Gene expression dynamics upon allopolyploidization: Global transcriptome analysis in synthetic hexaploid wheats and their parents

Akshaya Vasudevan

A thesis submitted to the University of Ottawa

in partial fulfilment of the requirements for the

Doctorate in Philosophy in Biology

Department of Biology

Faculty of Science

University of Ottawa

Abstract

The allohexaploid bread wheat (*Triticum aestivum* L.), evolved through a recent polyploidization event between tetraploid *Triticum turgidum* L. (AABB) and diploid *Aegilops tauschii* Coss. (DD), ~8,000 years ago. Contribution of only a subpopulation of *Ae. tauschii* to hexaploid evolution, followed by domestication and extensive breeding with the objective of higher yield gain and strict end-use quality determining the market classes of wheat, have created a genetic bottleneck. Synthetic hexaploid wheat (SHW) lines are generated to restore the diversity and exploit the genetic resource in the primary gene pool of wheat. However, there are challenges with recovering the phenotypes observed in the parental background in the hexaploid bread wheat in pre-breeding programs.

To understand and characterize the barriers in utilizing the progenitor genetic diversity, the transcriptome of four SHW lines and their corresponding tetraploid and diploid parents across ten tissues, totalling to 240 samples, was analysed. The comparison of expression bias of homoeologues present as >18,000 triads (1:1:1) between parental *in-silico* SHW-like scenarios and SHWs, indicated a large-scale suppression of D subgenome homoeoalleles in SHWs. Tissue-specificity was not observed in the homoeologues of a large proportion of the triads. The next largest fractions were triads where all homoeologues displayed the same tissue-specific expression followed by those where only one of the homoeologues was tissue-specific. Several SHW-tissues showed moderate relationship between tissue-specificity of the homoeologues and expression bias of the corresponding triad. The repression of the genes of the D subgenome was also validated in the differential expression analysis using the entire high-confidence gene set of hexaploid wheat. Qualitative analysis of the transcripts revealed all five splicing events with predominance of retained introns, and more differentially-spliced transcripts were associated with the D homoeoalleles in most SHW-tissue contexts.

The introgression patterns of the SHW-C66 into the elite bread wheat cultivar Carberry was analysed using a BC_1F_5 population. Large introgression of SHW-C66 were found closer to centromeric regions while smaller fragments were present towards the ends of the chromosomes. Correspondingly, the majority of the chromosomes showed higher recombination rates away from the centromere. The donor allele frequency was higher than the expected 25% for BC_1F_5 population in multiple regions of the A and B subgenomes but not in the D subgenome. In comparison, a preliminary analysis using an elite wheat × elite wheat doubled haploid population showed no subgenome-level variation in recombination rates or donor allele frequencies.

In this thesis work, both functional genomic and structural genomic investigations using a set of SHW parents and their derivative population with elite wheat cultivars have unearthed some key patterns that add to the collective knowledge needed to fully exploit genetic resources in broadening the genetic diversity in wheat improvement programs.

Acknowledgements

I immensely thank my supervisor and thesis advisor Dr. Sylvie Cloutier, Principal Research Scientist, Agriculture and Agri-Food Canada (AAFC) and Adjunct Professor, University of Ottawa, for the dream opportunity to join her worldclass genomics lab and her exceptional guidance, support and mentorship over the last four years. All the discussions with her have been intellectually stimulating, and has enriched my research mind and academic skills, including writing. She is a role model scientist, team lead and mentor and her entire professional career journey as a scientist of highest calibre has inspired me as a woman in STEM for positioning my research and academic career ahead of me. I will always be grateful.

I sincerely thank my thesis advisory committee members Dr. Jean-Sébastien Parent (Research Scientist, AAFC), Dr. Allyson M. MacLean (Associate Professor of Biology, University of Ottawa), and Dr. Nicolas Corradi (Professor of Biology, University of Ottawa), for all their intellectual inputs and feedbacks during the review meetings and PhD program. Their critical insights and suggestions have helped me to investigate the novel perspectives during my research work.

My special thanks to Dr. Martina Stromvik (Associate Professor of Plant Science, McGill University) for agreeing to serve as the external examiner for my thesis.

I am very grateful to senior research technicians of Dr. Sylvie Cloutier's lab, namely Tara Edwards and Madeleine Lévesque-Lemay for all their important assistance towards the experiments and additional support with manuscript review and editing.

I sincerely thank Dr. George Fedak (Former Principal Research Scientist, AAFC) and his senior technician Dawn Chi for generating the synthetic hexaploid wheat material.

I am also thankful to Dr. Frank You (Research Scientist, AAFC) and his lab members Dr. Chunfang Zhen (Biologist of Computational Biology) and Dr. Pingchuan Li (PostDoc in Bioinformatics) for their bioinformatics assistance. I am grateful to Dr. Ronald D. Smith, Associate Teaching Professor of Data Science, William & Mary, for clarifications on the code for likelihood ratio test and Dr. Steven Dreissig, Research Scientist, Martin-Luther-University Halle-Wittenberg, for readily clarifying my doubts while developing the scripts for introgression analysis.

I sincerely thank the current and former members of Dr. Cloutier's lab at AAFC for their support and friendship: Dr. Bijendra Khadka, Dr. Sridhar Ravichandran, Fizza Fatima, Dr. Nadeem Khan, Hamna Shazadee and Sampurna Bartaula.

I sincerely acknowledge the funding support for the project from the 4D Wheat: Diversity, Discovery, Design and Delivery project (with Drs. Pozniak and Cloutier as co-leads) - a Genome Canada, Agriculture and Agri-Food Canada and stakeholders of the wheat industry partnership with administrative support from Genome Prairie; and the Canadian Agricultural Partnership. I am grateful to the University of Ottawa for the financial support through the International Doctoral Scholarship and PhD Admission Scholarship.

I am very grateful to Dr. Raja Ragupathy (Research Scientist, AAFC) for his continued support and mentorship over the past seven years and for introducing me to Dr. Cloutier, his own PhD advisor.

My special thanks to Dr. Sylvie Cloutier and Dr. Gavin Humphreys for all their help to settle down, when I first arrived in Ottawa four years ago on a cool spring day in 2019, starting towards my dream of qualifying with a PhD.

I am sincerely thankful to my friends from India, T.V. Abiraami, M.A.S. Gowri, B. Sindhuja, Tanya Sharma and R. Vijayalakshmi for their constant support and encouragement during this journey.

Finally, I thank my dearest parents Leela Vasudevan and K. Vasudevan and sister Aardhra Vasudevan for their unconditional support, encouragement, and belief in me, to pursue my higher education dreams in a place 15,000 kms away from home. I am beyond grateful for everything they have done for me. I also fondly remember my grandfather V. Subbiah and uncle K. Murali who are not here, but would have loved to see me cross the finish line.

Dedicated to

My Parents
and My Sister

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

1DAA	One day after anthesis
A3SS	Alternative 3’ splicing site
A5SS	Alternative 5’ splicing site
AS	Alternative splicing
ATAC-seq	Assay for transposase-accessible chromatin with sequencing
BAM	Binary alignment map
BC_1F_1	Backcross 1 filial 1
BC_1F_2	Backcross 1 filial 2
BC_1F_5	Backcross 1 filial 5
BIL	Backcross introgression line
BLAST	Basic local alignment search tool
BWA	Burrows-Wheeler aligner
CIMMYT	International maize and wheat improvement center
CNPGS	Canadian national plant germplasm system
CO	Crossover
CO_2	Carbon dioxide
CRE	Cis-regulatory element
CRISPR	Clustered regularly interspaced short palindromic repeats
CWRS	Canadian western red spring
DNA	Deoxyribonucleic acid
DSB	Double strand breaks
FDR	False discovery rate
FHB	Fusarium head blight
GATK	Genome analysis toolkit
GBS	Genotyping-by-sequencing
GFF	General feature format
GRN	Gene regulatory network
GVCF	Genomic variant call format
HEB	Homoeologue expression bias

ICARDA	International centre for agricultural research in the dry areas
IWGSC	International wheat genome sequencing consortium
LogFC	Log fold change
LRT	Likelihood ratio test
MAF	Minor allele frequency
MQ	Mapping quality
MXE	Mutually exclusive exons
NAM	Nested association mapping
NCO	Non crossovers
PCR	Polymerase chain reaction
QD	Quality by depth
QTL	Quantitative trait locus/loci
QUAL	Phred-scaled quality score
RI	Retained intron
RIL	Recombinant inbred lines
RNA	Ribonucleic acid
RNAseq	RNA sequencing
RPKM	Reads per kilobase per million mapped reads
SE	Skipped exon
SHW	Synthetic hexaploid wheat
SNP	Single nucleotide polymorphism
STAR	Spliced transcripts alignment to a reference
UTR	Untranslated regions
VCF	Variant call format

CHAPTER 1

General introduction

The current global population is around eight billion and is projected to reach 9.7 billion by 2050 (UN-DESA 2022). Consequently, the global food demand is predicted to increase to 62% by 2050 (van Dijk et al. 2021). Anthropogenic climate change has many negative impacts, including more frequent climate extremes and deviations from climate normal, which have a profound influence on agricultural productivity (IPCC 2022). In addition, there is an increase in sensitivity of global agriculture to the change in climate (Liang et al. 2017; Ortiz-Bobea et al. 2021). Climate change also impacts crop production by modifying the pest and pathogen prevalence in a region, thereby altering their dynamics and bringing in new crop protection challenges (Bebber 2015; Chaloner et al. 2021). With the global population and the consequent food demand anticipated to increase in the next 30 years, coupled with the significant impacts of climate change on crop production, plant breeders are expected to continue breeding high-yielding cultivars resilient to both biotic and abiotic stressors (Challinor et al. 2014; Hickey et al. 2019). In this scenario, crop wild relatives, which are a major reservoir of genetic diversity, can be exploited to broaden the genetic base and introduce novel alleles governing key agronomic traits, including climate resilience, into elite germplasm (Zhang and Batley 2020).

Wheat is one of the three most important staple food crops in the world with a global production of ~771 million tonnes in 2021 (FAO 2021), and serves as a major calorie supplier to a large section of the global population. It is also an important source of protein, dietary fibre and vitamins (Shewry and Hey 2015). It is therefore vital that we continue to improve wheat production to prevent famines and malnourishment. However, the average genetic yield gap in wheat has been estimated at ~51%, and improvement of its genetic

diversity has the potential to narrow down this gap (Senapati et al. 2022). Domestication and subsequent artificial selection focussing on key traits, such as yield and quality in crop breeding, have resulted in a narrow genetic diversity in modern day cultivars.

Bread wheat (*Triticum aestivum* L.) is an allohexaploid species with the A, B and D subgenomes that arose ~8,000 years ago through polyploidization events considered recent on the evolutionary scale (Feuillet et al. 2008; Huang et al. 2002). The diversity of the D subgenome in bread wheat cultivars is narrow due to the limited number of hybridization events that occurred between wild emmer (AABB) and goatgrass (DD). In addition, the subsequent and simultaneous processes of domestication and, more recently, breeding have further contributed to the narrow diversity of the D subgenome. Synthetic hexaploid wheat (SHW) lines are derived from artificial hybridization between diverse accessions of *Triticum turgidum* L. and *Aegilops tauschii* Coss. followed by chromosome doubling. SHW lines are widely generated to harness the genetic diversity in the wild goatgrass accessions, specifically to utilise trait variations for abiotic stress tolerance, biotic stress resistance, yield components and quality parameters (Kishii 2019). SHW lines have been adopted in many wheat improvement programs worldwide. However, there are several challenges associated with recovering the desired parental phenotypes in the resynthesized bread wheat genomic background. To address the question of the impact of genomic background, including ploidy level on gene expression, the messenger RNA (mRNA) expression dynamics of SHW lines upon polyploidization and of their progenitors has been extensively investigated in this thesis. Further, this thesis explores the introgression patterns of SHW genetic material into elite wheat backgrounds to address a second phenomenon related to SHW, which is the potential bias in introgression that can also reduce the breeders' ability to recover the desirable phenotypes or traits.

1.1 Literature review

1.1.1 Wheat evolution

Bread wheat (*Triticum aestivum* L.) is an allohexaploid crop comprising the A, B and D homoeologous genomes (2n=6x=42). Archaeobotanical and molecular genetic evidence suggests that wheat was domesticated in the Near East, specifically in the region known as the 'Fertile Crescent' (Salamini et al. 2002). The allopolyploidization event(s) involving natural chance hybridizations between cultivated emmer wheat (*T. turgidum* L. ssp. *dicoccum*, 2n=4x=28; AABB genome) and goatgrass (*Ae. tauschii* Coss., 2n=2x=14; DD genome), followed by chromosomal doubling, ~8,000 years ago, resulted in the modern era bread wheat (*T*. *aestivum*; AABBDD genome; Figure 1.1) (Feuillet et al. 2008; Vergauwen and De Smet 2017). Various modifications at the genomic and epigenomic levels, resulting in control of homoeologous pairing, gene subfunctionalization, alteration in the composition and expression state of transposons, all contributed to the plasticity required for survival and establishment of the allopolyploid species (Feldman and Levy 2015).

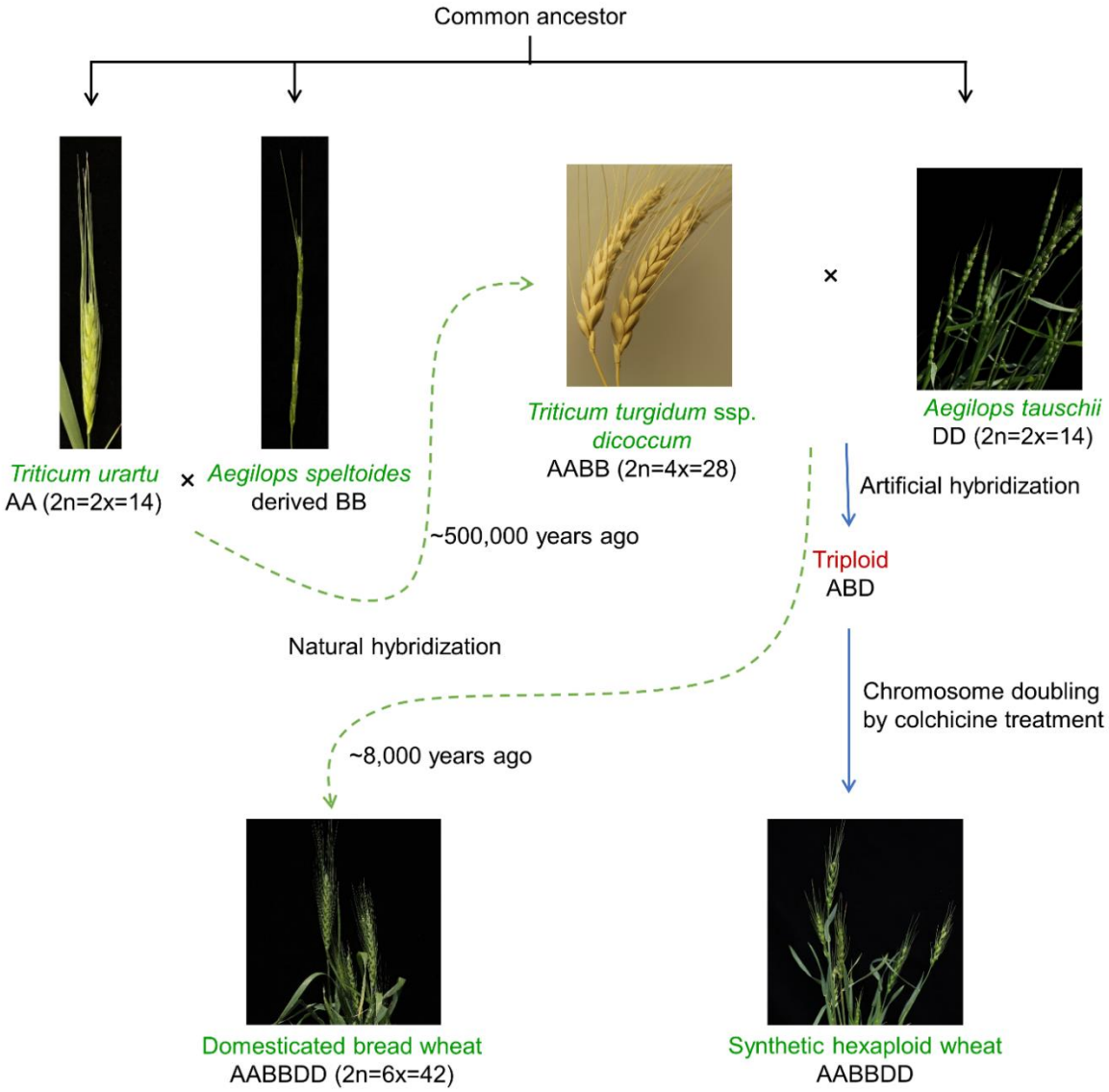


Figure 1.1 Evolution of hexaploid bread wheat; Photograph credits: Dr. Sylvie Cloutier.

1.1.2 Concept of pre-breeding

Selection of superior genotypes exhibiting desirable trait combinations is the foundation of plant breeding, and crop improvement programs are primarily dependent on trait genetic variation. In this context, wild relatives of crop species are a critical source of genetic diversity for a range of agriculturally important traits. In genebanks worldwide, there are

nearly seven million crop germplasm accessions (Reynolds et al. 2021), of which 110,000 accessions of almost 1,000 crop species are maintained by the Canadian National Plant Germplasm System (CNPGS) alone (Diederichsen and Davidson 2022). There are certain limitations when introducing novel variation from crop wild relatives, including linkage drag, infertility and poor cross-compatibility (Dempewolf et al. 2017). Pre-breeding is the bridging step for bringing the genetic diversity of crop wild relatives into breeding populations. The genomic regions governing the traits of interest in the wild crop-related species are introgressed into elite cultivar backgrounds through initial crossing, followed by multiple backcrossing cycles, often relying on DNA-marker assisted foreground- and background selections. This process is lengthy and challenging, and the end result can be fraught with unpredictability. Modern genome editing technologies offer an alternative strategy with the potential to revolutionize gene transfer from wild relatives to crops. This approach would be more precise, would not involve repeated crossings and their associated drawbacks, and therefore would save years of selections to recover adapted germplasm. The development of novel semi-adapted breeding germplasm with more plant breeder-accessible genetic variability, but without linkage drag, sterility and lengthy crossing time would constitute a breakthrough in pre-breeding (Moore 2015).

1.1.3 Synthetic hexaploid wheat

Only a subpopulation of *Ae. tauschii* Coss. lines from a restricted geographic range, have contributed to the development of hexaploid wheat (*T. aestivum* L.), representing only a subset of the existing diversity of this species, thereby causing an evolutionary bottleneck referred to as the founder effect (Rasheed et al. 2018; Wang et al. 2013). Hence, a large portion of the *Ae. tauschii* Coss. diversity remains unexplored. Synthetic hexaploid wheats (SHWs), derived from the artificial hybridization of *T. turgidum* L. ssp. *durum* (AABB) or other subspecies with *Ae. tauschii* Coss. (DD), followed by chromosome doubling in the F_1

generation with colchicine treatment or through spontaneous doubling, create valuable pre-breeding populations with potential to expand the genetic diversity of bread wheat. As expected, the genetic diversity of the D genome in SHWs has been reported to be similar to that of the A and B genomes, a contrast to the lower genetic diversity observed in the D genome of bread wheat (Akhunov et al. 2010; Bhatta et al. 2018a).

The International Maize and Wheat Improvement Centre (CIMMYT, Mexico), has generated more than 1,500 SHW since its inception in 1986, utilising around 900 *Ae. tauschii* accessions (Dreisigacker and Kishii 2019; Mujeeb-Kazi et al. 2008; Ogbonnaya et al. 2013). These SHW accessions are distributed worldwide for utilisation in wheat breeding. Considering the importance of SHW in improving wheat diversity, the National Institute of Agricultural Botany in the United Kingdom (https://www.niab.com), International Centre for Agricultural Research in the Dry Areas (ICARDA, Lebanon), United States Department of Agriculture (USDA) and wheat improvement programmes in China, also started producing SHWs as breeding germplasm (Ogbonnaya et al. 2013). Globally, as of 2020, 86 registered wheat cultivars were SHW-derived lines (Aberkane et al. 2020). At the phenotypic level, SHWs and their derivatives were reported to have increased grain size (Morgounov et al. 2018; Talbot et al. 2008) and yield (Jafarzadeh et al. 2016) when compared to locally adapted wheat cultivars. Large-effect quantitative trait loci (QTL) and alleles underlying pest and disease resistance have also been mapped in SHWs and SHW-derived germplasm (Bhatta et al. 2018b; Crespo-Herrera et al. 2019). SHWs can also serve as germplasm for developing climate-resilient wheat cultivars because they exhibit wider adaptation to heat, drought and combined abiotic stresses (Elbashir et al. 2017; Pradhan et al. 2012; Wan et al. 2023).

1.1.4 Whole genome duplication in plants

Whole genome duplication is capable of generating diversity, bringing about evolutionary innovations, increasing genome complexity and acting as a mechanism of instant plant speciation (De Storme and Mason 2014; Soltis et al. 2014). The evolutionary history of most angiosperms involve at least one polyploidization event (Jiao et al. 2011), and 30% of all crop species and 54% of monocot crops are polyploids (Salman-Minkov et al. 2016). Polyploidization can occur at intraspecific (autopolyploidy) and interspecific (allopolyploidy) levels. Autopolyploids comprise multiple sets of the same group of chromosomes while allopolyploids contain genomes of diverse species within their nucleus (Glover et al. 2016). Polyploids are formed either through sexual reproduction involving unreduced gametes from meiosis or chromosome doubling in the zygote (Kreiner et al. 2017; Szadkowski et al. 2011). The enhanced abilities of polyploids to endure abiotic and biotic stressors has been suggested by Van de Peer et al. (2021). It was recently shown that polyploid plant species are more prevalent in colder regions and higher latitudes (Rice et al. 2019), potentially due to their ability to withstand the environmental conditions compared to diploid species of the region (Levy and Feldman 2022). Polyploidization is significantly associated with crop domestication (Salman-Minkov et al. 2016), however it can occur either before or after domestication (Akagi et al. 2022). Further, the association of domestication with polyploids is attributed to their improved adaptability to agricultural environments, which is facilitated by the buffering capacity of the duplicated genomes (Dubcovsky and Dvorak 2007). This can be observed as a positive correlation between area and production of the wheat species (*T*. *monococcom* L., *T. turgidum* L. and *T*. *aestivum* L.) and their ploidy levels (Levy and Feldman 2022).

1.1.5 Genetic changes during polyploidization

In the model plant Arabidopsis, allopolyploids show transposon activation and chromosomal breaks with potential for translocation (del Pozo and Ramirez-Parra 2015). Homoeologous exchanges have also been reported in rice allotetraploids, leading to both genetic and phenotypic variability (Li et al. 2019a). *Brassica napus* L., a relatively recent allotetraploid (2n=4x=38) crop, originated from the interspecific hybridization between *B. rapa* L. (AA) and *B. oleracea* L. (CC) species. The polyploidization event led to both loss and gain of DNA fragments through non-reciprocal genetic exchanges (Gaeta et al. 2007), in addition to the homoeologous recombination at the first meiosis (Szadkowski et al. 2010). Resynthesized genomes experience depletion in repetitive DNA elements and highly redundant ribosomal RNA encoding genes. The initial generations experience aneuploidy, and stability is achieved by chromosomal dosage balance (Xiong et al. 2011). Transposition of DNA fragments in the genic region and loss of genes post polyploidization have been reported in *Glycine* sp. as well (Zhuang et al. 2022). Similar loss of genic regions has also been observed in allotetraploid *Cucumis* (Yu et al. 2021). The genome of *Brassica napus* L. synthetic allopolyploids with alternate C-genome donors have also shown restructured haplotype blocks caused by deletions and duplications (Zou et al. 2018). Both ribosomal DNA evolution and elimination are observed during resynthesis in *Nicotiana* allotetraploids as well (Renny-Byfield et al. 2011; Skalická et al. 2003).

The genetic alterations reported in wheat are similar to other plant species. Allopolyploidization events in wheat cause homoeologous exchanges in sub-telomeric regions of the chromosomes and are associated with DNA motifs similar to homologous recombination, and are contrastingly localized within the genic regions (Zhang et al. 2020). Previous reports have also shown loss of genes, and loss and amplification of repetitive genomic regions, after alloplyploidization (Levy and Feldman 2004). Similar to the

observations in *Brassica* and *Nicotiana* species, fewer rRNA genes were observed in the allopolyploids compared to their progenitors (Feldman and Levy 2012).

1.1.6 Epigenetic modification upon polyploidization

Epigenetic changes can occur upon interspecific hybridization. The epigenetic components of a plant genome include DNA methylation, histone modification, small non-coding RNAs (Calarco et al. 2012) and chromatin conformation. The three dimensional chromatin conformation is guided by genetic variants (Tang et al. 2015), DNA methylation and histone modifications (Boettiger et al. 2016). These components facilitate gene regulation even by distantly located loci (Delaneau et al. 2019). The epigenetic alterations, specifically methylation in plants, are heritable through meiosis (Quadrana and Colot 2016). Distinct epigenetic patterns can regulate gene transcription and alter the corresponding gene regulatory networks (Song and Chen 2015; Springer and Schmitz 2017), ultimately impacting phenotypes.

The epigenetic variation triggered by whole genome duplication has been described in several plant species. Upon allotetraploidization, the model plant Arabidopsis exhibits genome-wide demethylation, impacting gene expression (Madlung et al. 2005). In addition, studies of chromatin conformation in whole genome duplicates indicate that long-range inter-chromosomal interactions are more prevalent than intra-chromosomal interactions, and changes in histone 3 lysine 4 trimethylation (H3K4me3) and H3K27me3 have also been detected (Zhang et al. 2019). Variations in methylation patterns upon resynthesis were also observed in the young allotetraploid of the Brassicaceae family, i.e., *B. napus* L. (Lukens et al. 2006). Artificial autopolyploidization in rice (*Oryza sativa* L. ssp. *indica*) has revealed hypermethylation of DNA transposons linked to increased small interfering RNA (siRNA) production, leading to repression of neighbouring genes (Zhang et al. 2015b). Changes in

methylation patterns were also observed in synthetic rice allopolyploids generated from *O. sativa* L. ssp. *japonica* and *O. sativa* L. ssp. *indica* (Li et al. 2019a). The allotetraploid cotton species *Gossypium hirsutum* L. and *G. barabadense* L. showed increased methylation differences in their genic and intergenic regions compared to their diploid progenitors *G. arboreum* L. and *G. raimondii* L., and these stably inherited epigenetic marks were shown to govern distinct phenotypes, including response to photoperiod (Song et al. 2017). In addition, modification in three dimensional chromatin architecture post polyploidization and its significant impact on gene expression have been reported in cotton (Wang et al. 2018b), soybean (Wang et al. 2021a) and watermelon (Garcia-Lozano et al. 2021).

In wheat, variation in multiple epigenetic marks were observed with polyploidization. The study of methylation patterns in different ploidy backgrounds revealed D genome-mediated changes in methylation levels, especially in the CHG context (Yuan et al. 2020). Subgenome-specific silencing of nucleolus organizing regions, higher levels of CHG and CHH DNA methylation in promoter regions and subsequent increase in H3K27me3 and H3K9me2, were observed in resynthesized allotetraploids and allohexaploids derived from different *Triticum* and *Aegilops* sp. (Guo and Han 2014). In addition, variation in levels of small interfering RNAs (siRNAs) were also reported with a change in ploidy level (Kenan-Eichler et al. 2011), with additional influence based on the parental genomes involved (Jiao et al. 2018). When the histone marks on the hexaploid wheat and its tetraploid and diploid progenitors were assessed, H3K27me2 increased with ploidy, especially in the genomic regions showing transposon amplification (Liu et al. 2021). Further, the modification in histone marks by the introduction of the D subgenome, through inter-subgenome interactions, has also been unravelled (Lv et al. 2021). Recent studies have identified the change in chromatin interaction patterns after polyploidization in wheat, and the potential role of these changes on gene expression (Yuan et al. 2022).

1.1.7 Gene expression changes in polyploids

In principle, duplicated genomes tend to have more than one copy of non-redundant genes. However, the expression patterns of these are unique and not always additive, implying that the level of gene expression in allopolyploids is not equal to the average of the parental expression. In some cases, the non-additive expression bias is suggested to be due to expressional dominance or transgressive gene expression (Yoo et al. 2014). Transcriptomic studies in newly synthesized *Brassica napus* L. tetraploids have revealed non-additive gene expression in comparison to its diploid parents (Gaeta et al. 2009; Zhang et al. 2015a). The newly synthesized *Brassica napus* L. show alternative splicing of the primary transcripts of homoeologous genes, possibly bringing about new gene functions (Zhou et al. 2011). RNAseq analysis of leaf tissue samples of allotetraploid cotton and its diploid progenitors *G. arboreum* L. and *G. raimondii* L. revealed that ~40% of the genes were non-additively expressed (Yoo et al. 2013). The transcriptome of resynthesized allotetraploid wheat exhibited 60-80% of non-additive gene expression in roots and endosperm (Jiao et al. 2018) and 40-75% in leaf and inflorescence tissues (Zhang et al. 2016). In the case of SHW, although the proportion of non-additively expressed genes was less than 1% based on stem, spike, seedling and seed tissues, these genes played a significant role in plant growth and development (Chelaifa et al. 2013; Li et al. 2014a). The transcriptome analysis of a wider range of tissues in allohexaploids would provide further information about the nature of the genes that undergo transcriptional rewiring caused by the whole genome duplication induced “genome shock” (Fox et al. 2020).

1.1.8 Functional divergence of genes in polyploid background

In hexaploid wheat, ~35% of the genes are present in triads, with one copy in each homoeologous chromosome, and an additional 5.7% have extra paralogous copies in one

of the homoeologues (IWGSC 2018). The extensive RNA sequencing data analysis of wheat from varied tissues, environmental conditions, treatments and time points, suggested that the presence or absence of transposons and variation at the cis-regulatory elements, bring about novelty in gene function, including the regulatory neo-functionalization and sub-functionalization of duplicated homoeologues (Ramírez-González et al. 2018). Neo-functionalization refers to when one of the duplicated genes in the polyploids gains a new function, while in sub-functionalization, the actual role of the gene in the progenitors is partitioned among the duplicates. Recent studies in hexaploid wheat suggest that the homoeologous duplicates of MIKC-MADS box genes might have undergone neo-functionalization, leading to their new role in adaptation (Schilling et al. 2019). A significant proportion of the maize homoeologous duplicates have gained new functions through changes in their regulation, and exhibit novel tissue-specific expression patterns (Hughes et al. 2014). Interestingly, the study of *Brachypodium distachyon* L., *Oryza sativa* L. ssp. *japonica*, *Sorghum bicolor* (L.) Moench. segmental gene duplication events suggested that 33-50% of the genes retained were neo-functionalized, and these newly gained functions were specific to male tissues like anthers (Jiang and Assis 2019). Apart from protein coding DNA, regulatory sequences such as long non-coding RNAs (lnc-RNAs) can gain novel functions upon whole genome duplication or interspecific hybridization, as reported in allopolyploid cotton (Zhao et al. 2018).

1.1.9 Genetic diversity in the wild relatives of wheat

The evolutionary bottleneck of allohexaploidization has limited the genetic diversity in modern day wheat. Novel genetic variability exists among different ancestors for traits such as disease resistance, abiotic stress tolerance and other agronomically important traits. Their introgression into breeding lines has been achieved in some cases but much unexplored diversity remains to be discovered and introgressed.

With regards to agronomically important traits, variation associated with heading date, plant height and other spike-related traits are present among *Triticum urartu* Thumanjan ex Gandilyan (AA genome) accessions (Wang et al. 2017). When a collection of wheat wild relatives from Iran, including various *Aegilops* and *Triticum* species, was evaluated for different agronomic traits, such as days to heading, plant height, 1000-grain weight and grain yield per plant, the rich diversity for these traits in the germplasm was brought to light (Pour-Aboughadareh et al. 2018). Photosynthetic efficiency of a crop is a vital trait for crop improvement, because of its direct association with biomass and yield. When a set of 88 accessions of wheat wild relatives were characterized for photosynthesis-related traits, including CO_2-saturated rates of photosynthesis, maximum rate of carboxylation and Rubisco activity, high variability both between genera and accessions was observed (McAusland et al. 2020). As more studies are conducted, often requiring changes to crop architecture, such as indoor and vertical farming strategies, the recommended agronomic practices are also modified. Wild relatives are also a source of variability for improving crop architecture, as seen in wheat lines with introgressions from *Agropyron cristatum* (L.) Gaertner., showing reduced leaf size and plant height, suitable for high-density planting (Wang et al. 2022b), in addition to improvements in grain number per spike and 1000-grain weight (Song et al. 2016). One of the critical differences between wheat wild relatives and elite cultivars is the grain quality parameters, important for post-harvest processing. However, alien introgression lines have also been shown to provide variability for quality traits such as protein concentration, gluten strength and essential micronutrients (Johansson et al. 2020).

The wild relatives of wheat are also a source of resistance to a wide range of biotic stressors. Genomic regions conferring resistance to rust and powdery mildew have been mapped and successfully introgressed into bread wheat lines from *Ae. umbellulata* Zhuk.

(UU), *Ae. speltoides* Tausch. (SS) and many more species (Bansal et al. 2017; Petersen et al. 2015). When a set of 60 wild wheat relative accessions, including ten different *Triticum* and *Aegilops* species, were characterized for powdery mildew infection, they showed variability for infection types, suggesting the presence of novel resistance genes in the germplasm (Aliakbari Sadeghabad et al. 2020). Diverse accessions of *Ae*. *tauschii* Coss. also show resistance to the currently re-emerging pandemic disease called wheat blast (Cruppe et al. 2020; Latorre et al. 2023). Recently, the large effect gene *Fhb7* was identified from *Thinopyrum elongatum* (Host) D.R.Dewey, which upon introgression into wheat, provided resistance to Fusarium head blight and crown rot diseases (Wang et al. 2020). Strategies like AgRenSeq have been developed to rapidly identify and clone the vast repertoire of novel disease resistance genes in wheat wild relatives (Arora et al. 2019). Furthermore, the wild relatives can serve as a source of resistance to insect pests (El Bouhssini et al. 2013; Tadesse et al. 2022).

Wild crop relatives are an important resource conferring climate resilience to elite cultivars (Zhang and Batley 2020). Similar to agronomic and disease resistance traits, evaluation of wheat wild relative germplasm has unveiled wide variability for physiological traits related to drought tolerance as well (Pour-Aboughadareh et al. 2018). The diploid wheat ancestor *Triticum monococcum* L. (A^mA^m) has contributed to a salinity tolerance locus that provided ~25% yield gain under stress conditions (Munns et al. 2012). Alien introgression from *Ae. speltoides* Tausch. and *Daspyrum villosum* (L.) P. Candargy into wheat have resulted in better root system architecture, imparting drought tolerance (Djanaguiraman et al. 2019). The better adaptability of wild relatives to drought, heat and salinity conditions have also been widely reported (Langridge and Reynolds 2021; Molero et al. 2023; Nevo and Chen 2010).

1.1.10 Recombination and genome introgression

Meiosis, the cell division specific to reproductive lineage cells, is detected as far back as the last eukaryotic common ancestor in the evolutionary tree, and is observed across the eukaryota domain (Loidl 2016). The process of meiosis reduces the diploid set of chromosomes to a haploid set during gamete formation. Segregation of chromosomes and homologous recombination are the two critical steps in meiosis I responsible for maintaining the genome integrity and generating genetic diversity leading to the formation of novel allele combinations (Henderson and Bomblies 2021). During the process, another round of chromosome segregation occurs in meiosis II to finally give rise to four haploid nuclei, a critical step in sexual reproduction. One crossover (CO) event per chromosome is required for faithful chromosome segregation (Henderson and Bomblies 2021). However, variation in the number of CO events is observed both within and between species, although the rates are conserved and low, irrespective of the genome size (Henderson and Bomblies 2021; Lawrence et al. 2017). The number of recombinations seldom exceed three per homologous pair of chromosomes, in eukaryotes in general, and specifically in plants (Loidl 2016; Mercier et al. 2015).

Following the DNA replication in prophase I, a type II topoisomerase-like protein SPO-11, along with other proteins, creates multiple double strand breaks (DSBs) (Benyahya et al. 2020; Keeney et al. 1997) along the chromosomes, but only ~5% of these DSBs lead to COs in plants (Hyde et al. 2023). The 3’ ends of the single stranded DNA molecules created by the DSBs seek a sister chromatid or homologous chromosome as a template for repair, which may result in either of two fates. In the first instance, recombinases, including DMC1 and RAD51, guide the 3’ DNA ends towards the templates, and enable the formation of D-loops in the stranded DNA. This is followed by DNA synthesis and double Holliday junction formation, whose resolution can lead to COs

(Schwacha and Kleckner 1995; Wang and Copenhaver 2018). In the second fate, a synthesis-dependent strand annealing (SDSA) mechanism results in non-crossovers (NCOs) by disconnecting the invading strand and joining it to the end of the original break (McMahill et al. 2007; Wang and Copenhaver 2018).

There are several factors that influence the occurrence of COs in a genomic region. A study of interspecific hybrids in tomato revealed that the COs were localized in proximity to poly-A, poly-T and AT-rich motifs (Rommel Fuentes et al. 2020). In general, genomic regions enriched with AT, CTT and CNN motifs are more accessible to COs in plants (Lawrence et al. 2017). In *Zea mays* L., variation in CO frequencies was observed between male and female gametes and the genetic background also influenced the CO maturation (Luo et al. 2019). Further, structural variants in the genome, single nucleotide polymorphisms (SNPs) and heterozygosity are all known to influence recombination in plant species (Fuentes et al. 2022; Szymanska-Lejman et al. 2023). Additionally, epigenetic marks can affect the CO landscapes. DNA methylation is found to be inversely related to CO rates across plant species (Melamed-Bessudo and Levy 2012; Underwood et al. 2018; Yelina et al. 2015). Epigenetic marks like H3K27me3 and DNA methylation influence nucleosome occupancy patterns and consequently chromatin accessibility, thus impacting COs (Lloyd 2023; Tock et al. 2021). Recent studies in *Oryza sativa* L. also showed that three dimensional chromosome architecture determines the frequency of recombination in plant genomes (Golicz et al. 2020). In the majority of plant species, COs are suppressed in the centromeric regions and are mostly localized in the sub-telomeric regions (Fernandes et al. 2019). In summary, CO frequencies are influenced by DNA sequence-level variation, epigenetic marks and centromere-relative location of a genomic region on the chromosome.

Plant breeding is highly dependent on the biological process of recombination for generating and transferring genetic variability and for the identification of genomic regions

associated with traits of interest. In order to adapt to environmental challenges, several crop species have received genomic fragments from other relatives through a process known as adaptive evolution (Janzen et al. 2019). Crop improvement programs also utilise introgression from wild relatives or other elite germplasm for the introduction of desired alleles (Epstein et al. 2023). Introgression involves the crossing of elite crop cultivars with wild species harbouring a novel allele, followed by repeated backcrossing of the progeny to the elite parent. With several genetic and epigenetic factors influencing CO events, and considering that recombination is the basis of introgression, a critical understanding on the recombination/introgression patterns is paramount to facilitate the use of crop wild relative and pre-breeding germplasm development.

1.2 Purpose of the study

Progenitor species of wheat are a major source of variation for multiple agronomic, disease resistance, abiotic/climate-resilience and grain quality traits. However, the introgression of novel alleles from diverse progenitors through pre-breeding poses several challenges, including linkage drag and gene expression changes associated with background ploidy levels to name a few (Leigh et al. 2022). A comprehensive study to understand the complex changes in expression patterns of homoeologues both qualitatively and quantitatively, upon allopolyploidization, in the commonly used wheat pre-breeding material of SHW, was required. Further, after their development, SHW are crossed with elite wheat cultivars for transfer of desired alleles to breeding populations. Considering the impact of several genetic and epigenetic factors on the CO rates across the genome, and consequently the introgression patterns, the genome-wide introgression of the genetically diverse SHW alleles in an elite wheat background also needed investigation.

1.3 Objectives

The genome dynamics and genomic interference upon polyploidization result in several genetic and epigenetic alterations associated with changes in gene expression. Resynthesized allohexaploid wheat is a suitable model to examine the impact of the genome shock. In addition, SHWs are also the ideal pre-breeding material for investigating the introgression of diverse genetic material into elite genetic backgrounds. With these premises, the objectives of my research were as follows:

1. To understand the influence of genome interactions at various ploidy levels on gene expression in multiple synthetic hexaploid wheats (AABBDD), by determining the pattern and suppression of their sub-genome alleles, at the quantitative level, compared to that of their tetraploid (AABB) and diploid parents (DD) across several tissues.
2. To determine the qualitative differences in transcripts of SHWs in comparison to their tetraploid and diploid parents across tissues.
3. To characterize the introgression patterns of a cross between a SHW and an elite wheat cultivar.

1.4 Hypotheses

- The A and B genomes of hexaploid wheat have wider genetic diversity in contrast to its D genome in comparison to its diploid ancestor, and, the SHW lines generated to overcome this genetic bottleneck, undergo dynamic genomic changes that can be observed at the gene expression level.

- The introgression patterns observed in progenies while crossing the pre-breeding SHW material with elite wheat genotypes are different from the introgression patterns in elite wheat × elite wheat cross derivatives.

To answer these hypotheses and meet these objectives, the thesis focuses on the extensive transcriptome analysis carried out using RNAseq data from four SHW lines, their two tetraploid and diploid parents, across ten tissues. Introgression analysis was performed using a SHW × elite wheat backcross population compared to an elite × elite wheat doubled haploid population. In addition, the SNP dataset from the previously published 1,000 wheat exome project was lifted from IWGSC RefSeq v1.0 to RefSeq v2.1, in order to assist the necessary recalibration steps in the variant calling process needed for the analyses of the introgression patterns. The SNP lift-over process has been detailed in a separate chapter (Chapter #3).

CHAPTER 2

Global transcriptome analysis of allopolyploidization reveals large-scale repression of the D-subgenome in synthetic hexaploid wheat

Akshaya Vasudevan[1,2], Madeleine Lévesque-Lemay[1], Tara Edwards[1] and Sylvie Cloutier[1,2,*]

[1]Agriculture and Agri-Food Canada,
Ottawa Research and Development Centre,
Ottawa, Ontario, Canada K1A 0C6.

[2]Department of Biology,
University of Ottawa,
Ottawa, Ontario, Canada K1N 6N5.

[*]E-mail address of the corresponding author: sylviej.cloutier@agr.gc.ca

Note: This chapter has been published as the following article:

Vasudevan A, Lévesque-Lemay M, Edwards T, Cloutier S (2023) Global transcriptome analysis of allopolyploidization reveals large-scale repression of the D-subgenome in synthetic hexaploid wheat. Communications Biology 6(1):426

https://www.nature.com/articles/s42003-023-04781-7

S.C. designed the study. S.C., M.L.-L. and T.E. performed the experiments and collected data. A.V. and S.C. finalized the data analysis methods, carried out data analysis, interpreted the results and visualized them. A.V. and S.C. wrote the manuscript. All the authors read and approved the manuscript.

2.1 Abstract

Synthetic hexaploid wheat (SHW) lines are created as pre-breeding germplasm to diversify the D subgenome of hexaploid wheat and capitalize upon the untapped genetic diversity of the *Aegilops tauschii* gene pool. However, the phenotypes observed in the *Ae*. *tauschii* parents are not always recovered in the SHW lines, possibly due to inter-subgenome interactions. To elucidate this post-polyploidization genome reprogramming phenomenon, we performed RNAseq of four SHW lines and their corresponding tetraploid and diploid parents, across ten tissues and three biological replicates. Homoeologue expression bias (HEB) analysis using more than 18,000 triads suggests massive suppression of homoeoalleles of the D subgenome in SHWs. Comparative transcriptome analysis of the whole-genome gene set further corroborated this finding. Alternative splicing analysis of the high-confidence genes indicates an additional layer of complexity where all five splice events are identified, and retained intron is predominant. Homoeologue expression upon resynthesis of hexaploid wheat has implications to the usage and handling of this germplasm in breeding as it relates to capturing the effects of epistatic interaction across subgenomes upon polyploidization. Special considerations must be given to this germplasm in pre-breeding activities to consider the extent of the inter-subgenome interactions on gene expression and their impact on traits for crop improvement.

2.2 Introduction

Whole-genome duplication events are one of the prime drivers of speciation (Cui et al. 2006; De Storme and Mason 2014; Ren et al. 2018). The majority of angiosperms have undergone polyploidization in the course of their evolution, and notably, 30% of crop species are deemed to be polyploid based on the extent of duplicated loci in their genome (Salman-Minkov et al. 2016). The grass lineage underwent a minimum of three whole

genome duplication events (Jiao et al. 2014), basically making the whole Poaceae family polyploid. Whole genome duplication events are frequently found to be associated with significant evolutionary potential that promotes adaptability to changing environments (Alix et al. 2017; Fawcett et al. 2009). There are several intriguing questions related to polyploids, from the initial dynamic changes they undergo in the genome for stabilization, until their establishment as discrete populations (Doyle et al. 2008). Polyploidization creates extensive redundancy within the genome, thereby paving the way for novel alterations that promote the development of new phenotypes and/or adaptation (Van de Peer et al. 2021).

Bread wheat (*Triticum aestivum* L.), is an allohexaploid crop (2n=6x=42) comprised of the A, B and D subgenomes. The natural chance hybridizations that occurred ~8,000 years ago between cultivated emmer wheat (*Triticum turgidum* L. ssp. *dicoccum*, 2n=4x=28; AABB genome) and Tausch's goatgrass (*Aegilops tauschii* Coss., 2n=2x=14; DD genome), followed by chromosome doubling, resulted in the development of modern bread wheat (Feuillet et al. 2008; Huang et al. 2002). Presumably, only a limited number of *Ae. tauschii* plants contributed to the evolution of hexaploid wheat, leading to an evolutionary bottleneck called the founder effect (Cox 1997); a bottleneck that was further constricted by limited natural flow of genetic variation from diploid to hexaploid species (Akhunov et al. 2010; He et al. 2019) and subsequent domestication and breeding.

Wild relatives of crop species are a source of genetic diversity for multiple agriculturally important traits. However, there are certain limitations to their usage, including linkage drag, infertility and poor cross-compatibility (Dempewolf et al. 2017). Pre-breeding is a systematic approach that aims to recapture the genetic diversity of crop wild relatives and deploy it into breeding programs. Traditionally, the desired genomic segments from wild crop species are introgressed into elite cultivar backgrounds by repeated backcrossing with the support of state-of-the-art whole-genome genotyping tools for foreground and

background selections. SHWs generated from the artificial hybridization between *T. turgidum* ssp. *durum* (AABB) or other subspecies and *Ae. tauschii* (DD), followed by chromosome doubling, serve as effective pre-breeding populations to broaden the genetic diversity of the D subgenome of hexaploid wheat. SHWs developed by the International Maize and Wheat Improvement Centre (CIMMYT, Mexico) (Mujeeb-Kazi et al. 2008; Ogbonnaya et al. 2013) have been extensively utilised in wheat genetic diversity enrichment programs in several countries. Later, other research institutes and wheat improvement programs across the globe also produced SHWs using different sources of *Ae*. *tauschii* accessions to serve as base populations for commercial wheat breeding.

Ae. tauschii is a proven source of resistance to a broad range of biotic and abiotic stresses (Mirzaghaderi and Mason 2019). However, parental phenotypes are not always effectively recovered in the synthetic hexaploid lines, possibly due to the dynamic genetic and epigenetic changes and the consequent variation in gene expression (Dreisigacker and Kishii 2019). For instance, studies spanning the past two decades have shown that resistance to stem rust pathogen (*Puccinia graminis* f. sp. *tritici*) is suppressed in the hexaploid state by Med15, a component of the Mediator complex encoded by the D subgenome (Bai and Knott 1992; Hiebert et al. 2020). Upon polyploidization, the D homoeologue of the *High-Affinity K+ Transporter 1;5* (*HKT1;5*) shows reduced expression, when compared to diploid parental levels, but regains the diploid parent native expression level under salt stress condition (Yang et al. 2014). In certain phenotypes, like grain width, each of the homoeologues is associated differently with the trait of interest and their combined levels of expression determine the resultant phenotype (Hong et al. 2014).

The alteration across multiple layers of gene regulation upon polyploidization is also considered to improve adaptability of allohexaploid bread wheat when compared to its tetraploid and diploid parents (Dubcovsky and Dvorak 2007; Nieto Feliner et al. 2020). The

genetic and epigenetic alterations reported upon polyploidization in wheat are similar to other polyploid plant species. At the structural level, apart from homoeologous exchanges between genomes, both loss and amplification of repetitive DNA due to perceived genome stress have been observed (Levy and Feldman 2004). The recent whole genome sequencing of bread wheat sheds light on the transposon insertions and deletions in the intergenic regions of the A, B and D subgenomes, although the fraction of each transposon family among the homoeologous genomes does not significantly vary (Wicker et al. 2018). At the epigenetic level, in addition to changes in cytosine methylation status of the DNA, variability in histone methylation (Guo and Han 2014; Lv et al. 2021), euchromatin and heterochromatin patterns across chromosomes and prominent differences in the levels of epigenetic regulators like small RNAs were also observed (Feldman and Levy 2012; Jiao et al. 2018). These epigenetic marks on the regulatory elements of the genome modulate gene expression levels in hexaploid bread wheat (Li et al. 2019b).

The effect of some of these genomic and epigenomic changes occurring upon polyploidization are reflected in the transcriptome. Understanding the pattern of expression changes between the parental ploidy backgrounds and hexaploid wheat provides valuable information to inform interspecific crosses and assist in predicting the phenotypes recovered in the progenies. In this study, we used four SHW lines and their corresponding tetraploid (*T. turgidum*) and diploid (*Ae. tauschii*) parents for a comparative evaluation of transcriptome dynamics. We focused on unravelling the changes in expression patterns of homoeologous genes between the subgenomes in parental (2x and 4x) and synthetic hybrid backgrounds (6x). In addition to quantitative differences, the subtle qualitative differences in transcripts arising from alternative splicing events were also determined.

2.3 Results

2.3.1 HEB in SHW lines

In this experiment, RNAseq data was generated from four SHW lines and their two diploid and two tetraploid parents (Appendix 1), from three biological replicates across ten tissues (Appendix 2) for a total of 240 samples. After pre-processing of the sequencing reads for the 240 samples, we obtained a total of 4.9 B reads corresponding to an average of 20.4 M reads per sample, ranging from 4.8-70.7 M. The number of raw reads and processed reads for each sample was compiled (Appendices 3 and 4). Approximately 85% of the total reads were uniquely mapped to the Chinese Spring IWGSC RefSeq v2.1 (Zhu et al. 2021), ~12% mapped to multiple loci, and ~3% did not map at all (Appendix 5).

In order to understand the global pattern of gene expression dynamics upon polyploidization, specifically, HEB, and compare them between SHWs and their parental state, genes (homoeologues) belonging to triads, i.e., with one copy in all three sub-genomes (18,357 triads (IWGSC 2018)), were subjected to the likelihood ratio test (LRT) (Smith et al. 2019) computation. Here, the differences in gene length between the homoeologues are considered. The null hypothesis, i.e., no expression differences between the homoeologues, was tested for all comparisons. In the SHW lines the AB vs D homoeologues were compared. To test for the parental expression state, *in-silico* SHW-like scenarios with corresponding tetraploid and diploid parental expression levels were constituted. Expression biases, predominantly towards the D subgenome in parental level of expression, were observed across all ten tissues and SHWs (Figure 2.1). The number of triads showing significant bias towards the subgenome contributed by the diploid (D) parent were drastically reduced in the corresponding SHW lines across all tissues (Figure 2.1 and Appendix 6). The pistil tissue when anthers were green and immature had more than

10,000 triads biased towards the D subgenome, based on parental expression levels in all four SHWs. The heads collected at the boot stage showed a similar pattern in all SHWs but C44 (Figure 2.1).

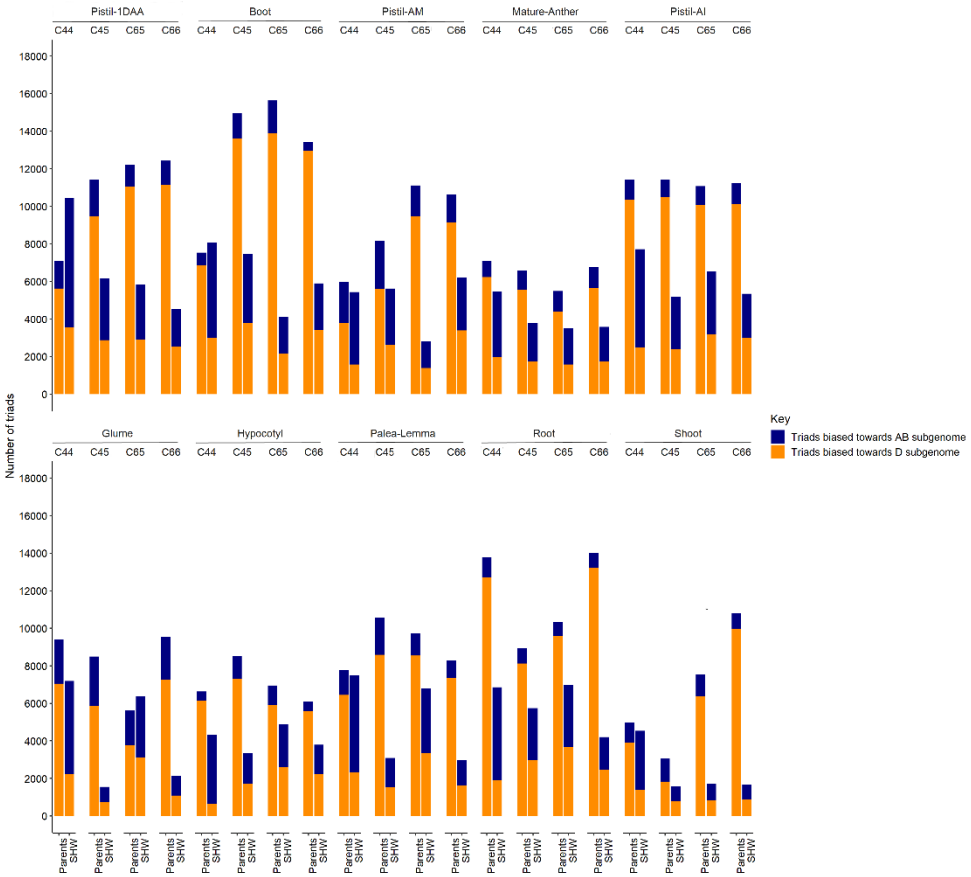


Figure 2.1 Number of triads showing significant bias towards AB and D subgenomes in all four SHW lines - C44, C45, C65 and C66 and in the natural state of expression in their parents. The navy blue and dark orange bars represent the number of triads showing significant expression bias towards AB and D subgenomes, respectively. Pistil-1DAA: pistil-one day after anthesis; Pistil-AM: pistil-when anthers are at mature stage; Pistil-AI: pistil-when anthers are at immature stage; Boot: head at boot stage.

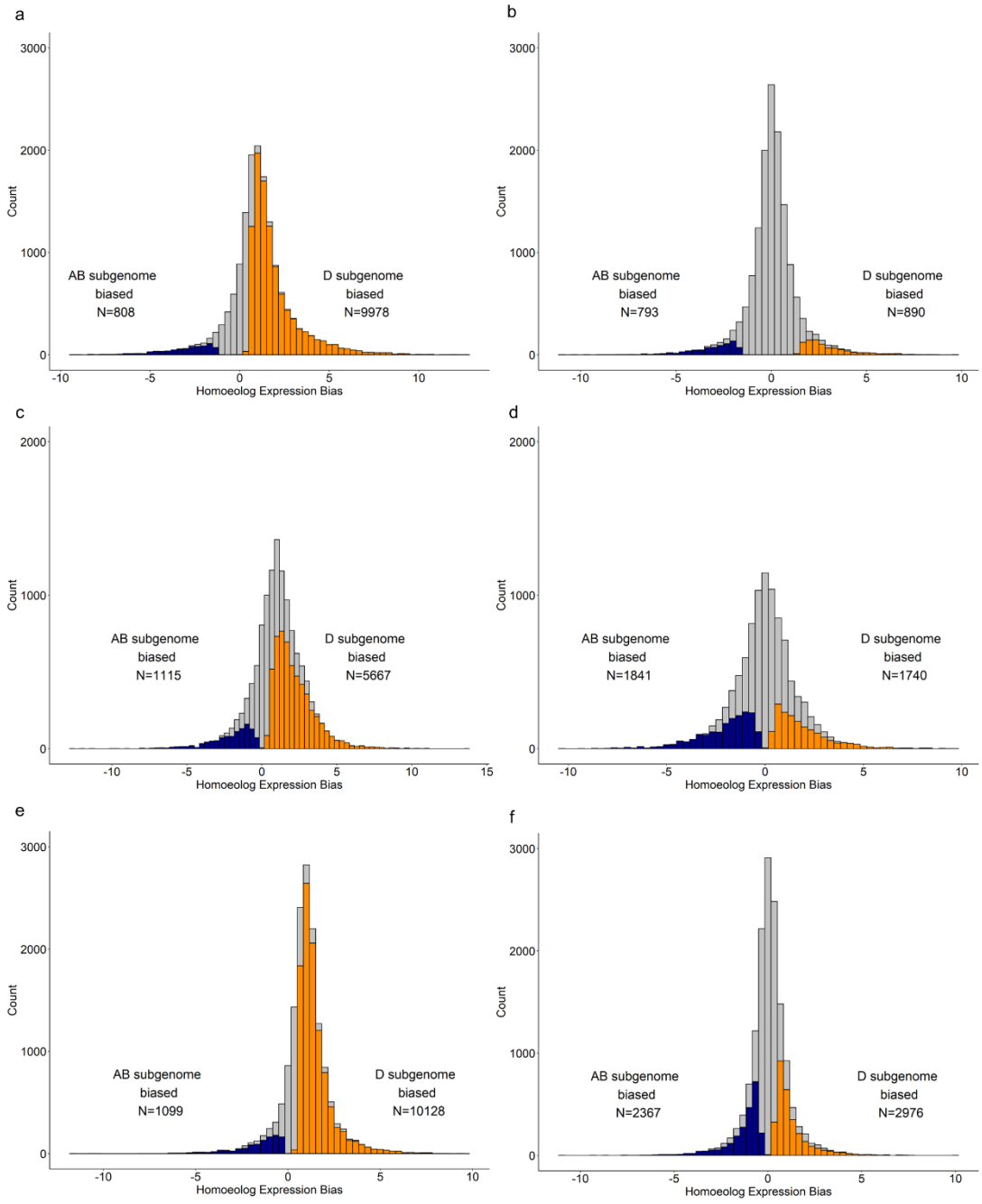


Figure 2.2 Distribution of homoeologue expression bias in SHW-C66. (a, c, e) Comparison of the expression bias of triads towards AB and D genomes of tetraploid (PI377655) and diploid (AS2386) parents; (b, d, f) Comparison of the expression bias of triads towards AB and D subgenomes of SHW-C66; (a & b) Shoot ; (c & d) Mature anther just prior to dehiscence; (e & f) Pistil when anthers are green and immature. The navy blue and dark orange bars represent the triads showing significant expression bias towards AB and D subgenomes, respectively.

In the shoot tissue of SHW-C66, 9,978 and 808 triads showed a significant bias towards the D and AB subgenomes, respectively, based on the parental level of expression. However, in the SHW background, only 890 triads were biased towards the D subgenome (Figure 2.2a and 2.2b). In anthers, 5,667 and 1,115 triads were biased towards the D and AB subgenomes, respectively, in the progenitor level of expression and more balanced in the hexaploid condition with 1,740 D-biased and 1,841 AB-biased triads (Figure 2.2c and 2.2d). Likewise, 10,128 triads biased towards the D subgenome in the parental expression levels were reduced to 2,976 in the pistils (collected when anthers were green and immature) of SHWs (Figure 2.2e and 2.2f). These results illustrate major shifts in expression bias among hundreds of gene sets, between the no-interaction scenario, i.e., the parental state of expression and the large-scale gene expression alteration caused by interactions of the subgenomes, i.e., in SHW lines. The number of triads significantly biased towards the D subgenome were higher in the parental genomic background, compared to their expression when they were introgressed into the synthetic hexaploid background. The distribution of the HEB for the other seven tissues for SHW-C66 is provided in Appendices 7-13.

In the no-interaction predicted *in-silico* scenario, the proportion of triads biased towards the D genome were at least 1.5 fold (LogFC 0.6) higher than those biased towards the AB genome, as observed in shoot tissue of SHW-C45, and up to 27 fold (LogFC 4.8) higher, as observed in the head at boot stage tissue of SHW-C66 (Figure 2.3). However, this number was drastically reduced in the actual hexaploid background and, an almost equal number of triads were biased towards the subgenomes of both parents (AB and D). On the contrary, the triads showing expression bias in the opposite direction, i.e., biased towards the AB genome increased in most tissue in the SHWs, with the most pronounced HEB being observed in SHW line C44 (Figure 2.3).

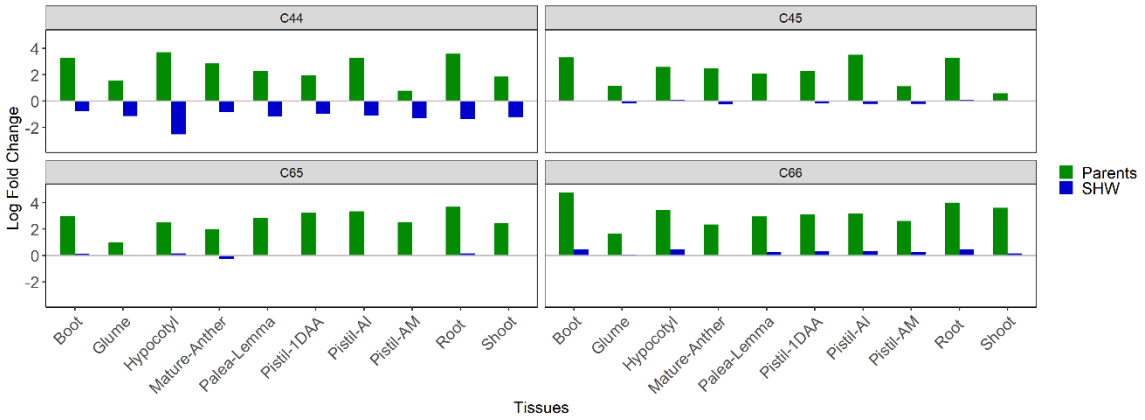


Figure 2.3 Comparison of ratios of triads significantly biased towards AB and D subgenomes within a genomic background (parents and SHW). The ratios are represented as log fold change (LogFC) values. A LogFC < 0 indicates more triads are biased towards the AB subgenome than the D subgenome and a LogFC > 0 indicates more triads are biased towards the D subgenome than the AB subgenome, within a genomic background. C44, C45, C65 and C66 are the four SHWs taken for the study. The green bars represent the ratios in the parental genomic background and the blue bars represent the SHW genomic background. Pistil-1DAA: pistil-one day after anthesis; Pistil-AM: pistil-when anthers are at mature stage; Pistil-AI: pistil-when anthers are at immature stage; Boot: head at boot stage.

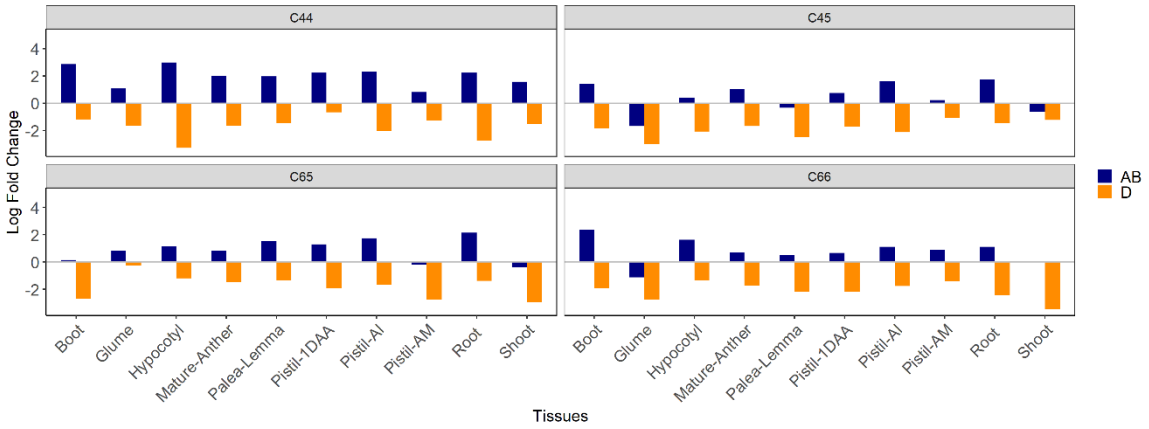


Figure 2.4 Comparison of ratios of triads significantly biased towards AB and D subgenomes between genomic backgrounds (parents vs SHWs).The ratios are presented as log fold change (LogFC) values. A LogFC > 0 indicates more triads are biased towards the AB subgenome in the SHW background than the parental background and a LogFC < 0 indicates more triads are biased towards the AB subgenome in the parental background than the SHW, similarly for D subgenome. C44, C45, C65 and C66 are the four SHWs used for the study. The navy blue and dark orange bars represent the triads showing significant expression bias towards AB and D subgenomes, respectively. Pistil-1DAA: pistil-one day after anthesis; Pistil-AM: pistil-when anthers are at mature stage; Pistil-AI: pistil-when anthers are at immature stage; Boot: head at boot stage.

The analysis of the trends within the same subgenome background revealed that the triads biased towards the AB subgenome were more numerous in the hexaploid state when compared to the parental state (Figure 2.4). All tissues of SHW-C44 had an increased proportion of triads showing expression bias towards the AB subgenome, with up to 7.9 fold (LogFC 2.98) more triads overexpressed in the hypocotyl tissue. In the other three SHW lines, the increase in triads biased towards the AB subgenome were up to 5.2 fold (head at boot stage of C66; LogFC 2.37). The proportion of triads biased towards the D subgenome were reduced by more than eight times as detected across multiple SHW-tissue contexts.

The average magnitude of HEB values for the AB and D biased triads for the SHW and parental background scenarios are summarized in Appendix 6. For the AB biased triads, the magnitude of HEB values ranged from -1.17 (C65 - pistil when anthers are green and immature) to -4.37 (C44 - hypocotyl), in the SHW background. The magnitude of HEB

of the AB biased triads varied from -1.22 (C65 - head at boot stage) to -3.56 (C66 - head at boot stage), in the parental background. Similarly, the magnitude of HEB values of the D biased triads extended from 1.16 (C44 - pistil when anthers are green and immature) to 3.12 (C45 - glume), in SHW background, and from 1.32 (C45 - head at boot stage) to 4.66 (C45 - shoot), in parental background. The average magnitude of HEB values of the AB biased triads were higher than those of the D biased triads across all SHW-tissue scenarios, except in the C65 SHW background in pistil when anthers are green and immature. However, no fixed pattern was observed in the parental background (Appendix 6).

2.3.2 Role of parental genotypes on HEB

The subset of triads showing similar expression bias patterns across different SHW lines was also analysed. The intersections among the four SHW lines and between the SHW lines sharing common tetraploid or diploid parents is represented in Figure 2.5 for the shoot tissue. The largest intersection represented the subset showing no significant bias in all four SHW lines, and this was true across all ten tissues (Figure 2.5; Appendices 14-22). Shoot tissue had the smallest subset of triads showing similar significant expression bias towards the AB (228 triads) and D (242 triads) subgenomes, across the four SHW lines (Figure 2.5). The largest intersection among the SHW lines was found in pistil when anthers are green and immature (Appendix 20) and pistil - one day after anthesis, for the AB biased (1,281) and D biased (1,066) triads, respectively (Appendix 14).

.

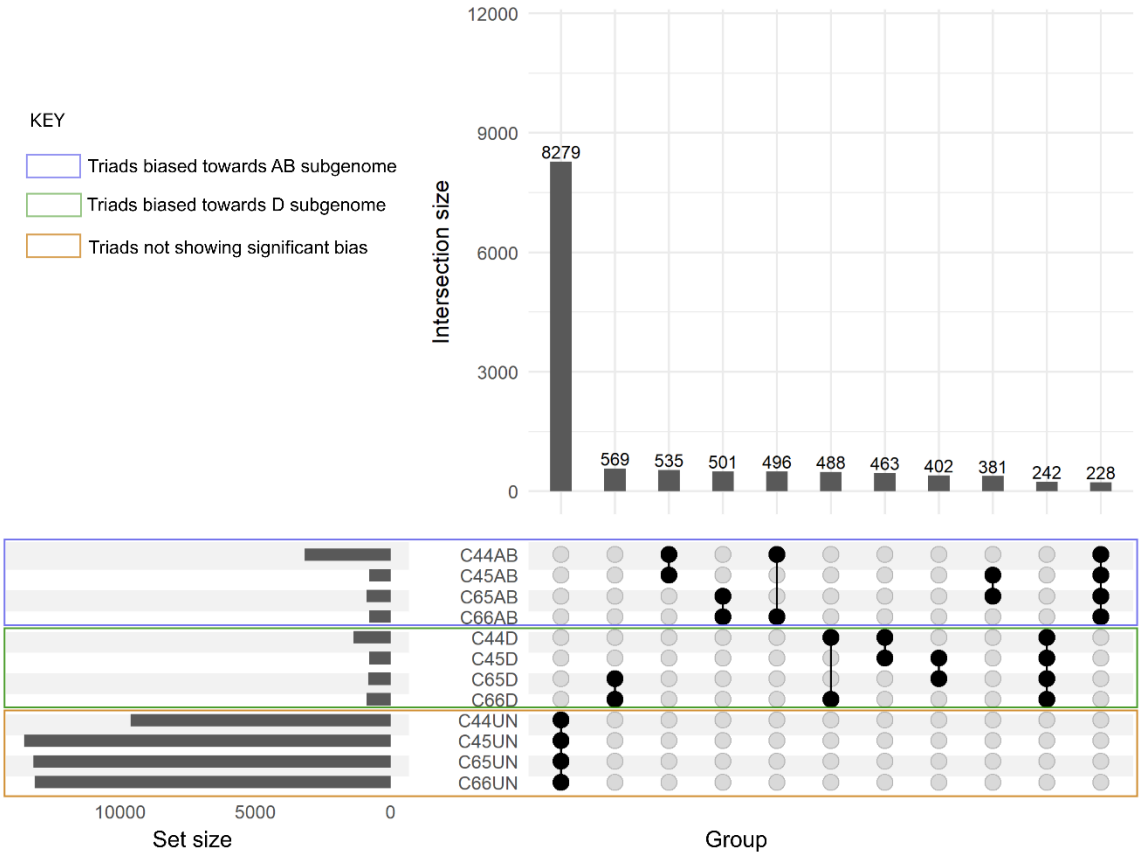


Figure 2.5 Subsets of triads showing similar expression patterns across all four SHW lines in shoot tissue. The grey horizontal bars under set size corresponding to C44AB indicates the number of triads biased towards the AB subgenome in the SHW-C44; C44D indicates the number of triads biased towards the D subgenome in the SHW-C44; C44UN indicates the number of triads showing no significant bias in the SHW-C44; Similarly, for SHWs C45, C65 and C66. The vertical bars under intersection size indicate the number of triads showing similar expression patterns between sets. The black dots highlight the two SHW sets compared for the corresponding intersection. The number over the bars represents the number of triads showing similar expression pattern. The triad sets showing significant bias towards the AB or D subgenomes and those not showing any significant bias are highlighted by the blue, green and brown rectangles, respectively.

2.3.3 Role of tissue of expression of homoeologues on HEB

Out of the 18,357 triads analysed for the expression pattern, 69% to 73% were not comprised of any tissue-specific homoeologues at all, in all four SHW lines (Appendix 23). The next largest fractions were triads with all three homoeologues showing same tissue

specific expression (8% to 11%) and triads with only one homoeologue showing tissue specificity with the other two being more broadly expressed (7% to 12%). Further, there were more triads with two homoeologues showing expression specificity towards the same tissue and one broadly expressed homoeologue, than one homoeologue alone displaying expression specificity in another tissue (Appendix 23).

Similar constitution was observed in individual SHW-tissue contexts, most commonly, with only one of the three homoeologues being tissue specific or all three homoeologues showing tissue specificity (Appendix 24). The triads with two of the three homoeologues alone showing tissue specificity were lowest in number across the ten tissues in all four SHW lines. In the root and mature anther tissues, the subset of triads comprised of all homoeologues showing tissue specificity were largest in all four SHW lines. Out of the larger subsets of triads with all three homoeologues showing same tissue specificity, 455 triads in the root were common among all four SHW lines. When the homoeologues of these triads were subjected to gene ontology analysis for understanding functional enrichment, biological process terms including regulation of root morphogenesis and regulation of root meristem growth were enriched in the root. In mature anthers, this number was 427 and the biological process terms including pollen germination, pollen sperm cell differentiation, pollen tube growth, and pollen wall assembly were enriched.

A large number of triads, comprised of either one, two or three tissue-specific homoeologues, showed no significant expression bias, similar to the observation in the entire triad dataset. When the triads showing AB- and D- bias and their tissue specificity patterns were interrogated specifically, a moderate strength of association with Cramer's-V statistic ranging between 0.20 and 0.57 was observed between bias pattern and tissue specificity of the homoeologues in most tissues in all four SHW lines. For instance, more triads showed D-subgenome bias when D-homoeologue alone was tissue-specific, greater

number of AB-biased triads were observed when A and B homoelogues showed tissue specificity, and more D-biased triads were observed when A and D or B and D homoeologues alone showed tissue specificity, in most SHW-tissue contexts. However, based on Chi-square test of independence, the association between expression bias and tissue specificity was significant only in mature anthers, root and hypocotyl tissues in all four SHW lines, and C66 showed significant association additionally in pistil-one day after anthesis, pistil-when anthers are green and immature and C44 displayed significant association in all tissues. Although the observed extent of associations were moderate, the relationship between homoeologues' tissue specificity and expression bias cannot be dismissed.

2.3.4 Impact of genome stabilization on HEB

To validate the observed patterns in more stabilized genomes, the expression bias was analysed in wheat lines that were subjected to domestication and breeding processes using the publicly available RNAseq data for the landrace Chinese Spring and the cultivar Azhurnaya from Ramírez-González et al. (2018). The number of HEB and the magnitude of the expression biases were much smaller in these genotypes compared to the SHW lines (Figure 2.6). In Azhurnaya, 2,087 to 4,244 triads were biased towards the AB subgenome and 2,032 to 3,739 triads were biased towards the D subgenome, in tissues reported in Figure 2.6. In Chinese Spring, 874 to 4,512 triads were biased towards the AB subgenome and 812 to 3,821 triads were biased towards the D subgenome, across the five tissues used in the analysis. The magnitude of biases was lower in the two domesticated wheat lines compared to the newly created SHW lines.

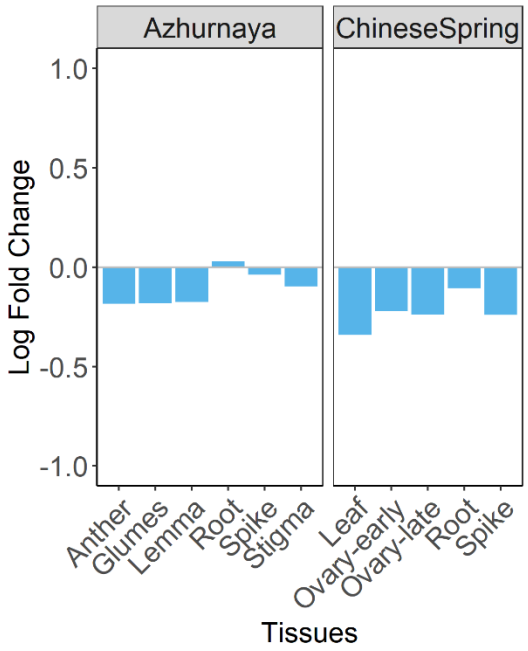


Figure 2.6 Ratio of triads significantly biased towards AB and D subgenomes in cultivar Azhurnaya and landrace Chinese Spring. The ratios are represented as log fold change (LogFC) values depicted by the blue bars. A LogFC < 0 indicates more triads are biased towards the AB subgenome than the D subgenome and a LogFC > 0 indicates more triads are biased towards the D subgenome than the AB subgenome.

The influence of domestication and breeding processes on the HEB patterns of the triads were analysed using the SHW, Chinese Spring and Azhurnaya results. Three of our SHW tissues could be matched to the developmental stages of samples in Ramírez-González et al. (2018), namely pistils when anthers are mature, root and head at boot stage, and were used in this analysis. The majority of the triads (29% - 55%) maintained similar expression bias patterns in the resynthesized wheat, the landrace (Chinese Spring) and the cultivar (Azhurnaya) in all three tissues (Appendix 25). Only 9% - 14% showed similar bias patterns in the SHW and Chinese Spring, and transitioned to a different bias state in Azhurnaya, possibly reflecting the impact of rigorous selection. Further, 9% - 30% of the triads, based on the different SHW-tissue scenarios, showed reversal to SHW bias state in Azhurnaya (Appendix 25). Taking into consideration the HEB differences for the triads

among the SHW lines, we also specifically investigated the subset of triads showing similar bias patterns across all four SHW lines. The major proportion of triads in this subset retained the same expression bias patterns in the SHWs, Chinese Spring and Azhurnaya, and the second largest proportion showed similar patterns in the SHWs and Azhurnaya and not in Chinese Spring, indicating the reversal of bias patterns during crop improvement. The triad set showing similar pattern changes including Chinese Spring and Azhurnaya was the smallest.

The HEB expression analysis across ten different tissues enabled the investigation of bias patterns of the triads across tissues, i.e., through the developmental stages. In all four SHW lines, the largest pattern observed was triads showing no significant bias across any tissues. There were anywhere from 1,449 to 3,014 triads in the SHW lines that showed this pattern (Figure 2.7, Appendix 26). In all four SHW lines, the next most commonly observed bias trend was where the triads did not show significant bias in any of the tissues but were biased towards the AB or D subgenomes only in the mature anther tissues (Figure 2.7, Appendix 26). Patterns showing only significant bias in one of the tissues such as pistil - one day after anthesis, pistil when anthers are green and immature and root, were observed in multiple triads. In addition, there were thousands of triad-specific patterns across tissues. Similar trends were also observed in Azhurnaya, and the subset of triads showing significant bias in anther tissues was larger, whereas in Chinese Spring, triads showing significant bias only in spike tissue were common.

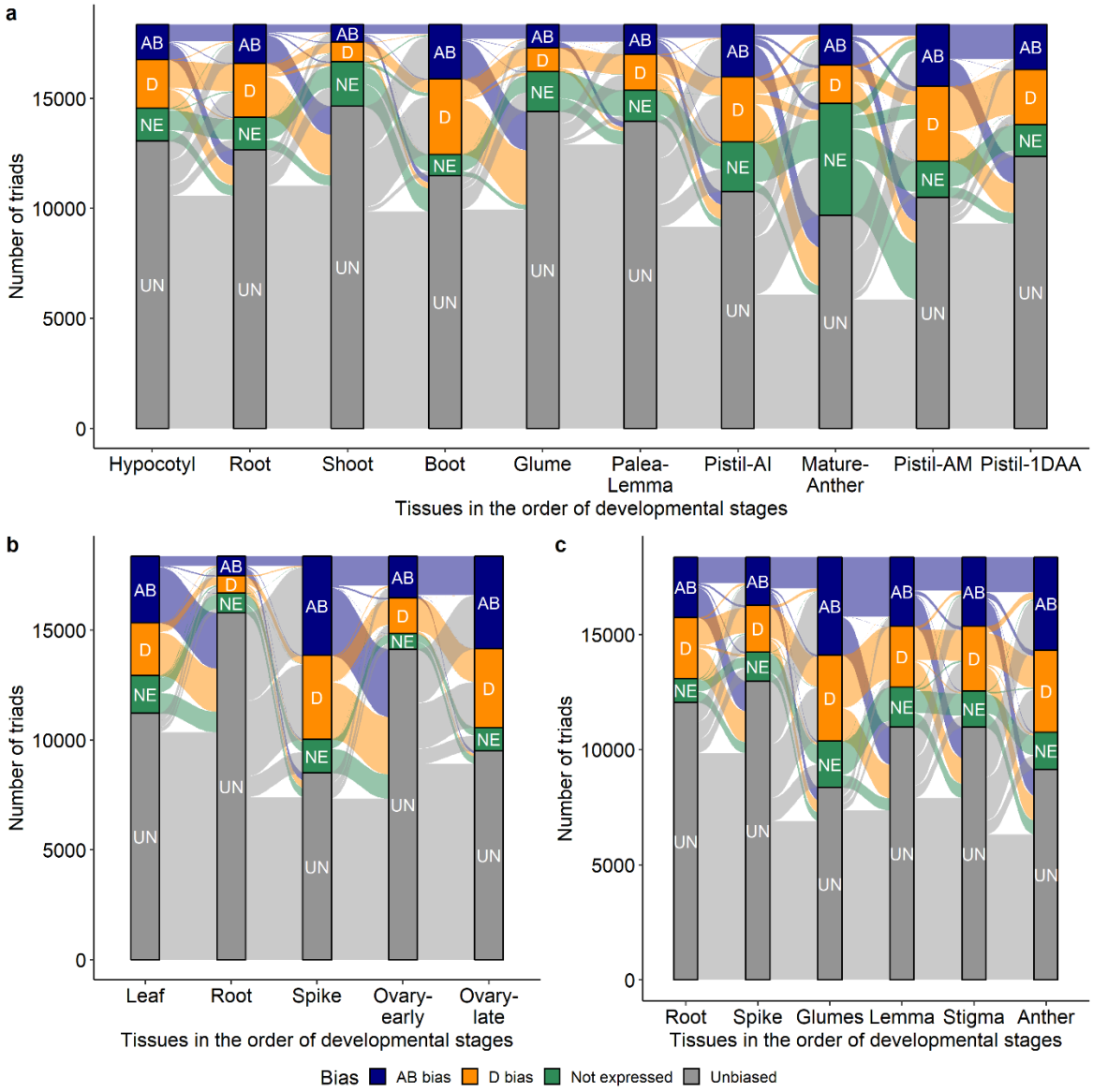


Figure 2.7 Expression bias trends of the triads across tissues. Alluvial plots representing the expression bias trends of the triads across the tissues in the (a) SHW - C66 (b) Landrace Chinese Spring and (c) Cultivar Azhurnaya; AB bias (AB), represented by navy blue bars: triads significantly biased towards the AB subgenome; D bias (D), represented by dark orange bars: triads significantly biased towards the D subgenome; Not expressed (NE), represented by green bars: the homoeologues of the triads are not expressed; Unbiased (UN), represented by grey bars: no significant expression bias observed.

2.3.5 Differential expression analysis of complete gene set

In addition, the differential expression analysis of the whole gene set between the parents and SHWs was investigated. The 106,913 high confidence gene models identified in the hexaploid wheat genome as per the IWGSC RefSeq annotation v2.1 (Zhu et al. 2021) were used in the analysis. In the genes from the A and B subgenomes, only a few hundred to a maximum of 1,015 genes were downregulated in the SHWs compared to their tetraploid parents in SHW-C44 (Figure 2.8). In contrast, more than 20,000 D subgenome genes were downregulated in all SHWs compared to their diploid parents (Figure 2.8). Most importantly, in all three subgenomes, only 357 to 755 genes and 141 to 372 genes exhibited upregulation in the hexaploid state, in the AB and D subgenomes, respectively.

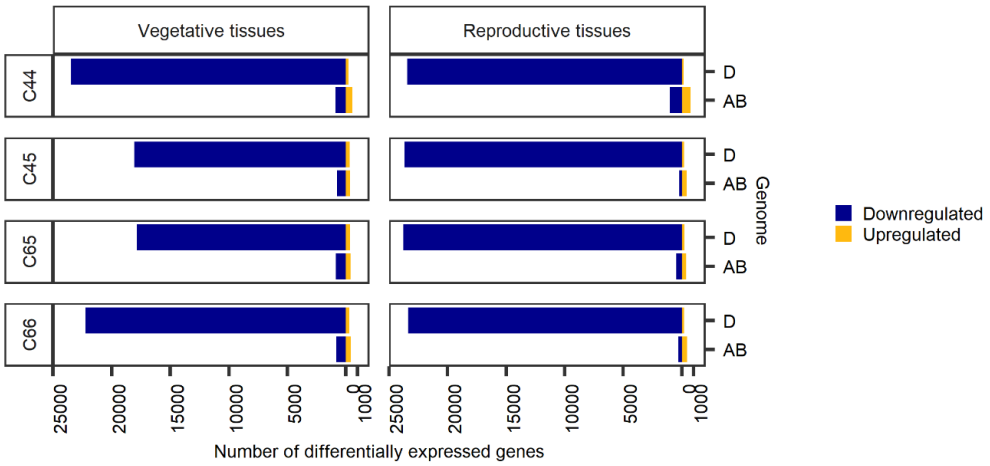


Figure 2.8 Differentially expressed genes of the AB and D subgenomes of the vegetative and reproductive tissue pools upon comparison of the SHW lines (C44, C45, C65 and C66) with their parents. The blue bars represent the number of downregulated genes and the yellow bars represent the number of upregulated genes.

2.3.6 Qualitative differences in transcripts between parents and SHWs

The RNAseq data was also mined to characterize the differences in splicing patterns at the different ploidy levels to shed light on the qualitative differences among transcripts. This analysis captured the splice variants from all genes whether differentially-expressed or not. Five major types of alternative splicing events, namely mutually exclusive exons (MXE), alternative 3' splicing site (A3SS), alternative 5' splicing site (A5SS), retained intron (RI) and skipped exon (SE) were investigated (Figure 2.9). As an example, the number of statistically significant alternatively spliced RNA isoforms detected in each genotype is illustrated for shoot tissue in the facets of the stacked bar chart (Figures 2.10a and 2.10b). *Retained intron* was the most common alternative splicing event (34.2%-79.5%) and *mutually exclusive exons* were least detected (0%-7.3%) in all scenarios. The mature anthers and pistil-when anthers are green and immature, showed the least alternative spliced events in comparison with other tissues. The number of alternatively spliced events detected in the other nine tissues are presented in Appendices 27-35. Among the multiple alternatively spliced events detected, 69%-78% corresponded to homoeologous genes in triads (Figure 2.10c and Appendices 27-35). In all comparisons, more alternative splicing events were associated with D homoeologues, except in the shoot tissue of C45 vs parents, where an equal proportion of events was associated with all three subgenomes (Figure 2.10c and Appendices 27-35). Overall, large-scale quantitative repression of the D subgenome was observed in SHW lines as measured by homoeologous expression biases and, qualitative changes were also more prominent in this subgenome as illustrated by the larger representation of alternative RNA splicing events.

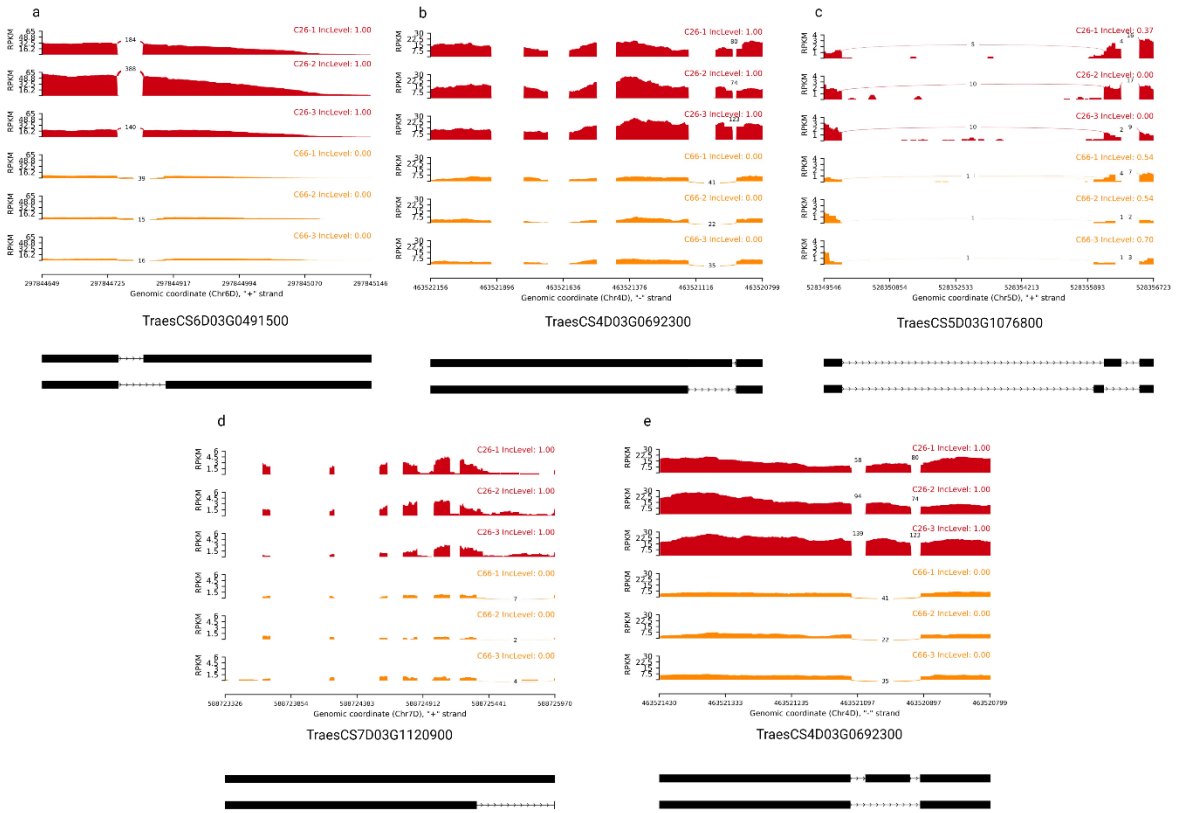


Figure 2.9 Sashimi plots representing examples of the five types of alternative splicing events in the shoot tissue of *Ae. tauschii* C26 line (diploid parent) vs SHW line C66. (a) Alternative 3' splicing site; (b) Alternative 5' splicing site; (c) Mutually exclusive exons; (d) Retained intron; (e) Skipped exon. The red plots represent the *Ae. tauschii* C26 parent transcripts and the yellow plots represent the SHW line C66 transcripts.

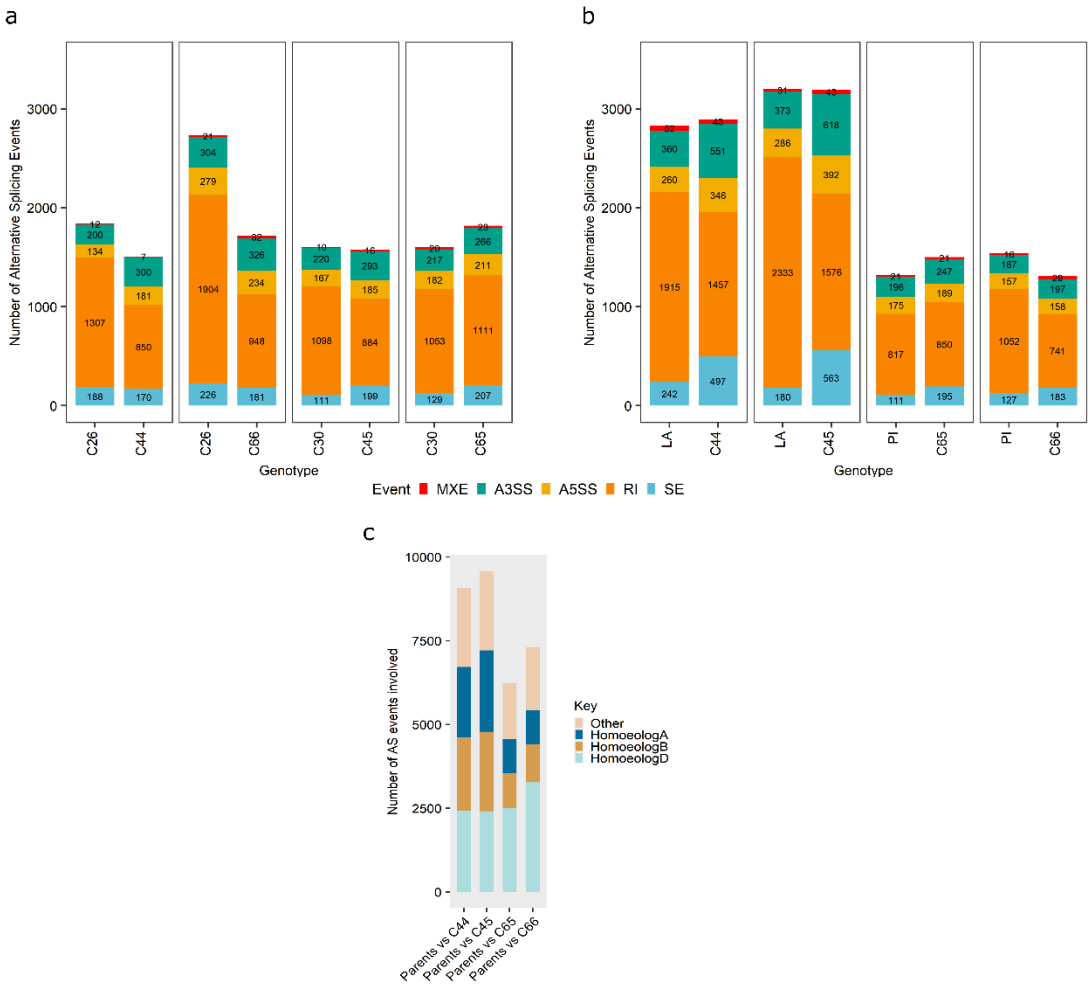


Figure 2.10 Number of alternative splicing events detected in the diploid (C26, C30)/tetraploid (LA, PI) parents compared to the SHWs (C44, C45, C65, C66) in shoot tissue. (a) Diploid parent vs SHW line; (b) Tetraploid parent vs SHW. MXE: mutually exclusive exons; A3SS: alternative 3' splicing; A5SS: alternative 5' splicing; RI: retained intron; SE: skipped exon; LA: Langdon; PI: PI377655; (c) Proportion of alternative splicing events involving the homoeologues in triads and other genes.

2.4 Discussion

The process of introgressing desirable novel alleles and haplotypes underlying phenotypes from genetically diverse parents known as pre-breeding is a viable strategy to develop cultivars with improved disease resistance, quality and agronomic traits, including climate resilience traits. However, hybridizations among diverse parents often result in poor penetrance and expressivity of traits sourced from donor parents. SHWs are being generated to access the genetic diversity present in the *Ae. tauschii* genepool (D genome donor) to overcome this allelic diversity bottleneck of bread wheat, wherein poor recovery of parental traits were reported when their genomic background changes from diploid (DD) to hexaploid (AABBDD) (Dreisigacker and Kishii 2019). Hence, a critical understanding of the genome-wide impact of polyploidization in crop species is needed to design strategies to enhance the potential of pre-breeding. In this study, we investigated and unravelled the global expression level alterations in SHW lines and their corresponding tetraploid and diploid progenitors to portray the transcriptome dynamics taking place during initial polyploidization events. The four parents used in the study were diverse in geographical origin. Specifically, *T. turgidum* cv. Langdon is from North America (Langdon) and PI377655 is from former Yugoslavia, while *Ae. tauschii* AS2386 and AS2399 are from Iran (Liu et al. 2013; Zhang et al. 2010). The primary SHW plants used in the experiment had been selfed for at least four generations; these selfings potentially facilitated the stabilization of the genome compared to nascent polyploids (Shaked et al. 2001), allowing the capture of heritable transcriptomic variations.

The triad gene sets (genes present in 1:1:1 condition in ABD homeologue chromosomes) form the vital portion of the wheat genome playing key functional roles, while the genes present as other homoeologue copy number variations perform other functions (Juery et al. 2021). In all ten developmentally distinct tissues in the four SHW lines, the

majority of the genes from the D subgenome were suppressed in SHWs, when compared with their parental levels of expression. Li et al. (2014b), in their transcriptome analysis of resynthesized wheat, have also shown the shift in expression dominance while comparing parental and hexaploid states. Changes in expression bias among homoeologues upon whole genome duplication have been observed in other allopolyploid crops upon resynthesis, including in rapeseed (Wu et al. 2018) and cotton (Yoo et al. 2013), reflecting the subgenome dominance in these species as well. However, in bread wheat, it was proposed that a balance is prevalent among the subgenomes, without any subgenome dominance (IWGSC 2018; Juery et al. 2021). Nevertheless, tissue-, growth stage-, and environmental cue- dependent dominance have been reported as the cause of developmental process and environmental adaptations (Pfeifer et al. 2014; Powell et al. 2017). In this study, based on the observations in the SHW lines, the expression level dominance in the parental states is reduced by polyploidization, navigating the evolutionary process towards achieving an overall balanced allohexaploid, while the homoeologue-specific variability is capitalized by the developmental or environmental requirements. In domesticated wheat genotypes, the expression of D homoeologues were found to be less repressed than the A and B homoeologues (Juery et al. 2021; Ramírez-González et al. 2018). Hence, it can be considered that the initial broad-scale repressive effect on D homoeologues is to establish a balance in expression between genomes derived from different parents (AB and D), and not for the subjugation of the D subgenome. Further, the expression bias patterns established in newly synthesized polyploids are maintained over multiple generations through the processes of domestication and breeding, as observed in the RNAseq data from more stabilized genomes such as Chinese Spring (a landrace) and Azhurnaya (a cultivar), sourced from the open-source databases corresponding to Ramírez-González et al. (2018). Hence, the initial novel expression patterns (repression of D

subgenome) observed in newly synthesized hexaploid wheat, are stabilized and inherited over multiple generations.

Polyploidization events involve genome reshuffling and can alter the gene dosage either immediately after duplication, or over a long period of time (million year scale) during evolution (Veitia and Birchler 2022). Here we showed the instantaneous effects (within five generations of selfing since initial crossing) of polyploidization in hexaploid wheat. The shift in expression balance can be attributed to gene dosage (allele copies) because biologically-significant stoichiometry of macromolecule complexes and regulatory genes impacting downstream loci are critical for the fitness of an organism (Birchler 2012). The difference among expression bias of the same gene sets between tissues may be attributed to the expression contribution of the duplicates that might differ in a tissue-specific pattern, especially in the reproductive cells (Birchler and Veitia 2021; Conant et al. 2014). Notably, five of the ten tissues used in the study possibly had a higher fraction of haploid gametophytic cells, referred to as the reproductive pool. In yet another dimension, polyploidization events increase the complexity of the gene regulatory network, through the introduction of inter-genomic trans-regulatory mechanisms (Hu and Wendel 2019). This could result in homoeoallele-specific expression as observed in previous studies in wheat (Powell et al. 2017) (2n=6x=42), high-ploidy sugarcane (Margarido et al. 2021) (2n=8x-12x=80-120) and peanut (Peng et al. 2020) (2n=4x=40). Further, the individual developmental state of the plant and the genotype may also have an influence over allele specificity (Correr et al. 2022; Margarido et al. 2021), further explaining the homoeoallele expression variation across the ten developmentally-distinct tissues and four genetically different SHW lines used in our investigation.

Gene expression dynamics is the translational component linking observed phenotypic variation and its underlying genotype. Several triads showing similar expression

bias patterns in the SHWs, Chinese Spring and Azhurnaya, indicate this subset remains unaffected by artificial selection. The other triads having altering HEB patterns in the resynthesized, domesticated and elite wheat backgrounds, imply the selection for specific expression bias patterns during domestication and crop improvement, which are potentially coupled to other genes regulating the characteristics of modern-day bread wheat. The limitation of this comparison is that the SHW lines used in the study are not the direct progenitors of neither Chinese Spring or Azhurnaya and the expression data are from different studies. Hence, further interrogation with multiple wheat landraces and elite wheat cultivars will shed more light on the selection for HEB.

While investigating the bias trends of triads across the tissues, the number of triads showing unbiased expression pattern across all tissues were the largest, aligning with the observation that major subsets of triads did not show significant bias in all four SHW lines in all ten tissues. The HEB specific to biotic (Powell et al. 2017) and abiotic (Dong et al. 2022) stress response have been reported in wheat and allopolyploid cotton. The results from this study, showing preponderance of triad-specific bias trends across developmental stages, suggest the spatial and temporal variability of HEB, leading to different bias trends through developmental stages, in addition to stress responsive patterns. In allopolyploid cotton, a study analysing the fiber development, revealed the changes in chromatin architecture in the different developmental stages and their influence on gene regulatory networks (GRNs) leading to HEB (Pei et al. 2022). This highlights the potential role of dynamic variation in 3D chromatin architecture in bringing about spatiotemporal differences in HEB. In addition, this also conveys how the expression plasticity in allopolyploids are leveraged throughout development.

Beyond the effect of the environmental stressors on expression bias, variability in expression bias in the different tissues has also been reported in several alloplyploid

species, such as rapeseed (Khan et al. 2022), cotton (Fang et al. 2017), coffee (Combes et al. 2011) and white clover (Griffiths et al. 2019). Previous work in hexaploid wheat has also shown that triads with varied expression bias patterns across tissues are more tissue specific (Ramírez-González et al. 2018). In this study using SHW transcriptomes, the moderate association between homoeologue expression bias and tissue-specificity of homoeologues in the triads, elucidate the potential role of expression specificity of homoeologues in driving the direction of the expression biases. Hence, the tissue, developmental stage, environmental cues and parental genotype, all determine the bias patterns of the triads through exploiting the additional plasticity in the hexaploid wheat genomes when compared to their tetraploid and diploid progenitors.

In addition to comparison of expression patterns among homoeoalleles, the differential expression analysis at the whole genome level between the parents and SHW lines performed to characterize the expression changes of the genes between different genomic backgrounds, revealed the large-scale repression of the D subgenome genes. The introgression of chromosomal segments from species in the secondary and tertiary gene pools has resulted in transcriptomic changes in both wheat (Coombes et al. 2021; Dong et al. 2020; Li et al. 2020; Rey et al. 2018) and other polyploid species (Bin et al. 2019; Shao et al. 2021). Specifically, transcriptomic analysis of wheat × *Ambylopyrum muticum* introgression lines indicated the suppressed expression of genes in the introgressed segments (Coombes et al. 2021). Similarly, a higher proportion of genes in the introgressed region were down regulated in the expression studies of wheat-barley addition lines (Rey et al. 2018). Introgression of alien chromatin segments into a species leads to the suppression of alien transcripts by the host genome; though how the host genome distinguishes the alien genome remains unclear. Combining the D genome from *Ae. tauschii* with the AB

genome in *T. turgidum* results in analogous responses, hinting that the AB genome acts as the native genome and the D genome as the introduced foreign chromatin in SHWs.

The outcome of a polyploidization event at the genomic, and consequently at the phenotypic level is dependent on the genetic background of the lines involved in the hybridization event (Gallois et al. 2018; Sharma et al. 2021). The assessment of multiple SHW lines derived from diverse tetraploid and diploid accessions showed variability in response to Fusarium head blight (FHB) infection, which could not be envisaged based on the FHB responses of the parent (Szabo-Hever et al. 2018). Similar variability based on genotypic background effects has been observed for yield-related traits of wheat lines with rye introgressions (Moskal et al. 2021).

The expression level differences from polyploidization to achieve dosage-balance, to establish homoeoallele-specific expression or in exhibiting genomic background effects, are all potentially the manifestation of the genetic and epigenetic changes that occur upon polyploidization. The chromosome-level analysis of synthetic lines and their tetraploid and diploid parents using molecular cytogenetic techniques have depicted structural differences in chromatin regions (Zhao et al. 2021a), and genome sequence level analysis have also revealed similar outcomes (Ozkan et al. 2001; Yu et al. 2017). In an epigenetic perspective, microRNAs (miRNAs), the key players in post transcriptional gene regulation, were expressed in a non-additive manner in resynthesized wheat (Li et al. 2014b). The modified expression of miRNAs possibly underlies the downstream shift in expression bias and repression in the mRNA data. Small interfering RNAs (siRNAs) derived from double stranded RNA molecules were involved in the methylation of DNA sequences, specifically transposable elements, and also in establishing repressive heterochromatin marks (Gutbrod and Martienssen 2020). Comparison of a newly synthesized hexaploid wheat and its *Ae. tauschii* parent revealed increased accumulation of siRNAs on the D homoeologues in the

hexaploid, in contrast to the observations in *T. turgidum* vs hexaploid A subgenomes (Li et al. 2014b). The increased accumulation of siRNAs could be the underlying cause for the repression of D homoeologues in hexaploids (Li et al. 2014b). DNA methylation is an epigenetic feature for gene expression regulation and genome stability through silencing of transposable elements (TEs). The ploidy-level changes drive modification of methylation patterns in CG, CHG and CHH sequence contexts in wheat, involving the TEs as well. As TEs occupy the major proportion of the wheat genome, such epigenetic reprogramming of the TE fraction modulates expression of genes in its vicinity because TEs harbour promoter and enhancer motifs (Yaakov and Kashkush 2011; Yuan et al. 2020). Among the multiple histone marks in the genome, an increase of the repressive histone-3 lysine-27 dimethylations (H3K27me2) was found to be positively correlated with the increase in ploidy in wheat (Liu et al. 2021). Furthermore, the H3K27me3 marks underlie the subgenome-specific activity of regulatory elements (Wang et al. 2021b), and the introduction of the D subgenome triggers distinct histone modifications and consequent inter-subgenome regulatory interactions (Lv et al. 2021). The chromatin accessibility, which is controlled by association between transcription factors and nucleosomes, is lower in the D subgenome of hexaploids compared to the *Ae. tauschii* background, also leading to extensive reduction in gene expression (Lu et al. 2020). In addition to these genetic and epigenetic changes, alteration of chromatin architecture upon polyploidization also warrants investigation in hexaploid wheat. For instance, changes in topologically-associated domains (TAD) and A/B compartmentalization have been reported in other polyploid crop species like watermelon (Garcia-Lozano et al. 2021) and cotton (Wang et al. 2018b).

Following gene transcription, precursor mRNAs are modified using the splicing machinery to produce mature RNAs. During this process, the splicing machinery can retain different combinations of introns and exons by alternative splicing (AS). This AS mechanism

can diversify the transcriptional and translational products of a gene, thereby altering its function in the developmental frameworks of space (different tissues) and time (vegetative vs reproductive stages) (Kazan 2003). There are tissue-specific isoforms, and abiotic (Laloum et al. 2018) and biotic (Huang et al. 2020) stress-induced AS events. However, the impact of polyploidization and the resulting genomic shock on AS have only recently been investigated. In a study of hexaploid wheat and its tetraploid and diploid parents and relatives, Yu et al. (2020) also found that retained intron was the most common AS mechanism observed. The dominance of intron retention has been observed across plant species, including *Arabidopsis* (Martín et al. 2021), rice, sorghum and maize (Min et al. 2015) and cotton (Wang et al. 2018a). The influence of polyploidization on differential splicing events has been observed in allopolyploid rapeseed, along with homoeologue-specific differences in AS (Zhou et al. 2011). The characterization of AS in the transcriptome derived from embryogenesis tissues in wheat and its progenitors, had a preponderance of alternative 3' splicing events, however, the difference has been attributed to AS-event specificity of the analysis tool and percent spliced in thresholds utilised (Gao et al. 2021).

In conclusion, the global transcriptome dynamics analysis in SHWs compared to its tetraploid and diploid parents revealed expression bias among homoeologues resulting in large-scale suppression of the D subgenome in the SHWs as inferred from more than 18,000 triads and expression from ten tissues. Qualitative changes observed among transcripts in the form of novel splice variants upon polyploidization also majorly impacted the D genome homoeologues. These quantitative (transcript abundance) and qualitative (splice variants) expression changes seem to be dependent on the genomic composition of the parents and the developmental stage (tissues) of the wheat lines. To harness the genetic diversity available in the D genome progenitors (*Ae. tauschii*) and other wild

relatives for wheat improvement through interspecific hybridization, the potentiality of modified expression in the progeny must be taken into consideration to recover the desired phenotype in the polyploid background.

2.5 Methods

2.5.1 Plant materials

Four SHW lines (C44, C45, C65 and C66) and their two tetraploid (PI377655 - C16, Langdon - C19) and two diploid (AS2386 - C26, AS2399 - C30) parents, for a total of eight genotypes, were used in the study (Appendix 1). The following ten tissue samples (Appendix 2) were collected from the above mentioned eight lines: (1) head at boot stage, (2) shoot collected at heading, (3) lemma and palea at heading, (4) glume from first and second floret of a spike, (5) pistil - when anthers are green and immature, (6) pistil - when anthers are yellow and just prior to dehiscence, (7) pistil - one day after anthesis, (8) yellow anther just prior to dehiscence, (9) hypocotyl and (10) root, where the latter two tissues were collected from young seedlings grown on Whatman filter papers. Three biological replicates per tissue from all the eight genotypes were used in the study (ten tissues × eight genotypes × three replicates = 240 samples).

2.5.2 RNA extraction and library preparation

The RNeasy Plant Mini Kit (Qiagen Inc., Germantown, MD, USA) was used for all tissues except for the mature anther and pistil when anthers are green and immature, which were isolated using miRNeasy mini kit (Qiagen Inc). The extracted RNAs were quantified initially with the Implen Nanophotometer (Implen Inc, Westlake Village, CA, USA) and the quality and concentration were assessed with the Agilent RNA 6000 Nano assay (Agilent Technologies, Santa Clara, CA, USA). The construction of cDNA libraries and sequencing

were carried out at the Centre d'expertise et de services Génome Québec (Montreal, QC, Canada). Briefly, mRNA enrichment was performed using NEBNext Poly(A) Magnetic Isolation Module (New England BioLabs, Ipswich, MA, USA), followed by cDNA synthesis and library preparation with the NEBNext RNA First Strand Synthesis, NEBNext Ultra Directional RNA Second Strand Synthesis modules and the NEBNext Ultra II DNA Library Prep Kit (New England BioLabs). Libraries were quantified using PicoGreen dsDNA assay kit, and quality was analysed using LabChip GX (PerkinElmer, MA, USA). The normalized libraries were clustered on an Illumina cBot, and 100-bp paired-end reads were generated on a HiSeq 4000 platform (Illumina Inc., CA, USA) for the mature anthers and the pistils collected when the anthers were still immature, i.e., prior to pollination. The other eight tissues were sequenced as 150-bp paired-end reads on the NovaSeq 6000 platform (Illumina Inc.).

2.5.3 Processing and alignment of sequencing reads

The sequencing reads were pre-processed using Trimmomatic (Bolger et al. 2014) and aligned to the recently updated version of the bread wheat genome pseudomolecules (RefSeq v2.1; Annotation v2.1) (Zhu et al. 2021) using Spliced Transcripts Alignment to a Reference (STAR) (Dobin et al. 2012), after genome indexing, with the maximum intron length set to 10,000 bp, the maximum number of allowed mismatches as six, and the other parameters left at their default settings.

2.5.4 Estimation of homoeologue expression bias (HEB) and likelihood ratio test (LRT)

In the bread wheat genome, nearly 35% of the genes are present as triads, with one homoeologous copy in each of the subgenomes (1:1:1 in A:B:D) (IWGSC 2018). A total of 18,474 triads were identified using the Refseq v1.0 annotation. However, after taking into account the revisions made in Refseq v2.1 annotation, a set of 18,357 triads, comprised of gene models annotated with high confidence (IWGSC 2018; Zhu et al. 2021), were used for this study of comparative transcriptomes. An expression matrix was generated from the gene count data, and the reads per kilobase per million mapped reads (RPKM) values were estimated from the raw count data using the *rpkm* function in edgeR (Robinson et al. 2009). Homoeologue expression bias (HEB), representing the number of fold bias and its direction, was estimated using the following formula (Smith et al. 2019).

$$HEB=\log_2\left(\frac{RPKM_D}{RPKM_{AB}}\right)$$

Where, $RPKM_{AB}$ and $RPKM_D$ represent the expression values from the AB and D subgenomes, respectively. For the RPKM expression values from the AB subgenome, the average between the expression levels of the A and B homoeologues of a triad gene set was utilised, considering the balanced expression levels of the A, B and D homoeologues, and the difference in gene length is accounted for by the normalization.

Further, the likelihood ratio test developed on MATLAB by Smith et al. (2019) and first used in Edger et al. (2017), was applied to identify the triads showing significant bias towards the tetraploid (AABB) or diploid (DD) genomes. Likelihood ratio tests were performed using the expression data from the SHW lines as well as the data from their corresponding tetraploid and diploid parents, creating an *in-silico* SHW without inter-subgenome interactions, in order to determine the shift in expression bias upon

polyploidization. The upstream input preparation and downstream summarization steps were performed with custom scripts in bash and R. In order to investigate the bias patterns of tissue-specific and broadly expressed genes, the Tau method (Kryuchkova-Mostacci and Robinson-Rechavi 2017) was used for all the genotypes with a cut-off of 0.8. The strength of association between the tissue specificity of the homoeologues and the expression bias of the triads was estimated using Cramer's-V statistic and the significance of association was determined using the Chi-square test of independence. The gene ontology analysis of the homoeologues in triads of root and mature anther tissues with A, B and D homoeologues showing same tissue specificity across all four SHW lines, was performed using the Triticeae-Gene Tribe tool kit (Chen et al. 2020). The R scripts used for input file preparation for LRT and tissue-specificity analyses are available at https://github.com/akshaya-v/SHW-Expression-bias.

To unravel the HEB in domesticated wheat lines, the RNAseq data from Ramírez-González et al. (2018) for the landrace Chinese Spring and the cultivar Azhurnaya were downloaded for the following tissues: Chines Spring: leaf (at 14 days), root (at 14 days), ovary (early anthesis), ovary (late anthesis) and spike (at booting); Azhurnaya: stigma and ovary (anthesis), glumes (milk grain stage), anther (anthesis), root (seedling), spike (30% spike out). The downstream read processing, alignment and HEB analysis was carried out as described above for the SHWs.

2.5.5 Differential expression analysis

The read count data from the ten tissues were grouped into two pools: vegetative (shoot, root, hypocotyl, glume and palea+lemma) and reproductive (head, pistil - when anthers are green, pistil - when anthers are yellow, pistil - one day after anthesis and yellow anther just prior to dehiscence). The logic was to group the tissues comprised of only vegetative cells

into one pool and the tissues comprised of both vegetative and reproductive cells into another pool. The differential expression analysis in the two pools was performed using the edgeR statistical package (Robinson et al. 2009), and a false discovery rate (FDR) of <0.05 and log fold change (LogFC) cut-off of ±2.00 were applied as thresholds for inferring differentially expressed genes.

2.5.6 Alternative splicing analysis

The alternative splicing analysis was carried out using rMATS (Shen et al. 2014). The BAM files generated using STAR were utilised. The allow-clipping option was enabled to accept alignments with soft clipping. The detection of novel splice sites not present in the annotation files were allowed using the novelSS option. The alternative splicing analysis was performed for all eight possible SHW vs diploid/tetraploid parent comparisons in the ten tissues. The sashimi plots were made using rmats2sashimiplot (rmats2sashimiplot 2015). The number and types of differentially spliced isoforms detected in parents or SHW lines were summarized for comparison. The alternative splice types were alternative 3' splice site, alternative 5' splice site, mutually exclusive exons, retained intron and skipped exon.

2.5.7 Statistics and reproducibility

Three biological replicates per tissue per genotype were utilised. There were ten tissues and eight genotypes, and hence, there were 240 tissue samples in total. The likelihood ratio test to detect significant HEB was performed with the null hypothesis that the homoeologues show equal levels of expression in AB and D subgenomes, and the alternate hypothesis that the homoeologues show unequal expression levels between subgenomes, as described in Smith et al. (2019). The differential expression analysis was performed using edgeR(Robinson et al. 2009) v3.38.4 with FDR <0.05 and logFC cut-off ±2.00. The

alternative splicing analysis was carried out with rMATS (Shen et al. 2014) using the default splicing difference cut-off of 0.01%.

2.5.8 Data availability

The raw RNAseq data has been deposited in the Short Read Archive (SRA) at NCBI under Bioproject PRJNA905376. The base data for the HEB figures, alternative splicing figures and tissue-specificity analyses results are available in Figshare (https://doi.org/10.6084/m9.figshare.c.6443621) (Vasudevan et al. 2023).

2.5.9 Code availability

The source code of the R scripts used for preparing input files and all other analyses are provided within the methods section.

CHAPTER 3

Lifting of the 1,000 wheat exome project SNPs from *Triticum aestivum* cv. Chinese Spring assembly RefSeq v1.0 to RefSeq v2.1

Akshaya Vasudevan[1,2] and Sylvie Cloutier[1,2,*]

[1]Agriculture and Agri-Food Canada,
Ottawa Research and Development Centre,
Ottawa, Ontario, Canada K1A 0C6.

[2]Department of Biology,
University of Ottawa,
Ottawa, Ontario, Canada K1N 6N5.

[*]E-mail address of the corresponding author: sylviej.cloutier@agr.gc.ca

Note: This chapter has been published as the following article:

Vasudevan A, Cloutier S (2023) Lifting of the 1,000 wheat exome project SNPs from *Triticum aestivum* cv. Chinese Spring assembly RefSeq v1. 0 to RefSeq v2. 1. BMC Research Notes 16(1):220.

https://bmcresnotes.biomedcentral.com/articles/10.1186/s13104-023-06496-8#Sec2

S.C. and A.V. conceived the idea and finalized the methods. A.V. performed the analysis. A.V. and S.C. wrote the paper.

3.1 Abstract

3.1.1 Objective

The 1,000 wheat exome project captured the single nucleotide variants in the coding regions of a diverse set of 890 wheat accessions to analyse the contribution of introgression to adaptation of wheat. However, this highly useful single nucleotide polymorphism (SNP) dataset is based on RefSeq v1.0 of the International Wheat Genome Sequencing Consortium (IWGSC) assembly of the bread wheat genome of Chinese Spring. This reference sequence has recently been updated using optical maps and long-read sequencing to produce the improved RefSeq v2.1. Our objective was to develop a reliable high-density SNP dataset positioned onto RefSeq v2.1 because it is the current standard reference sequence used by wheat researchers.

3.1.2 Results

The 3,039,822 SNPs originally positioned on RefSeq v1.0 were projected to v2.1 using Liftoff with four different flanking regions, and 2,946,536 SNPs were consistently lifted to the same location irrespective of the flanking region lengths. Of these, 2,799,166 were located on the ‘+’ ve strand. The distribution of the SNPs across the 21 chromosomes on RefSeq v2.1 was similar to that of RefSeq v1.0. Among the SNPs that were based on unanchored scaffolds in RefSeq v1.0, 11,938 were projected to one of the 21 pseudomolecules in the upgraded assembly. This SNP dataset constitutes a much-needed standardized resource for the wheat research community.

3.2 Introduction

As of January 2021, reference sequences had been generated for 798 land plants (Marks et al. 2021). With the advances in sequencing technologies, high-quality genome assemblies

of polyploid species are now more readily produced (Marks et al. 2021). Availability of gold-standard reference genomes is important for application in crop improvement through characterization of genomic regions underlying traits of interest and for the utilisation of genetic diversity from wild relatives (Bevan et al. 2017). Wheat (*Triticum aestivum* L.) is an allohexaploid species with a genome estimated at ~16 Gb that contains the A, B and D subgenomes (Arumuganathan and Earle 1991). The first assembly of bread wheat cultivar Chinese Spring, based on chromosome survey sequences and released in 2014 (IWGSC 2014), provided the baseline for the subsequent iterations. The requirement for a highly contiguous and well annotated genome eventually led to the development of the first pseudomolecule-level assembly of Chinese Spring (RefSeq v1.0) totalling 14.5 Gb, a feat that was accomplished by the International Wheat Genome Sequencing Consortium (IWGSC) using Illumina sequencing and DeNovo MAGIC assembly (IWGSC 2018). Using optical map and PacBio long-read sequencing data, the assembly quality was further improved to produce IWGSC RefSeq v2.1, in which multiple scaffolds were anchored, reoriented and relocated (Zhu et al. 2021). Consequently, the quality of the annotation was improved, and the number of high confidence genes in the annotation also increased from 105,534 genes in v1.1 to 106,913 in v2.1 (Zhu et al. 2021).

The current day hexaploid wheat evolved ~8000 years ago through chance hybridizations between domesticated emmer wheat (*Triticum turgidum* L. ssp. *dicoccum*) and goatgrass (*Aegilops tauschii* Coss.) (Feuillet et al. 2008; Huang et al. 2002). The limited polyploidization events, the subsequent domestication and breeding have restricted the genetic diversity in hexaploid wheat, especially in the D subgenome. Considering the need to understand gene flow from progenitors into hexaploid wheat and their role(s) in adaptation and evolution, an extensive 1,000 wheat exome project was designed and executed through an international collaboration (He et al. 2019). Here a diverse panel of

890 accessions comprised of hexaploid wheat cultivars, landraces and wild and domesticated tetraploid progenitors were taken up and subjected to exome capture and sequencing. Using the IWGSC RefSeq v1.0 as reference, single nucleotide polymorphisms (SNPs) were called, a haplotype map was generated, and the preponderance of introgressions in wheat adaptation was investigated by He et al. (2019). This publicly available SNP dataset is an important genomic resource to the international wheat research community. However, it is based on v1.0 which is no longer the standard reference genome for wheat research. To ensure its continued utility, projection of this extensive SNP dataset unto RefSeq v2.1 is urgently needed. Here we outline a strategy based on the repurposing of Liftoff (Shumate and Salzberg 2021) and present the process and outcome of the lifting-over of the 1,000 wheat exome SNP dataset to the IWGSC RefSeq v2.1 assembly.

3.3 Methods

Liftoff is a tool that was developed for projecting annotations from one reference genome assembly to another without the use of chain files (Shumate and Salzberg 2021). For this purpose, features are assigned in a hierarchy, for instance, gene as parent feature and exon as child feature. The parent feature (gene sequence) is aligned to the target, following which the location of the child features is deduced. Here we repurposed Liftoff for lifting-over SNPs, by assigning the SNP and the corresponding flanking regions of a chosen length as the child and parent features, respectively. In this process, the flanking regions of each SNP are aligned to the target genome and the "region" between them delineates the single nucleotide position of the SNP (Figure 3.1). To project the SNPs from the 1,000 exome project, the variant call format (VCF) file was downloaded from the Wheat URGI repository (https://wheat-urgi.versailles.inra.fr/Seq-Repository/Variations). The 3,039,822 SNPs in the VCF file were split into 16 parts for ease of computational analyses, and the result files were merged after processing. The VCF files were converted to GFF files using a

custom R script. For each SNP, the GFF file contained the parent feature, i.e., the flanking region start and end coordinates, and the child feature, i.e., the SNP coordinates *per se*. Each SNP and the flanking regions were assigned an ID, which associated the SNPs with their corresponding flanking regions. Strand orientation was not included in the 1,000 wheat exome VCF file. Since providing the plus strand alleles is the standard convention for publicly available SNP datasets (Nelson et al. 2012), the strand orientation for all SNPs in the input was labelled as '+'.

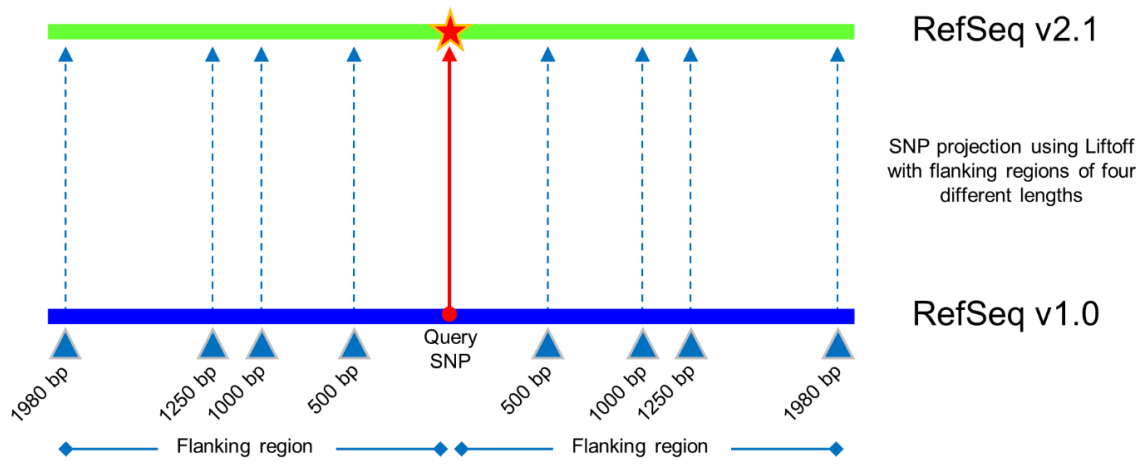


Figure 3.1 Strategy implemented for projecting SNPs from IWGSC RefSeq v1.0 to RefSeq v2.1 using Liftoff.

A flanking region of 1,980 bp on either side of the SNP was initially used because the average transcript length in the annotation RefSeq v1.1 was 3,961 bp; hence, this length can be argued to represent the average length of conserved regions in the genome. To test the ability of Liftoff to project the SNPs from v1.0 to v2.1, the following flanking region lengths were also assessed: 500 bp, 1,000 bp and 1,250 bp, and only the consensus set of SNPs lifted to the same location with all four flanking region lengths was retained.

Liftoff uses Minimap2 (Li 2018) for sequence alignment which requires sequences ≥ 100 bp; hence, all flanking region lengths tested met the criteria. Further, the chromosome-to-chromosome lift-over option was not used in Liftoff to allow SNPs to be projected onto different chromosomes or onto newly anchored scaffolds of RefSeq v2.1.

The new v2.1 SNP coordinates corresponding to the v1.0 SNPs of the 1,000 exome VCF file were assigned using their corresponding SNP IDs. The preparation of input files, merging and formatting of files and output summarization steps were performed using custom R scripts and bash. The functional annotation of the SNPs was performed using IWGSC RefSeq v2.1 on SnpEff (Cingolani et al. 2012). The database construction was carried out using both low- and high-confidence gene models from annotation v2.1 using the build option in SnpEff. With the VCF file as input, functional annotation was performed using the eff option with default parameters.

3.4 Results

The number of SNPs lifted-over using the four flanking region lengths is summarized in Appendices 36-39. The number of SNPs lifted-over ranged from 3,017,212 while using the longest flanking regions of 1,980 bp to 3,020,764 while using the shortest flanking regions of 500 bp. Comparisons across all GFF3 output files revealed that 2,946,536 SNPs were consistently projected to the same location regardless of the length of the flanking regions. Of these, 2,799,166 (95%) were located on the '+' ve strand (reference genome orientation), while 147,370 (5%) were in the '-' ve strand. In summary, 2,799,166 of the 3,039,822 SNPs (92%) of the 1,000 wheat exome project were lifted-over from RefSeq v1.0 to RefSeq v2.1.

A total of 35,284 out of the 39,822 SNPs originally positioned on the unanchored scaffolds (chrUn) in RefSeq v1.0 were lifted to the same loci regardless of the flanking

region lengths. Of those, 11,938 were assigned to a position and projected onto one of the 21 assembled chromosomes. The vast majority of the SNPs were lifted to the same location in both versions, but 5,949 SNPs ended up projected to a different chromosome from v1.0 to v2.1. This did not however affect the overall distribution pattern of SNPs across the 21 chromosomes which remained similar between RefSeq v1.0 and v2.1, with higher density of SNPs in the distal and interstitial regions of the chromosomes (Figure 3.2).

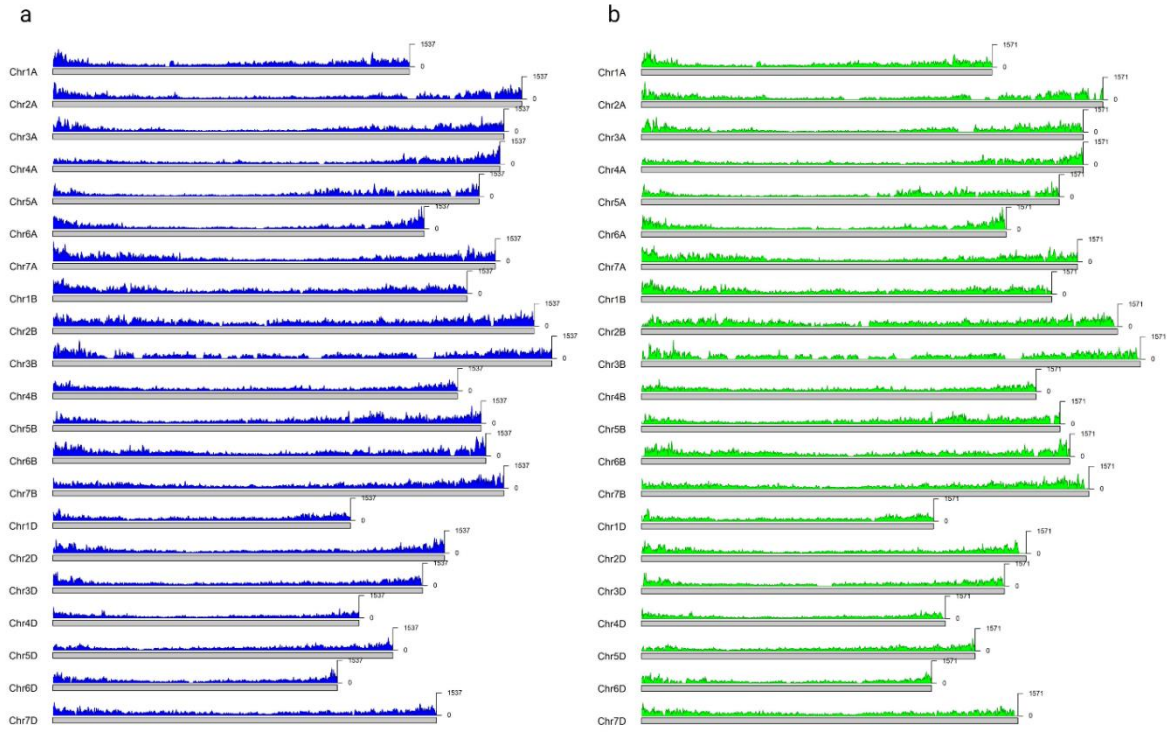


Figure 3.2 SNP density plots. (a) Distribution of 3,039,822 SNPs on IWGSC RefSeq v1.0; (b) Distribution of 2,799,166 SNPs on IWGSC RefSeq v2.1.

When the functional annotation of SNPs based on RefSeq v2.1 was carried out, ~5.4 million effects were predicted for the ~2.8 million SNPs. Effects are defined as the potential consequences of the variants on transcription and translation. For instance, a SNP

in a genic region might be present in either exons or introns in differentially spliced isoforms, which in turn will influence the outcome of the variant. Hence, the number of effects determined by the snpEff tool is higher than the number of SNPs. In the annotation by He et al. (2019) of the variants based on RefSeq v1.0 annotation, 1.4% of the SNPs were possibly deleterious, including splice site changes and start/stop gained or lost variants, and 10.5% of them were non-synonymous. Based on the RefSeq v2.1-dependent functional annotation, 1.5% and 8.7% of the SNPs were potentially deleterious and non-synonymous, respectively. With regards to the multiple effects of the SNPs lifted to RefSeq v2.1, the majority of the effects (89%) were predicted to have no impact on the proteome because they were either non-coding variants or associated with non-coding genes, as per the SnpEff categorization. A total of 4.7% of the effects were low impact such as synonymous variants, 5.8% were moderate impact with a possibility of affecting the encoded proteins, and 0.4% were high impact with significant effect on the coded proteins. Based on the functional classes, 58.6% of the effects were categorized as missense and 40.5% as silent. The number of deleterious ‘start lost’, ‘stop gained’ and ‘stop lost’ effects were 499, 15,176 and 1,583, respectively. In reference to genic regions, most SNPs were located in intergenic regions, followed by those located in the downstream and upstream regions. However, ~20% of the SNP effects were in the intronic and exonic regions of genes and 0.5% were localized at the splice site junctions (Table 3.1).

Table 3.1 Percent SNP effects based on their location respective to the nearest transcript.

Location	Count	Percentage
Downstream	1,240,165	22.77
Exon	557,842	10.24
Intergenic	1,887,968	34.66
Intron	594,000	10.90
Splice site acceptor	2,012	00.04
Splice site donor	1,901	00.04
Splice site region	28,984	00.53
Upstream	1,035,877	19.02
3'-UTR	70,492	01.29
5'-UTR	28,030	00.52

3.5 Discussion

Chain files are required for the widely-used VCF and annotation lift-over tools such as UCSC LiftOver (Hinrichs et al. 2006), CrossMap (Zhao et al. 2013) and Picard's LiftOverVcf (Broad-Institute 2022). Chain files encompass the pairwise alignment information of two sequences capturing the homologous blocks between genomes under comparison. Generating a chain file for large and complex genomes such as wheat is computationally intensive. Here we explored the repurposing of the Liftoff tool, originally designed for projecting coding sequences onto genome, to project millions of SNPs from the Chinese Spring RefSeq v1.0 onto the latest v2.1. Liftoff does not use the computationally intensive chain files. In this repurposing exercise, SNPs were defined as child feature, and SNP flanking regions of different lengths were concurrently used as parent features. Extracting and utilising only the SNPs lifted to the same loci consistently, regardless of the flanking region lengths, ensured that only the most reliable SNP coordinates were retained in the final dataset projected onto RefSeq v2.1.

Optical map and PacBio long-read sequencing were used to refine the Chinese Spring reference assembly and produce RefSeq v2.1 (Zhu et al. 2021). The utilisation of an

optical map led to the anchoring of 279 previously unanchored scaffolds, the orientation correction of 354 scaffolds and the relocation of 233 scaffolds (Zhu et al. 2021). This improved RefSeq v2.1 assembly validated 90% of the IWGSC RefSeq v1.0. As such, it was not unexpected to find that more than 99% of the SNPs were projected to the homologous position across the two assembly versions. The lifting of 11,938 SNPs from ChrUn in RefSeq v1.0 to one of the 21 assembled chromosomes is likely the outcome of the anchoring of previously unanchored scaffolds. A proportion of this lifting, such as lifting from Chr1A (v1.0) to Chr5A (v2.1), Chr2D (v1.0) to Chr4A (v2.1) etc., can be attributed to the relocation of scaffolds to other chromosomes across assemblies. In addition, during the assembly improvement process, small scaffolds were discarded (Zhu et al. 2021), which could have led to the lift-over of SNPs from these scaffolds to a highly homologous region in other chromosomes or onto ChrUn, or these could have remained unprojected. The lifting of 98% of genes from RefSeq v1.0 to RefSeq v2.1 with high accuracy was achieved only because Insertion Site-Based Polymorphisms (ISBP)-dependent complexity reduction strategy was utilized. The fact that the whole RefSeq v2.1 was used as the target might also be the reason for incorrect anchoring of a small set of SNPs (Zhu et al. 2021). Because of its superior quality, RefSeq v2.1 is now the preferred reference genome assembly version of the wheat community. The use of the SNP variants from v1.0 and their projection onto v2.1 on an *ad hoc* basis is impractical, time-consuming and could easily lead to errors. It was therefore urgent and crucial to project all SNPs onto RefSeq v2.1 to provide this valuable resource for the whole wheat research community. In conclusion, we successfully lifted the 1,000 wheat exome SNP data unto the IWGSC RefSeq v2.1. The public availability of this standardized genome-wide dataset is expected to spur its continued utilisation by the wheat research and breeding communities by ensuring consistency and reliability and, in doing so, to contribute to accelerating discoveries.

3.6 Data availability

The VCF file with the SNPs projected onto IWGSC RefeSeq v2.1 has been deposited in Zenodo (DOI: 10.5281/zenodo.7852690).

CHAPTER 4

Recombination landscape in an elite × synthetic hexaploid wheat (SHW) backcross population

Akshaya Vasudevan[1,2], Chunfang Zheng[1], Jennifer Ens[3], Frank You[1], Firdissa Bokore[4], Richard Cuthbert[4], Curtis J. Pozniak[3] and Sylvie Cloutier[1,2,*]

[1]Agriculture and Agri-Food Canada,
Ottawa Research and Development Centre,
Ottawa, Ontario, Canada K1A 0C6.

[2]Department of Biology,
University of Ottawa,
Ottawa, Ontario, Canada K1N 6N5.

[3]Crop Development Centre and Department of Plant Sciences,
University of Saskatchewan,
Saskatoon, Saskatchewan, Canada S7N 5A8.

[4]Agriculture and Agri-Food Canada,
Swift Current Research and Development Centre,
Swift Current, Saskatchewan, Canada S9H 3X2.

*E-mail address of the corresponding author: sylviej.cloutier@agr.gc.ca

Note: This chapter is yet to be published.

S.C. conceived the idea. S.C. and C.J.P. secured the funding. S.C. generated the genotypic data for the backcross population. F.B. and R.C. provided the genotypic data for the doubled haploid population. C.Z. and F.Y. processed the sequencing data initially. A.V. and S.C. strategized the variant calling with RefSeq v2.1 assembly, introgression, recombination and donor allele frequency analyses. A.V. performed the analyses.

4.1 Abstract

Triticum turgidum L. (AABB) and *Aegilops tauschii* Coss. (DD), the immediate progenitors of hexaploid wheat (*Triticum aestivum* L.), are the preliminary sources of genetic diversity for bread wheat improvement. SHW lines are generated to artificially introgress the progenitor diversity into hexaploid wheat. SHW lines are majorly harnessed through the introgression of their genetic fragments into elite bread wheat backgrounds. The crossover (CO) landscape across the chromosomes affect the introgression patterns. Considering the presence of the undomesticated progenitor version of the subgenome in SHW lines, their introgression patterns into elite wheat was investigated using the SHW-C66 × elite wheat Carberry BC_1F_5 population. Large introgressions were observed around the centromere and the introgression size became progressively smaller towards the telomeric regions. There appeared to be a preponderance for introgression in the A and B subgenomes compared to the D subgenome of Carberry, and distinct introgression patterns were observed across certain chromosomes. The recombination rate distribution was correlated with the observed introgression size distribution. The donor allele frequency was higher than the expected 25% for a BC_1F_5 population in many segments in the A and B subgenomes but not in the D subgenome. An exploratory comparison of the elite wheat Carberry × elite wheat AC Cadillac F_1-derived doubled haploid population did not indicate such subgenome level differences for either recombination or donor allele frequencies. The introduction of genetic diversity and recombination landscape variability with the use of SHW pre-breeding material was evident from this study. However, further investigations with other elite wheat × SHW derived populations and similarly generated elite wheat × elite wheat populations are needed to fully validate these results.

4.2 Introduction

Wild relatives of wheat are proven as a repertoire of novel alleles and hence, their introgression is used for the improvement of agronomic traits, disease resistance, abiotic stress tolerance and grain quality to meet the requirements of the growing population and to address the changing climate (Pour-Aboughadareh et al. 2021; Sansaloni et al. 2020). *Triticum turgidum* L. (AABB) and *Aegilops tauschii* Coss. (DD) are the immediate progenitors of current day bread wheat (*Triticum aestivum* L., AABBDD). As such, they belong to its primary gene pool (Gill 2022) and are a major source of genetic diversity to wheat improvement programs (Dreisigacker and Kishii 2019; Dreisigacker et al. 2008; Gaurav et al. 2022). The genetic diversity of the wheat A and B subgenomes is higher than that of the D subgenome due to gene flow from wild tetraploid progenitors (He et al. 2019). A single population of *Ae*. *tauschii* contributed to the evolutionary allopolyploidization events, thereby incorporating only a fraction of the diversity available within the species (Wang et al. 2013). To recover some of this unused genetic diversity, synthetic hexaploid wheat (SHW) lines were produced to serve as wheat pre-breeding materials (Ogbonnaya et al. 2013) i.e. *Ae*. *tauschii* accessions were crossed to *T*. *turgidum* accessions, followed by chromosome doubling. The similar genomic composition of SHW and elite wheat cultivars allows hybridization and subsequent utilisation of novel alleles, especially from *Ae*. *tauschii*.

Wild relatives offer a broad genetic diversity that can potentially benefit wheat but harnessing it remains challenging. The novel genetic variation in wild relatives can be brought into bread wheat genomic background either through chemical- or radiation-mediated translocation of chromosome fragments, or through introgression by the process of recombination during meiosis (King et al. 2022). The translocation strategy can lead to undesirable deletions and duplications resulting in the reduced fitness of offspring, and hence, it is less preferred (King et al. 2022). However, introgression, the transfer of genetic

material from the same or different species, is more dependent on the biological tool of ‘recombination’.

In meiosis, the homologous chromosomes are assembled across a linear axis and are held together by the synaptonemal complex (Mercier et al. 2015). This is followed by the formation of multiple double strand breaks (DSBs) across the genome by the heteromeric SPO11 catalytic complex (Hyde et al. 2023; Stacey et al. 2006). In the wheat genome, ~2,100 DSBs are formed per meiosis (Gardiner et al. 2019). Following DSB formation, DMC1 and RAD51 recombinases enable strand invasion between homologues leading to the formation of a displacement loop (D-loop) (Kurzbauer et al. 2012). These critical intermediaries are later resolved into crossover (CO) and non-crossover (NCO) events. In plants, less than one tenth of all DSBs results into COs (Mercier et al. 2015). In wheat, 41-52 COs were observed per meiosis (Gardiner et al. 2019), which is only ~2% of the total DSBs formed. The occurrence of a CO at one point on the chromosome interferes with nearby COs and this crossover interference leads to COs being widely-spaced along the chromosomes (Hillers 2004). Based on their sensitivity to interference, COs are classified into classes I and II. Class I COs, induced by proteins such as MER3, MSH4, HEI10, SHOC1/ZIP2, ZIP4, etc., constitute 85% of all COs (reviewed in Rafiei and Ronceret 2023). Class II COs are insensitive to interference and require MUS81 protein (Berchowitz et al. 2007; Mu et al. 2023). PPX1 encoded by *HCR1* (Nageswaran et al. 2021), repression of HEI10 (Kim et al. 2022) and ZYP1 - the transverse filaments of the synaptonemal complex (Capilla-Pérez et al. 2021; France et al. 2021) are anti-class I CO factors while FANCM (Crismani et al. 2012), FIGL1 (Girard et al. 2015) and RECQ4 (Séguéla-Arnaud et al. 2015) proteins act as anti-class II CO factors . In allohexaploid wheat, the *Ph1* locus on chromosome 5B, corresponding to a homolog of ZIP4 (Rey et al. 2017), and *Ph2* locus on chromosome 3D, encoding MSH7-3D (Serra et al. 2021), suppress recombination of

homoeologous chromosomes, a mechanism for maintaining diploid-like meiosis behaviour in an allohexaploid crop. In addition to all the underlying genetic pathways, chromatin signatures such as DNA methylation and accessibility also play crucial roles in determining the location of COs (Tock et al. 2021).

COs and recombination are tightly regulated biological processes that are key to creating novel alleles and breaking linkage drag with undesirable alleles. Similarly to other plant species, recombination in hexaploid wheat is largely localized to the distal regions of the chromosomes, with fewer recombinations in genes located in the interstitial and pericentromeric regions (Akhunov et al. 2003; Choulet et al. 2014). A previous study of introgression patterns of *Ae*. *tauschii* genome in hexaploid wheat showed increased introgression in the distal ends of the chromosome and a positive correlation with the recombination rate (Nyine et al. 2020). However, the backcross population in the study was grown in the field and subjected to selection based on yield, threshability and other traits that could have influenced the introgression patterns (Martin and Jiggins 2017; Nyine et al. 2020).

In this work we utilised a backcross population derived from a cross between the elite wheat cultivar Carberry and the SHW line C66 that was developed and advanced through single seed descent under controlled growth conditions, without selection for phenotypic traits. The absence of selection allows the characterisation of recombination and introgression patterns *per se* of *Ae*. *tauschii* genomes in hexaploid wheat. In addition, a preliminary comparison between recombination in this elite × SHW backcross population and that of an elite × elite doubled haploid population, sharing the same Carberry elite parent, was performed to gain additional perspectives on the influence of the progenitor subgenome (D) on recombination and introgression patterns.

4.3 Methods

4.3.1 Plant material

The SHW-C66 line is a derivative of tetraploid *T. turgidum* ssp. *dicoccum* PI37765 and *Ae*. *tauschii* AS2386. The SHW-C66 donor parent was crossed and backcrossed to the recurrent elite wheat cultivar Carberry to create a BC_1F_1. These initial crosses were followed by four generations of selfing and advancement by single seed descent to produce a population of 173 BC_1F_5 backcross introgression lines (BILs). The comparative Carberry × AC Cadillac doubled haploid population was generated using the maize pollen method and is described in Bokore et al. (2020).

4.3.2 Genotyping of the BILs and parents

The DNA of the BILs and parents was extracted from young leaves using the Qiagen DNeasy 96 Plant kit (Qiagen Inc., Germantown, MD, USA) and quantified using Quant-iT PicoGreen dsDNA assay kit (ThermoFisher Scientific Inc., Waltham, MA, USA) as per manufacturer's instructions. Genotyping-by-sequencing (GBS) libraries of the BILs were constructed after restriction digestion with *Pst*I/*Msp*I enzyme combinations at the Institut de biologie intégrative et des systèmes (IBIS) of Université Laval (Québec City, QC, Canada) as described in Colston-Nepali et al. (2019). Shotgun PCR-free libraries of the parents were prepared at the Centre d'expertise et de services Génome Québec (Montréal, QC, Canada). The sequencing of the parents and pooled population libraries was performed by the same service provider on a NovaSeq 6000 platform (Illumina Inc., CA, USA) using the 150-bp paired-end mode.

The genotyping method for the doubled haploid population was previously described (Bokore et al. 2020). Briefly, the DNA was extracted from leaves using the BioSprint 96 DNA

Plant Kit (Qiagen Inc.) and genotyped with the Illumina 90K Infinium iSelect single nucleotide polymorphism (SNP) wheat assay (Wang et al. 2014).

4.3.3 Bioinformatic analysis

The sequencing data from the parents and the individuals of the backcross population were processed similarly. The adapter sequences from the raw reads were removed using Trim Galore (Krueger et al. 2023). The processed reads were aligned to the International Wheat Genome Sequencing Consortium (IWGSC) RefSeq v2.1 (Zhu et al. 2021) with BWA-MEM (Li 2013). The variants were called using the Genome Analysis Toolkit (GATK) and the recommended best practices' workflow (McKenna et al. 2010) was followed. For the base recalibration step in the variant calling workflow, the variant dataset from the 1000 wheat exome project (He et al. 2019) that was lifted to IWGSC RefSeq v2.1 (Chapter 3) was utilised. The variants were called from individual samples using GATK's HaplotypeCaller. Finally, combined genotyping across samples was carried out using genotypeGVCFs, after which the variants were filtered using the following quality parameters: quality by depth (QD < 2), Phred-scaled quality score (QUAL < 30), Fisher strand value (FS > 60), root mean square mapping quality (MQ < 40), MQ rank sum test (MQRankSum < -12.5) and read position rank sum test (ReadPosRankSum < -8.0). The SNPs in the backcross population were filtered with a minor allele frequency (MAF) > 0.10 and a call rate > 0.8 (up to 20% missing alleles) using the VCFtools (Danecek et al. 2011).

Of the parental SNPs detected using the whole genome resequencing data, only the homozygous, biallelic and polymorphic ones were retained while only the polymorphic ones were retained from the GBS data detected SNPs in the BIL population. Finally, the subset of SNPs that were detected in both the parents and the BIL population were retained. Since the SNP density was very high, the entire SNP data set was partitioned into linkage

disequilibrium (LD) blocks using gpart (Kim et al. 2019), and a representative SNP for each block was kept. Initially, the LD blocks were delineated with the BigLD function using r^2 as the LD measure and an LD cut-off of 0.5. Then, the coordinates of the high confidence genes of the IWGSC RefSeq v2.1 were used to merge or split the LD blocks previously generated by BigLD, using the gpart function. The resulting SNPs were pruned by retaining one SNP per LD block, and SNPs with missing data were filtered out. The distribution of this curated SNP dataset across the RefSeq v2.1 genome was plotted with a bin size of 1 Mb using CMplot (Yin et al. 2021). The VCF file from this subset of SNPs was converted to a genotype matrix using the bcftools (Danecek et al. 2021).

The C66, Carberry, and heterozygous alleles in the population were coded as A, B and H, respectively. Considering that short structural variations, errors in read alignment and the high gene conversion rate in wheat can lead to false positive COs, the common allele in sliding windows of 155 SNPs with a step size of one were aggregated by taking the most frequent allele in the window as representative (Dreissig et al. 2020; Gardiner et al. 2019). This was accomplished using the SlidingWindow function of the evobiR package (Blackmon and Adams 2015), generating an aggregated SNP matrix that was utilised for downstream analyses.

The number of COs across the chromosomes were tallied as number of switches between the two parental or heterozygous alleles along the length of the chromosomes. To determine the introgression patterns in the BILs, a stretch of contiguous SHW-C66 alleles (haplotype block) along the chromosomes, flanked by two loci of Carberry alleles as boundaries marking the host genome, was considered as an introgression. The mid-point between the first SNP in the introgression region and the previous SNP in the chromosome and last SNP and its next SNP were considered as the introgression start and end junctions, respectively. For the first introgression fragment of a chromosome the midpoint

between the start of the chromosome and the first marker was considered as the start junction. The distance in bp between the start and end junctions was used as the introgression size. Further, the number of recombination between every pair of SNPs in each one Mb genome windows across the chromosomes was counted and the recombination rates in cM/Mb (recombination frequency × 100) were estimated and plotted. The analyses for the estimates of COs, introgression sizes, recombination rates and donor allele frequencies and the plotting of the results were performed on the open-access R platform (R Core Team 2022) using custom scripts.

For the doubled haploid population, the genotype matrix from Bokore et al. (2020) was taken, and the marker coordinates with reference to IWGSC RefSeq v2.1 (Zhu et al. 2021) were used. The coordinates of the markers on RefSeq v2.1 was determined by nucleotide BLAST (Zhang et al. 2000) alignment of RefSeq v1.0 marker regions to v2.1, using default parameters (match/mismatch scores = 1,-2; gap penalties: existence: 0, extension: 2.5 and evalue = 1e-6). Only the individuals with less than 1% missing data were retained and the markers with missing information in any of the individuals were removed.

4.4 Results

In this study, the SHW donor line C66 was crossed and backcrossed to the recurrent elite bread wheat cultivar Carberry, the BC_1F_1 plant was selfed and BC_1F_2 lines were advanced by single seed descent to the BC_1F_5 generation. This 173 BIL population was used to investigate the introgression patterns of the pre-breeding material into elite cultivar background. The parents were sequenced using whole genome shotgun short reads with a coverage of ~20x while the BILs were sequenced using GBS to 0.09-0.86x coverage. The initial filtering of the polymorphic, biallelic and homozygous SNPs yielded 28,148,727 SNPs between the parents, while the filtering for quality parameters, minor allele frequency and

call rate produced 730,508 SNPs in the backcross population. The intersect between the filtered parental SNP set and the BIL population SNP set contained 197,618 common SNPs. The LD analysis, using BigLD function of gpart, resulted in 41,124 LD blocks, excluding those in unanchored chromosome scaffolds (ChrUn) of the IWGSC RefSeq v2.1. Fine-tuning of the LD blocks based on annotated gene coordinates using the gpart function produced 37,888 LD blocks, from which the first SNP of each block was used as a tag for each haplotype-based block. These SNP markers representing haplotype blocks were used to generate the genotype matrix. Conversely, SNP markers with missing data in even one of the individuals were dropped from the matrix, resulting in a total of 37,393 SNPs, distributed on all 21 chromosomes of wheat (Figure 4.1), and that were used for downstream analyses.

The total number of estimated COs per individual in the backcross population ranged from 22 to 382, and the median was 151 (Appendix 40). The highest average number of COs (31) per chromosome was observed for chromosome 3B. The size distribution of the homozygous and heterozygous introgressions, deciphered from contiguous homozygous and heterozygous alleles, are represented in Figures 4.2 and 4.3, respectively. The homozygous introgressions ranged in size from 225 bp to ~790 Mb with an average size of ~58 Mb. As expected, the heterozygous introgressions were smaller and ranged in size between 248 bp to ~23 Mb with an average size of 0.6 Mb.

Figure 4.1 SNP density plot across the 21 wheat chromosomes. The distribution of the 37,393 SNPs from the BC_1F_5 population and parents, used for CO, introgression and recombination analyses.

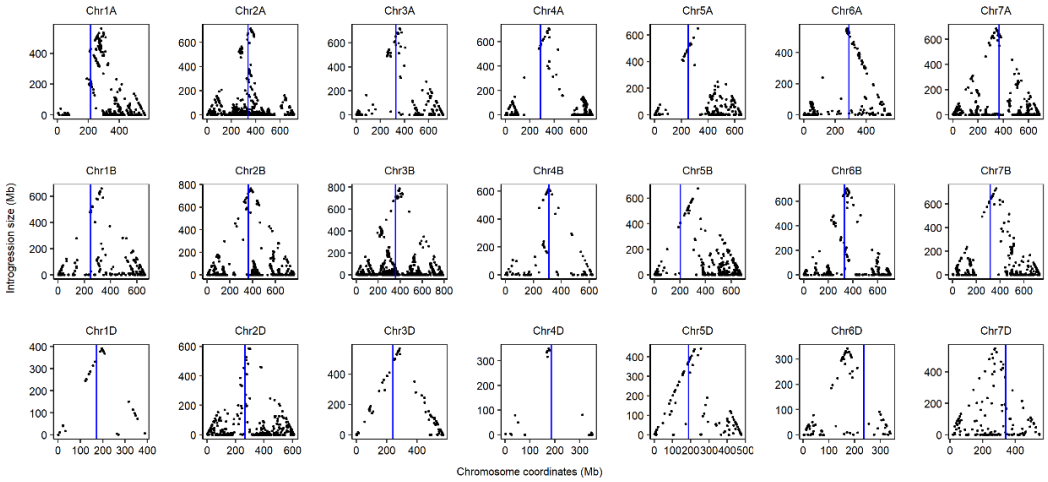


Figure 4.2 Size distribution of homozygous introgressions along the 21 chromosomes. The y-axis represents the introgression size (Mb) and the x-axis represents the chromosomal coordinates. The blue vertical lines depict the position of the centromeres.

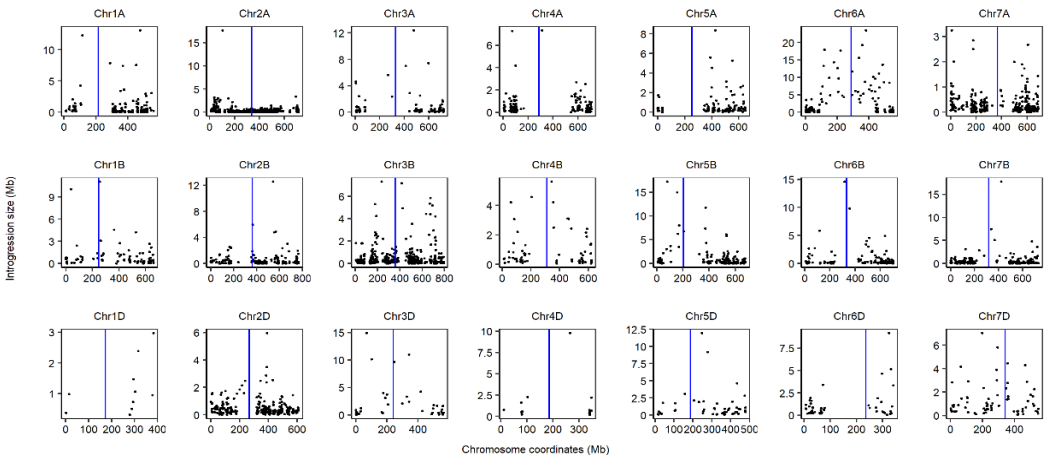


Figure 4.3 Size distribution of heterozygous introgressions along the 21 chromosomes. The y-axis represents the introgression size (Mb) and the x-axis represents the chromosomal coordinates. The blue vertical lines depict the position of the centromeres.

The distribution of the introgressions in each BIL is represented in the genome composition plots (Figure 4.4). Interestingly, these introgressions were distributed across the chromosomes and among the individuals for the A and B subgenomes. However, the extent of SHW introgression in the D subgenome was considerably less than the A and B subgenomes (Figure 4.4). In complete contrast, chromosome 6A was nearly entirely composed of SHW alleles with very few individuals showing Carberry alleles (Figure 4.4). While not as pronounced, several other chromosomes showed chromosome-specific distinctive introgression patterns, e.g., 1A, 2A, 7A, 2B, 3B and 7B (Figure 4.4). In general, smaller introgressions were found at the distal ends of chromosomes, while large introgressions were observed only in the centromeric and pericentromeric regions (Figure 4.2).

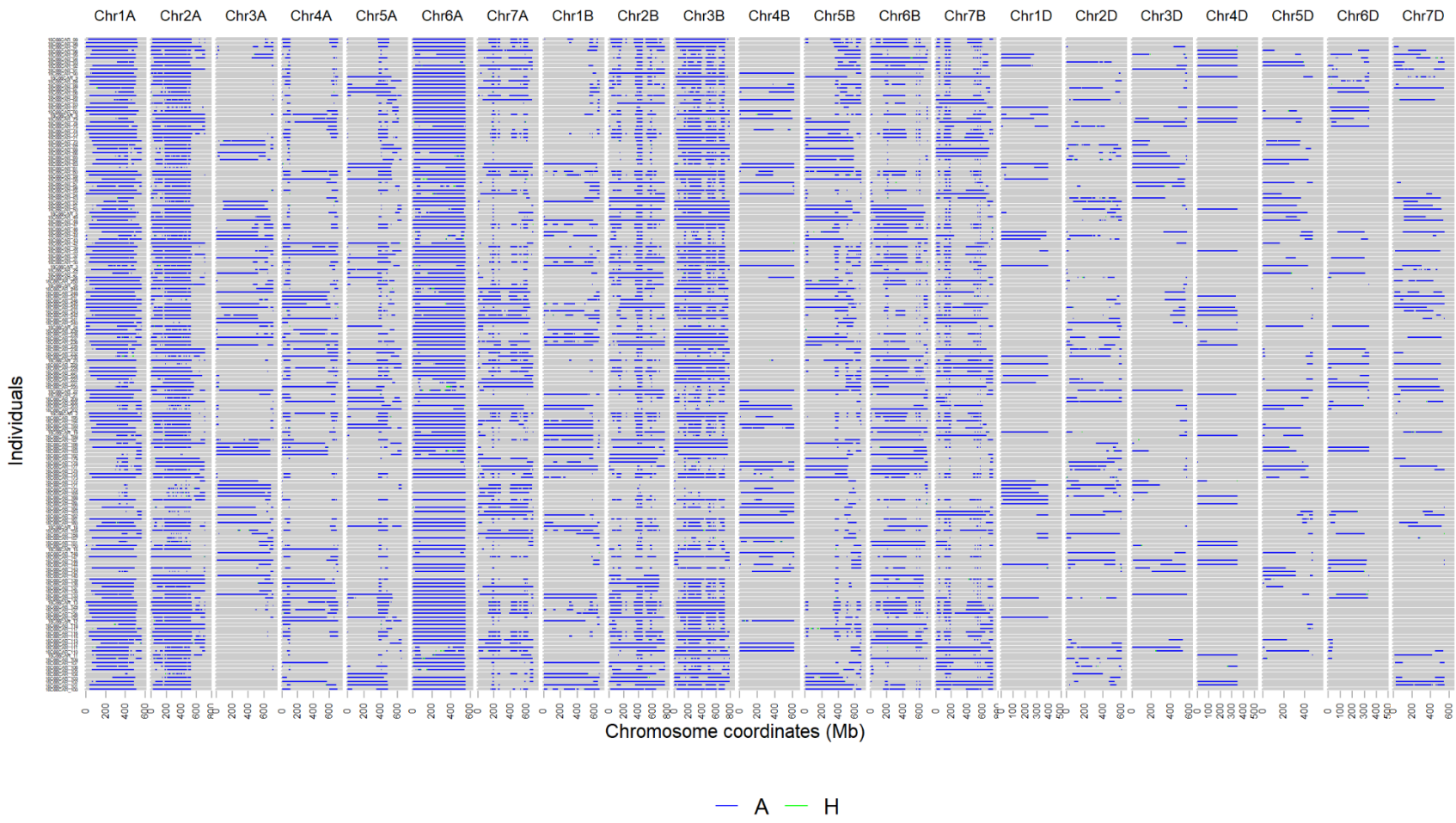


Figure 4.4 The genomic composition of the 21 chromosomes of the three subgenomes of the 173 individuals of the backcross population. The blue regions represent the homozygous SHW-C66 introgression fragments and the green regions represent the heterozygous regions.

The recombination rates across the genome were found to be higher in the telomeric and interstitial regions of the chromosomes. The recombination rate was lower in the pericentromeric regions of most chromosomes, except in chromosomes 2A, 7A, 2B and 3B (Figure 4.5). In agreement with the introgression distribution, the recombination rates were lower in the D subgenome than in the A and B subgenomes (Figure 4.5).

Similarly, the donor allele frequencies across all seven D subgenome chromosomes were lower than the expected 25% for a BC_1F_5 population and higher than the 25% threshold across the entire chromosome 6A (Figure 4.6). The donor allele frequencies were maintained at the same level over long stretches around the centromere, corresponding to the large introgressions in this region (Figure 4.6).

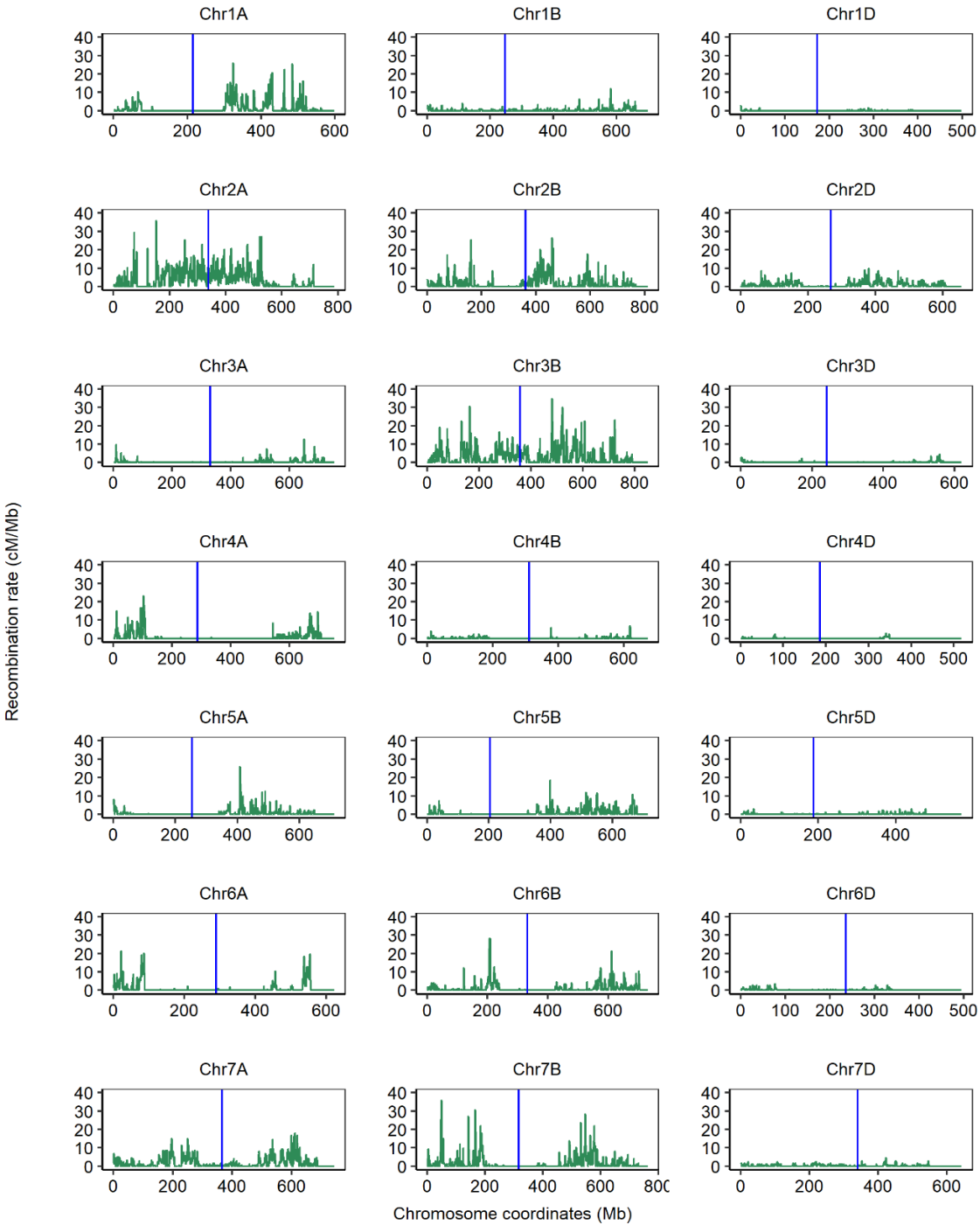


Figure 4.5 Distribution of recombination rates of the backcross population across the 21 chromosomes of the wheat genome. The x-axis represents the chromosome coordinates, and the y-axis represents the recombination rate in cM/Mb. The blue vertical lines depict the position of the centromeres.

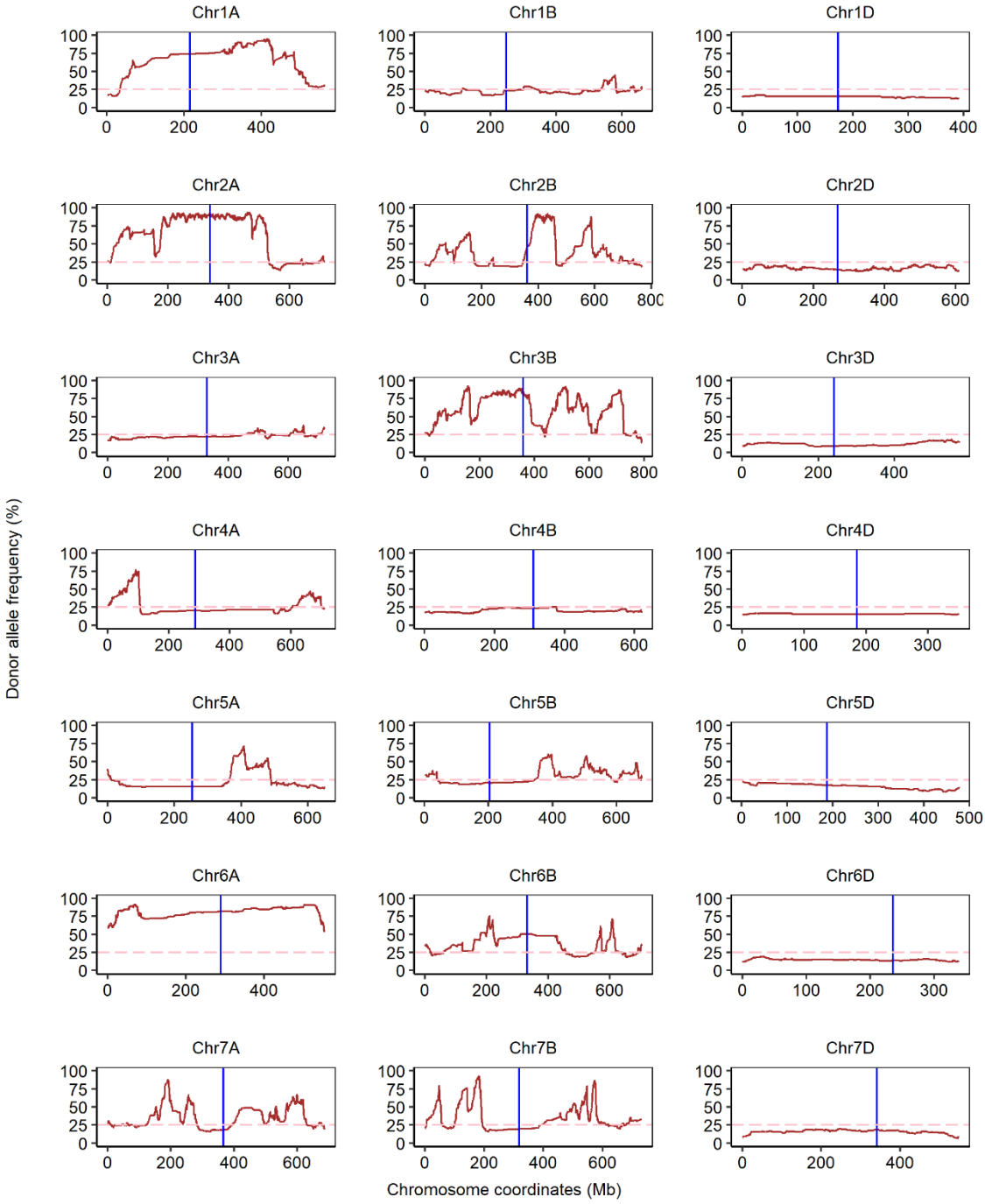


Figure 4.6 Donor allele frequencies across the 21 wheat chromosomes. The blue vertical lines depict the position of the centromeres. The pink horizontal dashed lines represent the expected donor allele frequencies of 25% for a BC_1F_5 population.

In the elite wheat Carberry × elite wheat AC Cadillac doubled haploid population, a data set representing 775 individuals genotyped with 6,804 SNP markers was available (Bokore et al. 2020). However, filtering of the markers with missing data would have resulted in very low marker density. Hence, only the individuals with genotype data for more than 99% of the markers were retained and, from that reduced population, only the markers without missing data were kept. This resulted in a total of 639 individuals and 4,627 markers. Out of these, coordinates on IWGSC Chinese Spring RefSeq v2.1 were identified for 4,505 markers, which were used for downstream analysis. The distribution of these markers across the RefSeq v2.1 genome are represented in Figure 4.7. The D subgenome had the lowest marker density with only 30 markers on chromosome 4D. Chromosome 2B had the highest marker density with 521 markers.

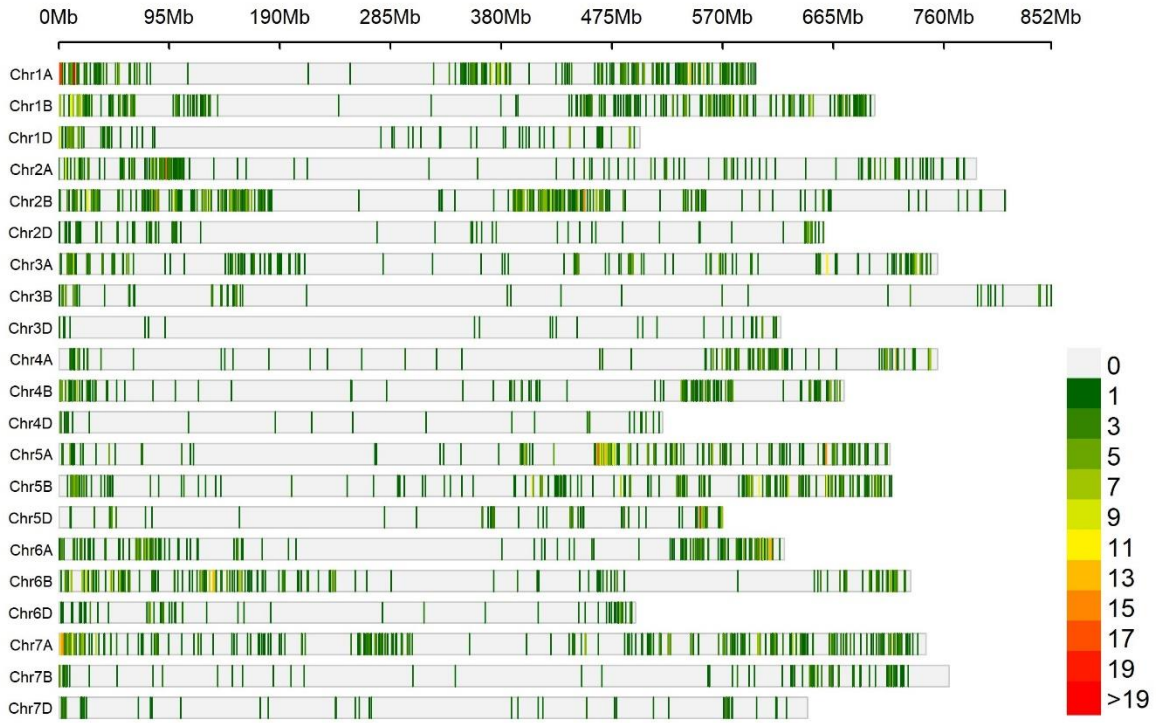


Figure 4.7 SNP density plot. Distribution of the 4,505 markers from the Carberry × AC Cadillac doubled haploid population across the 21 wheat chromosomes based on IWGSC RefSeq v2.1.

Using this marker dataset, it was possible to observe that small introgressions were detected near the distal ends of chromosome and that introgressions progressively became larger towards the centromeres. Here the term introgression refers to the chromosome fragments from AC Cadillac (Figure 4.8). In agreement with the introgression patterns, the recombination rate was higher within the distal regions of the chromosomes, but the amplitude of the patterns varied across chromosomes (Figure 4.9).

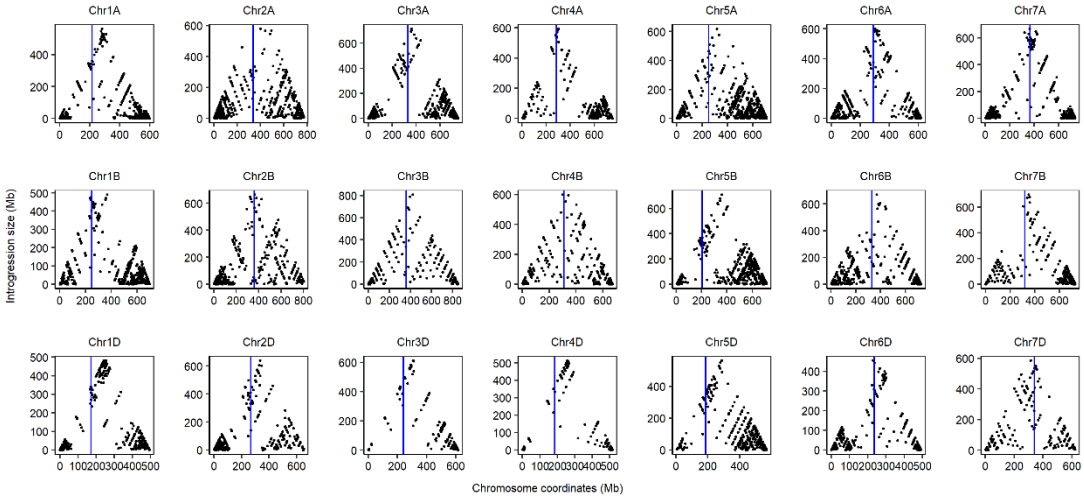


Figure 4.8 The distribution of homozygous introgression fragments along the 21 chromosomes, in the doubled haploid population. The y-axis represents the introgression size (Mb) and the x-axis represents the chromosomal coordinates. The blue vertical lines depict the position of the centromere.

As expected for a doubled haploid population, the donor allele frequency was 50% across the entire genome, i.e., the Carberry and AC Cadillac alleles were equally represented. Further, the composition of the three subgenomes revealed large segments of the Carberry and AC Cadillac genomes around the centromere and smaller segments towards the distal ends (Appendix 41). In contrast to the BC_1F_5 population, no difference was observed between the subgenomes for recombination and donor allele frequencies (Figures 4.9 and 4.10).

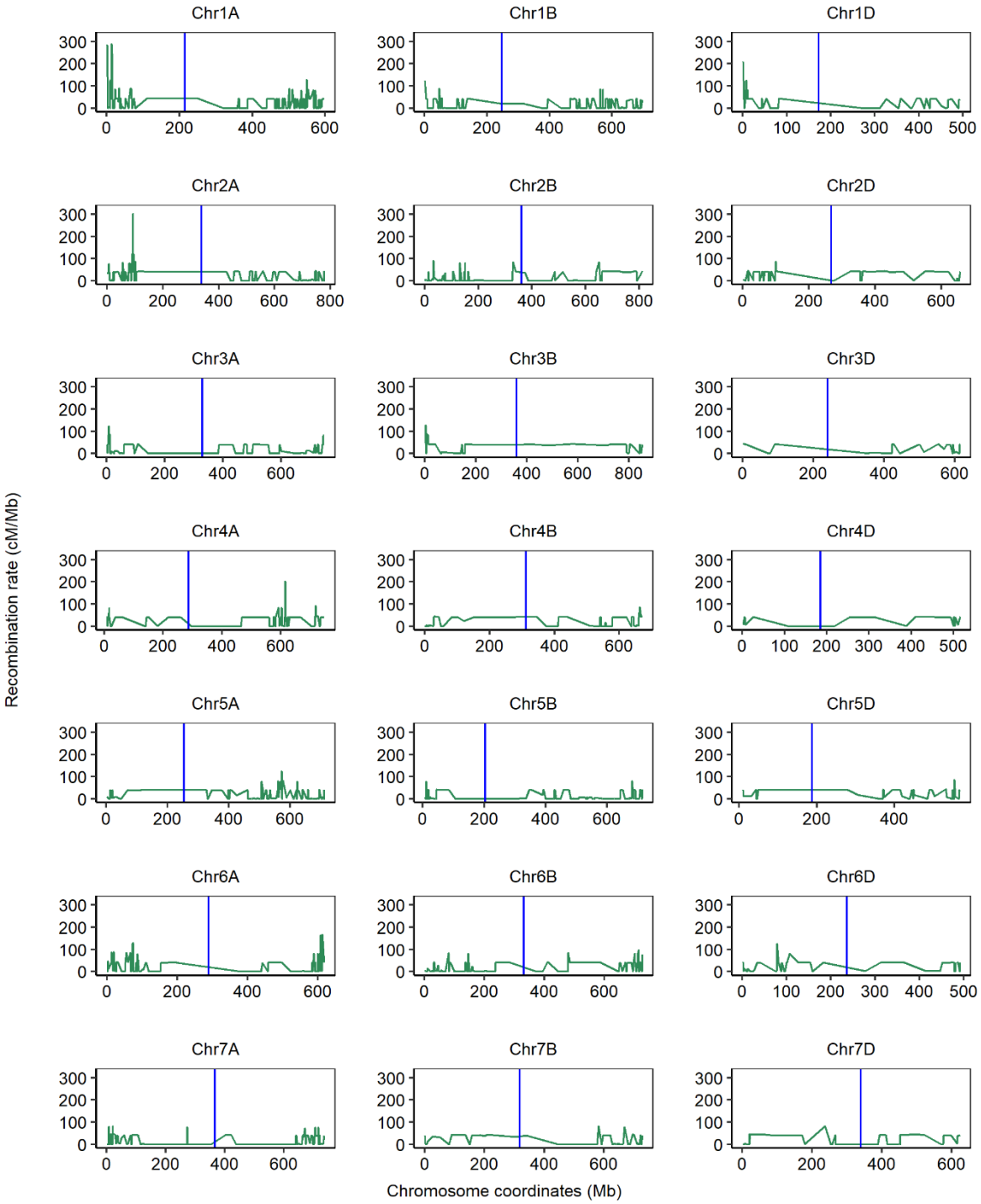


Figure 4.9 Distribution of the recombination rates of the doubled haploid population across the 21 chromosomes of the wheat genome. The x-axis represents the chromosome coordinates and the y-axis represents the recombination rate in cM/Mb. The blue vertical lines depict the centromere positions based on RefSeq v2.1.

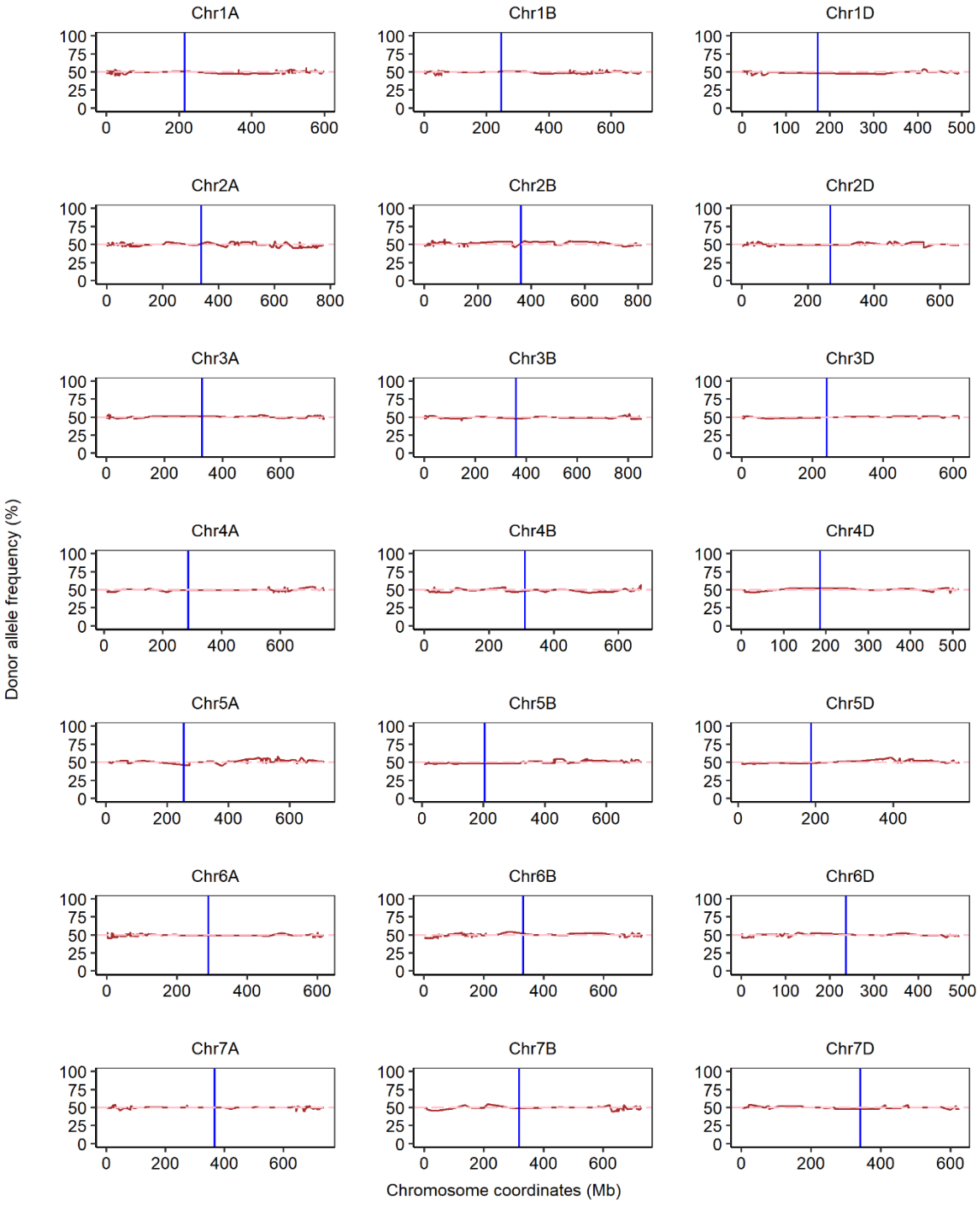


Figure 4.10 Allele frequency across the 21 wheat chromosomes. The blue vertical lines depict the position of the centromeres. The pink horizontal dashed lines represent the expected allele frequency of 50% for a doubled haploid population.

4.5 Discussion

Introduction of novel alleles into elite cultivar backgrounds is essential for achieving the goals of breeding programs such as developing climate-resilient and disease-resistant cultivars. In cross breeding, the introduction of allelic variants depends on meiotic recombination. This study provides information on the pattern of transfer of genetic material from a SHW line into an elite bread wheat accession. Specifically, SNP and haplotype-based estimates of COs, introgression sizes, recombination rates and donor allele frequencies were unearthed to provide insights into the non-uniform recombination landscapes among subgenomes. In addition, based on the already available datasets, a cursory comparison with an elite wheat × elite wheat-derived doubled haploid population with a common parent was carried out.

The backcross population recurrent parent Carberry is an elite cultivar belonging to the predominant Canadian Western Red Spring (CWRS) market class (DePauw et al. 2011), that occupies ~70% of the ~23 million acres of wheat grown annually on the Canadian Prairies. Carberry was derived from a cross between Alsen, a hard red spring wheat from North Dakota (Frohberg et al. 2006) and Superb, a Canadian hard red spring wheat (Townley-Smith et al. 2010). The donor SHW-C66 was derived from an interspecific cross between *Triticum turgidum* ssp. *dicoccum* PI377655 from former Yugoslavia and *Ae. tauschii* AS2386 from Iran (Zhang et al. 2010). The pedigrees of the recipient and donor parents highlight their contrasting geographical origins and genetic composition. In an analysis of a panel of more than 50,000 hexaploid wheat accessions comprised of several subgroups including landraces, elite lines, SHW and their derivatives, the elite wheat lines and SHW were farthest apart by genetic distance (Sansaloni et al. 2020).

The analysis of the switch between the parental alleles across the chromosomes indicated higher recombination rates within the distal and interstitial regions of the

chromosome compared to the pericentromeric regions. The introgression size in a genomic region is inversely related to the local recombination rates (Dreissig et al. 2020). This is reflected in the BILs by the presence of smaller introgressions towards the chromosomal ends and larger ones around the centromere because recombination is relatively repressed around the centromere (Danguy des Déserts et al. 2021; Tock et al. 2021). The higher recombination rate in the pericentromeric regions of chromosomes 2A, 2B, 3B and 7A are also reflected by the presence of smaller introgressions in this region. Heterozygous introgressions are overall smaller, suggesting successive meiotic recombination events in these regions, in nested fashion, at later generations of advancement of the population, decreasing their size (Dreissig et al. 2020).

The variation in recombination rates across chromosomes and the generation of specific introgression patterns can be attributed to several aspects. Environmental factors such as high temperature during meiosis have been shown to have an impact on the recombination landscape along the chromosomes in wheat (Coulton et al. 2020). However, every generation of the BC_1F_5 population was strictly produced in growth chambers with a 16h photoperiod at 22/18°C, thereby eliminating the potential impact of high temperatures. Earlier studies have indicated that specific DNA sequence motifs play a role in determining the CO location in Arabidopsis (Shilo et al. 2015), maize (He et al. 2017) and tomato (Demirci et al. 2017). In wheat, *TIR-MARINER* DNA transposon-related motifs were found to be associated with recombination (Darrier et al. 2017). Importantly, the epigenetic state of the chromatin, such as open chromatin signature, determines the recombination rate of that region of the chromosome (Wang et al. 2022a). For instance, in the model plant Arabidopsis, nucleosome occupancy was found to be associated with increased COs (Shilo et al. 2015) while histone 3 lysine 9 dimethylation (H3K9me2) and non-CG DNA methylation negatively impacted COs (Underwood et al. 2018). Similar impact of epigenetic chromatin

marks on recombination have been reported in crops such as soybean (Ma et al. 2023) and maize (Zhao et al. 2021b). In wheat, H3K27me3 marks in the distal regions of chromosomes have been suggested to be associated with increased CO activity in the area (Tock et al. 2021). In addition, in a BC_1F_5 population, the COs accumulated over six meiotic divisions (six generations of population advancement) can be theoretically captured. However, the reduction in heterozygosity during each round of selfing precludes the detection of all COs (Dreissig et al. 2020) because recombinants resulting from COs between regions with identical genetic composition cannot be detected.

The donor allele frequencies and the mosaic genome composition of donor and recurrent parent alleles in the individuals indicate that the A and B subgenomes are more receptive to the SHW donor alleles than the D subgenome of the elite bread wheat cultivar. The lower introgression/recombination rates in the D subgenome could be a consequence of the extent of the structural variations between the D subgenome of Carberry and that of SHW-C66 which carries the D genome of the *Ae. tauschii* accession AS2386. Indeed in Arabidopsis, repression of CO caused by structural variations, irrespective of their size and type, has been reported (Rowan et al. 2019). In contrast, chromosome 6A shows extensive introgression across its entire length. A previous study of introgression patterns of *Gossypium tomentosum* and *G. barbadense* in *G. hirsutum* lines highlighted the effects of genetic background on introgression patterns (Waghmare et al. 2016). In tomato, a recombinant inbred line population of 52 individuals, derived from an interspecific cross between *Solanum lycopersicum* × *S*. *pimpinellifolium*, revealed an 80% reduction in CO frequency compared to intraspecific crosses (Demirci et al. 2017). An introgression study using reciprocal backcross populations of *G. barbadense* and *G. hirsutum* also revealed different introgression hotspots and recalcitrant regions depending upon the recipient genomic backgrounds, suggesting the gradual accumulation of *G. hirsutum* chromatin

already in the *G. barbadense* cultivars over decades of crop improvement and the presence of enhanced fractions of homologous regions as a reason for increased introgression in the *G. barbadense* background (Adhikari et al. 2023). Further, when *G. hirsutum* was used as the donor parent, more donor alleles were retained in the A subgenome of the backcross population than in the D subgenome of *G*. *barbadense* (Adhikari et al. 2023). A lower donor allele frequency was similarly observed in the D subgenome than in the A and B subgenomes of Carberry in our study.

In the doubled haploid population derived from Carberry × AC Cadillac cross, high level of recombination was observed at the ends of the chromosomes, in agreement with the recombination landscapes reported in wheat (Akhunov et al. 2003; Choulet et al. 2014) and other crop species. However, the availability of only 4,505 polymorphic markers for the analysis and the non-uniform distribution of the markers across the chromosomes limited our ability to capture all the recombination events. A recent genetic diversity analysis using historical and modern Canadian spring wheat cultivars revealed high kinship co-efficient for most cultivar comparisons, indicative of allelic redundancy among cultivars, and diversity indices declining over the past four decades due to extensive selection for the market quality requirements (Semagn et al. 2021) dictated by the international wheat trade. Carberry and AC Cadillac had a kinship value of 0.648, indicating the high relatedness between the cultivars (Semagn et al. 2021). Consequently, the low polymorphism level prevented the detection of recombination events. Although recombination may have occurred at many loci that were neither detected nor did they introduce any genetic variability.

In contrast to the Carberry × SHW-C66 backcross population, fewer introgressions were detected in the pericentromeric and interstitial regions of the chromosomes. The distribution of the introgression sizes across the length of the chromosome was in

agreement with the distribution of recombination events. The donor allele frequencies of the entire genome was exactly as expected, i.e. 50% for doubled haploid population, compared to the 25% expected in the Carberry × SHW-C66 backcross population. The highly similar genomic composition of the two elite genotypes posed no resistance for recombination through the entirety of their genomes, and no subgenome-specific variations were observed for chromatin exchange between the two parental genotypes. The Appendix 41, illustrates the distribution of the parental genomic fragments in the doubled haploid population. It also shows shorter recombining regions in the distal regions of chromosomes. Evaluation of the recombination landscapes in four diverging populations of hexaploid wheat landraces revealed differences in recombination with increased genetic divergence (Danguy des Déserts et al. 2021). Upon analysing 13 recombinant inbred populations of wheat, Gardiner et al. (2019) showed that the use of a divergent parent in a cross resulted in novel CO profiles capable of breaking down undesirable linkage. Hence, the variation in the recombination landscape observed between the two populations is likely the result of the genomic divergence between the elite wheat and the SHW line.

The main limitation of our comparison between the elite wheat × SHW and elite wheat × elite wheat crosses is that different genotyping strategies (GBS vs array) were used, which captures genome-wide variation and gene space variation, respectively, as the Illumina iSelect 90k array was designed based on expressed sequence tags (Wang et al. 2014). In addition, there were differences in analysis steps such as allowing 20% missing data for the BILs and retaining individuals with missing data in less than 1% of the markers only for the doubled haploid population in the initial filtering steps. Applying the same cut-off of 20% in the doubled haploid population retained 767 individuals, however, the filtering of markers with missing data in the doubled haploid population would have resulted in only 338 markers, which would have made it unsuitable for CO analysis. Further, two different

population types, namely BC_1F_5 and doubled haploid populations, were compared. The multiple steps involved in doubled haploid generation, including haploid embryo formation, embryo rescue using artificial media and colchicine-induced chromosome doubling, can also result in biased selection of certain genomic regions and chromosomal aberrations (Ma et al. 2016; Shrestha et al. 2023). The backcross population allows the investigation of CO events accumulated over multiple meiosis cycles, specifically, six rounds of meiosis in a BC_1F_5 population. In contrast, the doubled haploid population captures the CO events that led to the generation of gametes over the single meiotic division (at F_1) that gave rise to the 639 individuals used in the study. Interestingly, two studies in wheat using different genotyping strategies and populations, specifically axiom 35k SNP array for recombinant inbred line populations (Gardiner et al. 2019) and TaBW410k SNP array for a collection of landraces (Danguy des Déserts et al. 2021), showed similar results of varying CO profiles with diverging genomic composition.

Overall, the wide genetic variability between the SHW and elite wheat germplasm highlights the diversity that the SHW material can bring into the elite wheat breeding population rather than using genetically similar elite cultivar × elite cultivar cross derivatives. Though elite cultivar × elite cultivar cross derivatives tend to be preferred to more readily achieve market requirements of quality parameters, the introduction of novel alleles from SHW lines conferring adaptive traits through pre-breeding and downstream polishing within the variety development pipeline could be advantageous considering the challenges posed by climate change and the need for increased diversity. The widespread recombination distribution suggests the utility of the pre-breeding material for broadening diversity with a potential for breaking linkage drag between genes located in genomic regions with conventionally low crossovers. However, the reduced introgression in the D subgenome demonstrates the need to further understand the transfer of genetic material from pre-

breeding to elite breeding populations using a range of genetic backgrounds. In addition, manipulation of CO determining factors such as knocking out anti-CO genes like FANCM encoding genes (Desjardins et al. 2022) or CRISPR/Cas9-mediated targeted induction of DSBs and modification of the epigenetic profiles can alter the number and distribution of CO events across the chromosomes, and thus have the potential to enhance introgression events in the hexaploid wheat genome (Taagen et al. 2020).

Although our analysis and comparison are based on the best available genotypic evidence, they should be interpreted with some degree of caution. Further analysis using other SHW × elite wheat and elite wheat × elite wheat populations of similar nature (both doubled haploids or both BILs) and similar genotyping platform (both using GBS or arrays) are expected to provide deeper insights on the introgression bias of the SHW chromatin into elite wheat backgrounds, creating new avenues for broadening the genetic diversity within wheat breeding programs, specifically as it pertains to climate resilience.

CHAPTER 5

General discussion

With the end of the last ice age ~12,000 years ago, humans started transitioning from a nomadic hunter-gatherer lifestyle to large scale food production systems from specific land areas. This switch involved domestication of wild crop species and the evolution of crop growing practices, bringing about the first agricultural revolution (Fuller et al. 2023). This laid the foundation for advancements in science and cultural changes. With an improved understanding in the foundational science of agriculture, *Homo sapiens* have experienced many of the challenges associated with food production and have taken responsive measures to maintain food supplies (Carey 2023). Currently, with the benefits of genomic revolution, we are working on tackling the challenges posed by climate change, population growth and other unprecedented events, to ensure food security. The legacy effects of domestication and subsequent rigorous plant breeding over centuries has eroded the genetic diversity of plant species which are consequently less adaptable to the rapidly changing biotic and abiotic stresses. Nowadays, scientists are turning to crop wild relatives to attempt to broaden the genetic diversity of modern day crop and germplasm to enhance the resilience of our agro-ecosystems.

The allohexaploid bread wheat evolution involved multiple steps of domestication and polyploidization involving diploid and tetraploid species. This complex evolutionary route has created a plethora of domesticated and wild species, as genetic resource for current day hexaploid wheat, and there are currently 784,753 unique accessions of *Triticum* and *Aegilops* species available in genetic resource databases (Guzzon et al. 2022). However, this entire genetic diversity resource is far from being fully exploited. One of the barriers in utilising the different gene pools for wheat improvement lies in the complexity of

inter and intra subgenome cis and trans interactions altering gene expression upon polyploidization, affecting the recovery of the parental phenotype in the hexaploid genomic background (Leigh et al. 2022) and the unpredictable mosaic introgression patterns of the divergent genomic fragments into elite wheat backgrounds.

This study utilises next generation sequencing data to understand the transcriptome and introgression dynamics to characterize potential barriers to exploiting the genetic diversity from the primary gene pool (*T. turgidum*, 2n = 4x = 28, AABB and *Ae. tauschii* 2n = 2x = 14, DD) of hexaploid wheat, at the expression (transcriptome) and recombination levels. These were addressed as follows:

- During generation of synthetic hexaploid wheat (SHW): extensive interrogation of gene expression changes across tissues and parental genotypes upon polyploidization in SHW (Chapter #2).

- While using SHW as pre-breeding donor material: determining the introgression pattern of SHW into elite wheat background (Chapter #4).

Both findings are critical for the deployment of primary gene pool germplasm, because the application of SHW lines involves direct advancement of this germplasm along the breeding pipeline or crossing with other elite wheat cultivars for introgression purposes and generating synthetic-derived lines.

5.1 Dynamics of gene expression upon resynthesis

Cis-regulatory elements (CRE), are 12-15 bp long motifs of DNA, capable of altering the expression of genes by affecting the DNA binding specificity of transcription factors (Yocca and Edger 2022). The spatial, temporal and quantitative expression of a gene is regulated by combinations of CREs and their activity (Schmitz et al. 2021). The distribution of the

CREs is associated with accessible chromatin regions, which are defined as open chromatin susceptible to nucleases activity (Klemm et al. 2019; Lu et al. 2019; Ricci et al. 2019). Subgenome level differences are observed in the distribution of accessible chromatin regions imparting expression changes on the linked genes (Lu et al. 2019). The trans-acting factors/transcription factors are also core regulators of transcriptional networks. Cis and trans effects constitute ~42% and 58% of fibre domestication related regulatory changes in allotetraploid cotton (Bao et al. 2019). The impact of trans-regulation is higher than cis-regulation within species in allopolyploid cotton, potentially because polyploidization introduces trans-regulatory factors capable of acting across subgenomes to create novel regulatory patterns (Bao et al. 2019; Hu and Wendel 2019). Hence, the variation in cis- and trans- regulatory elements between the parental and SHW genomes could similarly be hypothesized as one of the factors underlying the gene expression changes observed upon polyploidization in wheat. By extension, the tissue-specific variability of the cis-element activity and transcription factor footprints may be responsible for the observed differences in expression bias at the tissue level (Marand et al. 2023).

In wheat, interrogations of gene regulation at the tissue level revealed the enrichment of footprints from TFs in a tissue-specific manner (Zhao et al. 2023). Although certain TFs were broadly transcribed and translated across tissues, their activity at some loci was controlled by epigenetic marks such as H3K27me3 (Zhao et al. 2023). Interestingly, a recent study in allohexaploid wheat has shown the presence of cell-type specific TFs and TF-dependent regulons in roots and their constituent cell types. Furthermore, asymmetric transcription of homologous genes was higher when using single-cell level transcriptome compared to bulk RNA sequencing, because the heterogeneous nature of the latter reduces the resolution and the asymmetry among homologues which invariably changes through developmental stages (Zhang et al. 2023). Although bulk

RNAseq was used in this thesis work, the variation in expression bias of triads across different tissues sampled at discrete developmental stages has been captured, possibly because most of these tissues, such as glumes, palea+lemma and hypocotyl, were majorly comprised of homogeneous cells.

Alternative splicing is a co-transcriptional regulatory mechanism of gene expression, contributing to transcript and protein diversity. Alternative splicing mechanisms are suggested to play a role in adaptive evolution over short evolutionary time scales (Singh and Ahi 2022). In agreement with other plant species, retained intron events were predominantly detected in the comparison between the SHWs and their parents. Although retained introns are most frequent, the biological significance of their transcripts could be intermediary or regulatory considering that their predominance is not always reflected in the translatome (Petrillo 2023). If transcripts with retained introns are considered dead-ends leading to their degradation (Petrillo 2023), they may also play a role in passing on only homoeologue-specific transcripts to the translatome. In another outlook, the splicing regulatory sites on the mRNA and splicing regulatory elements (SREs) are cis- and trans-acting factors, respectively. Changes in their expression can modify the splicing patterns, and in a loop-like fashion, the alternatively spliced transcripts can influence cis and trans elements (Muhammad et al. 2023). Reciprocal and with-replacement homoeologous exchanges (HEs) have been reported in resynthesized allotetraploid wheat (Zhang et al. 2020). The HE could alter the dosage of splicing factor encoding genes, which in turn, could result in alternative splicing profiles (Zhang et al. 2022). Although allohexaploid wheat possesses the *Pairing homoeologous* (*Ph*) loci, the negative control of HE is not always complete (He et al. 2017), suggesting its potential influence on gene expression and splicing regulation.

5.2 Introgression of SHW genomic fragments

Introgression breeding involves the transfer of genomic segments from related to crop species to improve traits of interest. This usually involves initial crossing and repeated backcrossings. Introgression from species with homoeologous genomes requires additional strategies such as employing high-crossability genotypes, such as Crocus, *Ph1* mutants, alien addition and alien substitution lines (King et al. 2022). SHW serves as a bridge for gene flow between the tetraploid and diploid progenitor species and hexaploid elite bread wheat. In contrast to elite cultivar × elite cultivar crosses in wheat, the location of the introgressions can be deciphered when using SHWs as donor parents, because of the high sequence diversity between the parents.

A study using a population derived from domesticated and wild barley indicated that meiosis associated gene *REC8* influenced crossover (CO) numbers and patterns (Dreissig et al. 2020). In hexaploid wheat, 14 genomic regions across 12 chromosomes were identified to be associated with CO frequency as estimated using a wheat nested associated mapping (NAM) population (Jordan et al. 2018). In a wheat study using recombinant inbred lines (RILs), the ATP-dependent RNA helicase *RecQ-7* located on chromosome 2A was suggested to play a role in gene conversion and CO frequencies (Gardiner et al. 2019). The SHW lines studied herein could potentially harbour diverse alleles of these and other undiscovered loci regulating recombination patterns across the chromosomes. The specific introgression patterns observed in certain chromosomes throughout most of the backcross introgression lines (BILs) could also reflect the presence of introgression hotspots, which can be validated further using the BILs derived from crosses of elite wheat with other SHW lines.

Overall, this study provides a detailed investigation of the mRNA transcriptional variation using resynthesized wheat and brings out the importance of parental genotypic background and combination, tissue of study, expression specificity and co-transcriptional regulation through splicing, and their prospective consequences on phenotype recovery. Hence, utilisation of these and other SHW materials in breeding programs to transfer genes of interest, especially those catalogued in the *Ae*. *tauschii* donor, should involve expression level evaluation using a transcriptome atlas, presenting the extent of influence of the inter-subgenome interactions. The preliminary study of the introgression patterns might serve as a reference in determining the efficiency of introgression of genomic regions of interest based on their chromosomal location.

5.3 Future perspectives

With the gene expression component of the resynthesized allopolyploids and their parents provided in this study, the next logical step would be to investigate the different layers of gene regulation. Firstly, the regulatory elements at the DNA sequence level can be studied using the Assay for Transposase-Accessible Chromatin with sequencing (ATAC-seq) to deduce the chromatin accessibility in the different tissues, collating with transcriptomic data, and interrogating DNA-sequence level evidence in these regions for the presence of CREs. It will be intriguing to compare the accessible chromatin regions in the tetraploid and diploid parents and the SHW lines to explore the trans subgenome interactions. Secondly, considering the importance of the epigenetic marks for the regulation of gene expression and the action of regulatory elements, it will be valuable to analyse the methylation and chromatin immunoprecipitation data to characterize the change in epigenetic marks upon polyploidization. At a higher resolution, the genome-wide chromatin interaction changes upon polyploidization can also be deduced with techniques like Hi-C, because chromatin interactions on gene expression in allopolyploidization have been detected in studies using

domesticated hexaploid wheat and tetraploid and diploid progenitors (Yuan et al. 2022). With respect to the introgression studies, with advancements in long-read sequencing technologies and their reduction in costs (Marx 2023), whole genome assemblies of the parents can be generated and the confidence of the parental alleles used to deduce the introgressions can be further increased.

A recent addition to the molecular genetic toolbox is the genome editing technology based on the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 system (Wang and Doudna 2023). Improvements in the CRISPR systems in crop plants now allow the simultaneous targeting of multiple genes in a single pathway (Pan et al. 2021) and the editing of regulatory elements (Song et al. 2022) and epigenetics marks, such as DNA methylation (Liu et al. 2022). Considering the large-scale effects of the parental genotypes on gene expression, if traits of interest from the accessions in the primary gene pool of wheat are not recovered even under different combinations of parental genomes in the SHWs, targeted editing of the regulatory elements involved in the gene expression may be a feasible alternative to enable the phenotype recovery. *De novo* domestication of resynthesized polyploids has also been suggested as a novel approach for crop improvement, with the advantage of the ability to simultaneously harness multiple beneficial traits from the progenitors, such as abiotic and biotic stress resistance, and introducing domestication and market quality related traits through genome editing (Teng et al. 2023).

This present study has clearly shown how using pre-breeding material derived from progenitor species, namely *T. turgidum* and *Ae. tauschii*, introduces genetic diversity in hexaploid bread wheat, which is visualized as novel introgression patterns. Analysis of additional backcross populations will allow further validation of these results and potentiates the mapping of new alleles underlying the CO and introgression pattern phenotypes brought in by the pre-breeding material. It will also open the door for probing the molecular

signatures of introgression hotspots. This is an important contribution towards wheat improvement programs for revitalizing the CO landscapes and improving outcomes of introgression breeding.

The developed BIL population can be utilised for quantitative trait locus (QTL) mapping of agronomic, disease resistance and physiological traits. Once the QTLs are identified, contribution of the introgressions to the traits of interest can be deduced. Based on this solid evidence for the role of progenitor genomic fragments in influencing the traits of interest, DNA markers can be developed for application in marker-assisted breeding. Multiple BIL populations can be assembled as a training population for genomic prediction of quantitative traits in wheat breeding programs (Bassi et al. 2016). The increased diversity and readily-defined haplotypes in the BILs will improve haplotype-based genomic prediction accuracy (Difabachew et al. 2023).

On a concluding note, the process of whole genome duplication involves various changes in genetic and epigenetic features of the genome, resulting in diversification – the purpose of polyploidization in an evolutionary context (Heslop-Harrison et al. 2022). It has been shown time and again that this diversification has led to increased adaptability and survival of the polyploids in wide-ranging environments (Dubcovsky and Dvorak 2007; Van de Peer et al. 2021). Hence, emergence of novel phenotypes and their impact on conquering niche environments are, in an evolutionary perspective, the rule rather than exception upon polyploidization. Therefore, recovery of the parental phenotype in the background of neopolyploids will always remain a challenge, and therefore a topic for new avenues for research.

CHAPTER 6

References

Aberkane H, Payne T, Kishi M, Smale M, Amri A, Jamora N (2020) Transferring diversity of goat grass to farmers' fields through the development of synthetic hexaploid wheat. Food Security 12:1017-1033

Adhikari J, Chandnani R, Vitrakoti D, Khanal S, Ployaram W, Paterson AH (2023) Comparative transmission genetics of introgressed chromatin in reciprocal advanced backcross populations in Gossypium (cotton) polyploids. Heredity 130:209-222

Akagi T, Jung K, Masuda K, Shimizu KK (2022) Polyploidy before and after domestication of crop species. Current Opinion in Plant Biology 69:102255

Akhunov ED, Akhunova AR, Anderson OD, Anderson JA, Blake N, Clegg MT, Coleman-Derr D, Conley EJ, Crossman CC, Deal KR (2010) Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC Genomics 11:702

Akhunov ED, Goodyear AW, Geng S, Qi L-L, Echalier B, Gill BS, Gustafson JP, Lazo G, Chao S, Anderson OD (2003) The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Research 13:753-763

Aliakbari Sadeghabad A, Dadkhodaie A, Heidari B (2020) Phenotypic and genetic diversity of leaf rust resistance in wheat wild relatives. Journal of Phytopathology 168:428-438

Alix K, Gérard PR, Schwarzacher T, Heslop-Harrison JS (2017) Polyploidy and interspecific hybridization: partners for adaptation, speciation and evolution in plants. Annals of Botany 120:183-194

Arora S, Steuernagel B, Gaurav K, Chandramohan S, Long Y, Matny O, Johnson R, Enk J, Periyannan S, Singh N, Asyraf Md Hatta M, Athiyannan N, Cheema J, Yu G, Kangara N, Ghosh S, Szabo LJ, Poland J, Bariana H, Jones JDG, Bentley AR, Ayliffe M, Olson E, Xu SS, Steffenson BJ, Lagudah E, Wulff BBH (2019) Resistance gene cloning from a wild crop relative by sequence capture and association genetics. Nature Biotechnology 37:139-143

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Molecular Biology Reporter 9:208-218

Bai D, Knott D (1992) Suppression of rust resistance in bread wheat (*Triticum aestivum* L.) by D-genome chromosomes. Genome 35:276-282

Bansal M, Kaur S, Dhaliwal HS, Bains NS, Bariana HS, Chhuneja P, Bansal UK (2017) Mapping of *Aegilops umbellulata*-derived leaf rust and stripe rust resistance loci in wheat. Plant Pathology 66:38-44

Bao Y, Hu G, Grover CE, Conover J, Yuan D, Wendel JF (2019) Unraveling cis and trans regulatory evolution during cotton domestication. Nature Communications 10:5399

Bassi FM, Bentley AR, Charmet G, Ortiz R, Crossa J (2016) Breeding schemes for the implementation of genomic selection in wheat (Triticum spp.). Plant Science 242:23-36

Bebber DP (2015) Range-expanding pests and pathogens in a warming world. Annual Review of Phytopathology 53:335-356

Benyahya F, Nadaud I, Da Ines O, Rimbert H, White C, Sourdille P (2020) SPO11.2 is essential for programmed double-strand break formation during meiosis in bread wheat (*Triticum aestivum* L.). The Plant Journal 104:30-43

Berchowitz LE, Francis KE, Bey AL, Copenhaver GP (2007) The Role of AtMUS81 in Interference-Insensitive Crossovers in *A*. *thaliana*. PLOS Genetics 3:e132

Bevan MW, Uauy C, Wulff BBH, Zhou J, Krasileva K, Clark MD (2017) Genomic innovation for crop improvement. Nature 543:346-354

Bhatta M, Morgounov A, Belamkar V, Poland J, Baenziger PS (2018a) Unlocking the novel genetic diversity and population structure of synthetic hexaploid wheat. BMC Genomics 19:591

Bhatta M, Morgounov A, Belamkar V, Yorgancılar A, Baenziger PS (2018b) Genome-wide association study reveals favorable alleles associated with common bunt resistance in synthetic hexaploid wheat. Euphytica 214:200

Bin Z, Qi P, Dongao H, Pan Z, Bowei C, Xianhong G, Zaiyun L (2019) Transcriptional aneuploidy responses of *Brassica rapa-oleracea* monosomic alien addition lines (MAALs) derived from natural allopolyploid *B. napus*. Frontiers in Genetics 10:67

Birchler JA (2012) Insights from paleogenomic and population studies into the consequences of dosage sensitive gene expression in plants. Current Opinion in Plant Biology 15:544-548

Birchler JA, Veitia R (2021) One hundred years of gene balance: How stoichiometric issues affect gene expression, genome evolution, and quantitative traits. Cytogenetic and Genome Research 161:529-550

Blackmon H, Adams R (2015) evobiR: Comparative and population genetic analyses. R package version 1

Boettiger AN, Bintu B, Moffitt JR, Wang S, Beliveau BJ, Fudenberg G, Imakaev M, Mirny LA, Wu C-t, Zhuang X (2016) Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature 529:418

Bokore FE, Knox RE, Cuthbert RD, Pozniak CJ, McCallum BD, N'Diaye A, DePauw RM, Campbell HL, Munro C, Singh A, Hiebert CW, McCartney CA, Sharpe AG, Singh AK, Spaner D, Fowler DB, Ruan Y, Berraies S, Meyer B (2020) Mapping quantitative trait loci associated with leaf rust resistance in five spring wheat populations using single nucleotide polymorphism markers. PLOS ONE 15:e0230855

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

Broad-Institute (2022) http://broadinstitute.github.io/picard/.

Calarco Joseph P, Borges F, Donoghue Mark TA, Van Ex F, Jullien Pauline E, Lopes T, Gardner R, Berger F, Feijó José A, Becker Jörg D, Martienssen Robert A (2012) Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. Cell 151:194-205

Capilla-Pérez L, Durand S, Hurel A, Lian Q, Chambon A, Taochy C, Solier V, Grelon M, Mercier R (2021) The synaptonemal complex imposes crossover interference and heterochiasmy in *Arabidopsis*. Proceedings of the National Academy of Sciences 118:e2023613118

Carey J (2023) Unearthing the origins of agriculture. Proceedings of the National Academy of Sciences 120:e2304407120

Challinor AJ, Watson J, Lobell DB, Howden SM, Smith DR, Chhetri N (2014) A meta-analysis of crop yield under climate change and adaptation. Nature Climate Change 4:287

Chaloner TM, Gurr SJ, Bebber DP (2021) Plant pathogen infection risk tracks global crop yields under climate change. Nature Climate Change 11:710-715

Chelaifa H, Chague V, Chalabi S, Mestiri I, Arnaud D, Deffains D, Lu Y, Belcram H, Huteau V, Chiquet J (2013) Prevalence of gene expression additivity in genetically stable wheat allohexaploids. New Phytologist 197:730-736

Chen Y, Song W, Xie X, Wang Z, Guan P, Peng H, Jiao Y, Ni Z, Sun Q, Guo W (2020) A collinearity-incorporating homology inference strategy for connecting emerging assemblies in the Triticeae tribe as a pilot practice in the plant pangenomic era. Molecular Plant 13:1694-1708

Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, Pingault L, Sourdille P, Couloux A, Paux E, Leroy P, Mangenot S, Guilhot N, Le Gouis J, Balfourier F, Alaux M, Jamilloux V, Poulain J, Durand C, Bellec A, Gaspin C, Safar J, Dolezel J, Rogers J, Vandepoele K, Aury J-M, Mayer K, Berges H, Quesneville H, Wincker P, Feuillet C (2014) Structural and functional partitioning of bread wheat chromosome 3B. Science 345:1249721

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly 6:80-92

Colston-Nepali L, Tigano A, Boyle B, Friesen V (2019) Hybridization does not currently pose conservation concerns to murres in the Atlantic. Conservation Genetics 20:1465-1470

Combes M-C, Cenci A, Baraille H, Bertrand B, Lashermes P (2011) Homeologous gene expression in response to growing temperature in a recent allopolyploid (*Coffea arabica* L.). Journal of Heredity 103:36-46

Conant GC, Birchler JA, Pires JC (2014) Dosage, duplication, and diploidization: clarifying the interplay of multiple models for duplicate gene evolution over time. Current Opinion in Plant Biology 19:91-98

Coombes B, Fellers J, Grewal S, Rusholme-Pilcher R, Hubbart-Edwards S, Yang C-y, Joynson R, King IP, King J, Hall A (2021) Whole genome sequencing uncovers the structural and transcriptomic landscape of hexaploid wheat/*Am. muticum* introgression lines. Plant Biotechnology Journal 21:482-496

Correr FH, Furtado A, Garcia AAF, Henry RJ, Margarido GRA (2022) Allele expression biases in mixed-ploid sugarcane accessions. bioRxiv 12:8778

Coulton A, Burridge AJ, Edwards KJ (2020) Examining the effects of temperature on recombination in wheat. Frontiers in Plant Science 11:230

Cox TS (1997) Deepening the wheat gene pool. Journal of Crop Production 1:1-25

Crespo-Herrera L, Singh RP, Reynolds M, Huerta-Espino J (2019) Genetics of greenbug resistance in synthetic hexaploid wheat derived germplasm. Frontiers in Plant Science 10:782

Crismani W, Girard C, Froger N, Pradillo M, Santos JL, Chelysheva L, Copenhaver GP, Horlow C, Mercier R (2012) FANCM limits meiotic crossovers. Science 336:1588-1590

Cruppe G, Cruz CD, Peterson G, Pedley K, Asif M, Fritz A, Calderon L, Lemes da Silva C, Todd T, Kuhnem P, Singh PK, Singh RP, Braun H-J, Barma NCD, Valent B (2020) Novel sources of wheat head blast resistance in modern breeding lines and wheat wild relatives. Plant Disease 104:35-43

Cui L, Wall PK, Leebens-Mack JH, Lindsay BG, Soltis DE, Doyle JJ, Soltis PS, Carlson JE, Arumuganathan K, Barakat A (2006) Widespread genome duplications throughout the history of flowering plants. Genome research 16:738-749

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Group GPA (2011) The variant call format and VCFtools. Bioinformatics 27:2156-2158

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H (2021) Twelve years of SAMtools and BCFtools. GigaScience 10

Danguy des Déserts A, Bouchet S, Sourdille P, Servin B (2021) Evolution of recombination landscapes in diverging populations of bread wheat. Genome Biology and Evolution 13:evab152

Darrier B, Rimbert H, Balfourier F, Pingault L, Josselin A-A, Servin B, Navarro J, Choulet F, Paux E, Sourdille P (2017) High-resolution mapping of crossover events in the hexaploid wheat genome suggests a universal recombination mechanism. Genetics 206:1373-1388

De Storme N, Mason A (2014) Plant speciation through chromosome instability and ploidy change: cellular mechanisms, molecular factors and evolutionary relevance. Current Plant Biology 1:10-33

del Pozo JC, Ramirez-Parra E (2015) Whole genome duplications in plants: an overview from *Arabidopsi*s. Journal of Experimental Botany 66:6991-7003

Delaneau O, Zazhytska M, Borel C, Giannuzzi G, Rey G, Howald C, Kumar S, Ongen H, Popadin K, Marbach D (2019) Chromatin three-dimensional interactions mediate genetic effects on gene expression. Science 364:eaat8266

Demirci S, van Dijk ADJ, Sanchez Perez G, Aflitos SA, de Ridder D, Peters SA (2017) Distribution, position and genomic characteristics of crossovers in tomato recombinant inbred lines derived from an interspecific cross between *Solanum lycopersicum* and *Solanum pimpinellifolium*. The Plant Journal 89:554-564

Dempewolf H, Baute G, Anderson J, Kilian B, Smith C, Guarino L (2017) Past and future use of wild relatives in crop breeding. Crop Science 57:1070-1082

DePauw RM, Knox RE, McCaig TN, Clarke FR, Clarke JM (2011) Carberry hard red spring wheat. Canadian Journal of Plant Science 91:529-534

Desjardins SD, Simmonds J, Guterman I, Kanyuka K, Burridge AJ, Tock AJ, Sanchez-Moran E, Franklin FCH, Henderson IR, Edwards KJ, Uauy C, Higgins JD (2022) FANCM promotes class I interfering crossovers and suppresses class II non-interfering crossovers in wheat meiosis. Nature Communications 13:3644

Diederichsen A, Davidson C (2022) Canada's Country Report for the Third Report on the State of the World's Plant Genetic Resources for Food and Agriculture.

Difabachew YF, Frisch M, Langstroff AL, Stahl A, Wittkop B, Snowdon RJ, Koch M, Kirchhoff M, Cselényi L, Wolf M, Förster J, Weber S, Okoye UJ, Zenke-Philippi C (2023) Genomic prediction with haplotype blocks in wheat. Frontiers in Plant Science 14

Djanaguiraman M, Prasad PVV, Kumari J, Sehgal SK, Friebe B, Djalovic I, Chen Y, Siddique KHM, Gill BS (2019) Alien chromosome segment from *Aegilops speltoides* and *Dasypyrum villosum* increases drought tolerance in wheat via profuse and deep root system. BMC Plant Biology 19:242

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR (2012) STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29:15-21

Dong Y, Hu G, Grover CE, Miller ER, Zhu S, Wendel JF (2022) Parental legacy versus regulatory innovation in salt stress responsiveness of allopolyploid cotton (*Gossypium*) species 111:872-887

Dong Z, Ma C, Tian X, Zhu C, Wang G, Lv Y, Friebe B, Li H, Liu W (2020) Genome-wide impacts of alien chromatin introgression on wheat gene transcriptions. Scientific Reports 10:4801

Doyle JJ, Flagel LE, Paterson AH, Rapp RA, Soltis DE, Soltis PS, Wendel JF (2008) Evolutionary genetics of genome merger and doubling in plants. Annual Review of Genetics 42:443-461

Dreisigacker S, Kishii M (2019) The use of synthetic hexaploid wheat in the CIMMYT Global Wheat Program. CIMMYT https://repository.cimmyt.org/handle/10883/20692

Dreisigacker S, Kishii M, Lage J, Warburton M (2008) Use of synthetic hexaploid wheat to increase diversity for CIMMYT bread wheat improvement. Australian Journal of Agricultural Research 59:413-420

Dreissig S, Maurer A, Sharma R, Milne L, Flavell AJ, Schmutzer T, Pillen K (2020) Natural variation in meiotic recombination rate shapes introgression patterns in intraspecific hybrids between wild and domesticated barley. New Phytologist 228:1852-1863

Dubcovsky J, Dvorak J (2007) Genome Plasticity a Key Factor in the Success of Polyploid Wheat Under Domestication. Science 316:1862-1866

Edger PP, Smith R, McKain MR, Cooley AM, Vallejo-Marin M, Yuan Y, Bewick AJ, Ji L, Platts AE, Bowman MJ, Childs KL, Washburn JD, Schmitz RJ, Smith GD, Pires JC, Puzey JR (2017) Subgenome Dominance in an Interspecific Hybrid, Synthetic Allopolyploid, and a 140-Year-Old Naturally Established Neo-Allopolyploid Monkeyflower. The Plant Cell 29:2150-2167

El Bouhssini M, Ogbonnaya FC, Chen M, Lhaloui S, Rihawi F, Dabbous A (2013) Sources of resistance in primary synthetic hexaploid wheat (*Triticum aestivum* L.) to insect pests: Hessian fly, Russian wheat aphid and Sunn pest in the fertile crescent. Genetic Resources and Crop Evolution 60:621-627

Elbashir AAE, Gorafi YSA, Tahir ISA, Kim J-S, Tsujimoto H (2017) Wheat multiple synthetic derivatives: a new source for heat stress tolerance adaptive traits. Breeding Science 67:248-256

Epstein R, Sajai N, Zelkowski M, Zhou A, Robbins KR, Pawlowski WP (2023) Exploring impact of recombination landscapes on breeding outcomes. Proceedings of the National Academy of Sciences 120:e2205785119

Fang L, Guan X, Zhang T (2017) Asymmetric evolution and domestication in allotetraploid cotton (*Gossypium hirsutum* L.). The Crop Journal 5:159-165

FAO (2021) https://www.fao.org/faostat/en/#home.

Fawcett JA, Maere S, Van de Peer Y (2009) Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. Proceedings of the National Academy of Sciences 106:5737

Feldman M, Levy AA (2012) Genome evolution due to allopolyploidization in wheat. Genetics 192:763-774

Feldman M, Levy AA (2015) Origin and evolution of wheat and related Triticeae species. In: Molnár-Láng M, Ceoloni C, Doležel J (eds) Alien introgression in wheat: cytogenetics, molecular biology, and genomics. Springer International Publishing, Cham, Germany, pp 21-76

Fernandes JB, Wlodzimierz P, Henderson IR (2019) Meiotic recombination within plant centromeres. Current Opinion in Plant Biology 48:26-35

Feuillet C, Langridge P, Waugh R (2008) Cereal breeding takes a walk on the wild side. Trends in Genetics 24:24-32

Fox DT, Soltis DE, Soltis PS, Ashman T-L, Van de Peer Y (2020) Polyploidy: A Biological Force From Cells to Ecosystems. Trends in Cell Biology 30:688-694

France MG, Enderle J, Röhrig S, Puchta H, Franklin FCH, Higgins JD (2021) ZYP1 is required for obligate cross-over formation and cross-over interference in *Arabidopsis*. Proceedings of the National Academy of Sciences 118:e2021671118

Frohberg RC, Stack RW, Olson T, Miller JD, Mergoum M (2006) Registration of 'Alsen' wheat. Crop Science 46:2311-2312

Fuentes RR, Nieuwenhuis R, Chouaref J, Hesselink T, Dooijeweert Wv, Broeck HCvd, Schijlen E, Fransz P, Stam M, Jong Hd, Trivino SD, Ridder Dd, Dijk ADJv, Peters SA (2022) Pollen sequencing reveals barriers and aberrant patterns of recombination in interspecific tomato hybrids. bioRxiv:2022.2007.2002.498571

Fuller DQ, Denham T, Allaby R (2023) Plant domestication and agricultural ecologies. Current Biology 33:R636-R649

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC (2007) Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. The Plant Cell 19:3403-3417

Gaeta RT, Yoo S-Y, Pires J, Doerge RW, Chen ZJ, Osborn TC (2009) Analysis of gene expression in resynthesized *Brassica napus* allopolyploids using arabidopsis 70mer oligo microarrays. PLoS One 4:e4760

Gallois J-L, Moury B, German-Retana S (2018) Role of the genetic background in resistance to plant viruses. International Journal of Molecular Sciences 19:2856

Gao P, Quilichini TD, Zhai C, Qin L, Nilsen KT, Li Q, Sharpe AG, Kochian LV, Zou J, Reddy ASN, Wei Y, Pozniak C, Patterson N, Gillmor CS, Datla R, Xiang D (2021) Alternative splicing dynamics and evolutionary divergence during embryogenesis in wheat species. Plant Biotechnology Journal 19:1624-1643

Garcia-Lozano M, Natarajan P, Levi A, Katam R, Lopez-Ortiz C, Nimmakayala P, Reddy UK (2021) Altered chromatin conformation and transcriptional regulation in watermelon following genome doubling. The Plant Journal 106:588-600

Gardiner L-J, Wingen LU, Bailey P, Joynson R, Brabbs T, Wright J, Higgins JD, Hall N, Griffiths S, Clavijo BJ, Hall A (2019) Analysis of the recombination landscape of hexaploid bread wheat reveals genes controlling recombination and gene conversion frequency. Genome Biology 20:69

Gaurav K, Arora S, Silva P, Sánchez-Martín J, Horsnell R, Gao L, Brar GS, Widrig V, John Raupp W, Singh N, Wu S, Kale SM, Chinoy C, Nicholson P, Quiroz-Chávez J, Simmonds J, Hayta S, Smedley MA, Harwood W, Pearce S, Gilbert D, Kangara N, Gardener C, Forner-Martínez M, Liu J, Yu G, Boden SA, Pascucci A, Ghosh S, Hafeez AN, O'Hara T, Waites J, Cheema J, Steuernagel B, Patpour M, Justesen AF, Liu S, Rudd JC, Avni R, Sharon A,

Steiner B, Kirana RP, Buerstmayr H, Mehrabi AA, Nasyrova FY, Chayut N, Matny O, Steffenson BJ, Sandhu N, Chhuneja P, Lagudah E, Elkot AF, Tyrrell S, Bian X, Davey RP, Simonsen M, Schauser L, Tiwari VK, Randy Kutcher H, Hucl P, Li A, Liu D-C, Mao L, Xu S, Brown-Guedira G, Faris J, Dvorak J, Luo M-C, Krasileva K, Lux T, Artmeier S, Mayer KFX, Uauy C, Mascher M, Bentley AR, Keller B, Poland J, Wulff BBH (2022) Population genomic analysis of *Aegilops tauschii* identifies targets for bread wheat improvement. Nature Biotechnology 40:422-431

Gill BS (2022) A century of cytogenetic and genome analysis: Impact on wheat crop improvement. In: Reynolds MP, Braun H-J (eds) Wheat improvement: Food security in a changing climate. Springer International Publishing, Cham, pp 277-297

Girard C, Chelysheva L, Choinard S, Froger N, Macaisne N, Lehmemdi A, Mazel J, Crismani W, Mercier R (2015) AAA-ATPase FIDGETIN-LIKE 1 and helicase FANCM antagonize meiotic crossovers by distinct mechanisms. PLOS Genetics 11:e1005369

Glover NM, Redestig H, Dessimoz C (2016) Homoeologs: what are they and how do we infer them? Trends in Plant Science 21:609-621

Golicz AA, Bhalla PL, Edwards D, Singh MB (2020) Rice 3D chromatin structure correlates with sequence variation and meiotic recombination rate. Communications Biology 3:235

Griffiths AG, Moraga R, Tausen M, Gupta V, Bilton TP, Campbell MA, Ashby R, Nagy I, Khan A, Larking A, Anderson C, Franzmayr B, Hancock K, Scott A, Ellison NW, Cox MP, Asp T, Mailund T, Schierup MH, Andersen SU (2019) Breaking free: The genomics of allopolyploidy-facilitated niche expansion in white clover. The Plant Cell 31:1466-1487

Guo X, Han F (2014) Asymmetric epigenetic modification and elimination of rDNA sequences by polyploidization in wheat. The Plant Cell 26:4311-4327

Gutbrod MJ, Martienssen RA (2020) Conserved chromosomal functions of RNA interference. Nature Reviews Genetics 21:311-331

Guzzon F, Gianella M, Giovannini P, Payne TS (2022) Conserving Wheat Genetic Resources. In: Reynolds MP, Braun H-J (eds) Wheat Improvement: Food Security in a Changing Climate. Springer International Publishing, Cham, pp 299-318

He F, Pasam R, Shi F, Kant S, Keeble-Gagnere G, Kay P, Forrest K, Fritz A, Hucl P, Wiebe K, Knox R, Cuthbert R, Pozniak C, Akhunova A, Morrell PL, Davies JP, Webb SR, Spangenberg G, Hayes B, Daetwyler H, Tibbits J, Hayden M, Akhunov E (2019) Exome sequencing highlights the role of wild-relative introgression in shaping the adaptive landscape of the wheat genome. Nature Genetics 51:896-904

He Y, Wang M, Dukowic-Schulze S, Zhou A, Tiang C-L, Shilo S, Sidhu GK, Eichten S, Bradbury P, Springer NM, Buckler ES, Levy AA, Sun Q, Pillardy J, Kianian PMA, Kianian SF, Chen C, Pawlowski WP (2017) Genomic features shaping the landscape of meiotic double-strand-break hotspots in maize. Proceedings of the National Academy of Sciences 114:12231-12236

Henderson IR, Bomblies K (2021) Evolution and plasticity of genome-wide meiotic recombination rates. Annual Review of Genetics 55:23-43

Heslop-Harrison JS, Schwarzacher T, Liu Q (2022) Polyploidy: its consequences and enabling role in plant diversification and evolution. Annals of Botany 131:1-10

Hickey LT, Hafeez AN, Robinson H, Jackson SA, Leal-Bertioli SC, Tester M, Gao C, Godwin ID, Hayes BJ, Wulff BBH (2019) Breeding crops to feed 10 billion. Nature Biotechnology 37:744-754

Hiebert CW, Moscou MJ, Hewitt T, Steuernagel B, Hernández-Pinzón I, Green P, Pujol V, Zhang P, Rouse MN, Jin Y, McIntosh RA, Upadhyaya N, Zhang J, Bhavani S, Vrána J, Karafiátová M, Huang L, Fetch T, Doležel J, Wulff BBH, Lagudah E, Spielmeyer W (2020) Stem rust resistance in wheat is suppressed by a subunit of the mediator complex. Nature Communications 11:1123

Hillers KJ (2004) Crossover interference. Current Biology 14:R1036-R1037

Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, Hillman-Jackson J, Kuhn RM, Pedersen JS, Pohl A, Raney BJ, Rosenbloom KR, Siepel A, Smith KE, Sugnet CW, Sultan-Qurraie A, Thomas DJ, Trumbower H, Weber RJ, Weirauch M, Zweig AS, Haussler D, Kent WJ (2006) The UCSC Genome Browser Database: update 2006. Nucleic Acids Research 34:D590-D598

Hong Y, Chen L, Du L-p, Su Z, Wang J, Ye X, Qi L, Zhang Z (2014) Transcript suppression of TaGW2 increased grain width and weight in bread wheat. Functional & Integrative Genomics 14:341-349

Hu G, Wendel JF (2019) Cis–trans controls and regulatory novelty accompanying allopolyploidization. New Phytologist 221:1691-1700

Huang J, Lu X, Wu H, Xie Y, Peng Q, Gu L, Wu J, Wang Y, Reddy ASN, Dong S (2020) Phytophthora effectors modulate genome-wide alternative splicing of host mRNAs to reprogram plant immunity. Molecular Plant 13:1470-1484

Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, Gornicki P (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum*/*Aegilops* complex and the evolutionary history of polyploid wheat. Proceedings of the National Academy of Sciences 99:8133-8138

Hughes TE, Langdale JA, Kelly S (2014) The impact of widespread regulatory neofunctionalization on homeolog gene evolution following whole-genome duplication in maize. Genome Research 24:1348-1355

Hyde L, Osman K, Winfield M, Sanchez-Moran E, Higgins JD, Henderson IR, Sparks C, Franklin FCH, Edwards KJ (2023) Identification, characterization, and rescue of CRISPR/Cas9 generated wheat SPO11-1 mutants. Plant Biotechnology Journal 21:405-418

IPCC (2022) 2022: Technical Summary. In: Climate Change 2022: Impacts, Adaptation and Vulnerability. . In: H.-O. Pörtner DCR, E.S. Poloczanska, K. Mintenbeck, M. Tignor, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem (ed) Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, pp 37-118

IWGSC (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345:1251788

IWGSC (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361:eaar7191

Jafarzadeh J, Bonnett D, Jannink J-L, Akdemir D, Dreisigacker S, Sorrells ME (2016) Breeding value of primary synthetic wheat genotypes for grain yield. PLoS One 11:e0162860

Janzen GM, Wang L, Hufford MB (2019) The extent of adaptive wild introgression in crops. New Phytologist 221:1279-1288

Jiang X, Assis R (2019) Rapid functional divergence after small-scale gene duplication in grasses. BMC Evolutionary Biology 19:97

Jiao W, Yuan J, Jiang S, Liu Y, Wang L, Liu M, Zheng D, Ye W, Wang X, Chen ZJ (2018) Asymmetrical changes of gene expression, small RNAs and chromatin in two resynthesized wheat allotetraploids. The Plant Journal 93:828-842

Jiao Y, Li J, Tang H, Paterson AH (2014) Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. The Plant Cell 26:2792-2802

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473:97-100

Johansson E, Henriksson T, Prieto-Linde ML, Andersson S, Ashraf R, Rahmatov M (2020) Diverse wheat-alien introgression lines as a basis for durable resistance and quality characteristics in bread wheat. Frontiers in Plant Science 11:1067

Jordan KW, Wang S, He F, Chao S, Lun Y, Paux E, Sourdille P, Sherman J, Akhunova A, Blake NK, Pumphrey MO, Glover K, Dubcovsky J, Talbert L, Akhunov ED (2018) The genetic architecture of genome-wide recombination rate variation in allopolyploid wheat revealed by nested association mapping. The Plant Journal 95:1039-1054

Juery C, Concia L, De Oliveira R, Papon N, Ramírez-González R, Benhamed M, Uauy C, Choulet F, Paux E (2021) New insights into homoeologous copy number variations in the hexaploid wheat genome. The Plant Genome 14:e20069

Kazan K (2003) Alternative splicing and proteome diversity in plants: the tip of the iceberg has just emerged. Trends in Plant Science 8:468-471

Keeney S, Giroux CN, Kleckner N (1997) Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. Cell 88:375-384

Kenan-Eichler M, Leshkowitz D, Tal L, Noor E, Melamed-Bessudo C, Feldman M, Levy AA (2011) Wheat Hybridization and Polyploidization Results in Deregulation of Small RNAs. Genetics 188:263-272

Khan D, Ziegler DJ, Kalichuk JL, Hoi V, Huynh N, Hajihassani A, Parkin IAP, Robinson SJ, Belmonte MF (2022) Gene expression profiling reveals transcription factor networks and subgenome bias during Brassica napus seed development 109:477-489

Kim J, Park J, Kim H, Son N, Kim E-J, Kim J, Byun D, Lee Y, Park YM, Nageswaran DC, Kuo P, Rose T, Dang TVT, Hwang I, Lambing C, Henderson IR, Choi K (2022) *Arabidopsis* HEAT SHOCK FACTOR BINDING PROTEIN is required to limit meiotic crossovers and *HEI10* transcription. The EMBO Journal 41:e109958

Kim SA, Brossard M, Roshandel D, Paterson AD, Bull SB, Yoo YJ (2019) gpart: human genome partitioning and visualization of high-density SNP data by identifying haplotype blocks. Bioinformatics 35:4419-4421

King J, Grewal S, Fellers JP, King IP (2022) Exploring untapped wheat genetic resources to boost food security. In: Reynolds MP, Braun H-J (eds) Wheat improvement: Food security in a changing climate. Springer International Publishing, Cham, pp 319-340

Kishii M (2019) An update of recent use of *Aegilops* species in wheat breeding. Frontiers in Plant Science 10:585

Klemm SL, Shipony Z, Greenleaf WJ (2019) Chromatin accessibility and the regulatory epigenome. Nature Reviews Genetics 20:207-220

Kreiner JM, Kron P, Husband BC (2017) Evolutionary Dynamics of Unreduced Gametes. Trends in Genetics 33:583-593

Krueger F, James F, Ewels P, Afyounian E, Weinstein M, Schuster-Boeckler B, Hulselmans G, Clamons S (2023) FelixKrueger/TrimGalore https://zenodo.org/badge/latestdoi/62039322.

Kryuchkova-Mostacci N, Robinson-Rechavi MJBib (2017) A benchmark of gene expression tissue-specificity metrics 18:205-214

Kurzbauer M-T, Uanschou C, Chen D, Schlögelhofer P (2012) The recombinases DMC1 and RAD51 are functionally and spatially separated during meiosis in *Arabidopsis*. The Plant Cell 24:2058-2070

Laloum T, Martín G, Duque P (2018) Alternative splicing control of abiotic stress responses. Trends in Plant Science 23:140-150

Langridge P, Reynolds M (2021) Breeding for drought and heat tolerance in wheat. Theoretical and Applied Genetics 134:1753-1769

Latorre SM, Were VM, Foster AJ, Langner T, Malmgren A, Harant A, Asuke S, Reyes-Avila S, Gupta DR, Jensen C, Ma W, Mahmud NU, Mehebub MS, Mulenga RM, Muzahid ANM, Paul SK, Rabby SMF, Rahat AAM, Ryder L, Shrestha R-K, Sichilima S, Soanes DM, Singh PK, Bentley AR, Saunders DGO, Tosa Y, Croll D, Lamour KH, Islam T, Tembo B, Win J, Talbot NJ, Burbano HA, Kamoun S (2023) Genomic surveillance uncovers a pandemic clonal lineage of the wheat blast fungus. PLOS Biology 21:e3002052

Lawrence EJ, Griffin CH, Henderson IR (2017) Modification of meiotic recombination by natural variation in plants. Journal of Experimental Botany 68:5471-5483

Leigh FJ, Wright TIC, Horsnell RA, Dyer S, Bentley AR (2022) Progenitor species hold untapped diversity for potential climate-responsive traits for use in wheat breeding and crop improvement. Heredity 128:291-303

Levy AA, Feldman M (2004) Genetic and epigenetic reprogramming of the wheat genome upon allopolyploidization. Biological Journal of the Linnean Society 82:607-613

Levy AA, Feldman M (2022) Evolution and origin of bread wheat. The Plant Cell 34:2549-2567

Li A, Liu D, Wu J, Zhao X, Hao M, Geng S, Yan J, Jiang X, Zhang L, Wu J, Yin L, Zhang R, Wu L, Zheng Y, Mao L (2014a) mRNA and small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. The Plant Cell 26:1878-1900

Li A, Liu D, Wu J, Zhao X, Hao M, Geng S, Yan J, Jiang X, Zhang L, Wu J, Yin L, Zhang R, Wu L, Zheng Y, Mao L (2014b) mRNA and Small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in nascent hexaploid wheat. The Plant Cell 26:1878-1900

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:13033997

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094-3100

Li H, Dong Z, Ma C, Xia Q, Tian X, Sehgal S, Koo D-H, Friebe B, Ma P, Liu W (2020) A spontaneous wheat-Aegilops longissima translocation carrying *Pm66* confers resistance to powdery mildew. Theoretical and Applied Genetics 133:1149-1159

Li N, Xu C, Zhang A, Lv R, Meng X, Lin X, Gong L, Wendel JF, Liu B (2019a) DNA methylation repatterning accompanying hybridization, whole genome doubling and homoeolog exchange in nascent segmental rice allotetraploids. New Phytologist 223:979-992

Li Z, Wang M, Lin K, Xie Y, Guo J, Ye L, Zhuang Y, Teng W, Ran X, Tong Y, Xue Y, Zhang W, Zhang Y (2019b) The bread wheat epigenomic map reveals distinct chromatin architectural and evolutionary features of functional genetic elements. Genome Biology 20:139

Liang X-Z, Wu Y, Chambers RG, Schmoldt DL, Gao W, Liu C, Liu Y-A, Sun C, Kennedy JA (2017) Determining climate effects on US total agricultural productivity. Proceedings of the National Academy of Sciences 114:E2285-E2292

Liu M, Zhang C, Yuan C, Zhang L, Huang L, Wu J, Wang J, Zheng Y, Zhang H, Liu D, Fu D (2013) Stripe rust resistance in *Aegilops tauschii* germplasm. Crop Science 53:2014-2020

Liu S, Sretenovic S, Fan T, Cheng Y, Li G, Qi A, Tang X, Xu Y, Guo W, Zhong Z, He Y, Liang Y, Han Q, Zheng X, Gu X, Qi Y, Zhang Y (2022) Hypercompact CRISPR–Cas12j2 (CasΦ) enables genome editing, gene activation, and epigenome editing in plants. Plant Communications 3:100453

Liu Y, Yuan J, Jia G, Ye W, Jeffrey Chen Z, Song Q (2021) Histone H3K27 dimethylation landscapes contribute to genome stability and genetic recombination during wheat polyploidization. The Plant Journal 105:678-690

Lloyd A (2023) Crossover patterning in plants. Plant Reproduction 36:55-72

Loidl J (2016) Conservation and variability of meiosis across the Eukaryotes. Annual Review of Genetics 50:293-316

Lu F-H, McKenzie N, Gardiner L-J, Luo M-C, Hall A, Bevan MW (2020) Reduced chromatin accessibility underlies gene expression differences in homologous chromosome arms of diploid *Aegilops tauschii* and hexaploid wheat. GigaScience 9:giaa070

Lu Z, Marand AP, Ricci WA, Ethridge CL, Zhang X, Schmitz RJ (2019) The prevalence, evolution and chromatin signatures of plant regulatory elements. Nature Plants 5:1250-1259

Lukens LN, Pires JC, Leon E, Vogelzang R, Oslach L, Osborn T (2006) Patterns of sequence loss and cytosine methylation within a population of newly resynthesized *Brassica napus* allopolyploids. Plant Physiology 140:336-348

Luo C, Li X, Zhang Q, Yan J (2019) Single gametophyte sequencing reveals that crossover events differ between sexes in maize. Nature Communications 10:785

Lv Z, Li Z, Wang M, Zhao F, Zhang W, Li C, Gong L, Zhang Y, Mason AS, Liu B (2021) Conservation and trans-regulation of histone modification in the A and B subgenomes of polyploid wheat during domestication and ploidy transition. BMC Biology 19:42

Ma J, Deng M, Lv S-Y, Yang Q, Jiang Q-T, Qi P-F, Li W, Chen G-Y, Lan X-J, Wei Y-M (2016) Identification of QTLs associated with tissue culture response of mature wheat embryos. SpringerPlus 5:1552

Ma X, Fan L, Zhang Z, Yang X, Liu Y, Ma Y, Pan Y, Zhou G, Zhang M, Ning H, Kong F, Ma J, Liu S, Tian Z (2023) Global dissection of the recombination landscape in soybean using a high-density 600K SoySNP array. Plant Biotechnology Journal 21:606-620

Madlung A, Tyagi AP, Watson B, Jiang H, Kagochi T, Doerge RW, Martienssen R, Comai L (2005) Genomic changes in synthetic Arabidopsis polyploids. The Plant Journal 41:221-230

Marand AP, Eveland AL, Kaufmann K, Springer NM (2023) cis-Regulatory Elements in Plant Development, Adaptation, and Evolution. Annual Review of Plant Biology 74:111-137

Margarido GRA, Correr FH, Furtado A, Botha FC, Henry RJ (2021) Limited allele-specific gene expression in highly polyploid sugarcane. Genome Research 32:297-308

Marks RA, Hotaling S, Frandsen PB, VanBuren R (2021) Representation and participation across 20 years of plant genome sequencing. Nature Plants 7:1571-1578

Martín G, Márquez Y, Mantica F, Duque P, Irimia M (2021) Alternative splicing landscapes in Arabidopsis thaliana across tissues and stress conditions highlight major functional differences with animals. Genome Biology 22:35

Martin SH, Jiggins CD (2017) Interpreting the genomic landscape of introgression. Current Opinion in Genetics & Development 47:69-74

Marx V (2023) Method of the year: long-read sequencing. Nature Methods 20:6-11

McAusland L, Vialet-Chabrand S, Jauregui I, Burridge A, Hubbart-Edwards S, Fryer MJ, King IP, King J, Pyke K, Edwards KJ, Carmo-Silva E, Lawson T, Murchie EH (2020) Variation in key leaf photosynthetic traits across wheat wild relatives is accession dependent not species dependent. New Phytologist 228:1767-1780

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research 20:1297-1303

McMahill MS, Sham CW, Bishop DK (2007) Synthesis-dependent strand annealing in meiosis. PLOS Biology 5:e299

Melamed-Bessudo C, Levy AA (2012) Deficiency in DNA methylation increases meiotic crossover rates in euchromatic but not in heterochromatic regions in Arabidopsis. Proceedings of the National Academy of Sciences 109:E981-E988

Mercier R, Mézard C, Jenczewski E, Macaisne N, Grelon M (2015) The molecular biology of meiosis in plants. Annual Review of Plant Biology 66:297-327

Min XJ, Powell B, Braessler J, Meinken J, Yu F, Sablok G (2015) Genome-wide cataloging and analysis of alternatively spliced genes in cereal crops. BMC Genomics 16:721

Mirzaghaderi G, Mason AS (2019) Broadening the bread wheat D genome. Theoretical and Applied Genetics 132:1295-1307

Molero G, Coombes B, Joynson R, Pinto F, Piñera-Chávez FJ, Rivera-Amado C, Hall A, Reynolds MP (2023) Exotic alleles contribute to heat tolerance in wheat under field conditions. Communications Biology 6:21

Moore G (2015) Strategic pre-breeding for wheat improvement. Nature Plants 1:15018

Morgounov A, Abugalieva A, Akan K, Akın B, Baenziger S, Bhatta M, Dababat AA, Demir L, Dutbayev Y, El Bouhssini M (2018) High-yielding winter synthetic hexaploid wheats resistant to multiple diseases and pests. Plant Genetic Resources 16:273-278

Moskal K, Kowalik S, Podyma W, Łapiński B, Boczkowska M (2021) The pros and cons of rye chromatin introgression into wheat genome. Agronomy 11:456

Mu N, Li Y, Li S, Shi W, Shen Y, Yang H, Zhang F, Tang D, Du G, You A, Cheng Z (2023) MUS81 is required for atypical recombination intermediate resolution but not crossover designation in rice. New Phytologist 237:2422-2434

Muhammad S, Xu X, Zhou W, Wu L (2023) Alternative splicing: An efficient regulatory approach towards plant developmental plasticity. WIREs RNA 14:e1758

Mujeeb-Kazi A, Gul A, Farooq M, Rizwan S, Ahmad I (2008) Rebirth of synthetic hexaploids with global implications for wheat improvement. Australian Journal of Agricultural Research 59:391-398

Munns R, James RA, Xu B, Athman A, Conn SJ, Jordans C, Byrt CS, Hare RA, Tyerman SD, Tester M, Plett D, Gilliham M (2012) Wheat grain yield on saline soils is improved by an ancestral Na+ transporter gene. Nature Biotechnology 30:360

Nageswaran DC, Kim J, Lambing C, Kim J, Park J, Kim E-J, Cho HS, Kim H, Byun D, Park YM, Kuo P, Lee S, Tock AJ, Zhao X, Hwang I, Choi K, Henderson IR (2021) *HIGH CROSSOVER RATE1* encodes PROTEIN PHOSPHATASE X1 and restricts meiotic crossovers in *Arabidopsis*. Nature Plants 7:452-467

Nelson SC, Doheny KF, Laurie CC, Mirel DB (2012) Is 'forward' the same as 'plus'?…and other adventures in SNP allele nomenclature. Trends in Genetics 28:361-363

Nevo E, Chen G (2010) Drought and salt tolerances in wild relatives for wheat and barley improvement. Plant, Cell & Environment 33:670-685

Nieto Feliner G, Casacuberta J, Wendel JF (2020) Genomics of evolutionary novelty in hybrids and polyploids. Frontiers in Genetics 11:792

Nyine M, Adhikari E, Clinesmith M, Jordan KW, Fritz AK, Akhunov E (2020) Genomic patterns of introgression in interspecific populations created by crossing wheat with its wild relative. G3 Genes Genomes Genetics 10:3651-3661

Ogbonnaya FC, Abdalla O, Mujeeb-Kazi A, Kazi AG, Xu SS, Gosman N, Lagudah ES, Bonnett D, Sorrells ME, Tsujimoto H (2013) Synthetic hexaploids: harnessing species of the primary gene pool for wheat improvement. Plant Breed Rev 37:35-122

Ortiz-Bobea A, Ault TR, Carrillo CM, Chambers RG, Lobell DB (2021) Anthropogenic climate change has slowed global agricultural productivity growth. Nature Climate Change 11:306-312

Ozkan H, Levy AA, Feldman M (2001) Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops–Triticum*) group. The Plant Cell 13:1735-1747

Pan C, Wu X, Markel K, Malzahn AA, Kundagrami N, Sretenovic S, Zhang Y, Cheng Y, Shih PM, Qi Y (2021) CRISPR–Act3.0 for highly efficient multiplexed gene activation in plants. Nature Plants 7:942-953

Pei L, Huang X, Liu Z, Tian X, You J, Li J, Fang DD, Lindsey K, Zhu L, Zhang X, Wang M (2022) Dynamic 3D genome architecture of cotton fiber reveals subgenome-coordinated chromatin topology for 4-staged single-cell differentiation. Genome Biology 23:45

Peng Z, Chen H, Tan L, Shu H, Varshney RK, Zhou Z, Zhao Z, Luo Z, Chitikineni A, Wang L, Maku J, López Y, Gallo M, Zhou H, Wang J (2020) Natural polymorphisms in a pair of

NSP2 homoeologs can cause loss of nodulation in peanut. Journal of Experimental Botany 72:1104-1118

Petersen S, Lyerly JH, Worthington ML, Parks WR, Cowger C, Marshall DS, Brown-Guedira G, Murphy JP (2015) Mapping of powdery mildew resistance gene *Pm53* introgressed from *Aegilops speltoides* into soft red winter wheat. Theoretical and Applied Genetics 128:303-312

Petrillo E (2023) Do not panic: An intron-centric guide to alternative splicing. The Plant Cell 35:1752-1761

Pfeifer M, Kugler KG, Sandve SR, Zhan B, Rudi H, Hvidsten TR, Mayer KFX, Olsen O-A (2014) Genome interplay in the grain transcriptome of hexaploid bread wheat. Science 345:1250091

Pour-Aboughadareh A, Ahmadi J, Mehrabi A, Moghaddam M, Etminan A (2018) Evaluation of agro-morphological diversity in wild relatives of wheat collected in Iran. Journal of Agricultural Science and Technology 19:943-956

Pour-Aboughadareh A, Kianersi F, Poczai P, Moradkhani H (2021) Potential of wild relatives of wheat: Ideal genetic resources for future breeding programs. Agronomy 11:1656

Powell JJ, Fitzgerald TL, Stiller J, Berkman PJ, Gardiner DM, Manners JM, Henry RJ, Kazan K (2017) The defence-associated transcriptome of hexaploid wheat displays homoeolog expression and induction bias. Plant Biotechnology Journal 15:533-543

Pradhan GP, Prasad PVV, Fritz AK, Kirkham MB, Gill BS (2012) Effects of drought and high temperature stress on synthetic hexaploid wheat. Functional Plant Biology 39:190-198

Quadrana L, Colot V (2016) Plant transgenerational epigenetics. Annual Review of Genetics 50:467-491

R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rafiei N, Ronceret A (2023) Crossover interference mechanism: New lessons from plants. Frontiers in Cell and Developmental Biology 11:1156766

Ramírez-González RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F, Pasha A, Khedikar Y, Robinson SJ, Cory AT, Florio T, Concia L, Juery C, Schoonbeek H, Steuernagel B, Xiang D, Ridout CJ, Chalhoub B, Mayer KFX, Benhamed M, Latrasse D, Bendahmane A, Wulff BBH, Appels R, Tiwari V, Datla R, Choulet F, Pozniak CJ, Provart NJ, Sharpe AG, Paux E, Spannagl M, Bräutigam A, Uauy C (2018) The transcriptional landscape of polyploid wheat. Science 361:eaar6089

Rasheed A, Ogbonnaya FC, Lagudah E, Appels R, He Z (2018) The goat grass genome's role in wheat improvement. Nature Plants 4:56-58

Ren R, Wang H, Guo C, Zhang N, Zeng L, Chen Y, Ma H, Qi J (2018) Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. Molecular Plant 11:414-428

Renny-Byfield S, Chester M, Kovařík A, Le Comber SC, Grandbastien M-A, Deloger M, Nichols RA, Macas J, Novák P, Chase MW, Leitch AR (2011) Next generation sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs. Molecular Biology and Evolution 28:2843-2854

Rey E, Abrouk M, Keeble-Gagnère G, Karafiátová M, Vrána J, Balzergue S, Soubigou-Taconnat L, Brunaud V, Martin-Magniette M-L, Endo TR, Bartoš J, Consortium IWGS, Appels R, Doležel J (2018) Transcriptome reprogramming due to the introduction of a barley telosome into bread wheat affects more barley genes than wheat. Plant Biotechnology Journal 16:1767-1777

Rey M-D, Martín AC, Higgins J, Swarbreck D, Uauy C, Shaw P, Moore G (2017) Exploiting the *ZIP4* homologue within the wheat *Ph1* locus has identified two lines exhibiting homoeologous crossover in wheat-wild relative hybrids. Molecular Breeding 37:95

Reynolds M, Atkin OK, Bennett M, Cooper M, Dodd IC, Foulkes MJ, Frohberg C, Hammer G, Henderson IR, Huang B, Korzun V, McCouch SR, Messina CD, Pogson BJ, Slafer GA, Taylor NL, Wittich PE (2021) Addressing research bottlenecks to crop productivity. Trends in Plant Science 26:607-630

Ricci WA, Lu Z, Ji L, Marand AP, Ethridge CL, Murphy NG, Noshay JM, Galli M, Mejía-Guerra MK, Colomé-Tatché M, Johannes F, Rowley MJ, Corces VG, Zhai J, Scanlon MJ, Buckler ES, Gallavotti A, Springer NM, Schmitz RJ, Zhang X (2019) Widespread long-range cis-regulatory elements in the maize genome. Nature Plants 5:1237-1249

Rice A, Šmarda P, Novosolov M, Drori M, Glick L, Sabath N, Meiri S, Belmaker J, Mayrose I (2019) The global biogeography of polyploid plants. Nature Ecology & Evolution 3:265-273

rmats2sashimiplot (2015) https://github.com/Xinglab/rmats2sashimiplot

Robinson MD, McCarthy DJ, Smyth GK (2009) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-140

Rommel Fuentes R, Hesselink T, Nieuwenhuis R, Bakker L, Schijlen E, van Dooijeweert W, Diaz Trivino S, de Haan JR, Sanchez Perez G, Zhang X, Fransz P, de Jong H, van Dijk ADJ, de Ridder D, Peters SA (2020) Meiotic recombination profiling of interspecific hybrid F1 tomato pollen by linked read sequencing. The Plant Journal 102:480-492

Rowan BA, Heavens D, Feuerborn TR, Tock AJ, Henderson IR, Weigel D (2019) An ultra high-density *Arabidopsis thaliana* crossover map that refines the influences of structural variation and epigenetic features. Genetics 213:771-787

Salamini F, Özkan H, Brandolini A, Schäfer-Pregl R, Martin W (2002) Genetics and geography of wild cereal domestication in the near east. Nature Reviews Genetics 3:429-441

Salman-Minkov A, Sabath N, Mayrose I (2016) Whole-genome duplication as a key factor in crop domestication. Nature Plants 2:16115

Sansaloni C, Franco J, Santos B, Percival-Alwyn L, Singh S, Petroli C, Campos J, Dreher K, Payne T, Marshall D, Kilian B, Milne I, Raubach S, Shaw P, Stephen G, Carling J, Pierre CS, Burgueño J, Crosa J, Li H, Guzman C, Kehel Z, Amri A, Kilian A, Wenzl P, Uauy C, Banziger M, Caccamo M, Pixley K (2020) Diversity analysis of 80,000 wheat accessions reveals consequences and opportunities of selection footprints. Nature Communications 11:4572

Schilling S, Kennedy A, Pan S, Jermiin LS, Melzer R (2019) Genome-wide analysis of MIKC-type MADS-box genes in wheat: pervasive duplications may have facilitated adaptation to different environmental conditions. bioRxiv:585232

Schmitz RJ, Grotewold E, Stam M (2021) Cis-regulatory sequences in plants: Their importance, discovery, and future challenges. The Plant Cell 34:718-741

Schwacha A, Kleckner N (1995) Identification of double holliday junctions as intermediates in meiotic recombination. Cell 83:783-791

Séguéla-Arnaud M, Crismani W, Larchevêque C, Mazel J, Froger N, Choinard S, Lemhemdi A, Macaisne N, Van Leene J, Gevaert K, De Jaeger G, Chelysheva L, Mercier R (2015) Multiple mechanisms limit meiotic crossovers: TOP3α and two BLM homologs antagonize crossovers in parallel to FANCM. Proceedings of the National Academy of Sciences 112:4713-4718

Semagn K, Iqbal M, Alachiotis N, N'Diaye A, Pozniak C, Spaner D (2021) Genetic diversity and selective sweeps in historical and modern Canadian spring wheat cultivars using the 90K SNP array. Scientific Reports 11:23773

Senapati N, Semenov MA, Halford NG, Hawkesford MJ, Asseng S, Cooper M, Ewert F, van Ittersum MK, Martre P, Olesen JE, Reynolds M, Rötter RP, Webber H (2022) Global wheat production could benefit from closing the genetic yield gap. Nature Food 3:532-541

Serra H, Svačina R, Baumann U, Whitford R, Sutton T, Bartoš J, Sourdille P (2021) *Ph2* encodes the mismatch repair protein MSH7-3D that inhibits wheat homoeologous recombination. Nature Communications 12:803

Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA (2001) Sequence Elimination and Cytosine Methylation Are Rapid and Reproducible Responses of the Genome to Wide Hybridization and Allopolyploidy in Wheat. The Plant Cell 13:1749-1759

Shao Y, Pan Q, Zhang D, Kang L, Li Z (2021) Global gene expression perturbations in rapeseed due to the introduction of alien radish chromosomes. Journal of Genetics 100:25

Sharma S, Schulthess AW, Bassi FM, Badaeva ED, Neumann K, Graner A, Özkan H, Werner P, Knüpffer H, Kilian B (2021) Introducing beneficial alleles from plant genetic resources into the wheat germplasm. Biology 10:982

Shen S, Park JW, Lu Z-x, Lin L, Henry MD, Wu YN, Zhou Q, Xing Y (2014) rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proceedings of the National Academy of Sciences 111:E5593-E5601

Shewry PR, Hey SJ (2015) The contribution of wheat to human diet and health. Food and Energy Security 4:178-202

Shilo S, Melamed-Bessudo C, Dorone Y, Barkai N, Levy AA (2015) DNA crossover motifs associated with epigenetic modifications delineate open chromatin regions in *Arabidopsis*. The Plant Cell 27:2427-2436

Shrestha S, Koo D-H, Evers B, Wu S, Walkowiak S, Hucl P, Pozniak C, Fritz A, Poland J (2023) Wheat doubled haploids have a marked prevalence of chromosomal aberrations. The Plant Genome 16:e20309

Shumate A, Salzberg SL (2021) Liftoff: accurate mapping of gene annotations. Bioinformatics 37:1639-1643

Singh P, Ahi EP (2022) The importance of alternative splicing in adaptive evolution. Molecular Ecology 31:1928-1938

Skalická K, Lim KY, Matyášek R, Koukalová B, Leitch AR, Kovařík A (2003) Rapid evolution of parental rDNA in a synthetic tobacco allotetraploid line. American Journal of Botany 90:988-996

Smith RD, Kinser TJ, Smith GDC, Puzey JR (2019) A likelihood ratio test for changes in homeolog expression bias. BMC bioinformatics 20:1-11

Soltis DE, Visger CJ, Soltis PS (2014) The polyploidy revolution then…and now: Stebbins revisited. American Journal of Botany 101:1057-1078

Song L, Lu Y, Zhang J, Pan C, Yang X, Li X, Liu W, Li L (2016) Cytological and molecular analysis of wheat – *Agropyron cristatum* translocation lines with 6P chromosome fragments conferring superior agronomic traits in common wheat. Genome 59:840-850

Song Q, Chen ZJ (2015) Epigenetic and developmental regulation in plant polyploids. Current Opinion in Plant Biology 24:101-109

Song Q, Zhang T, Stelly DM, Chen ZJ (2017) Epigenomic and functional analyses reveal roles of epialleles in the loss of photoperiod sensitivity during domestication of allotetraploid cottons. Genome Biology 18:99

Song X, Meng X, Guo H, Cheng Q, Jing Y, Chen M, Liu G, Wang B, Wang Y, Li J, Yu H (2022) Targeting a gene regulatory element enhances rice grain yield by decoupling panicle number and size. Nature Biotechnology 40:1403-1411

Springer NM, Schmitz RJ (2017) Exploiting induced and natural epigenetic variation for crop improvement. Nature Reviews Genetics 18:563-575

Stacey NJ, Kuromori T, Azumi Y, Roberts G, Breuer C, Wada T, Maxwell A, Roberts K, Sugimoto-Shirasu K (2006) Arabidopsis SPO11-2 functions with SPO11-1 in meiotic recombination. The Plant Journal 48:206-216

Szabo-Hever A, Zhang Q, Friesen TL, Zhong S, Elias EM, Cai X, Jin Y, Faris JD, Chao S, Xu SS (2018) Genetic diversity and resistance to Fusarium head blight in synthetic

hexaploid wheat derived from *Aegilops tauschii* and diverse *Triticum turgidum* subspecies. Frontiers in Plant Science 9:1829

Szadkowski E, Eber F, Huteau V, Lodé M, Coriton O, Jenczewski E, Chèvre AM (2011) Polyploid formation pathways have an impact on genetic rearrangements in resynthesized Brassica napus. New Phytologist 191:884-894

Szadkowski E, Eber F, Huteau V, Lodé M, Huneau C, Belcram H, Coriton O, Manzanares-Dauleux MJ, Delourme R, King GJ, Chalhoub B, Jenczewski E, Chèvre A-M (2010) The first meiosis of resynthesized Brassica napus, a genome blender. New Phytologist 186:102-112

Szymanska-Lejman M, Dziegielewski W, Dluzewska J, Kbiri N, Bieluszewska A, Poethig RS, Ziolkowski PA (2023) The effect of DNA polymorphisms and natural variation on crossover hotspot activity in Arabidopsis hybrids. Nature Communications 14:33

Taagen E, Bogdanove AJ, Sorrells ME (2020) Counting on crossovers: Controlled recombination for plant breeding. Trends in Plant Science 25:455-465

Tadesse W, El-Hanafi S, El-Fakhouri K, Imseg I, Ezzahra Rachdad F, El-Gataa Z, El Bouhssini M (2022) Wheat breeding for Hessian fly resistance at ICARDA. The Crop Journal 10:1528-1535

Talbot S, Ogbonnaya F, Chalmers K, Mather D, Appels R, Eastwood R, Lagudah E, Langridge P, Mackay M, McIntyre L (2008) Is synthetic hexaploid wheat a useful germplasm source for increasing grain size and yield in bread wheat breeding? Proceedings of the 11th International Wheat Genetics Symposium, 24–29 Aug 2008, pp 315-318

Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, Trzaskoma P, Magalska A, Wlodarczyk J, Ruszczycki B (2015) CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. Cell 163:1611-1627

Teng Y, Su M, Liu L, Chen S, Liu X (2023) Creating and De Novo Improvement of New Allopolyploid Crops for Future Agriculture. Critical Reviews in Plant Sciences 42:53-64

Tock AJ, Holland DM, Jiang W, Osman K, Sanchez-Moran E, Higgins JD, Edwards KJ, Uauy C, Franklin FCH, Henderson IR (2021) Crossover-active regions of the wheat genome are distinguished by DMC1, the chromosome axis, H3K27me3, and signatures of adaptation. Genome Research 31:1614-1628

Townley-Smith T, Humphreys D, Czarnecki E, Lukow O, McCallum B, Fetch T, Gilbert J, Menzies J, Brown P (2010) Superb hard red spring wheat. Canadian Journal of Plant Science 90:347-352

UN-DESA (2022) United Nations Department of Economic and Social Affairs, Population Division. World Population Prospects 2022: Summary of Results. UN DESA/POP/2022/TR/NO 3

Underwood CJ, Choi K, Lambing C, Zhao X, Serra H, Borges F, Simorowski J, Ernst E, Jacob Y, Henderson IR (2018) Epigenetic activation of meiotic recombination near *Arabidopsis thaliana* centromeres via loss of H3K9me2 and non-CG DNA methylation. Genome Research 28:519-531

Van de Peer Y, Ashman T-L, Soltis PS, Soltis DE (2021) Polyploidy: an evolutionary and ecological force in stressful times. The Plant Cell 33:11-26

van Dijk M, Morley T, Rau ML, Saghai Y (2021) A meta-analysis of projected global food demand and population at risk of hunger for the period 2010–2050. Nature Food 2:494-501

Vasudevan A, Lévesque-Lemay M, Edwards T, Cloutier S (2023) Global transcriptome analysis of allopolyploidization reveals large-scale repression of the D-subgenome in synthetic hexaploid wheat. Figshare Repository https://doi.org/10.6084/m9.figshare.c.6443621

Veitia RA, Birchler JA (2022) Gene-dosage issues: a recurrent theme in whole genome duplication events. Trends in Genetics 38:1-3

Vergauwen D, De Smet I (2017) From early farmers to Norman Borlaug—the making of modern wheat. Current Biology 27:R858-R862

Waghmare VN, Rong J, Rogers CJ, Bowers JE, Chee PW, Gannaway JR, Katageri I, Paterson AH (2016) Comparative transmission genetics of introgressed chromatin in *Gossypium* (cotton) polyploids. American Journal of Botany 103:719-729

Wan H, Yang F, Li J, Wang Q, Liu Z, Tang Y, Yang W (2023) Genetic improvement and application practices of synthetic hexaploid wheat. Genes 14:283

Wang H, Sun S, Ge W, Zhao L, Hou B, Wang K, Lyu Z, Chen L, Xu S, Guo J, Li M, Su P, Li X, Wang G, Bo C, Fang X, Zhuang W, Cheng X, Wu J, Dong L, Chen W, Li W, Xiao G, Zhao J, Hao Y, Xu Y, Gao Y, Liu W, Liu Y, Yin H, Li J, Li X, Zhao Y, Wang X, Ni F, Ma X, Li A, Xu SS, Bai G, Nevo E, Gao C, Ohm H, Kong L (2020) Horizontal gene transfer of *Fhb7* from fungus underlies *Fusarium* head blight resistance in wheat. Science 368:eaba5435

Wang J, Luo M-C, Chen Z, You FM, Wei Y, Zheng Y, Dvorak J (2013) Aegilops tauschii single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. New Phytologist 198:925-937

Wang JY, Doudna JA (2023) CRISPR technology: A decade of genome editing is only the beginning. Science 379:eadd8643

Wang L, Jia G, Jiang X, Cao S, Chen ZJ, Song Q (2021a) Altered chromatin architecture and gene expression during polyploidization and domestication of soybean. The Plant Cell 33:1430-1446

Wang M, Li Z, Zhang Ye, Zhang Y, Xie Y, Ye L, Zhuang Y, Lin K, Zhao F, Guo J, Teng W, Zhang W, Tong Y, Xue Y, Zhang Y (2021b) An atlas of wheat epigenetic regulatory elements reveals subgenome divergence in the regulation of development and stress responses. The Plant Cell 33:865-881

Wang M, Shilo S, Zhou A, Zelkowski M, Olson MA, Azuri I, Shoshani-Hechel N, Melamed-Bessudo C, Marand AP, Jiang J, Schnable JC, Underwood CJ, Henderson IR, Sun Q, Pillardy J, Kianian PMA, Kianian SF, Chen C, Levy AA, Pawlowski WP (2022a) Machine

learning reveals conserved chromatin patterns determining meiotic recombination sites in plants. bioRxiv:2022.2007.2011.499557

Wang M, Wang P, Liang F, Ye Z, Li J, Shen C, Pei L, Wang F, Hu J, Tu L, Lindsey K, He D, Zhang X (2018a) A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation. New Phytologist 217:163-178

Wang M, Wang P, Lin M, Ye Z, Li G, Tu L, Shen C, Li J, Yang Q, Zhang X (2018b) Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. Nature Plants 4:90-97

Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, Mastrangelo AM, Whan A, Stephen S, Barker G, Wieseke R, Plieske J, Consortium IWGS, Lillemo M, Mather D, Appels R, Dolferus R, Brown-Guedira G, Korol A, Akhunova AR, Feuillet C, Salse J, Morgante M, Pozniak C, Luo M-C, Dvorak J, Morell M, Dubcovsky J, Ganal M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KJ, Hayden M, Akhunov E (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnology Journal 12:787-796

Wang X, Han B, Sun Y, Kang X, Zhang M, Han H, Zhou S, Liu W, Lu Y, Yang X, Li X, Zhang J, Liu X, Li L (2022b) Introgression of chromosome 1P from *Agropyron cristatum* reduces leaf size and plant height to improve the plant architecture of common wheat. Theoretical and Applied Genetics 135:1951-1963

Wang X, Luo G, Yang W, Li Y, Sun J, Zhan K, Liu D, Zhang A (2017) Genetic diversity, population structure and marker-trait associations for agronomic and grain traits in wild diploid wheat *Triticum urartu*. BMC Plant Biology 17:112

Wang Y, Copenhaver GP (2018) Meiotic Recombination: Mixing It Up in Plants. Annual Review of Plant Biology 69:577-609

Wicker T, Gundlach H, Spannagl M, Uauy C, Borrill P, Ramírez-González RH, De Oliveira R, Mayer KFX, Paux E, Choulet F, Consortium IWGS (2018) Impact of transposable elements on genome structure and evolution in bread wheat. Genome Biology 19:103

Wu J, Lin L, Xu M, Chen P, Liu D, Sun Q, Ran L, Wang Y (2018) Homoeolog expression bias and expression level dominance in resynthesized allopolyploid *Brassica napus*. BMC Genomics 19:586

Xiong Z, Gaeta RT, Pires JC (2011) Homoeologous shuffling and chromosome compensation maintain genome balance in resynthesized allopolyploid *Brassica napus*. Proceedings of the National Academy of Sciences 108:7908-7913

Yaakov B, Kashkush K (2011) Massive alterations of the methylation patterns around DNA transposons in the first four generations of a newly formed wheat allohexaploid. Genome 54:42-49

Yang C, Zhao L, Zhang H, Yang Z, Wang H, Wen S, Zhang C, Rustgi S, von Wettstein D, Liu B (2014) Evolution of physiological responses to salt stress in hexaploid wheat. Proceedings of the National Academy of Sciences 111:11882-11887

Yelina NE, Lambing C, Hardcastle TJ, Zhao X, Santos B, Henderson IR (2015) DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in Arabidopsis. Genes & Development 29:2183-2202

Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan X, Zhu M, Zhao S, Li X, Liu X (2021) rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated Tool for Genome-wide Association Study. Genomics, Proteomics & Bioinformatics 19:619-628

Yocca AE, Edger PP (2022) Current status and future perspectives on the evolution of cis-regulatory elements in plants. Current Opinion in Plant Biology 65:102139

Yoo M-J, Liu X, Pires JC, Soltis PS, Soltis DE (2014) Nonadditive gene expression in polyploids. Annual Review of Genetics 48:485-517

Yoo M, Szadkowski E, Wendel J (2013) Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity 110:171-180

Yu K, Feng M, Yang G, Sun L, Qin Z, Cao J, Wen J, Li H, Zhou Y, Chen X, Peng H, Yao Y, Hu Z, Guo W, Sun Q, Ni Z, Adams K, Xin M (2020) Changes in alternative splicing in response to domestication and polyploidization in wheat. Plant Physiology 184:1955-1968

Yu M, Guan L-L, Chen G-Y, Pu Z-E, Hou D-B (2017) Allopolyploidy-induced rapid genomic changes in newly generated synthetic hexaploid wheat. Biotechnology & Biotechnological Equipment 31:236-242

Yu X, Wang P, Li J, Zhao Q, Ji C, Zhu Z, Zhai Y, Qin X, Zhou J, Yu H, Cheng X, Isshiki S, Jahn M, Doyle JJ, Ottosen C-O, Bai Y, Cai Q, Cheng C, Lou Q, Huang S, Chen J (2021) Whole-genome sequence of synthesized allopolyploids in *Cucumis* reveals insights into the genome evolution of allopolyploidization. Advanced Science 8:2004222

Yuan J, Jiao W, Liu Y, Ye W, Wang X, Liu B, Song Q, Chen ZJ (2020) Dynamic and reversible DNA methylation changes induced by genome separation and merger of polyploid wheat. BMC Biology 18:171

Yuan J, Sun H, Wang Y, Li L, Chen S, Jiao W, Jia G, Wang L, Mao J, Ni Z, Wang X, Song Q (2022) Open chromatin interaction maps reveal functional regulatory elements and chromatin architecture variations during wheat evolution. Genome Biology 23:34

Zhang D, Pan Q, Cui C, Tan C, Ge X, Shao Y, Li Z (2015a) Genome-specific differential gene expressions in resynthesized *Brassica* allotetraploids from pair-wise crosses of three cultivated diploids revealed by RNA-seq. Frontiers in Plant Science 6:957

Zhang F, Batley J (2020) Exploring the application of wild species for crop improvement in a changing climate. Current Opinion in Plant Biology 56:218-222

Zhang H, Gou X, Zhang A, Wang X, Zhao N, Dong Y, Li L, Liu B (2016) Transcriptome shock invokes disruption of parental expression-conserved genes in tetraploid wheat. Scientific Reports 6:26363

Zhang H, Zheng R, Wang Y, Zhang Y, Hong P, Fang Y, Li G, Fang Y (2019) The effects of Arabidopsis genome duplication on the chromatin organization and transcriptional regulation. Nucleic Acids Research 47:7857-7869

Zhang J, Liu Y, Xia E-H, Yao Q-Y, Liu X-D, Gao L-Z (2015b) Autotetraploid rice methylome analysis reveals methylation variation of transposable elements and their effects on gene expression. Proceedings of the National Academy of Sciences 112:E7022-E7029

Zhang L-Q, Liu D-C, Zheng Y-L, Yan Z-H, Dai S-F, Li Y-F, Jiang Q, Ye Y-Q, Yen Y (2010) Frequent occurrence of unreduced gametes in *Triticum turgidum–Aegilops tauschii* hybrids. Euphytica 172:285-294

Zhang L, He C, Lai Y, Wang Y, Kang L, Liu A, Lan C, Su H, Gao Y, Li Z, Yang F, Li Q, Mao H, Chen D, Chen W, Kaufmann K, Yan W (2023) Asymmetric gene expression and cell-type-specific regulatory networks in the root of bread wheat revealed by single-cell multiomics analysis. Genome Biology 24:65

Zhang Z, Gou X, Xun H, Bian Y, Ma X, Li J, Li N, Gong L, Feldman M, Liu B, Levy AA (2020) Homoeologous exchanges occur through intragenic recombination generating novel transcripts and proteins in wheat and other polyploids. Proceedings of the National Academy of Sciences 117:14561-14571

Zhang Z, Schwartz S, Wagner L, Miller W (2000) A Greedy Algorithm for Aligning DNA Sequences. Journal of Computational Biology 7:203-214

Zhang Z, Xun H, Lv R, Gou X, Ma X, Li J, Zhao J, Li N, Gong L, Liu B (2022) Effects of homoeologous exchange on gene expression and alternative splicing in a newly formed allotetraploid wheat. The Plant Journal 111:1267-1282

Zhao H, Sun Z, Wang J, Huang H, Kocher J-P, Wang L (2013) CrossMap: a versatile tool for coordinate conversion between genome assemblies. Bioinformatics 30:1006-1007

Zhao L, Xie D, Fan C, Zhang S, Huang L, Ning S, Jiang B, Zhang L, Yuan Z, Liu D, Hao M (2021a) Chromosome stability of synthetic-natural wheat hybrids. Frontiers in Plant Science 12:654382

Zhao L, Yang Y, Chen J, Lin X, Zhang H, Wang H, Wang H, Bie X, Jiang J, Feng X, Fu X, Zhang X, Du Z, Xiao J (2023) Dynamic chromatin regulatory programs during embryogenesis of hexaploid wheat. Genome Biology 24:7

Zhao M, Ku J-C, Liu B, Yang D, Yin L, Ferrell TJ, Stoll CE, Guo W, Zhang X, Wang D, Wang C-JR, Lisch D (2021b) The *mop1* mutation affects the recombination landscape in maize. Proceedings of the National Academy of Sciences 118:e2009475118

Zhao T, Tao X, Feng S, Wang L, Hong H, Ma W, Shang G, Guo S, He Y, Zhou B, Guan X (2018) LncRNAs in polyploid cotton interspecific hybrids are derived from transposon neofunctionalization. Genome Biology 19:195

Zhou R, Moshgabadi N, Adams KL (2011) Extensive changes to alternative splicing patterns following allopolyploidy in natural and resynthesized polyploids. Proceedings of the National Academy of Sciences 108:16122-16127

Zhu T, Wang L, Rimbert H, Rodriguez JC, Deal KR, De Oliveira R, Choulet F, Keeble-Gagnère G, Tibbits J, Rogers J, Eversole K, Appels R, Gu YQ, Mascher M, Dvorak J, Luo M-C (2021) Optical maps refine the bread wheat *Triticum aestivum* cv. Chinese Spring genome assembly. The Plant Journal 107:303-314

Zhuang Y, Wang X, Li X, Hu J, Fan L, Landis JB, Cannon SB, Grimwood J, Schmutz J, Jackson SA, Doyle JJ, Zhang XS, Zhang D, Ma J (2022) Phylogenomics of the genus Glycine sheds light on polyploid evolution and life-strategy transition. Nature Plants 8:233-244

Zou J, Hu D, Mason AS, Shen X, Wang X, Wang N, Grandke F, Wang M, Chang S, Snowdon RJ (2018) Genetic changes in a novel breeding population of *Brassica napus* synthesized from hundreds of crosses between *B. rapa* and *B. carinata.* Plant Biotechnology Journal 16:507-519

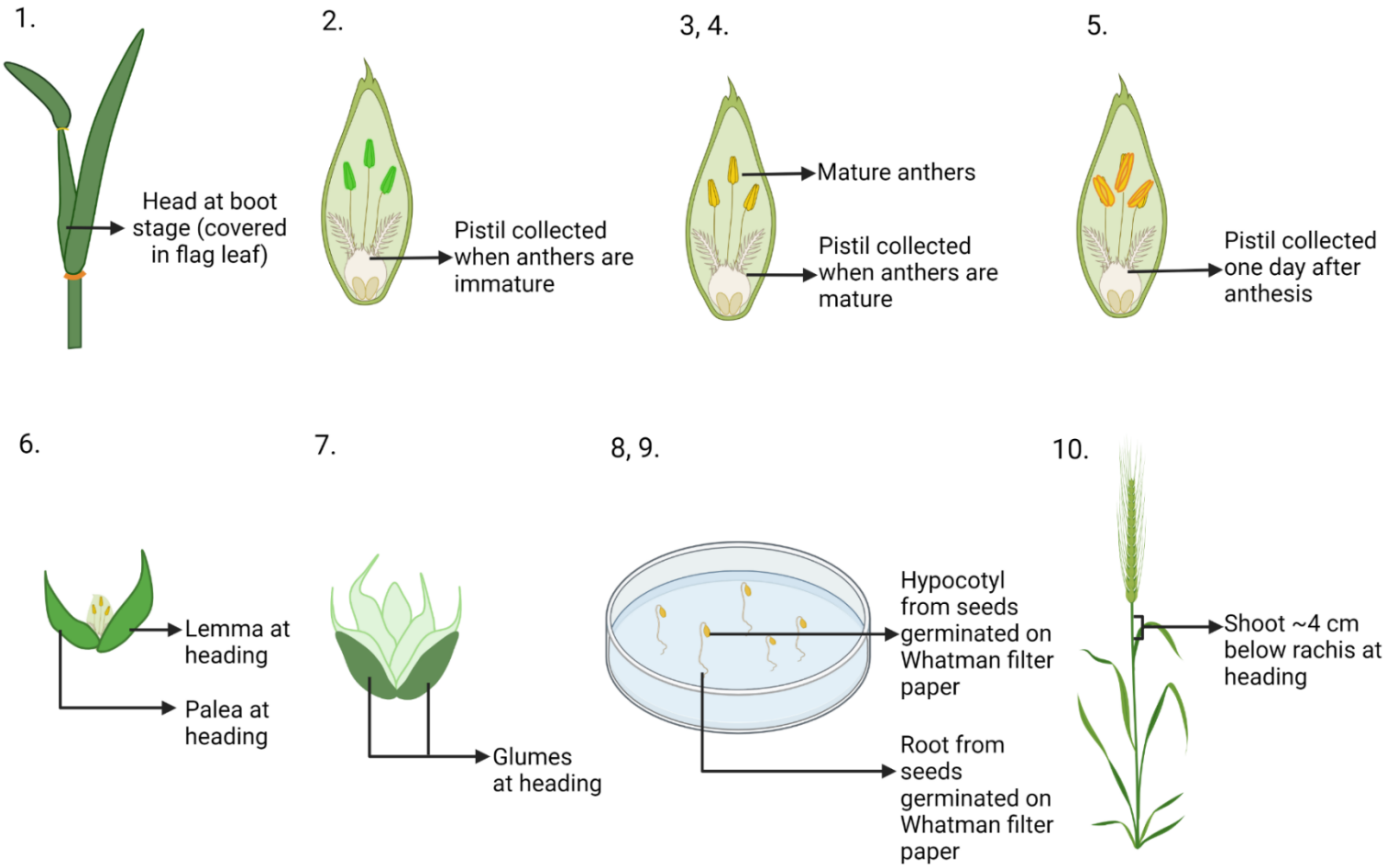
CHAPTER 7

Appendices

7.1 Supplementary materials for Chapter 2

Appendix 1 Synthetic hexaploid lines and their corresponding diploid and tetraploid parents used in the study.

		Male parent	
		AS2386 (DD) - C26 *Ae*. *tauschii*	AS2399 (DD) - C30 *Ae*. *tauschii*
Female parent	PI377655 (AABB) *T. turgidum* ssp. *dicoccum*	C66 (AABBDD)	C65 (AABBDD)
	Langdon (AABB) *T*. *turgidum* ssp. *durum*	C44 (AABBDD)	C45 (AABBDD)



Appendix 2 Pictorial representation of the ten different tissues used in the study. (1) Head at boot stage; (2) Pistil-when anthers are green and immature; (3) Mature anthers; (4) Pistil-when anthers are yellow and just prior to dehiscence.; (5) Pistil-one day after anthesis; (6) Palea+Lemma; (7) Glumes; (8) Hypocotyl; (9) Root; (10) Shoot. Figure created with Biorender.com.

Appendix 3 Number of raw reads in the 240 samples

		Diploid - AS2386	Diploid - AS2399	SHW - C44	SHW - C45	SHW - C65	SHW - C66	Tetraploid - Langdon	Tetraploid - PI377655
Pistil-1DAA	Rep1	22,195,307	21,995,116	33,587,073	26,493,898	38,641,474	26,548,645	28,846,234	28,035,524
	Rep2	22,995,789	34,757,108	36,490,086	33,720,903	39,157,764	23,082,570	44,131,554	22,652,205
	Rep3	24,015,698	22,989,427	91,812,896	23,633,788	27,469,412	31,711,344	26,122,325	22,785,314
Boot	Rep1	76,018,937	32,301,846	28,027,166	38,784,598	18,707,397	38,612,808	33,243,732	24,941,487
	Rep2	26,200,871	73,864,787	21,492,555	26,939,056	61,995,968	20,344,851	45,296,952	25,427,653
	Rep3	29,151,714	68,049,299	28,500,457	85,145,293	69,852,158	22,442,149	27,178,086	23,280,574
Glume	Rep1	35,420,234	29,310,601	20,759,066	28,314,129	47,565,899	28,268,183	38,976,192	64,828,969
	Rep2	21,937,111	23,713,143	22,781,410	47,640,557	20,144,197	44,869,236	27,482,753	22,573,351
	Rep3	29,713,083	21,744,437	23,737,425	26,490,383	21,575,767	31,696,463	40,205,086	23,904,763
Hypocotyl	Rep1	30,449,989	20,003,076	13,082,000	48,991,348	18,179,165	18,007,903	28,016,934	38,735,285
	Rep2	22,841,340	31,214,263	23,833,152	51,412,571	27,737,105	29,903,001	42,518,094	27,349,705
	Rep3	63,882,306	46,891,825	50,655,072	28,343,485	51,303,621	22,989,560	39,910,639	40,895,942
Palea+Lemma	Rep1	24,328,003	32,518,917	24,198,386	42,344,790	20,249,452	23,162,717	64,847,457	33,799,385
	Rep2	26,014,180	21,748,599	22,482,334	30,774,636	25,410,801	24,318,988	25,346,626	20,559,596
	Rep3	27,805,863	21,900,549	21,547,690	41,084,241	29,369,761	23,520,086	22,848,870	40,133,310
Pistil-AM	Rep1	26,023,695	28,445,422	56,573,952	27,251,070	24,266,368	24,290,980	53,144,692	24,637,191
	Rep2	23,690,579	20,529,663	12,878,054	23,573,721	26,393,617	24,437,509	30,271,276	23,713,978
	Rep3	20,806,267	27,419,849	25,902,215	24,955,332	27,162,721	23,635,128	38,392,332	32,237,698
Root	Rep1	37,598,774	28,541,828	33,567,318	26,045,053	25,478,138	19,511,358	34,691,430	26,340,471
	Rep2	29,137,833	25,129,915	23,934,237	35,683,705	29,225,962	21,364,532	21,883,804	24,330,645
	Rep3	47,832,348	25,090,283	22,864,680	30,691,902	21,624,952	26,808,534	25,119,702	28,174,617
Shoot	Rep1	27,411,093	23,796,164	30,077,914	25,784,399	27,149,920	35,503,678	42,150,240	23,856,393
	Rep2	29,526,120	23,787,924	23,346,192	25,126,541	40,080,997	19,545,212	54,426,481	25,778,462
	Rep3	27,553,375	12,778,098	53,706,831	28,990,179	48,394,047	34,000,995	36,586,301	24,547,842
Anther	Rep1	14,858,969	12,068,339	13,332,346	12,464,913	15,245,079	14,566,281	14,138,343	12,871,028
	Rep2	13,950,637	14,561,991	12,961,043	13,591,105	11,329,218	14,192,473	14,466,506	12,874,863
	Rep3	13,939,729	13,903,778	12,948,472	14,120,579	13,776,126	13,739,453	12,189,781	14,738,894
Pistil-AI	Rep1	13,188,605	13,449,612	13,036,625	12,154,908	13,265,009	13,370,423	13,956,148	13,240,572
	Rep2	12,789,103	13,009,865	12,653,047	12,378,022	12,666,519	13,168,331	13,237,777	15,109,353
	Rep3	11,175,795	14,047,307	13,584,399	13,336,772	12,082,013	6,573,063	13,008,756	11,463,936

Pistil - 1DAA: Pistil - one day after anthesis; Pistil-AM: pistil-when anthers are at mature stage; Pistil-AI: pistil-when anthers are at immature stage; Boot: Head at boot stage

Appendix 4 Number of reads in the 240 samples, after processing

		Diploid - AS2386	Diploid - AS2399	SHW - C44	SHW - C45	SHW - C65	SHW - C66	Tetraploid - Langdon	Tetraploid - PI377655
Pistil-1DAA	**Rep1**	16,652,462	15,620,498	23,790,476	18,114,304	26,722,162	18,910,673	21,160,261	20,875,899
	Rep2	15,274,606	24,308,623	27,552,871	22,883,456	26,607,670	16,101,400	32,282,843	16,014,445
	Rep3	17,596,827	16,943,904	70,743,378	16,474,019	19,708,725	21,376,446	18,355,411	15,236,838
Boot	**Rep1**	57,664,218	24,949,933	21,275,022	29,714,612	13,089,654	30,169,811	23,571,404	17,661,315
	Rep2	19,408,624	57,391,546	15,926,482	20,617,391	47,873,378	14,793,083	34,287,497	18,371,225
	Rep3	20,664,589	53,646,495	20,195,131	64,951,442	49,579,406	15,987,495	19,319,412	17,171,614
Glume	**Rep1**	26,815,121	20,940,544	15,246,336	19,648,878	32,319,338	21,431,871	29,463,422	48,193,103
	Rep2	16,181,955	17,665,911	17,013,063	34,431,917	14,395,762	34,300,068	19,643,543	16,772,106
	Rep3	23,384,600	15,494,944	16,957,425	18,430,477	15,674,851	22,514,609	28,535,092	18,511,746
Hypocotyl	**Rep1**	22,561,521	14,100,352	10,041,404	37,561,278	13,827,431	13,993,188	19,525,603	27,837,920
	Rep2	16,526,306	22,602,975	17,221,309	40,381,040	20,276,202	21,450,701	30,098,228	20,226,843
	Rep3	48,387,770	34,862,966	36,242,751	21,403,064	35,854,302	16,338,510	28,766,799	29,606,341
Palea+Lemma	**Rep1**	18,391,019	25,039,738	17,712,620	29,284,744	14,637,765	17,059,434	48,637,322	24,787,679
	Rep2	20,019,127	15,967,411	16,328,130	21,710,695	18,812,481	17,763,568	17,905,772	14,222,274
	Rep3	19,870,649	16,053,756	16,023,215	32,167,434	21,649,321	17,532,475	17,107,285	28,181,077
Pistil-AM	**Rep1**	18,349,055	20,387,882	39,204,701	19,597,028	17,808,171	18,540,363	39,515,015	17,642,079
	Rep2	17,009,153	14,338,346	9,774,662	17,003,537	18,099,276	16,707,667	20,147,366	16,541,649
	Rep3	14,373,624	19,829,020	19,493,466	18,377,313	19,459,379	17,142,187	28,539,125	23,726,105
Root	**Rep1**	25,851,418	21,188,494	24,812,414	18,371,284	19,203,581	13,541,267	26,718,856	19,765,758
	Rep2	21,390,326	18,620,533	17,297,014	26,998,387	22,692,101	15,503,742	15,431,234	17,525,916
	Rep3	33,999,451	17,916,134	15,909,748	21,996,230	15,535,210	19,776,405	19,374,245	19,934,284
Shoot	**Rep1**	19,523,100	18,255,513	22,491,360	18,934,613	19,607,143	27,386,269	30,091,216	16,814,046
	Rep2	22,614,844	17,088,763	17,308,713	17,897,686	28,082,616	14,874,879	38,629,216	18,534,622
	Rep3	21,063,835	10,026,717	38,010,200	21,020,090	34,583,788	24,084,686	27,829,956	17,347,074
Anther	**Rep1**	10,862,366	8,457,261	10,832,242	9,653,328	12,421,104	10,966,326	10,320,422	10,231,104
	Rep2	11,112,928	11,739,559	9,546,590	10,147,261	9,136,176	11,468,430	11,543,432	10,230,761
	Rep3	10,378,299	9,800,090	9,542,973	11,388,733	9,912,320	9,921,095	9,810,303	11,785,351
Pistil-AI	**Rep1**	9,854,619	11,066,191	10,647,010	9,158,324	10,770,855	9,972,559	10,392,686	10,512,209
	Rep2	9,304,179	9,562,473	9,243,986	10,093,944	10,334,726	9,781,472	10,698,688	11,338,157
	Rep3	8,967,274	10,234,854	11,120,788	9,712,338	9,793,567	4,829,128	9,814,128	9,472,780

Pistil-1DAA: pistil-one day after anthesis; Pistil-AM: pistil-when anthers are at mature stage; Pistil-AI: pistil-when anthers are at immature stage; Boot: Head at boot stage

Appendix 5 Summary of proportion of uniquely mapped and multi-mapped reads

Genotype	Source material	Tissue	Rep	Uniquely mapped reads (%)	Multi-mapped (%)	Too many multi-mapped* (%)	Total multi-mapped$ (%)
AS2386	Diploid parent	Anther	1	83.07	13.69	1.22	14.91
AS2386	Diploid parent	Anther	2	84.01	13.21	1.19	14.40
AS2386	Diploid parent	Anther	3	82.94	13.49	1.32	14.81
AS2386	Diploid parent	Glume	1	84.42	12.72	0.53	13.25
AS2386	Diploid parent	Glume	3	83.45	13.74	0.50	14.24
AS2386	Diploid parent	Glume	2	85.43	11.07	1.15	12.22
AS2386	Diploid parent	Boot	3	88.18	8.95	0.36	9.31
AS2386	Diploid parent	Boot	2	87.34	9.74	0.75	10.49
AS2386	Diploid parent	Boot	1	88.09	9.32	0.38	9.70
AS2386	Diploid parent	Hypocotyl	1	85.23	11.11	1.21	12.32
AS2386	Diploid parent	Hypocotyl	3	87.49	9.46	0.58	10.04
AS2386	Diploid parent	Hypocotyl	2	85.63	10.90	0.95	11.85
AS2386	Diploid parent	Palea+Lemma	3	85.55	10.33	0.98	11.31
AS2386	Diploid parent	Palea+Lemma	1	86.50	10.11	0.90	11.01
AS2386	Diploid parent	Palea+Lemma	2	86.34	10.64	0.67	11.31
AS2386	Diploid parent	Pistil-1DAA	3	87.59	7.47	2.24	9.71
AS2386	Diploid parent	Pistil-1DAA	2	85.98	7.95	3.05	11.00
AS2386	Diploid parent	Pistil-1DAA	1	89.04	7.23	1.33	8.56
AS2386	Diploid parent	Pistil-AM	3	87.70	7.84	1.83	9.67
AS2386	Diploid parent	Pistil-AM	1	86.76	7.94	2.51	10.45
AS2386	Diploid parent	Pistil-AM	2	87.66	7.85	1.66	9.51
AS2386	Diploid parent	Pistil-AI	3	83.79	13.42	0.54	13.96
AS2386	Diploid parent	Pistil-AI	2	83.12	13.82	0.61	14.43
AS2386	Diploid parent	Pistil-AI	1	82.82	13.86	0.61	14.47
AS2386	Diploid parent	Root	5	88.36	6.68	2.19	8.87
AS2386	Diploid parent	Root	4	88.78	6.64	1.99	8.63
AS2386	Diploid parent	Root	2	89.52	6.83	0.93	7.76
AS2386	Diploid parent	Shoot	2	89.12	8.15	0.31	8.46
AS2386	Diploid parent	Shoot	1	86.94	9.44	0.88	10.32
AS2386	Diploid parent	Shoot	3	86.90	10.25	0.51	10.76
AS2399	Diploid parent	Anther	4	83.23	14.42	0.57	14.99
AS2399	Diploid parent	Anther	2	81.54	14.89	0.93	15.82
AS2399	Diploid parent	Anther	3	80.38	13.59	1.65	15.24
AS2399	Diploid parent	Glume	3	83.82	13.01	0.60	13.61
AS2399	Diploid parent	Glume	2	85.70	10.47	1.05	11.52
AS2399	Diploid parent	Glume	1	83.42	13.20	0.74	13.94

Genotype	Source material	Tissue	Rep	Uniquely mapped reads (%)	Multi-mapped (%)	Too many multi-mapped[*] (%)	Total multi-mapped[$] (%)
AS2399	Diploid parent	Boot	3	88.17	9.14	0.39	9.53
AS2399	Diploid parent	Boot	1	87.08	9.79	0.93	10.72
AS2399	Diploid parent	Boot	2	86.15	11.28	0.34	11.62
AS2399	Diploid parent	Hypocotyl	1	83.53	12.81	1.17	13.98
AS2399	Diploid parent	Hypocotyl	3	85.57	11.38	0.67	12.05
AS2399	Diploid parent	Hypocotyl	2	83.78	12.60	0.97	13.57
AS2399	Diploid parent	Palea+Lemma	1	85.94	11.05	0.37	11.42
AS2399	Diploid parent	Palea+Lemma	2	86.17	10.13	0.81	10.94
AS2399	Diploid parent	Palea+Lemma	3	86.84	9.74	0.71	10.45
AS2399	Diploid parent	Pistil-1DAA	1	86.10	7.47	3.64	11.11
AS2399	Diploid parent	Pistil-1DAA	3	88.03	7.38	2.18	9.56
AS2399	Diploid parent	Pistil-1DAA	2	85.31	7.91	3.71	11.62
AS2399	Diploid parent	Pistil-AM	2	84.95	8.02	4.04	12.06
AS2399	Diploid parent	Pistil-AM	1	85.55	7.82	3.55	11.37
AS2399	Diploid parent	Pistil-AM	3	86.51	7.62	3.24	10.86
AS2399	Diploid parent	Pistil-AI	2	82.28	13.75	0.89	14.64
AS2399	Diploid parent	Pistil-AI	1	83.05	13.89	0.52	14.41
AS2399	Diploid parent	Pistil-AI	3	83.54	13.72	0.50	14.22
AS2399	Diploid parent	Root	1	88.70	6.39	2.29	8.68
AS2399	Diploid parent	Root	4	89.31	7.42	0.70	8.12
AS2399	Diploid parent	Root	3	88.88	6.47	2.20	8.67
AS2399	Diploid parent	Shoot	2	84.25	12.07	0.78	12.85
AS2399	Diploid parent	Shoot	1	83.29	13.83	0.56	14.39
AS2399	Diploid parent	Shoot	3	87.57	9.57	0.53	10.10
C44	SHW	Anther	1	82.15	13.41	0.83	14.24
C44	SHW	Anther	6	81.42	13.70	0.90	14.60
C44	SHW	Anther	8	82.04	13.12	1.31	14.43
C44	SHW	Glume	3	84.89	11.03	0.96	11.99
C44	SHW	Glume	2	85.48	10.13	1.10	11.23
C44	SHW	Glume	1	84.42	11.73	0.94	12.67
C44	SHW	Boot	1	88.37	8.29	0.53	8.82
C44	SHW	Boot	2	88.02	8.53	0.79	9.32
C44	SHW	Boot	3	87.33	9.08	0.67	9.75
C44	SHW	Hypocotyl	1	85.67	10.72	0.68	11.40
C44	SHW	Hypocotyl	3	85.35	10.25	1.24	11.49
C44	SHW	Hypocotyl	2	86.52	9.25	1.12	10.37
C44	SHW	Palea+Lemma	3	85.77	10.87	0.61	11.48
C44	SHW	Palea+Lemma	1	85.34	10.63	0.77	11.40

Genotype	Source material	Tissue	Rep	Uniquely mapped reads (%)	Multi-mapped (%)	Too many multi-mapped* (%)	Total multi-mapped$ (%)
C44	SHW	Palea+Lemma	2	85.40	11.01	0.58	11.59
C44	SHW	Pistil-1DAA	2	88.49	6.82	1.83	8.65
C44	SHW	Pistil-1DAA	1	87.50	7.20	2.17	9.37
C44	SHW	Pistil-1DAA	3	89.60	6.85	0.82	7.67
C44	SHW	Pistil-AM	1	86.18	7.82	2.70	10.52
C44	SHW	Pistil-AM	2	89.61	7.02	0.61	7.63
C44	SHW	Pistil-AM	3	86.58	7.16	3.25	10.41
C44	SHW	Pistil-AI	4	83.76	13.04	0.63	13.67
C44	SHW	Pistil-AI	3	83.18	13.26	0.70	13.96
C44	SHW	Pistil-AI	2	83.64	13.00	0.63	13.63
C44	SHW	Root	1	88.52	7.28	0.66	7.94
C44	SHW	Root	3	86.41	7.23	2.84	10.07
C44	SHW	Root	2	88.69	7.08	0.96	8.04
C44	SHW	Shoot	2	86.75	9.31	0.89	10.20
C44	SHW	Shoot	3	85.68	10.24	0.89	11.13
C44	SHW	Shoot	1	84.31	12.06	1.02	13.08
C45	SHW	Anther	2	81.86	13.66	0.75	14.41
C45	SHW	Anther	5	83.46	12.90	0.61	13.51
C45	SHW	Anther	1	82.71	13.29	0.83	14.12
C45	SHW	Glume	3	78.84	17.61	0.63	18.24
C45	SHW	Glume	1	83.85	12.27	0.64	12.91
C45	SHW	Glume	2	84.55	11.29	0.88	12.17
C45	SHW	Boot	2	86.66	10.34	0.44	10.78
C45	SHW	Boot	3	87.99	9.03	0.38	9.41
C45	SHW	Boot	1	85.51	11.40	0.45	11.85
C45	SHW	Hypocotyl	1	87.90	8.51	0.59	9.10
C45	SHW	Hypocotyl	2	88.68	7.93	0.41	8.34
C45	SHW	Hypocotyl	3	86.47	10.05	0.55	10.60
C45	SHW	Palea+Lemma	3	84.51	11.99	0.82	12.81
C45	SHW	Palea+Lemma	2	84.89	10.65	0.93	11.58
C45	SHW	Palea+Lemma	1	83.48	12.61	0.72	13.33
C45	SHW	Pistil-1DAA	3	87.12	7.54	2.15	9.69
C45	SHW	Pistil-1DAA	2	84.16	7.72	4.57	12.29
C45	SHW	Pistil-1DAA	1	85.30	7.78	3.53	11.31
C45	SHW	Pistil-AM	2	83.85	7.56	5.34	12.90
C45	SHW	Pistil-AM	1	84.61	7.65	4.52	12.17
C45	SHW	Pistil-AM	3	88.15	7.36	1.66	9.02
C45	SHW	Pistil-AI	3	83.19	13.41	0.55	13.96

Genotype	Source material	Tissue	Rep	Uniquely mapped reads (%)	Multi-mapped (%)	Too many multi-mapped[*] (%)	Total multi-mapped[$] (%)
C45	SHW	Pistil-AI	2	83.13	13.24	0.67	13.91
C45	SHW	Pistil-AI	1	83.14	13.14	0.75	13.89
C45	SHW	Root	3	88.71	7.37	0.98	8.35
C45	SHW	Root	2	88.25	6.91	1.04	7.95
C45	SHW	Root	1	88.48	7.43	0.94	8.37
C45	SHW	Shoot	2	85.61	10.46	0.82	11.28
C45	SHW	Shoot	1	84.08	11.55	1.30	12.85
C45	SHW	Shoot	3	86.77	9.35	0.81	10.16
C65	SHW	Anther	4	84.09	13.10	0.34	13.44
C65	SHW	Anther	1	84.15	13.09	0.41	13.50
C65	SHW	Anther	3	83.91	12.89	0.46	13.35
C65	SHW	Glume	2	83.86	12.11	1.00	13.11
C65	SHW	Glume	3	82.30	13.63	1.03	14.66
C65	SHW	Glume	1	83.43	11.71	1.12	12.83
C65	SHW	Boot	1	87.09	9.56	0.42	9.98
C65	SHW	Boot	2	86.95	9.86	0.59	10.45
C65	SHW	Boot	3	85.03	11.52	0.68	12.20
C65	SHW	Hypocotyl	2	84.61	11.08	1.53	12.61
C65	SHW	Hypocotyl	1	85.58	11.20	0.65	11.85
C65	SHW	Hypocotyl	3	85.84	9.95	1.13	11.08
C65	SHW	Palea+Lemma	2	86.60	9.52	1.10	10.62
C65	SHW	Palea+Lemma	1	84.49	11.18	1.18	12.36
C65	SHW	Palea+Lemma	3	83.38	12.70	1.11	13.81
C65	SHW	Pistil-1DAA	1	85.14	7.65	3.90	11.55
C65	SHW	Pistil-1DAA	2	86.35	7.68	2.74	10.42
C65	SHW	Pistil-1DAA	3	88.30	7.43	1.53	8.96
C65	SHW	Pistil-AM	2	87.42	7.94	1.65	9.59
C65	SHW	Pistil-AM	1	84.54	7.55	4.73	12.28
C65	SHW	Pistil-AM	3	82.00	7.80	6.72	14.52
C65	SHW	Pistil-AI	3	83.41	13.12	0.58	13.70
C65	SHW	Pistil-AI	1	82.74	13.60	0.66	14.26
C65	SHW	Pistil-AI	2	82.81	13.24	0.73	13.97
C65	SHW	Root	2	90.02	6.63	0.87	7.50
C65	SHW	Root	3	87.81	7.23	2.04	9.27
C65	SHW	Root	1	89.31	6.99	1.13	8.12
C65	SHW	Shoot	1	78.86	17.40	1.12	18.52
C65	SHW	Shoot	2	82.96	12.59	1.33	13.92
C65	SHW	Shoot	3	84.72	11.39	0.88	12.27

Genotype	Source material	Tissue	Rep	Uniquely mapped reads (%)	Multi-mapped (%)	Too many multi-mapped[*] (%)	Total multi-mapped[$] (%)
C66	SHW	Anther	1	83.57	13.15	0.43	13.58
C66	SHW	Anther	2	83.69	13.32	0.40	13.72
C66	SHW	Anther	4	83.65	12.96	0.53	13.49
C66	SHW	Glume	3	78.74	17.81	0.70	18.51
C66	SHW	Glume	2	86.57	10.20	0.36	10.56
C66	SHW	Glume	1	85.17	11.16	0.76	11.92
C66	SHW	Boot	2	86.07	10.54	0.64	11.18
C66	SHW	Boot	1	86.61	10.58	0.37	10.95
C66	SHW	Boot	3	85.73	10.93	0.58	11.51
C66	SHW	Hypocotyl	2	83.79	12.08	1.29	13.37
C66	SHW	Hypocotyl	1	81.67	14.51	0.74	15.25
C66	SHW	Hypocotyl	3	81.80	13.30	1.86	15.16
C66	SHW	Palea+Lemma	2	84.62	11.27	1.00	12.27
C66	SHW	Palea+Lemma	3	84.10	12.39	0.82	13.21
C66	SHW	Palea+Lemma	1	83.95	12.33	0.87	13.20
C66	SHW	Pistil-1DAA	1	84.88	7.46	4.28	11.74
C66	SHW	Pistil-1DAA	2	86.41	7.71	2.82	10.53
C66	SHW	Pistil-1DAA	3	84.97	7.83	3.82	11.65
C66	SHW	Pistil-AM	3	87.19	7.54	2.42	9.96
C66	SHW	Pistil-AM	2	84.45	7.91	4.20	12.11
C66	SHW	Pistil-AM	1	87.26	7.42	2.51	9.93
C66	SHW	Pistil-AI	1	83.61	13.13	0.66	13.79
C66	SHW	Pistil-AI	3	82.76	13.70	0.62	14.32
C66	SHW	Pistil-AI	2	82.43	13.73	0.69	14.42
C66	SHW	Root	1	86.17	7.16	1.01	8.17
C66	SHW	Root	3	88.51	6.99	1.62	8.61
C66	SHW	Root	2	88.62	6.77	1.76	8.53
C66	SHW	Shoot	3	84.61	11.53	0.73	12.26
C66	SHW	Shoot	1	86.57	9.85	0.46	10.31
C66	SHW	Shoot	2	83.72	12.81	0.83	13.64
Langdon	Tetraploid parent	Glume	1	84.25	11.86	0.44	12.30
Langdon	Tetraploid parent	Glume	2	83.44	11.69	0.81	12.50
Langdon	Tetraploid parent	Glume	3	82.89	12.83	0.71	13.54
Langdon	Tetraploid parent	Boot	2	84.90	12.05	0.32	12.37
Langdon	Tetraploid parent	Boot	1	85.75	10.48	0.55	11.03
Langdon	Tetraploid parent	Boot	3	85.56	10.63	0.68	11.31
Langdon	Tetraploid parent	Hypocotyl	2	83.77	11.69	1.18	12.87
Langdon	Tetraploid parent	Hypocotyl	3	85.22	11.13	0.58	11.71

Genotype	Source material	Tissue	Rep	Uniquely mapped reads (%)	Multi-mapped (%)	Too many multi-mapped[*] (%)	Total multi-mapped[$] (%)
Langdon	Tetraploid parent	Hypocotyl	1	86.27	9.27	1.15	10.42
Langdon	Tetraploid parent	Palea+Lemma	1	84.15	12.03	0.48	12.51
Langdon	Tetraploid parent	Palea+Lemma	3	86.68	9.32	0.94	10.26
Langdon	Tetraploid parent	Palea+Lemma	2	84.25	11.48	0.75	12.23
Langdon	Tetraploid parent	Pistil-1DAA	2	87.76	7.15	2.01	9.16
Langdon	Tetraploid parent	Pistil-1DAA	1	87.10	7.13	2.59	9.72
Langdon	Tetraploid parent	Pistil-1DAA	3	85.99	7.57	2.76	10.33
Langdon	Tetraploid parent	Pistil-AM	2	85.70	7.61	3.21	10.82
Langdon	Tetraploid parent	Pistil-AM	1	88.66	7.02	1.21	8.23
Langdon	Tetraploid parent	Pistil-AM	3	89.36	7.08	0.58	7.66
Langdon	Tetraploid parent	Root	3	89.37	6.96	0.74	7.70
Langdon	Tetraploid parent	Root	1	89.54	7.11	0.50	7.61
Langdon	Tetraploid parent	Root	2	88.60	7.46	0.68	8.14
Langdon	Tetraploid parent	Shoot	2	80.25	15.64	0.99	16.63
Langdon	Tetraploid parent	Shoot	1	82.92	11.06	1.57	12.63
Langdon	Tetraploid parent	Shoot	3	84.79	11.67	0.53	12.20
Langdon	Tetraploid parent	Anther	5	82.93	13.31	0.48	13.79
Langdon	Tetraploid parent	Anther	2	82.21	13.42	0.56	13.98
Langdon	Tetraploid parent	Anther	3	82.88	13.22	0.48	13.70
Langdon	Tetraploid parent	Pistil-AI	1	81.58	12.61	1.28	13.89
Langdon	Tetraploid parent	Pistil-AI	2	82.75	12.36	1.07	13.43
Langdon	Tetraploid parent	Pistil-AI	4	83.08	12.35	0.96	13.31
PI377655	Tetraploid parent	Glume	3	85.34	11.69	0.32	12.01
PI377655	Tetraploid parent	Glume	1	86.14	10.32	0.49	10.81
PI377655	Tetraploid parent	Glume	2	86.42	9.48	1.11	10.59
PI377655	Tetraploid parent	Boot	1	83.87	11.71	0.45	12.16
PI377655	Tetraploid parent	Boot	3	86.20	9.42	1.59	11.01
PI377655	Tetraploid parent	Boot	2	82.08	14.53	0.53	15.06
PI377655	Tetraploid parent	Hypocotyl	3	85.68	9.90	1.09	10.99
PI377655	Tetraploid parent	Hypocotyl	1	85.28	10.56	1.20	11.76
PI377655	Tetraploid parent	Hypocotyl	2	83.98	12.25	1.03	13.28
PI377655	Tetraploid parent	Palea+Lemma	1	82.16	14.07	0.88	14.95
PI377655	Tetraploid parent	Palea+Lemma	3	83.28	13.25	0.67	13.92
PI377655	Tetraploid parent	Palea+Lemma	2	83.56	12.51	0.95	13.46
PI377655	Tetraploid parent	Pistil-1DAA	1	88.36	7.28	1.43	8.71
PI377655	Tetraploid parent	Pistil-1DAA	3	85.35	7.71	3.20	10.91
PI377655	Tetraploid parent	Pistil-1DAA	2	80.77	7.50	7.75	15.25
PI377655	Tetraploid parent	Pistil-AM	3	89.17	7.32	0.61	7.93

Genotype	Source material	Tissue	Rep	Uniquely mapped reads (%)	Multi-mapped (%)	Too many multi-mapped* (%)	Total multi-mapped$ (%)
PI377655	Tetraploid parent	Pistil-AM	1	84.80	7.72	3.87	11.59
PI377655	Tetraploid parent	Pistil-AM	2	86.22	7.93	2.56	10.49
PI377655	Tetraploid parent	Root	4	88.54	7.42	1.01	8.43
PI377655	Tetraploid parent	Root	1	88.59	7.10	1.01	8.11
PI377655	Tetraploid parent	Root	2	88.70	7.36	0.82	8.18
PI377655	Tetraploid parent	Shoot	1	80.87	14.33	1.23	15.56
PI377655	Tetraploid parent	Shoot	2	84.28	11.30	0.95	12.25
PI377655	Tetraploid parent	Shoot	3	85.13	10.55	1.13	11.68
PI377655	Tetraploid parent	Anther	4	83.37	13.09	0.53	13.62
PI377655	Tetraploid parent	Anther	2	83.73	12.73	0.50	13.23
PI377655	Tetraploid parent	Anther	3	83.55	13.07	0.41	13.48
PI377655	Tetraploid parent	Pistil-AI	2	83.38	12.94	0.55	13.49
PI377655	Tetraploid parent	Pistil-AI	4	83.16	12.42	0.93	13.35
PI377655	Tetraploid parent	Pistil-AI	3	82.77	13.02	0.78	13.80

*Too many multi-mapped reads imply reads mapping to more than 10 loci $Total multi-mapped implies the sum of multi-mapped and too many multi-mapped reads; Pistil-1DAA: pistil-one day after anthesis; Pistil-AM: pistil-when anthers are at mature stage; Pistil-AI: pistil-when anthers are at immature stage; Boot: Head at boot stage.

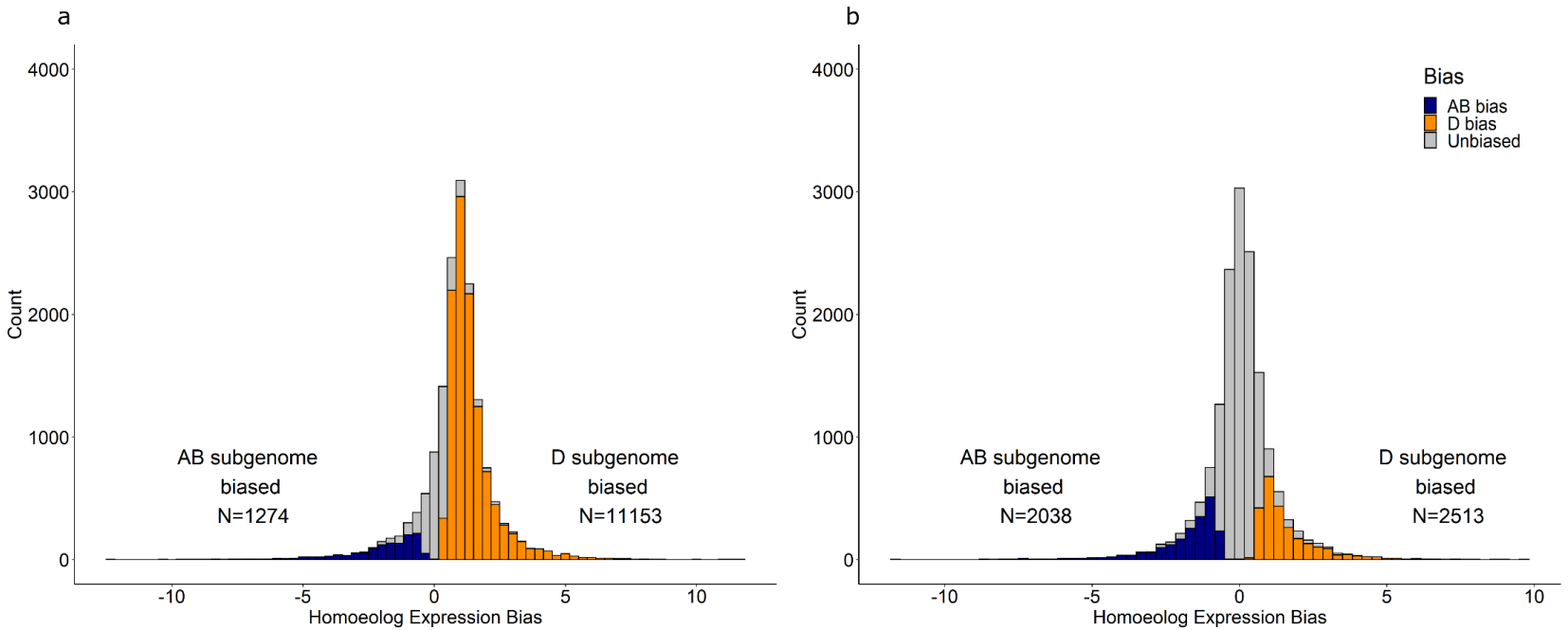
Appendix 6 Number of triads biased towards the AB and D subgenomes in the parental level of expression and the SHW background

Tissues	Bias	C44		C45		C65		C66	
		SHW	Parents	SHW	Parents	SHW	Parents	SHW	Parents
Pistil-1DAA	**AB biased**	6,886 (-2.7577)	1,471 (-1.7018)	3,286 (-1.4999)	1,963 (-1.5472)	2,941 (-1.5634)	1,184 (-2.4306)	2,038 (-1.8335)	1,274 (-2.0341)
	D biased	3,556 (1.1739)	5,626 (2.1351)	2,875 (1.413)	9,476 (1.6018)	2,899 (1.4843)	11,041 (1.401)	2,513 (1.6338)	11,153 (1.4293)
Boot	**AB biased**	5,069 (-2.7606)	700 (-3.2717)	3,666 (-1.2864)	1,369 (-1.9462)	1,957 (-1.8329)	1,760 (-1.2221)	2,478 (-1.4125)	478 (-3.5611)
	D biased	2,992 (1.2135)	6,839 (1.889)	3,790 (1.233)	13,604 (1.3177)	2,163 (1.6844)	13,878 (1.3188)	3,422 (1.3084)	12,944 (1.3496)
Pistil-AM	**AB biased**	3,860 (-3.7535)	2,203 (-1.29)	3,008 (-1.5487)	2,550 (-1.3416)	1,426 (-2.2148)	1,645 (-1.941)	2,809 (-1.4906)	1,496 (-1.7724)
	D biased	1,572 (1.8123)	3,784 (2.2471)	2,617 (1.4366)	5,608 (1.7844)	1,392 (2.1096)	9,476 (1.4874)	3,392 (1.3367)	9,125 (1.5867)
Anther	**AB biased**	3,484 (-2.5037)	865 (-2.3239)	2,064 (-1.8394)	1,000 (-2.1385)	1,927 (-1.9912)	1,092 (-2.2004)	1,841 (-2.0856)	1,115 (-2.1407)
	D biased	1,975 (1.5327)	6,234 (2.4059)	1,730 (1.7007)	5,575 (2.3361)	1,576 (1.971)	4,402 (2.2806)	1,740 (1.9145)	5,667 (2.1912)
Pistil-AI	**AB biased**	5,243 (-2.5754)	1,047 (-1.7311)	2,788 (-1.3753)	920 (-1.9668)	3,343 (-1.171)	988 (-1.9739)	2,367 (-1.3552)	1,099 (-1.7060)
	D biased	2,477 (1.1562)	10,365 (1.5026)	2,396 (1.3264)	10,495 (1.4122)	3,177 (1.1735)	10,086 (1.4445)	2,976 (1.2564)	10,128 (1.5192)
Glume	**AB biased**	4,940 (-2.988)	2,357 (-2.1738)	815 (-3.2174)	2,614 (-1.8089)	3,281 (-1.599)	1,853 (-2.0502)	1,051 (-2.8733)	2,279 (-1.9382)
	D biased	2,244 (1.5222)	7,044 (1.7993)	735 (3.1225)	5,881 (2.0119)	3,104 (1.5195)	3,764 (2.3942)	1,090 (2.7251)	7,264 (1.7225)
Hypocotyl	**AB biased**	3,704 (-4.3743)	471 (-3.1401)	1,632 (-2.153)	1,216 (-2.0073)	2,308 (-1.6095)	1,042 (-2.3845)	1,584 (-1.9159)	512 (-3.3352)
	D biased	640 (2.6087)	6,160 (2.554)	1,723 (1.8929)	7,301 (2.0876)	2,591 (1.5255)	5,907 (1.864)	2,208 (1.7135)	5,597 (2.1047)
Palea+Lemma	**AB biased**	5,166 (-2.9745)	1,310 (-2.3041)	1,565 (-2.2685)	1,985 (-1.8371)	3,423 (-1.4091)	1,188 (-2.2121)	1,348 (-2.2346)	950 (-2.5973)
	D biased	2,330 (1.4397)	6,448 (2.3956)	1,527 (2.1279)	8,586 (2.1801)	3,362 (1.3694)	8,547 (2.1358)	1,633 (2.0951)	7,344 (2.2196)
Root	**AB biased**	4,963 (-3.5897)	1,053 (-2.4667)	2,765 (-1.5269)	824 (-2.5347)	3,293 (-1.3203)	738 (-2.5328)	1,754 (-1.8702)	817 (-2.7019)
	D biased	1,902 (1.5716)	12,714 (1.5124)	2,984 (1.3928)	8,122 (2.0916)	3,674 (1.2486)	9,590 (1.9212)	2,457 (1.6176)	13,208 (1.4689)

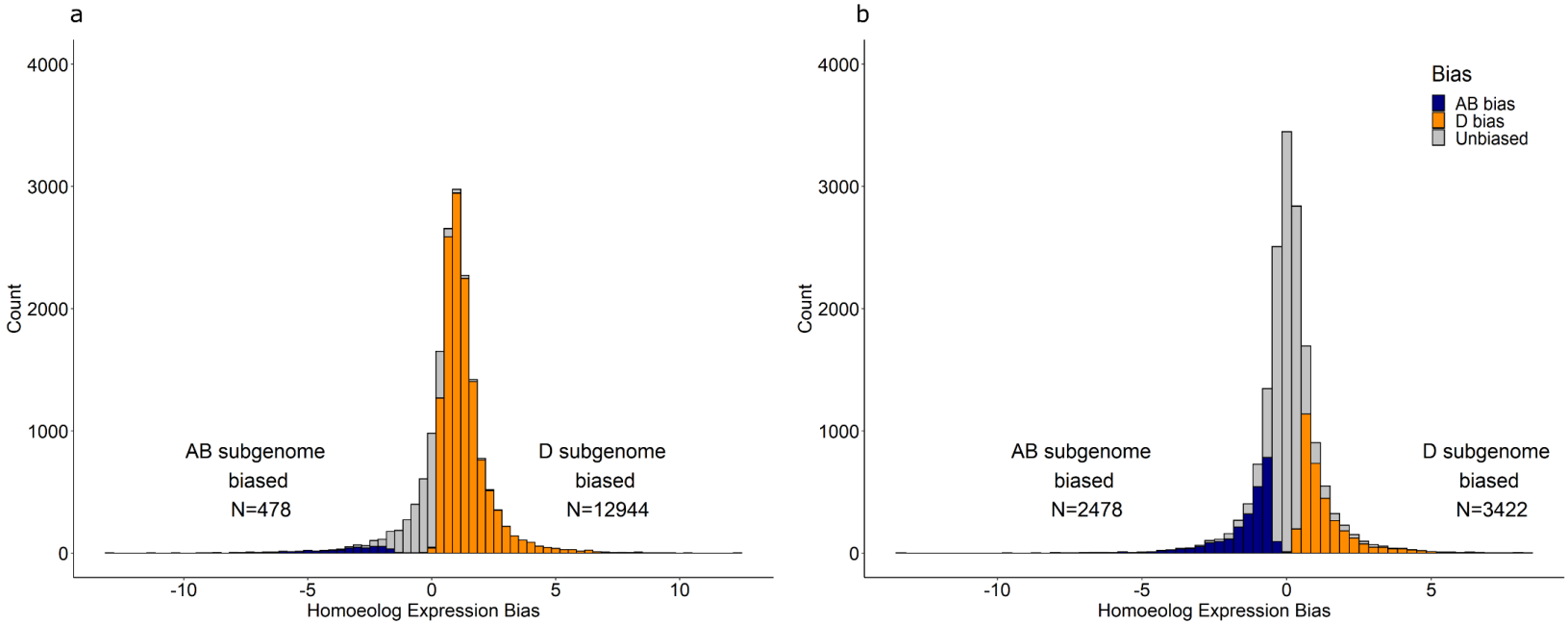
Tissues	Bias	C44		C45		C65		C66	
		SHW	Parents	SHW	Parents	SHW	Parents	SHW	Parents
Shoot	**AB biased**	3,167 (-4.2085)	1,071 (-2.6505)	795 (-3.0994)	1,221 (-2.157)	881 (-3.1496)	1,155 (-2.1834)	793 (-3.1758)	808 (-3.1095)
	D biased	1,376 (2.1535)	3,907 (3.6611)	786 (2.9369)	1,830 (4.6597)	826 (3.0063)	6,387 (2.3872)	890 (2.9798)	9,978 (1.991)

Pistil-1DAA: pistil-one day after anthesis; Pistil-AM: pistil-when anthers are at mature stage; Pistil-AI: pistil-when anthers are at immature stage; Boot: Head at boot stage. Parents - Parental level of expression; SHW - Synthetic hexaploid wheat.

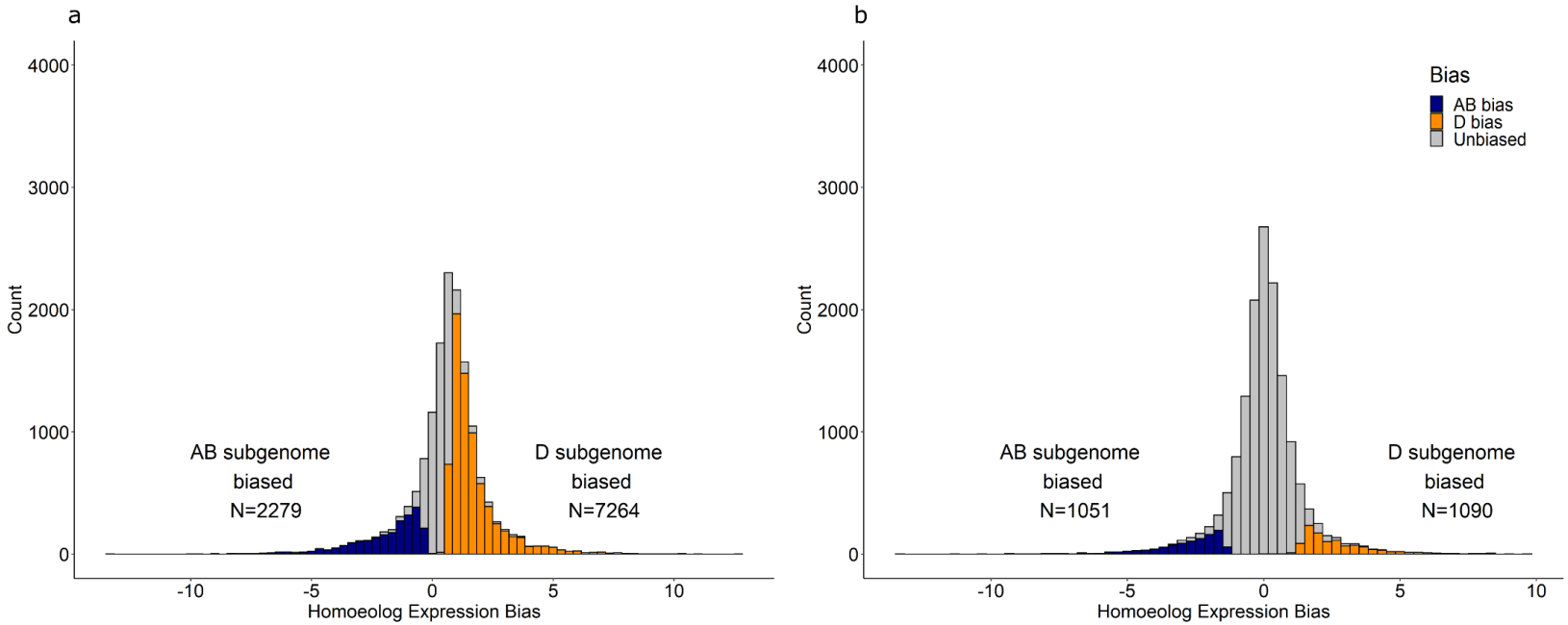
The number of triads showing significant bias towards the AB and D subgenomes in the SHW and parental expression levels (*in-silico* SHW scenario) are summarized in this table. The values in the brackets in each cell represent the average value of the magnitude of HEB in the particular SHW-tissue scenario i.e., $\frac{1}{n}\sum_{i=1}^{n}\log_2\left(\frac{RPKM_D}{RPKM_{AB}}\right)$, where n is the number of triads showing significant bias towards the specific subgenome, and $RPKM_D$ and $RPKM_{AB}$ represent the expression values from the D and AB subgenomes, respectively.



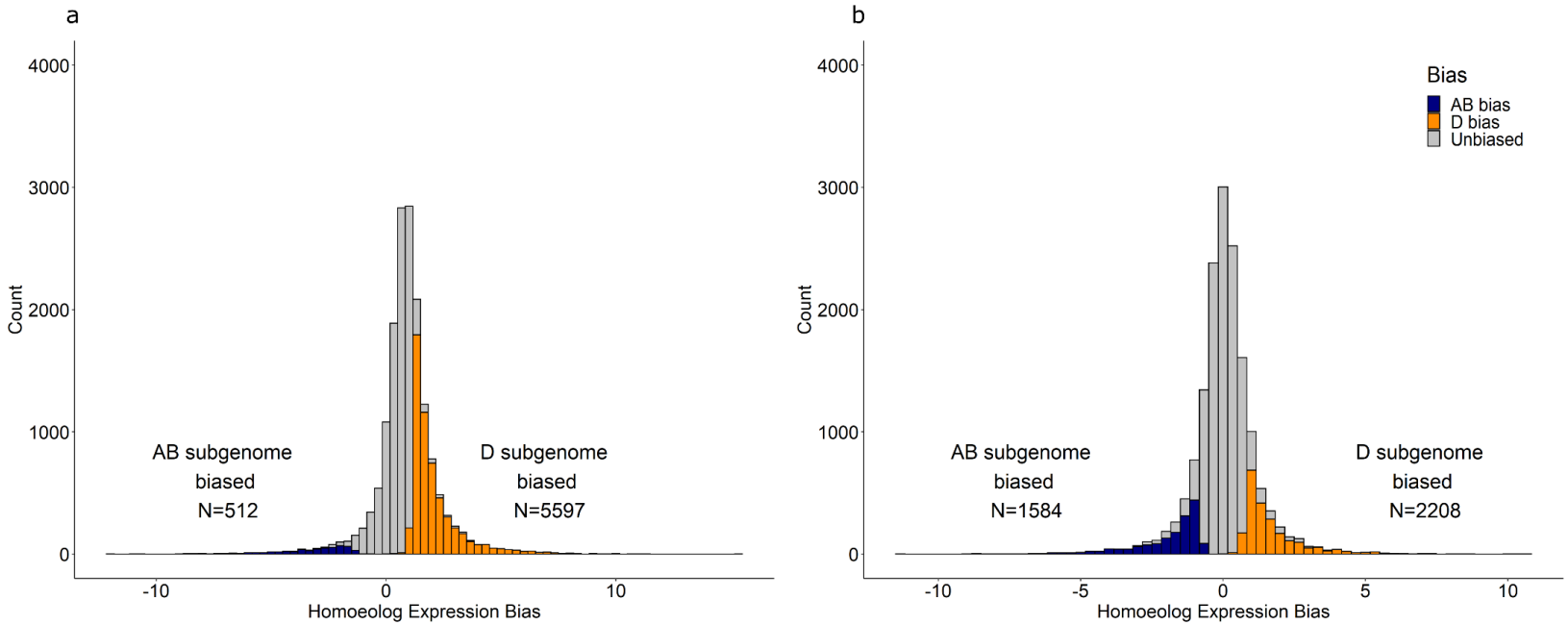
Appendix 7 Distribution of homoeologue expression bias in pistil-one day after anthesis of SHW-C66. (a) Comparison of the expression bias of triads towards AB and D genomes of tetraploid (PI377655) and diploid (AS2386) parents; (b) Comparison of the expression bias of triads towards AB and D subgenomes of SHW C66.



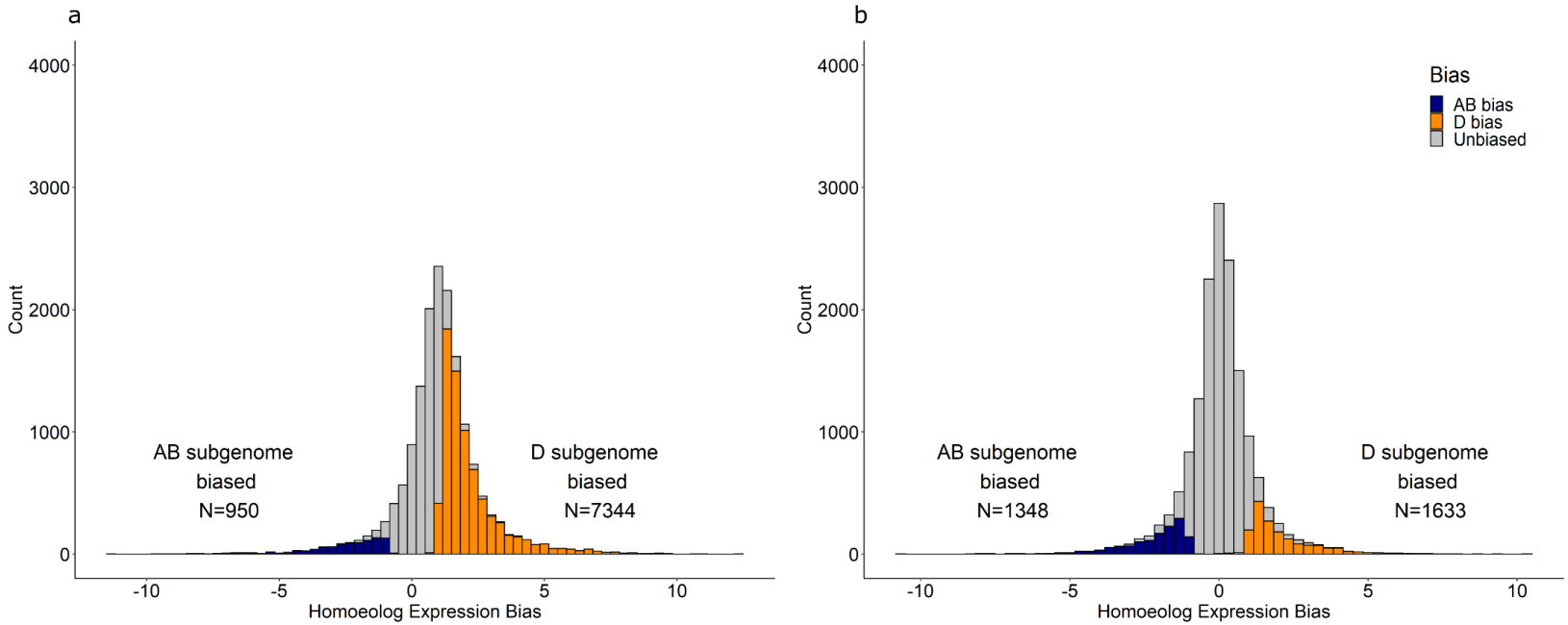
Appendix 8 Distribution of homoeologue expression bias in head collected at boot stage of SHW-C66. (a) Comparison of the expression bias of triads towards AB and D genomes of tetraploid (PI377655) and diploid (AS2386) parents; (b) Comparison of the expression bias of triads towards AB and D subgenomes of SHW C66.



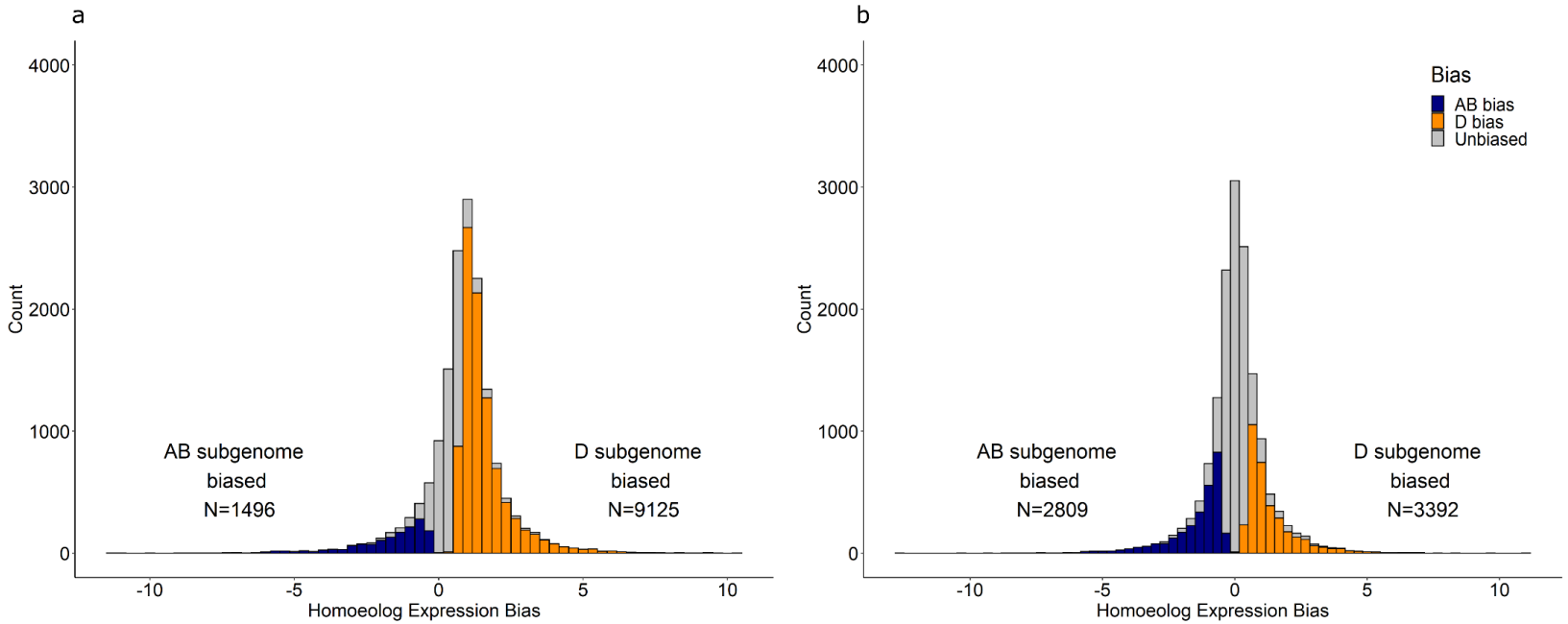
Appendix 9 Distribution of homoeologue expression bias in glume SHW-C66. (a) Comparison of the expression bias of triads towards AB and D genomes of tetraploid (PI377655) and diploid (AS2386) parents; (b) Comparison of the expression bias of triads towards AB and D subgenomes of SHW C66.



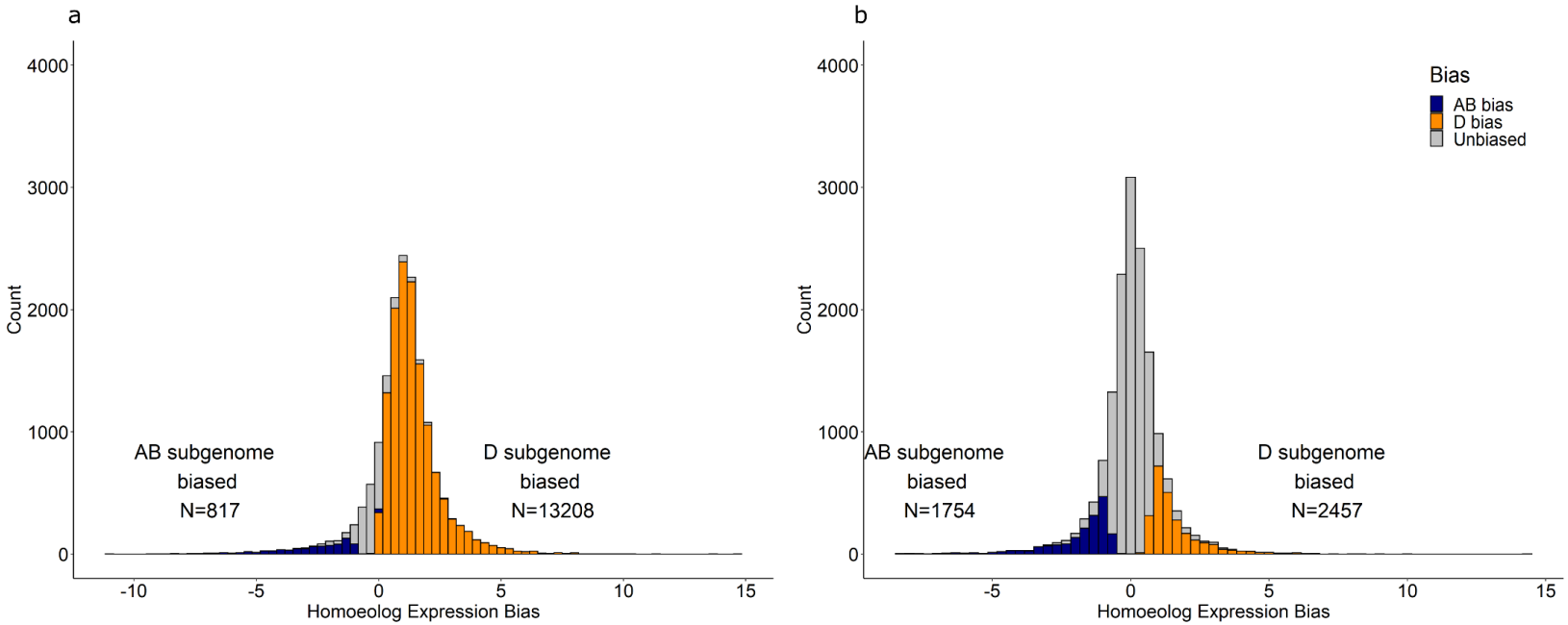
Appendix 10 Distribution of homoeologue expression bias in hypocotyl of SHW-C66. (a) Comparison of the expression bias of triads towards AB and D genomes of tetraploid (PI377655) and diploid (AS2386) parents; (b) Comparison of the expression bias of triads towards AB and D subgenomes of SHW C66.



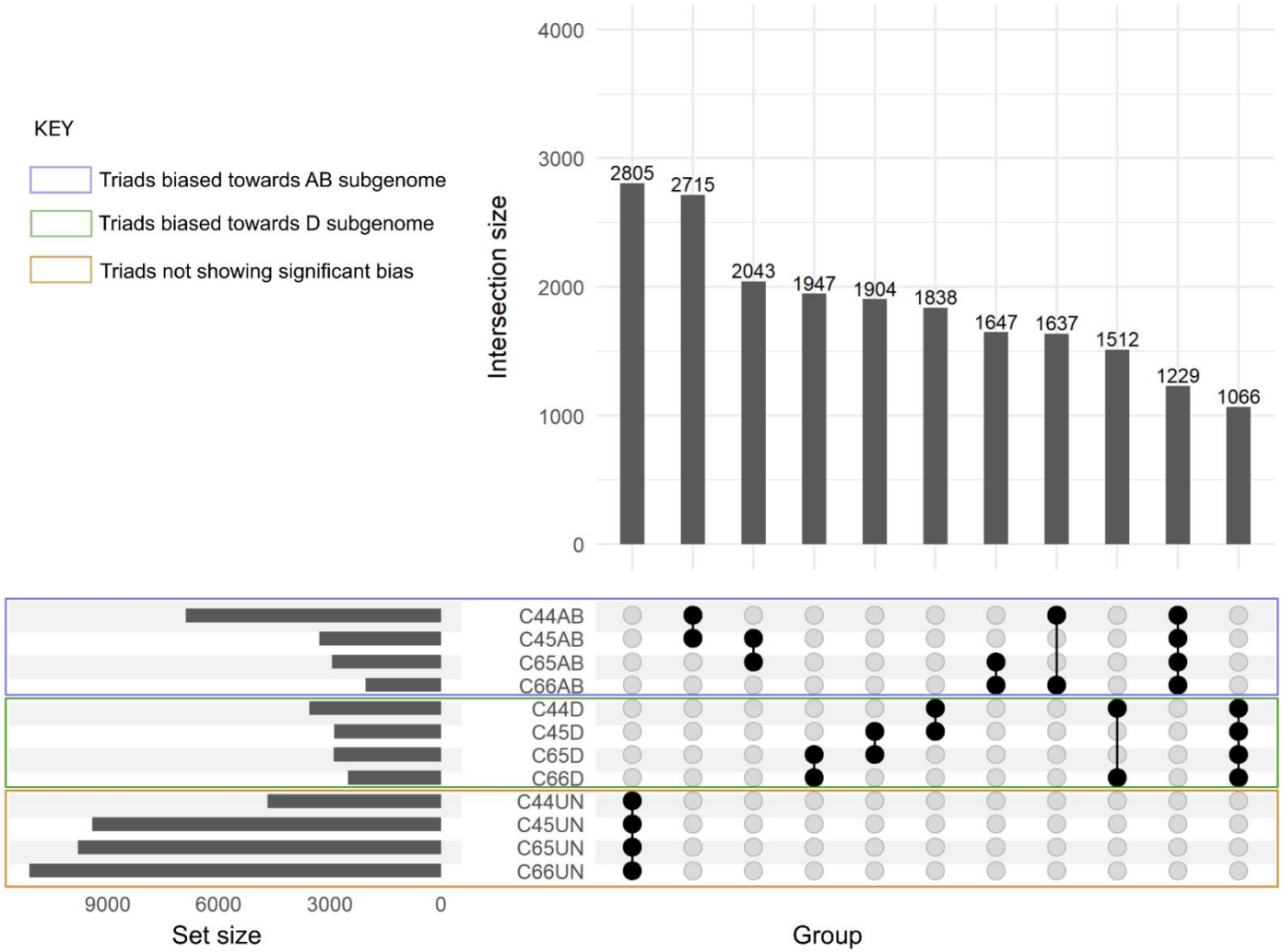
Appendix 11 Distribution of homoeologue expression bias in palea+lemma SHW-C66. (a) Comparison of the expression bias of triads towards AB and D genomes of tetraploid (PI377655) and diploid (AS2386) parents; (b) Comparison of the expression bias of triads towards AB and D subgenomes of SHW C66.



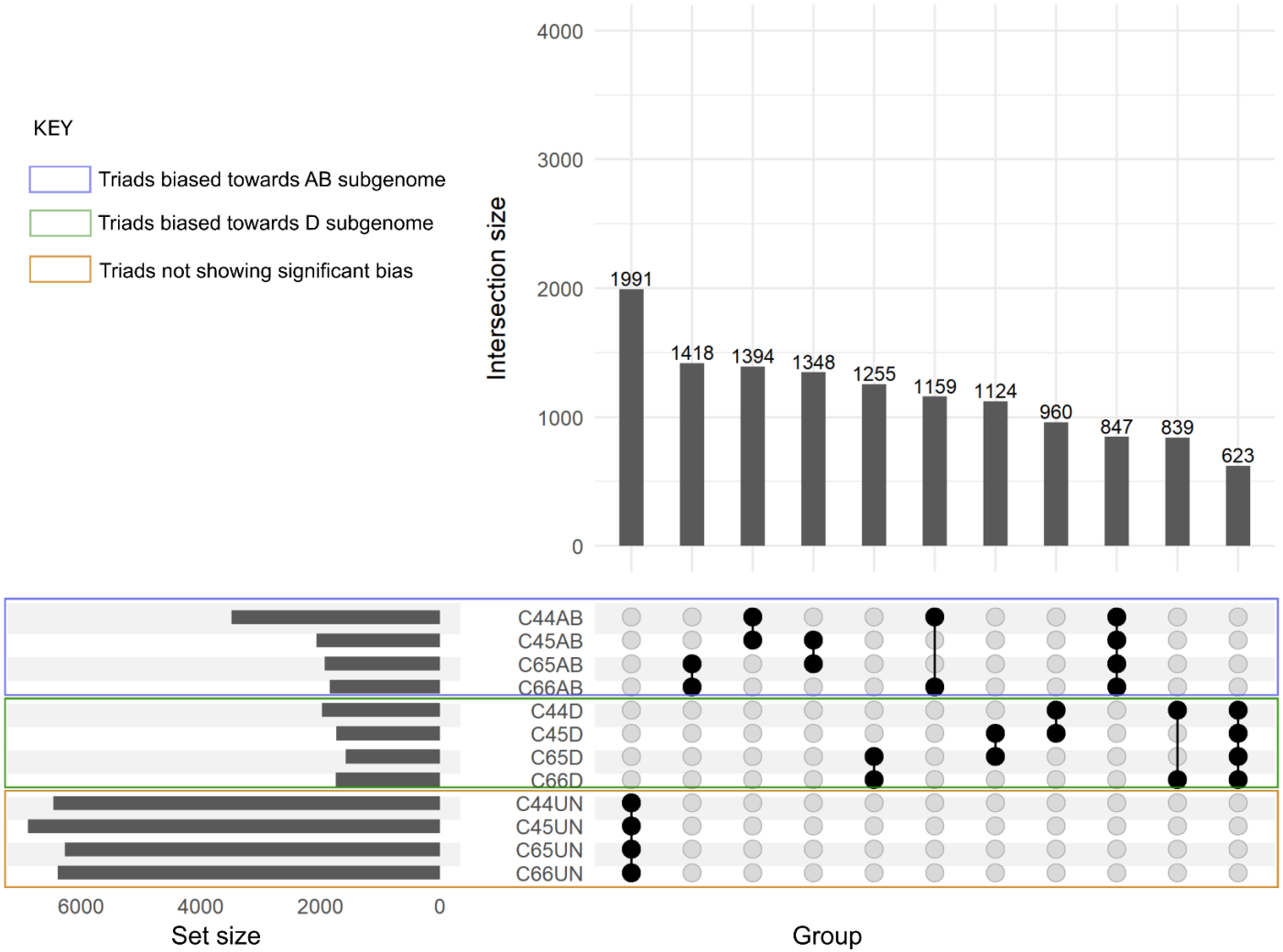
Appendix 12 Distribution of homoeologue expression bias in pistil-when anthers are yellow and just prior to dehiscence, of SHW-C66. (a) Comparison of the expression bias of triads towards AB and D genomes of tetraploid (PI377655) and diploid (AS2386) parents; (b) Comparison of the expression bias of triads towards AB and D subgenomes of SHW C66.



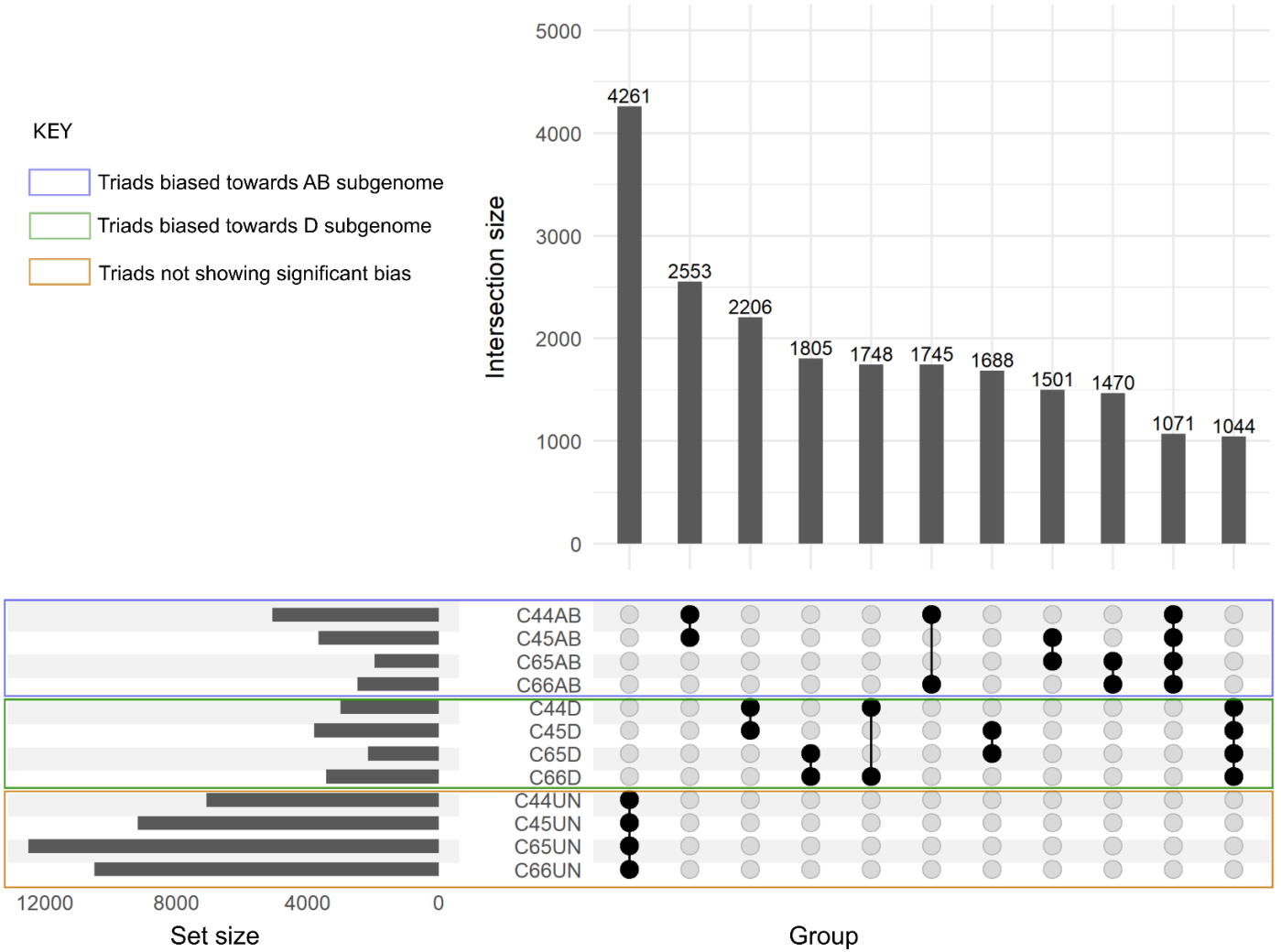
Appendix 13 Distribution of homoeologue expression bias in root of SHW-C66. (a) Comparison of the expression bias of triads towards AB and D genomes of tetraploid (PI377655) and diploid (AS2386) parents; (b) Comparison of the expression bias of triads towards AB and D subgenomes of SHW C66.



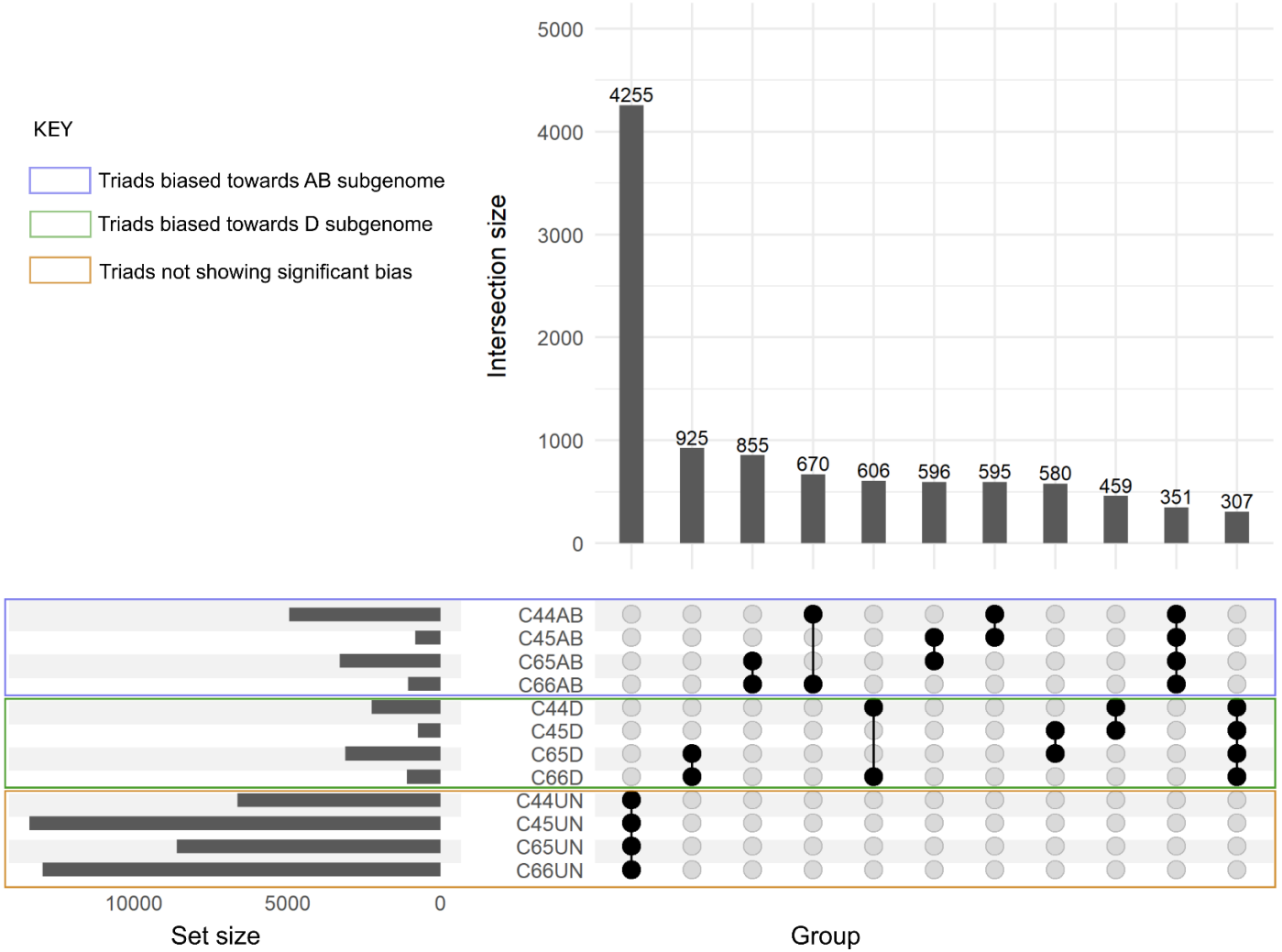
Appendix 14 The subset of triads showing similar expression patterns across all four SHW lines in pistil-one day after anthesis.



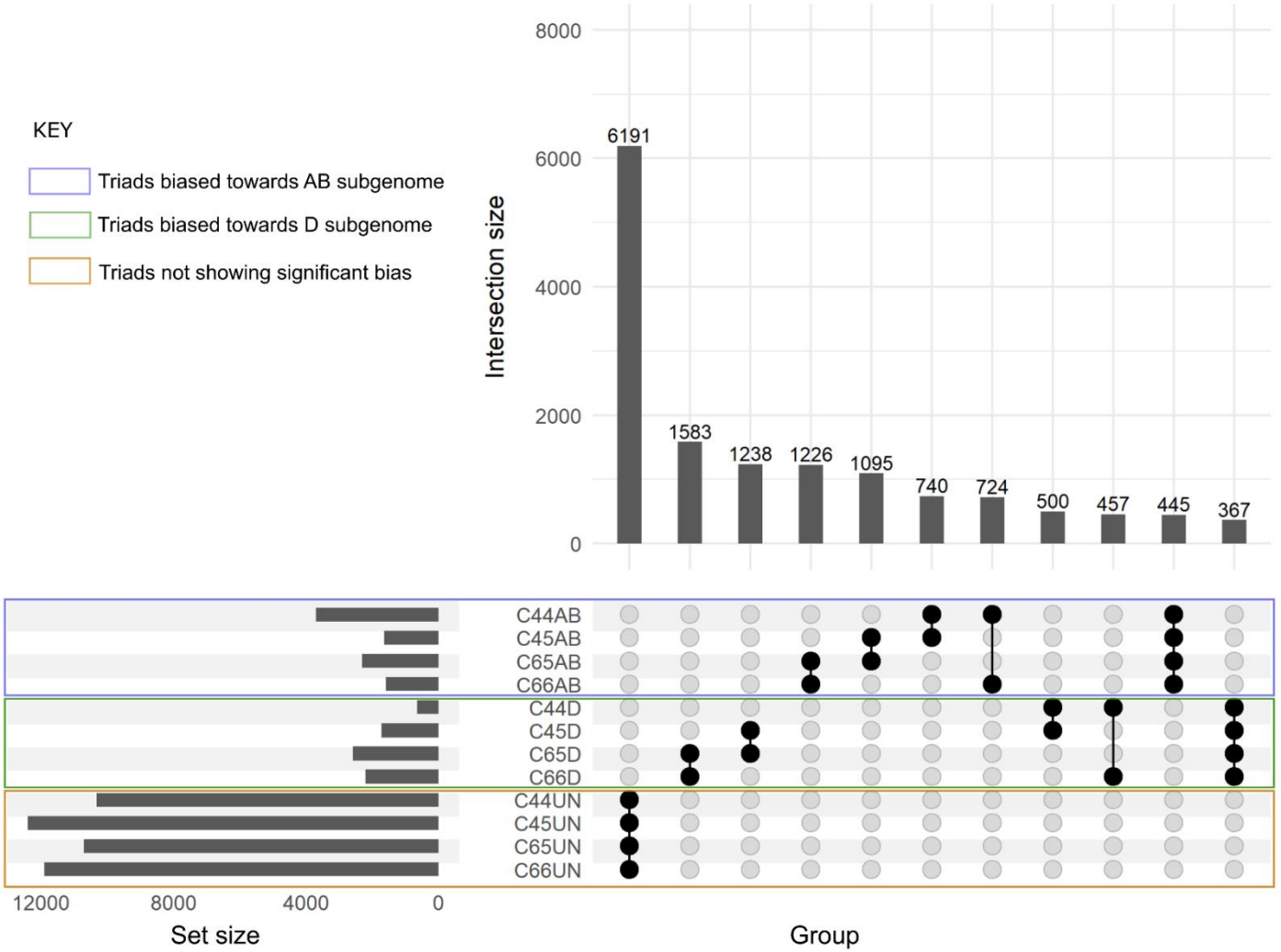
Appendix 15 The subset of triads showing similar expression patterns across all four SHW lines in mature anther.



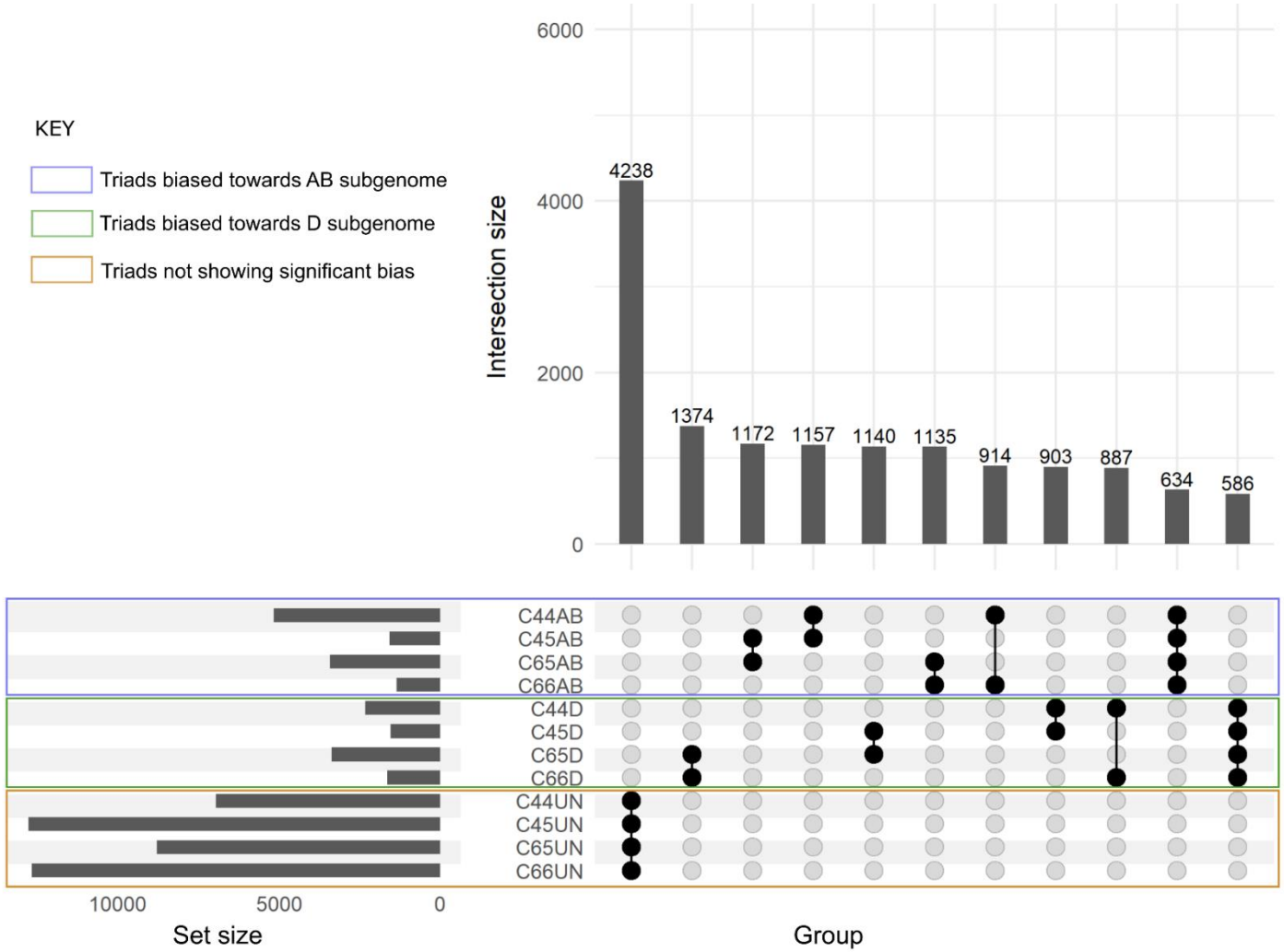
Appendix 16 The subset of triads showing similar expression patterns across all four SHW lines in head collected at boot stage.



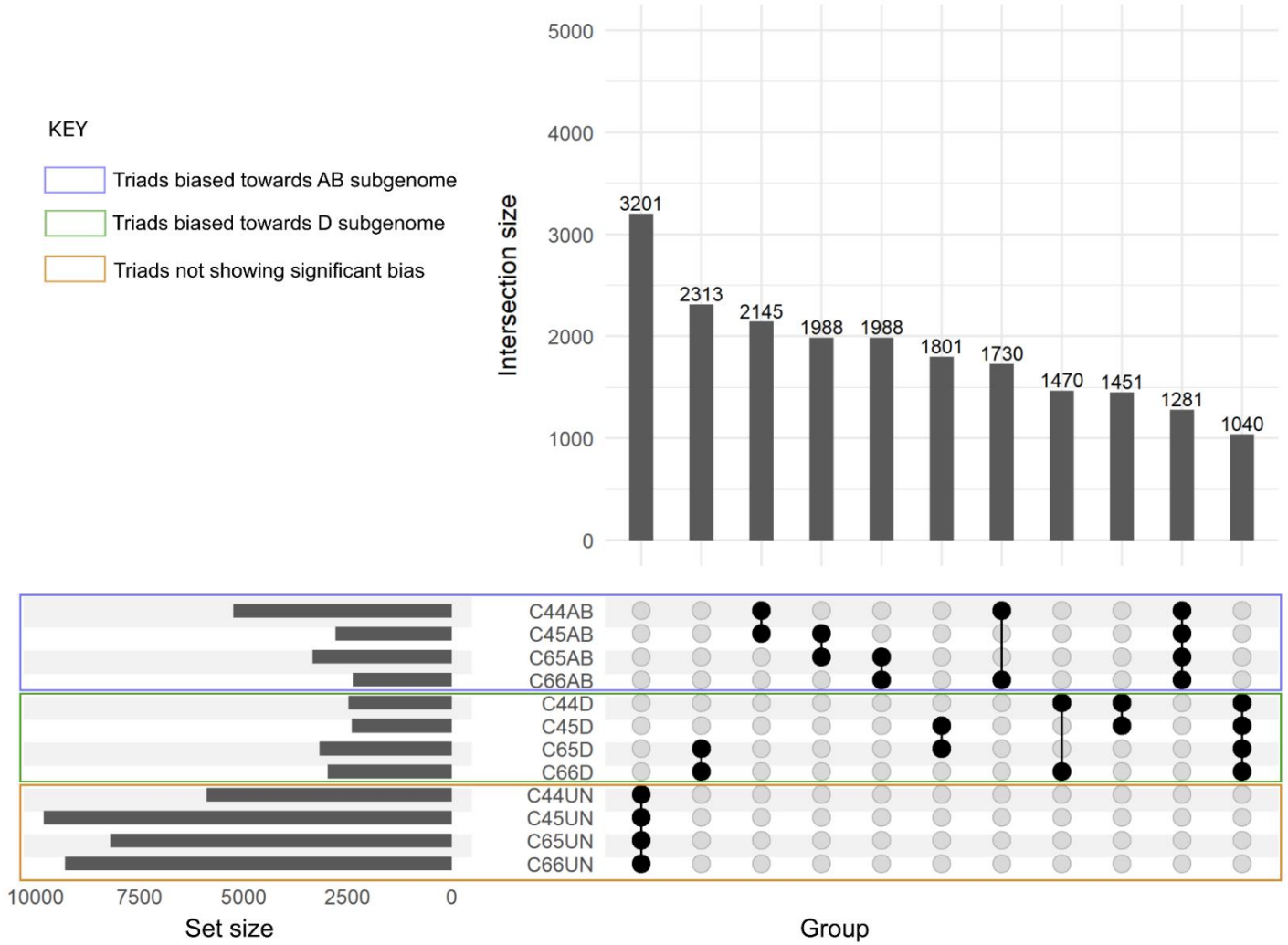
Appendix 17 The subset of triads showing similar expression patterns across all four SHW lines in glume.



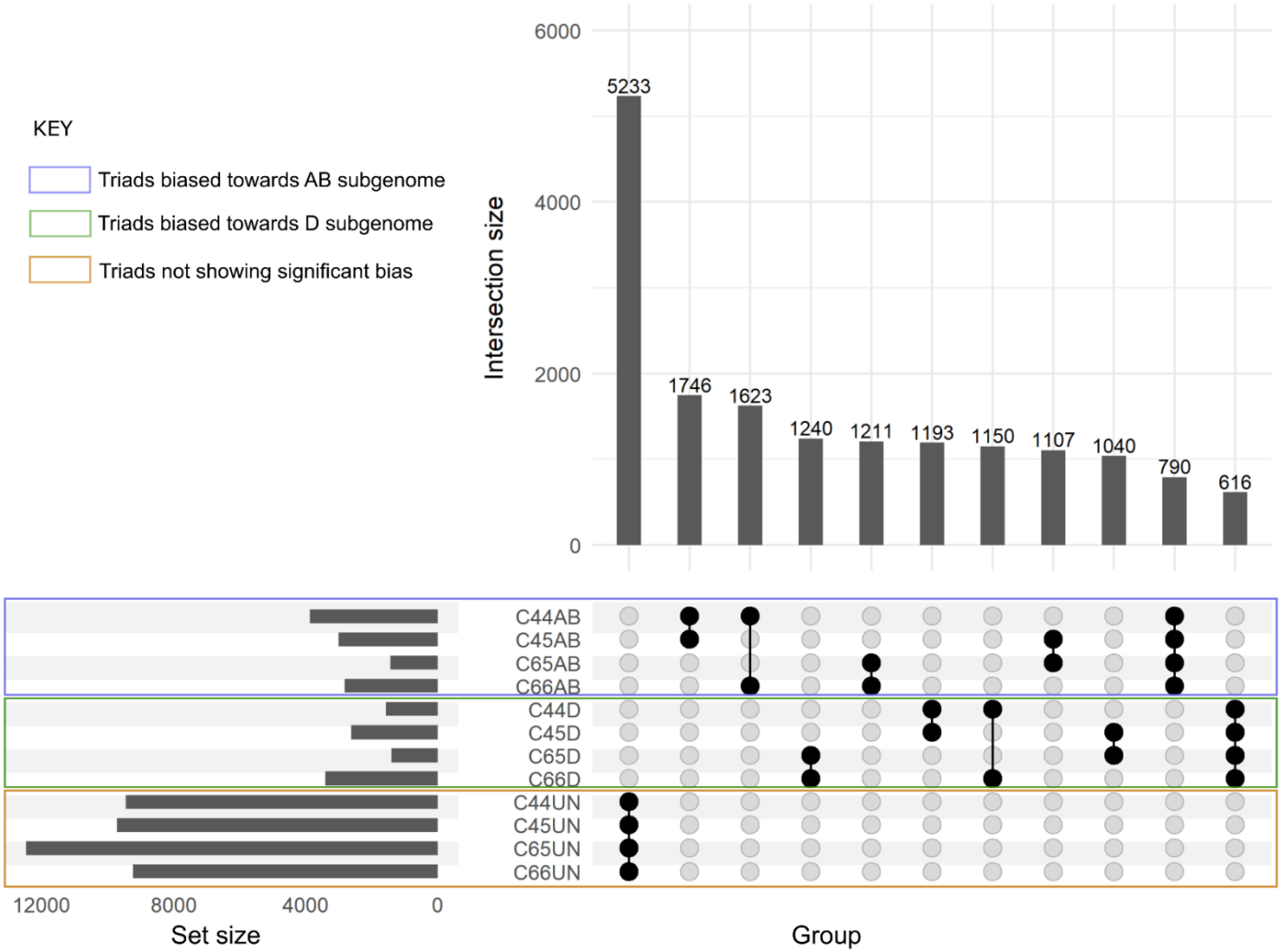
Appendix 18 The subset of triads showing similar expression patterns across all four SHW lines in hypocotyl.



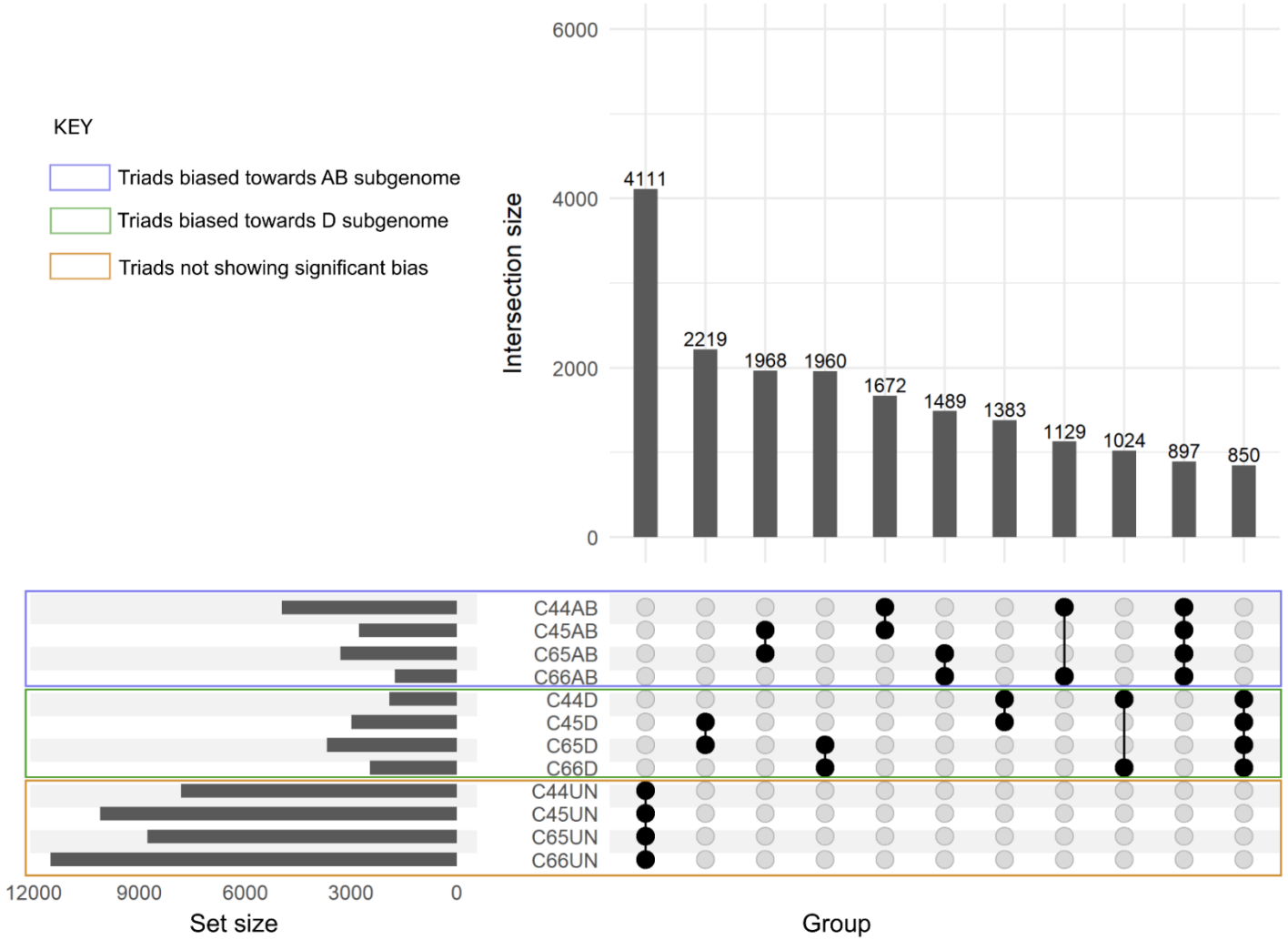
Appendix 19 The subset of triads showing similar expression patterns across all four SHW lines in palea+lemma.



Appendix 20 The subset of triads showing similar expression patterns across all four SHW lines in pistil-when anthers are green and immature.



Appendix 21 The subset of triads showing similar expression patterns across all four SHW lines in pistil-when anthers are yellow and just prior to dehiscence.



Appendix 22 The subset of triads showing similar expression patterns across all four SHW lines in root.

Appendix 23 Summary of tissue specificity of homoeologues constituting the triads in the four SHW lines

	Number of triads (absolute values)				Relative percentage*			
Homoeologue tissue specificity	**C44**	**C45**	**C65**	**C66**	**C44**	**C45**	**C65**	**C66**
All homoeologues not tissue specific[1]	12,751	13,368	13,241	13,391	69.5	72.8	72.1	72.9
All homoeologues same tissue specificity[2]	1,556	1,938	1,868	1,798	8.5	10.6	10.2	9.8
D tissue specific, A and B not tissue specific[3]	1204	381	399	394	6.6	2.1	2.2	2.1
B tissue specific, A and D not tissue specific[3]	516	476	495	523	2.8	2.6	2.7	2.8
A tissue specific, B and D not tissue specific[3]	473	422	482	481	2.6	2.3	2.6	2.6
A and B same tissue specificity, D other tissue specificity[4]	119	137	98	99	0.6	0.7	0.5	0.5
A and B same tissue specificity, D not tissue specific[5]	351	227	245	228	1.9	1.2	1.3	1.2
A and D same tissue specificity, B other tissue specificity[4]	115	153	168	158	0.6	0.8	0.9	0.9
A and D same tissue specificity, B not tissue specific[5]	259	263	277	277	1.4	1.4	1.5	1.5
B and D same tissue specificity, A other tissue specificity[4]	114	156	169	167	0.6	0.8	0.9	0.9
B and D same tissue specificity, A not tissue specific[5]	269	259	293	279	1.5	1.4	1.6	1.5
All homoeologues showing varied tissue specificity[6]	630	577	622	562	3.4	3.1	3.4	3.1
Total	18,357	18,357	18,357	18,357	100.0	100.0	100.0	100.0

[1]All three homoeologues of the triads are broadly expressed without any tissue specificity; 2All three homoeologues show expression specificity to the same tissue; 3One of the three homoeologues is tissue specific and the other two are broadly expressed; 4Two of the three homoeologues show expression specificity in the same tissue and the third homoeologue shows expression specificity in another tissue; 5Two of the three homoeologues show expression specificity in the same tissue and the third homoeologue is broadly expressed; 6All homoeologues showed varied tissue specificity or wider expression. *Percentage of triads relative to the total number of triads (18,357).

Appendix 24 Summary of number of triads with varied tissue specificity of homoeologues and expression bias patterns across the ten tissues in the four SHW lines

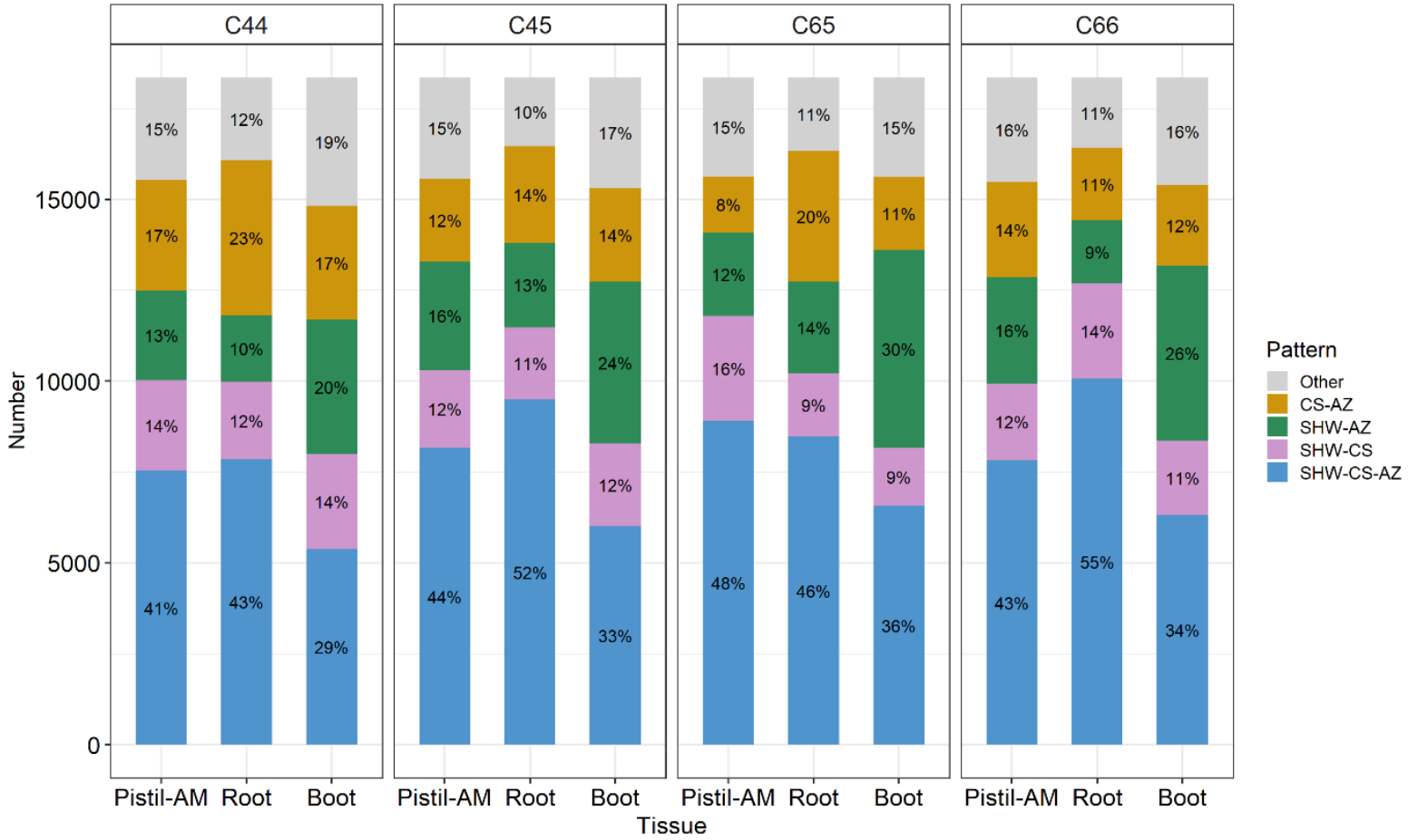
SHW	Homoeologues showing tissue specificity in triads	Bias pattern	Pistil-1DAA	Pistil-AI	Palea+ Lemma	Pistil-AM	Root	Shoot	Anther	Boot	Glume	Hypocotyl
C44	**A,B,D**	**Total**	**13**	**23**	**6**	**18**	**653**	**24**	**600**	**140**	**14**	**65**
C44	A,B,D	AB-bias	5	6	1	5	244	6	236	33	5	16
C44	A,B,D	D-bias	7	2	2	1	125	6	143	32	4	6
C44	A,B,D	UN-biased	1	15	3	12	284	12	221	75	5	43
C44	**A and B**	**Total**	**16**	**20**	**6**	**20**	**219**	**15**	**77**	**49**	**8**	**40**
C44	A and B	AB-bias	9	3	1	2	59	4	17	3	1	8
C44	A and B	D-bias	2	2	0	1	13	2	3	3	1	3
C44	A and B	UN-biased	5	15	5	17	147	9	57	43	6	29
C44	**A and D**	**Total**	**14**	**21**	**4**	**11**	**146**	**29**	**72**	**37**	**12**	**28**
C44	A and D	AB-bias	7	4	2	2	41	6	29	7	5	4
C44	A and D	D-bias	3	4	1	2	42	7	18	9	1	3
C44	A and D	UN-biased	4	13	1	7	63	16	25	21	6	21
C44	**B and D**	**Total**	**19**	**11**	**7**	**10**	**134**	**28**	**84**	**47**	**10**	**33**
C44	B and D	AB-bias	4	2	1	4	46	5	27	6	4	7
C44	B and D	D-bias	8	5	2	1	33	7	26	7	2	3
C44	B and D	UN-biased	7	4	4	5	55	16	31	34	4	23
C44	**A only**	**Total**	**60**	**80**	**59**	**82**	**231**	**92**	**176**	**112**	**72**	**100**
C44	A only	AB-bias	6	12	8	5	57	10	34	8	8	13
C44	A only	D-bias	11	19	8	16	16	15	9	16	13	4
C44	A only	UN-biased	43	49	43	61	158	67	133	88	51	83
C44	**B only**	**Total**	**77**	**78**	**47**	**71**	**266**	**78**	**188**	**90**	**60**	**140**
C44	B only	AB-bias	20	10	9	9	51	8	27	13	6	17
C44	B only	D-bias	11	14	4	14	38	15	25	6	11	13
C44	B only	UN-biased	46	54	34	48	177	55	136	71	43	110
C44	**D only**	**Total**	**158**	**225**	**116**	**163**	**241**	**189**	**204**	**159**	**121**	**164**
C44	D only	AB-bias	104	181	76	110	114	85	100	72	91	98
C44	D only	D-bias	20	4	8	9	33	31	30	20	6	7
C44	D only	UN-biased	34	40	32	44	94	73	74	67	24	59
C45	**A,B,D**	**Total**	**58**	**33**	**21**	**8**	**838**	**52**	**527**	**268**	**10**	**123**
C45	A,B,D	AB-bias	16	2	3	0	178	6	159	43	2	24
C45	A,B,D	D-bias	17	2	7	4	236	5	146	64	1	23
C45	A,B,D	UN-biased	25	29	11	4	424	41	222	161	7	76
C45	**A and B**	**Total**	**25**	**21**	**8**	**10**	**127**	**37**	**37**	**50**	**6**	**43**
C45	A and B	AB-bias	7	4	1	3	33	5	13	7	2	12

SHW	Homoeologues showing tissue specificity in triads	Bias pattern	Pistil-1DAA	Pistil-AI	Palea+ Lemma	Pistil-AM	Root	Shoot	Anther	Boot	Glume	Hypocotyl
C45	A and B	D-bias	3	1	1	1	17	6	8	6	1	8
C45	A and B	UN-biased	15	16	6	6	77	26	16	37	3	23
C45	**A and D**	**Total**	**25**	**22**	**10**	**5**	**157**	**51**	**46**	**57**	**5**	**38**
C45	A and D	AB-bias	3	5	2	2	28	8	9	11	0	9
C45	A and D	D-bias	9	3	0	1	52	12	14	16	0	11
C45	A and D	UN-biased	13	14	8	2	77	31	23	30	5	18
C45	**B and D**	**Total**	**24**	**14**	**9**	**5**	**165**	**38**	**39**	**61**	**6**	**54**
C45	B and D	AB-bias	6	2	2	1	24	3	8	8	0	3
C45	B and D	D-bias	4	2	3	3	64	9	15	14	0	15
C45	B and D	UN-biased	14	10	4	1	77	26	16	39	6	36
C45	**A only**	**Total**	**89**	**81**	**78**	**61**	**199**	**99**	**110**	**124**	**42**	**114**
C45	A only	AB-bias	15	8	13	7	36	8	31	29	5	17
C45	A only	D-bias	15	20	14	13	40	10	11	22	7	14
C45	A only	UN-biased	59	53	51	41	123	81	68	73	30	83
C45	**B only**	**Total**	**86**	**80**	**81**	**53**	**237**	**130**	**99**	**135**	**34**	**126**
C45	B only	AB-bias	18	8	3	8	42	13	18	20	0	26
C45	B only	D-bias	14	20	18	9	43	13	14	29	4	15
C45	B only	UN-biased	54	52	60	36	152	104	67	86	30	85
C45	**D only**	**Total**	**69**	**64**	**76**	**50**	**218**	**105**	**86**	**105**	**52**	**104**
C45	D only	AB-bias	6	15	9	11	30	11	9	14	10	11
C45	D only	D-bias	19	9	12	6	61	14	33	26	9	22
C45	D only	UN-biased	44	40	55	33	127	80	44	65	33	71
C65	**A,B,D**	**Total**	**82**	**22**	**10**	**15**	**856**	**53**	**468**	**237**	**8**	**117**
C65	A,B,D	AB-bias	19	2	1	4	189	3	141	27	2	29
C65	A,B,D	D-bias	16	4	4	2	251	6	107	41	1	21
C65	A,B,D	UN-biased	47	16	5	9	416	44	220	169	5	67
C65	**A and B**	**Total**	**22**	**11**	**6**	**13**	**142**	**38**	**33**	**33**	**9**	**36**
C65	A and B	AB-bias	6	3	1	3	38	3	12	6	1	10
C65	A and B	D-bias	4	2	0	1	25	6	7	4	3	7
C65	A and B	UN-biased	12	6	5	9	79	29	14	23	5	19
C65	**A and D**	**Total**	**34**	**25**	**10**	**8**	**181**	**41**	**48**	**57**	**4**	**37**
C65	A and D	AB-bias	5	1	2	3	37	4	9	10	1	10
C65	A and D	D-bias	8	4	3	0	59	12	24	10	1	11
C65	A and D	UN-biased	21	20	5	5	85	25	15	37	2	16
C65	**B and D**	**Total**	**32**	**19**	**8**	**10**	**192**	**33**	**44**	**68**	**6**	**50**
C65	B and D	AB-bias	5	4	1	2	35	5	8	8	1	4
C65	B and D	D-bias	11	2	1	3	83	11	21	13	1	18

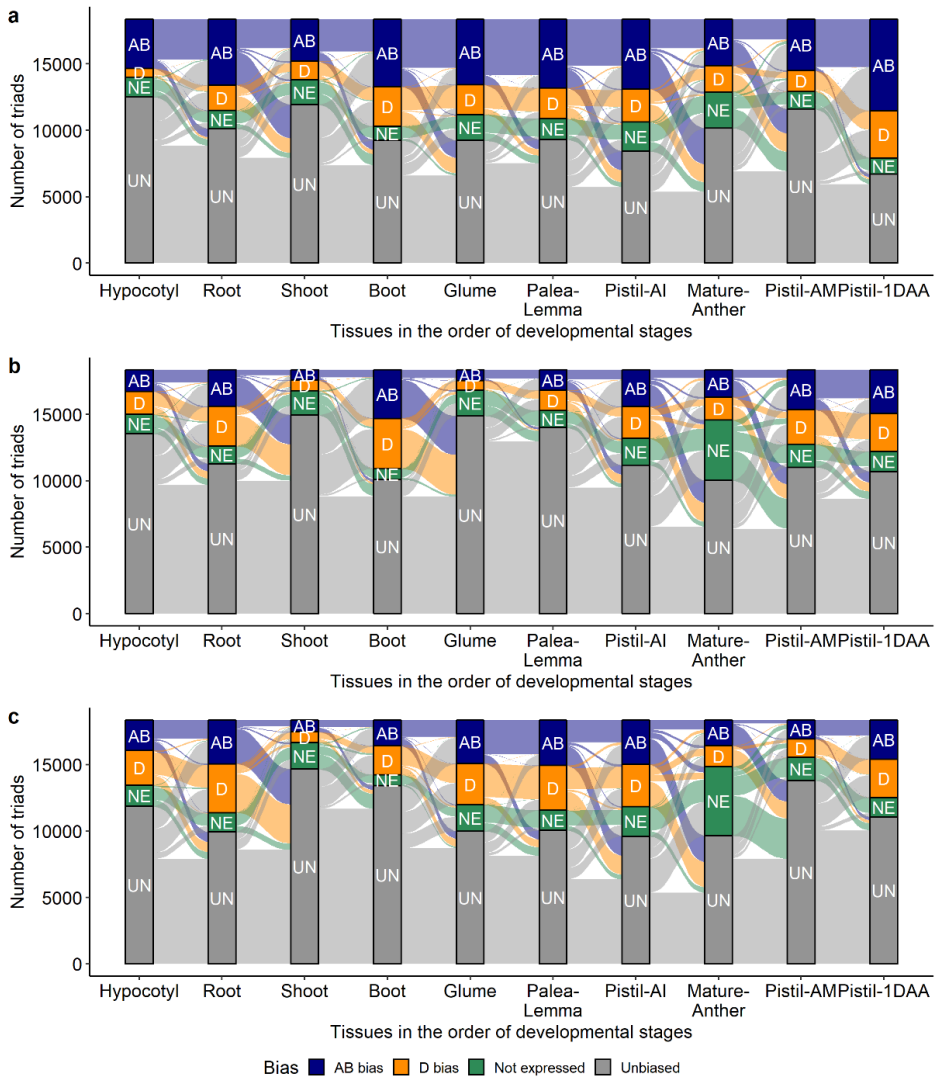
SHW	Homoeologues showing tissue specificity in triads	Bias pattern	Pistil-1DAA	Pistil-AI	Palea+ Lemma	Pistil-AM	Root	Shoot	Anther	Boot	Glume	Hypocotyl
C65	B and D	UN-biased	16	13	6	5	74	17	15	47	4	28
C65	**A only**	**Total**	**88**	**97**	**81**	**54**	**221**	**117**	**120**	**173**	**45**	**133**
C65	A only	AB-bias	15	8	8	6	50	16	28	34	4	23
C65	A only	D-bias	10	27	17	7	46	17	13	31	12	31
C65	A only	UN-biased	63	62	56	41	125	84	79	108	29	79
C65	**B only**	**Total**	**100**	**84**	**68**	**67**	**244**	**106**	**110**	**137**	**63**	**125**
C65	B only	AB-bias	14	12	10	7	46	10	26	21	6	30
C65	B only	D-bias	9	16	15	5	55	6	10	20	14	20
C65	B only	UN-biased	77	56	43	55	143	90	74	96	43	75
C65	**D only**	**Total**	**78**	**75**	**55**	**52**	**224**	**94**	**94**	**128**	**51**	**101**
C65	D only	AB-bias	9	12	9	9	38	11	10	19	11	23
C65	D only	D-bias	17	12	10	7	67	17	34	27	11	20
C65	D only	UN-biased	52	51	36	36	119	66	50	82	29	58
C66	**A,B,D**	**Total**	**65**	**26**	**11**	**15**	**781**	**29**	**503**	**254**	**5**	**109**
C66	A,B,D	AB-bias	13	6	0	1	118	3	123	30	1	26
C66	A,B,D	D-bias	16	3	3	7	184	6	136	56	2	16
C66	A,B,D	UN-biased	36	17	8	7	479	20	244	168	2	67
C66	**A and B**	**Total**	**24**	**24**	**10**	**6**	**120**	**29**	**36**	**33**	**6**	**39**
C66	A and B	AB-bias	9	2	1	1	27	4	13	4	1	8
C66	A and B	D-bias	1	2	1	1	18	4	7	3	1	5
C66	A and B	UN-biased	14	20	8	4	75	21	16	26	4	26
C66	**A and D**	**Total**	**23**	**22**	**9**	**17**	**156**	**32**	**48**	**73**	**11**	**44**
C66	A and D	AB-bias	4	3	1	3	20	1	11	13	6	4
C66	A and D	D-bias	4	3	2	2	42	9	21	20	0	11
C66	A and D	UN-biased	15	16	6	12	94	22	16	40	5	29
C66	**B and D**	**Total**	**27**	**20**	**10**	**13**	**174**	**34**	**39**	**69**	**8**	**52**
C66	B and D	AB-bias	4	2	1	3	23	3	11	5	0	3
C66	B and D	D-bias	12	2	3	2	49	4	18	15	1	16
C66	B and D	UN-biased	11	16	6	8	102	27	10	49	7	33
C66	**A only**	**Total**	**100**	**95**	**61**	**68**	**205**	**107**	**111**	**126**	**53**	**130**
C66	A only	AB-bias	10	7	9	6	34	10	31	20	5	18
C66	A only	D-bias	20	23	10	14	30	11	12	20	7	18
C66	A only	UN-biased	70	65	42	48	141	86	68	86	41	94
C66	**B only**	**Total**	**116**	**85**	**46**	**58**	**229**	**110**	**113**	**135**	**52**	**144**
C66	B only	AB-bias	25	10	2	6	41	12	24	15	3	18
C66	B only	D-bias	12	25	10	10	32	11	16	28	4	24
C66	B only	UN-biased	79	50	34	42	156	87	73	92	45	102

SHW	Homoeologues showing tissue specificity in triads	Bias pattern	Pistil-1DAA	Pistil-AI	Palea+ Lemma	Pistil-AM	Root	Shoot	Anther	Boot	Glume	Hypocotyl
C66	**D only**	**Total**	**83**	**86**	**44**	**55**	**211**	**84**	**81**	**121**	**38**	**93**
C66	D only	AB-bias	11	25	4	7	19	9	7	19	11	15
C66	D only	D-bias	20	8	4	12	54	22	25	17	3	19
C66	D only	UN-biased	52	53	36	36	138	53	49	85	24	59

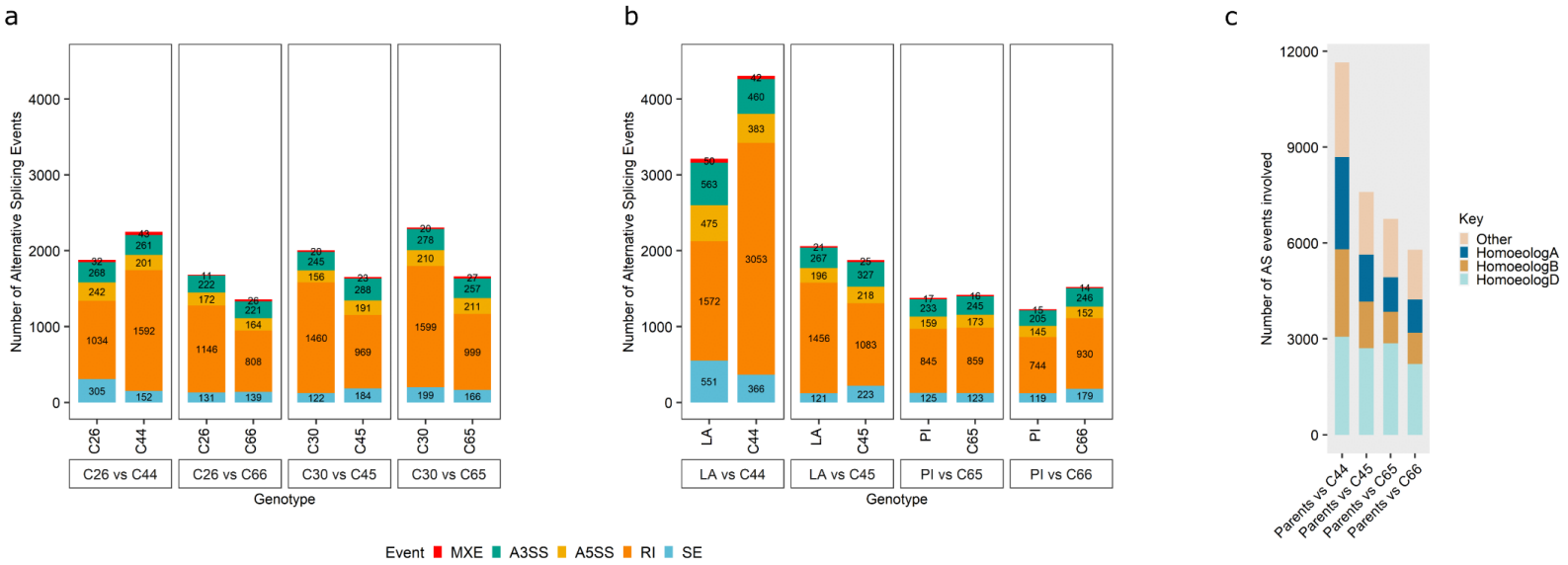
Pistil-1DAA: Pistil - one day after anthesis; Pistil-AM: pistil-when anthers are at mature stage; Pistil-AI: pistil-when anthers are at immature stage; Boot: Head at boot stage.



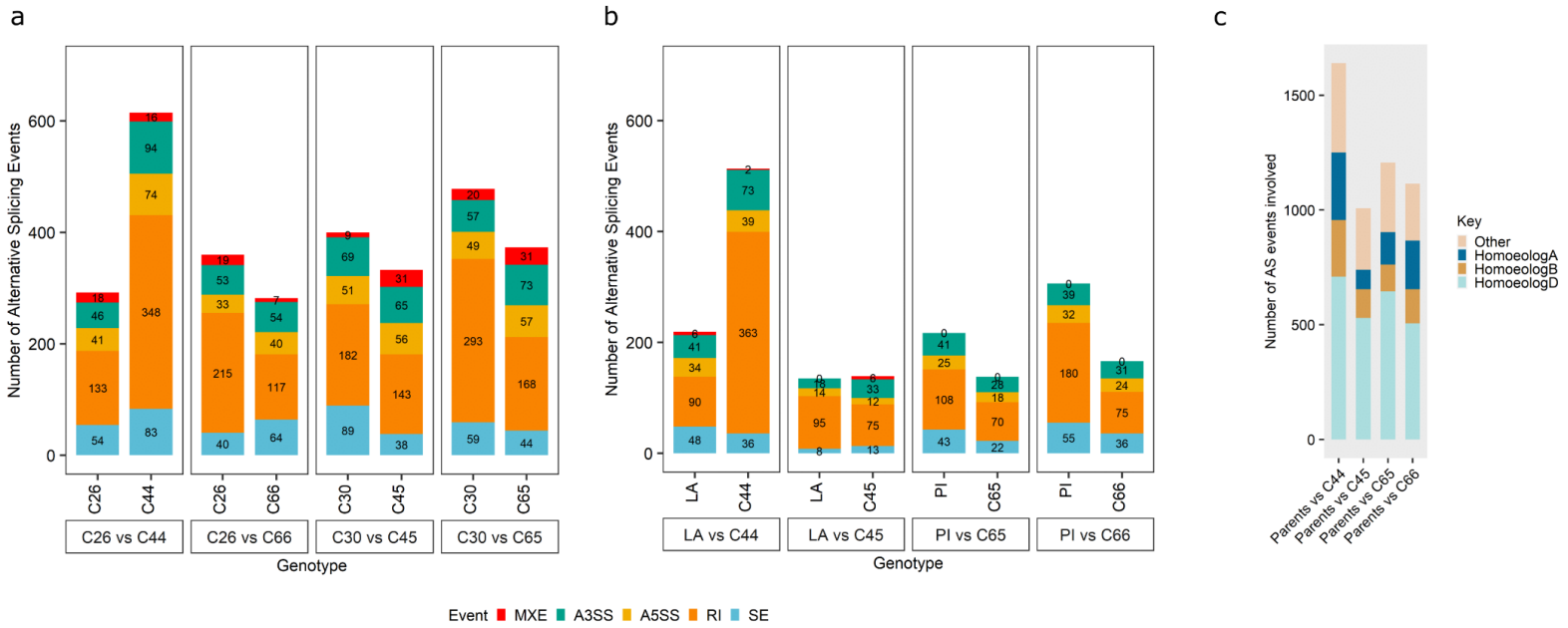
Appendix 25 The proportion of triads showing varied bias patterns between the SHW, the landrace Chinese Spring and cultivar Azhurnaya. The different patterns represent the following: CS-AZ - triads showing similar pattern in Chinese Spring and Azhurnaya and not in the SHW line; SHW-AZ - triads showing reversal of bias in Azhurnaya to the same observed in SHW; SHW-CS - triads showing similar pattern in SHW and Azhurnaya and not in Chinese Spring; SHW-CS-AZ - triads showing similar patterns in SHW, Chinese Spring and Azhurnaya.



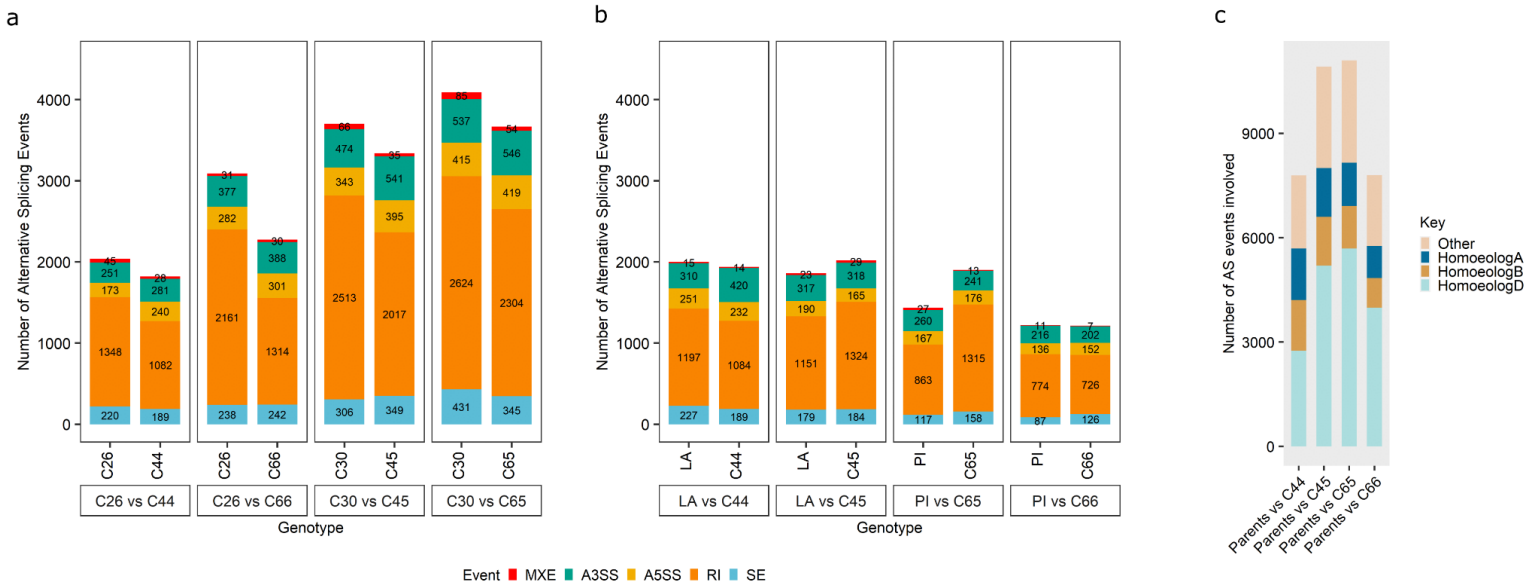
Appendix 26 Alluvial plots representing the expression bias trends of the triads across the tissues in the SHWs (a) SHW - C44 (b) SHW - C45 and (c) SHW - C65.



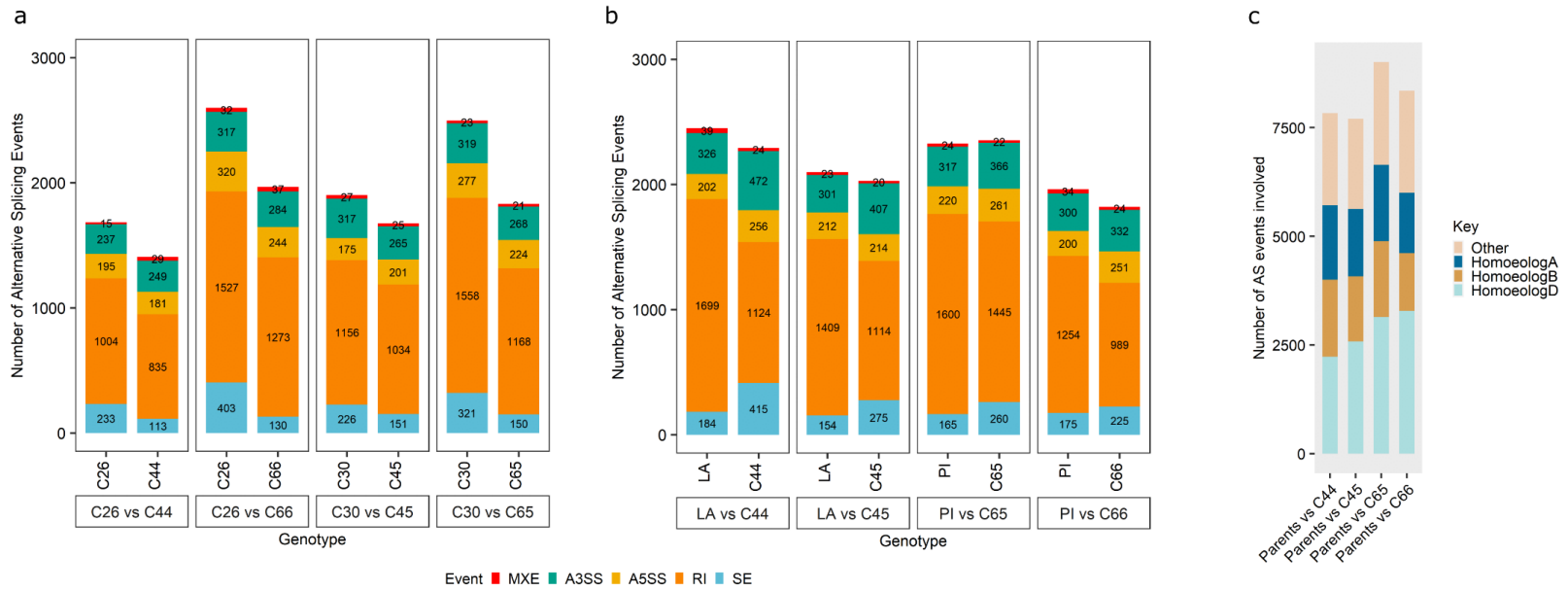
Appendix 27 Number of alternative splicing events detected in the Diploid/Tetraploid parent vs SHW comparison in pistil-one day after anthesis.



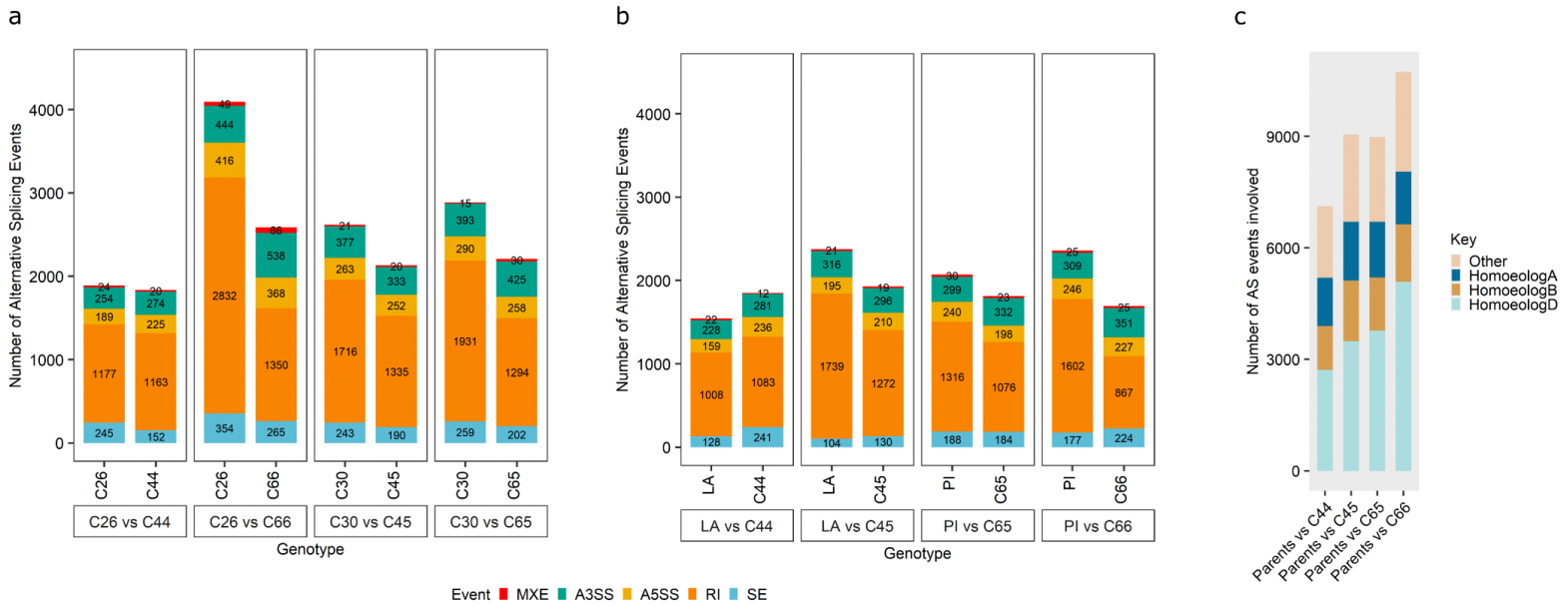
Appendix 28 Number of alternative splicing events detected in the Diploid/Tetraploid parent vs SHW comparison in mature anther.



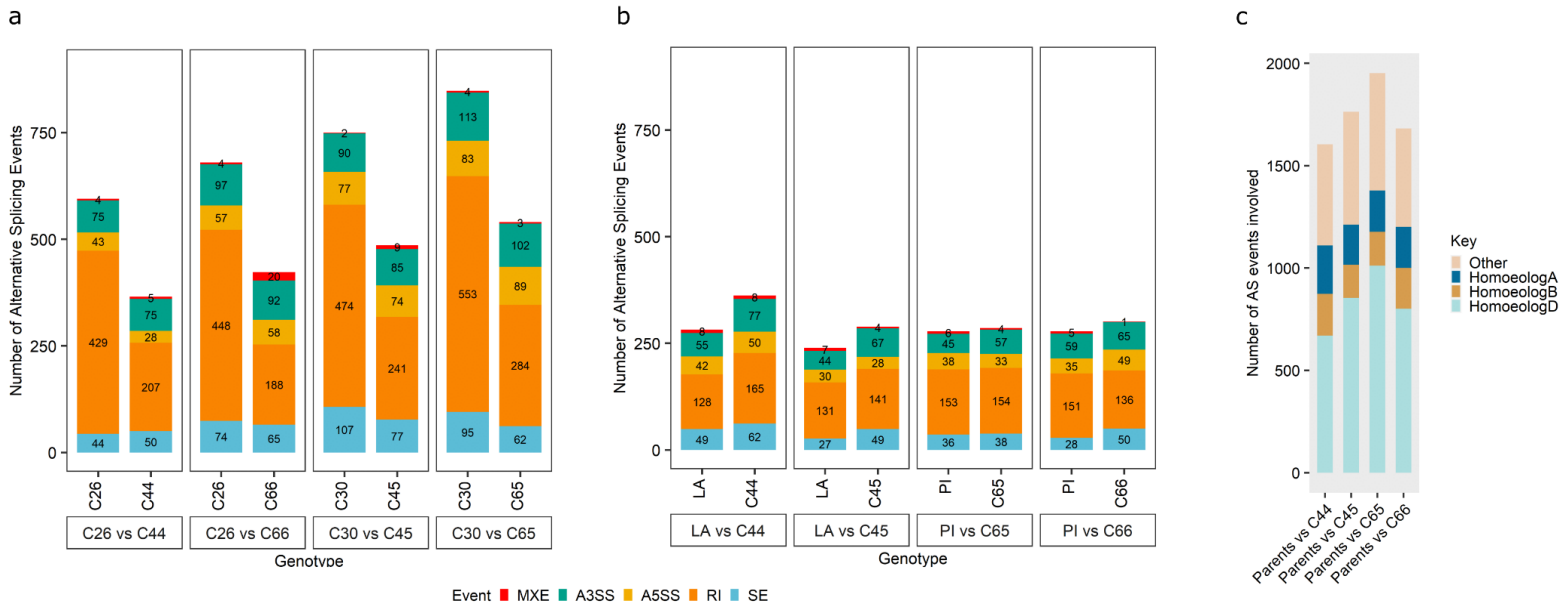
Appendix 29 Number of alternative splicing events detected in the Diploid/Tetraploid parent vs SHW comparison in head at boot stage.



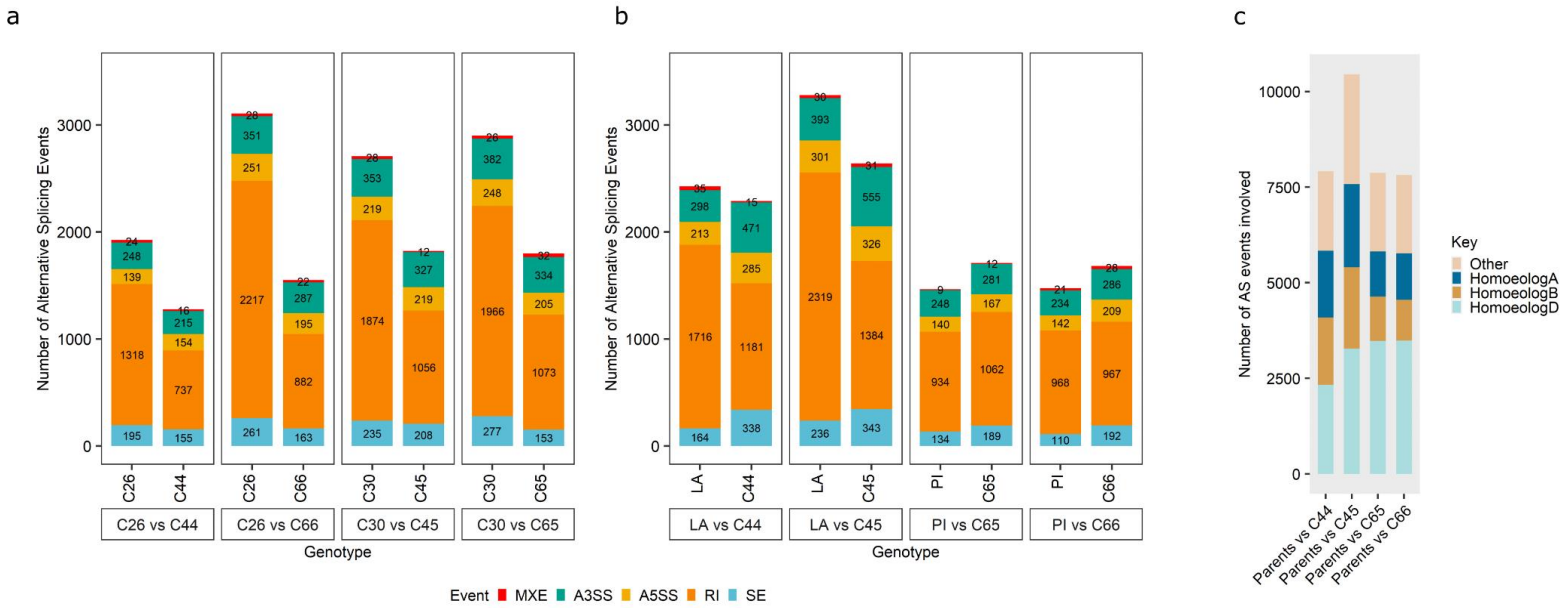
Appendix 30 Number of alternative splicing events detected in the Diploid/Tetraploid parent vs SHW comparison in glume.



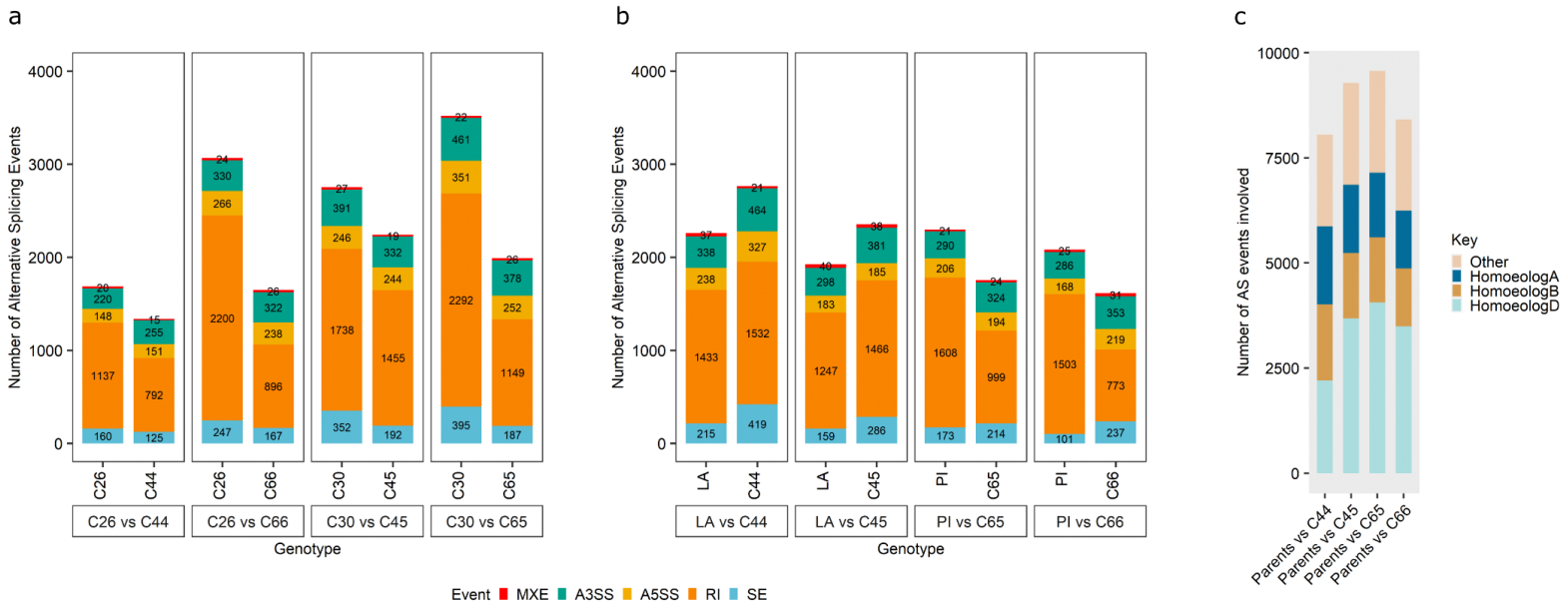
Appendix 31 Number of alternative splicing events detected in the Diploid/Tetraploid parent vs SHW comparison in hypocotyl.



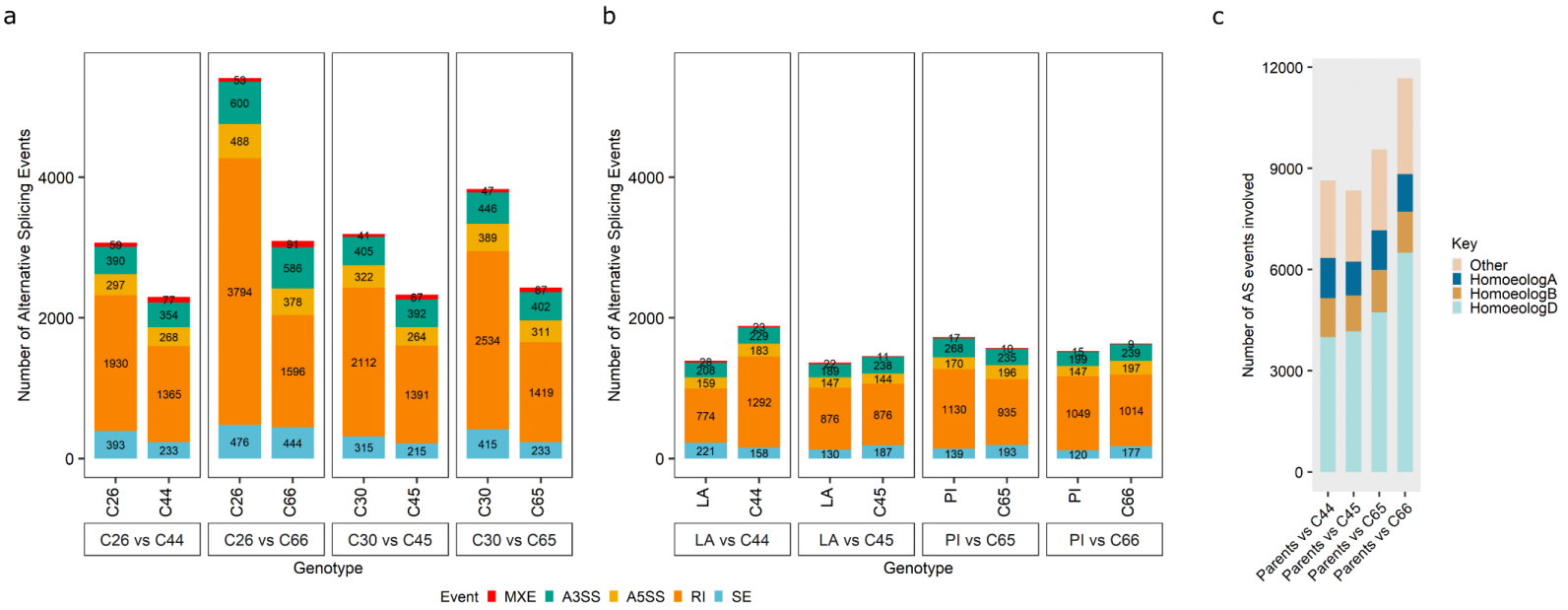
Appendix 32 Number of alternative splicing events detected in the Diploid/Tetraploid parent vs SHW comparison in pistil-when anthers are green and immature.



Appendix 33 Number of alternative splicing events detected in the Diploid/Tetraploid parent vs SHW comparison in pistil-when anthers are yellow and just prior to dehiscence.



Appendix 34 Number of alternative splicing events detected in the Diploid/Tetraploid parent vs SHW comparison in palea+lemma.



Appendix 35 Number of alternative splicing events detected in the Diploid/Tetraploid parent vs SHW comparison in root.

7.2 Supplementary materials for Chapter 3

Appendix 36 Summary of SNPs lifted from RefSeq v1.0 to RefSeq v2.1 using flanking regions of 500 bp length

Chr	Part1 500bp	Part2 500bp	Part3 500bp	Part4 500bp	Part5 500bp	Part6 500bp	Part7 500bp	Part8 500bp	Part9 500bp	Part10 500bp	Part11 500bp	Part12 500bp	Part13 500bp	Part14 500bp	Part15 500bp	Part16 500bp	
Chr1A	135,417	156	214	35	90	215	26	191	104	66	121	77	179	104	479	140	
Chr1B	61,422	120,067	62	245	29	174	122	148	103	159	79	205	61	170	911	565	
Chr1D	90	75,369	21,172	34	266	95	164	28	99	41	185	53	308	59	324	20	
Chr2A	153	55	147,859	87	191	194	45	133	146	70	94	76	141	174	875	440	
Chr2B	95	210	27,817	196,559	19,250	155	204	61	93	200	85	281	37	271	1,194	190	
Chr2D	60	232	185	99	133,643	132	233	34	238	92	255	52	295	65	733	161	
Chr3A	184	77	233	46	42,671	93,707	123	131	144	71	166	70	165	127	326	103	
Chr3B	85	269	32	272	92	102,216	85,516	99	134	204	46	255	71	215	109	1,694	
Chr3D	49	198	87	107	363	178	110,364	2,990	194	51	210	102	285	85	497	1,021	
Chr4A	119	56	142	128	233	106	211	125,409	155	41	103	62	135	173	185	750	
Chr4B	30	230	45	147	53	139	148	68,248	52,925	113	102	185	32	225	44	942	
Chr4D	62	196	88	21	231	57	175	41	71,284	40	137	36	364	52	1,136	1,773	
Chr5A	450	56	180	11	90	131	53	129	71,578	53,414	170	45	169	83	74	307	
Chr5B	71	218	67	212	32	189	157	94	113	143,375	21,995	278	34	244	63	303	
Chr5D	86	182	120	66	281	102	173	31	196	106	101,793	51	255	112	272	562	
Chr6A	129	53	92	23	70	112	26	105	101	44	72,166	35,064	155	111	38	2,127	
Chr6B	75	208	78	251	86	169	129	105	88	224	27	161,447	18,723	244	107	938	
Chr6D	48	144	85	39	198	49	160	60	142	31	179	111	78,862	37	1,305	3,998	
Chr7A	225	35	214	33	107	161	54	119	132	123	144	49	98,086	79,122	130	797	
Chr7B	52	286	86	349	60	227	176	169	195	231	80	281	56	116,751	59,109	1,568	
Chr7D	60	276	88	98	313	133	314	54	159	86	323	137	283	164	127,233	543	
ChrUnknown	54	11	47	31	222	100	34	297	65	74	108	49	30	52	3,657	20,796	
Actual number in input	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	39,822	3,039,822
Total lifted	199,016	198,584	198,993	198,893	198,571	198,741	198,607	198,676	198,388	198,856	198,568	198,966	198,726	198,640	198,801	39,738	3,020,764
Number of SNPs not lifted	984	1,416	1,007	1,107	1,429	1,259	1,393	1,324	1,612	1,144	1,432	1,034	1,274	1,360	1,199	84	19,058
Number of SNPs lifted to the same chromosome	196,839	195,436	196,848	196,559	195,564	195,923	195,880	196,647	195,787	196,789	195,954	196,511	195,671	195,873	189,999	20,796	2,957,076
Number of SNPs lifted to the other chromosomes	2,177	3,148	2,145	2,334	3,007	2,818	2,727	2,029	2,601	2,067	2,614	2,455	3,055	2,767	8,802	18,942	63,688

The green colour coded cells represent the chromosomes present in the input file. For instance, part1 contained SNPs from chromosomes 1A and 1B.

Appendix 37 Summary of SNPs lifted from RefSeq v1.0 to RefSeq v2.1 using flanking regions of 1000 bp length

Chr	Part1 1000bp	Part2 1000bp	Part3 1000bp	Part4 1000bp	Part5 1000bp	Part6 1000bp	Part7 1000bp	Part8 1000bp	Part9 1000bp	Part10 1000bp	Part11 1000bp	Part12 1000bp	Part13 1000bp	Part14 1000bp	Part15 1000bp	Part16 1000bp	
Chr1A	135,583	129	176	22	93	169	21	152	89	62	108	33	167	116	494	118	
Chr1B	61,373	120,177	20	222	21	181	132	123	35	151	44	210	44	145	926	533	
Chr1D	64	75,287	21,134	60	245	94	131	28	128	70	190	50	337	95	306	30	
Chr2A	122	54	148,007	117	152	152	33	130	139	80	98	70	131	166	872	441	
Chr2B	73	202	27,841	196,645	19,225	139	193	90	72	216	63	230	25	233	1,183	186	
Chr2D	60	249	155	143	133,675	113	250	47	178	98	242	74	288	98	725	147	
Chr3A	175	91	242	44	42,698	93,741	109	105	115	59	133	91	138	158	331	136	
Chr3B	98	180	29	232	82	102,190	85,458	85	113	167	49	215	71	195	94	1684	
Chr3D	55	198	93	70	306	187	110,375	2,968	165	55	263	133	301	121	518	1026	
Chr4A	93	50	132	109	258	143	206	125,541	144	42	74	77	124	117	140	753	
Chr4B	21	187	47	144	78	128	126	68,211	52,971	156	89	226	43	207	24	952	
Chr4D	73	263	79	17	228	68	205	57	71,335	44	216	39	382	67	1,127	1779	
Chr5A	426	41	178	9	97	110	29	98	71,744	53,320	126	51	193	78	75	317	
Chr5B	69	210	83	213	13	203	192	92	94	143,368	21,954	242	34	261	79	289	
Chr5D	67	214	143	70	282	120	156	17	181	74	101,726	40	321	102	300	556	
Chr6A	102	44	87	5	68	110	21	114	51	35	72,278	35,072	148	98	40	2119	
Chr6B	84	195	51	198	51	199	136	87	89	252	58	161,417	18,717	192	90	939	
Chr6D	55	161	84	24	188	52	178	32	129	17	217	85	78,783	41	1,285	3999	
Chr7A	163	77	132	39	98	143	35	97	107	118	128	49	98,026	79,118	110	807	
Chr7B	40	140	77	306	87	170	202	137	122	239	49	283	45	116,664	59,074	1590	
Chr7D	67	277	96	91	294	151	270	48	155	67	259	99	296	172	127,250	524	
ChrUnknown	61	6	25	23	217	66	26	287	71	43	98	48	13	39	3,664	20,810	
Actual number in input	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	39,822	3,039,822
Total lifted	198,924	198,432	198,911	198,803	198,456	198,629	198,484	198,546	198,227	198,733	198,462	198,834	198,627	198,483	198,707	39,735	3,018,993
Number of SNPs not lifted	1,076	1,568	1,089	1,197	1,544	1,371	1,516	1,454	1,773	1,267	1,538	1,166	1,373	1,517	1,293	87	20,829
Number of SNPs lifted to the same chromosome	196,956	195,464	196,982	196,645	195,598	195,931	195,833	196,720	196,050	196,688	195,958	196,489	195,526	195,782	189,988	20,810	2,957,420
Number of SNPs lifted to the other chromosomes	1,968	2,968	1,929	2,158	2,858	2,698	2,651	1,826	2,177	2,045	2,504	2,345	3,101	2,701	8,719	18,925	61,573

The green colour coded cells represent the chromosomes present in the input file. For instance, part1 contained SNPs from chromosomes 1A and 1B.

Appendix 38 Summary of SNPs lifted from RefSeq v1.0 to RefSeq v2.1 using flanking regions of 1250 bp length

Chr	Part1 1250bp	Part2 1250bp	Part3 1250bp	Part4 1250bp	Part5 1250bp	Part6 1250bp	Part7 1250bp	Part8 1250bp	Part9 1250bp	Part10 1250bp	Part11 1250bp	Part12 1250bp	Part13 1250bp	Part14 1250bp	Part15 1250bp	Part16 1250bp	
Chr1A	135,620	136	163	35	82	201	30	144	58	29	105	39	185	95	467	131	
Chr1B	61,346	120,187	23	241	43	160	155	128	44	142	70	212	62	161	975	537	
Chr1D	74	75,177	21,144	48	260	121	134	33	149	55	179	54	339	89	331	25	
Chr2A	145	59	148,025	83	129	161	42	117	126	92	72	58	88	164	860	438	
Chr2B	63	173	27,852	196,580	19,230	162	175	90	103	228	60	244	21	240	1,163	181	
Chr2D	69	192	154	116	133,590	117	232	45	185	116	189	59	373	99	726	161	
Chr3A	169	89	229	45	42,685	93,758	114	96	109	48	108	106	166	96	325	98	
Chr3B	82	169	22	239	108	101,962	85,469	73	115	185	44	246	59	172	100	1,696	
Chr3D	45	254	115	91	299	217	110,392	2,962	155	46	257	152	287	99	500	994	
Chr4A	107	45	142	97	250	134	188	125,467	162	34	97	94	98	110	147	759	
Chr4B	24	204	51	121	57	141	127	68,141	52,967	156	74	167	40	195	25	953	
Chr4D	84	236	89	32	219	64	188	49	71,285	24	201	64	335	97	1,185	1,779	
Chr5A	410	40	139	14	116	128	33	137	71,759	53,313	115	34	229	90	80	322	
Chr5B	68	229	51	229	18	197	130	116	58	143,421	21,959	220	29	247	77	306	
Chr5D	57	223	127	84	279	104	137	33	143	82	101,645	59	344	75	299	545	
Chr6A	93	38	124	4	68	82	21	108	58	58	72,298	35,064	154	96	51	2,126	
Chr6B	72	162	72	217	39	212	134	86	85	223	74	161,345	18,735	227	73	909	
Chr6D	28	154	75	31	235	48	177	36	115	12	247	91	78,727	52	1,255	4,025	
Chr7A	161	73	117	33	129	140	33	134	120	111	126	50	97,972	79,179	112	783	
Chr7B	40	160	43	287	43	201	212	151	127	198	48	289	57	116,588	59,097	1,554	
Chr7D	91	371	88	123	310	192	301	81	183	72	346	87	274	205	127,131	530	
ChrUnknown	54	2	23	21	211	63	20	281	71	43	96	48	13	52	3,682	20,879	
Actual number in input	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	39,822	3,039,822
Total lifted	198,902	198,373	198,868	198,771	198,400	198,565	198,444	198,508	198,177	198,688	198,410	198,782	198,587	198,428	198,661	39,731	3,018,295
Number of SNPs not lifted	1,098	1,627	1,132	1,229	1,600	1,435	1,556	1,492	1,823	1,312	1,590	1,218	1,413	1,572	1,339	91	21,527
Number of SNPs lifted to the same chromosome	196,966	195,364	197,021	196,580	195,505	195,720	195,861	196,570	196,011	196,734	195,902	196,409	195,434	195,767	189,910	20,879	2,956,633
Number of SNPs lifted to the other chromosomes	1,936	3,009	1,847	2,191	2,895	2,845	2,583	1,938	2,166	1,954	2,508	2,373	3,153	2,661	8,751	18,852	61,662

The green colour coded cells represent the chromosomes present in the input file. For instance, part1 contained SNPs from chromosomes 1A and 1B.

Appendix 39 Summary of SNPs lifted from RefSeq v1.0 to RefSeq v2.1 using flanking regions of 1980 bp length

Chr	Part1 1980bp	Part2 1980bp	Part3 1980bp	Part4 1980bp	Part5 1980bp	Part6 1980bp	Part7 1980bp	Part8 1980bp	Part9 1980bp	Part10 1980bp	Part11 1980bp	Part12 1980bp	Part13 1980bp	Part14 1980bp	Part15 1980bp	Part16 1980bp	
Chr1A	135,578	105	198	22	100	173	11	141	85	48	91	47	161	117	494	135	
Chr1B	61,349	120,109	40	231	35	180	155	160	85	130	57	185	79	143	929	530	
Chr1D	73	75,172	21,116	59	257	98	154	25	128	43	213	46	366	85	331	35	
Chr2A	154	54	147,938	70	147	133	61	107	124	60	102	62	99	139	848	426	
Chr2B	72	193	27,824	196,657	19,252	136	221	85	57	179	47	200	19	155	1,167	163	
Chr2D	87	187	150	87	133,523	125	237	34	157	108	187	81	311	124	735	164	
Chr3A	152	96	227	42	42,724	93,813	113	121	115	50	136	107	149	105	332	92	
Chr3B	97	163	57	191	85	102,133	85,375	78	76	160	57	279	77	170	115	1,704	
Chr3D	46	256	150	83	367	163	110,315	2,968	216	72	303	136	303	134	480	1,020	
Chr4A	114	47	116	52	249	98	211	125,516	137	28	102	71	117	100	104	755	
Chr4B	27	241	35	143	50	111	124	68,122	53,007	229	89	150	50	220	16	973	
Chr4D	64	221	66	42	141	78	186	38	71,174	20	182	55	302	117	1,163	1,784	
Chr5A	441	35	154	16	67	114	26	101	71,716	53,355	132	35	154	77	70	290	
Chr5B	53	184	47	259	15	203	101	55	74	143,269	21,965	193	24	174	52	277	
Chr5D	61	236	102	79	328	115	139	64	148	83	101,721	46	294	112	284	552	
Chr6A	81	25	132	10	33	88	25	100	90	63	72,213	35,057	134	81	55	2,129	
Chr6B	51	127	14	211	27	189	160	88	63	243	52	161,332	18,749	166	67	899	
Chr6D	37	148	83	29	245	56	193	41	141	25	181	76	78,676	42	1,285	3,997	
Chr7A	136	85	145	45	135	120	32	111	131	89	121	56	97,992	79,175	125	796	
Chr7B	68	164	77	229	52	178	215	128	112	228	49	348	38	116,705	59,080	1,571	
Chr7D	88	428	109	134	307	159	300	83	188	86	278	126	407	185	127,167	505	
ChrUnknown	25	7	35	21	210	29	15	277	53	37	70	34	8	29	3,688	20,895	
Actual number in input	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	200,000	39,822	3,039,822
Total lifted	198,854	198,283	198,815	198,712	198,349	198,492	198,369	198,443	198,077	198,605	198,348	198,722	198,509	198,355	198,587	39,692	3,017,212
Number of SNPs not lifted	1,146	1,717	1,185	1,288	1,651	1,508	1,631	1,557	1,923	1,395	1,652	1,278	1,491	1,645	1,413	130	22,610
Number of SNPs lifted to the same chromosome	196,927	195,281	196,878	196,657	195,499	195,946	195,690	196,606	195,897	196,624	195,899	196,389	195,417	195,880	189,935	20,895	2,956,420
Number of SNPs lifted to the other chromosomes	1,927	3,002	1,937	2,055	2,850	2,546	2,679	1,837	2,180	1,981	2,449	2,333	3,092	2,475	8,652	18,797	60,792

The green colour coded cells represent the chromosomes present in the input file. For instance, part1 contained SNPs from chromosomes 1A and 1B.

7.3 Supplementary materials for Chapter 4

Appendix 40 Number of COs in the BILs across the 21 chromosomes of the wheat genome.

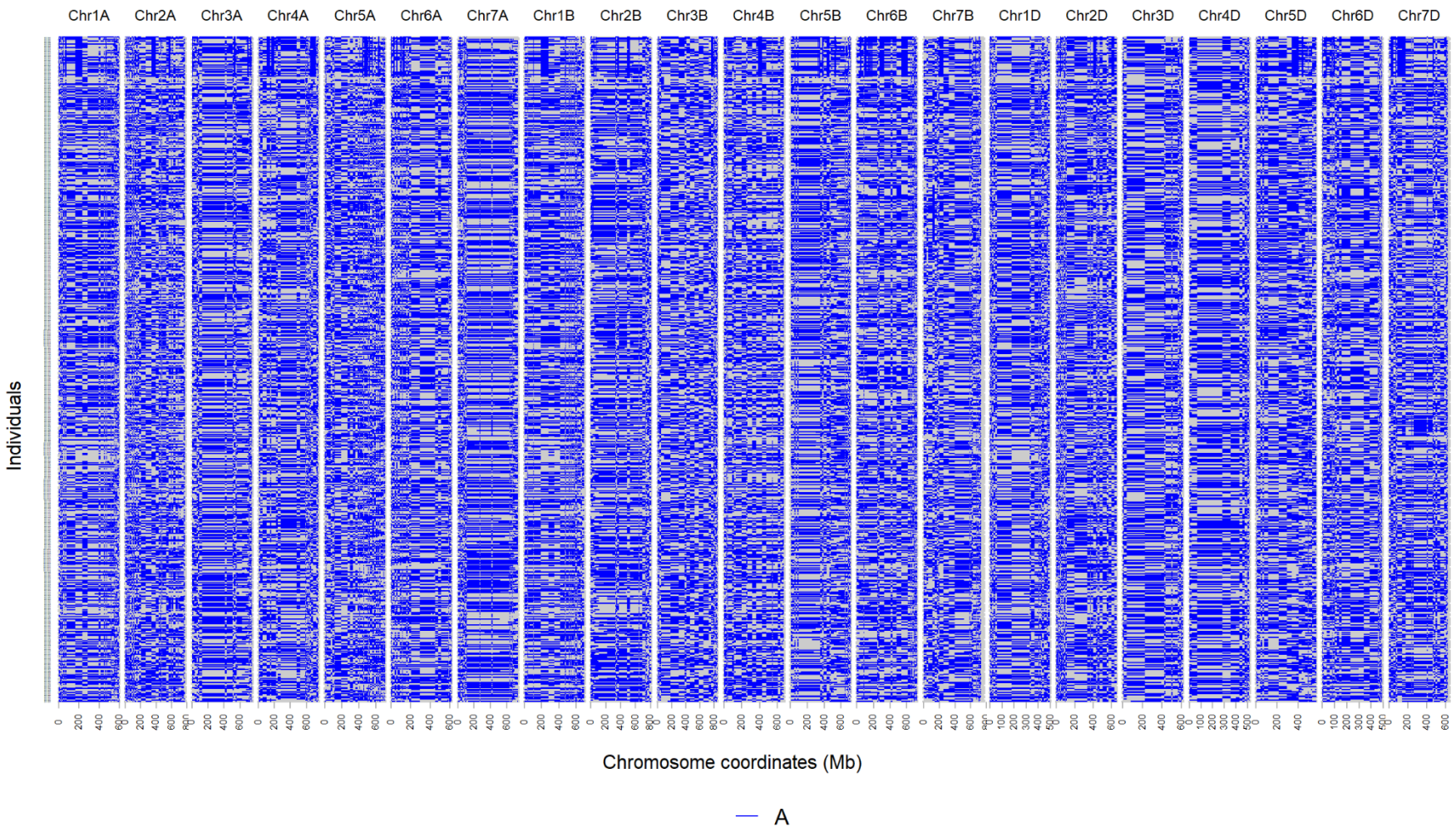
BILs	Chr1A	Chr1B	Chr1D	Chr2A	Chr2B	Chr2D	Chr3A	Chr3B	Chr3D	Chr4A	Chr4B	Chr4D	Chr5A	Chr5B	Chr5D	Chr6A	Chr6B	Chr6D	Chr7A	Chr7B	Chr7D	Total
18C66CAR_100	4	19	0	4	4	0	2	12	0	6	0	0	3	2	0	0	12	0	20	21	0	109
18C66CAR_101	4	1	0	6	5	14	0	45	0	12	0	1	1	3	0	0	12	0	20	5	17	146
18C66CAR_102	6	4	0	6	0	65	0	0	0	1	0	0	1	2	0	0	9	0	7	0	3	104
18C66CAR_103	3	0	0	44	10	4	0	28	1	3	0	0	1	17	0	0	5	0	4	19	4	143
18C66CAR_104	4	4	0	214	3	3	0	8	2	0	0	0	1	3	0	0	2	0	11	5	0	260
18C66CAR_105	4	0	0	28	26	3	2	45	0	3	0	0	2	6	0	23	8	0	6	0	0	156
18C66CAR_106	6	0	0	36	20	34	1	59	0	10	0	2	8	2	0	3	13	0	8	15	0	217
18C66CAR_107	10	0	0	6	0	20	0	1	0	1	0	0	4	15	0	9	14	0	9	19	3	111
18C66CAR_109	22	0	0	92	50	13	0	51	0	6	0	0	2	0	2	8	9	1	3	0	0	259
18C66CAR_11	16	0	0	21	14	0	0	0	0	5	0	0	2	4	2	22	17	0	16	55	5	179
18C66CAR_110	16	0	0	29	14	2	0	23	3	9	5	0	8	1	3	36	22	1	36	2	2	212
18C66CAR_111	18	0	0	51	3	19	0	8	3	0	0	0	0	11	2	22	44	2	1	2	0	186
18C66CAR_112	38	0	0	8	1	16	0	43	0	4	0	0	3	2	0	0	1	3	10	24	0	153
18C66CAR_113	9	0	0	13	11	2	0	38	2	5	0	0	7	5	2	0	2	1	5	17	0	119
18C66CAR_115	1	10	0	15	27	0	10	6	0	6	0	0	7	17	0	0	10	0	14	12	0	135
18C66CAR_116	2	4	0	4	7	0	6	10	0	3	0	0	3	24	0	0	17	0	10	9	0	99
18C66CAR_117	24	0	0	1	13	0	0	31	0	6	0	0	16	16	4	0	12	0	0	14	0	137
18C66CAR_118	11	1	0	2	11	0	0	16	0	2	0	0	8	9	2	0	8	0	2	13	0	85
18C66CAR_12	1	1	0	15	10	0	0	29	0	0	32	0	11	2	0	0	4	0	16	20	0	141
18C66CAR_120	7	2	0	6	13	0	1	15	0	3	0	0	8	8	0	0	15	0	50	20	0	148
18C66CAR_126	2	8	0	1	9	0	0	12	0	5	0	0	2	13	0	0	19	0	8	9	0	88
18C66CAR_127	2	17	0	11	14	0	8	8	0	4	0	0	8	21	0	0	9	0	29	8	0	139
18C66CAR_129	4	19	0	5	28	0	9	2	0	3	0	0	3	18	0	0	9	0	20	18	0	138
18C66CAR_13	1	7	0	5	14	0	12	2	0	5	0	0	10	16	0	0	13	0	31	17	0	133
18C66CAR_132	2	1	1	13	9	50	1	43	2	9	0	0	7	4	0	0	23	3	20	13	0	201
18C66CAR_133	18	4	0	11	5	0	3	34	0	3	0	0	1	0	3	0	4	8	3	8	0	105
18C66CAR_135	2	0	0	1	16	0	4	47	0	3	0	0	4	4	0	6	1	2	18	20	0	128
18C66CAR_137	10	0	0	42	17	0	0	46	0	5	0	0	4	4	2	4	14	0	1	16	0	165
18C66CAR_138	6	0	0	3	2	0	10	32	0	5	0	0	4	16	1	1	1	0	40	16	0	137
18C66CAR_139	2	2	0	2	14	0	3	18	0	10	0	0	7	6	10	0	10	0	25	22	0	131
18C66CAR_140	0	0	0	5	5	0	2	0	0	0	2	0	0	0	3	0	1	2	2	0	0	22
18C66CAR_142	34	0	0	95	2	2	1	4	3	2	3	0	0	0	2	11	0	2	0	0	0	161
18C66CAR_143	8	2	2	5	15	6	4	32	3	10	3	0	13	5	1	1	15	2	8	19	0	154
18C66CAR_144	0	4	1	0	0	7	2	1	3	2	19	0	0	4	0	0	0	1	0	0	2	46

BILs	Chr1A	Chr1B	Chr1D	Chr2A	Chr2B	Chr2D	Chr3A	Chr3B	Chr3D	Chr4A	Chr4B	Chr4D	Chr5A	Chr5B	Chr5D	Chr6A	Chr6B	Chr6D	Chr7A	Chr7B	Chr7D	Total
18C66CAR_146	0	3	0	0	0	3	2	3	3	5	0	0	0	1	1	0	0	1	0	0	2	24
18C66CAR_147	2	5	2	7	16	0	0	7	0	4	3	0	3	27	0	0	19	0	26	16	0	137
18C66CAR_148	0	4	0	2	0	2	2	4	2	0	0	0	0	2	2	12	0	0	0	0	0	32
18C66CAR_15	2	0	0	58	24	0	1	75	0	2	0	0	16	6	0	0	16	0	10	15	0	225
18C66CAR_150	13	1	0	11	19	0	4	36	3	29	2	0	3	4	0	0	2	0	37	16	0	180
18C66CAR_151	26	0	0	150	18	0	2	73	0	2	22	0	0	0	0	6	2	0	3	14	0	318
18C66CAR_157	10	1	2	120	32	2	2	75	0	0	0	0	0	5	0	15	2	0	11	13	5	295
18C66CAR_158	14	2	2	101	19	1	7	56	0	0	0	0	0	3	2	0	0	3	2	3	2	217
18C66CAR_159	3	10	0	6	22	0	9	11	0	9	10	0	3	7	0	0	8	0	7	14	0	119
18C66CAR_16	5	4	0	22	12	0	0	44	2	14	0	0	6	0	0	0	6	0	2	19	4	140
18C66CAR_160	23	4	0	29	10	0	4	30	0	6	0	0	3	3	3	0	0	3	11	19	4	152
18C66CAR_161	1	2	0	10	24	4	2	14	0	13	0	0	8	7	7	0	8	0	25	18	0	143
18C66CAR_162	9	5	2	13	0	2	5	0	0	0	2	0	0	2	5	0	0	3	5	4	1	58
18C66CAR_163	13	0	0	145	22	2	10	10	0	0	2	0	0	0	1	2	0	3	3	16	2	231
18C66CAR_164	2	0	0	30	17	0	4	39	0	23	1	0	12	4	1	14	12	0	35	16	1	211
18C66CAR_166	42	0	0	0	21	0	2	45	0	0	2	0	0	4	0	0	8	0	1	8	0	133
18C66CAR_167	2	2	2	12	10	0	2	16	3	8	12	2	2	7	0	0	16	0	31	18	0	145
18C66CAR_168	20	0	0	131	8	7	5	28	0	1	0	0	0	0	0	4	0	0	2	6	0	212
18C66CAR_169	20	0	2	163	20	2	9	44	4	1	1	0	0	7	0	0	1	0	36	4	0	314
18C66CAR_170	8	0	1	177	16	4	6	26	0	0	2	0	2	4	0	0	4	0	44	6	0	300
18C66CAR_171	22	0	0	70	10	10	2	36	3	0	0	0	1	0	0	0	3	0	15	0	0	172
18C66CAR_172	0	0	2	26	0	5	3	12	3	1	0	10	0	2	0	0	0	0	2	6	0	72
18C66CAR_173	4	14	0	12	22	21	1	23	1	6	6	0	15	1	6	0	15	0	22	30	0	199
18C66CAR_174	1	2	0	3	27	0	1	14	0	9	6	0	2	2	0	0	10	0	27	19	0	123
18C66CAR_175	51	2	0	55	20	12	0	46	2	0	9	0	0	2	6	0	0	0	5	9	1	220
18C66CAR_177	41	1	0	3	40	16	0	55	1	0	0	0	0	1	2	14	1	0	5	10	4	194
18C66CAR_178	26	13	0	3	1	4	0	42	0	0	9	1	0	0	3	6	3	0	3	5	0	119
18C66CAR_179	18	3	0	4	3	3	0	73	2	7	0	0	9	2	0	2	7	0	15	16	1	165
18C66CAR_182	33	1	0	40	1	3	1	3	0	1	0	0	2	3	0	13	2	0	1	1	0	105
18C66CAR_183	1	0	0	0	2	22	2	2	2	2	0	0	2	12	1	37	3	0	0	0	0	88
18C66CAR_184	3	2	0	20	7	2	3	2	0	26	0	0	8	8	2	0	6	0	21	10	0	120
18C66CAR_186	28	1	0	2	20	4	1	1	2	2	0	0	1	4	0	6	0	0	4	12	0	88
18C66CAR_187	2	13	0	13	3	0	21	1	2	13	0	0	6	4	0	0	4	4	22	21	0	129
18C66CAR_189	19	3	0	33	14	0	0	51	2	2	2	0	1	0	4	6	8	0	5	2	0	152
18C66CAR_19	5	2	0	12	5	11	0	48	1	4	0	0	5	2	0	23	10	0	28	22	26	204
18C66CAR_192	6	2	0	8	12	0	3	35	0	43	16	0	3	2	2	22	3	0	32	5	0	194
18C66CAR_193	8	1	0	63	4	0	2	58	0	2	0	0	1	1	1	0	0	0	3	18	0	162
18C66CAR_194	1	12	0	11	12	0	3	12	2	18	0	0	3	8	0	0	23	0	17	10	0	132

BILs	Chr1A	Chr1B	Chr1D	Chr2A	Chr2B	Chr2D	Chr3A	Chr3B	Chr3D	Chr4A	Chr4B	Chr4D	Chr5A	Chr5B	Chr5D	Chr6A	Chr6B	Chr6D	Chr7A	Chr7B	Chr7D	Total
18C66CAR_195	1	2	0	3	24	0	1	14	0	7	5	0	24	13	0	0	15	0	20	14	0	143
18C66CAR_2	15	4	0	5	25	0	1	15	0	10	0	0	2	30	0	0	14	0	22	14	4	161
18C66CAR_202	31	2	0	179	36	0	4	32	1	0	1	0	0	1	2	5	0	2	2	5	0	303
18C66CAR_203	45	1	0	170	20	0	2	56	1	0	2	0	4	0	1	13	18	2	1	13	0	349
18C66CAR_205	0	0	0	0	16	11	1	78	0	5	2	0	4	3	0	8	0	3	0	8	11	150
18C66CAR_209	9	0	0	160	29	3	2	36	0	2	0	0	4	0	3	11	1	2	0	5	0	267
18C66CAR_21	12	7	0	3	34	0	3	35	0	8	0	2	8	3	0	4	9	0	32	21	1	182
18C66CAR_22	0	0	0	52	2	20	1	59	1	0	0	0	3	6	3	18	5	1	8	8	2	189
18C66CAR_220	18	0	0	1	46	0	2	48	0	1	0	0	12	1	0	26	5	1	4	5	3	173
18C66CAR_221	22	0	1	3	36	5	1	0	0	17	0	0	4	4	0	24	3	0	0	17	0	137
18C66CAR_222	17	0	0	12	10	19	0	40	0	3	0	0	2	6	4	22	1	2	4	23	2	167
18C66CAR_224	2	0	2	5	10	6	2	0	0	32	0	0	6	7	0	0	13	1	31	9	0	126
18C66CAR_227	2	4	0	3	10	0	2	8	0	6	0	0	7	17	0	0	21	0	16	7	0	103
18C66CAR_228	4	0	0	5	21	0	0	16	0	22	0	0	24	7	0	0	7	1	25	7	0	139
18C66CAR_229	6	0	0	90	0	2	0	0	0	0	0	0	0	2	1	44	3	0	3	0	4	155
18C66CAR_23	2	6	0	4	24	0	3	5	0	5	6	0	4	9	0	0	15	0	27	15	0	125
18C66CAR_230	60	0	1	207	44	5	0	53	0	0	0	0	3	0	2	0	1	2	3	1	0	382
18C66CAR_232	19	0	0	0	16	4	0	10	0	2	0	0	4	8	4	21	6	1	47	15	6	163
18C66CAR_234	12	0	0	2	7	43	1	50	0	8	0	0	0	1	1	0	4	3	2	2	0	136
18C66CAR_235	7	4	0	27	4	24	2	1	0	12	0	0	2	6	0	5	15	0	15	14	0	138
18C66CAR_236	2	23	0	3	5	1	0	2	0	2	0	17	4	13	0	0	13	0	11	14	0	110
18C66CAR_237	74	0	0	77	38	9	2	2	0	16	0	0	0	5	0	0	0	0	4	14	2	243
18C66CAR_238	9	9	0	16	7	0	15	7	0	16	0	0	5	11	0	0	18	0	14	12	0	139
18C66CAR_239	10	3	0	112	47	9	2	2	0	4	0	0	2	3	0	1	8	0	2	6	0	211
18C66CAR_24	8	0	0	18	24	2	5	48	2	7	0	0	1	7	3	0	18	2	6	14	0	165
18C66CAR_240	5	3	0	64	8	5	0	14	4	2	0	1	0	3	0	14	1	0	1	2	0	127
18C66CAR_241	0	1	0	8	13	14	3	23	0	7	0	0	15	4	1	0	11	0	6	2	4	112
18C66CAR_242	6	12	0	7	33	14	8	3	3	24	0	0	12	14	1	17	10	0	32	34	2	232
18C66CAR_243	1	8	0	6	10	0	2	33	2	16	0	0	8	2	1	7	15	0	9	21	12	153
18C66CAR_244	4	0	0	14	5	20	6	25	3	3	0	3	4	9	0	0	16	0	28	12	2	154
18C66CAR_245	0	13	0	9	5	4	3	1	0	9	0	0	10	10	0	20	8	0	1	13	0	106
18C66CAR_246	2	10	0	152	5	3	10	15	3	1	0	0	6	1	0	13	2	0	37	2	2	264
18C66CAR_247	6	1	0	7	16	2	0	2	0	9	0	1	17	8	0	0	11	0	21	23	0	124
18C66CAR_248	1	1	0	23	7	0	0	39	2	1	0	0	14	7	0	22	4	0	15	6	8	150
18C66CAR_249	7	0	0	21	10	2	1	15	2	3	0	0	20	24	1	20	9	0	1	15	0	151
18C66CAR_25	1	2	0	10	12	0	3	13	5	4	0	0	5	6	0	0	11	1	30	15	0	118
18C66CAR_250	5	0	0	10	18	1	3	35	0	10	0	0	20	21	1	1	19	0	25	12	17	198
18C66CAR_26	0	0	0	36	4	48	0	34	4	12	1	0	4	2	0	0	26	2	21	15	0	209

BILs	Chr1A	Chr1B	Chr1D	Chr2A	Chr2B	Chr2D	Chr3A	Chr3B	Chr3D	Chr4A	Chr4B	Chr4D	Chr5A	Chr5B	Chr5D	Chr6A	Chr6B	Chr6D	Chr7A	Chr7B	Chr7D	Total
18C66CAR_27	6	2	0	29	2	2	4	31	0	8	0	0	11	14	2	25	4	8	6	15	0	169
18C66CAR_29	2	2	0	9	22	3	6	39	0	25	0	0	7	4	0	3	12	0	19	16	37	206
18C66CAR_3	30	2	0	45	9	0	1	54	0	13	3	0	3	3	2	34	15	0	8	19	3	244
18C66CAR_30	2	22	0	6	8	0	4	8	0	3	0	0	4	12	0	0	11	0	14	13	0	107
18C66CAR_31	13	10	0	9	18	0	2	33	0	3	0	0	1	6	2	0	35	3	40	18	0	193
18C66CAR_32	1	4	0	5	28	0	5	10	0	0	0	0	6	16	0	0	18	0	21	6	0	120
18C66CAR_33	2	2	1	16	16	1	5	35	0	3	4	0	14	11	0	5	12	0	41	19	3	190
18C66CAR_39	2	10	0	6	20	0	6	6	0	7	0	0	7	14	0	0	10	0	20	14	0	122
18C66CAR_42	6	0	0	0	6	0	0	46	2	0	2	0	6	2	1	33	15	0	8	11	0	138
18C66CAR_43	10	0	1	21	29	13	0	74	0	6	0	0	0	1	3	27	4	2	6	31	0	228
18C66CAR_44	13	5	2	99	44	41	3	62	0	14	0	0	0	1	1	35	6	0	4	5	5	340
18C66CAR_45	37	1	1	37	33	2	2	34	2	2	0	0	0	10	1	18	4	3	3	13	1	204
18C66CAR_46	1	2	0	3	19	2	13	26	0	3	0	0	4	15	0	0	8	0	16	14	0	126
18C66CAR_47	4	9	0	4	10	0	5	18	0	12	0	0	15	19	0	0	15	0	22	12	0	145
18C66CAR_48	8	1	0	13	0	60	2	0	0	10	0	0	3	1	1	0	1	0	6	17	2	125
18C66CAR_49	11	1	0	12	12	30	6	0	0	4	2	0	8	5	2	11	3	0	21	26	3	157
18C66CAR_5	8	0	1	21	15	2	0	12	0	8	0	0	5	7	1	18	4	0	6	26	0	134
18C66CAR_50	21	0	0	6	1	5	3	7	4	4	0	0	8	2	0	25	2	0	3	12	5	108
18C66CAR_51	4	0	0	2	24	4	2	0	4	5	0	0	3	1	2	0	2	0	8	13	3	77
18C66CAR_52	32	1	0	93	0	26	4	0	0	2	0	4	0	5	0	7	1	0	2	13	2	192
18C66CAR_53	7	1	0	28	0	47	0	42	1	9	0	0	0	7	1	0	17	0	8	13	1	182
18C66CAR_54	0	1	0	16	6	0	0	64	0	11	1	0	6	5	1	9	5	0	15	7	4	151
18C66CAR_55	3	2	0	15	13	0	2	42	0	19	2	0	6	7	0	0	15	0	24	14	0	164
18C66CAR_56	25	1	0	31	4	5	0	54	6	14	1	0	2	0	0	29	9	0	6	21	0	208
18C66CAR_57	0	2	0	12	0	0	0	49	1	8	0	0	20	8	3	0	24	0	27	22	4	180
18C66CAR_58	2	0	1	13	6	0	0	55	2	6	2	0	1	0	0	24	10	0	17	22	0	161
18C66CAR_59	1	26	2	4	18	0	0	11	8	3	8	0	3	27	0	0	16	0	9	22	0	158
18C66CAR_60	2	33	0	3	22	0	0	7	0	8	5	0	3	7	0	0	21	0	18	15	0	144
18C66CAR_61	8	5	2	126	26	0	0	21	2	3	0	0	2	5	0	12	8	0	3	5	0	228
18C66CAR_63	20	1	0	57	18	0	0	44	1	21	2	0	2	0	2	6	11	0	2	6	0	193
18C66CAR_64	7	0	0	28	10	66	7	26	1	2	0	0	10	4	2	8	12	0	3	19	0	205
18C66CAR_65	16	0	0	121	44	5	5	4	4	7	0	0	4	2	0	20	12	0	0	2	0	246
18C66CAR_66	10	0	0	18	8	38	2	53	1	5	0	0	13	0	0	16	14	0	2	2	0	182
18C66CAR_68	23	0	0	34	14	3	6	3	0	11	0	0	5	1	2	0	8	0	0	1	0	111
18C66CAR_69	17	0	0	40	5	23	4	2	1	3	0	0	5	1	8	0	16	0	7	16	0	148
18C66CAR_70	1	0	0	5	9	0	27	19	1	4	0	0	2	10	6	11	12	0	22	1	0	130
18C66CAR_71	2	2	0	9	14	0	3	23	0	6	0	0	14	3	0	0	12	0	20	12	0	120
18C66CAR_73	2	11	0	10	22	0	2	14	0	10	0	0	3	16	0	0	13	0	20	9	0	132

BILs	Chr1A	Chr1B	Chr1D	Chr2A	Chr2B	Chr2D	Chr3A	Chr3B	Chr3D	Chr4A	Chr4B	Chr4D	Chr5A	Chr5B	Chr5D	Chr6A	Chr6B	Chr6D	Chr7A	Chr7B	Chr7D	Total
18C66CAR_74	1	3	0	3	16	0	0	8	0	8	8	0	2	4	0	0	10	0	20	13	0	96
18C66CAR_75	10	0	3	20	21	67	0	49	1	8	0	0	7	0	0	21	2	1	15	32	0	257
18C66CAR_76	2	0	0	3	7	2	0	31	2	16	0	1	9	1	2	18	12	2	26	16	0	150
18C66CAR_79	11	0	1	3	15	0	0	93	5	5	1	0	4	2	1	6	2	4	6	21	0	180
18C66CAR_8	2	2	4	2	16	0	2	7	0	8	8	0	10	19	0	0	12	0	16	21	0	129
18C66CAR_80	1	1	0	14	15	0	0	24	1	4	1	0	10	4	10	0	7	1	32	15	0	140
18C66CAR_81	7	1	0	176	46	0	0	64	1	0	0	0	2	4	0	0	2	1	0	13	0	317
18C66CAR_83	4	3	0	9	1	2	0	28	0	15	1	0	9	10	0	3	16	0	20	11	0	132
18C66CAR_84	38	1	0	61	37	8	1	66	0	1	2	0	4	2	0	2	0	1	3	2	4	233
18C66CAR_85	6	6	0	11	25	0	1	16	0	6	0	0	7	19	0	0	14	0	29	11	0	151
18C66CAR_86	9	1	0	92	26	5	0	71	0	5	2	0	4	4	0	0	6	1	6	18	0	250
18C66CAR_87	25	6	0	41	16	23	0	60	1	15	0	0	6	10	0	0	0	1	3	27	2	236
18C66CAR_88	3	3	0	7	23	0	2	11	0	28	7	0	3	7	0	0	12	15	28	45	4	198
18C66CAR_89	2	1	0	6	14	3	1	13	2	23	0	0	28	8	0	0	21	13	12	20	0	167
18C66CAR_9	1	0	0	3	12	0	0	9	0	12	0	2	1	2	1	0	1	6	27	8	43	128
18C66CAR_90	7	10	0	17	0	4	0	46	1	24	2	0	2	2	0	21	14	1	16	14	0	181
18C66CAR_91	2	0	0	0	24	0	0	23	5	16	0	0	4	2	6	2	1	4	39	13	17	158
18C66CAR_92	8	2	5	5	10	0	1	39	1	7	0	0	0	6	1	11	15	1	3	15	1	131
18C66CAR_93	19	0	0	4	16	49	0	70	0	6	0	0	2	15	2	49	0	2	8	26	2	270
18C66CAR_94	2	0	1	7	25	0	3	24	2	20	0	1	9	2	0	0	11	31	26	24	17	205
18C66CAR_95	4	0	1	20	14	0	7	63	10	6	0	0	2	2	2	29	10	2	12	21	4	209
18C66CAR_96	10	4	0	8	23	0	5	17	0	10	0	0	2	4	1	5	7	2	15	17	3	133
18C66CAR_97	2	2	0	3	10	1	7	14	2	15	0	6	2	10	0	0	14	0	14	21	2	125
18C66CAR_98	1	6	0	12	24	0	7	6	0	3	0	0	7	26	0	0	11	0	20	14	0	137
18C66CAR_99	4	17	0	6	16	0	6	4	0	3	0	0	5	16	0	0	9	0	18	7	0	111



Appendix 41 The genomic composition of the A subgenome of the doubled haploid population individuals. The blue regions represent the homozygous AC Cadillac fragments (A).