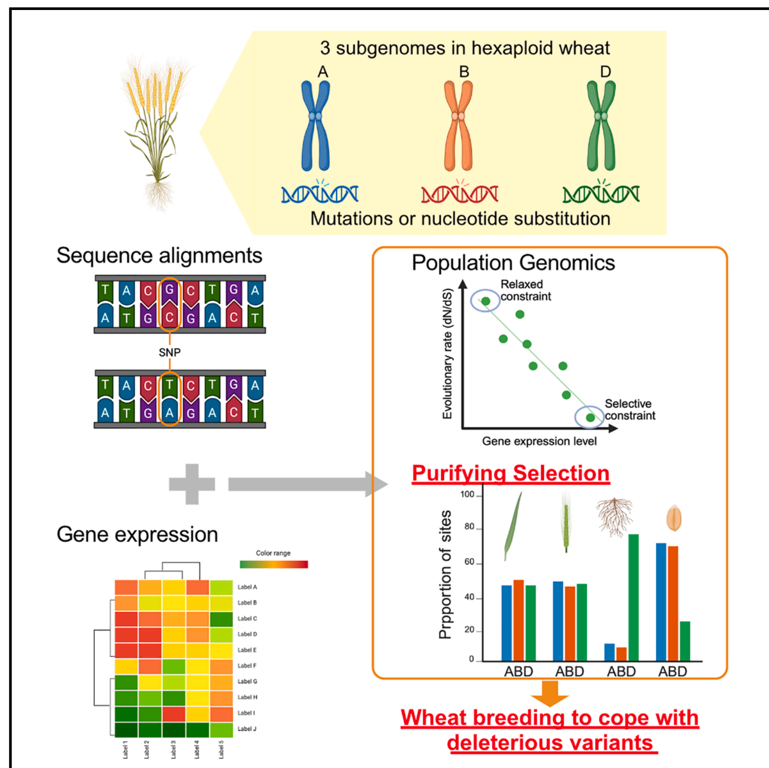
Cell Reports

Article

Purifying selection on deleterious variants affected by the combination of subgenomes and gene expression in bread wheat

Graphical abstract



Authors

Gwyneth Halstead-Nussloch, Moeko Okada, Georg Haberer, ..., Manuel Spannagl, Timothy Paape, Kentaro K. Shimizu

Correspondence

timothy.paape@usda.gov (T.P.), kentaro.shimizu@uzh.ch (K.K.S.)

In brief

Halstead-Nussloch et al. quantify genome-wide purifying selection on deleterious variants using high-quality annotation of 10 genomes of bread wheat. Strong purifying selection is observed in the category of genes that are highly expressed in root tissues and are in the D subgenome, suggesting a target of breeding to mitigate deleterious variants.

Highlights

- Genome-wide analysis of polymorphisms and purifying selection in bread wheat
- *De novo* assemblies and transcriptome data of the 10+ Wheat Genome Project
- Strong purifying selection in the D homoeologs with high expression in the root
- These findings will contribute to wheat breeding to cope with deleterious variants

Halstead-Nussloch et al., 2025, Cell Reports *45*, 116785
February 24, 2025 © 2025 The Author(s). Published by Elsevier Inc.
https://doi.org/10.1016/j.celrep.2025.116785

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Purifying selection on deleterious variants affected by the combination of subgenomes and gene expression in bread wheat

Gwyneth Halstead-Nussloch,[1,12] Moeko Okada,[1,2,3,12] Georg Haberer,[4] Thomas Lux,[4] A.S.M. Faridul Islam,[5] Masaomi Hatakeyama,[1,6] Roman Briskine,[1,7] Benjamen White,[8] Anthony Hall,[8] Curtis Pozniak,[9] 10+ Wheat Genome Project, Manuel Spannagl,[4] Timothy Paape,[1,10,11,*] and Kentaro K. Shimizu[1,2,13,*]
[1]Department of Evolutionary Biology and Environmental Studies, University of Zurich, 8057 Zurich, Switzerland
[2]Kihara Institute for Biological Research, Yokohama City University, Yokohama 244-0813, Japan
[3]Graduate School of Science and Technology, Niigata University, Niigata 950-2181, Japan
[4]Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany
[5]Institute for Advancing Health Through Agriculture, Texas A&M, College Station, TX, USA
[6]Functional Genomics Center Zurich, 8057 Zurich, Switzerland
[7]Science IT, University of Zurich, 8057 Zurich, Switzerland
[8]Earlham Institute, Norwich Research Park, Norwich NR4 7UH, UK
[9]Crop Development Centre and Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, Canada
[10]USDA-ARS Responsive Agriculture Food Systems Research Unit, College Station, TX, USA
[11]Department of Soil and Crop Sciences, Texas A&M, College Station, TX, USA
[12]These authors contributed equally
[13]Lead contact
*Correspondence: timothy.paape@usda.gov (T.P.), kentaro.shimizu@uzh.ch (K.K.S.)
https://doi.org/10.1016/j.celrep.2025.116785

SUMMARY

The genome-wide pattern of purifying selection on deleterious variants has attracted broad interest in breeding and evolutionary biology. Little is known about purifying selection in polyploid species with potential functional redundancy of duplicated homoeologs. Here, we examine the pattern of polymorphisms and the combined effects of both subgenome and gene expression on estimates of purifying selection in the allohexaploid bread wheat. We utilize the *de novo* annotated assemblies and transcriptome data of the 10+ Wheat Genome Project. Purifying selection is detected after filtering chromosomes harboring segregating foreign introgression by modern breeding. Interestingly, strong purifying selection is observed in the category of genes that are highly expressed in root tissues and are in the D subgenome, which is thought to contribute to the range expansion after allopolyploidization. We suggest that the combination of expressed tissues and subgenomes will inform the target of future wheat breeding to mitigate the effect of deleterious variants.

INTRODUCTION

Population genomic data are key to examining the genome-wide pattern of polymorphisms and selection that underlie genome evolution and domestication.[1] The patterns of deleterious variants under purifying selection, along with neutral and adaptive variants, have attracted broad interest because of their importance in crop breeding as well as their evolutionary effects on natural species.[2,3] Deleterious variants are considered to underlie heterosis or heterozygote advantages to overcome inbreeding depression.[4–6] Increasing crop yields by mitigating the effects of deleterious variants is one of the major goals of genomics-assisted breeding and future breeding termed Breeding 4.0.[7–9] Two groups of methods are used to study deleterious variants under purifying selection.[10] The first group is site-based and identifies mutations at phylogenetically conserved nucleotides or high-impact mutations based on gene annotations.[11,12] The second group requires genome-wide polymorphism data to estimate the strength of purifying selection using the site frequency spectrum. Purifying selection on nonsynonymous (or amino acid replacement) nucleotide substitutions makes their frequency lower (singletons in particular) compared to synonymous (or silent) mutations that are considered effectively neutral and thus can be exploited to estimate the distribution of fitness effects (DFE).[10,13,14] The DFE analyses show the proportions of neutral ($N_e s$ 0–1, where N_e is the effective population size, and s is the selection coefficient), slightly deleterious ($N_e s$ 1–10) to strongly deleterious nucleotide substitutions ($N_e s > 10$) estimated from genome-wide polymorphism data.[10,13,14] In many wild and crop plant species, DFE analyses in maize, rice, and cassava[15,16] showed that the majority of nonsynonymous mutations are under purifying selection rather than under

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

neutral evolution. Furthermore, the proportions of variants under neutrality are reduced in subsets of genes that are highly expressed.[17–20] These studies show that the expression level is a major determinant of the evolutionary rate through the strength of purifying selection due to greater selective constraint on highly expressed genes.[21]

Polyploid species are widespread among natural and crop plants[22,23] and are classified into autopolyploid and allopolyploid species, which were derived from genome duplication within a species or between species, respectively. The genomes of allopolyploid species are composed of subgenomes derived from different progenitor species, and duplicated homoeologous genes on different subgenomes can confer functional redundancy. Theoretical studies suggested that masking of phenotypic effects by redundancy can affect the pattern of polymorphisms and selection in allopolyploid species.[1,24] In functional genetics studies, often there is no change in qualitative phenotype by the disruption of a single homoeolog, which can be interpreted as the weakening of purifying selection on deleterious mutations in allopolyploid species due to sheltering of these mutations by duplicated homoeologs.[25] The gradual loss of homoeologs in ancient polyploidy is also consistent with the functional redundancy of homoeologs.[22] In contrast, the maintenance of homoeologs may be important for dosage balance and diversification.[25] Maize is a diploid species but experienced whole genome duplication relatively recently (about 5–12 million years ago). Its chromosomal regions derived from two subgenomes were reconstructed, and the different strengths of purifying selection were detected in the two subgenomes using the DFE.[26] It was suggested that the fractionation (gene loss after polyploidization due to redundancy of duplicated genes) was biased to one of the two subgenomes, in which reduced gene expression led to a weaker purifying selection.[26] Cotton is a relatively old polyploid species (1–2 million years ago) and similarly showed different strengths of purifying selection in different subgenomes.[27] Recent advances in genome sequencing and bioinformatics have opened the way for high-quality genome assembly and polymorphism analysis of allopolyploid species by distinguishing highly similar homoeologous sequences.[23,25,28–31] In model allopolyploid species with relatively small genomes, such as *Capsella bursa-pastoris* and *Arabidopsis kamchatica*, the DFE analysis suggested that purifying selection on deleterious mutations could be detected in allopolyploid species despite the presence of duplicated homoeologs.[25,28,30] In contrast to maize or cotton, no clear difference in the strength of purifying selection was detected between the subgenomes in *C. bursa-pastoris* and *A. kamchatica* despite the difference between diploid progenitor species.[28,30,32]

Bread wheat (*Triticum aestivum*), also called common wheat, is an allohexaploid species (AABBDD genome) used for making bread and noodles. The first polyploidization events (estimated < 0.5 million years ago) yielded the tetraploid AABB genome.[33–35] This was followed by the second event about 8,000 years ago in the Neolithic period that incorporated the D subgenome from a wild diploid species *Aegilops tauschii*, which produced hexaploid *T. aestivum*.[33–35] The low level of nucleotide diversity in the D subgenome indicated a bottleneck at the polyploid speciation event, in which only a small number of haplotypes of *Ae. tauschii* contributed to the formation of the new hexaploid.[36,37] Yet, the acquisition of the D genome from *Ae. tauschii* is considered to have played a key role in the global spread of bread wheat, providing it with broad adaptability to diverse environments.[33–35] This contrasts with diploid and tetraploid domesticated wheat species, which are typically cultivated in more restricted regions such as the Transcaucasus and Mediterranean.[33–35] The D genome is thought to contribute beneficial traits such as disease resistance[38] and salt stress tolerance through subgenome-biased gene expression.[39–41] Purifying selection in bread wheat was detected by genotype-by-sequencing and exon sequence capture,[42,43] which showed a low frequency of nonsynonymous polymorphisms.[42–44] However, compared with the quantitative studies of purifying selection using the DFE in maize and other diploid crop species,[10] little is known about purifying selection on deleterious variants of bread wheat because of the complexity of its large allohexaploid genome (∼15 Gb). Recently, *de novo* chromosome-level assembly[31,45,46] and annotation using pan-transcriptome data[47] of 10 cultivars of bread wheat by the 10+ Wheat Genome Project provided an ideal dataset for genome-wide polymorphism analysis. Because all cultivars were assembled *de novo*, they are free from the bias of variant calling that can arise when using only a single reference genome. The chromosome-scale *de novo* assemblies revealed that many introgressed chromosomal segments from other species are segregating in bread wheat, which we call segregating foreign introgressions.[31,45,48] While some of the segregating foreign introgressions of the 10 cultivars can be traced to known interspecific crosses during breeding in recent decades, some introgressions are also found in old cultivars,[45,48] suggesting that some introgressions existed before modern breeding.

In this study, we constructed coding sequence alignments of the newly annotated 10 genomes of bread wheat, including rye or barley as an outgroup (Figure 1). Furthermore, because introgressions are observed in many species, such as the ones from Neanderthal to modern humans, and can affect the pattern of polymorphisms and signature of selection,[49] we generated datasets excluding chromosomes harboring large segregating foreign introgression identified by the genome-wide survey by Keilwagen et al.[48] to examine the effect of these introgressions on the pattern of polymorphisms and signature of selection. These alignment datasets based on each *de novo* assembled genome sequence allow us to identify differences in the strength of purifying selection among subgenomes of wheat during allopolyploidization and following the expansion of its geographical distribution.[33–35] Although previous studies showed the effect of subgenome and gene expression on purifying selection separately in polyploid species, little is known about their joint effects.[33–35] By integrating polymorphism and transcriptome data (Figure 1), we addressed the following questions: how do the chromosomes harboring segregating foreign introgressions affect the pattern of polymorphisms and signatures of selection in the newly annotated 10 genomes of bread wheat? Are there joint effects of subgenome and gene expression on the signature of purifying selection? What is the pattern of purifying selection of the D subgenome that is known to have contributed to range expansion to broader environments?

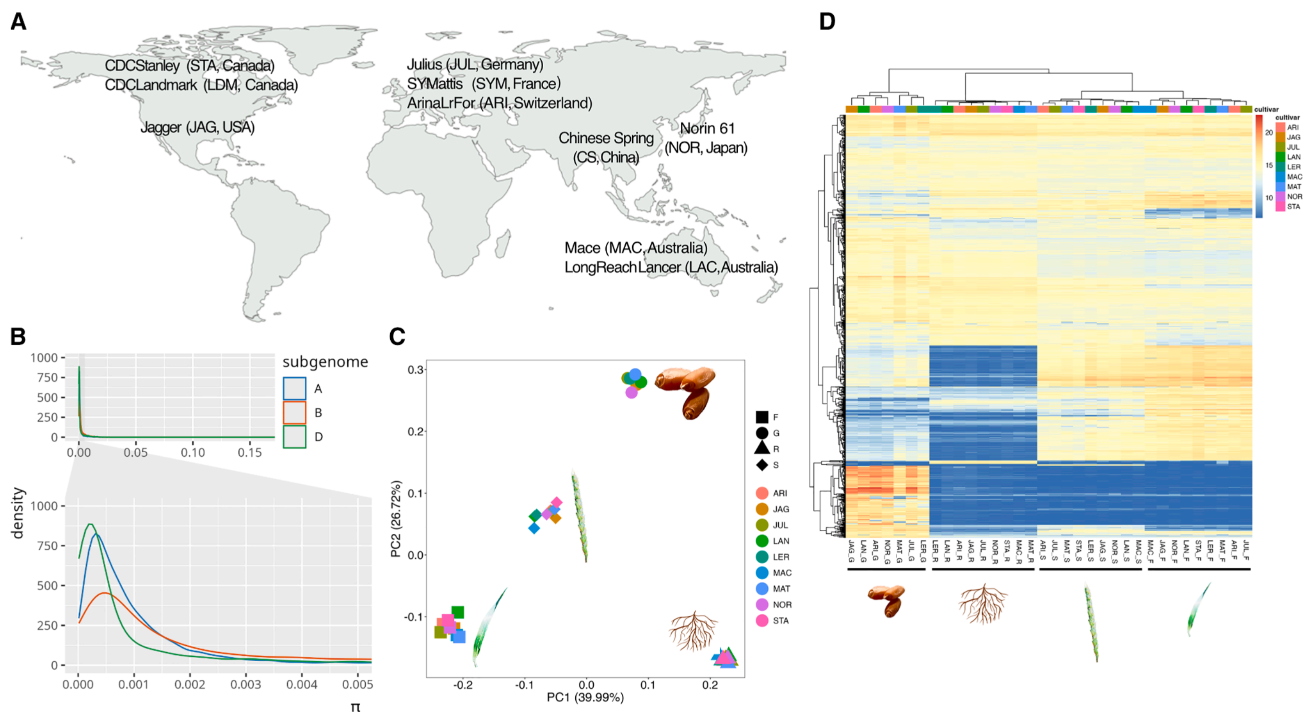


Figure 1. The genome-wide polymorphism and transcriptome data of bread wheat
(A) Country of origin of the 10 cultivars and the abbreviation of each cultivar.
(B) Density plot of the nucleotide diversity π of the three subgenomes A (blue), B (red), and D (green) with rye as the outgroup. Zoom of the range between 0.000 and 0.005 highlights the difference between the three subgenomes.
(C) Principal component analysis of the transcriptome data of grain, flag leaf, root, and spike. Tissues are indicated by icons and shapes, and cultivars are shown by color. Principal components 1 and 2 (PC1 and PC2) are shown.
(D) Two-dimensional clustering of the transcriptome data for the homoeologous gene sets analyzed in this study.
See also Tables S1 and S2.

RESULTS

Sequence alignments

The 10 cultivars used for *de novo* assembly, annotated by full-length cDNA sequences[45–47] were derived from 8 countries across Europe, Asia, North America, and Australia (Figure 1A). As an outgroup, we used the genomes of either rye (*Secale cereale*)[50] or barley (*Hordeum vulgare*),[51] the former of which is more closely related to wheat. The outgroup sequence was used to guide the identification of homoeologous sequences, together forming orthologous groups (OGs), and to identify fixed and segregating variants. We excluded alignments in which no sequences were available from one or more of the 10 cultivars, and generated alignments of the sequences that were present in all three subgenomes in a 1:1:1 ratio and had an orthologous sequence identified in an outgroup (see STAR Methods). After filtering, a total of 10,602 OG alignments were analyzed with rye as an outgroup and 10,604 with barley as an outgroup. These sequence alignments represent a highly accurate, albeit conservative, dataset for genome-wide polymorphism and divergence analyses of bread wheat (dataset 1 = full dataset).

To examine the effect of segregating foreign introgressions on the pattern of polymorphisms and signature of selection, we further constructed two more datasets. The introgressions that were derived from other species and segregating among the 10 genomes identified in the genome-wide scan by Keilwagen et al. are all located on chromosomes 2A, 2B, 2D, and 3D.[48] Dataset 2 encompasses 6 of the 7 trios of chromosomes and excludes the trio of whole chromosome 2 that includes most of these introgressions (i.e., without 2A, 2B, and 2D). In dataset 3, we excluded all 7 introgressions and included 5 out of 7 trios of whole chromosome sequences (i.e., without 2A, 2B, 2D, 3A, 3B, and 3D).

Level of polymorphisms among subgenomes and the effect of segregating foreign introgressions

We calculated standard population genetic indices for the total dataset of the ten wheat cultivars in the full dataset (dataset 1). Total nucleotide diversity (π) and Watterson's theta (θ_W) were highest in the B subgenome, followed by the A subgenome and then by the D subgenome (Figure 1B; Tables S1 and S2), consistent with previous reports.[42,43,45] Tajima's *D* was most negative in the D subgenome, indicating an excess of rare alleles. The low level of polymorphism and negative Tajima's *D* value within the D subgenome is consistent with a bottleneck following hexaploidization and subsequent population expansion.[42,43,45] Mean d_N/d_S values were about 0.20 in all the subgenomes, which suggests purifying selection on nonsynonymous

mutations (Table S1) ($d_N/d_S > 1$ indicates more amino acid changing substitutions and positive selection, and $d_N/d_S < 1$ indicates fewer amino acid changes and purifying selection).

Next, we examined the effect of the choice of the outgroup (rye or barley) on these population genetic indices. The effect sizes of all comparisons showed a negligible effect (Cliff's delta[52–54] with lower and upper bounds for 95% confidence interval < 0.11, Bayesian Cohen's d[55] with 95% confidence interval < 0.2) (Tables S1 and S3). Corrected p values were not significant except for d_N/d_S (Tables S1 and S3). These analyses indicate that the OG formation using either outgroup was consistent. Below, we will mainly present the results using rye, which is phylogenetically closer to wheat. Finally, we examined the difference between datasets, i.e., the exclusion of chromosomes with the segregating foreign introgressions. Based on Cliff's delta, the effect of most comparisons was negligible (Table S4). Small effect was detected in the comparisons of Tajima's D, θ_W, π, and π_S in the B subgenome between datasets 1 and 2 and between datasets 1 and 3 (0.11 < Cliff's delta < 0.28[54]), which were also supported by corrected p values (Table S4). According to Cohen's d, the effect size was also negligible except for the comparison of Tajima's D of the B subgenome between datasets 1 and 2 (Tables S1 and S4). These results are consistent with the fact that the two introgressions in the B subgenome introduce more singletons in the full dataset (i.e., only found in a single cultivar), which increases rare variants in the full dataset (Tables S1 and S4). Mean d_N/d_S values were very similar among the three datasets (Tables S1 and S4). These analyses suggest that the segregating foreign introgressions increased the genome-wide polymorphism levels of the elite cultivars of the 10+ Wheat Genome Project, but the effect on overall genetic diversity was relatively small (Table S4).

Next, genealogies were constructed for each subgenome for the three datasets (Figures S1–S3). Rye or barley was separated by a long branch from all bread wheat cultivars, consistent with their outgroup status (Figures S1–S3). The longest segregating foreign introgression reported for the 10+ genome[45] spans most of chromosome 2B (nearly 427 Mb) derived from *Triticum timopheevii* into the LongReach Lancer (LAC) cultivar. In the genealogy of the B subgenome using the full dataset (Figure S1B), LAC had a relatively long branch and was located at the base of wheat cultivars. In contrast, the genealogy of dataset 2 without the trio of chromosome 2 (Figure S3B) did not show this pattern, suggesting the effect of the large introgression on the genealogy. We could not find a clear geographic pattern in the genealogy, except for the clustering of two East Asian cultivars, Norin 61 and Chinese Spring, in some genealogies (e.g., the B subgenome of the whole dataset in Figures S1B and S4).

Correlation of population genetics statistics between homoeologs

Differences in selection among subgenomes, especially between A and B subgenomes, may not be apparent at chromosome scales, but may be detectable in correlations of population genetics statistics of homoeolog pairs (coding sequences) from each subgenome. To examine whether homoeologous genes in different subgenomes showed a similar or different pattern of polymorphisms and signatures of selection,[30,45] we calculated the correlation of each of the population genetics statistics between the A, B, and D homoeologs (Table S5). Tajima's D, a standard neutrality test to detect positive and balancing selection, showed a low correlation among the homoeologs (<0.1 in all combinations) (Table S5). The correlation of nucleotide diversity (π) was similarly low ($\leq$0.25; Table S5). These results suggest that each homoeolog experienced generally different histories,[30,45] despite similar overall subgenome-wide means in genetic diversity (Table S5).

Considering that the subgenomes A and B merged first and shared longer histories as an allotetraploid species, a higher correlation between the A and B subgenomes of these statistics may be expected. However, among the three comparisons (AB, AD, and BD) of π and θ_W, the correlations of π and θ_W between B and D homoeologs were the highest in all three datasets (Table S5). The similarity between B and D homoeologs may be attributed to different patterns of selections, hybridizations, or uncharacterized introgression events.[56,57]

Derived allele frequencies affected by introgressions

We examined the distribution of site frequency spectra, which can reveal further information about selection acting on a population. We used the rye and barley outgroups to polarize nonsynonymous ("replacement" substitutions that result in a nucleotide change in a codon that changes the amino acid sequence) and synonymous ("silent," no change in amino acid) polymorphic sites to construct a derived allele frequency (DAF) spectrum for each subgenome. In all three subgenomes, the proportions of low-frequency nonsynonymous variants, i.e., a nucleotide substitution in only one or two individual cultivars, were higher than synonymous variants. This is most conspicuously seen as the higher proportion of nonsynonymous singletons (allele frequency 1 in the x axis of Figure 2) than synonymous singletons (Figure 2 and S4). This suggests that purifying selection is acting to keep deleterious amino acid-changing alleles rarer.[43]

When the entire chromosomes were used (dataset 1), the shape of the DAF spectrum contained conspicuous peaks (highest at frequencies of 4 in the D subgenome) (Figure 2A). This suggested that the segregation of foreign introgressions caused an excess of intermediate frequency alleles in the center of the DAF. Using dataset 3 without chromosomes 2 and 3, intermediate frequency peaks in the DAF spectrum were absent (Figure 2C). This pattern is consistent with the introgression in chromosome 2D that is shared by four cultivars.[48] The intermediate frequency peaks in the DAF disappeared when using dataset 2, which excluded the trio of chromosome 2 but included the trio of chromosome 3 (Figure 2B). The similar results of datasets 2 and 3 are consistent with the contribution of the introgression in the trio of chromosomes 2 to the DAF spectrum, while the effect of introgression in the trio of chromosome 3 was small in the D subgenome. The analysis using barley as the outgroup showed very similar results (Figure S4). These analyses showed that segregating foreign introgressions had a large effect on the DAF of the 10 genome data. In contrast, the higher proportion of nonsynonymous singletons compared to synonymous singletons was found to be clearly similar to the entire chromosome dataset, suggesting purifying selection (see the section "purifying selection acting in subgenomes of hexaploid wheat").

CellPress
OPEN ACCESS

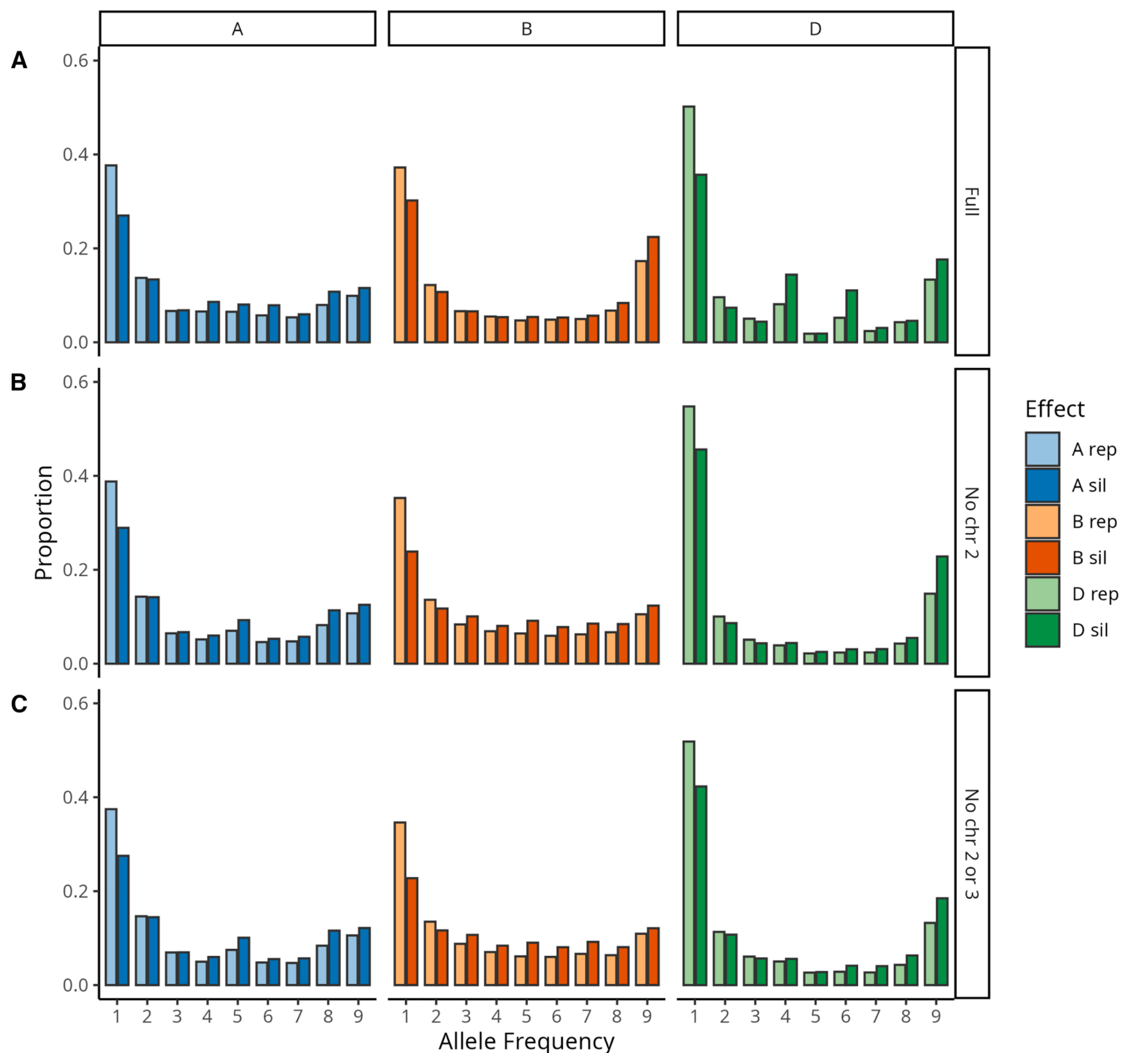


Figure 2. Derived allele frequency spectrum with rye as an outgroup
(A) Dataset 1 (full dataset), (B) dataset 2 (without the trio of chromosome 2), and (C) dataset 3 (without trios of chromosomes 2 and 3). Results of the A (blue), B (red), and D (green) subgenomes are shown. Polymorphic sites were polarized using rye as an outgroup, and sites were classified as either a nonsynonymous replacement (bars with light color, rep) or a synonymous silent (bars with dark color, sil) substitution.
See also Figure S4.

Furthermore, we examined the effect of these introgressions on the estimation of the proportion of adaptive substitutions (α) that were fixed by positive selection. In contrast to zero or negative values, positive α values indicate adaptive substitutions among multiple sequence alignments,[58] but it is known to be sensitive to previous demographic changes.[14] Using the dataset of entire chromosomes, the average α of B homoeologs was positive (Figure S5). However, in datasets 2 and 3, in which chromosomes harboring segregating foreign introgressions were removed, α values were not positive, and those of B homoeologs had near-zero mean α values (Figure S5). These results suggest that the introgressions inflated the estimates of α when using the complete set of chromosomes. This would be expected because the introgressions inherently result in greater divergence from the outgroup used in the alignments (rye or barley), which did not have homologous introgressions.

Purifying selection acting in subgenomes of hexaploid wheat

Based on the allele frequency spectrum (Figure 2) and genetic diversity estimates (Table S1), the higher proportion of low-frequency nonsynonymous mutations (i.e., singletons) compared with synonymous mutations in all three datasets suggested genome-wide purifying selection. To more quantitatively address whether purifying selection acted on the homoeologs of hexaploid wheat, we estimated the strength of purifying selection using the DFE.[13] The DFE shows the proportion of substitutions that range from neutral (N_es 0–1), slightly deleterious (N_es 1–10), to strongly deleterious (N_es > 10).[13,14] A high proportion

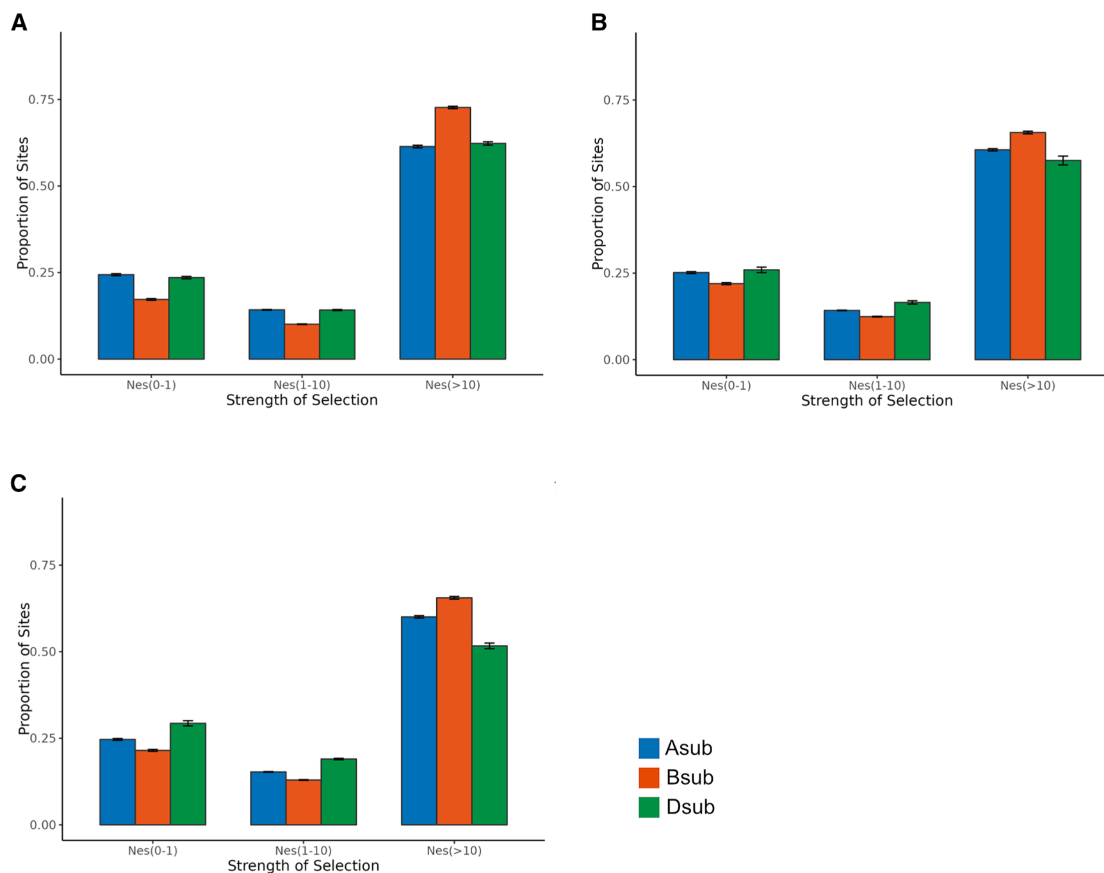


Figure 3. Distribution of fitness effects

DFE estimation for each subgenome using rye as an outgroup with (A) dataset 1 (full dataset), (B) dataset 2 (without the trio of chromosome 2), and (C) dataset 3 (without trios of chromosomes 2 and 3). N_es 0–1 are considered effectively neutral substitutions, N_es 1–10 slightly deleterious, and $N_es > 10$ strongly deleterious. A, B, and D homoeologs are shown by blue, red, and green, respectively. The DFE was estimated from 10^6 MCMC replicates. Error bars are standard deviations from the mean of 1,000 resampled MCMC replicates.

See also Figure S6.

of variants in the neutral category (N_es 0–1) is often used synonymously with "relaxed selective constraint," while high proportions of the $N_es > 10$ category suggest purifying selection owing to strong constraints on protein evolution. Together, the signature of purifying selection is detected when the DFE indicates a simultaneous lower proportion of the neutral category but high proportions of the deleterious categories. The results using the entire dataset (dataset 1) and datasets 2 and 3 without chromosomes with known segregating foreign introgressions were similar. Among all of the A, B, and D homoeologs, approximately 15%–30% of substitutions were classified as effectively neutral, 10%–20% as slightly deleterious, and >50% under strong purifying selection (Figures 3 and S6), providing evidence of purifying selection in all three subgenomes. The B homoeologs consistently showed a lower proportion of effectively neutral sites together with a higher proportion of deleterious categories than A and D homoeologs (Figures 3 and S6). These results suggest stronger purifying selection on B homoeologs, consistent with its relatively high genetic diversity estimates. Removal of homoeologs within introgressed chromosomes had minimal effect, increasing the proportion of sites under neutrality mostly within D homoeologs (Figures 3 and S6), where introgressions may contribute more significantly to rare and intermediate frequency substitutions. The use of rye (Figure 3) or barley (Figure S6) as the outgroup affected the results little.

Purifying selection is affected by gene expression levels

Gene expression is a major determinant of the strength of purifying selection, with highly expressed genes more often under strong selective constraint.[21] Thus, we included gene expression data from the second phase of the 10+ Wheat Genome Project[47] to test whether selective constraint varies at the homoeolog level in wheat when gene expression level is considered. In each of the A, B, and D homoeologs, we found that highly expressed genes (in the top 5% of gene expression in the subgenome) contained a lower proportion of sites classified as neutral (N_es 0–1) than homoeologs in each of the three subgenomes when all tissue data were analyzed (Figures 4A–4C). In contrast to the decreased proportion of effectively neutral sites of highly expressed genes, the proportion of strongly deleterious sites was increased,

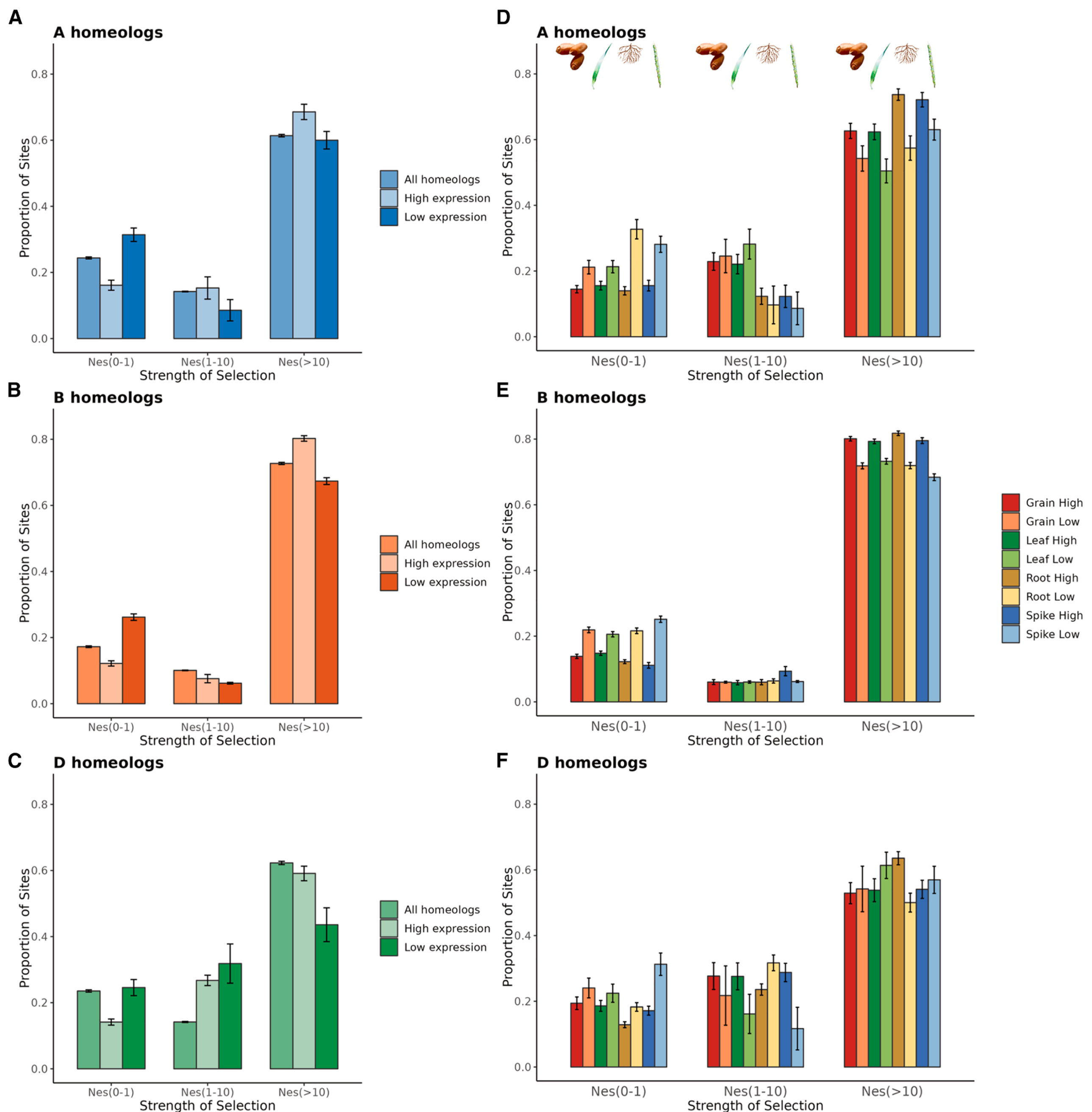


Figure 4. Distribution of fitness effects and expression levels

(A–C) DFE of all homoeologous copies within a subgenome alongside those that have the highest and lowest expression as determined by 5% cutoff using all tissue types. A, B, and D homoeologs are shown by blue, red, and green, respectively.

(D–F) DFE of the highest and lowest expressing homoeologs within specific tissue types. Four different tissues are indicated by colors as well as icons above. N_es 0–1 are effectively neutral substitutions, N_es 1–10 are slightly deleterious, and $N_es > 10$ are considered strongly deleterious. Tissues are indicated by icons and shapes corresponding to Figures 1C and 1D.

(A and D) A homoeologs, (B and E) B homoeologs, and (C and F) D homoeologs. Rye was used as the outgroup.

See also Figure S7; Table S6. The DFE was estimated from 10^6 MCMC replicates. Error bars are standard deviations from the mean of 1,000 resampled MCMC replicates.

particularly in A and B homoeologs. In the D subgenome, slightly deleterious sites were increased instead of strongly deleterious sites, consistent with weaker purifying selection. In contrast to the highly expressed genes, the genes with low expression (lower 5% of gene expression in the subgenome) showed weaker signatures of purifying selection. These results suggest

that genes that were highly expressed contained a higher proportion of sites under strong purifying selection (and strong selective constraint), and those that produced very few transcripts were under more relaxed constraint. Furthermore, the data provided additional support for purifying selection working in bread wheat, a polyploid crop species.

Tissue-specific expression is also a strong predictor of selective constraint.[17] Transcriptomics data of the wheat pangenome showed that gene expression clustered by tissue type (Figures 1C and 1D) more so than by cultivar. Taking advantage of gene expression data of different tissue types, we further analyzed the DFE using the highly and lowly expressed genes in grain, leaf, root, and spike. Among neutral categories of the four tissues, the difference between genes with high and low expression was pronounced in the spike tissue in all three subgenomes. The genes expressed in spikes may be more constrained, which could be explained by artificial selection on spike-related traits. Interestingly, we found that the subgenomes showed different patterns for different tissues. The proportion of the neutral sites in the category of the genes in the D subgenome with high expression in the root was very low (Figure 4F), which is lower than those in the other three tissues of the D subgenome, and among the lowest in all tissues of the three subgenomes. In addition, its proportion of the strongly deleterious category (N_es >10) was much higher than that with low expression in the root (Figure 4F). Very similar results were obtained with barley as the outgroup (Figure S7). Furthermore, a linear model (ANOVA) showed a significant interaction of the effects of tissue-specific expression and subgenomes, as well as the effect of each on the proportions (Table S6).

DISCUSSION

Globally, bread wheat is the most widely cultivated crop species,[59] but its complex allohexaploid genome with its huge genome size of ∼15 Gb was an obstacle to studying genome-wide patterns of polymorphisms and selection. Many studies have used reduced representation libraries, such as exon capture methods for cost-efficient sequencing[42,43] or relatively low coverage resequencing,[60,61] but these datasets and analyses were less complete compared with other major crop species, such as maize and rice, where high-coverage sequencing was used for diversity analyses.[9,62] Here, we analyzed a unique polymorphism dataset using the *de novo* assemblies of wheat homoeologs aligned together using rye and barley outgroups. An obvious limitation of the data is that the 10 samples focused on modern cultivars encompassing segregating chromosomal segments introgressed from different species either by modern breeding or unintentionally.[31,45,48] We found that the chromosomes harboring foreign segregating introgressions had large effects on analyses such as genealogy and the estimation of the proportion of substitutions fixed by positive selection (α), but a relatively small effect on nucleotide diversity and the DFE. Introgressed chromosomal segments appear to be a widespread feature of many species, including bread wheat[31,45,48] and human lineages,[49] and their complete removal is not feasible in many analyses. In addition, Norin 61 and Chinese Spring from Asia are shown to have diverged from European, American, and Australian modern cultivars.[45] Indeed, Norin 61 is a modern cultivar bred from East Asian landraces[31,63] and thus contributes to genetic variation that predates modern breeding. The advantages of the *de novo* assembly for genetic diversity analyses include the lack of ascertainment bias of using a single reference genome and the lack of mapping errors that can be pronounced in allopolyploid species,[64–66] particularly when repetitive regions are rampant in the genome.

In polyploid species, it is often thought that deleterious variants may be masked by the redundancy of duplicated homoeologous genes.[1,25] If one assumes that nonsynonymous mutations had no fitness effects due to complete redundancy, 100% of the nonsynonymous substitutions would be classified as neutral. However, we found that only approximately 15%–30% of nonsynonymous substitutions are under neutrality in the wheat gene alignments. This proportion of neutral variants is comparable to that reported using wild polyploid species. About 20% of the nonsynonymous substitutions of the allopolyploid species *A. kamchatica* and about 30% of the 0-fold nonsynonymous substitutions of *C. bursa-pastoris* were in the neutral category.[28,30] The comparison suggests a similar pattern of purifying selection in natural and crop allopolyploid species, which should be validated by the study of other polyploid species. Furthermore, selfing is known to decrease the efficacy of selection,[67–69] and selfing is common in polyploid species, including bread wheat.[67–71] Yet, the majority of the nonsynonymous mutations in bread wheat were under purifying selection. In the current study on wheat, the DFE analyses indicate that most nonsynonymous mutations have a slight but significant deleterious effect, and then their frequencies become lower than neutral. Together with previous reports, such as high-impact mutations,[42,43] our genome-wide quantification of purifying selection using the DFE suggested that deleterious variants under purifying selection are segregating in bread wheat despite being a hexaploid species. These results highlight the role of hidden deleterious variation in modern wheat cultivars. This provides a molecular basis for the effectiveness of hybrid breeding in wheat.[72–74] These many slightly deleterious mutations in wheat support the use of hybrid breeding to mask these mutations and enhance yield and vigor. Furthermore, it enables the development of tools to identify and purge deleterious alleles, contributing to modern cultivar improvement through genetic optimization.

The DFE analysis showed that the proportion of sites under neutrality (N_es 1–10) in the B subgenome was less than that of the A and D subgenomes, suggesting stronger purifying selection in the B subgenome. A potential hypothesis explaining the subgenome difference in purifying selection is subgenome expression dominance, which was proposed in previous analyses of genetic diversity in maize.[26] Maize experienced genome duplication 5–12 million years ago, and its current diploid genome is composed of the dominant subgenome (maize 1) and the recessive subgenome (maize 2). It was suggested that stronger purifying selection in the maize 1 subgenome was attributed to higher gene expression levels relative to maize 2.[26] However, we suggest that subgenome expression dominance is not a major mechanism of the different DFE patterns in bread wheat. In expression studies of bread wheat, there is no clear difference in the expression levels between the A, B, and D

subgenomes.[47,65,75,76] Although we showed a clear relationship between expression level and purifying selection in each subgenome, the expression level cannot be a major force of subgenome difference in purifying selection. An alternative hypothesis on the subgenome difference in purifying selection is the difference in genetic variation inherited from progenitor species through polyploidization. Recent genomic studies of polyploid species suggested the importance of variation inherited from progenitor species in addition to mutations at the time or following polyploidization.[25,28] In addition, the high outcrossing mating of the diploid *Ae*. *speltoides* that is closely related to the B subgenome ancestor may have contributed to the high variation of the B subgenome among hexaploid wheat lines,[43] and subsequently affected the efficacy of selection.

We unexpectedly found that genes that are located in the D subgenome and expressed highly in roots were under strong purifying selection compared with those of the other three tissues studied. To our knowledge, this is the first study to report the combined effect of subgenomes and tissue-specific expression on purifying selection. The D subgenome is known to have contributed to the drastic range expansion of bread wheat by conferring a broad environmental resilience, including temperature, precipitation, and salinity tolerance.[33,34] Bread wheat inherited an increased number of lateral roots that can enhance water and nutrient uptake from *Ae. tauschii* (DD genome).[40] The D homoeolog of *LATERAL ORGAN BOUNDARIES-DOMAIN16/ASYMMETRIC LEAVES2-LIKE18* gene (*TaLBD16-D*) is highly expressed, and the overexpression of this gene in transgenic wheat showed increased lateral root number.[40] In addition, salt stress tolerance was attributed to the expression of the D homoeolog of an ion transporter gene *HKT1;5-D* and *Salt Overly Sensitive1-D* (*TaSOS1-D*) in stressed roots.[39,41] We speculate that root tissues are critical for environmental tolerance and that many root-expressed genes may be under selective constraint. Further research on the details of these D homoeologs that show particularly high expression in roots, and whether this biased expression contributes to environmental responses, is required to test this hypothesis.

We found low correlations of Tajima's *D,* a neutrality test to detect positive and balancing selection between the A, B, and D homoeologs in all three datasets (0.04–0.09). The correlation was similarly low in the wild polyploid species *A*. *kamchatica* (0.04).[30,77] In addition, the chromosomal regions under selection were different among the two subgenomes of *C*. *bursa-pastoris*.[32] These results support similarity in the pattern of positive and balancing selection among natural and crop polyploids. We suggest that allopolyploid species can encompass homoeologs that experienced different evolutionary forces, which can contribute to the environmental robustness to cope with a broad range of conditions.[25,78,79]

A major goal of future breeding is to mitigate the effect of numerous deleterious variants, possibly using hybrid breeding and genome editing in addition to traditional crossing.[7–9] The different patterns of purifying selection among subgenomes and among genes expressed in different tissues can inform which subgenome or genes should be prioritized. For example, deleterious variants of the D subgenome expressed in roots may be restored by genome editing, or their effects may be mitigated by hybrid breeding. In hybrid breeding of the D subgenome, it would be effective to use landraces or the wild progenitor *Ae. tauschii* through synthetic polyploids because deleterious alleles may be fixed due to the low variation of the D subgenome. Moreover, studies of loci under selective sweeps[42,44] can be informative to restore adaptive variants lacking in landraces and wild species. To further understand and tackle deleterious variants under purifying selection in wheat, it will be important to combine the presented bread wheat data with recently increasing data of diploid and tetraploid progenitor species (e.g., *Ae. tauschii* as the donor of the D genome[80]) as well as landraces.

Limitations of the study

Our analysis of genome-wide selection has caveats and limitations. Firstly, we found that chromosomes harboring segregating foreign introgressions, as identified by Keilwagen et al.[48] had a major effect on the estimation of the proportion of adaptive substitutions, but had little effect on the DFE estimates. We indeed found consistently stronger purifying selection in the B subgenome, with or without the chromosome trios encompassing the introgressions. This result is consistent with theoretical studies showing that the DFE may be affected less by past demography than the estimation of the proportion of substitutions (α),[13,14] although it does not entirely exclude the effect of demography. Furthermore, because homoeologs of the same subgenome expressed in different tissues are expected to be affected similarly by past demography, we suggest that the relative difference in the strength of purifying selection among homoeologs expressed in different tissues is less affected by demography. Secondly, in this study, we quantified purifying selection using nonsynonymous mutations that are abundant, but it is likely that purifying selection on gene deletion or high-impact mutations is stronger, as shown in studies of these types of mutations in other species.[28,30,43]

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Kentaro K. Shimizu (kentaro.shimizu@uzh.ch).

Materials availability

The alignment files with outgroups were submitted in the Zenodo repository (doi: https://doi.org/10.5281/zenodo.13370482).

Data and code availability

- The genome sequence and gene annotations of all wheat cultivars can be viewed in and downloaded from Ensembl Plants (https://plants.ensembl.org/index.html).
- All raw data used in this study are available at the European Nucleotide Archive under accession no. PRJEB51827 (https://www.ebi.ac.uk/ena/browser/view/PRJEB51827).
- Codes are available in the GitHub repository https://github.com/brisk022/wheat_evo_analysis, and a version of record is saved at https://doi.org/10.5281/zenodo.17657312 (DOI: https://doi.org/10.5281/zenodo.17657313).

Consortia

The members of the 10+ Wheat Genome Project are Sean Walkowiak, Valentyna Klymiuk, Brook Byrns, Kirby Nilsen, Jennifer Ens, Krystalee Wiebe, Amidou N' Diaye, Pierre J. Hucl, Curtis Pozniak, Bin Xiao Fu, Liangliang Gao, Emily

Delorean, Dal-Hoe Koo, Allen K. Fritz, Jesse A. Poland, Cecile Monat, Axel Himmelbach, Anne Fiebig, Sudharsan Padmarasu, Uwe Scholz, Martin Mascher, Nils Stein, Georg Haberer, Heidrun Gundlach, Klaus F. X. Mayer, Manuel Spannagl, Mulualem T. Kassa, Pierre Fobert, Sateesh Kagale, Jemima Brinton, Ricardo H. Ramirez-Gonzalez, Michael Bevan, Neil McKenzie, Burkhard Steuernagel, Cristobal Uauy, Markus C. Kolodziej, Simon G. Krattinger, Beat Keller, Thomas Wicker, Dinushika Thambugala, Curt A. McCartney, Venkat Bandi, Jorge Nunez Siri, Carl Gutwin, Catharine Aquino, Masaomi Hatakeyama, Dario Copetti, Gwyneth Halstead-Nussloch, Timothy Paape, Rie Shimizu-Inatsugi, Kentaro K. Shimizu, Tomohiro Ban, Kanako Kawaura, Toshiaki Tameshige, Hiroyuki Tsuji, Luca Venturini, Matthew Clark, BernardoClavijo, Christine Fosker, Gonzalo Garcia Accinelli, Darren Heavens, Ksenia Krasileva, David Swarbreck, Jonathan Wright, Anthony Hall, Keith A. Gardner, Nick Fradgley, Lawrence Percival-Alwyn, James Cockram, Juan Gutierrez-Gonzalez, Gary Muehlbauer, Chu Shin Koh, Andrew G. Sharpe, Jasline Deek, Alejandro C. Costamagna, Hiroyuki Kanamori, Fuminori Kobayashi, Tsuyoshi Tanaka, Jianzhong Wu, Hirokazu Handa, Tony Kuo, Jun Sese, Kazuki Murata, Yusuke Nabeka, Shuhei Nasuda, Philomin Juliana, Ravi Singh, Hikmet Budak, Ian Small, Joanna Melonek, Sylvie Cloutier, Gabriel Keeble-Gagnère, Josquin Tibbets, Erik Legg, Arvind Bharti, Peter Langridge, KenChalmers, and Assaf Distelfeld.

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation grants 310030_212551, 31003A_182318, and CRSII5_18357 (K.K.S.); JSPS KAKENHI grant numbers 22H02316, 23K23582, and 22H05179 (K.K.S.); 22K15161 and 25K18217 (M.O.), Core Research for Evolutionary Science and Technology (grant number JPMJCR16O3) (K.K.S.); FOREST Program (grant number JPMJFR243H, Japan) (M.O.) of Japan Science and Technology Agency; University Research Priority Program in Evolution in Action of the University of Zurich (K.K.S.); and The CTAG2 project cofunded by Genome Canada, Genome Prairie, Saskatchewan Ministry of Agriculture, Western Grains Research Foundation, Saskatchewan Wheat Development Commission, Alberta Grains, and the Manitoba Crop Alliance (C.P.).

AUTHOR CONTRIBUTIONS

Conceptualization, G.H.N., T.P., and K.K.S.; data generation and formatting, G.H., T.L., B.W., A.H., C.P., 10+ Wheat Genome Project, and M.S.; data analysis, G.H.N., A.S.M.F.I., M.O., M.H., and R.B.; supervision, T.P. and K.K.S.; writing – original draft, G.H.N., T.P., and K.K.S.; and writing – review & editing, M.O., T.P., and K.K.S.

DECLARATION OF INTERESTS

All authors declare they have no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- METHOD DETAILS
  - Orthologous group identification and codon-based alignment
  - Population genetic analysis and divergence estimates
  - Genealogical analysis
  - Distribution of fitness effects and proportion of adaptive substitutions
  - Gene expression data for DFE analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2025.116785.

Received: June 25, 2025
Revised: October 27, 2025
Accepted: December 1, 2025
Published: January 26, 2026

REFERENCES

1. Hough, J., Williamson, R.J., and Wright, S.I. (2013). Patterns of selection in plant genomes. Annu. Rev. Ecol. Evol. Syst. *44*, 31–49. https://doi.org/10.1146/ANNUREV-ECOLSYS-110512-135851/1.
2. Wright, S.I., and Gaut, B.S. (2005). Molecular Population Genetics and the Search for Adaptive Evolution in Plants. Mol. Biol. Evol. *22*, 506–519. https://doi.org/10.1093/MOLBEV/MSI035.
3. Turner-Hissong, S.D., Mabry, M.E., Beissinger, T.M., Ross-Ibarra, J., and Pires, J.C. (2020). Evolutionary insights into plant breeding. Curr. Opin. Plant Biol. *54*, 93–100. https://doi.org/10.1016/J.PBI.2020.03.003.
4. Charlesworth, D., and Willis, J.H. (2009). The genetics of inbreeding depression. Nat. Rev. Genet. *10*, 783–796. https://doi.org/10.1038/nrg2664.
5. Birchler, J.A. (2013). Genetic Rules of Heterosis in Plants (Polyploid and Hybrid Genomics), pp. 313–321. https://doi.org/10.1002/9781118552872.CH19.
6. Birchler, J.A. (2015). Heterosis: The genetic basis of hybrid vigour. Nat. Plants *1*, 15020–15022. https://doi.org/10.1038/nplants.2015.20.
7. Wallace, J.G., Rodgers-Melnick, E., and Buckler, E.S. (2018). On the Road to Breeding 4.0: Unraveling the Good, the Bad, and the Boring of Crop Quantitative Genomics. Annu. Rev. Genet. *52*, 421–444. https://doi.org/10.1146/ANNUREV-GENET-120116-024846.
8. Varshney, R.K., Bohra, A., Yu, J., Graner, A., Zhang, Q., and Sorrells, M.E. (2021). Designing Future Crops: Genomics-Assisted Breeding Comes of Age. Trends Plant Sci. *26*, 631–649. https://doi.org/10.1016/j.tplants.2021.03.010.
9. Sun, S., Wang, B., Li, C., Xu, G., Yang, J., Hufford, M.B., Ross-Ibarra, J., Wang, H., and Wang, L. (2023). Unraveling Prevalence and Effects of Deleterious Mutations in Maize Elite Lines across Decades of Modern Breeding. Mol. Biol. Evol. *40*, msad170. https://doi.org/10.1093/MOLBEV/MSAD170.
10. Chen, J., Bataillon, T., Glémin, S., and Lascoux, M. (2022). What does the distribution of fitness effects of new mutations reflect? Insights from plants. New Phytol. *233*, 1613–1619. https://doi.org/10.1111/NPH.17826.
11. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly (Austin) *6*, 80–92. https://doi.org/10.4161/FLY.19695.
12. Simons, J.M., Herbert, T.C., Kauffman, C., Batete, M.Y., Simpson, A.T., Katsuki, Y., Le, D., Amundson, D., Buescher, E.M., Weil, C., et al. (2022). Systematic prediction of EMS-induced mutations in a sorghum mutant population. Plant Direct *6*, e404. https://doi.org/10.1002/PLD3.404.
13. Eyre-Walker, A., and Keightley, P.D. (2007). The distribution of fitness effects of new mutations. Nat. Rev. Genet. *8*, 610–618. https://doi.org/10.1038/nrg2146.
14. Eyre-Walker, A., and Keightley, P.D. (2009). Estimating the Rate of Adaptive Molecular Evolution in the Presence of Slightly Deleterious Mutations and Population Size Change. Mol. Biol. Evol. *26*, 2097–2108. https://doi.org/10.1093/MOLBEV/MSP119.
15. Chen, J., Glémin, S., and Lascoux, M. (2017). Genetic Diversity and the Efficacy of Purifying Selection across Plant and Animal Species. Mol. Biol. Evol. *34*, 1417–1428. https://doi.org/10.1093/MOLBEV/MSX088.
16. Ramu, P., Esuma, W., Kawuki, R., Rabbi, I.Y., Egesi, C., Bredeson, J.V., Bart, R.S., Verma, J., Buckler, E.S., and Lu, F. (2017). Cassava haplotype map highlights fixation of deleterious mutations during clonal propagation. Nat. Genet. *49*, 959–963. https://doi.org/10.1038/ng.3845.

17. Yang, L., and Gaut, B.S. (2011). Factors that contribute to variation in evolutionary rate among Arabidopsis genes. Mol. Biol. Evol. *28*, 2359–2369. https://doi.org/10.1093/MOLBEV/MSR058.

18. Gaut, B., Yang, L., Takuno, S., and Eguiarte, L.E. (2011). Variation in the rates of molecular evolution in plants. Annu. Rev. Ecol. Evol. Syst. *42*, 245–266. https://doi.org/10.1146/ANNUREV-ECOLSYS-102710-145119/CITE/REFWORKS.

19. Paape, T., Bataillon, T., Zhou, P., J Y Kono, T., Briskine, R., Young, N.D., and Tiffin, P. (2013). Selection, genome-wide fitness effects and evolutionary rates in the model legume Medicago truncatula. Mol. Ecol. *22*, 3525–3538. https://doi.org/10.1111/MEC.12329.

20. Monnahan, P., Kolář, F., Baduel, P., Sailer, C., Koch, J., Horvath, R., Laenen, B., Schmickl, R., Paajanen, P., Šrámková, G., et al. (2019). Pervasive population genomic consequences of genome duplication in Arabidopsis arenosa. Nat. Ecol. Evol. *3*, 457–468. https://doi.org/10.1038/s41559-019-0807-4.

21. Zhang, J., and Yang, J.R. (2015). Determinants of the rate of protein sequence evolution. Nat. Rev. Genet. *16*, 409–420. https://doi.org/10.1038/nrg3950.

22. Van De Peer, Y., Mizrachi, E., and Marchal, K. (2017). The evolutionary significance of polyploidy. Nat. Rev. Genet. *18*, 411–424. https://doi.org/10.1038/nrg.2017.26.

23. Akagi, T., Jung, K., Masuda, K., and Shimizu, K.K. (2022). Polyploidy before and after domestication of crop species. Curr. Opin. Plant Biol. *69*, 102255. https://doi.org/10.1016/J.PBI.2022.102255.

24. Clo, J., and Kolář, F. (2022). Inbreeding depression in polyploid species: a meta-analysis. Biol. Lett. *18*, 20220477. https://doi.org/10.1098/RSBL.2022.0477.

25. Shimizu, K.K. (2022). Robustness and the generalist niche of polyploid species: Genome shock or gradual evolution? Curr. Opin. Plant Biol. *69*, 102292. https://doi.org/10.1016/J.PBI.2022.102292.

26. Pophaly, S.D., and Tellier, A. (2015). Population Level Purifying Selection and Gene Expression Shape Subgenome Evolution in Maize. Mol. Biol. Evol. *32*, 3226–3235. https://doi.org/10.1093/MOLBEV/MSV191.

27. Conover, J.L., and Wendel, J.F. (2022). Deleterious Mutations Accumulate Faster in Allopolyploid Than Diploid Cotton (Gossypium) and Unequally between Subgenomes. Mol. Biol. Evol. *39*, msac024. https://doi.org/10.1093/molbev/msac024.

28. Douglas, G.M., Gos, G., Steige, K.A., Salcedo, A., Holm, K., Josephs, E.B., Arunkumar, R., Ågren, J.A., Hazzouri, K.M., Wang, W., et al. (2015). Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid Capsella bursa-pastoris. Proc. Natl. Acad. Sci. USA *112*, 2806–2811. https://doi.org/10.1073/PNAS.1412277112.

29. Avni, R., Nave, M., Barad, O., Baruch, K., Twardziok, S.O., Gundlach, H., Hale, I., Mascher, M., Spannagl, M., Wiebe, K., et al. (2017). Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. Science *357*, 93–97. https://doi.org/10.1126/SCIENCE.AAN0032.

30. Paape, T., Briskine, R.V., Halstead-Nussloch, G., Lischer, H.E.L., Shimizu-Inatsugi, R., Hatakeyama, M., Tanaka, K., Nishiyama, T., Sabirov, R., Sese, J., and Shimizu, K.K. (2018). Patterns of polymorphism and selection in the subgenomes of the allopolyploid Arabidopsis kamchatica. Nat. Commun. *9*, 3909–3913. https://doi.org/10.1038/s41467-018-06108-1.

31. Shimizu, K.K., Copetti, D., Okada, M., Wicker, T., Tameshige, T., Hatakeyama, M., Shimizu-Inatsugi, R., Aquino, C., Nishimura, K., Kobayashi, F., et al. (2021). De Novo Genome Assembly of the Japanese Wheat Cultivar Norin 61 Highlights Functional Variation in Flowering Time and Fusarium-Resistant Genes in East Asian Genotypes. Plant Cell Physiol. *62*, 8–27. https://doi.org/10.1093/PCP/PCAA152.

32. Kryvokhyzha, D., Salcedo, A., Eriksson, M.C., Duan, T., Tawari, N., Chen, J., Guerrina, M., Kreiner, J.M., Kent, T.V., Lagercrantz, U., et al. (2019). Parental legacy, demography, and admixture influenced the evolution of the two subgenomes of the tetraploid Capsella bursa-pastoris (Brassicaceae). PLoS Genet. *15*, e1007949. https://doi.org/10.1371/JOURNAL.PGEN.1007949.

33. Dubcovsky, J., and Dvorak, J. (2007). Genome plasticity a key factor in the success of polyploid wheat under domestication. Science *316*, 1862–1866. https://doi.org/10.1126/SCIENCE.1143986.

34. Matsuoka, Y. (2011). Evolution of Polyploid Triticum Wheats under Cultivation: The Role of Domestication, Natural Hybridization and Allopolyploid Speciation in their Diversification. Plant Cell Physiol. *52*, 750–764. https://doi.org/10.1093/PCP/PCR018.

35. Okada, M., and Shimizu, K.K. (2024). Wheat: Recent Advances in Genomics to Explore the History of Domestication and Polyploidization. 155–180. https://doi.org/10.1007/978-981-97-6887-5_10.

36. Cavalet-Giorsa, E., González-Muñoz, A., Athiyannan, N., Holden, S., Salhi, A., Gardener, C., Quiroz-Chávez, J., Rustamova, S.M., Elkot, A.F., Patpour, M., et al. (2024). Origin and evolution of the bread wheat D genome. Nature *633*, 848–855. https://doi.org/10.1038/s41586-024-07808-z.

37. Wang, Z., Wang, W., He, Y., Xie, X., Yang, Z., Zhang, X., Niu, J., Peng, H., Yao, Y., Xie, C., et al. (2024). On the evolution and genetic diversity of the bread wheat D genome. Mol. Plant *17*, 1672–1686. https://doi.org/10.1016/J.MOLP.2024.09.007.

38. Asuke, S., Umehara, Y., Inoue, Y., Vy, T.T.P., Iwakawa, M., Matsuoka, Y., Kato, K., and Tosa, Y. (2021). Origin and Dynamics of Rwt6, a Wheat Gene for Resistance to Nonadapted Pathotypes of Pyricularia oryzae. Phytopathology *111*, 2023–2029. https://doi.org/10.1094/PHYTO-02-21-0080-R.

39. Yang, C., Zhao, L., Zhang, H., Yang, Z., Wang, H., Wen, S., Zhang, C., Rustgi, S., Von Wettstein, D., and Liu, B. (2014). Evolution of physiological responses to salt stress in hexaploid wheat. Proc. Natl. Acad. Sci. USA *111*, 11882–11887. https://doi.org/10.1073/PNAS.1412839111.

40. Wang, H., Hu, Z., Huang, K., Han, Y., Zhao, A., Han, H., Song, L., Fan, C., Li, R., Xin, M., et al. (2018). Three genomes differentially contribute to the seedling lateral root number in allohexaploid wheat: evidence from phenotype evolution and gene expression. Plant J. *95*, 976–987. https://doi.org/10.1111/TPJ.14005.

41. Zheng, M., Li, J., Zeng, C., Liu, X., Chu, W., Lin, J., Wang, F., Wang, W., Guo, W., Xin, M., et al. (2022). Subgenome-biased expression and functional diversification of a Na+/H+ antiporter homoeologs in salt tolerance of polyploid wheat. Front. Plant Sci. *13*, 1072009. https://doi.org/10.3389/FPLS.2022.1072009.

42. Jordan, K.W., Wang, S., Lun, Y., Gardiner, L.J., MacLachlan, R., Hucl, P., Wiebe, K., Wong, D., Forrest, K.L., and IWGS Consortium; and et al.. (2015). A haplotype map of allohexaploid wheat reveals distinct patterns of selection on homoeologous genomes. Genome Biol. *16*, 48. https://doi.org/10.1186/S13059-015-0606-4.

43. He, F., Pasam, R., Shi, F., Kant, S., Keeble-Gagnere, G., Kay, P., Forrest, K., Fritz, A., Hucl, P., Wiebe, K., 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. https://doi.org/10.1038/s41588-019-0382-2.

44. Cavanagh, C.R., Chao, S., Wang, S., Huang, B.E., Stephen, S., Kiani, S., Forrest, K., Saintenac, C., Brown-Guedira, G.L., Akhunova, A., et al. (2013). Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc. Natl. Acad. Sci. USA *110*, 8057–8062. https://doi.org/10.1073/PNAS.1217133110.

45. Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M.T., Brinton, J., Ramirez-Gonzalez, R.H., Kolodziej, M.C., Delorean, E., Thambugala, D., et al. (2020). Multiple wheat genomes reveal global variation in modern breeding. Nature *588*, 277–283. https://doi.org/10.1038/s41586-020-2961-x.

46. International Wheat Genome Sequencing Consortium IWGSC (2018). Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science *361*, eaar7191. https://doi.org/10.1126/SCIENCE.AAR7191.

47. White, B., Lux, T., Rusholme-Pilcher, R., Juhász, A., Kaithakottil, G., Duncan, S., Simmonds, J., Rees, H., Wright, J., Colmer, J., et al. (2025). De novo annotation reveals transcriptomic complexity across the hexaploid wheat pan-genome. Nat. Commun. *16*, 8538. https://doi.org/10.1038/S41467-025-64046-1.

48. Keilwagen, J., Lehnert, H., Berner, T., Badaeva, E., Himmelbach, A., Börner, A., and Kilian, B. (2022). Detecting major introgressions in wheat and their putative origins using coverage analysis. Sci. Rep. *12*, 1908–1911. https://doi.org/10.1038/s41598-022-05865-w.

49. Kelso, J., and Prüfer, K. (2014). Ancient humans and the origin of modern humans. Curr. Opin. Genet. Dev. *29*, 133–138. https://doi.org/10.1016/J.GDE.2014.09.004.

50. Rabanus-Wallace, M.T., Hackauf, B., Mascher, M., Lux, T., Wicker, T., Gundlach, H., Baez, M., Houben, A., Mayer, K.F.X., Guo, L., et al. (2021). Chromosome-scale genome assembly provides insights into rye biology, evolution and agronomic potential. Nat. Genet. *53*, 564–573. https://doi.org/10.1038/s41588-021-00807-0.

51. Mascher, M., Wicker, T., Jenkins, J., Plott, C., Lux, T., Koh, C.S., Ens, J., Gundlach, H., Boston, L.B., Tulpová, Z., et al. (2021). Long-read sequence assembly: a technical evaluation in barley. Plant Cell *33*, 1888–1906. https://doi.org/10.1093/PLCELL/KOAB077.

52. Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. Psychol. Bull. *114*, 494–509. https://doi.org/10.1037/0033-2909.114.3.494.

53. Cliff, N. (2014). Ordinal Methods for Behavioral Data Analysis (Psychology Press). https://doi.org/10.4324/9781315806730.

54. Vargha, A., and Delaney, H.D. (2000). A Critique and Improvement of the CL Common Language Effect Size Statistics of McGraw and Wong. J. Educ. Behav. Stat. *25*, 101–132. https://doi.org/10.3102/10769986025002101.

55. Cohen, J. (2013). Statistical Power Analysis for the Behavioral Sciences. Statistical Power Analysis for the Behavioral Sciences. https://doi.org/10.4324/9780203771587.

56. Marcussen, T., Sandve, S.R., Heier, L., Spannagl, M., Pfeifer, M., International Wheat Genome Sequencing Consortium; Jakobsen, K.S., Wulff, B.B.H., Steuernagel, B., Mayer, K.F.X., and Olsen, O.A. (2014). Ancient hybridizations among the ancestral genomes of bread wheat. Science *345*, 1250092. https://doi.org/10.1126/SCIENCE.1250092.

57. Glémin, S., Scornavacca, C., Dainat, J., Burgarella, C., Viader, V., Ardisson, M., Sarah, G., Santoni, S., David, J., and Ranwez, V. (2019). Pervasive hybridizations in the history of wheat relatives. Sci. Adv. *5*. https://doi.org/10.1126/SCIADV.AAV9188.

58. Eyre-Walker, A. (2006). The genomic rate of adaptive evolution. Trends Ecol. Evol. *21*, 569–575. https://doi.org/10.1016/J.TREE.2006.06.015.

59. FAOstat (2022). Food and Agriculture Organization of the United Nations. http://faostat.fao.org. http://faostat.fao.org.

60. Zhou, Y., Zhao, X., Li, Y., Xu, J., Bi, A., Kang, L., Xu, D., Chen, H., Wang, Y., Wang, Y.G., et al. (2020). Triticum population sequencing provides insights into wheat adaptation. Nat. Genet. *52*, 1412–1422. https://doi.org/10.1038/s41588-020-00722-w.

61. Zhao, X., Guo, Y., Kang, L., Yin, C., Bi, A., Xu, D., Zhang, Z., Zhang, J., Yang, X., Xu, J., et al. (2023). Population genomics unravels the Holocene history of bread wheat and its relatives. Nat. Plants *9*, 403–419. https://doi.org/10.1038/s41477-023-01367-3.

62. Lv, Q., Li, W., Sun, Z., Ouyang, N., Jing, X., He, Q., Wu, J., Zheng, J., Zheng, J., Tang, S., et al. (2020). Resequencing of 1,143 indica rice accessions reveals important genetic variations and different heterosis patterns. Nat. Commun. *11*, 4778. https://doi.org/10.1038/s41467-020-18608-0.

63. Tsujimoto, H. (2021). Gene-Mining Asian Wheat to Feed the Population in the 21st Century. Plant Cell Physiol. *62*, 1–2. https://doi.org/10.1093/PCP/PCAA158.

64. Akama, S., Shimizu-Inatsugi, R., Shimizu, K.K., and Sese, J. (2014). Genome-wide quantification of homeolog expression ratio revealed nonstochastic gene regulation in synthetic allopolyploid Arabidopsis. Nucleic Acids Res. *42*, e46. https://doi.org/10.1093/NAR/GKT1376.

65. Kuo, T.C.Y., Hatakeyama, M., Tameshige, T., Shimizu, K.K., and Sese, J. (2020). Homeolog expression quantification methods for allopolyploids. Brief. Bioinform. *21*, 395–407. https://doi.org/10.1093/BIB/BBY121.

66. Hu, G., Grover, C.E., Arick, M.A., Liu, M., Peterson, D.G., and Wendel, J.F. (2021). Homoeologous gene expression and co-expression network analyses and evolutionary inference in allopolyploids. Brief. Bioinform. *22*, 1819–1835. https://doi.org/10.1093/BIB/BBAA035.

67. Slotte, T., Foxe, J.P., Hazzouri, K.M., and Wright, S.I. (2010). Genome-Wide Evidence for Efficient Positive and Purifying Selection in Capsella grandiflora, a Plant Species with a Large Effective Population Size. Mol. Biol. Evol. *27*, 1813–1821. https://doi.org/10.1093/MOLBEV/MSQ062.

68. Shimizu, K.K., and Tsuchimatsu, T. (2015). Evolution of Selfing: Recurrent Patterns in Molecular Adaptation. Annu. Rev. Ecol. Evol. Syst. *46*, 593–622. https://doi.org/10.1146/ANNUREV-ECOLSYS-112414-054249/1.

69. Yew, C.L., Tsuchimatsu, T., Shimizu-Inatsugi, R., Yasuda, S., Hatakeyama, M., Kakui, H., Ohta, T., Suwabe, K., Watanabe, M., Takayama, S., and Shimizu, K.K. (2023). Dominance in self-compatibility between subgenomes of allopolyploid Arabidopsis kamchatica shown by transgenic restoration of self-incompatibility. Nat. Commun. *14*, 7618. https://doi.org/10.1038/S41467-023-43275-2.

70. Ledyard, S.G. (1950). Variation and Evolution in Plants (Columbia University Press).

71. Robertson, K., Goldberg, E.E., and Igić, B. (2011). COMPARATIVE EVIDENCE FOR THE CORRELATED EVOLUTION OF POLYPLOIDY AND SELF-COMPATIBILITY IN SOLANACEAE. Evolution *65*, 139–155. https://doi.org/10.1111/J.1558-5646.2010.01099.X.

72. Gupta, P.K., Balyan, H.S., Gahlaut, V., Saripalli, G., Pal, B., Basnet, B.R., and Joshi, A.K. (2019). Hybrid wheat: past, present and future. Theor. Appl. Genet. *132*, 2463–2483. https://doi.org/10.1007/S00122-019-03397-Y.

73. Zhao, Y., Thorwarth, P., Jiang, Y., Philipp, N., Schulthess, A.W., Gils, M., Boeven, P.H.G., Longin, C.F.H., Schacht, J., Ebmeyer, E., et al. (2021). Unlocking big data doubled the accuracy in predicting the grain yield in hybrid wheat. Sci. Adv. *7*, 9106–9117. https://doi.org/10.1126/SCIADV.ABF9106.

74. ter Steeg, E.M.S., Struik, P.C., Visser, R.G.F., and Lindhout, P. (2022). Crucial factors for the feasibility of commercial hybrid breeding in food crops. Nat. Plants *8*, 463–473. https://doi.org/10.1038/s41477-022-01142-w.

75. Ramírez-González, R.H., Borrill, P., Lang, D., Harrington, S.A., Brinton, J., Venturini, L., Davey, M., Jacobs, J., Van Ex, F., Pasha, A., et al. (2018). The transcriptional landscape of polyploid wheat. Science *361*. https://doi.org/10.1126/SCIENCE.AAR6089.

76. Sun, J., Okada, M., Tameshige, T., Shimizu-Inatsugi, R., Akiyama, R., Nagano, A.J., Sese, J., and Shimizu, K.K. (2023). A low-coverage 3′ RNA-seq to detect homeolog expression in polyploid wheat. NAR Genom. Bioinform. *5*, lqad067. https://doi.org/10.1093/NARGAB/LQAD067.

77. Paape, T., Hatakeyama, M., Shimizu-Inatsugi, R., Cereghetti, T., Onda, Y., Kenta, T., Sese, J., and Shimizu, K.K. (2016). Conserved but Attenuated Parental Gene Expression in Allopolyploids: Constitutive Zinc Hyperaccumulation in the Allotetraploid Arabidopsis kamchatica. Mol. Biol. Evol. *33*, 2781–2800. https://doi.org/10.1093/MOLBEV/MSW141.

78. Ledyard, S.G. (1971). Chromosomal Evolution in Higher Plants (Edward Arnold ELBS).

79. Akiyama, R., Goto, T., Tameshige, T., Sugisaka, J., Kuroki, K., Sun, J., Akita, J., Hatakeyama, M., Kudoh, H., Kenta, T., et al. (2023). Seasonal pigment fluctuation in diploid and polyploid Arabidopsis revealed by machine learning-based phenotyping method PlantServation. Nat. Commun. *14*, 5792. https://doi.org/10.1038/S41467-023-41260-3.

80. Gaurav, K., Arora, S., Silva, P., Sánchez-Martín, J., Horsnell, R., Gao, L., Brar, G.S., Widrig, V., John Raupp, W., Singh, N., et al. (2021). Population genomic analysis of Aegilops tauschii identifies targets for bread wheat improvement. Nat. Biotechnol. *40*, 422–431. https://doi.org/10.1038/s41587-021-01058-4.

81. Abrouk, M., Athiyannan, N., Müller, T., Pailles, Y., Stritt, C., Roulin, A.C., Chu, C., Liu, S., Morita, T., Handa, H., et al. (2021). Population genomics and haplotype analysis in spelt and bread wheat identifies a gene regulating glume color. Commun. Biol. *4*, 1–11. https://doi.org/10.1038/s42003-021-01908-6.
82. Halstead-Nussloch, G., Tanaka, T., Copetti, D., Paape, T., Kobayashi, F., Hatakeyama, M., Kanamori, H., Wu, J., Mascher, M., Kawaura, K., et al. (2021). Multiple Wheat Genomes Reveal Novel Gli-2 Sublocus Location and Variation of Celiac Disease Epitopes in Duplicated α-Gliadin Genes. Front. Plant Sci. *12*, 715985. https://doi.org/10.3389/FPLS.2021.715985.
83. Emms, D.M., and Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/S13059-019-1832-Y.
84. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797. https://doi.org/10.1093/NAR/GKH340.
85. Thornton, K. (2003). libsequence: a C++ class library for evolutionary genetic analysis. Bioinformatics *19*, 2325–2327. https://doi.org/10.1093/BIOINFORMATICS/BTG316.
86. Gossmann, T.I., Song, B.H., Windsor, A.J., Mitchell-Olds, T., Dixon, C.J., Kapralov, M.V., Filatov, D.A., and Eyre-Walker, A. (2010). Genome Wide Analyses Reveal Little Evidence for Adaptive Evolution in Many Plant Species. Mol. Biol. Evol. *27*, 1822–1832. https://doi.org/10.1093/MOLBEV/MSQ079.
87. Romano, J., Kromrey, J.D., Coraggio, J., and Skowronek, J. (2006). Appropriate statistics for ordinal level data: Should we really be using t-test and Cohen'sd for evaluating group differences on the NSSE and other surveys. In In Annual Meeting of the Florida Association of Institutional Research, p. 177.
88. Nakagawa, S., and Cuthill, I.C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol. Rev. *82*, 591–605. https://doi.org/10.1111/J.1469-185X.2007.00027.X.
89. Ferguson, C.J. (2009). An Effect Size Primer: A Guide for Clinicians and Researchers. Prof Psychol Res Pr *40*, 532–538. https://doi.org/10.1037/A0015808.
90. Meissel, K., and Yao, E.S. (2024). Using Cliff's Delta as a Non-Parametric Effect Size Measure: An Accessible Web App and R Tutorial. Practical Assess. Res. Eval. *29*, 1–12. https://doi.org/10.7275/PARE.1977.
91. Torchiano, M. Effsize - a package for efficient effect size computation. 2016 https://doi.org/10.5281/ZENODO.1480624.
92. Bürkner, P.C. (2021). Bayesian Item Response Modeling in R with brms and Stan. J. Stat. Softw. *100*, 1–54. https://doi.org/10.18637/JSS.V100.I05.
93. Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One *11*, e0163962. https://doi.org/10.1371/JOURNAL.PONE.0163962.
94. Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.Y. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. *8*, 28–36. https://doi.org/10.1111/2041-210X.12628.
95. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/S13059-014-0550-8.
96. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915. https://doi.org/10.1038/s41587-019-0201-4.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Coding sequence alignments	This paper	https://zenodo.org/records/13370482
Raw sequence data	White et al. 2025	https://www.ebi.ac.uk/ena/browser/view/PRJEB51827
de novo assemblies and annotations of 10 wheat cultivars	White et al. 2025	https://doi.org/10.1101/2024.01.09.574802 https://plants.ensembl.org/index.html
Rye genome data	Rabanus-Wallace et al. 2021	https://doi.org/10.1038/s41588-021-00807-0
Barley genome data	Mascher et al. 2021	https://doi.org/10.1093/PLCELL/KOAB077
Software and algorithms		
OrthoFinder	Emms and Kelly 2019	https://doi.org/10.1186/S13059-019-1832-Y
Muscle aligner	Edger 2004	https://doi.org/10.1093/NAR/GKH340
compute	Thornton 2003	https://doi.org/10.1093/BIOINFORMATICS/BTG316
seqkit	Shen et al. 2016	https://doi.org/10.1371/JOURNAL.PONE.0163962
DoFE	Eyre-Walker and Keightley 2009	https://doi.org/10.1093/MOLBEV/MSP119
HISAT2	Kim et al. 2019	https://doi.org/10.1038/s41587-019-0201-4
DeSEQ2	Love et al. 2014	https://doi.org/10.1186/S13059-014-0550-8

METHOD DETAILS

Orthologous group identification and codon-based alignment

Identification of orthologous groups (OGs) was conducted similar to the approach,[45] in which OGs contain one-to-one orthologs from the 10 wheat lines ArinaLrFor, CDC Stanley, CDC Landmark, Chinese Spring, Jagger, Julius, LongReach Lancer, Norin 61, Mace and SY Mattis plus an outgroup of either rye, *S. cereale*,[50] or barley, *H. vulgare*.[51] Because we were interested in the genome-wide selection in bread wheat, we did not include the assembly of Spelt (*T. aestivum* ssp. *spelta* accession PI190962,[45] which was likely derived from the hybridization of wild and domesticated wheat.[81,82] For constructing accurate and conservative datasets, we focused on OGs with exactly one gene copy for the A-, B-, and D-subgenomes in each wheat line (called 30-lets by White et al. (2025)[47]) and exactly one gene from the outgroup. In this dataset, OGs with copy-number variation (i.e., increase or decrease) were excluded to avoid their potential effect on the estimation of the amount of variation and the strength of selection. First, OGs with copy number increase were excluded because it is known to affect the pattern of variation through gene conversion particularly in tandem duplicates and can affect purifying selection due to redundancy. Tandemly duplicated gene copies can result in one or more of the duplicated copies experiencing relaxed selective constraint where mutations increase compared to the original parent copy. Additionally, sub-functionalization may result in duplicated genes retaining some of the parental gene function, but may experience increased mutations due to drift. These duplicated genes can bias the OGs because one subgenome may have tandemly duplicated copies of a gene, while the others do not. In a sense, this is the problem we are already addressing but at the whole genome level – comparison of relaxed constraint and purifying selection of duplicated gene copies. Second, OGs with a non-functional allele (or copy number decrease) were excluded because they are considered non-essential genes and thus can have a different pattern of purifying selection. White et al. (2025)[47] reported that a considerable proportion of the OGs had a non-functional allele (36.61% of the OGs, called "shell" genes in contrast to "core" genes conserved in all the 10 cultivars by White et al. (2025)[47]) and thus removed from the present analysis. Orthologs were derived from an OrthoFinder[83] run for which each of the three subgenomes A, B, and D of each wheat line were provided as 'separate species' to mimic a diploid state for hexaploid wheat. Thus, orthologs and homoeologs could be identified in one run, with the diploid outgroup. These alignments had no missing accessions and negligible amounts of missing data represented by an N. Codon-based alignments were computed from back-translated protein alignments generated by the Muscle aligner.[84] All wheat genes in one group are assigned to pseudochromosomes, no chrUn genes were included. Alignments in total 10,602 and 10,604 OGs with rye and barley as an outgroup were used as Dataset 1 (Full Dataset). Datasets 2 and 3 were constructed by removing the trio of whole chromosome 2 (2A, 2B, 2D) and the trios of whole chromosome 2 and 3 (2A, 2B, 2D, 3A, 3B, 3D). respectively, because these chromosomes encompass known large segregating foreign introgressions. The 2A introgression is known to encompass the *Lr37-Yr17-Sr38* gene cluster from *Ae. ventricosa*, the 2B introgression *Sr36* from *T. timophevii*, 3D introgression *Lr24* and *Sr24* from *Thinopyrum ponticum*.[45] Entire chromosomes rather than the introgressed segments were removed because the removal of introgession region alone would affect the gradient of selection or genetic diversity from the telomere to the centromere.

Population genetic analysis and divergence estimates

The alignments of the coding regions of the ten cultivars, which were generated as described above, were used to calculate within-species average pairwise diversity (π), Watterson's theta (θ_W), and Tajima's *D* using *compute*[85] for each of the three subgenomes. Within cultivar, segregating sites (S) were used to filter out alignments which may be poor. Homoeolog alignments with more than 5% segregating sites were removed. Similar to previous population genetic analysis focusing on a single species, a constant mutation rate at the species level was assumed.[86] The alignments with an outgroup sequence were used to classify the synonymous (silent) and non-synonymous (replacement) polymorphic sites using *polydNdS*, the fixed synonymous and non-synonymous sites using *MKtest* and the minor and derived allele frequency spectra using SFS.pl.[85] Evolutionary rates were estimated as the proportion of nonsynonymous substitutions per nonsynonymous site (d_N) to synonymous substitutions per synonymous site (d_S) or $\omega = d_N/d_S$. The Wilcoxon rank-sum test with Benjamini–Hochberg correction, Cliff's delta[52–54,87–90] with 95% confidence intervals, and Bayesian effect size estimates (raw difference and Cohen's *d*[53]) were applied to compare differences in population genetic statistics (Tajima's *D*, θ_W, d_N/d_S, π, π_{NS}, and π_S) among subgenomes, outgroups, and datasets, respectively. The effsize package[91] was used for calculating Cliff's delta and the brms package[92] for the Bayesian analysis.

Genealogical analysis

For genealogies, either barley or rye was included as an outgroup. 100 randomly selected orthologous groups (OGs) were chosen and sequences within the A-, B-, or D-subgenomes were concatenated using seqkit.[93] Neighbor-joining trees were constructed using the Tamura-Nei model and bootstrap support was calculated using 100 replicates using Geneious prime. Trees are visualized using the R package 'ggtree'.[94]

Distribution of fitness effects and proportion of adaptive substitutions

To estimate the DFEs and the proportion of adaptive substitutions (α) of the A, B and D-subgenomes, we used the likelihood method implemented in the software DoFE 3.0 (http://www.sussex.ac.uk/lifesci/eyre-walkerlab/resources). Minor allele frequencies, number of polymorphic and fixed sites, and length of nonsynonymous and synonymous sites within each homoeolog were calculated from *polydNdS*, then summed for all genes for each subgenome individually, and used as input for DoFE v3.0 to estimate both the distribution of fitness effects (DFE) and proportion of adaptive substitutions (α) per subgenome for different sets of OGs.[14] We chose a method that takes into account slightly deleterious alleles segregating in the population as well as population size changes to calculate less biased estimates of α based on the shape parameters of the DFE.[14] The program was run for 1×10^6 Markov Chain Monte Carlo (MCMC) replicates, and sampled every 1000 replicates after a burn-in of 100,000 replicates. To estimate DFE, we used folded allele frequency spectra and included the number of non-synonymous (D_n) and synonymous (D_s) differences between rye or barley homoeologs as outgroup orthologs.

Gene expression data for DFE analysis

Gene expression levels in six samples (grain, leaf (flag leaf), root, spike, whole aerial organs sampled at dawn, and those at dusk) were quantified in White et al. (2025).[47] Reads were aligned to Chinese Spring RefSeq 1.1 using HISAT2 v2.0.4, and normalized counts per million mapped reads (CPM) were generated using DeSEQ2.[46,95,96] For each condition, the geometric mean of CPM values for each identified homoeolog within an orthologous group (OG; described above) was calculated across replicates. Using the distribution of the mean CPMs, a cutoff of 5% was used to identify genes with the highest and lowest expression values within a given tissue or set of tissues. To examine the effect of the difference of expressed tissues on purifying selection, four of them, i.e., grain, leaf, root, and spike, were used to represent spatially different tissues/organs. We conducted a linear model analysis using JMP to test for an interaction between subgenome and gene expression using either the neutral category using the equation: (DFE value N_es 0–1) $\sim$ Genome + Tissue + Genome x Tissue + Expression + Genome x Expression + Tissue x Expression + Genome x Tissue x Expression. We repeated this analysis but using the values in the DFE category for DFE value $N_es > 10$. The input DFE values that we used came from the DoFE software output where 1000 DFE values for either N_es 0–1 or $N_es > 10$ were the dependent variable, and covariates were Genome (A, B, and D), tissue (leaf, spike, grain, and root), and expression level (high or low). We use ANOVA to report the significance of each of the covariates and their interactions (Table S6). We also used Tukey's honestly significant difference (HSD) test to examine pairwise comparisons among subgenome, tissue, and expression level combinations.

Principal component analysis (PCA) and hierarchical clustering were conducted using the R software ver. 4.3.2. The CPM values of four tissues (grains, leaves, roots, and spikes) were normalized using DESeq2[95] before analysis. PCA was performed using the prcomp() function with the default settings. The top 1000 variable genes were used for hierarchical clustering using the pheatmap() function with the default settings.