Institut für Nutzpflanzenwissenschaften und Ressourcenschutz (INRES)

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

Transcriptomic regulation of hybrid vigor in immortalized backcross populations of maize (*Zea mays* L.)

Dissertation

zur Erlangung des Grades

Doktorin der Agrarwissenschaften (Dr. agr.)

der Agrar-, Ernährungs- und Ingenieurwissenschaftlichen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn

von

Marion Pitz

aus

Emmerich am Rhein

Bonn, 2025

Referent: Prof. Dr. Frank Hochholdinger

Korreferent: Prof. Hans-Peter Piepho

Tag der mündlichen Prüfung: 12.06.2025

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

Content

List of figures

Abbreviations

AD	Allele depth
BLINK	Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway
BPH	Better-parent heterosis
B73v5	B73 reference genome version 5
C	Celsius
cM	centiMorgan
DAG	Days after germination
DP	Depth (of snp calling coverage)
eQTL	Expression quantitative trait locus
F_1	First filial generation
FarmCPU	Fixed and random model Circulating Probability Unification
FDR	False discovery rate
FS	Fisher strand
GAPIT	Genome Association and Prediction Integrated Tool
GATK	Genome Analysis Toolkit
Gb	Giga bases
GC	Guanine-cytosine content
GQ	Genotype quality
GWAS	Genome wide association study
IBM-RIL	Intermated B73xMo17 recombinant inbred line
I-RIL	Intermated recombinant inbred line
LOD	Logarithm of odds
Log_2FC	Logarithm with base 2 of the fold change in gene expression
MDS	Multidimensional scaling
MLMM	Multi-Locus Mixed Model
MPH	Mid-parent heterosis
MPV	Mid-parent value
MQ	Mapping quality
mya	Million years ago
N	Number (sample size)
NGS	Next generation sequencing
PAV	Presence-Absence Variation
QTL	Quantitative trait locus
QUAL	Variant site quality score
R	Software
RIL	Recombinant inbred line
RNA	Ribonucleic acid
SNP	Single nucleotide polymorphism
SPE	Single parent expression
TMM	Trimmed mean of M-values
TSG	TWAS and SPE genes
TWAS	Transcriptome wide association study
Vcf	Variant call format

1. Summary/Zusammenfassung

1.1 Summary

Heterosis is the phenomenon that heterozygous F_1-hybrids outperform their distinct homozygous parental inbred lines for many agronomically important traits. Despite its economic importance and utilization for alomst 100 years, the molecular mechanisms underlying heterosis are not fully understood. It has been demonstrated that genes with expression differences between parents and their hybrid progeny are involved in heterosis. However, quantitative associations between gene expression differences and heterosis are scarce, and the regulation of these differences is not yet established. This thesis aimed to close these knowledge gaps and advance the molecular understanding of heterosis. We analyzed 112 lines of the intermated B73xMo17 recombinant inbred line (IBM-RIL) population of maize and their backcrosses to B73 and Mo17. These backcross hybrids contain heterozygous and homozygous genomic regions and allow for the identification of genomic locations regulating specific gene expression patterns.

In chapter 3 of this thesis we investigated single parent expression (SPE). These genes, which are active in only one of the parents and in the hybrid, explained up to 29% of the heterotic variance in the backcross hybrids. This pattern of expression complementation in hybrids is consistent with the dominance model of heterosis. Moreover, expression quantitative loci (eQTL), regulating SPE genes are predominantly located in heterozygous regions of the genome, highlighting the importance of the genomic architecture of regulatory elements for gene expression. As a consequence, heterozygosity leads to a higher number of active genes in the backcross hybrids by SPE complementation. Finally, we identified an SPE gene that regulates lateral root density in hybrids. Notably, the activity of this gene depends on the presence of a Mo17 allele in the eQTL that regulates it. This highlights the pivotal role of distantly located regulatory elements for the activity of heterosis-associated genes.

In chapter 4 we analyzed non-additive gene expression, a pattern in which genes in the hybrid are expressed significantly different from the average of the two parents. In most instances, these genes exceeded the mean of the parental expression. Non-additive gene expression explained up to 27% of the heterotic variance. Consistent with our observation for SPE genes, complementation of non-additive genes is consistent with the dominance model and is regulated almost exclusively by eQTL in heterozygous genomic regions. Moreover, regulation of non-additively expressed genes depends on the genetic origin of the higher expressed parent.

In summary, gene expression complementation contributes substantially to heterosis, likely by increasing the number of active genes by SPE or by higher expression levels of non-additive genes in the hybrids. Heterozygosity and the genetic architecture of hybrids might be aspects of how genetic variation is translated into vigorous hybrids via the regulation of heterosis-associated gene expression patterns.

1.2 Zusammenfassung

Heterosis beschreibt das Phänomen, dass heterozygote F_1-Hybriden in vielen landwirtschaftlich wichtigen Merkmalen ihren homozygoten Elterninzuchtlinien überlegen sind. Trotz der wirtschaftlichen Bedeutung und Nutzung von Heterosis seit über 100 Jahren sind die zugrunde liegenden molekularen Mechanismen nicht vollständig geklärt. Gene mit unterschiedlicher Expression zwischen Eltern und Hybrid, stehen vermutlich im Zusammenhang mit Heterosis. Es gibt jedoch nur wenige Studien über die quantitativen Zusammenhänge zwischen Expressionsunterschieden und Heterosis, und die Regulation dieser Unterschiede ist unklar. Ziel dieser Arbeit war es, diese Wissenslücken zu schließen und das molekulare Verständnis von Heterosis zu verbessern. Wir analysierten dazu 112 Maislinien der rekombinanten B73xMo17 Inzuchtpopulation (IBM-RIL) und deren Rückkreuzungen mit B73 und Mo17. Diese Rückkreuzungshybride bestehen aus heterozygoten und homozygoten Regionen und ermöglichen die Identifizierung von genomischen Positionen, die spezifische Genexpressionsmuster regulieren.

In Kapitel 3 dieser Arbeit untersuchten wir „single parent expression“ (SPE) Gene. Diese Gene, die nur in einem der Elternteile und im Hybrid aktiv sind, erklärten bis zu 29% der heterotischen Varianz in den Rückkreuzungshybriden. Die Komplementierung der Expression in Hybriden durch SPE steht im Einklang mit dem Dominanzmodell der Heterosis. Darüber hinaus befinden sich die „expression quantitative trait loci“ (eQTL), die genomischen Positionen, die SPE-Gene regulieren, überwiegend in heterozygoten Regionen des Genoms. Dies unterstreicht, wie wichtig die genomische Architektur der regulatorischen Elemente für die Genexpression ist. Folglich führt die Heterozygotie zu einer höheren Anzahl aktiver Gene in den Rückkreuzungshybriden durch SPE-Komplementierung. Schließlich haben wir ein SPE-Gen identifiziert, das die Seitenwurzeldichte in Hybriden reguliert. Bemerkenswert ist, dass dieses Gen nur aktiv ist, wenn ein Mo17-Allel im regulierenden eQTL vorliegt. Wir konnten somit zeigen, wie zentral die Rolle von weit entfernten regulatorischen Elementen für die Aktivität von Heterosis-assoziierten Genen ist.

In Kapitel 4 analysierten wir die nicht-additive Genexpression. Dises Muster liegt vor, wenn sich die Expression eines Gens im Hybrid signifikant vom Durchschnitt der Eltern unterscheidet. In den meisten Fällen liegt die Expression diese Gene über dem Mittelwert der elterlichen Expression. Die nicht-additive Genexpression erklärte bis zu 27% der heterotischen Varianz. In Übereinstimmung mit unseren Beobachtungen bei den SPE-Genen ist die Komplementierung nicht-additiver Gene mit dem Dominanzmodell konsistent. Sie wird ebenfalls fast ausschließlich durch eQTL in heterozygoten genomischen Regionen reguliert. Darüber hinaus hängt die Regulation nicht-additiver Gene von der genetischen Herkunft des Elternteils mit höherer Expression ab.

Zusammenfassend trägt die Komplementierung der Genexpression wesentlich zur Heterosis bei. Wahrscheinlich durch eine Erhöhung der Anzahl aktiver Gene durch SPE sowie durch höhere Expression der nicht-additiven Gene in den Hybriden. Die Heterozygotie und die genetische Architektur von Hybriden könnten Aspekte davon sein, wie genetische Variation über die Regulation von Heterosis-assoziierten Genexpressionsmustern in leistungsfähige Hybride umgesetzt wird.

2. Introduction

2.1 Maize – a versatile component of global agri-food systems

Maize (*Zea mays* L.) is a versatile component of the global agri-food system and one of the most important staple cereals, reaching a global production of 1,163 million tons in 2022[1]. Together, the three major staple cereals, maize, rice and wheat, provide an estimated 42% of the global human calorie consumption[2,3]. Global maize production has increased by 46% since 2010[4], due to multiple uses including animal feed and increasing industrial uses such as biofuels.

The concentration on a few crops, such as maize leads to shorter crop rotations. In combination with high input of fertilizers, this leads to vulnerability of the agricultural system[5]. At the same time, maize is particularly resistant to the effects of climate change. It has relatively low sensitivity to early summer drought and high yield stability[6]. Due to C4 carbon fixation, maize has one of the most favorable water footprints relative to the nutritional energy (0.41 L of water/kcal) among crops[7]. In addition, diversity in photoperiod sensitivity[8] allowed maize, originally a short-day plant, to thrive under continental long-day conditions. As a consequence, maize is now grown on a geographically wider range than any other important crop[9]. Maize is also used for various products. It is primarily used as animal feed (56%), followed by non-food/industrial use (20%) and food (13%). Regional differences regarding these preferences range from 66% of maize used for food in southern and eastern Africa to 76% animal feed in Europe[2]. With increasing economic development in the Global South, this usage is expected to increase even further[2].

2.2 Plant description – What made maize a model organism

Maize originates from southern Mexico, where it was domesticated from its wild ancestor teosinte (ssp. *Parviglumis)* around 9.000 years ago[10]. It is an annual monocotyledonous crop of the *Poaceae* family, also known as grasses[11]. In addition to the outstanding economic importance, several morphological and genetic aspects of maize made it an ideal model organism for biological studies[12]. The tall grass shows monoecious floral development, where female and male flowers are physically separated[12]. The male germ cells are produced in the tassel that develops at the top of the stem. The female germ cells form in one or more ears, which develop at the base of leaves around the middle of the stem, five to six nodes below the tassel[12,13]. This characteristic allows for controlled pollinations, which is essential for genetic experiments. Controlled pollinations also facilitated the development of hybrids, which largely contributed to the agricultural success of maize[9,14]. The ear structure, with many large seeds in rows, enables discoveries in connection with kernel colour, like the discovery of jumping

genes or transposons[15]. This discovery paved the way for the development of a large range of populations with induced transposon mutations for gene analyses[16–18]. For studying the genetic basis of quantitative traits, a range of mapping populations exists[19–21]. In addition, the relatedness between the grasses and the resulting collinearity is exploited in evolutionary studies[13] and to infer knowledge on orthologous genes. Due to all these aspects, maize has remained a model organism with diverse research applications. Today it ranges from methodological research on genome editing with CRISPR/Cas9 and genomic selection to foundational research on maize morphology and stress resistance[22].

2.3 Maize roots

The maize root system is responsible for nutrient and water uptake as well as stable anchorage in the soil. All of these functions are important to meet global agricultural challenges[23,24]. The root system of maize is defined by early embryonic roots (primary and seminal) (Figure 1.1) and later post-embryonic (crown, brace, lateral) roots[25]. Initially, the primary root develops at the basal pole of the embryo, followed by the initiation of seminal roots that develop at the root-shoot junction, known as the scrutellar node (Figure 1.1).

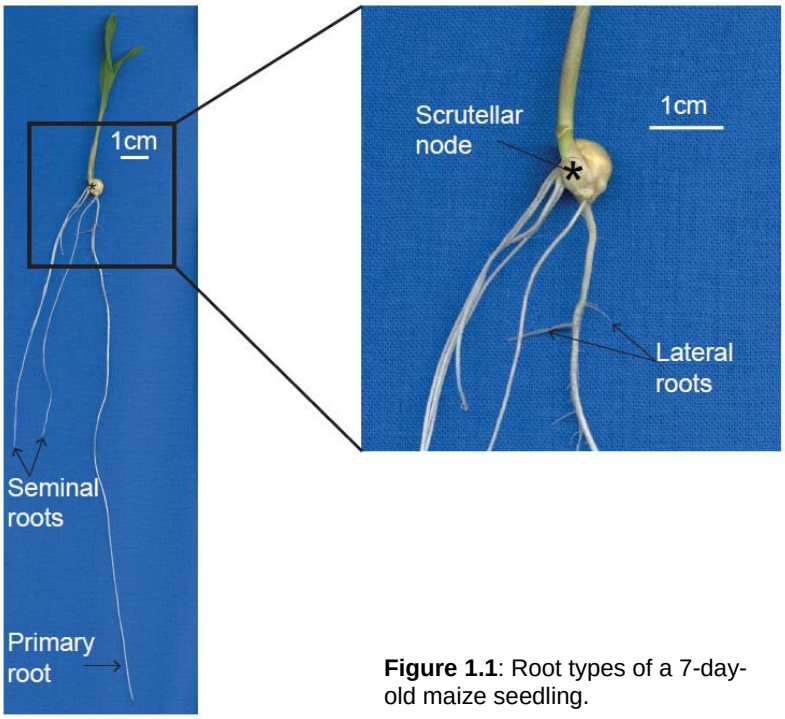


Figure 1.1: Root types of a 7-day-old maize seedling.

About two weeks after germination, crown roots initiate at belowground shoot nodes and form the post-embryonic root system together with brace roots that initiate at aboveground shoot nodes. Lateral roots develop on all root types of the embryonic and post-embryonic root system[25-28]. The embryonic root system establishes early seedling vigour. In mature maize at flowering, the root system predominantly consists of shoot borne roots that provide enhanced lodging resistance and take up nutrients and water[25,27]. In the embryonic root system, lateral roots absorb the majority of water, while primary and seminal roots mainly transport the absorbed water to the shoot[29]. Upon maturation, the majority of water uptake switches to the crown roots and their laterals[30]. Efforts have been made to understand and adapt the root system to agricultural challenges[23,24]. The early root system reacts to environmental cues and is especially important for drought adaptation. Upon water deficit, there is an integrative and

coordinated response of all root types, characterized by rapid changes in hydraulic conductance and high plasticity to resume growth under normalized conditions[31]. Natural variation towards water as well as nutrient acquisition has been utilized to identify genetic features associated with beneficial traits, that can be used for breeding improvements[27]. It has been shown that reducing the number of seminal roots in favour of increasing the lateral root density can help the embryonic root system in drought adaptation[32]. Recently, also interactions between the plant and the microbiome have gained increasing recognition[33].

2.4 Heterosis – a transformative phenomenon

The development of hybrids largely contributed to the agricultural success of maize[9,14]. The underlying phenomenon is called heterosis. It describes the observation that F_1-hybrid offspring of two distinct inbred parental lines outperform the parental average in agronomically important traits[34,35]. Heterosis is often investigated in above ground traits, such as yield, biomass and disease resistance[36] and can also be measured in plant height or cob size[37]. Typical quantifications of heterosis are better-parent heterosis (BPH), referring to the difference between hybrid and better performing parent and mid-parent heterosis (MPH), referring to the difference between hybrid and parental average. Many traits can show substantial heterosis, while the correlation between heterosis for different traits on the same plant is generally weak[38]. The importance of root related traits for overall plant performance have previously been established. In maize roots, heterosis manifests as early as 5 to 7 days after germination (DAG) (Figure 1.2). It becomes measurable in terms of a consistent increase in primary root length and an increase in the number of seminal roots. An average MPH of 51% was measured for the density of lateral root primordia[39] (compare Figure 1.2). Heterosis influences maize roots also at later developmental stages and is especially prominent under water stress[40,41]. Other crops, such as wheat and rice benefit from heterosis in root related traits as well[42–44].

While the seed of inbred lines can be saved by farmers and planted with a comparably stable phenotype across generations, this is not the case for hybrid seed. Self-fertilization of fully heterozygous hybrids results in recombination and an average loss of 50% heterozygosity in each generation, leading to genetically diverse plants with generally lower yields[45,46]. Therefore, hybrid seed must be produced each season by crossing the distinct parents. This way, hybrids provide a form of indirect biological intellectual property[47] and thereby pose an incentive for increased private sector involvement in seed production on markets with no plant variety protection, as in many Sub-Saharan African countries[48,49].

Largely due to the yield and performance advantage, hybrid maize has been adopted in many areas globally[50]. Since the 1930th, when most breeders introduced hybrid varieties, grain yields

of maize increased more than six-fold (reviewed in[14,37,51]). Hybrid breeding in rice increased food security in China, where the average yield increased by 47% within 15 years of hybrid introduction[52]. The adoption of hybrid canola in Canada indicates that famers deem the advantages of hybrid seed worth the extra cost[53] and decrease in independence.

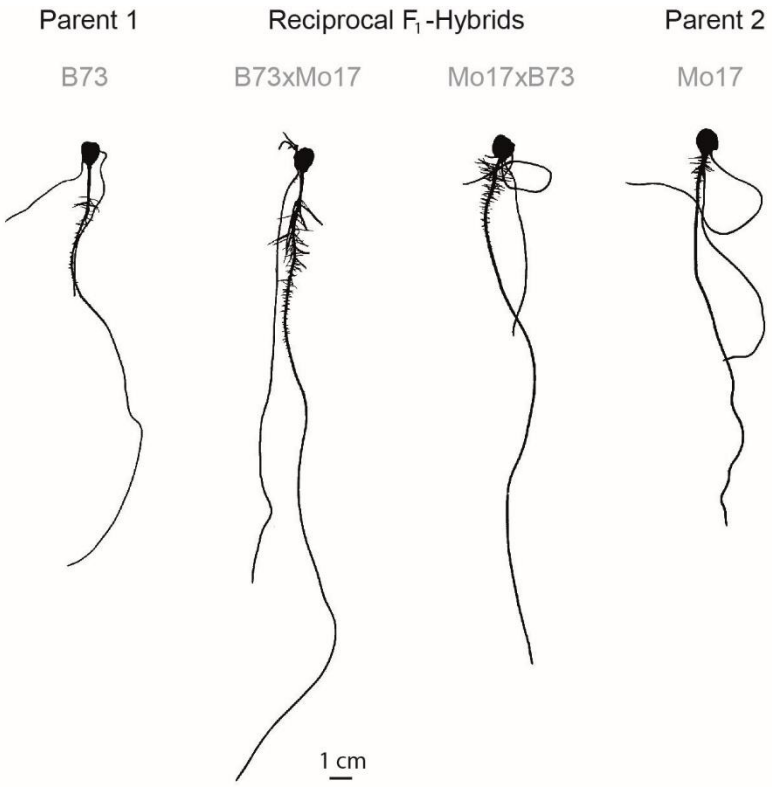


Figure 1.2: Heterosis during early seedling root system development. Root scans of 7-day-old B73 and Mo17 seedlings and their reciprocal hybrids.

Nevertheless, the yield increases in maize and rice were accompanied by changes in agricultural practices, often involving mechanization, fertilization and chemical plant protection[14,52], which can negatively affect the environment. But hybrids also have the potential of tackling these environmental impacts with their greater yield stability regarding marginal environments and the possibility to combine dominant genes faster[47]. Given the wide range of agricultural, environmental and socio-economic implications of hybrids, knowledge of the underlying heterosis phenomenon should be of interest, not only to scientists and breeders, but decision to makers across various disciplines.

2.5 Evolution of the maize genome

Maize displays an exceptional degree of structural genomic diversity between different genotypes[54], facilitated by the predominantly out-crossing nature[12]. For illustration, there are on average about as many nucleotide polymorphisms between humans and chimpanzees, as between two maize lines[54]. This diversity is reflected by the 50 different fully sequenced reference genomes of distinct maize lines (https://www.maizegdb.org/genome, checked:

14.01.2025). The first fully sequenced genome of a maize inbred line was B73 in 2009[55], which has been updated four times by sequencing and annotation advances[56–58]. The maize genome is diploid and consists of 10 chromosomes. Compared to other grass species, it is of medium size (~2.5 Gb), bigger than rice or sorghum, but smaller than those of barley and wheat[12].

Contributing to the complexity of the maize genome is the history of polyploidy events that led to duplications in the genome. The last whole genome duplication resulted from an allopolyploid hybridization around 5 million years ago (mya)[59]. Preceding this hybridization, the progenitors of maize split from one another and from sorghum around 12 mya[59] (Figure 1.3). Therefore, the sorghum genome did not undergo the same duplication as maize. After the hybridization of the two maize progenitors, breakage, fusion and reassembly resulted in the 10 diploid chromosomes of modern maize[60] (Figure 1.3). Consequently, this genome can be divided into two subgenomes, Maize1 and Maize2, which are each homologous to the entire genome of sorghum (Figure 1.3). During continuous evolution and purifying selection frequently only one of the gene copies was retained, with a bias towards the Maize1 copy[61,62]. Generally, those genes that were retained and are syntenic to sorghum were conserved through evolution and therefore often have conserved roles in determining phenotype and fitness[63]. In addition, a portion of genes does not have syntenic homologs in sorghum or other related grass species[64].

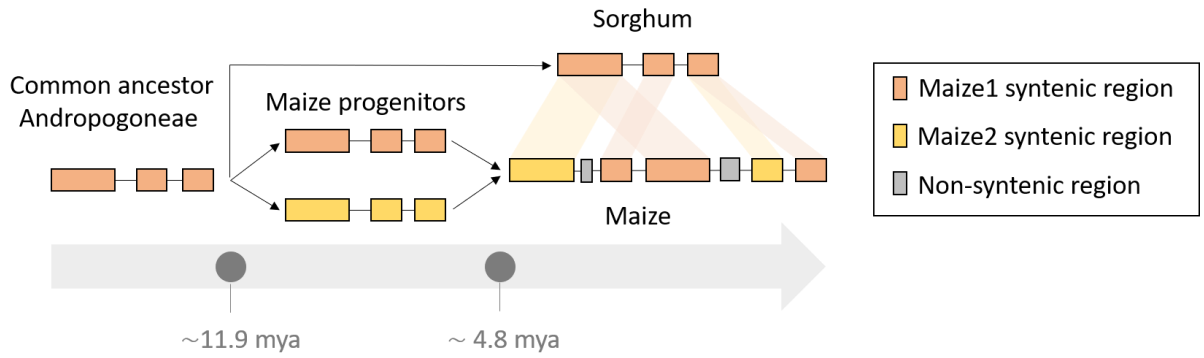


Figure 1.3: Schematic depiction of maize and sorghum synteny. Genomic regions are depicted as boxes. Maize progenitors and sorghum separated ca. 11.9 million years ago (mya). The hybridization of two maize progenitors resulted in the modern maize genome with duplicated genomic regions that are syntenic to sorghum and additional non-syntenic regions.

They were probably transposed by a copy-and-paste mechanism to non-syntenic positions, after the divergence of maize and sorghum[61,63,64]. This process can alter their functionality[63]. The non-syntenic genes can be described as evolutionary younger, due to their more recent history, in contrast to the evolutionary older or conserved syntenic genes[63,65]. While on a macro-scale ~15% of maize genomic regions were classified as non-syntenic blocks[66], on the gene level around 1/3 of genes are non-syntenic[64]. These non-syntenic genes are underrepresented among curated genes and less likely to show a mutant phenotype[64], hindering the unambiguous identification of their functions. Nevertheless, non-syntenic genes

have repeatedly been associated with disease resistance[65]. Remarkably, while there is little variation in presence-absence variation (PAV) or transcriptional variation among syntenic genes between different maize inbred lines, this is not the case for non-syntenic genes. As non-syntenic genes were under different evolutionary constraints, they may be a source of variation regarding agronomic traits[66]. In further support of this notion, non-syntenic regions were correlated with the level of heterosis, while syntenic regions were not[67].

2.6 Heterotic groups of the B73xMo17 hybrid and mapping population

Early on, researchers and breeders observed that the phylogenetic distance between parental lines is positively correlated with heterosis[68]. This observation has led to the clustering of parents into heterotic groups so that they exhibit especially high levels of heterosis when combined with another heterotic group. There are typical female and typical male heterotic groups and advantageous combinations of them are called heterotic pattern[69,70]. Other crops, such as rice, also benefit from the classification of parental lines into heterotic groups[70,71]. Consequently, breeders selected inbred lines not only on their general performance, but also their combining ability in heterotic patterns. This limits the variability within the heterotic groups and probably facilitates their conservation[72]. A better understanding of the underlying principles of heterosis might therefore directly influence breeding approaches.

The B73xMo17 hybrid represents the common heterotic pattern of crossing a parent from the stiff stalk and non-stiff stalk heterotic groups[73,74]. This hybrid was used commercially in the 1980s and as one of the last hybrids developed in the public sector, became a popular research hybrid. Today, these genetic resources are still relevant commercially, as large proportions of contemporary germplasm are descended from B73[75,76]. In addition, B73 and Mo17 are among the most widely used lines in research[12]. There is a high level of variation between B73 and Mo17, such as single nucleotide polymorphisms (SNPs) and structural variants or PAV. For example, PAVs of genomic segments longer than 500 Kbp, only present in one of the genomes, sum up to more than 25 Mbp in total[77].

This variation facilitated the success of the intermated B73xMo17 recombinant inbred line (IBM-RIL) population[19]. Recombinant inbred lines (RILs) are generated by crossing two distinct homozygous lines, followed by repeated rounds of selfing. This results in a set of homozygous lines whose genomes are mosaics of the original parents[78]. In addition, intermated RILs (I-RILs) are crossed for several generations before selfing[79], generating even denser breakpoints. In case of the IBM-RIL syn. 4 population, four generations of intercrossing were performed resulting in ~300 highly homozygous lines with a high degree of variance in phenotypic traits and genomic regions[19]. These RILs or I-RILs can be backcrossed to their original parents to create backcross populations showing high diversity in heterozygosity and heterosis (chapter

3). In contrast to backcross populations generated directly from F_1-offspring, RIL backcross populations are obtained by crossing fully homozygous lines. This means that they can be regenerated eternally and therefore are often termed immortalized backcross populations[80]. The genotypic and phenotypic diversity of RILs and backcross populations can be used to identify genomic locations that are associated with a phenotypic trait, also called quantitative trait locus (QTL)[81,82] (Figure 1.4). Additionally, RILs and backcross populations can be used for candidate gene identification or heterosis studies[80,83]. If instead of a phenotypic trait, the expression of a gene is associated with a genomic location, an expression QTL (eQTL) is identified[84] (Figure 1.4). Depending on their distance to the gene of interest, they can be classified into *cis*- (in close proximity) or *trans*-regulating (at a distance, for example on a different chromosome) (Figure 1.4). In QTL studies, the resolution may not be sufficient to identify single genes or exact genetic position, but instead regions of several Mbp are identified. Transcriptome wide association studies (TWAS) directly correlate gene expression values with phenotypic traits (Figure 1.4). It was shown, that TWAS can successfully complement results from genomic analyses and identify confirmed candidate genes[85].

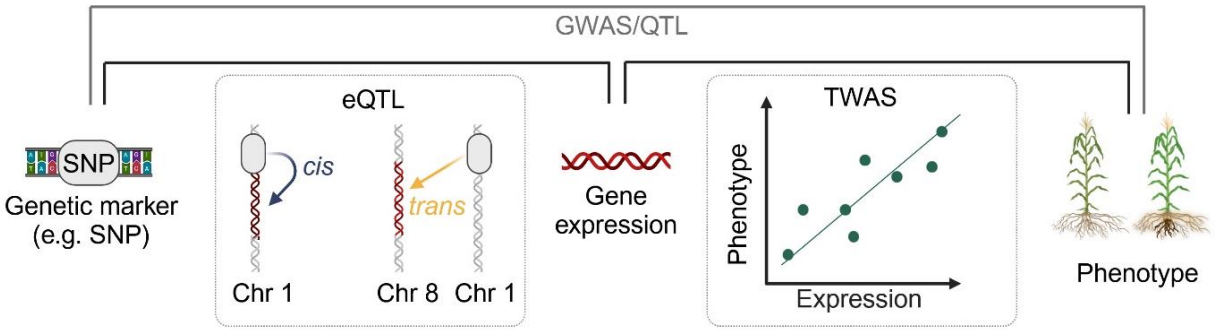


Figure 1.4: Whole genome analyses - overview. Genetic markers that are associated with quantitative traits are identified via Genome Wide Association Studies (GWAS) or as Quantitative Trait Loci (QTL). Genetic features that regulate gene expression can be identified as expression QTL (eQTL). Those can regulate nearby genes in *cis* or at a distance in *trans*. Genes affecting phenotypic quantitative traits by their expression can be identified via Transcriptome Wide Association Studies (TWAS), which correlate the expression of a gene with the phenotype. Adapted from https://BioRender.com/j69i246[85].

2.7 Heterosis – explanatory mechanisms

Despite the global success and long-standing exploitation of heterosis, no unifying model combining all observations in relation to heterosis has been identified[86]. Nevertheless, during the last 100 years, several mechanisms have been proposed which evolved with the technological advancements and general research progress made in molecular biology. The most important historical mechanisms to explain heterosis are the dominance and overdominance model together with epistasis[87].

The dominance model assumes that many slightly deleterious alleles are complemented in the hybrid by dominant or at least stronger alleles[88,89]. Consequently, it should be theoretically possible to obtain the same phenotype in an inbred line that combines all the advantageous

alleles in homozygous form. While this is virtually unattainable, substantial progress has been made in the performance of parental inbred lines. At the same time, no dramatic decline of heterosis was observed[51]. This makes it unlikely that dominance is the sole explanation for heterosis[37]. In the second historical model of overdominance, two different alleles at the same locus cause heterosis by their interaction in hybrids, which is superior to the interaction of homozygous alleles in the parental lines[90]. While there have been single gene evidence examples for overdominance[91,92], they do not account for the full scale of heterosis and such observations are rare compared to the occurrence of heterosis. Since current inbred lines outperform early heterozygous hybrids, overdominance cannot be the sole mechanism underlying heterosis[86]. Another concept suggests that, like leaf area is a product of width and length, genes or effects of genes might interact with each other and thereby further increase heterosis[93], a concept known as epistasis.

Since these early concepts, huge strides have been made in agricultural sciences by the molecular understanding of genes and their functions. Driven by the observation of progressive heterosis in polyploids, the gene-balance theory was proposed. The authors suggested that subunits of regulatory complexes have to be present in the optimal ratio for efficient gene expression[94]. A similar effect would be expected for gene products, acting downstream of one another in a cascade[95]. Those ratios of gene products would then be most optimal in case of equal genome numbers in polyploids. On a similar note, a study proposed an additional model, in which divergent gene expression levels between parents are buffered towards an often advantageous mid-parental level in the hybrid[96]. A positive association between this expression pattern and yield heterosis was found in commercial maize hybrids, supporting this theory[97]. With the advent and ongoing improvement of high-throughput RNA-sequencing, understanding of the transcriptome was revolutionized[98,99]. It was since established that hybrids and their parental inbred lines vary in their expression profiles in terms of allele-specific, differential or non-additive expression[97,100–104]. Genes which are expressed at higher or lower levels in the hybrid than the parental average are called non-additively expressed. Genes with this expression pattern are specific to the tissue, developmental stage and genotype[102,105–107], hindering the establishment of their connection to heterosis. Nevertheless, some general features emerged. Differential expression and allele-specific expression in hybrids are probably linked[103] and influenced by the parent of origin, with emphasis on,maternal effects[106,108,109]. Transcriptome studies also revealed single parent expression (SPE) complementation in hybrids, where a gene that is active in only one of the two parental lines is also active in the hybrid[102]. This leads to higher numbers of active genes in the hybrid compared to the parents[102,110]. This pattern is in line with the dominance model of heterosis[102]. It was shown, that the number of SPE genes is positively associated with MPH of several phenotypic traits[111]. Notably, SPE genes were significantly more often non-syntenic than

syntenic, linking them to the adaptability of hybrids to different environments[112–114]. Next to SPE as a possible source contributing to heterosis, regulatory patterns of gene expression have gained increasing recognition in relation to heterosis[108]. For example, *trans*-regulation was associated with paternal dominance in IBM-RIL hybrids implicating strong parental imprinting of gene expression[115]. In addition to the transcriptome, also epigenetics, the metabolome and proteome and more recently also the microbiome are studied in associated with hybrid vigour[86,116,117]. As mentioned previously, structural variants as well as PAVs, especially related to non-syntenic regions contribute to heterosis[67].

Generally, researchers agree, that not one explanatory mechanism is responsible for heterosis, but several aspects contribute together and are not mutually exclusive. On the genetic level, a wide range of genes that differ between tissue and trait most likely contribute to heterosis of different traits. A better understanding of this phenomenon is of outstanding economic importance and might directly influence maize breeding programs.

2.8 Aims of the study

The general objective of this work was to quantify how single parent expression complementation and non-additive gene expression in maize hybrids contribute to phenotypic heterosis and how these gene expression patterns are regulated.

Specifically, we tested the following hypotheses:

1. Non-additively expressed genes and SPE are mainly located in heterozygous regions of the hybrid genomes and the number of these expression pattern genes is associated with the degree of heterozygosity in the hybrids.
2. Substantial proportions of phenotypic heterotic variance can be attributed to SPE complementation and non-additively expressed genes.
3. Closely as well as distantly located expression QTL regulate gene expression patterns that contribute to hybrid vigour.
4. The parental genetic origin of expression pattern genes is associated with their regulation.
5. Association of phenotypic traits with gene expression patterns (Transcriptome wide association analysis: TWAS) can identify candidate genes related to hybrid vigour in roots.

3. Regulation of heterosis-associated gene expression complementation in maize hybrids

Authors

Marion Pitz[1#], Jutta A. Baldauf[1#], Hans-Peter Piepho[2], Peng Yu[3], Heiko Schoof[4], Annaliese S. Mason[5], Guoliang Li[6], Frank Hochholdinger[1*]

Affiliations

[1]Institute of Crop Science and Resource Conservation, Crop Functional Genomics, University of Bonn, 53113 Bonn, Germany

[2]Institute of Crop Science, Biostatistics Unit, University of Hohenheim, 70599 Stuttgart, Germany

[3]Institute of Crop Science and Resource Conservation, Emmy Noether Group Root Functional Biology, University of Bonn, 53113 Bonn, Germany

[4]Institute of Crop Science and Resource Conservation, Crop Bioinformatics, University of Bonn, 53115 Bonn, Germany

[5]Institute of Crop Science and Resource Conservation, Plant Breeding Department, University of Bonn, 53115, Bonn, Germany

[6]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, 06466 Seeland, Germany

# Contributed equally to this work

* Corresponding author: Frank Hochholdinger

E-Mail: hochholdinger@uni-bonn.de

Preprint: https://doi.org/10.1101/2024.10.30.620956

This manuscript is under review in *Genome Biology*.

Abstract

Classical concepts of heterosis attribute the superiority of F_1-hybrids over their homozygous parents to the complementation of unfavorable by beneficial alleles (dominance) or to heterozygote advantage (overdominance). Here we analyzed 112 intermated B73xMo17 recombinant inbred lines of maize and backcrossed them to their original parents to obtain hybrids with an average heterozygosity of ~50%. This genetic architecture allowed us to study the influence of homozygous and heterozygous genomic regions on gene expression in hybrids. We demonstrated, that up to 29% of the heterotic variance in these hybrids is explained by single parent expression (SPE) complementation. In this mode of expression, consistent with the dominance model, genes are expressed in only one of the parents and in the hybrid. Furthermore, we demonstrated that eQTL regulating SPE genes are predominantly located in heterozygous regions of the genome. Thus, we demonstrated that dominance of SPE genes is important for gene activity, while heterozygosity is instrumental for the regulation of these genes. Finally, we identified an SPE gene that regulates lateral root density in hybrids. Remarkably, the activity of this gene depends on the presence of a Mo17 allele in an eQTL that regulates this gene, supporting the notion that the genetic constitution of distant regulatory elements plays a key role for the activity of heterosis-associated genes. In summary, the prevalence of dominance at the level of gene activity and overdominance at the level of gene regulation reconciles these classical genetic concepts and explains how they could both contribute to heterosis.

Introduction

Heterozygous F_1-hybrids are more vigorous and show higher fitness and biomass production compared to their homozygous parental inbred lines[1], a phenomenon called heterosis[2]. The introduction of hybrids in plant breeding was one of the landmark innovations of agriculture. Among crops, outcrossing species such as maize display the highest degree of heterosis[3]. Heterosis is typically measured for above-ground traits related to yield and biomass[4], but root traits, which have been largely neglected in the past, are also important for future crop improvement[5]. In maize roots, heterosis manifests as early as 5 to 7 days after germination for traits such as primary root length or lateral root density[6].

One classical genetic mechanism to explain heterosis is the dominance model, which assumes that many slightly deleterious alleles are complemented in the hybrid by dominant or at least stronger alleles[7]. On the gene expression level, single parent expression (SPE) complementation, where the hybrid expresses a gene that is only active in one of its parents is consistent with the dominance model[8,9]. SPE complementation results in hundreds of additionally active genes in hybrids. It probably plays a role in translating parental diversity into phenotypic heterosis and has been linked to hybrid vigor[10]. It was further established that SPE genes are significantly enriched among evolutionarily younger, non-syntenic genes and might function in the adaptation of hybrids to different environments[10–12]. Additionally, variation in transcriptional regulation of *cis*- and *trans*-acting factors has been highlighted in relation to hybrid performance[13]. *Trans*-regulated gene expression in hybrids was associated with paternal alleles in maize[14]. Moreover, an association between SPE and *cis*-regulation was suggested[12].

Intermated recombinant inbred lines (I-RILs) are the result of crossing two distinct inbred lines, followed by several generations of intercrossing and subsequent self-pollination of their progeny[15]. In maize, the intermated B73×Mo17 recombinant inbred line (IBM-RIL) syn. 4 population, a set of ~ 300 highly homozygous intercrossed RILs (4 generations of intercrossing), shows a high diversity of phenotypes and of genomic regions contributed by the two parental genotypes[16]. By backcrossing RILs to their original parents, backcross populations can be generated which show varying degrees of heterozygosity and heterosis. Genetically and phenotypically diverse RILs and backcross populations are an important resource for QTL mapping, as well as for candidate gene identification and heterosis studies[17-20].

We used the IBM-RIL population and two backcross populations to study how varying heterozygosity and the regulation of SPE complementation influences the manifestation of heterosis in seedling root development. We demonstrated, that SPE genes are predominantly

regulated from heterozygous regions of the genome and that depending on the genetic constitution of the active parent, *cis*- or *trans*- regulation of SPE is prevalent. We hypothesize based on our findings, that differences in regulation of the parental lines of a hybrid contribute to heterosis.

Results

Transcriptome profiling and sample evaluation via SNP calling reveals cross-over locations and IBM-RIL specific regions

After seed propagation, we selected 112 IBM-RILs and their B73×IBM-RIL and Mo17×IBM-RIL hybrids (Figure 1A) based on seed availability together with the reciprocal hybrids Mo17×B73 and B73×Mo17 as well as the parents B73 and Mo17 (Figure 1B) for subsequent phenotyping and RNA-sequencing of young primary roots. After SNP calling and quality control (see methods), we classified the IBM-RIL genomes into B73 and Mo17 regions. During this process, we masked IBM-RIL specific regions which contained SNPs not present in Mo17 or B73 and are thus likely the result of contamination from other genotypes. In total, 834 out of 1152 sequenced samples passed all quality thresholds and were subjected to further analyses. They consist of 2-3 biological replicates of 85 B73×IBM-RIL and 82 Mo17×IBM-RIL backcross hybrids and their IBM-RIL parents together with 23 (B73×Mo17) and 24 (Mo17×B73) biological replicates of the full reference hybrids and 47 biological replicates of B73 and 42 of Mo17 (Data file S1). The categorization of the IBM-RIL genomes revealed the putative recombination breakpoints between Mo17 and B73 regions (Figure S1).

We explored the sample quality and the relationship among the 834 RNA-seq samples in a multidimensional scaling (MDS) plot (Figure 1C). The different genotypes (B73, Mo17, B73×Mo17, Mo17×B73) and populations (IBM-RILs, B73×IBM-RILs, Mo17×IBM-RILs) form distinct clusters, which are clearly separated from each other. As expected, the clusters of the two reciprocal reference hybrids B73×Mo17 and Mo17×B73 overlap because their genomes are identical. These hybrid samples are located in between their parental inbred lines on dimension 1 of the MDS plot (Figure 1C). Similarly, the IBM-RILs are located between their original parental inbred lines B73 and Mo17. Finally, as expected the two backcross populations B73×IBM-RIL and Mo17×IBM-RIL are located between their parental lines (Figure 1C).

Root traits display heterosis in backcross populations

For each parent-hybrid combination, we determined lateral root density (Figure S2A), total number of root tips (Figure S2B), total root length (Figure S2C) and total root volume (Figure S2D). The inbred line B73 outperformed the inbred line Mo17 in all measured root traits. The B73×Mo17 hybrid displayed higher values than the Mo17×B73 hybrid for the total number of root tips, total root length and total root volume. In contrast, Mo17×B73 displayed a higher lateral root density than B73×Mo17 (Figure S2). In addition, we estimated mid-parent heterosis (MPH) for all measured root traits. The level of MPH of the fully heterozygous B73×Mo17 and

Mo17×B73 hybrids (from 49% for lateral root density to 127% for the number of root tips) exceeded the average of the partially heterozygous B73×IBM-RIL and Mo17×IBM-RIL hybrids (from 25% for lateral root density to 62% for total root volume) in all traits. Nevertheless, most of the IBM-RIL backcross hybrids displayed substantial MPH (Figure S2, Data file S2).

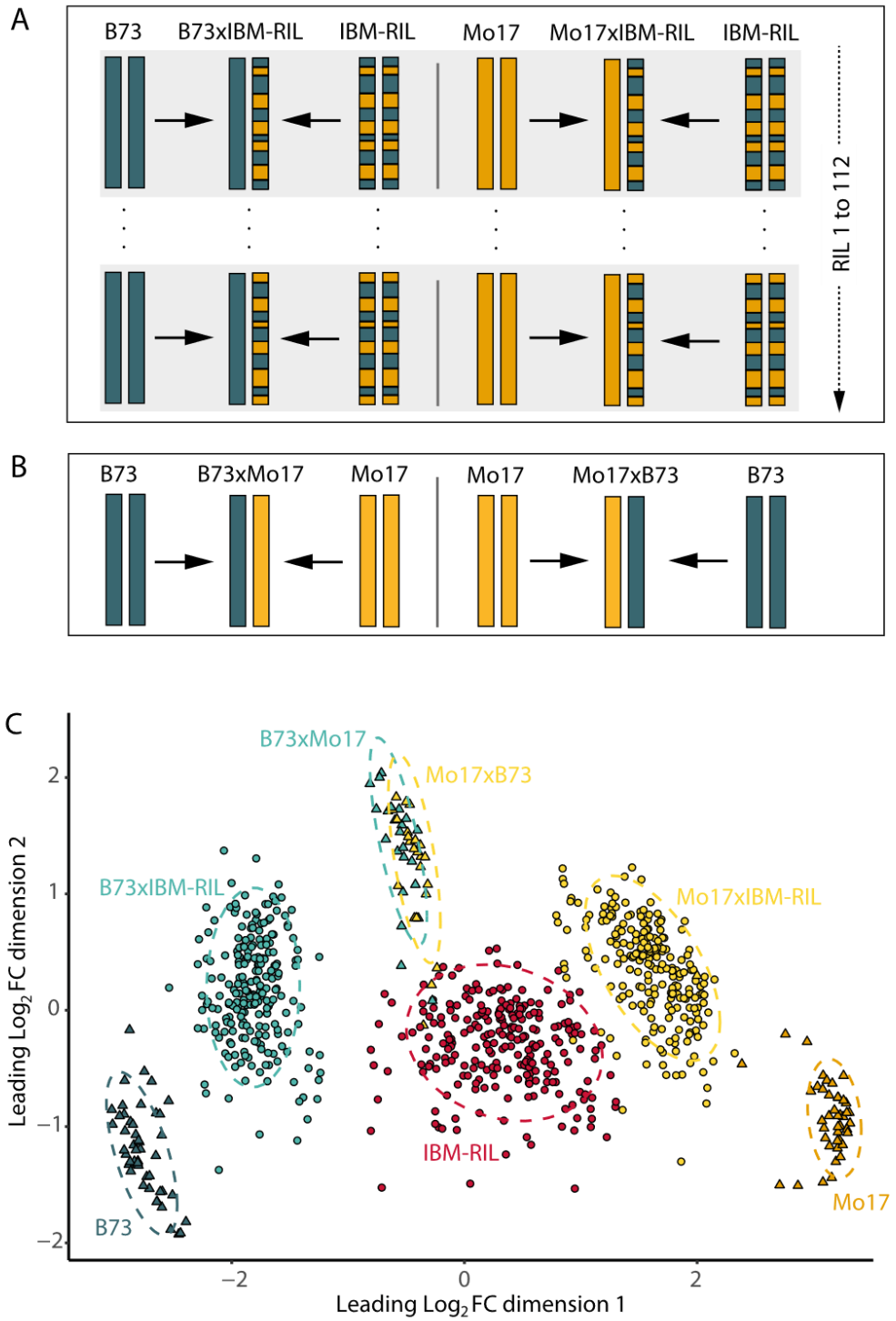


Figure 1: Plant genetic material and sample information. **A** Schematic depiction of the 112 IBM-RILs and their B73- and Mo17-backcross hybrids. **B** Schematic depiction of the reference genotypes. **C** Multidimensional Scaling Plot of 834 high-quality RNA-seq samples. Each genotype group is highlighted by a different colour. The reference genotypes B73, Mo17, B73xMo17 and Mo17xB73 are depicted as triangles, IBM-RILs and their backcrosses as circles. The leading Log_2FC between two samples can be interpreted as expression difference between those two samples.

Heterozygosity promotes single parent expression complementation

Genes which are active in one parental inbred line and in the F_1-hybrid offspring, but inactive in the second parent, display single-parent expression (SPE). In the highly heterozygous reference hybrid B73×Mo17 we identified 1297 SPE **B73**/Mo17 genes, active in the hybrid and the maternal inbred line B73 (pattern 1, Figure 2A; allele contributed by active parent in bold). In the reciprocal hybrid Mo17×B73 we identified 1241 SPE **Mo17**/B73 genes active in the hybrid and only in the maternal inbred line Mo17 (pattern 5, Figure 2B). In addition, we identified 1228 SPE B73/**Mo17** (pattern 2) and 1253 SPE Mo17/**B73** (pattern 6) genes active in the paternal inbred line and the hybrid, but not the corresponding maternal parent. Between both reference hybrids, 85% of the SPE genes were conserved (Data file S3).

Within the two IBM-RIL backcross populations we identified eight different SPE patterns, four in each backcross population, depending on the genomic composition of the gene in the paternal IBM-RIL. SPE genes of all patterns are expressed in the hybrid. But we can distinguish them by the active parent (Figure 2A, B; allele contributed by active parent in bold). Pattern 1 (B73×IBM-RILs): Genes lie within heterozygous regions of the hybrid and the expressed parent is B73; Pattern 2 (B73×IBM-RILs): Genes lie within heterozygous regions of the hybrid and the expressed parent is the IBM-RIL, which contributes the Mo17 allele for these genes; Pattern 3 (B73×IBM-RILs): Genes lie within homozygous regions of the hybrid and the expressed parent is B73; Pattern 4 (B73×IBM-RILs): Genes lie within homozygous regions of the hybrid and the expressed parent is the IBM-RIL, which contributes the B73 allele for these genes; Pattern 5 (Mo17×IBM-RILs): Genes lie within heterozygous regions of the hybrid and the expressed parent is Mo17; Pattern 6 (Mo17×IBM-RILs): Genes lie within heterozygous regions of the hybrid and the expressed parent is the IBM-RIL, which contributes the B73 allele for these genes; Pattern 7 (Mo17×IBM-RILs): Genes lie within homozygous regions of the hybrid and the expressed parent is Mo17; and Pattern 8 (Mo17×IBM-RILs): Genes lie within homozygous regions of the hybrid and the expressed parent is the IBM-RIL, which contributes the Mo17 allele for these genes. (Figure 2A, B).

We observed in the B73×IBM-RIL and Mo17×IBM-RIL backcrosses on average 1229 (Figure 2C) and 1247 (Figure 2D) SPE genes per hybrid, which is ~50% of the number of SPE genes in B73×Mo17 and Mo17×B73. Across B73×IBM-RIL and Mo17×IBM-RIL backcross hybrids, SPE patterns in heterozygous regions (Figure 2C, D: patterns 1, 2, 5, 6) occurred more often than in homozygous regions (Figure 2C, D: patterns 3, 4, 7, 8).

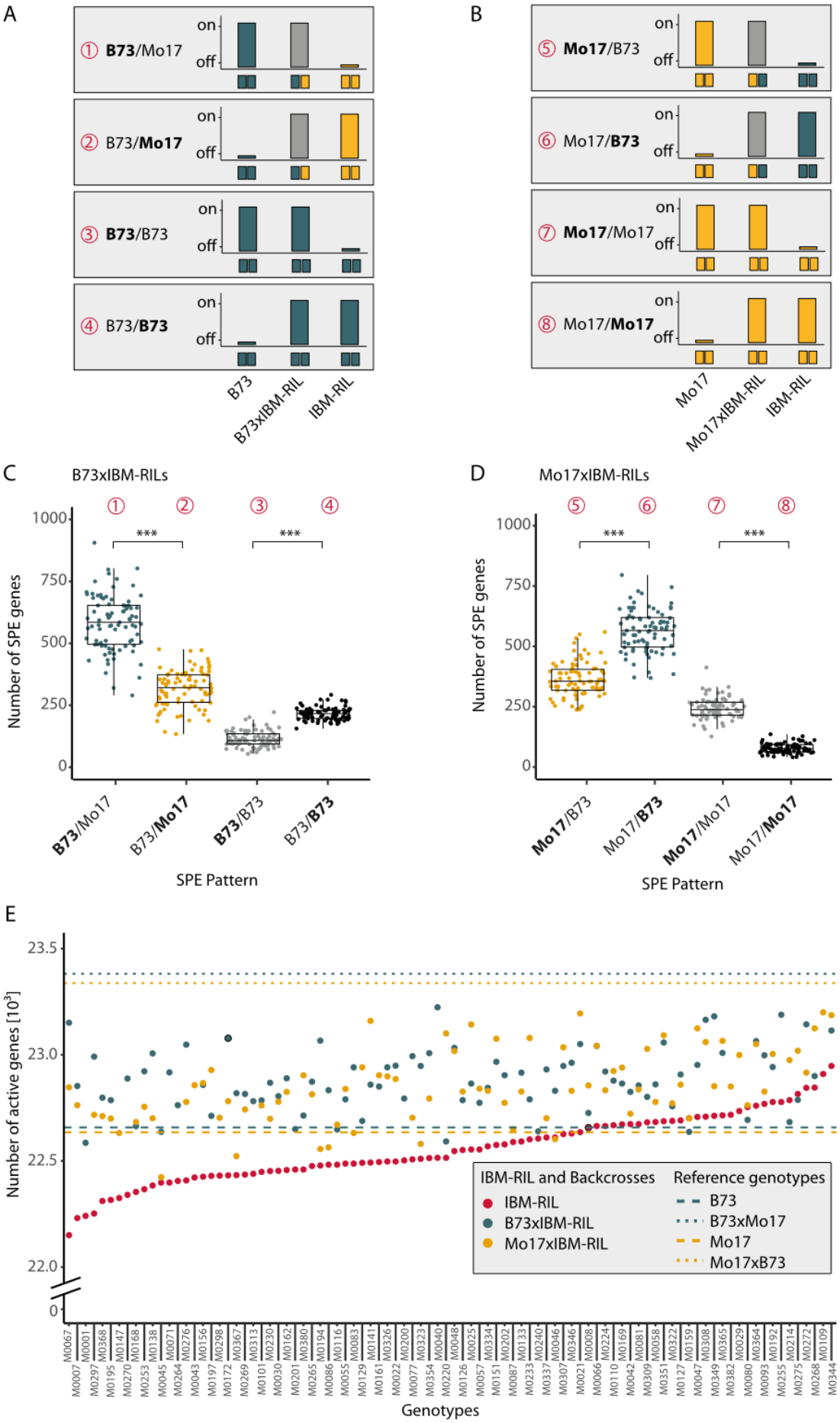

A
① B73/Mo17
② B73/Mo17
③ B73/B73
④ B73/B73
on
off
B73
B73xIBM-RIL
IBM-RIL
B
⑤ Mo17/B73
⑥ Mo17/B73
⑦ Mo17/Mo17
⑧ Mo17/Mo17
Mo17
Mo17xIBM-RIL
IBM-RIL
C
B73xIBM-RILs
Number of SPE genes
SPE Pattern
B73/Mo17
B73/Mo17
B73/B73
B73/B73
***
D
Mo17xIBM-RILs
Number of SPE genes
SPE Pattern
Mo17/B73
Mo17/B73
Mo17/Mo17
Mo17/Mo17
E
Number of active genes [10³]
Genotypes
IBM-RIL and Backcrosses
IBM-RIL
B73xIBM-RIL
Mo17xIBM-RIL
Reference genotypes
B73
B73xMo17
Mo17
Mo17xB73

Figure 2: Single-parent expression (SPE) complementation. **A** SPE pattern present in B73xIBM-RIL backcross hybrids relative to the genomic composition of the paternal IBM-RIL and the active parent (Pattern 1-4). **B** SPE pattern in Mo17xIBM-RIL backcross hybrids relative to the genomic composition of the paternal IBM-RIL and the active parent (5-8). A simplified depiction of the activity (on/off) of SPE genes is shown for each SPE pattern above the genotypic composition of the respective pattern for each parent and hybrid. Pattern 1, 2 and 5, 6 are also present in B73xMo17 and Mo17xB73 respectively. **C** Boxplots displaying the total number of SPE genes for each of the possible SPE pattern (1-4) in the B73xIBM-RIL hybrids (N= 85). **D** Boxplots displaying the total number of SPE genes for each of the possible SPE pattern (5-8) in the Mo17xIBM-RIL hybrids (N=82). Asterisks indicate significant differences ($\alpha<0.0001$, p-values given as <0.0001 in other words, zero) identified by a gaussian mixed model with the hybrid as random effect, the SPE pattern and non-SPE pattern as a fixed factor and a diagonal variance component for the SPE pattern. **E** For each IBM-RIL line (red) and the corresponding B73xIBM-RIL (blue) and Mo17xIBM-RIL (yellow) backcross hybrids the number of active genes is displayed. The dashed lines represent the number of active genes in the inbred lines B73 (blue) and Mo17 (yellow). The dotted lines represent the number of active genes in the reciprocal hybrids B73xMo17 (blue) and Mo17xB73 (yellow).and Mo17xB73 (yellow).

We identified more SPE genes in heterozygous regions with B73 as the active parent (Figure 2C, D: pattern 1 and 6) than with Mo17 as the active parent (Figure 2C, D: pattern 2 and 5). In regions homozygous for B73, more genes with paternal activity were observed (Figure 2C: patterns 4 vs 3), whereas in regions homozygous for Mo17 more genes with maternal activity were determined (Figure 2D: pattern 7 vs 8).

As a consequence of expression complementation of genes expressed in only one parent, which are then also active in the hybrid (SPE), we observed that hybrids express more genes than their parental lines (Figure 2E). While the inbred lines B73 and Mo17 express 22 658 and 22 635 genes, respectively, their reciprocal F_1-hybrid offspring B73×Mo17 and Mo17×B73 display 736 and 692 more active genes than their parental average. The corresponding B73×IBM-RIL and Mo17×IBM-RIL backcrosses expressed on average 278 and 253 more genes than their parents. Furthermore, the number of active genes in the backcross hybrids is positively correlated with the fraction of heterozygous genomic regions in the hybrid (Figure S3). These results indicate that the level of heterozygosity is important for expression complementation in hybrids.

Contribution of SPE to heterosis

To estimate the proportion of heterosis variance explained by the number of SPE genes underlying MPH of root traits, we determined in each backcross population the coefficient of determination p_{HET}. Across the four examined root traits, 12% (total root volume) to 29% (total number of root tips) of the heterotic variance was explained by the number of SPE genes across the four different patterns in the B73×IBM-RIL backcross hybrids. In contrast, in the Mo17×IBM-RILs only between 4% (lateral root density) and 9% (total root volume) of the heterotic variance was explained by the total number of SPE genes. For the traits of total number of root tips and total root length in the Mo17×IBM-RILs we observed a negative proportion (-8% and -0.5%) of explained heterotic variance of the number of SPE genes on MPH (Table 1).

Table 1: Proportion of heterotic variance explained by the number of SPE genes on mid-parent heterosis for different root phenotypes

Trait	B73xIBM-RILs			Mo17xIBM-RILs		
	σ^2_{Het}	σ^2_G	p_{Het}	σ^2_{Het}	σ^2_G	p_{Het}
No. Of root tips	0.232	0.325	0.29 (29%)	0.205	0.190	-0.08 (-8%)
Total root volume	2.366	2.701	0.12 (12%)	1.022	1.121	0.09 (9%)
Total root length	2.715	3.561	0.24 (24%)	2.202	2.191	-0.01 (-1%)
Lateral root density	5.461	6.805	0.20 (20%)	6.675	6.925	0.04 (4%)

σ^2_{Het} = unexplained genetic variance of heterosis effect, not associated with SPE genes; σ^2_G = total genetic variance among the hybrid genotypes; p_{Het} = Coefficient of determination: proportion of the heterotic variance explained by the number of SPE genes.

Trans-regulatory elements are more frequent in Mo17 than B73

An expression quantitative trait locus (eQTL) is a position in the genome that is significantly associated with expression variation of a gene and thus likely regulates this gene. We identified 13,778 eQTL for 13,434 protein-coding genes. While most genes were regulated by a single eQTL, 334 genes were regulated by two eQTL and five by three eQTL (Data file S4). We categorized all eQTL into either *cis*- or *trans*-regulating by their location relative to the starting position of the corresponding gene. We defined *cis*-regulating eQTL as located at a distance of <2.5 Mbp from the start codon of their target gene on the same chromosome. In contrast, most *trans*-regulating eQTL were located on chromosomes other than those of their target gene (83%). The remaining *trans* eQTL were located on the same chromosome as their target at a distance of ≥2.5 Mbp (and the target gene was outside of the eQTL confidence interval). The logarithm of odds (LOD) is the significance measure of an eQTL being present using the Haley-Knott regression. The median LOD was 15.5 for *trans*- and 16.3 for *cis*-eQTL (Data file S4). Effect sizes usually estimate how big the effect of each identified locus is on the phenotype, or in this case on gene expression. As LOD values and effect size estimates of the eQTL on the gene expression are correlated[21], this indicates similar effect sizes of *cis*- and *trans*-acting eQTL. In total, 88% (12,057/13,778) of all identified eQTL acted in *cis*, while 12% (1,701/13,778) acted in *trans* on their target genes. For genes active in the B73 reference genotype 7% (771) of their eQTL were acting in *trans* while for genes active in the Mo17 reference genotype almost twice as many eQTL acting in *trans* were identified (13%, 1,439). This suggests that Mo17 contains more *trans*-acting eQTL than B73. The values for all other genotypes show *trans*-acting eQTL ratios between the B73 and Mo17 values (Table 2).

Heterozygous SPE genes with Mo17 activity are *trans*-regulated disproportionately more often

In both fully heterozygous reference hybrids B73×Mo17 and Mo17×B73, we detected eQTL for 85% of the SPE genes (Data file S3). Overall, 95% of eQTL for SPE genes with B73

contributing the active parental allele were *cis*-regulating (Figure S4). By contrast, SPE genes with Mo17 as the active allele were predominantly (58 - 59%) *trans*-regulated (Figure S4).

Table 2: *Trans* eQTL for active genes

Genotype	***Trans*-regulating eQTL [in % of eQTL for active genes]**	**Number of *trans*-regulated genes**
B73	7	771
B73_Mo17	11	1365
Mo17	13	1439
Mo17_B73	11	1363
IBM-RILs*	10	1107
B73xIBM-RILs*	9	1066
Mo17xIBM-RILs*	12	1385

*Values are the mean of the respective category

Similarly, in B73×IBM-RIL and Mo17×IBM-RIL backcross hybrids (Figure S5), heterozygous SPE genes where B73 is the active allele are regulated almost exclusively by *cis* eQTL (on average 95% and 96%) (Figure S5A: pattern 1; Figure S5B: pattern 6). By contrast, on average only 60% (Figure S5A pattern 2) and 62% (Figure S5B pattern 5) of heterozygous SPE genes with an active Mo17 allele are *cis*-regulated. Hence, in general heterozygous SPE genes with an active Mo17 allele are significantly more frequently *trans*-regulated than those with B73 as the active parental allele (Table 2).

Homozygous SPE genes are partially *trans*-regulated

SPE genes located in homozygous regions showed different ratios of *cis*- to *trans*-eQTL regulation, based on the active parent. While in homozygous B73 regions (occurring in B73×IBM-RILs), SPE with maternally active alleles (**B73**/B73, Figure S5A, pattern 3) were primarily *cis*-regulated (87%), those with paternally active alleles (B73/**B73,** Figure S5A pattern 4) were primarily *trans*-regulated (67%). In homozygous Mo17 regions (Mo17×IBM-RILs) we saw the opposite. SPE genes with maternally active alleles (**Mo17**/Mo17, Figure S5B pattern 7) showed more *trans*-regulating eQTL (58%) and for SPE with paternal activity (Mo17/**Mo17,** Figure S5B pattern 8) the majority of eQTL (81%) were *cis*-regulating. Interestingly, the SPE patterns 4 and 7, which are primarily regulated by *trans*-eQTL, also had higher absolute numbers of SPE genes (Figure 2A, B). The observed proportions of *trans*-regulation for SPE patterns 1 to 8 were significantly different from the ratios in non-SPE genes, of which 91% were *cis*-regulated (Figure S5).

SPE genes are predominantly regulated in heterozygous regions

Overall, we observed that for heterozygous SPE genes, both *cis*- and *trans*-regulating eQTL are predominantly located in heterozygous genomic regions (Figure 3). We observed that *cis*-

regulating eQTL located in hetero- and homozygous regions regulate SPE genes which are also hetero- or homozygous as the corresponding eQTL. This is expected, as *cis*-eQTL are located in close proximity to their gene (Figure 3A, Figure 3B). In contrast, target genes of *trans*-regulating eQTL are randomly distributed across the genome. Accordingly, *trans*-acting eQTL regulate SPEs in both homozygous and heterozygous regions with equal frequency, independent of whether the eQTL is homo- or heterozygous (Figure 3A, Figure 3B). Together with the association of Mo17 activity with *trans*-regulation, this leads to the observed ratios of *cis*- to *trans*-regulation in homozygous SPE patterns. For instance, eQTL regulating homozygous genes in *trans* (Figure 3A and 3B, patterns 4 and 7) were located almost exclusively in heterozygous regions. In both cases, the active parent carries the Mo17 allele at the eQTL position which is responsible for the gene activity of those SPE genes (Figure S6 A, B). This further explains the higher proportion of *trans*-regulation of pattern 4 and 7 as well as the generally higher number of SPE genes for these patterns compared to pattern 3 and 8. In summary, SPE patterns with parental B73 activity showed a slightly higher proportion of regulation by *cis*-acting eQTL compared with non-SPE genes, while the SPE patterns with Mo17 activity showed substantially lower regulation by *cis*-acting eQTL and thus higher *trans*-regulation (Figure 3, Figure S5).

Trans-regulated genes and SPE genes are often non-syntenic

The maize genome contains genes with orthologs of positional synteny in sorghum, suggesting these genes are highly conserved across evolution, and a set of genes without syntenic partners (non-syntenic) in sorghum and other grass species, which are therefore most likely evolutionarily younger or changed their position during evolution[22]. In the B73v5 maize genome, 40% (15 612) of genes can be classified as non-syntenic. Among the active genes in this study, 27% (7622) were non-syntenic. By contrast, on average 58% of SPE genes in the hybrids are non-syntenic.

Interestingly, genes with *trans*-eQTL were more likely to be non-syntenic (70%) than genes with *cis*-eQTL (22%) across all expressed genes (Table S1). Both *cis*- and *trans*-regulated SPE genes were more frequently non-syntenic than non-SPE genes (Figure 4). Additionally, non-syntenic genes were particularly prominent among the t*rans*-regulated SPE genes with Mo17 activity. For example, 96% of *trans*-regulated pattern 2 genes (B73/**Mo17**) were non-syntenic (Figure 4). It should be noted that these are relative values: the absolute numbers of genes in patterns 3 and 8 are low in general, and patterns 1 and 6 are not often *trans*-regulated (Figure 3).

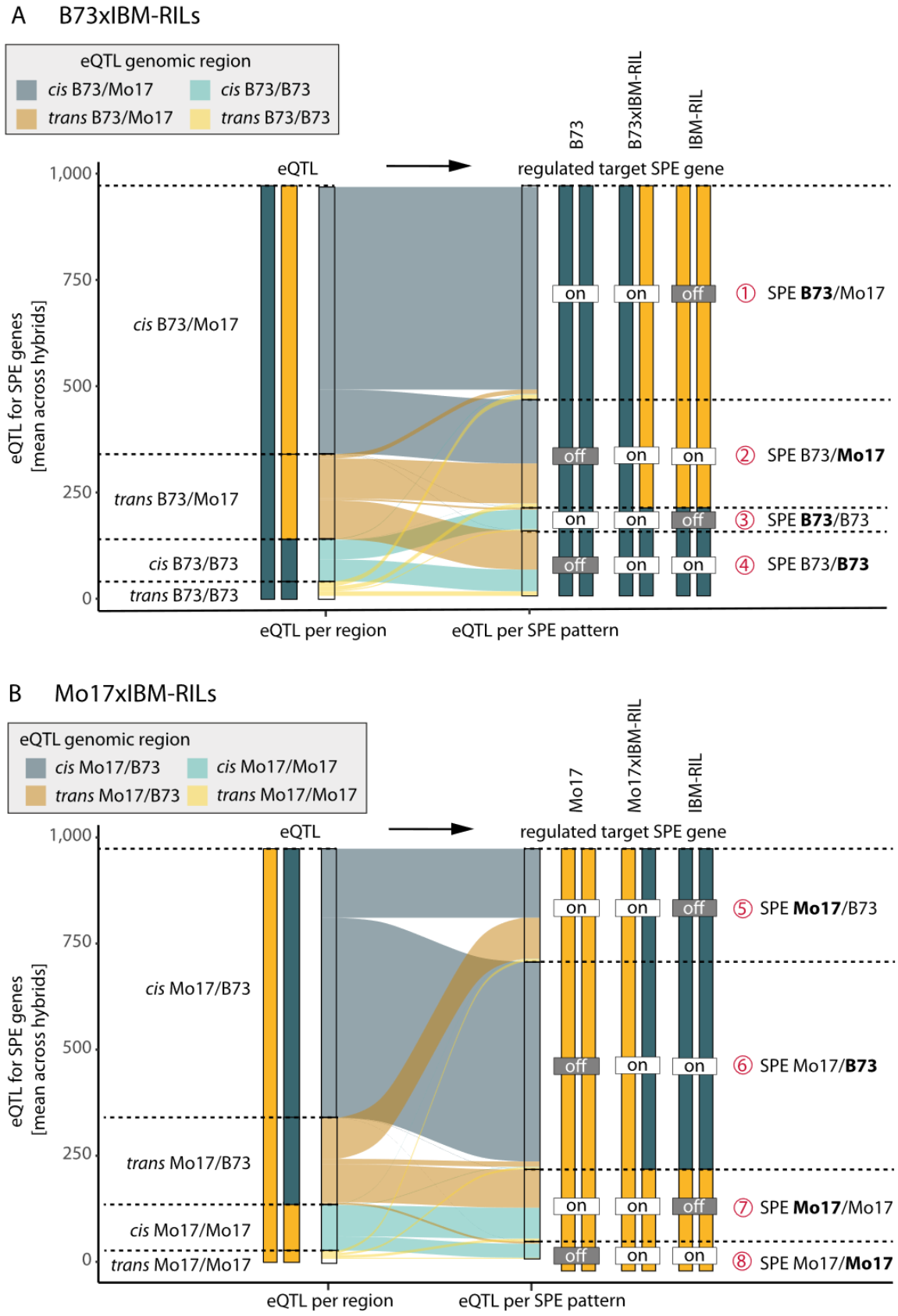


Figure 3: The average number of eQTL per genomic region for different SPE pattern in **A** B73xIBM-RILs and **B** Mo17xIBM-RILs. The width of the connecting bands corresponds to the average number of eQTL. The shade indicates the genomic region of the eQTL, (dark = heterozygous, light = homozygous) and blue colour corresponds to *cis* and orange to *trans*-regulation. The genotypes at the eQTL position and regulated SPE genes are indicated as bars (yellow = Mo17 allele, turquoise = B73 allele), and the gene activity for SPE is shown as (on/off).

SPE gene expression influences lateral root density in different ways

We performed a transcriptome wide association analysis (TWAS) to identify genes whose expression values associate with phenotypic traits. We analyzed four root traits (lateral root density, number of root tips, total root length, total root volume) and the two IBM-RIL backcross hybrid populations B73×IBM-RIL and Mo17×IBM-RIL separately, as different mechanisms of expression regulation and expression levels might be observed between the populations. We found 18 positively and 17 negatively correlated genes across the two populations. Only one gene was identified in both populations (Data file S5). Therefore, different genes might control hybrid vigor of roots in the different populations.

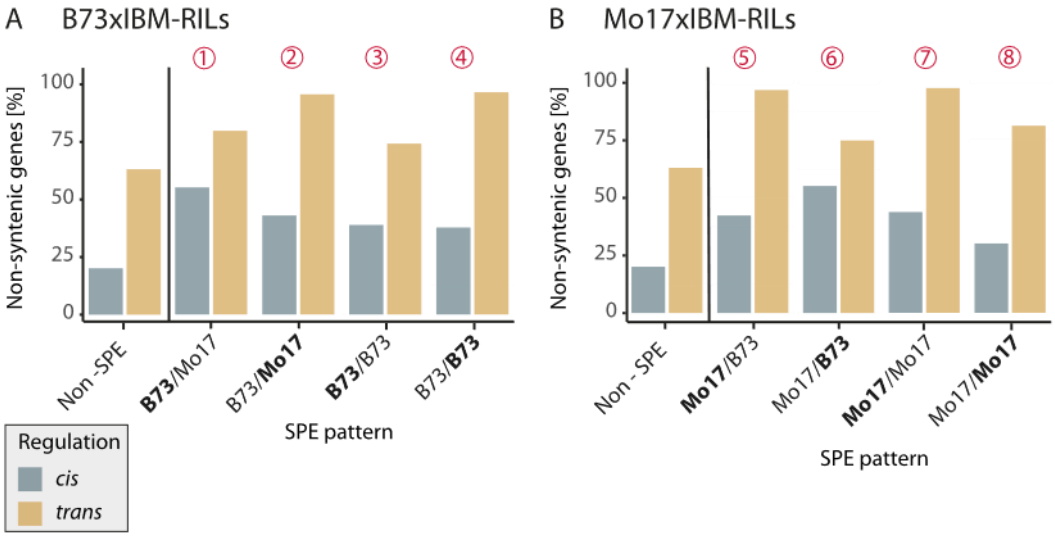


Figure 4: Synteny of SPE pattern genes. The average proportion of non-syntenic genes (vs. syntenic genes) among *cis*- (blue) and *trans* (yellow)-regulated SPE genes. Percentages are an average of genes in the respective category across all **A** B73xIBM-RILs (N=85) or **B** Mo17xIBM-RILs (N=82).

Among the 35 TWAS genes whose expression correlated with phenotypic traits, 7 showed SPE complementation in more than 10 hybrids. We designated these TWAS & SPE (TSG) genes (Data file S6). All of the TSG were identified in the B73×IBM-RIL hybrid population. For lateral root density, 10 TWAS genes were identified by the BLINK and MLMM models of GAPIT in total, and 4 of these genes also showed an SPE pattern (TSG) (Data file S6, Figure 5A). We followed up on the two TSG which showed an SPE pattern in most hybrids (TSG 1 & TSG 2) (Figure 5B, C). TSG 1 (Zm00001eb349930) showed an SPE pattern in 30 of the 85 B73×IBM-RIL hybrids (Figure 5B, Data file S6), predominantly the SPE B73/Mo17. The gene was *cis*-regulated and non-syntenic. B73×IBM-RIL hybrids with an active TSG 1 showing an SPE pattern had significantly lower lateral root density ($\alpha \leq 0.05$, Figure 5B).

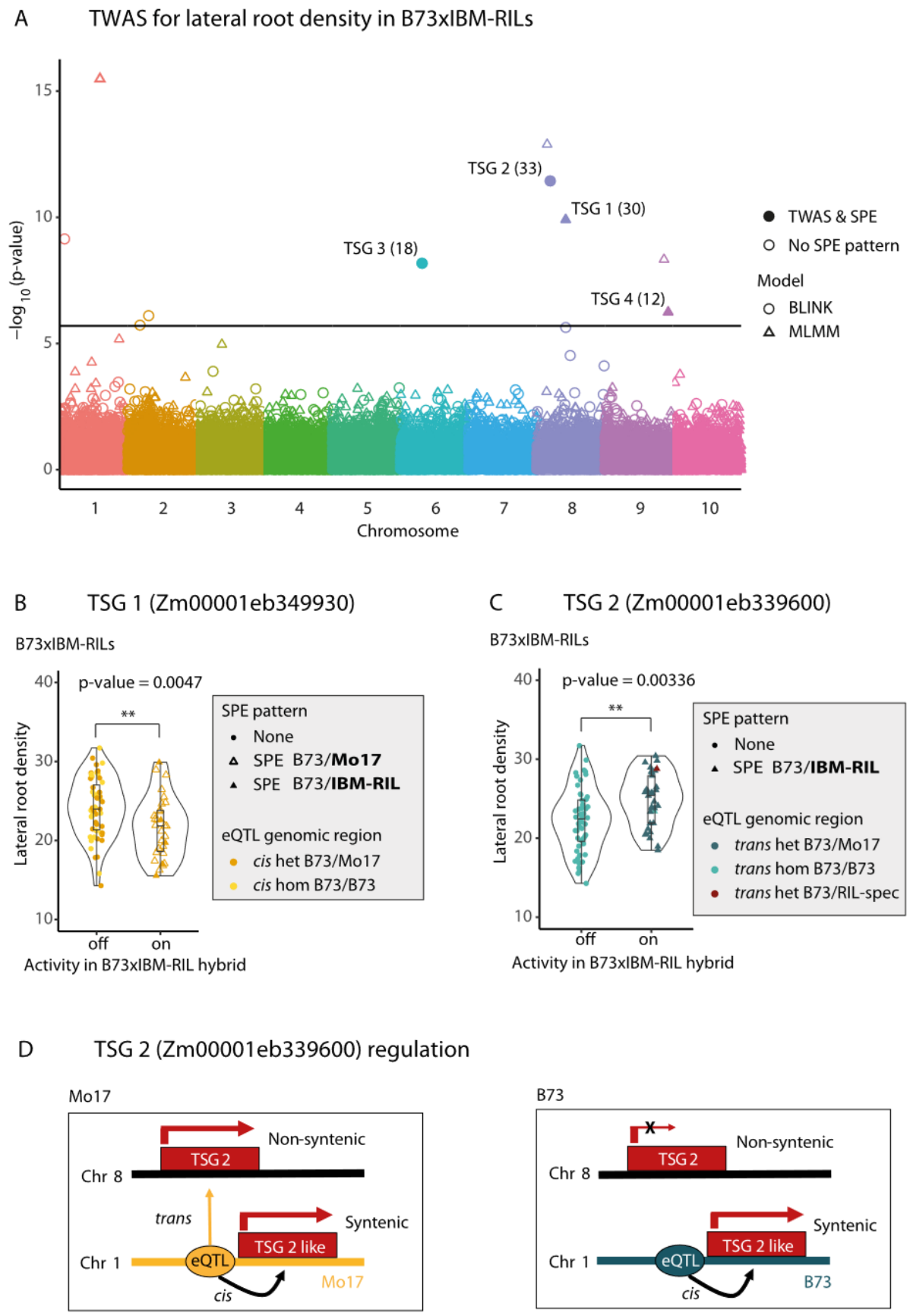


Figure 5: Transcriptome Wide Association Analysis (TWAS). **A** Manhattan plot of TWAS analysis for lateral root density in B73xIBM-RIL hybrids. Each point shows a gene with the p-value of the association between the gene's expression with lateral root density. Results for the analysis with the BLINK model are shown as circles, results of

the MLMM model are shown as triangles. Filled points indicate significant TWAS genes, that also show SPE pattern (TSG), with the number of hybrids where the gene shows an SPE in brackets. The horizontal line indicates the fdr = 0.05 by Bonferroni correction. **B**, **C** P-values correspond to Student's t-test independent from the TWAS analysis based on gene activity. Points show lateral root density of each B73xIBM-RIL with either inactive (off) or active (on) **B** TSG 1 Zm0001eb349930 or **C** TSG 2 Zm0001eb339600 in the hybrids. Shape corresponds to the SPE pattern and colour to the genotype at the eQTL position (yellow = *cis* regulating, blue = *trans* regulating, dark = heterozygous regions, light = homozygous regions). The colour shows the genotype of the hybrid at the eQTL position. **D** shows a model of gene expression of the non-syntenic gene TSG 2 and the syntenic paralog of TSG 2, TSG 2 like, which are regulated from the same eQTL. Depiction shows activity in Mo17 and B73 inbred lines. In heterozygous genotypes, TSG 2 is active, when at least one Mo17 allele is present at the eQTL position. TSG 2 like is active in all genotypes.

A different effect was visible for TSG 2 (Zm00001eb339600), which showed B73/**IBM-RIL** pattern in 33 B73×IBM-RIL hybrids (Figure 5C Data file S6). As there were no SNPs near the gene to classify the paternal allele of the gene itself, the SPE pattern was called B73/**IBM-RIL**, instead of B73/**Mo17** or B73/**B73**. Interestingly, TSG 2 was *trans*-regulated and the genotype at the eQTL position could be determined. In the B73×IBM-RILs, the hybrids with an active TSG 2 had a significantly ($\alpha \leq 0.05$) higher lateral root density and showed the heterozygous B73/Mo17 genotype at the eQTL position (Figure 5C). When searching for sequences similar to TSG 2 (Zm00001eb339600) in the B73 reference genome, we identified a paralog (Zm00001eb039610) and called it TSG 2-like. TSG 2-like is located close to the eQTL position of TSG 2 (<2.5 Mbp) (Data file S7, Figure 5D). While the TWAS gene TSG 2 is non-syntenic and *trans*-regulated, its paralog TSG 2-like at the eQTL position is *cis*-regulated from the same eQTL position and syntenic (Figure 5D). The syntenic TSG 2-like gene is expressed in Mo17 and in B73 and expressed in all genotypes. However, the non-syntenic TSG 2 is only active in Mo17, but not in B73 (Figure 5D) and subsequently only active in those hybrids where at least one Mo17 allele is present at the eQTL position (Figure 5C).

Discussion

SPE complementation, where the hybrid expresses a gene that is only active in one of its parents, was suggested to contribute to the translation of parental diversity into phenotypic heterosis[10]. This concept is in line with the dominance model of heterosis which assumes that many slightly deleterious alleles of the parents are complemented in the hybrid by dominant alleles[7]. The number of SPE genes identified in this study is similar to previous studies for B73xMo17 and Mo17xB73 seedling roots[11,12,23]. In the IBM-RIL backcross hybrids, we identified on average a substantially lower phenotypic mid-parent heterosis, fewer SPE genes compared to the fully heterozygous hybrids and a lower average heterozygosity. We also observed that SPE genes are mainly located in heterozygous regions of the genome, thus explaining the lower numbers of SPE in the backcross hybrids (Figure 2). Hence, the association of the number of SPE genes with heterosis is not only conditioned by genetic diversity[10] but also by the degree of heterozygosity. As a result of SPE complementation, the number of active genes increased with heterozygosity (Figure S3).

We demonstrated that the number of genes showing SPE complementation in homozygous and heterozygous regions of the genome (Figure 2A-D) explains up to 29% of the heterotic variance in the phenotype of the hybrids (Table 1). As different genes are likely responsible for heterosis in different traits[24], our finding suggests that SPE as a group of genes are in general responsible for heterosis, but to a different extent for different traits.

It is for instance likely that SPE genes influence heterosis in B73xIBM-RILs but to a smaller degree in Mo17xIBM-RILs, which in general generate less vigorous hybrids. In our study we detected different ratios of *cis* and *trans* regulation among active genes in inbred lines (Mo17:13% *trans*-regulation; B73; 7% *trans*-regulation; Table 2). In the reciprocal hybrids, this difference was significantly amplified for SPE genes (Mo17 active: ~60% *trans*; B73 active: ~5% *trans*; Figure S4), similar to the IBM-RIL backcross populations (Figure S5).

Previously, contrasting *cis*- and *trans*-regulation was associated with parental alleles in another eQTL study, using IBM-RIL backcrosses from a smaller subset of the IBM-RIL population[14]. In that study, 86% of *trans*-regulation was associated with paternal dominance, where the expression level of the paternal allele was adopted[14]. In our study, investigation of the eQTL positions revealed that most SPE genes are regulated by eQTL located in heterozygous regions. Substantial proportions of SPE genes located in homozygous regions are regulated in *trans*- from eQTL in heterozygous regions, leading to higher numbers of these specific SPE patterns (Figure 3, pattern 4 and 7). We showed that for many *trans*-regulated SPE genes, the presence of the Mo17 allele in the eQTL is required for gene activity in the SPE pattern genes in the hybrid (Figure 3, Figure S6). Thus, we did not observe paternal dominance of *trans*-

regulated genes regarding SPE genes, but rather a Mo17 dominance of *trans*-regulated SPE genes.

Our present data confirms findings that non-syntenic genes are enriched among SPE genes[11] and are correlated with heterosis[25] (Figure 4). We further expanded this concept by demonstrating that non-syntenic genes are enriched among *trans*-regulated genes. Non-syntenic genes have been associated with disease resistance genes[26] and were suggested to function in environmental adaptation of plants and help hybrids to cope with abiotic stress[23]. We identified a *trans*-regulated non-syntenic SPE candidate gene TSG 2 (Zm00001eb339600), which controls lateral root density in hybrids, whose expression was induced by the Mo17 allele at the eQTL position (Figure 5 B). This gene has a syntenic paralog Zm00001eb039610 (TSG 2-like), which is located close to the eQTL position (<2.5Mbp) (Figure 5C, Data file S7). Interestingly, this paralog is *cis*-regulated from the same eQTL position as TSG 2 (Figure 5C) and expressed in B73, Mo17 as well as heterozygous genotypes. We hypothesize that there might be a regulatory connection of the *trans*-regulated non-syntenic gene to the syntenic paralog, in the Mo17 genotype. Regulatory interactions of paralogous genes have been previously reported. For instance, the paralogous genes *rcts* and *rctl* are regulated by the same transcription factor[27] and the syntenic gene *rtcs* recruited younger non-syntenic genes during seminal root evolution[28]. A regulatory connection of *trans*-regulated non-syntenic genes with their syntenic *cis*-regulated paralogs in the Mo17 genotype could explain the high number of *trans*-regulating eQTL among non-syntenic SPE patterns associated with the Mo17 allele. The regulatory differences (B73: *cis*, Mo17: *trans*) between the parents of a hybrid and their different contributions to SPE pattern might be an aspect of how phylogenetic distance is contributing to heterosis.

Among the TWAS genes whose expression correlated with phenotypic traits, we identified genes in the B73xIBM-RIL population for lateral root density, which displayed SPE in a substantial number of parent-hybrid combinations (Data file S6, Figure 5). We surveyed the two candidate genes which displayed SPE in the highest number of parent-hybrid combinations in more detail. They showed significantly different lateral root density, based on the activity or inactivity of the gene in the hybrid, and thus the presence or absence of SPE. For the *cis*-regulated gene TSG 1 (Zm00001eb349930) (Figure 5B), a lower lateral root density was observed upon gene activity. In contrast, gene TSG 2 (Zm00001eb339600) (Figure 5C) is *trans*-regulated and displayed a higher lateral root density, upon gene activity, which is induced by the Mo17 allele at the eQTL position.

Thus, we observe regulatory effects of single genes displaying SPE on lateral root density, which might help maize to adapt to changing local environmental conditions such as water

availability where disparate lateral root densities are beneficial[29]. In summary, we also demonstrated that the association of the number of SPE genes with heterosis is not only conditioned by genetic diversity but also by the degree of heterozygosity. Additionally, SPE mediated phenotypic heterosis, as well as the regulation of SPE genes in IBM-RIL backcross populations depends on the genetic background of the population.

Material and Methods

Plant genetic resources

The IBM-RIL syn. 4 population[16] represents a collection of highly diversified genotypes with respect to the genomic regions contributed by their two parental inbred lines B73 and Mo17 (Figure 1A). To study the phenotypic and transcriptomic plasticity of maize F_1-hybrids relative to their parents, a random subset of 112 IBM-RILs, was backcrossed to their original parental inbred lines B73 and Mo17. In all crosses, B73 and Mo17 were selected as the female parent to secure a homogenous phenotype of all plants on which the pollinated ears will develop and thus similar seed quality. Each of the two F_1-backcross hybrids of a specific IBM-RIL (hereafter named B73×IBM-RIL and Mo17×IBM-RIL backcross hybrids) show contrasting homozygous and heterozygous genomic regions and genes (Figure 1A). For example, if a region between two recombination breakpoints is homozygous in a B73×IBM-RIL backcross, it is heterozygous in the corresponding Mo17×IBM-RIL backcross hybrid and vice versa.

Experimental design

We studied the phenotypic and transcriptomic plasticity of the IBM-RIL backcross hybrids relative to their parental inbred lines. A selection of 112 IBM-RILs were used as paternal inbred lines, corresponding to 112 B73×IBM-RIL and 112 Mo17×IBM-RIL backcross hybrids (Figure 1A). To optimally fit the experimental design and increase the precision of subsequent pairwise comparisons with the common parental inbred lines and reference hybrids both common maternal inbred lines B73 and Mo17 and the two reference hybrids B73×Mo17 or Mo17×B73, respectively, were included (Figure 1B). For each sample, 25 kernels of the same genotype (parent or hybrid) were surface sterilized in 10% H_2O_2 for 20 min, rinsed with distilled water and afterwards pre-germinated in filter paper rolls with five kernels each in a climate chamber with a 16 h light (26 °C), 8 h dark (21 °C) cycle in distilled water[11]. After three days, eight seedlings per genotype with approximately the same length of primary root and, if already present, shoot length were selected and transferred into a row of an aeroponic growth system for four additional days. Each aeroponic growth system (“Elite Klone Machine 96”, TurboKlone, USA) was composed of 12 rows each with eight planting sites. Thus, we could fit 12 different genotypes into one aeroponic growth system and eight systems at the same time into our climate chamber (16 h light, 26 °C; 8 h dark, 21 °C) (Figure S7). We analyzed three independent biological replicates, each comprising all IBM-RIL inbred lines and hybrids. Due to the large number of samples and space limitations, the different genotypes of each biological replicate were grown in four batches distributed across four weeks, also called alpha-design with incomplete blocks[30]. Within each batch, the eight aeroponic growth systems were randomly assigned to eight positions in the climate chamber. Three successive rows of an

aeroponic growth system represent one triplet. To each triplet an IBM-RIL and its corresponding B73 and Mo17 backcross hybrids, or both common maternal inbred lines B73 and Mo17 and one of the two reference hybrids B73×Mo17 or Mo17×B73, respectively were assigned. The randomization process was conducted at the replicate level for the triplets, whereas it was ensured, that in each batch two reference triplets each of B73, Mo17, B73xMo17 and B73, Mo17, Mo17xB73 were distributed. Thus, in each batch we surveyed 30 IBM-RIL triplets and both reference hybrids and the common inbred lines B73 and Mo17 in two additional triplets (Figure S7). So that in total, 3 samples of each IBM-RIL and each backcross hybrid, 48 samples (biological replicates) of the maternal inbred lines B73 and Mo17, and 24 samples of the reciprocal hybrids B73×Mo17 and Mo17×B73 as reference hybrids were analyses. In other words, each independent biological replicate contained 384 samples (in total: 384 x 3 replicates = 1,152 samples). The number of individual samples per replicate was designed to also fit one sequencing run on the NovaSeq 6000 S4 flow cell machine (Illumina, San Diego, USA), described later.

Root phenotyping and sampling for RNA-seq

Seven days after germination, all seedlings per sample (maximum eight seedlings) were removed from the aeroponic growth system and the seedling root system was scanned using an Epson Expression 12000XL scanner (Epson, Meerbusch, Germany) with up to four plants per image. The resulting images were cropped to create single plant images that only showed the root system and the maize kernel. We used the RootPainter software client (version 0.1.0) and server component (version 0.2.7) to train a convolutional neural network to recognize and segment roots in images[31]. We then analyzed the segmented images in a batch using RhizoVision Explorer (version 2.0.3)[32]. After inspection and cleanup, we determined the total root length, total root volume and number of root tips for each plant for subsequent analysis (Details in Supplement Material SM1).

After imaging the seedlings, the primary root was separated from the kernel to collect (i) the proximal first centimeter with emerged lateral roots in 80% Ethanol to count the number of lateral roots per cm as density and (ii) the distal region of the primary root, composed of the root tip and the meristematic zone followed by the elongation zone, in liquid nitrogen for subsequent RNA extraction.

Analysis of phenotypic data

To evaluate the phenotypic data of each genotype a linear mixed model (baseline model 1) with a fixed effect for block (three replicates as levels) and genotype (263 levels) was fitted. According to the layout of our experimental design (Figure S7) we included random effects for

row, triplet, system and batch effect in the model. The residual error assesses the within-row variance among plants.

$$Y_{jklmnp(i)} = \mu + g_i + b_j + p_k + s_l + t_m + r_n + e_{jklmnp(i)} \quad (1)$$

$Y_{jklmnp(i)}$ represents the mean phenotypic value of a specific trait of interest of the respective genotype *i*, μ represents the intercept, g_i represents the fixed effect for genotype *i*, b_j represents the fixed effect for block *j*, p_k represents the random effect for batch *k* nested within block *j*, s_l represents the random effect for system *l* nested within batch *k* and block *j*, t_m represents the random effect for triplet *m* nested within system *l*, batch *k* and block *j*, r_n represents the random effect for row *n* nested within triplet *m*, system *l*, batch *k* and block *j*, and $e_{jklmnp(i)}$ represents the random error effect for plant *p* of genotype *i* in block *j*, batch *k*, system *l*, triplet *m* and row *n*. To fulfil the assumptions of linear models, the phenotypic values for the traits “total root length”, “total root volume” and “total number of root tips” had to be square root-transformed. An offset of 0.5 was added to each phenotypic value before transformation. The resulting modelled means were transformed back to their original scale for visual inspection (Figure S2). Modelled means on the transformed scale (and original scale in case of lateral root density) were used for TWAS analysis (described below).

RNA-sequencing and preparation of alignments

For subsequent RNA extraction and sequencing, a maximum of eight primary roots of each genotype grown in the same row of an aeroponic system were pooled. These root samples were manually ground in liquid nitrogen and total RNA was isolated with the RNeasy Plant Mini Kit (QIAGEN, Venlo, the Netherlands). RNA quality was assessed with a Bioanalyzer (RNA ScreenTape + TapeStation Analysis Software 3.2, Agilent Technologies, Santa Clara, CA, USA) by the Next Generation Sequencing (NGS) Core Facility in Bonn, Germany (https://btc.uni-bonn.de/ngs/), which subsequently constructed cDNA libraries for RNA-seq according to the TruSeq stranded mRNA library preparation protocol (Illumina, San Diego, USA). Sequencing was performed on a NovaSeq 6000 S4 flow cell machine (Illumina, San Diego, USA), generating 100-bp paired-end reads. This allowed for processing all 384 samples of a single replicate in one flow-cell and each batch of 96 samples on one lane. The obtained reads are reversely stranded. The raw reads were trimmed and filtered using Trimmomatic (version 0.39) in paired-end mode with the following settings: ILLUMINACLIP:adapters/TruSeq3-PE-2.fa:2:30:10:8:True, LEADING:3, TRAILING:3, MAXINFO:30:0.8, and MINLEN:40. With this step, remaining adapter sequences were removed, low quality bases from the start and end of the reads were cropped and adaptive quality trimming was performed. After these quality control, reads with a minimum length of 40 bases were retained and resulting single-end reads were excluded[33]. The maize reference

genome B73 version 5 (B73v5, ftp.ensemblgenomes.org/pub/plants/release-52/fasta/zea_mays/dna/Zea_mays.Zm-B73-REFERENCE-NAM-5.0.dna.toplevel.fa.gz) was indexed with exon information from the corresponding annotation file (http://ftp.ensemblgenomes.org/pub/plants/release-52/gff3/zea_mays/Zea_mays.Zm-B73-REFERENCE-NAM-5.0.52.gff3.gz). The trimmed reads were aligned to the indexed reference genome using Hisat2 (version 2.2.1)[34] with the appropriate input file settings and intron lengths: -q –phred 33 –rna-strandedness RF --min-intronlen 20 –max-intronlen 60000. The data was then saved in BAM format using the samtools view command from htslib (version 1.14)[35]. Picard tools (version 2.27.1; http://broadinstitute.github.io/picard/) was used to remove duplicates using MarkDuplicates.

Reads aligned to exons of genes were counted using htseq-count (version 2.0.1), with specifications to only count uniquely mapped reads[36]. Samples with less than 5 million counted reads were excluded.

Preparation of alignments for SNP calling

For SNP calling between the genotypes of this study and the B73 reference genome, the read alignments were processed using the HaplotypeCaller of GATK (version 4.2.6.1) with respect to GATK's best practices for RNA-seq data (https://gatk.broadinstitute.org/hc/en-us/articles/360035531192-RNAseq-short-variant-discovery-SNPs-Indels-, checked on 12/14/2023). First, Picard's AddOrReplaceReadGroups (version 2.27.1) was used to add readgroup information to the alignments of all samples. The replicate number was set as RGLB, the RGPL field was set to 'ILLUMINA', the RGPU field was set to 'unknown', and the RGSM field was filled with the sample name. The samtools view (version 1.14) command was used to filter for uniquely mapped reads by only including reads with mapping quality of 60 or higher and to format and index the alignments. Second, GATK's SplitNCigarReads was used to split alignments at positions with N in the CIGAR field, such as intron-spanning alignments[37].

SNP calling between B73 and Mo17 samples and B73v5 reference for sample evaluation

In brief, the GATK HaplotypeCaller was used to identify variants between the Mo17 samples and the B73v5 reference genome. The frequency of the B73 and Mo17 alleles at each SNP locus was previously identified in a similar manner[38,39]. The ratio of homozygous loci was calculated and samples with less than 95% homozygosity across expectedly homozygous loci were excluded (details in Supplementary Material SM2). A total of 175 RNA samples were excluded because they did not meet the criteria of being homozygous in ≥95% of the supposedly homozygous loci. Additionally, 10 samples were excluded beforehand due to their library size being <5 million read counts. Moreover, 17 samples were excluded because only one out of three replicates was left for the respective genotype. Since downstream analyses

include the comparisons between both parents and their resulting hybrids, we had to further exclude 90 hybrid RNA-seq samples because all corresponding paternal IBM-RIL RNA-seq samples were excluded. In addition, 8 IBM-RIL RNA-seq samples were excluded because the corresponding hybrid samples were missing. Finally, 852 RNA-samples remained for subsequent SNP calling as described below (Data file S1).

SNP calling between all high-confidence samples and the B73v5 reference

In the second SNP calling, SNPs between each sample and the B75v5 reference were called. We included the previously identified variants between our Mo17 samples and the B73v5 reference, as well as variants from our B73 samples vs. the B73v5 reference. They were filtered based on several criteria. The mapping quality (MQ), variant site quality (QUAL), Fisher strand (FS) and allele depth (AD) of SNP alleles were used with different thresholds for InDels and SNPs with respect to GATKs' guide on hard-filtering short variants (https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants, checked on 12/18/2023). For the SNPs the filters QD >2, SOR <3, MQ >40, QUAL >30, FS <60 and FORMAT/AD[0:1] >5 were applied. For the indels, the filters QD >5, QUAL >30, FS <200 and FORMAT/AD[0:1] >5 were applied. The base qualities of each sample were then recalibrated using the filtered SNPs and indels as known-sites with GATK (version 4.2.6.1). BaseRecalibrator was run to generate recalibration tables, which were then applied to the aligned reads with ApplyBQSR (https://gatk.broadinstitute.org/hc/en-us/articles/360035531192-RNAseq-short-variant-discovery-SNPs-Indels-, checked on 12/14/2023). The HaplotypeCaller was run with the recalibrated samples in BP_RESOLUTION mode and reported the SNP sites at each individual position. The resulting variant files were then filtered for positions with a coverage (DP) of ≥1, to eliminate loci without any information. The variant files from Mo17 and B73 samples were combined by GenomicsDBImport. The samples of a triplet (IBM-RIL samples plus corresponding B73×IBM-RIL and Mo17×IBM-RIL hybrids) were combined with the Mo17 and B73 samples, resulting in one database per triplet. Genotyping of all samples within each database was performed using the GenotypeGVCFs function. Since the Mo17 and B73 samples are present in each database, we ensured that genotyping was performed on the loci differentiating B73 and Mo17 in each database[37]. The genotyping data was then filtered for SNPs with QD >2, SOR <3, MQ >40, QUAL >30 and FS <60 using bcftools (version 1.17). A list of high confidence SNPs was created in R using the results from the HaplotypeCaller of the Mo17 and B73 samples (Data file S8). For these loci, it was established that the genotyping results of the HaplotypeCaller (B73v5 reference allele vs. non-reference allele) correspond to the B73 and Mo17 alleles of the germplasm of this study (reference = B73, non-reference = Mo17): The HaplotypeCaller reports for each SNP locus of each sample the most likely genotype, which we term genotype-call in the following

and the corresponding genotype quality (GQ), a measure for the confidence of the genotype-calls. Only genotype-calls of bi-allelic loci with a GQ of ≥10 were considered. The loci were filtered to include only those where ≥90% but a minimum of three remaining genotype-calls in Mo17 samples are homozygous for the non-reference allele, and 90% but a minimum of three remaining genotype-calls in B73 samples are homozygous for the reference allele. Next to the high confidence B73 vs Mo17 SNPs, we identified SNPs which did not belong to B73 or Mo17 as IBM-RIL specific (homozygous or heterozygous and regardless of GQ). Loci genotyped with a GQ of <10 were filtered. Only loci that were either in the high confidence SNP list of B73 vs. Mo17 alleles or which were IBM-RIL-specific (for masking putative IBM-RIL specific regions) are considered further.

Classification of IBM-RIL genomic regions

The filtered SNP data were used to classify each IBM-RIL genome into B73 or Mo17 regions and to mask regions which were not B73 or Mo17. A distance-function was used to calculate the distance between the IBM-RIL specific loci. Loci with a distance of <2.5 Mbp were grouped together as a block. Blocks containing a minimum of 10 IBM-RIL-specific loci, with ≥5 of those being homozygous, were identified and masked as IBM-RIL-specific third origin regions. The start and end positions of these regions were recorded and loci within those regions were dropped. Next, a sliding window approach was used to eliminate singular loci that did not match their surrounding loci. A window of 15 loci was used, and ≥11 had to be homozygous for the Mo17 allele for the window to be considered a Mo17 window. For a B73 window, ≥12 out of 15 loci must have homozygous B73 alleles. The values for the windows were obtained by computing the minimum number of matching loci in a 15-loci window across B73 and Mo17 samples. Otherwise, the window was considered ambiguous[40]. Loci within an ambiguous window were dropped, as well as loci which were classified differently from their window. The previously mentioned distance-function was utilised to calculate the distance between the remaining loci. Loci that carried the same allele and which were less than 0.5 Mbp apart were grouped together as a block, and all blocks were retained. The start and end positions of these blocks were recorded as the Mo17 and B73 regions within each IBM-RIL (Data file S9). Two IBM-RILs which had more than 50% of their genomes consisting of IBM-RIL specific regions from a third parental origin were excluded along with their hybrids (Data file S1), leaving 834 samples for final analyses. The data set of each triplet reported by the HaplotypeCaller was filtered to only include loci within the B73 or Mo17 regions of the IBM-RILs and within exons of protein-coding genes. We checked for all protein-coding genes whether they were located in a B73 or Mo17 region or masked as neither a B73 or Mo17 region, or whether they were located in a genomic region without SNP information. This verification was performed for each IBM-RIL separately. Centromere locations of the 10 chromosomes were taken from the

genome assembly of MaizeGDB by selecting the “Knobs, centromeres and telomeres” information https://jbrowse.maizegdb.org[41]. The proportion of heterozygous to homozygous regions was calculated for each backcross hybrid by dividing the total lengths of classified heterozygous regions (B73 regions of the IBM-RIL for the Mo17×IBM-RIL and Mo17 regions of the IBM-RIL for B73×IBM-RIL) by the total lengths of all classified regions (not considering IBM-RIL specific masked regions and regions without SNP information).

Multidimensional Scaling (MDS)

To evaluate the quality and the structure of the RNA-seq samples in this study, a multidimensional scalig (MDS) plot was used. The active genes of the 834 filtered samples were compared by the plotMDS() Bioconductor package limma (version 3.50.3)[42] in R.

Analysis of expression complementation

The activity/expression status of each gene was determined as previously described[10] based on thresholding normalized read counts. In short, fitting a generalized additive model (R package mgcv)[43] using guanine-cytosine (GC) content and log-transformed gene length as explanatory variables to account for artifactual read count differences across genes[44] resulted in a predictive count for each gene. The inverse of the predictive count was used as a multiplicative gene-specific normalization factor. In addition, sample-wise scale factors using the trimmed mean of M-values (TMM) method were estimated to adjust for differences between library sizes[45]. Each raw read count was multiplied by the product of the corresponding gene-specific normalization factor and the TMM scale factor to obtain a normalized count. The average expression level of each gene was represented by the mean normalized count across all replicates of each genotype in our data set. After estimation of the density distribution, the 0.25 quantile of the non-zero average expression levels was set as the threshold for calling the activity status of each gene in each sample. Thus, a gene was called active if the average expression level across all replicates was greater than the threshold and otherwise inactive for each genotype. Genes active in only one parent but also the hybrid are designated single parent expression (SPE) or SPE complementation genes, as the expression of only one parent is complemented in the hybrid. We identified these, by comparing the activity of each gene in the hybrids and their corresponding parents. From the classification of the IBM-RIL regions, we can deduct the genotypes of the SPE genes in the hybrids. Based on these genotypes, we classified our SPE genes into those within heterozygous (B73/Mo17 in B73×IBM-RILs, Mo17/B73 in Mo17×IBM-RILs) or homozygous regions (B73/B73 in B73×IBM-RILs, Mo17/Mo17 in Mo17×IBM-RILs). We further distinguished the SPE by which parent was active and indicated the active parent in bold. So, for example a SPE gene in a heterozygous region of a B73×IBM-RIL, where the IBM-RIL parent is active, but the B73 is not, the pattern would

be B73/**Mo17** (Figure 2A). For a SPE gene in a homozygous region of a Mo17×IBM-RIL, where the Mo17 parent is active, but the IBM-RIL is not, the pattern would be **Mo17**/Mo17 (Figure 2B).

Proportion of heterotic variance explained by the number of SPE genes underlying mid-parent heterosis (MPH) of root phenotypes

To estimate the fraction of the heterotic variance explained by the number of genes displaying SPE patterns, we propose the parameter $\mathrm{p}_{Het} = 1 - \frac{\sigma^2_{Het}}{\sigma^2_G}$, where σ^2_{Het} defines the total genetic variance across the hybrid genotypes and σ^2_G is the genetic variance of the heterosis effect not associated with the number of SPE genes[46,47]

For each backcross population, the genetic variance of the mid-parent heterosis (MPH) effects σ^2_{Het} was estimated separately in a "full" regression model (2) based on an extension of the baseline model (1). For this purpose, we defined for each parental genotype (i.e. each the IBM-RIL and the two common parental inbred lines B73 and Mo17) covariates (x_{i1-x}). These covariates were initially all set to 0 for each observation. For observations on the parental genotypes, the corresponding covariate for that specific parent was set to 1. For the observation on the hybrids, the two covariates corresponding to its two parents were set to 0.5. Thus, collectively these x_{i1-x} covariates model the effect of the *per se* performance of the parents and the mid-parent values of the hybrids.

MPH was modelled by a regression on the number of SPE genes. For this purpose, the number of SPE genes was set to 0 for all parental genotypes. This was done to be able to include them in the overall model. However, the parental genotypes have no impact on the regression, because their effect is fully absorbed by the covariates for the parental genotype effects.

As the MPH effects of the hybrids were not expected to fall on the regression line, we allowed for deviations from the regression by adding a random effect for hybrids. This was implemented by fitting the random effect z*genotype, where z is a continuous dummy covariable with z=0 for the parental genotypes and z=1 for the hybrids. This dummy variable acts as a switch that turns the random effect off for parental genotypes and on for hybrids [48].

$$Y_{jklmnp(ih)} = \mu + \beta_1 x_{i1} + \cdots + \beta_q x_{iq} + b_j + \gamma_1 sa_h + \gamma_1 sb_h + \gamma_1 sc_h + \gamma_1 sd_h + z_i * g_i + p_k + s_l + t_m + r_n + e_{jklmnp(i)} \quad (2)$$

$Y_{jklmnp(ih)}$ represents the parental effect value contributing to MPH of the corresponding hybrid *h* of a specific trait of interest. μ represents the intercept, x_{iq} represents the parental covariables of parent *q* for genotype *i*, b_j represents the fixed effect for block *j*, sa_h , sb_h , sc_h and sd_h represent the covariables for the number of genes displaying pattern 1-4 (Figure 2A) or 5-8 (Figure 2B) for hybrid *h*, respectively. $z_i * g_i$ represents the random effect for hybrid *h*

(corresponding to genotype *i*), z_i is a dummy variable with $z_i = 0$ for parents and $z_i = 1$ for hybrids[48]. Variable p_k represents the random effect for batch *k* nested within block *j*, s_l represents the random effect for system *l* nested within batch *k* and block *j*, t_m represents the random effect for triplet *m* nested within system *l*, batch *k* and block *j*, r_n represents the random effect for row *n* nested within triplet *m*, system *l*, batch *k* and block *j*, and $e_{jklmnp(i)}$ represents the random error effect for plant *p* of genotype *i* in block *j*, batch *k*, system *l*, triplet *m* and row *n*.

The analysis was implemented in R (version 4.0.1) using the lme4 package (version 1.1-29). In contrast to the baseline model, the fixed effect for genotype was replaced by individually defined covariates of the parental genotypes and fixed effects for the number of SPE genes. To determine σ_G^2 a "null" model (3) excluding the fixed effects of the covariates accounting for the number of SPE genes, was fitted.

$$Y_{jklmnp(ih)} = \mu + \beta_1 x_{i1} + \cdots + \beta_q x_{iq} + b_j + z_i * g_i + p_k + s_l + t_m + r_n + e_{jklmnp(i)} \quad (3)$$

where the notations are the same as in model (2).

Expression Quantitative Trait Loci (eQTL) analysis

An eQTL analysis was performed with the R/qtl2 package (version 0.22)[49] to identify positions that were significantly associated with gene expression values based on the masked and filtered SNP data. For each of the three cross-types (IBM-RIL, B73×RILs, Mo17×RILs) the classified and filtered SNP loci within B73 or Mo17 regions in the IBM-RILs were taken as marker input data. The positions of these SNP loci were used as preliminary genomic positions, as well as physical positions. The estimated expression means obtained from the model coefficients within the differential expression analysis of each genotype and gene were used as the phenotype data input in R/qtl2. As additional specifications to write the control file, the cross type was set to "rilself" for RILs by selfing, the alleles were set to "A" and "B" for B73 and Mo17 and the genotype codes were set to A/A=1 and B/B=2 to specify the transformation of homozygous alleles into numeric values (https://github.com/agroot-ibed/r-qtl2-analysis, updated on 12/19/2023, checked on 12/19/2023). Samples with more than 19% missing genotypes were dropped, as well as duplicated genotypes and markers with more than 60% missing genotype information. The genetic map was estimated from the physical positions and genotype information by the est_map() function with parameters maxit = 2000, error_prob = 0.001 and tol = 0.0001. The reduce_markers() function was used to retain only markers that were ≥1 centiMorgan (cM) apart to avoid retaining an excess of redundant markers. Pseudomarkers were inserted at a distance of 1 cM to the existing markers. A hidden Markov model calculated the genotype probabilities at all positions, with error_prob = 0.001. This was

followed by a genome scan, which was done by a Haley-Knott regression[50] to establish the association between genotype and expression phenotype with a linear model. In simple words, within each eQTL analysis, each marker is tested to see whether there is an association with the expression of single genes, the result is an LOD curve. In order to find out whether the highest LOD value is significant, a permutation was carried out and all significant peaks were saved. In more detail, to calculate the adjusted *p*-values for the resulting logarithm of odds (LOD) scores for a single gene, 10 000 permutations were done, reshuffling the expression data randomly and recording the maximum LOD score of each permutation. Selecting a significance threshold of $\alpha \leq 0.001$, we used the 99.9^{th} percentile of the ordered LOD maxima as the threshold to detect a significant eQTL for the gene[51]. The genomic map was converted to a physical map with the interp_map() function. By selecting the respective threshold, the physical position, confidence interval and LOD of significant peaks was obtained by the find_peaks() function[49]. The exact adjusted *p*-values were determined by calculating the percentile of permutation maxima, higher than the respective LOD[51]. This process was repeated for all (37782) active genes. To subsequently also correct for the testing of multiple genes, the false discovery rate was used on the adjusted *p*-values of the LOD peaks of all genes with p.adjust() function setting method to “FDR” and n to the total number of genes plus the number of second and third significant peaks. Peaks with an FDR ≤ 0.001 were considered significant. Start and end position of genes were added from the annotation file. The same procedure was performed on all three cross-type data sets (IBM-RIL, B73×IBM-RIL, Mo17×IBM-RIL). The resulting eQTL peaks were combined and distinct eQTLs were selected: in cases where multiple eQTLs were identified for a gene, we assessed whether the different peak positions corresponded to different regulatory elements. If eQTLs for the same gene were ≥25 Mbp apart or on different chromosomes and their positions did not lie within the confidence intervals of each other, they were considered to be different from each other and were retained. If multiple eQTLs for the same gene did not differ by the specified standards but were in close proximity to each other, only the eQTL with the shortest confidence interval or the highest LOD in case of equal confidence intervals was retained in the merged list. The eQTLs were classified into *cis* and *trans* eQTLs based on their distance from the start of their respective gene. *Trans*-regulating eQTLs were defined as located at a distance of at least 2.5 Mbp from the start of the gene and where their confidence interval did not include the start of the gene. *Cis*-regulating eQTLs were defined as located in proximity to the start of the gene (<2.5 Mbp) or located such that their confidence interval includes the start of the gene (Data file S4).

Transcriptome Wide Association Analysis (TWAS)

A TWAS was conducted to associate gene expression levels of active genes with phenotypic traits. The active genes were filtered to select those genes which were active in at least 5% of

all genotypes (14). We used the MLMM[52], BLINK[53] and FarmCPU[54] models, implemented in the R package GAPIT (version 3)[55], including the three first principal components for the initial identification of genes and in case of the MLMM the variance-covariance matrix between individuals as kinship. Each population (IBM-RILs, B73×IBM-RILs, Mo17×IBM-RILs) was analyzed separately. For use in GAPIT, expression values were rescaled to values between 0 and 2 for each population. The presented TWAS + SPE candidate genes (TSG) were additionally investigated using Student's t-test.

Determination of syntenic and non-syntenic genes

The syntenic and non-syntenic genes were determined by comparison against a published list of syntenic grass genes[56]. Those genes with *cis*-regulatory eQTL were compared against those with *trans*-regulatory eQTL using the Fishers' exact test in R with fisher.test().

Acknowledgements

We would like to thank KWS for the propagation of the IBM-RIL syn. 4 seeds. We thank Helmut Rehkopf (University of Bonn) for his support in propagating the genetic material for this study and the experimental station "Auf dem Hügel" of the University of Bonn. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) Next Generation Sequencing Competence Network (NGS-CN; project 423957469) grant HO 2249/18-1 to F.H. NGS analyses were carried out at the West German Genome Center (WGGC) site in Bonn. The authors acknowledge access to the bonna cluster hosted by the University of Bonn along with the support provided by its High-Performance Computing & Analytics Lab.

Author contributions

M. P. and J.A.B. carried out the experiments, conducted the statistical analyses, interpreted the data, and drafted the manuscript. A.S.M, G.L, H.S. gave advice regarding population genetics, TWAS analysis and bioinformatics handling of the data. H.-P.P. provided help with the experimental design for the RNA-seq experiment and helped with the statistical analyses. F.H. conceived and coordinated the study and participated in data interpretation and drafting the manuscript. All authors edited the manuscript and approved the final draft.

Competing interests

The authors declare no competing interests.

Data availability

Sequence data used in this study have been deposited in NCBI Bioproject with the ID PRJNA923128.

References

1. **Hochholdinger F & Yu P** (2025). Molecular concepts to explain heterosis in crops. *Trends in Plant Science* **30**; 10.1016/j.tplants.2024.07.018.
2. **Shull GH** (1914). Duplicate genes for capsule-form in *Bursa bursa-pastoris*. *Zeitschrift für Induktive Abstammungs- und Vererbungslehre* **12,** 97–149; 10.1007/bf01837282.
3. **Hochholdinger F & Baldauf JA** (2018). Heterosis in plants. *Current Biology* **28,** R1089-R1092; 10.1016/j.cub.2018.06.041.
4. **Paril J, Reif J, Fournier-Level A, Pourkheirandish M** (2024). Heterosis in crop improvement. *The Plant Journal* **117,** 23–32; 10.1111/tpj.16488.
5. **Hochholdinger F** (2016). Untapping root system architecture for crop improvement. *Journal of Experimental Botany* **67,** 4431–4433; 10.1093/jxb/erw262.
6. **Hoecker N, Keller B, Piepho H-P, Hochholdinger F** (2006). Manifestation of heterosis during early maize (*Zea mays* L.) root development. *Theoretical and Applied Genetics* **112,** 421–429; 10.1007/s00122-005-0139-4.
7. **Jones DF** (1917). Dominance of linked factors as a means of accounting for heterosis. *Proceedings of the National Academy of Sciences of the United States of America* **3,** 310–312; 10.1073/pnas.3.4.310.
8. **Paschold A, Jia Y, Marcon C, Lund S, Larson NB, Yeh C-T, Ossowski S, Lanz C, Nettleton D, Schnable PS** *et al.* (2012). Complementation contributes to transcriptome complexity in maize (*Zea mays L.*) hybrids relative to their inbred parents. *Genome Research* **22,** 2445–2454; 10.1101/gr.138461.112.
9. **Paschold A, Larson NB, Marcon C, Schnable JC, Yeh C-T, Lanz C, Nettleton D, Piepho H-P, Schnable PS, Hochholdinger F** (2014). Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. *The Plant Cell* **26,** 3939–3948; 10.1105/tpc.114.130948.
10. **Baldauf JA, Liu M, Vedder L, Yu P, Piepho H-P, Schoof H, Nettleton D, Hochholdinger F** (2022). Single-parent expression complementation contributes to phenotypic heterosis in maize hybrids. *Plant Physiology* **189,** 1625–1638; 10.1093/plphys/kiac180.
11. **Baldauf JA, Marcon C, Lithio A, Vedder L, Altrogge L, Piepho H-P, Schoof H, Nettleton D, Hochholdinger F** (2018). Single-parent expression is a general mechanism driving extensive complementation of non-syntenic genes in maize hybrids. *Current Biology* **28,** 431-437.e4; 10.1016/j.cub.2017.12.027.
12. **Li Z, Zhou P, Della Coletta R, Zhang T, Brohammer AB, H O'Connor C, Vaillancourt B, Lipzen A, Daum C, Barry K** *et al.* (2021). Single-parent expression drives dynamic gene expression complementation in maize hybrids. *The Plant Journal* **105,** 93–107; 10.1111/tpj.15042.
13. **Botet R & Keurentjes JJB** (2020). The role of transcriptional regulation in hybrid vigor. *Frontiers in Plant Science* **11,** 410; 10.3389/fpls.2020.00410.
14. **Swanson-Wagner RA, DeCook R, Jia Y, Bancroft T, Ji T, Zhao X, Nettleton D, Schnable PS** (2009). Paternal dominance of *trans*-eQTL influences gene expression patterns in maize hybrids. *Science* **326,** 1118–1120; 10.1126/science.1178294.
15. **Liu SC, Kowalski SP, Lan TH, Feldmann KA, Paterson AH** (1996). Genome-wide high-resolution mapping by recurrent intermating using *Arabidopsis thaliana* as a model. *Genetics* **142,** 247–258; 10.1093/genetics/142.1.247.
16. **Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A** (2002). Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. *Plant Molecular Biology* **48,** 453–461; 10.1023/a:1014893521186.

17. **Rahman H, Pekic S, Lazic-Jancic V, Quarrie SA, Shah SMA, Pervez A, Shah MM** (2011). Molecular mapping of quantitative trait loci for drought tolerance in maize plants. *Genetics and Molecular Research* **10,** 889–901; 10.4238/vol10-2gmr1139.

18. **Pan Q, Xu Y, Li K, Peng Y, Zhan W, Li W, Li L, Yan J** (2017). The genetic basis of plant architecture in 10 maize recombinant inbred line populations. *Plant Physiology* **175,** 858–873; 10.1104/pp.17.00709.

19. **Huo X, Wang J, Zhang L** (2023). Combined QTL mapping on bi-parental immortalized heterozygous populations to detect the genetic architecture on heterosis. *Frontiers in Plant Science* **14,** 1157778; 10.3389/fpls.2023.1157778.

20. **Yang H, Zhang Z, Zhang N, Li T, Wang J, Zhang Q, Xue J, Zhu W, Xu S** (2024). QTL mapping for plant height and ear height using bi-parental immortalized heterozygous populations in maize. *Frontiers in Plant Science* **15,** 1371394; 10.3389/fpls.2024.1371394.

21. **Göring HH, Terwilliger JD, Blangero J** (2001). Large upward bias in estimation of locus-specific effects from genomewide scans. *American Journal of Human Genetics* **69,** 1357–1369; 10.1086/324471.

22. **Schnable JC & Lyons E** (2011). Comparative genomics with maize and other grasses: from genes to genomes! *Maydica* **56**.

23. **Marcon C, Paschold A, Malik WA, Lithio A, Baldauf JA, Altrogge L, Opitz N, Lanz C, Schoof H, Nettleton D** *et al.* (2017). Stability of single-parent gene expression complementation in maize hybrids upon water deficit stress. *Plant Physiology* **173,** 1247–1257; 10.1104/pp.16.01045.

24. **Springer NM & Stupar RM** (2007). Allelic variation and heterosis in maize: how do two halves make more than a whole? *Genome Research* **17,** 264–275; 10.1101/gr.5347007.

25. **Wang B, Hou M, Shi J, Ku L, Song W, Li C, Ning Q, Li X, Li C, Zhao B** *et al.* (2023). De novo genome assembly and analyses of 12 founder inbred lines provide insights into maize heterosis. *Nature Genetics* **55,** 312–323; 10.1038/s41588-022-01283-w.

26. **Salvi S** (2017). An evo-devo perspective on root genetic variation in cereals. *Journal of Experimental Botany* **68,** 351–354; 10.1093/jxb/erw505.

27. **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 Phytologist* **207,** 1123–1133; 10.1111/nph.13420.

28. **Tai H, Opitz N, Lithio A, Lu X, Nettleton D, Hochholdinger F** (2017). Non-syntenic genes drive RTCS-dependent regulation of the embryo transcriptome during formation of seminal root primordia in maize (*Zea mays* L.). *Journal of Experimental Botany* **68,** 403–414; 10.1093/jxb/erw422.

29. **Yu P, Li C, Li M, He X, Wang D, Li H, Marcon C, Li Y, Perez-Limón S, Chen X** *et al.* (2024). Seedling root system adaptation to water availability during maize domestication and global expansion. *Nature Genetics* **56,** 1245–1256; 10.1038/s41588-024-01761-3.

30. **Piepho H-P, Büchse A, Truberg B** (2006). On the use of multiple lattice designs and α -designs in plant breeding trials. *Plant Breeding* **125,** 523–528; 10.1111/j.1439-0523.2006.01267.x.

31. **Smith AG, Han E, Petersen J, Olsen NAF, Giese C, Athmann M, Dresbøll DB, Thorup-Kristensen K** (2022). RootPainter: deep learning segmentation of biological images with corrective annotation. *New Phytologist* **236,** 774–791; 10.1111/nph.18387.

32. **Seethepalli A, Dhakal K, Griffiths M, Guo H, Freschet GT, York LM** (2021). RhizoVision Explorer: open-source software for root image analysis and measurement standardization. *AoB Plants* **13,** plab056; 10.1093/aobpla/plab056.

33. **Bolger AM, Lohse M, Usadel B** (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120; 10.1093/bioinformatics/btu170.

34. **Kim D, Langmead B, Salzberg SL** (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **12,** 357–360; 10.1038/nmeth.3317.

35. **Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM** *et al.* (2021). Twelve years of SAMtools and BCFtools. *GigaScience* **10**; 10.1093/gigascience/giab008.

36. **Anders S, Pyl PT, Huber W** (2015). HTSeq—a python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169; 10.1093/bioinformatics/btu638.

37. **van der Auwera G & O'Connor BD** (2020). *Genomics in the cloud: using Docker, GATK, and WDL in Terra.* 1st ed. Beijing: O'Reilly Media, Incorporated.

38. **Vedder L** (2024). MaizeSNP. DOI:10.5281/ZENODO.10684044.

39. **Baldauf JA, Vedder L, Schoof H, Hochholdinger F** (2020). Robust non-syntenic gene expression patterns in diverse maize hybrids during root development. *Journal of Experimental Botany* **71,** 865–876; 10.1093/jxb/erz452.

40. **Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T** *et al.* (2009). High-throughput genotyping by whole-genome resequencing. *Genome Research* **19,** 1068–1076; 10.1101/gr.089516.108.

41. **Hufford MB, Seetharam AS, Woodhouse MR, Chougule KM, Ou S, Liu J, Ricci WA, Guo T, Olson A, Qiu Y** *et al.* (2021). De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373,** 655–662; 10.1126/science.abg5289.

42. **Ritchie ME, Phipson B, Di Wu, Hu Y, Law CW, Shi W, Smyth GK** (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43,** e47; 10.1093/nar/gkv007.

43. **Wood SN** (2017). *Generalized additive models. An introduction with R.* 2nd ed. Boca Raton: CRC Press/Taylor & Francis Group.

44. **Lithio A & Nettleton D** (2015). Hierarchical modeling and differential expression analysis for RNA-seq experiments with inbred and hybrid genotypes. *Journal of Agricultural, Biological, and Environmental Statistics* **20,** 598–613; 10.1007/s13253-015-0232-3.

45. **Robinson MD & Oshlack A** (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11,** R25; 10.1186/gb-2010-11-3-r25.

46. **Feldmann MJ, Piepho H-P, Bridges WC, Knapp SJ** (2021). Average semivariance yields accurate estimates of the fraction of marker-associated genetic variance and heritability in complex trait analyses. *PLOS Genetics* **17,** e1009762; 10.1371/journal.pgen.1009762.

47. **Piepho H-P** (2019). A coefficient of determination (R2) for generalized linear mixed models. *Biometrical Journal* **61,** 860–872; 10.1002/bimj.201800270.

48. **Piepho H-P, Williams ER, Fleck M** (2006). A note on the analysis of designed experiments with complex treatment structure. *HortScience* **41,** 446–452; 10.21273/HORTSCI.41.2.446.

49. **Broman KW, Gatti DM, Simecek P, Furlotte NA, Prins P, Sen Ś, Yandell BS, Churchill GA** (2019). R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multiparent populations. *Genetics* **211,** 495–502; 10.1534/genetics.118.301595.

50. **Haley CS & Knott SA** (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69,** 315–324; 10.1038/hdy.1992.131.

51. **Lystig TC** (2003). Adjusted p values for genome-wide scans. *Genetics* **164,** 1683–1687; 10.1093/genetics/164.4.1683.

52. **Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M** (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics* **44,** 825–830; 10.1038/ng.2314.

53. **Huang M, Liu X, Zhou Y, Summers RM, Zhang Z** (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* **8**; 10.1093/gigascience/giy154.

54. **Liu X, Huang M, Fan B, Buckler ES, Zhang Z** (2016). Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLOS Genetics* **12,** e1005767; 10.1371/journal.pgen.1005767.

55. **Wang J & Zhang Z** (2021). GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genomics, Proteomics & Bioinformatics* **19,** 629–640; 10.1016/j.gpb.2021.08.005.

56. **Zhang Y, Ngu DW, Carvalho D, Liang Z, Qiu Y, Roston RL, Schnable JC** (2017). Differentially regulated orthologs in sorghum and the subgenomes of maize. *The Plant Cell* **29,** 1938–1951; 10.1105/tpc.17.00354.

4. Non-additive gene expression contributing to heterosis in partially heterozygous maize hybrids is predominantly regulated from heterozygous regions

Authors

Marion Pitz[1#], Jutta A. Baldauf[1#], Hans-Peter Piepho[2], Frank Hochholdinger[1]*

Affiliations

[1]Institute of Crop Science and Resource Conservation, Crop Functional Genomics, University of Bonn, 53113 Bonn, Germany

[2]Institute of Crop Science, Biostatistics Unit, University of Hohenheim, 70599 Stuttgart, Germany

# Contributed equally to this work

* Corresponding author: Frank Hochholdinger

E-Mail: hochholdinger@uni-bonn.de

This manuscript is under review in *New Pyhtologist.*

Abstract

Hybrid plants often perform better than their homozygous parents, a phenomenon that is commonly referred to as heterosis. Heterosis is widely utilized in modern agriculture, although its molecular basis is not very well understood. In this study, we backcrossed an intermated recombinant inbred line population of maize with its parental inbred lines B73 and Mo17. The resulting hybrids exhibited different degrees of heterozygosity and heterosis. Non-additive gene expression was consistently higher than mid-parent expression across these partially heterozygous backcross hybrids. We surveyed the regulation of these non-additively expressed genes in the hybrids by investigating expression quantitative trait loci (eQTL). We demonstrated that non-additively expressed genes explain up to 27% of heterotic variance. Furthermore, our data demonstrated that non-additively expressed genes are regulated almost exclusively from heterozygous regions of the genome. Moreover, we observed that non-additive expression patterns are distinctly regulated depending on the genetic origin of the higher expressed parent. As a consequence, these regulatory regimes lead to higher gene activity in most non-additively expressed genes in the hybrids and might contribute to the hybrid performance. The results of this study reconcile the dominance and overdominance model of heterosis by demonstrating that expression of many non-additively expressed genes is dominant while their regulation from heterozygous regions is consistent with the overdominance model of heterosis.

Introduction

The term heterosis describes the observation that hybrid progeny of genetically distinct parents display superior agricultural performance[1]. The introduction of hybrids in maize breeding in the 1930s is considered as one of the landmark innovations of modern agriculture and has contributed to an enormous increase in yield[2–4]. It has been observed that the phylogenetic distance between the parental inbred lines is positively associated with heterosis[5]. The observation that specific combinations of parents result in especially high levels of heterosis has resulted in the definition of typical female and male heterotic groups[6]. Other crops, such as rice, also benefit from the classification of genotypes into heterotic groups and their combination as heterotic pattern[7,8].

Heterosis is observed in all parts of the plant throughout development, but is typically investigated for above-ground traits related to yield[9]. In maize roots, which play an important role for the overall performance of plants, heterosis becomes apparent 5 to 7 days after germination[10].

Classical genetic concepts to explain heterosis include the dominance and overdominance model. The dominance model postulates that heterosis is caused by complementation of slightly deleterious alleles at many loci in the hybrid by dominant or at least stronger alleles[11]. The overdominance model postulates that two different alleles at the same locus cause heterosis by their interaction and that the heterozygous state itself is advantageous to the homozygous situation of the parents[5]. Despite examples of single genes displaying overdominance[12,13], none of these models alone can fully explain heterosis[4,14,15].

Genes with differential expression between two maize lines can show a variety of expression levels in the resulting hybrid. They can display additive expression, reflecting the average expression of their parents, or deviate from this pattern and display non-additive expression[16]. Depending on the surveyed tissues, developmental stages and genotypes, maize displays a highly variable degree of non-additive gene expression[17–20]. Reciprocal maize hybrids of B73 and Mo17 generally share the same non-additive pattern[21,22]. In early primary roots and developing ear shoots of the same hybrids, a trend towards adoption of high parent expression, rather than low parent expression was observed[19,21,23], but not further investigated in detail. Both additive and non-additive expression have been considered to contribute to heterosis[17,21,22,24,25]. The observation that non-additive genes are conserved under stress conditions and mostly belong to evolutionary less conserved non-syntenic genes suggests that they are involved in adaptation to different environments or stress conditions[21,26].

Gene expression differences are the result of alterations in gene regulation. Regulatory elements can be classified as *cis*, if they are positioned close to the regulated gene, and *trans*,

if the element is located at a different position, often on a different chromosome[27]. A possible connection between transcriptional variation in the regulation of *cis*- and *trans*-acting factors and hybrid performance was discussed[28] and an association of *trans*-regulated gene expression in hybrids with paternal alleles was shown in maize[29].

Recombinant inbred line (RIL) populations as well as backcross populations have been extensively used in genetics for quantitative trait locus (QTL) mapping, candidate gene identification and heterosis studies[30–33]. They can further be used to identify expression QTL (eQTL). These are genomic regions associated with variation in gene expression across the mapping population and provide direct insights into the regulation of gene expression[27].

In this study, we analyzed the transcriptomes of the maize intermated B73 and Mo17 (IBM) RIL syn. 4 population[34] and their partially homozygous and heterozygous backcross hybrid populations with the original parents B73 and Mo17 (Figure S1). We demonstrated that non-additive gene expression patterns influence the manifestation of heterosis in seedling root development. We further showed that regulatory elements of non-additive genes are predominantly located in heterozygous regions, suggesting that heterozygosity on the regulatory level promotes a higher expression in the hybrid than the parental average. Depending on their parental genetic origin, these regulatory elements act either predominantly in *cis* or *trans*, possibly influencing the formation of heterotic patterns.

Results

Transcriptome analysis of two maize intermated recombinant inbred line backcross populations

To study the regulation of non-additive gene expression in maize hybrids, i. e. patterns that significantly deviate from the average of the parental values (mid parent value; MPV), we generated two partially heterozygous backcross populations by crossing 112 IBM-RIL syn. 4 recombinant inbred lines to the maternal IBM-RIL parents B73 and Mo17 (Figure S1A). The backcross populations, obtained by crossing the IBM-RILs to their original parents, vary in their heterozygosity as well as heterosis (Figure S1A).

We subjected 1-week-old primary roots of these backcross hybrids, their parents (IBM-RILs and B73 or Mo17) and the fully heterozygous hybrids B73xMo17 and Mo17xB73 to RNA-sequencing and root phenotyping. After quality assessment, 2-3 biological replicates of 85 B73xIBM-RIL and 82 Mo17xIBM-RIL backcross hybrids remained for downstream analyses. To obtain higher accuracy in the pairwise comparisons, we included more replicates of the fully heterozygous reciprocal hybrids B73xMo17 (23 biological replicates) and Mo17xB73 (24 biological replicates) and the parental inbred lines B73 (47 biological replicates) and Mo17 (42 biological replicates) in our analyses (Figure S1B).

Most non-additively expressed genes are expressed above the mid-parent value in hybrids

Non-additively expressed genes in hybrids are expressed significantly higher or lower than the mid-parent value (MPV). To study non-additive gene expression, we determined genes with significantly different expression in the hybrid compared to the MPV (FDR ≤0.05, $|Log_2FC| > 1$).

In the parent-hybrid triplets of the fully heterozygous reference hybrids, we investigated 24,241 (B73xMo17) and 24,203 (Mo17xB73) genes active in at least one genotype. Among those, 22,453 (93%; B73xMo17) and 22,621 (93%; Mo17xB73) were additively expressed, of which 83% (B73xMo17: 18,604) and 82% (Mo17xB73: 18,630) did not show any expression difference between the parents (Figure 1A). The remaining additively expressed genes adopted the MPV of their differentially expressed parents with either B73 or Mo17 being the high parent (Figure 1A). Among the 1,788 (B73xMo17) and 1,582 (Mo17xB73) non-additively expressed genes (Figure 1B), 93% (B73xMo17) and 97% (Mo17xB73) showed a higher expression level than the MPV (Figure 1B). Most of these genes (1,584 (95%) in B73xMo17 and 1,494 (93%) in Mo17xB73) showed significantly different expression between the parents (DEGs: FDR ≤0.05, $|Log_2FC| > 1$) (Figure 1B).

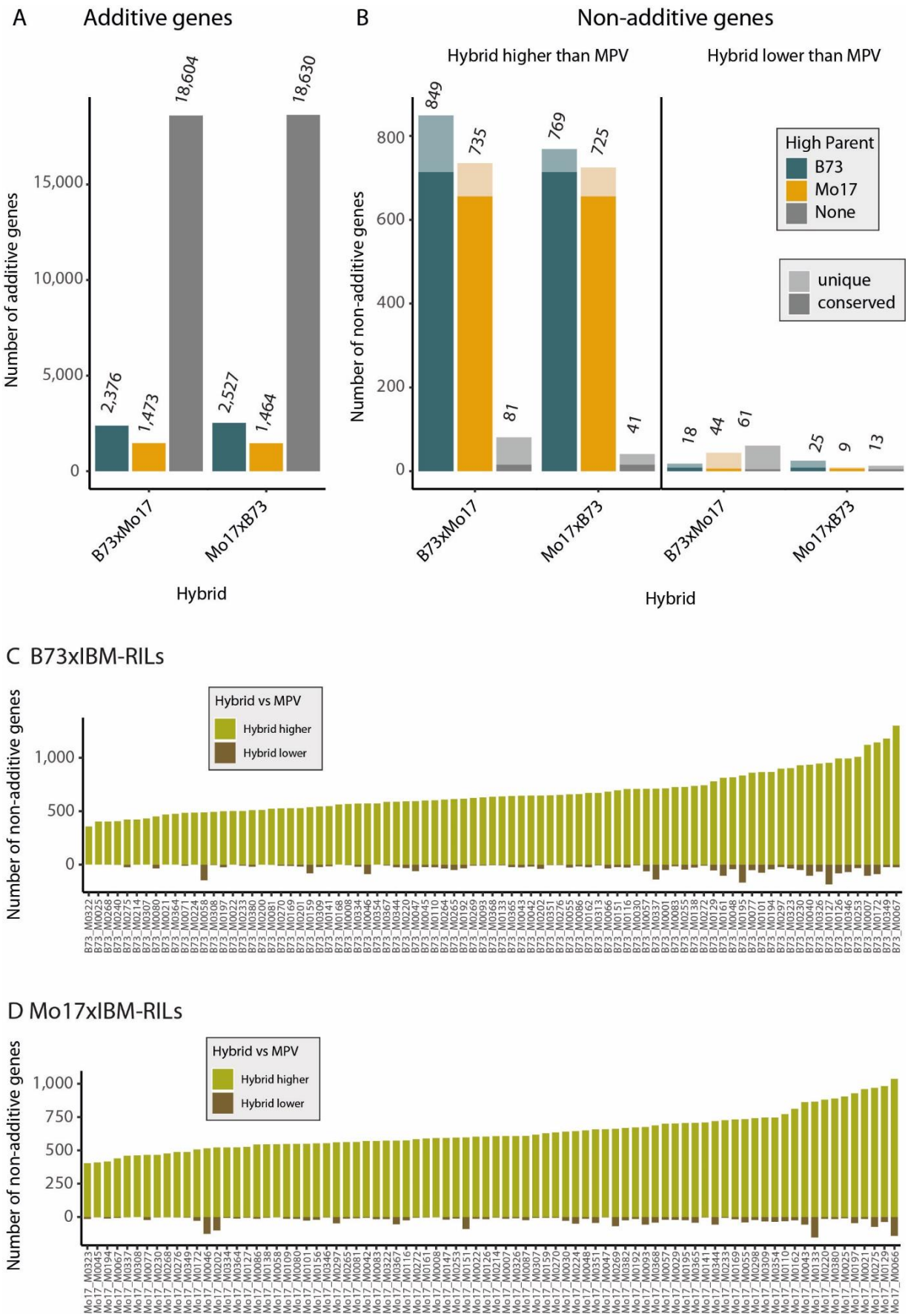


Figure 1: Distribution of gene expression pattern. **A** Numbers of additive genes (difference between (mid-parent value of gene expression) MPV and hybrid not significant, |Log2FC|>1, p<0.05) in B73xMo17 and Mo17xB73 hybrids. **B** Numbers of non-additive genes with either significantly higher than MPV expression or lower than MPV

expression in the hybrid (|Log2FC|>1, p<0.05). Colors in **A** and **B** indicate parental gene expression difference whith blue = higher expression in B73 than Mo17 (|Log2FC|>1, p<0.05), yellow = higher expression in Mo17 than B73 and grey = no significant difference. In **B** darker shade indicates conserved expression between the reciprocal hybrids. **C** Numbers of non-additive genes in B73xIBM-RIL and **D** Mo17xIBM-RIL hybrids. Different colors indicate, whether the hybrid expression value is significantly (|Log2FC|>,p<0.05) higher (grey, positive values) or lower (brown, negative scale) than the MPV.

B73 was the high parent in 54% (B73xMo17) and 51% (Mo17xB73) of genes with differentially expressed parents in the highly heterozygous reference hybrids (Figure 1B "Hybrid higher than MPV", Data file S1). Among those, 79% (714) of genes with B73 as the high parent and 82% (656) of genes with Mo17 as high parent were conserved between the reciprocal reference hybrids (Figure 1B, Data file S1). Nevertheless, expression of 94% (B73xMo17) and 98% (Mo17xB73) of these genes was within the range of their parents. The average numbers of non-additively expressed genes among the backcross-populations in general were 847 in B73xIBM-RILs and 807 in Mo17xIBM-RILs, which is approximately half the number (47% and 51%) of non-additive genes in the fully heterozygous B73xMo17 and Mo17xB73 hybrids. For further analyses, the active genes in the IBM-RIL backcross populations were classified into heterozygous and homozygous, based on SNPs present in these genes, or surrounding regions. Among those, we identified on average 19,042 additive genes in B73xIBM-RIL backcrosses and 18,985 in Mo17xIBM-RIL backcrosses (Data file S2). For most non-additively expressed genes, we observed a prevalence for higher expression compared to the MPV across all B73xIBM-RILs (Figure 1C) and Mo17xIBM-RILs (Figure1D). In B73xIBM-RILs, on average 668 (95%) genes were expressed higher than the MPV (Figure1C), while in Mo17xIBM-RILs on average 637 (96%) displayed this expression pattern (Figure1D). In contrast, on average only 35 (5%; B73xIBM-RIL) and 26 (4%; Mo17xIBM-RIL) of genes were expressed lower in the hybrid compared to the MPV. Hence, the trend of non-additively expressed genes to exceed the MPV as observed in the reference hybrids (Figure 1B), is also conserved in both IBM-RIL backcross populations (Figure 1C-D). We therefore focused our downstream analyses on non-additively expressed genes displaying above MPV expression in the hybrids with parents displaying contrasting expression.

Heterozygosity drives non-additive gene expression in backcross hybrids

In the B73xIBM-RIL and Mo17xIBM-RIL backcross hybrids, non-additively expressed genes are either located in homo- or heterozygous regions of the genome (Figure 2A and 2B). In B73xIBM-RIL hybrids, most non-additively expressed genes with higher than MPV expression (on average 487, 82%) were located in heterozygous genomic regions (B73/Mo17, Figure 2C, pattern 1 and 2), while less genes (on average 105, 18%) were observed in homozygous genomic regions (B73/B73, pattern 3 and 4) (Figure 2C). In heterozygous regions of B73xIBM-RIL backcross hybrids we observed more genes with B73 as high parent (66%; Figure 2C pattern 1 vs 2).

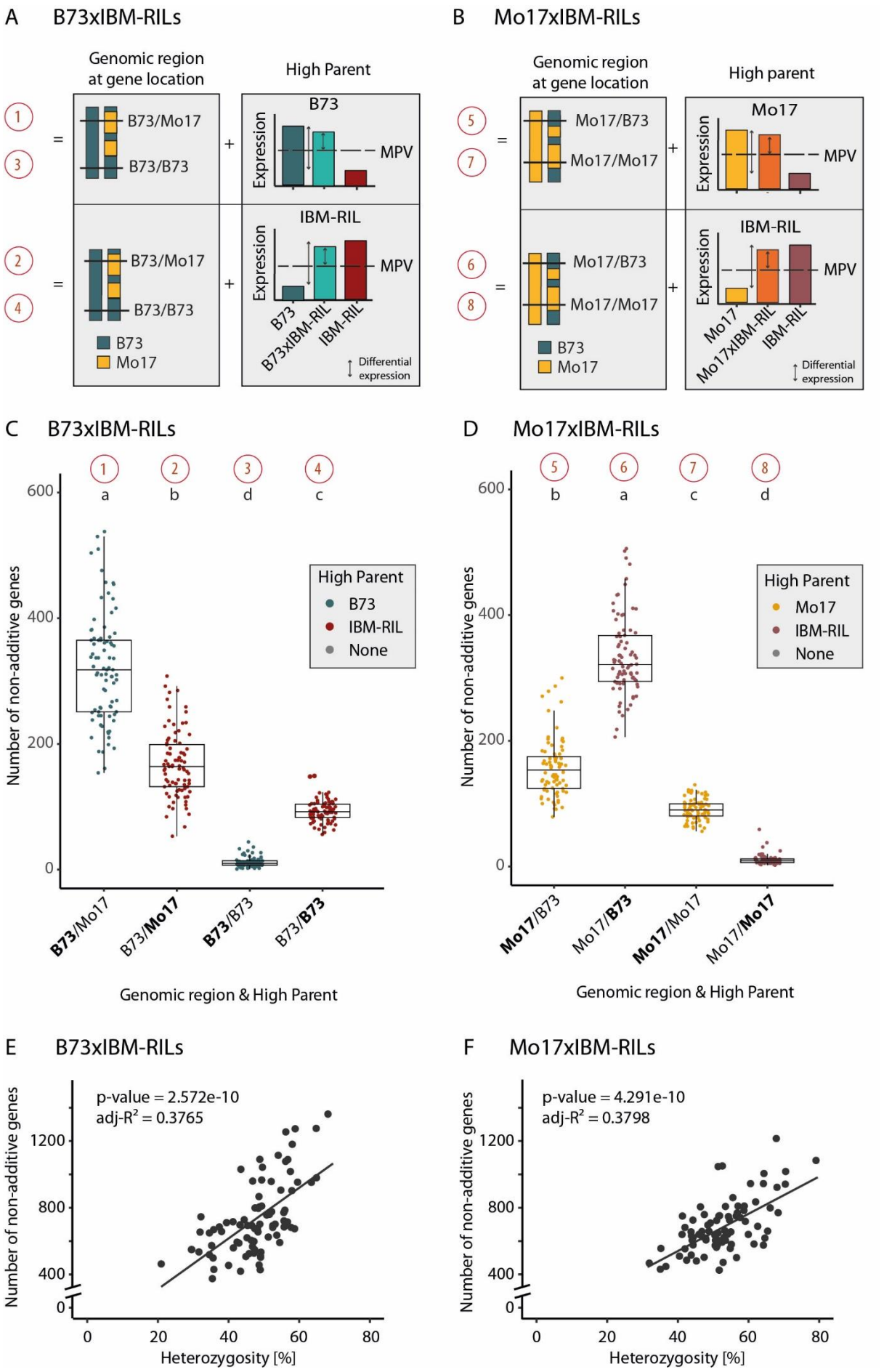

A B73xIBM-RILs
Genomic region at gene location
High Parent
B73/Mo17
B73/B73
B73
MPV
IBM-RIL
Expression
B73xIBM-RIL
Differential expression
B73
Mo17
B
Mo17xIBM-RILs
High parent
Mo17/B73
Mo17/Mo17
Mo17
Mo17xIBM-RIL
C B73xIBM-RILs
D Mo17xIBM-RILs
Number of non-additive genes
High Parent
None
Genomic region & High Parent
E B73xIBM-RILs
F Mo17xIBM-RILs
p-value = 2.572e-10
adj-R² = 0.3765
p-value = 4.291e-10
adj-R² = 0.3798
Heterozygosity [%]

Figure 2: Heterozygous genomic regions drive non-additive gene expression in backcross hybrids. **A** Schematic depiction of the genomic region of a gene in the hybrid and the possible expresison pattern with higher than MPV in B73xIBM-RILs and **B** Mo17xIBM-RILs. **C** Numbers of non-additive genes in B73xIBM-RIL hybrids differentiated iby expression pattern. **D** Numbers of non-additive genes in B73xIBM-RIL hybrids differentiated by expression pattern. Colors in C and C indicate the parent with higher expression. The high parent is also indicated on the x-axis in bold letters. The red numbers correspond to the graphical description in **A** and **B**. For **E** B73xIBM-RIL hybrids and **F** Mo17xIBM-RIL hybrids, the heterozygosity of each hybrid in % of all classified regions are displayed againts the corresponding number of non-additive genes. The p-value and adjusted R^2 are indicated for a linear regression with the heterozygosity as indepentent and the number of non-additive genes as dependent variable.

In homozygous B73/B73 regions of B73xIBM-RILs backcross hybrids, we almost exclusively observed non-additive genes with the paternal IBM-RILs as high parent (90%, Figure 2C, pattern 4 vs. 3).

Similarly, in Mo17xIBM-RILs, non-additively expressed genes with higher expression levels than the MPV were preferentially (on average 492; 83%) detected in heterozygous genomic regions (Mo17/B73, pattern 5 and 6), while less genes (101; 17%) were detected in homozygous genomic regions (Mo17/Mo17, pattern 7 and 8) (Figure 2D). Again, in heterozygous regions of Mo17xIBM-RIL backcross hybrids we observed more genes with B73 as high parent (69%; Figure 2D pattern 5 vs 6). In homozygous regions, we almost exclusively observed non-additive genes with the maternal high parent Mo17 in Mo17xIBM-RILs (90%; Figure 2C, pattern 7 vs. 8).

In summary, in both backcross hybrid populations, B73xIBM-RIL and Mo17xIBM-RIL, non-additive genes with higher than MPV expression were predominantly located in heterozygous regions. Furthermore, the degree of heterozygosity of the backcross hybrids was significantly positively associated with the number of non-additive genes (Figure 2E-F), suggesting that heterozygosity is a major driver of non-additive gene expression.

Up to 27% of heterotic variance in root traits can be explained by non-additive genes in Mo17xIBM-RILs

We further used a linear modelling approach to calculate the proportion of heterotic variance that can be attributed to the number of non-additively expressed genes in patterns 1-8 (Table 1). We determined that up to 27% of heterotic variance (variance in hybrid phenotypes that is not accounted for by experimental factors or parental values) can be attributed to non-additively expressed genes (Table 1). The different traits in the different populations show contrasting values for the heterotic variance attributable to non-additive genes. This indicates an important but variable role of non-additive genes for heterosis of different traits and populations (Table 1).

Table 1: Proportion of heterotic variance explained by the number of non-additive pattern 1-8 genes on mid-parent heterosis for different root phenotypes

Trait	B73xIBM-RILs			Mo17xIBM-RILs		
	σ^2_{Het}	σ^2_G	p_{Het}	σ^2_{Het}	σ^2_G	p_{Het}
No. Of root tips	0.277	0.332	17% (0.167)	0.155	0.194	20% (0.199)
Total root volume	2.498	2.705	8% (0.076)	0.818	1.123	27% (0.272)
Total root length	3.094	3.567	13% (0.132)	1.786	2.194	19% (0.186)
Lateral root density	5.407	6.804	21% (0.205)	7.366	6.925	-7% (-0.063)

σ^2_{Het} = unexplained genetic variance of heterosis effect, not associated with non-additive genes; σ^2_G = total genetic variance among the hybrid genotypes; p_{Het} = Coefficient of determination: Proportion of the heterotic variance explained by the number of non-additive genes.

Non-additive genes with Mo17 as high parent are predominantly regulated in *trans*

To study the regulation of non-additively expressed genes, we identified expression quantitative trait loci (eQTL), which are significantly associated with the expression of these genes and thus likely regulate their activity. For 93% of the non-additively expressed genes in B73xMo17 and for 96% of these genes in Mo17xB73 we identified at least one eQTL (Data file S2). We further distinguished the eQTL into *cis*-regulating in case of close proximity to the regulated gene (<2.5 Mbp) and *trans*-regulating in case of distal regulation (>2.5 Mbp, in most instances on a different chromosome). In both reference hybrids we observed that if B73 was the high parent, genes were primarily *cis*-regulated by eQTL (Figure 3A B73xMo17: 96%; Figure 3B: Mo17xB73: 96%) while when Mo17 was the high parent, genes were preferentially regulated by eQTL in *trans* in the reciprocal hybrids (Figure 3A: B73xMo17: 69%; Figure 3B: Mo17xB73: 72%).

For IBM-RIL backcross hybrids, we identified eQTL for 90% (633/699) of non-additively expressed genes with higher than MPV expression on average. In heterozygous regions of B73xIBM-RIL and Mo17xIBM-RIL hybrids, we observed a similar pattern as in the fully heterozygous B73xMo17 and Mo17xB73 hybrids: genes with the high parent contributing the B73 allele were almost exclusively controlled by *cis*-regulating eQTL (Figure 3C pattern 1: B73xIBM-RILs: 96%; Figure 3D pattern 6: Mo17xIBM-RILs: 97%), whereas genes with the Mo17 allele as high parent showed a similar number of *cis*- and *trans*-regulating eQTL (Figure 3C, pattern 2: B73xIBM-RILs: 49% *cis*; Figure 3D, pattern 5: Mo17xIBM-RILs: 49% *cis*). Among additive genes, on average 94% and 91% were *cis*-regulated (Figure 3C, D).

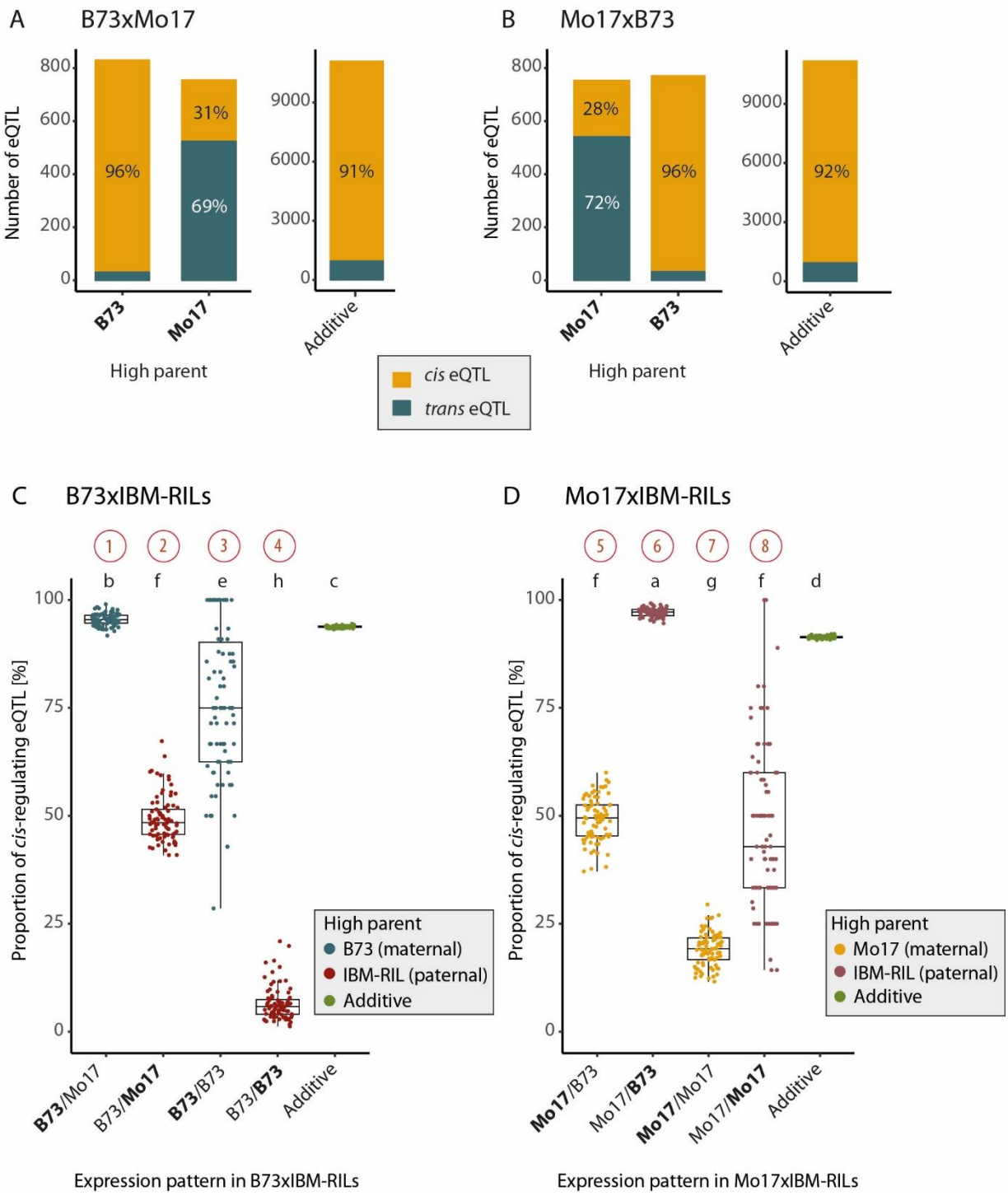


Figure 3: Proportion of *cis*- and *trans*- regulation in expression pattern of non-additive and additive genes. **A** Bars show the number of *cis*- and *trans*- regulating eQTL of the genes with different expression pattern in the B73xMo17 hybrid and **B** the Mo17xB73 hybrid. Percentages indicate the proportion of *cis*- or *trans*-regulation. Boxplots display the proportion of *cis*-regulation among non-additive expression pattern and additive genes in **C** B73xIBM-RIL and **D** Mo17xIBM-RIL hybrids. Different letters indicate significantly different proportions ($\alpha < 0.05$), identified with a gaussian mixed model with the hybrid as random effect, the non-additive expression pattern and additive genes as a fixed factor and a diagonal variance component for the fixed factors.

Heterozygous eQTL regulate heterozygous and homozygous non-additive genes

Next, we distinguished between heterozygous and homozygous eQTL in the backcross hybrids. Since *cis*-acting eQTL are located in close proximity to their target gene, we observed

that *cis*-acting eQTL generally have the same zygosity, i.e. homo- or heterozygote, as the non-additive gene they regulate (Figure 4A, B). Thus, most *cis*-acting eQTL regulating non-additive genes are located in heterozygous regions for B73xIBM-RILs (Figure 4A, Figure S2A) and Mo17xIBM-RILs (Figure 4B, Figure S2B) because most non-additively expressed genes are located in heterozygous genomic regions (Figure 2C, D). *Trans*-acting eQTL are also almost exclusively located in heterozygous regions, but they regulate heterozygous and homozygous non-additively expressed genes at a similar rate (Figure 4A pattern 2 and 4, Figure 4B, pattern 5 and 7). Thus, non-additive genes with the expression patterns 1-8 (Figure 2) are almost exclusively regulated in heterozygous regions (B73xIBM-RILs: 94%, Mo17xIBM-RILs: 95%). In homozygous regions in B73xIBM-RILs, nearly all non-additive genes displayed pattern 4 (Figure 2C) and eQTL for those genes were predominantly *trans*-regulating (94%; Figure 3A). These eQTL were located almost exclusively in heterozygous regions (Figure 4A, pattern 4). This indicates for pattern 4 that although these genes were homozygous for the B73 allele, they were regulated by eQTL carrying also the Mo17 allele at the heterozygous location of the eQTL (Figure 5, pattern 4). Thus, genes with the same allele in both parents display differential expression. In this case the IBM-RIL parent, providing the Mo17 allele at the eQTL position, was higher expressed. In contrast, genes in homozygous regions in Mo17xIBM-RILs had the maternal Mo17 as the high parent (pattern 7) but were also *trans*-regulated (81% *trans*-regulated; Figure 3B). In this pattern, the genes were located in homozygous Mo17/Mo17 regions, while the eQTL were located in heterozygous Mo17/B73 regions (Figure 4B, pattern 7) with the IBM-RIL providing the B73 allele at the eQTL. At the same time, the expression (as per definition of the pattern 7) was lower in the IBM-RIL than in the Mo17 parental line (Figure 5, pattern 7).

In summary, non-additive genes in heterozygous regions with the Mo17 allele provided by the high parent are more often *trans*-regulated, compared with those with B73 as the high parent. We further demonstrated that non-additive genes in homozygous regions are regulated by *trans*-eQTL from heterozygous regions, where the Mo17 allele at the eQTL is responsible for increased gene expression.

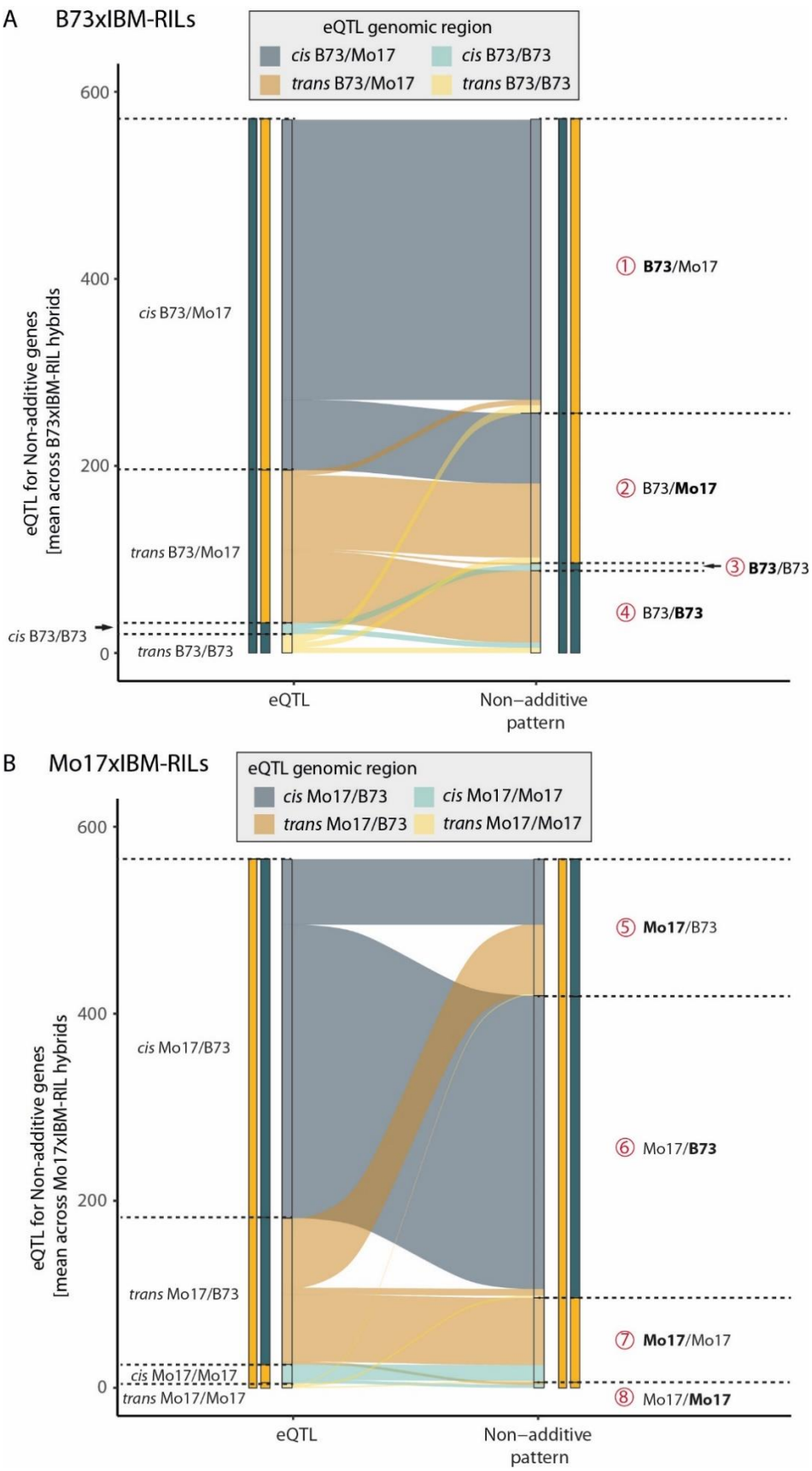


Figure 4: The distribution and location of eQTL of non-additive genes in homo- or heterozygous regions in **A** B73xIBM-RIL hybrids or **B** Mo17xIBM-RIL hybrids. Non-additive gene can be expressed higher (|Log2FC|>1, p<0.05) in the maternal parent or the paternal parent (indicated in bold). And corresponding eQTL can be *cis*-regulating from heterozygous (dark blue) or homozygous (light blue) regions or *trans*-regulating from heterozygous (dark yellow) or homozygous (light yellow) regions.

Discussion

The dominance model of heterosis suggests that deleterious alleles are complemented by beneficial alleles at many loci in the hybrid[11], while the overdominance model explains heterosis by beneficial interactions of alleles in heterozygous genes in the hybrid [35]. It was suggested that heterosis is controlled by several genetic mechanisms, with varying contribution depending on the species and trait under analysis[36]. Results of transcriptome studies supported the notion that specific gene expression patterns can contribute to heterosis, although no direct correlation between differential expression or non-additive expression and heterosis has been identified before[3]. Here we demonstrated, how non-additively expressed genes and their regulation contribute to heterosis. For this purpose, we analysed gene expression profiles of the fully heterozygous reference hybrids B73xMo17 and Mo17xB73 and populations of partially heterozygous B73xIBM-RIL and Mo17xIBMRIL hybrids.

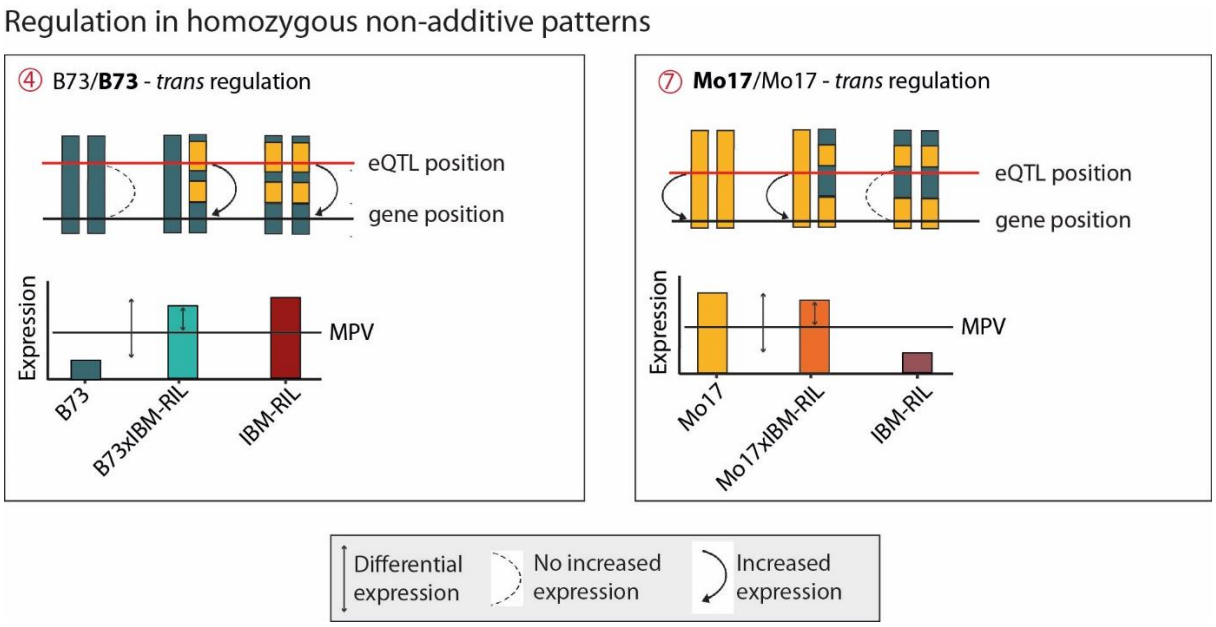


Figure 5: Schematic depiction of the *trans*-regulation of non-additive genes in homozygous regions and how expression is initiated by the Mo17 allele at the eQTL position in both cases.

In general, we observed that most genes in fully and partially heterozygous hybrids are additively expressed, i.e. their expression is not different from the mid-parent value. This is in line with previous studies where in general >90% of genes in maize hybrids were additively expressed[19,22]. This maintenance of the *status quo* for most genes was suggested to be beneficial for the hybrid [22] and is in line with the gene balance hypothesis[37], stating that quantitative traits are influenced by gene dosages of different alleles of different genes[36,38]. The number of non-additively expressed genes in the reference crosses in the present study are consistent with previous observations[39]. As expected, in the backcross hybrids we observed only approximately half the number of non-additively expressed genes compared to

the reference crosses because they are predominantly observed in heterozygous regions of the genome. We demonstrated that most non-additive genes were expressed higher than mid-parental expression. This pattern was universal across all backcross and reference hybrids. Previous studies showed a similar pattern, in which more non-additive genes adopted the high-parent expression in the same maize[19], cotton[40] or coffee[41] hybrids. Recently, a thermodynamic model of transcription factor binding in hybrids suggested this to be a general mechanism in case of parental expression divergence based on above-average occupancy of promoters[42]. This trend of expression complementation is consistent with the dominance model of heterosis[11]. Similar observations were made in Arabidopsis, where different pathways were complemented in a high-parent expression pattern and connected to hybrid adaptability across developmental periods[43].

We demonstrated that up to 27% of heterotic variance in phenotypic root traits can be explained by the number of non-additively expressed genes in Mo17xIBM-RILs (Table 1). Similarly, genes displaying single parent expression (SPE), a pattern where a gene is expressed in the hybrid but in only one of the parents, influenced heterosis significantly up to 29% in the backcross population B73xIBM-RILs[44]. Remarkably, while SPE genes contributed mainly in B73xIBM-RIL backcross hybrids to heterosis, non-additively expressed genes contributed substantially to heterosis in both Mo17xIBM-RIL and B73xIBM-RIL backcross hybrids, with the exception of lateral root density[44] (Figure 6). Thus, non-additive genes and SPE genes appear to contribute to heterosis in a genotype- and trait-specific manner, with variable contribution.

Using IBM-RIL backcross populations allows to study the regulation of non-additively expressed genes via eQTL analyses. Additionally, the partially hetero- and homozygous nature of our backcross populations revealed aspects of non-additive gene expression regulation that cannot be studied in fully heterozygous hybrids: In both IBM-RIL backcross hybrid populations, eQTL regulating non-additive gene expression were almost exclusively located in heterozygous regions (B73xIBM-RILs: 94%, Mo17xIBM-RILs: 95%), no matter which allele (B73/Mo17) was contributed by the higher expressed parent. This is in line with the observed numbers of non-additive genes. In the partially homozygous backcross populations, about half as many non-additive genes were identified, compared to the fully heterozygous hybrids. On average, only ~50% of the genome of the backcross hybrids are heterozygous and non-additive expression is regulated in those heterozygous regions. While single-parent expression complementation[45,46] and non-additive expression is consistent with the dominance model, these expression patterns are regulated by heterozygous eQTL interactions, hence suggesting overdominance[44]. These observations are consistent with the notion that the dominance and overdominance models are not mutually exclusive[36]. In our study dominance prevailed on the

level of gene expression, while overdominance was observed predominantly on the level of gene regulation.

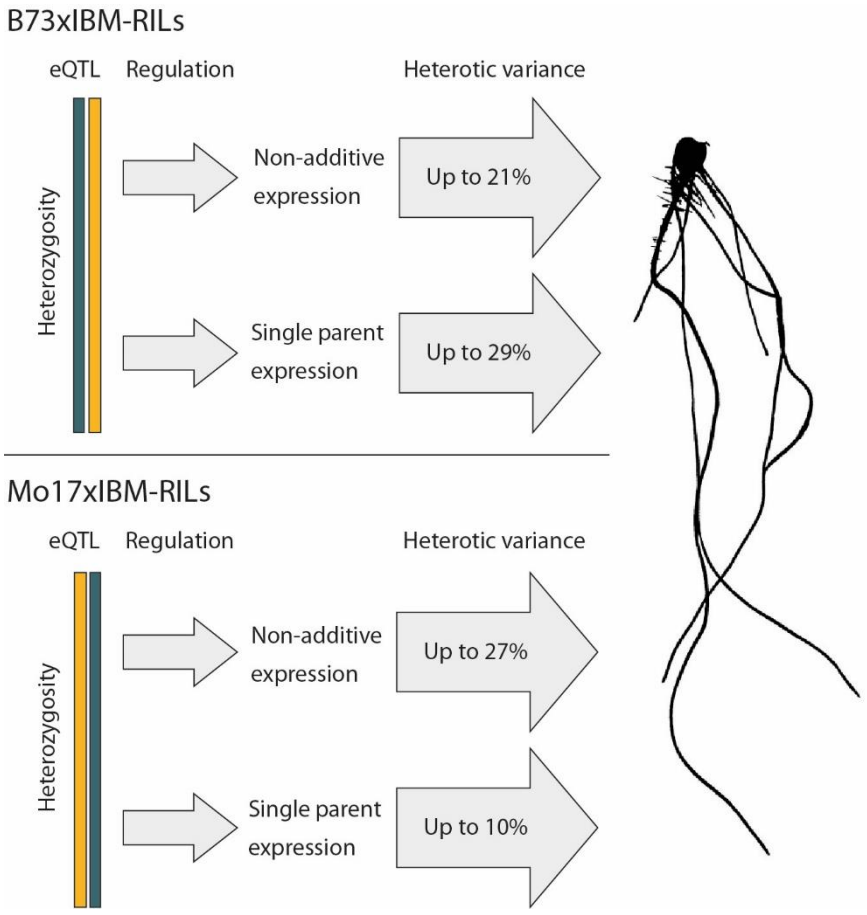


Figure 6: Contribution of non-additive (this study) and single parent expression complementation (Pitz *et al*., 2024) to heterotic variance of root traits in B73xIBM-RIL and Mo17xIBM-RIL backcross populations. Both, non-additive and single parent expression complementation are predominantly regulated by eQTL in heterozygous regions of the genome.

A possible explanation for the observed differences in contribution of heterozygosity and non-additive expression between the different backcross populations might be connected to the different regulation modes of non-additive expression patterns: We discovered that non-additive genes with Mo17 as the high parent were predominantly *trans*-regulated (~70%), while those with B73 as high-parent were almost exclusively *cis*-regulated (~95%) (Figure 3 A-B). In a previous study, *trans*-regulation was associated with paternal dominance[29]. Although Mo17 is traditionally the paternal parent, we observed that *trans*-regulation associated with Mo17 expression was independent of the maternal or paternal origin. A similar observation was made for genes that display single-parent expression (SPE), where the expression of a gene in only one parent is complemented in the hybrid [44]. In summary, Mo17 dominance, in terms of higher expression or activity, is largely *trans*-regulated, while B73 dominance is preferentially *cis*-regulated. The same tendency was observed for SPE genes with either Mo17 (*trans*-regulation) or B73 (*cis*-regulation) as the active parent[44]. Based on our findings, we suggest

that one of the factors contributing to the outstanding hybrid performance of B73 and Mo17 hybrids might be the distinct regulation (*cis* vs *trans*) of non-additive and SPE genes of these genotypes. The observed combination of *cis*- and *trans*-regulated gene expression results in a complementation of higher than mid-parental gene expression. Due to their conservation under different conditions, non-additive expression was proposed to be beneficial to the hybrid under different environmental conditions[26]. Additionally, complementation of gene expression and function during plant development was suggested to contribute to heterosis in Arabidopsis[26,43]. The presented differences in the high parent regulation regime might explain the different results obtained for the two different populations, as not all genes can be complemented in the backcross hybrids.

While higher expression is not necessarily beneficial for every phenotypic trait and for every gene and condition, non-additive gene expression was suggested to be beneficial for hybrids to thrive under different environmental cues[26,43]. We demonstrated that eQTL associated with non-additively expressed genes are mainly located in heterozygous regions, leading to a complementation of higher than mid-parent expression across non-additive genes, showing how heterozygosity on the regulatory level influences complementation of gene expression. We further showed that genes displaying non-additive expression patterns contribute to heterosis and that their regulation might be a new aspect, necessary to translate phylogenetic distance into vigorous hybrids.

Material and Methods

Plant Material

For our study, we backcrossed a subset of 112 IBM-RILs of the maize intermated recombinant inbred line population (IBM-RIL syn. 4)[34] as males to their original parents B73 and Mo17 as females. The IBM-RIL syn. 4 population was generated by crossing the maize inbred lines B73 and Mo17, followed by four generations of intercrossing and subsequent self-pollination of their progeny[47]. Different IBM-RILs are highly homozygous and diverse regarding their genomic regions contributed by B73 and Mo17. B73xIBM-RIL and Mo17xIBM-RIL backcross hybrids are partially homozygous and heterozygous. Backcross hybrids of a specific IBM-RIL show contrasting homozygous and heterozygous genomic regions. We additionally included the parental inbred lines B73 and Mo17, as well as their reciprocal hybrids B73xMo17 and Mo17xB73 as fully heterozygous reference hybrids.

Experimental design and harvesting of plant material

We germinated all genotypes in an alpha-design with incomplete blocks as described in detail in Pitz *et al.* (2024)[44] containing 3 biological replicates of each of the 112 IBM-RILs (336 samples) and each of the 112 backcross hybrids B73xIBM-RIL and Mo17xIBM-RIL (672 samples). Moreover, we included 48 biological replicates of the maternal inbred lines B73 and Mo17 and 24 biological replicates of the reciprocal hybrids B73xMo17 and Mo17xB73 as reference hybrids. In total, we analyzed 1,152 samples. For each sample, we sterilized 25 kernels with 10% H_2O_2 and pre-germinated them in germination paper rolls and placed them in distilled water in a climate chamber with 16 h light (26 °C) and 8 h dark period (21°C). After three days, we selected up to 8 seedlings per sample based on similar primary root length, and placed them into an aeroponic growth system (“Elite Klone Machine 96”, TurboKlone, USA). After another 4 days, we collected the distal part of the primary root, which included the root tip as well as the meristematic and elongation zone. We stored the roots immediately in liquid nitrogen until RNA extraction[44].

RNA-sequencing processing and SNP calling

We ground each sample, consisting of up to eight primary roots of the same genotype, in liquid nitrogen before RNA extraction with the RNeasy Plant Mini Kit (QIAGEN, Venlo, the Netherlands). The next generation sequencing core facility in Bonn, Germany (https://btc.uni-bonn.de/ngs/) assessed the RNA quality, using a Bioanalyzer (RNA ScreenTape + TapeStation Analysis Software 3.2, Agilent Technologies, Santa Clara, CA, USA), constructed the cDNA libraries, necessary for RNA-sequencing following the protocol for TruSeq reversely-stranded mRNA (Illumina, San Diego, USA) and sequenced 100-bp paired-end reads on a

NovaSeq 6000 S4 flow cell machine (Illumina, San Diego, USA). We trimmed and filtered the raw-reads with Trimmomatic (version 0.39) in paired-end mode with the following settings: ILLUMINACLIP:adapters/TruSeq3-PE-2.fa:2:30:10:8:True, LEADING:3, TRAILING:3, MAXINFO:30:0.8, and MINLEN:40[48]. We aligned the trimmed reads to the Maize reference genome (B73v5, ftp.ensemblgenomes.org/pub/plants/release-52/fasta/zea_mays/dna/Zea_mays.Zm-B73-REFERENCE-NAM-5.0.dna.toplevel.fa.gz) after indexing (exon information from http://ftp.ensemblgenomes.org/pub/plants/release-52/gff3/zea_mays/Zea_mays.Zm-B73-REFERENCE-NAM-5.0.52.gff3.gz) with Hisat2 (version 2.2.1; Kim et al. 2015) with the settings: -q –phred 33 –rna-strandedness RF --min-intronlen 20 –max-intronlen 60000. We then used samtools from htslib (version 1.14)[49] and Picard tools (version 2.27.1; http://broadinstitute.github.io/picard/) for formatting and duplicate removal. We then counted uniquely mapped reads with htseq-count (version 2.0.1)[50] and excluded samples with a library size <5 million counted reads.

We prepared the alignments for SNP calling with GATKs HaplotypeCaller by adding readgroup information (Picard AddOrReplaceReadGroups, version 2.27.1), filtering for uniquely mapped reads (mapping quality >=60) and formatting (samtools view, version 1.14). Subsequently, we split alignments at positions with N in the CIGAR files, for example intron-spanning reads, by using GATKs SplitNCigarReads (version 2.4.6.1)[51,52].

We performed SNP calling in two steps[44]. In brief, we performed a first SNP calling, based on which we excluded samples with low homozygosity within regions, that should be homozygous or genotypes in the hybrids not matching the parental genotypes. We identified SNPs between our Mo17 inbred line samples and the B73v5 reference genome. Additionally, we identified SNPs between B73 inbred line samples and the B73v5 reference genome to exclude those loci where both, B73 and Mo17 lines show a non-reference allele. We then counted the frequency of the B73 and Mo17 allele at each SNP locus in each sample ([53], adapted from[39]). First, we calculated the ratio of homozygous loci in the parental inbred lines. In case of <95% homozygous loci, we excluded the sample. The same was done for the homozygous regions of the partially homozygous backcross hybrids. Thus, we excluded 175 samples. Additionally, 10 samples had a library size of <5 million read counts and 17 samples were excluded, because they were the only one left out of three replicates of a genotype. We further generated a variant call file (vcf) of known SNPs to be used in the second SNP calling. Subsequently, we had to exclude 90 hybrid samples, because the paternal inbred was excluded and 8 IBM-RIL paternal inbred lines, because all corresponding hybrids were excluded. This left 852 samples for the second SNP calling. For this, we filtered the variant call files (vcf) from the first SNP calling (https://gatk.broadinstitute.org/hc/en-us/articles/360035890471-Hard-filtering-germline-short-variants, checked on 12/18/2023) and used them for recalibration of base

qualities with GATKs BaseRecalibrator and ApplyBSQR in the samples (https://gatk.broadinstitute.org/hc/en-us/articles/360035531192-RNAseq-short-variant-discovery-SNPs-Indels-, checked on 12/14/2023). We then used the HaplotypeCaller in BP_RESOLUTION mode. To subsequently eliminate positions without information, we filtered the result for positions with a coverage (DP) of ≥1. We obtained on database per triplet (IBM-RIL, B73xIBM-RIL, Mo17xIBM-RIL) by combining the respective samples. To ensure genotyping of all possible SNP loci (not just the ones present in the respective triplet), we added the B73 and Mo17 samples to each triplets database. We filtered the resulting genotyped loci for SNPs with QD >2, SOR <3, MQ >40, QUAL >30 and FS <60 using bcftools (version 1.17). To determine a list of high confidence B73 versus Mo17 SNPs, we confirmed the B73 allele in 90% but ≥3 genotyped B73 samples and the Mo17 allele in 90% but at least 3 genotyped Mo17 samples with high (>10) GenotypeCall quality (GQ). Additionally, we identified SNPs which were not present in any B73 or Mo17 samples IBM-RIL specific (homozygous or heterozygous and regardless of GQ).

Classification of genomic regions and genes, determination of heterozygosity

We used the filtered SNP data (GQ>10) to classify each IBM-RIL genome into B73 or Mo17 regions and to mask regions which were not B73 or Mo17. In short, we implemented a distance-function to group IBM-RIL specific loci with a distance of <2.5 Mbp into blocks. Blocks of at least 10 IBM-RIL-specific loci were masked as third origin regions. We then used a sliding window approach of 15 consecutive loci to eliminate singular loci that did not match their surrounding loci[54]. We used the previously mentioned distance-function to group loci that carry the same allele and which were less than 0.5 Mbp apart as a block, and all blocks were retained. We excluded two IBM-RILs which had more than 50% of their genomes consisting of IBM-RIL specific regions from a third parental origin and their respective hybrids (details in[44]) Thus, 834 samples remained for final analyses. We filtered the data set of each triplet to only include loci within the B73 or Mo17 regions of the IBM-RILs and within exons of protein-coding genes and we classified the genes as within a B73, within a Mo17 region, an unclassified region without SNP information, or masked them as IBM-RIL specific. This verification was performed for each IBM-RIL separately. We calculated the proportion of heterozygous to homozygous regions for each backcross hybrid by dividing the total lengths of classified heterozygous regions (B73 regions of the IBM-RIL for the Mo17×IBM-RIL and Mo17 regions of the IBM-RIL for B73×IBM-RIL) by the total lengths of all classified regions (not considering IBM-RIL specific masked regions and regions without SNP information) (Data S3).

Determination of differential and non-additive gene expression patterns

We obtained differentially expressed genes by processing raw read counts with the Bioconductor package limma (version 3.50.3)[55] in R (version 4.1.1). For each triplet combination composed of both parents and their hybrid offspring only genes which are active in at least one of the we considered three genotypes. In addition, we filtered lowly expressed genes by the filterByExpr() function of the Bioconductor package edgeR (version 3.36.0). We used the function CalcNormFactors() of limma to calculate normalization factors of the raw counts, which were later used by the voomWithQualityWeights() function of limma to obtain sample- and gene- specific weights. We implemented the following model to estimate the gene expression across samples and genes[55,56] :

$$Y_{jkln(i)} = g_i + p_k + s_l + e_{jkln(i)} \quad (1)$$

We modelled the expression value of a specific gene of the respective genotype *i* as $Y_{jkl(i)}$. The fixed effect for genotype *i* was represented by g_i. The remaining terms correspond to the experimental design, where the fixed effect for batch *k* nested within block *j* was included by p_k and represents an incomplete block within a replicate. The random effect for system *l* nested within batch *k* and block *j* was included as s_l and represents one of eight growth systems within each block. The random error effect for row *n* of genotype *i* in block *j* and batch *k* and system *l* was represented by $e_{jkln(i)}$. We subjected the fixed effects to the function voomWithQualityWeights() of limma, with s_l serving as a block to obtain consolidated weights for library size and heterogeneity in sample quality as well as observational variance[57]. Since the limma package does not provide a framework for random effects, we used limmas' duplicateCorrelation() function to approximate the effect, with s_l as a block[58]. Both, the voomWithQualityWeights() and duplicateCorrelation() function were run twice, and we updated the resulting consensus value and used it in the lmFit() function. We made two contrasts for detecting differential expression: First, we compared each hybrid value to the mean of both parents and second, we compared both parents to each other using the contrast.fit() function. We computed moderated t-statistics using an empirical Bayes method by the eBayes() function. We considered genes differentially expressed if they had an absolute log2FC >1 and an adjusted *p*-value ≤0.05[59,60].

Determination of the contribution of non-additive expression pattern to heterotic variance

We quantified the contribution of non-additively expressed genes with higher than MPV expression to heterosis. Therefore, we calculated the total heterotic variance, accounting for experimental factors and parental effects (σ^2_G), using a linear mixed model (2). By extending the model and including the numbers of non-additive pattern genes (σ^2_{Het}), we were able to calculate the proportion of variance attributable to non-additive genes (p_{Het}) (3). This approach was adapted from the evaluation of SPE contribution to heterosis from[44]). We performed separate analyses for the B73xIBM-RIL and Mo17xIBM-RIL population. The phenotypic data for "total root length", "total root volume" and "total number of root tips" was obtained from image analysis of the whole root system of each plant. "Lateral root density" was obtained by counting the number of lateral roots emerged from the proximal first centimeter of the primary root[44]. We square-root transformed the values of the traits "total root length", "total root volume" and "total number of root tips", to fulfil the modelling assumptions.

The base-model in each population for calculating σ^2_G was fitted as follows. We defined covariates to differentiate the IBM-RILs and B73 and Mo17 as parental genotypes from their hybrids, but include them in the model simultaneously. We set the covariates initially to zero, but for observations corresponding to a parental genotype, we set the covariate for this genotype to one. We set the parental covariates of both parents to 0.5 for observations on the hybrid. Thus, ultimately the terms $\beta_1 x_{i1} + \cdots + \beta_q x_{iq}$ model the general parental performance as well as the hybrid mid-parent values.

$$Y_{jklmnp(i)} = \mu + \beta_1 x_{i1} + \cdots + \beta_q x_{iq} + b_j + z_i * g_i + p_k + s_l + t_m + r_n + e_{jklmnp(i)} \quad (2)$$

We defined $Y_{jklmnp(i)}$ as the parental effect on MPH of the respective hybrid genotype *i* for a phenotypic trait. We defined μ as the intercept, x_{iq} as the parental covariables of parent *q* for genotype *i*, b_j as the fixed effect for block *j*. Further, we included $z_i * g_i$ as random effect of genotype *i*, whereas z_i is a dummy variable and set to $z_i = 0$ for parents and $z_i = 1$ for hybrids[61]. We defined b_j as the fixed effect for block *j*, which represents one of three complete replicates. We added p_k as random effect for batch *k* nested within block *j* and s_l as the random effect for system *l* nested within batch *k* and block *j*, which represents one of eight growth systems within each batch. t_m represents the random effect for one of four triplets *m* nested within system *l*, batch *k* and block *j*, which each consisted of an IBM-RIL and both corresponding backcrosses or the reference inbreds B73 and Mo17 and a reciprocal hybrid. We let r_n represent the random effect for a row of plants with the same genotype *n* nested within triplet *m*, system *l*, batch *k* and block *j*. The random error effect $e_{jklmnp(i)}$ corresponds to plant *p* of genotype *i* in block *j*, batch *k*, system *l*, triplet *m* and row *n*.

Next, we included the numbers of non-additively expressed genes in the model. It should be noted, that each pattern (1-4 in B73xIBM-RILs, 5-8 in Mo17xIBM-RILs) was represented by their own covariate. The numbers of non-additive genes were set to 0 for parental genotypes.

$$Y_{jklmnp(i)} = \mu + \beta_1 x_{i1} + \cdots + \beta_q x_{iq} + b_j + \gamma_1 sa_i + \gamma_1 sb_i + \gamma_1 sc_i + \gamma_1 sd_i + z_i * g_i + p_k + s_l + t_m + r_n + e_{jklmnp(i)} \quad (3)$$

Thus, sa_i , sb_i , sc_i and sd_i represent the covariables for non-additively expressed genes in pattern 1-4 (Figure 2A) or 5-8 (Figure 2B) for hybrid *i.* We used the lme4 package (version 1.1-29) within R (version 4.0.1) for this analysis.

eQTL analysis

We performed an eQTL analysis with the R/qtl2 package (version 0.22)[62] to identify positions that were significantly associated with gene expression values based on the masked and filtered SNP data as described in detail in[44]. For each of the three cross-types (IBM-RIL, B73×RILs, Mo17×RILs) we took the classified and filtered SNP loci within B73 or Mo17 regions in the IBM-RILs as marker input data. We used the positions of these SNP loci as preliminary genomic positions, as well as physical positions. As the phenotype data input in R/qtl2 we used the estimated expression means obtained from the model coefficients within the differential expression analysis of each genotype and gene. We removed samples with more than 19% missing genotypes, duplicated genotypes and markers with more than 60% missing genotype information. We estimated the genetic map from the physical positions and genotype information and kept markers that were ≥1 cM apart to avoid retaining an excess of redundant markers. We used a hidden Markov model and and Haley-Knott regression[63] to establish the association between genotype and expression phenotype with a linear model. In simple words, within each eQTL analysis, each marker is tested to see whether there is an association with a single genes expression, the result is an LOD curve. In order to find out whether the highest LOD value is significant, we performed a permutation analysis with 10,000 permutations and all significant peaks (α ≤0.001) were saved[62,64,65]. This process was repeated for all (37,782) active genes. To subsequently also correct for the testing of multiple genes, we considered genes with FDR ≤0.001 significant. We performed this procedure on all three cross-type data sets (IBM-RIL, B73×IBM-RIL, Mo17×IBM-RIL). We combined the resulting eQTL peaks and distinct eQTLs were selected: in cases where multiple eQTLs were identified for a gene, we assessed whether the different peak positions corresponded to different regulatory elements (≥25 Mbp apart or on different chromosomes, not within the confidence intervals of each other). If multiple eQTLs for the same gene did not differ by the specified standards, we only retained the eQTL with the shortest confidence interval or the highest LOD in case of equal confidence intervals. We categorized the eQTLs into *cis* and *trans* eQTLs based on their distance from

the start of their respective gene. We defined *Trans*-regulating eQTLs as located at a distance of at least 2.5 Mbp from the start of the gene and where their confidence interval did not include the start of the gene. We classified c*is*-regulating eQTLs as located in proximity to the start of the gene (<2.5 Mbp) or located such that their confidence interval includes the start of the gene[44].

Acknowledgements

We would like to thank KWS for the propagation of the IBM-RIL syn. 4 seeds. We thank Helmut Rehkopf (University of Bonn) for his support in propagating the genetic material for this study and the experimental station "Auf dem Hügel" of the University of Bonn. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) Next Generation Sequencing Competence Network (NGS-CN; project 423957469) grant HO 2249/18-1 to F.H. NGS analyses were carried out at the West German Genome Center (WGGC) site in Bonn. The authors acknowledge access to the bonna cluster hosted by the University of Bonn along with the support provided by its High-Performance Computing & Analytics Lab.

Competing interests

The authors declare no competing interests.

Author contributions

M. P. and J.A.B. carried out the experiments, conducted the statistical analyses, interpreted the data, and drafted the manuscript. H.-P.P. provided help with the experimental design for the RNA-seq experiment and helped with the statistical analyses. F.H. conceived and coordinated the study and participated in data interpretation and drafting the manuscript. All authors edited the manuscript and approved the final draft.

Data availability

NCBI Bioproject ID PRJNA923128

References

1. **Shull GH** (1914). Duplicate genes for capsule-form in *Bursa bursa-pastoris*. *Zeitschrift für Induktive Abstammungs- und Vererbungslehre* **12,** 97–149; 10.1007/bf01837282.
2. **Duvick DN** (2005). The contribution of breeding to yield advances in maize (*Zea mays* L.) **86,** 83–145; 10.1016/S0065-2113(05)86002-X.
3. **Hochholdinger F & Baldauf JA** (2018). Heterosis in plants. *Current Biology* **28,** R1089-R1092; 10.1016/j.cub.2018.06.041.
4. **Hochholdinger F & Yu P** (2025). Molecular concepts to explain heterosis in crops. *Trends in Plant Science* **30**; 10.1016/j.tplants.2024.07.018.
5. **East EM** (1936). Heterosis. *Genetics* **21,** 375–397; 10.1093/genetics/21.4.375.
6. **Reif JC, Hallauer AR, Melchinger AE** (2005). Heterosis and heterotic patterns in maize. *Maydica* **50,** 215–223.
7. **Beukert U, Li Z, Liu G, Zhao Y, Ramachandra N, Mirdita V, Pita F, Pillen K, Reif JC** (2017). Genome-based identification of heterotic patterns in rice. *Rice* **10,** 22; 10.1186/s12284-017-0163-4.
8. **Melchinger AE & Gumber RK** (1998). Overview of heterosis and heterotic groups in agronomic crops. *Concepts and Breeding of Heterosis 2015,* 29–44; 10.2135/cssaspecpub25.c3.
9. **Paril J, Reif J, Fournier-Level A, Pourkheirandish M** (2024). Heterosis in crop improvement. *The Plant Journal* **117,** 23–32; 10.1111/tpj.16488.
10. **Hoecker N, Keller B, Piepho H-P, Hochholdinger F** (2006). Manifestation of heterosis during early maize (*Zea mays* L.) root development. *Theoretical and Applied Genetics* **112,** 421–429; 10.1007/s00122-005-0139-4.
11. **Jones DF** (1917). Dominance of linked factors as a means of accounting for heterosis. *Proceedings of the National Academy of Sciences of the United States of America* **3,** 310–312; 10.1073/pnas.3.4.310.
12. **Lin T, Zhou C, Chen G, Yu J, Wu W, Ge Y, Liu X, Li J, Jiang X, Tang W** *et al.* (2020). Heterosis-associated genes confer high yield in super hybrid rice. *Theoretical and Applied Genetics* **133,** 3287–3297; 10.1007/s00122-020-03669-y.
13. **Krieger U, Lippman ZB, Zamir D** (2010). The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nature Genetics* **42,** 459–463; 10.1038/ng.550.
14. **Chen ZJ & Birchler JA** (2013). *Polyploid and hybrid genomics.* Ames, Iowa: Wiley-Blackwell.
15. **Duvick DN** (2001). Biotechnology in the 1930s: the development of hybrid maize. *Nature Reviews Genetics* **2,** 69–74; 10.1038/35047587.
16. **Hochholdinger F & Hoecker N** (2007). Towards the molecular basis of heterosis. *Trends in plant science* **12,** 427–432; 10.1016/j.tplants.2007.08.005.
17. **Hoecker N, Keller B, Muthreich N, Chollet D, Descombes P, Piepho H-P, Hochholdinger F** (2008). Comparison of maize (*Zea mays* L.) F_1-hybrid and parental inbred line primary root transcriptomes suggests organ-specific patterns of nonadditive gene expression and conserved expression trends. *Genetics* **179,** 1275–1283; 10.1534/genetics.108.088278.
18. **Zhang X, Hirsch CN, Sekhon RS, Leon N** de, **Kaeppler SM** (2016). Evidence for maternal control of seed size in maize from phenotypic and transcriptional analysis. *Journal of Experimental Botany* **67,** 1907–1917; 10.1093/jxb/erw006.
19. **Paschold A, Jia Y, Marcon C, Lund S, Larson NB, Yeh C-T, Ossowski S, Lanz C, Nettleton D, Schnable PS** *et al.* (2012). Complementation contributes to transcriptome complexity in maize (*Zea mays L.*) hybrids relative to their inbred parents. *Genome Research* **22,** 2445–2454; 10.1101/gr.138461.112.

20. **Uzarowska A, Keller B, Piepho H-P, Schwarz G, Ingvardsen C, Wenzel G, Lübberstedt T** (2007). Comparative expression profiling in meristems of inbred-hybrid triplets of maize based on morphological investigations of heterosis for plant height. *Plant Molecular Biology* **63,** 21–34; 10.1007/s11103-006-9069-z.

21. **Baldauf JA, Marcon C, Paschold A, Hochholdinger F** (2016). Nonsyntenic genes drive tissue-specific dynamics of differential, nonadditive, and allelic expression patterns in maize hybrids. *Plant Physiology* **171,** 1144–1155; 10.1104/pp.16.00262.

22. **Stupar RM & Springer NM** (2006). *Cis*-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F_1 hybrid. *Genetics* **173,** 2199–2210; 10.1534/genetics.106.060699.

23. **Qin J, Scheuring CF, Wei G, Zhi H, Zhang M, Huang JJ, Zhou X, Galbraith DW, Zhang H-B** (2013). Identification and characterization of a repertoire of genes differentially expressed in developing top ear shoots between a superior hybrid and its parental inbreds in *Zea mays* L. *Molecular Genetics and Genomics* **288,** 691–705; 10.1007/s00438-013-0781-5.

24. **Stupar RM, Gardiner JM, Oldre AG, Haun WJ, Chandler VL, Springer NM** (2008). Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis. *BMC Plant Biology* **8,** 33; 10.1186/1471-2229-8-33.

25. **Guo M, Rupe MA, Yang X, Crasta O, Zinselmeier C, Smith OS, Bowen B** (2006). Genome-wide transcript analysis of maize hybrids: allelic additive gene expression and yield heterosis. *Theoretical and applied genetics* **113,** 831–845; 10.1007/s00122-006-0335-x.

26. **Marcon C, Paschold A, Malik WA, Lithio A, Baldauf JA, Altrogge L, Opitz N, Lanz C, Schoof H, Nettleton D** *et al.* (2017). Stability of single-parent gene expression complementation in maize hybrids upon water deficit stress. *Plant Physiology* **173,** 1247–1257; 10.1104/pp.16.01045.

27. **Jansen RC & Nap JP** (2001). Genetical genomics: the added value from segregation. *Trends in Genetics* **17,** 388–391; 10.1016/S0168-9525(01)02310-1.

28. **Botet R & Keurentjes JJB** (2020). The role of transcriptional regulation in hybrid vigor. *Frontiers in Plant Science* **11,** 410; 10.3389/fpls.2020.00410.

29. **Swanson-Wagner RA, DeCook R, Jia Y, Bancroft T, Ji T, Zhao X, Nettleton D, Schnable PS** (2009). Paternal dominance of *trans*-eQTL influences gene expression patterns in maize hybrids. *Science* **326,** 1118–1120; 10.1126/science.1178294.

30. **Rahman H, Pekic S, Lazic-Jancic V, Quarrie SA, Shah SMA, Pervez A, Shah MM** (2011). Molecular mapping of quantitative trait loci for drought tolerance in maize plants. *Genetics and Molecular Research* **10,** 889–901; 10.4238/vol10-2gmr1139.

31. **Pan Q, Xu Y, Li K, Peng Y, Zhan W, Li W, Li L, Yan J** (2017). The genetic basis of plant architecture in 10 maize recombinant inbred line populations. *Plant Physiology* **175,** 858–873; 10.1104/pp.17.00709.

32. **Huo X, Wang J, Zhang L** (2023). Combined QTL mapping on bi-parental immortalized heterozygous populations to detect the genetic architecture on heterosis. *Frontiers in Plant Science* **14,** 1157778; 10.3389/fpls.2023.1157778.

33. **Yang H, Zhang Z, Zhang N, Li T, Wang J, Zhang Q, Xue J, Zhu W, Xu S** (2024). QTL mapping for plant height and ear height using bi-parental immortalized heterozygous populations in maize. *Frontiers in Plant Science* **15,** 1371394; 10.3389/fpls.2024.1371394.

34. **Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A** (2002). Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. *Plant Molecular Biology* **48,** 453–461; 10.1023/a:1014893521186.

35. **Shull GH** (1908). The composition of a field of maize. *Journal of Heredity* **os-4,** 296–301; 10.1093/jhered/os-4.1.296.

36. **Schnable PS & Springer NM** (2013). Progress toward understanding heterosis in crop plants. *Annual review of plant biology* **64,** 71–88; 10.1146/annurev-arplant-042110-103827.

37. **Birchler JA & Veitia RA** (2010). The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytologist* **186,** 54–62; 10.1111/j.1469-8137.2009.03087.x.

38. **Yao H, Dogra Gray A, Auger DL, Birchler JA** (2013). Genomic dosage effects on heterosis in triploid maize. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 2665–2669; 10.1073/pnas.1221966110.

39. **Baldauf JA, Vedder L, Schoof H, Hochholdinger F** (2020). Robust non-syntenic gene expression patterns in diverse maize hybrids during root development. *Journal of Experimental Botany* **71,** 865–876; 10.1093/jxb/erz452.

40. **Yoo M-J, Szadkowski E, Wendel JF** (2013). Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* **110,** 171–180; 10.1038/hdy.2012.94.

41. **Combes M-C, Hueber Y, Dereeper A, Rialle S, Herrera J-C, Lashermes P** (2015). Regulatory divergence between parental alleles determines gene expression patterns in hybrids. *Genome Biology and Evolution* **7,** 1110–1121; 10.1093/gbe/evv057.

42. **Janko K, Eisner J, Cigler P, Tichopád T** (2024). Unifying framework explaining how parental regulatory divergence can drive gene expression in hybrids and allopolyploids. *Nature Communications* **15,** 8714; 10.1038/s41467-024-52546-5.

43. **Liu W, He G, Deng XW** (2021). Biological pathway expression complementation contributes to biomass heterosis in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **118,** e2023278118; 10.1073/pnas.2023278118.

44. **Pitz M, Baldauf J, Piepho H-P, Yu P, Schoof H, Mason AS, Li G, Hochholdinger F** (2024). Regulation of heterosis-associated gene expression complementation in maize hybrids. *bioRxiv,* 2024.10.30.620956; 10.1101/2024.10.30.620956.

45. **Baldauf JA, Liu M, Vedder L, Yu P, Piepho H-P, Schoof H, Nettleton D, Hochholdinger F** (2022). Single-parent expression complementation contributes to phenotypic heterosis in maize hybrids. *Plant Physiology* **189,** 1625–1638; 10.1093/plphys/kiac180.

46. **Baldauf JA, Marcon C, Lithio A, Vedder L, Altrogge L, Piepho H-P, Schoof H, Nettleton D, Hochholdinger F** (2018). Single-parent expression is a general mechanism driving extensive complementation of non-syntenic genes in maize hybrids. *Current Biology* **28,** 431-437.e4; 10.1016/j.cub.2017.12.027.

47. **Liu SC, Kowalski SP, Lan TH, Feldmann KA, Paterson AH** (1996). Genome-wide high-resolution mapping by recurrent intermating using *Arabidopsis thaliana* as a model. *Genetics* **142,** 247–258; 10.1093/genetics/142.1.247.

48. **Bolger AM, Lohse M, Usadel B** (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120; 10.1093/bioinformatics/btu170.

49. **Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM** *et al.* (2021). Twelve years of SAMtools and BCFtools. *GigaScience* **10**; 10.1093/gigascience/giab008.

50. **Anders S, Pyl PT, Huber W** (2015). HTSeq—a python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169; 10.1093/bioinformatics/btu638.

51. **van der Auwera G & O'Connor BD** (2020). *Genomics in the cloud: using Docker, GATK, and WDL in Terra.* 1st ed. Beijing: O'Reilly Media, Incorporated.

52. **GATK** (2023). RNAseq short variant discovery (SNPs + Indels). https://gatk.broadinstitute.org/hc/en-us/articles/360035531192-RNAseq-short-variant-discovery-SNPs-Indels- (visited on 14.12.2023).

53. **Vedder L** (2024). MaizeSNP. DOI:10.5281/ZENODO.10684044.

54. **Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T** *et al.* (2009). High-throughput genotyping by whole-genome resequencing. *Genome Research* **19,** 1068–1076; 10.1101/gr.089516.108.

55. **Ritchie ME, Phipson B, Di Wu, Hu Y, Law CW, Shi W, Smyth GK** (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43,** e47; 10.1093/nar/gkv007.

56. **Law CW, Chen Y, Shi W, Smyth GK** (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology* **15,** R29; 10.1186/gb-2014-15-2-r29.

57. **Liu R, Holik AZ, Su S, Jansz N, Chen K, Leong HS, Blewitt ME, Asselin-Labat M-L, Smyth GK, Ritchie ME** (2015). Why weight? Modelling sample and observational level variability improves power in RNA-seq analyses. *Nucleic Acids Research* **43,** e97; 10.1093/nar/gkv412.

58. **Smyth GK, Michaud J, Scott HS** (2005). Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21,** 2067–2075; 10.1093/bioinformatics/bti270.

59. **Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK** (2016). Robust hyperparameter estimation protects agains hypervariable genes and improves power to detect differential expression. *The annals of applied statistics* **10,** 946–963; 10.1214/16-AOAS920.

60. **Benjamini Y & Hochberg Y** (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57,** 289–300; 10.1111/j.2517-6161.1995.tb02031.x.

61. **Piepho H-P, Williams ER, Fleck M** (2006). A note on the analysis of designed experiments with complex treatment structure. *HortScience* **41,** 446–452; 10.21273/HORTSCI.41.2.446.

62. **Broman KW, Gatti DM, Simecek P, Furlotte NA, Prins P, Sen Ś, Yandell BS, Churchill GA** (2019). R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multiparent populations. *Genetics* **211,** 495–502; 10.1534/genetics.118.301595.

63. **Haley CS & Knott SA** (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69,** 315–324; 10.1038/hdy.1992.131.

64. **Lystig TC** (2003). Adjusted p values for genome-wide scans. *Genetics* **164,** 1683–1687; 10.1093/genetics/164.4.1683.

65. **Broman K** (2023). R/qtl2 user guide. https://kbroman.org/qtl2/assets/vignettes/user_guide.html (visited on 19.12.2023)

5. Discussion

The agricultural performance of F_1-hybrids often surpasses the performance of the two distinct parental inbred lines. This effect is called heterosis[68,90]. The heterosis effect is utilized in many crops, particularly in maize, with wide ranging socio-economic implications. Despite ongoing research, extensive knowledge gaps remain to completely explain the molecular basis of this effect[118]. With the advance of transcriptomic studies, several types of expression patterns between parents and hybrid have been studied in relation to heterosis in maize[97,102,111,119]. Among those, single parent expression (SPE) was repeatedly suggested to positively influence hybrid performance. In this pattern, a gene is active in only one parent, but also active in the hybrid by complementation, leading to a higher number of active genes in the hybrid[110–113]. The role of additive and non-additive expression patterns is less precisely defined[97,120]. The quantitative contribution of genes showing non-additive pattern to heterosis as well as the regulation of SPE and non-additive genes remain largely elusive.

In the present study, we outcrossed 112 lines of the maize intermated B73xMo17 recombinant inbred line (IBM-RIL) syn. 4 population[19] to their original parental inbred lines B73 and Mo17. The generation of these populations had several advantages: They allowed us to quantify the contribution of SPE and non-additive expression pattern to heterosis manifestation in seedling root development. Additionally, we identified locations of regulatory elements for gene expression by eQTL analyses. The B73xMo17 hybrid shows exceptional heterosis, with B73 as female parent and Mo17 as male hybrid parent[74,76]. This hybrid also represents the common heterotic pattern of crossing a stiff stalk and a non-stiff stalk line[73,74]. This hybrid was used commercially in the 1980s. Nevertheless, these genetic resources are still relevant today, as large proportions of contemporary germplasm are descended from B73[75,76].

5.1 Recombination in IBM-RIL population

Using SNP calling, we identified crossovers between B73 and Mo17 regions in the IBM-RILs. In addition to SNPs clearly indicating Mo17 or B73 origin, we unexpectedly identified a third category of SNPs. These did not occur in B73 or Mo17 reference genotypes, therefore we designated them IBM-RIL specific SNPs. They were not evenly distributed across the genome, but clustered in certain IBM-RIL specific regions. The most straightforward explanation is a contamination of the IBM-RIL germplasm with one or more additional lines by outcrossing. The cross-pollinating nature of maize makes the generation of recombinant inbred lines prone to contamination by outcrossing with foreign pollen, which can occur at any generation[121]. We therefore masked the IBM-RIL specific regions to minimize their influence on the presented results.

Regarding the recombination frequency, we generally detected only few cross-overs in the centromeric regions of the ten chromosomes. This is consistent with previous observations[122]. In addition, regions without sufficient SNP information were frequently close to the centromeres. SNP information can be low due to low transcription, little variance between the B73v5 reference and the investigated samples, gaps in the reference genome or absence of B73v5 genomic regions in Mo17. On average, there were 71.5 recombination breakpoints in each IBM-RIL (chapter 3), thus increasing the original RFLP marker-based mapping resolution[19]. The number and distribution of recombinations indicates a sound categorization of the IBM-RILs into B73 and Mo17 regions[123]. The definition of these cross-over sites together with the knowledge of possible IBM-RIL specific regions will be useful for further studies on the widely used IBM-RIL population.

5.2 Non-additive and single parent expression complementation

Across all our investigated hybrids, we observed a general trend that non-additive genes are expressed higher than the mid-parent level. A similar pattern was observed in other maize hybrids[102] and in other crops such as cotton[124] or coffee[125]. In these studies, the non-additive genes rather adopted the expression level of the higher expressed parent than that of the lower expressed parent. In an extreme case, expression complementation can lead to a higher number of active genes in a hybrid. This is the case when a gene is active in only one parent, but also active in the hybrid by complementation. This pattern is designated single parent expression (SPE)[102]. All backcross and reference hybrids showed substantial numbers of SPE pattern genes. This finding is supported by previous studies on B73xMo17 and other genotypes in seedling roots[110,113,114].

In conclusion, our findings imply that hybrids not only have more active genes due to SPE, but that non-additive genes are also expressed at higher levels. Complementation of gene expression is in support of the dominance model of heterosis. This early theory explains heterosis by the complementation of many slightly deleterious alleles by beneficial alleles in the hybrid[88].

5.3 Heterosis is trait- and background specific

We identified that up to 29% of heterotic variance can be explained by the number of SPE genes and up to 27% by the number of non-additively expressed genes (chapter 3, 4). A link between the number of SPE genes and heterosis was previously suggested in a panel of distinct genotypes[111]. In contrast to previous studies on SPE[111,114], the present study used only two different parental inbred lines, demonstrating that the associations of expression patterns with heterosis are not only conditioned by the effects of diverse genotypes. We further

demonstrated that not only SPE but also non-additively expressed genes contribute substantially to heterosis and that this contribution can be quantified.

Heterosis is trait specific and it was concluded that different mechanisms or genes are responsible for heterosis of different traits in the same plant[96]. Our findings support this interpretation, as we identified different contributions of gene expression patterns (SPE and non-additive genes) to heterosis, depending on the trait (chapter 3, 4). We propose to expand this concept: Different mechanisms or genes are likely responsible for heterosis in different populations or genetic backgrounds as well. SPE genes explained higher proportions of heterotic variance in B73xIBM-RILs, while non-additive genes contributed higher proportions to all traits in Mo17xIBM-RILs except lateral root density. This interpretation is supported by the fact that different sets of SPE genes were detected in diverse parent-hybrid combinations[110]. We further identified contrasting genes between the B73xIBM-RIL and Mo17xIBM-RIL population (chapter 3) in our TWAS analysis. TWAS genes show a direct correlation between their expression and a phenotypic trait[85]. An overlap between genes identified by TWAS and SPE patterns was predominantly observed for the B73xIBM-RIL population. In summary, our data indicates a substantial but highly background and trait specific influence of non-additive genes and SPE genes on heterosis with partially complementing effects.

5.4 Most expression quantitative trait loci (eQTL) regulate nearby genes in *cis*

The regulation of non-additive and SPE genes is of particular interest, as these patterns are major contributors to heterosis (chapter 3, 4). With the identification of expression quantitative trait loci (eQTL), we determined genomic regions that are significantly associated with gene expression levels across a population, providing direct insights into the regulation of genes. Expression of genes can be regulated by nearby *cis*-regulating elements or more distantly located *trans*-regulating elements[84]. Of the eQTL regulating the active genes, 88% were classified as *cis*-regulating (chapter 3). Other studies identified similar fractions when large effect eQTL were considered[126,127], or reported more *trans*- than *cis*-regulating eQTL when smaller effects were considered, which usually belonged to *trans*-eQTL[115,126]. In other words, the *trans*-eQTL had less influence on expression variability. Our results indicate generally large effects of both *cis*- and *trans*-regulating eQTL. Still, we cannot exclude the possibility that small effect eQTL, especially those acting in *trans*, are present but were not detected or not considered significant in our eQTL analysis.

5.5 Regulatory differences are based on the genetic background of parents and hybrids

Comparing the active genes in each genotype, we detected a higher proportion of *trans*-regulation in the inbred line Mo17 (13%) compared to B73 (7%) (chapter 3), with the hybrids showing intermediate proportions. This indicates different regulatory trends in the two inbred lines, which are largely inherited by the hybrid offspring. In the reciprocal hybrids, this regulatory difference was strongly amplified for SPE genes with Mo17 as the active parental allele (~60% trans) compared to SPE genes with B73 as the active parent (~5% trans) (chapter 3). For non-additive genes, it was even more pronounced with ~70% *trans*-regulation if Mo17 was the high-parent and ~95% *cis*-regulation if B73 was the high parent (chapter 4). In the IBM-RIL backcross populations, significant differences depending on the active or high-parent genotype of SPE and non-additive patterns respectively, support the findings from the reciprocal hybrids. Generally, it was observed that the phylogenetic distance between the parental lines is positively correlated with heterosis[68]. However, genetic diversity alone is not enough to explain heterosis, but specific combinations of parents result in distinct levels of heterosis[69]. The different regulatory preferences (B73: *cis*, Mo17: *trans*) of hybrid parents and their contributions to expression pattern is likely one factor of how phylogenetic distance is related to heterosis.

5.6 *Trans*-regulated non-syntenic SPE genes and their possible functions

We can estimate the evolutionary age of maize genes by the presence or absence of syntenic orthologs (genes with similar position and sequence) in sorghum (*Sorghum bicolor* L.). While syntenic genes are evolutionary conserved, the non-syntenic genes are younger as they were altered after the separation of maize and sorghum, most likely by gene duplications[128]. It was established that SPE genes are significantly enriched among the non-syntenic genes and depleted among syntenic genes[110,111,114]. The presented data confirms and extends these findings (chapter 3). In addition, *trans*-regulated genes are also enriched within the non-syntenic genes (chapter 3). In particular, *trans*-regulated genes with Mo17 activity are present among the non-syntenic genes. In several organisms, non-syntenic genes have been associated with disease resistance[65]. It has also been suggested that non-syntenic SPE genes function in plant environmental adaptation, helping hybrids to better cope with abiotic stress[113]. In contrast to syntenic genes, non-syntenic genes show substantial genetic and transcriptional variation within the maize breeding pool[66], which was also demonstrated in this study. Taken together with the fact that non-syntenic regions are correlated with the level of heterosis[67], our results highlight the importance of non-syntenic genes for heterosis.

5.7 Candidate genes

We identified two candidate genes controlling lateral root density in the backcross hybrids via transcriptome wide association studies (TWAS). One gene (TSG-1: Zm00001eb349930) was *cis*-regulated and associated with lower lateral root density upon gene activity. The other gene (TSG 2: Zm00001eb339600) was *trans*-regulated and showed higher lateral root density in genotypes where the candidate gene was active (chapter 3). In both cases, the genotype at the eQTL position determined the activity of the gene. This further highlights the pivotal role of closely, as well as distantly, located regulatory elements and their genotype for the activity of heterosis-associated genes. Root architecture adaptation to shifting local conditions is especially critical for maize to thrive in dynamic environments[129]. The regulatory effects of single genes on lateral root density that we have discovered here might contribute to viability of maize under varying water availability. Lateral roots take up the majority of water at the seedling stage[29] and gene expression complementation[113] and higher lateral root density were suggested to be beneficial upon water stress[32].

5.8 Gene regulation of non-syntenic and syntenic paralogs are connected

We investigated one of the non-syntenic *trans*-regulated SPE genes in more detail. The expression of the candidate gene TSG 2 (Zm00001eb339600) was associated with lateral root density in hybrids. We identified a syntenic paralog of this gene, which in contrast to TSG 2 is *cis*-regulated. Unexpectedly, both genes are regulated from the same eQTL position (chapter 3). While activity of TSG 2 depends on the Mo17 allele at the eQTL position, the *cis*-regulated syntenic paralog was expressed across all genotypes. These findings lead us to propose a regulatory connection between the *trans*-regulated non-syntenic gene and the syntenic paralog, present in the Mo17 line. This notion is supported by previous findings of regulatory interactions between paralogous genes, such as the co-regulation of *rcts* and *rctl* by the same transcription factor[130]. Additionally, it was proposed that non-syntenic genes were brought under the regulatory regime of the syntenic gene *rtcs* which contributed to the development of maize seminal roots[131]. We observed a high proportion of *trans*-regulating eQTL among non-syntenic SPE pattern genes, especially among those associated with the Mo17 allele. A regulatory influence on *trans*-regulated non-syntenic genes by their syntenic *cis*-regulated paralogs which is retained in the Mo17 genotype would explain this finding. Such connections would further strengthen the notion that regulatory differences between parental genotypes are translating phylogenetic distance into heterosis.

5.9 Implications for hybrid breeding

The two inbred lines B73 and Mo17 had contrasting breeding objectives and originated from different breeding pools. The B73 line belongs to the Iowa stiff stalk synthetic/BSSS group, typically used as female parent, while Mo17 belongs to the Lancaster group and is a typical male parent[74]. Those differences might have unintentionally resulted in the observed differences in regulatory preference (chapter 3, 4). It is possible that due to selection, Mo17 retained the activity of many *trans*-regulated non-syntenic genes that are not active in B73 (chapter 3). Non-syntenic genes have been associated with adaptation to stress or disease resistance[65,113] and our results indicate that they might be connected to some extend to their syntenic *cis*-regulated paralogs, increasing the possibilities for transcriptional activity in the resulting hybrids. SPE probably translates genetic diversity into phenotypic heterosis[111], a notion that is further supported by this study. Therefore, large proportions of the Mo17 parental contribution to the hybrid would be the activation of these *trans*-regulated non-syntenic genes, which are likely involved in the adaptability of hybrids. This is in line with the general properties of the two inbred lines. First, B73 outperforms Mo17 in cob and grain yield-related traits[132]. Second, B73 was bred for synchronous flowering and valued for its outstanding hybrid performance but is susceptible to the European corn borer. In contrast, Mo17 is resistant to leaf diseases, smut and common rust[74]. Moreover, commercial maize lines of the B73 heterotic subgroup were likely selected towards the B73 founder alleles[75]. It is therefore likely that in addition to phenotypic traits, regulatory preferences were passed on within this heterotic subgroup as well, influencing contemporary hybrid maize breeding pools.

5.10 Heterozygosity is necessary for expression pattern regulation

A key finding of our study is the strong association between heterozygosity and the number of SPE (chapter 3) and non-additively expressed genes (chapter 4). These in turn can explain large proportions of heterotic variance. It was discovered, based on genetic markers in rice, that heterozygosity is not strongly associated with heterosis and that any association is highly trait and background specific[133]. The relatively low association between heterosis and heterozygosity was suggested to reflect the complexity of heterosis[133]. The indirect influence of heterozygosity on heterosis via SPE and non-additive expression patterns is one aspect of this complexity and likely a reason for the limited direct association between heterozygosity and heterosis (observed also in this study, data not shown).

Our findings on the regulation of expression patterns in hybrids advance the understanding of heterosis. While expression complementation of SPE and non-additive genes follows the dominance model, their regulation, and therefore their occurrence, is based on heterozygous eQTL. The dominance model explains heterosis by complementation of alleles at different loci

throughout the genome[88,89]. Other interpretations, such as the overdominance model, associate heterosis with the presence of heterozygosity itself[90]. Our observations can also be described on the molecular level: Differential expression between the parents is caused by differences in their regulation. In case of a combination of their regulatory alleles, expression either adopts the mid-parent level or a level above the mid-parental expression. These findings underscore, that different mechanisms to explain heterosis are not mutually exclusive, nor can any single mechanism fully explain heterosis[72]. Instead, different mechanisms influence heterosis on different levels.

5.11 Conclusion and outlook

The aim of this thesis was to understand how heterosis-associated gene expression patterns are regulated. One of the key findings was that non-additive and single parent expression (SPE) pattern genes can explain large proportions of heterotic variance for root phenotypic traits. Remarkably, eQTL regulating those patterns are located almost exclusively within heterozygous regions, highlighting the important but indirect influence of heterozygosity on heterosis.

We further argued that in addition to trait- and tissue-specific genes associated with heterosis, distinct population-specific gene sets are responsible for heterosis. In our study, we also observed a difference in regulation between B73 (predominantly *cis*) and Mo17 (predominantly *trans*) associated gene expression patterns. Whether such differences in regulation are also observed for other parental combinations and whether the accumulation of *trans*-regulation in Mo17 is caused by co-regulation of evolutionary younger non-syntenic with their older syntenic counterparts will be of interest for future research. Finally, we identified an SPE gene that regulates lateral root density in hybrids. Notably, the activity of this gene depends on the presence of a Mo17 allele in the eQTL that regulates it, thereby supporting the notion that the genetic constitution of distantly located regulatory elements play a pivotal role for the function of heterosis-associated genes.

Better understanding of heterosis is of outstanding importance. The increasing prevalance of hybrids on global markets and the application of hybrid breeding to various crops has a wide range of impacts on agricultural practices as well as socio-economic implications. Those range from their association with high input agriculture to increased food security and breeding advantages. Therefore, detailed knowledge on heterosis is of utmost concern to decision makers. Nevertheless, many aspects of this phenomenon remain largely elusive. Transcriptomic knowledge, as provided in this study contributes to closing the knowledge gap on the molecular basis of heterosis and the regulation of heterosis-associated gene expression. We emphasized the effects of *trans*-regulatory eQTL on single candidate genes

as well as regulatory preferences of parental genotypes. Both can be directly considered in future hybrid breeding and research. In addition, large proportions of current parental breeding pools are derivatives of the B73 line and the B73xMo17 hybrid is one of the most prominent hybrids used in heterosis research. Therefore, our results can also be applied to other heterosis studies and will probably lead to more research on the regulation of expression between inbred lines of different heterotic patterns and especially expression complementation.

6. References

1. **FAOSTAT** (2024). https://www.fao.org/faostat/en/#data/QCL/visualize (visited on 07.03.2024).
2. **Erenstein O, Jaleta M, Sonder K, Mottaleb K, Prasanna BM** (2022). Global maize production, consumption and trade: trends and R&D implications. *Food Security* **14,** 1295–1319; 10.1007/s12571-022-01288-7.
3. **Food and Agriculture Organization of the United Nations (FAO)** (2019). *World food and agriculture - statistical pocketbook 2019.* Rome: FAO.
4. **Food and Agriculture Organization of the United Nations (FAO)** (2024). *Agricultural production statistics 2010–2023.* FAOSTAT Analytical Briefs, 96. Rome.
5. **Federal Ministry of Food and Agriculture (BMEL)** (2019). 2035 Arable farming strategy.
6. **Deutsches Maiskomitee e.V.** (2020). Maize – a key crop for the implementation of the arable farming strategy 2035. https://www.maiskomitee.de/Downloadcenter (visited on 11.02.2025).
7. **Mekonnen MM & Hoekstra AY** (2011). The green, blue and grey water footprint of crops and derived crop products. *Hydrology and Earth System Sciences* **15,** 1577–1600; 10.5194/hess-15-1577-2011.
8. **Ellis RH, Summerfield RJ, Edmeades GO, Roberts EH** (1992). Photoperiod, temperature, and the interval from sowing to tassel initiation in diverse cultivars of maize. *Crop Science* **32,** 1225–1232; 10.2135/cropsci1992.0011183X003200050033x.
9. **Hake S & Ross-Ibarra J** (2015). Genetic, evolutionary and plant breeding insights from the domestication of maize. *eLife* **4**; 10.7554/eLife.05861.
10. **Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez G J, Buckler E, Doebley J** (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 6080–6084; 10.1073/pnas.052125199.
11. **Kellogg EA** (2001). Evolutionary history of the grasses. *Plant Physiology* **125,** 1198–1205; 10.1104/pp.125.3.1198.
12. **Strable J & Scanlon MJ** (2009). Maize *(Zea mays*): a model organism for basic and applied research in plant biology. *Cold Spring Harbor Protocols* **2009,** pdb.emo132; 10.1101/pdb.emo132.
13. **Nannas NJ & Dawe RK** (2015). Genetic and genomic toolbox of *Zea mays*. *Genetics* **199,** 655–669; 10.1534/genetics.114.165183.
14. **Duvick DN** (2005). The contribution of breeding to yield advances in maize (*Zea mays* L.) **86,** 83–145; 10.1016/S0065-2113(05)86002-X.
15. **McClintock B** (1950). The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America* **36,** 344–355; 10.1073/pnas.36.6.344.
16. **Marcon C, Altrogge L, Win YN, Stöcker T, Gardiner JM, Portwood JL, Opitz N, Kortz A, Baldauf JA, Hunter CT** *et al.* (2020). BonnMu: a sequence-indexed resource of transposon-induced maize mutations for functional genomics studies. *Plant Physiology* **184,** 620–631; 10.1104/pp.20.00478.
17. **Liu P, McCarty DR, Koch KE** (2016). Transposon mutagenesis and analysis of mutants in UniformMu maize (*Zea mays*). *Current Protocols in Plant Biology* **1,** 451–465; 10.1002/cppb.20029.
18. **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 Physiology* **181,** 1404–1414; 10.1104/pp.19.00894.

19. **Lee M, Sharopova N, Beavis WD, Grant D, Katt M, Blair D, Hallauer A** (2002). Expanding the genetic map of maize with the intermated B73 x Mo17 (IBM) population. *Plant Molecular Biology* **48,** 453–461; 10.1023/a:1014893521186.

20. **Dell'Acqua M, Gatti DM, Pea G, Cattonaro F, Coppens F, Magris G, Hlaing AL, Aung HH, Nelissen H, Baute J** *et al.* (2015). Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. *Genome Biology* **16,** 167; 10.1186/s13059-015-0716-z.

21. **Yu J, Holland JB, McMullen MD, Buckler ES** (2008). Genetic design and statistical power of nested association mapping in maize. *Genetics* **178,** 539–551; 10.1534/genetics.107.074245.

22. **Butrón A, Santiago R, Gowda M** (2023). Editorial: model organisms in plant science: maize. *Frontiers in Plant Science* **14,** 1147857; 10.3389/fpls.2023.1147857.

23. **Lynch JP** (2019). Root phenotypes for improved nutrient capture: an underexploited opportunity for global agriculture. *New Phytologist* **223,** 548–564; 10.1111/nph.15738.

24. **Lynch JP** (2007). Roots of the second green revolution. *Australian Journal of Botany* **55,** 493; 10.1071/BT06118.

25. **Rogers ED & Benfey PN** (2014). Regulation of plant root system architecture: implications for crop advancement. *Current Opinion in Biotechnology*.

26. **G Viana W, Scharwies JD, Dinneny JR** (2022). Deconstructing the root system of grasses through an exploration of development, anatomy and function. *Plant, Cell & Environment* **45,** 602–619; 10.1111/pce.14270.

27. **Hochholdinger F & Tuberosa R** (2009). Genetic and genomic dissection of maize root development and architecture. *Current Opinion in Plant Biology* **12,** 172–177; 10.1016/j.pbi.2008.12.002.

28. **Hochholdinger F, Park WJ, Sauer M, Woll K** (2004). From weeds to crops: genetic analysis of root development in cereals. *Trends in Plant Science* **9,** 42–48; 10.1016/j.tplants.2003.11.003.

29. **Ahmed MA, Zarebanadkouki M, Kaestner A, Carminati A** (2016). Measurements of water uptake of maize roots: the key function of lateral roots. *Plant and Soil* **398,** 59–77; 10.1007/s11104-015-2639-6.

30. **Ahmed MA, Zarebanadkouki M, Meunier F, Javaux M, Kaestner A, Carminati A** (2018). Root type matters: measurement of water uptake by seminal, crown, and lateral roots in maize. *Journal of Experimental Botany* **69,** 1199–1206; 10.1093/jxb/erx439.

31. **Protto V, Bauget F, Rishmawi L, Nacry P, Maurel C** (2024). Primary, seminal and lateral roots of maize show type-specific growth and hydraulic responses to water deficit. *Plant Physiology* **194,** 2564–2579; 10.1093/plphys/kiad675.

32. **Yu P, Li C, Li M, He X, Wang D, Li H, Marcon C, Li Y, Perez-Limón S, Chen X** *et al.* (2024). Seedling root system adaptation to water availability during maize domestication and global expansion. *Nature Genetics* **56,** 1245–1256; 10.1038/s41588-024-01761-3.

33. **Michl K, Berg G, Cernava T** (2023). The microbiome of cereal plants: the current state of knowledge and the potential for future applications. *Environmental microbiome* **18,** 28; 10.1186/s40793-023-00484-y.

34. **Shull GH** (1914). Duplicate genes for capsule-form in *Bursa bursa-pastoris*. *Zeitschrift für Induktive Abstammungs- und Vererbungslehre* **12,** 97–149; 10.1007/bf01837282.

35. **East EM & Jones DF** (1919). *Inbreeding and outbreeding: their genetic and sociological significance.* Philadelphia: J. B. Lippincott Company.

36. **Paril J, Reif J, Fournier-Level A, Pourkheirandish M** (2024). Heterosis in crop improvement. *The Plant Journal* **117,** 23–32; 10.1111/tpj.16488.

37. **Hochholdinger F & Baldauf JA** (2018). Heterosis in plants. *Current Biology* **28,** R1089-R1092; 10.1016/j.cub.2018.06.041.

38. **Flint-Garcia SA, Buckler ES, Tiffin P, Ersoz E, Springer NM** (2009). Heterosis is prevalent for multiple traits in diverse maize germplasm. *PLOS ONE* **4,** e7433; 10.1371/journal.pone.0007433.

39. **Hoecker N, Keller B, Piepho H-P, Hochholdinger F** (2006). Manifestation of heterosis during early maize (*Zea mays* L.) root development. *Theoretical and Applied Genetics* **112,** 421–429; 10.1007/s00122-005-0139-4.

40. **Chairi F, Elazab A, Sanchez-Bragado R, Araus JL, Serret MD** (2016). Heterosis for water status in maize seedlings. *Agricultural Water Management* **164,** 100–109; 10.1016/j.agwat.2015.08.005.

41. **Kamphorst SH, Amaral Júnior, Antônio Teixeira do, Vergara-Diaz O, Gracia-Romero A, Fernandez-Gallego JA, Chang-Espino MC, Buchaillot ML, Rezzouk FZ, Lima VJ** de, **Serret MD** *et al.* (2022). Heterosis and reciprocal effects for physiological and morphological traits of popcorn plants under different water conditions. *Agricultural Water Management* **261,** 107371; 10.1016/j.agwat.2021.107371.

42. **Wang Z, Ni Z, Wu H, Nie X, Sun Q** (2006). Heterosis in root development and differential gene expression between hybrids and their parental inbreds in wheat (*Triticum aestivum* L.). *Theoretical and applied genetics* **113,** 1283–1294; 10.1007/s00122-006-0382-3.

43. **Zhai R, Feng Y, Wang H, Zhan X, Shen X, Wu W, Zhang Y, Chen D, Dai G, Yang Z** *et al.* (2013). Transcriptome analysis of rice root heterosis by RNA-Seq. *BMC Genomics* **14,** 19; 10.1186/1471-2164-14-19.

44. **Sharma S, DeMason DA, Ehdaie B, Lukaszewski AJ, Waines JG** (2010). Dosage effect of the short arm of chromosome 1 of rye on root morphology and anatomy in bread wheat. *Journal of Experimental Botany* **61,** 2623–2633; 10.1093/jxb/erq097.

45. **Morris ML, Risopoulos J, Beck D** (1999). *Genetic change in farmer-recycled maize seed: a review of the evidence.* CIMMYT Economics Working Paper No. 99-07. Mexico: CYMMIT.

46. **Chase SS & Nanda DK** (1969). Rapid inbreeding in maize. *Economic Botany* **23,** 165–173; 10.1007/BF02860622.

47. **Schulthess, A.W., Zhao, Y., Reif, J.C.** Genomic selection in hybrid breeding. In: *Genomic Selection in Hybrid Breeding,* pp. 149–183.

48. **El-Namaky RA & Demont M** (2013). Hybrid rice in Africa: challenges and prospects. In: *Realizing Africa's rice promise.* 1st ed. UK: CABI, pp. 173–178.

49. **Oppong Abebrese S & Yeboah A** (2021). Hybrid rice in Africa: progress, prospects, and challenges. In: *Recent Advances in Rice Research,* edited by M. Rahman Ansari: IntechOpen.

50. **Byerlee D** (2020). The globalization of hybrid maize, 1921–70. *Journal of Global History* **15,** 101–122; 10.1017/S1740022819000354.

51. **Duvick DN** (2001). Biotechnology in the 1930s: the development of hybrid maize. *Nature Reviews Genetics* **2,** 69–74; 10.1038/35047587.

52. **Xie F & Hardy B.** Accelerating hybrid rice development. International Rice Research Institute (IRRI), 2009.

53. **Erickson LR & Atnaseo C** (2011). 4.10 - Transgenic crops with producer-oriented traits: development, application, and impact. In: *Comprehensive Biotechnology,* edited by M. Moo-Young. 2nd ed. Burlington: Academic Press, pp. 121–131.

54. **Buckler ES, Gaut BS, McMullen MD** (2006). Molecular and functional diversity of maize. *Current Opinion in Plant Biology* **9,** 172–176; 10.1016/j.pbi.2006.01.013.

55. **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; 10.1126/science.1178534.

56. **Hufford MB, Seetharam AS, Woodhouse MR, Chougule KM, Ou S, Liu J, Ricci WA, Guo T, Olson A, Qiu Y** *et al.* (2021). De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373,** 655–662; 10.1126/science.abg5289.

57. **Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin C-S** *et al.* (2017). Improved maize reference genome with single-molecule technologies. *Nature* **546,** 524–527; 10.1038/nature22971.

58. **Law M, Childs KL, Campbell MS, Stein JC, Olson AJ, Holt C, Panchy N, Lei J, Jiao D, Andorf CM** *et al.* (2015). Automated update, revision, and quality control of the maize genome annotations using MAKER-P improves the B73 RefGen_v3 gene models and identifies new genes. *Plant Physiology* **167,** 25–39; 10.1104/pp.114.245027.

59. **Swigonová Z, Lai J, Ma J, Ramakrishna W, Llaca V, Bennetzen JL, Messing J** (2004). Close split of sorghum and maize genome progenitors. *Genome research* **14,** 1916–1923; 10.1101/gr.2332504.

60. **Wei F, Coe E, Nelson W, Bharti AK, Engler F, Butler E, Kim H, Goicoechea JL, Chen M, Lee S** *et al.* (2007). Physical and genetic structure of the maize genome reflects its complex evolutionary history. *PLOS Genetics* **3,** e123; 10.1371/journal.pgen.0030123.

61. **Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M** (2010). Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. *PLOS Biology* **8,** e1000409; 10.1371/journal.pbio.1000409.

62. **Schnable JC, Springer NM, Freeling M** (2011). Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proceedings of the National Academy of Sciences of the United States of America* **108,** 4069–4074; 10.1073/pnas.1101368108.

63. **Schnable JC** (2015). Genome evolution in maize: from genomes back to genes. *Annual review of plant biology* **66,** 329–343; 10.1146/annurev-arplant-043014-115604.

64. **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; 10.1371/journal.pone.0017855.

65. **Salvi S** (2017). An evo-devo perspective on root genetic variation in cereals. *Journal of Experimental Botany* **68,** 351–354; 10.1093/jxb/erw505.

66. **Brohammer AB, Kono TJY, Springer NM, McGaugh SE, Hirsch CN** (2018). The limited role of differential fractionation in genome content variation and function in maize (*Zea mays* L.) inbred lines. *The Plant Journal* **93,** 131–141; 10.1111/tpj.13765.

67. **Wang B, Hou M, Shi J, Ku L, Song W, Li C, Ning Q, Li X, Li C, Zhao B** *et al.* (2023). De novo genome assembly and analyses of 12 founder inbred lines provide insights into maize heterosis. *Nature Genetics* **55,** 312–323; 10.1038/s41588-022-01283-w.

68. **East EM** (1936). Heterosis. *Genetics* **21,** 375–397; 10.1093/genetics/21.4.375.

69. **Reif JC, Hallauer AR, Melchinger AE** (2005). Heterosis and heterotic patterns in maize. *Maydica* **50,** 215–223.

70. **Melchinger AE & Gumber RK** (1998). Overview of heterosis and heterotic groups in agronomic crops. *Concepts and Breeding of Heterosis,* 29–44; 10.2135/cssaspecpub25.c3.

71. **Beukert U, Li Z, Liu G, Zhao Y, Ramachandra N, Mirdita V, Pita F, Pillen K, Reif JC** (2017). Genome-based identification of heterotic patterns in rice. *Rice* **10,** 22; 10.1186/s12284-017-0163-4.

72. **Schnable PS & Springer NM** (2013). Progress toward understanding heterosis in crop plants. *Annual review of plant biology* **64,** 71–88; 10.1146/annurev-arplant-042110-103827.

73. **Mikel MA & Dudley JW** (2006). Evolution of north american dent corn from public to proprietary germplasm. *Crop Science* **46,** 1193–1205; 10.2135/cropsci2005.10-0371.

74. **Troyer F** (2001). Temperate corn - background, behavior, and breeding. In: *Specialty corns,* edited by A. R. Hallauer. 2nd ed.: CRC Press.

75. **White MR, Mikel MA, Leon N** de, **Kaeppler SM** (2020). Diversity and heterotic patterns in North American proprietary dent maize germplasm. *Crop Science* **60,** 100–114; 10.1002/csc2.20050.

76. **Mikel MA** (2008). Genetic diversity and improvement of contemporary proprietary North American dent corn. *Crop Science* **48,** 1686–1695; 10.2135/cropsci2008.01.0039.

77. **Sun S, Zhou Y, Chen J, Shi J, Zhao H, Zhao H, Song W, Zhang M, Cui Y, Dong X** *et al.* (2018). Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nature Genetics* **50,** 1289–1295; 10.1038/s41588-018-0182-0.

78. **Broman KW** (2005). The genomes of recombinant inbred lines. *Genetics* **169,** 1133–1146; 10.1534/genetics.104.035212.

79. **Liu SC, Kowalski SP, Lan TH, Feldmann KA, Paterson AH** (1996). Genome-wide high-resolution mapping by recurrent intermating using *Arabidopsis thaliana* as a model. *Genetics* **142,** 247–258; 10.1093/genetics/142.1.247.

80. **Huo X, Wang J, Zhang L** (2023). Combined QTL mapping on bi-parental immortalized heterozygous populations to detect the genetic architecture on heterosis. *Frontiers in Plant Science* **14,** 1157778; 10.3389/fpls.2023.1157778.

81. **Rahman H, Pekic S, Lazic-Jancic V, Quarrie SA, Shah SMA, Pervez A, Shah MM** (2011). Molecular mapping of quantitative trait loci for drought tolerance in maize plants. *Genetics and Molecular Research* **10,** 889–901; 10.4238/vol10-2gmr1139.

82. **Pan Q, Xu Y, Li K, Peng Y, Zhan W, Li W, Li L, Yan J** (2017). The genetic basis of plant architecture in 10 maize recombinant inbred line populations. *Plant Physiology* **175,** 858–873; 10.1104/pp.17.00709.

83. **Yang H, Zhang Z, Zhang N, Li T, Wang J, Zhang Q, Xue J, Zhu W, Xu S** (2024). QTL mapping for plant height and ear height using bi-parental immortalized heterozygous populations in maize. *Frontiers in Plant Science* **15,** 1371394; 10.3389/fpls.2024.1371394.

84. **Jansen RC & Nap JP** (2001). Genetical genomics: the added value from segregation. *Trends in Genetics* **17,** 388–391; 10.1016/S0168-9525(01)02310-1.

85. **Torres-Rodríguez JV, Li D, Schnable JC** (2024). Evolving best practices for transcriptome-wide association studies accelerate discovery of gene-phenotype links. *Current Opinion in Plant Biology* **83,** 102670; 10.1016/j.pbi.2024.102670.

86. **Hochholdinger F & Yu P** (2025). Molecular concepts to explain heterosis in crops. *Trends in Plant Science* **30**; 10.1016/j.tplants.2024.07.018.

87. **Baldauf JA & Hochholdinger F** (2023). Molecular dissection of heterosis in cereal roots and their rhizosphere. *Theoretical and Applied Genetics* **136,** 173; 10.1007/s00122-023-04419-6.

88. **Jones DF** (1917). Dominance of linked factors as a means of accounting for heterosis. *Proceedings of the National Academy of Sciences of the United States of America* **3,** 310–312; 10.1073/pnas.3.4.310.

89. **Davenport CB** (1908). Degeneration, albinism and inbreeding. *Science* **28,** 454–455; 10.1126/science.28.718.454.c.

90. **Shull GH** (1908). The composition of a field of maize. *Journal of Heredity* **os-4,** 296–301; 10.1093/jhered/os-4.1.296.

91. **Lin T, Zhou C, Chen G, Yu J, Wu W, Ge Y, Liu X, Li J, Jiang X, Tang W** *et al.* (2020). Heterosis-associated genes confer high yield in super hybrid rice. *Theoretical and Applied Genetics* **133,** 3287–3297; 10.1007/s00122-020-03669-y.

92. **Krieger U, Lippman ZB, Zamir D** (2010). The flowering gene SINGLE FLOWER TRUSS drives heterosis for yield in tomato. *Nature Genetics* **42,** 459–463; 10.1038/ng.550.

93. **Richey FD** (1942). Mock-dominance and hybrid vigor. *Science* **96**, 280-281; 10.1126/science.96.2490.280

94. **Birchler JA, Riddle NC, Auger DL, Veitia RA** (2005). Dosage balance in gene regulation: biological implications. *Trends in Genetics* **21,** 219–226; 10.1016/j.tig.2005.02.010.

95. **Birchler JA & Veitia RA** (2010). The gene balance hypothesis: implications for gene regulation, quantitative traits and evolution. *New Phytologist* **186,** 54–62; 10.1111/j.1469-8137.2009.03087.x.

96. **Springer NM & Stupar RM** (2007). Allelic variation and heterosis in maize: how do two halves make more than a whole? *Genome Research* **17,** 264–275; 10.1101/gr.5347007.

97. **Guo M, Rupe MA, Yang X, Crasta O, Zinselmeier C, Smith OS, Bowen B** (2006). Genome-wide transcript analysis of maize hybrids: allelic additive gene expression and yield heterosis. *Theoretical and applied genetics* **113,** 831–845; 10.1007/s00122-006-0335-x.

98. **Wang Z, Gerstein M, Snyder M** (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10,** 57–63; 10.1038/nrg2484.

99. **Tyagi P, Singh D, Mathur S, Singh A, Ranjan R** (2022). Upcoming progress of transcriptomics studies on plants: An overview. *Frontiers in Plant Science* **13,** 1030890; 10.3389/fpls.2022.1030890.

100. **Swanson-Wagner RA, Jia Y, DeCook R, Borsuk LA, Nettleton D, Schnable PS** (2006). All possible modes of gene action are observed in a global comparison of gene expression in a maize F_1 hybrid and its inbred parents. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 6805–6810; 10.1073/pnas.0510430103.

101. **Stupar RM & Springer NM** (2006). *Cis*-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F_1 hybrid. *Genetics* **173,** 2199–2210; 10.1534/genetics.106.060699.

102. **Paschold A, Jia Y, Marcon C, Lund S, Larson NB, Yeh C-T, Ossowski S, Lanz C, Nettleton D, Schnable PS** *et al.* (2012). Complementation contributes to transcriptome complexity in maize (*Zea mays L.*) hybrids relative to their inbred parents. *Genome Research* **22,** 2445–2454; 10.1101/gr.138461.112.

103. **Song G, Guo Z, Liu Z, Cheng Q, Qu X, Chen R, Jiang D, Liu C, Wang W, Sun Y** *et al.* (2013). Global RNA sequencing reveals that genotype-dependent allele-specific expression contributes to differential expression in rice F_1 hybrids. *BMC Plant Biology* **13,** 221; 10.1186/1471-2229-13-221.

104. **Baldauf JA, Marcon C, Paschold A, Hochholdinger F** (2016). Nonsyntenic genes drive tissue-specific dynamics of differential, nonadditive, and allelic expression patterns in maize hybrids. *Plant Physiology* **171,** 1144–1155; 10.1104/pp.16.00262.

105. **Hoecker N, Keller B, Muthreich N, Chollet D, Descombes P, Piepho H-P, Hochholdinger F** (2008). Comparison of maize (*Zea mays* L.) F_1-hybrid and parental inbred line primary root transcriptomes suggests organ-specific patterns of nonadditive gene expression and conserved expression trends. *Genetics* **179,** 1275–1283; 10.1534/genetics.108.088278.

106. **Zhang X, Hirsch CN, Sekhon RS, Leon N** de, **Kaeppler SM** (2016). Evidence for maternal control of seed size in maize from phenotypic and transcriptional analysis. *Journal of Experimental Botany* **67,** 1907–1917; 10.1093/jxb/erw006.

107. **Uzarowska A, Keller B, Piepho H-P, Schwarz G, Ingvardsen C, Wenzel G, Lübberstedt T** (2007). Comparative expression profiling in meristems of inbred-hybrid triplets of maize based on morphological investigations of heterosis for plant height. *Plant Molecular Biology* **63,** 21–34; 10.1007/s11103-006-9069-z.

108. **Botet R & Keurentjes JJB** (2020). The role of transcriptional regulation in hybrid vigor. *Frontiers in Plant Science* **11,** 410; 10.3389/fpls.2020.00410.

109. **Springer NM & Stupar RM** (2007). Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. *The Plant Cell* **19,** 2391–2402; 10.1105/tpc.107.052258.

110. **Baldauf JA, Marcon C, Lithio A, Vedder L, Altrogge L, Piepho H-P, Schoof H, Nettleton D, Hochholdinger F** (2018). Single-parent expression is a general mechanism driving extensive complementation of non-syntenic genes in maize hybrids. *Current Biology* **28,** 431-437.e4; 10.1016/j.cub.2017.12.027.

111. **Baldauf JA, Liu M, Vedder L, Yu P, Piepho H-P, Schoof H, Nettleton D, Hochholdinger F** (2022). Single-parent expression complementation contributes to phenotypic heterosis in maize hybrids. *Plant Physiology* **189,** 1625–1638; 10.1093/plphys/kiac180.

112. **Paschold A, Larson NB, Marcon C, Schnable JC, Yeh C-T, Lanz C, Nettleton D, Piepho H-P, Schnable PS, Hochholdinger F** (2014). Nonsyntenic genes drive highly dynamic complementation of gene expression in maize hybrids. *The Plant Cell* **26,** 3939–3948; 10.1105/tpc.114.130948.

113. **Marcon C, Paschold A, Malik WA, Lithio A, Baldauf JA, Altrogge L, Opitz N, Lanz C, Schoof H, Nettleton D** *et al.* (2017). Stability of single-parent gene expression complementation in maize hybrids upon water deficit stress. *Plant Physiology* **173,** 1247–1257; 10.1104/pp.16.01045.

114. **Li Z, Zhou P, Della Coletta R, Zhang T, Brohammer AB, H O'Connor C, Vaillancourt B, Lipzen A, Daum C, Barry K** *et al.* (2021). Single-parent expression drives dynamic gene expression complementation in maize hybrids. *The Plant Journal* **105,** 93–107; 10.1111/tpj.15042.

115. **Swanson-Wagner RA, DeCook R, Jia Y, Bancroft T, Ji T, Zhao X, Nettleton D, Schnable PS** (2009). Paternal dominance of *trans*-eQTL influences gene expression patterns in maize hybrids. *Science* **326,** 1118–1120; 10.1126/science.1178294.

116. **Marcon C, Schützenmeister A, Schütz W, Madlung J, Piepho H-P, Hochholdinger F** (2010). Nonadditive protein accumulation patterns in Maize (*Zea mays L*.) hybrids during embryo development. *Journal of Proteome Research* **9,** 6511–6522; 10.1021/pr100718d.

117. **Rehman Au, Dang T, Qamar S, Ilyas A, Fatema R, Kafle M, Hussain Z, Masood S, Iqbal S, Shahzad K** (2021). Revisiting plant heterosis - from field scale to molecules. *Genes* **12,** 1688; 10.3390/genes12111688.

118. **Hochholdinger F & Yu P** (2025). Molecular concepts to explain heterosis in crops. *Trends in Plant Science* **30,** 95–104; 10.1016/j.tplants.2024.07.018.

119. **Stupar RM & Springer NM** (2006). *Cis*-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F_1 hybrid. *Genetics* **173,** 2199–2210; 10.1534/genetics.106.060699.

120. **Stupar RM, Gardiner JM, Oldre AG, Haun WJ, Chandler VL, Springer NM** (2008). Gene expression analyses in maize inbreds and hybrids with varying levels of heterosis. *BMC Plant Biology* **8,** 33; 10.1186/1471-2229-8-33.

121. **Pan Q, Ali F, Yang X, Li J, Yan J** (2012). Exploring the genetic characteristics of two recombinant inbred line populations via high-density SNP markers in maize. *PLOS ONE* **7,** e52777; 10.1371/journal.pone.0052777.

122. **Mather K** (1939). Crossing over and heterochromatin in the X chromosome of *drosophila melanogaster*. *Genetics* **24,** 413–435; 10.1093/genetics/24.3.413.

123. **Sidhu GK, Fang C, Olson MA, Falque M, Martin OC, Pawlowski WP** (2015). Recombination patterns in maize reveal limits to crossover homeostasis. *Proceedings of the National Academy of Sciences of the United States of America* **112,** 15982–15987; 10.1073/pnas.1514265112.

124. **Yoo M-J, Szadkowski E, Wendel JF** (2013). Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* **110,** 171–180; 10.1038/hdy.2012.94.

125. **Combes M-C, Hueber Y, Dereeper A, Rialle S, Herrera J-C, Lashermes P** (2015). Regulatory divergence between parental alleles determines gene expression patterns in hybrids. *Genome Biology and Evolution* **7,** 1110–1121; 10.1093/gbe/evv057.

126. **Li L, Petsch K, Shimizu R, Liu S, Xu WW, Ying K, Yu J, Scanlon MJ, Schnable PS, Timmermans MCP** *et al.* (2013). Mendelian and non-mendelian regulation of gene expression in maize. *PLOS Genetics* **9,** e1003202; 10.1371/journal.pgen.1003202.

127. **Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G** *et al.* (2003). Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422,** 297–302; 10.1038/nature01434.

128. **Schnable JC & Lyons E** (2011). Comparative genomics with maize and other grasses: from genes to genomes! *Maydica* **56**.

129. **Bishopp A & Lynch JP** (2015). The hidden half of crop yields. *Nature Plants* **1,** 15117; 10.1038/nplants.2015.117.

130. **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 Phytologist* **207,** 1123–1133; 10.1111/nph.13420.

131. **Tai H, Opitz N, Lithio A, Lu X, Nettleton D, Hochholdinger F** (2017). Non-syntenic genes drive RTCS-dependent regulation of the embryo transcriptome during formation of seminal root primordia in maize (*Zea mays* L.). *Journal of Experimental Botany* **68,** 403–414; 10.1093/jxb/erw422.

132. **Jansen C, Leon N** de, **Lauter N, Hirsch C, Ruff L, Lübberstedt T** (2013). Genetic and morphometric analysis of cob architecture and biomass-related traits in the intermated B73 × Mo17 recombinant inbred lines of maize. *BioEnergy Research* **6,** 903–916; 10.1007/s12155-013-9319-2.

133. **Zhang Q, Zhou ZQ, Yang GP, Xu CG, Liu KD, Saghai Maroof MA** (1996). Molecular marker heterozygosity and hybrid performance in indica and japonica rice. *Theoretical and applied genetics* **93,** 1218–1224; 10.1007/BF0022345

7. Appendix

7.1 Supplementary information of chapter 3

Supplement Material SM1: Analysis of image-based phenotypic data and lateral root density

References listed in chapter 3

After scanning the up to 8 maize seedlings per sample, the images were adjusted to have a minimum size of 600 x 600 px by adding a black frame around smaller images, using a custom Python script. The RootPainter software client (version 0.1.0) and server component (version 0.2.7) were used to train a convolutional neural network to recognize roots in images. The training dataset is a subset of the images, generated with the following settings: a maximum of 2 tiles per image, all images, and a target height and width of 700 pixels. The corrective annotation method of the software was used to mark the roots in the training dataset and the network was trained based on this data. Subsequently, all images were segmented using the trained model and converted to black recognized roots and white background[31]. The converted images were then analyzed in a batch using RhizoVision Explorer (version 2.0.3). The software was set to analyze only the largest root component to exclude roots from neighboring plants that could not be removed by image cropping. However, this approach resulted in incomplete analysis of some images with gaps in roots. To address this issue, an additional run was performed with settings to include all root objects bigger than 60 mm^2 [32]. The images with different results were examined and the correct results were saved for further analysis. Among the measured root parameters are the total root lengths and volume of all roots and of roots specific customary diameter ranges. One customary diameter range was set to 2.5 and above, so that only the seed falls into this category. The length and volume of roots from this category were subtracted from the total root length and volume. In addition, the number of total root tips in each image was measured. Technical outliers, like images with a brown background instead of a blue one and images where the roots were covered by name tags, were removed from the data.

For measuring the lateral root density, the uppermost cm of the primary root with emerged lateral roots was collected of each seedling and stored in 80% ethanol until counting. For each root piece, the number of lateral roots per cm was counted and used as lateral root density.

Supplement Material SM2: SNP calling between B73 and Mo17 for sample evaluation

In short, the GATK HaplotypeCaller was used to identify variants between the Mo17 samples and the B73v5 reference genome. Loci where the B73 samples also showed a variant were

excluded. The resulting SNPs were compiled into a list of potential SNP loci. The aligned reads of each RNA-seq sample were then compared to this list, and the frequency of the B73 and Mo17 alleles at each SNP locus was determined ([38], adapted from[39]). The resulting allele frequencies were utilised to identify high quality samples.

In detail, to prepare the list of potential SNP loci, all Mo17 samples were combined into one file, using samtools merge and an index was created after renaming the RGSM (sample name field of readgroup information) for all samples to Mo17. The GATK HaplotypeCaller was run, setting the standard-min-confidence-threshold-for-calling to 20 and dont-use-soft-clipped-bases to true. This was done in parallel on each chromosome, by specifying intervals of whole chromosomes. The chromosome data was combined using GATK GatherVcfs, resulting in a file of Mo17 variants vs the B73v5 reference genome. The same was done for the B73 samples vs the B73v5 reference genome, because the B73 samples used in this study, might not be completely identical with the reference genome. These differences should be ignored. Therefore, the locations, where the B73 samples also show variance to the B73 genome were excluded from the list of possible SNP locations. This was done by a custom python script (Python3)[38]. The list of Mo17 variants was further filtered to contain only SNPs (no InDels) and only those SNPs which are homozygous for the SNP allele (putative Mo17) with bcftools view -i 'TYPE="snp" && GT="AA"' (from htslib version 1.14)[35].

To determine the allele frequencies of the B73 and Mo17 alleles at each of the SNP loci, all alignment files, (prepared as described in Material and Methods of chapter 3) were separately analyzed by a custom python script (Python 2.7). The script checks at each identified Mo17 SNP position, how often the reference (putative B73) allele is present in the alignments and how often the SNP allele (putative Mo17) is present at the same site. These counts were then filtered for loci within protein-coding genes. The protein-coding genes were extracted from the maize B73 annotation file (http://ftp.ensemblgenomes.org/pub/plants/release-52/gff3/zea_mays/Zea_mays.Zm-B73-REFERENCE-NAM-5.0.52.gff3.gz) via a custom python script (Python3)[38]. It had to be ascertained for the SNP loci in protein-coding genes that the reported reference and SNP allele counts correspond to the B73 and Mo17 allele of the germplasm used in this study (reference = B73, SNP = Mo17). This was done by retaining only loci, where the in Mo17 samples the SNP count is strongly predominant and in B73 samples the reference counts are strongly predominant. Therefore, the average of the Mo17 allele count and B73 allele count across the Mo17 samples was calculated. Only loci with minimum 95% Mo17 counts (and max 5% B73) on average were kept. The same was done for the B73 samples, where loci with at least 95% B73 counts were kept. The loci were further checked to have no third allele. Additionally, only those loci, where the average of the reference count in B73 samples is larger than 0.25, are kept. Considering the 48 B73 samples, that means a total count of 12 within all samples. This was done to make sure that the reference (B73) allele is

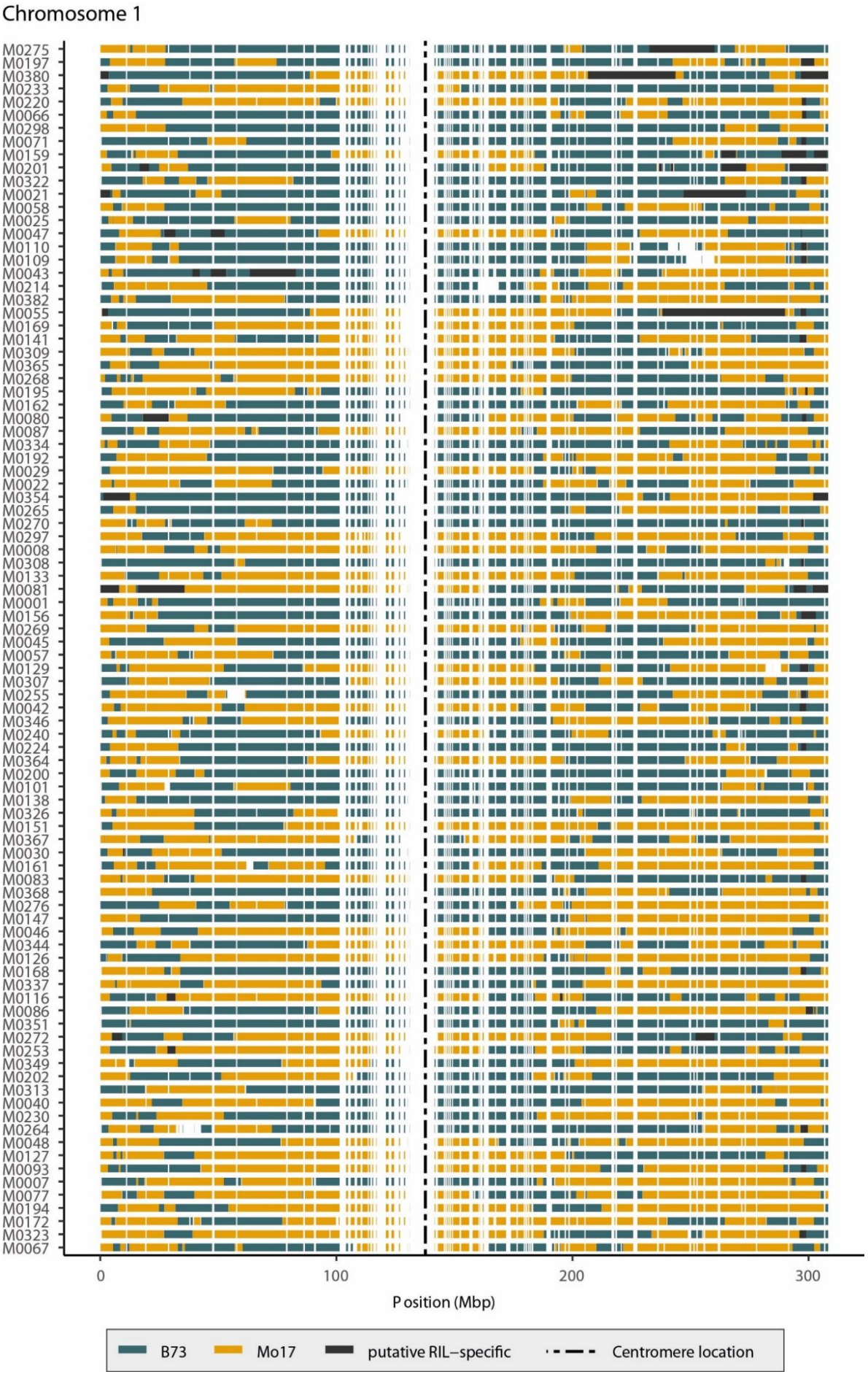
confirmed by sufficient coverage in our samples. (The SNP allele is already confirmed, because if there was no coverage for it, the SNP could not have been called).

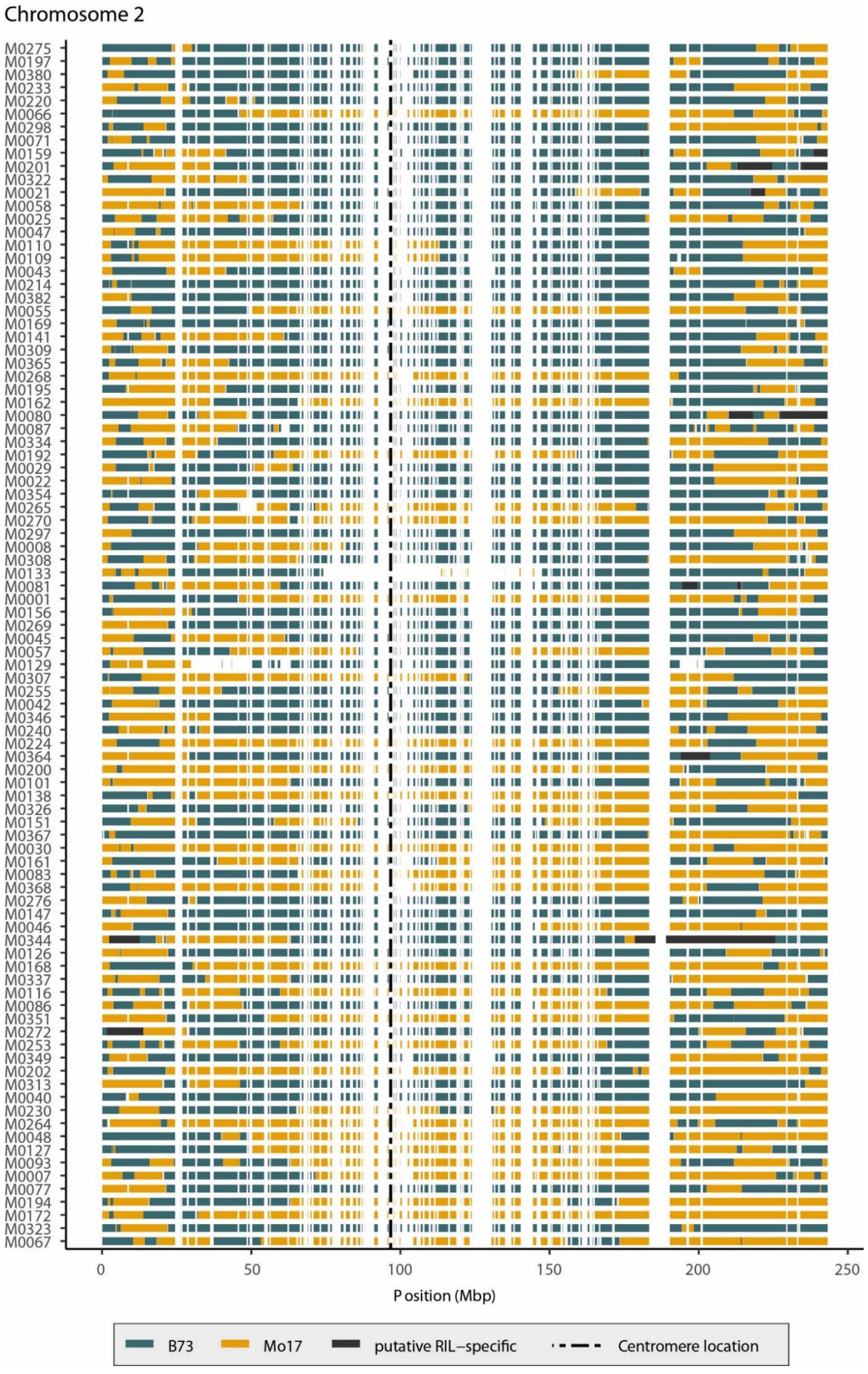
Due to some samples clustering in a different group (e.g. B73xIBM-RIL sample in Mo17xIBM-RIL group), visible in the MDS plot (Figure 1C, chapter 3), all samples were investigated to check, if their information is consistent within the three replicates of a genotype and a triplet. First, the homozygosity in the inbred IBM-RIL samples was calculated. Only loci with a minimum coverage of 0.5 counts per million (cpm) were considered. The percentage of homozygous B73 and Mo17 loci was calculated, whereas a locus was considered homozygous if 95% of all counts belong to the major allele. The sum of the percentages of homozygous B73 and Mo17 loci is calculated and is expected to be close to 100%. This was done to check whether all the pooled roots within an RNA-seq sample came from the correct genotype. Thus, samples with less than 95% homozygous loci were excluded. Furthermore, IBM-RILs with only one remaining replicate were excluded.

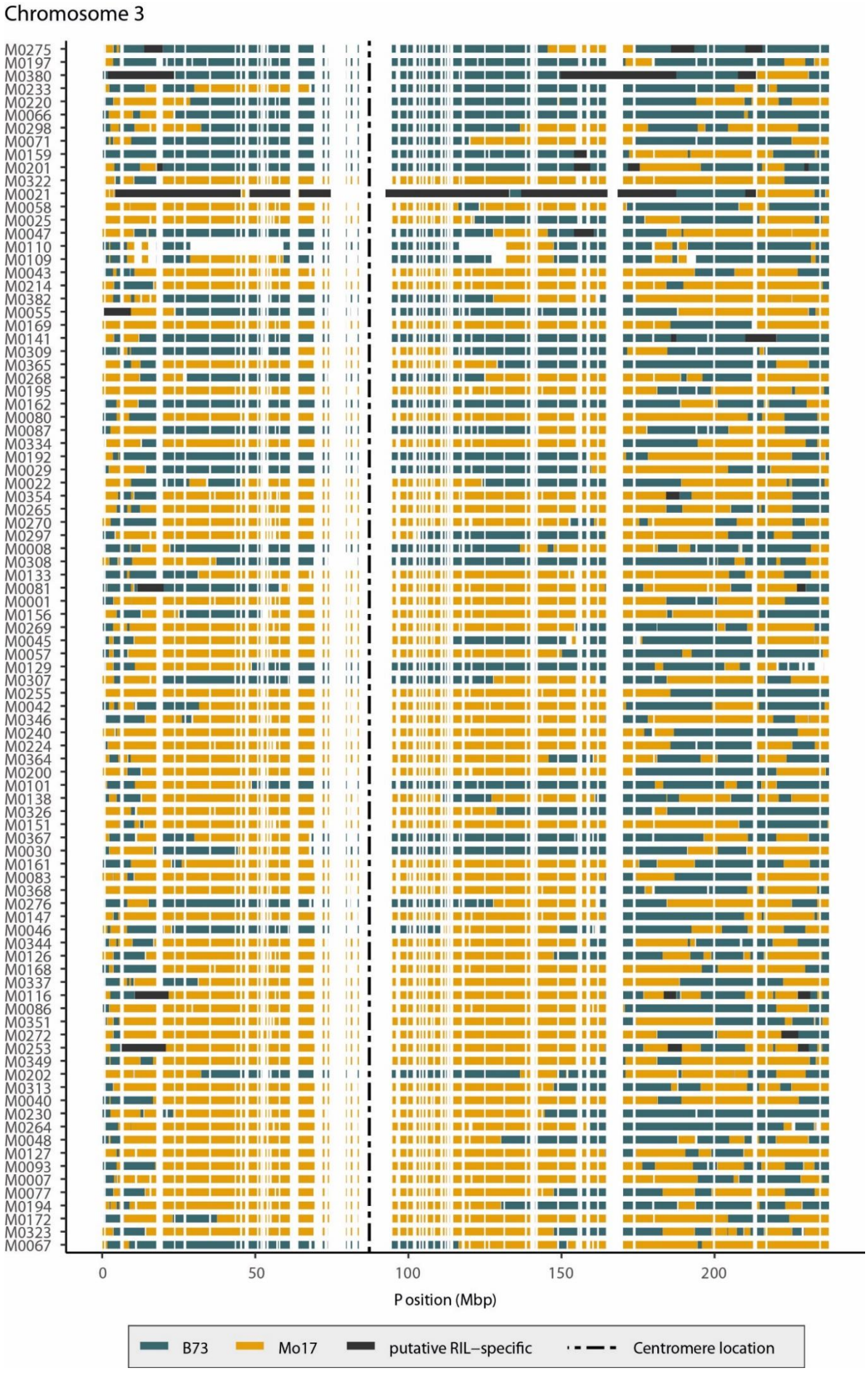
The remaining IBM-RIL samples were then compared to their B73xIBM-RIL and Mo17xRIL hybrid samples. A B73xIBM-RIL hybrid for example, should be homozygous at the loci where the IBM-RIL has the homozygous B73 allele and respectively, Mo17xIBM-RIL hybrids should be homozygous for the Mo17 allele at the loci where the IBM-RIL is homozygous Mo17. The loci were filtered for those with consistent allele information across the samples of the same IBM-RIL. Loci that were homozygous for Mo17 were selected in the Mo17xIBM-RIL hybrid of the same triplet and the homozygosity was calculated as described for the IBM-RIL samples. The same was done with B73 loci in the B73xIBM-RIL hybrids. Again, samples with less than 95% homozygous loci were excluded. In total, 175 samples were excluded, because they were not homozygous in at least 95% of supposedly homozygous loci. Beforehand, 10 samples had been excluded due to their library size of <5 million read counts. As a result, 17 samples were excluded because they were the only replicate left of the respective genotype. In our analyses, we always compared both parents and the resulting hybrid, thus 90 hybrid samples were excluded, because the respective paternal IBM-RIL samples were all excluded and 8 IBM-RIL samples were excluded, because all corresponding hybrid samples were excluded. A total of 852 samples remained for the final SNP calling of all samples against the B73v5 reference genome.

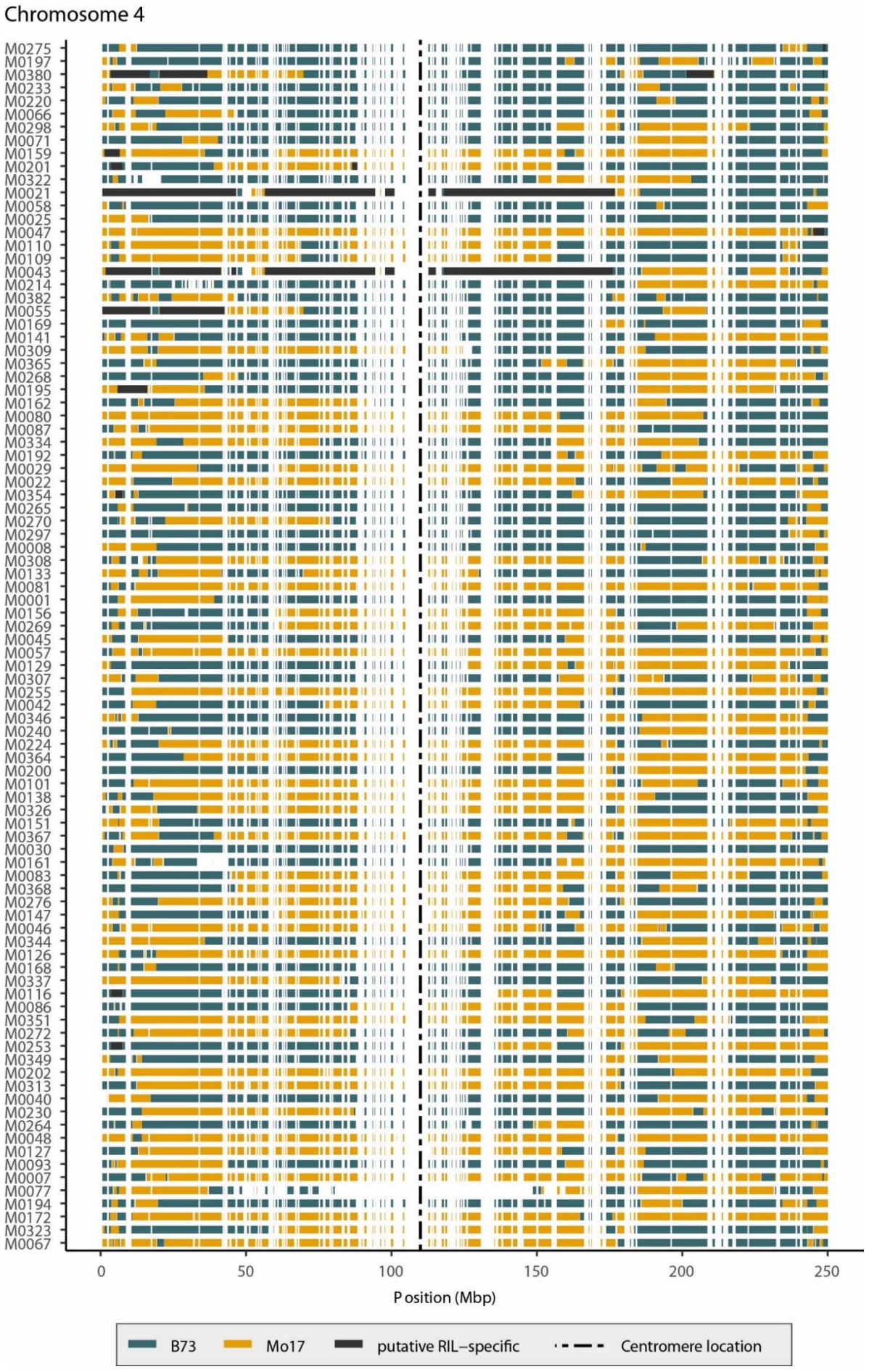
Table S1: Synteny of all active genes with eQTL

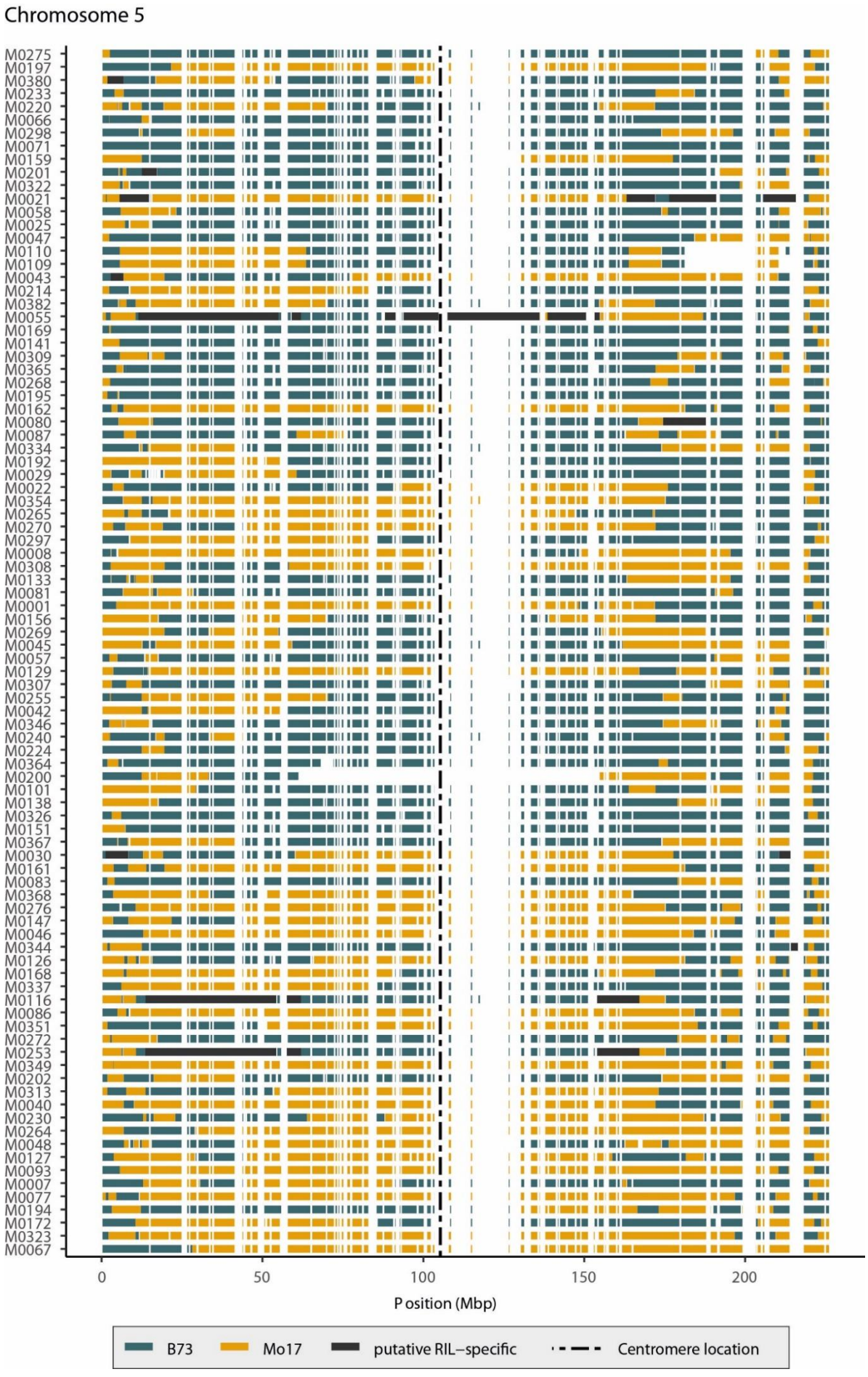
Regulation	Syntenic	Non-syntenic
Genes with *cis*-eQTL	9292 (78%)	2694 (22%)
Genes with *trans*-eQTL	496 (30%)	1131 (70%)

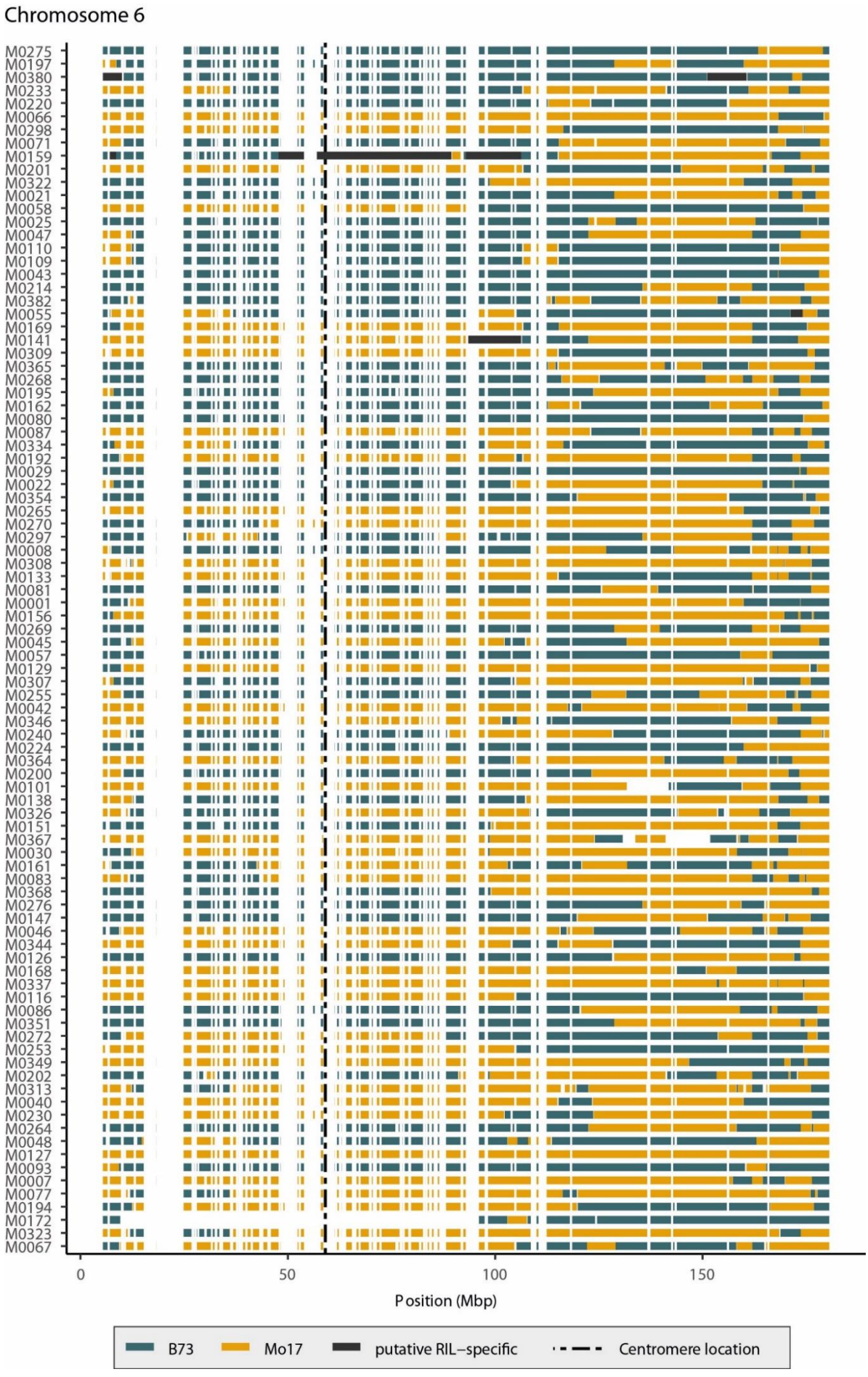

Chromosome 1
M0275
M0197
M0380
M0233
M0220
M0066
M0298
M0071
M0159
M0201
M0322
M0021
M0058
M0025
M0047
M0110
M0109
M0043
M0214
M0382
M0055
M0169
M0141
M0309
M0365
M0268
M0195
M0162
M0080
M0087
M0334
M0192
M0029
M0022
M0354
M0265
M0270
M0297
M0008
M0308
M0133
M0081
M0001
M0156
M0269
M0045
M0057
M0129
M0307
M0255
M0042
M0346
M0240
M0224
M0364
M0200
M0101
M0138
M0326
M0151
M0367
M0030
M0161
M0083
M0368
M0276
M0147
M0046
M0344
M0126
M0168
M0337
M0116
M0086
M0351
M0272
M0253
M0349
M0202
M0313
M0040
M0230
M0264
M0048
M0127
M0093
M0007
M0077
M0194
M0172
M0323
M0067
0
100
200
300
Position (Mbp)
B73
Mo17
putative RIL−specific
Centromere location

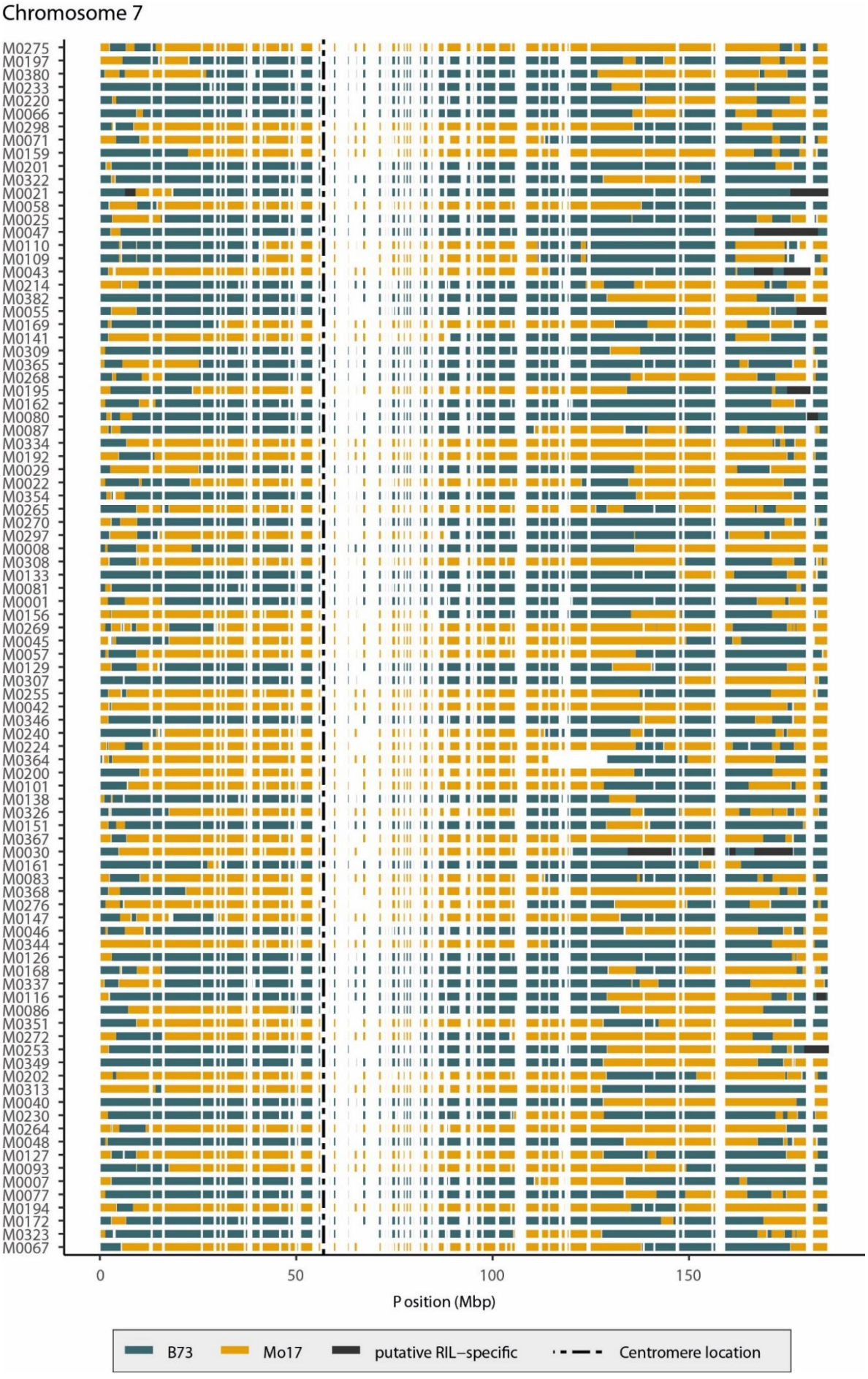

Chromosome 2
M0275
M0197
M0380
M0233
M0220
M0066
M0298
M0071
M0159
M0201
M0322
M0021
M0058
M0025
M0047
M0110
M0109
M0043
M0214
M0382
M0055
M0169
M0141
M0309
M0365
M0268
M0195
M0162
M0080
M0087
M0334
M0192
M0029
M0022
M0354
M0265
M0270
M0297
M0008
M0308
M0133
M0081
M0001
M0156
M0269
M0045
M0057
M0129
M0307
M0255
M0042
M0346
M0240
M0224
M0364
M0200
M0101
M0138
M0326
M0151
M0367
M0030
M0161
M0083
M0368
M0276
M0147
M0046
M0344
M0126
M0168
M0337
M0116
M0086
M0351
M0272
M0253
M0349
M0202
M0313
M0040
M0230
M0264
M0048
M0127
M0093
M0007
M0077
M0194
M0172
M0323
M0067
0
50
100
150
200
250
Position (Mbp)
B73
Mo17
putative RIL−specific
Centromere location

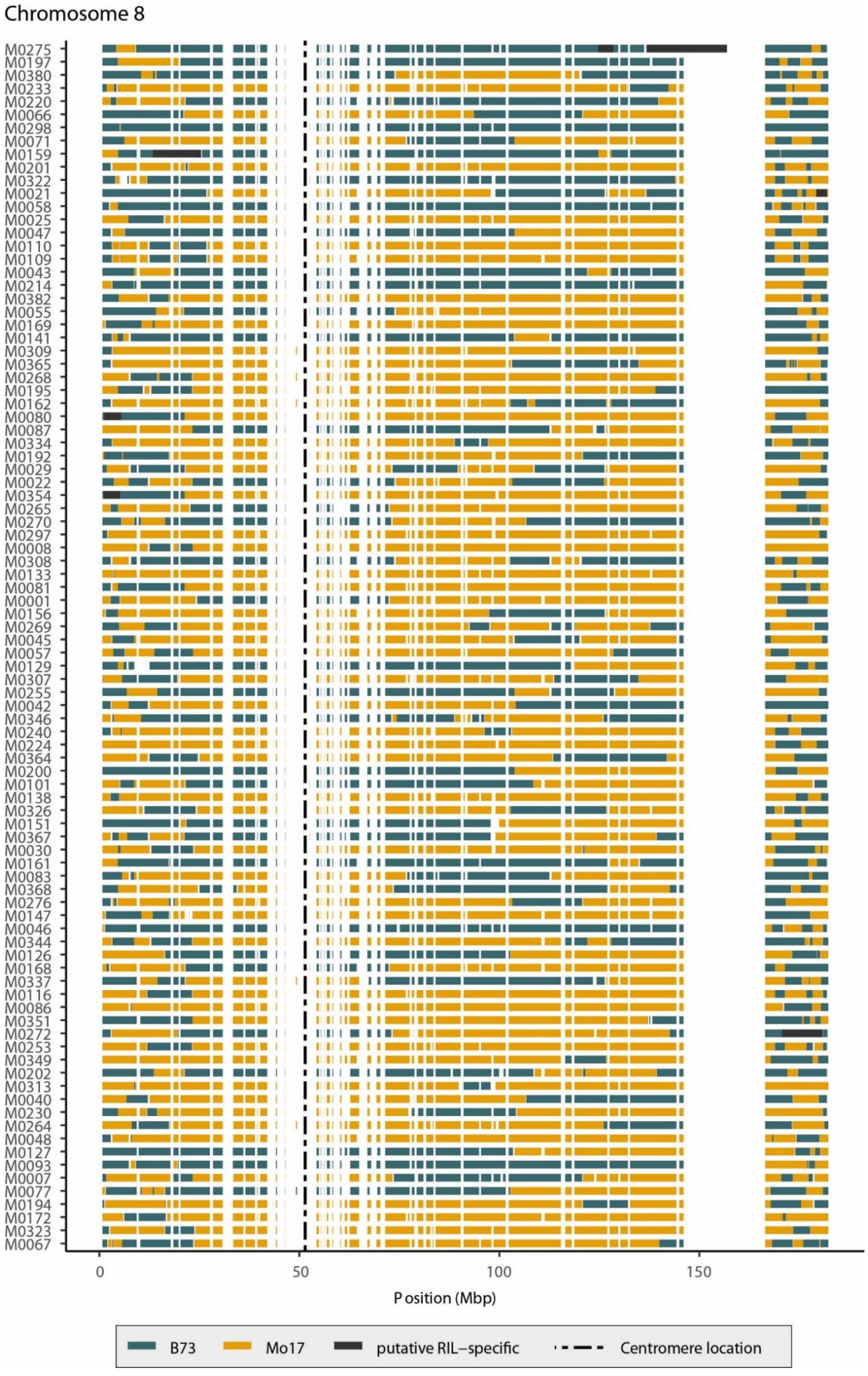

Chromosome 3
M0275
M0197
M0380
M0233
M0220
M0066
M0298
M0071
M0159
M0201
M0322
M0021
M0058
M0025
M0047
M0110
M0109
M0043
M0214
M0382
M0055
M0169
M0141
M0309
M0365
M0268
M0195
M0162
M0080
M0087
M0334
M0192
M0029
M0022
M0354
M0265
M0270
M0297
M0008
M0308
M0133
M0081
M0001
M0156
M0269
M0045
M0057
M0129
M0307
M0255
M0042
M0346
M0240
M0224
M0364
M0200
M0101
M0138
M0326
M0151
M0367
M0030
M0161
M0083
M0368
M0276
M0147
M0046
M0344
M0126
M0168
M0337
M0116
M0086
M0351
M0272
M0253
M0349
M0202
M0313
M0040
M0230
M0264
M0048
M0127
M0093
M0007
M0077
M0194
M0172
M0323
M0067
0
50
100
150
200
Position (Mbp)
B73
Mo17
putative RIL−specific
Centromere location

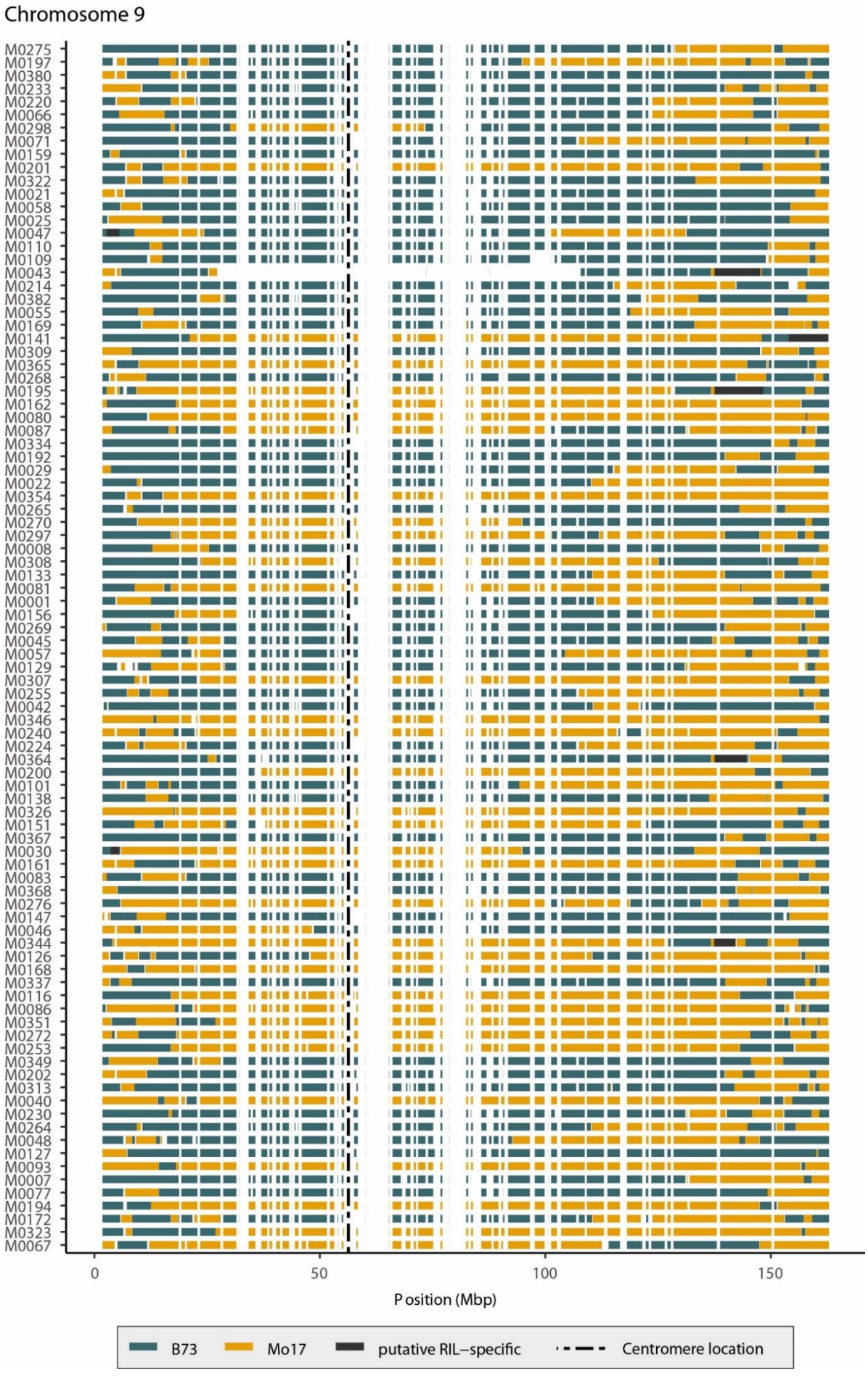

Chromosome 4
M0275
M0197
M0380
M0233
M0220
M0066
M0298
M0071
M0159
M0201
M0322
M0021
M0058
M0025
M0047
M0110
M0109
M0043
M0214
M0382
M0055
M0169
M0141
M0309
M0365
M0268
M0195
M0162
M0080
M0087
M0334
M0192
M0029
M0022
M0354
M0265
M0270
M0297
M0008
M0308
M0133
M0081
M0001
M0156
M0269
M0045
M0057
M0129
M0307
M0255
M0042
M0346
M0240
M0224
M0364
M0200
M0101
M0138
M0326
M0151
M0367
M0030
M0161
M0083
M0368
M0276
M0147
M0046
M0344
M0126
M0168
M0337
M0116
M0086
M0351
M0272
M0253
M0349
M0202
M0313
M0040
M0230
M0264
M0048
M0127
M0093
M0007
M0077
M0194
M0172
M0323
M0067
0
50
100
150
200
250
Position (Mbp)
B73
Mo17
putative RIL−specific
Centromere location


Chromosome 5
M0275
M0197
M0380
M0233
M0220
M0066
M0298
M0071
M0159
M0201
M0322
M0021
M0058
M0025
M0047
M0110
M0109
M0043
M0214
M0382
M0055
M0169
M0141
M0309
M0365
M0268
M0195
M0162
M0080
M0087
M0334
M0192
M0029
M0022
M0354
M0265
M0270
M0297
M0008
M0308
M0133
M0081
M0001
M0156
M0269
M0045
M0057
M0129
M0307
M0255
M0042
M0346
M0240
M0224
M0364
M0200
M0101
M0138
M0326
M0151
M0367
M0030
M0161
M0083
M0368
M0276
M0147
M0046
M0344
M0126
M0168
M0337
M0116
M0086
M0351
M0272
M0253
M0349
M0202
M0313
M0040
M0230
M0264
M0048
M0127
M0093
M0007
M0077
M0194
M0172
M0323
M0067
0
50
100
150
200
Position (Mbp)
B73
Mo17
putative RIL−specific
Centromere location


Chromosome 6
M0275
M0197
M0380
M0233
M0220
M0066
M0298
M0071
M0159
M0201
M0322
M0021
M0058
M0025
M0047
M0110
M0109
M0043
M0214
M0382
M0055
M0169
M0141
M0309
M0365
M0268
M0195
M0162
M0080
M0087
M0334
M0192
M0029
M0022
M0354
M0265
M0270
M0297
M0008
M0308
M0133
M0081
M0001
M0156
M0269
M0045
M0057
M0129
M0307
M0255
M0042
M0346
M0240
M0224
M0364
M0200
M0101
M0138
M0326
M0151
M0367
M0030
M0161
M0083
M0368
M0276
M0147
M0046
M0344
M0126
M0168
M0337
M0116
M0086
M0351
M0272
M0253
M0349
M0202
M0313
M0040
M0230
M0264
M0048
M0127
M0093
M0007
M0077
M0194
M0172
M0323
M0067
0
50
100
150
Position (Mbp)
B73
Mo17
putative RIL–specific
Centromere location


Chromosome 7
M0275
M0197
M0380
M0233
M0220
M0066
M0298
M0071
M0159
M0201
M0322
M0021
M0058
M0025
M0047
M0110
M0109
M0043
M0214
M0382
M0055
M0169
M0141
M0309
M0365
M0268
M0195
M0162
M0080
M0087
M0334
M0192
M0029
M0022
M0354
M0265
M0270
M0297
M0008
M0308
M0133
M0081
M0001
M0156
M0269
M0045
M0057
M0129
M0307
M0255
M0042
M0346
M0240
M0224
M0364
M0200
M0101
M0138
M0326
M0151
M0367
M0030
M0161
M0083
M0368
M0276
M0147
M0046
M0344
M0126
M0168
M0337
M0116
M0086
M0351
M0272
M0253
M0349
M0202
M0313
M0040
M0230
M0264
M0048
M0127
M0093
M0007
M0077
M0194
M0172
M0323
M0067
0
50
100
150
Position (Mbp)
B73
Mo17
putative RIL−specific
Centromere location


Chromosome 8
M0275
M0197
M0380
M0233
M0220
M0066
M0298
M0071
M0159
M0201
M0322
M0021
M0058
M0025
M0047
M0110
M0109
M0043
M0214
M0382
M0055
M0169
M0141
M0309
M0365
M0268
M0195
M0162
M0080
M0087
M0334
M0192
M0029
M0022
M0354
M0265
M0270
M0297
M0008
M0308
M0133
M0081
M0001
M0156
M0269
M0045
M0057
M0129
M0307
M0255
M0042
M0346
M0240
M0224
M0364
M0200
M0101
M0138
M0326
M0151
M0367
M0030
M0161
M0083
M0368
M0276
M0147
M0046
M0344
M0126
M0168
M0337
M0116
M0086
M0351
M0272
M0253
M0349
M0202
M0313
M0040
M0230
M0264
M0048
M0127
M0093
M0007
M0077
M0194
M0172
M0323
M0067
0
50
100
150
Position (Mbp)
B73
Mo17
putative RIL−specific
Centromere location


Chromosome 9
M0275
M0197
M0380
M0233
M0220
M0066
M0298
M0071
M0159
M0201
M0322
M0021
M0058
M0025
M0047
M0110
M0109
M0043
M0214
M0382
M0055
M0169
M0141
M0309
M0365
M0268
M0195
M0162
M0080
M0087
M0334
M0192
M0029
M0022
M0354
M0265
M0270
M0297
M0008
M0308
M0133
M0081
M0001
M0156
M0269
M0045
M0057
M0129
M0307
M0255
M0042
M0346
M0240
M0224
M0364
M0200
M0101
M0138
M0326
M0151
M0367
M0030
M0161
M0083
M0368
M0276
M0147
M0046
M0344
M0126
M0168
M0337
M0116
M0086
M0351
M0272
M0253
M0349
M0202
M0313
M0040
M0230
M0264
M0048
M0127
M0093
M0007
M0077
M0194
M0172
M0323
M0067
0
50
100
150
Position (Mbp)
B73
Mo17
putative RIL−specific
Centromere location

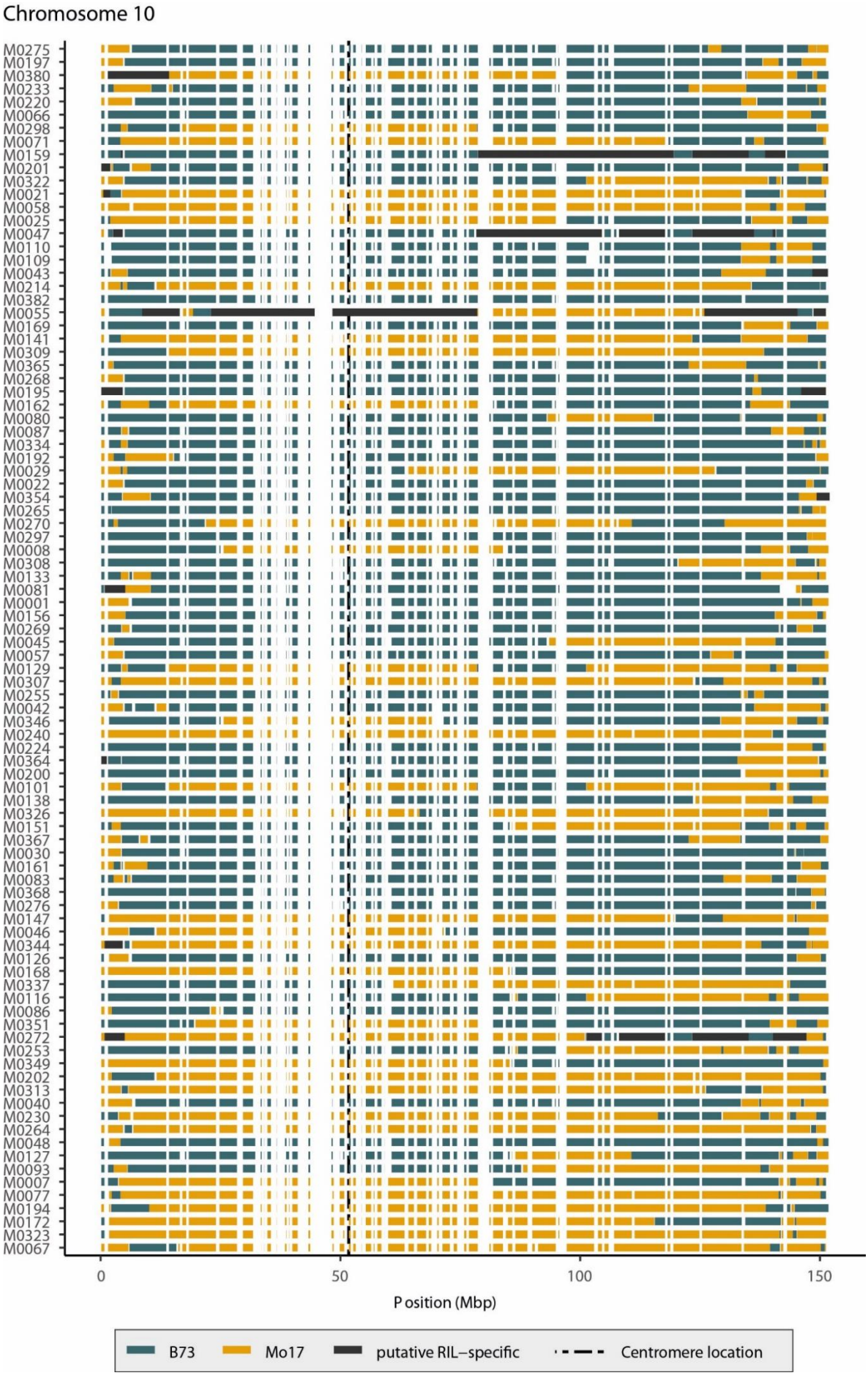

Chromosome 10
M0275
M0197
M0380
M0233
M0220
M0066
M0298
M0071
M0159
M0201
M0322
M0021
M0058
M0025
M0047
M0110
M0109
M0043
M0214
M0382
M0055
M0169
M0141
M0309
M0365
M0268
M0195
M0162
M0080
M0087
M0334
M0192
M0029
M0022
M0354
M0265
M0270
M0297
M0008
M0308
M0133
M0081
M0001
M0156
M0269
M0045
M0057
M0129
M0307
M0255
M0042
M0346
M0240
M0224
M0364
M0200
M0101
M0138
M0326
M0151
M0367
M0030
M0161
M0083
M0368
M0276
M0147
M0046
M0344
M0126
M0168
M0337
M0116
M0086
M0351
M0272
M0253
M0349
M0202
M0313
M0040
M0230
M0264
M0048
M0127
M0093
M0007
M0077
M0194
M0172
M0323
M0067
0
50
100
150
Position (Mbp)
B73
Mo17
putative RIL−specific
Centromere location

Figure S1: Map of genomic regions in IBM-RILs. Each page shows the genomic regions, of all 94 IBM-RILs for one chromosome on physical scale. Regions of B73 are shown in blue, Mo17 in yellow, putative IBM-RIL specifc regions, which were masked, are shown in black. White spaces indicate that no SNPs were found in this region. The centromere location is indicated by a vertical dashed line.

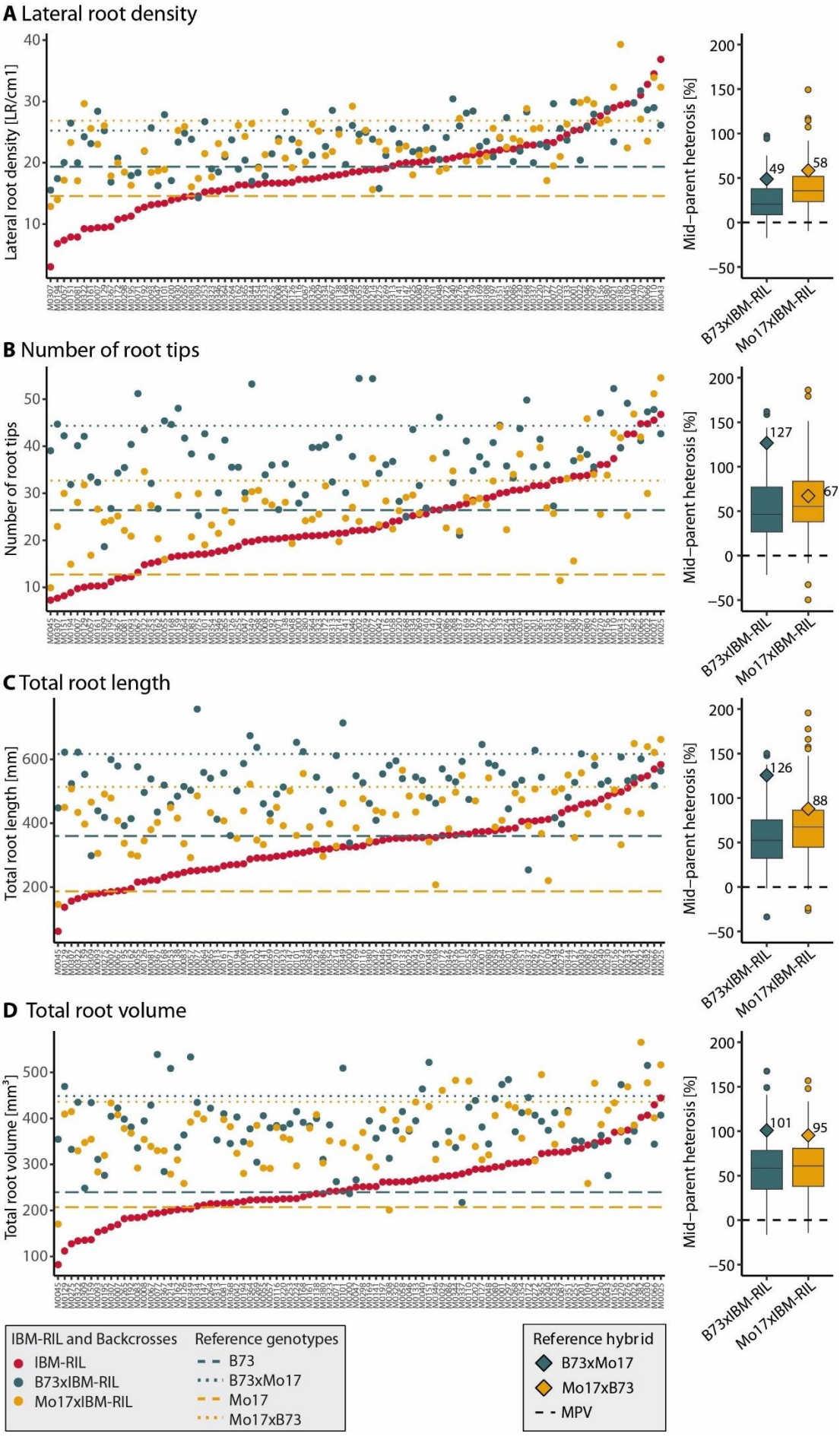

A Lateral root density
Lateral root density [LR/cm1]
Mid−parent heterosis [%]
49
58
B73xIBM-RIL
Mo17xIBM-RIL
B Number of root tips
Number of root tips
127
67
C Total root length
Total root length [mm]
126
88
D Total root volume
Total root volume [mm³]
101
95
IBM-RIL and Backcrosses
IBM-RIL
B73xIBM-RIL
Mo17xIBM-RIL
Reference genotypes
B73
B73xMo17
Mo17
Mo17xB73
Reference hybrid
B73xMo17
Mo17xB73
MPV

Figure S2: Phenotypic values and mid-parent heterosis (MPH) of **A** lateral root density, **B** Number of root tips, **C** Total root length, **D** Total root volume. On the left panel, the estimated means for each genotype are shown. The IBM-RILs (red) and the B73xIBM-RIL (blue) and Mo17xIBM-RIL (yellow) backcross hybrids are shown as points. The reference genotypes B73 (blue) and Mo17 (yellow) are shown as dashed lines and the reciprocal reference hybrids B73xMo17 (blue) and Mo17xB73 (yellow) as dotted lines. On the right panel, the mid-parent heterosis in percent of the parental mean is shown as boxplots for B73xIBM-RILs and Mo17xIBM-RILs. The MPH for the reference hybrids B73xMo17 (blue) and Mo17xB73 (yellow) is shown as diamond shaped points and their exact values are indicated. The dashed line indicates an MPV of 0.

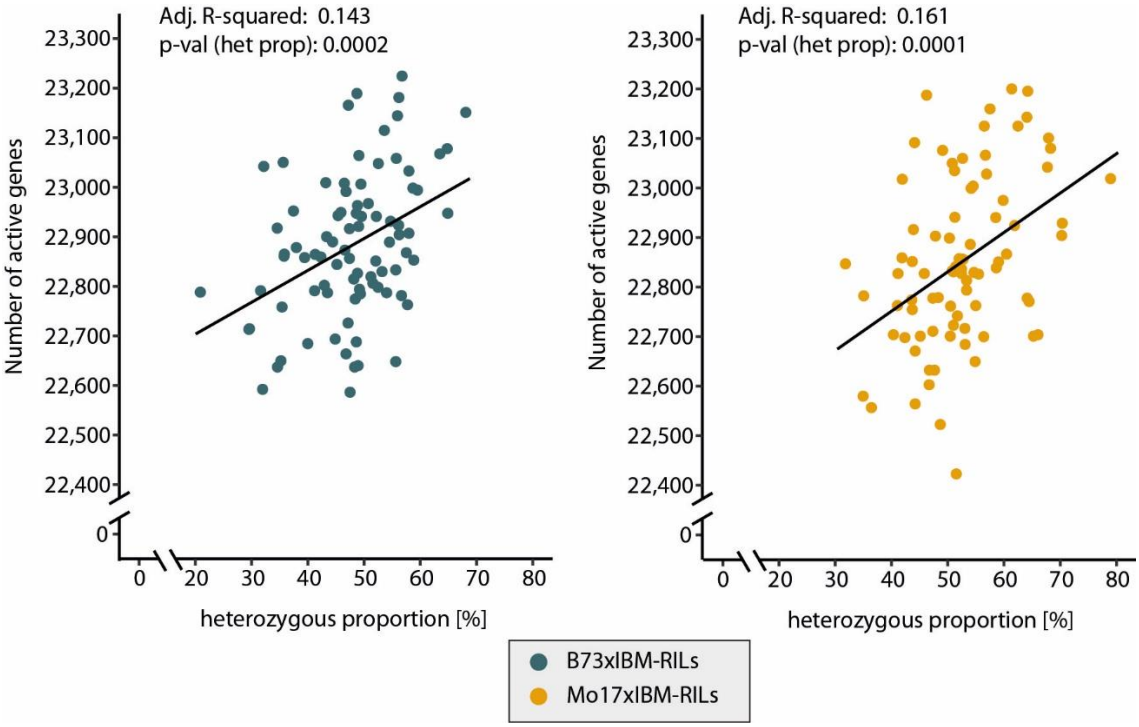


Figure S3: Correlation of heterozygosity and the total number of active genes in B73xIBM-RILs (left) and Mo17xIBM-RILs (right). The heterozygous proportion on the x-axis was calculated from the classified IBM-RIL regions. A linear regression with an intercept and the heterozygous proportion as covariate was fitted and the adjusted R-squared and p-value for the slope of the heterozygosity are indicated above the regression lines.

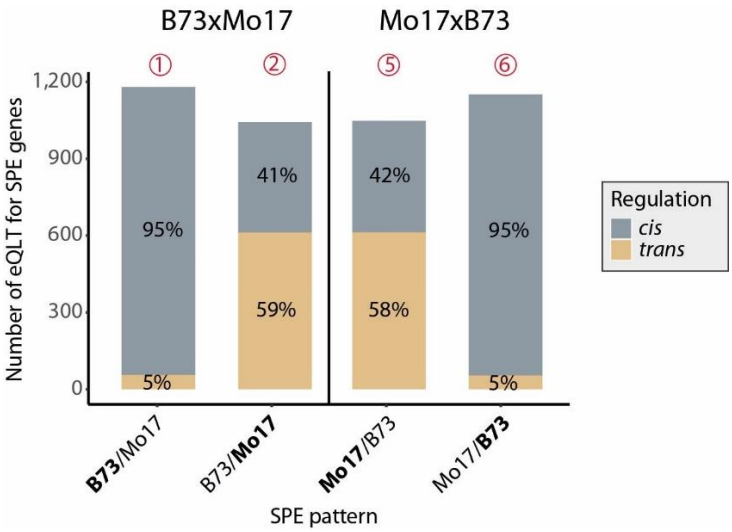


Figure S4: Regulation of SPE pattern genes in fully heterozygous reference hybrids B73xMo17 (left side) and Mo17xB73 (right side). The number of *cis*-(blue) and *trans*-(yellow) acting eQTL are given as bars, with the percentages of *cis* and *trans*-acting eQTL indicated per SPE pattern. The numbers above correspond to SPE pattern as indicated in Figure 2.

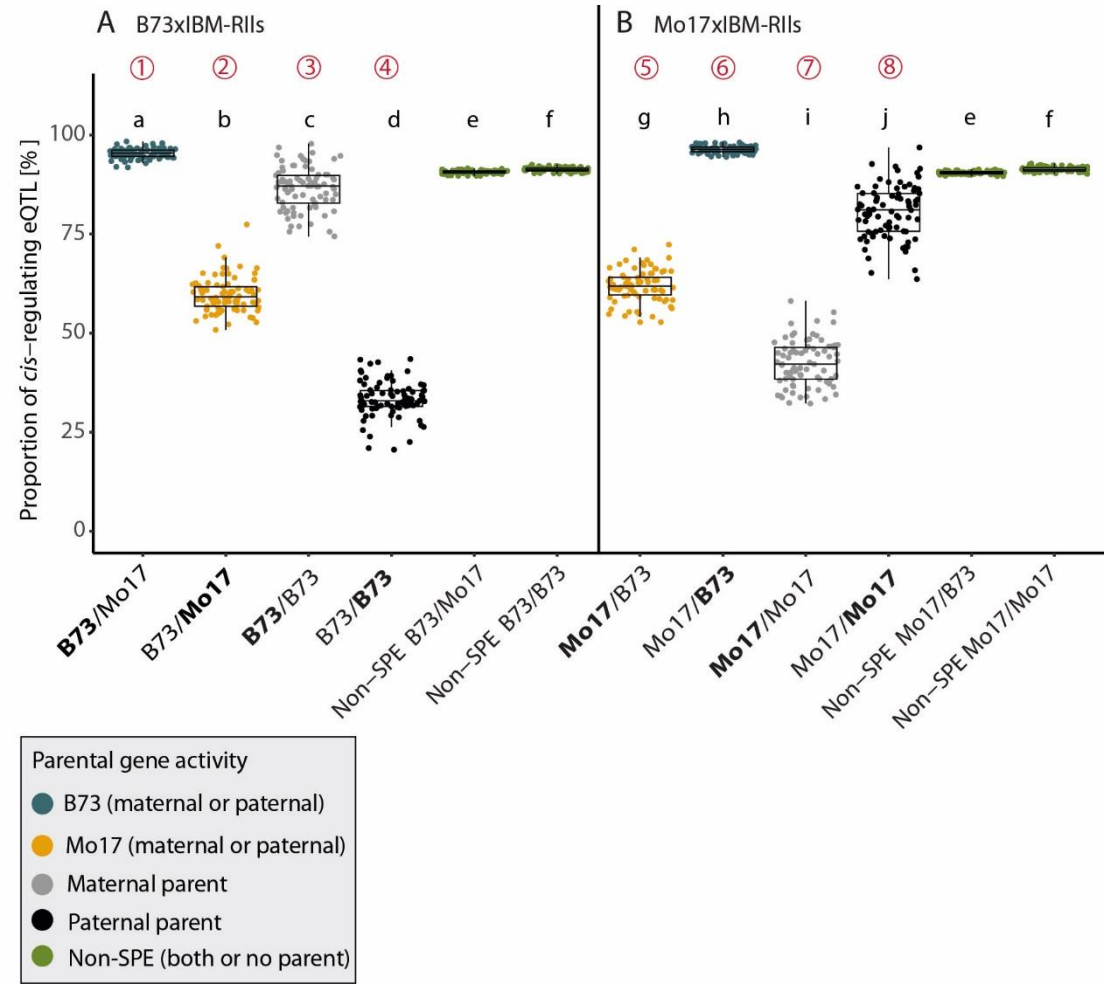


Figure S5: Proportion of *cis*- and *trans*- regulation in SPE pattern genes. Boxplots display the proportion of *cis*-regulation among SPE pattern and non-SPE pattern genes in **A** B73xIBM-RIL and **B** Mo17xIBM-RIL hybrids. Different letters indicate significantly different proportions (a<0.05), identified with a gaussian mixed model with the hybrid as random effect, the SPE pattern and non-SPE pattern as a fixed factor and a diagonal variance component for the SPE and non-SPE pattern.

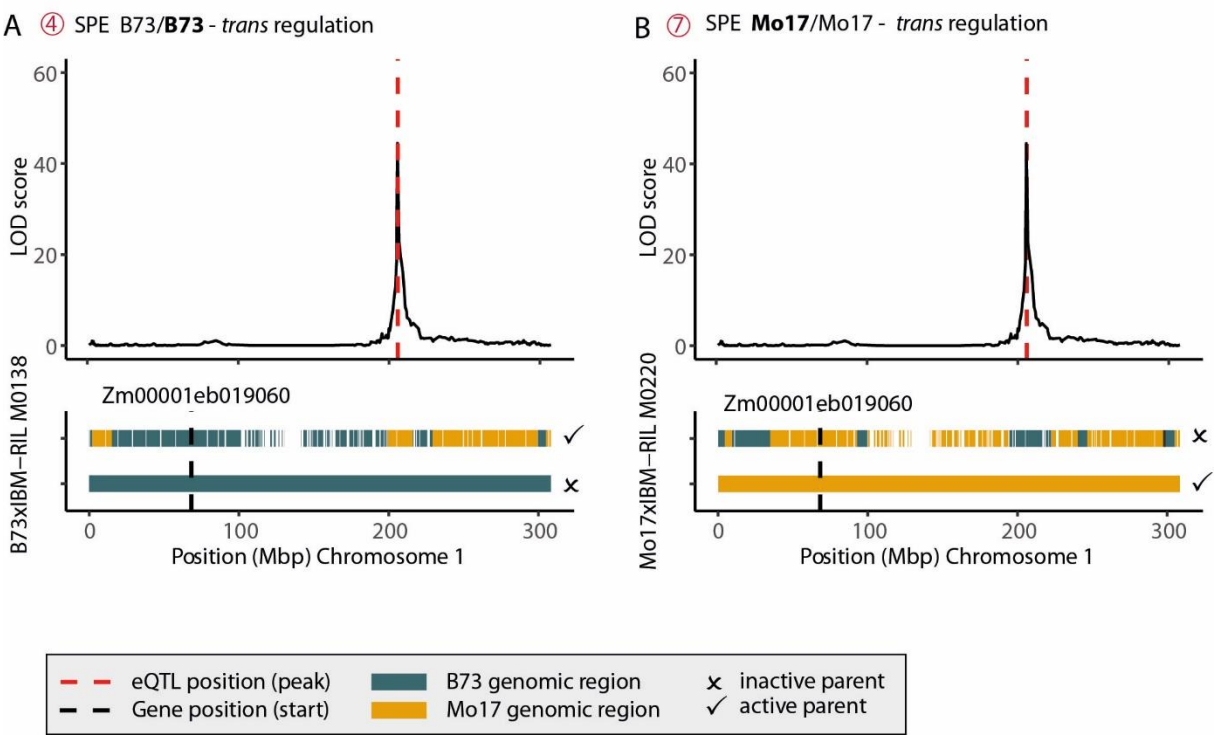


Figure S6: Examples for gene regulation pattern. **A** Homozygous SPE pattern 4 (B73/**B73**) gene in a B73xIBM-RIL, which is *trans* regulated from a heterozygous eQTL. **B** homozygous SPE pattern 7 (**Mo17**/Mo17) gene in a Mo17xIBM-RIL, which is *trans*-regulated homozygous gene from heterozygous eQTL.

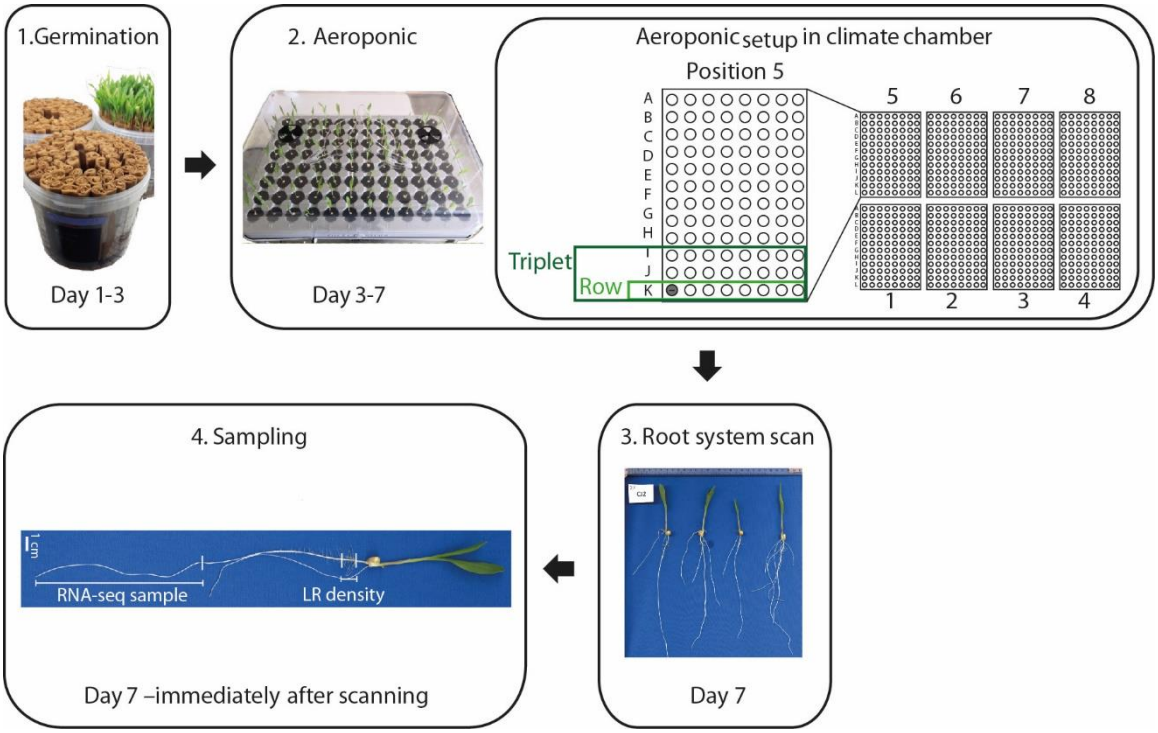


Figure S7: Experimental workflow incl. layout of the experimental design. Schematic depiction of the plant growing process from paper rolls until sampling of roots for RNA sequencing. The second box shows the distribution of the aeroponic systems in the climate chamber and the assignment of genotype triplets and individual genotypes within each system.

Supplementary data files for Chapter 3 in the digital appendix

Data_S1_exclusion_table.xlsx	Details of retained and excluded samples.
Data_S2_Loci_listB73vsMo17.xlsx	List and details of identified loci with different alleles in B73 and Mo17 samples.
Data_S3_IBM-RIL_haplotype_regions.xlsx	Specifications of start and end position of each identified Mo17, B73 or IBM-RIL specific region in each IBM-RIL.
Data_S4_eQTL_details.xlsx	Details on identified eQTL
Data_S5_MPH_pheno.xlsx	Phenotypic mid-parent heterosis values of each hybrid.
Data_S6_pattern_ref.xlsx	SPE pattern of all genes in reference hybrids and their eQTL.
Data_S7_TWAS_genes.xlsx	All genes and details identified in the TWAS analysis.
Data_S8_TSG.xlsx	Details on SPE pattern across hybrids of TWAS candidate genes.
Data_S9_Zm00001eb339600_B73.txt	BLAST result for candidate gene.

7.2 Supplementary information of chapter 4

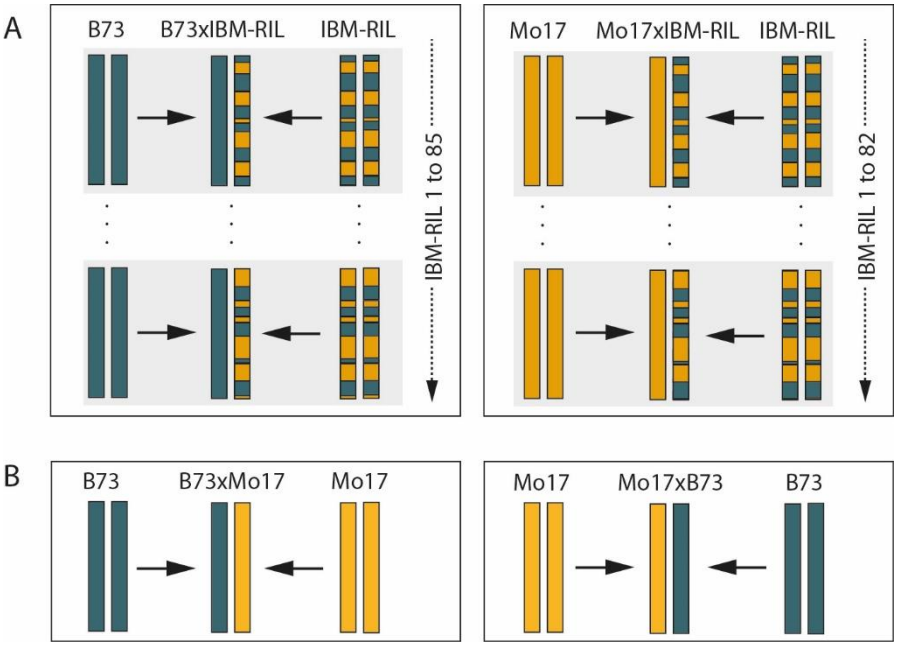


Figure S1: Investigated parent and hybrid genotypes. Modified after Pitz et al. 2024. Schematic depiction of **A** IBM-RIL backcross population composition and **B** reference hybrid composition.

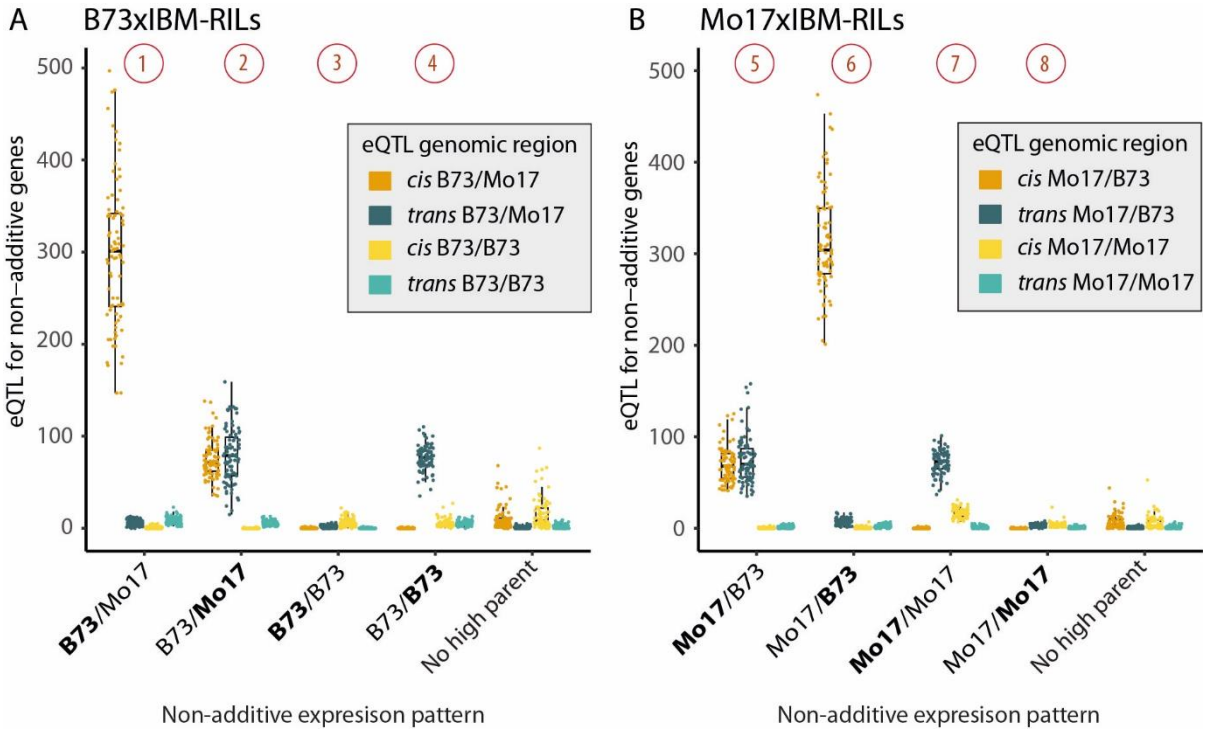


Figure S2: Boxplots showing eQTL of non-additive genes. **A** B73xIBM-RIL hybrids or **B** Mo17xIBM-RIL hybrids. Non-additive expression pattern is shown on the x-axis, with the genotype as indicated to distinguish hetero- and homozygous genes and the higher expressed parent (|Log2FC|>1, p<0.05) indicated in bold, or no signifcant difference in the parents (No high parent). The associated eQTL can be *cis*-regulating from heterozygous (dark blue) or homozygous (light blue) regions or *trans*-regulating from heterozygous (dark yellow) or homozygous (light yellow) regions.

Supplementary data files for Chapter 4 in the digital appendix

Data_S1_NAG_refs.xlsx	Non-additive pattern of all genes in reference hybrids.
Data_S2_overview.xlsx	Summarised details of additive and non-additive genes in all hybrids.
Data_S3_het_prop.xlsx	Proportions of heterozygous regions in the hybrids.

8. Publications

8.1 Publications related to this thesis

- **Regulation of heterosis-associated gene expression complementation in maize hybrids**

  Marion Pitz[#], Jutta A. Baldauf[#], Hans-Peter Piepho, Peng Yu, Heiko Schoof, Annaliese S. Mason, Guoliang Li, Frank Hochholdinger

  Genome Biology. (Under review) (#: co-first authors).

  Own contribution: I participated in and co-organized the growing and harvesting of plant material and RNA-extraction. I conducted the phenotypic data collection and analyis as well as the bioinformatic data preparation of the RNA-seq data (alignment, readcount quantification, SNP calling, classification of IBM-RILs in genomic regions). Further, I performed the eQTL and TWAS analyses and identified the intersections of those resuls with the SPE genes and the syntenic and non-syntenic genes. I interpreted the resulting data, designed the figures and wrote the corresponding parts of the manuscript.

- **Non-additive gene expression contributing to heterosis in partially heterozygous maize hybrids is predominantly regulated from heterozygous regions**

  Marion Pitz[#], Jutta A. Baldauf[#], Hans-Peter Piepho, Frank Hochholdinger

  New Phytologist. (Under review) (#: co-first authors).

  Own contribution: Both papers originate from the same set of raw data. Regarding this paper specifically, I determined and classified the additively and non-additively expressed genes and their intersections with the eQTL, performed all statistical analyses and interpreted the results. I wrote the manuscript and designed the figures.

8.2 Publications unrelated to this thesis

- **Expanding the BonnMu sequence-indexed repository of transposon induced maize (Zea mays L.) mutations in dent and flint germplasm**

  Yan Nain Win[#], Tyll Stöcker[#], Xuelian Du, Alexa Brox, Marion Pitz, Alina Klaus, Hans-Peter Piepho, Heiko Schoof, Frank Hochholdinger, Caroline Marcon.

  The Plant Journal (2024).

- **Transcriptomic diversity in seedling roots of European flint maize in response to cold**

  Felix P. Frey[#], Marion Pitz[#], Chris-Carolin Schön, Frank Hochholdinger
  BMC Genomics (2020).

8.3 Presentations at conferences

- In preparation: **Regulation of heterosis-associated gene expression complementation in maize hybrids**
  67th Annual Maize Genetics Meeting 06.-09.03.2025 St. Louis, Missouri, USA (*Oral presentation*)

- **How hybrid expression patterns are regulated and how they influence heterosis in maize seedling roots**
  6th European Maize Meeting 18.-20.09.2024 Antwep, Belgium (*Oral presentation*).

- **How hybrid expression patterns are regulated and how they influence heterosis in maize seedling roots**
  German society of plant nutrition 56th annual conference 02.-04.09.2024 Bonn, Germany (*Poster Presentation*)

- **“The more, the better?”: Investigating differential, non-additive and allelic expression in association with the phenotypic manifestation of heterosis**
  5th European Maize Meeting 14.-16.06.2023 Bologna, Italy (*Poster presentation*)

- **Gene expression complementation and its association with the phenotypic manifestation of heterosis in maize hybrids**
  Botanik Tagung: International Conference of the German Society for Plant Sciences 28.08.-01.09.2022 Bonn, Germany (*Poster presentation*)

Acknowledgement

First of all, I would like to thank my supervisor Prof. Dr. Frank Hochholdinger for the chance to work on this PhD project. Thank you very much for your constant support, advice and guidance as well as the various opportunities to advance my scientific and professional skills, for example at international conferences. I was happy to rejoin the Crop Functional Genomics group which I valued for the friendly, helpful and familiar atmosphere. It was a step that I never regretted and I look forward to staying a little longer with you.

Many thanks also to the collaborators of this project, especially Peng Yu (University Bonn), Heiko Schoof (University Bonn), Guoliang Li (IPK Germany) as well as Annaliese S. Mason (University Bonn) and Hans-Peter Piepho (University of Hohenheim) , who additionally agreed to form my examination committee next to Janina Dierks (University of Bonn).

I would like to thank the Bonn International Graduate School (BIGS) – Land and Food for financial support and learning opportunities.

A warm thank you goes to all my colleagues of the Crop Functional Genomics and Chemical Signalling lab for their help, encouragements and support.

- Thanks to Jutta Baldauf, for welcoming me to the project and mentoring me. It was incredibly valuable to work with you and learn from you.
- Thanks to my fellow (former) PhD students: Alina, Annika, Cay, Ling, Liuyang, Luca, Manjeera, Mauritz, Michelle, Taheb, Verena, Wenxin, Xiaofang, Xiaolin, Xuelian, Yue and Zamiga as well as the many visiting guests. Thank you for your helpful input and discussions, in particular within the RNA-extravaganza group.
- Thanks to the post-docs Caro, Li, Marcel, Micha, Yan, Xiaoming, Yaping, Zhihui, José and Anna.
- Thanks to Christine Jessen and Ellen Kreitz for their administrative skills and patience in their support.
- Thank you Alena, Alexa, Britta, Helmut, Ute and Selina for your valuable work in fixing anything and helping with everything.

Finally, I would like to thank my family and friends, especially my parents and Kathi for their ongoing support during the whole time of this PhD project.