Institut für Nutzpflanzenwissenschaften und Ressourcenschutz (INRES)

Bioinformatics software and analyses for inference of gene functional relationships in crops

Dissertation

zur Erlangung des Grades

Doktor der Agrarwissenschaften (Dr. agr.)

der Agrar-, Ernährungs- und Ingenieurwissenschaftlichen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn

von

Tyll Gordian Stöcker

aus

Bonn

Bonn 2025

Referent: Prof. Dr. Heiko Schoof

Korreferent: PD Dr. Caroline Marcon

Tag der mündlichen Prüfung: 13. Juni 2025

Angefertigt mit Genehmigung der Agrar-, Ernährungs- und Ingenieurwissenschaftlichen Fakultät der Universität Bonn

Abstract

Associating interesting phenotypic traits with causal genes and elucidating the underlying gene functional relationships is of great importance to crop breeding and improvement efforts. To this end, the research presented in this cumulative thesis has included the use of established approaches, such as differential transcriptomics and bulked-segregant analysis, as well as the development of novel bioinformatics solutions for both the automatic analysis of insertional mutagenesis libraries and the *in silico* prediction of trait-associated gene families.

An efficient bioinformatics software solution called MuWU (Mutant-Seq Workflow Utility) for the identification and annotation of heritable mutations caused by *Mutator* transposon insertion sites was developed and proved central to the establishment and continuous extension of the world's largest sequence-indexed insertional mutagenesis library in maize – *BonnMu*.

Demonstrating the feasibility of a new custom computational pipeline for identification and characterization of underlying causal genes of phenotypes generated as part of *BonnMu* or via similar efforts, the first acyl-CoA reductase (FAR) in a barley wax mutant was reported. With this, the wax biosynthetic pathway in barley (currently the world's fourth most cultivated cereal) was further elucidated. Identifying (cuticular) wax-related genes aids crop breeding by offering new avenues to improve resilience against a lack of water caused by droughts or heat stress.

Analyses of gene functional relationships in crops were also performed with the inclusion of environmental stimuli. Using poplar hybrids (*Populus x cancescens*) the global gene regulatory effects of abiotic stresses were investigated, with special emphasis on the hydrophobic biopolymer suberin. The results indicated that the genetic enhancement of poplar root suberization could be a worthwhile strategy for genetically optimized poplar trees. These could be suitable for cultivation on water-limited and/or salty land that would be unusable for food production, thus being a great benefit to agroforestry.

Lastly, A2TEA (Automated Assessment of Trait-specific Evolutionary Adaptations) an entirely new approach and accompanying software suite for predicting genes important for stress tolerance via inference of evolutionary adaptation events was developed.

In summary, this thesis presents new bioinformatics tools for unraveling gene functional relationships and novel insights into important agronomic traits central to future breeding efforts.

Zusammenfassung

Die Verknüpfung interessanter phänotypischer Merkmale mit kausalen Genen und die Aufklärung der zugrundeliegenden Gen-Funktionsbeziehungen ist von großer Bedeutung für die Züchtung und Verbesserung von Nutzpflanzen. Zu diesem Zweck wurden bei den in dieser kumulativen Dissertation vorgestellten Forschungsarbeiten sowohl etablierte Ansätze wie z.B. differenzielle Transkriptomik und Bulked-Segregant-Analyse eingesetzt als auch neue bioinformatische Lösungen für die automatische Analyse von Insertionsmutagenese Bibliotheken und die *in silico* Vorhersage von Genfamilien, die mit Merkmalen assoziiert sind, entwickelt.

Eine effiziente Bioinformatik-Softwarelösung namens MuWU (Mutant-Seq Workflow Utility) zur Identifizierung und Annotation von vererbbaren Mutationen, die durch *Mutator*-Transposon-Insertionsstellen verursacht werden, wurde entwickelt und erwies sich als essenziell für den Aufbau und die kontinuierliche Erweiterung der weltweit größten sequenzindizierten Insertionsmutagenese Bibliothek in Mais – *BonnMu*.
Zur Demonstration der Nützlichkeit einer neuen maßgeschneiderten computergestützten Pipeline für die Identifizierung und Charakterisierung der zugrundeliegenden kausalen Gene von Phänotypen, die im Rahmen von *BonnMu* oder durch ähnliche Bemühungen erzeugt wurden, wurde die erste Acyl-CoA-Reduktase (FAR) in einer Wachsmutante in Gerste identifiziert. Hierdurch wurde die Wachsbiosynthese in Gerste (der derzeit viertmeist angebauten Getreideart der Welt) weiter aufgeklärt. Die Identifizierung von Genen, die mit (cuticulärem) Wachs in Zusammenhang stehen, hilft der Pflanzenzüchtung, indem neue Möglichkeiten zur Verbesserung der Widerstandsfähigkeit gegenüber Wassermangel durch Dürre oder Hitzestress eröffnet werden.
Analysen von Gen-Funktionsbeziehungen in Nutzpflanzen wurden auch unter Berücksichtigung von Umweltreizen durchgeführt. Anhand von Pappelhybriden (*Populus x canescens*) wurden die globalen genregulatorischen Auswirkungen von abiotischem Stress untersucht, wobei der Schwerpunkt auf dem hydrophoben Biopolymer Suberin lag. Es wurde aufgezeigt, dass sich die genetische Verbesserung der Pappel-Wurzelsuberisierung als lohnende Strategie für genetisch optimierte Pappelbäume lohnen würde. Diese könnten sich für den Anbau auf wasserarmen und/oder salzhaltigen Böden eignen, die für die Nahrungsmittelproduktion unbrauchbar wären, und können somit für die Agroforstwirtschaft von großem Nutzen sein.
Schließlich wurde mit A2TEA (Automated Assessment of Trait-specific Evolutionary Adaptations) ein neuer Ansatz und eine dazugehörige Software-Suite für die Vorhersage von Genen entwickelt, die für die Stresstoleranz wichtig sind, indem auf evolutionäre Anpassungsereignisse geschlossen wird.

Zusammenfassend ergibt sich, dass in dieser Dissertation sowohl neue Bioinformatik-Tools für die Entschlüsselung von Gen-Funktionsbeziehungen als auch neue Erkenntnisse über wichtige agronomische Merkmale vorgestellt werden, die für künftige Züchtungsbemühungen von zentraler Bedeutung sind.

Contents

Publications related to this thesis

First–Author publications:

- **MuWU: Mutant-seq library analysis and annotation**
  Tyll Stöcker, Lena Altrogge, Caroline Marcon, Yan Naing Win, Frank Hochholdinger, Heiko Schoof
  Bioinformatics 38.3 (2022): 837-838
  DOI: https://doi.org/10.1093/bioinformatics/btab679

  Own contribution:
  Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing;

- **A2TEA: Identifying trait-specific evolutionary adaptations**
  Tyll Stöcker, Carolin Uebermuth-Feldhaus, Florian Boecker, Heiko Schoof
  F1000Research 11.1137 v2 (2023): 1137.
  DOI: https://doi.org/10.12688/f1000research.126463.2

  Own contribution:
  Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing;

- **Expanding the *BonnMu* sequence-indexed repository of transposon induced maize (*Zea mays* L.) mutations in dent and flint germplasm**
  Yan Naing Win, Tyll Stöcker, Xuelian Du, Alexa Brox, Marion Pitz, Alina Klaus, Hans-Peter Piepho, Heiko Schoof, Frank Hochholdinger, Caroline Marcon
  The Plant Journal 120.5 (2024): 2253-2268
  (*at the time of review of this dissertation this publication was still under review itself*)
  DOI: https://doi.org/10.1111/tpj.17088

  Own contribution:
  Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing;

Publications related to this thesis

Co–Author publications:

- ***BonnMu*: A Sequence-Indexed Resource of Transposon-Induced Maize Mutations for Functional Genomics Studies**
  Caroline Marcon, Lena Altrogge, Yan Naing Win, Tyll Stöcker, Jack M. Gardiner, John L. Portwood, II, Nina Opitz, Annika Kortz, Jutta A. Baldauf, Charles T. Hunter, Donald R. McCarty, Karen E. Koch, Heiko Schoof, Frank Hochholdinger
  Plant physiology 184.2 (2020): 620-631
  DOI: https://doi.org/10.1104/pp.20.00478

  Own contribution:
  Together with LA and HS, I conceived and carried out the bioinformatics analyses. LA and I wrote the software used.

- ***Populus* × *canescens* root suberization in reaction to osmotic and salt stress is limited to the developing younger root tip region**
  Paul Grünhofer, Tyll Stöcker, Yayu Guo, Ruili Li, Jinxing Lin, Kosala Ranathunge, Heiko Schoof, Lukas Schreiber
  Physiologia Plantarum 174.5 (2022): e13765
  DOI: https://doi.org/10.1111/ppl.13765

  Own contribution: PG and I evaluated the transcriptomics data and wrote the manuscript.

- **Isolation and characterization of the gene *HvFAR1* encoding acyl-CoA reductase from the *cer-za.227* mutant of barley (*Hordeum vulgare*) and analysis of the cuticular barrier functions**
  Yannic Müller, Payal Patwari, Tyll Stöcker, Viktoria Zeisler-Diehl, Ulrike Steiner, Chiara Campoli, Lea Grewe1, Magdalena Kuczkowska, Maya Marita Dierig, Sarah Jose, Alistair M. Hetherington, Ivan Acosta, Heiko Schoof, Lukas Schreiber, Peter Dörmann
  New Phytologist 239.5 (2023): 1903-1918
  DOI: https://doi.org/10.1111/nph.19063

  Own contribution:
  Together with HS, I conceived the bioinformatics analyses; I performed the bioinformatics analyses and together with YM, PP and PD wrote the manuscript.

Further publications:

- **ENHANCED GRAVITROPISM 2 encodes a STERILE ALPHA MOTIF–containing protein that controls root growth angle in barley and wheat**
  Gwendolyn K. Kirschner, Serena Rosignoli, Li Guo, Isaia Vardanega, Jafargholi Imani, Janine Altmüller, Sara G. Milner, Raffaella Balzano, Kerstin A. Nagel, Daniel Pflugfelder, Cristian Forestan,Riccardo Bovina, Robert Koller, Tyll G. Stöcker, Martin Mascher, James Simmonds, Cristobal Uauy,Heiko Schoof, Roberto Tuberosa, Silvio Salvi, and Frank Hochholdinger
  Proceedings of the National Academy of Sciences 118.35 (2021): e2101526118

- **Transcriptome profiling at the transition to the reproductive stage uncovers stage and tissue-specific genes in wheat**
  Salma Benaouda, Tyll Stöcker, Heiko Schoof, Jens Léon, Agim Ballvora
  BMC Plant Biology 23.1 (2023): 25

- **NCBench: providing an open, reproducible, transparent, adaptable, and continuous benchmark approach for DNA-sequencing-based variant calling**
  Friederike Hanssen, Gisela Gabernet, Nicholas H Smith, Christian Mertes, Avirup Guha Neogi, Leon Brandhoff, Anna Ossowski, Janine Altmueller, Kerstin Becker, Andreas Petzold, Marc Sturm, Tyll Stöcker, Sugirthan Sivalingam, Fabian Brand, Axel Schmid, Andreas Buness, Alexander J Probst, Susanne Motameny, Johannes Köster
  F1000Research 12 (2024): 1125

- **Changes in qualitative wax composition but not in quantitative wax amount enhance cuticular transpiration**
  Paul Grünhofer, Lena Herzig, Qihui Zhang, Simon Vitt, Tyll Stöcker, Yaron Malkowsky, Tobias Bruegmann, Matthias Fladung, Lukas Schreiber
  Plant, Cell & Environment 47.1 (2023): 91-105

- **Suberin deficiency and its effect on the transport physiology of young poplar roots**
  Paul Grünhofer, Ines Heimerich, Svenja Pohl, Marlene Oertel, Hongjun Meng, Lin Zi, Kevin Lucignano, Syed Nadeem Hussain Bokhari, Yayu Guo, Ruili Li, Jinxing Lin, Matthias Fladung, Tino Kreszies, Tyll Stöcker, Heiko Schoof, Lukas Schreiber
  New Phytologist 242.1 (2024): 137-153

- **Cold mediates maize root hair developmental plasticity via epidermis-specific transcriptomic responses**
  Yaping Zhou, Mauritz Leonard Sommer, Annika Meyer, Danning Wang, Alina Klaus, Tyll Stöcker, Caroline Marcon, Heiko Schoof, Georg Haberer, Chris-Carolin Schön, Peng Yu, Frank Hochholdinger
  Plant Physiology 196.3 (2024): 2105-2120

Conference participation

Attended Conferences with presentation:

- **A2TEA: Identifying trait-specific evolutionary adaptations**
  Tyll Stöcker; Carolin Uebermuth-Feldhaus; Florian Boecker; Heiko Schoof
  German Conference on Bioinformatics 2022 (*Poster Presentation*)

Attended Conferences without presentation:

- German Conference on Bioinformatics 2019
- Plant & Animal Genome 2020
- Research in Computational Molecular Biology (Recomb) 2021
- International Conference of the German Society for Plant Sciences 2022

Further Presentations:

- **MuWU - a Snakemake workflow for Mu-seq library annotation and analysis**
  Tyll Stöcker, Lena Altrogge, Caroline Marcon, Yan Naing Win, Frank Hochholdinger, Heiko Schoof
  West German Genome Center, Webinar Series 2021 (*Presentation*)

List of Figures

continues on next page

List of Figures continued from previous page

continues on next page

List of Figures continued from previous page

List of Tables

Lists of Supplemental Data

Chapter	Information/Link
Chapter 2	MuWU GitHub Repository
	MuWU v1.5 Zenodo DOI
Chapter 5	A2TEA.Workflow GitHub Repository
	A2TEA.Workflow v1.1 Zenodo DOI
	A2TEA.Workflow Test Data Zenodo DOI
	A2TEA.WebApp GitHub Repository
	A2TEA.WebApp v1.1.5 Zenodo DOI
	A2TEA.WebApp Test Data (Poaceae reduced) Zenodo DOI
	A2TEA.WebApp Test Data (Brassicaceae) Zenodo DOI

1 Introduction

1.1 Gene Functional Relationships – an Overview

The study of gene functional relationships is central to our understanding of crop biology and thereby, the core breeding goal of safeguarding a stable worldwide food supply. Comprehending the function of genes in crucial food crops such as rice, wheat, and maize is particularly important; for example, these three alone provide roughly half of the calories consumed by humans (Ranum et al., 2014; Poole et al., 2021).

The interplay between genes and their functions is at the core of genetics/ evolutionary biology research and thus a brief historical overview is reasonable. The study of these relationships began with the seminal work of Gregor Mendel, who, in the mid-19th century began to conduct variation experiments on common pea plants which ultimately led him to propose the core principles of heredity. Most importantly, that the inheritance of traits is grounded in discrete units of heredity – now commonly referred to as genes (Tateno 2013). Mendel's experiments demonstrated that the inheritance of several traits in peas is determined by segregation and an independent assortment of genes. He further observed that the ratio of traits in the offspring of two parents was predictable and followed certain laws (Mendel 1866). Mendel's work laid the foundation for the study of gene functional relationships, and since then, countless advances have been made in our understanding of the relationship between genes and their functions. Novel insights have led to an ever-increasing assortment of vernacular and methods to describe and study them.

Chief among them are the umbrella terms "Forward" and "Reverse" genetics which describe two complementary approaches used to identify phenotype-altering mutations and define the functions of genes in biological processes. In plants, forward genetics involves screening of specimens with either naturally occurring or induced mutations to identify those with desirable traits, such as increased yield or disease resistance. Once a plant with a desirable trait is identified, further work is performed to identify the causative mutation. One could argue that the first step of canonical forward genetics is, in fact, a long-standing approach that predates Mendel's first formalization of the field and has its roots in man's first attempts at the domestication of plants and animals (Jankowicz-Cieslak et al., 2015); e.g., useful traits such as reduced seed shattering, were selected from crop populations and used for subsequent propagation, thereby constituting the first cultivars (Jankowicz-Cieslak et al., 2015; Maity et al., 2021).

While spontaneous mutations are the ultimate source of genetic variation that under-

lies naturally occurring heritable biological diversity (Huang et al., 2016), it was the discovery that mutations could also be artificially induced at much higher frequencies in the 1920s (Muller 1928; Stadler 1928), which paved the way for the establishment of thousands of novel crop cultivars. This, together with the revolutionary insight that DNA constitutes the heritable material, ultimately made reverse genetics possible, in which specific genomic alterations are recovered and tested for gene-functional variation.

Laboratory-based functional analyses provided insights into the molecular fundamentals; however, it was the emergence of the necessary computing power and bioinformatics as a field that truly paved the way for large-scale gene function research (and many other domains). One of the most important contributions to modern biological research, the development of efficient sequence similarity analyses/searches, has enabled researchers to compare genetic and proteomic sequences in a variety of ways. At the core of these is the concept of homology: two sequences are more similar than expected by chance, which indicates a shared evolutionary history. Since protein functions are evolutionarily conserved under selective pressure, this conservation is mirrored in its function-giving structure and underlying amino acid sequence. Homology is therefore strongly linked to function, although it has to be mentioned that the precise amount of similarity needed to cause similar structure and ultimately function is still a field of active research; for example, a function is particularly difficult to assign to a sequence since any given protein structure can have multiple and "function" is a fundamentally ambiguous concept with tissue- and concentration-specific impacts (Pearson 2013). However, the indispensable value of sequence similarity analyses is reflected in them being central to many areas of research, such as phylogenomics, forensics, and function prediction, but also in that the underlying publications of seminal computational tools rank among the most cited papers of all time (e.g., BLAST (Altschul et al., 1990) and CLUSTALW/X (Larkin et al., 2007)).

One of the key developments in our understanding of gene functional relationships has been the recognition of genome and gene duplications as providers of "raw" material for new genes and functions to arise. Whereas genome duplications (instances of an organism's entire genome being copied) cause copies of all genes, gene duplications are defined as singular or group copy events that lead to multiple copies of one or several genes within the genome. The first recognition of the importance of these processes can in large part be attributed to Susumu Ohno, who proposed in his formative work on the topic ("Evolution by Gene Duplication") that duplication events serve as a mechanism for the evolutionary functional diversification of genes (Ohno 2013). Ohno's hypothesis "regarding the importance of gene duplication in the generation of evolutionary novelty has steadily gained support as we have entered the genome-sequencing era" (Liberles et al., 2010), as it has been generally accepted that duplications serve as a source of additional genetic material on which forces such as mutation, drift, and selection can act, ultimately making new evolutionary opportunities possible (Crow & Wagner, 2006). Summarized by Owens et al. (2013) Ohno's model of gene duplication "predicts that new genes are formed from the asymmetric functional divergence of a newly arisen, redundant duplicate locus". This is

because gene duplicates are under relaxed evolutionary constraints and are thus enabled to acquire new functions or new expression patterns, thereby constituting a major driver of molecular evolution (Ohno 2013; Taylor & Raes, 2004). Genes can diverge in function through a variety of processes, such as "accumulation (fixation) of coding sequence changes, which may influence binding interactions and catalysis, through the evolution of splice variants, and through spatial, temporal, and concentration-level changes in the expression of the protein product" (Liberles et al., 2010).

While the inference of any individual gene's functions is intrinsically valuable for high-yielding crop breeding and production strategies, it is crucial to recognize that evolutionary (and domestication) pressures gave rise to these functions in the first place. Since favorable resistance traits are quantitative in nature and governed by complex interactions of many genes, it is an understanding of the underlying adaptive evolutionary history that will ultimately allow us to comprehend them in full. As such, deepening our understanding of the origin of favorable functions will only aid our crop breeding efforts, as we face the ever-increasing challenges of both a growing world population and climate change.

1.2 Adaptive Evolution of Gene Functions in Plants and Crops

While the core principles of adaptation seem to be conserved across all life, kingdoms differ in the relative importance that specific adaptive processes play in their evolutionary past. In plants in particular, gene family expansion is an important driver of adaptation processes (Stöcker et al., 2023) and is mainly due to whole-genome duplication events (Adams et al., 2003; Moore & Purugganan, 2005); for example, it has been shown that all angiosperms have experienced at least one common genome duplication before the monocot-dicot divergence (Bowers et al., 2003).

However, plant gene duplications also occur on a smaller scale. Unequal crossing-over is a process that occurs during meiosis, in which homologous chromosomes pair, and genetic material is exchanged, which can result in the duplication and tandem (consecutive) arrangement of specific genes or gene segments (Tartof 1988). Single-gene duplicates can also be transposed into distant genomic locations via DNA- or RNA-based mechanisms. The former is facilitated by class II (DNA) transposons that relocate duplications to another genomic position; plant species possess different types of transposable elements (TEs) that facilitate this process, with helitrons in maize (Lai et al., 2005) or packmules in rice (Jiang et al., 2005) representing prominent examples. RNA-based (retro-)transposition has been reported to usually create a single-exon retrocopy from the original gene by reverse transcription of a spliced messenger RNA (Wang et al., 2012). Since the retrocopy is based on a transcript and consequently transposed without any

promoter region, it will be functional only if a new promoter is acquired (Brosius 1991; Kaessmann et al., 2009). Categorization of emerging homologous genes is typically done by distinguishing between paralogs, arising by duplication of a gene in the same species, and orthologs, which are defined as homologous genes that originated in a common ancestor by way of speciation (Koonin, 2005).

The mechanisms of gene retention and loss following duplication, as well as the ultimate fate and function of a duplicated gene, are manifold in plants (reviewed by Panchy et al., 2016), with most gene duplicates not persisting, as they are lost or silenced through selective pressure on just one of the duplicates (Adams & Wendel, 2005). Mechanisms facilitating either pseudogenization through the accumulation of deleterious point mutations (Schnable et al., 2011), epigenetic, transposon, or RNAi-induced silencing (Adams et al., 2003; Adams & Wendel, 2005; Sémon & Wolfe, 2007; Kashkush et al., 2003), and eventual loss of a gene copy are all involved in maintaining the coherence of regulatory networks and stable levels of expression (Blanc & Wolfe, 2004; Duarte et al., 2006; Thomas et al., 2006).

However, the existence of a further gene copy also offers opportunities for adaptation, and its fate is not limited to the aforementioned processes (Grotewold et al., 2015). Retention of gene duplicates can for instance occur as a result of selection for an ecological niche and it has been shown that genes involved in the response to abiotic stress in plants are mostly retained after polyploidization events (Maere et al., 2005).

In terms of function/s, various mechanisms exist for the retention of gene duplicates (reviewed by Panchy et al., 2016). The most basic forms are gene dosage effects, as well as neo- and subfunctionalization. An increased amount of gene product caused by two copies of a gene can be advantageous and is frequently reported for housekeeping genes, that is, highly expressed genes, such as those coding for histones or rRNAs (Zhang, 2003). A gene copy can acquire a new function because of the nonsynonymous exchange of several amino acids and can be advantageous if, for instance, strong directional selection pressure exists and increased functional diversity constitutes an advantage (Lan et al., 2009). Lastly, a split of function, in which each duplicate retains only part of the original (Force et al., 1999; Lynch et al., 2001), is a process that mainly takes place under neutral evolutionary conditions during which no positive selection for either duplicate exists (Force et al., 1999; Zhang, 2003). One common form of such a split is the observed differential expression between tissues or conditions as a consequence of complementary, degenerative mutations in the regulatory sequences of two gene copies (Force et al., 1999, Adams et al., 2003, Adams & Wendel, 2005, Roulin et al., 2013).

All these constitute potentially interesting ways in which plant species may have adapted to particular environmental conditions. A well-studied example is the MADS-box gene family, which has been linked to the evolution of flowering time, various tissue shapes and flower colors in plants (Alvarez-Buylla et al., 2000; Becker & Theißen 2003; Hsu et

al., 2021), which likely facilitated co-evolution with pollinators (Mondragón-Palomino & Theißen 2008). It is important to consider the various regulatory processes that shape the complexity of gene functional relationships in plants and other eukaryotes. The control of genes via tissue-, time-, and stimulus-specific expression is central to functioning organisms and (along with post-translational modifications) ultimately determines how much effect (if any) a particular gene's protein product can have. As the control of expression is itself interlinked with gene products and the protein-biosynthesis machinery, we are faced with a self-regulating system with no clear start-or end – adaptive processes thus extending to this regulatory layer of gene function. Notable mechanisms of expression control in plants include alternative splicing (Syed et al., 2012), epigenetic modifications (Feng & Jacobsen, 2011), transcription factors (Singh et al., 2002), and microRNAs (Djami-Tchatchou et al., 2017).

In brief, duplication and the subsequent retention of gene copies, as well as genomic restructuring, promote rapid evolution in plants owing to the elevated available diversity. Gene duplications, gene family expansions, as well as the regulation of gene expression, are key mechanisms for the adaptation of crops to different environments and changing conditions. Research on the adaptation processes and both the origin and causality of (novel) gene functions is important for the continuous improvement of crop varieties.

1.3 Approaches to study Gene Functional Relationships in Crops

State-of-the-art research leverages the high throughput capabilities of next-generation sequencing (NGS). Smart integration of multiple datasets, such as expression profiles, protein-protein interaction networks, and comparative genomics, allows for a holistic view of gene functional relationships and the transfer of previously generated knowledge to other organisms. Valuable insights into the complex biology of plants/crops are achieved in this way which aids in identifying the genes that are most important for the control of specific traits.

Artificially creating mutations in genomes is a key strategy for obtaining interesting genotypes and validating the hypotheses of phenotypic consequences in reverse genetics. Although the power of targeted gene knockouts, such as CRISPR/Cas-based gene editing, cannot be overstated and has already been used to great success in crop research and improvement (Zhang et al., 2021), untargeted large-scale loss-of-function mutant libraries offer the generation of a large number of mutants in a relatively short period of time as well as a range of mutations, including those that may not be predicted to affect gene function. Sequence-indexed insertional libraries, in which insertions of additional DNA sequences cause non-functional proteins or stop transcription altogether, are crucial tools for plant researchers worldwide, both in forward and reverse genetic investigations

(Stöcker et al., 2022). Many mutant collections exist, for example, based on the use of TEs in maize (Marcon et al., 2020; Liang et al., 2019; McCarty et al., 2005) and rice (Hirochika et al., 2004), or transgenic T-DNA insertion lines in Arabidopsis (Alonso et al., 2003). Still, inferral of the causal mutation/s in a mutagenized plant and understanding the implications for the organism on the transcriptomic, proteomic, and metabolomic levels is a complex task that has birthed a plethora of different analyses. Therefore, a large variety of approaches exist to find or characterize candidate genes that might serve as the genetic foundation for better e.g., more resilient crops (Stöcker et al., 2023). Described by Stöcker et al. (2023), this includes forward genetics techniques, such as identifying the genes responsible for advantageous mutant phenotypes (Kirschner et al., 2021) or discovering common regulators for multiple stresses using traditional transcriptomics, most commonly differential expression analysis (DEA; Sham et al., 2015). Common methods include Quantitative Trait Locus (QTL) mapping and/or genome-wide association studies (GWAS) with the potential integration of expression data (Guo et al., 2022). Also used in candidate gene predictions are machine learning-based approaches that use transcriptomic or phenomic data (Shaik & Ramakroshna 2014; Braun et al., 2020), as well as combinations of expression data with functional information and clustering techniques (Sewelam et al., 2020; Kar et al., 2021). This also extends to comparative genomic approaches that combine data from several plants (Shinozuka et al., 2012).

The wide assortment of underlying methods includes set analyses of DEGs, frequently in conjunction with pathway or Gene Ontology (GO) term enrichment analyses (Supek et al., 2017), co-expression networks (Aoki et al., 2007), weighted correlation network analysis (WGCNA; Langfelder et al., 2008), bulked-segregant analysis (BSA; Michelmore et al., 1991), and k-means clustering (Likas et al., 2003). With regard to the contents of this dissertation, the most important of these research approaches, besides the already discussed domain of insertional mutagenesis, are BSA, DEA, and the field of comparative genomics.

BSA is a mapping technique used in plant genetics to identify causal genes for phenotypes of interest, such as disease resistance (Hyten et al., 2009) or root growth angle (Kirschner et al., 2021). BSA works by pooling specimens with a particular (mutant) phenotype and comparing the DNA of the pool with another (usually wildtype) pool lacking the trait to identify genetic markers that are linked to the phenotype of interest (Zou et al., 2016). In combination with NGS, BSA has proven to be valuable for the characterization of crop mutants in various species, such as cotton, wheat, and soybean (Forrest et al., 2014; Chen et al., 2015; Thakare et al., 2017).

DEA is a quantitative comparison analysis that leverages our capability to sequence RNA and thereby compare the number of gene transcripts (used as a proxy for final protein product concentrations) between experimental conditions. In crop research, it facilitates the study of important biological processes, such as stress response, development, or metabolism, and how stresses or mutations impact the overall transcriptomic landscape of any specific cultivar.

Lastly, comparative genomics, from a high-level perspective, is the study of genomic sim-

ilarities and differences between different species or populations (Morrell et al., 2012). Facilitated by the ever-increasing genomic resources for plant and crop species, with even high-quality cultivar-specific genomes becoming available, exciting research regarding conserved or unique genes, chromosomal restructuring, and evolutionary history is made possible. To provide an example of interesting insights enabled by comparative genomics in plant research, one must look no further than gene family sizes. Research has shown that the majority of gene families are in fact conserved over long evolutionary time periods, including even the diversification of all angiosperms and non-flowering plants (Panchy et al., 2016). Interestingly, lineage-specific variations in gene family size, which are common among taxa, coexist with this conservation of gene families (Rensing et al., 2008; Flagel et al., 2009; Panchy et al., 2016). This implies that since relatively few original gene families emerged, a large portion of the high level of diversity and phenotypic variation observed in terrestrial plants may have primarily originated as a result of gene duplication and subsequent adaptive specialization (Panchy et al., 2016; Stöcker et al., 2023).

1.4 Aims of the Thesis – Goals of the Scientific Work

As outlined in the previous sections, the association of interesting phenotypic traits with causal genes and elucidating the sometimes complex mechanistic underpinnings is of great importance to crop breeding and improvement efforts. To this end, my work has included the use of established approaches, such as differential transcriptomics and bulked-segregant analysis, as well as the development of novel bioinformatics solutions for both the automatic analysis of insertional mutagenesis libraries and the *in silico* prediction of trait-associated gene families.
The publications included in the chapters of this dissertation follow the narrative structure outlined below:

1. Chapter 2 details my work on an efficient software solution (Stöcker et al., 2022) for the identification and annotation of heritable mutations caused by TE insertion sites, as part of a larger insertional mutagenesis effort in maize (Marcon et al., 2020; a follow-up publication is currently under review: Win & Stöcker et al., 2024).

2. As previously detailed, the logical follow-up of generating phenotypes is the identification and characterization of the underlying causal genes. Chapter 3 reports the first acyl-CoA reductase (FAR) in a barley wax (so-called *eceriferum*) mutant (Müller et al., 2023). In this project, my bioinformatics analyses allowed us to identify the genetic variants underlying the altered phenotype.

3. Chapter 4 extends my analyses of gene functional relationship to interactions of crops with their environment – special focus was put on the hydrophobic biopolymer suberin. Using poplar hybrids (Grünhofer et al., 2022) we investigated the global gene regulatory effects of abiotic stresses.

4. Finally, Chapter 5 describes a novel approach and accompanying software suite for predicting genes important for stress tolerance via inference of evolutionary adaptation events (Stöcker et al., 2023). This last project leads to a gratifying ring closure because the validation of the first predictions with this new software is currently underway based on mutants generated from the project described in Chapter 2.

1.5 References

Adams, K. L., Cronn, R., Percifield, R., & Wendel, J. F. (2003). Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. *Proceedings of the National Academy of sciences, 100*(8), 4649-4654.

Adams, K. L., & Wendel, J. F. (2005). Polyploidy and genome evolution in plants. *Current opinion in plant biology, 8*(2), 135-141.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology, 215*(3), 403-410.

Alvarez-Buylla, E. R., Liljegren, S. J., Pelaz, S., Gold, S. E., Burgeff, C., Ditta, G. S., ... & Yanofsky, M. F. (2000). MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. *The Plant Journal, 24*(4), 457-466.

Becker, A., & Theißen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Molecular phylogenetics and evolution, 29*(3), 464-489.

Blanc, G., & Wolfe, K. H. (2004). Functional divergence of duplicated genes formed by polyploidy during Arabidopsis evolution. *The Plant Cell, 16*(7), 1679-1691.

Bowers, J. E., Chapman, B. A., Rong, J., & Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature, 422*(6930), 433-438.

Brosius, J. 1991. Retroposons—seeds of evolution. *Science, 251*(4995), 753-753.

Chen, W., Yao, J., Chu, L., Yuan, Z., Li, Y., & Zhang, Y. (2015). Genetic mapping of the nulliplex-branch gene (*gb_nb1*) in cotton using next-generation sequencing. *Theoretical and Applied Genetics, 128*, 539-547.

Crow, K. D., & Wagner, G. P. (2005). What is the role of genome duplication in the evolution of complexity and diversity? *Molecular biology and evolution, 23*(5), 887-892.

Djami-Tchatchou, A. T., Sanan-Mishra, N., Ntushelo, K., & Dubery, I. A. (2017). Functional roles of microRNAs in agronomically important plants—potential as targets for crop improvement and protection. *Frontiers in plant science, 8*, 242053.

Duarte, J. M., Cui, L., Wall, P. K., Zhang, Q., Zhang, X., Leebens-Mack, J., ... & DePamphilis, C. W. (2006). Expression pattern shifts following duplication indicative of subfunctionalization and neofunctionalization in regulatory genes of Arabidopsis. *Molecular biology and evolution, 23*(2), 469-478.

Feng, S., & Jacobsen, S. E. (2011). Epigenetic modifications in plants: an evolutionary perspective. *Current opinion in plant biology, 14*(2), 179-186.

Flagel, L. E., & Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytologist, 183*(3), 557-564.

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., & Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics, 151*(4), 1531-1545.

Forrest, K., Pujol, V., Bulli, P., Pumphrey, M., Wellings, C., Herrera-Foessel, S., ... & Spielmeyer, W. (2014). Development of a SNP marker assay for the *Lr67* gene of wheat using a genotyping by sequencing approach. *Molecular Breeding, 34*, 2109-2118.

Grotewold, E., Chappell, J., & Kellogg, E. A. (2015). *Plant genes, genomes and genetics*. John Wiley & Sons.

Grünhofer, P., Stöcker, T., Guo, Y., Li, R., Lin, J., Ranathunge, K., ... & Schreiber, L. (2022). *Populus × canescens* root suberization in reaction to osmotic and salt stress is limited to the developing younger root tip region. *Physiologia Plantarum, 174*(5), e13765.

Hsu, H. F., Chen, W. H., Shen, Y. H., Hsu, W. H., Mao, W. T., & Yang, C. H. (2021). Multifunctional evolution of B and *AGL6* MADS box genes in orchids. *Nature communications, 12*(1), 902.

Huang, W., Lyman, R. F., Lyman, R. A., Carbone, M. A., Harbison, S. T., Magwire, M. M., & Mackay, T. F. (2016). Spontaneous mutations and the origin and maintenance of quantitative genetic variation. *Elife, 5*, e14625.

Hyten, D. L., Smith, J. R., Frederick, R. D., Tucker, M. L., Song, Q., & Cregan, P. B. (2009). Bulked segregant analysis using the GoldenGate assay to locate the *Rpp3* locus that confers resistance to soybean rust in soybean. *Crop Science, 49*(1), 265-271.

Jankowicz-Cieslak, J., & Till, B. J. (2015). Forward and reverse genetics in crop breeding. *Advances in plant breeding strategies: breeding, biotechnology and molecular tools*, 215-240.

Jiang, N., Bao, Z., Zhang, X., Eddy, S. R., & Wessler, S. R. (2004). Pack-MULE transposable elements mediate gene evolution in plants. *Nature, 431*(7008), 569-573.

Kaessmann, H., N. Vinckenbosch & M. Long. 2009. RNA-based gene duplication: mechanistic and evolutionary insights. *Nat. Rev. Genet. 10*: 19– 31.

Kashkush, K., Feldman, M., & Levy, A. A. (2003). Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature genetics, 33*(1), 102-106.

Koonin, E. V. (2005). Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet., 39*, 309-338.

Kumar, A., & Bennetzen, J. L. (1999). Plant retrotransposons. *Annual review of genetics, 33*(1), 479-532.

Lai, J., Li, Y., Messing, J., & Dooner, H. K. (2005). Gene movement by Helitron transposons contributes to the haplotype variability of maize. *Proceedings of the National Academy of Sciences, 102*(25), 9068-9073.

Lan, T., Yang, Z. L., Yang, X., Liu, Y. J., Wang, X. R., & Zeng, Q. Y. (2009). Extensive functional diversification of the Populus glutathione S-transferase supergene family. *The Plant Cell, 21*(12), 3749-3766.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., ... & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *bioinformatics, 23*(21), 2947-2948.

Liberles, D. A., Kolesov, G., & Dittmar, K. (2010). Understanding gene duplication through biochemistry and population genetics. *Evolution after gene duplication*, 1-21.

Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., & Van de Peer, Y. (2005). Modeling gene and genome duplications in eukaryotes. *Proceedings of the National Academy of Sciences, 102*(15), 5454-5459.

Maity, A., Lamichaney, A., Joshi, D. C., Bajwa, A., Subramanian, N., Walsh, M., & Bagavathiannan, M. (2021). Seed shattering: A trait of evolutionary importance in plants. *Frontiers in Plant Science, 12*, 657773.

Marcon, C., Altrogge, L., Win, Y. N., Stöcker, T., Gardiner, J. M., Portwood, J. L., ... & Hochholdinger, F. (2020). *BonnMu*: a sequence-indexed resource of transposon-induced maize mutations for functional genomics studies. *Plant physiology, 184*(2), 620-631.

Mendel, G. (1866). *Experiments on Plant Hybrids.* Verhandlungen des naturforschenden Vereines in Brünn, 4, 3-47.

Mondragón-Palomino, M., & Theißen, G. (2008). MADS about the evolution of orchid flowers. *Trends in plant science, 13*(2), 51-59.

Morrell, P. L., Buckler, E. S., & Ross-Ibarra, J. (2012). Crop genomics: advances and applications. *Nature Reviews Genetics, 13*(2), 85-96.

Muller, H. J. (1928). The production of mutations by X-rays. *Proceedings of the National Academy of Sciences*, 14(9), 714-726.

Müller, Y., Patwari, P., Stöcker, T., Zeisler-Diehl, V., Steiner, U., Campoli, C., ... & Dörmann, P. (2023). Isolation and characterization of the gene *HvFAR1* encoding acyl-CoA reductase from the *cer-za. 227* mutant of barley (*Hordeum vulgare*) and analysis of the cuticular barrier functions. *New Phytologist, 239* (5), 1903-1918

Ohno, S. (2013). *Evolution by gene duplication.* Springer Science & Business Media.

Owens, S. M., Harberson, N. A., & Moore, R. C. (2013). Asymmetric functional divergence of young, dispersed gene duplicates in Arabidopsis thaliana. *Journal of molecular evolution, 76*, 13-27.

Panchy, N., Lehti-Shiu, M., & Shiu, S. H. (2016). Evolution of gene duplication in plants. *Plant physiology*, 171(4), 2294-2316.

Pearson, W. R. (2013). An introduction to sequence similarity ("homology") searching. *Current protocols in bioinformatics*, 42(1), 3-1.

Poole, N., Donovan, J., & Erenstein, O. (2021). Agri-nutrition research: revisiting the contribution of maize and wheat to human nutrition and health. *Food Policy, 100*, 101976.

Ranum, P., Peña-Rosas, J. P., & Garcia-Casal, M. N. (2014). Global maize production, utilization, and consumption. *Annals of the new York academy of sciences, 1312*(1), 105-112.

Rensing, S. A., Lang, D., Zimmer, A. D., Terry, A., Salamov, A., Shapiro, H., ... & Boore, J. L. (2008). The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science, 319*(5859), 64-69.

Roulin, A., Auer, P. L., Libault, M., Schlueter, J., Farmer, A., May, G., ... & Jackson, S. A. (2013). The fate of duplicated genes in a polyploid plant genome. *The Plant Journal, 73*(1), 143-153.

Schnable, J. C., Springer, N. M., & 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, 108*(10), 4069-4074.

Sémon, M., & Wolfe, K. H. (2007). Consequences of genome duplication. *Current opinion in genetics & development, 17*(6), 505-512.

Shinozuka, H., Cogan, N. O., Spangenberg, G. C., & Forster, J. W. (2012). Quantitative Trait Locus (QTL) meta-analysis and comparative genomics for candidate gene prediction in perennial ryegrass (Lolium perenne L.). **BMC genetics**, 13, 1-12.

Singh, K. B., Foley, R. C., & Oñate-Sánchez, L. (2002). Transcription factors in plant defense and stress responses. *Current opinion in plant biology, 5*(5), 430-436.

Stadler, L. J. (1928). Mutations in barley induced by X-rays and radium. *Science, 68*(1756), 186-187.

Stöcker, T., Altrogge, L., Marcon, C., Win, Y. N., Hochholdinger, F., & Schoof, H. (2022). MuWU: Mutant-seq library analysis and annotation. *Bioinformatics*, 38(3), 837-838.

Stöcker, T., Uebermuth-Feldhaus, C., Boecker, F., & Schoof, H. (2023). A2TEA: Identifying trait-specific evolutionary adaptations *F1000Research, 11*, 1137

Syed, N. H., Kalyna, M., Marquez, Y., Barta, A., & Brown, J. W. (2012). Alternative splicing in plants–coming of age. *Trends in plant science, 17*(10), 616-623.

Tartof, K. D. (1988). Unequal crossing over then and now. *Genetics, 120*(1), 1.

Tateno, Y. (2013). Development of Information Biology (I). *Interdisciplinary Bio Central, 5*(1), 2-1.

Thakare, D. S., Chimote, V. P., Adsul, A., Deshmukh, M. P., & Pulate, S. C. (2017). Molecular tagging of pod shattering tolerance trait in soybean [*Glycine max* (L.) Merrill] genotype MACS-450. *Legume Research-An International Journal, 40*(2), 224-231.

Thomas, B. C., Pedersen, B., & Freeling, M. (2006). Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. *Genome research, 16*(7), 934-946.

Wang, Y., Wang, X., & Paterson, A. H. (2012). Genome and gene duplications and gene expression divergence: a view from plants. *Annals of the New York Academy of Sciences, 1256*(1), 1-14.

Zhang, D., Zhang, Z., Unver, T., & Zhang, B. (2021). CRISPR/Cas: A powerful tool for gene function study and crop improvement. *Journal of Advanced Research, 29*, 207-221.

Zhang, J. (2003). Evolution by gene duplication: an update. *Trends in ecology & evolution, 18*(6), 292-298.

Zou, C., Wang, P., & Xu, Y. (2016). Bulked sample analysis in genetics, genomics and crop improvement. *Plant biotechnology journal, 14*(10), 1941-1955.

Bioinformatics, 38(3), 2022, 837–838
doi: 10.1093/bioinformatics/btab679
Advance Access Publication Date: 29 September 2021
Applications Note

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.

Downloaded from https://academic.oup.com/bioinformatics/article/38/3/837/6377773 by Universitaet Bonn user on 04 April 2023

Mu-seq experiment results in 2*n* 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/article/38/3/837/6377773 by Universitaet Bonn user on 04 April 2023

Check for updates

Breakthrough Technologies

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 Universitaet Bonn user on 04 April 2023

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 Universitaet Bonn user on 04 April 2023

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 Universitaet Bonn user on 04 April 2023

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.

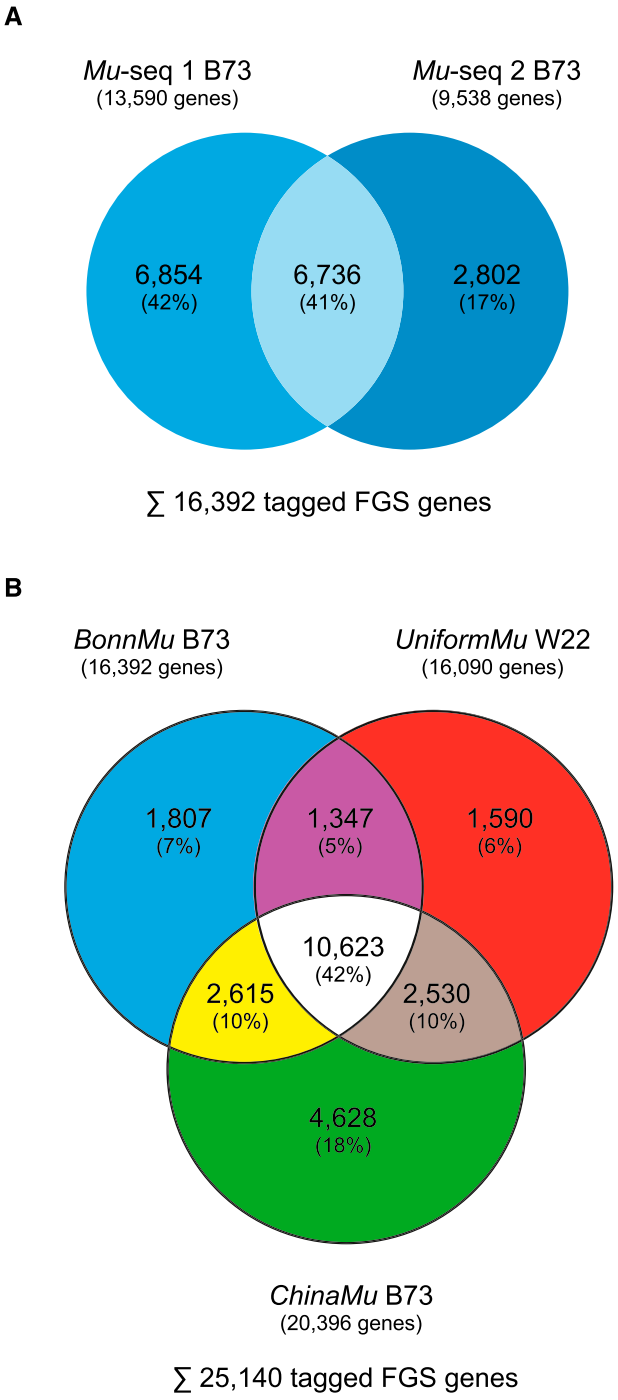


Figure 1. Overlap of FGS genes tagged by *Mu* transposon insertions. A, Tagged FGS genes among the two *Mu*-seq libraries in the B73 background. B, FGS genes with *Mu* insertions in the three available *Mu*-tagged resources in the B73 (blue, *BonnMu*; green, *ChinaMu*) and W22 inbred backgrounds (red,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

Downloaded from https://academic.oup.com/plphys/article/184/2/620/6117865 by Universitaet Bonn user on 04 April 2023

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.

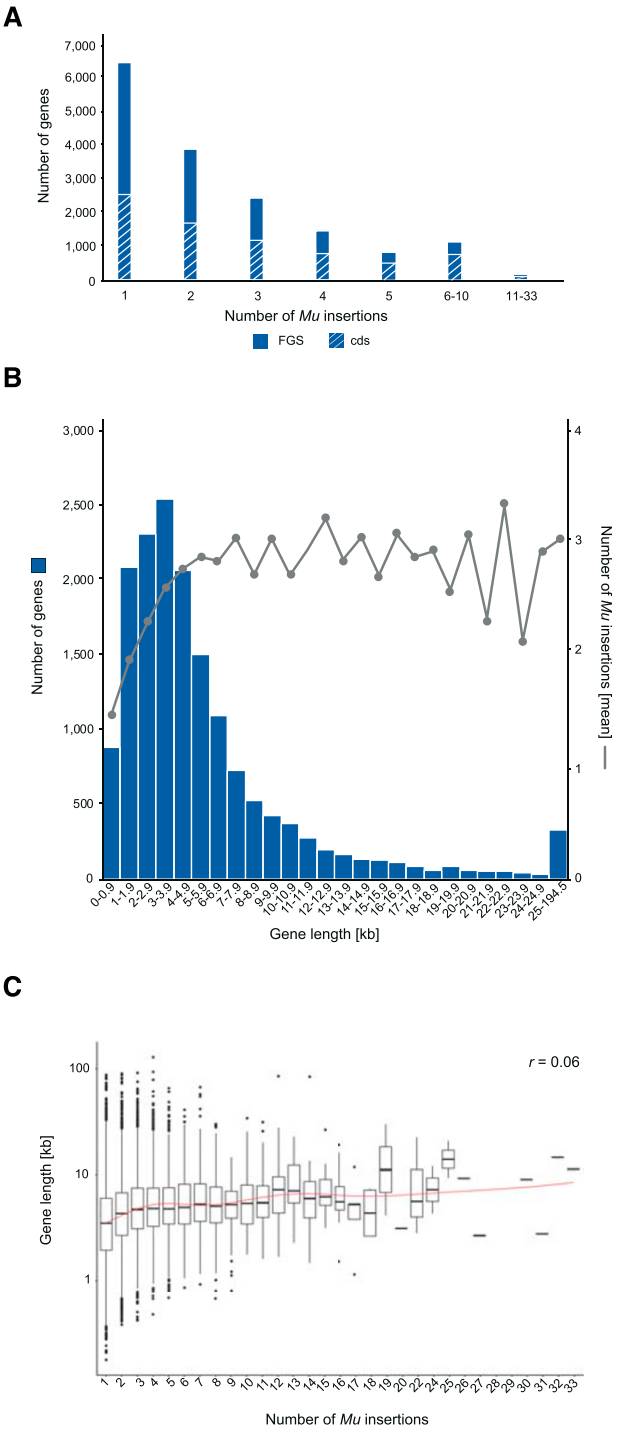


Figure 2. Number of *Mu* insertions and its correlation to the length of affected FGS genes of B73. A, Distribution and number of *Mu* insertions among 16,392 affected FGS genes and the subset of 7,582 genes harboring insertions in the cds of FGS genes. B, Number of tagged genes and associated mean number 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 Universitaet Bonn user on 04 April 2023

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.

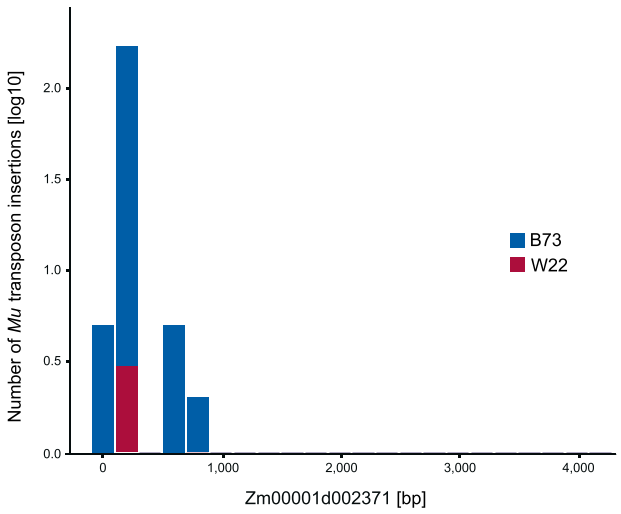


Figure 3. A *Mu* insertion hotspot in the representative gene Zm00001d002371 affected in the B73 and W22 genomes. The number of *Mu* insertions is plotted against location in Zm00001d002371.

Downloaded from https://academic.oup.com/plphys/article/184/2/620/6117865 by Universitaet Bonn user on 04 April 2023

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 (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.

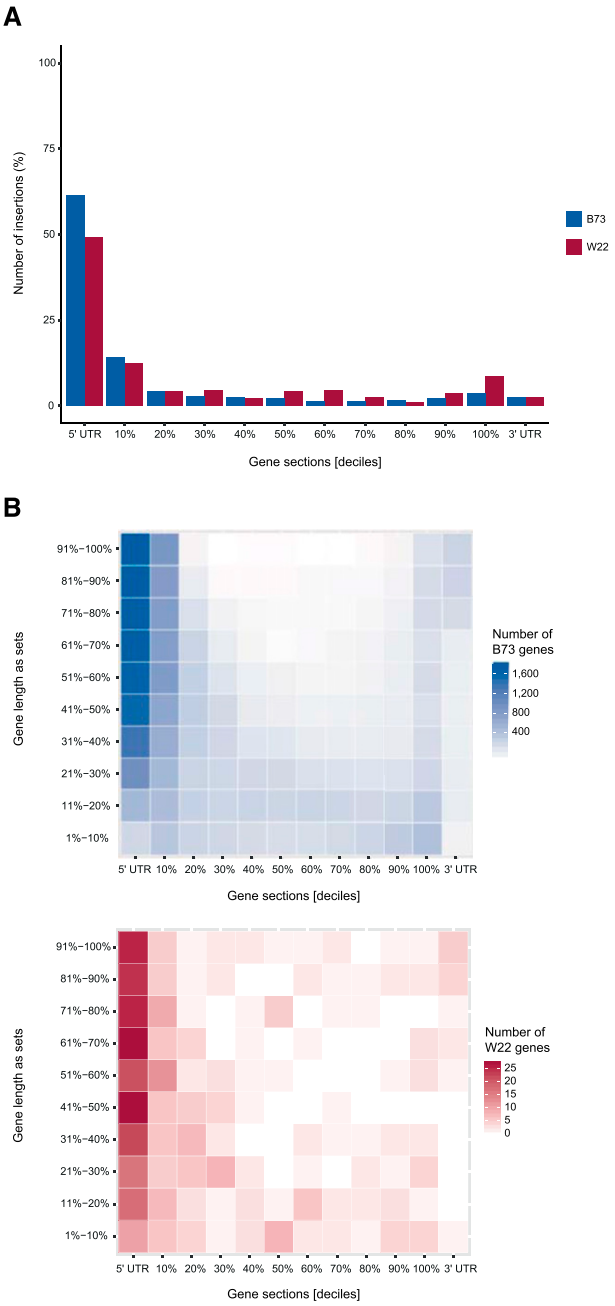
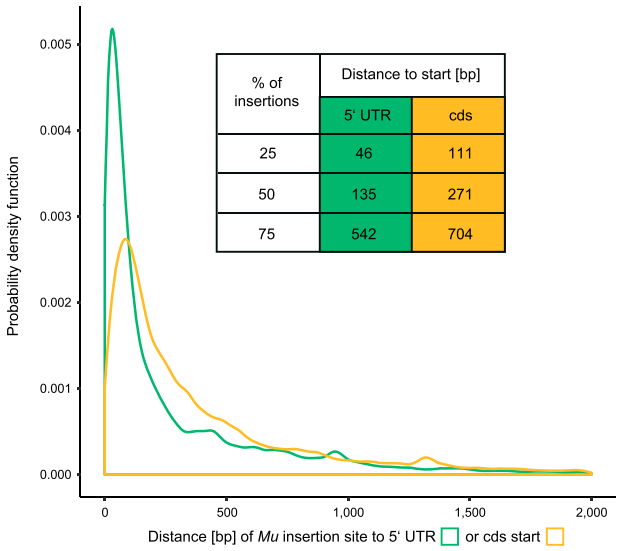


Figure 4. Distribution of *Mu* transposon insertion sites in FGS genes. A, Proportion of *Mu* insertions located in each decile of the cds, 5′ UTR, and 3′ UTR (B73, blue; W22, red). B, All FGS genes targeted by *Mu* transposons in B73 and W22 (SRP028545) were divided into gene sets based on their length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

Downloaded from https://academic.oup.com/plphys/article/184/2/620/6117865 by Universitaet Bonn user on 04 April 2023



% of insertions	Distance to start [bp]	
	5′ UTR	cds
25	46	111
50	135	271
75	542	704

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 Universitaet Bonn user on 04 April 2023

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 Universitaet Bonn user on 04 April 2023

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 Universitaet Bonn user on 04 April 2023

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 Universitaet Bonn user on 04 April 2023

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 Universitaet Bonn user on 04 April 2023

Check for updates

New Phytologist

Research

Isolation and characterization of the gene *HvFAR1* encoding acyl-CoA reductase from the *cer-za.227* mutant of barley (*Hordeum vulgare*) and analysis of the cuticular barrier functions

Yannic Müller[1]*, Payal Patwari[1]*, Tyll Stöcker[2], Viktoria Zeisler-Diehl[3], Ulrike Steiner[2], Chiara Campoli[4,5], Lea Grewe[1], Magdalena Kuczkowska[1], Maya Marita Dierig[1], Sarah Jose[6], Alistair M. Hetherington[6], Ivan F. Acosta[7], Heiko Schoof[2], Lukas Schreiber[3] and Peter Dörmann[1]

[1]Institute of Molecular Physiology and Biotechnology of Plants, University of Bonn, 53115 Bonn, Germany; [2]Institute of Crop Science and Resource Conservation, University of Bonn, 53115 Bonn, Germany; [3]Institute for Cellular and Molecular Botany, University of Bonn, 53115 Bonn, Germany; [4]Division of Plant Sciences, School of Life Sciences, University of Dundee, Dundee, DD2 5DA, UK; [5]James Hutton Institute, Dundee, DD2 5DA, UK; [6]School of Biological Sciences, University of Bristol, Bristol, BS8 1TQ, UK; [7]Max-Planck Institute for Plant Breeding Research, 50829 Cologne, Germany

Author for correspondence:
Peter Dörmann
Email: doermann@uni-bonn.de

Received: *7 February 2023*
Accepted: *16 May 2023*

New Phytologist (2023) **239:** 1903–1918
doi: 10.1111/nph.19063

Key words: barley, cuticle, cutin, eceriferum, permeability, wax.

Summary

- The cuticle is a protective layer covering aerial plant organs. We studied the function of waxes for the establishment of the cuticular barrier in barley (*Hordeum vulgare*). The barley *eceriferum* mutants *cer-za.227* and *cer-ye.267* display reduced wax loads, but the genes affected, and the consequences of the wax changes for the barrier function remained unknown.
- Cuticular waxes and permeabilities were measured in *cer-za.227* and *cer-ye.267*. The mutant loci were isolated by bulked segregant RNA sequencing. New *cer-za* alleles were generated by genome editing. The CER-ZA protein was characterized after expression in yeast and Arabidopsis *cer4-3*.
- *Cer-za.227* carries a mutation in *HORVU5Hr1G089230* encoding acyl-CoA reductase (FAR1). The *cer-ye.267* mutation is located to *HORVU4Hr1G063420* encoding β-ketoacyl-CoA synthase (KAS1) and is allelic to *cer-zh.54*. The amounts of intracuticular waxes were strongly decreased in *cer-ye.267*. The cuticular water loss and permeability of *cer-za.227* were similar to wild-type (WT), but were increased in *cer-ye.267*. Removal of epicuticular waxes revealed that intracuticular, but not epicuticular waxes are required to regulate cuticular transpiration.
- The differential decrease in intracuticular waxes between *cer-za.227* and *cer-ye.267*, and the removal of epicuticular waxes indicate that the cuticular barrier function mostly depends on the presence of intracuticular waxes.

Introduction

The cuticle forms the apoplastic barrier between aerial plant organs and the environment (Nawrath, 2006). It represents the first boundary of the plant and enhances the resistance against water loss, UV light, and pathogen attack and establishes a self-cleaning surface (Koch *et al.*, 2008; Lee & Suh, 2015). The protective capabilities of the cuticle depend on the presence of waxes which are embedded in (intracuticular) or overlaid onto (epicuticular) the cutin matrix forming crystalline structures on the surface (Nawrath, 2006; Koch *et al.*, 2008). Cuticular waxes are a mixture of very-long-chain fatty acid (VLCFA)-derivatives obtained from plastidial fatty acid *de novo* biosynthesis, often containing triterpenoids or sterols (Bernard & Joubès, 2013; Batsale *et al.*, 2021). Acyl-CoAs are transported to the endoplasmic reticulum (ER), where they are elongated by the fatty acid elongation (FAE) complex to give rise to VLCFA-CoAs. The FAE complex consists of four enzymes, β-ketoacyl-CoA synthase (KCS), β-ketoacyl-CoA reductase, β-hydroxyacyl-CoA dehydratase, and enoyl-CoA reductase. Different KCS subunits determine the chain length specificity of the FAE complexes (Millar & Kunst, 1997). VLCFA-CoAs are hydrolyzed by thioesterases to release free fatty acids, or they are reduced to aldehydes and decarbonylated to yield odd chain length alkanes, secondary alcohols, and ketones. In a third pathway, VLCFA is reduced to even chain length primary alcohols, which involves the formation of an aldehyde intermediate. Finally, waxes are exported from the plasma membrane, through the cell wall, and incorporated into the cuticle (Kunst & Samuels, 2009).

The identification of wax (*eceriferum*, *cer*) mutants was instrumental for the elucidation of the wax biosynthetic pathway. In *Arabidopsis thaliana*, 21 *cer* mutants were isolated with altered

*These authors contributed equally to this work.

© 2023 The Authors
New Phytologist © 2023 New Phytologist Foundation
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.

New Phytologist

wax loads or compositions on the stems (Koornneef *et al.*, 1989). Nevertheless, knowledge about the function of the individual wax components and fractions is still incomplete. The accumulation of cuticular waxes under drought for UV irradiation is a common adaptive mechanism for plants (Long *et al.*, 2003; Patwari *et al.*, 2019); therefore, identifying cuticular wax-related genes in crops broadens the scope for breeding strategies to enhance resilience against water loss during drought and heat stress (Mansour *et al.*, 2018). Barley (*Hordeum vulgare*) is the fourth most cultivated cereal world-wide (www.statistica.com), and an excellent genetic model thanks to the availability of genetic resources, including cultivars, landraces, wild relatives, and mutants (Mascher *et al.*, 2017). Numerous mutants have been isolated, mostly through UV, radioactive, or chemical mutagenesis (Lundqvist & Lundqvist, 1988; Lundqvist, 2014). A total of 1580 *cer* mutants of barley were isolated and assigned to 79 loci (Lundqvist & Lundqvist, 1988). The *cer-za* mutation with 78 alleles represents the fourth largest group behind *cer-c*, *cer-q*, and *cer-u* (Lundqvist & Lundqvist, 1988). Less frequent mutations are *cer-zh* (11 alleles) and *cer-ye* (5 alleles). Few *cer* genes in barley have been identified. The *cer-c*, *cer-q*, and *cer-u* loci are a cluster of genes on chromosome 2HS, coding for, respectively, a β-diketone synthase, a lipid/carboxyl transferase, and a P450 protein, and are crucial for the biosynthesis of β-diketones, the major component of the wax bloom on barley spikes and leaf sheaths (Schneider *et al.*, 2016). *CER-ZH* was identified as *HvKCS1*, required for fatty acid elongation (Li *et al.*, 2018). Recently, the HvSHINE1 transcription factor involved in the regulation of wax biosynthesis was found to be encoded by *CER-X* (McAllister *et al.*, 2022), while *CER-G* and *CER-S* were characterized as encoding two additional transcription factors, *HvYDA1* and *HvBRX-Solo*, regulating both cuticular integrity and cell patterning (Liu *et al.*, 2022).

To study the relevance of waxes as a cuticular barrier, we selected two *cer* mutants with a strong reduction in total leaf wax load, *cer-za.227* and *cer-ye.267* (Larsson & Svenningsson, 1986; Rostás *et al.*, 2008). Bulked segregant RNA sequencing (BSR-Seq) approaches revealed that *cer-za.227* contains a mutation in an acyl-CoA reductase (*HvFAR1*), while *cer-ye.267* is allelic *to cer-zh.54* which is mutated in *HvKCS1*. The analysis of different fractions of cuticular waxes revealed that *cer-ye.267*, but not *cer-za.227*, shows a significantly increased cuticular permeability, which can be attributed to alterations in the amounts of intracuticular waxes.

Materials and Methods

Plant material and cultivation conditions

Barley (*Hordeum vulgare* L.) seeds were obtained from the NordGen seed bank (www.nordgen.org; Alnarp, Sweden; Supporting Information Table S1). The *cer-za.227* mutant was isolated after mutagenesis of the cultivar Foma (Lundqvist & Lundqvist, 1988; Bregitzer *et al.*, 2013). Gene isolation by mapping in the original mutant is hampered by the presence of background mutations. This issue was addressed by the generation of near-isogenic lines (NILs) via repeated backcrossing of the original mutants to the cultivar Bowman (Druka *et al.*, 2011). A seven-times backcrossed *cer-za.227* line (BW157) was obtained from the Bowman NIL collection. Three further alleles in their original backgrounds, *cer-za.232* (Foma), *cer-za.318* (Foma), and *cer-za.173* (Bonus), were obtained (Bregitzer *et al.*, 2013; Table S1). For the *cer-ye.267* mutant (originally in cultivar Foma), we used the *cer-ye.267* NIL (BW136) which had been backcrossed to Bowman seven times (Druka *et al.*, 2011). Three allelic lines in the original backgrounds, *cer-ye.582* (Foma), *cer-ye.792* (Bonus), and *cer-ye.1395* (Bonus), were obtained (Table S1). Barley transformation was carried on in the early flowering line Golden Promise (GP-fast, Ppd-H1; Gol *et al.*, 2021).

The seeds were germinated on moistened filter paper in darkness for 3 d, after which the seedlings were transferred to pots with soil (Einheitserde ED73; Patzer, Sinntal-Altengronau, Germany) and vermiculite (3 : 1). The recessive mutants *cer-ye.267* (BW136) and *cer-zh.54* (Bonus) were crossed according to published protocols (Harwood, 2019).

Seeds of the *Arabidopsis thaliana* (L. Heynh.) *cer4-3* mutant were obtained from Gillian Dean (University of Vancouver, Canada; Rowland *et al.*, 2006). *Arabidopsis* seeds were surface-sterilized with chlorine gas and cultivated on Murashige and Skoog medium with 1% agarose and 1% sucrose for 2 wk before transfer to soil. *Nicotiana benthamiana* (Domin) seeds were directly sown into soil. All plants were cultivated in growth chambers at 21°C with 55% relative humidity and 16 h of light (150 μmol m^{-2} s^{-1}).

Cuticular wax analysis

Waxes were analyzed as described previously (Haas & Rentschler, 1984; Zeisler & Schreiber, 2016). Briefly, waxes were extracted from the second leaf of 14-d-old barley plants which were cut at 2 cm from the point of emergence (Richardson *et al.*, 2007), or from stems of 4-wk-old *Arabidopsis* plants. The plant organs were dipped into chloroform for 10 s, and the internal standard was added (10 μg tetracosane). The waxes were silylated and measured using gas chromatography (GC; Zeisler & Schreiber, 2016). For the selective removal of the epicuticular waxes, barley leaf sections 2 cm in length were treated with collodion (Haas & Rentschler, 1984). The epicuticular waxes were extracted from the strips, and the remaining (intracuticular) waxes were extracted from the stripped leaf. In addition, wax esters were measured by highly sensitive mass spectrometry (MS; Patwari *et al.*, 2019). Briefly, total waxes in chloroform were supplemented with internal standard (18:0ol-17:0), purified using a solid phase extraction and measured by direct infusion MS/MS (Patwari *et al.*, 2019).

Analysis of cuticular permeability and foliar water loss

The entire length of the second leaf of three-leaf stage barley plants was pressure sprayed for 3 s with a solution of 50 μM Metribuzin in 0.1% Brij-L4 (Merck, Darmstadt, Germany). The quantum yield of chlorophyll fluorescence of photosystem II

© 2023 The Authors
New Phytologist © 2023 New Phytologist Foundation

14698137, 2023, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19063, Wiley Online Library on [25/03/2024]. 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

New Phytologist

Research 1905

14698137, 2023, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19063, Wiley Online Library on [25/03/2024]. 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

(quantum yield of PSII), a measure for the penetration of the photosynthetic inhibitor into the leaf, was monitored with a pulse-amplitude modulation fluorometer (Walz, Effeltrich, Germany). Leaves were dark-adapted, and the quantum yield of PSII was measured during the treatment in 5-min intervals for 3 h by applying PSII-saturating light flashes.

Foliar water loss was determined as published (Grünhofer *et al.*, 2022). Total water permeance of the leaf (predominantly representing transpiration through the stomata) was measured with an SC-1 Leaf Porometer (Decagon Devices Inc., Pulman, WA, USA). Residual water permeance (representing transpiration through the cuticle) of the detached leaves was gravimetrically measured in constant intervals throughout an incubation at 2% humidity and 25°C. The leaves were finally dried to determine their dry weights.

Determination of contact angles and scanning electron microscopy

Leaf-water drop contact angles were determined with a Drop Shape Analyzer (DSA25; KRÜSS, Hamburg, Germany). For imaging epicuticular wax crystals, leaf sections of 14-d-old barley plants were mounted on carbon-coated aluminum stubs and dried in a desiccator. The dry samples were sputter-coated with gold (5 nm; Automatic Sputter Coater; Ingenieurbüro Peter Liebscher, Wetzlar, Germany). Images were acquired using a scanning electron microscope (Phenom-World, Eindhoven, the Netherlands; or TESCAN GmbH, Dortmund, Germany).

Mutagenesis of the *CER-ZA* locus

Two 20-bp sequences containing protospacer adjacent motifs (PAMs) on exons 5 and 8 of *CER-ZA* were targeted by the design of two single-guide RNAs (sgRNA1, 5′-GGAAACCCCTCG AAAACAGA; sgRNA2, 5′-GAACTTGCAGAGGCTAAGAG) using Cas-Designer (Park *et al.*, 2015). The sgRNA sequences were assembled from two oligonucleotides each (sgRNA1, bn4388, bn4389; sgRNA2, bn4390, bn4391; Table S2) and cloned into pMG625 and pMG627 and then introduced into pMP217 via Golden Gate cloning using *Bsa*I/*Bbs*I. The recipient plasmid pMP217 (derivative of pMGE599) harbors hygromycin B and Cas9 endonuclease genes, both driven by the maize ubiquitin promoter. The construct pMP217-sgRNA1-sgRNA2 was transferred into barley embryos (Golden Promise) as recently described (Amanda *et al.*, 2022). Transgenic calli were selected on hygromycin B and the plantlets transferred into soil. The two PAM sites were screened for mutations by a PCR amplification of their genomic DNA (sgRNA1, bn4748, bn4749; sgRNA2, bn4752, bn4753; Table S2). The PCR products were ligated into pJET1.2 (Thermo Fisher Scientific, Dreieich, Germany). Eight *E. coli* clones per PAM site and per plant were sequenced.

Semiquantitative RT-PCR and subcellular localization

A strip of leaf epidermis was carefully removed from 10-d-old Bowman seedlings using forceps (Weyers & Travis, 1981). Total RNA was isolated from the epidermal strip, the stripped leaf, and the roots (NucleoSpin RNA Plant; Macherey-Nagel, Düren, Germany) and used for cDNA synthesis (RevertAid First Strand cDNA Synthesis Kit; Thermo Fisher Scientific). RT-PCR was performed for *CER-ZA* (*HORVU5Hr1G089230*, bn4254, bn4255), *CER-YE* (*HORVU4Hr1G063420*, bn4275, bn4276) and the housekeeping gene *ACT2* (*AY145451.1*, bn3698, bn3699). Primers were designed using Primer3 (Table S2; Untergasser *et al.*, 2012).

The coding sequence of CER-ZA was PCR-amplified from the cDNA of Bowman leaves (bn4045, bn4046; Table S2) and ligated into the *Bam*HI/*Sal*I sites of pLH9000-GFP-WSD1 carrying the *35S* promoter (Patwari *et al.*, 2019). The pLH9000-GFP-CER-ZA construct, the ER marker (pCB-DsRed-HDEL; Patwari *et al.*, 2019), and pMP19 (Voinnet *et al.*, 2000) were transferred into *Agrobacterium* GV3101. The *Agrobacterium* cells were infiltrated into the abaxial side of *N. benthamiana* leaves and observed with a spinning disc confocal microscope (IX73; Olympus Optical, Tokyo, Japan; DSU filter turret Lambda 10-3, Sutter Instrument, Novato, CA, USA; Linear laser system 400 Series, Oxford Instruments, Oxford, UK). DsRed fluorescence was excited at 561 nm, and its emission was observed with a center wavelength of 607.36 nm. GFP fluorescence was excited at 488 nm, and its emission was filtered with a center wavelength of 525.3 nm. A depth of 290 nm was covered in 30 planes.

Expression of CER-ZA in *Saccharomyces cerevisiae* and *Arabidopsis*

The coding sequence of CER-ZA was PCR-amplified from Bowman cDNA (bn4277, bn4171, Table S2). The amplicon was ligated into pJet1.2, released with *Eco*RI/*Sal*I, and ligated into the yeast vector pDR196 (Rentsch *et al.*, 1995). The pDR196-CER-ZA construct was introduced into *Saccharomyces cerevisiae* BY4741 (Euroscarf, Oberursel, Germany). After growth at 28°C for 72 h, the cells were harvested by centrifugation (3500 ***g***, 10 min). Lipids were extracted with chloroform/methanol (2 : 1) in the presence of the internal standard (1-heptadecanol, 5 µg). The organic phase was harvested after centrifugation, the solvent was evaporated, and the fatty acids were transmethylated with 1 M methanolic HCl (80°C, 1 h). Lipids were extracted with hexane/0.9% NaCl (1 : 1) and separated by thin-layer chromatography in hexane/diethyl ether/acetic acid (55 : 45 : 0.5) on silica plates. After staining with 8-anilinonaphthalene-1-sulfonic acid, the free alcohols were isolated from the silica with hexane/0.9% NaCl (1 : 1). After centrifugation, the organic phase was collected and the solvent evaporated. Alcohols were dissolved in chloroform and silylated with pyridine and N,O-bis-(trimethylsilyl)-trifluoroacetamide (BSTFA; 30 min, 70°C). The silylated alcohols were analyzed using gas chromatography–mass spectrometry (GC–MS) (helium flow, 1.05 ml min^{-1}; split/splitless injection; inlet temperature, 250°C; septum flow, 3 ml min^{-1}; initial oven temperature, 120°C for 2 min, ramped at 10°C min^{-1} to 180°C, held for 2 min, increased to 310°C, held for 10 min, decrease of 5°C min^{-1} to 120°C).

© 2023 The Authors
New Phytologist © 2023 New Phytologist Foundation

1906 Research

New Phytologist

The promoter and terminator regions of *CER4* were PCR-amplified from *Arabidopsis* Col-0 genomic DNA (Rowland *et al.*, 2006). The *CER-ZA* sequence was amplified from Bowman leaf cDNA (bn3955, bn3956, Table S2). The binary vector pBin-35S-eGFP-DsRed (Patwari *et al.*, 2019) was digested with *SmaI/MluI* to remove the entire eGFP cassette and ligated with the three *BsaI*-digested fragments (*CER4* promoter, *CER-ZA* cDNA, *CER4* terminator) using Golden Gate cloning. The construct pBIN-AtCER4pro::CER-ZA-AtCER4term-DsRed was transferred into *Agrobacterium* GV3101 and transferred into the *Arabidopsis cer4-3* mutant (Zhang *et al.*, 2006). Four independent homozygous transgenic lines were selected in the T_3 generation based on red seed fluorescence.

Results

Alterations in water repellence and surface structure in *cer-za.227* and *cer-ye.267*

After spraying with water, water droplets were not retained on the leaf surfaces of wild-type (WT) barley cultivar Bowman, but accumulated on *cer-za.227* and *cer-ye.267* (Fig. 1a). The contact angles of the water droplets on WT Bowman leaves are very large due to the high hydrophobicity of the wax layer, with values > 150°. The contact angles of water droplets on *cer-za.227* (122.2 ± 2.6°) and *cer-ye.267* (119.9 ± 7.4°) were decreased, indicating that the wettability of the mutant leaves was increased.

Scanning electron microscopy revealed that Bowman leaves are densely covered with platelet-shaped wax crystals. The number of crystals was reduced, and the crystals were unevenly distributed on the surfaces of *cer-za.227* and *cer-ye.267* (Fig. 1b). The morphology of the crystals of *cer-za.227* was altered, with fewer platelet-shaped structures accompanying an increase in tubular forms. The structure of wax crystals in *cer-ye.267* was not altered, although the density of wax crystals appeared lower (Fig. 1). Because *cer-za.227* and *cer-ye.267* had similar wax amounts (Fig. 2a), the auto-assembly of waxes into crystals might be affected.

Changes in wax load and composition in *cer-za.227* and *cer-ye.267*

The total wax load and composition were determined in the leaves of Bowman and the *cer-za.227* (BW157) and *cer-ye.267* (BW136) mutants (Fig. 2a). The epicuticular (surface) and the intracuticular (embedded into the cutin matrix) waxes were separately extracted and measured. In Bowman, the total wax load was 11.5 μg cm^{-2}, 4/5 of which were epicuticular waxes. Total waxes declined to 3.2 μg cm^{-2} in *cer-za.227* (28%, all percentages compared with Bowman), and 3.4 μg cm^{-2} in *cer-ye.267* (30%). In *cer-za.227*, the epicuticular waxes were particularly decreased (1.8 μg cm^{-2}, 19%), while the intracuticular waxes were less affected (1.4 μg cm^{-2}, 60%) (Fig. 2a). In *cer-ye.267*, the epicuticular waxes were less affected (2.3, 26%), and the intracuticular waxes were strongly decreased (1.1 μg cm^{-2}, 46%). The decrease in intracuticular waxes was more pronounced in *cer-ye.267* compared with *cer-za.227* (Fig. 2a).

Most wax classes were decreased in the two mutants (Fig. 2b). The cuticular waxes of barley leaves are dominated by primary alcohols, mostly 1-hexacosanol (26:0ol) (Nødskov Giese, 1975). Alcohols (26:0ol) were strongly reduced in *cer-za.227* and *cer-ye.267* (Fig. S1). The alkane content (mostly C33) was increased in *cer-za.227*, but it was decreased in *cer-ye.267*. The contents of aldehydes, particularly 26:0al and 28:0al, and esters (C40, C42, C44, and C46) were reduced in *cer-za.227* and *cer-ye.267*, except the C40 ester which was increased in *cer-ye.267* (Fig. S1). Highly sensitive wax ester measurements by direct infusion MS/MS analysis confirmed that the amounts of wax esters were decreased in *cer-za.227*, particularly those containing 26:0ol (Fig. S2). The total wax loads were also decreased by 50–70% in the three alleles of *cer-za* (*cer-za.173*, *cer-za.232*, and *cer-za.318*) and *cer-ye* (*cer-ye.582*, *cer-*

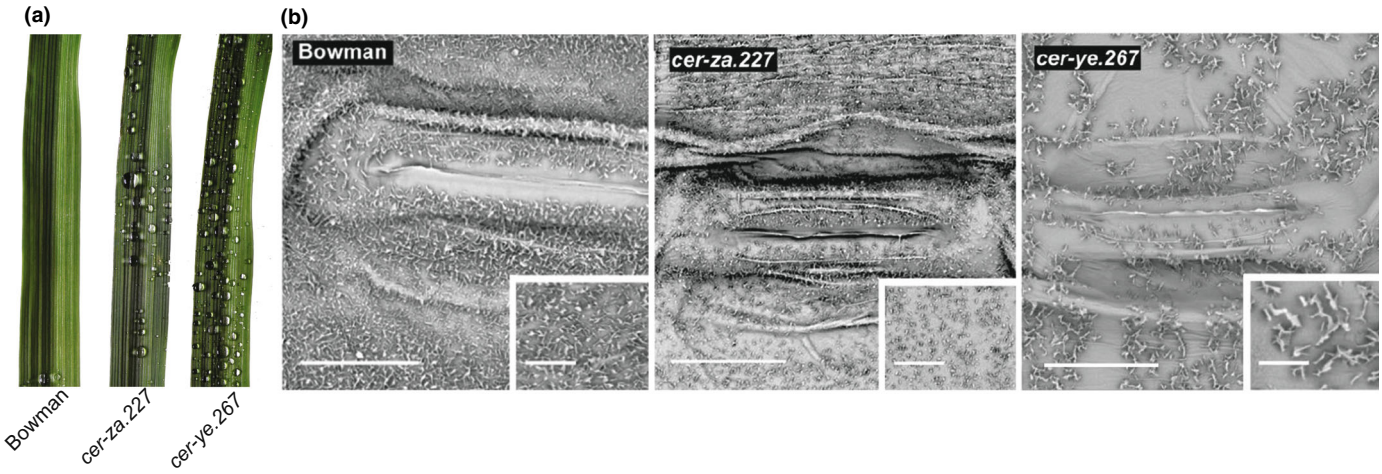


Fig. 1 Wettability and surface structures of barley cultivars Bowman, *cer-za.227* and *cer-ye.267*. (a) Water droplets accumulate on the surfaces of *cer-za.227* and *cer-ye.267* leaves but roll off the leaf surface of the Bowman cultivar. (b) Scanning electron microscopy of leaf surfaces reveals a dense coverage with plate-shaped crystalline waxes on Bowman leaves, but a low number of wax crystals on *cer-za.227* and *cer-ye.267* leaves. Bars: (b) 10 μm; (insets) 8 μm.

© 2023 The Authors
New Phytologist © 2023 New Phytologist Foundation

14698137, 2023, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19063, Wiley Online Library on [25/03/2024]. 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

New Phytologist

Research 1907

14698137, 2023, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19063, Wiley Online Library on [25/03/2024]. 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

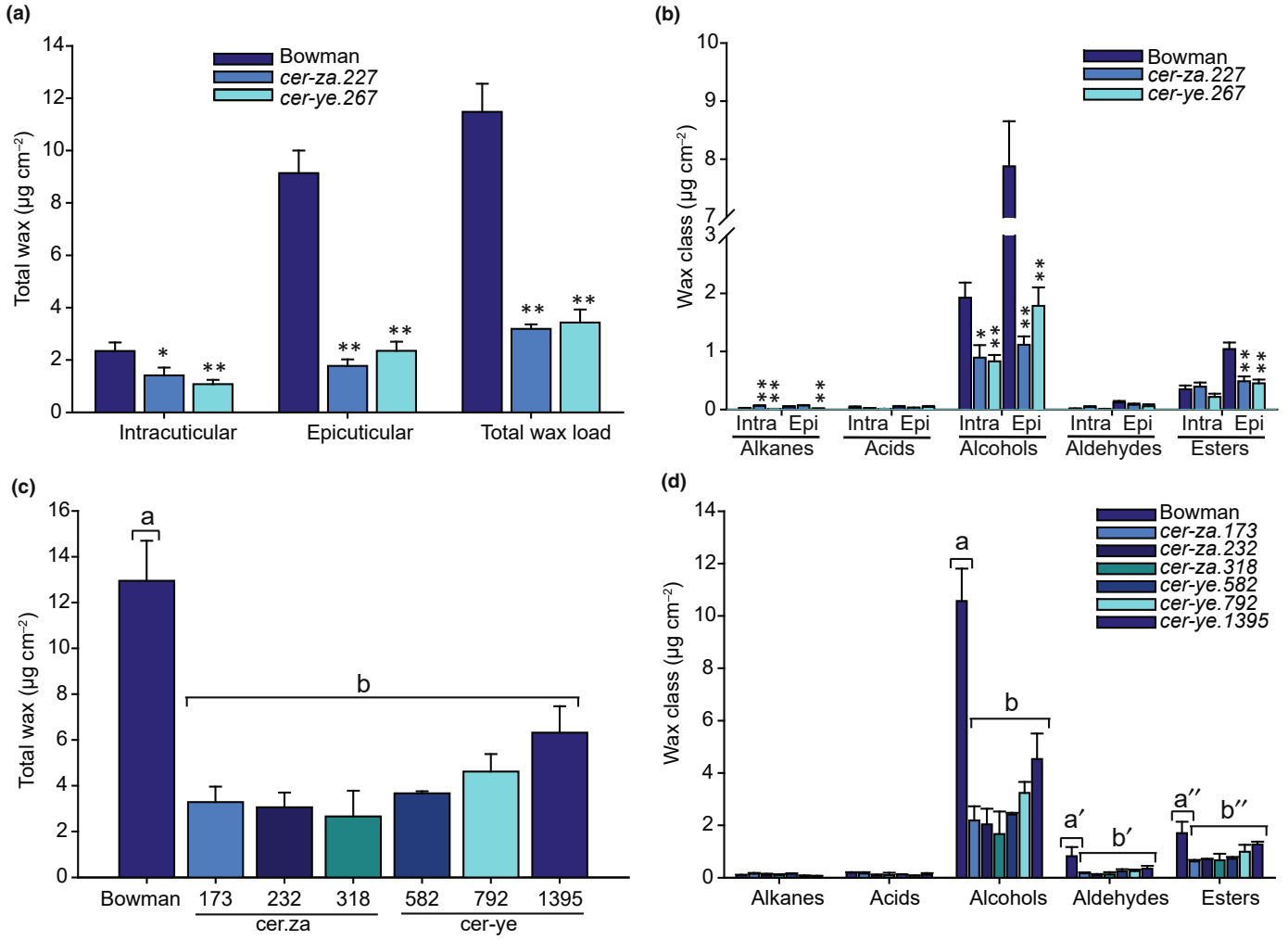


Fig. 2 Cuticular wax load and composition of *cer-za* and *cer-ye* mutant alleles. Waxes were extracted from the leaves, and the wax components were quantified using gas chromatography (GC-FID). (a) Total waxes of the epicuticular (isolated) and intracuticular (remaining) wax fractions, extracted from the leaves of barley cultivars Bowman, *cer-za.227*, and *cer-ye.267*. (b) Different wax classes in the epicuticular and intracuticular wax fractions. (d) Wax classes from different *cer-za* and *cer-ye* alleles. (c) Total cuticular waxes from *cer-za* and *cer-ye* alleles. The wax loads and composition of the cultivar Bowman were very similar to those of Foma (*cer-za.232*, *cer-za.318*, *cer-ye.267*, and *cer-ye.582*) and Bonus (*cer-za.173*, *cer-ye.792*, and *cer-ye.1395*). Mean ± SD; $n = 3$; (a, b): *t*-test, significant differences to Bowman are indicated (**, $P < 0.01$; *, $P < 0.05$) (c, d) ANOVA, different letters indicate significant differences with $P < 0.05$. In (d), a, b, a′, b′ and a″, b″ refer to the comparison of the amounts of alcohols, aldehydes and esters from the different lines, respectively.

ye.792, and *cer-ye.1395*) compared with Bowman (Fig. S1). The total wax load and composition of the different cultivars (Bowman, Bonus, Foma, and Golden Promise) were highly similar. The changes in the wax load and composition of the three additional *cer-za* and *cer-ye* alleles tested were very similar to *cer-za.227 and cer-ye.267*, respectively (Fig. 2c,d).

The *cer-za.227* mutation is located to *HORVU5Hr1G089230*

BSR-Seq represents a powerful method to identify mutant genes in small populations (Dong *et al.*, 2018; Wu *et al.*, 2018). The *CER-ZA* locus was mapped in a F_2 population from a cross of *cer-za.227* with Bowman (Methods S1). Using BSR-Seq, we identified an elevated mutant allele SNP frequency in the region of 580–660 Mb on chromosome 5, in agreement with previous results (Druka *et al.*, 2011; Bregitzer *et al.*, 2013). This region (Fig. S3) carries 49 genes with high-confidence SNPs compared with the Morex sequence (Mascher *et al.*, 2017). The number of candidate genes was narrowed down by comparing their expression levels between the Bowman control and the *cer-za.227* bulks, removing genes with similarly abundant RNA reads (Table S3). The analysis of the remainder of the genes for their predicted functions allowed us to identify *HORVU5Hr1G089230* as a candidate for CER-ZA. Computation of the Δ(SNP-index; Takagi *et al.*, 2013) and the *G′* values (Magwene *et al.*, 2011) of the most likely causal SNPs positioned *HORVU5Hr1G089230* among the three most likely genes for CER-ZA. Furthermore, the transcript abundance of *HORVU5Hr1G089230* was clearly downregulated in the *cer-za.227* bulk compared with Bowman ($\log_2$ fold change: −2.2; false discovery rate (FDR) 1.704×10^{-20}).

© 2023 The Authors
New Phytologist © 2023 New Phytologist Foundation

New Phytologist

HORVU5Hr1G089230 encodes a protein of 498 amino acids with a calculated mass of 55.95 kDa. According to AHRD prediction, it encodes a FAR likely involved in wax biosynthesis. 'NAD_-binding_4' and 'sterile' domains were predicted by Pfam, and a Rossmann fold within the NAD_binding_4 domain (amino acids 11–57) with NADP as potential cofactor was identified using Cofactory. The sterile domain (also known as FAR_C) is shared by members of the FAR family. A transmembrane domain was predicted at the C terminus by TMHMM, in accordance with DeepLoc, which indicated that *HORVU5Hr1G089230* is membrane-bound and ER-localized. Therefore, *CER-ZA* is likely *HORVU5Hr1G089230*, which is predicted to encode a membrane-bound, ER-localized protein with FAR activity. Protein BLAST searches in barley revealed that HORVU5Hr1G089230 belongs to a family of 22 FAR proteins (Table S4). Sequences of FAR proteins from different plants were used to construct a phylogenetic tree (Fig. S4). HORVU5Hr1G089230 clusters with FAR1 from the grasses *Aegilops* and *Brachypodium*. The FAR sequences form four clades with a split of predicted ER-localized FARs from monocots (clade I) or dicots (clade II), and a separation of plastid-localized FARs into clades III (dicots) and IV (monocots), the latter in agreement with previous results (Zhang *et al.*, 2021).

The BSR-Seq approach revealed 10 SNPs in *HORVU5Hr1G089230* of *cer-za.227* compared with Morex_V2 (Fig. 3a; Table S5). Because the sequence around the *cer-za.227* mutation is presumably derived from Foma, most of the SNPs (1–4 and 6–10) are due to polymorphisms between Foma and Morex. Only SNP5 was specific for *cer-za.227*. SNP5 caused a guanine deletion at position 584526708 of chromosome 5, associated with a frameshift in exon 4 which causes the exchange of Asn153Ile, a premature stop codon, and the truncation of the polypeptide after 153 amino acids (Fig. 3a,b). Comparison of the RNA-Seq sequence of the other alleles with their respective WT cultivars revealed the presence of further SNPs. The *cer-za.318* allele displayed the same mutation as *cer-za.227*; therefore, these two SNPs might be derived from the same mutagenesis event. The *cer-za.232* mutant carried a thymine/adenine exchange at position 584526341 at the exon5/intron5 splicing site after Lys229, leading to the translation of the intron sequence and a premature stop codon after 253 amino acids. Finally, *cer-za.173* contains a cytosine/adenine exchange at position 584525054, resulting in a Glu331 stop mutation in exon 8. The amino acid sequences of HORVU5Hr1G089230 in Bonus and Morex are identical, while the Foma sequence differs by two amino acids (Asn34Asp and Ala627Tyr). The protein structures (modeled with Alphafold) revealed the presence of a Rossmann fold and are highly similar for all cultivars (www.uniprot.org) (Fig. 3b). The HORVU5Hr1G089230 polypeptides of *cer-za.227* and *cer-za.318* are truncated, lacking a large part of the C terminus. Similarly, the protein sequences of *cer-za.232* and *cer-za.173* are truncated, and the structures are strongly affected (Fig. 3c).

Generation of *cer-za* alleles by genome editing

To prove that *HORVU5Hr1G089230* is *CER-ZA*, we mutagenized *HORVU5Hr1G089230* by genome editing. Barley Golden Promise was transformed with constructs harboring the Cas9 endonuclease and two sgRNA sequences targeted to PAM sites in exons 5 and 8. Two independent lines, *cer-za.2001* and *cer-za.2002*, were regenerated. Sequencing of the regions around the PAM site in exon 5 revealed that *cer-za.2001* is heteroallelic, with a thymine insertion in one and a thymine deletion in the other allele, while *cer-za.2002* is homoallelic, with thymine deletions (Fig. 4a). After spraying with water, water droplets rolled off the leaves of Golden Promise, but they adhered to the leaves of the *cer-za* mutants obtained by genome editing, similar to *cer-za.227* (Fig. 4a). The total wax load was decreased from 16.3 $\mu g\,cm^{-2}$ in Golden Promise to 5.7 $\mu g\,cm^{-2}$ in *cer-za.2001* and 6.7 $\mu g\,cm^{-2}$ in *cer-za.2002*, and this was mainly based on the decrease in 26:0ol (Fig. 4a). Therefore, *cer-za.2001* and *cer-za.2002* display changes in wettability, wax load, and composition analogous to *cer-za.227*, demonstrating that *HORVU5Hr1G089230* is *cer-za.227*.

Expression of CER-ZA in yeast and *Arabidopsis cer4-3*

CER-ZA was expressed in *Saccharomyces cerevisiae* to study its enzymatic function. After expression, lipids were extracted from the cells and analyzed using GC–MS. The *CER-ZA*-expressing cells accumulated primary alcohols, hexadecanol (16:0ol), octadecanol (18:0ol), and hexacosanol (26:0ol) (Fig. 4b). CER-ZA is therefore capable of producing primary alcohols, in particular 26:0ol, indicating that it harbors alcohol-forming FAR activity.

The *AtFAR3/CER4* gene product from *Arabidopsis* produces 24:0ol and 26:0ol alcohols (Rowland *et al.*, 2006). As a consequence, waxes on the stems of the *cer4* mutant are deficient in primary alcohols. To study the *in planta* function, *CER-ZA* was introduced into the *Arabidopsis cer4-3* mutant under control of the *AtCER4* promoter to confer its epidermal expression. Four independent homozygous *CER-ZA*-expressing *cer4-3* plants were selected, and the wax load and composition on the stems were measured (Fig. 4c). The wax ester content was increased in all *CER-ZA-cer4-3* lines compared with *cer4-3*, albeit without reaching WT levels. Concomitantly, the primary alcohol content was doubled in *CER-ZA-cer4-3* lines compared with *cer4-3*, but it also did not reach WT levels. The amounts of alkanes and aldehydes in the transgenic lines were slightly increased compared with WT and *cer4-3*, while the secondary alcohol contents were unchanged. The expression of *CER-ZA* in *cer4-3* therefore results in the accumulation of primary alcohols, indicating that it harbors FAR activity *in planta*.

The *cer-ye.267* mutation localizes to *HORVU4Hr1G063420* and is allelic to *cer-zh.54*

The computation of the Δ(SNP-index) (Takagi *et al.*, 2013) and the G' values (Magwene *et al.*, 2011) for the SNPs between the Bowman and *cer-ye.267* bulks led to the identification of regions on chromosomes 2H, 4H, and 5H with G' values above the threshold (FDR = 0.01; Fig. S5). We focused on chromosome 4 where this mutation was previously reported to be located (Druka *et al.*, 2011). The region of 490–530 Mbp contains seven genes

© 2023 The Authors
New Phytologist © 2023 New Phytologist Foundation

14698137, 2023, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19063, Wiley Online Library on [25/03/2024]. 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

New Phytologist

Research 1905

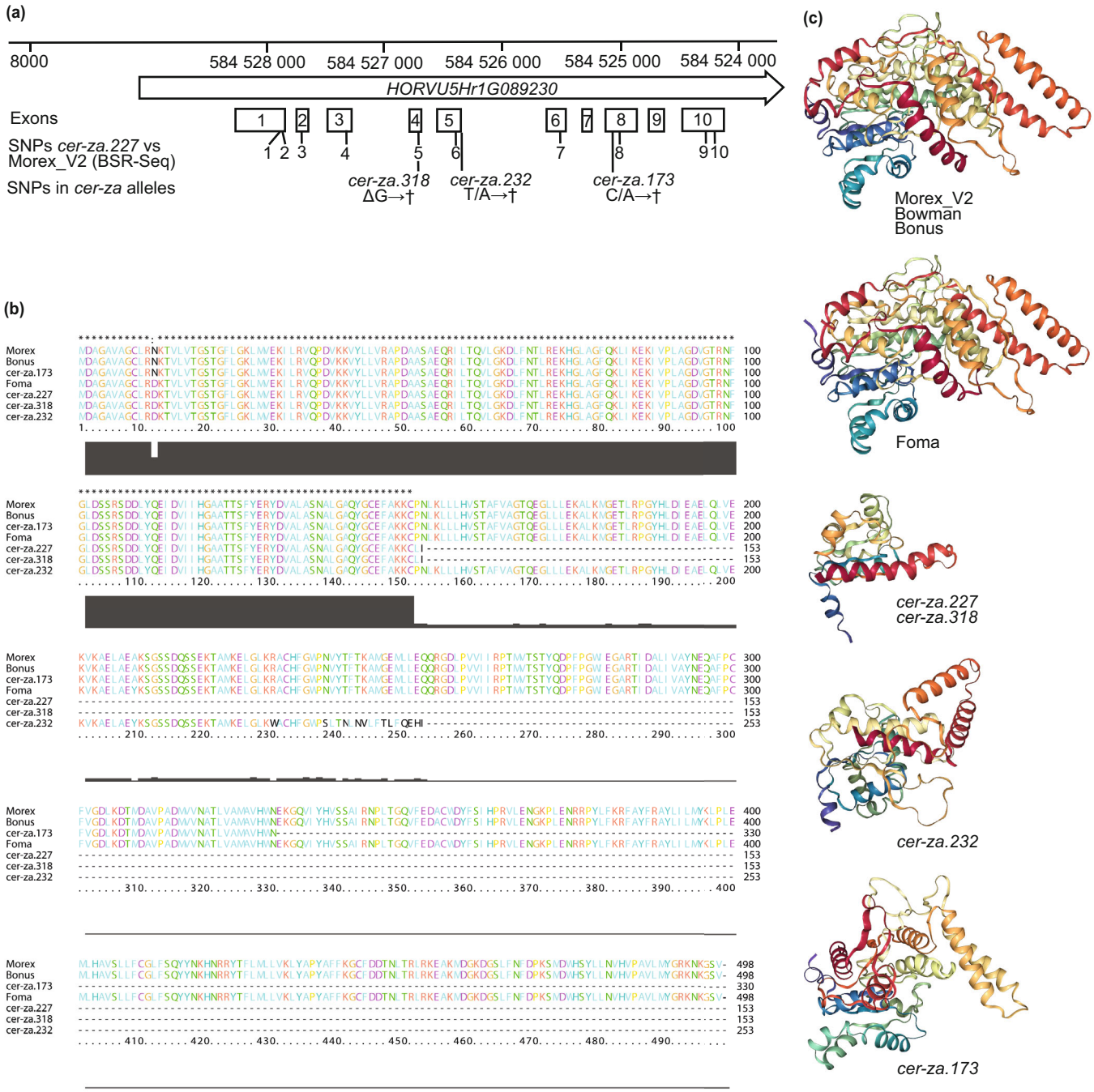


Fig. 3 *cer-za* mutations are associated with SNPs in the gene *HORVU5Hr1G089230*. (a) Representation of the *HORVU5Hr1G089230* locus. The diagram shows its position on chromosome 5H, the exon/intron structure, the SNPs between *cer-za.232* and the reference sequence of Morex_V2, the SNPs between *cer-za.173* and Bonus, and the SNPs between *cer-za.232* and *cer-za.318* and Foma. The deletion of G (ΔG) in *cer-za.227* and *cer-za.318* in exon 8 results in a premature stop codon (†). The T-to-A exchange at the splicing site after exon 5 in *cer-za.232*, and the C-to-A exchange in exon 8 of *cer-za.173*, also cause the introduction of premature stop codons. (b) Protein sequences based on SNPs identified busing BSR-Seq for the different *cer*-za alleles in comparison with Morex, Foma, and Bonus. The amino acid sequences of Morex, Bowman, and Bonus are identical. Foma differs from this sequence by two amino acids. Bars indicate the degree of identity between all sequences at a given position. (c) Protein models calculated by Alphafold. Proteins of *cer-za.227*, *cer-za.318*, *cer-za.232*, and *cer-za.173* are truncated, lacking considerable parts of the C-terminal structural elements.

with high-confidence SNPs (Table S6). Combining this approach with the screening for differentially induced genes indicated that *HORVU4Hr1G063420* is a likely candidate for *cer-ye.267*. This gene covers 2979 bp with an intron-less coding sequence of 1638 bp, encoding a protein of 545 amino acids. The *cer-ye.267* sequence harbors a thymine/adenine exchange compared with Morex_V2 and Foma at position 1940 (Fig. 5a), which results in a Phe495Tyr exchange. RNA-Seq analysis of the additional *cer-ye* alleles revealed that *cer-ye.792* and *cer-ye.1395* carry an identical SNP of cytosine/adenine at position 1068

© 2023 The Authors
New Phytologist © 2023 New Phytologist Foundation

14698137, 2023, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19063, Wiley Online Library on [25/03/2024]. 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

1910 Research

New Phytologist

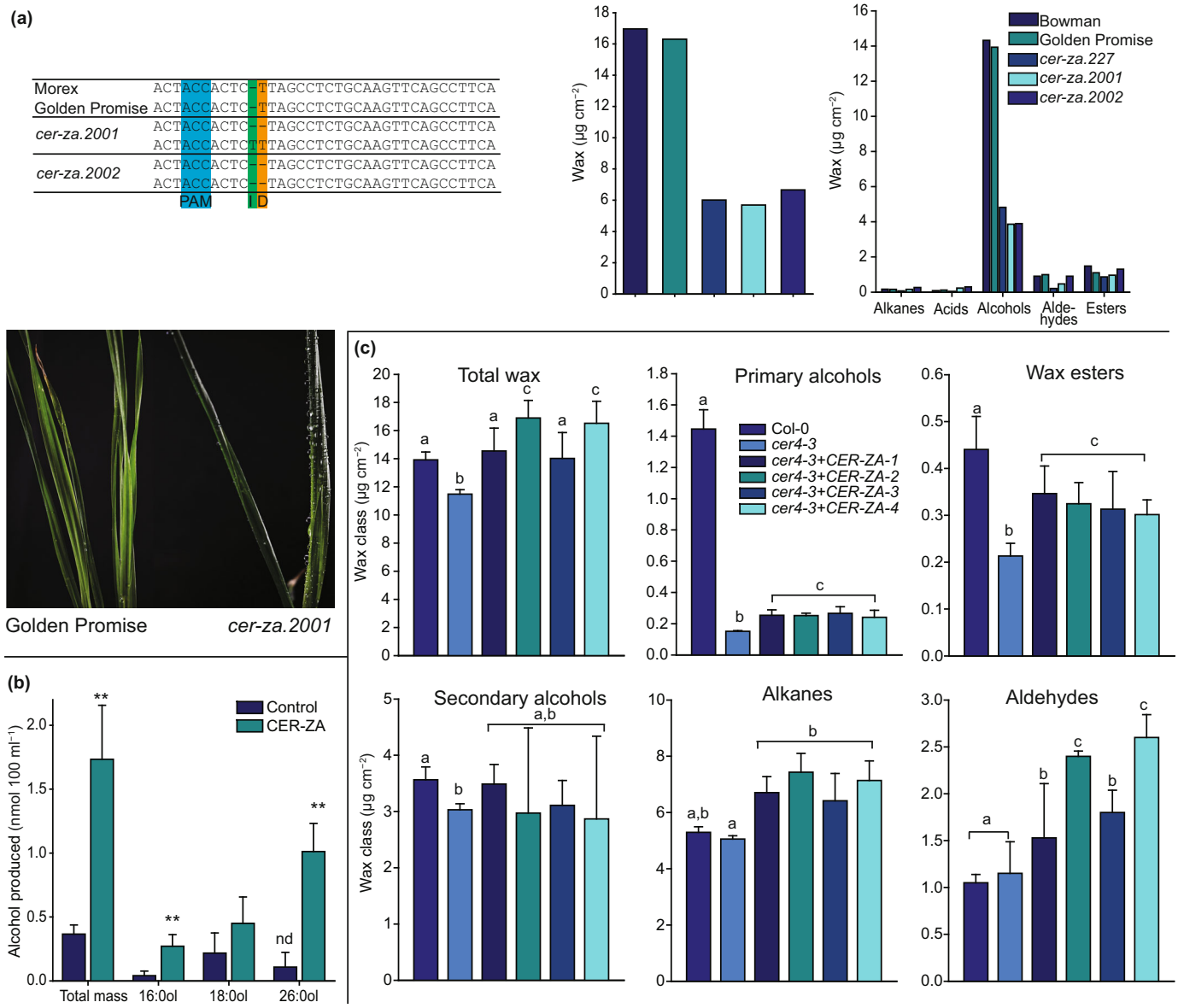


Fig. 4 *CER-ZA* gene product harbors acyl-CoA reductase activity. (a) The *eceriferum* mutant plants *cer-za.2001* and *cer-za.2002* generated by genome editing. Alignment of the DNA sequences around the protospacer adjacent motif (PAM, blue) in exon 5 of *HORVU5Hr1G089230*. Sequences of barley cultivars Morex, Golden Promise, and the alleles of the two mutant lines. The sequence of *cer-za.2001* shows an insertion (I, green) of a T in one allele and a deletion (D, orange) of a T the other one. The sequence of *cer-za.2002* is homoallelic with a deletion of a T. The water droplets adhere to the leaves of the *cer-za* alleles obtained by genome editing (shown here: *cer-za.2001*), in contrast to Golden Promise. Total amount and composition of cuticular waxes on the leaves of *cer-za.227*, *cer-za.2001*, and *cer-za.2002*. Only single measurements of waxes of the genome-edited lines *cer-za.2001* and *cer-za.2002* were possible due to the limitation of leaf material. (b) The *CER-ZA* sequence was expressed in *Saccharomyces cerevisiae*. Lipids were isolated from cells obtained from 100 ml (OD_{600} = 1) and, after conversion into trimethylsilyl ethers, measured using gas chromatography–mass spectrometry (GC–MS) with total ion chromatograms (TIC). Three primary alcohols accumulated in *CER-ZA* -expressing cells: 16:0ol, 18:0ol, and a very-long-chain alcohol identified as 26:0ol. 26:0ol showed the strongest increase. Note that the peak eluting at the retention time of 26:0ol is a contaminant with a different mass spectrum (nd, not detected). Mean ± SD; $n = 3$; *t*-test; significant differences to the control (empty vector); **, $P \leq 0.01$. (c) Wax analysis of the stems of four independent *Arabidopsis thaliana cer4-3* mutant plants expressing *CER-ZA* from barley. Cuticular waxes were isolated from the stems of transformed *cer4-3* plants and measured using GC–MS. The panels show total wax coverage and the distribution of wax classes. Mean ± SD; $n = 3$; ANOVA; different letters indicate significant differences; $P < 0.05$.

compared with both Morex and Bonus, causing a premature stop codon, and a truncated protein (Fig. 5b). No SNP was found in the *HORVU4Hr1G063420* sequence of *cer-ye.582* compared with Morex or Bonus. Fig. 2 shows that *cer-ye.582* has the same wax phenotype as *cer-ye.267*; therefore, *cer-ye.267* might carry a mutation in the regulatory sequences (promoter or intron), but these sequences were not covered within the RNA-Seq approach.

HORVU4Hr1G063420 was previously identified as the gene underlying the *cer-zh.54* mutation (Li *et al.*, 2018). *CER-ZH* encodes HvKCS1, and consequently, the *cer-zh.54* mutant is wax

© 2023 The Authors
New Phytologist © 2023 New Phytologist Foundation

14698137, 2023, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19063, Wiley Online Library on [25/03/2024]. 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

New Phytologist

Research 191

14698137, 2023, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19063, Wiley Online Library on [25/03/2024]. 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

deficient and shows the typical adhesion of water droplets on the leaf. To confirm allelism, *cer-ye.267* and *cer-zh.54* were crossed, and F_1 plants were analyzed after spraying the leaves with water. A dense accumulation of water droplets was observed on the leaves of the F_1 plants, demonstrating that *cer-ye.267* and *cer-zh.54* are allelic (Fig. 5c).

Expression of CER-ZA and CER-YE and the subcellular localization of CER-ZA

CER-ZA/HvFAR1 and *CER-YE/HvKCS1* are likely expressed in the leaf epidermis, in analogy with numerous genes involved in wax biosynthesis (Suh *et al.*, 2005). The expression was studied using semiquantitative RT-PCR after the isolation of an epidermal strip from the leaves of Bowman, compared with the residual stripped leaf and the roots. Bands of 196 and 238 bp corresponding to *CER-ZA* and *CER-YE*, respectively, were observed in the epidermal tissue after electrophoresis, but were absent from the roots or residual leaf material (Fig. 6a). This is consistent with RNA-Seq data (EoRNA database; Milne *et al.*, 2021). *CER-ZA/HvFAR1* (locus BART1_0-p38542) displays low transcript abundance in the roots, apical meristems, spikes, anthers/microspores grains/seeds (including embryo, lemma, palea, lodicule, and rachis), moderate expression in the leaves, and coleoptiles, and high expression in the epidermis. Similarly, *CER-YE/HvKCS1* (BART1_0-p30122) is highly expressed in the leaves and epidermis and shows considerable expression in the root maturation zones, lemma, palea and stat-like/multicellular microspores, and low expression in the roots and seeds. Therefore, *CER-ZA* and *CER-YE* are highly expressed in the leaf epidermis, while *CER-YE* is expressed in additional tissues.

CER-ZA was predicted to be ER-localized using DeepLoc. In addition, CER-ZA carries a C-terminal KNKGSV sequence related to the prototypic dilysine ER-retrieval motif KXKXX, also found in the ER-resident fatty acid desaturase FAD3 of *Arabidopsis* (McCartney *et al.*, 2004). To study its subcellular localization experimentally, the CER-ZA sequence was fused to the C terminus of the green fluorescent protein (GFP) and transiently expressed in *N. benthamiana* leaves. A DsRed-HDEL construct was co-infiltrated as a marker for the ER. The GFP and DsRed fluorescence were observed using confocal fluorescence microscopy. After co-infiltration, the two signals were observed in a net-like structure in epidermal cells (Fig. 6b). The merged signals revealed a clear overlap, confirming that CER-ZA localizes to the ER.

Mutation of *cer-ye.267*, but not *cer-za.227*, affects the function of the cuticular barrier

To study the consequences of the wax changes for the barrier function, the permeation properties of the cuticles of *cer-za.227* and *cer-ye.267* were measured after treatment with the photosynthetic inhibitor Metribuzin (Fig. 7a). The quantum yield of PSII was used as a measure for the penetration rate of Metribuzin across the cuticular barrier. After the application of Metribuzin to the leaves, the quantum yield of PSII decreased within 25 min from 0.7 to *c.* 0.48 in Bowman and *cer-za.227*, indicating that the penetration rates were similar. The quantum yield of PSII drastically declined in *cer-ye.267* reaching values as low as *c.* 0.25, then it slowly recovered to *c.* 0.4, but was always lower than in Bowman and *cer-za.227*. Therefore, Metribuzin penetrated the cuticle of *cer-ye.267* much faster than that of Bowman or *cer-za.227*.

Water loss through the epidermis can be explained by transpiration through stomata and by water evaporation through the cuticle. Under normal conditions with open stomata, transpiration through the stomata is predominant. Therefore, the total water permeance of a leaf is mostly caused by transpiration through the stomata. The total water permeance was determined by measuring the water flow for the adaxial and abaxial leaf sites with a porometer. The total water permeance did not differ between Bowman, *cer-za.227*, and *cer-ye.267*, indicating that transpiration through the stomata was not affected (Fig. 7b). This was in agreement with the finding that stomatal sizes and densities were very similar to Bowman as revealed by microscopic inspection of the leaf surfaces. Next, the water permeance of detached leaves was measured and plotted against the relative water deficit (RWD; Fig. S6). The water permeance was initially high for all lines, but then rapidly declined due to stomatal closure to reach final values of residual permeance P_{min}, which were *c.* 95% lower. The P_{min}, which is a proxy for the cuticular transpiration rate, was calculated from the averages of the measurements after stomatal closure in the range of RWD of 0.15–0.55 (Figs 7c, S6). The P_{min} of *cer-za.227* (2.14 ± 0.56 nm s^{-1}) was slightly but not significantly higher compared with Bowman (1.63 ± 0.18 nm s^{-1}; 95% significance level, *t*-test, $P = 0.086$), whereas P_{min} of *cer-ye.267* (2.35 ± 0.35 nm s^{-1}; $P = 0.0005$) was even higher and significantly increased compared with Bowman. Therefore, the penetration rate of Metribuzin and the cuticular transpiration are not significantly changed in *cer-za.227*, but both are compromised in *cer-ye.267*.

To study the roles of intracuticular and epicuticular waxes for water permeability directly, the water loss of a barley leaf was recorded starting 1 h after detachment. At this time, the stomata were closed, and the residual water loss is a measure of cuticular transpiration. After one additional h, the epicuticular waxes were removed with collodion, and the measurements were continued. The rate of water loss did not change after removal of the epicuticular waxes (Fig. 7d). Therefore, cuticular transpiration rates do not depend on the presence of epicuticular waxes. Fig. 7(e) shows an electron micrograph of the leaf surface with the removal of epicuticular waxes on the left side, leaving behind the layer of intracuticular waxes, while the right side remained untreated. This experiment demonstrates that intracuticular waxes rather than epicuticular waxes are crucial for the cuticular barrier function of barley.

Discussion

The cuticle establishes a barrier between the plant and the environment, providing essential protection against abiotic and biotic stresses. The composition of the cuticular wax layer has been studied in detail, but the contribution of the different wax fractions to the barrier functions remains unclear. The two barley mutants

© 2023 The Authors
New Phytologist © 2023 New Phytologist Foundation

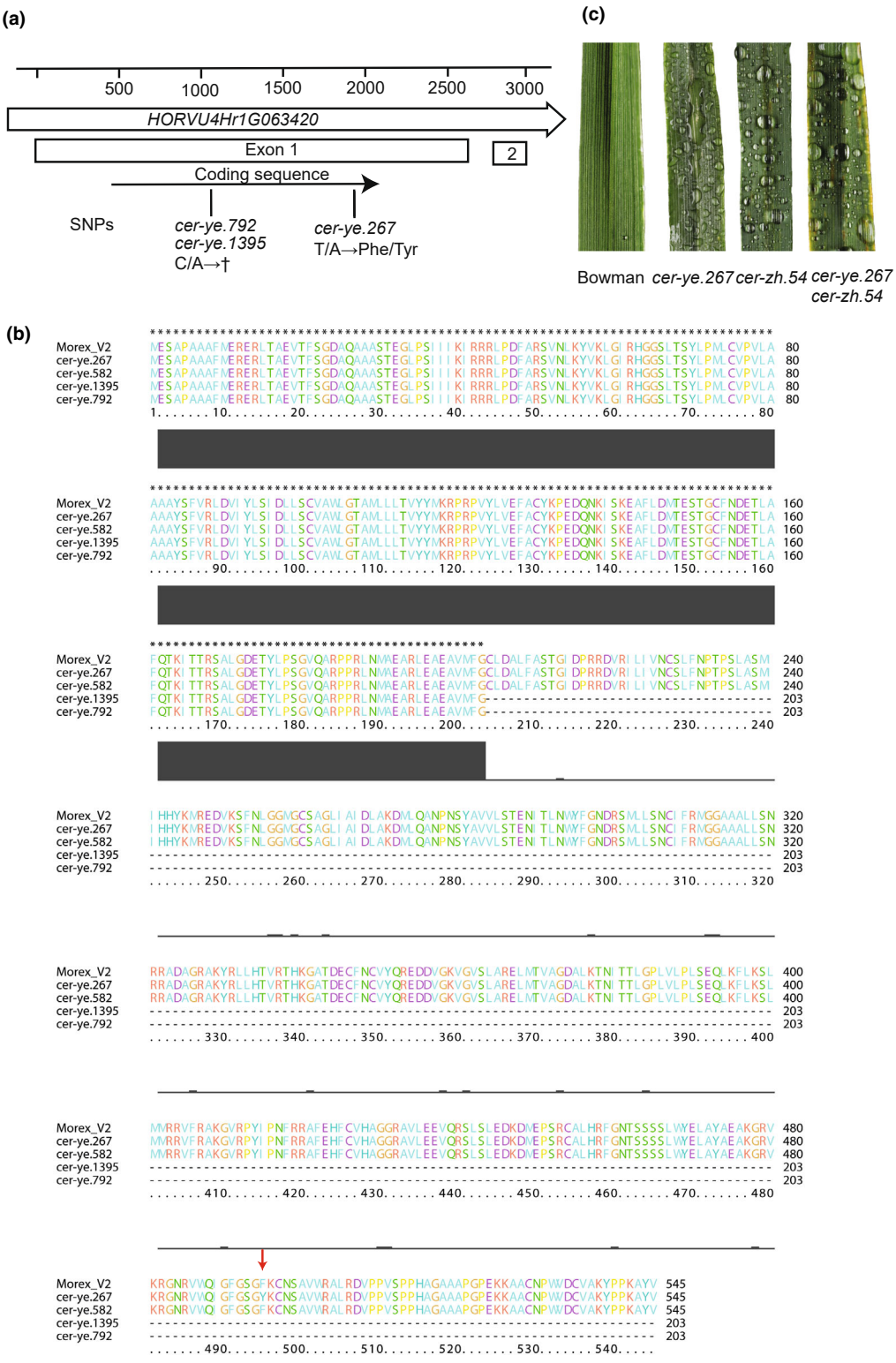


Fig. 5 *cer-ye* mutations are caused by SNPs in the gene *HORVU4Hr1G063420*. (a) Representation of the *HORVU4Hr1G063420* locus. The diagram shows the position on chromosome 4H, the exon/intron structure, and the SNPs between *cer-ye.267*, *cer-ye.792*, and *cer-za.1395* and the reference sequence of Morex_V2. The *cer-ye.267* mutation of T/A causes an amino acid exchange of Phe/Tyr. The *cer-ye.792* and *cer-ye.1395* mutants carry the same mutation of C/A resulting in a premature stop codon (†). (b) Protein sequences based on SNPs identified using BSR-Seq for Morex and the different *cer*-za alleles, *cer-ye.267*, *cer-ye.582*, *cer-ye.1395*, and *cer-ye.792*. The amino acid sequences of barley cultivars Morex, Bowman, and Bonus are identical. Bars indicate the degree of identity between all sequences. The position of the amino acid exchange Phe/Tyr in *cer-ye.267* is depicted with a red arrow. (c) Allelism test for *cer-ye.267* (BW136) and *cer-zh.54* (Bonus). The two mutants *cer-ye.267* and *cer-zh.54* were crossed, and leaves from the plants of the F_1 generation were sprayed with water to determine the wettability of the leaf surface. No water droplets adhere to leaves of Bowman (shown as control) or Bonus.

© 2023 The Authors
New Phytologist © 2023 New Phytologist Foundation

14698137, 2023, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19063, Wiley Online Library on [25/03/2024]. 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

New Phytologist

Research 1915

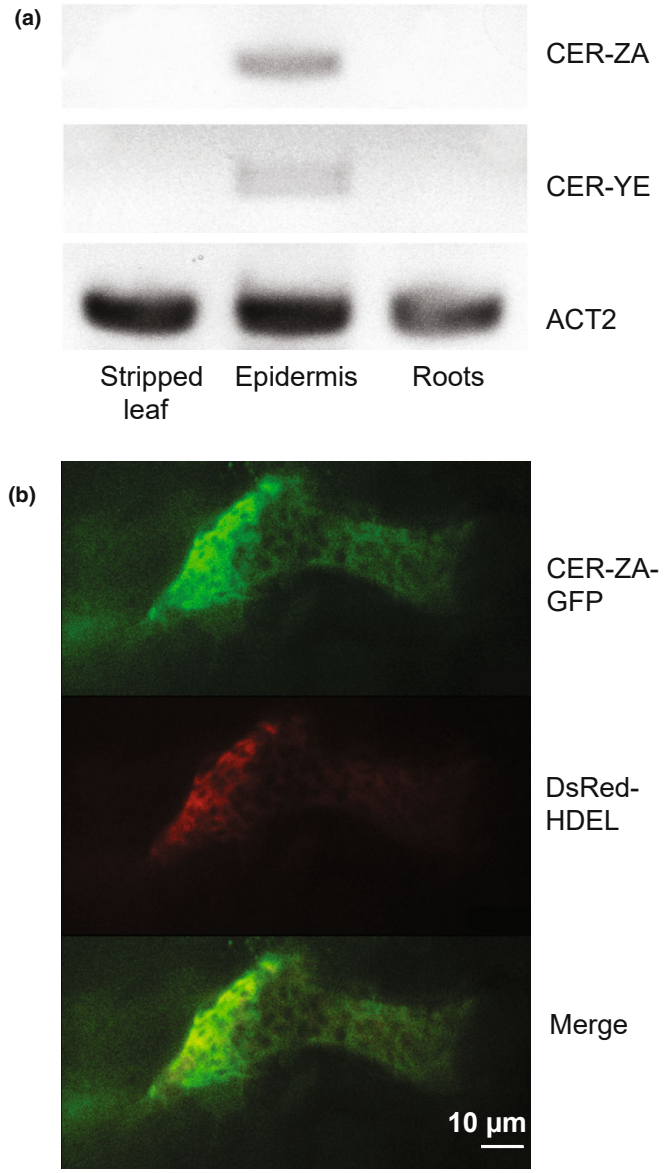


Fig. 6 Expression analysis of barley cultivars *CER-ZA* and *CER-YE* and subcellular localization of CER-ZA. (a) Expression of CER-ZA and CER-YE was compared using a semiquantitative RT-PCR of RNA isolated from epidermal strips of leaves, the stripped leaf, and the root tissue. The *ACT2* gene was used as the control. RT-PCR bands were separated using agarose gel electrophoresis and stained with Midori green. (b) Subcellular localization of CER-ZA. The CER-ZA sequence was fused to the C terminus of GFP, and the GFP-CER-ZA construct expressed in *Nicotiana benthamiana* leaves together with the endplasmic reticulum (ER) marker DsRed-HDEL. Epidermal leaf cells were observed using confocal fluorescence microscopy.

cer-za.227 and *cer-ye.267* reveal strong decreases in total wax load, but unique changes in wax composition and the epicuticular and intracuticular distribution. The *CER-ZA* gene was identified as *HORVU5Hr1G089230*, encoding HvFAR1, while the mutation in *cer-ye.267* is allelic to *cer-zh.54*, which corresponds to *HORVU4Hr1G063420*, encoding HvKCS1. In contrast to *cer-za.227*, the *cer-ye.267* mutation affected the barrier functions, which are therefore presumably established by unique properties in wax distribution rather than a decrease in total wax load.

Identification of the *CER-ZA* and *CER-YE* genes

The gene *HORVU5Hr1G089230* was identified as CER-ZA by BSR-Seq and confirmed by the generation of novel *cer-za* alleles. *CER-ZA* represents the first *FAR* characterized in barley. A total of 22 and eight FAR sequences were annotated in barley and Arabidopsis, respectively (Table S4; Aarts *et al.*, 1997; Doan *et al.*, 2009, 2012). FAR enzymes were characterized in wheat and *Brachypodium* (Wang *et al.*, 2015b, 2018; Chai *et al.*, 2018). Expression in yeast confirmed that CER-ZA harbors alcohol-forming FAR activity similar to AtFAR3/CER4 (Rowland *et al.*, 2006). CER-ZA localizes to the ER, in analogy with the FARs in *Arabidopsis* (Rowland *et al.*, 2006) and wheat (Wang *et al.*, 2015b; Chai *et al.*, 2018). Some *FARs*, including *AtFAR1*, *AtFAR4*, and *AtFAR5*, are expressed in the leaves and roots and therefore also contribute to suberin biosynthesis (Domergue *et al.*, 2010). *CER-ZA* is highly expressed in the leaf epidermis, but not in the roots (Fig. 6a), similar to *AtFAR3/CER4*, suggesting that the two proteins are functional orthologs. The expression of *CER-ZA* in the *Arabidopsis cer4-3* mutant led to increases in the amounts of primary alcohols and wax esters, however without reaching WT levels; therefore, *CER-ZA* partially restores the wax deficiency of *cer4-3*, similar to *TaFAR2*, *TaFAR3*, *TaFAR3*, *TaFAR5*, *BdFAR1*, *BdFAR2*, and *BdFAR3* from wheat and *Brachypodium* (Wang *et al.*, 2015b, 2016, 2018). Possibly, the heterologous FARs from monocots localize to different subcompartments of the ER, resulting in an incomplete complementation of *cer4* (Wang *et al.*, 2015a).

The *cer-ye.267* mutation (BGS 448) was located in *HORVU4Hr1G063420* (Fig. S5). The same locus was previously linked to *cer-zh.54*, and the *cer-ye.267* and *cer-zh.54* mutations were confirmed to be allelic (Fig. 6c). CER-ZH (HvKCS1) is a member of a family of 33 KCS sequences in barley (Li *et al.*, 2018; Tong *et al.*, 2021), but only one other KCS (HvKCS6/EMR1) was functionally characterized (Weidenbach *et al.*, 2015). KCS enzymes are involved in fatty acid metabolism, cutin and suberin biosynthesis, and developmental processes (Lee *et al.*, 2009; Voisin *et al.*, 2009). Expression in yeast showed that CER-ZH elongates C16/C18 saturated and monounsaturated fatty acids up to C22 carbon atoms, which can presumably be elongated by other KCSs (Li *et al.*, 2018).

Wax composition of *cer-za.227* and *cer-ye.267*

The *cer-za.227* and *cer-ye.267* mutations decreased the total wax load by *c.* 70% (Fig. 2), in agreement with previous results using *cer-za.126* and *cer-zh.54* (Rostás *et al.*, 2008; Li *et al.*, 2018). The amounts of epicuticular waxes were strongly decreased. Intracuticular waxes were less affected in *cer-za.227*, and they were strongly decreased in *cer-ye.267* (Fig. 2). The most prominent change in the two mutants was the decline in the amount of 26:0ol. In *cer-za.227*, the amounts of aldehydes (26:0al) and wax esters were decreased, while alkanes (mostly C33) accumulated. Primary alcohols and wax esters are derived from the reductive pathway, whose initial step is catalyzed by FAR (Haslam *et al.*, 2017). A restriction in the production of primary alcohols

© 2023 The Authors
New Phytologist © 2023 New Phytologist Foundation

14698137, 2023, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19063, Wiley Online Library on [25/03/2024]. 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

1914 Research

New Phytologist

consequently affects substrate availability for wax ester biosynthesis. Alkane production is increased, possibly due to an increase in the carbon flux into the decarbonylation pathway, which gives rise to aldehyde and alkane production. The decrease in the amounts of aldehydes in *cer-za.227* might be explained by rapid decarbonylation of aldehydes to alkanes, or by a reduced production of aldehydes through the reductive pathway as a side product of FAR. The amounts of all wax classes, including alcohols,

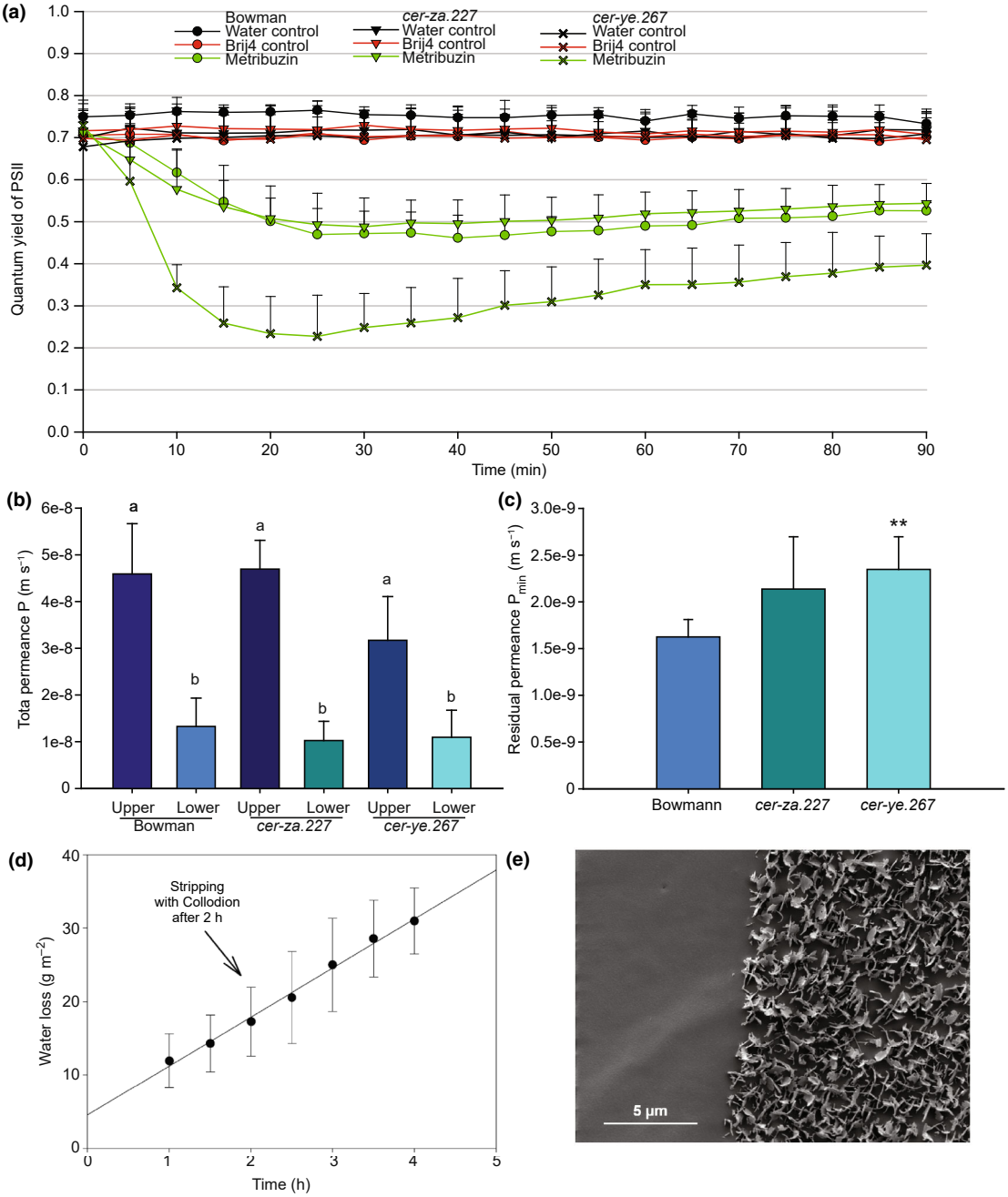


Fig. 7 Permeability and water loss through the cuticle of the leaves of the barley *cer-za.227* and *cer-ye.267* mutants. (a) The quantum yield of Photosystem II (PSII) was recorded as a measure of the penetration of the photosynthetic inhibitor Metribuzin across the cuticular barrier of Bowman, *cer-za.227*, and *cer-za.267*. Brij 4 and water controls were added to exclude effects by components of the spray solution. Mean ± SD; $n = 3$. (b) Stomatal transpiration rates given as total permeance P (m s^{-1}) of the adaxial and abaxial sites of five leaves each of Bowman, *cer-za.227* and *cer-za.267* determined using a leaf porometer. Means ± SD; $n = 5$; ANOVA; different letters indicate differences with $P < 0.05$. (c) Residual permeance P_{min} (transpiration through the cuticle). The P_{min} values were calculated from the means of individual measurements after stomatal closing, as shown in Supporting Information Fig. S6. Data show means and the 95% confidence intervals (CI) for Bowman ($n = 35$), *cer-za.227* ($n = 50$), and *cer-ye.267* ($n = 50$); *t*-test; **, $P < 0.01$. (d) Water loss of a leaf (cultivar Golden Promise) was recorded 1 h after detachment when stomata were closed, thus representing residual water permeance (cuticular transpiration). After one additional h, the epicuticular waxes were stripped off with collodion, and the measurements continued. Mean ± SD, $n = 3$. (e) Scanning electron micrograph of the leaf surface (Golden Promise). Epicuticular waxes have been stripped with collodium on the left side of the leaf, leaving behind the intracuticular waxes, while the right side, which was not treated, displays the coverage with epicticular waxes.

© 2023 The Authors
New Phytologist © 2023 New Phytologist Foundation

14698137, 2023, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19063, Wiley Online Library on [25/03/2024]. 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

New Phytologist

Research 1915

aldehydes, esters, alkanes and acids, were decreased in cer-ye.267, in analogy with *cer-zh.54* (Larsson & Svenningsson, 1986; Li *et al.*, 2018). The *cer-ye.267* and *cer-zh.54* lines carry mutations in HvKCS1, involved in acyl chain elongation. The two mutants displayed changes in chain length distribution of wax esters, because long-chain esters were decreased compared with short ones (Fig. S1j; Li *et al.*, 2018). It is possible that the HvKCS1 activity is redundant with other HvKCS proteins in barley (Li *et al.*, 2018). Furthermore, the amount of alkanes was decreased in *cer-ye.267* rather than increased as in *cer-za.267*. Although these two plants carry mutations in genes of wax biosynthetic enzymes, they result in a different distribution between intracuticular and epicuticular waxes. This might be caused by differences in the wax composition, which might alter their auto-assembly and affect the ratio of intracuticular to epicuticular waxes (Koch *et al.*, 2008).

Wax crystal formation and self-cleaning effect

Primary alcohols in the epicuticular waxes play an important role for the formation of plate-shaped crystals, which render the surface super-hydrophobic (Koch *et al.*, 2006). Bowman leaves are densely covered with plate-like crystals. The number of crystals was reduced in *cer-za.228* and *cer-ye.267*, which both displayed decreased amounts of 26:0ol, in agreement with previous results for *cer-za.126* and *cer-zh.54* (Rostás *et al.*, 2008; Li *et al.*, 2018). The crystals provide an efficient self-cleaning effect, resulting in decreased water droplet angles and compromised water repellence. Therefore, deficiency in epicuticular wax in *cer-za.227* and *cer-ye.267* affects their self-cleaning properties and water repellence.

Water loss and cuticular permeability in *cer-za.227* and *cer-ye.267*

The intracuticular waxes in *cer-za.227* are decreased to *c.* 60% of Bowman levels, and the remaining amount was still sufficient to establish a functional barrier for Metribuzin and to maintain a slightly but not significantly increased cuticular transpiration rate (Fig. 7). The amount of intracuticular waxes in *cer-ye.267* is decreased to *c.* 46% of Bowman levels, and the permeability for Metribuzin and the cuticular transpiration were significantly increased (Fig. 7). Direct measurements of residual water loss of a leaf after collodion treatment demonstrated that intracuticular rather than epicuticular waxes are relevant for establishing the cuticular transpiration barrier (Fig. 7d,e). These results are in good agreement with previous findings, which showed that cuticular barrier functions are primarily established by intracuticular waxes (Jetter & Riederer, 2016; Zeisler & Schreiber, 2016; Zeisler-Diehl *et al.*, 2018). No simple and linear relationship between cuticular barrier properties and wax amounts was found for different species (Riederer & Schreiber, 2001), or for the same species (Sadler *et al.*, 2016; Grünhofer *et al.*, 2022). The cuticular permeability remained unchanged or it increased in a set of seven cuticle mutants of Arabidopsis, with decreases or increases in wax or cutin amounts. Therefore, the coordinated deposition and the semicrystalline arrangement of the highly ordered wax barrier within the cuticle might have been disturbed in the mutants with increased permeabilities. This could also be the reason for the altered cuticular permeability of *cer-ye.267* (Fig. 7). In contrast to the total wax load, alkanes and esters were proposed to contribute to the establishment of a functional transport barrier (Larsson & Svenningsson, 1986; W. Wang *et al.*, 2015; Bueno *et al.*, 2019; Li *et al.*, 2020; Wu *et al.*, 2022). The amount of alkanes was increased in *cer-za.227*, possibly contributing to the barrier function, but it was decreased in *cer-ye.267* (Fig. S2). Although the wax ester content was decreased, this effect was restricted to the epicuticular waxes. The *cer-ye.267* and *cer-zh.54* mutants showed a shift in chain length distribution of wax esters (Fig. S1j; Li *et al.*, 2018) possibly contributing to the modulation of cuticular permeability. Conversely, the strong reduction in the amounts of primary alcohols (26:0ol) did not affect the cuticular barrier function in *cer-za.227*, and it therefore cannot be the cause for increased permeability and cuticular transpiration in *cer-ye.267*. Instead, primary alcohols, which are abundant in Poaceae, presumably fulfill other functions in crystal formation and water repellence and light reflection (Fig. S2; Koch *et al.*, 2006).

In conclusion, the two barley lines *cer-za.227* and *cer-ye.267*, which carry mutations in two genes involved in wax biosynthesis, *HvFAR1* and *HvKCS1*, respectively, reveal a strong reduction in total wax amounts, distinct changes in composition, and amounts of intracuticular to epicuticular waxes. These changes caused an increase in cuticular permeability (Fig. 7a) in *cer-ye.267*, but not in *cer-za.227*. Therefore, the decrease in cuticular barrier function in *cer-ye.267* compared with *cer-za.227* is presumably explained by the differences in wax composition and a reduction in intracuticular waxes. The precise role of the individual wax lipids for the establishment of the (intra)cuticular barrier, however, remains to be discovered.

Acknowledgements

Funding for this project was provided by the Deutsche Forschungsgemeinschaft (DFG Research Training Group, GRK2064, 'Water use efficiency and drought stress responses in Arabidopsis and /barley'; and Germany's Excellence Strategy, EXC 2070, PhenoRob grant no. 390732324), DFG program Major Research Instrumentation (Scanning Electron Microscope with EDX, grant no. 471591895), and by University of Bonn. CC acknowledges funding from the Biotechnological and Biological Sciences Research Council (BB/M018466/1) to Robbie Waugh. We would like to thank Marlies Becker and Christiane Buchholz for barley transformation and Penny von Wettstein-Knowles (University of Copenhagen) for her insightful guidance on the work with the barley mutants. Open Access funding enabled and organized by Projekt DEAL.

Competing interests

None declared.

© 2023 The Authors
New Phytologist © 2023 New Phytologist Foundation

14698137, 2023, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19063, Wiley Online Library on [25/03/2024]. 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

New Phytologist

Author contributions

YM, PP, US, HS, LS, and PD conceived and designed experiments. YM, PP, VZ-D, LG, MK, and MMD performed the experiments. YM, PP, TS, US, HS, LS, and PD analyzed the data. TS, US, CC, SJ, AMH, IA, HS, and LS contributed bioinformatic, genetic resources. YM, PP, TS, and PD wrote the manuscript with contributions from all authors. YM and PP contributed equally to this work.

ORCID

Ivan F. Acosta https://orcid.org/0000-0001-7080-3384
Chiara Campoli https://orcid.org/0000-0001-7809-2927
Peter Dörmann https://orcid.org/0000-0002-5845-9370
Lea Grewe https://orcid.org/0000-0001-7593-7840
Alistair M. Hetherington https://orcid.org/0000-0001-6060-9203
Sarah Jose https://orcid.org/0000-0003-2737-6473
Yannic Müller https://orcid.org/0000-0002-6974-4951
Payal Patwari https://orcid.org/0000-0003-1356-6639
Heiko Schoof https://orcid.org/0000-0002-1527-3752
Lukas Schreiber https://orcid.org/0000-0001-7003-9929
Tyll Stöcker https://orcid.org/0000-0001-7184-9472
Viktoria Zeisler-Diehl https://orcid.org/0000-0001-7050-9716

Data availability

Raw sequencing data (transcriptomic sequencing, RNA-Seq) have been deposited at the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) with accession no. PRJNA890449.

References

Aarts MG, Hodge R, Kalantidis K, Florack D, Wilson ZA, Mulligan BJ, Stiekema WJ, Scott R, Pereira A. 1997. The Arabidopsis MALE STERILITY 2 protein shares similarity with reductases in elongation/condensation complexes. *The Plant Journal* **12**: 615–623.

Amanda D, Frey FP, Neumann U, Przybyl M, Šimura J, Zhang Y, Chen Z, Gallavotti A, Fernie AR, Ljung K *et al.* 2022. Auxin boosts energy generation pathways to fuel pollen maturation in barley. *Current Biology* **32**: 1798–1811.

Batsale M, Bahammou D, Fouillen L, Mongrand S, Joubès J, Domergue F. 2021. Biosynthesis and functions of very-long-chain fatty acids in the responses of plants to abiotic and biotic stresses. *Cells* **10**: 1284.

Bernard A, Joubès J. 2013. Arabidopsis cuticular waxes: advances in synthesis, export and regulation. *Progress in Lipid Research* **52**: 110–129.

Bregitzer P, Lundqvist U, Carollo BV. 2013. Descriptions of barley genetic stocks for 2013. *Barley Genetics Newsletter* **43**: 48–223.

Bueno A, Alfarhan A, Arand K, Burghardt M, Deininger A-C, Hedrich R, Leide J, Seufert P, Staiger S, Riederer M. 2019. Effects of temperature on the cuticular transpiration barrier of two desert plants with water-spender and water-saver strategies. *Journal of Experimental Botany* **70**: 1613–1625.

Chai G, Li C, Xu F, Li Y, Shi X, Wang Y, Wang Z. 2018. Three endoplasmic reticulum-associated fatty acyl-coenzyme a reductases were involved in the production of primary alcohols in hexaploid wheat (*Triticum aestivum* L.). *BMC Plant Biology* **18**: 41.

Doan TTP, Carlsson AS, Hamberg M, Bülow L, Stymne S, Olsson P. 2009. Functional expression of five Arabidopsis fatty acyl-CoA reductase genes in *Escherichia coli*. *Journal of Plant Physiology* **166**: 787–796.

Doan TTP, Domergue F, Fournier AE, Vishwanath SJ, Rowland O, Moreau P, Wood CC, Carlsson AS, Hamberg M, Hofvander P. 2012. Biochemical characterization of a chloroplast localized fatty acid reductase from *Arabidopsis thaliana*. *Biochimica et Biophysica Acta (BBA) – Bioenergetics* **1821**: 1244–1255.

Domergue F, Vishwanath SJ, Joubès J, Ono J, Lee JA, Bourdon M, Alhattab R, Lowe C, Pascal S, Lessire R *et al.* 2010. Three Arabidopsis fatty acyl-coenzyme A reductases, FAR1, FAR4, and FAR5, generate primary fatty alcohols associated with suberin deposition. *Plant Physiology* **153**: 1539–1554.

Dong W, Wu D, Li G, Wu D, Wang Z. 2018. Next-generation sequencing from bulked segregant analysis identifies a dwarfism gene in watermelon. *Scientific Reports* **8**: 2908.

Druka A, Franckowiak J, Lundqvist U, Bonar N, Alexander J, Houston K, Radovic S, Shahinnia F, Vendramin V, Morgante M *et al.* 2011. Genetic dissection of barley morphology and development. *Plant Physiology* **155**: 617–627.

Gol L, Haraldsson EB, von Korff M. 2021. Ppd-H1 integrates drought stress signals to control spike development and flowering time in barley. *Journal of Experimental Botany* **72**: 122–136.

Grünhofer P, Herzig L, Sent S, Zeisler-Diehl VV, Schreiber L. 2022. Increased cuticular wax deposition does not change residual foliar transpiration. *Plant, Cell & Environment* **45**: 1157–1171.

Haas K, Rentschler I. 1984. Discrimination between epicuticular and intracuticular wax in blackberry leaves: ultrastructural and chemical evidence. *Plant Science Letters* **36**: 143–147.

Harwood WA. 2019. An introduction to barley: the crop and the model. *Methods in Molecular Biology* **1900**: 1–5.

Haslam TM, Gerelle WK, Graham SW, Kunst L. 2017. The unique role of the ECERIFERUM2-LIKE clade of the BAHD acyltransferase superfamily in cuticular wax metabolism. *Plants* **6**: 23.

Jetter R, Riederer M. 2016. Localization of the transpiration barrier in the epi- and intracuticular waxes of eight plant species: water transport resistances are associated with fatty acyl rather than alicyclic components. *Plant Physiology* **170**: 921–934.

Koch K, Barthlott W, Koch S, Hommes A, Wandelt K, Mamdouh W, De-Feyter S, Broekmann P. 2006. Structural analysis of wheat wax (*Triticum aestivum*, c.v. 'Naturastar' L.): from the molecular level to three dimensional crystals. *Planta* **223**: 258–270.

Koch K, Bhushan B, Barthlott W. 2008. Diversity of structure, morphology and wetting of plant surfaces. *Soft Matter* **4**: 1943.

Koornneef M, Hanhart CJ, Thiel F. 1989. A genetic and phenotypic description of *eceriferum* (*cer*) mutants in *Arabidopsis thaliana*. *Journal of Heredity* **80**: 118–122.

Kunst L, Samuels L. 2009. Plant cuticles shine. advances in wax biosynthesis and export. *Current Opinion in Plant Biology* **12**: 721–727.

Larsson S, Svenningsson M. 1986. Cuticular transpiration and epicuticular lipids of primary leaves of barley (*Hordeum vulgare*). *Physiologia Plantarum* **68**: 13–19.

Lee S-B, Jung S-J, Go Y-S, Kim H-U, Kim J-K, Cho H-J, Park OK, Suh M-C. 2009. Two Arabidopsis 3-ketoacyl CoA synthase genes, KCS20 and KCS2/DAISY, are functionally redundant in cuticular wax and root suberin biosynthesis, but differentially controlled by osmotic stress. *The Plant Journal* **60**: 462–475.

Lee SB, Suh MC. 2015. Advances in the understanding of cuticular waxes in *Arabidopsis thaliana* and crop species. *Plant Cell Reports* **34**: 557–572.

Li C, Haslam TM, Krüger A, Schneider LM, Mishina K, Samuels L, Yang H, Kunst L, Schaffrath U, Nawrath C *et al.* 2018. The β-ketoacyl-CoA synthase HvKCS1, encoded by Cer-zh, plays a key role in synthesis of barley leaf wax and germination of barley powdery mildew. *Plant and Cell Physiology* **59**: 806–822.

Li H, Guo Y, Cui Q, Zhang Z, Yan X, Ahammed GJ, Yang X, Yang J, Wei C, Zhang X. 2020. Alkanes (C29 and C31)-mediated intracuticular wax accumulation contributes to melatonin- and ABA-induced drought tolerance in watermelon. *Journal of Plant Growth Regulation* **39**: 1441–1450.

© 2023 The Authors
New Phytologist © 2023 New Phytologist Foundation

14698137, 2023, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19063, Wiley Online Library on [25/03/2024]. 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

New Phytologist

Research 1917

Liu L, Jose SB, Campoli C, Bayer MM, Sánchez-Diaz MA, McAllister T, Zhou Y, Eskan M, Milne L, Schreiber M *et al.* 2022. Conserved signalling components coordinate epidermal patterning and cuticle deposition in barley. *Nature Communications* **13**: 6050.

Long LM, Patel HP, Cory WC, Stapleton AE. 2003. The maize epicuticular wax layer provides UV protection. *Functional Plant Biology* **30**: 75–81.

Lundqvist U. 2014. Scandinavian mutation research in barley – a historical review. *Hereditas* **151**: 123–131.

Lundqvist U, Lundqvist A. 1988. Mutagen specificity in barley for 1580 *eceriferum* mutants localized to 79 loci. *Hereditas* **108**: 1–12.

Magwene PM, Willis JH, Kelly JK. 2011. The statistics of bulk segregant analysis using next generation sequencing. *PLoS Computational Biology* **7**: e1002255.

Mansour E, Moustafa ESA, Qabil N, Abdelsalam A, Wafa HA, Kenawy AE, Casas AM, Igartua E. 2018. Assessing different barley growth habits under Egyptian conditions for enhancing resilience to climate change. *Field Crops Research* **224**: 67–75.

Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J *et al.* 2017. A chromosome conformation capture ordered sequence of the barley genome. *Nature* **544**: 427–433.

McAllister T, Campoli C, Eskan M, Liu L, McKim SM. 2022. A gene encoding a SHINE1/WAX INDUCER1 transcription factor controls cuticular wax in barley. *Agronomy* **12**: 1088.

McCartney AW, Dyer JM, Dhanoa PK, Kim PK, Andrews DW, McNew JA, Mullen RT. 2004. Membrane-bound fatty acid desaturases are inserted co-translationally into the ER and contain different ER retrieval motifs at their carboxy termini. *The Plant Journal* **37**: 156–173.

Millar AA, Kunst L. 1997. Very-long-chain fatty acid biosynthesis is controlled through the expression and specificity of the condensing enzyme. *The Plant Journal* **12**: 121–131.

Milne L, Bayer M, Rapazote-Flores P, Mayer C-D, Waugh R, Simpson CG. 2021. EORNA, a barley gene and transcript abundance database. *Scientific Data* **8**: 90.

Nawrath C. 2006. Unraveling the complex network of cuticular structure and function. *Current Opinion in Plant Biology* **9**: 281–287.

Nødskov Giese B. 1975. Effects of light and temperature on the composition of epicuticular wax of barley leaves. *Phytochemistry* **14**: 921–929.

Park J, Bae S, Kim J-S. 2015. Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. *Bioinformatics* **31**: 4014–4016.

Patwari P, Salewski V, Gutbrod K, Kreszies T, Dresen-Scholz B, Peisker H, Steiner U, Meyer AJ, Schreiber L, Dörmann P. 2019. Surface wax esters contribute to drought tolerance in Arabidopsis. *The Plant Journal* **98**: 727–744.

Rentsch D, Laloi M, Rouhara I, Schmelzer E, Delrot S, Frommer WB. 1995. NTR1 encodes a high affinity oligopeptide transporter in *Arabidopsis*. *FEBS Letters* **370**: 264–268.

Richardson A, Boscari A, Schreiber L, Kerstiens G, Jarvis M, Herzyk P, Fricke W. 2007. Cloning and expression analysis of candidate genes involved in wax deposition along the growing barley (*Hordeum vulgare*) leaf. *Planta* **226**: 1459–1473.

Riederer M, Schreiber L. 2001. Protecting against water loss: analysis of the barrier properties of plant cuticles. *Journal of Experimental Botany* **52**: 2023–2032.

Rostás M, Ruf D, Zabka V, Hildebrandt U. 2008. Plant surface wax affects parasitoid's response to host footprints. *Die Naturwissenschaften* **95**: 997–1002.

Rowland O, Zheng H, Hepworth SR, Lam P, Jetter R, Kunst L. 2006. *CER4* encodes an alcohol-forming fatty acyl-coenzyme A reductase involved in cuticular wax production in Arabidopsis. *Plant Physiology* **142**: 866–877.

Sadler C, Schroll B, Zeisler V, Waßmann F, Franke R, Schreiber L. 2016. Wax and cutin mutants of *Arabidopsis*: quantitative characterization of the cuticular transport barrier in relation to chemical composition. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1861**: 1336–1344.

Schneider LM, Adamski NM, Christensen CE, Stuart DB, Vautrin S, Hansson M, Uauy C, von Wettstein-Knowles P. 2016. The *Cer-cqu* gene cluster determines three key players in a β-diketone synthase polyketide pathway synthesizing aliphatics in epicuticular waxes. *Journal of Experimental Botany* **67**: 2715–2730.

Suh MC, Samuels AL, Jetter R, Kunst L, Pollard M, Ohlrogge J, Beisson F. 2005. Cuticular lipid composition, surface structure, and gene expression in Arabidopsis stem epidermis. *Plant Physiology* **139**: 1649–1665.

Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, Takuno S *et al.* 2013. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *The Plant Journal* **74**: 174–183.

Tong T, Fang Y-X, Zhang Z, Zheng J, Zhang X, Li J, Niu C, Xue D, Zhang X. 2021. Genome-wide identification and expression pattern analysis of the KCS gene family in barley. *Plant Growth Regulation* **93**: 89–103.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3 – new capabilities and interfaces. *Nucleic Acids Research* **40**: e115.

Voinnet O, Lederer C, Baulcombe DC. 2000. A viral movement protein prevents spread of the gene silencing signal in *Nicotiana benthamiana*. *Cell* **103**: 157–167.

Voisin D, Nawrath C, Kurdyukov S, Franke RB, Reina-Pinto JJ, Efremova N, Will I, Schreiber L, Yephremov A. 2009. Dissection of the complex phenotype in cuticular mutants of Arabidopsis reveals a role of SERRATE as a mediator. *PLoS Genetics* **5**: e1000703.

Wang M, Wang Y, Wu H, Xu J, Li T, Hegebarth D, Jetter R, Chen L, Wang Z. 2016. Three TaFAR genes function in the biosynthesis of primary alcohols and the response to abiotic stresses in *Triticum aestivum*. *Scientific Reports* **6**: 25008.

Wang W, Zhang Y, Xu C, Ren J, Liu X, Black K, Gai X, Wang Q, Ren H. 2015. Cucumber ECERIFERUM1 (CsCER1), which influences the cuticle properties and drought tolerance of cucumber, plays a key role in VLC alkanes biosynthesis. *Plant Molecular Biology* **87**: 219–233.

Wang Y, Sun Y, You Q, Luo W, Wang C, Zhao S, Chai G, Li T, Shi X, Li C *et al.* 2018. Three fatty acyl-coenzyme A reductases, BdFAR1, BdFAR2 and BdFAR3, are involved in cuticular wax primary alcohol biosynthesis in *Brachypodium distachyon*. *Plant and Cell Physiology* **59**: 527–543.

Wang Y, Wang M, Sun Y, Hegebarth D, Li T, Jetter R, Wang Z. 2015a. Molecular characterization of TaFAR1 involved in primary alcohol biosynthesis of cuticular wax in hexaploid wheat. *Plant and Cell Physiology* **56**: 1944–1961.

Wang Y, Wang M, Sun Y, Wang Y, Li T, Chai G, Jiang W, Shan L, Li C, Xiao E *et al.* 2015b. FAR5, a fatty acyl-coenzyme A reductase, is involved in primary alcohol biosynthesis of the leaf blade cuticular wax in wheat (*Triticum aestivum* L.). *Journal of Experimental Botany* **66**: 1165–1178.

Weidenbach D, Jansen M, Bodewein T, Nagel KA, Schaffrath U. 2015. Shoot and root phenotyping of the barley mutant kcs6 (3-ketoacyl-CoA synthase6) depleted in epicuticular waxes under water limitation. *Plant Signaling & Behavior* **10**: 1–3.

Weyers JDB, Travis AJ. 1981. Selection and preparation of leaf epidermis for experiments on stomatal physiology. *Journal of Experimental Botany* **32**: 837–850.

Wu H, Le Liu CY, Liu T, Jiang Q, Wei Z, Li C, Wang Z. 2022. Tomato SlCER1-1 catalyzes the synthesis of wax alkanes which increases the drought tolerance and fruit storability. *Horticulture Research* **9**: uhac004.

Wu P, Xie J, Hu J, Qiu D, Liu Z, Li J, Li M, Zhang H, Yang L, Liu H *et al.* 2018. Development of molecular markers linked to powdery mildew resistance gene Pm4b by combining SNP discovery from transcriptome sequencing data with bulked segregant analysis (BSR-Seq) in wheat. *Frontiers in Plant Science* **9**: 95.

Zeisler V, Schreiber L. 2016. Epicuticular wax on cherry laurel (*Prunus laurocerasus*) leaves does not constitute the cuticular transpiration barrier. *Planta* **243**: 65–81.

Zeisler-Diehl V, Müller Y, Schreiber L. 2018. Epicuticular wax on leaf cuticles does not establish the transpiration barrier, which is essentially formed by intracuticular wax. *Journal of Plant Physiology* **227**: 66–74.

Zhang S, Wu S, Niu C, Liu D, Yan T, Tian Y, Liu S, Xie K, Li Z, Wang Y *et al.* 2021. *ZmMs25* encoding a plastid-localized fatty acyl reductase is critical for anther and pollen development in maize. *Journal of Experimental Botany* **72**: 4298–4318.

Zhang X, Henriques R, Lin S-S, Niu Q-W, Chua N-H. 2006. Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nature Protocols* **1**: 641–646.

© 2023 The Authors
New Phytologist © 2023 New Phytologist Foundation

14698137, 2023, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19063, Wiley Online Library on [25/03/2024]. 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

1918 Research

New Phytologist

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Cuticular wax composition in leaves of Bowman and the *cer-za.227* and *cer-ye.267* mutants.

Fig. S2 Wax ester composition in leaves of the Bowman and the *cer-za.227* mutant.

Fig. S3 Identification of the *CER-ZA* locus using bulked segregant RNA sequencing (BSR-Seq).

Fig. S4 Phylogenetic tree of amino acid sequences with similarity to acyl-CoA reductases (FAR) from different plant species.

Fig. S5 Identification of the *CER-YE* locus using bulked segregant RNA sequencing (BSR-Seq).

Fig. S6 Residual permeance P ($m\,s^{-1}$) giving the cuticular transpiration of five detached leaves plotted against the relative water deficit (RWD).

Methods S1 Bulked segregant analysis RNA sequencing (BSR-Seq) and bioinformatic and phylogenetic analyses.

Table S1 Cultivars and mutant lines of barley (*Hordeum vulgare*) used in this study.

Table S2 Oligonucleotides used in this study.

Table S3 Expression patterns derived from BSR-Seq experiments in Bowman and *cer-za.227*.

Table S4 Paralogous sequences of *HORVU5Hr1G089230* in barley.

Table S5 Single-nucleotide polymorphisms (SNPs) in *HORVU5Hr1G089230* in the cultivars Morex and Foma and in the *cer-za* mutant alleles.

Table S6 Expression patterns derived from BSR-Seq experiments in Bowman and *cer-ye.267*.

Please note: Wiley is not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

© 2023 The Authors
New Phytologist © 2023 New Phytologist Foundation

14698137, 2023, 5, Downloaded from https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.19063, Wiley Online Library on [25/03/2024]. 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

Received: 1 July 2022 | Revised: 5 August 2022 | Accepted: 12 August 2022

DOI: 10.1111/ppl.13765

ECOPHYSIOLOGY, STRESS AND ADAPTATION

Physiologia Plantarum

Populus × *canescens* root suberization in reaction to osmotic and salt stress is limited to the developing younger root tip region

Paul Grünhofer[1] | **Tyll Stöcker**[2] | **Yayu Guo**[3,4,5] | **Ruili Li**[3,4,5] | **Jinxing Lin**[3,4,5] | **Kosala Ranathunge**[6] | **Heiko Schoof**[2] | **Lukas Schreiber**[1]

[1]Department of Ecophysiology, Institute of Cellular and Molecular Botany, University of Bonn, Bonn, Germany

[2]Department of Crop Bioinformatics, Institute of Crop Science and Resource Conservation, University of Bonn, Bonn, Germany

[3]Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, Beijing Forestry University, Beijing, China

[4]College of Biological Science and Technology, Beijing Forestry University, Beijing, China

[5]Institute of Tree Development and Genome Editing, Beijing Forestry University, Beijing, China

[6]UWA School of Biological Sciences, The University of Western Australia, Perth, Western Australia, Australia

Correspondence
Paul Grünhofer, Department of Ecophysiology, Institute of Cellular and Molecular Botany, University of Bonn, Kirschallee 1, 53115 Bonn, Germany.
Email: p.gruenhofer@uni-bonn.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Numbers: SCHR17/1, Project Number 391657309; National Natural Science Foundation of China, Grant/Award Numbers: 31761133009, 31970182; Programme of Introducing Talents of Discipline to Universities, 111 Project, Grant/Award Number: B13007

Edited by K. Noguchi

Abstract

Populus is a valuable and fast-growing tree species commonly cultivated for economic and scientific purposes. But most of the poplar species are sensitive to drought and salt stress. Thus, we compared the physiological effects of osmotic stress (PEG8000) and salt treatment (NaCl) on poplar roots to identify potential strategies for future breeding or genetic engineering approaches. We investigated root anatomy using epifluorescence microscopy, changes in root suberin composition and amount using gas chromatography, transcriptional reprogramming using RNA sequencing, and modifications of root transport physiology using a pressure chamber. Poplar roots reacted to the imposed stress conditions, especially in the developing younger root tip region, with remarkable differences between both types of stress. Overall, the increase in suberin content was surprisingly small, but the expression of key suberin biosynthesis genes was strongly induced. Significant reductions of the radial water transport in roots were only observed for the osmotic and not the hydrostatic hydraulic conductivity. Our data indicate that the genetic enhancement of root suberization processes in poplar might be a promising target to convey increased tolerance, especially against toxic sodium chloride.

1 | INTRODUCTION

The anthropogenic climate change, which has occurred over the past 50 years and is still ongoing (Rosenzweig et al., 2008), has inevitable consequences threatening global agriculture (Boyer, 1982; Challinor et al., 2014) and agroforestry (Allen et al., 2010; Choat et al., 2012). It is no secret that even more extreme weather conditions in the future, for example, prolonged periods of drought or flooding, represent a substantial risk for crop productivity (Kang et al., 2009; Polle et al., 2019) and will impose a plethora of challenges on future generations. Dry periods are often coinciding with heat (Mittler, 2006) and salt accumulation in the upper soil layers (Polle & Chen, 2015).

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.
© 2022 The Authors. *Physiologia Plantarum* published by John Wiley & Sons Ltd on behalf of Scandinavian Plant Physiology Society.

Therefore, comprehensive studies of more than one abiotic stress may be of great scientific benefit (Munns, 2002). In terms of osmotic and salt stress, a water potential of −0.8 MPa and a sodium chloride (NaCl) concentration of 150 mM are considered of moderate intensity and are frequently observed under natural conditions (Osthoff et al., 2019). When thinking about optimizing plant productivity to meet the demands of a planet subjected to severe environmental changes, the idea of cultivating trees on saline soils that may not be used for food production appears to be of particular interest (Polle & Chen, 2015).

In this context, *Populus* might be a promising tree genus, as it is employed both in science and economic purposes thanks to several useful attributes. About 30 associated species, subdivided into six sections with diverse habitats, display a great variety of morphological features and physiological peculiarities (Eckenwalder, 1996; Isebrands & Richardson, 2014). Aside from phenotypic heterogeneity (Grünhofer, Herzig, & Schreiber, 2021; Pearce et al., 2005), natural hybridization and interspecific crossability (Cervera et al., 2005; Ronald, 1982), and ease of vegetative propagation (Müller et al., 2013), poplars are generally renowned for their rapid growth (Dillen et al., 2010) and high scientific and economic value (Jansson & Douglas, 2007; Sannigrahi & Ragauskas, 2010). This has resulted in the development of efficient breeding and transformation protocols (Fillatti et al., 1987; Leplé et al., 1992) as well as in the sequencing and exploitation of several of its genomes since 2006 (Lin et al., 2018; Ma et al., 2019; Mader et al., 2016; Qiu et al., 2019; Tuskan et al., 2006; Zhang et al., 2020).

A wide variety of stress conditions has already been empirically tested on poplars, often with more than just one species or clone, which greatly facilitates the utilization of beneficial features for future biotechnological approaches. Investigated abiotic stresses and stimuli include: (1) salinity (Chen & Polle, 2010; Tang et al., 2010); (2) osmotic and water deficit stress (Leng et al., 2013; Silim et al., 2009); (3) heavy metal exposure (Stoláriková et al., 2012; Zacchini et al., 2009); (4) hypoxia (Kreuzwieser et al., 2009; Peng et al., 2017); (5) irradiance (Chen et al., 2006; Niinemets et al., 1999); (6) temperature (Ranganathan et al., 2016; Zanewich et al., 2018); and (7) chemical treatment (Kamaluddin & Zwiazek, 2003; Wan & Zwiazek, 1999).

Root suberization processes might represent a promising target for future genetic manipulation of salinity and drought tolerance, as they were repeatedly identified in stress-induced reactions in many relatively short-lived crops (Grünhofer, Schreiber, & Kreszies, 2021; Reinhardt & Rost, 1995; Schreiber et al., 2005). Increased suberization of endo- and exodermal cell layers is thought to decrease the plasma membrane surface area available for ion absorption and is speculated to aid in the restriction of uncontrolled water loss (Enstone et al., 2003). However, despite a few pioneer studies (Bagniewska-Zadworna et al., 2014; Lux et al., 2004), information on root suberization in perennial tree species is still scarce (Brunner et al., 2015; Polle et al., 2019). The abundant supply of transcriptomic studies with *Populus* root tissue (Bogeat-Triboulot et al., 2007; Brinker et al., 2010; Cohen et al., 2010; Ma et al., 2013; Royer et al., 2016; Zheng et al., 2015) only partially satisfies this interest, because suberization genes were not a core target, samples were taken only hours after stress application, or corresponding physiological data was missing entirely. In general, studies on long-term abiotic stress application, that is, weeks or months, are less common (Polle et al., 2019).

In this study, we intended to compare the physiological effects of prolonged (7-day-long) osmotic (PEG8000) and salt (NaCl) stress with matched water potentials on root suberization processes of poplars.

2 | MATERIALS AND METHODS

For a more detailed description of the methods applied for hydroponic *P.* × *canescens* plant cultivation, histochemical, and analytical analyses of apoplastic transport barriers in poplar roots, and means of root transport physiological investigation, the reader is advised to Grünhofer, Guo, et al. (2021).

2.1 | Plant material and cultivation conditions

All experiments have been conducted with the fully sequenced *P.* × *canescens* (Aiton) Sm. clone "84K" (*P. alba* × *P. tremula* var. *glandulosa*) (Qiu et al., 2019). The plants were constantly growing in a climate chamber under long-day conditions (16 h light/8 h dark period) at a light intensity of approximately 100 µmol m^{-2} s^{-1}. The mean temperature and mean relative humidity were 21°C and 50% during illumination, and 19°C and 67% throughout the night, respectively. Tissue culture-propagated plants (6- to 8-week-old) were transplanted into soil (Einheitserde Classic Type Topf 1.5, Einheitserde Werksverband e.V.) and acclimatized to climate chamber conditions. After a further 8–10 weeks of growth in soil (total plant age of 14–18 weeks), they were used in hydroponic experiments.

Each hydroponic experiment lasted 5 weeks and was composed of an initial 2-week-long rooting phase of stem cuttings in stagnant tap water, followed by 2 weeks of cultivation in aerated ½ Hoagland nutrient solution (Hoagland & Arnon, 1950). In the fifth week, one set of plants continued to grow in ½ Hoagland solution (control) while the other sets were exposed to different stress treatments and intensities (½ Hoagland solution adjusted with osmotic stress or salt stress). The nutrient solution was replaced weekly, and five stem cuttings were combined in one hydroponic container to yield one biological replicate.

2.2 | Stress treatments

Osmotic stress conditions were initiated by adjusting the osmotic potential (Ψ_s) of the nutrient solution to −0.4, or −0.6, or −0.8 MPa by adding 17.5%, 21.8%, or 25.5% (w/w) polyethylene glycol 8000 (PEG8000, Carl Roth), respectively (Michel, 1983). The Ψ_s of the salt (NaCl) stress treatments were matched to the PEG8000 series using the Van't Hoff equation; $\Psi_s = -RTiC$, in which R = universal gas constant, T = absolute temperature, i = Van't Hoff factor of sodium chloride (i = 2), and C = osmolarity of the medium. The resulting salt concentrations that matched Ψ_s were 80, 120, or 160 mM, respectively. The osmotic potentials of all nutrient solutions were confirmed

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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

by measuring them using a WP4C Dewpoint PotentiaMeter (Decagon Devices).

2.3 | Estimation of physiological parameters

To physiologically characterize the effects of the different treatments, the chlorophyll contents of leaves (Force A device, Dualex Scientific) and the shoot and root lengths were measured at plant harvest. The intermediate treatments of −0.6 MPa PEG8000 and 120 mM NaCl were used for further studies of root transport measurements (hydraulic conductivity and apoplastic PTS (trisodium, 3-hydroxy-5,8,10-pyrenetrisulfonate) bypass flow), RNA sequencing, stomatal conductance measurements (LI-600 Leaf Porometer, Li-Cor), and osmotic potential measurements of xylem sap as well as in leaf and root tissues (freezing point osmometer, OSMOMAT 030, gonotec).

2.4 | Histochemical observation of Casparian bands and suberin lamellae

Casparian bands were detected using 0.1% (w/v) berberine hemi-sulfate and 0.5% (w/v) aniline blue (Brundrett et al., 1988). Suberin lamellae were stained with 0.01% (w/v) fluorol yellow 088 (Brundrett et al., 1991). Adventitious roots with lengths close to the calculated mean (±3 cm) of each replicate and at least six roots of each treatment were used per staining procedure. Before staining, selected roots were divided into 1 cm segments and subsequently cut into 30 µm thick cross-sections using a cryostat microtome (Microm HM 500 M, Microm International GmbH). Epifluorescence microscopy photographs were taken with a Canon EOS 600D camera using an ultraviolet (UV) filter set (excitation filter BP 365, dichroic mirror FT 395, barrier filter LP 397; Zeiss). Pictures were analyzed (e.g., diameter determination from cross-sections) and adjusted in brightness with ImageJ (Abramoff et al., 2004). Thus it is important to mention that color intensity does not reflect suberin quantity but only suberin localization. The investigated root segments were expressed as relative lengths of the whole root, with 0% representing the root tip and 100% the root base according to Kreszies et al. (2019). Two functional developmental zones were defined based on the investigation of roots cultivated in control conditions for 5 weeks (Grünhofer, Guo, et al., 2021). Zone A, showing no endodermal suberization, ranged from 0% to 27.5% of the relative total root length. Zone B, exhibiting patchy endodermal suberization, spanned the remaining 27.5%–100% distance of the root.

2.5 | Chemical analysis of root suberin

Ten to 20 individual adventitious roots of a given biological replicate, again selected to be close to the calculated mean, were shaved to remove lateral roots using a razor blade, separated into the two functional zones (A: 0%–27.5%; B: 27.5%–100%), and pooled for suberin analysis. The root zones were treated with 0.5% (w/v) cellulase and 0.5% (w/v) pectinase for 2 weeks, borate buffer for 1 day, and 1:1 (v/v) chloroform: methanol for further 2 weeks (Zeier & Schreiber, 1997). All solutions were replaced after 3 days. At the end of this procedure, the remaining root tissue consisting of mainly polymerized and nonextractable cell wall components was dried on polytetrafluoroethylene over activated silica gel and weighed. Samples were transesterified with BF_3-methanol (Zeier & Schreiber, 1998) and complemented with 10 µg of internal standard (Dotriacontane, Fluka). After chloroform extraction of the released suberin monomers, the sample volume was reduced under a gentle stream of nitrogen and derivatized with 20 µl pyridine (Sigma Aldrich) and 20 µl BSTFA (N,O-Bis[trimethylsilyl]trifluoroacetamide, Macherey-Nagel). Splitless gas chromatography (50°C for 1 min, a temperature increase of 25°C min^{-1} up to 200°C, 1 min at 200°C, 10°C min^{-1} up to 320°C, and final hold for 8 min at 320°C; Delude et al., 2017) with subsequent flame ionization detection (GC-FID: 6890 N, Agilent Technologies) or mass spectrometry (GC-MS: 7890B-5977A, Agilent Technology) was performed. The resulting amounts of identified suberin monomers were related to the endodermal surface area (A_{en}) of each zone, calculated based on a truncated cone shape: $A_{en} = \pi\,(R + r\sqrt{(R-r)^2 + h^2})$ (R, endodermis radius at the basal side of root zone; r, endodermis radius at the apical side of root zone; h, length of the individual root zone; r and R were estimated from the diameter of the cross-sections of roots which were used for histochemical studies for each treatment).

2.6 | Transport physiology

To estimate the effects of osmotic and salt stress on transport physiology, only the intermediate treatments of PEG8000 (−0.6 MPa) and NaCl (120 mM) were selected in addition to the control. This was done due to the observation that the intermediate stress intensities revealed observable phenotypic stress reactions without significantly affecting plant vitality. Intact root systems were mounted to the pressure chamber to investigate the xylem sap exudation rate in the absence and the presence of applied pneumatic pressure (Krishnamurthy et al., 2011; Miyamoto et al., 2001), which is in accordance with the composite transport model of roots (Ranathunge et al., 2017; Steudle et al., 1993).

To at least partially reverse short-term fine adjustable stress adaptations (e.g., a decrease of the osmotic potential of roots or gating of aquaporins) and focus on more permanent long-term structural changes (e.g., suberization), stress-exposed stem cuttings were transferred to control conditions (termed "recovery" here-after) 12 h before the transport measurements of roots as previously described by Krishnamurthy et al. (2011) and Knipfer et al. (2020). After truncating the newly developed shoot below the first developed leaf underwater, the remaining stem cutting was fixed to the pressure chamber using silicone seals (Xantopren blue, Heraeus). The root system was placed either in the same nutrient solution used for plant growth under control cultivation to measure both the osmotic and hydrostatic hydraulic conductivity (Lp_r(OS) or Lp_r(HY), respectively), or in 0.01% (w/v) PTS (trisodium, 3-hydroxy-5,8,10-pyrenetrisulfonate) in distilled water for apoplastic bypass flow determination (Hanson et al., 1985;

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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

Ranathunge & Schreiber, 2011). In both experiments, exuded xylem sap was collected in Eppendorf tubes.

To estimate the hydraulic conductivity of roots, we determined the volume flow rate (J_v) by transforming measured weights into volume assuming a density of 1 g cm^{-3}. After normalization of the volume flow rate to the exposed root surface area (J_{vr}), the Lp_r(OS) could be calculated using $Lp_r(OS) = J_{vr}\,(\Delta\pi\,\sigma_{sr})^{-1}$. Here, J_{vr} denotes the normalized volume flow rate at 0 MPa applied pneumatic pressure, $\Delta\pi$ is the osmotic potential gradient between the nutrient solution and the exuded xylem sap acting as the sole driving force, and σ_{sr} is the reflection coefficient of nutrient ions in the solution. The latter was previously measured to be 0.47 for the same poplar clone and nutrient solution with whole individual adventitious roots using a pressure probe (Grünhofer, Guo, et al., 2021). The surface area of roots was determined by scanning them after staining with 0.03% (w/v) toluidine blue O (Merck, Germany) and multiplication by π to account for their round geometry. In contrast, the Lp_r(HY) was estimated based on the slopes of the normalized volume flow with added pneumatic pressure (P) between 0.2 and 0.4 MPa. In this region, J_{vr} was linear when plotted against P. Considering the combined osmotic and hydrostatic driving forces allowed the calculation of the hydrostatic hydraulic conductivity as follows: $Lp_r(HY) = J_{vr}\,(P + \Delta\pi\,\sigma_{sr})^{-1}$.

To determine the apoplastic bypass flow, we collected xylem sap by steadily applying a pneumatic pressure of 0.2 MPa to the chamber. The concentration of the apoplastic tracer PTS in the exuded xylem sap was measured using a fluorescence spectrophotometer (Cary Eclipse, Agilent Technologies; excitation wavelength 404 nm and emission wavelength 512 nm at 800 V PMT voltage). Calibration was achieved with a linear 10-point calibration curve in the range of 2×10^{-9} to 1×10^{-6} mol l^{-1}. Apoplastic bypass flow was then estimated by relating the measured PTS concentration of the xylem sap to the concentration in the root medium.

2.7 | RNA sequencing and analysis

Similar to the physiological investigations of transport, only the intermediate stress conditions of PEG8000 (−0.6 MPa) and NaCl (120 mM) were analyzed by RNA sequencing. It was previously shown with barley that the patchy suberized root zone (25%–50% relative distance in this study) reacted most significantly to the imposed osmotic stress conditions of −0.8 MPa (Kreszies et al., 2019). However, in our study, pronounced developmental effects were observed especially at 10%–27.5% distance from the root tip. Thus, we harvested a root zone of 20%–50% to fit both observations. Five adventitious roots were pooled together in 2 ml Eppendorf tubes, frozen in liquid nitrogen, and milled (Retsch MM400, Retsch GmbH) with steel beads. RNA was then isolated according to the manufacturer's instructions using the RNeasyPlus Universal Mini Kit (Qiagen) and sequenced on an Illumina HiSeq X Ten platform (BGI Tech Solutions, China) using a PE 150 protocol. Raw sequencing data are deposited at the sequence read archive, SRP354749.

Quality assessment of the read libraries was performed using FastQC v0.11.8 (Andrews, 2010). Trimmomatic version 0.38 (Bolger et al., 2014) was used to remove low-quality reads and remaining adapter sequences from each read dataset. Specifically, a sliding window approach was used, in which a read was clipped if the average quality in a window of 4 bp fell below a phred quality score of 20. Only reads with a length of ≥40 bp were retained for further analyses. The splice-aware STAR aligner version 2.7.3a (Dobin et al., 2013) was used to align the remaining reads against a genome index of the reference sequence and annotation of genotype 84K (Qiu et al., 2019). Multimapping reads that mapped to more than one position were excluded from subsequent steps by considering only reads mapping a single location. On average, 35 million reads per sample aligned to unique positions in the 84K reference genome gene set composed of 85,755 predicted gene models. The aligned paired-end reads were ordered according to their position and transformed to .bam files with the software samtools version 1.9 (Li et al., 2009). We employed featureCounts version 1.6.4 (Liao et al., 2014) to obtain aggregate counts of aligned reads at the exonlevel and to construct gene- and transcript-level matrices of these counts comprising all samples. Differentially expressed features were identified with the package "edgeR" version 3.26.4 (Robinson et al., 2010) using the R programming language (R Core Team, 2013). Differential expression analysis was based on comparing genes and transcripts between control, PEG8000 (−0.6 MPa), and NaCl (120 mM) treatments. Expression counts were normalized with library size using the trimmed mean of M-values method of edgeR. Expression levels of genes and transcripts were estimated by a generalized linear model using the negative binomial (NB) distribution to model the read counts for each gene in each sample. By computing dispersion estimates of the NB distribution, we controlled for variability between biological replicates. We used the glmQLFTest approach to account for uncertainty in these dispersion estimates. Only features passing combined thresholds of a false discovery rate (FDR) < 0.05 (Benjamini-Hochberg) and a $|\log_2$ fold change$| \geq 1$ were considered as differentially expressed genes (DEGs).

We used the Mercator online annotation tool (https://plabipd.de/portal/mercator4) to generate MapMan mappings for the protein set of 84K and visualized transcript level expression changes of pathways between treatments with MapMan4 (Schwacke et al., 2019). In addition to this untargeted approach, we performed an extensive screening of the available *Arabidopsis* and *Populus* literature to acquire a set of known and interesting genes primarily related to osmotic and salt stress reactions. We further performed a de novo functional annotation of the 84K gene models with human-readable descriptions, including GO terms using AHRD (https://github.com/groupschoof/AHRD).

Old poplar gene identifiers "POPTR…" were converted into current identifiers "Potri…" (v3; v4.1) using PopGenIE (Sjödin et al., 2009). Gene names without gene identifiers suggested in the literature were searched in Phytozome (Goodstein et al., 2012) and Ensemble Plants (Bolser et al., 2016).

2.8 | Phylogenetic analysis

Orthofinder version 2.4.0 (Emms & Kelly, 2019) was used to infer orthologous groups ("Orthogroups") between the longest representative

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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

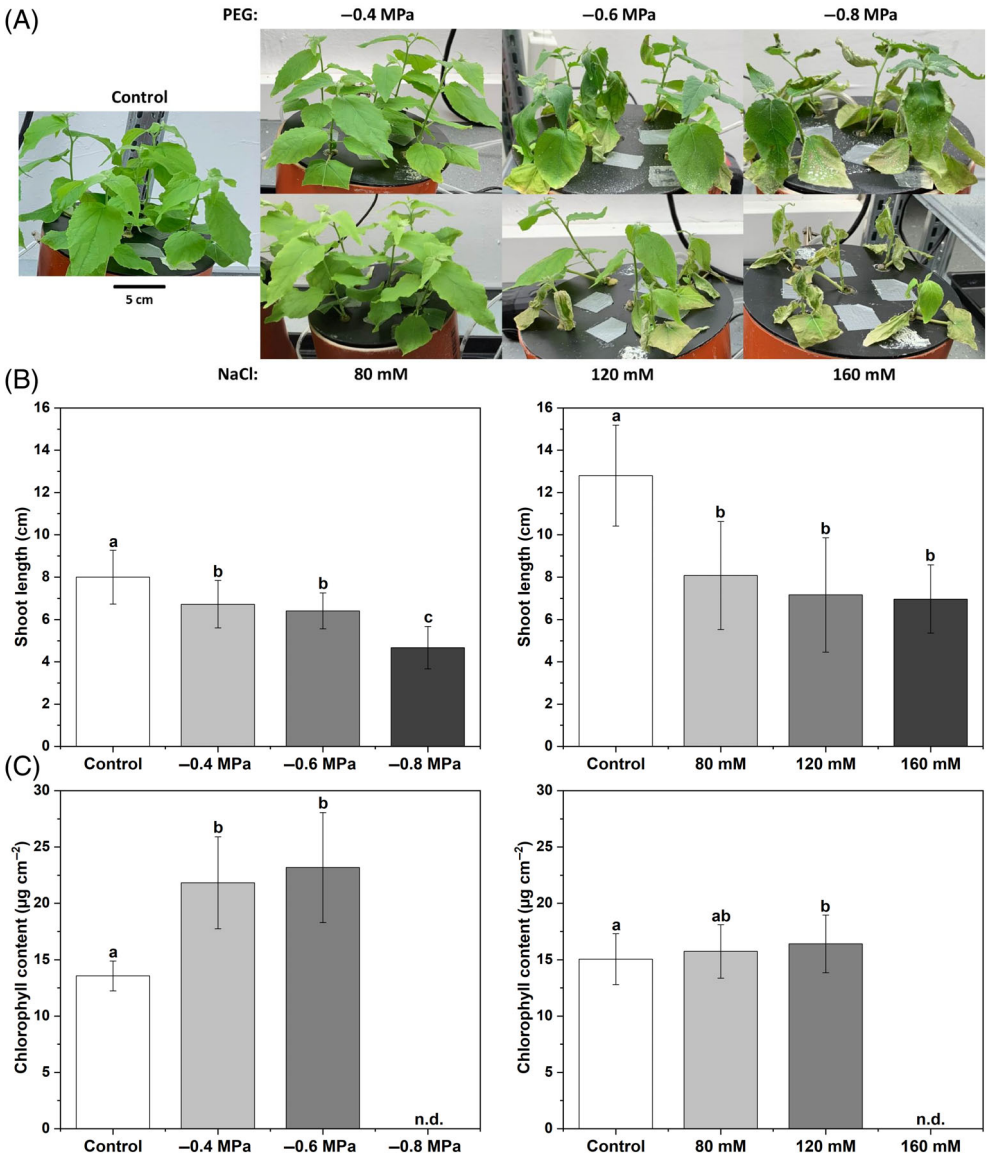


FIGURE 1 Phenotypic characterization of hydroponically cultivated poplar shoots subjected to different osmotic and salt stress intensities. (A) Photographs taken at the day of harvest, photographs of the polyethylene glycol series (−0.4, −0.6, and − 0.8 MPa) originate from Grünhofer, Guo, et al. (2021). (B) Shoot lengths of plants cultivated for 5 weeks, including 1 week of stress treatments (left, osmotic stress; right, salt stress); $n = 15–20$ shoots. (C) Chlorophyll contents measured on three leaves per shoot on the day of harvest (left, osmotic stress; right, salt stress); $n = 40–60$ leaves per treatment; n.d. = not determined. Means with standard deviations are shown. Differential letters indicate significant differences (ANOVA) at $p < 0.05$.

isoforms of the proteomes of *P.* × *canescens* genotype 84K (Qiu et al., 2019), *P.* × *canescens* genotype INRA 717-1B4 (sPta717 v1.1) (Xue et al., 2015), *Populus trichocarpa* (v4.1 Phytozome) (Goodstein et al., 2012), *Arabidopsis thaliana* (TAIR10/Araport11) (Cheng et al., 2017) and *Hordeum vulgare* (IBSC v2.0) (Mascher et al., 2017).

2.9 | Statistical analysis

At least three or more biological replicates were investigated in each experiment. OriginPro 20 (OriginLab Corporation) was used for the statistical analysis of data. After evaluation of normal distribution, two-sample *t* test (indicated by asterisks) or one-way ANOVA with Fisher's LSD post hoc test (indicated by differential letters) were performed for all experiments. Most statistical tests were carried out at a significance level of $p < 0.05$. Due to the frequently reported very high variability of hydraulic conductivity measurements (Kreszies et al., 2019; Miyamoto et al., 2001; Ranathunge, Lin, et al., 2011; Steudle & Meshcheryakov, 1996; Zimmermann et al., 2000), stricter testing of significance was performed at $p < 0.01$ only for these experiments. To visualize the data, means with standard deviations are shown.

3 | RESULTS

3.1 | Plant physiology in reaction to stress

After 1 week of stress exposure, shoots of the lowest stress intensity (−0.4 MPa, 80 mM) showed no visible phenotypic effects compared to the control (Figure 1A). In contrast, shoots of the intermediate (−0.6 MPa, 120 mM) and highest (−0.8 MPa, 160 mM) stress intensity suffered from considerable damage (Figure 1A). Effects of the salinity treatment were more pronounced than that of the corresponding osmotic stress treatment, especially in older and larger leaves. Shoot lengths of all treatments decreased with increasing

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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

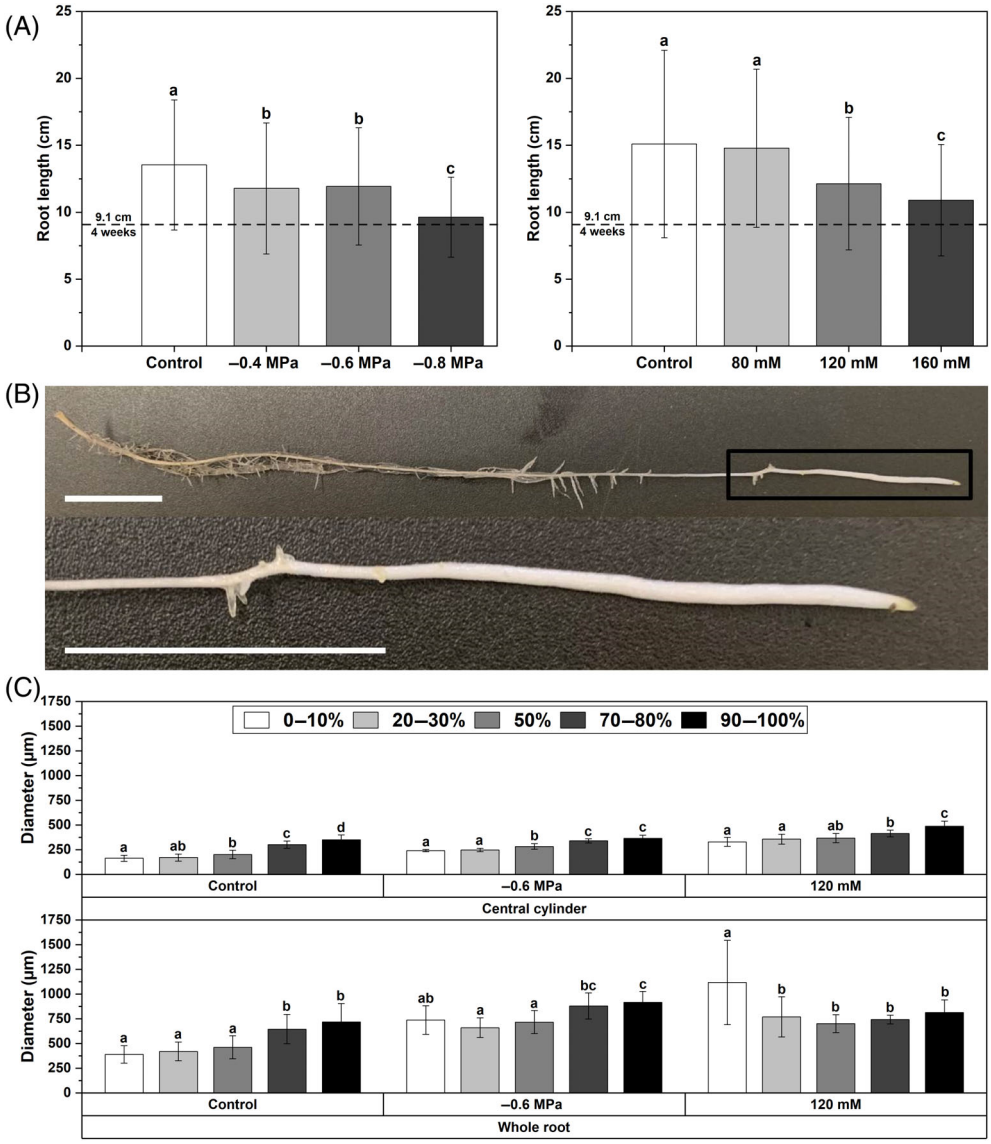


FIGURE 2 Phenotypic characterization of hydroponically cultivated poplar roots subjected to different osmotic and salt stress intensities. (A) Root lengths of plants cultivated for 5 weeks, including 1 week of stress treatments (left, osmotic stress; right, salt stress), the reference line indicates the average root length after 4 weeks of growth. Roots <5 cm were excluded from the analysis; $n = 182–248$ roots per treatment. (B) Representative photographs of the root tip swelling phenotype taken shortly after harvest, scales bars = 2 cm. (C) Central cylinder and whole root diameters were estimated by microscopy; $n = 6$ roots per treatment. Means with SDs are shown. Differential letters indicate significant differences (ANOVA) at $p < 0.05$.

stress intensity (Figure 1B). Only osmotic and not salt stress significantly increased the chlorophyll content of leaves by almost twofold (Figure 1C). As leaves of the highest stress intensity generally suffered from severe tissue damage, the chlorophyll contents were not determined.

In addition to the shoot lengths, also the root lengths decreased with increasing stress intensity in both treatment groups (Figure 2A). Roots subjected to the highest stress intensities were hardly growing at all within the fifth week. They increased in length by only 6% (−0.8 MPa) and 20% (160 mM) if compared to the mean root length measured after 4 weeks of cultivation, the exact time point of stress application.

A phenotypic alteration, never observed in control conditions, was root tip swelling as a reaction to all imposed stresses (−0.4, −0.6, −0.8 MPa and 80, 120, 160 mM; Figure 2B). Swollen root tips (occurring only in 0%–30% relative root length) were found more often with increasing stress intensity and also more often in the NaCl than in the PEG8000 treatment. This phenomenon was studied in more detail in the intermediate stress conditions, where about 70% of roots ($n = 116$ individual roots) showed a swelling reaction. When investigating the whole root as well as the central cylinder diameter, it was clear that the tissue responsible for the swelling reaction of the root tip was the cortex rather than any tissue layer of the central cylinder (see especially 0%–10% and 20%–30% relative root length; Figure 3C).

3.2 | Histochemical observation of Casparian bands and suberin lamellae

No pronounced differences were observed in the development of endodermal Casparian bands upon salt or osmotic stress treatment (no data shown), which were previously identified to be already fully developed at 10%–20% relative root length after cultivation in control conditions (Grünhofer, Guo, et al., 2021).

In contrast, the onset of endodermal suberization (at 27.5% in control conditions) shifted toward the root tip with increasing osmotic

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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

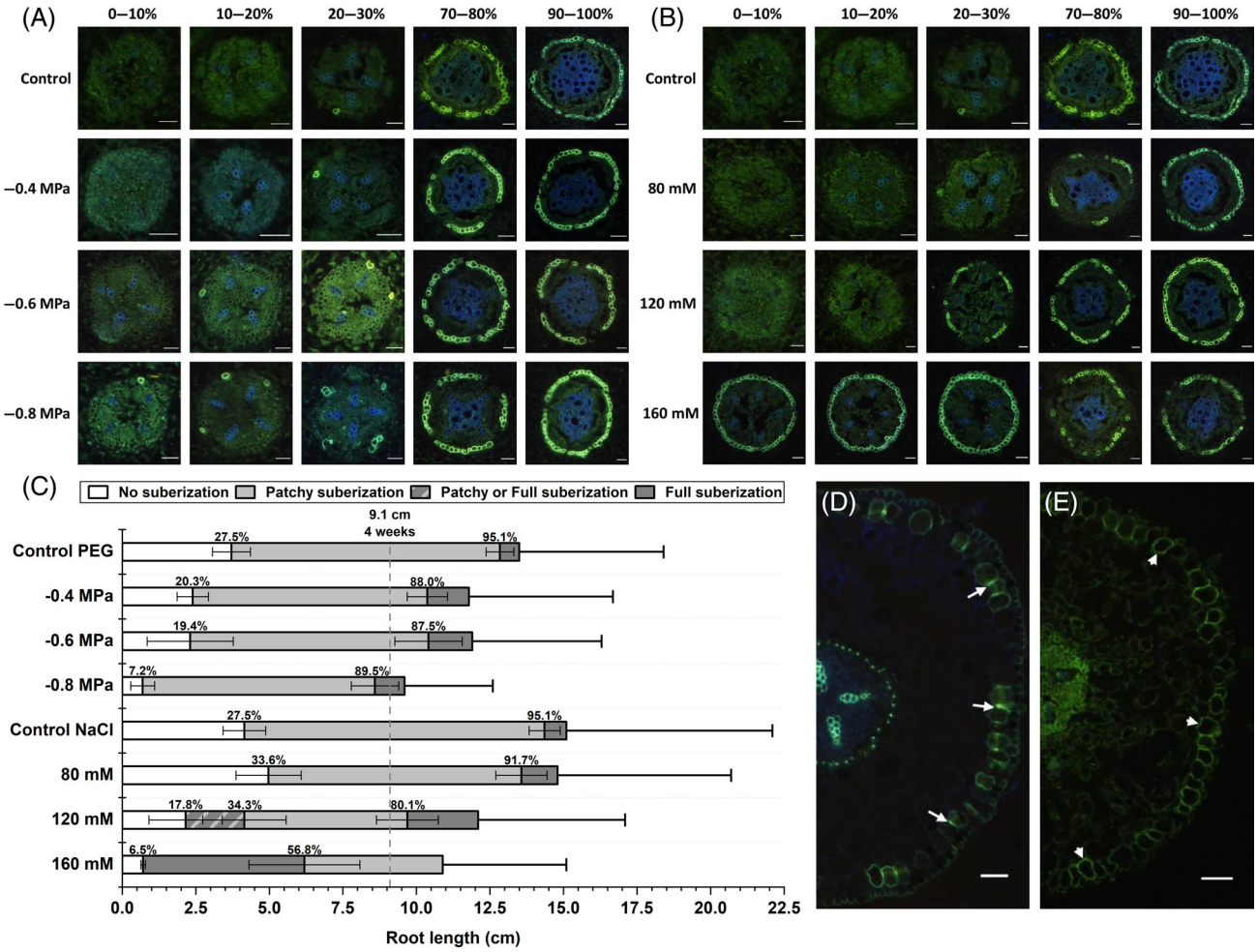


FIGURE 3 Results of the histochemical analysis of hydroponically cultivated poplar roots subjected to different osmotic and salt stress intensities. Endodermal suberization after osmotic (A) or salt (B) stress treatment stained with fluorol yellow 088. The relative distances from the root tip (0%) to the root base (100%) are given. A functional zone of no suberization (zone a, 0%–27.5%) and patchy suberization (zone B, 27.5%–100%) was defined based on observations in control conditions, scale bars = 50 μm. (C) Different endodermal suberization after osmotic (top) and salt (bottom) stress treatment. Of special interest are the noncontinuous suberization patterns of the 120 and 160 mM salt treatment. n = at least six roots per treatment. Means with standard deviations are shown. (D,E) Representative photographs of hypodermal Casparian band (arrows), stained with berberine-aniline blue (D), and hypodermal suberin lamellae (arrowheads) development (E) observed in a few roots in response to osmotic and salt stress. These features were found more frequently with increasing stress intensity and also more often in the salinity treatments if equal osmotic potentials are compared, scale bars = 100 μm.

stress intensity, but no pronounced effects could be observed in the basal parts of the root (Figure 3A). This was similar for the salinity treatments, where it was also the root tip and not the basal root zone that showed significant increases in suberization with increasing stress intensity (Figure 3B). In the case of 120 and 160 mM NaCl, a full endodermal suberization could already be observed in the apical half of some roots, which was again followed by a patchy suberization in the basal half (Figure 3C). As this noncontinuous and exclusive full root tip suberization pattern of the endodermis has, to the best of the author's knowledge, not been described before, this observation was confirmed with 10 further roots of the 160 mM NaCl treatment originating from five independent replicates.

Distinct reactions to the imposed stress conditions were found not only in the endodermis. With increasing stress intensity in both types of stress, Casparian band and suberin lamellae development were also observed in the hypodermis of a few roots investigated (Figure 3D,E). However, this was also strictly limited to the root tip and never observed in root segments more basal than 30% relative distance. Similar to the root swelling phenotype (Figure 2B), this feature was found more frequently with increasing stress intensity and also more often in the salinity treatments if equal osmotic potentials are compared. Nonetheless, it is important to mention that root swelling and hypodermal modifications did not necessarily coincide in the same root.

3.3 | Chemical analysis of root suberin

When the root dry weights after enzymatic and chloroform: methanol treatment (leaving behind only polymerized and nonextractable cell wall components) were related to the root lengths, significant

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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

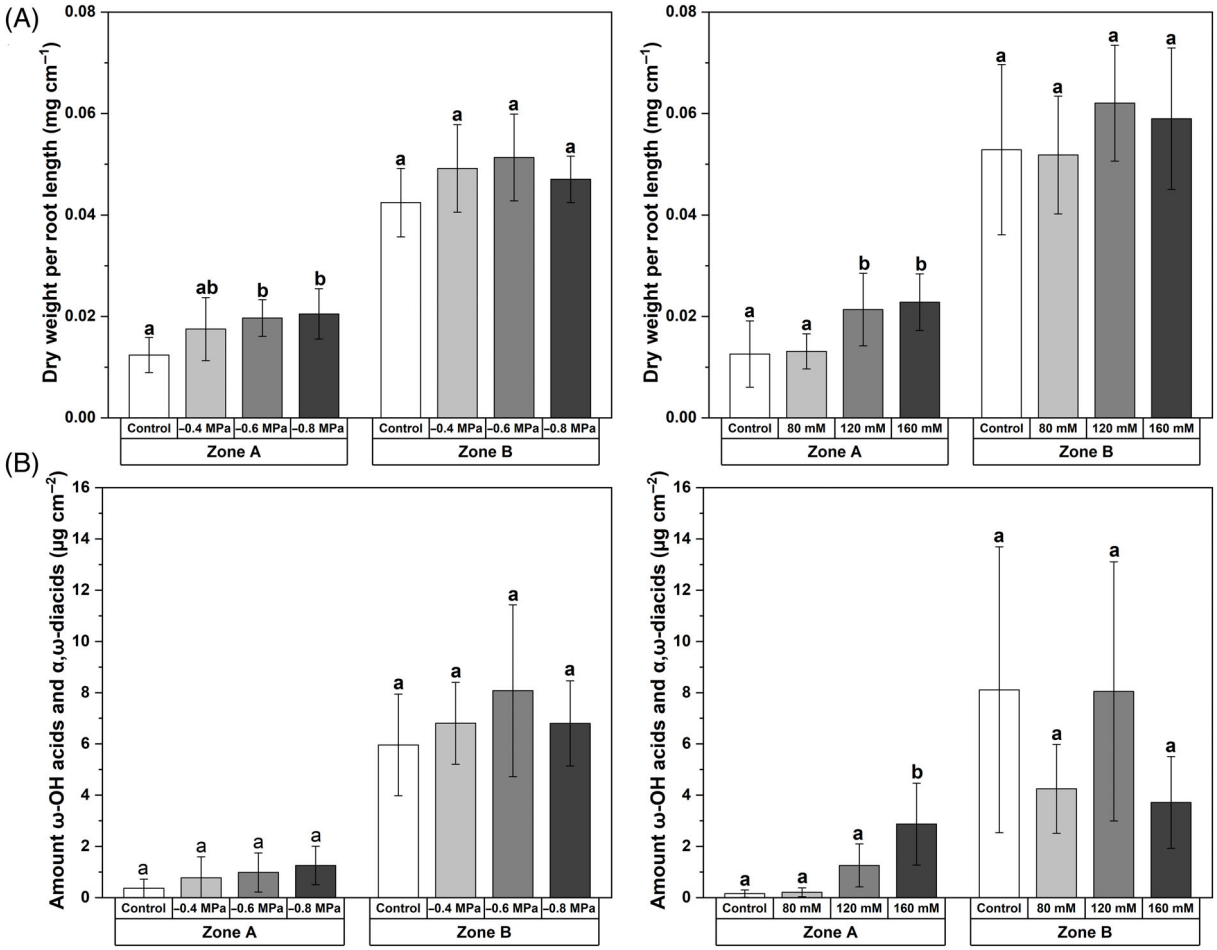


FIGURE 4 Results of the chemical analysis of hydroponically cultivated poplar roots subjected to different osmotic and salt stress intensities. (A) Dry-weights of polymerized and nonextractable cell wall components (left, osmotic stress; right, salt stress) measured after enzymatic and chloroform:methanol treatments. (B) Sums of the suberin diagnostic functional groups ω-hydroxy acids (ω-OH acids) and α,ω-dicarboxylic acids (α,ω-diacids) divided into functional root zones A and B are shown (left, osmotic stress; right, salt stress), suberin amounts were related to the endodermal surface area; $n = 5$ replicates per treatment. Means with standard deviations are shown. Differential letters indicate significant differences (ANOVA) at $p < 0.05$.

increases were only observed in zone A in either treatment group (Figure 4A). This increase was most pronounced in the intermediate and high stress intensities.

The results of the aliphatic suberin analysis by GC/MS fit very well to the suberization patterns observed in microscopy. The suberin diagnostic functional groups of ω-hydroxy acids and α,ω-dicarboxylic acids, related to the endodermal surface area, showed a trend of constantly increasing aliphatic suberin amounts with increasing stress intensities only in the apical zone A (Figure 4B). This trend was statistically confirmed in the 160 mM salt treatment. In contrast, suberin amounts of zone B showed a high standard deviation but no significant differences. In addition, a more detailed investigation of all aliphatic functional groups (Figure S1) as well as the chain-length distribution of suberin monomers (Figure S2) revealed considerable differences only in zone A. Relative proportions of functional groups and chain-lengths of the intermediate (−0.6 MPa, 120 mM) and high (−0.8 MPa, 160 mM) stress intensities in zone A were approaching the composition of zone B if compared to the control and low intensity (−0.4 MPa, 80 mM) stress treatments. Other than that, the relative constitution of zone B was highly stable, irrespective of imposed stress conditions.

In contrast to the moderate increases of aliphatic suberin in zone A, which due to the sporadic formation of an exodermis may not exclusively be attributed to endodermal suberization, no clear trend of the aromatic suberin fraction (only composed of cis- and trans-ferulic acid) could be observed in either zone or treatment (Figure S3A). If absolute aliphatic and aromatic suberin amounts are compared, the aromatic contribution to the suberin polyester appears negligible. Similarly, no clear trend was found for co-extracted benzoic acid derivatives (Figure S3B).

3.4 | Transport physiology

Based on these combined findings, we focused on the intermediate stress conditions (−0.6 MPa, 120 mM) for the following experiments (transport physiology and RNA sequencing), as they represented a

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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

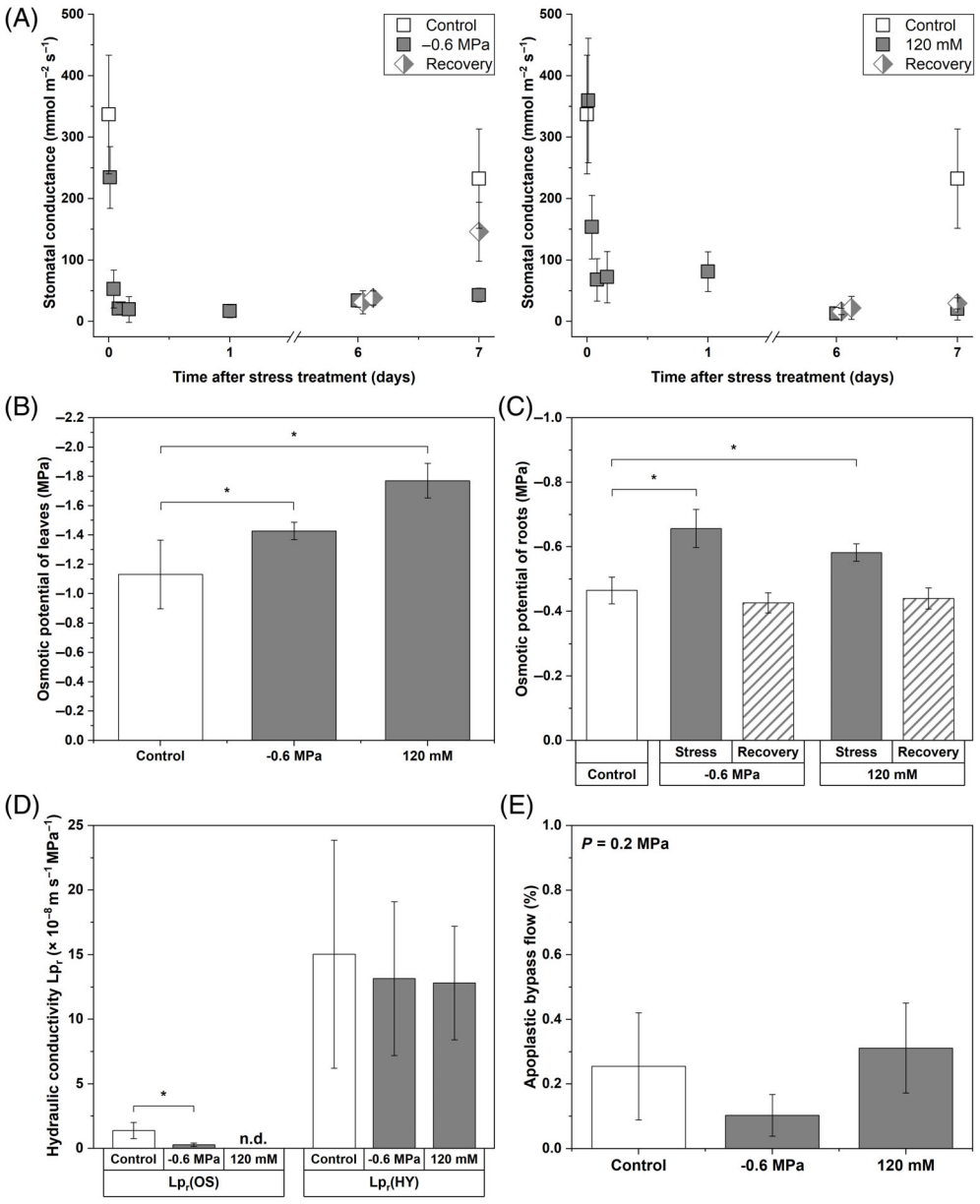


FIGURE 5 Transport physiological investigations of hydroponically cultivated poplar leaves and root systems subjected to different osmotic and salt stress intensities. Only the selected intermediate stress intensities of −0.6 Mpa osmotic stress and 120 mM salt stress were evaluated. (A) Stomatal conductances measured throughout the 1 week of stress treatment during the fifth week of cultivation; $n = 4–18$ leaves per treatment (B) osmotic potentials of leaves ($n = 3–7$ leaves per treatment) and (C) osmotic potentials of roots ($n = 5–7$ roots per treatment) were measured at the day of harvest. (D) Hydraulic conductivity and (E) apoplastic PTS bypass flow of whole recovered root systems were measured using the pressure chamber; recovery = stressed root systems transferred back into normal ½ Hoagland solution 12 h before measurements; $n = 8–10$ (D) and 7–12 (E) root systems per treatment; n. d. = not determined. Means with standard deviations are shown. Asterisks indicate significant differences (*t* test) at $p < 0.05$ (B,C) or $p < 0.01$ (D,E).

good compromise between sufficient stress intensity, observable phenotypic stress reactions, and yet remaining plant vitality.

To investigate the reversibility of certain short-term stress adjustments within 12 h before transport physiological experiments, the stomatal conductances of leaves, as well as the osmotic potentials of leaf and root tissues, were monitored. In both treatments, the stomatal conductance decreased steeply already 15 min after stress exposure (Figure 5A). When stressed plants were put back into normal ½ Hoagland solution for recovery on the sixth day, the stomatal conductances of plants exposed to the osmotic stress, but not to the salt stress treatment, started to increase again within 12 h. This coincided with a significantly higher decrease in osmotic potential of salt-stressed leaves at harvest (Figure 5B). Further beneficial effects of the recovery treatment were observed in the osmotic potentials of roots (Figure 5C). Initially, decreased osmotic potentials due to both osmotic and salt stress reached control levels after only 12 h of recovery.

Osmotic stress significantly decreased the osmotic hydraulic conductivity by 80% (Figure 5D). But despite the 12 h recovery, no Lp_r(OS) could be determined after exposure to salt. This was only possible for the Lp_r(HY) determination with applied pneumatic pressure, where no significant differences induced by either stress condition could be observed. Also, the apoplastic bypass flow was not significantly altered (Figure 5E). Nonetheless, considerable differences were identified for the osmotic potentials of exuded xylem sap (Figure S4). Here, the dilution and even filtration of exudate with increasing pneumatic pressures were clearly visible.

3.5 | RNA sequencing and analysis

A separated clustering of all osmotic and salt stress samples was observed when compared to the control treatment, indicating shared

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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

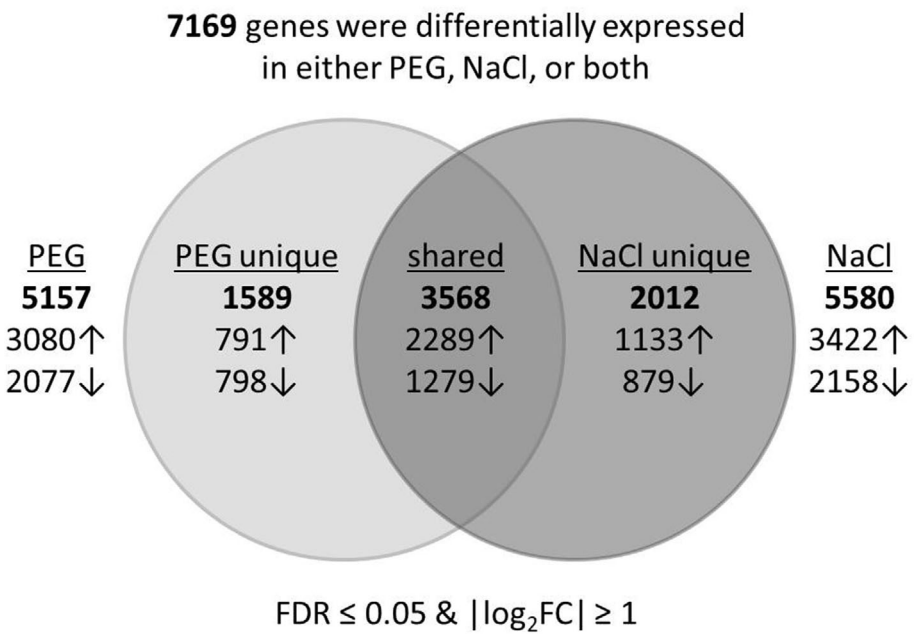


FIGURE 6 Results of the differential expression analysis of hydroponically cultivated poplar roots subjected to different osmotic and salt stress intensities. Only the selected intermediate treatments of −0.6 MPa PEG and 120 mM NaCl were investigated and related to the control treatment. Unique = significant genes (false discovery rate < 0.05) were differentially expressed ($|\log_2 FC| \geq 1$) exclusively in the respective stress treatment; shared = significant genes were differentially expressed in both stress treatments; $n = 3$ replicates per treatment

similarities between the stress treatments (Figure S5). The transcriptome of the 84K root zone (20–50% relative root length) consisted of 46,107 expressed genes, covering ~54% of the 85,755 gene models of the 84K genome. A set of 7169 genes was differentially expressed (FDR < 0.05 and $|\log_2 FC| \geq 1$) either uniquely in osmotic stress (1589), uniquely in salt stress (2012), or shared (3568) in both treatments (Figure 6). Of these 7169 DEGs, 59% were up- and 41% were downregulated upon stress exposure. The shared DEGs identified to be upregulated (2289, 64%) or downregulated (1279, 36%) were always regulated in the same direction (up or down) and often also in similar $\log_2 FC$ intensity (Table 1; see also Tables S1–S7 or Table S8). Using MapMan4, it was possible to observe that a lot of the downregulated genes belonged especially to the bin "9 Secondary metabolism," "13 Cell division," "17 Protein biosynthesis," "20 Cytoskeleton organization," and also partly to "24 Solute transport." In contrast, many genes associated with the bins "1 Photosynthesis," "5 Lipid metabolism," "9 Secondary metabolism," "10 Redox homeostasis," "18 Protein modification," "19 Protein homeostasis," "21 Cell wall organization," and also partly with "24 Solute transport" were significantly upregulated. When the uniquely expressed genes of either treatment condition are compared to the shared set of DEGs, these generally followed the same trends.

In agreement with the previously described *Populus* genome duplication (Tuskan et al., 2006) and potential gene number increases due to hybridization, every *A. thaliana* gene may have up to two orthologs in *P. trichocarpa* and up to four orthologs in the hybrid *P.* × *canescens*. Especially many genes described to belong to the secondary metabolism and cell wall organization pathways of suberin, phenylpropanoid, cutin, and cuticular wax biosynthesis were significantly upregulated (Table 2). This was different for Casparian band and aquaporin genes, whose expression was not significantly altered in most cases (Tables S2 and S3). Instead, many genes contributing to osmoprotection and the ROS (reactive oxygen species) scavenging system were significantly upregulated in response to both types of stress (Tables S4 and S5). The core salt-overly-sensitive signaling pathway (SOS1, SOS2, and SOS3) was hardly significantly induced after 1 week of either PEG8000 or NaCl stress, whereas many other genes especially associated with drought and salt stress reactions (including transcription factors of the WRKY, MYB, NAC, ERF, and HSF family) showed significant upregulation (Tables S6 and S7).

Results generated in this study regarding phylogenetic relationships, annotations, expression results, associated literature of known genes of interest as well as the pathway analysis were carefully joined to generate one-stop reference tables (Tables 1, 2, and S1–S10). This combined view of the data represents a novel resource for poplar researchers, ready to be mined for further studies.

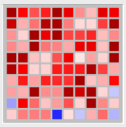
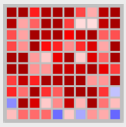
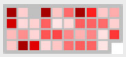
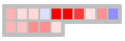
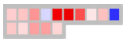
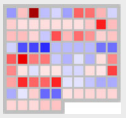
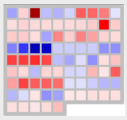
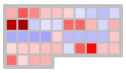
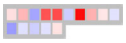
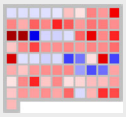
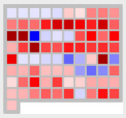
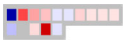
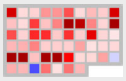
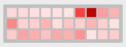
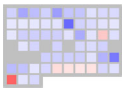
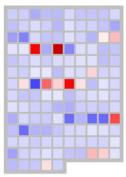
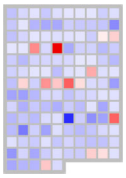
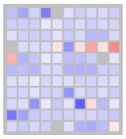
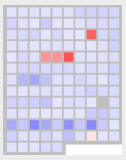
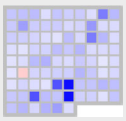
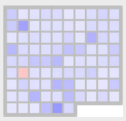
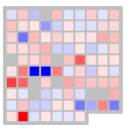
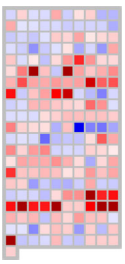
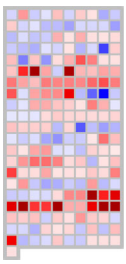
4 | DISCUSSION

4.1 | Plant physiology in reaction to stress

All phenotypic observations of shoot organs (Figure 1), considering both types of imposed stress conditions (osmotic and salt stress), have previously been described (Bolu & Polle, 2004; Liu et al., 2019; Syaiful et al., 2014). This includes the shoot collapse after stress application, reduced shoot growth, and increased chlorophyll contents after exposure to PEG8000 but not NaCl. Comparable to our results, the vitality of a *P. alba* × *P. tomentosa* hybrid also significantly declined with osmotic potentials and salt concentrations being around −0.8 MPa and 150 mM, respectively (Watanabe et al., 2000). Although initial reactions to both stress types are highly similar and mainly caused by water relations, salinity treatment is known to provoke more severe long-term effects. This is due to NaCl accumulation in older and larger leaves if compared to its matched osmotic stress treatment (Munns, 2002), as was also reflected in the osmotic potentials of leaves (Figure 5B) in this study.

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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

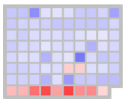
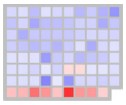
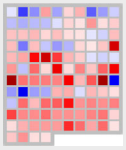
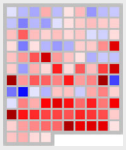
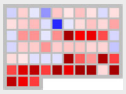
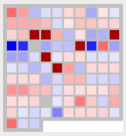
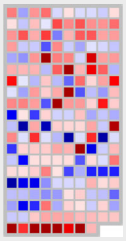
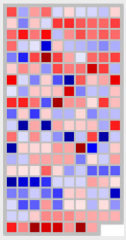
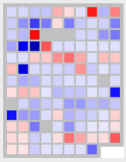
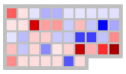
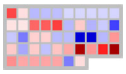
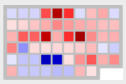
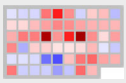
TABLE 1 Differentially expressed genes visualized using MapMan4

MapMan bin	PEG unique	PEG shared	NaCl shared	NaCl unique
log_2FC				
1. Photosynthesis				
2. Cellular respiration				
3. Carbohydrate metabolism				
4. Amino acid metabolism				
5. Lipid metabolism				
9. Secondary metabolism				
10. Redox homeostasis				
13. Cell division				
17. Protein biosynthesis				
18. Protein modification				
19. Protein homeostasis				

(Continues)

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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 (Continued)

MapMan bin	PEG unique	PEG shared	NaCl shared	NaCl unique
20. Cytoskeleton organization				
21. Cell wall organization				
22. Vesicle trafficking				
24. Solute transport				
25. Nutrient uptake				
26. External stimuli				

Note: Polyethylene glycol (PEG) (−0.6 MPa) and NaCl (120 mM) unique genes were differentially expressed (false discovery rate <0.05 and $|\log_2 FC| \geq 1$) exclusively in the corresponding treatment, whereas shared genes were differentially expressed in both stress conditions. Only a selection of available MapMan bins is shown.

However, besides a decrease in root length, often reported in response to many different abiotic stress factors (Grünhofer, Schreiber, & Kreszies, 2021), the root tip swelling phenotype observed in this study (Figure 2) is generally less investigated. The available literature suggests that root tip swelling can be caused by increased starch deposition (Zawaski et al., 2012), which might be linked to stress-responsive carbohydrate relocation. Increased sugar breakdown in swollen root tips could assist in the accumulation of osmotically active solutes. Alternatively, root tip swelling was suggested to be caused by stress-induced hydrogen peroxide production and increased activity of the ROS scavenging system upon stress encounter (Mahmood et al., 2016; Sun et al., 2018; Tamás et al., 2012). Trees appear to utilize complex interactions of osmoprotective coping mechanisms to protect their cell integrity and withstand osmotically active stress conditions. This includes the accumulation of proline, soluble sugars, or even sodium as a metabolically cheap osmolyte (Dluzniewska et al., 2007; Janz et al., 2012; Ottow et al., 2005; Pallara et al., 2012; Watanabe et al., 2000). To increase the content of osmotically active oligosaccharides, accumulated starch needs to be broken down by increased amylase activity (Doyle et al., 2007). In addition, dehydrins and osmotins were found to be crucial for plant performance in primary stress adaptation and following homeostasis, respectively (Brinker et al., 2010). All of the above fits very well with our observation. Proline synthesis genes were not significantly increased, but proline catabolism gene expression was decreased (Table S4). In parallel, the transcription of amylases and oligosaccharide synthases was increased (Table S4), which were previously identified to play potential roles not only in osmoprotection but also ROS scavenging (Cohen et al., 2010; Nishizawa et al., 2008).

Similarly, many genes associated with the antioxidant system also identified in comparable osmotic or salt stress transcription studies with poplar (Brosché et al., 2005; Cohen et al., 2010; Janz et al., 2012; Ma et al., 2013; Zheng et al., 2015), were significantly upregulated in our study (Table S5). Thus, based on the transcriptomics data, both stress adaptation processes (osmotic adaptation and/or ROS scavenging) could be responsible for the root tip swelling phenotype. Nonetheless, other potentially contributing factors, for example, cell wall mechanics and cell turgor behavior, cannot be ruled out at this point. Future research including physiological experiments is needed to investigate this remarkable phenotypic modification.

4.2 | Histochemical and chemical analysis

Aside from these hypotheses, a denser cytoplasm in swollen root tips could help establish an effective barrier to solutes (Chen et al., 2011),

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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 Differential expression of genes associated with the suberin, phenylpropanoid, cutin, and cuticular wax biosynthesis

Orthogroup	Description	*P.* × *canescens* Gene ID	log_2FC PEG	log_2FC NaCl	*A. thaliana* Gene ID	*P. trichocarpa* Gene ID	References
PATE FAT FATB	Fatty acyl-ACP thioesterase	Pop_G12G008837	2.6	2.4	NA	Potri.015G099600; Potri.012G101500	Rains et al., 2017; Soler et al., 2007
		Pop_A12G067979	2.2	2.3			
		Pop_A15G049117	1.8	2.5			
		Pop_G15G016747	n.s.	2.0			
LACS1 CER8	Long chain acyl-CoA synthetase	Pop_A02G012579	2.8	2.7	AT2G47240	Potri.002G192400	Fich et al., 2016; Li-Beisson et al., 2013; Philippe et al., 2020; Rains et al., 2017
		Pop_G02G024130	2.7	3.1			
LACS2	Long chain acyl-CoA synthetase	Pop_G09G077207	1.9	2.1	AT1G49430	Potri.009G109900	Fich et al., 2016; Lopes et al., 2020; Philippe et al., 2020; Rains et al., 2017; Soler et al., 2007
FAD4	Fatty acid desaturase	Pop_G01G024941	−3.4	−3.3	AT4G27030; AT2G22890; AT1G62190	Potri.001G424700	Rains et al., 2017
FAD8	Fatty acid desaturase	Pop_A06G089411	−1.3	−1.4	AT5G05580; AT3G11170	Potri.006G101500; Potri.010G187800; Potri.008G069600	Rains et al., 2017
		Pop_A08G063468	0.8	n.s.			
		Pop_G08G057951	1.0	n.s.			
		Pop_G10G047669	−1.4	−1.1			
KCS1	3-ketoacyl-CoA synthase	Pop_A02G012441	2.2	3.0	AT1G01120	Potri.014G104300; Potri.002G178000	Lopes et al., 2020; Rains et al., 2017; Todd et al., 1999
		Pop_G02G024284	2.0	3.3			
KCS2/11/20	3-ketoacyl-CoA synthase	Pop_A08G046170	3.0	3.6	AT2G26640; AT5G43760; AT1G04220	Potri.006G249200; Potri.018G032200; Potri.008G160000; Potri.010G079400; Potri.010G079300; Potri.010G080200; Potri.010G080333; Potri.010G079700; Potri.010G079500; Potri.010G080400	Franke et al., 2009; Lee et al., 2009; Lopes et al., 2020; Paul et al., 2006; Rains et al., 2017; Soler et al., 2007; Trenkamp et al., 2004; Ursache et al., 2021
		Pop_A10G047242	n.s.	5.3			
		Pop_A10G047245	2.6	3.7			
		Pop_G08G022033	2.4	3.0			
		Pop_G10G060718	2.3	3.6			
KCS5/6 CER6/60 CUT1	3-ketoacyl-CoA synthase	Pop_A10G069355	n.s.	3.0	AT1G68530; AT1G25450	Potri.008G120300; Potri.010G125300	Fiebig et al., 2000; Hooker et al., 2002; Li-Beisson et al., 2013; Millar et al., 1999; Rains et al., 2017
FAR3-like	Fatty acyl-CoA reductase	Pop_A04G070695	n.s.	3.8	NA	Potri.004G185100; Potri.009G145000	Rains et al., 2017
		Pop_G04G060328	2.4	2.7			
		Pop_A09G076863	n.s.	4.0			
		Pop_G09G051094	n.s.	3.3			
		Pop_G09G051095	n.s.	4.2			

(Continues)

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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 (Continued)

Orthogroup	Description	*P.* × *canescens* Gene ID	log_2FC PEG	log_2FC NaCl	*A. thaliana* Gene ID	*P. trichocarpa* Gene ID	References
FAR3/3-like CER4	Fatty acyl-CoA reductase	Pop_A04G070693	n.s.	2.8	AT4G33790	Potri.004G185000; Potri.009G144900	Li-Beisson et al., 2013; Lopes et al., 2020; Rains et al., 2017
		Pop_G09G051093	n.s.	2.8			
		Pop_A09G076864	n.s.	2.8			
CYP72A7/A14/A14-like/A15/A59	Cytochrome P450 monooxygenase	Pop_A11G021606	1.2	1.5	AT3G14620; AT3G14610; AT3G14630; AT3G14680; AT3G14660; AT3G14650; AT3G14690; AT3G14640; AT1G17060	Potri.011G101500; Potri.011G101250; Potri.011G099800; Potri.011G101401; Potri.011G099400; Potri.011G101700; Potri.011G099200; Potri.011G098800	Lopes et al., 2020; Soler et al., 2007
		Pop_A11G021607	1.3	1.6			
		Pop_A11G021610	1.9	2.0			
		Pop_G11G077839	1.0	1.2			
CYP75B1	Flavonoid 3-hydroxylase	Pop_G09G031627	−1.6	n.s.	NA	Potri.001G274600; Potri.009G069100	Rains et al., 2017
CYP86A1	Fatty acid ω-hydroxylase	Pop_A01G056971	2.5	3.2	AT5G58860	Potri.001G249700; Potri.009G043700	Benveniste et al., 1998; Höfer et al., 2008; Li et al., 2007; Lopes et al., 2020; Molina et al., 2009; Rains et al., 2017; Soler et al., 2007; Ursache et al., 2021
		Pop_A09G014397	n.s.	2.6			
		Pop_G01G020985	2.6	3.5			
		Pop_G09G014223	n.s.	2.8			
CYP86A2/A4/A7/A8	Fatty acid ω-hydroxylase	Pop_A03G035534	n.s.	4.1	AT1G63710; AT4G00360; AT1G01600; AT2G45970	Potri.003G129100; Potri.014G085800	Fich et al., 2016; Philippe et al., 2020; Rains et al., 2017
		Pop_A14G044766	n.s.	2.1			
		Pop_G03G050618	n.s.	3.9			
		Pop_G14G023567	3.1	3.1			
CYP86B1/B2	Fatty acid ω-hydroxylase	Pop_A05G011576	2.1	2.4	AT5G08250; AT5G23190	Potri.005G092200; Potri.007G072100	Compagnon et al., 2009; Lopes et al., 2020; Molina et al., 2009; Rains et al., 2017; Soler et al., 2007
		Pop_A07G006292	2.2	3.0			
		Pop_G05G073588	2.1	2.3			
		Pop_UnG079503	2.3	3.3			
CYP94B1/B3	Cytochrome P450 monooxygenase	Pop_A12G068047	−2.2	−2.1	AT3G48520; AT5G63450	Potri.012G096800	Krishnamurthy et al., 2021
GPAT1	Glycerol-3-phosphate acyltransferase	Pop_G05G008468	3.6	3.4	AT1G06520	Potri.005G202200	Lopes et al., 2020
GPAT4/8	Glycerol-3-phosphate acyltransferase	Pop_A14G044764	2.2	3.1	AT4G00400; AT1G01610	Potri.014G085500	Li et al., 2007; Rains et al., 2017; Soler et al., 2007
		Pop_G14G023565	2.2	2.9			
GPAT5/7	Glycerol-3-phosphate acyltransferase	Pop_A08G063348	2.1	2.8	AT5G06090; AT3G11325; AT3G11430	Potri.008G058200; Potri.010G201200	Beisson et al., 2007; Lopes et al., 2020; Rains et al., 2017; Soler et al., 2007; Ursache et al., 2021; Yang et al., 2010; Yang et al., 2012
		Pop_A10G046995	2.1	2.6			
		Pop_G08G057812	2.0	3.0			

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts- U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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 (Continued)

Orthogroup	Description	*P.* × *canescens* Gene ID	log_2FC PEG	log_2FC NaCl	*A. thaliana* Gene ID	*P. trichocarpa* Gene ID	References
		Pop_G10G007149	2.2	2.8			
GPAT6	Glycerol-3-phosphate acyltransferase	Pop_A06G089454	2.7	3.9	AT2G38110	Potri.016G113100; Potri.006G097800	Fich et al., 2016; Philippe et al., 2020; Rains et al., 2017
		Pop_A16G090182	n.s.	2.7			
		Pop_G06G051703	n.s.	3.5			
		Pop_G16G068536	2.4	2.7			
4CL1/2	4-coumarate ligase	Pop_G06G053838	1.4	2.0	AT1G51680; AT3G21240	Potri.018G094200; Potri.006G169700; Potri.003G188500; Potri.001G036900	Fraser & Chapple, 2011; Rains et al., 2017; Soler et al., 2007
C4H REF3 CYP73A5	Cinnamate 4-hydroxylase	Pop_A13G031180	n.s.	1.2	AT2G30490	Potri.013G157800; Potri.013G157900; Potri.019G130700	Fraser & Chapple, 2011; Soler et al., 2007
C3H REF8 CYP98A3	Coumaroyal-3-hydroxylase	Pop_G06G089809	n.s.	0.9	AT2G40890	Potri.006G033300; Potri.016G031000; Potri.016G031100	Fraser & Chapple, 2011; Rains et al., 2017
		Pop_G16G025588	1.7	1.9			
PAL1/2	Phenylalanine amonia lyase	Pop_A06G085807	n.s.	1.2	AT2G37040; AT3G53260; AT5G04230; AT3G10340	Potri.006G126800; Potri.016G091100; Potri.008G038200; Potri.010G224200; Potri.010G224100	Fraser & Chapple, 2011; Soler et al., 2007
		Pop_A10G046757	n.s.	1.1			
		Pop_A10G046758	2.5	2.6			
		Pop_G06G051426	n.s.	1.3			
		Pop_G10G007753	1.3	1.6			
HCT	Shikimate O-hydroxycinnamoxyl transferase	Pop_G03G055670	0.9	1.4	AT5G48930	Potri.003G183900; Potri.001G042900	Fraser & Chapple, 2011; Rains et al., 2017
CCR1/2	Cinnamoyl CoA reductase	Pop_G01G075493	3.6	3.1	AT1G80820; AT1G15950	Potri.001G045100; Potri.001G045000; Potri.001G045500; Potri.001G046100; Potri.001G046400; Potri.003G181400	Fraser & Chapple, 2011; Rains et al., 2017; Soler et al., 2007
F5H FAH1 CYP84A1	Ferulate 5-hydroxylase	Pop_A05G073116	n.s.	2.1	AT5G04330; AT4G36220	Potri.007G016400; Potri.005G117500	Fraser & Chapple, 2011; Soler et al., 2007
		Pop_A07G022645	1.5	2.2			
		Pop_G05G017772	n.s.	1.4			
		Pop_G07G062561	n.s.	1.5			
COMT1	Caffeic acid/5-hydroxyferulic acid O-methyltransferase	Pop_A12G067093	n.s.	1.5	AT5G54160	Potri.015G003100; Potri.012G006400	Fraser & Chapple, 2011
		Pop_A15G076619	−1.7	n.s.			
		Pop_G12G076215	n.s.	1.3			

(Continues)

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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 (Continued)

Orthogroup	Description	*P.* × *canescens* Gene ID	log_2FC PEG	log_2FC NaCl	*A. thaliana* Gene ID	*P. trichocarpa* Gene ID	References
CAD-C/D	Cinnamyl alcohol dehydrogenase	Pop_A09G033803	n.s.	1.3	AT3G19450; AT4G34230	Potri.009G095800	Fraser & Chapple, 2011
		Pop_G09G077377	n.s.	0.8			
Prx52	Peroxidase	Pop_A13G010041	2.4	1.8	AT5G05340	Potri.013G083600	Rains et al., 2017
		Pop_G13G071450	2.2	1.5			
PER10	Peroxidase	Pop_A09G026771	−1.3	−1.7	AT1G49570	Potri.004G144600; Potri.009G106400	Ursache et al., 2021
		Pop_G04G028151	−3.7	−3.2			
PER11	Peroxidase	Pop_A08G063917	2.4	3.3	AT1G68850	Potri.010G134500; Potri.008G110600	Lopes et al., 2020; Ursache et al., 2021
		Pop_A10G069096	2.1	2.8			
		Pop_G08G058348	n.s.	2.6			
		Pop_G10G048153	2.3	2.7			
PER23	Peroxidase	Pop_A01G029520	n.s.	3.1	AT2G38380; AT2G38390; AT4G08780; AT4G08770; AT3G32980; AT3G49120; AT3G49110	Potri.003G214900; Potri.001G011200; Potri.001G011300; Potri.001G011000; Potri.001G012901; Potri.001G013000	Ursache et al., 2021
		Pop_A03G014623	3.0	4.5			
		Pop_G03G076294	n.s.	5.2			
PER59	Peroxidase	Pop_A03G014627	−2.1	−2.2	AT5G19890	Potri.003G214500	Ursache et al., 2021
		Pop_G03G076289	−1.6	−1.6			
LAC2	Laccase	Pop_G01G013289	−1.9	n.s.	AT2G29130; AT5G60020	Potri.006G087500; Potri.006G087100; Potri.001G184300; Potri.009G156600; Potri.009G156800; Potri.009G034500; Potri.001G054600; Potri.011G120200; Potri.011G120300; Potri.001G401100; Potri.001G401300; Potri.001G341600	Ursache et al., 2021
LAC3/13	Laccase	Pop_A19G055152	−2.3	−1.6	AT2G30210; AT5G07130	Potri.013G152700; Potri.019G124300	Rains et al., 2017; Ursache et al., 2021
LAC12	Laccase	Pop_A08G058002	n.s.	1.2	AT5G05390; AT2G40370	Potri.008G073800; Potri.010G183500; Potri.008G073700; Potri.010G183600	Soler et al., 2007; Ursache et al., 2021
ASFT		Pop_A01G032194	1.8	2.0	AT5G41040	Potri.001G326300	

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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 (Continued)

Orthogroup	Description	*P.* × *canescens* Gene ID	log_2FC PEG	log_2FC NaCl	*A. thaliana* Gene ID	*P. trichocarpa* Gene ID	References
BAHD HCBT	Aliphatic suberin feruloyl transferase	Pop_G01G090317	2.0	2.2			Cheng et al., 2013; Gou et al., 2009; Molina et al., 2009; Rains et al., 2017; Soler et al., 2007; Ursache et al., 2021
BAHD	BAHD acyltransferase	Pop_A10G047527	n.s.	4.1	AT1G24430	Potri.010G054002; Potri.010G053800	Soler et al., 2007
		Pop_G10G027153	n.s.	4.0			
		Pop_G10G027154	−1.7	n.s.			
		Pop_G10G027155	n.s.	3.9			
HXXXD-type acyltransferase	Acyltransferase	Pop_A01G003980	2.1	3.0	AT5G02890	Potri.003G082100; Potri.001G152500	Rains et al., 2017
		Pop_G01G004748	2.7	3.5			
HXXXD-type acyltransferase	Acyltransferase	Pop_A07G062775	n.s.	2.0	AT1G32910; AT5G16410; AT1G78990	Potri.007G003800	Rains et al., 2017
		Pop_G07G063287	n.s.	1.8			
HXXXD-type acyltransferase	Acyltransferase	Pop_A04G026494	n.s.	3.7	AT3G29590; AT5G39090; AT5G39050; AT5G39080; AT3G29670; AT5G61160; AT3G29690; AT3G29680; AT3G29720; AT3G29635; AT3G29636	Potri.009G019700; Potri.004G120700; Potri.004G103300; Potri.004G103200; Potri.004G103100; Potri.001G455150; Potri.004G093800; Potri.004G096300; Potri.004G096425; Potri.004G096132; Potri.004G096066; Potri.017G094700; Potri.001G395300; Potri.001G395700; Potri.001G395400; Potri.001G395900; Potri.010G208100; Potri.004G109300; Potri.004G110400; Potri.004G110640; Potri.004G110700; Potri.004G109800	Rains et al., 2017
		Pop_G01G075120	1.9	2.1			
ABCG1/2/6/16/20	ABC transporter	Pop_A02G012230	2.2	2.9	AT3G55090; AT2G39350; AT2G37360; AT3G53510; AT5G13580	Potri.014G080200; Potri.002G156900; Potri.010G213400; Potri.008G047900	Lopes et al., 2020; Philippe et al., 2020; Rains et al., 2017; Shanmugarajah et al., 2019; Shiono et al., 2014; Soler et al., 2007; Xin & Herburger, 2020; Yadav et al., 2014
		Pop_A10G047178	2.1	2.5			
		Pop_A14G044706	2.1	2.5			
		Pop_G02G030909	1.8	2.3			

(Continues)

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts- U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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 (Continued)

Orthogroup	Description	*P.* × *canescens* Gene ID	log_2FC PEG	log_2FC NaCl	*A. thaliana* Gene ID	*P. trichocarpa* Gene ID	References
		Pop_G10G007316	2.6	2.7			
		Pop_G14G023679	1.9	2.6			
ABCG23	ABC transporter	Pop_A09G028823	2.1	2.5	AT5G19410	Potri.009G070100	Lopes et al., 2020; Rains et al., 2017
		Pop_G09G031639	1.7	2.1			
ABCG32	ABC transporter	Pop_UnG083200	2.0	2.5	AT2G26910	Potri.018G074500	Fich et al., 2016; Philippe et al., 2020; Rains et al., 2017
ABCG40 PDR12	ABC transporter	Pop_A03G019891	2.6	2.6	AT1G15520	Potri.006G248500; Potri.018G032900; Potri.001G175700; Potri.003G183200; Potri.001G048700; Potri.001G048800; Potri.001G049000; Potri.001G048900; Potri.003G178900; Potri.003G179000	Rains et al., 2017
		Pop_A06G063069	n.s.	−2.3			
		Pop_G01G074551	1.5	n.s.			
		Pop_G03G055723	2.8	3.2			
LTP5 ELTP	Lipid transfer protein	Pop_A05G002152	2.3	3.0	AT1G05450; AT2G48140	Potri.002G050200; Potri.005G212000	Rains et al., 2017; Ursache et al., 2021
LTPG2	GPI-anchored lipid transfer protein	Pop_A09G076714	n.s.	3.2	AT2G27130; AT3G43720	Potri.009G158100; Potri.004G196000	Li-Beisson et al., 2013
		Pop_G09G022435	n.s.	2.7			
LTPG5	GPI-anchored lipid transfer protein	Pop_A02G065611	1.0	n.s.	AT3G22600; AT4G14815	Potri.010G085300; Potri.002G050500; Potri.005G211800; Potri.010G085400; Potri.008G155100	Rains et al., 2017
		Pop_A05G002150	1.6	1.6			
		Pop_A08G046226	2.5	2.8			
		Pop_A10G048659	3.3	3.6			
		Pop_A10G048660	8.3	8.0			
		Pop_G02G005233	2.4	2.0			
		Pop_G05G008584	3.3	3.0			
		Pop_G08G022096	1.7	2.0			
		Pop_G10G001585	6.8	7.4			
		Pop_G10G001592	2.6	2.2			
LTPG15	GPI-anchored lipid transfer protein	Pop_G02G005195	2.3	2.5	NA	Potri.002G050300	Rains et al., 2017
GDSL	Esterase/lipase	Pop_G09G026118	n.s.	4.0	AT5G22810	Potri.009G151000	Lopes et al., 2020; Soler et al., 2007
GDSL	Esterase/lipase	Pop_A10G046613	n.s.	2.5	AT5G03590; AT5G03600; AT5G03610;	Potri.001G406500; Potri.010G237000	Lopes et al., 2020
		Pop_A10G046619	1.5	n.s.			

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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 (Continued)

Orthogroup	Description	*P.* × *canescens* Gene ID	log_2FC PEG	log_2FC NaCl	*A. thaliana* Gene ID	*P. trichocarpa* Gene ID	References
		Pop_G10G006166	1.5	1.5	AT3G09930; AT2G36325		
GDSL GELP38	Esterase/lipase	Pop_A12G019390	2.3	3.0	AT1G74460	Potri.012G068700	Lopes et al., 2020; Soler et al., 2007; Ursache et al., 2021
		Pop_G12G068302	2.4	3.2			
GDSL GELP51/73	Esterase/lipase	Pop_A07G022795	2.4	3.3	AT2G23540; AT3G50400	Potri.007G036300	Lopes et al., 2020; Soler et al., 2007; Ursache et al., 2021
		Pop_G07G062532	2.0	3.0			
GDSL GELP96	Esterase/lipase	Pop_A04G029313	1.1	1.6	AT5G37690	Potri.017G130100; Potri.004G086700	Lopes et al., 2020; Soler et al., 2007; Ursache et al., 2021
		Pop_G04G066624	n.s.	2.7			
BDG1/2/3	Lipase or alpha/beta-hydrolase	Pop_A01G059914	2.4	3.4	AT4G24140; AT1G64670; AT5G41900	Potri.003G147400; Potri.001G083400	Berhin et al., 2019; Fich et al., 2016; Lopes et al., 2020; Philippe et al., 2020; Ursache et al., 2021; Yeats et al., 2014
		Pop_A03G014942	n.s.	3.3			
		Pop_G01G002914	2.5	3.4			
		Pop_G03G013684	n.s.	2.8			
PLIP1	Lipase or alpha/beta-hydrolase	Pop_A02G012356	1.8	1.8	AT3G61680	Potri.014G097200; Potri.002G169600	Rains et al., 2017
		Pop_A14G044894	n.s.	1.8			
		Pop_G14G078045	1.5	1.7			
LOX1	Lipoxygenase	Pop_A10G048618	n.s.	−1.5	AT3G22400	Potri.008G151500; Potri.010G089500	Soler et al., 2007
		Pop_G10G001636	−1.3	−1.1			
LOX1	Lipoxygenase	Pop_G05G072291	−0.7	−0.7	AT1G55020	Potri.013G022100; Potri.005G032800; Potri.005G032650; Potri.013G022000; Potri.005G032700; Potri.005G032600; Potri.005G032400	Rains et al., 2017; Soler et al., 2007
CER7	Exoribonuclease positively regulating CER3 transcription	Pop_A01G057007	n.s.	−1.1	AT3G12990; AT3G60500	Potri.001G253200; Potri.009G047500	Li-Beisson et al., 2013
		Pop_G01G020944	n.s.	−1.0			
		Pop_A09G014440	n.s.	−0.9			
		Pop_G09G014179	−1.0	−1.5			
WRKY43	Transcription factor	Pop_G14G078059	3.0	3.5	AT2G46130; AT1G64000; AT5G41570	Potri.002G164900; Potri.014G090700	Geilen et al., 2017; Lopes et al., 2020; Rains et al., 2017; Soler et al., 2007
MYB9/107	Transcription factor	Pop_A19G052829	2.3	1.7	AT3G02940; AT5G16770	Potri.019G050900	Lashbrooke et al., 2016; Philippe et al., 2020; Rains et al., 2017; To et al., 2020
		Pop_G19G007588	1.5	1.6			
MYB41	Transcription factor	Pop_G19G084088	3.2	n.s.	AT4G28110	Potri.019G118200; Potri.013G148600	Capote et al., 2018; Kosma et al., 2014; Philippe et al., 2020

(Continues)

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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 (Continued)

Orthogroup	Description	*P.* × *canescens* Gene ID	log_2FC PEG	log_2FC NaCl	*A. thaliana* Gene ID	*P. trichocarpa* Gene ID	References
MYB93	Transcription factor	Pop_A02G088589	2.0	2.6	AT1G34670	Potri.005G164900; Potri.002G096800	Legay et al., 2016; Lopes et al., 2020; Rains et al., 2017; To et al., 2020
		Pop_A05G055977	n.s.	1.7			
		Pop_G02G065864	2.4	3.0			
ANAC058 NAC103	Transcription factor	Pop_A12G073877	2.4	2.7	AT3G18400	Potri.015G046800; Potri.012G056300	Lopes et al., 2020; Rains et al., 2017; Soler et al., 2007; Verdaguer et al., 2016
		Pop_G12G035036	1.8	2.9			
ERF38	Transcription factor	Pop_A14G000394	−1.4	−1.2	AT2G35700; AT2G44940; AT3G60490	Potri.002G141200; Potri.014G055700	Lasserre et al., 2008
		Pop_G14G044481	−2.1	−1.5			
PtERF38	Transcription factor	Pop_A06G079383	n.s.	n.s.	NA	Potri.006G138900; Potri.018G038100	Cheng et al., 2019
		Pop_A18G012844	1.7	1.7			
		Pop_G06G051489	1.4	n.s.			
		Pop_G18G078566	n.s.	1.6			

Note: Sorting of orthogroups was performed in rough accordance to the suberin biosynthesis pathway depicted in Rains et al. (2017). The *Arabidopsis thaliana* gene names are given as orthogroup description except where "Pt" clearly indicates *Populus* gene names; this table has been shortened from not significantly regulated (in both −0.6 MPa PEG and 120 mM NaCl treatment) and not available genes, for the extended version see Table S1; n.s., not significant (FDR < 0.05); NA, not available (either not mapped to the 84K reference genome, or no match in orthofinder)

which is also one of the main functions commonly described for the exodermis (Figure 3D,E). An exodermis is strictly defined as a hypodermis with Casparian bands (Perumalla & Peterson, 1986). It has previously been shown that severe osmotic (Kreszies et al., 2020) and salt stress (Reinhardt & Rost, 1995) can induce a suberized exodermis in species where it is missing in control conditions. This, in turn, could restrict the uptake of ions by blocking intermicrofibrillar spaces or plasmodesmata (Meyer & Peterson, 2013) and reducing the membrane surface area available for ion absorption (Enstone et al., 2003). Since an exodermis formation was observed only in apical positions (<30% relative root length) of a fraction of roots, and since no pronounced differences were observed for endodermal Casparian band development, this could explain that genes known to take part in the Casparian band development were not differentially regulated in response to stress (Table S2).

In contrast, many genes coding for the biosynthesis of phenylpropanoids and of lipophilic apoplastic barriers were significantly upregulated (Tables 2 and S1). This correlated with the microscopically observed and analytically confirmed enhanced suberization of the root tip (zone A) (Figures 3, 4, S1, and S2). However, based on expectations from comparable studies with different monocotyledonous species (Grünhofer, Schreiber, & Kreszies, 2021), it was surprising how little suberin deposition was increased in *P.* × *canescens* overall. Especially in the patchy zone B, which was shown to be the most responsive in several cultivars of barley cultivated in the very same laboratory under comparable osmotic stress conditions (Kreszies et al., 2019; Kreszies et al., 2020). This might be explained by the hypothesis that roots are subject to secondary growth and periderm formation in dicotyledonous trees, which ultimately replaces the endodermis and all outer tissue layers (Esau, 1977; Machado et al., 2013). Consequently, only the modification of the root tip in its primary developmental state should be of biological relevance.

A stress-induced shift of suberization toward the root tip has been described (Perumalla & Peterson, 1986; Reinhardt & Rost, 1995). However, aside from asymmetrical unilateral endodermal suberin deposition observed in roots of maize (Líška et al., 2016), the noncontinuous and exclusive full suberization of the root tip in response to salt stress observed here (Figure 3B,C) was never reported before. This represents a salt-specific reaction since the corresponding osmotic stress intensities produced less severe phenotypic alterations (Figure 3A,C). Previous observations have shown that cell wall suberization led to an effective ion barrier rather than an impediment to water movement (Ranathunge & Schreiber, 2011; Ranathunge, Schreiber, & Franke, 2011; Steudle, 2000) and root tips have been described to be the most conductive root zones for ions (Foster & Miklavcic, 2016). Regarding water relations, it has been discussed before that, for example, osmotic adjustments, closing of stomata, gating or decreased expression of aquaporins, and decline of growth might be more important water conservation processes than the suberization of root organs (Brunner et al., 2015; Steudle, 2000).

4.3 | Transport physiology

This is supported by our results studying responses of water and PTS transport in roots exposed to osmotic and salt stress (Figure 5).

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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

Stomatal conductances and osmotic potentials in roots and leaves decreased in response to stress (Figure 5A–C). But no significant differences were observed in the hydrostatic hydraulic conductivity, which is indicative of the combined radial apoplastic and cell-to-cell pathways for water uptake (Steudle, 2000), and apoplastic bypass flow (Figure 5D,E). PTS bypass due to ion size, weight, and molecule charge by no means is a measure of apoplastic water movement (Ranathunge & Schreiber, 2011; Voicu & Zwiazek, 2004). Instead, these results might suggest that the observed root tip suberization, especially after salinity treatment, is not sufficient to fully restrict sodium and chloride diffusion in roots, which is supported by the strong decline in plant vitality (Figure 1A).

The failure of Lp_r(OS) determination after salt stress treatment is not surprising and has been reported before (Knipfer et al., 2020; Krishnamurthy et al., 2011). The significant reduction of only Lp_r(OS), which is indicative of the cell-to-cell (symplastic and transcellular) pathway for water uptake (Steudle, 2000), and not Lp_r(HY) after osmotic stress, might point toward a considerable involvement of aquaporins in contributing to the observed abiotic stress responses. For poplars, members of the PIP2 family, and especially PIP2;5, were frequently identified as the major water conductive aquaporins (Almeida-Rodriguez et al., 2010; Almeida-Rodriguez et al., 2011; Marjanović et al., 2005; Secchi et al., 2009; Xu et al., 2016). Thorough research has already shown that stress intensity and stress duration are both crucial factors determining aquaporin regulation. Mild stress increased the hydraulic conductivity to transiently enhance water transport properties, potentially to avoid xylem embolisms, whereas severe stress led to hydraulic conductivity reductions in different species of poplars (Ranganathan et al., 2017; Siemens & Zwiazek, 2003, 2004). This was also indicated in transcriptomic studies, where aquaporin gene expression was increased after mild osmotic stresses (>−0.5 MPa; Jiang et al., 2020; Royer et al., 2016) but downregulated after exposure to 150 mM salt (Brinker et al., 2010). In addition, aquaporin expression can vary in temporal dependency and differ significantly between plant organs investigated (Bae et al., 2011; Gambetta et al., 2017; Liu et al., 2014). Only a few significant changes in aquaporin expression after 7 days of medium stress were observed here (Table S3). But besides the aquaporin gene transcription, also protein abundance and functional modifications need to be considered as factors influencing root hydraulic adaptations (Gambetta et al., 2017; Ranganathan et al., 2018).

4.4 | Additional transcriptomic remodeling

Changes in gene expression over time were also shown for members of the SOS pathway in poplar roots subjected to salinity. SOS1 and SOS2 expression decreased again after a strong initial induction, yet SOS3 expression steadily increased after 24 h of stress exposure (Tang et al., 2010), which fits to the results of our transcriptomic analysis after prolonged stress treatment (Table S6). SOS2 and SOS3 not only stimulate sodium exclusion via SOS1 but may also inhibit sodium uptake via the, in our study also significantly downregulated, sodium and potassium transporter HKT1 (Yoon et al., 2018).

In consensus with literature-based expectations, almost all investigated transcription factors known to take part in the mediation of abiotic stress responses were significantly induced even after 7 days of stress treatment (Table S7).

The pronounced upregulation of numerous genes in roots associated with the photosynthetic machinery in leaves remains obscure to us (Table 1). Expression of these transcripts in stress-exposed root tissue suggests that the corresponding protein products seem to possess roles also in nonphotosynthetic tissues and plastids, an observation that, while unexpected, has been reported sporadically in plants before (Jankangram, 2013; Singh et al., 2008). How far these genes may influence root development and energy metabolism (Nakata et al., 2018) under the influence of abiotic stress is an interesting puzzle and warrants further investigation.

5 | CONCLUSION

Besides important roles of osmoprotective pathways and ROS scavenging, our combined data set supports the idea that enhancing root suberization might be a promising target for the generation or breeding of genetically optimized poplars to be cultivated on water-limited and saline soils usually unusable for food production. Of course, secondary effects of increased suberization on root biomechanics have to be considered during this process. Albeit many similar reactions in response to both types of stress (osmotic and salt stress), the exclusive root tip suberization and stronger upregulation of suberin biosynthesis genes in response to salt, if compared to osmotic stress, indicates an especially important function of suberin in facilitating tolerance against increased concentrations of toxic NaCl.

AUTHOR CONTRIBUTIONS

Lukas Schreiber, Jinxing Lin, and Paul Grünhofer designed the experiments. Paul Grünhofer performed the experiments. Paul Grünhofer and Lukas Schreiber evaluated the physiological experiments. Paul Grünhofer and Tyll Stöcker evaluated the transcriptomics data. Paul Grünhofer and Tyll Stöcker wrote the manuscript. Yayu Guo, Ruili Li, Jinxing Lin, Kosala Ranathunge, Heiko Schoof, and Lukas Schreiber revised the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

Financial support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; SCHR17/1; Project Number 391657309), the National Natural Science Foundation of China (31761133009, 31970182), and the Programme of Introducing Talents of Discipline to Universities (111 project, B13007) is highly appreciated. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The raw sequencing data have been deposited at the National Center for Biotechnology Information (NCBI) sequence read archive (SRA study: SRP354749). Further data that support the findings of this

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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

study are available from the corresponding author upon reasonable request."

ORCID

Paul Grünhofer https://orcid.org/0000-0002-6298-5656

REFERENCES

Abramoff, M.D., Magalhães, P.J. & Ram, S.J. (2004) Image processing with ImageJ. *Biophotonics International*, 11, 36–42.

Allen, C.D., Macalady, A.K., Chenchouni, H., Bachelet, D., McDowell, N., Vennetier, M. et al. (2010) A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management*, 259, 660–684.

Almeida-Rodriguez, A.M., Cooke, J.E.K., Yeh, F. & Zwiazek, J.J. (2010) Functional characterization of drought-responsive aquaporins in *Populus balsamifera* and *Populus simonii* × *balsamifera* clones with different drought resistance strategies. *Physiologia Plantarum*, 140, 321–333.

Almeida-Rodriguez, A.M., Hacke, U.G. & Laur, J. (2011) Influence of evaporative demand on aquaporin expression and root hydraulics of hybrid poplar. *Plant, Cell & Environment*, 34, 1318–1331.

Andrews, S. (2010) *FastQC: a quality control tool for high throughput sequence data*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Bae, E.-K., Lee, H., Lee, J.-S. & Noh, E.-W. (2011) Drought, salt and wounding stress induce the expression of the plasma membrane intrinsic protein 1 gene in poplar (*Populus alba* × *P. tremula* var. *glandulosa*). *Gene*, 483, 43–48.

Bagniewska-Zadworna, A., Stelmasik, A. & Minicka, J. (2014) From birth to death—*Populus trichocarpa* fibrous roots functional anatomy. *Biologia Plantarum*, 58, 551–560.

Beisson, F., Li, Y., Bonaventure, G., Pollard, M. & Ohlrogge, J.B. (2007) The acyltransferase GPAT5 is required for the synthesis of suberin in seed coat and root of *Arabidopsis*. *The Plant Cell*, 19, 351–368.

Benveniste, I., Tijet, N., Adas, F., Philipps, G., Salaün, J.P. & Durst, F. (1998) CYP86A1 from *Arabidopsis thaliana* encodes a cytochrome P450-dependent fatty acid omega-hydroxylase. *Biochemical and Biophysical Research Communications*, 243, 688–693.

Berhin, A., de Bellis, D., Franke, R.B., Buono, R.A., Nowack, M.K. & Nawrath, C. (2019) The root cap cuticle: a cell wall structure for seedling establishment and lateral root formation. *Cell*, 176, 1367–1378.e8.

Bogeat-Triboulot, M.-B., Brosché, M., Renaut, J., Jouve, L., Le Thiec, D., Fayyaz, P. et al. (2007) Gradual soil water depletion results in reversible changes of gene expression, protein profiles, ecophysiology, and growth performance in *Populus euphratica*, a poplar growing in arid regions. *Plant Physiology*, 143, 876–892.

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

Bolser, D., Staines, D.M., Pritchard, E. & Kersey, P. (2016) Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomics data. In: Edwards, D. (Ed.) *Plant bioinformatics: methods in molecular biology*, Vol. 1374. New York: Humana Press, pp. 115–140.

Bolu, W.H. & Polle, A. (2004) Growth and stress reactions in roots and shoots of a salt-sensitive poplar species (*Populus* × *canescens*). *Tropical Ecology*, 45, 161–171.

Boyer, J.S. (1982) Plant productivity and environment. *Science*, 218, 443–448.

Brinker, M., Brosché, M., Vinocur, B., Abo-Ogiala, A., Fayyaz, P., Janz, D. et al. (2010) Linking the salt transcriptome with physiological responses of a salt-resistant *Populus* species as a strategy to identify genes important for stress acclimation. *Plant Physiology*, 154, 1697–1709.

Brosché, M., Vinocur, B., Alatalo, E.R., Lamminmäki, A., Teichmann, T., Ottow, E.A. et al. (2005) Gene expression and metabolite profiling of *Populus euphratica* growing in the Negev desert. *Genome Biology*, 6, R101.

Brundrett, M.C., Enstone, D.E. & Peterson, C.A. (1988) A Berberine-aniline blue fluorescent staining procedure for suberin, lignin, and callose in plant tissue. *Protoplasma*, 146, 133–142.

Brundrett, M.C., Kendrick, B. & Peterson, C.A. (1991) Efficient lipid staining in plant material with Sudan red 7B or fluoral yellow 088 in polyethylene glycol-glycerol. *Biotechnic & Histochemistry*, 66, 111–116.

Brunner, I., Herzog, C., Dawes, M.A., Arend, M. & Sperisen, C. (2015) How tree roots respond to drought. *Frontiers in Plant Science*, 6, 547.

Capote, T., Barbosa, P., Usié, A., Ramos, A.M., Inácio, V., Ordás, R. et al. (2018) ChIP-Seq reveals that QsMYB1 directly targets genes involved in lignin and suberin biosynthesis pathways in cork oak (*Quercus suber*). *BMC Plant Biology*, 18, 198.

Cervera, M.T., Storme, V., Soto, A., Ivens, B., van Montagu, M., Rajora, O.P. et al. (2005) Intraspecific and interspecific genetic and phylogenetic relationships in the genus *Populus* based on AFLP markers. *Theoretical and Applied Genetics*, 111, 1440–1456.

Challinor, A.J., Watson, J., Lobell, D.B., Howden, S.M., Smith, D.R. & Chhetri, N. (2014) A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change*, 4, 287–291.

Chen, T., Cai, X., Wu, X., Karahara, I., Schreiber, L. & Lin, J. (2011) Casparian strip development and its potential function in salt tolerance. *Plant Signaling & Behavior*, 6, 1499–1502.

Chen, Y.P., Chen, Y.N., Li, W.H. & Xu, C.C. (2006) Characterization of photosynthesis of *Populus euphratica* grown in the arid region. *Photosynthetica*, 44, 622–626.

Chen, S. & Polle, A. (2010) Salinity tolerance of *Populus*. *Plant Biology*, 12, 317–333.

Cheng, A.-X., Gou, J.-Y., Yu, X.-H., Yang, H., Fang, X., Chen, X.-Y. et al. (2013) Characterization and ectopic expression of a *Populus* hydroxyacid hydroxycinnamoyltransferase. *Molecular Plant*, 6, 1889–1903.

Cheng, C.-Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S. & Town, C.D. (2017) Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *The Plant Journal*, 89, 789–804.

Cheng, Z., Zhang, X., Zhao, K., Yao, W., Li, R., Zhou, B. et al. (2019) Overexpression of *ERF38* gene enhances salt and osmotic tolerance in transgenic poplar. *Frontiers in Plant Science*, 10, 1375.

Choat, B., Jansen, S., Brodribb, T.J., Cochard, H., Delzon, S., Bhaskar, R. et al. (2012) Global convergence in the vulnerability of forests to drought. *Nature*, 491, 752–755.

Cohen, D., Bogeat-Triboulot, M.-B., Tisserant, E., Balzergue, S., Martin-Magniette, M.-L., Lelandais, G. et al. (2010) Comparative transcriptomics of drought responses in *Populus*: a meta-analysis of genome-wide expression profiling in mature leaves and root apices across two genotypes. *BMC Genomics*, 11, 630.

Compagnon, V., Diehl, P., Benveniste, I., Meyer, D., Schaller, H., Schreiber, L. et al. (2009) CYP86B1 is required for very long chain omega-hydroxyacid and alpha, omega-dicarboxylic acid synthesis in root and seed suberin polyester. *Plant Physiology*, 150, 1831–1843.

Delude, C., Vishwanath, S.J., Rowland, O. & Domergue, F. (2017) Root aliphatic suberin analysis using non-extraction or solvent-extraction methods. *Bio-Protocol*, 7, e2331.

Dillen, S.Y., Rood, S.B. & Ceulemans, R. (2010) Growth and physiology. In: Jansson, S., Bhalerao, R. & Groover, A. (Eds.) *Genetics and genomics of Populus*. Berlin, Heidelberg, New York: Springer, pp. 39–63.

Dluzniewska, P., Gessler, A., Dietrich, H., Schnitzler, J.-P., Teuber, M. & Rennenberg, H. (2007) Nitrogen uptake and metabolism in *Populus* × *canescens* as affected by salinity. *New Phytologist*, 173, 279–293.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S. et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29, 15–21.

Doyle, E.A., Lane, A.M., Sides, J.M., Mudgett, M.B. & Monroe, J.D. (2007) An α-amylase (At4g25000) in *Arabidopsis* leaves is secreted and

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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

induced by biotic and abiotic stress. *Plant, Cell & Environment*, 30, 388–398.

Eckenwalder, J.E. (1996) Systematics and evolution of *Populus*. In: Stettler, R.F., Bradshaw, T., Heilman, P. & Hinckley, T. (Eds.) *Biology of Populus and its implications for management and conservation*. Ottawa: NRC Research Press, pp. 7–32.

Emms, D.M. & Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20, 238.

Enstone, D.E., Peterson, C.A. & Ma, F. (2003) Root endodermis and exodermis: structure, function, and responses to the environment. *Journal of Plant Growth Regulation*, 21, 335–351.

Esau, K. (1977) *Anatomy of seed plants*, 2nd edition. New York: Wiley.

Fich, E.A., Segerson, N.A. & Rose, J.K.C. (2016) The plant polyester cutin: biosynthesis, structure, and biological roles. *Annual Review of Plant Biology*, 67, 207–233.

Fiebig, A., Mayfield, J.A., Miley, N.L., Chau, S., Fischer, R.L. & Preuss, D. (2000) Alterations in *CER6*, a gene identical to *CUT1*, differentially affect long-chain lipid content on the surface of pollen and stems. *The Plant Cell*, 12, 2001–2008.

Fillatti, J.J., Sellmer, J., McCown, B., Haissig, B. & Comati, L. (1987) *Agrobacterium* mediated transformation and regeneration of *Populus*. *Molecular Genetics and Genomics*, 206, 192–199.

Foster, K.J. & Miklavcic, S.J. (2016) Modeling root zone effects on preferred pathways for the passive transport of ions and water in plant roots. *Frontiers in Plant Science*, 7, 914.

Franke, R., Höfer, R., Briesen, I., Emsermann, M., Efremova, N., Yephremov, A. et al. (2009) The *DAISY* gene from *Arabidopsis* encodes a fatty acid elongase condensing enzyme involved in the biosynthesis of aliphatic suberin in roots and the chalaza-micropyle region of seeds. *The Plant Journal*, 57, 80–95.

Fraser, C.M. & Chapple, C. (2011) The phenylpropanoid pathway in *Arabidopsis*. *The Arabidopsis Book/American Society of Plant Biologists*, 9, e0152.

Gambetta, G.A., Knipfer, T., Fricke, W. & McElrone, A.J. (2017) Aquaporins and root water uptake. In: Chaumont, F. & Tyerman, S.D. (Eds.) *Plant aquaporins: from transport to signaling. Signaling and communication in plants*. Cham: Springer, pp. 133–153.

Geilen, K., Heilmann, M., Hillmer, S. & Böhmer, M. (2017) WRKY43 regulates polyunsaturated fatty acid content and seed germination under unfavourable growth conditions. *Scientific Reports*, 7, 14235.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J. et al. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, 40, D1178–D1186.

Gou, J.-Y., Yu, X.-H. & Liu, C.-J. (2009) A hydroxycinnamoyltransferase responsible for synthesizing suberin aromatics in *Arabidopsis*. *Proceedings of the National Academy of Sciences*, 106, 18855–18860.

Grünhofer, P., Guo, Y., Li, R., Lin, J. & Schreiber, L. (2021) Hydroponic cultivation conditions allowing the reproducible investigation of poplar root suberization and water transport. *Plant Methods*, 17, 129.

Grünhofer, P., Herzig, L. & Schreiber, L. (2021) Leaf morphology, wax composition, and residual (cuticular) transpiration of four poplar clones. *Trees*, 36, 645–658.

Grünhofer, P., Schreiber, L. & Kreszies, T. (2021) Suberin in monocotyledonous crop plants: structure and function in response to abiotic stresses. In: Baluška, F. & Mukherjee, S. (Eds.) *Rhizobiology: molecular physiology of plant roots. Signaling and communication in plants*. Cham: Springer International Publishing, pp. 333–378.

Hanson, P.J., Sucoff, E.I. & Markhart, A.H. (1985) Quantifying apoplastic flux through red pine root systems using trisodium, 3-hydroxy-5,-8,10-pyrenetrisulfonate. *Plant Physiology*, 77, 21–24.

Hoagland, D.R. & Arnon, D.I. (1950) *The water-culture method for growing plants without soil. Circular, 347*. Berkeley, CA: University of California.

Höfer, R., Briesen, I., Beck, M., Pinot, F., Schreiber, L. & Franke, R. (2008) The *Arabidopsis* cytochrome P450 *CYP86A1* encodes a fatty acid omega-hydroxylase involved in suberin monomer biosynthesis. *Journal of Experimental Botany*, 59, 2347–2360.

Hooker, T.S., Millar, A.A. & Kunst, L. (2002) Significance of the expression of the CER6 condensing enzyme for cuticular wax production in *Arabidopsis*. *Plant Physiology*, 129, 1568–1580.

Isebrands, J.G. & Richardson, J. (2014) *Poplars and willows: trees for society and the environment*. Wallingford: CABI.

Jankangram, W. (2013) Analysis of gene expression in roots of rice seedlings: differing responses of three cultivars to salt-stress. *Asia-Pacific Journal of Science and Technology*, 18, 749–758.

Jansson, S. & Douglas, C.J. (2007) *Populus*: a model system for plant biology. *Annual Review of Plant Biology*, 58, 435–458.

Janz, D., Lautner, S., Wildhagen, H., Behnke, K., Schnitzler, J.-P., Rennenberg, H. et al. (2012) Salt stress induces the formation of a novel type of "pressure wood" in two *Populus* species. *New Phytologist*, 194, 129–141.

Jiang, C., Song, X., He, H., Chu, L., Zhou, H., Zhao, Y. et al. (2020) Genome-wide identification of plasma membrane aquaporin gene family in *Populus* and functional identification of *PIP1;1* involved in osmotic stress. *Environmental and Experimental Botany*, 179, 104200.

Kamaluddin, M. & Zwiazek, J.J. (2003) Fluoride inhibits root water transport and affects leaf expansion and gas exchange in aspen (*Populus tremuloides*) seedlings. *Physiologia Plantarum*, 117, 368–375.

Kang, Y., Khan, S. & Ma, X. (2009) Climate change impacts on crop yield, crop water productivity and food security—a review. *Progress in Natural Science*, 19, 1665–1674.

Knipfer, T., Danjou, M., Vionne, C. & Fricke, W. (2020) Salt stress reduces root water uptake in barley (*Hordeum vulgare* L.) through modification of the transcellular transport path. *Plant, Cell & Environment*, 44, 1–18.

Kosma, D.K., Murmu, J., Razeq, F.M., Santos, P., Bourgault, R., Molina, I. et al. (2014) AtMYB41 activates ectopic suberin synthesis and assembly in multiple plant species and cell types. *The Plant Journal*, 80, 216–229.

Kreszies, T., Eggels, S., Kreszies, V., Osthoff, A., Shellakkutti, N., Baldauf, J. A. et al. (2020) Seminal roots of wild and cultivated barley differentially respond to osmotic stress in gene expression, suberization, and hydraulic conductivity. *Plant, Cell & Environment*, 43, 344–357.

Kreszies, T., Shellakkutti, N., Osthoff, A., Yu, P., Baldauf, J.A., Zeisler-Diehl, V.V. et al. (2019) Osmotic stress enhances suberization of apoplastic barriers in barley seminal roots: analysis of chemical, transcriptomic and physiological responses. *New Phytologist*, 221, 180–194.

Kreuzwieser, J., Hauberg, J., Howell, K.A., Carroll, A., Rennenberg, H., Millar, A.H. et al. (2009) Differential response of gray poplar leaves and roots underpins stress adaptation during hypoxia. *Plant Physiology*, 149, 461–473.

Krishnamurthy, P., Ranathunge, K., Nayak, S., Schreiber, L. & Mathew, M. K. (2011) Root apoplastic barriers block Na^+ transport to shoots in rice (*Oryza sativa* L.). *Journal of Experimental Botany*, 62, 4215–4228.

Krishnamurthy, P., Vishal, B., Bhal, A. & Kumar, P.P. (2021) WRKY9 transcription factor regulates cytochrome P450 genes *CYP94B3* and *CYP86B1*, leading to increased root suberin and salt tolerance in *Arabidopsis*. *Physiologia Plantarum*, 172, 1673–1687.

Lashbrooke, J., Cohen, H., Levy-Samocha, D., Tzfadia, O., Panizel, I., Zeisler, V. et al. (2016) MYB107 and MYB9 homologs regulate suberin deposition in angiosperms. *The Plant Cell*, 28, 2097–2116.

Lasserre, E., Jobet, E., Llauro, C. & Delseny, M. (2008) *AtERF38* (*At2g35700*), an AP2/ERF family transcription factor gene from *Arabidopsis thaliana*, is expressed in specific cell types of roots, stems and seeds that undergo suberization. *Plant Physiology and Biochemistry*, 46, 1051–1061.

Lee, S.-B., Jung, S.-J., Go, Y.-S., Kim, H.-U., Kim, J.-K., Cho, H.-J. et al. (2009) Two *Arabidopsis* 3-ketoacyl CoA synthase genes, *KCS20* and *KCS2/DAISY*, are functionally redundant in cuticular wax and root suberin biosynthesis, but differentially controlled by osmotic stress. *The Plant Journal*, 60, 462–475.

Legay, S., Guerriero, G., André, C., Guignard, C., Cocco, E., Charton, S. et al. (2016) MdMyb93 is a regulator of suberin deposition in russeted apple fruit skins. *New Phytologist*, 212, 977–991.

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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

Leng, H., Lu, M. & Wan, X. (2013) Variation in embolism occurrence and repair along the stem in drought-stressed and re-watered seedlings of a poplar clone. *Physiologia Plantarum*, 147, 329–339.

Leplé, J.-C., Brasileiro, A.C.M., Michel, M.F., Delmotte, F. & Jouanin, L. (1992) Transgenic poplars: expression of chimeric genes using four different constructs. *Plant Cell Reports*, 11, 137–141.

Li, Y., Beisson, F., Koo, A.J.K., Molina, I., Pollard, M. & Ohlrogge, J. (2007) Identification of acyltransferases required for cutin biosynthesis and production of cutin with suberin-like monomers. *Proceedings of the National Academy of Sciences*, 104, 18339–18344.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079.

Liao, Y., Smyth, G.K. & Shi, W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30, 923–930.

Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M.X., Arondel, V., Bates, P.D. et al. (2013) Acyl-lipid metabolism. *The Arabidopsis Book/American Society of Plant Biologists*, 11, e0161.

Lin, Y.-C., Wang, J., Delhomme, N., Schiffthaler, B., Sundström, G., Zuccolo, A. et al. (2018) Functional and evolutionary genomic inferences in *Populus* through genome and population sequencing of American and European aspen. *Proceedings of the National Academy of Sciences of The United States of America*, 115, E10970–E10978.

Líška, D., Martinka, M., Kohanová, J. & Lux, A. (2016) Asymmetrical development of root endodermis and exodermis in reaction to abiotic stresses. *Annals of Botany*, 118, 667–674.

Liu, J., Equiza, M.A., Navarro-Rodenas, A., Lee, S.H. & Zwiazek, J.J. (2014) Hydraulic adjustments in aspen (*Populus tremuloides*) seedlings following defoliation involve root and leaf aquaporins. *Planta*, 240, 553–564.

Liu, J.-G., Han, X., Yang, T., Cui, W.-H., Wu, A.-M., Fu, C.-X. et al. (2019) Genome-wide transcriptional adaptation to salt stress in *Populus*. *BMC Plant Biology*, 19, 1–14.

Lopes, S.T., Sobral, D., Costa, B., Perdiguero, P., Chaves, I., Costa, A. et al. (2020) Phellem versus xylem: genome-wide transcriptomic analysis reveals novel regulators of cork formation in cork oak. *Tree Physiology*, 40, 129–141.

Lux, A., Sottníková, A., Opatrná, J. & Greger, M. (2004) Differences in structure of adventitious roots in *Salix* clones with contrasting characteristics of cadmium accumulation and sensitivity. *Physiologia Plantarum*, 120, 537–545.

Ma, J., Wan, D., Duan, B., Bai, X., Bai, Q., Chen, N. et al. (2019) Genome sequence and genetic transformation of a widely distributed and cultivated poplar. *Plant Biotechnology Journal*, 17, 451–460.

Ma, T., Wang, J., Zhou, G., Yue, Z., Hu, Q., Chen, Y. et al. (2013) Genomic insights into salt adaptation in a desert poplar. *Nature Communications*, 4, 1–8.

Machado, A., Pereira, H. & Teixeira, R.T. (2013) Anatomy and development of the endodermis and phellem of *Quercus suber* L. roots. *Microscopy and Microanalysis*, 19, 525–534.

Mader, M., Le Paslier, M.-C., Bounon, R., Bérard, A., Rampant, P.F., Fladung, M. et al. (2016) Whole-genome draft assembly of *Populus tremula* × *P. alba* clone INRA 717-1B4. *Silvae Genetica*, 65, 74–79.

Mahmood, K., El-Kereamy, A., Kim, S.-H., Nambara, E. & Rothstein, S.J. (2016) ANAC032 positively regulates age-dependent and stress-induced senescence in *Arabidopsis thaliana*. *Plant and Cell Physiology*, 57, 2029–2046.

Marjanović, Z., Uehlein, N., Kaldenhoff, R., Zwiazek, J.J., Weiss, M., Hampp, R. et al. (2005) Aquaporins in poplar: what a difference a symbiont makes! *Planta*, 222, 258–268.

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S.O., Wicker, T. et al. (2017) A chromosome conformation capture ordered sequence of the barley genome. *Nature*, 544, 427–433.

Meyer, C.J. & Peterson, C.A. (2013) Structure and function of three suberized cell layers: epidermis, exodermis, and endodermis. In: Eshel, A. & Beeckman, T. (Eds.) *Plant roots: the hidden half*, Fourth edition. New York, USA: CRC Press, pp. 5.1–5.20.

Michel, B.E. (1983) Evaluation of the water potentials of solutions of polyethylene glycol 8000 both in the absence and presence of other solutes. *Plant Physiology*, 72, 66–70.

Millar, A.A., Clemens, S., Zachgo, S., Giblin, M.E., Taylor, D.C. & Kunst, L. (1999) *CUT1*, an *Arabidopsis* gene required for cuticular wax biosynthesis and pollen fertility, encodes a very-long-chain-fatty acid condensing enzyme. *The Plant Cell*, 11, 825–838.

Mittler, R. (2006) Abiotic stress, the field environment and stress combination. *Trends in Plant Science*, 11, 15–19.

Miyamoto, N., Steudle, E., Hirasawa, T. & Lafitte, R. (2001) Hydraulic conductivity of rice roots. *Journal of Experimental Botany*, 52, 1835–1846.

Molina, I., Li-Beisson, Y., Beisson, F., Ohlrogge, J.B. & Pollard, M. (2009) Identification of an *Arabidopsis* feruloyl-coenzyme A transferase required for suberin synthesis. *Plant Physiology*, 151, 1317–1328.

Müller, A., Volmer, K., Mishra-Knyrim, M. & Polle, A. (2013) Growing poplars for research with and without mycorrhizas. *Frontiers in Plant Science*, 4, 1–11.

Munns, R. (2002) Comparative physiology of salt and water stress. *Plant, Cell & Environment*, 25, 239–250.

Nakata, M.T., Sato, M., Wakazaki, M., Sato, N., Kojima, K., Sekine, A. et al. (2018) Plastid translation is essential for lateral root stem cell patterning in *Arabidopsis thaliana*. *Biology Open*, 7, 1–13.

Niinemets, Ü., Kull, O. & Tenhunen, J.D. (1999) Variability in leaf morphology and chemical composition as a function of canopy light environment in coexisting deciduous trees. *International Journal of Plant Sciences*, 160, 837–848.

Nishizawa, A., Yabuta, Y. & Shigeoka, S. (2008) Galactinol and raffinose constitute a novel function to protect plants from oxidative damage. *Plant Physiology*, 147, 1251–1263.

Osthoff, A., Donà Dalle Rose, P., Baldauf, J.A., Piepho, H.-P. & Hochholdinger, F. (2019) Transcriptomic reprogramming of barley seminal roots by combined water deficit and salt stress. *BMC Genomics*, 20, 325.

Ottow, E.A., Brinker, M., Teichmann, T., Fritz, E., Kaiser, W., Brosché, M. et al. (2005) *Populus euphratica* displays apoplastic sodium accumulation, osmotic adjustment by decreases in calcium and soluble carbohydrates, and develops leaf succulence under salt stress. *Plant Physiology*, 139, 1762–1772.

Pallara, G., Giovannelli, A., Traversi, M.L., Camussi, A. & Racchi, M.L. (2012) Effect of water deficit on expression of stress-related genes in the cambial region of two contrasting poplar clones. *Journal of Plant Growth Regulation*, 31, 102–112.

Paul, S., Gable, K., Beaudoin, F., Cahoon, E., Jaworski, J., Napier, J.A. et al. (2006) Members of the *Arabidopsis* FAE1-like 3-ketoacyl-CoA synthase gene family substitute for the Elop proteins of *Saccharomyces cerevisiae*. *The Journal of Biological Chemistry*, 281, 9018–9029.

Pearce, D.W., Millard, S., Bray, D.F. & Rood, S.B. (2005) Stomatal characteristics of riparian poplar species in a semi-arid environment. *Tree Physiology*, 26, 211–218.

Peng, Y., Zhou, Z., Tong, R., Hu, X. & Du, K. (2017) Anatomy and ultrastructure adaptations to soil flooding of two full-sib poplar clones differing in flood-tolerance. *Flora*, 233, 90–98.

Perumalla, C.J. & Peterson, C.A. (1986) Deposition of Casparian bands and suberin lamellae in the exodermis and endodermis of young corn and onion roots. *Canadian Journal of Botany*, 64, 1873–1878.

Philippe, G., Sørensen, I., Jiao, C., Sun, X., Fei, Z., Domozych, D.S. et al. (2020) Cutin and suberin: assembly and origins of specialized lipidic cell wall scaffolds. *Current Opinion in Plant Biology*, 55, 11–20.

Polle, A. & Chen, S. (2015) On the salty side of life: molecular, physiological and anatomical adaptation and acclimation of trees to extreme habitats. *Plant, Cell & Environment*, 38, 1794–1816.

Polle, A., Chen, S.L., Eckert, C. & Harfouche, A. (2019) Engineering drought resistance in forest trees. *Frontiers in Plant Science*, 9, 1875.

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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

Qiu, D., Bai, S., Ma, J., Zhang, L., Shao, F., Zhang, K. et al. (2019) The genome of *Populus alba* × *Populus tremula* var. *glandulosa* clone 84K. *DNA Research*, 26, 423–431.

R Core Team. (2013) *R: A language and environment for statistical computing*. https://www.r-project.org/

Rains, M.K., Gardiyehewa de Silva, N.D. & Molina, I. (2017) Reconstructing the suberin pathway in poplar by chemical and transcriptomic analysis of bark tissues. *Tree Physiology*, 38, 340–361.

Ranathunge, K., Kim, Y.X., Wassmann, F., Kreszies, T., Zeisler, V. & Schreiber, L. (2017) The composite water and solute transport of barley (*Hordeum vulgare*) roots: effect of suberized barriers. *Annals of Botany*, 119, 629–643.

Ranathunge, K., Lin, J., Steudle, E. & Schreiber, L. (2011) Stagnant deoxygenated growth enhances root suberization and lignifications, but differentially affects water and NaCl permeabilities in rice (*Oryza sativa* L.) roots. *Plant, Cell & Environment*, 34, 1223–1240.

Ranathunge, K. & Schreiber, L. (2011) Water and solute permeabilities of *Arabidopsis* roots in relation to the amount and composition of aliphatic suberin. *Journal of Experimental Botany*, 62, 1961–1974.

Ranathunge, K., Schreiber, L. & Franke, R. (2011) Suberin research in the genomics era - new interest for an old polymer. *Plant Science*, 180, 399–413.

Ranganathan, K., Cooke, J.E.K., El Kayal, W., Equiza, M.A., Vaziriyeganeh, M. & Zwiazek, J.J. (2017) Over-expression of PIP2;5 aquaporin alleviates gas exchange and growth inhibition in poplars exposed to mild osmotic stress with polyethylene glycol. *Acta Physiologiae Plantarum*, 39, 1–9.

Ranganathan, K., El Kayal, W., Cooke, J.E.K. & Zwiazek, J.J. (2016) Responses of hybrid aspen over-expressing a PIP2;5 aquaporin to low root temperature. *Journal of Plant Physiology*, 192, 98–104.

Ranganathan, K., Vaziri, M. & Zwiazek, J.J. (2018) Regulation of aquaporins in plants under stress. *Biological Research*, 51, 4.

Reinhardt, D.H. & Rost, T.L. (1995) Salinity accelerates endodermal development and induces an exodermis in cotton seedling roots. *Environmental and Experimental Botany*, 35, 563–574.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K. (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26, 139–140.

Ronald, W.G. (1982) Intersectional hybridization of *Populus* sections, Leuce-Aigeiros and Leuce-Tacamahaca. *Silvae Genetica*, 31, 94–99.

Rosenzweig, C., Karoly, D., Vicarelli, M., Neofotis, P., Wu, Q., Casassa, G. et al. (2008) Attributing physical and biological impacts to anthropogenic climate change. *Nature*, 453, 353–357.

Royer, M., Cohen, D., Aubry, N., Vendramin, V., Scalabrin, S., Cattonaro, F. et al. (2016) The build-up of osmotic stress responses within the growing root apex using kinematics and RNA-sequencing. *Journal of Experimental Botany*, 67, 5961–5973.

Sannigrahi, P. & Ragauskas, A.J. (2010) Poplar as a feedstock for biofuels: a review of compositional characteristics. *Biofuels, Bioproducts and Biorefining*, 4, 209–226.

Schreiber, L., Franke, R. & Hartmann, K. (2005) Effects of NO_3 deficiency and NaCl stress on suberin deposition in rhizo- and hypodermal (RHCW) and endodermal cell walls (ECW) of castor bean (*Ricinus communis* L.) roots. *Plant and Soil*, 269, 333–339.

Schwacke, R., Ponce-Soto, G.Y., Krause, K., Bolger, A.M., Arsova, B., Hallab, A. et al. (2019) MapMan4: a refined protein classification and annotation framework applicable to multi-omics data analysis. *Molecular Plant*, 12, 879–892.

Secchi, F., Maciver, B., Zeidel, M.L. & Zwieniecki, M.A. (2009) Functional analysis of putative genes encoding the PIP2 water channel subfamily in *Populus trichocarpa*. *Tree Physiology*, 29, 1467–1477.

Shanmugarajah, K., Linka, N., Gräfe, K., Smits, S.H.J., Weber, A.P.M., Zeier, J. et al. (2019) ABCG1 contributes to suberin formation in *Arabidopsis thaliana* roots. *Scientific Reports*, 9, 11381.

Shiono, K., Ando, M., Nishiuchi, S., Takahashi, H., Watanabe, K., Nakamura, M. et al. (2014) RCN1/OsABCG5, an ATP-binding cassette (ABC) transporter, is required for hypodermal suberization of roots in rice (*Oryza sativa*). *The Plant Journal*, 80, 40–51.

Siemens, J.A. & Zwiazek, J.J. (2003) Effects of water deficit stress and recovery on the root water relations of trembling aspen (*Populus tremuloides*) seedlings. *Plant Science*, 165, 113–120.

Siemens, J.A. & Zwiazek, J.J. (2004) Changes in root water flow properties of solution culture-grown trembling aspen (*Populus tremuloides*) seedlings under different intensities of water-deficit stress. *Physiologia Plantarum*, 121, 44–49.

Silim, S., Nash, R., Reynard, D., White, B. & Schroeder, W. (2009) Leaf gas exchange and water potential responses to drought in nine poplar (*Populus* spp.) clones with contrasting drought tolerance. *Trees*, 23, 959–969.

Singh, P., Kaloudas, D. & Raines, C.A. (2008) Expression analysis of the *Arabidopsis* CP12 gene family suggests novel roles for these proteins in roots and floral tissues. *Journal of Experimental Botany*, 59, 3975–3985.

Sjödin, A., Street, N.R., Sandberg, G., Gustafsson, P. & Jansson, S. (2009) The *Populus* genome integrative explorer (PopGenIE): a new resource for exploring the *Populus* genome. *New Phytologist*, 182, 1013–1025.

Soler, M., Serra, O., Molinas, M., Huguet, G., Fluch, S. & Figueras, M. (2007) A genomic approach to suberin biosynthesis and cork differentiation. *Plant Physiology*, 144, 419–431.

Steudle, E. (2000) Water uptake by roots: effects of water deficit. *Journal of Experimental Botany*, 51, 1531–1542.

Steudle, E. & Meshcheryakov, A.B. (1996) Hydraulic and osmotic properties of oak roots. *Journal of Experimental Botany*, 47, 387–401.

Steudle, E., Murrmann, M. & Peterson, C.A. (1993) Transport of water and solutes across maize roots modified by puncturing the endodermis (further evidence for the composite transport model of the root). *Plant Physiology*, 103, 335–349.

Stoláriková, M., Vaculík, M., Lux, A., Baccio, D., Minnocci, A., Andreucci, A. et al. (2012) Anatomical differences of poplar (*Populus* × *euramericana* clone I-214) roots exposed to zinc excess. *Biologia*, 67, 483–489.

Sun, L., Ma, Y., Wang, H., Huang, W., Wang, X., Han, L. et al. (2018) Overexpression of *PtABCC1* contributes to mercury tolerance and accumulation in *Arabidopsis* and poplar. *Biochemical and Biophysical Research Communications*, 497, 997–1002.

Syaiful, S.A., Dungga, N.E., Riadi, M. & Ridwan, I. (2014) Seed priming with PEG 8000 for improving drought stress tolerance of soybean (*Glycine max*). *International Journal of Agriculture System*, 1, 19–26.

Tamás, L., Bočová, B., Huttová, J., Liptáková, Ľ., Mistrík, I., Valentovičová, K. et al. (2012) Impact of the auxin signaling inhibitor p-chlorophenoxyisobutyric acid on short-term Cd-induced hydrogen peroxide production and growth response in barley root tip. *Journal of Plant Physiology*, 169, 1375–1381.

Tang, R.-J., Liu, H., Bao, Y., Lv, Q.-D., Yang, L. & Zhang, H.-X. (2010) The woody plant poplar has a functionally conserved salt overly sensitive pathway in response to salinity stress. *Plant Molecular Biology*, 74, 367–380.

To, A., Joubès, J., Thueux, J., Kazaz, S., Lepiniec, L. & Baud, S. (2020) AtMYB92 enhances fatty acid synthesis and suberin deposition in leaves of *Nicotiana benthamiana*. *The Plant Journal*, 103, 660–676.

Todd, J., Post-Beittenmiller, D. & Jaworski, J.G. (1999) *KCS1* encodes a fatty acid elongase 3-ketoacyl-CoA synthase affecting wax biosynthesis in *Arabidopsis thaliana*. *The Plant Journal*, 17, 119–130.

Trenkamp, S., Martin, W. & Tietjen, K. (2004) Specific and differential inhibition of very-long-chain fatty acid elongases from *Arabidopsis thaliana* by different herbicides. *Proceedings of the National Academy of Sciences*, 101, 11903–11908.

Tuskan, G.A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U. et al. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, 313, 1596–1604.

Ursache, R., de Jesus Vieira Teixeira, C., Dénervaud, T.V., Gully, K., de Bellis, D., Schmid-Siegert, E. et al. (2021) GDSL-domain proteins have

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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

key roles in suberin polymerization and degradation. *Nature Plants*, 7, 353–364.

Verdaguer, R., Soler, M., Serra, O., Garrote, A., Fernández, S., Company-Arumí, D. et al. (2016) Silencing of the potato *StNAC103* gene enhances the accumulation of suberin polyester and associated wax in tuber skin. *Journal of Experimental Botany*, 67, 5415–5427.

Voicu, M.C. & Zwiazek, J.J. (2004) Cycloheximide inhibits root water flow and stomatal conductance in aspen (*Populus tremuloides*) seedlings. *Plant, Cell & Environment*, 27, 199–208.

Wan, X. & Zwiazek, J.J. (1999) Mercuric chloride effects on root water transport in Aspen seedlings. *Plant Physiology*, 121, 939–946.

Watanabe, S., Kojima, K., Ide, Y. & Sasaki, S. (2000) Effects of saline and osmotic stress on proline and sugar accumulation in *Populus euphratica* in vitro. *Plant Cell, Tissue and Organ Culture*, 63, 199–206.

Xin, A. & Herburger, K. (2020) Mini review: transport of hydrophobic polymers into the plant apoplast. *Frontiers in Plant Science*, 11, 590990.

Xu, H., Cooke, J.E.K., Kemppainen, M., Pardo, A.G. & Zwiazek, J.J. (2016) Hydraulic conductivity and aquaporin transcription in roots of trembling aspen (*Populus tremuloides*) seedlings colonized by Laccaria bicolor. *Mycorrhiza*, 26, 441–451.

Xue, L.-J., Alabady, M.S., Mohebbi, M. & Tsai, C.-J. (2015) Exploiting genome variation to improve next-generation sequencing data analysis and genome editing efficiency in *Populus tremula* × *alba* 717-1B4. *Tree Genetics & Genomes*, 11, 82.

Yadav, V., Molina, I., Ranathunge, K., Castillo, I.Q., Rothstein, S.J. & Reed, J. W. (2014) ABCG transporters are required for suberin and pollen wall extracellular barriers in *Arabidopsis*. *Plant Cell*, 26, 3569–3588.

Yang, W., Pollard, M., Li-Beisson, Y., Beisson, F., Feig, M. & Ohlrogge, J. (2010) A distinct type of glycerol-3-phosphate acyltransferase with *sn*-2 preference and phosphatase activity producing 2-monoacylglycerol. *Proceedings of the National Academy of Sciences*, 107, 12040–12045.

Yang, W., Simpson, J.P., Li-Beisson, Y., Beisson, F., Pollard, M. & Ohlrogge, J.B. (2012) A land-plant-specific glycerol-3-phosphate acyltransferase family in *Arabidopsis*: substrate specificity, *sn*-2 preference, and evolution. *Plant Physiology*, 160, 638–652.

Yeats, T.H., Huang, W., Chatterjee, S., Viart, H.M.-F., Clausen, M.H., Stark, R.E. et al. (2014) Tomato cutin deficient 1 (CD1) and putative orthologs comprise an ancient family of cutin synthase-like (CUS) proteins that are conserved among land plants. *The Plant Journal*, 77, 667–675.

Yoon, S.-K., Bae, E.-K., Lee, H., Choi, Y.-I., Han, M., Choi, H. et al. (2018) Downregulation of stress-associated protein 1 (*PagSAP1*) increases salt stress tolerance in poplar (*Populus alba* × *P. glandulosa*). *Trees*, 32, 823–833.

Zacchini, M., Pietrini, F., Scarascia, M.G., Iori, V., Pietrosanti, L. & Massacci, A. (2009) Metal tolerance, accumulation and translocation in poplar and willow clones treated with cadmium in hydroponics. *Water, Air, & Soil Pollution*, 197, 23–34.

Zanewich, K.P., Pearce, D.W. & Rood, S.B. (2018) Heterosis in poplar involves phenotypic stability: cottonwood hybrids outperform their parental species at suboptimal temperatures. *Tree Physiology*, 38, 789–800.

Zawaski, C., Ma, C., Strauss, S.H., French, D., Meilan, R. & Busov, V.B. (2012) *PHOTOPERIOD RESPONSE 1* (*PHOR1*)-like genes regulate shoot/root growth, starch accumulation, and wood formation in *Populus*. *Journal of Experimental Botany*, 63, 5623–5634.

Zeier, J. & Schreiber, L. (1997) Chemical composition of hypodermal and endodermal cell walls and xylem vessels isolated from *Clivia miniata*. *Plant Physiology*, 113, 1223–1231.

Zeier, J. & Schreiber, L. (1998) Comparative investigation of primary and tertiary endodermal cell walls isolated from the roots of five monocotyledoneous species: chemical composition in relation to fine structure. *Planta*, 206, 349–361.

Zhang, Z., Chen, Y., Zhang, J., Ma, X., Li, Y., Li, M. et al. (2020) Improved genome assembly provides new insights into genome evolution in a desert poplar (*Populus euphratica*). *Molecular Ecology Resources*, 20, 781–794.

Zheng, L., Meng, Y., Ma, J., Zhao, X., Cheng, T., Ji, J. et al. (2015) Transcriptomic analysis reveals importance of ROS and phytohormones in response to short-term salinity stress in *Populus tomentosa*. *Frontiers in Plant Science*, 6, 678.

Zimmermann, H.M., Hartmann, K.-D., Schreiber, L. & Steudle, E. (2000) Chemical composition of apoplastic transport barriers in relation to radial hydraulic conductivity of corn roots (*Zea mays* L.). *Planta*, 210, 302–311.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Grünhofer, P., Stöcker, T., Guo, Y., Li, R., Lin, J., Ranathunge, K. et al. (2022) *Populus* × *canescens* root suberization in reaction to osmotic and salt stress is limited to the developing younger root tip region. *Physiologia Plantarum*, 174(5), e13765. Available from: https://doi.org/10.1111/ppl.13765

13993054, 2022, 5, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/ppl.13765 by Universitäts-U Landesbibliothek Bonn, Wiley Online Library on [04/04/2023]. 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

F1000Research F1000Research 2023, 11:1137 Last updated: 12 MAR 2024

Check for updates

SOFTWARE TOOL ARTICLE

REVISED A2TEA: Identifying trait-specific evolutionary adaptations [version 2; peer review: 2 approved]

Tyll Stöcker, Carolin Uebermuth-Feldhaus, Florian Boecker, Heiko Schoof

Crop Bioinformatics, University of Bonn, Bonn, NRW, 53115, Germany

v2 **First published:** 05 Oct 2022, **11**:1137
https://doi.org/10.12688/f1000research.126463.1
Latest published: 12 Apr 2023, **11**:1137
https://doi.org/10.12688/f1000research.126463.2

Abstract

Background: Plants differ in their ability to cope with external stresses (e.g., drought tolerance). Genome duplications are an important mechanism to enable plant adaptation. This leads to characteristic footprints in the genome, such as protein family expansion. We explore genetic diversity and uncover evolutionary adaptation to stresses by exploiting genome comparisons between stress tolerant and sensitive species and RNA-Seq data sets from stress experiments. Expanded gene families that are stress-responsive based on differential expression analysis could hint at species or clade-specific adaptation, making these gene families exciting candidates for follow-up tolerance studies and crop improvement.
Software: Integration of such cross-species omics data is a challenging task, requiring various steps of transformation and filtering. Ultimately, visualization is crucial for quality control and interpretation. To address this, we developed A2TEA: Automated Assessment of Trait-specific Evolutionary Adaptations, a Snakemake workflow for detecting adaptation footprints in silico. It functions as a one-stop processing pipeline, integrating protein family, phylogeny, expression, and protein function analyses. The pipeline is accompanied by an R Shiny web application that allows exploring, highlighting, and exporting the results interactively. This allows the user to formulate hypotheses regarding the genomic adaptations of one or a subset of the investigated species to a given stress.
Conclusions: While our research focus is on crops, the pipeline is entirely independent of the underlying species and can be used with any set of species. We demonstrate pipeline efficiency on real-world datasets and discuss the implementation and limits of our analysis workflow as well as planned extensions to its current state. The A2TEA workflow and web application are publicly available at: https://github.com/tgstoecker/A2TEA.Workflow and https://github.com/tgstoecker/A2TEA.WebApp, respectively.

Keywords

plants, crops, adaptation, evolution, stress, workflow, software

Open Peer Review

Approval Status ✓ ✓

	1	2
version 2 (revision) 12 Apr 2023	✓ view	✓ view
version 1 05 Oct 2022	? view	? view

1. **Manuel Aranda**, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
Octavio Salazar Moya, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

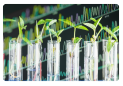
2. **Rahul Siddharthan**, The Institute of Mathematical Sciences, Chennai, India

Any reports and responses or comments on the article can be found at the end of the article.

F1000Research F1000Research 2023, 11:1137 Last updated: 12 MAR 2024

This article is included in the Genomics and Genetics gateway.

This article is included in the Bioinformatics gateway.



This article is included in the Plant Computational and Quantitative Genomics collection.

Corresponding authors: Tyll Stöcker (tyll.stoecker@gmail.com), Heiko Schoof (schoof@uni-bonn.de)

Author roles: Stöcker T: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Uebermuth-Feldhaus C**: Conceptualization, Formal Analysis, Investigation, Methodology, Writing – Review & Editing; **Boecker F**: Software, Writing – Review & Editing; **Schoof H**: Conceptualization, Funding Acquisition, Resources, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2023 Stöcker T *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Stöcker T, Uebermuth-Feldhaus C, Boecker F and Schoof H. **A2TEA: Identifying trait-specific evolutionary adaptations [version 2; peer review: 2 approved]** F1000Research 2023, **11**:1137 https://doi.org/10.12688/f1000research.126463.2

First published: 05 Oct 2022, **11**:1137 https://doi.org/10.12688/f1000research.126463.1

F1000Research 2023, 11:1137 Last updated: 12 MAR 2024

REVISED **Amendments from Version 1**

Dear readers, based on the reviewer's comments, we have made improvements to the manuscript, and we have also added new features to our software.

Fixed editing errors: garbled text, missing citations, improved visualizations.

Paragraph 1 of the introduction has been rewritten to more clearly separate prior approaches and the need to combine them as we do with A2TEA.

New features of the software:

The pipeline has become more flexible since it is no longer required to provide RNA-seq data for all species – this allows species to be included in the phylogenetic analyses for which no RNA-seq data for the stress/treatment under investigation is available.

Tables in the WebApp can now be downloaded in their un-/filtered entirety.

Updates to plots such as better labels, positioning of cutoff lines behind plotted data, and removal of small 0 bars in the log2FC layer of the tea plots.

Any further responses from the reviewers can be found at the end of the article

Introduction

While genomic resources for crop species continuously increase, with more and more high-quality reference genome sequences and transcriptome datasets becoming available, the lack of integrated trait and functional information limits the ability to interpret genomic-scale datasets and discover genotype-phenotype associations. Methods such as differential expression analysis have led to the discovery of many candidate genes e.g., involved in tolerance to stresses or, more generally, central to the physiological reaction pattern towards a specific experimental treatment.[1,2] However, the interpretation of large lists of candidate genes is challenging and does not provide insight into evolutionary adaptation to the treatment under investigation. In contrast, comparative genomics approaches allow the identification of genomic footprints of adaptation,[3] such as protein family expansions.[4] Gene duplication is a major driver of molecular evolution,[5,6] and in plants, whole-genome duplication events are frequent (reviewed in Ref. 7), but tandem and transposon-mediated duplications also play a role (reviewed in Ref. 8). Most gene duplicates are lost or silenced,[9] but retained duplicates may hint at some evolutionary advantage and may be targets of adaptation.[10] However, associating evolutionary retention with functions relating to complex traits of a species is not possible without considering further information, such as insight into condition-specific gene usage, e.g., in the form of expression data. These approaches thus have clear limitations when used in isolation.

While many efforts have focused on individual (model) species and the outlined singular methodological approaches, the increasing availability of omics data for many more or less related genomes opens opportunities to explore genetic diversity through multi-genome comparisons. To overcome the aforementioned limitations, we propose a pipeline that combines differential expression analysis and comparative genomics to prioritize genes that were targets of evolutionary adaptation, thereby facilitating their application in crop improvement. Especially for the adaptation of regulatory networks, duplication allows for neo- or subfunctionalization,[11] which form an evolutionary scenario that can be observed based on our integration of phylogeny and differential expression under treatment/stress data. However, we also consider differential expression under a given treatment in any species as a functional link to the given treatment, even if there is not sufficient data to confirm neo- or subfunctionalization. This allows us to filter for gene family expansions functionally linked to the given treatment, as not all adaptations and thus retained duplicates in a genome need to relate to e.g., tolerance to the given stress, other traits not under analysis will also show adaptation and thus protein family expansions. Our approach allows for a more comprehensive understanding of trait adaptation (in both plants and other organisms) and can guide the development of strategies for improving crops (Figure 1).

The challenge for this multi-genome approach is the cross-species integration of multiple types of omics data, which requires several software tools and various custom steps of transformation and filtering. To promote the exploration of genomics and transcriptomics data and the association of genotype with phenotype data in order to address adaptation, we developed **A2TEA** (**A**utomated **A**ssessment of **T**rait-specific **E**volutionary **A**daptations). Our software aims at identifying candidate genes for stress adaptation in plant species and enables GUI-based exploration of the results but is suitable for gene family expansion analysis integrated with differential expression data in any set of genomes. It is composed of a Snakemake workflow and an R (Shiny) package working in tandem to automate and ease all bioinformatics and analysis tasks involved.

F1000Research 2023, 11:1137 Last updated: 12 MAR 2024

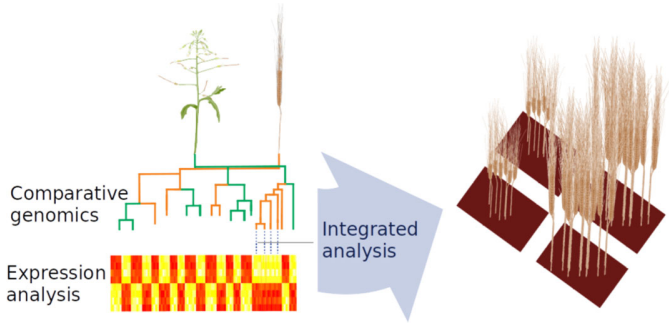


Figure 1. Identification of interesting gene families for crop improvement by integration of differential gene expression with gene family expansion.

The A2TEA.Workflow functions as a one-stop processing pipeline, integrating the prediction of gene families in the form of orthologous groups (OGs) with the analysis of their phylogeny, protein function, and expression, using RNA-seq data from all species. It allows the user to formulate adaptation hypotheses as specific scenarios of gene family expansion in one or several of the genomes, for example, based on a classification of species as stress-tolerant or sensitive, or to identify clade-specific adaptations. As input, the workflow requires for each species a protein FASTA file for orthologous group prediction and RNA-seq reads suitable for differential expression analysis (control vs. treatment), together with either a genomic FASTA file with appropriate gene annotation or a transcriptomic/cDNA FASTA file. Functional information for each species can be provided by the user or can be optionally inferred by our tool AHRD (https://github.com/groupschoof/AHRD) during runtime. The single compressed output is ready for analysis with the R programming language[12] as we took care to create well-structured objects and easy-to-parse outputs. In addition, in order to facilitate immediate and easy exploration and visualization of the results, we created the A2TEA.WebApp written in R Shiny,[13] which allows exploring, highlighting, and exporting the results interactively.

The A2TEA.Workflow combines state-of-the-art bioinformatics software with custom integration steps to combine inferred gene family expansion events with expression results and functional associations.

The workflow is designed as a complete solution starting with raw data and performing upstream quality controls and data transformations automatically. We also took care to allow for a high degree of customizability - e.g., RNA-seq analysis can be performed either alignment-based or using pseudoalignment, and tool-specific parameters can be tweaked in one central config file. Importantly, the workflow is designed to answer biological questions and as such requires the definition of hypotheses in form of combinations of the species of interest. For each hypothesis, the user needs to adjust parameters related to the definition and cutoffs of expansion events. This allows computation of results for several combinations in parallel and facilitates the investigation of many hypotheses downstream e.g., expansion in all tolerant species, in only a specific species, or in all species of a clade.

The A2TEA.WebApp provides an interactive web interface to explore, filter, and visualize the previously generated results via a straightforward tab-structured dashboard design. We took care to create a user-friendly mouse-controlled experience in order to extend the usability from bioinformaticians to experimentalists. The user first uploads the output file of the workflow and chooses the specific “hypothesis” to investigate. This generates a general information tab providing an overview of phylogeny, expression, and set sizes of orthologous groups (OGs) passing the thresholds. The user is then able to switch to dedicated analysis tabs relating to 1) filtering and analyzing OGs with associated data, 2) set size comparisons and tests, and 3) gene ontology (GO) term enrichment analyses. Reactively rendered tables and visualizations are dynamically populated with links to databases such as Ensembl[14] and AmiGO[15] to allow for an immediate follow-up exploration of interesting genes. Tables and graphs can be exported in a variety of formats. The web application also provides a bookmarking system that facilitates the collection and export of the most interesting genes and OGs.

To extend the usability of the workflow by allowing for further species-specific exploration of gene and geneset functional enrichments we integrated the creation of GeneTonic input data files into the A2TEA.Workflow. GeneTonic is a web application that serves as a comprehensive toolkit for streamlining the interpretation of functional enrichment analyses from RNA-seq data.[16] As our workflow is built on Snakemake,[17] the addition of further analyses or outputs allows for modular expansion of its current state. We also intend to add further analyses and features to the A2TEA.WebApp web application.

F1000Research 2023, 11:1137 Last updated: 12 MAR 2024

A2TEA combines best practices in both choice of tools as well as reproducibility and offers a one-stop solution for the integration of genome comparisons with expression and functional data to unravel candidate genes for natural adaptation, e.g. in stress-tolerant plant species. The web application empowers users to explore stress-specific gene family expansions combined with transcriptomic data from their own or published stress experiments by providing interactive visualizations, statistical tests, and dynamically generated database queries.

Both the A2TEA.Workflow and A2TEA.WebApp are freely available at https://github.com/tgstoecker/A2TEA.Workflow and https://github.com/tgstoecker/A2TEA.WebApp, respectively, and archived in Zenodo.[59,60] For demonstration purposes, we also made a public instance of the web application available at https://tgstoecker.shinyapps.io/A2TEA-WebApp.

Methods

Implementation

The A2TEA.Workflow is written in Python and makes use of the Snakemake workflow framework. It leverages the bioconda project channel[18] of the conda package manager to handle software installation and dependency management. Another tool from our lab, AHRD, is integrated as a Git submodule and can be optionally used to infer protein function annotation for any of the species under investigation.

The typical use case for running the workflow consists of cloning the GitHub repository, configuring it to specific needs, and then starting the analyses with either installation of software and dependencies during runtime or usage of a Docker/Singularity container (Figure 2). Modification of the workflow is performed by changing dedicated configuration files controlling samples, species, hypotheses, and tool-specific options. With "hypotheses" we refer to the definition

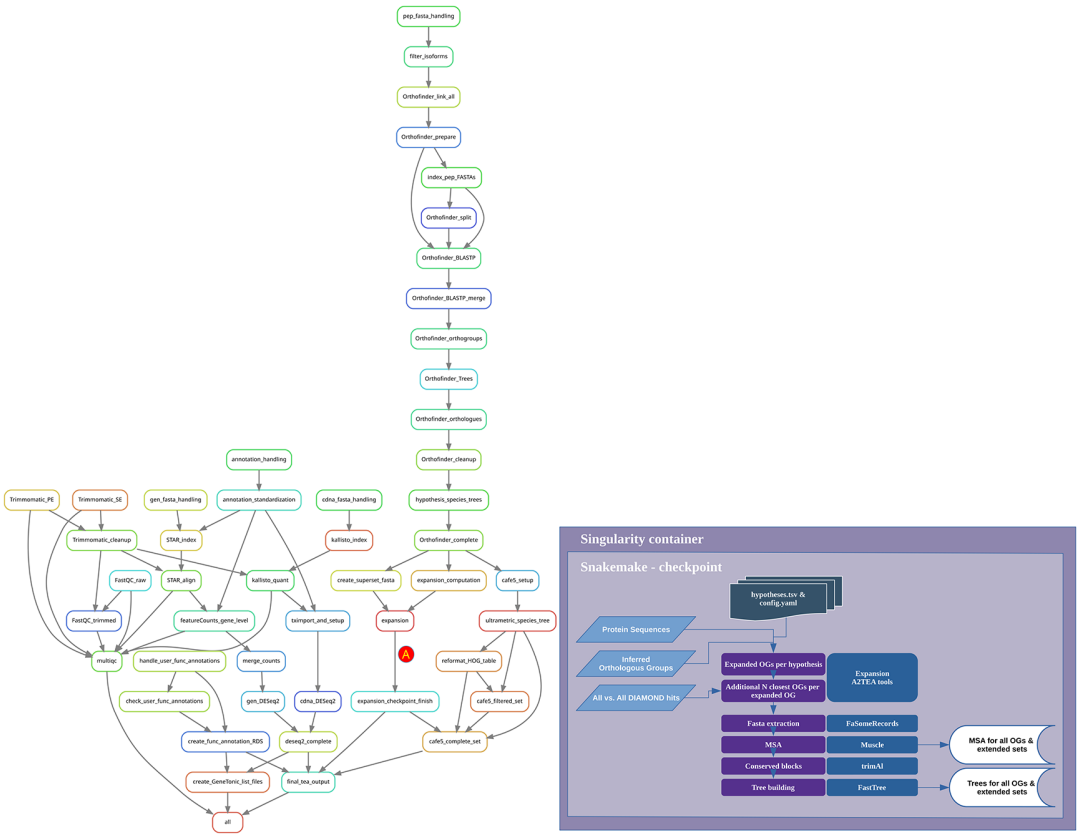


Figure 2. Overview of the A2TEA.Workflow. Workflow diagram of the A2TEA.Workflow displayed as Snakemake rulegraph. After computation of expanded orthologous groups (OGs) (rule expansion - marked with A) the directed acyclic graph (DAG) is re-evaluated since the results are not known beforehand. This Snakemake checkpoint then uses the reciprocal best hits computed by Orthofinder to find the N most similar additional OGs per OG, where N is a variable set by the user. For each OG and additional set of 1 to N additional OGs, multiple sequence alignments and phylogenetic trees are built and used in the downstream steps.

of "gene family expansion" in the set of species under investigation. Several hypotheses can be run in parallel. This multi-hypothesis structure permits the investigation of several defined biological questions, for instance, gene family expansion in stress-tolerant compared to stress-sensitive species. For each hypothesis, we always require the definition of a set of one or more species that should be checked for expansion compared to a second set of one or more species that should not show expansion. For each hypothesis, the user is able to set several options, such as the ratio or the minimal number of genes in a species, to qualify as an expanded OG. The hypotheses.tsv file is structured column-wise with both an index number and a "name" variable used to identify the choices throughout the workflow. Generally, the connection between files and workflow rules is achieved by the species names (e.g., "Arabidopsis_thaliana"). Many hypotheses can be computed in a single workflow with a single final output object that contains all results. This facilitates easy comparisons in the downstream web application, which is especially useful to check the parameter choices for the definition of gene family expansion or when it is necessary to work with unclear trait classification of some species.

The final output generated by the workflow is a single .RData file that can be loaded into an active R environment with the load() command. This provides several separate objects containing all results in a compact form factor:

- HYPOTHESES.a2tea - List object with one S4 object per hypothesis. Each S4 object contains several layers of nested information. E.g., HYPOTHESES.a2tea$hypothesis_2@expanded_OGs$N0.HOG0001225 refers to a specific expanded OG and S4 data object that contains:
  - blast_table (complete BLAST/DIAMOND[19,20] results for OG genes & extended hits)
  - add_OG_analysis (includes multiple sequence alignment (MSA), phylogenetic tree, and gene info for expanded OG and additional OGs based on best BLAST/DIAMOND hits)
- HOG_level_list - List object with one tibble per hypothesis. Information includes OG, number of genes per species, boolean expansion info, number of significant differentially expressed genes (DEGs), and more. The last N list element is a non-redundant superset of all species analyzed over all formulated hypotheses. This makes it easy to create a comparison set e.g., conserved OGs of all species to which the hypothesis subset can then be compared.
- HOG_DE.a2tea - Tibble of DESeq2[21] results for all genes + additional columns.
- A2TEA.fa.seqs - Non-redundant list object containing corresponding amino acid FASTA sequences of all genes/transcripts in the final analysis (this includes those of expanded OGs + those in additional BLAST hits & additional OGs based on user-chosen parameters).
- SFA/SFA_OG_level - Gene/transcript level tables that contain functional predictions (human readable descriptions & GO terms inferred by AHRD).
- hypotheses - A copy of the user-defined hypotheses definitions for the underlying workflow run.
- all_speciesTree - Phylogenetic tree of all species in the workflow run (a non-redundant superset of hypotheses) as inferred by Orthofinder/STAG/Stride.

The .RData output can be investigated inside an R session or via the A2TEA.WebApp, which was specifically designed to allow for interactive inspection, visualization, filtering, and export of the results and subsets. We feature a tutorial for its usage and details on how to work with the results of an A2TEA.Workflow analysis run in the Use Case section and in the project's pkgdown site.

The A2TEA.WebApp is written in the R programming language[12] and uses the Shiny[13] framework to facilitate interactivity with the data. It expects the user to upload a .RData file created by the A2TEA.Workflow. The web application comes with a test dataset that can be loaded with a single click so that interested users can try out its functionality before having to finish an A2TEA.Workflow run.

We developed the web application following community standards and have set up a continuous integration system with GitHub actions that performs build checks of both the package itself and the associated pkgdown site[22] hosted via GitHub pages.

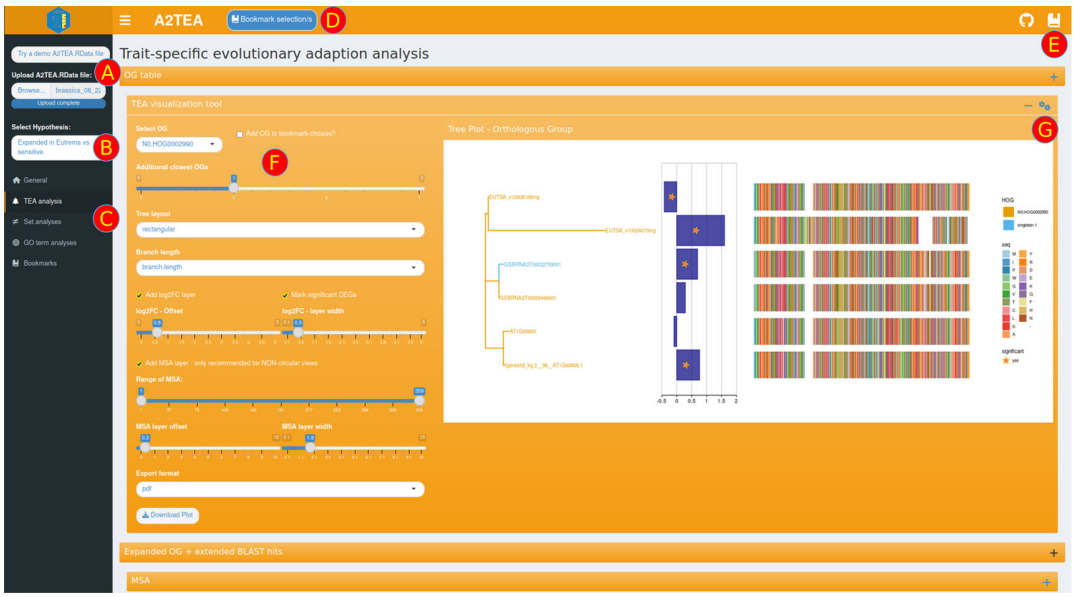


Figure 3. Screenshot of the trait evolutionary analysis tab in the A2TEA.WebApp. The user can either decide to load the included test dataset or upload a .RData result object (A). Other options in the sidebar menu are a selector of the hypothesis (meaning: species comparison) to display or change to another analysis tab (C). Genes, transcripts, or orthologous groups (OGs) can be marked in tables or boxes ticked and then bookmarked with a dedicated button (D). Bookmarks have their dedicated tab but can also be displayed as a sidebar window anywhere for quick reference purposes (E). Analysis- or plot-specific parameters (F) are displayed to the left of the visualization, and a box-specific sidebar window for aesthetic parameters can be opened by clicking the gears icon (G). The underlying dataset investigates drought tolerance among four Brassicaceae species. Displayed here are the maximum likelihood phylogenies of a gene family showing potential subfunctionalization of *Eutrema salsugineum* homologs (top two genes in the tree). Blue bars show log2 (fold change) of gene expression between drought and control conditions, stars mark adj. $p \leq 0.1$ significance cutoff, and the multiple sequence alignment of amino acid sequences is displayed to the right. We provide this particular OG as a bookmarked subset .RData file (see Underlying Data).

The interface is structured in tabs with shinydashboard[23] and shinydashboardPlus,[24] providing the layout infrastructure (shown in Figure 3). The main functionality includes a selector to choose which hypothesis to display (Figure 3B), a sidebar menu that enables the user to switch between different analysis types (Figure 3C), and tool-specific options for parameters, visuals, and export (Figure 3F).

We designed the interface to allow the focus to be put on an individual analysis or plot to gain insight from the data. Plots and tables are contained in collapsible boxes, leaving it up to the user to decide how much information should be displayed at once. Additionally, we tried to separate important parameters from purely aesthetic choices in the plot options, with main options always visible at the side and aesthetic choices reactively displayed after the user switches a box toggle.

Since the exploration of data can be a lengthy process with many iteration cycles, we looked for ways of aiding the user in storing the observations made. Following the example set by the GeneTonic web application,[16] we integrated a bookmarking system that temporarily stores interesting genes/transcripts and OGs. For this, the user needs to mark the respective ID in one of the tables and then click the dedicated bookmarking button displayed at the top of the interface (Figure 3D). All bookmarks are rendered in two reference tables, both in a dedicated tab as well as a pop-up window that can be displayed on every analysis tab. This quick reference is convenient when performing filtering operations in the tables or choosing an interesting OG to display. While these tables can of course, be downloaded, we also implemented a subsetting feature on the bookmarks tab that creates a smaller .RData file with information only pertaining to the bookmarks and associated data. These smaller subsets are fully functioning complete inputs and can be loaded into the application again at a later time - for re-plotting purposes for instance. With this feature, it is straightforward to extract, store and share all information about interesting genes, transcripts, or OGs.

Operation

The A2TEA.Workflow has been primarily designed for use within Linux and requires a standard bash environment with working installations of Snakemake[17] and conda/mamba (Ref. 18, https://github.com/mamba-org/mamba). The former

facilitates compatibility with common cluster setups such as SLURM or LSF. Instructions for a minimal setup are described in the project's README.

For each species, the A2TEA.Workflow should be provided with input RNA-Seq reads (both paired-/single-end possible) suitable for differential expression analysis (control vs. treatment), either a genomic or transcriptomic FASTA file, annotations, and a peptide FASTA file. Since the latest release (v1.1), RNA-Seq data is not required for a workflow run because downstream inferences can still be made by the user in cases where, for one or several species, no expression data for the investigated conditions are available. The user can provide functional information per species, or it can be optionally inferred by our tool AHRD during the workflow. Control of the workflow is handled by several configuration files, which the user needs to adapt to their specific inputs and scientific questions.

The samples.tsv table needs to list all RNA-seq FASTQ files with the columns providing additional information based on which the workflow can infer associations such as species, replicate, and the correct steps to perform. For instance, by leaving out the column for the second paired sample, it is automatically inferred that single-end options have to be used (trimming, mapping, etc.). Operations such as recognizing that files are gzipped and need to be handled appropriately are performed automatically as well. The species.tsv table functions similarly and needs to provide per species information on the FASTA and annotation files, the ploidy of the species, and the location of a file providing the functional information, in the form of GO terms, per protein. If no functional information can be provided the user can choose to add "AHRD" instead of a file path which will trigger a sub-workflow during computation that will create an appropriate file via our functional annotation tool AHRD. Based on whether the user provides a genomic or cDNA FASTA file for a species, the workflow will perform either traditional alignment with STAR[25] or pseudoalignment with kallisto.[26]

The config.yaml controls parameters such as thread usage for individual steps, tool-specific parameters, and parameters relating specifically to the A2TEA.Workflow. Two other very important choices that have to be considered are whether or not automatic filtering for the longest representative isoforms of the peptide FASTA files should be performed and whether gene or transcript level quantification is wanted. Choosing automatic isoform filtering will create a subset peptide FASTA file with only the longest isoform per gene; the header will be shortened to just the gene name identifier. This option must be used in conjunction with gene level quantification since otherwise matching both types of data is not possible.

The notion behind the hypotheses.tsv table is outlined in the Implementation section due to its central importance to the expansion calculation. Here, we briefly want to present some of the available choices the user can consider. Besides defining sets of species that should be analyzed for expansion compared to other sets of species, the user is able to specify the required numerical differences between the two and which OGs to disregard immediately. For instance, "Nmin_expanded_in" takes as input an integer value that defines the minimum number of the investigated species that need to fulfill the expansion criteria in order for the gene family to be called "expanded". These criteria are for instance "min_expansion_factor" and "min_expansion_difference" with which either the minimum factor of expansion or the minimum number of additional genes compared to the non-expanded set of species can be defined. To complement these broad cutoffs, the workflow also integrates a hypothesis-specific CAFE analysis,[27] with which changes in gene family size are analyzed in a way that accounts for phylogenetic history and provides a statistical foundation for evolutionary inferences.

After all choices have been made, the workflow can be started with a single Snakemake command. A2TEA.Workflow will then perform all previously listed steps and merge results into the final output file described in the Implementation section (Figure 2). The user is then able to investigate the integrated and condensed results.

We offer several ways of starting an A2TEA.WebApp instance for downstream investigation of the data: 1) installation with R devtools from our GitHub repository, 2) a docker container with the latest release installed, and lastly, 3) a demo instance hosted on shinyapps.io (https://tgstoecker.shinyapps.io/A2TEA-WebApp/). As the A2TEA.WebApp is an interactive tool with an explorative focus and no strict work order, we illustrate its core operative features in the dedicated Use cases section of this manuscript.

Use cases

In this section, we will illustrate the functionality of the A2TEA.WebApp, using the A2TEA.Workflow results of a three-species analysis of *Hordeum vulgare* (barley), *Zea mays* (maize), and *Oryza sativa japonica* (rice) that investigates adaptive processes in barley to drought stress. Details on the files used as well as their respective publication and SRA accession numbers are listed in detail in both GitHub repositories and the Source data section.[56,57,58]

F1000Research 2023, 11:1137 Last updated: 12 MAR 2024

We integrated this dataset into the workflow and the web application to illustrate the software's setup and to allow for a quick exploration of the tools' functionalities. After cloning the A2TEA.Workflow repository, an additional script can be run (get_test_data.sh) that quickly sets up the experiment by downloading the required input files. Peptide FASTA files are reduced to 2000 proteins; the transcriptomic data is subsampled to 2M reads to allow for a quicker runtime. The functional annotations are precomputed by AHRD. The differential expression analysis is set to be performed on the gene level and two comparisons are performed as defined in the hypotheses.tsv table. These are "Expanded in barley compared to rice and maize" and "Expanded in barley compared to maize". For both, expansion is defined as "number of genes species A $\geq 2 \times$ number of genes of species B".

The final output produced by the workflow is also integrated into the current release of the A2TEA.WebApp and can be loaded via clicking the "Try a demo A2TEA.RData file" at the top of the interface.

Initial inspection of integrated data

The general analysis tab is the default view inside the A2TEA.WebApp. Once input is loaded, reactive information boxes display the number of species, the number of expanded OGs, and the number of DEGs for the currently selected hypothesis. Changing the hypothesis (Figure 3B) e.g., to the second hypothesis in our test set ("Expanded in barley compared to maize"), changes the statistics and all other sets/plots to reflect only the species considered in the hypothesis. Two tables display gene-level differential expression results and functional annotation information (human readable descriptions and GO terms), which allow, for example, the exploration of genes related to a particular function. Also displayed are an inferred phylogenetic tree of the species in the hypothesis subset and an intersection plot (Venn/UpSet) which displays the number of conserved (OG with $\geq$1 gene from every species), overlapping, or species unique OGs and singleton genes. Importantly, a table describing the details of the currently displayed hypothesis is also displayed. All of this facilitates a broad overview of the data and allows the user to spot errors such as faulty hypothesis definitions or cutoffs that are too strict.

Exploring expansion events with annotated phylogenetic trees

The main feature of the TEA (trait-specific evolutionary adaptation) tab is a comprehensive toolkit for the visualization of maximum likelihood phylogenies of expanded OGs and associated information such as the log2(fold change) of the displayed genes and an MSA of the respective protein sequences (Figure 3). The MSA can be added as a geometric layer to the tree plots[28] or displayed separately with additional options such as a conservation bar (Figure 3A). To make an informed decision of which OGs are most worthwhile to investigate closer, a table showcasing the total and significantly differentially expressed genes per OG is also provided. With this, the user is enabled to apply several filters, for example, to select all expanded OGs that possess at least 1 DEG and more than 4 genes from *Hordeum vulgare*. The last table on the tab provides insight into the reciprocal BLAST/DIAMOND hits for the currently chosen OG and the additional most similar OGs. Notably, this table also provides the identifiers given to the proteins by Orthofinder,[29] making it easy to relate insight gained in the web application back to other outputs created by Orthofinder in the A2TEA.Workflow, such as the list of putative xenologs.

Comparing sets of orthologous groups

To describe adaptive processes at a larger scale, we also integrated functionality to visualize distributions of user-defined OG sets and test for their over-representation; e.g., "What is the frequency of OGs that show expansion and at least 1 DEG in *Hordeum vulgare* in all conserved OGs?" and "Is this set over-represented within the background distribution of conserved OGs with at least 1 DEG from any species?". We took care to make answering such questions very accessible by providing the user with text-based choices of which sets to plot or compare. Currently integrated are an enrichment analysis suite allowing for Fisher-Tests and a corresponding circular set plot (Figure 4B) that visualizes the chosen sets. Also provided is a tool for comparing the size distributions of the OGs (Figure 4C) with which group size effects can be checked; e.g., "Do we see differences in the number of DEGs in OGs of a certain size range between the set of interest and the background set?".

Performing functional enrichment tests

The last analysis tab provides options for performing GO term over-representation analysis based on the topGO R package.[30] Functions that occur more often than expected can be identified by setting several parameters that specify the set of OGs the user wants to analyze. With our test data, the user could, for instance, be interested in enriched molecular functions of OGs that are expanded in *Hordeum vulgare* and also possess at least 2 DEGs of *Hordeum vulgare*. Once computed, a table is displayed that shows the top significantly enriched GO terms and also contains dynamically created links for these to AmiGO2.[15] A second table contains information on the corresponding OGs and genes so that the user can follow up on a particular enriched GO term and inspect the underlying data. We also provide two visualizations that summarize the results. The first is a GO enrichment dotplot (Figure 4D) straightforwardly showcasing the overall results,

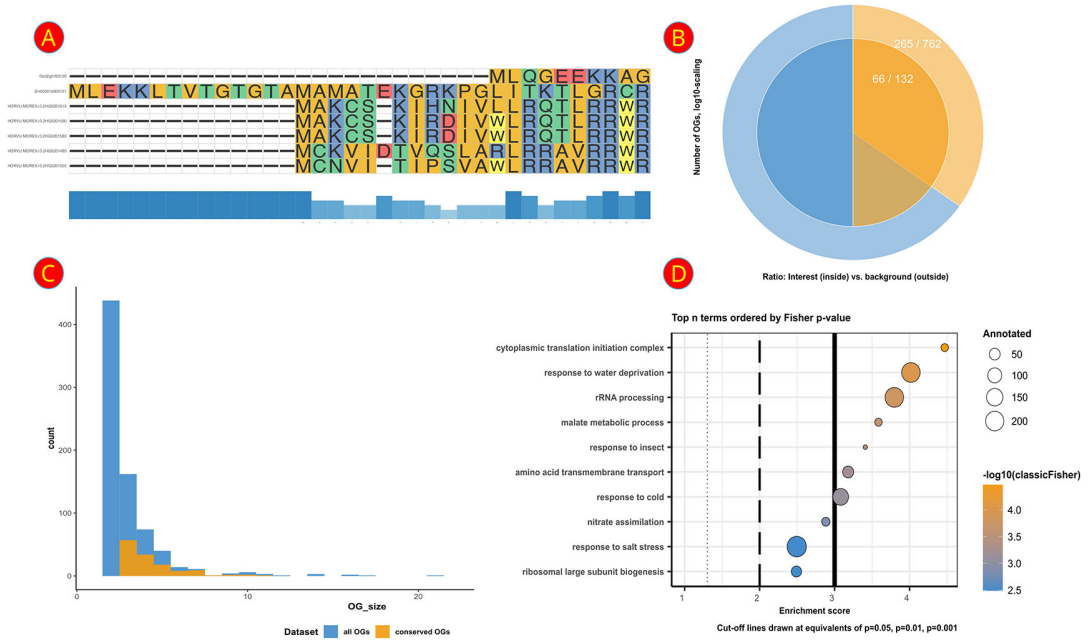


Figure 4. Overview of several analysis plots featured in the A2TEA.WebApp. (A) Multiple sequence alignment of an expanded orthologous group (OG) + additional most similar singletons or OGs. Bars at the bottom represent the degree of conservation. (B) Visual representation of a hypergeometric test for over-representation - colors/layers represent sets in the urn model. The outer ring shows the background set (light blue; "complete background") and the subset in the background set (light orange; "success in background"). The inner ring displays the set of interest (blue; "sample size") and the subset in the set of interest (orange; "success in sample"). (C) Barplot of the total number of OGs per group size (number of genes) between all OGs (blue) and only those OGs that are conserved among the analyzed species (orange). (D) Dotplot of the top over-represented biological processes in the subset of OGs of interest compared to the background; dot size indicates the number of OGs with the respective GO term; color displays negative log [base 10] of the p-values from the enrichment test.

and the second is a GO subgraph of selected top N enriched GO terms. With the latter, we provide the user with an insightful way of investigating how the significant GO terms are distributed over the GO graph.

Export options, bookmarking & ending a session

Tables can be downloaded as .tsv files, and plots are exportable into various formats, such as.pdf, .png, or.svg, allowing the user to easily save and share the observations and results. However, even a relatively small set of species, like the three Poaceae species in our test data, lead to several OGs that are worthwhile to investigate, substantiating the need for the bookmarking system outlined in the Implementation section. It quickly becomes very valuable to bookmark, e.g., all OGs annotated with the top 5 enriched BP GO terms in the OGs expanded in *Hordeum vulgare* if the intention is to return to the analysis later or to generate a list to use with another tool quickly. Relating this to the previous sub-sections, we want to emphasize that bookmarking is integral to using the A2TEA.WebApp and is fully featured on all analysis tabs except the "Set analyses" since here individual genes or OGs are not the focus. To further aid users in the bookmarking process, we also added informative pop-up messages to indicate for instance, that all selected genes/OGs have already been saved. Since the bookmarks can also be used to export a completely functional .RData subset file, only the most relevant information is kept while the processing speed is increased, and all relevant results of the integrative effort are kept. If, for instance, during the analysis, it turned out that hypothesis 2 in our example data ("Expanded in barley compared to maize") is, in fact, not of interest anymore, subsetting the .RData file to interesting OGs of hypothesis 1 completely removes the unneeded "bloat" of hypothesis 2. Similarly, the user could create 2 .RData files (one for each hypothesis) and run a custom script on each separately, efficiently producing hypothesis-specific results.

Discussion

Classic transcriptomic studies produce large lists of gene regulatory information for which, traditionally, pathway or GO term analyses are used to discover the overall molecular trends caused by the experimental treatments.[31] We propose that we can identify novel genes relevant for stress adaptation by comparing same-stress experiments of several plant species with different levels of stress adaptation in combination with evolutionary footprints in the form of protein family expansion. As illustrated in the Methods & Use Cases sections, our novel software tool A2TEA facilitates the

identification of genes associated with the evolution of a trait in a species or a group of related species. Based on the rediscovery of known genes related to the trait, we believe that also novel genes discovered through A2TEA are related to the trait, but experimental verification is in progress, see below. As an example, Figure 3 presents a possible subfunctionalization of gene duplicates in *Eutrema salsugineum*, discovered from data of drought tolerance among four Brassicaceae species (details see Underlying Data). The *A. thaliana* homolog is involved in drought stress response.[32]

Several approaches have been employed to identify potential candidate genes that could provide a genetic basis for more resilient crops. This includes forward genetics approaches such as identifying causative genes for advantageous mutant phenotypes,[33] finding common regulators for several stresses via traditional transcriptomics,[34] usage of Quantitative Trait Locus (QTL) mapping and Genome-wide association studies (GWAS) incl. potential integration with expression data,[35] the combination of expression data with functional information and clustering methods[36,37] and also machine learning based approaches that employ transcriptomic or phenomic data as the basis of their candidate gene predictions.[38,39]

The underlying methods are manifold and include approaches such as Bulked-Segregant analysis,[40] k-means clustering,[41] WGCNA,[42] co-expression networks[43] and set analyses of DEGs often in combination with pathway or GO term enrichment analyses.[31] While most studies share the approach of reducing a list of regulated genes via secondary criteria, to our knowledge, A2TEA is the first openly available tool that specifically combines stress-specific expression data from several species with gene family expansion to unravel candidate genes for stress adaptation in stress-tolerant species.

With A2TEA, we present software that simplifies the complex bioinformatics workflow for the user and provides an interactive web interface for analysis of the results. By using Snakemake as a bioinformatic workflow manager, we remove the need for step-by-step handling of raw data (including software setup and dependencies necessary for computations) and ensure FAIR (findable, accessible, interoperable, and reusable) computational analysis standards.[44] The downstream analysis and visualization framework makes the navigation of the resulting large sets of tabular data faster, more intuitive, and more practicable for scientists without programming skills.[16,31] It offers a variety of summary statistics on the levels of gene family expansion, differential expression, and functional enrichment to ensure quality control. The Shiny framework provides interactivity regarding the visualization and the analysis of the results, and this interactivity highly facilitates the exploration of scientific questions.

Based on user experiences with the web application, we have included analyses and visualizations to allow detecting problems in the bioinformatic predictions, e.g., of orthologous groups (OGs). In order to spot potential misassignments of Orthofinder, close homologs to members of an OG are detected by similarity search and displayed with phylogenetic trees and multiple sequence alignments. A typical case is the non-inclusion of a singleton gene of a species due to a significant portion of protein sequence missing in the annotation, caused e.g., by gaps or sequencing errors in the genome sequence or errors in gene prediction. Similarly, false expansions based on a putative paralog that has only very limited alignment overlap with other members of the OG can be detected. These could be actual duplicates but degenerated through pseudogenization or partial duplication, e.g., the action of transposable elements.

We designed A2TEA with extendability in mind. Both the Snakemake-based A2TEA.Workflow and the A2TEA Shiny App can be easily expanded in a modular fashion to integrate novel features. We are currently testing several additional visualization and testing options. This includes the option for positive selection tests concerning a particular OG, e.g., by calculating the ratio of non-synonymous amino-acid substitutions over synonymous amino-acid substitutions (dN/dS)), distribution comparisons between random and actual DEG-containing OGs, and visualizations for the analysis of general gene/transcript regulation trends. The GO term enrichment functionality is aimed at discovering general trends in the adaptation to the particular stress under investigation. At the moment, the implemented enrichment tests provide options for single over-representation analysis as implemented in the R topGO package.[30] It will be interesting to evaluate and potentially implement further options for functional enrichment analysis in A2TEA, such as modern ensembl approaches[45] or simplification strategies that aid in summarization.[46] Lastly, we intend to implement the option to download a comprehensive RMarkdown/Quarto report summarizing plots and statistics for all bookmarked genes and OGs. This has been demonstrated to be a significant step forward in guaranteeing the portability of results once an interactive session is concluded.[16]

While our research focus is on crops, from a software perspective, A2TEA is entirely independent of the underlying species and can be used with any set of species. This opens the question of how feasible applying the A2TEA methodology to species from other kingdoms might be. Our motivation for developing A2TEA is primarily rooted in the notion that genome duplication played a major role in the evolutionary past of plants.[8] Plant comparative genomics research has shown that gene families are mostly conserved across great evolutionary timescales, comprising even the

diversification of all angiosperms and nonflowering plants.[47,48] Fascinatingly, this conservation of gene families is combined with lineage-specific fluctuations in gene family size, which are frequent among taxa.[8,47–50] This suggests that since comparatively few novel gene families arose, much of the great diversity and phenotypic variation seen in land plants may have arisen primarily due to duplication and adaptive specialization of already existing genes.[48]

While whole genome duplication events are expected and reported less frequently in the animal kingdom and thus gene duplication as a driver of protein family expansion does not play as prominent a role in animals as in plants,[51,52] protein family expansion is still an important driver of adaptation.[53] We expect that A2TEA will be useful in non-plant species, even if protein family expansion only represents a small portion of adaptive changes, with other sources of variation, like alternative splicing playing a potentially more important role.[54]

Currently, we are investigating several publicly available genomic and transcriptomic datasets from various groups of plant species with A2TEA. While we expect to detect candidate genes relevant to adaptation to the stress being investigated, this assumption is based on the rediscovery of known genes. One important follow-up step is thus to experimentally verify the impact of selected candidate genes in vivo. To this end, we perform stress experiments in plants bearing knockout mutations in candidate genes predicted by A2TEA, using sequence-indexed mutant collections such as BonnMu.[55] This will allow us to assess the phenotypic impact of these mutations and, thus, the role of these genes in the tolerance of the stress. While testing all candidates will not be feasible, the rate of genes relevant to the trait under investigation among tested candidates will represent an estimate of the prediction performance.

Conclusions

With the availability of multiple genome sequence and RNA-seq data sets, it is now possible to combine comparative evolutionary analyses, in our case protein family expansion, with differential expression to predict genes involved in adaptive traits. However, running the required bioinformatics analyses and data integration tasks as well as summarizing and visualizing the results, remains challenging. A2TEA only requires standard data files as input, follows best practice software standards for both reproducibility and portability, and provides a user-friendly web application for interactive exploration and selection of the most promising candidate genes. We show that genes known to be involved in stress tolerance can be detected in datasets of stress-tolerant and stress-sensitive plants, but we expect A2TEA to be useful in a broader scope when analyzing protein families and their expression in multiple genomes as the parameters for selecting interesting families are very flexible. A2TEA follows a positive trend in modern research software development that provides easy installation and execution through the use of container and workflow technologies as well as interactive visualization and exploration tools for the generated results. Combined, this facilitates better reproducibility, communication, and shareability of comprehensive analyses.

Data availability

Source data

Poaceae test data:

Transcriptomics:

Hordeum vulgare: SRR6782243, SRR6782247, SRR6782257, SRR6782249, SRR6782250, SRR6782254

Zea mays: SRR2043219, SRR2043217, SRR2043190, SRR2043220, SRR2043226, SRR2043227

Oryza sativa japonica: SRR5134063, SRR5134064, SRR5134065, SRR5134066

These correspond to the following studies relating on drought stress:

Hordeum vulgare: https://doi.org/10.1186/s12864-019-5634-0

Zea mays: https://doi.org/10.1104/pp.16.01045

Oryza sativa japonica: https://doi.org/10.3389/fpls.2017.00580

Assemblies & annotations hosted on EnsemblPlants:

Hordeum vulgare: cDNA FASTA, GTF, Peptide FASTA

Zea mays: Genome FASTA, GTF, Peptide FASTA

Oryza sativa japonica: cDNA FASTA, GTF, Peptide FASTA

An archived version of the complete grasses test data (reduced as used in the examples) is deposited here:

https://zenodo.org/record/7089022[56]

Data used in the Brassicaceae example:

Transcriptomics:

Eutrema salsugineum: SRR7624684, SRR7624685, SRR7624692, SRR7624687, SRR7624721, SRR7624722

Arabidopsis lyrata: SRR7624680, SRR7624702, SRR7624703, SRR7624732, SRR7624733, SRR7624742

Arabidopsis thaliana: SRR7624694, SRR7624696, SRR7624697, SRR7624710, SRR7624714, SRR7624723

Brassica napus: SRR12429701, SRR12429702, SRR12429703, SRR12429698, SRR12429699, SRR12429700

These correspond to the following studies on drought stress response: *Eutrema salsugineum, Arabidopsis lyrata, Arabidopsis thaliana*: https://doi.org/10.1111/nph.15841

Brassica napus: PRJNA656507

Assemblies & annotations hosted on EnsemblPlants:

Eutrema salsugineum: Genome FASTA, GTF, Peptide FASTA

Arabidopsis lyrata; Genome FASTA, GTF, Peptide FASTA

Arabidopsis thaliana: Genome FASTA, GTF, Peptide FASTA

Brassica napus: Genome FASTA, GTF, Peptide FASTA

Underlying data

The results generated by the A2TEA.Workflow which are also used for demonstrating the A2TEA.WebApp's functionality presented in this work are available at https://zenodo.org/record/7089608[57] and https://zenodo.org/record/7089606.[58]

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Software availability

Both the A2TEA.Workflow and the A2TEA.WebApp are available as MIT licensed open source softwares.

- Software available from: https://tgstoecker.github.io/A2TEA.WebApp
- Source code available from: https://github.com/tgstoecker/A2TEA.Workflow, https://github.com/tgstoecker/A2TEA.WebApp
- Archived source code at time of publication: https://zenodo.org/record/7725859,[59] https://zenodo.org/record/7750290[60]
- License: MIT

References

1. Muktar MS, Lübeck J, Strahwald J, *et al.*: **Selection and validation of potato candidate genes for maturity corrected resistance to Phytophthora infestans based on differential expression combined with SNP association and linkage mapping.** *Front. Genet.* 2015; **6**: 294.
PubMed Abstract | Publisher Full Text | Free Full Text

2. Lovell JT, Mullen JL, Lowry DB, *et al.*: **Exploiting differential gene expression and epistasis to discover candidate genes for drought-associated QTLs in Arabidopsis thaliana.** *Plant Cell.* 2015; **27**(4): 969–983.
PubMed Abstract | Publisher Full Text | Free Full Text

3. Lin D, Lan L, Zheng T, *et al.*: **Comparative genomics reveals recent adaptive evolution in Himalayan giant honeybee Apis laboriosa.** *Genome Biol. Evol.* 2021; **13**(10): evab227.
Publisher Full Text

4. Wang J, Hu H, Liang X, *et al.*: **High-quality genome assembly and comparative genomic profiling of yellowhorn (*Xanthoceras sorbifolia*) revealed environmental adaptation footprints and seed oil contents variations.** *Front. Plant Sci.* 2023; **14**: 976.
PubMed Abstract | Publisher Full Text | Free Full Text

5. Ohno S: *Evolution by gene duplication.* Springer Science & Business Media; 2013.

6. Taylor JS, Raes J: **Duplication and divergence: The evolution.** *Annu. Rev. Genet.* 2004; **38**: 615–643.
PubMed Abstract | Publisher Full Text

7. Qiao X, Zhang S, Paterson AH: **Pervasive genome duplications across the plant tree of life and their links to major evolutionary innovations and transitions.** *Comput. Struct. Biotechnol. J.* 2022; **20**: 3248–3256.
PubMed Abstract | Publisher Full Text

8. Panchy N, Lehti-Shiu M, Shiu S-H: **Evolution of gene duplication in plants.** *Plant Physiol.* 2016; **171**(4): 2294–2316.
PubMed Abstract | Publisher Full Text

9. Adams KL, Wendel JF: **Polyploidy and genome evolution in plants.** *Curr. Opin. Plant Biol.* 2005; **8**(2): 135–141.
Publisher Full Text

10. Maere S, De Bodt S, Raes J, *et al.*: **Modeling gene and genome duplications in eukaryotes.** *Proc. Natl. Acad. Sci.* 2005; **102**(15): 5454–5459.
PubMed Abstract | Publisher Full Text

11. Voordeckers K, Pougach K, Verstrepen KJ: **How do regulatory networks evolve and expand throughout evolution?** *Curr. Opin. Biotechnol.* 2015; **34**: 180–188.
PubMed Abstract | Publisher Full Text

12. R Core Team: *R: A Language and Environment for Statistical Computing.* 2022.
Reference Source

13. Chang W, Cheng J, Allaire JJ, *et al.*: *shiny: Web Application Framework for R.* 2022. R package version 1.7.2.
Reference Source

14. Cunningham F, Allen JE, Allen J, *et al.*: **Ensembl 2022.** *Nucleic Acids Res.* 2022; **50**(D1): D988–D995.

15. Carbon S, Ireland A, Mungall CJ, *et al.*: **Amigo: online access to ontology and annotation data.** *Bioinformatics.* 2009; **25**(2): 288–289.
PubMed Abstract | Publisher Full Text

16. Marini F, Ludt A, Linke J, *et al.*: **Genetonic: an r/bioconductor package for streamlining the interpretation of rna-seq data.** *BMC Bioinform.* 2021; **22**(1): 1–19.

17. Mölder F, Jablonski KP, Letcher B, *et al.*: **Sustainable data analysis with snakemake.** *F1000Res.* 2021; **10**: 33.
PubMed Abstract | Publisher Full Text

18. Grüning B, Dale R, Sjödin A, *et al.*: **Bioconda: sustainable and comprehensive software distribution for the life sciences.** *Nat. Methods.* 2018; **15**(7): 475–476.
PubMed Abstract | Publisher Full Text

19. Altschul SF, Gish W, Miller W, *et al.*: **Basic local alignment search tool.** *J. Mol. Biol.* 1990; **215**(3): 403–410.
PubMed Abstract | Publisher Full Text

20. Buchfink B, Xie C, Huson DH: **Fast and sensitive protein alignment using DIAMOND.** *Nat. Methods.* 2015; **12**(1): 59–60.
PubMed Abstract | Publisher Full Text

21. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.** *Genome Biol.* 2014; **15**(12): 1–21.
PubMed Abstract | Publisher Full Text | Free Full Text

22. Wickham H, Hesselberth J, Salmon M: *pkgdown: Make Static HTML Documentation for a Package.* 2022. R package version 2.0.6.
Reference Source

23. Chang W, Ribeiro BB: *shinydashboard: Create Dashboards with 'Shiny'.* 2021. R package version 0.7.2.
Reference Source

24. Granjon D: *shinydashboardPlus: Add More 'AdminLTE2' Components to 'shinydashboard'.* 2021. R package version 2.0.3.
Reference Source

25. Dobin A, Davis CA, Schlesinger F, *et al.*: **Star: ultrafast universal rna-seq aligner.** *Bioinformatics.* 2013; **29**(1): 15–21.
PubMed Abstract | Publisher Full Text

26. Bray NL, Pimentel H, Melsted P, *et al.*: **Near-optimal probabilistic rna-seq quantification.** *Nat. Biotechnol.* 2016; **34**(5): 525–527.
Publisher Full Text

27. Mendes FK, Vanderpool D, Fulton B, *et al.*: **Cafe 5 models variation in evolutionary rates among gene families.** *Bioinformatics.* 2021; **36**(22-23): 5516–5518.

28. Yu G, Smith DK, Zhu H, *et al.*: **ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data.** *Methods Ecol. Evol.* 2017; **8**(1): 28–36.
Publisher Full Text

29. Emms DM, Kelly S: **Orthofinder: phylogenetic orthology inference for comparative genomics.** *Genome Biol.* 2019; **20**(1): 1–14.

30. Alexa A, Rahnenführer J: **Gene set enrichment analysis with topgo.** *Bioconductor Improv.* 2009; **27**: 1–26.

31. Supek F, Škunca N: **Visualizing go annotations.** *The Gene Ontology Handbook.* New York, NY: Humana Press; 2017; pp. 207–220.

32. Ling Q, Jarvis P: *Current Biology, and undefined 2015. Regulation of chloroplast protein import by the ubiquitin e3 ligase sp1 is important for stress tolerance in plants.* Elsevier; 2015.
Reference Source

33. Kirschner GK, Rosignoli S, Guo L, *et al.*: **Enhanced gravitropism 2 encodes a sterile alpha motif–containing protein that controls root growth angle in barley and wheat.** *Proc. Natl. Acad. Sci.* 2021; **118**(35): e2101526118.
PubMed Abstract | Publisher Full Text

34. Sham A, Moustafa K, Al-Ameri S, *et al.*: **Identification of arabidopsis candidate genes in response to biotic and abiotic stresses using comparative microarrays.** *PLoS One.* 2015; **10**(5): e0125666.
PubMed Abstract | Publisher Full Text

35. Guo T, Yang J, Li D, *et al.*: **Integrating gwas, qtl, mapping and rna-seq to identify candidate genes for seed vigor in rice (oryza sativa l.).** *Mol. Breed.* 2019; **39**(6): 1–16.

36. Sewelam N, Brilhaus D, Bräutigam A, *et al.*: **Molecular plant responses to combined abiotic stresses put a spotlight on unknown and abundant genes.** *J. Exp. Bot.* 2020; **71**(16): 5098–5112.
PubMed Abstract | Publisher Full Text

37. Kar S, Mai H-J, Khalouf H, *et al.*: **Comparative transcriptomics of lowland rice varieties uncovers novel candidate genes for adaptive iron excess tolerance.** *Plant Cell Physiol.* 2021; **62**(4): 624–640.
PubMed Abstract | Publisher Full Text

38. Shaik R, Ramakrishna W: **Machine learning approaches distinguish multiple stress conditions using stress-responsive genes and identify candidate genes for broad resistance in rice.** *Plant Physiol.* 2014; **164**(1): 481–495.
PubMed Abstract | Publisher Full Text

39. Braun IR, Yanarella CF, Lawrence-Dill CJ: **Computing on phenotypic descriptions for candidate gene discovery and crop improvement.** 2020.
Publisher Full Text | Reference Source | Reference Source

40. Michelmore RW, Paran I, Kesseli RV: **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.* 1991; **88**: 9828–9832. 00278424.
Publisher Full Text | Reference Source

41. Likas A, Vlassis N, Verbeek JJ: **The global k-means clustering algorithm.** *Pattern Recogn.* 2003; **36**(2): 451–461.
Publisher Full Text

42. Langfelder P, Horvath S: **Wgcna: an r package for weighted correlation network analysis.** *BMC Bioinform.* 2008; **9**(1): 1–13.

43. Aoki K, Ogata Y, Shibata D: **Approaches for extracting practical information from gene co-expression networks in plant biology.** *Plant Cell Physiol.* 2007; **48**(3): 381–390.
PubMed Abstract | Publisher Full Text

44. Wratten L, Wilm A, Göke J: **Reproducible, scalable, and shareable analysis pipelines with bioinformatics workflow managers.** *Nat. Methods.* 2021; **18**(10): 1161–1168.
PubMed Abstract | Publisher Full Text

45. Alhamdoosh M, Ng M, Wilson NJ, *et al.*: **Combining multiple tools outperforms individual methods in gene set enrichment analyses.** *Bioinformatics.* 2017; **33**(3): 414–424.
PubMed Abstract | Publisher Full Text

46. Zuguang G, Hübschmann D: **Simplify enrichment: A bioconductor package for clustering and visualizing functional enrichment results.** *Genom. Proteom. Bioinf.* 2022.
Publisher Full Text

47. Rensing SA, Lang D, Zimmer AD, *et al.*: **The physcomitrella genome reveals evolutionary insights into the conquest of land by plants.** *Science.* 2008; **319**(5859): 64–69.
PubMed Abstract | Publisher Full Text

48. Flagel LE, Wendel JF: **Gene duplication and evolutionary novelty in plants.** *New Phytol.* 2009; **183**(3): 557–564.
Publisher Full Text

49. Velasco R, Zharkikh A, Troggio M, *et al.*: **A high quality draft consensus sequence of the genome of a heterozygous grapevine variety.** *PLoS One.* 2007; **2**(12): e1326.
PubMed Abstract | Publisher Full Text

50. Ming R, Hou S, Feng Y, *et al.*: **The draft genome of the transgenic tropical fruit tree papaya (carica papaya linnaeus).** *Nature.* 2008; **452**(7190): 991–996.
PubMed Abstract | Publisher Full Text

51. Mable BK: **'why polyploidy is rarer in animals than in plants': myths and mechanisms.** *Biol. J. Linn. Soc.* 2004; **82**(4): 453–466.
Publisher Full Text

52. Murat F, Van de Peer Y, Salse J: **Decoding plant and animal genome plasticity from differential paleo-evolutionary patterns and processes.** *Genome Biol. Evol.* 2012; **4**(9): 917–928.
PubMed Abstract | Publisher Full Text

53. Demuth JP, De Bie T, Stajich JE, *et al.*: **The evolution of mammalian gene families.** *PLoS One.* 2006; **1**(1): e85.
PubMed Abstract | Publisher Full Text

54. Kim E, Magen A, Ast G: **Different levels of alternative splicing among eukaryotes.** *Nucleic Acids Res.* 2007; **35**(1): 125–131.
PubMed Abstract | Publisher Full Text

55. Marcon C, Altrogge L, Win YN, *et al.*: **Bonnmu: a sequence-indexed resource of transposon-induced maize mutations for functional genomics studies.** *Plant Physiol.* 2020; **184**(2): 620–631.
PubMed Abstract | Publisher Full Text

56. Stöcker T: **A2TEA.Workflow test data (v1.0.0) [Data set]. Zenodo.** 2022.
Publisher Full Text

57. Stöcker T: **A2TEA.Workflow Poaceae reduced example data (v1.0.0) [Data set]. Zenodo.** 2022.
Publisher Full Text

58. Stöcker T: **A2TEA Brassicaceae example data (v.1.0.0) [Data set]. Zenodo.** 2022.
Publisher Full Text

59. Stöcker T: **tgstoecker/A2TEA.Workflow: No transcriptomes required (v1.1.0). Zenodo.** 2023.
Publisher Full Text

60. Stöcker T: **tgstoecker/A2TEA.WebApp: v1.1.5 (v1.1.5). Zenodo.** 2023.
Publisher Full Text

6 General Discussion

Modern crop breeding and improvement are fundamentally based on understanding the genomic underpinnings that shape traits such as yield, resistances, or nutritional quality (Atlaf et al., 2021; Wani et al., 2022; Ku et al., 2020). Both basic research and continuous breeding efforts are of great importance since the exponential increase of the worldwide population – estimated to reach 9 billion people by 2050 – will require an increase of between 70% to 100% more food (Godfray et al., 2010). This is all the more critical when considering that climate change caused prolonged drought and further degradation processes will continuously result in large areas of land becoming non-arable (Prăvălie et al., 2018). The world faces an immense and complex challenge needing to counter production losses caused by extreme weather events while aiming for a zero net-expansion of agricultural land and both significantly reducing and offsetting greenhouse gas emissions by reforesting at least 585 million hectares (World Resources Report; Searchinger et al., 2014).

While production technology has increased yields, this is probably maxed out in industrialized farming, but important potential remains in genetics/the yield potential of plants (Godfray et al., 2010). Increased use of genetically modified crops also promises to support climate change mitigation with e.g., an increased European adoption estimated to "result in a reduction equivalent to 7.5% of the total agricultural greenhouse gas emissions of Europe" (Kovak et al., 2022). Increasing genomic resources, understanding mechanistic underpinnings of plant traits, and developing novel approaches for crop improvement therefore constitute scientific avenues to contribute to safeguarding and improving the global food supply. From a basic science, molecular biology perspective these are all ultimately rooted in the study of gene functional relationships.

To this end, the projects presented in this thesis – while employing different methodologies – all investigated the relationship between genotype and phenotype in crops to contribute to securing a global food supply and paving new ways of crop improvement. Creating large pools of mutant plants and determining the germinal target sites of the transposable element mutagen allowed for the creation of an expansive resource of interesting phenotypes with linked genomic information (Chapter 2). Identification of causal gene/s for any given interesting phenotype is a logical follow-up and was successfully performed and validated extensively in a wax mutant (Chapter 3). Crop yields in the field are heavily impacted by environmental factors and by various biotic and abiotic stresses. Thereby, it becomes necessary to also study gene regulatory (and ultimately physiological) changes that shape the genotype-phenotype relationship. Presented in this thesis is a study on the impacts of both salt and drought stress on the transcriptome of poplar roots (Chapter 4). Lastly, Chapter 5 details a novel computational approach

for predicting genes important for stress resistance. This approach combines transcriptomic and evolutionary data to find gene families that show signs of trait adaptation. As a purely *in silico* approach, validation is still needed but will be performed with the mutants of the *BonnMu* resource described in Chapter 2.

6.1 *BonnMu*/MuWU – Creating a Fundamental Resource for the Study of Gene Functional Relationships in a Major Crop

In Chapter 2, the European sequence-indexed insertional library in maize (*BonnMu*) as well as an accompanying bioinformatics analysis software were presented. The *BonnMu* mutant collection is an effort that we expanded after its first publication in 2020. Consistently, we were able to continuously add genes tagged by germinal *Mu* insertions to the resource. Since then, *BonnMu* has achieved a gene-mutation coverage of 83% of all high-confidence gene models in the maize genome (B73v5; Win & Stöcker et al., 2024). *Zea mays* is a worldwide leading crop as a food, feed, fuel, and fiber source and additionally possesses the broadest cultivation range of all crops with great diversity in sequence, structure, and morphology (Tenaillon and Charcosset, 2011). *BonnMu* has garnered significant interest from researchers around the globe for easy access to mutants useful in forward and reverse genetics research. Integral to this effort was the formalization of the aforementioned computational workflow tool – the presented *Mu*-seq workflow utility (MuWU; Stöcker et al., 2022). By trivializing the recurrent processing of the underlying sequencing approach of *Mu* insertional mutagenesis, the identification and annotation of *Mu* insertion sites in newly generated *Mu*-seq libraries was automated. Notably, MuWU is not specific for *Mutator* insertions but is capable of automatic analysis of any insertion event that causes target site duplications and is, therefore, species agnostic and useful for similar research projects in the future (Stöcker et al., 2022).

Analysis of each new *BonnMu* sequencing library has confirmed the longstanding reports of *Mu* elements preferentially inserting in regions with epigenetic marks distinctive for open chromatin and in particular around transcription start site (TSS) and 5' UTR of genes (Dietrich et al., 2002; Liu et al., 2009; Liang et al., 2019; Marcon et al., 2020; Win & Stöcker et al., 2024). In a subsequent publication, currently in review, we also worked on gaining new insights on *Mu* elements and extending the information provided for each insertion (Win & Stöcker et al., 2024). The simplicity of the *Mu* insertion pattern has recently been challenged (Zhang et al., 2020) with a more specific but as of yet unclear tethering mechanism underlying the association with near-TSS/5' UTR regions being hypothesized. Also, a clear association with highly expressed genes and genes with conserved biological ("house-keeping") function was indicated. Building on these notions, we investigated whether a mechanism associated with evolutionary novelty might be an influencer of this higher insertion frequency or if merely the amount of expression commonly found in these housekeeping genes is in fact the main driver behind this dis-

tribution pattern. While our results remained inconclusive, further such investigations might also shed light on the reasons behind the remaining set of genes that has never been tagged by a *Mu* insertion in any of the *BonnMu* sequencing libraries generated so far.

Zea mays is well-suited for an analysis of this kind, as its evolutionary past has been unraveled enough to distinguish between delimited "sets" of genes. The maize genome underwent a whole genome duplication (WGD) event between 5-12 mya ago (Swigonova et al., 2004) and subsequently, the then n=20 genome consisting of two subgenomes underwent several chromosome breakages and fusions, which reduced the chromosome count back to n=10, with each subgenome losing parts of its DNA (Schnable et al., 2011; Hughes et al., 2014). In total, the maize genome underwent 62 rearrangement events, 39 inversions, 14 inversions, and eight translocations (Wei et al., 2007). Additionally, the WGD event still serves as the main source for gene diversification due to mutations in the homologs and due to TE insertions "reshuffling" redundant genetic material (Guo et al., 2019). This leads to novel genes that are not syntenic to Sorghum and are therefore deemed as non-syntenic. Redundant genes are gradually lost; mainly by an intrachromosomal recombination deletion mechanism, referred to as "fractionation" (Woodhouse et al., 2010; Liang & Schnable, 2018). This process is biased towards just one subgenome in maize, while the other genome loses fewer genes. Further research has shown that for each pair of chromosomes, one retained more syntenic genes, than the other (Schnable et al., 2011). The subgenomes were subsequently classified as subgenome 1, containing chromosomes with more syntenic genes, and subgenome 2, containing chromosomes with less syntenic genes to Sorghum (Schnable et al., 2011). Additionally, it could be seen that the two subgenomes have different expression rates. Genes on subgenome 2 seem to be lower expressed than subgenome 1, which is underlined by the finding that mutations on subgenome 1 account for more phenotypic variation (Renny-Byfield et al., 2017). Combining all of these insights we investigated correlations between synteny and *Mu* insertions on the subgenomes of maize.

While our results (unpublished) and previous research suggests that all TE families show little maize subgenome bias (Anderson et al., 2019; Stitzer et al., 2021) increasing evidence also suggests that TE-mediated DNA-methylation changes following hybridization events may be one of the key factors involved in fractionation and subgenome dominance (Edger et al., 2017). Therefore, investigating the interaction of both chromatin and methylation signals together with one of the most active TE families that also displays a distinct preference for genic regions may reveal novel mechanistic insights into maize (sub-)genome differentiation.

In addition to trying to better understand the forces behind the *Mu* insertion distribution pattern, we have also aimed at differentiating between different kinds of *Mu* elements. Since there are in fact multiple *Mu* species with sequence alterations in both their main transposon cassette as well as their TIR region (Lisch, 2002), this could offer three advantages: First, a finer granularity might reveal *Mu* species-specific level distributions

that could reveal a better picture of the chromatin insertion drivers than the generalized "bulk" *Mu* type investigations performed up until now. Second, since the nature of the *BonnMu* resource requires the acquired stocks to be screened via PCR for validation of the insertions, differentiating between *Mu* species leads to more specific primers being available. Third, since new *Mu* elements are still being described (Tan et al., 2011; Li et al.; 2013) as of yet unknown *Mu* elements could be discovered that might offer insights into the sequence divergence of *Mu* in our stocks.

There are also further interesting analyses that might be worth pursuing; e.g., since it has become evident that *Mu* insertions mark open chromatin signals and tend to cluster at the 5' UTR of genes, such notable accumulations at positions not yet associated with any particular genetic element might harbor as of yet unidentified genes and could be investigated via an expression-insertion overlap.

6.2 Establishment of a BSR-Seq Bioinformatics Pipeline for the Identification of Mutant Phenotype causing Genes

The various strategies (radioactive, chemical, biological) employed in untargeted mutagenesis (Shelake et al., 2019) come with distinct disadvantages for follow-up analyses and breeding efforts. In a simple scenario, a single causal gene for an induced phenotype of interest needs to be identified in follow-up experiments. This is made difficult by the usually large amount of induced genomic alterations that all could be the cause of the functional divergence that is observed. Other scenarios exist, such as several genes' protein products having changed or their net amount altered; this can lead to more complex situations in which combinations of mutations are responsible for the observed phenotype (Østergaard & Yanofsky, 2004). In the *BonnMu* project, the approach includes mapping TE insertion sites, which (in theory) reduces the challenge of finding the causal gene/s significantly. However, for any F_2-family there is still an average of 37 germinal mutations (Marcon et al., 2020) to contend with, all of which could be the reason behind the altered functioning of the plant's morphology/physiology. Also, the potential for a novel, non-*Mu* mutation or genetic material from the used *Mutator* line being the ultimate cause is always a possibility.

A reliable approach to locating one or several causal loci is Bulked segregant analysis which combines a backcrossing strategy with bioinformatics analyses which was explained in depth in Chapter 1 of this thesis (Michelmore et al., 1991). This approach is based on the insight that a complete linkage of one or several markers with the mutated gene/s would lead to only one or a distinct set of marker alleles being detected in the mutant plant/s. A mapping population is thus typically created that consists of wild type and mutant plants whose DNA samples are pooled respectively and subsequently analyzed in terms of their allele frequencies. By backcrossing the plants several times with a wild

type cultivar with the aim of producing a mutant and non-mutant bulk, the genomic differences are continuously reduced and effectively "narrow in" on the causal region of the mutant-phenotype individuals. A common and cost-effective variation of this approach employs RNA-Seq data and uses single nucleotide polymorphisms (SNPs) as the analyzed markers – Bulked segregant analysis RNA-seq (BSR-Seq; Liu et al., 2012). It bears special mention that employing BSR-Seq comes with the inherent limitation of only generating data representing (depending on the protocol) parts of the transcriptome. In cases in which SNPs are only used as markers to identify regions that are most likely to harbor, e.g., a causal *Mu* insertion, this is unproblematic. However, if the expectation of the causal mutation is a SNP and the intention is to identify this mutation directly, the lack of data outside the transcribed and sequenced RNA regions comes with the risk of not being able to spot it (e.g., the causal SNP is located in an intron or upstream of the 5' UTR). Therefore in many cases, after identifying a high-confident, causative locus, more steps such as fine-mapping or screening a final set of genes are necessary to pinpoint the ultimate gene/s responsible for the altered phenotype (Li et al., 2013; Li et al., 2016; Huang et al., 2017; Shi et al., 2019).

While several such investigations are currently underway with maize mutants generated as part of the *BonnMu* effort, Chapter 3 presents a study employing BSR-Seq in barley with a focus on wax (*eceriferum*) mutants (Müller et al., 2023). In brief, in our study, we identified two barley genes (*CER-ZA* and *CER-YE*) encoding enzymes active in fatty acids and wax synthesis (acyl-CoA reductase "FAR"; β-ketoacyl-CoA synthase "KCS"). We used the mutants *cer-za* and *cer-ye* to characterize changes in their cuticular wax composition and allocation as compared to the control barley plants. Our study confirms the previous knowledge on the role of epicuticular and intracuticular wax in the limitation of water loss: the intracuticular wax is the main water transport barrier, while the epicuticular wax plays only a minor (if any) role. The most notable novel insight of this study is that we were able to discover and report on the *CER-ZA* gene (*HORVU5Hr1G089230*), the first FAR gene isolated from barley (orthologous by function to Arabidopsis *AtFAR3*), which demonstrates a novel function in wax synthesis in barley. All subsequent steps (isolation of the gene; ectopic expression in yeast and Arabidopsis; the use of genome editing to create knockouts) were contingent on identifying this strong candidate gene that explained the observed mutant phenotype. Since the publication focuses on the aforementioned characterization analyses, in the following, the bioinformatics analysis will be discussed. Employing a BSR-Seq strategy with eight times backcrossed and thereby near-isogenic lines (NILs) was ultimately the key to discovering the causal gene. This likely allowed the bioinformatics steps of variant calling and analysis on the generated RNA-Seq data to produce decidedly differentiating signals between the mutant and wildtype bulks. The pipeline established in this study combined various tools of the genome analysis toolkit (GATK; Van der Auwera et al., 2013; Poplin et al., 2017) with the analysis and visualization capabilities of the QTLseqR R package (Mansfeld & Grumet, 2018). Once sequenced, the bulks underwent a rigorous quality control and variant calling pipeline, since one of the main challenges

at this point stems from sequencing errors and depth biases distorting the identification of the polymorphisms. Specific to the chosen RNA-Seq-based approach, is the chance of allele-specific expression comprising the reported ratios between the bulks at any given position (Wang et al., 2013). The common mathematical approaches for evaluating the remaining high-confident SNPs between the bulks; i.e., ΔSNP-index (Takagi et al., 2013) and G' values (Magwene et al., 2011) require careful analysis of SNP ratio distributions, organism-specific parameters (e.g., genome size) and filtering steps in order to provide meaningful results. BSA-/BSR-Seq are thus not "push-button" analyses but rather require careful evaluation at each step to generate correct and insightful results.

All this however, only provided confirmation that the causal locus was located in the 580–660 Mb QTL on chromosome 5, which had already been established prior (Druka et al., 2011; Bregitzer et al., 2013). Ultimately, it was the integration of several types of data that led to the discovery of *HORVU5Hr1G089230*; filtering the list of genes with the highest bulk ratio differences for only those with expression changes between the bulks and screening the remainder for their predicted functions. Although ultimately worthwhile, this approach can only be described as very fortunate. While frameshift mutations can lead to reduced expression (Brault & Miller, 1991), mutations leading to non-synonymous base exchanges can lead to nonsense or significantly altered protein products, which in turn causes altered function and ultimately phenotype, while expression remains the same (Brassac et al., 2021). Therefore performing an expression filter as described was simply a guess that turned out to be correct. Similarly fortunate was the functional inference via orthology transfer. *Arabidopsis thaliana* is the plant species with the richest genomics datasets available, and the generation of this data has often been motivated by the promise of mapping the results to crop species (International Arabidopsis Informatics Consortium, 2010). It is thus satisfying to see this promise fulfilled and to benefit from the groundwork of the plant functional genomics pioneers.

The conventional (and usually required) approach to identifying the causal genetic variants responsible for a trait of interest in a QTL region is "fine mapping". The term describes a collection of methods, most of which can be described as high-resolution genetic mapping techniques used to narrow down a QTL to a smaller region of the genome (Spain & Barrett, 2015). Some of the most common approaches are: extending the NILs by further backcrosses (Steiner et al., 2019), extending computational models in genomic best linear unbiased prediction (GBLUP), or GWAS with genomic data of NILs (Peiffer et al., 2014) or recombinant inbred lines (RILs;Pan et al., 2018), joint-linkage mapping of QTL (Lu et al., 2010), PCR based high-resolution melting (HRM; Laila et al., 2019), mapping-by-sequencing (MBS; Jamann et al., 2015) and specific-locus amplified fragment sequencing (SLAF-seq; Xu et al., 2015). While these follow-up analyses were not necessary, they offer tried avenues of identifying causal genes behind interesting phenotypes.

The presented, successful wax mutant study served as the basis for several *BonnMu* forward-genetics projects, as the established BSR-Seq-based bioinformatics work-

flow has since been used in three separate mutant characterization projects, all in various stages of completion.

6.3 A2TEA – a novel Software Approach for Predicting Genes important for Stress Tolerance via Inference of Evolutionary Adaptation Events

In Chapter 5, a novel approach and accompanying software suite (data workflow and analysis web application) for predicting genes important in stress resistance was presented. Uncovering the genetic basis of tolerance allows utilizing this information in breeding/crop optimization. As our publication described, the development of A2TEA (Automated Assessment of Trait-specific Evolutionary Adaptations) was primarily rooted in the notion that genome duplication played a major role in the evolution of plants. More specifically, the observation that although comparatively few novel gene families arose in their evolutionary past, much of the astounding phenotypic differences observable in land plants may be largely explainable by duplication first and subsequent adaptive specialization of already existing genes (Flagel & Wendel, 2009; Stöcker et al., 2023). Initially designed to identify families where expansion yields an adaptive advantage, more general evolutionary scenarios can be explored.

Orthology remains difficult to determine based on sequence similarity alone (Mao et al., 2006), so integration of evolutionary (phylogenetic trees, gene families) and functional (differential expression) information is advantageous. This can form the basis for transfer of information across species also without a focus on family expansions. It is important to consider that further ways of approaching adaptation research have been effectively employed; e.g., those building on top of different genomic concepts for grouping genes and not relying on cross-species transfer. For instance, Battlay et al. were able to show recently that large haploblocks (putative inversions) contribute to local, rapid adaptation in the invasive weed Ambrosia artemisiifolia (Battlay et al., 2023). These were proposed to harbor cassettes of genes that are both co-selected, enriched for certain biological functions and ultimately segregate akin to single-alleles with a strong effect.

While we made an active effort of programming A2TEA in a decidedly accessible way that would allow broad applicability to any set of species and emphasized this potential in the publication, it is important to discuss if an expansion-based approach is also applicable in genomes where WGD events are not commonplace. Consistent with the importance of the early and increased establishment of mating opportunities it has been argued that WGD events and polyploidy should most commonly occur in organisms that share certain traits such as self-fertilization, a large number of gametes and a perennial lifestyle (Mable, 2004). A combination of these traits distinguishes plants e.g., from the animal kingdom in which fewer WGD events and polyploidy are both expected and re-

ported. More obstacles for WGD in animals comes from expected imbalances between sex chromosomes and autosomes, disruption of developmental processes by gene expression dosage effects, and generally lower developmental plasticity in animals than plants as they do not possess pluripotent meristem cells (Muller, 1925; Orr, 1990; Mable, 2004; Mable et al., 2011; Wertheim 2013). It is thus likely that in non-plant organisms other methods of evolutionary adaptation such as alternative splicing or pathway/network rewiring could play a bigger proportional role (Stöcker et al., 2023). Nevertheless, in eukaryotes such as hexapods or teleost fishes, clear examples exist of WGD as a source of evolutionary novelty and being a fundamental source of adaptive evolution (Li et al., 2018; Moriyama & Koshiba-Takeuchi, 2018). Also, one must not disregard other potent sources of novelties such as tandem duplications or co-option of single-copy genes. Therefore the application of A2TEA to adaptation research in species from other kingdoms could be feasible if specifically duplication history in their evolutionary past is suspected to have contributed to interesting trait adaptation.

The key takeaway is that an understanding of gene functional relationships – and in this particular case how function meets an adaptation need arises from duplicated genetic material in evolution – are of interest to not only basic science but also can demonstratively offer new avenues in the applied field of plant breeding. Perhaps the most notable and species spanning example of this can be found in the MADS-box family of genes. This highly conserved family underwent several gene duplication events and the resulting diversification of gene functions has been shown to be important in controlling flowering time as well as tissue shape and flower color (Alvarez-Buylla et al., 2000; Becker & Theißen 2003; Hsu et al., 2021). These insights are essential to many breeding programs in both monocots and dicots as influencing heading date or flowering development is both of interest to the breeding programs themselves as well as the target cultivars (Jeon et al., 2000; Tomes et al., 2023).

6.4 Future Perspectives & Conclusion

As technology advances the microscopic lens through which plant molecular biology is analyzed is continuously refined. So quick are these developments that some of the computational approaches described in this thesis are already starting to be outdated compared to the latest available bioinformatics approaches. Modern, sequencing-based research methodologies allow plant molecular biology to be unraveled even at the level of single cells (Shaw et al., 2021). However, single-cell sequencing in plants was long held back compared to human and animal applications due to several reasons. Firstly, the cell walls found in plants render the required steps of single-cell suspension considerably more difficult (Xu, et al., 2022). Secondly, for many/most plant species, well-established protoplast isolation protocols are still lacking and forced the research community to contend with Arabidopsis focussed studies for several years (Conde et al., 2021). Thirdly, often

encountered circumstances of plant research – minimal amounts of starting material, and loss and/or contamination of samples – make ultimate analysis more difficult than in well-established animal organisms (Xu and Zhou, 2018). Still, continuous progress in dealing with these issues will allow for an even more fine-grained understanding of many developmental processes and genetic responses to environmental stimuli.

These latest technologies, such as single-cell approaches or the establishment of high-throughput protein sequencing (Brinkerhoff et al., 2021) promise to offer both exciting and quite easily achievable novel insights. This mirrors how each leap in microscope generation facilitated new levels of understanding by allowing higher magnifications and ultimately the generation of new types of data/signals. However, it is imperative to not overlook the large amounts of already generated data. The data revolution caused by the staggering progress of information technology has left its mark on molecular biology in that it created a paradoxical situation in which more analyzable data exist than often are even used. It is not an uncommon observation that datasets are generated e.g., for the analysis of expression changes of only a subset of genes while the data would offer more information on other parts of, if not the whole transcriptome under the experimental conditions. Fortunately, most respected journals require the deposition of such data in public databases as part of the publication process for the benefit of the research community. The power of using previously generated data has been demonstrated many times and ranges from small re-analyses with novel insights to monumental efforts such as the Serratus project that uses the petabytes of accessible sequencing data to continuously advance the discovery and surveillance of viruses (Edgar et al., 2022).

Once integrated, processed, and visualized, sequencing data enable "low hanging fruit" insights to be gleaned instantaneously; for instance via plant focussed web services such as the Bioanalytic Resource's electronic fluorescent pictograph browsers (Winter et al., 2017) and Ensembl (Bolser et al., 2016), or organism-specific community websites like MaizeGDB (Portwood et al., 2018). "How does this gene's expression change in this tissue with these environmental circumstances acting upon it?" – combined with expanding and improving functional information, truly powerful insights can be gleaned without the need for performing additional experiments. Relating this to the works presented in this thesis, the listed resources were, for example, repeatedly employed to analyze the underlying genetic impacts causing particular mutant phenotypes of the *BonnMu* resource (Marcon et al., 2020; Win & Stöcker et al., 2024). The growing availability of previously generated data for integrative analyses was also both the requirement and motivating force behind my multi-year software project A2TEA (Stöcker et al., 2023), which demonstrated a novel data integration strategy. Thus a major learning spanning the totality of the research presented in this thesis is how proper data management facilitates follow-up research being initiated. For instance the *BonnMu* effort was significantly aided by the automation capabilities of the MuWU software (Stöcker et al., 2022). Similarly, the comprehensive table of expression changes created for *Populus x canescens* (Grünhofer et al., 2022) was central to formulating the hypotheses of a successful follow up research project

also investigating poplar root suberization (Grünhofer et al., 2024). This underlines the importance of Bioinformatics adhering to best practices in data stewardship which can then facilitate valuable new knowledge gains within the agri-science context.

Yet all of the aforementioned is ultimately only feasible if modern molecular biology data is managed appropriately. The "quality" of research data is a dimension that has only quite recently received truly globe-spanning and exhaustive attention. Adherence to the FAIR principle (findable, accessible, interoperable, and reusable) data standards promises to ensure high quality and more straightforward integrative science and research directions (Wilkinson et al., 2016).

Both not chasing the latest methodologies for the sake of their novelty alone as well as sustainable data practices, are important, however, a third aspect often overlooked is actionability. In the domain of agriculture, more specifically crop science, basic research investigating e.g., root mutants with steeper growth angles (Kirschner et al., 2021) or modified wax composition of leaves (Müller et al., 2023) represent avenues with which, in theory, optimized, more drought resilient crops can be bred. Insightful science is achieved by recognizing that "phenotype" for such complex traits is a result of both genetic and environmental factors. Therefore research projects such as the one presented in Chapter 4 (Grünhofer et al., 2022), investigating the similarities and differences of physiological reactions to various environmental conditions and stresses are central to our understanding of the complexity of gene function under specific conditions. In the study, we exposed poplar roots to osmotic or salt stress conditions and were able to show that while both phenotypic alterations and global transcriptomic expression patterns in reaction to both types of stress were highly comparable, there were also remarkable differences setting them apart. We created physiological and metabolomic reaction tables so that the expression data we obtained could easily be mined for biological insights also by other researchers, thereby extending the aforementioned resources and providing easily accessible insights. Taken together, our results suggest that the genetic enhancement of poplar root suberization could lead to genetically optimized poplar trees which might then be suitable for cultivation on water-limited and/or salty land that would be unusable for food production, thus being of great benefit to agroforestry. The informative value of our results is limited though, as they are based on 19-23 week old poplar plants and thus cannot represent reliable transcriptomic and physiological insights of adult poplars.

This highlights the need for validation of results generated *in silico* not only in a laboratory but also in realistic agronomic settings, such as the field or greenhouse. Ultimately, the principle that can be derived is that transdisciplinary approaches and interoperable data standards for crop improvement are greatly beneficial since genes of interest determined by softwares such as MuWU or A2TEA (or more precisely: associated genotypes) need to be tested under real-world farming environments. Several national large-scale initiatives have begun to alleviate this need. In Germany, the National Research Data Infrastructure (NFDI) aims to create a permanent digital repository of knowledge for

the sciences. Here the FAIRagro consortium (Specka et al., 2023), together with further consortia with touch points with the inherently interdisciplinary field of agriculture, service the need for agricultural research data management and exchange outlined above. In the United States, the Agricultural Genome to Phenome Initiative (AG2PI; Tuggle et al., 2022) specifically aims at assembling transdisciplinary communities for large-scale genome-to-phenome research and development efforts.

Integrative, smart science adhering to FAIR guidelines is required to move molecular biology forward in a speedy fashion so that each new generation of scientists is capable of efficiently building on the ever-expanding knowledge and data available to them. By following the outlined principles and using the tools and insights that are the outcomes of this thesis, future research in the (molecular) crop sciences will be facilitated which will help the scientific effort to overcome the looming resource crises of the 21st century.

6.5 References

Altaf, A., Gull, S., Shah, A. Z., Faheem, M., Saeed, A., Khan, I., ... & Zhu, M. (2021). Advanced genetic strategies for improving rice yield. *Journal of Global Innovations in Agricultural Sciences, 9*, 167-171.

Battlay, P., Wilson, J., Bieker, V. C., Lee, C., Prapas, D., Petersen, B., ... & Hodgins, K. A. (2023). Large haploblocks underlie rapid adaptation in the invasive weed *Ambrosia artemisiifolia*. *Nature communications, 14*(1), 1717.

Bolser, D., Staines, D. M., Pritchard, E., & Kersey, P. (2016). Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomics data. *Plant bioinformatics: Methods and protocols*, 115-140.

Brassac, J., Muqaddasi, Q. H., Plieske, J., Ganal, M. W., & Röder, M. S. (2021). Linkage mapping identifies a non-synonymous mutation in *FLOWERING LOCUS T (FT-B1)* increasing spikelet number per spike. *Scientific Reports, 11*(1), 1585.

Brault, V., & Miller, W. A. (1992). Translational frameshifting mediated by a viral sequence in plant cells. *Proceedings of the National Academy of Sciences, 89*(6), 2262-2266.

Brinkerhoff, H., Kang, A. S., Liu, J., Aksimentiev, A., & Dekker, C. (2021). Multiple rereads of single proteins at single–amino acid resolution using nanopores. *Science, 374*(6574), 1509-1513.

Conde, D., Triozzi, P. M., Balmant, K. M., Doty, A. L., Miranda, M., Boullosa, A., ... & Kirst, M. (2021). A robust method of nuclei isolation for single-cell RNA sequencing of solid tissues from the plant genus *Populus*. *PLoS One, 16*(5), e0251149.

Dietrich, C. R., Cui, F., Packila, M. L., Li, J., Ashlock, D. A., Nikolau, B. J., & 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*(2), 697-716.

Edgar, R. C., Taylor, J., Lin, V., Altman, T., Barbera, P., Meleshko, D., ... & Babaian, A. (2022). Petabase-scale sequence alignment catalyses viral discovery. *Nature, 602*(7895), 142-147.

Edger, P. P., Smith, R., McKain, M. R., Cooley, A. M., Vallejo-Marin, M., Yuan, Y., ... & Puzey, J. R. (2017). Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *The Plant Cell, 29*(9), 2150-2167.

Flagel, L. E., & Wendel, J. F. (2009). Gene duplication and evolutionary novelty in plants. *New Phytologist, 183*(3), 557-564.

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., ... & Toulmin, C. (2010). Food security: the challenge of feeding 9 billion people. *science, 327*(5967), 812-818.

Grünhofer, P., Heimerich, I., Pohl, S., Oertel, M., Meng, H., Zi, L., ... & Schreiber, L. (2024). Suberin deficiency and its effect on the transport physiology of young poplar roots. *New Phytologist*.

Guo, H., Jiao, Y., Tan, X., Wang, X., Huang, X., Jin, H., & Paterson, A. H. (2019). Gene duplication and genetic innovation in cereal genomes. *Genome research, 29*(2), 261-269.

Huang, Z., Peng, G., Liu, X., Deora, A., Falk, K. C., Gossen, B. D., ... & Yu, F. (2017). Fine mapping of a clubroot resistance gene in Chinese cabbage using SNP markers identified from bulked segregant RNA sequencing. *Frontiers in plant science, 8*, 284684.

Hughes, T. E., Langdale, J. A., & Kelly, S. (2014). The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. *Genome research, 24*(8), 1348-1355.

International Arabidopsis Informatics Consortium. (2010). An international bioinformatics infrastructure to underpin the Arabidopsis community. *The Plant Cell, 22*(8), 2530-2536.

Jamann, T. M., Balint-Kurti, P. J., & Holland, J. B. (2015). QTL mapping using high-throughput sequencing. *Plant functional genomics: methods and protocols*, 257-285.

Jeon, J. S., Lee, S., Jung, K. H., Yang, W. S., Yi, G. H., Oh, B. G., & An, G. (2000). Production of transgenic rice plants showing reduced heading date and plant height by ectopic expression of rice MADS-box genes. *Molecular Breeding, 6*, 581-592.

Kovak, E., Blaustein-Rejto, D., & Qaim, M. (2022). Genetically modified crops support climate change mitigation. *Trends in Plant Science, 27(7)*, 627-629.

Ku, H. K., & Ha, S. H. (2020). Improving nutritional and functional quality by genome editing of crops: status and perspectives. *Frontiers in plant science, 11*, 577313.

Laila, R., Park, J. I., Robin, A. H. K., Natarajan, S., Vijayakumar, H., Shirasawa, K., ... & Nou, I. S. (2019). Mapping of a novel clubroot resistance QTL using ddRAD-seq in Chinese cabbage (*Brassica rapa* L.). *BMC plant biology, 19*, 1-9.

Li, L., Li, D., Liu, S., Ma, X., Dietrich, C. R., Hu, H. C., ... & Schnable, P. S. (2013). 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*(12), e82333.

Li, Y., Harris, L., & Dooner, H. K. (2013). TED, an autonomous and rare maize transposon of the *Mutator* superfamily with a high gametophytic excision frequency. *The Plant Cell, 25*(9), 3251-3265.

Li, L., Hey, S., Liu, S., Liu, Q., McNinch, C., Hu, H. C., ... & Hochholdinger, F. (2016). Characterization of maize roothairless6 which encodes a D-type cellulose synthase and controls the switch from bulge formation to tip growth. *Scientific Reports, 6*(1), 34395.

Li, Z., Tiley, G. P., Galuska, S. R., Reardon, C. R., Kidder, T. I., Rundell, R. J., & Barker, M. S. (2018). Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proceedings of the National Academy of Sciences, 115*(18), 4713-4718.

Liang, Z., & Schnable, J. C. (2018). Functional divergence between subgenomes and gene pairs after whole genome duplications. *Molecular plant, 11*(3), 388-397.

Liang, L., Zhou, L., Tang, Y., Li, N., Song, T., Shao, W., ... & Song, R. (2019). A sequence-indexed *Mutator* insertional library for maize functional genomics study. *Plant physiology, 181*(4), 1404-1414.

Lisch, D. (2002). *Mutator* transposons. *Trends in plant science, 7*(11), 498-504.

Liu, S., Yeh, C. T., Ji, T., Ying, K., Wu, H., Tang, H. M., ... & Schnable, P. S. (2009). *Mu* transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. *PLoS genetics, 5*(11), e1000733.

Lu, Y., Zhang, S., Shah, T., Xie, C., Hao, Z., Li, X., ... & Xu, Y. (2010). Joint linkage–linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. *Proceedings of the National Academy of Sciences, 107*(45), 19585-19590.

Mable, B. K. (2004). 'Why polyploidy is rarer in animals than in plants': myths and mechanisms. *Biological Journal of the Linnean Society, 82*(4), 453-466.

Mable, B. K., Alexandrou, M. A., & Taylor, M. I. (2011). Genome duplication in amphibians and fish: an extended synthesis. *Journal of zoology, 284*(3), 151-182.

Mansfeld, B. N., & Grumet, R. (2018). QTLseqr: An R package for bulk segregant analysis with next-generation sequencing. The plant genome, 11(2), 180006.

Mao, F., Su, Z., Olman, V., Dam, P., Liu, Z., & Xu, Y. (2006). Mapping of orthologous genes in the context of biological pathways: An application of integer programming. *Proceedings of the National Academy of Sciences, 103*(1), 129-134.

Michelmore, R. W., Paran, I., & Kesseli, R. (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. *Proceedings of the national academy of sciences, 88*(21), 9828-9832.

Moriyama, Y., & Koshiba-Takeuchi, K. (2018). Significance of whole-genome duplications on the emergence of evolutionary novelties. *Briefings in functional genomics, 17*(5), 329-338.

Muller, H. J. (1925). Why polyploidy is rarer in animals than in plants. *The American Naturalist*, 59(663), 346-353.

Orr, H. A. (1990). "Why polyploidy is rarer in animals than in plants" revisited. *The American Naturalist, 136*(6), 759-770.

Østergaard, L., & Yanofsky, M. F. (2004). Establishing gene function by mutagenesis in Arabidopsis thaliana. *The Plant Journal, 39*(5), 682-696.

Pan, L., He, J., Zhao, T., Xing, G., Wang, Y., Yu, D., ... & Gai, J. (2018). Efficient QTL detection of flowering date in a soybean RIL population using the novel restricted two-stage multi-locus GWAS procedure. *Theoretical and Applied Genetics, 131*, 2581-2599.

Peiffer, J. A., Romay, M. C., Gore, M. A., Flint-Garcia, S. A., Zhang, Z., Millard, M. J., ... & Buckler, E. S. (2014). The genetic architecture of maize height. Genetics, 196(4), 1337-1356.

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., ... & Banks, E. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. *BioRxiv*, 201178.

Portwood, J. L., Woodhouse, M. R., Cannon, E. K., Gardiner, J. M., Harper, L. C., Schaeffer, M. L., ... & Andorf, C. M. (2019). MaizeGDB 2018: the maize multi-genome genetics and genomics database. *Nucleic acids research, 47*(D1), D1146-D1154.

Prăvălie, R., Patriche, C., Borrelli, P., Panagos, P., Roșca, B., Dumitraşcu, M., ... & Bandoc, G. (2021). Arable lands under the pressure of multiple land degradation processes. A global perspective. *Environmental research, 194*, 110697.

Renny-Byfield, S., Rodgers-Melnick, E., & Ross-Ibarra, J. (2017). Gene fractionation and function in the ancient subgenomes of maize. *Molecular biology and evolution, 34*(8), 1825-1832.

Schnable, J. C., Springer, N. M., & 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, 108*(10), 4069-4074.

Searchinger, T., Hanson, C., Ranganathan, J., Lipinski, B., Waite, R., Winterbottom, R., ... & Ari, T. B. (2014). *Creating a sustainable food future. A menu of solutions to sustainably feed more than 9 billion people by 2050. World resources report 2013-14: interim findings* (pp. 154-p). World Resources Institute (WRI); World Bank Groupe-Banque Mondiale; United Nations Environment Programme (UNEP); United Nations Development Programme (UNDP); Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD); Institut National de la Recherche Agronomique (INRA).

Shaw, R., Tian, X., & Xu, J. (2021). Single-cell transcriptome analysis in plants: advances and challenges. Molecular Plant, 14(1), 115-126.

Shelake, R. M., Pramanik, D., & Kim, J. Y. (2019). Evolution of plant mutagenesis tools: a shifting paradigm from random to targeted genome editing. *Plant Biotechnology Reports, 13*, 423-445.

Shi, L., Jiang, C., He, Q., Habekuß, A., Ordon, F., Luan, H., ... & Yang, P. (2019). Bulked segregant RNA-sequencing (BSR-seq) identified a novel rare allele of eIF4E effective against multiple isolates of BaYMV/BaMMV. *Theoretical and Applied Genetics, 132*, 1777-1788.

Spain, S. L., & Barrett, J. C. (2015). Strategies for fine-mapping complex traits. *Human molecular genetics, 24*(R1), R111-R119.

Specka, X., Martini, D., Weiland, C., Arend, D., Asseng, S., Boehm, F., ... & Ewert, F. (2023). FAIRagro: Ein Konsortium in der Nationalen Forschungsdateninfrastruktur (NFDI) für Forschungsdaten in der Agrosystemforschung: Herausforderungen und Lösungsansätze für den Aufbau einer FAIRen Forschungsdateninfrastruktur. *Informatik Spektrum, 46*(1), 24-35.

Steiner, B., Buerstmayr, M., Wagner, C., Danler, A., Eshonkulov, B., Ehn, M., & Buerstmayr, H. (2019). Fine-mapping of the Fusarium head blight resistance QTL *Qfhs.ifa-5A* identifies two resistance QTL associated with anther extrusion. *Theoretical and applied genetics, 132*, 2039-2053.

Swigoňová, Z., Lai, J., Ma, J., Ramakrishna, W., Llaca, V., Bennetzen, J. L., & Messing, J. (2004). Close split of sorghum and maize genome progenitors. *Genome research, 14*(10a), 1916-1923.

Tan, B. C., Chen, Z., Shen, Y., Zhang, Y., Lai, J., & Sun, S. S. (2011). Identification of an active new *mutator* transposable element in maize. *G3: Genes| Genomes| Genetics, 1*(4), 293-302.

Tenaillon, M. I., & Charcosset, A. (2011). A European perspective on maize history. *Comptes rendus biologies, 334*(3), 221-228.

Tomes, S., Gunaseelan, K., Schaffer, R. J., & Varkonyi-Gasic, E. (2023). A MADS-box gene-induced early flowering pear (*Pyrus communis* L.) for accelerated pear breeding. *Frontiers in Plant Science, 14*, 1235963.

Tuggle, C. K., Clarke, J., Dekkers, J. C., Ertl, D., Lawrence-Dill, C. J., Lyons, E., ... & Schnable, P. S. (2022). The Agricultural Genome to Phenome Initiative (AG2PI): creating a shared vision across crop and livestock research communities. *Genome biology, 23*, 1-11.

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., ... & DePristo, M. A. (2013). From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. *Current protocols in bioinformatics, 43*(1), 11-10.

Wani, S. H., Choudhary, M., Barmukh, R., Bagaria, P. K., Samantara, K., Razzaq, A., ... & Varshney, R. K. (2022). Molecular mechanisms, genetic mapping, and genome editing for insect pest resistance in field crops. *Theoretical and Applied Genetics, 135*(11), 3875-3895.

Wei, F., Coe, E. D., Nelson, W., Bharti, A. K., Engler, F., Butler, E., ... & Wing, R. A. (2007). Physical and genetic structure of the maize genome reflects its complex evolutionary history. *PLoS genetics, 3*(7), e123.

Wertheim, B., Beukeboom, L. W., & Van de Zande, L. (2013). Polyploidy in animals: effects of gene expression on sex determination, evolution and ecology. *Cytogenetic and genome research, 140*(2-4), 256-269.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data, 3*(1), 1-9.

Win, Y. N., Stöcker, T., Du, X., Brox, A., Pitz, M., Klaus, A., ... & Marcon, C. (2024). Expanding the *BonnMu* sequence-indexed repository of transposon induced maize (*Zea mays* L.) mutations in dent and flint germplasm. *bioRxiv*, 2024-02.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V., & Provart, N. J. (2007). An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. *PloS one, 2*(8), e718.

Woodhouse, M. R., Schnable, J. C., Pedersen, B. S., 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 homeologs. *PLoS biology, 8*(6), e1000409.

Xu, X., Lu, L., Zhu, B., Xu, Q., Qi, X., & Chen, X. (2015). QTL mapping of cucumber fruit flesh thickness by SLAF-seq. Scientific reports, 5(1), 15829.

Xu, Y., & Zhou, X. (2018). Applications of single-cell sequencing for multiomics. *Computational Systems Biology: Methods and Protocols*, 327-374.

Xu, Z., Wang, Q., Zhu, X., Wang, G., Qin, Y., Ding, F., ... & Jin, S. (2022). Plant Single Cell Transcriptome Hub (PsctH): an integrated online tool to explore the plant single-cell transcriptome landscape. *Plant Biotechnology Journal, 20*(1), 10.

Zhang, X., Zhao, M., McCarty, D. R., & Lisch, D. (2020). Transposable elements employ distinct integration strategies with respect to transcriptional landscapes in eukaryotic genomes. *Nucleic acids research, 48*(12), 6685-6698.

Supplementary Figures for First-Author Publications

Supplementary figures for MuWU: Mutant-seq library analysis and annotation

MuWU: Mutant-seq library analysis and annotation
Tyll Stöcker, Lena Altrogge, Caroline Marcon, Yan Naing Win, Frank Hochholdinger, Heiko Schoof
Bioinformatics 38.3 (2022): 837-838.

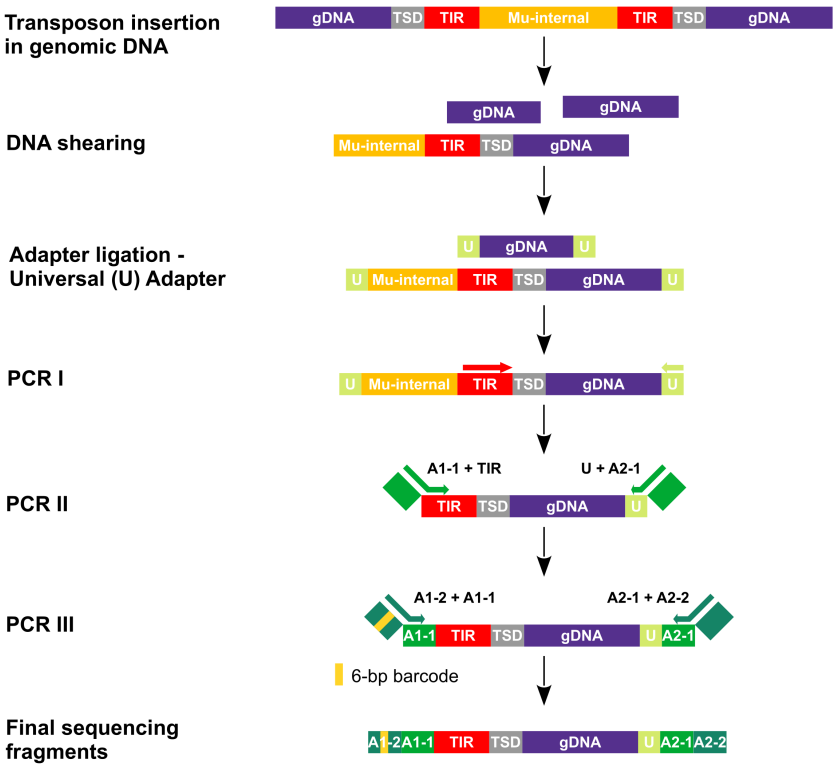


Figure S1: Overview of Mu-seq library construction.
Mu-seq library construction (Marcon et al., 2020) starts with genomic DNA isolation from pools of 576 maize families, followed by DNA shearing and ligation of a universal (U) adapter. PCR I enriches Mu-TIR flanking DNA. PCR II+III incorporate Illumina sequencing adapters A1 and A2 respectively and a 6-bp barcode enabling multiplexing of 48 pools in a single library.

	Col_01	Col_02	Col_03	Col_04	Col_05	Col_06	Col_07	Col_08	Col_09	Col_10	Col_11	Col_12	Col_13	Col_14	Col_15	Col_16	Col_17	Col_18	Col_19	Col_20	Col_21	Col_22	Col_23	Col_24
Row_01	B-0074	B-0075	B-0076	B-0650	B-0078	B-0079	B-0080	B-0081	B-0082	B-0083	B-0084	B-0085	B-0086	B-0087	B-0088	B-0089	B-0090	B-0091	B-0092	B-0093	B-0094	B-0095	B-0096	B-0097
Row_02	B-0098	B-0099	B-0100	B-0101	B-0102	B-0103	B-0104	B-0105	B-0106	B-0107	B-0108	B-0109	B-0110	B-0111	B-0112	B-0113	B-0114	B-0115	B-0116	B-0117	B-0118	B-0119	B-0120	B-0121
Row_03	B-0122	B-0123	B-0124	B-0125	B-0126	B-0127	B-0128	B-0129	B-0130	B-0131	B-0132	B-0133	B-0134	B-0135	B-0136	B-0137	B-0138	B-0139	B-0140	B-0141	B-0142	B-0143	B-0144	B-0145
Row_04	B-0146	B-0147	B-0148	B-0149	B-0150	B-0151	B-0152	B-0153	B-0154	B-0155	B-0156	B-0157	B-0158	B-0159	B-0160	B-0161	B-0162	B-0163	B-0164	B-0165	B-0166	B-0167	B-0168	B-0169
Row_05	B-0170	B-0171	B-0172	B-0173	B-0174	B-0175	B-0176	B-0177	B-0178	B-0179	B-0180	B-0181	B-0182	B-0183	B-0184	B-0185	B-0186	B-0187	B-0188	B-0189	B-0190	B-0191	B-0192	B-0193
Row_06	B-0194	B-0195	B-0196	B-0197	B-0198	B-0199	B-0200	B-0201	B-0202	B-0203	B-0204	B-0205	B-0206	B-0207	B-0208	B-0209	B-0210	B-0211	B-0212	B-0213	B-0214	B-0215	B-0216	B-0217
Row_07	B-0218	B-0219	B-0220	B-0221	B-0222	B-0223	B-0224	B-0225	B-0226	B-0227	B-0228	B-0229	B-0230	B-0231	B-0232	B-0233	B-0234	B-0235	B-0236	B-0237	B-0238	B-0239	B-0240	B-0241
Row_08	B-0242	B-0243	B-0244	B-0245	B-0246	B-0247	B-0248	B-0249	B-0250	B-0251	B-0252	B-0253	B-0254	B-0255	B-0256	B-0257	B-0258	B-0259	B-0260	B-0261	B-0262	B-0263	B-0264	B-0265
Row_09	B-0266	B-0267	B-0268	B-0269	B-0270	B-0271	B-0272	B-0273	B-0274	B-0275	B-0276	B-0277	B-0278	B-0279	B-0280	B-0281	B-0282	B-0283	B-0284	B-0285	B-0286	B-0287	B-0288	B-0289
Row_10	B-0290	B-0291	B-0292	B-0293	B-0294	B-0295	B-0296	B-0297	B-0298	B-0299	B-0300	B-0301	B-0302	B-0303	B-0304	B-0305	B-0306	B-0307	B-0308	B-0309	B-0310	B-0311	B-0312	B-0313
Row_11	B-0314	B-0315	B-0316	B-0317	B-0318	B-0319	B-0320	B-0321	B-0322	B-0323	B-0324	B-0325	B-0326	B-0327	B-0328	B-0329	B-0330	B-0331	B-0332	B-0333	B-0334	B-0335	B-0336	B-0337
Row_12	B-0338	B-0339	B-0340	B-0341	B-0342	B-0343	B-0344	B-0345	B-0346	B-0347	B-0348	B-0349	B-0350	B-0351	B-0352	B-0353	B-0354	B-0355	B-0356	B-0357	B-0358	B-0359	B-0360	B-0361
Row_13	B-0362	B-0363	B-0364	B-0365	B-0366	B-0367	B-0368	B-0369	B-0370	B-0371	B-0372	B-0373	B-0374	B-0375	B-0376	B-0377	B-0378	B-0379	B-0380	B-0381	B-0382	B-0383	B-0384	B-0385
Row_14	B-0386	B-0387	B-0388	B-0389	B-0390	B-0391	B-0392	B-0393	B-0394	B-0395	B-0396	B-0397	B-0398	B-0399	B-0400	B-0401	B-0402	B-0403	B-0404	B-0405	B-0406	B-0407	B-0408	B-0409
Row_15	B-0410	B-0411	B-0412	B-0413	B-0414	B-0415	B-0416	B-0417	B-0418	B-0419	B-0420	B-0421	B-0422	B-0423	B-0424	B-0425	B-0426	B-0427	B-0428	B-0429	B-0430	B-0431	B-0432	B-0433
Row_16	B-0434	B-0435	B-0436	B-0437	B-0438	B-0439	B-0440	B-0441	B-0442	B-0443	B-0444	B-0445	B-0446	B-0447	B-0448	B-0449	B-0450	B-0451	B-0452	B-0453	B-0454	B-0455	B-0456	B-0457
Row_17	B-0458	B-0459	B-0460	B-0461	B-0462	B-0463	B-0464	B-0465	B-0466	B-0467	B-0468	B-0469	B-0470	B-0471	B-0472	B-0473	B-0474	B-0475	B-0476	B-0477	B-0478	B-0479	B-0480	B-0481
Row_18	B-0482	B-0483	B-0484	B-0485	B-0486	B-0487	B-0488	B-0489	B-0490	B-0491	B-0492	B-0493	B-0494	B-0495	B-0496	B-0497	B-0498	B-0499	B-0500	B-0501	B-0502	B-0503	B-0504	B-0505
Row_19	B-0506	B-0507	B-0508	B-0509	B-0510	B-0511	B-0512	B-0513	B-0514	B-0515	B-0516	B-0517	B-0518	B-0519	B-0520	B-0521	B-0522	B-0523	B-0524	B-0525	B-0526	B-0527	B-0528	B-0529
Row_20	B-0530	B-0531	B-0532	B-0533	B-0534	B-0535	B-0536	B-0537	B-0538	B-0539	B-0540	B-0541	B-0542	B-0543	B-0544	B-0545	B-0546	B-0547	B-0548	B-0549	B-0550	B-0551	B-0552	B-0553
Row_21	B-0554	B-0555	B-0556	B-0557	B-0558	B-0559	B-0560	B-0561	B-0562	B-0563	B-0564	B-0565	B-0566	B-0567	B-0568	B-0569	B-0570	B-0571	B-0572	B-0573	B-0574	B-0575	B-0576	B-0577
Row_22	B-0578	B-0579	B-0580	B-0581	B-0582	B-0583	B-0584	B-0585	B-0586	B-0587	B-0588	B-0589	B-0590	B-0591	B-0592	B-0593	B-0594	B-0595	B-0596	B-0597	B-0598	B-0599	B-0600	B-0601
Row_23	B-0602	B-0603	B-0604	B-0605	B-0606	B-0607	B-0608	B-0609	B-0610	B-0611	B-0612	B-0613	B-0614	B-0615	B-0616	B-0617	B-0618	B-0619	B-0620	B-0621	B-0622	B-0623	B-0624	B-0625
Row_24	B-0626	B-0627	B-0628	B-0629	B-0630	B-0631	B-0632	B-0633	B-0634	B-0635	B-0636	B-0637	B-0638	B-0639	B-0640	B-0641	B-0642	B-0643	B-0644	B-0645	B-0646	B-0647	B-0648	B-0649

Figure S2: Pooling design of mutagenized maize families according to a grid. Library specific seed stock matrixes are used by MuWU to annotate every germinal insertion by checking the specific combination of row and column pool for each insertion event.

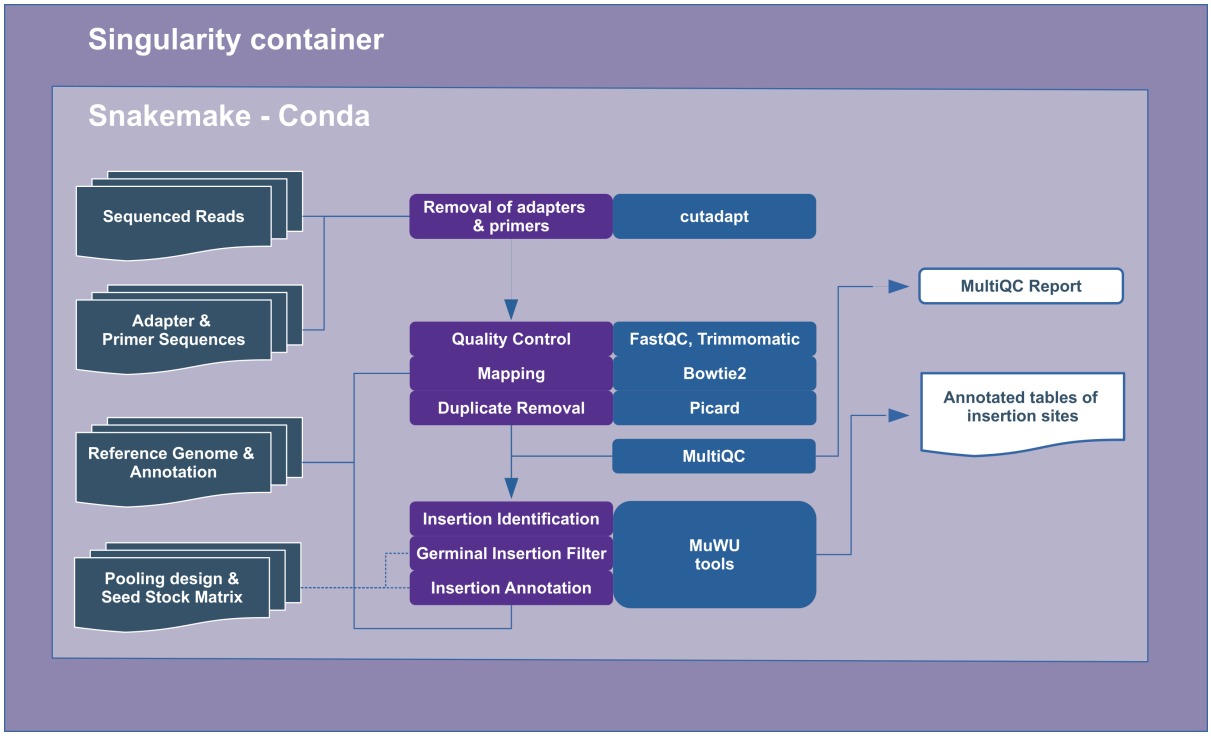


Figure S3: Workflow diagram for MuWU.
Required input files:

- Reference Genome: un-/gzipped FASTA or GenBank (.gbff, .dat) - file location or URL
- Annotation: un-/gzipped GFF3, GTF or GenBank (.gbff, .dat) format - file location or URL
- DNA sequencing reads: un-/gzipped single/paired end FASTQ files
- Primer/Adapter sequences: either pasted into config.yaml or provided as separate FASTA file/s
- GRID method: Pooling Design/Stock Matrix table: .xlsx file (see Supplemental Figure 2.); GENERIC method: samples.tsv table listing samples

Acknowledgement

First, I want to express my utmost gratitude towards my supervisor, Heiko Schoof, for his invaluable expertise and unwavering guidance throughout my research journey. Heiko's vast knowledge and experience in multiple fields have been instrumental in helping me navigate the intricacies of science. Moreover, I am grateful to Heiko for providing me with the freedom to choose both my preferred projects and work conditions.
I greatly appreciated the family atmosphere and weekly team breakfasts in the Crop Bioinformatics group.
Furthermore, I would like to thank all collaborators, especially Caroline Marcon, Frank Hochholdinger, Lukas Schreiber, and Peter Dörmann.

Sincere thanks also to all my colleagues and partner labs during my time as a Ph.D. student:

- Thanks to my fellow (former) Ph.D. candidates Lena, Florian, Lucia, Caro, Annika, Paul, and Yannic.
- Thanks to my student Marius who I had the pleasure to supervise during his master's thesis.
- Thanks also to Michael and Karin for their technical and administrative support.

For all the personal support and the constant readiness to take my mind off things, I am deeply grateful to my family and friends.

Commonly known is the story about the anxious and stressed Ph.D. candidate, typing in the last words just a minute before actually printing the thesis. Further, worries about the future and lack of direction as the undertaking of a multi-year effort draws to a close might plague the psyche of the (hopefully) soon-to-be doctorate holder. As knowledge of such and worse circumstances shaped my mindset from the very start of my research adventures, I am quite proud of my pleasantly stress-free write-up.

These last years, during which I contributed to molecular crop science research, were formative both on a personal and intellectual level. I am immensely grateful for this. Yet, after climbing the ivory tower of academia for ten years, I felt it was time to leave the lecture halls behind me.

And so I conclude my academic adventures with reminding myself of the Literary studies that started it all by quoting the Bard:

In nature's infinite book of secrecy
A little I can read.