Understanding the introgression process from *Aegilops tauschii* into hexaploid wheat through identity by descent analysis and its effect on genetic diversity

Moses Nyine[1], Elina Adhikari[1], Marshall Clinesmith[2], Katherine Jordan[1], Allan K. Fritz[2*], Eduard Akhunov[1*]

[1] Department of Plant Pathology, Kansas State University, Manhattan, KS 66506
[2] Department of Agronomy, Kansas State University, Manhattan, KS 66506

*Corresponding authors: Allan K. Fritz (akf@ksu.edu); Eduard Akhunov (eakhunov@ksu.edu)

Key words: *Aegilops tauschii,* domestication, genetic diversity, hexaploid wheat, identity by descent, imputation, introgression

Abstract

Wild relatives of hexaploid wheat (*Triticum aestivum*) are the reservoirs of novel allelic diversity with great potential to improve many agronomic traits in wheat. Here, we investigated the genome-wide patterns and efficiency of *Aegilops tauschii* allele introgression into the winter wheat cultivars. The introgression population of 351 $BC_1F_{3:5}$ lines was selected based on phenology and development characteristics from crosses between six hexaploid wheat lines and 21 wheat-*Ae. tauschii* octoploids. Complexity reduced genomic library sequencing was used to develop SNP markers and infer the regions of identity-by-descent and the boundaries of the introgressed segments. Using a diverse panel of 116 *Ae. tauschii* accessions, it was possible to infer that introgression lines had single or multiple IBD segments from accessions of diverse geographic origin. Introgression frequency was high at the ends of chromosomes and low in the large pericentromeric 2/3 of the chromosome arms characterized by low crossover rate. While the effect of selection for free-threshing genotypes was evident around the domestication gene *Tg*, reduction in the frequency of introgression was limited to relatively small regions flanking the gene. These results suggest that the effects of phenotypic selection on the introgressed wild relative's alleles at the early generations of population development are strongly influenced by the distribution of crossover frequency across genome, consistent with the Hill-Robertson effect. Our study offers insights into the introgression population development to ensure retention of genetic

diversity across entire genome and presents a resource that will be valuable for deploying wild relative diversity in breeding programs to create climate resilient and disease resistant varieties with improved yield and quality traits.

Introduction

Wheat production is constrained by several biotic and abiotic factors, yet the demand for wheat is expected to double by 2050. A yield increase of ~2.4 % per year has been projected as required to close the gap between the current production level and an increasing demand (Ray *et al.* 2013). While this goal can be achieved by improving agronomic practices, expanding the production area and/or deployment of high-yielding wheat varieties, the first two alternatives are unsustainable because land is a limited resource and most effective agronomic practices are costly. Accelerated wheat improvement through extensive deployment of available genomics tools and genetic resources, including close and distant wild relatives of wheat, is viewed as the most effective and sustainable alternative to increasing yield.

Allohexaploid wheat, *Triticum aestivum* (2n = 6x = 42, AABBDD) resulted from hybridization of *Triticum turgidum* (2n = 4x = 28, AABB) and *Aegilops tauschii* ssp. *strangulata* (2n = 2x = 14, DD) (Kihara 1944; Luo *et al.* 2007; Wang *et al.* 2013). Domestication of wheat followed by continuous selection by early farmers led to the 'domestication syndrome' that resulted in fixation or loss of alleles from the populations of wild relatives (Peng *et al.* 2003; Haudry *et al.* 2007; Dvorak *et al.* 2012). The wild diploid and tetraploid relatives that carry homoeologous genomes such as *Triticum turgidum* ssp. *dicoccoides* (AB genomes), *T. monococcum* (A genome), *T. urartu* (A genome), *Ae. tauschii* ssp. *tauschii* (D genome), and *Ae. tauschii* ssp. s*trangulat*a (D genome) are the primary sources of genes for improvement of common wheat (Gill and Raupp 1987; Qi *et al.* 2007). Secondary sources of resistance genes are close relatives of hexaploid wheat such as *T. timopheevii* (2n = 4x = 28, AAGG) and *Ae. speltoides* (2n = 2x = 14, SS). Introgression of beneficial alleles from these wild relatives was achieved by homoeologous recombination between the chromosomes of common wheat and wild relatives in the absence of *Ph1* gene controlling the pairing between homoeologs (Sears 1977). Introgression involves either direct crosses between common wheat and wild relatives or crosses between common wheat and synthetic wheat lines that are generated by hybridizing tetraploid wheat and wild diploids (Qi *et al.* 2007; Ogbonnaya *et al.* 2013).

Direct crossing of wild relatives to common wheat followed by backcrosses to the recurrent common wheat parent has been reported as a faster approach for introducing traits (Alonso and Kimber 1984; Gill and Raupp 1987). This approach was successfully used to transfer resistance to Hessian fly, greenbug and leaf rust into wheat (Gill and Raupp 1987). The D genome from *Ae. tauschii* has been associated with important adaptation traits such as drought and salinity stress tolerance, increased yield by influencing various yield components as well as grain, flour and dough quality (Ogbonnaya *et al.* 2013; Jones *et al.* 2013). Several genes that confer resistance to stem rust and leaf rust have been identified on D chromosomes in bread wheat landraces or their relatives (Liu *et al.* 2013; Periyannan *et al.* 2013). Other members of the *Triticeae* family have been utilized as a tertiary genetic pool for wheat improvement via non-homoeologous recombination (alien introgression). Many major disease resistance and environmental adaptation genes have been introgressed into wheat by translocation of chromosome segments and addition of full chromosomes or chromosome arms from rye, *Agropyron*, *Ae. ventricosa* and other species (Ayliffe *et al.* 2008; Liu *et al.* 2011; McIntosh *et al.* 2015; Cruz *et al.* 2016).

Improvement of wheat via alien genome introgression and homoeologous recombination can be challenging due to linkage drag of unwanted alleles from the wild relatives' genomes that can negatively impact agronomic traits (Anugrahwati et al. 2008). Genes affecting plant growth, development and domestication traits, such as dwarfing gene *Rht1,* photoperiod response gene *Ppd-D1*, tenacious glume gene *Tg*, and domestication gene Q have been mapped on chromosome arms 4DS, 2DS, 2BS and 5AL (Peng *et al.* 1999; Jantasuriyarat *et al.* 2004; Simons *et al.* 2006; Beales *et al.* 2007; Sood *et al.* 2009). Quantitative trait loci (QTL) from chromosome 4A, 4B, 6B and 7B have also been reported to confer free-threshability in wheat recombinant inbred lines (Jantasuriyarat *et al.* 2004; Peleg *et al.* 2011). The analyses of SNPs around genes affecting domestication traits showed substantial reduction of genetic diversity, which was indicative of strong positive selection for the domesticated allelic variants (He *et al.* 2019; Wang *et al.* 2019).

Genotyping approaches based on next-generation sequencing of complexity-reduced genomic libraries substantially accelerated analysis of genetic diversity in large crop genomes (Elshire *et al.* 2011; Saintenac *et al.* 2011; Poland *et al.* 2012; Saintenac *et al.* 2013; Jordan *et al.* 2015, 2018). The high proportion of missing data in low-coverage sequencing datasets was compensated by the availability of the whole genome sequence (The International Wheat Genome Sequencing Consortium (IWGSC) 2018) that facilitated accurate genotype imputation. Imputation of ungenotyped SNP markers from a reference panel into a target

population takes advantage of regions of identity-by-descent (IBD), thus allowing the interpolation of SNPs into the target population (Browning and Browning 2013). The power and resolution of association studies have been shown to improve after imputation (Browning and Browning 2012; Jordan *et al.* 2015; Nyine *et al.* 2019).

In this study, we developed the populations of winter wheat lines carrying introgression from a diverse set of *Ae. tauschii* accessions selected to represent broad genetic and geographic diversity of the species. The boundaries of introgressed segments in wheat genome were detected using the IBD analyses based on the SNP datasets generated by complexity-reduced sequencing of 378 introgression population lines and 116 *Ae. tauschii* accessions. The distribution of introgressed segments across the genome was investigated to assess its overall effect on genetic diversity, and evaluate the impact of recombination rate variation and early selection for uniform phenological and developmental characteristics on the introgression frequency in different parts of the wheat genome. The effect of selection against non-adaptive traits contributed by *Ae. tauschii* was investigated around the domestication gene *Tg* controlling tenacious glume trait (Sood *et al.* 2009).

Materials and methods

The study population consisted of 351 $BC_1F_{3:5}$ *Ae. tauschii* introgression lines developed by crossing synthetic *Ae. tauschii*-wheat octoploid lines with hexaploid wheat recurrent parents. The octoploid lines were developed by crossing six hexaploid wheat parents with 21 *Ae. tauschii* accessions (Supporting Information Table S1). The resulting F_1 hybrid plants regenerated from rescued embryos were treated with colchicine to generate the synthetic octoploids (Dale *et al.* 2017). The synthetic octoploids were then backcrossed once to the respective hexaploid wheat parents or to another wheat line. The BC_1F_1 plants were selfed and advanced by single seed descent to the BC_1F_3 generation. Seeds from individual BC_1F_3 plants were bulked and grown in single rows in the field at the Kansas State University Ashland Research Farm near Manhattan, KS in the 2016-17 growing season. Thirty-one families were represented in this material. The number of lines per family ranged from 42 to 137 and resulted in a total of 2,861 lines that were planted. The 351 lines used in this research were selected from this set of materials. Selection criteria included production of sufficient seed to allow yield testing, general fitness, threshability to allow mechanical harvest and phenology similar to the elite hexaploid parent(s). In addition, 116 diverse *Ae. tauschii* accessions representing *Ae. tauschii* ssp. *tauschii* and *Ae. tauschii* ssp. *strangulata* from

different geographical locations were used as the reference panel in the study (Supporting Information Table S2).

Sequencing complexity-reduced genomic libraries

DNA from *Ae. tauschii* introgression population and the reference panel samples was extracted using DNeasy 96 Plant DNA extraction kit (Qiagen) following the manufacture's protocol. The quality and concentration of the DNA was assessed using PicoGreen® dsDNA assay kit (Life Technologies). Input DNA was normalized to 400 ng (20ul of 20ng/ul) using Qiagility robot (Qiagen). Genotyping by sequencing (GBS) libraries were constructed using the modified protocol previously described by Saintenac *et al.* (2013), and subjected to size selection using Pippin Prep system (Sage Scientific) to enrich for 270-330 bp fragments. In total, five libraries were produced, representing 80 barcoded accessions each. Each library was sequenced on Illumina NextSeq 500 using a 1 x 75 bp kit for the introgression lines and 1 x 100 bp kit for the reference panel following the Illumina protocol. TASSEL 5.0 GBS v2 pipeline (Glaubitz *et al.* 2014) was used to generate SNPs from the fastq files of the introgression lines and the reference panel. In brief, the raw GBS sequence reads were aligned to the Chinese Spring reference sequence v1.0 (The International Wheat Genome Sequencing Consortium (IWGSC) 2018) using Burrow's Wheeler Alignment (BWA) software. TASSEL 5.0 GBS v2 default parameters were used in all steps (Glaubitz *et al.* 2014).

SNP genotyping and imputation

SNPs for the reference *Ae. tauschii* panel with minor allele frequency (MAF) less than 0.02 and maximum missingness greater than 70 % were filtered out using vcf-filter tools. The missing SNPs were imputed using the program Beagle v.5.0 (Browning and Browning 2013) with default parameters (File S1). SNPs from *Ae. tauschii* derived introgression population were filtered in two steps. First, SNPs from all subgenomes (A, B and D) with minor allele frequency (MAF) less than 0.05 and maximum missingness greater than 30 % were filtered out using vcf-filter tools. The missing SNP were imputed using the program Beagle v.5.0 with default parameters. In the next step, all A and B genome SNPs, and D genome SNPs with MAF less than 0.01 were excluded from the raw vcf file using vcf-filter tools. The program conform-gt (https://faculty.washington.edu/browning/conform-gt.html) was used to check the concordance of D genome SNP positions between the introgression population and

the reference panel based on the Chinese Spring genome coordinates (IWGSC, 2018). Missing and ungenotyped SNPs in the D genome of the introgression population were imputed from the reference panel using Beagle v.5.0 (File S2).

Principal component analysis (PCA)

The population structure of the diverse *Ae. tauschii* accessions and the introgression population was analyzed using the 11,624 D genome SNPs segregating in both populations (File S3, S4). SNP dataset was converted to the hapmap format and imported into TASSEL v.5.0, which was used to calculate the principal components. The first two components were plotted to show the distribution and clustering of the reference panel accessions in relation to the 21 parental *Ae. tauschii* accessions and the entire introgression population. In addition, a total of 13,719 SNPs (File S5), including 4,016, 4,142, 5,112 and 449 from A, B, D genomes and unanchored scaffolds, respectively, were used to evaluate the distribution of *Ae. tauschii*-derived introgression lines on the first two principal components using wheat parents as grouping factors.

Genetic diversity

To evaluate the effect of introgression on genetic diversity, the mean number of base differences for each SNP site in all pairwise comparisons (π) among *Ae. tauschii* accessions, introgression lines and hexaploid wheat lines were calculated using vcftools and summarized in R (R Development Core Team 2011). The π values for each chromosome were interpolated using the R function 'approx' (method='linear', rule=1) and plotted using R package 'ggplot2'.

Recombination hotspots

The imputed D-genome SNPs were split into subsets based on families. A combination of custom Perl and R scripts (Nyine *et al.* 2018), were used to convert the SNP alleles to 0, 1, and 2, of which, 0 is homozygous major allele, 1 is heterozygous and 2 is homozygous minor allele. Regions containing monomorphic SNPs were eliminated by the R script. A total of 16 families each having at least 10 progenies plus the respective parents were used in this analysis. A separate custom Perl script was used to count the number of allele phase transitions in each chromosome per individual and recode the flanking SNP

positions as break points (Jordan *et al.* 2018). The number of recombination breakpoints (RBP) per 10 kb sliding window in each chromosome per family was obtained using bedmap option from BEDOPS v2.4.35 (Neph *et al.* 2012). The total RBP per 10 kb window across the 16 families was obtained and the 99th percentile plotted using R-package ggplot2. All windows with total RBP below the 99th percentile of recombination events were masked by adding 0 before the line plots were generated. The centromere position in each chromosome was marked based on the Chinese Spring reference genome (The International Wheat Genome Sequencing Consortium (IWGSC) 2018; Su *et al.* 2019). Kruskal Wallis test was used to test for significant differences in the distribution of recombination breakpoints in each family.

In order to investigate the effect of sequence divergence and structural re-arrangements on recombination, we compared hexaploid wheat (Chinese Spring) and the diploid relative, *Ae. tauschii* ssp. *strangulata* (AL8/79) D genomes at protein level. High confidence D genome gene protein sequences from Chinese Spring v.1.0 and *Ae. tauschii* v.4.0 (Luo *et al.* 2017) were used. The annotation of the *Ae. tauschii* genome was downloaded from http://aegilops.wheat.ucdavis.edu/ATGSP/annotation/. Local protein BLAST databases were created for each dataset using BLAST2+. Reciprocal blastp was performed between the two species' genome proteins using default parameters. A Perl script was used to filter out blast hits with percent identity less than 95 and gap opens greater than 0. A file consisting of species chromosome identity, gene name, gene start and end positions was generated from the respective gff3 file. MCScanX software (Wang *et al.* 2012) was used to generate the dot plot and dual synteny plot that were used to compare the structural differences between the genome of *T. aesitvum* and *Ae. tauschii*.

The difference in recombination rate between *Ae. tauschii* ssp. *strangulata* and *Ae. tauschii* ssp. *tauschii* introgression lines was ascertained by the pairwise comparison of families derived from each subspecies using equal number of SNPs from the same genomic loci. The correlation between total RBP and genetic distance was calculated between the introgression lines and their respective hexaploid wheat and *Ae. tauschii* parents. Scatter plots for genetic distance versus total RBP were generated for each family using ggplot2.

Identity by Descent detection (IBD)

Introgression of *Ae. tauschii* genome in hexaploid wheat was inferred using IBD. SNPs from each chromosome were separated and used as input genotype (gt) data for IBD detection. The program Beagle v.4.1 was used to detect IBD segments between introgression

lines, hexaploid wheat and *Ae. tauschii* parents using default parameters. The R-package ggplot2 was used to generate a density plot of IBD segment start per chromosome to show the distribution pattern. All chromosomes were scaled by dividing the IBD values by the individual chromosome length and then multiplied by 100. Using a sliding window of 0.5 Mb and a 1 % fraction of overlap between features, IBD segments shared between introgression lines and *Ae. tauschii* parents were counted in each window using the bedmap tool provided in BEDOPS v2.4.35 and a line graph was plotted using ggplot2.

The efficiency of introgression was estimated as a percentage of observed proportion of *Ae. tauschii* genome in the introgression lines as inferred by IBD to the expected proportion of *Ae. tauschii* in $BC_1F_{3:5}$. Assuming that recombination events between *Ae. tauschii* and hexaploid wheat D genomes occurred normally in each chromosome, the expected proportion of *Ae. tauschii* genome in the $BC_1F_{3:5}$ introgression lines was approximated at 25 %. The observed proportion of introgression was obtained by dividing the total length of IBD segments from *Ae. tauschii* shared with each line by the genome size of *Ae. tauschii* (4.3 Gb) and multiplied by 100. The result was then divided by 25 and multiplied by 100 to get the percentage introgression efficiency. The average, standard deviation, minimum and maximum IBD length shared between introgression lines, introgression lines and hexaploid wheat, introgression lines and *Ae. tauschii* parents were determined, and divided by the chromosome size.

The relationship between IBD and the domestication gene tenacious glume (*Tg*) on chromosome arm 2DS was explored. The IBD count per 1 kb sliding window was used to compare the frequency of introgression in the *Tg* region. Genes within the *Tg* region (21.8 Mb to 23.3 Mb) and their functional annotation were extracted from the Chinese Spring reference gene annotation file. Introgression lines were phenotyped for tenacious glume trait. The results were used to confirm the presence or absence of wild type alleles depending on whether the introgression segment spanned the *Tg* gene region or not. Genome-wide association analysis of tenacious glume trait with the 11,624 SNP markers was done using GAPIT function in R. A mixed linear model was used and the population structure was controlled using the first three principal components calculated from the markers. A Manhattan plot of negative $\log_{10}$ of false discovery rate (FDR) transformed P-values from the D chromosomes was generated in R using 'qqman' package.

Data availability

All supplemental material and relevant data are available at FigShare.

Results

Genotyping and SNP imputation

A total of 314,783,044 high quality NGS reads with barcodes were generated with an average of 2,713,647 reads per sample from the diverse *Ae. tauschii* accessions (Supporting Information Table S2). Eighty-six percent (86 %) of the reads were aligned to the Chinese Spring reference sequence v.1.0 (The International Wheat Genome Sequencing Consortium (IWGSC) 2018) with an average of 2,336,299 reads per sample. The number of SNP sites generated from the TASSEL v. 5.0 GBS v.2 pipeline was 120,877. After filtering out SNPs with MAF less than 0.02, and maximum missingness greater than 70%, the number of retained SNPs was 86,031.

Similarly, 1,080,452,138 high quality reads with barcodes were generated with an average of 2,904,441 reads per sample from the introgression population (Supporting Information Table S1). Ninety-six percent (96 %) of the reads were aligned to the Chinese Spring reference with an average of 2,801,376 reads per sample. The number of unfiltered SNPs generated by the TASSEL v.5.0 GBS v.2 pipeline was 275,286. A total of 58,932 SNPs from the A, B, and D genomes were retained after filtering out SNPs with MAF less than 0.05 and maximum missingness greater than 30%. The number of SNPs from the D genome was 37.6 % of the filtered SNP dataset. The second filtering performed on the D genome SNPs to remove sites with MAF less than 0.01 resulted in 41,228 SNPs, out of which, 7,749 also segregated in the diverse set of *Ae. tauschii* accessions (henceforth, reference panel). Using the program Beagle v.5.0 (Browning and Browning 2013), 78,282 SNPs were imputed from the reference panel into the *Ae. tauschii*-derived introgression population.

Principle component analysis

Population structure of *Ae. tauschii* and introgression populations based on genetic markers reflects the allele diversity in the species. The 137 *Ae. tauschii* accessions formed three distinct clusters when the first two PCs calculated from 11,624 SNPs were plotted (Fig. 1). One cluster consisting of accessions known to belong to *Ae. tauschii* ssp. *strangulata* or lineage 2 (L2), was clearly distinct from the rest (Wang *et al.* 2013). The remaining two clusters belonged to *Ae. tauschii* ssp. *tauschii* or lineage 1 (L1a and L1b). Cluster L1a was the most heterogeneous with accessions coming from Afghanistan (AFG), Turkmenistan (TKM), Iran (IRN), Pakistan (PAK) and Tajikistan (TJK), (Table S3). Fifteen of the *Ae. tauschii* parents used to generate the introgression population belonged to this cluster. More

than two thirds of the accessions in cluster L1b were from Turkey (TUR) with only a few admixtures from Armenia (ARM), IRN, TJK and PAK. Three parents of the introgression population were present in this cluster. Cluster L2 consisted of *Ae. tauschii* accessions mostly collected from Iran (IRN), although a few accessions from Azerbaijan (AZE), Turkmenistan (TKM) and TUR were present. Three parents of the introgression population parents clustered in this group and two of them (TA1642, TA2378) are known to belong to *Ae. tauschii* ssp. *strangulata* or lineage 2 (Wang *et al.* 2013; Singh *et al.* 2019).

The broad geographic distribution of *Ae. tauschii* accessions used to generate the introgression population increases the chances of transferring alleles adaptive to different agroecological zones. When the introgression lines were plotted on the first two PCs together with *Ae. tauschii* accessions and hexaploid wheat parents, cluster L1a and L1b collapsed into one cluster (Fig. 2). Cluster L2 remained independent while the introgression lines and hexaploid wheat parents formed another cluster. The introgression lines showed a wide distribution on the two PCs relative to the *Ae. tauschii* and hexaploid wheat parents. Many introgression lines clustered closer to hexaploid wheat parents indicating that the greater proportion of genome in the $BC_1F_{3:5}$ lines comes from hexaploid wheat. This trend is likely associated with the loss of the introgressed segments as a result of backcrossing to the hexaploid parents and selection during population development. When the introgression lines were compared with the hexaploid wheat parents using 13,719 SNPs from all three sub-genomes, clustering was consistent with the pedigree (Fig. 3). In each cluster, admixed introgression lines were observed because of the shared *Ae. tauschii* parents.

Genetic diversity

While most domesticated species experienced loss of genetic diversity due to population bottleneck and selection for alleles controlling domestication traits, their respective wild ancestors often maintain high levels of genetic diversity (Akhunov *et al.* 2010; Xu *et al.* 2012; Hufford *et al.* 2012). To assess the effect of wild relative introgression on genetic diversity in wheat, we estimated SNP diversity (π) in the populations of *Ae. tauschii*, hexaploid wheat parents and the introgression lines. The average π value for the diverse *Ae. tauschii* accessions was 0.33, suggesting that these accessions represented a rich allelic diversity that could be valuable for wheat improvement. A cross-population diversity comparison showed a low average genetic diversity in the wheat D genome across all chromosomes (Table 1). The lowest diversity was found in the hexaploid wheat parents with

the chromosome mean ranging from 0.004 to 0.014 as compared to *Ae. tauschii* parents that ranged from 0.108 to 0.114. For most chromosome regions, the levels of genetic diversity in the introgression population were intermediate between the levels of diversity in the parental populations of wheat and *Ae. tauschii* but tended towards the *Ae. tauschii* with maximum mean π of 0.12 on chromosome 4D (Fig. 4 and Fig. S2). Analysis of variance showed significant differences in π values between *Ae. tauschii*, hexaploid wheat and introgression lines ($P < 0.001$), but not between chromosomes ($P = 1$). The genetic diversity of the introgression lines for most regions of chromosome 4D and 5D were higher than those of *Ae. tauschii* parents (Fig. S2). Taken together, these results indicate that *Ae. tauschii* introgression lines substantially increased the genetic diversity of the recurrent hexaploid wheat parents.

Effect of recombination rate on introgression

One of the factors affecting the distribution of recombination events across genome are structural re-arrangements (Stapley *et al.* 2017). Using the comparative dot-plot analysis of gene order along the chromosomes, we observed that more than 99% of the genes from *T. aestivum* were perfectly collinear to those of *Ae. tauschii* ssp. *strangulata* suggesting lack of major structural re-arrangements between the D genomes of bread wheat and its diploid ancestor (Fig. 5A). However, some small-scale inversions were observed on chromosomes 2D, 4D and 6D in the regions near the centromeres, and four genes were found in non-syntenic positions between the wheat (1D and 5D) and *Ae. tauschii* (1D, 4D and 5D) chromosomes (Fig. 5B).

Efficiency of introgression could also be strongly influenced by the distribution of recombination rate along the chromosomes. Consistent with previous observations (Jordan *et al.* 2018), a high frequency of recombination events was observed towards the telomeres but the patterns were chromosome specific. Kruskal Wallis test showed that across all families, the number of recombination breakpoints (RBP) was significantly different between chromosomes ($P < 0.001$) at 95 % confidence level. At 99th percentile, chromosome 2D, 5D and 7D had the highest number of regions with elevated recombination rate, while 1D and 5D had the highest total RBP per 10 kb window (Table 2, Table S4). Most chromosomes showed no evidence of recombination in the regions near the centromere, except for chromosomes 1D and 6D (Fig. 6).

The frequency of recombination points between two different species are determined by the genetic distance or sequence diversity between them. It is expected that, as the genetic

distance between hexaploid wheat and the *Ae. tauschii* accessions increases, the total RBP should reduce within the introgression lines resulting from such a cross. All families showed negative Pearson's correlation coefficients for total RBP and genetic distance between introgression lines and *Ae*. *tauschii* parents (Fig. 7). As expected, introgression lines with many introgression segments from *Ae. tauschii* had many total RBP and were genetically closer to *Ae. tauschii* than wheat. A reduction in the total RBP was observed in families FAM92, FAM93 and FAM96 derived from wheat and *Ae. tauschii* parents TA1642 and TA2378, which belong to *Ae. tauschii* ssp. *strangulata.* This observation was contrary to what was expected. Further analysis comparing FAM93 derived from *Ae. tauschii* ssp. *strangulata* to seven *Ae. tauschii* ssp. *tauchsii*-derived families (FAM97, FAM98, FAM99, FAM106, FAM109, FAM112 and FAM116) and FAM92 derived from *Ae. tauschii* ssp. *strangulata* using same and equal number of SNPs that were segregating between the parental lines also showed the same trend (Fig S3). The *t*-test statistics did not reveal significant difference in recombination rate between FAM93 and FAM92 ($P = 0.469$) at 95 % confidence level. However, significant differences were observed between *Ae. tauschii* ssp. *strangulata*-derived family FAM93 and all *Ae. tauschii* ssp. *tauschii*-derived families ($P < 0.001$). These observations can be explained by the low level of SNP diversity between the wheat D genome and *Ae. tauschii* ssp. *strangulata*, which is considered to be the donor of the wheat D genome (Dvorak *et al.* 1998), resulting in underestimation of the total number of crossovers in the FAM92, FAM93 and FAM96 families. It is also possible that increase in the levels of interhomolog polymorphism can stimulate recombination. In *Arabidopsis*, increase in crossovers was observed when heterozygous regions are juxtaposed with homozygous regions (Ziolkowski *et al.* 2015), suggesting that the genomic distribution of interhomolog divergence have substantial effect on distribution of recombination rate.

Identity by descent analysis shows low introgression frequency in the pericentromeric regions

The proportion of wild relative genome in the elite wheat lines can influence many traits but the location of introgressions is key in determining the effects. In this study, IBD was used to infer introgression of *Ae. tauschii* genome into hexaploid wheat lines. A density plot of IBD segments along the chromosomes of the introgression population showed a U-shaped distribution (Fig. 8). The frequency of IBD segments positively correlated with the distribution of recombination rate (Jordan *et al.* 2018) and increased from the centromeres towards the telomeric regions of the chromosomes. There was no chromosome preference

404 during introgression. Variation in the number of introgressions per line were observed across
405 chromosomes with the percentage proportion of *Ae. tauschii* genome in the introgression
406 lines ranging from 0.075 % to 13.5 % (Table S5). The efficiency of introgression as inferred
407 by IBD ranged from 0.3 % to 54.1 % based on the expected 25 % *Ae. tauschii* genome in the
408 $BC_1F_{3:5}$ lines. Some lines had single or multiple introgression per chromosome. The IBD
409 segments shared between the introgression lines and wheat parents were on average 2.4 folds
410 longer than those shared with the *Ae. tauschii* parents (Table 3), but not significantly different
411 at 95% confidence level based on the *t*-test statistics (P = 0.066). The average percent length
412 of IBD segments shared between introgression lines and *Ae. tauschii* parents varied from
413 2.69 % to 6.98 % with a minimum of 0.28 % and a maximum of 41.33 %. Similarly, the
414 average percent length of IBD segments shared between the chromosomes of introgression
415 lines and hexaploid wheat parents ranged between 6.16 % and 26.63 % with a minimum of
416 0.44 % and a maximum of 86.18 %. The IBD segments shared between introgression lines
417 reached up to 100 % on chromosomes 3D, 4D and 6D.
418
419 **Relationship between IBD segments and tenacious glume gene**
420 Free-threshing is one of the traits that led to the domestication of wheat and it is
421 controlled by *Q* and *Tg* genes (Jantasuriyarat *et al.* 2004; Simons *et al.* 2006). In-depth
422 analysis of chromosome 2D was carried out to understand the relationship between IBD
423 segments and domestication gene *Tg* (Sood et al. 2009). The analysis was based on the
424 hypothesis that introgression lines that were free-threshing had shared IBD segments with
425 wheat on 2DS where *Tg* gene is expected and no shared IBD segments with *Ae. tauschii*. To
426 test this hypothesis, the sequences of microsatellite markers *Xgwm*455, *Xgwm*296, *Xgwm*261
427 and *Xwmc*503 linked to *Tg* were aligned to the Chinese Spring reference v.1.0 to determine
428 their location on 2DS. Marker *Xwmc*503 closest to *Tg* gene mapped at 19.6 Mb on 2DS
429 (Table S6). Based on Sood *et al.* (2009) genetic map, the *Tg* gene is located 2.2 cM away
430 from marker *Xwmc*503, implying that the *Tg* gene is located approximately at position 21.8
431 Mb. A count of IBD segments within 1-kb sliding windows showed a sharp decline in IBD
432 segments shared between introgression lines and *Ae. tauschii* parents within the *Tg* gene
433 region (Fig. 9A). The IBD segments shared between the introgression lines and hexaploid
434 wheat parents increased in the *Tg* gene region indicating a selection pressure for free-
435 threshing trait during population development. The lowest decline in IBD segments count
436 was observed at 23.3 Mb. There were 40 high confidence genes within the 21.8 Mb to 23.3
437 Mb interval (Table S7) including two transcription factors from the bZIP and GRAS families.

To verify the impact of introgression on free-threshing, we phenotyped the introgression lines for tenacious glume trait and compared the results with IBD map. All lines that had introgression segments spanning the *Tg* gene region on 2DS were positive for tenacious glume trait (Table S8). Some lines, which had the introgression segment boundary close to the *Tg* region also scored positive for tenacious glume trait (false negative), but a majority were negative as expected. The presence of some false negatives could be explained by the inability of the Beagle program to accurately determine the exact boundary of the introgression in some cases.

Genome-wide association studies are used to determine the non-random association of marker alleles to the trait of interest. Using a mixed linear model while controlling for the population structure, we observed that majority of the significant SNPs associated with tenacious glume trait in the introgression population were located on chromosome arm 2DS (Fig. 9B), which was consistent with IBD analysis. At a threshold FDR q-value of 0.05, 31 SNPs near the *Tg* locus on 2DS showed significant association with the trait and the closest SNPs to the *Tg* locus were chr2D_19242994 and chr2D_22955732 located downstream and upstream of the locus, respectively (Table S9).

Discussion

Wheat improvement through breeding is a continuous process that delivers new varieties to the farmers to ensure sustainable food production under changing environmental conditions and increasing world population. For several decades, breeding efforts have been directed towards improving agronomic, yield and disease resistance traits but of late climate change is becoming one of the major crop production constraints. Wild relatives adapted to various agroecological climates are the sources of alleles that can protect wheat by making it resilient to climate change. In this study, we tracked the introgression of diverse *Ae. tauschii* subspecies genome segments in hexaploid wheat and assessed its impact on genetic diversity.

Loss of genetic diversity associated with domestication and breeding (Haudry *et al.* 2007; Akhunov *et al.* 2010; Ozkan *et al.* 2011; Xu *et al.* 2012; Hufford *et al.* 2012) can potentially reduce the adaptive potential of cultivated wheat. Wild relatives of wheat were shown to be valuable source of allelic diversity for improving disease resistance, drought tolerance and quality traits (Uauy *et al.* 2006; Sohail *et al.* 2011; Saintenac *et al.* 2013; Periyannan *et al.* 2013; Chen *et al.* 2015). The wild diploid ancestor of the wheat D genome, *Ae. tauschii*, was shown to have two main lineages that experienced limited contact (Wang *et al.* 2013). The level of D genome genetic diversity in the wheat lines derived from crosses

with *Ae. tauschii* ssp. *strangulata* from lineage 2 was shown to be lower than in the lines derived from *Ae. tauschii* ssp. *tauschii* from lineage 1, consistent with the origin of the wheat D genome from *Ae. tauschii* ssp. *strangulata* (Dvorak *et al.* 2012; Wang *et al.* 2013). Therefore, the introgression lines developed in our study carry genomic segments from *Ae. tauschii* accessions coming from geographically diverse locations that represent different climatic conditions and biotic pressure under which wheat grows. These lines present a valuable genetic resource for breeding climate resilient, disease resistant and nutritionally high-quality wheat.

Introgression from wild relatives into wheat is challenged by hybrid incompatibility, embryo abortion and infertility (Gill and Raupp 1987), and was successfully accomplished from the direct ancestors of the wheat D and AB genomes using synthetic hexaploids or octoploids (Miranda *et al.* 2006; Dreisigacker *et al.* 2008; Dale *et al.* 2017). High density genotyping data generated by sequencing now permits high-resolution haplotype analysis of diverse populations and accurate imputation of missing data using reference panels (Jordan *et al.* 2015; Nyine *et al.* 2019). In this study, imputation with the reference panel of *Ae. tauschii* accessions was used to increase the number of SNPs on the D genome of the introgression lines. This resource enabled identification of introgression segments from a wild relative by inferring IBD regions. A similar approach was used in maize to identify 23 regions showing IBD with the foundation parents using MaizeSNP50 BeadChip (Liu *et al.* 2015).

The lack of major structural rearrangements differentiating the wheat D genome from its diploid ancestor underlies successful gene and trait transfer from *Ae. tauschii* into bread wheat. The high level of structural similarity between these two genomes facilitated the hybridization and recombination between them (Gill and Raupp 1987; Dvorak *et al.* 1998; Akhunov *et al.* 2010; Luo *et al.* 2017), with our results being consistent with the previously made observations. Relatively small inversions near the centromeric regions of chromosomes 2D, 4D and 6D had little impact on introgression efficiency compared to other centromeric regions without inversions. The low frequency of crossovers in these regions does not allow for estimating the effect of structural rearrangements on recombination and introgression.

The frequency of IBD regions along the chromosomes showed a U-shaped distribution with lower incidence of regions derived from *Ae. tauschii* in the pericentromeric regions. The introgression frequency correlated negatively with the length of IBD regions and positively with the frequency of crossovers indicating that longer introgressed segments in the low-recombining pericentromeric regions had lower chance of being inherited in the progeny of crosses between *Ae. tauschii*-derived octoploids and wheat. These chromosomal

patterns of introgression efficiency and length suggest that introgression was strongly affected by the distribution of recombination rate along chromosomes. This outcome was not expected as we assumed that selected lines in the $BC_1F_{3:4}$ generation plants should have equal probability of inheriting introgressed regions across entire genome. However, it is likely that selection applied at $BC_1F_{3:4}$ generation to maintain uniform phenology, threshability, flowering time and developmental characteristics inadvertently eliminated many lines carrying large introgressed regions in the pericentromeric regions. According to theory, introgressions that carry alleles having a negative impact on the selected traits will be removed from the population, with the size of the affected region defined by the recombination rate (Hill and Robertson 1966). It appears that negative interaction between alleles located within large introgressions in the low-recombining pericentromeric region and alleles of the adapted recurrent parent affected targeted phenotypes resulting in removal of these plants during population development. The limited number of recombination events at the BC_1F_2 generation, especially in the large pericentromeric regions of wheat chromosomes, resulted in linkage drag that affected substantial proportion of the genome.

On the contrary, terminal regions of wheat chromosomes showed the high rate of introgression consistent with the theoretical predictions of the effect of selection on linked variation (Hill and Robertson 1966). The importance of recombination in separating the negatively selected alleles from the background was clearly demonstrated for the *Tg* locus controlling free-threshing trait in wheat (Jantasuriyarat *et al.* 2004; Sood *et al.* 2009). Since this gene is located in the high-recombining terminal region of chromosome, we did not observe substantial effect of selection against the wild-type allele on the frequency of introgression from *Ae. tauschii*. The high recombination rate even allowed for mapping the *Tg* gene locus to the 1.5 Mb genomic interval, which was confirmed by genome-wide association analysis. Taken together, these results indicate that the unintended consequence of selection applied during the early stages of introgression population development is the low rate of introgression in the low-recombining regions of the wheat genome.

With the development of new genomic resources for wheat and its wild relatives (Ling *et al.* 2013; Avni *et al.* 2017; Luo *et al.* 2017; The International Wheat Genome Sequencing Consortium (IWGSC) 2018; Arora *et al.* 2019), the importance of introgression populations for wheat improvement is increasing. Recombination will be one of the main factors that will influence the efficiency of introgression in these populations. Our study suggests that any form of selection applied during population development quickly eliminates large portions of the donor genome, especially in regions of low recombination. The loss of the wild relative's

alleles in the pericentromeric regions in this study was somewhat surprising given that phenotypic selection was only applied at the $BC_1F_{3:4}$ stage and field conditions for selection were poor. This clearly suggests that multiple genes with strong combined effect on adaptive traits are present in these regions and identification of any beneficial alleles in these regions will be complicated by linkage drag.

It is common practice for germplasm development programs to subject material to selection pressure from early stages of population development. This is consistent with the goal of identifying high performing materials relatively quickly to support commercial breeding. The application of early selection allows rapid exploitation of beneficial alleles in the regions of high recombination. This is a worthy objective but the current results are a clear justification for a two-tiered approach to germplasm development if such programs are to fully exploit the diversity present in donor material.

Exploration and exploitation of diversity in regions of lower recombination requires a parallel approach to germplasm development that complements the efforts to rapidly exploit diversity present in the highly recombining regions. The first step is to ensure that maximum diversity is maintained in the introgression materials. This could be achieved by genotyping early generation populations to select subsets of lines carrying introgressions covering the genome. Low selection pressure and marker-assisted population management will retain introgressed regions. The drive toward fixation that occurs with additional selfing generations calls for methods to maintain heterozygosity, such as random mating through the use of genetic male sterility or chemical hybridizing agents, that are warranted in self-pollinated species. This would enhance effective recombination and increase the probability of freeing beneficial alleles from the influence of linked deleterious alleles in regions of low recombination. Failure to engage such strategies will result in the near-immediate loss of introgressed diversity, reducing the potential long-term impact of germplasm development programs. Recently, genetic factors controlling crossover frequency across genome and in the pericentromeric regions of wheat chromosomes have been identified (Jordan *et al.* 2018; Gardiner *et al.* 2019). The discovery of these genetic factors could also facilitate strategies to further increase the efficiency of introgression, and selection for favorable introgressed alleles in the low recombining regions.

Acknowledgements

This project was supported by the Agriculture and Food Research Initiative Competitive Grant 2016-67013-24473 from the USDA National Institute of Food and Agriculture. We

would like to thank Alina Akhunova and KSU Integrated Genomics Facility for sequencing genomic libraries, and Jon Raupp from Wheat Genetics Resources Center for providing seeds of 21 accessions of *Aegilops tauschii* used for developing introgression population.

References

Akhunov, E. D., A. R. Akhunova, O. D. Anderson, J. a Anderson, N. Blake *et al.*, 2010 Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC Genomics 11: 702.

Alonso, L., and G. Kimber, 1984 Use of restitution nuclei to introduce alien genetic variation into hexaploid wheat. Zeitschrift für Pflanzenzüchtung 92: 185–189.

Arora, S., B. Steuernagel, K. Gaurav, S. Chandramohan, Y. Long *et al.*, 2019 Resistance gene cloning from a wild crop relative by sequence capture and association genetics. Nat. Biotechnol. 37: 139–143.

Avni, R., M. Nave, O. Barad, K. Baruch, S. O. Twardziok *et al.*, 2017 Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science 97: 93–97.

Ayliffe, M., R. Singh, and E. Lagudah, 2008 Durable resistance to wheat stem rust needed. Curr. Opin. Plant Biol. 11: 187–92.

Beales, J., A. Turner, S. Griffiths, J. W. Snape, and D. a Laurie, 2007 A pseudo-response regulator is misexpressed in the photoperiod insensitive Ppd-D1a mutant of wheat (Triticum aestivum L.). Theor. Appl. Genet. 115: 721–33.

Browning, S. R., and B. L. Browning, 2012 Identity by descent between distant relatives: detection and applications. Annu. Rev. Genet. 46: 617–33.

Browning, B. L., and S. R. Browning, 2013 Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics 194: 459–71.

Chen, S., M. N. Rouse, W. Zhang, Y. Jin, E. Akhunov *et al.*, 2015 Fine mapping and characterization of Sr21, a temperature-sensitive diploid wheat resistance gene effective against the Puccinia graminis f. sp. tritici Ug99 race group. Theor. Appl. Genet. 128: 645–56.

Cruz, C. D., G. L. Peterson, W. W. Bockus, P. Kankanala, J. Dubcovsky *et al.*, 2016 The 2NS translocation from Aegilops ventricosa confers resistance to the Triticum pathotype of Magnaporthe oryzae. Crop Sci. 56:.

Dale, Z., H. Jie, H. Luyu, Z. Cancan, Z. Yun *et al.*, 2017 An advanced backcross population through synthetic octaploid wheat as a “Bridge”: Development and QTL detection for

seed dormancy. Front. Plant Sci. 8: 1–10.

Dreisigacker, S., M. Kishii, J. Lage, and M. Warburton, 2008 Use of synthetic hexaploid wheat to increase diversity for CIMMYT bread wheat improvement. Aust. J. Agric. Res. 59: 413–420.

Dvorak, J., K. R. Deal, M.-C. Luo, F. M. You, K. von Borstel *et al.*, 2012 The origin of spelt and free-threshing hexaploid wheat. J. Hered. 103: 426–41.

Dvorak, J., M. Luo, and E. D. Akhunov, 2011 N . I . Vavilov ' s Theory of Centres of Diversity in the Light of Current Understanding of Wheat Diversity , Domestication and Evolution. Czech. J. Genet. Plant Breed. 47: 1–8.

Dvorak, J., M. C. Luo, Z. L. Yang, and H. B. Zhang, 1998 The structure of the Aegilops tauschii genepool and the evolution of hexaploid wheat. Theor. Appl. Genet. 97: 657–670.

Elshire, R. J., J. C. Glaubitz, Q. Sun, J. a Poland, K. Kawamoto *et al.*, 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19379.

Gardiner, L., L. U. Wingen, P. Bailey, R. Joynson, T. Brabbs *et al.*, 2019 Analysis of the recombination landscape of hexaploid bread wheat reveals genes controlling recombination and gene conversion frequency. Genome Biol. 20: 69.

Ge, S. De, 2001 Size of Donor Chromosome Segments Around Introgressed Loci and Reduction of Linkage Drag in Marker-Assisted Backcross Programs. Genetics.

Gill, B. S., and W. J. Raupp, 1987 Direct Genetic Transfers from Aegilops squarrosa L. to Hexaploid Wheat. Crop Sci. 27: 445–450.

Glaubitz, J. C., T. M. Casstevens, F. Lu, J. Harriman, R. J. Elshire *et al.*, 2014 TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLoS One 9: e90346.

Haudry, a, a Cenci, C. Ravel, T. Bataillon, D. Brunel *et al.*, 2007 Grinding up wheat: a massive loss of nucleotide diversity since domestication. Mol. Biol. Evol. 24: 1506–17.

He, F., R. Pasam, F. Shi, S. Kant, G. Keeble-Gagnere *et al.*, 2019 Exome sequencing highlights the role of wild relative introgression in shaping the adaptive landscape of the wheat genome. Nat. Genet. 51: 896–904.

Hill, W., and Robertson, 1966 The effect of linkage on limits to artificial selection. Genet. Res. 8: 269–294.

Hufford, M. B., X. Xu, J. van Heerwaarden, T. Pyhäjärvi, J.-M. Chia *et al.*, 2012 Comparative population genomics of maize domestication and improvement. Nat. Genet. 44: 808–11.

Jantasuriyarat, C., M. I. Vales, C. J. W. Watson, and O. Riera-Lizarazu, 2004 Identification and mapping of genetic loci affecting the free-threshing habit and spike compactness in wheat (Triticum aestivum L.). Theor. Appl. Genet. 108: 261–273.

Jones, H., N. Gosman, R. Horsnell, G. A. Rose, L. A. Everest *et al.*, 2013 Strategy for exploiting exotic germplasm using genetic, morphological, and environmental diversity: the Aegilops tauschii Coss. example. Theor. Appl. Genet. 126: 1793–808.

Jordan, K. W., S. Wang, F. He, S. Chao, Y. Lun *et al.*, 2018 The genetic architecture of genome-wide recombination rate variation in allopolyploid wheat revealed by nested association mapping. Plant J. 95: 1039–1054.

Jordan, K., S. Wang, Y. Lun, L. Gardiner, R. MacLachlan *et al.*, 2015 A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. Genome Biol. 16: 48.

Kihara, H., 1944 Discovery of the DD-analyser, one of the ancestors of Triticum vulgare. Agricuture Hortic. 19: 889–890.

Ling, H.-Q., S. Zhao, D. Liu, J. Wang, H. Sun *et al.*, 2013 Draft genome of the wheat A-genome progenitor Triticum urartu. Nature 496: 87–90.

Liu, Z., R. L. Bowden, and G. Bai, 2013 Molecular markers for leaf rust resistance gene Lr42 in wheat. Crop Sci. 53: 1566–1570.

Liu, C., Z. Hao, D. Zhang, C. Xie, M. Li *et al.*, 2015 Genetic properties of 240 maize inbred lines and identity-by-descent segments revealed by high-density SNP markers. Mol. Breed. 35: 146.

Liu, S., L. X. Yu, R. P. Singh, Y. Jin, M. E. Sorrells *et al.*, 2010 Diagnostic and co-dominant PCR markers for wheat stem rust resistance genes Sr25 and Sr26. Theor. Appl. Genet. 120: 691–697.

Luo, M.-C., Y. Q. Gu, D. Puiu, H. Wang, S. O. Twardziok *et al.*, 2017 Genome sequence of the progenitor of the wheat D genome Aegilops tauschii. Nature 551: 498–502.

Luo, M.-C., Z.-L. Yang, F. M. You, T. Kawahara, J. G. Waines *et al.*, 2007 The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. Theor. Appl. Genet. 114: 947–59.

McIntosh, R. a., P. Zhang, C. Cowger, R. Parks, E. S. Lagudah *et al.*, 2011 Rye-derived powdery mildew resistance gene Pm8 in wheat is suppressed by the Pm3 locus. Theor. Appl. Genet. 123: 359–367.

Miranda, L. M., J. P. Murphy, D. Marshall, and S. Leath, 2006 Pm34: a new powdery mildew resistance gene transferred from Aegilops tauschii Coss. to common wheat (Triticum

aestivum L.). Theor. Appl. Genet. 113: 1497–504.

Neph, S., M. S. Kuehn, A. P. Reynolds, E. Haugen, R. E. Thurman *et al.*, 2012 BEDOPS: High-performance genomic feature operations. Bioinformatics 28: 1919–1920.

Nyine, M., B. Uwimana, N. Blavet, E. Hřibová, H. Vanrespaille *et al.*, 2018 Genomic Prediction in a Multiploid Crop: Genotype by Environment Interaction and Allele Dosage Effects on Predictive Ability in Banana. Plant Genome 11:.

Nyine, M., S. Wang, K. Kiani, K. Jordan, S. Liu *et al.*, 2019 Genotype Imputation in Winter Wheat Using First-Generation Haplotype Map SNPs Improves Genome-Wide Association Mapping and Genomic Prediction of Traits. G3: Genes|Genomes|Genetics 9: 125–133.

Ogbonnaya, F. C., O. Abdalla, A. Mujeeb-Kazi, A. G. Kazi, S. S. Xu *et al.*, 2013 Synthetic Hexaploids: Harnessing Species of the Primary Gene Pool for Wheat Improvement. Plant Breed. Rev. 35–122.

Ozkan, H., G. Willcox, A. Graner, F. Salamini, and B. Kilian, 2011 Geographic distribution and domestication of wild emmer wheat (Triticum dicoccoides). Genet. Resour. Crop Evol. 58: 11–53.

Peleg, Z., T. Fahima, A. B. Korol, S. Abbo, and Y. Saranga, 2011 Genetic analysis of wheat domestication and evolution under domestication. J. Exp. Bot. 62: 5051–61.

Peng, J., D. E. Richards, N. M. Hartley, G. P. Murphy, K. M. Devos *et al.*, 1999 “Green revolution” genes encode mutant gibberellin response modulators. Nature 400: 256–61.

Peng, J., Y. Ronin, T. Fahima, M. S. Röder, Y. Li *et al.*, 2003 Domestication quantitative trait loci in Triticum dicoccoides, the progenitor of wheat. Proc. Natl. Acad. Sci. U. S. A. 100: 2489–94.

Periyannan, S., J. Moore, M. Ayliffe, U. Bansal, X. Wang *et al.*, 2013 The Gene Sr33, an Ortholog of Barley Mla Genes, Encodes Resistance to Wheat Stem Rust Race Ug99. Science 341: 786–8.

Poland, J. A., P. J. Brown, M. E. Sorrells, and J.-L. Jannink, 2012 Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7: e32253.

Qi, L., B. Friebe, P. Zhang, and B. S. Gill, 2007 Homoeologous recombination, chromosome engineering and crop improvement. Chromosom. Res. 15: 3–19.

R Development Core Team, R., 2011 R: A Language and Environment for Statistical Computing (R. D. C. Team, Ed.). R Found. Stat. Comput. 1: 409.

Ray, D. K., N. D. Mueller, P. C. West, and J. A. Foley, 2013 Yield Trends Are Insufficient to

Double Global Crop Production by 2050. PLoS One 8:.

Saintenac, C., D. Jiang, and E. D. Akhunov, 2011 Targeted analysis of nucleotide and copy number variation by exon capture in allotetraploid wheat genome. Genome Biol. 12: R88.

Saintenac, C., D. Jiang, S. Wang, and E. Akhunov, 2013 Sequence-based mapping of the polyploid wheat genome. G3 (Bethesda). 3: 1105–14.

Sears, E. R., 1977 AN INDUCED MUTANT WITH HOMOEOLOGOUS PAIRING IN COMMON WHEAT. Can. J. Genet. Cytol. 19: 585–593.

Simons, K. J., J. P. Fellers, H. N. Trick, Z. Zhang, Y.-S. Tai *et al.*, 2006 Molecular characterization of the major wheat domestication gene Q. Genetics 172: 547–55.

Singh, N., S. Wu, V. Tiwari, S. Sehgal, J. Raupp *et al.*, 2019 Genomic analysis confirms population structure and identifies inter-lineage hybrids in Aegilops tauschii. Front. Plant Sci. 10: 1–13.

Sohail, Q., T. Inoue, H. Tanaka, A. E. Eltayeb, Y. Matsuoka *et al.*, 2011 Applicability of Aegilops tauschii drought tolerance traits to breeding of hexaploid wheat. Breed. Sci. 61: 347–57.

Sood, S., V. Kuraparthy, G. Bai, and B. S. Gill, 2009 The major threshability genes soft glume (sog) and tenacious glume (Tg), of diploid and polyploid wheat, trace their origin to independent mutations at non-orthologous loci. Theor. Appl. Genet. 119: 341–51.

Stapley, J., P. G. D. Feulner, S. E. Johnston, A. W. Santure, and C. M. Smadja, 2017 Variation in recombination frequency and distribution across eukaryotes: Patterns and processes. Philos. Trans. R. Soc. B Biol. Sci. 372: 20160455.

Su, H., Y. Liu, C. Liu, Q. Shi, Y. Huang *et al.*, 2019 Centromere Satellite Repeats Have Undergone Rapid Changes in Polyploid Wheat Subgenomes. Plant Cell 31: tpc.00133.2019.

The International Wheat Genome Sequencing Consortium (IWGSC), 2018 Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361: eaar7191.

Uauy, C., A. Distelfeld, T. Fahima, A. Blechl, and J. Dubcovsky, 2006 A NAC Gene Regulating Senescence Improves Grain Protein, Zinc, and Iron Content in Wheat. Science 314: 1298–1301.

Wang, J., M.-C. Luo, Z. Chen, F. M. You, Y. Wei *et al.*, 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 Phytol. 198: 925–

937.
Wang, W., Q. Pan, B. Tian, F. He, Y. Chen *et al.*, 2019 Gene editing of the wheat homologs of TONNEAU1–recruiting motif encoding gene affects grain shape and weight in wheat. Plant J. https://doi.org/10.1111/tpj.14440.
Wang, Y., H. Tang, J. D. Debarry, X. Tan, J. Li *et al.*, 2012 MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40: 1–14.
Xu, X., X. Liu, S. Ge, J. D. Jensen, F. Hu *et al.*, 2012 Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. Nat. Biotechnol. 30: 105–11.
Ziolkowski, P. A., L. E. Berchowitz, C. Lambing, N. E. Yelina, X. Zhao *et al.*, 2015 Juxtaposition of heterozygous and homozygous regions causes reciprocal crossover remodelling via interference during Arabidopsis meiosis. Elife 4: 1–29.

Figure legends

Fig. 1. Distribution of 116 *Ae. tauschii* accessions (red) used as reference panel and the 21 *Ae. tauschii* accessions (magenta) used to generate the introgression lines on the first two principal components. L1a and L1b accessions belong to *Ae. tauschii* ssp. *tauschii* while L2 accessions belong to *Ae. tauschii* ssp. *strangulata.*

Fig. 2. Distribution of 116 *Ae. tauschii* accessions (AT) used as reference panel, the 21 *Ae. tauschii* accessions (ILP_AT) used to generate the introgression lines, hexaploid wheat parents (ILP_HW) and the 351 introgression lines (IL) on the first two principal components.

Fig. 3. Distribution of introgression lines and the hexaploid wheat parents on the first two principal components based on SNP markers from A, B, D genomes and unanchored scaffolds.

Fig. 4. Variation in nucleotide diversity for chromosome 2D based on pi values interpolated using R function ‘approx’. The blue dashed line indicates the position of SSR marker *Xwmc*503 linked to tenacious glume gene indicated by the magenta dashed line.

Fig. 5. Comparison of *T. aestivum* and *Ae. tauschii* genomes at protein sequence level. A is a dot plot showing the collinearity between genes and the deviation of the dots from the main diagonal indicate inversion. B shows the synteny between the two species genomes.

Fig. 6. Distribution of recombination hotspots per chromosome at 99th percentile.

Fig. 7. Scatter plots showing the correlation between total recombination breakpoints and genetic distance per introgression line from *Ae. tauschii* and hexaploid wheat parents in a family.

Fig. 8. Density plots of identity by descent segments start positions along the seven D chromosomes of the introgression lines derived from hexploid wheat and *Ae. tauschii*.

Fig. 9. Location of Tg locus on chromosome arm 2DS as inferred by identity by descent (IBD) analysis and genome-wide association study. A. Frequency of introgression from *Ae. tauschii* into hexploid wheat as inferred by IBD in chromosome arm 2DS region containing tenacious glume (*Tg*) gene. The IBD segments were counted per 1-kb sliding window. The blue line shows the position of marker *Xmwc*503, magenta line indicates the most likely position of *Tg* gene based on Sood et al. (2009) and the red lines shows the chromosome region with the lowest IBD frequency (*Tg* locus boundary). B. Manhattan plot showing the position of significant SNPs on 2DS and the red line shows the SNPs that are significant at an FDR q-value of 0.001.

Supporting information

Supplemental tables

Table S1. Summary of GBS data for introgression lines, *Ae. tuaschii* and hexaploid wheat parents.

Table S2. Summary of GBS data for 116 *Ae. tauschii* accessions used as a reference panel.

Table S3. Origin of *Ae. tauschii* accessions used as reference panel, the source of 21 *Ae. tauschii* used as introgression parents and their grouping based on the first two principal components.

Table S4. Frequency of total recombination breakpoint from 16 introgression population families.

Table S5. Efficiency of *Ae. tauschii* introgression in wheat as inferred by identity by descent.

Table S6. Location of microsatellite markers linked to tenacious glume (*Tg*) gene on the Chinese Spring reference v1.

Table S7. High confidence genes within chromosome arm 2DS interval known to control tenacious glume trait.

Table S8. Tenacious glume scores for the introgression lines with and without introgression from *Ae. tauschii* parents on chromosome arm 2DS where the *Tg* gene is located.

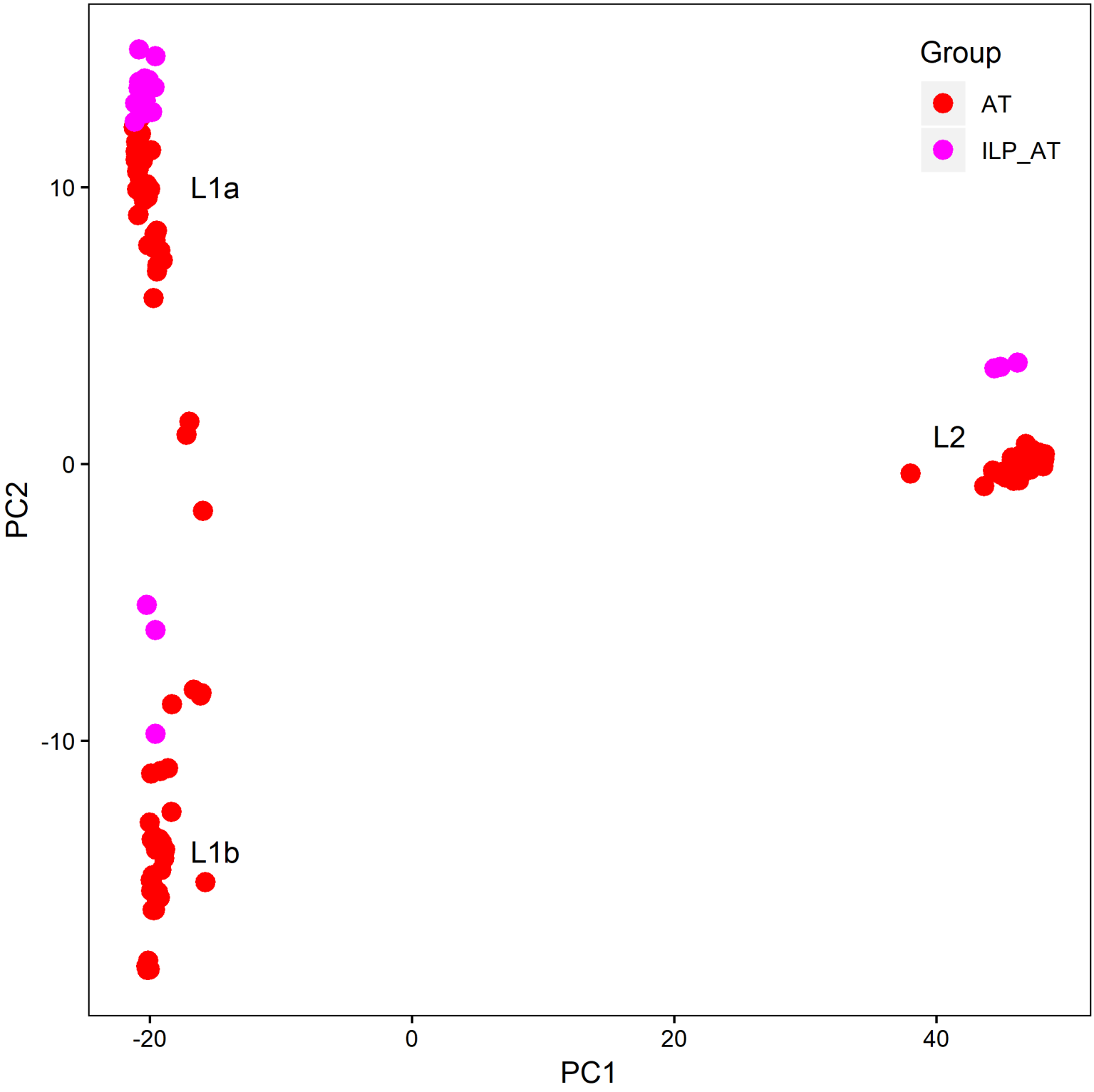
Table S9. SNPs on chromosome arm 2DS closest to Tg locus significantly associated with tenacious glume trait.

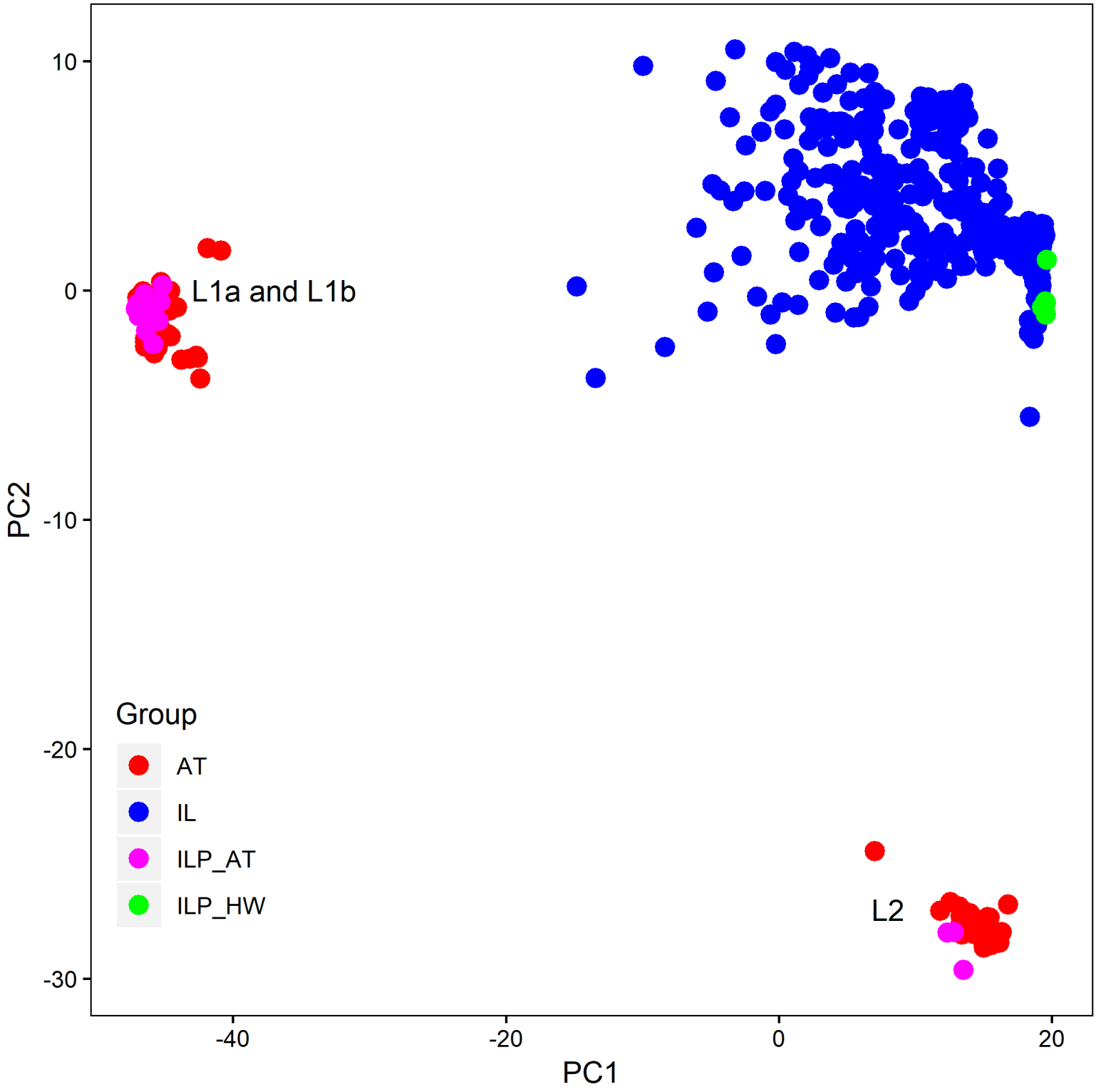
Supplemental figures

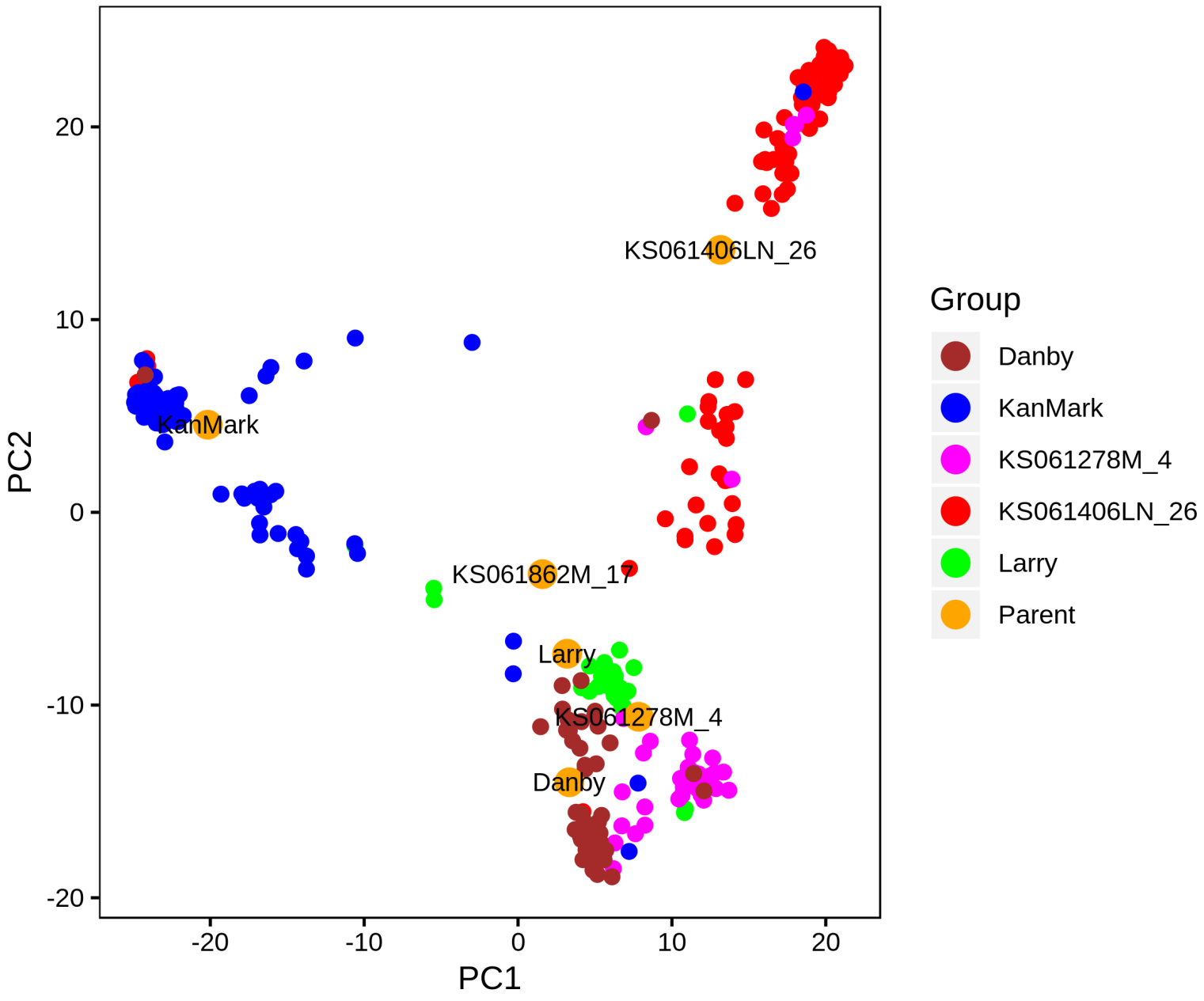
Fig. S1. A plot of identity by descent (IBD) count shared between the introgression lines and *Ae. tauschii* parents in 0.5 Mb sliding windows.

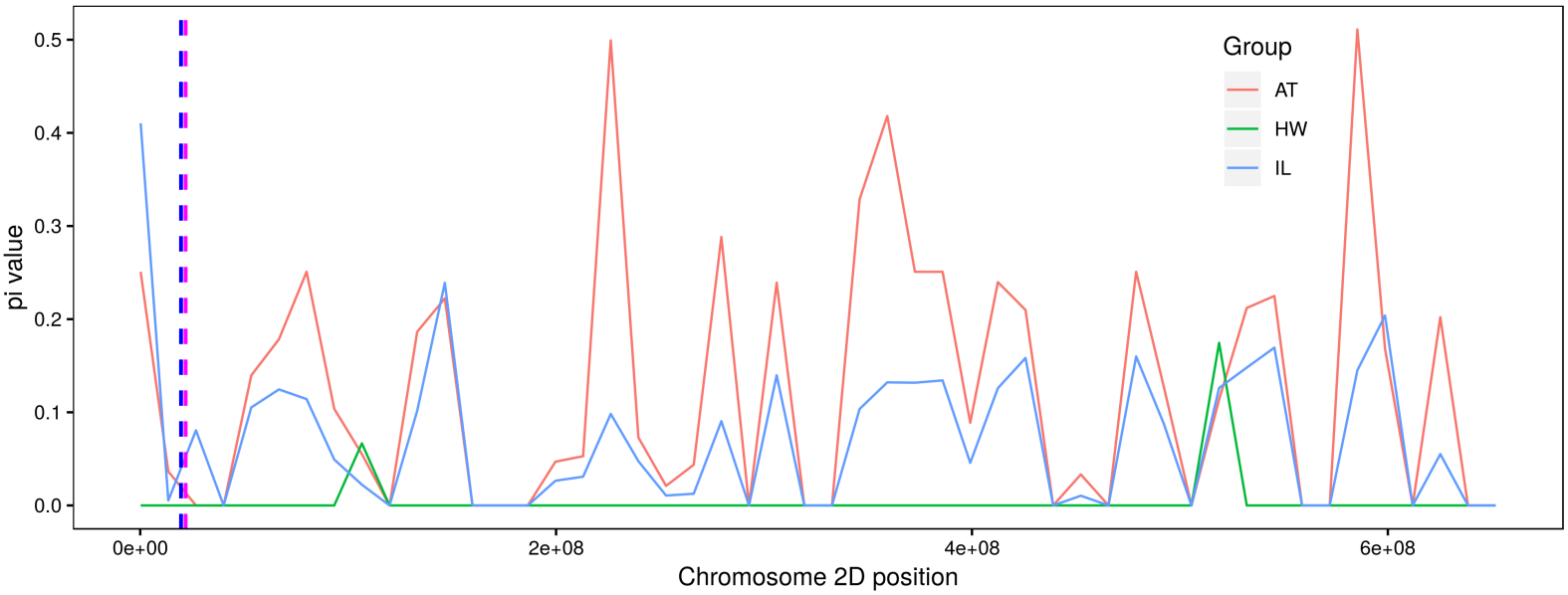
Fig. S2. Variation in nucleotide diversity per chromosome based on pi values interpolated using R function 'approx'.

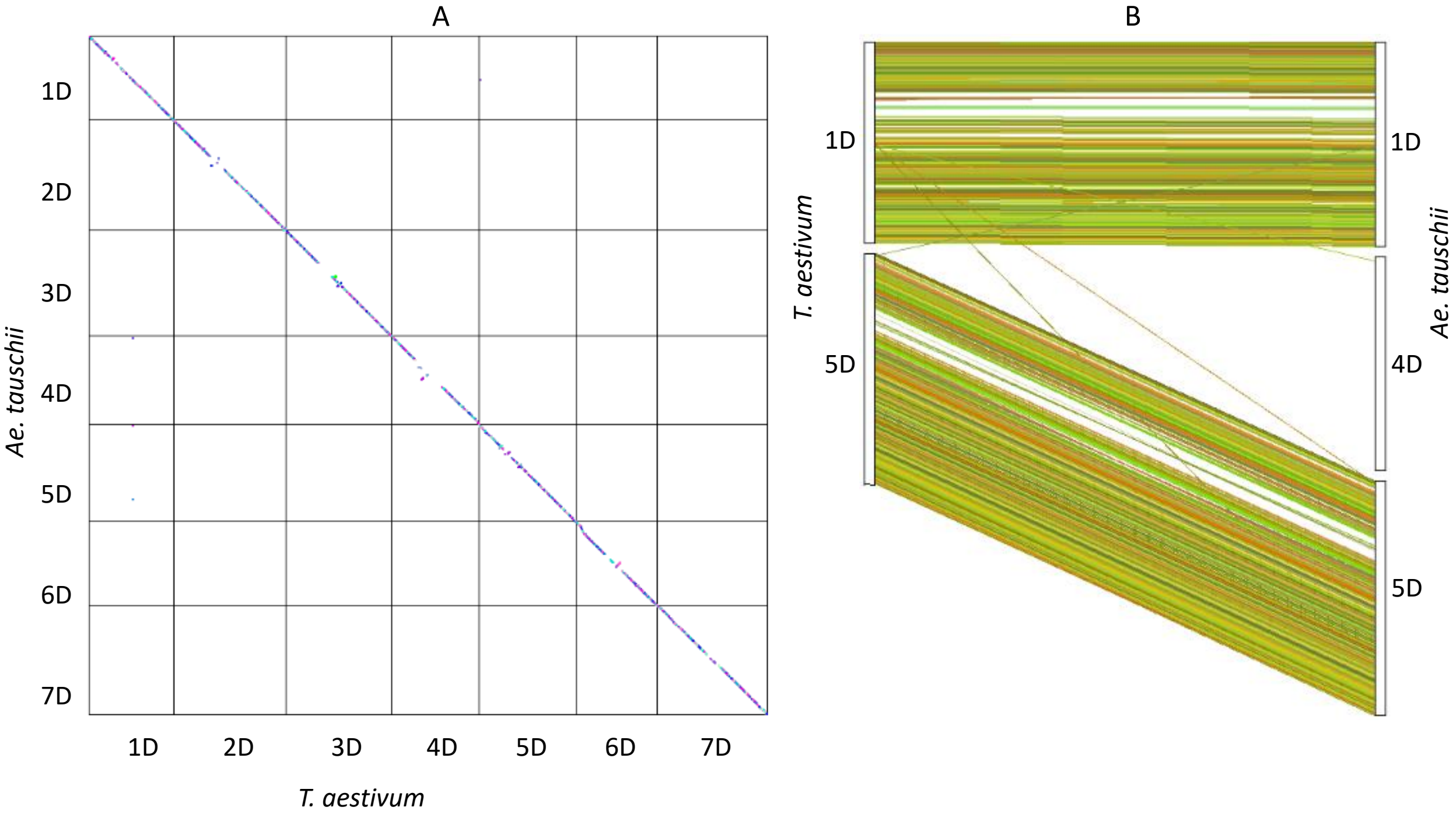
Fig. S3. Relationship between total recombination breaks and genetic distance in *Ae. tauschii* ssp. *strangulata* derived family FAM93 compared with *Ae. tauschii* ssp. *tauschii* derived families.

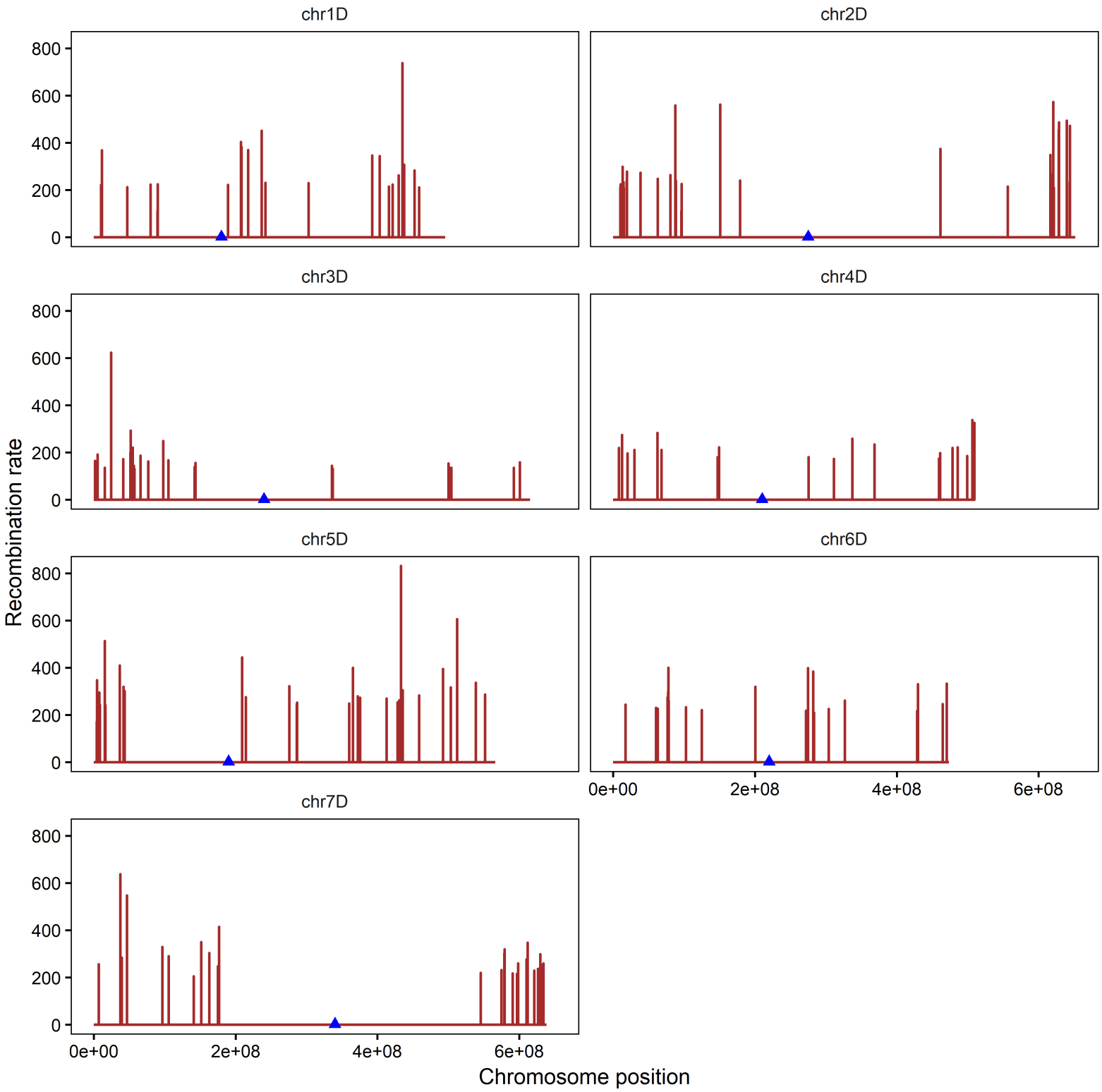

Group
AT
ILP_AT
L1a
L2
L1b
10
0
-10
PC2
-20
0
20
40
PC1

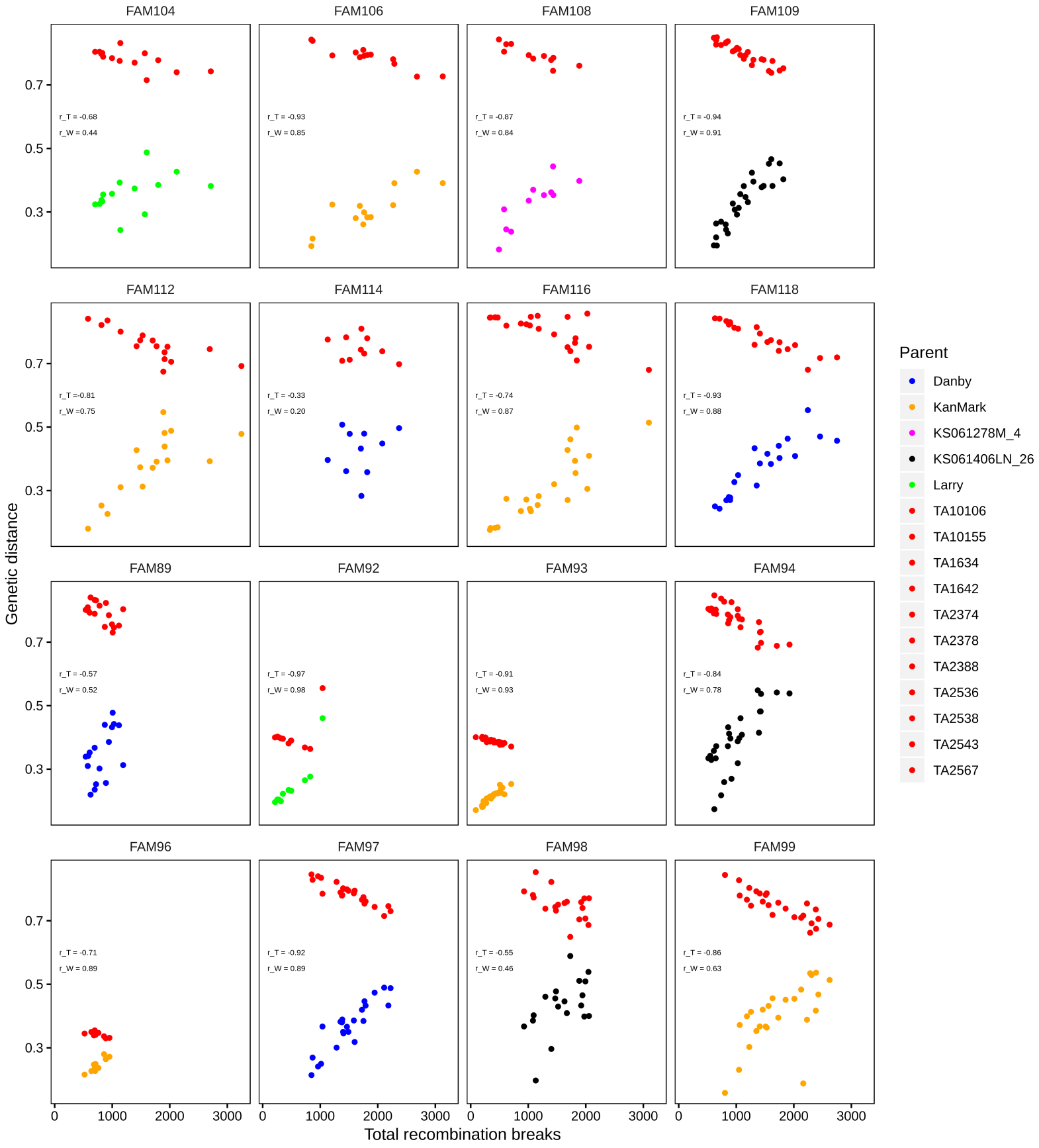

10
0
-10
-20
-30
PC2
-40
-20
0
20
PC1
L1a and L1b
L2
Group
AT
IL
ILP_AT
ILP_HW

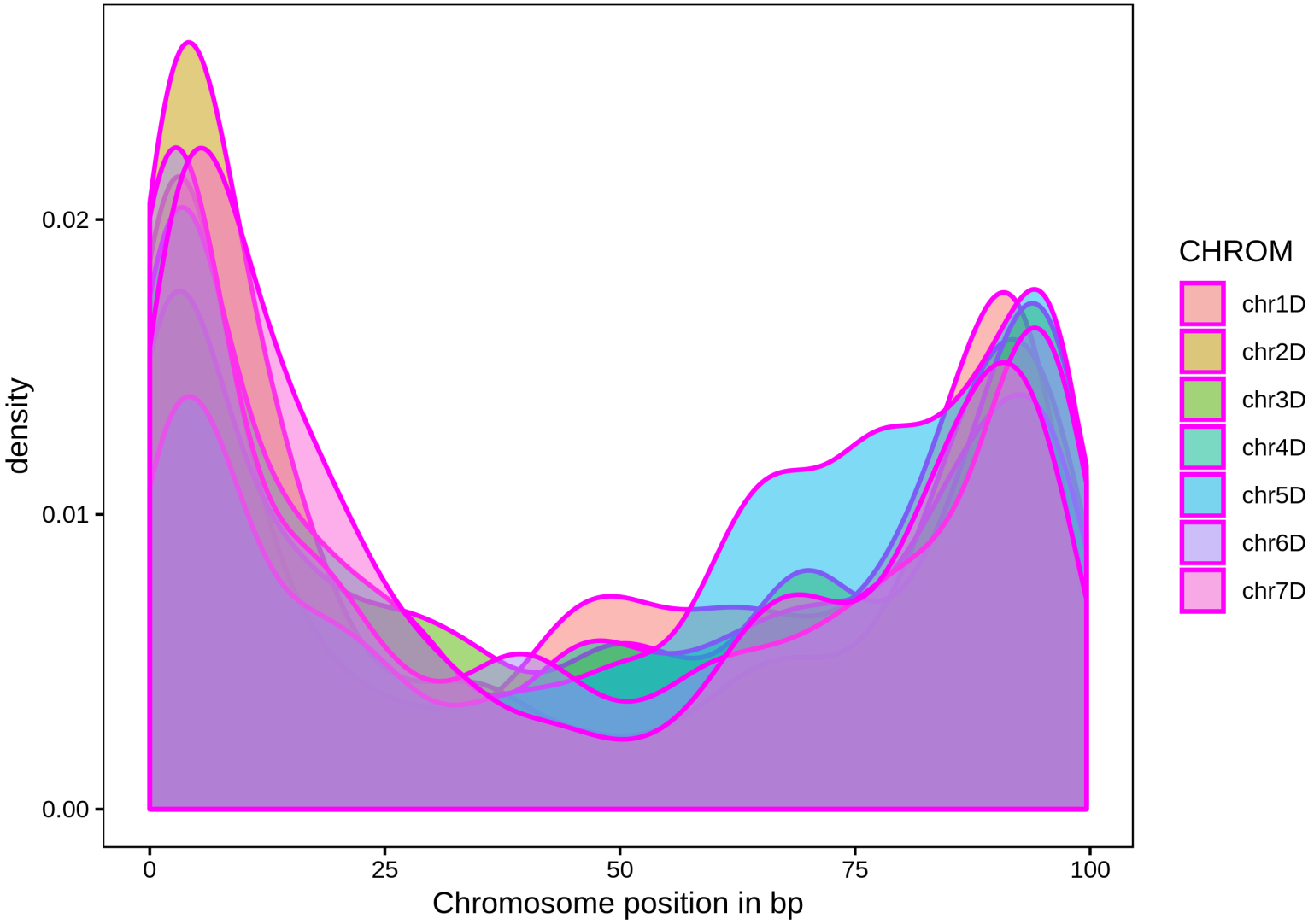

KS061406LN_26
KanMark
KS061862M_17
Larry
KS061278M_4
Danby
PC2
PC1
20
10
0
-10
-20
-20
-10
0
10
20
Group
Danby
KanMark
KS061278M_4
KS061406LN_26
Larry
Parent

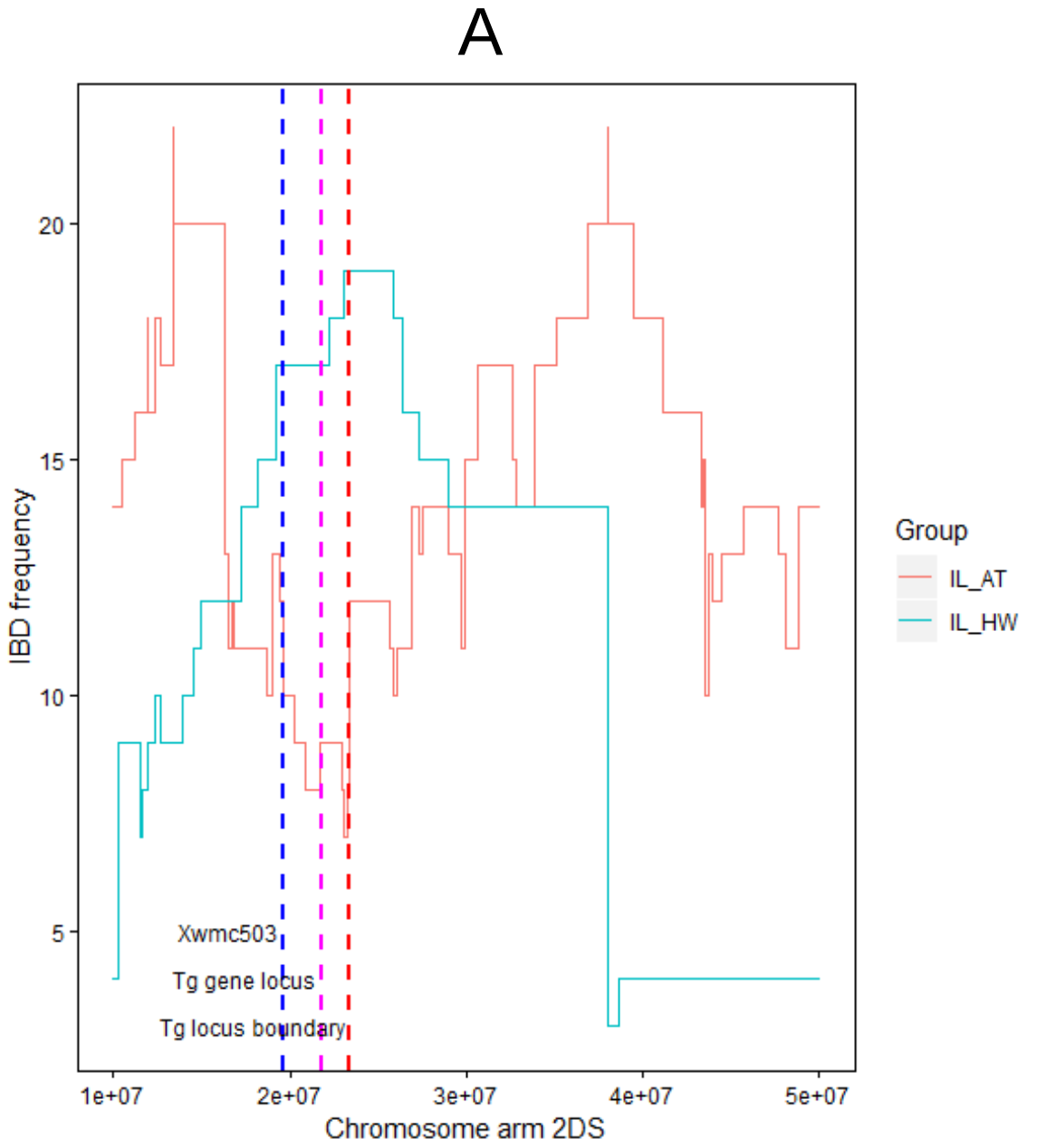

Group
AT
HW
IL
pi value
0.0
0.1
0.2
0.3
0.4
0.5
0e+00
2e+08
4e+08
6e+08
Chromosome 2D position


A
1D
2D
3D
4D
5D
6D
7D
Ae. tauschii
1D
2D
3D
4D
5D
6D
7D
T. aestivum
B
T. aestivum
1D
5D
Ae. tauschii
1D
4D
5D


chr1D
chr2D
chr3D
chr4D
chr5D
chr6D
chr7D
Recombination rate
Chromosome position
0
200
400
600
800
0e+00
2e+08
4e+08
6e+08


FAM104
r_T = -0.68
r_W = 0.44
FAM106
r_T = -0.93
r_W = 0.85
FAM108
r_T = -0.87
r_W = 0.84
FAM109
r_T = -0.94
r_W = 0.91
FAM112
r_T =-0.81
r_W =0.75
FAM114
r_T = -0.33
r_W = 0.20
FAM116
r_T = -0.74
r_W = 0.87
FAM118
r_T = -0.93
r_W = 0.88
FAM89
r_T = -0.57
r_W = 0.52
FAM92
r_T = -0.97
r_W = 0.98
FAM93
r_T = -0.91
r_W = 0.93
FAM94
r_T = -0.84
r_W = 0.78
FAM96
r_T = -0.71
r_W = 0.89
FAM97
r_T = -0.92
r_W = 0.89
FAM98
r_T = -0.55
r_W = 0.46
FAM99
r_T = -0.86
r_W = 0.63
Genetic distance
0.7
0.5
0.3
0
1000
2000
3000
Total recombination breaks
Parent
Danby
KanMark
KS061278M_4
KS061406LN_26
Larry
TA10106
TA10155
TA1634
TA1642
TA2374
TA2378
TA2388
TA2536
TA2538
TA2543
TA2567


density
0.02
0.01
0.00
0
25
50
75
100
Chromosome position in bp
CHROM
chr1D
chr2D
chr3D
chr4D
chr5D
chr6D
chr7D


A
IBD frequency
20
15
10
5
Xwmc503
Tg gene locus
Tg locus boundary
1e+07
2e+07
3e+07
4e+07
5e+07
Chromosome arm 2DS
Group
IL_AT
IL_HW

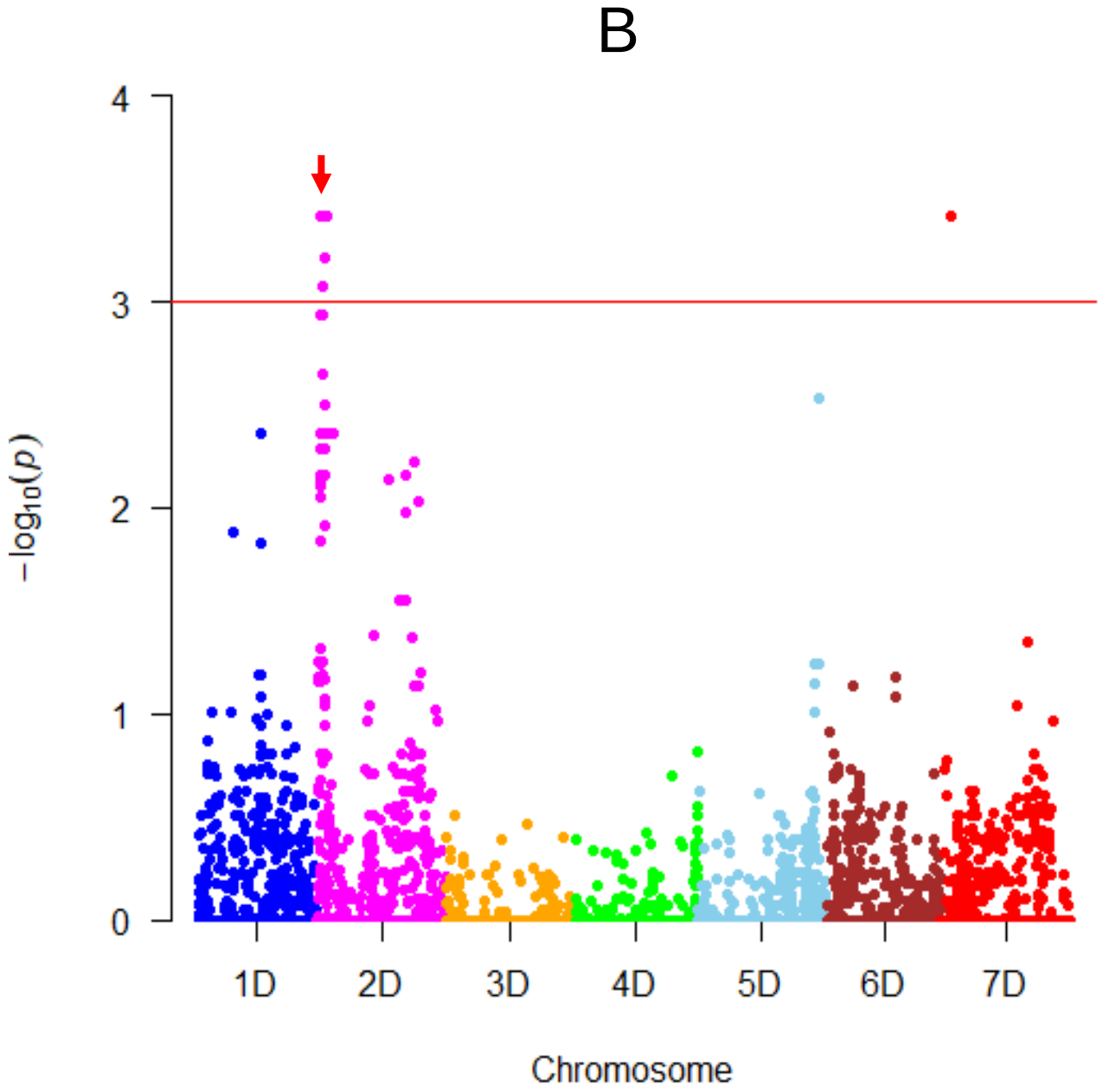

B
$-\log_{10}(p)$
4
3
2
1
0
1D
2D
3D
4D
5D
6D
7D
Chromosome

Table 1. A comparison of genetic diversity of *Ae. tauschii* derived introgression lines and their parents.

Group	Min.	Mean	Max.
chr1D_AT	0.000	0.098	0.512
chr1D_HW	0.000	0.014	0.545
chr1D_IL	0.000	0.083	0.501
chr2D_AT	0.000	0.107	0.512
chr2D_HW	0.000	0.013	0.546
chr2D_IL	0.000	0.065	0.501
chr3D_AT	0.000	0.110	0.512
chr3D_HW	0.000	0.009	0.545
chr3D_IL	0.000	0.047	0.499
chr4D_AT	0.000	0.112	0.512
chr4D_HW	0.000	0.004	0.485
chr4D_IL	0.000	0.120	0.501
chr5D_AT	0.000	0.105	0.512
chr5D_HW	0.000	0.011	0.545
chr5D_IL	0.000	0.115	0.501
chr6D_AT	0.000	0.114	0.512
chr6D_HW	0.000	0.010	0.545
chr6D_IL	0.000	0.090	0.501
chr7D_AT	0.000	0.108	0.512
chr7D_HW	0.000	0.010	0.545
chr7D_IL	0.000	0.079	0.501

AT are *Aegilops tauschii* parents, HW are hexaploid wheat parents and IL are the introgression lines

Table 2. Summary of recombination hotspots at 99th percentile of total recombination breakpoints from 16 families of the introgression population.

Chromosome	No. of windows	Min recombination	Max recombination
chr1D	23	211	738
chr2D	29	210	573
chr3D	26	131	623
chr4D	21	171	338
chr5D	28	241	831
chr6D	22	210	400
chr7D	32	205	637

Table 3. Percentage length of identity by descent segments shared between introgression lines, hexaploid wheat and *Ae. tauschii* accessions.

Chromosome	Mean (%)	SD (%)	Min (%)	Max (%)
chr1D[a]	4.84	5.45	0.32	26.97
chr1D[b]	12.07	13.00	0.54	47.72
chr1D[c]	11.57	16.17	0.31	95.23
chr2D[a]	3.88	4.60	0.33	41.33
chr2D[b]	6.61	10.90	0.90	55.89
chr2D[c]	11.92	19.32	0.30	91.72
chr3D[a]	2.69	2.57	0.28	21.12
chr3D[b]	8.70	12.01	0.95	38.88
chr3D[c]	12.12	20.24	0.26	100.00
chr4D[a]	6.98	6.36	0.49	38.26
chr4D[b]	26.63	21.77	2.55	86.18
chr4D[c]	17.78	23.49	0.33	100.00
chr5D[a]	3.35	3.40	0.46	24.50
chr5D[b]	6.74	8.68	0.88	58.99
chr5D[c]	8.30	11.90	0.41	85.68
chr6D[a]	4.88	4.03	0.33	20.73
chr6D[b]	6.16	10.86	0.58	69.88
chr6D[c]	12.88	20.78	0.33	100.00
chr7D[a]	3.07	4.01	0.29	24.99
chr7D[b]	6.76	9.78	0.44	48.29
chr7D[c]	8.50	13.97	0.24	91.27

[a] Introgression lines x *Ae. tauschii*

[b] Introgression lines x Hexaploid wheat

[c] Introgression line x Introgression line