification in wild-type and *rum1* primary roots. Cross-sections of the central region of 10-d-old wild-type (A) and *rum1* (B) primary roots. Phloroglucinol–HCl-stained lignified cell walls were stained in red. Lignin was detected in protoxylem and early metaxylem elements (black arrow), in enlarged pith cells (black arrowhead) surrounding the late metaxylem elements (x) in *rum1* and in endodermis cells forming Casparian strips (blue arrow). (C) Pith cells surrounding late metaxylem elements in the mutant *rum1* display significantly thicker cell walls than their wild-type counterparts (n=30, ***$p \leq 0.001$). (D) Cell wall thickness of endodermis cells (n=18) is not significantly different between wild-type and *rum1* primary roots. (A,B) scale bar: 100 µm. (This figure is available in colour at *JXB* online.)

In summary, in the present study the expression of several genes involved in auxin signal transduction was shown to be RUM1-dependent. In other species, orthologues of several of these genes are known to function in root development. Hence, genetic dissection of these candidate genes may provide a better understanding of the molecular processes involved in maize root formation.

RUM1 controls auxin dependent xylem development

The *Arabidopsis* MP/ARF5 protein, which controls vascular development (Przemeck *et al.*, 1996; Hardtke and Berleth, 1998) is orthologous to the paralogous maize auxin signal transduction regulators ARF8 and ARF37 (von Behrens *et al.*, 2011; Wang *et al.*, 2012). In the present study, transcription of both *arf8* and *arf37* was almost completely repressed in the mutant *rum1*. These findings support the notion that transcription of *arf* genes is controlled by Aux/IAA proteins (Lau *et al.*, 2011). To survey if *arf8* and *arf37* repression affects xylem organization in *rum1* primary roots, a comparative histological analysis of *rum1* versus wild-type vasculature was undertaken. These experiments led to the observation that the fate of the vascular system in *rum1* primary roots depended on its developmental status.

The vascular system of the mutant *rum1* formed soon after germination at a distance of 10 mm from the root apex displayed only minor defects in xylem organization (Fig. 3A′, C′, G′, M′). It could be hypothesized that, at this early stage of primary root development, the lack of *rum1* expression is compensated by another member of the maize *Aux/IAA* gene family. This notion is supported by a recent expression analysis of the maize *Aux/IAA* gene family reporting very low gene expression levels of *rum1* but much higher levels of other *Aux/IAA* genes such as *IAA14* in the meristematic zone of maize primary roots (Ludwig *et al.*, 2013). Moreover, at this developmental stage, auxin synthesized in the root apex (Feldmann, 1980) could compensate for the significantly reduced auxin amounts transported from the shoot to the root (Woll *et al.*, 2005). Subsequently formed xylem elements at a distance of 20–40 mm from the root apex were in severe disarray and their surrounding pith parenchyma cells were significantly enlarged (Fig. 3B′, D′–F′, H′–J′; N′–O′). This observation suggests that RUM1 controls the development of distinct cells-types such as xylem elements and pith cell parenchyma. In *Arabidopsis*, it was reported that a mutually inhibitory feedback loop between auxin and cytokinin sets up the boundary between xylem and neighbouring procambial parenchyma cells (Bishopp *et al.*, 2011). Significant alterations in auxin and cytokinin signalling can affect the positions of these domains. Since the enlarged pith cells around the metaxylem elements are lignified, such as xylem cells (Fig. 5B), one could assume that these pith cells have

altered their cell identity. However, in *Arabidopsis*, protoxylem cells are defined by high auxin signalling while the surrounding procambial cells are characterized by low auxin and high cytokinin signalling (Bishopp *et al.*, 2011). For instance, *Arabidopsis* mutants with significantly impaired auxin signalling such as *axr3* do not form xylem cells at all (Leyser *et al.*, 1996). Therefore, if one assumes a similar mechanism in maize, one would rather expect a reduction of the xylem elements in a mutant with compromised auxin signalling such as in *rum1*. Remarkably, the phenotype of disorganized xylem elements and enlarged lignified surrounding cells is only transiently observed during a defined developmental period. In continuously proliferating older primary roots of 60 or 80 mm in length, no xylem elements were formed later in development (60 mm: Fig. 3K′–L′; 80 mm: Fig 3P′–T′). Hence, at this developmental stage, *rum1* roots display the same low auxin signalling phenotype as observed in the *Arabidopsis axr3* mutant (Leyser *et al.*, 1996).

It was demonstrated previously that the *Aux/IAA* mutant *rum1* displayed significantly reduced polar auxin transport (Woll *et al.*, 2005). A link between reduced polar auxin transport in *rum1* roots and the observed defects in vasculature formation was demonstrated by a treatment of wild-type roots with the polar auxin transport inhibitor NPA. NPA-treated wild-type roots showed the similar disarray in vascular organization as observed in the mutant *rum1*. This supports the notion that polar auxin transport is important for the differentiation of xylem elements and the surrounding pith cells in maize (Jansen *et al.*, 2012). By contrast, αNAA was not able to restore the wild-type phenotype in *rum1* primary roots. This is in line with the earlier observation that auxin treatment cannot induce lateral roots in the mutant *rum1* (Woll *et al.*, 2005), and supports the molecular model explaining RUM1 function (von Behrens *et al.*, 2011). According to this model, the rum1 mutant protein constitutively represses downstream auxin responsive genes even at high auxin levels (von Behrens *et al.*, 2011). The same effect and phenotype can be obtained in wild-type plants treated with NPA which inhibits polar auxin transport and thereby reduces cellular auxin levels.

A similar defect in vascular development has been demonstrated in tomato *IAA15* (Deng *et al.*, 2012). However, while *rum1* specifically controls vascular development in roots but not in the shoot, tomato *IAA15* regulates xylem development in the stem (Deng *et al.*, 2012).

Excessive lignification of pith cell-walls in the primary root of the mutant rum1

The hydrophobic cell wall polymer lignin is deposited in endodermis and pith cells surrounding xylem elements to make them impermeable to water and to provide structural support (Vermerris *et al.*, 2010). Histochemical staining revealed excessive lignin deposition in enlarged pith cells of the mutant *rum1* while lignin deposition in endodermis cells was not significantly affected. Comparative RNA-Seq experiments of young maize primary roots revealed strong up-regulation of three lignin biosynthesis genes in mutant *rum1* roots. Recent studies revealed complex transcriptional networks that control lignin biosynthesis in addition to a plethora of developmental and environmental cues including various types of stress and auxin (Zhao and Dixon, 2011). The maize mutant *lrt1* (*lateralrootless1*) displays a similar phenotype as *rum1* (Hochholdinger and Feix, 1998). Both mutants fail to initiate lateral roots from the primary root. The *lrt1* gene has not yet been cloned but it has been demonstrated that it is not allelic with *rum1* (Woll *et al.*, 2005). A comparative proteome analysis of 9-d-old wild-type and *lrt1* primary roots revealed four proteins involved in lignin biosynthesis (including Caffeoyl-CoA-3-*O*-methyltransferase, CCoA-OMT) that were up-regulated in *lrt1* primary roots compared with wild-type primary roots (Hochholdinger *et al.*, 2004b). Similarly, a comparative transcriptome analysis of wild-type versus *rum1* pericycle cells revealed preferential expression of CCoA-OMT in mutant pericycle cells (Woll *et al.*, 2005).

Moreover, it has been demonstrated that inhibition of auxin transport in *Arabidopsis* resulted in dramatically increased numbers of cells with thickened secondary cell walls in stems which suggests that cell wall thickness is also regulated by auxin (Mattsson *et al.*, 1999). Therefore, drastically decreased polar auxin transport in *rum1* might lead to the mis-regulation of lignin biosynthesis genes resulting in thickened secondary cell walls in enlarged pith cells.

In summary, the data presented here indicate that the transcriptional repressor RUM1, in addition to its previously demonstrated function in regulating lateral and seminal root initiation, also controls vascular and pith cell differentiation and lignin deposition in these cells.

Supplementary data

Supplementary data can be found at *JXB* online.

Supplementary Fig. S1. Phylogenetic tree of nine maize and eight *Arabidopsis plt* genes generated by ClustalV.

Supplementary Table S1. Summary of the RNA-Seq data of *rum1* and the wild type, and alignments to the B73 reference genome (ZmB73_RefGen_v2).

Supplementary Table S2. List of the 22 833 expressed genes and their characteristics.

Supplementary Table S3. Differentially expressed genes (Fc ≥2; FDR ≤1%) were functionally annotated using the MapMan software. Genes that were assigned to more than one functional class or subgroup are labelled in red. Fold changes (*rum1*/WT) are given as logarithmic (log_2) values.

Supplementary Table S4. Determination of overrepresented and underrepresented functional classes among differentially expressed genes.

Supplementary Table S5. Sequences of oligonucleotide primers used for qRT-PCR analyses.

Acknowledgements

We would like to thank Jörg Hagmann (Max Planck Institute for Developmental Biology, Tübingen) for processing the raw data, Yongqiang Guo (State Intellectual Property Office, China) for help with editing some of the supplementary data, and Eva Husakova (Charles University

Prague, Czech Republic) for suggestions on the histological analyses. This project was supported by a DFG (Deutsche Forschungsgemeinschaft) grant to FH (SFB446), a grant from the US National Science Foundation to PSS (IOS-1027527), and a CSC (China Scholarship Council) fellowship to YZ.

References

Aida M, Beis D, Heidstra R, Willemsen V, Blilou I, Galinha C, Nussaume L, Noh YS, Amasino R, Scheres B. 2004. The *PLETHORA* genes mediate patterning of the Arabidopsis root stem cell niche. *Cell* **119,** 109–120.

Aiken RM, Smucker AJ. 1996. Root system regulation of whole plant growth. *Annual Review of Phytopathology* **34,** 325–346.

Bishopp A, Help H, El-Showk S, Weijers D, Scheres B, Benková E, Friml J, Mähonën AP, Helariutta Y. 2011. A mutually inhibitory interaction between auxin and cytokinin specifies vascular pattern in roots. *Current Biology* **21,** 917–926.

Bullard JH, Purdom E, Hansen KD, Dudoit S. 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11,** 94.

Deng W, Yan F, Liu M, Wang X, Li Z. 2012. Down-regulation of *SlIAA15* in tomato altered stem xylem development and production of volatile compounds in leaf exudates. *Plant Signaling & Behavior* **7,** 911–913.

Feldmann LJ. 1980. Auxin biosynthesis and metabolism in isolated roots of *Zea mays*. *Physiologia Plantarum* **49,** 145–150.

Goh T, Joi S, Mimura T, Fukaki H. 2012. The establishment of asymmetry in Arabidopsis lateral root founder cells is regulated by LBD16/ASL18 and related LBD/ASL proteins. *Development* **139,** 883–893.

Gray WM, Kepinski S, Rouse D, Leyser O, Estelle M. 2001. Auxin regulates SCF(TIR1)-dependent degradation of AUX/IAA proteins. *Nature* **414,** 271–276.

Hardtke CS, Berleth T. 1998. The Arabidopsis gene *MONOPTEROS* encodes a transcription factor mediating embryo axis formation and vascular development. *EMBO Journal* **17,** 1405–1411.

Hochholdinger F, Feix G. 1998. Early post-embryonic root formation is specifically affected in the maize mutant *lrt1*. *The Plant Journal* **16,** 247–255.

Hochholdinger F, Guo L, Schnable PS. 2004*b*. Lateral roots affect the proteome of the primary root of maize (*Zea mays* L.). *Plant Molecular Biology* **56,** 397–412.

Hochholdinger F, Woll K, Sauer M, Dembinsky D. 2004*a*. Genetic dissection of root formation in maize (*Zea mays* L.) reveals root-type specific developmental programmes. *Annals of Botany* **93,** 359–368.

Jansen L, Roberts I, De Rycke R, Beeckman T. 2012. Phloem-associated auxin response maxima determine radial positioning of lateral roots in maize. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences* **367,** 1525–1533.

Lau S, De Smet I, Kolb M, Meinhardt H, Jürgens G. 2011. Auxin triggers a genetic switch. *Nature Cell Biology* **13,** 611–615.

Li YJ, Fu YR, Huang JG, Wu CA, Zheng CC. 2011. Transcript profiling during the early development of the maize brace root via Solexa sequencing. *The FEBS Journal* **278,** 156–166.

Leyser HM, Pickett FB, Dharmasiri S, Estelle M. 1996. Mutations in the *AXR3* gene of Arabidopsis result in altered auxin response including ectopic expression from the SAUR-AC1 promoter. *The Plant Journal* **10,** 403–413.

Liu S, Yeh CT, Tang HM, Nettleton D, Schnable PS. 2012. Gene mapping via bulked segregant RNA-Seq (BSR-Seq). *PLoS One* **7,** e36406.

Ludwig Y, Zhang Y, Hochholdinger F. 2013. The maize (*Zea mays* L.) *AUXIN/INDOLE-3-ACETIC ACID* gene family: phylogeny, synteny, and unique root-type and tissue-specific expression patterns during development. *PLoS One* **8,** e78859

Lund SP, Nettleton D, McCarthy DJ, Smyth GK. 2012. Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. *Statistical Applications in Genetics and Molecular Biology* **11,** 8.

Majer C, Hochholdinger F. 2011. Defining the boundaries: structure and function of LOB domain proteins. *Trends in Plant Science* **16,** 47–52.

Marchant A, Bhalerao R, Casimiro I, Eklöf J, Casero PJ, Bennett M, Sandberg G. 2002. AUX1 promotes lateral root formation by facilitating indole-3-acetic acid distribution between sink and source tissues in the Arabidopsis seedling. *The Plant Cell* **14,** 589–597.

Mattsson J, Sung ZR, Berleth T. 1999. Responses of plant vascular systems to auxin transport inhibition. *Development* **126,** 2979–2991.

Mockaitis K, Estelle M. 2008. Auxin receptors and plant development: a new signaling paradigm. *Annual Review of Cell and Developmental Biology* **24,** 55–80.

Nettleton D, Hwang JTG, Caldo RA, Wise RP. 2006. Estimating the number of true null hypotheses from a histogram of *p*-values. *Journal of Agricultural, Biological, and Environmental Statistics* **11,** 337–356.

Okushima Y, Fukaki H, Onoda M, Theologis A, Tasaka M. 2007. ARF7 and ARF19 regulate lateral root formation via direct activation of *LBD/ASL* genes in Arabidopsis. *The Plant Cell* **19,** 118–130.

Overvoorde P, Fukaki H, Beeckman T. 2010. Auxin control of root development. *Cold Spring Harbor Perspectives in Biology* **2,** a001537.

Peret B, De Rybel B, Casimiro I, Benkova E, Swarup R, Laplaze L, Beeckman T, Bennett MJ. 2009*a*. Arabidopsis lateral root development: an emerging story. *Trends in Plant Science* **14,** 399–408.

Peret B, Larrieu A, Bennett MJ. 2009*b*. Lateral root emergence: a difficult birth. *Journal of Experimental Botany* **60,** 3637–3643.

Poroyko V, Hejlek LG, Spollen WG, Springer GK, Nguyen HT, Sharp RE, Bohnert HJ. 2005. The maize root transcriptome by serial analysis of gene expression. *Plant Physiology* **138,** 1700–1710.

Prasad K, Grigg SP, Barkoulas M, *et al*. 2011. Arabidopsis PLETHORA transcription factors control phyllotaxis. *Current Biology* **21,** 1123–1128.

Przemeck GKH, Mattsson J, Hardtke CS, Sung ZR, Berleth T. 1996. Studies on the role of the Arabidopsis gene *MONOPTEROS* in vascular development and plant cell axialization. *Planta* **200,** 229–237.

Schnable PS, Ware D, Fulton RS, *et al*. 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science* **326,** 1112–1115.

Srinivasan C, Liu Z, Heidmann I, *et al*. 2007. Heterologous expression of the BABY BOOM AP2/ERF transcription factor enhances the regeneration capacity of tobacco (*Nicotiana tabacum* L.). *Planta* **225,** 341–351.

Storey JD. 2002. *A direct approach to false discovery rates*. *Journal of the Royal Statistical Society: Series* **B 64,** 479–498.

Swarup K, Benkova E, Swarup R, *et al*. 2008. The auxin influx carrier LAX3 promotes lateral root emergence. *Nature Cell Biology* **10,** 946–954.

Tan X, Calderon-Villalobos LI, Sharon M, Zheng C, Robinson CV, Estelle M, Zheng N. 2007. Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **446,** 640–645.

Taramino G, Sauer M, Stauffer Jr. JL, Multani D, Niu X, Sakai H, Hochholdinger F. 2007. The maize (*Zea mays* L.) *rtcs* gene encodes a LOB domain protein that is a key regulator of embryonic seminal and post-embryonic shoot-borne root initiation. *The Plant Journal* **50,** 649–659.

Thimm O, Bläsing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Muller LA, Rhee SY, Stitt M. 2004. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal* **37,** 914–939.

Tian Q, Uhlir NJ, Reed JW. 2002. Arabidopsis SHY2/IAA3 inhibits auxin-regulated gene expression. *The Plant Cell* **14,** 301–319.

Uggla C, Moritz T, Sandberg G, Sundberg B. 1996. Auxin as a positional signal in pattern formation in plants. *Proceedings of the National Academy Sciences, USA* **93,** 9282–9286.

Vermerris W, Sherman DM, McIntyre LM. 2010. Phenotypic plasticity in cell walls of maize brown midrib mutants is limited by lignin composition. *Journal of Experimental Botany* **61,** 2479–2490.

von Behrens I, Komatsu M, Zhang YX, Berendzen KW, Niu XM, Sakai H, Taramino G, Hochholdinger F. 2011. *Rootless with undetectable meristem 1* encodes a monocot-specific AUX/IAA protein that controls embryonic seminal and post-embryonic lateral root initiation in maize. *The Plant Journal* **66,** 341–353.

Wang Y, Deng D, Shi Y, Miao N, Bian Y, Yin Z. 2012. Diversification, phylogeny and evolution of auxin response factor (ARF) family: insights

gained from analyzing maize *ARF* genes. *Molecular Biology Reports* **39,** 2401–2415.

Wang YJ, Deng DX, Bian YL, Lv YP, Xie Q. 2010. Genome-wide analysis of primary auxin-responsive *Aux/IAA* gene family in maize (*Zea mays*. L.). *Molecular Biology Reports* **37,** 3991–4001.

Wilson JW, Wilson PMW, Walker ES. 1991. Patterns of tracheary differentiation in lettuce pith explants - positional control and temperature effects. *Annals of Botany* **68,** 109–128.

Woll K, Borsuk LA, Stransky H, Nettleton D, Schnable PS, Hochholdinger F. 2005. Isolation, characterization, and pericycle-specific transcriptome analyses of the novel maize lateral and seminal root initiation mutant *rum1*. *Plant Physiology* **139,** 1255–1267.

Woodward AW, Bartel B. 2005. Auxin: regulation, action, and interaction. *Annals of Botany* **95,** 707–735.

Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26,** 873–881.

Xie Q, Frugis G, Colgan D, Chua NH. 2000. Arabidopsis NAC1 transduces auxin signal downstream of TIR1 to promote lateral root development. *Genes & Development* **14,** 3024–3036.

Xie Q, Guo HS, Dallman G, Fang S, Weissman AM, Chua NH. 2002. SINAT5 promotes ubiquitin-related degradation of NAC1 to attenuate auxin signals. *Nature* **419,** 167–170.

Zhao Q and Dixon RA. 2011. Transcriptional networks for lignin biosynthesis: more complex than we thought? *Trends in Plant Science* **16,** 227–233.

den Blüten der Samenpflanzen. Wenn ein Farnblatt aus dem Kohleflöz unter Bochum auch genauso aussieht wie eines vom Waldrand tausend Meter darüber an der frischen Luft, so ist es eben nicht gleich. 200 Millionen Jahre gingen auch an dem Farn nicht spurlos vorüber, die grobe Form muss wohl optimal gewesen sein, und ist immer noch sehr erfolgreich, aber die Chemie innen hat sich ganz sicher massiv verändert. Formulierungen wie „Farne kommen bis heute zahlreich auf der Erde vor, obwohl sie entwicklungsbiologisch älter als 400 Millionen Jahre sind“ [2] sind einfach schräg, nicht nur das „obwohl“ ist falsch, sondern auch die Unterstellung, dass Farne „entwicklungsbiologisch“ anders alt sind als Blütenpflanzen: Genaugenommen sind alle Lebewesen entwicklungsbiologisch gleich alt, denn sie stammen alle von einer einzigen Erfindung – der Selbstreplikation des ersten Moleküls – vielleicht einer RNA – und später der ersten Zelle ab.

[1] V. Radhika et al., PLoS One 2012, doi:10.1371/journal.pone.0048050
[2] J.-W. Kellmann, A. Overmeyer, http://www.ice.mpg.de/ext/968.html

Axel Brennicke, Universität Ulm

PFLANZENZÜCHTUNG

Heterosis: Maishybride exprimieren mehr Gene als ihre Eltern

Die molekularen Grundlagen der Überlegenheit von Hybriden im Vergleich zu ihren reinerbigen Eltern, welche als Heterosis bezeichnet wird, sind bislang nur unzureichend verstanden. In einer neuen Studie konnte gezeigt werden, dass Maishybride einige hundert Gene mehr exprimieren als ihre Eltern. Dies könnte ein Ansatz zur gezielten Vorauswahl geeigneter Maislinien in der Züchtung bieten und so auf lange Sicht Zeit und Kosten sparen.

Die Überlegenheit von Pflanzenhybriden gegenüber ihren reinerbigen Eltern ist seit Jahrhunderten bekannt und wurde bereits von Joseph Gottlieb Köhlreuter (1733–1806) und Charles Darwin (1809–1882) beschrieben (Abbildung 1). George Harrison Shull legte 1908 mit seinen Arbeiten über Mais die Grundlage zur Ausnutzung dieses Phänomens in der Pflanzenzüchtung [1]. Shull bezeichnete die Überlegenheit von Maishybriden gegenüber ihren homozygoten Elternlinien als *Heterosis*. Heute wird Mais in der landwirtschaftlichen Praxis der Industrienationen ausschließlich als Hybridmais angebaut. Heterosis ist deshalb von überragender landwirtschaftlicher Bedeutung. Im Jahr 2011 wurden weltweit mit Hybrid-Körnermais 865 Millionen Tonnen Ertrag und damit mehr als bei Reis und Weizen erzielt. Dennoch ist über die molekularen Grundlagen, die zur Überlegenheit der Hybride führen, nur wenig bekannt. Bereits kurz nach den bahnbrechenden Arbeiten Shulls wurden erste genetische Erklärungsversuche der Heterosis postuliert. Die von Donald F. Jones im Jahr 1917 vorgeschlagene Dominanzhypothese besagt, dass Heterosis durch das Zusammenspiel verschiedener Gene zu erklären ist. Nach dieser Hypothese sind vorteilhafte Allele in der Hybridpflanze dominant und komplementieren somit das unterlegene zweite Allel [2]. Komplementation bedeutet in diesem Zusammenhang, dass Eigenschaften der vorteilhaften Allele der einen Elternpflanze in der Hybridpflanze die ungünstigeren Eigenschaften des jeweils unterlegenen Allels der anderen Elternpflanze ausgleichen. Somit summieren sich die besseren Allele der beiden Eltern und deren Eigenschaften und führen zu einer überlegenen Hybridpflanze. Diese Hypothese wurde zu einer Zeit formuliert, in der weder Träger noch Struktur der Erbinformation bekannt waren. Inzwischen wissen wir nicht nur, dass die Erbinformation in DNA-Molekülen gespeichert ist, sondern es wurden auch ganze Genome – so auch bei Mais – entschlüsselt. Das Maisgenom besitzt rund 40.000 Gene. Jede Maissorte besitzt im Prinzip für jedes Gen zwei identische, aber für die Sorte spezifische Allele, so dass es für jedes Maisgen eine unüberschaubare Vielfalt von natürlichen Allelen gibt. Wegen der großen Zahl von Genen und Allelvarianten sind Prognosen über günstige Kombinationen von Sorten deshalb in der Praxis fast unmöglich.

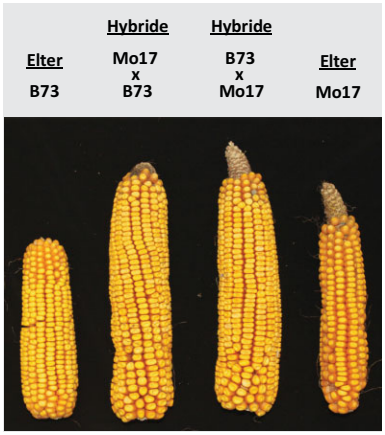
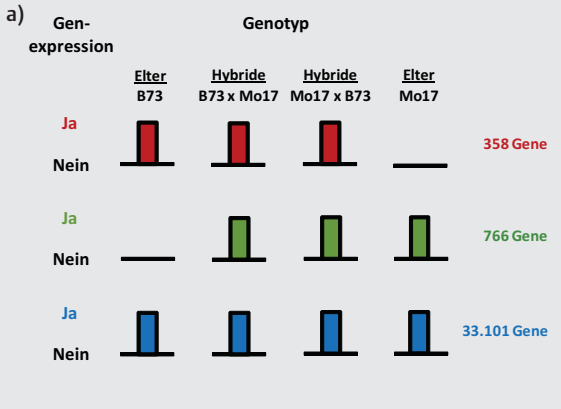


ABB. 1 Hybridmais ist leistungsfähiger als seine homozygoten Elternlinien, was sich zum Beispiel in der Kolbengröße und dem Ertrag zeigt.

Durch Kombination moderner Sequenzierungsmethoden mit sta-

© 2013 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim



b)

Genotyp	B73	B73 x Mo17	Mo17 x B73	Mo17
Aktive Gene	33.460	34.225	34.226	33.874

ABB. 2 a) Während die meisten der exprimierten Gene (33.101) in allen vier Genotypen aktiv sind, gibt es mehrere hundert Gene (358 und 766), die nur in einer der beiden Elternlinien, aber beiden Hybriden aktiv sind und somit Komplementation auf Genexpressionsebene zeigen. b) Die Expression hunderter Gene in nur einer der Elternlinien – aber in beiden Hybriden – führt dazu, dass in Hybriden mehr Gene aktiv sind als in den Elternlinien.

tistischen Methoden konnte nun in Maiswurzeln ein Spezialfall der Dominanzhypothese auf Ebene der Genexpression nachgewiesen werden [3]. Dabei wurde gezeigt, dass es in den Maisinzuchtlinien B73 und Mo17 jeweils hunderte von Genen gibt, die nur in einer der beiden Inzuchtlinien, aber in beiden reziproken Hybriden exprimiert werden (Abbildung 2a). Es handelt sich dabei also um Komplementation der Genexpression in Hybriden, die dazu führt, dass in Maishybriden hunderte Gene mehr aktiv sind als in den Elternlinien (Abbildung 2b). Diese zusätzlich aktiven Gene könnten, obwohl jedes einzelne vermutlich nur einen geringen Beitrag leistet, in der Summe mit dafür verantwortlich sein, dass Hybride leistungsfähiger als ihre Eltern sind. Die Identifizierung und Analyse von wenigen hundert Genen, die zwischen verschiedenen Linien aktiv beziehungsweise inaktiv sind, könnte zukünftig die Auswahl geeigneter Linien für die Hybridzüchtung erleichtern und Pflanzenzüchtern Zeit und Kosten sparen.

EXKURS: MACHT GENTECHNIK HYBRIDMAIS ÜBERFLÜSSIG?

In der öffentlichen Diskussion wird gentechnisch veränderter Mais oft mit höheren Erträgen in Verbindung gebracht. Auf den ersten Blick könnten transgene Maissorten also eine Alternative zur Züchtung von Hybridmais sein.

Dies ist allerdings nicht richtig. Während Hybridzüchtung auf Ertragssteigerung abzielt, sorgen transgene Maissorten in der Regel für Ertragssicherheit. Ertragssicherheit bedeutet, dass gentechnisch veränderte Pflanzen toleranter gegenüber ungünstigen Umweltbedingungen wie beispielsweise Trockenheit oder resistent gegenüber Schädlingen wie dem Maiszünsler oder dem Maiswurzelbohrer sind. Der konventionelle Ansatz der Hybridmaiszüchtung (Ausnutzung von Heterosis) wird seit über 80 Jahren in der Maiszüchtung eingesetzt und kommt auch bei allen transgenen Sorten zum Einsatz. Das heißt, alle transgenen Sorten sind gleichzeitig Hybride, die somit hohe Erträge (Heterosis) und Ertragssicherheit (Resistenzen) kombinieren. In den USA, dem weltgrößten Maisproduzenten, waren im Jahr 2012 bereits 88% des angebauten Mais gentechnisch verändert, weltweit 31%, die Tendenz ist steigend.

[1] G. F. Shull, Rep. Am. Breed. Assoc. 1908, 5, 51–59.
[2] D. F. Jones, Genetics 1917, 2, 466–479.
[3] A. Paschold, Y. Jia, C. Marcon, S. Lund, N. B. Larson, C. T. Yeh, S. Ossowski, C. Lanz, D. Nettleton, P. S. Schnable, F. Hochholdinger, Genome Res. 2012, 22, 2445–2454.

Frank Hochholdinger, Universität Bonn
hochholdinger@uni-bonn.de;
Anja Paschold, Universität Bonn
paschold@uni-bonn.de;
Caroline Marcon, Universität Bonn
marcon@uni-bonn.de

ARTENSCHUTZ

Geiersterben auf dem indischen Subkontinent beendet?

Keine andere Tierart beseitigt Kadaver so gründlich wie die Geier. Fällt diese „Gesundheitspolizei" aus, erwachsen daraus Gefahren für den Menschen. Besserung scheint in Sicht.

Während in den 1980er Jahren noch millionenfach Geier den Himmel über dem Indischen Subkontinent bevölkerten, setzte in der zweiten Hälfte der 1990er Jahre eine katastrophale Abnahme auf rund 60.000 ein. Dieser Rückgang betraf alle drei vorkommenden Geierarten, am stärksten aber den Bengalgeier (*Gyps bengalensis*). In einem Nationalpark im Nordwesten Indiens beobachtete man, dass die Geier einfach aus dem Geäst fielen und starben.

Nun haben Geier ja keinen besonders guten Ruf und in Europa kommen auch nur noch wenige Geier vor – also geht es auch ohne sie. Doch in Indien ist die Lage anders. Dort sind Rinder heilige Tiere und dürfen vom Menschen nicht geschlachtet und gegessen werden. Die Kadaver von verendeten „Heiligen Kühen" werden vor den Städten gesammelt und den Geiern überlassen – bisher. Mit der Dezimierung der Geier blieben immer mehr Kadaver liegen, häuften

© 2013 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim