

Supplementary Materials for

The transcriptional landscape of polyploid wheat

R. H. Ramírez-González*, P. Borrill*†, D. Lang, S. A. Harrington, J. Brinton, L. Venturini, M. Davey, J. Jacobs, F. van Ex, A. Pasha, Y. Khedikar, S. J. Robinson, A. T. Cory, T. Florio, L. Concia, C. Juery, H. Schoonbeek, B. Steuernagel, D. Xiang, C. J. Ridout, B. Chalhoub, K. F. X. Mayer, M. Benhamed, D. Latrasse, A. Bendahmane, International Wheat Genome Sequencing Consortium, B. B. H. Wulff, R. Appels, V. Tiwari, R. Datla, F. Choulet, C. J. Pozniak, N. J. Provart, A. G. Sharpe, E. Paux, M. Spannagl, A. Bräutigam, C. Uauy†

*These authors contributed equally to this work.

†Corresponding author. Email: cristobal.uauy@jic.ac.uk (C.U.); philippa.borrill@jic.ac.uk (P.B.)

Published 17 August 2018, *Science* **XXX**, eaar6089 (2018)

DOI: 10.1126/science.aar6089

This PDF file includes:

Additional Materials and Methods

Figs. S1 to S24

Tables S4, S7, S11 to S13, S15, S16, S22 to S25, S27, S32, S34, and S37

Captions for Tables S1 to S3, S5, S6, S8 to S10, S14, S17 to S21, S26, S28 to S31, S33, S35, and S36

IWGSC Collaborator List

Captions for Data S1 and S2

References

Other Supplementary Material for this manuscript includes the following:

(available at www.sciencemag.org/content/361/6403/eaar6089/suppl/DC1)

Tables S1 to S3, S5, S6, S8 to S10, S14, S17 to S21, S26, S28 to S31, S33, S35, and S36

Data S1 and S2

Additional Materials and Methods

RNA-Seq samples

The additional 321 RNA-Seq samples from six studies used for this analysis included a developmental time-course across 70 tissues timepoint combinations, a pathogen-associated molecular pattern (PAMP) triggered immune response study, and four tissue studies (Table S1). Methods for these new studies are outlined below.

- *Developmental time course*: Wheat plants from Bayer Crop Science Ukrainian spring wheat cultivar Azhurnaya (available at http://genbank.vurv.cz/ewdb/asp/ewdb_d2.asp?accn=168030) were grown in growth cabinets with 16:8 hours day:night length at 25:15 °C. Three biological replicates, consisting of five individual plants each, were sampled at the developmental times and tissues outlined in Table S1. All tissues were harvested between 7.5 and 8.5 h into the day and immediately frozen in liquid nitrogen upon collection. Seedling root samples were collected from wheat plants cultivated in agar. RNA was extracted from 100 mg fresh weight material using the Spectrum Plant Total RNA Kit (Sigma-Aldrich) and DNA contamination removed using the RNase-Free DNase Set (Qiagen Cat. 79254), and quality checks with the BioAnalyser. Sequencing libraries were prepared with 250-350 bp insert size and sequenced on the Illumina HiSeq 2500 using 2 x 125 bp paired-end strand-specific chemistry v4. Raw data is deposited as PRJEB25639.
- *PAMP-triggered immune response*: Chinese Spring wheat plants were grown for 3 weeks in a growth cabinet with 16:8 hours day:night length at 23:18 °C. Elicitation with PAMPs was modified from (70). For each of the three biological replicates, three 2 cm sections were cut from leaf 2 and 3, placed in a 2 mL tube with sterile water and vacuum-infiltrated three times for 1 min. The following day water was removed and replaced by fresh water or PAMPs dissolved in water at 1 g/L for chitin (Nacosy, YSK, Japan) or 500 nM flg22 (www.peptron.com). Samples were drained and flash frozen in liquid Nitrogen after 30 or 180 min prior to pulverisation with 2 stainless steel balls in a Genogrinder (SPEX). RNA was extracted using the RNeasy plant kit (Qiagen), the concentration determined on a nanodrop8000 (Thermo scientific) spectrophotometer and quality assessed with a RNA 6000 Nano chip on a Bioanalyzer2100 (Agilent). After removal of genomic DNA with DNase Turbo DNA-Free (Ambion), 1 µg of RNA was converted to cDNA with SuperscriptIV (ThermoFisher) for confirmation of induction of known PAMP-inducible genes (70). Sequencing was performed by Novogene (Beijing, China) using the Illumina HiSeq4000 platform 2 x 150 bp paired-end chemistry. Raw data is deposited in NCBI as BioProject PRJEB23056.
- *Chinese Spring spike*: Chinese Spring wheat plants were grown under controlled conditions with 16:8 hours day:night length at 22:20 °C. For each of the two biological replicates, spike and inflorescence tissues were collected at the growth stages specified in Table S1. RNA was extracted using the RNeasy Micro kit (Qiagen, Cat 1927) and mRNAs amplified using MessageAmp aRNA kit (Ambion, Cat 1750) according to the manufacturer's instructions. Illumina RNA-seq libraries were prepared using the aRNA and the TruSeq RNA kit (version 1, rev A). Paired-end reads were obtained using the Illumina HiSeq 2000 with four libraries pooled per lane. Raw data is deposited in NCBI as BioProject PRJNA436817.
- *Chinese Spring tissues*: Chinese Spring wheat plants were grown in growth pouches supplied with 50% Hoagland's solution in growth chambers with 12:12 hours day:night length at constant 20 °C for 14 days. On the 14th day, plants at the three-leaf stage (Zadok stage 13) were

selected for RNA extraction. Tissue samples were collected from leaf and root, frozen in liquid nitrogen, and stored at -80°C until processed. For spike tissue, plants were grown in the greenhouse in two litre pots and spike tissue was collected at 50% anthesis and frozen in liquid nitrogen. For RNA extraction and RNA-Seq library preparation, tissues were ground in liquid nitrogen and total RNA was extracted using the Qiagen RNeasy Plant Mini Kit (Qiagen) according to manufacturer's protocol. During isolation, RNA samples were treated with DNase I to remove contaminating DNA. RNA integrity was evaluated on an Agilent Bioanalyser RNA 6000 nano chip, and was quantified using the Qubit® Broad Range (BR) assay kit (ThermoFisher). Individually barcoded cDNA libraries were prepared using the Truseq v2 unstranded kit (Illumina) according to the manufacturer's protocol. Library integrity was checked on an Agilent Bioanalyser using the High Sensitivity DNA analysis kit and library quantification was performed using the Qubit® High Sensitivity (HS) DNA assay kit. Individually barcoded libraries were diluted to 10 ng/μl, and sequenced by Genome Quebec (Montreal, QC, Canada) using either the Illumina HiSeq4000 platform 2 x 150 bp paired-end chemistry or the Illumina HiSeq2000 platform 2 x 100 bp paired-end chemistry. Raw data is deposited in NCBI as SRP133837.

- *Developing spike*: Data collected from the 2011 and 2012 glasshouse experiments were benchmarked using penultimate internode auricle distance (AD) phenotypes against pollen cell division stage, in the developing spike in response to withholding water from the pots in which the plants were maintained (71). The rate of soil drying was related to level of leaf turgor conferred by contrasting watering regimes (+/- water). The AD phenotype (from AD=0 cm to full head emergence) was also related to the Zadoks classification of the respective plant. Samples were collected in biological triplicates over the course of 10 days following the start of water exclusion at AD=0 from the pots in both the 2011 and 2012 trials. The doubled haploid lines used for the experiments were selected from progeny from a Westonia x Kauz cross, based on molecular markers indicating that they were either more closely related to the Westonia parent (D08-299) or Kauz (D02-105) while minimizing differences in overall plant phenology (71). Plants at different stages of development, based on the AD phenotype, were used for RNA extraction by removing the leaf tissue around the developing spike and snap-freezing the spike tissue. RNA extraction followed standard protocols. Total RNA preparations were converted to cDNA and polyA-plus selected cDNAs used for preparing libraries (not strand specific) for paired-end sequencing using Illumina technology. Raw data is deposited as PRJEB25640.
- *Aneuploidy controls*: Chinese Spring control plants were grown in growth chambers at 22°C and 16:8 h day:night length. RNA was extracted at seventh leaf stage from the fourth leaves and roots of four/three biological replicates according to (72). Samples were then quantified and evaluated for their quality on an Agilent Bioanalyzer. The 15μg cDNA libraries were prepared by TruSeq RNA sample preparation kits (Illumina). cDNA libraries were indexed (as additional genotypes were also run on the same flow-cell) and sequenced on an Illumina HiSeq2000 (100 bp single-end read run). Raw data is deposited as PRJEB25593.

In total, the 850 RNA-Seq samples used in this study are derived from over 30 different wheat cultivars and germplasm stocks with Azhurnaya (209 samples) and Chinese Spring (123 non-stressed samples) representing the predominant cultivars (Table S1).

Homoeolog specific mapping by kallisto in nulli-tetrasomic lines

We confirmed homoeolog specific mapping (15) of kallisto (14) using a series of criterion. First, we analyzed expression of HC genes expressed >0.5 TPM in nulli-tetrasomic wheat lines from the publicly available study SRP028357 (49). We found that the mean expression of genes on the deleted chromosome was 5.6% of the level in samples with that chromosome indicating stringent homoeolog-specific mapping (Fig. S1).

We then selected only 1:1:1 triads whose triad expression sum (A+B+D) was > 1.0 TPM in root or leaf samples in the wild type control. We selected only triads on chromosome 1, and removed any triad for which the A, B, or D genome expression level was 0. We then calculated the percentage of reads mismapping for each genome in the nulli-tetrasomic lines. For example, for the A genome the percentage mismapping is calculated as:

$$(\text{TPM of A genome in nulliA} / \text{TPM of A genome in wild type control}) * 100$$

We found that the mean expression of these 1:1:1 triads was 3.9% in nulli-tetrasomic lines with respect to the wildtype controls. The distribution of mismapping is positively skewed therefore the mean is not the most representative way to show the data (Fig. S2). The median expression in nulli-tetrasomic lines with respect to wildtype controls is much lower (0.68%) than the mean (3.9%).

To investigate whether mismapping affects different genomes (A, B, and D) or tissues to differing degrees we used the same nulli-tetrasomic lines, but this time we analyzed the data from the leaf and the root samples separately. We found that in both tissues the level of mismapping was low (median 1.15% in leaf, median 0.36% in root), but that in the leaf the D genome had more mismapping (median 1.68%) than the A and B genome (medians 0.32% and 1.32% respectively; Kruskal-Wallis with Dunn multiple comparison P adj <0.001; Fig. S2A). The difference between the A and B genome in the leaf was also statistically significant (Kruskal-Wallis with Dunn multiple comparison P adj <0.001). However, in the root (Fig. S2B) the only statically significant difference between genomes in mismapping was between the B (median 0.29%) and D (median 0.40%; Kruskal-Wallis with Dunn multiple comparison P adj <0.001). Multiple testing adjustments were carried out using the Benjamini-Yekutieli correction (67).

Although we identified a slightly higher level of mismapping to the D genome > B genome > A genome in the leaf samples (D vs A 1.36% higher median mismapping), this is reduced to 0.11% in the equivalent root samples (D genome vs B genome). The subtle D genome bias we identified was two-fold. First, we observed that the relative contribution of the D genome to the overall transcript abundance of triads was higher than that of the A and B genome for 11 out of 15 tissues. This was consistent in both leaves and root samples in Chinese Spring (CS) and Azhurnaya (Table S5). Second, we identified a lower frequency of D-homoeolog suppression (Table S6) compared to A- or B-homoeolog suppression across all tissues, including the equivalent leaf and root samples as in the nulli-tetrasomic analysis. Given that the subtle D-genome bias was consistent in roots and leaves, even though roots show only 0.11% higher mismapping to the D genome, suggests that it is most likely not due to the slightly higher mismapping of the D genome in the leaf samples. We also used the data from the nulli-tetrasomic lines to examine whether balanced triads showed more mismapping compared to dominant or suppressed triads. We found no evidence for balanced triads having higher mismapping in either the leaf (Fig. S2C) or root (Fig. S2D).

Theoretical read mapping accuracy between homoeologs using SNP distributions

To put the 3.9% mean mismapping seen in the nulli-tetrasomic lines compared to wild type in context we next calculated the theoretically expected mismapping rate based on the distribution of SNPs throughout the Chinese Spring RefSeqv1.0 genome sequence. First, we assigned homoeologous SNPs as either homoeolog-specific or semi-specific for a given genome. Homoeolog-specific SNPs are those SNPs which are unique to a single homoeolog and are therefore used by kallisto to assign a read-specifically. The same position in the two other genomes would be considered semi-specific as the position would discriminate against the first genome, but could not distinguish between the second and third genome. It is important to note that two semi-specific SNPs can also generate a homoeolog-specific haplotype when combined. If we assume that two semi-specific SNPs would generate a homoeolog-specific haplotype when combined (i.e. the SNPs were semi-specific between different genomes) we would have the “best-case” scenario where the distance between semi-specific SNPs would be the distance required to distinguish homoeologs. If we are more stringent and we require that we have two homoeolog-specific SNPs to generate a homoeolog-specific haplotype then we can generate a second distance metric for the distance between homoeolog-specific SNPs (the “worst-case” scenario). Based on these assignments, we calculated the average distance between these two SNP types across all the HC triads expressed in Chinese Spring (CS) or Azhurnaya, and compared these distances with the effective read length of the CS (~200 bp) and Azhurnaya (~250 bp) RNA-Seq samples. The effective read length used by kallisto is twice the single read length for paired end (PE) samples (14). The 209 Azhurnaya RNA-Seq samples are 125 bp PE, meaning that the effective read length is ~250 bp. In the case of CS, over 91% of RNA-Seq samples (113/123) are 100 bp PE or longer meaning that their effective read length is ~200 bp (or more).

We found that over 94.7% of homoeolog specific SNPs were closer than 200 bp in the CS HC expressed genes, with this value rising to 99.7% for semi-specific SNPs. This means that 5.3% of homoeolog-specific SNPs and less than 0.4% of semi-specific SNPs have a distance greater than 200 bp in CS and would probably lead to ambiguous read mapping. For Azhurnaya (using a 250 bp cutoff as determined by the effective read length) we find that less than 3.5% of homoeolog specific and less than 0.15% of semi-specific SNPs have a distance greater than 250 bp. This suggests that a very small fraction of the transcriptome reference will lead to possible ambiguous mapping of reads.

Genome of origin effect on gene expression

To assess whether genome of origin has an influence on gene expression we carried out hierarchical clustering. After applying the *initial 850 filter*, the 1:1:1 syntenic triads in Azhurnaya which were expressed >0.5 TPM in at least 3 samples were selected (17,481 triads). The A, B, and D genome were considered separately for each tissue/age time-point (70 tissue time-points in total). Expression levels were normalized by a log transformation: $\log_2(\text{TPM}+1)$. Hierarchical clustering was carried out using the R function “hclust” using the “Euclidean” distance method and the clustering method “average”. The R package “pvclust” (73) was used to estimate uncertainty in the clustering using the same parameters with 1,000 bootstraps. The tissue types largely explained the pattern of clustering, although at a fine scale the genome of origin influenced clustering (Fig. S4).

Expression complexity

To determine the expression complexity of the Azhurnaya transcriptome, the average TPM for each of the 22 intermediate tissues was calculated based on the genes expressed using the *initial 850 filter* criterion. The relative contribution of each gene to the total transcripts within each of the 22 intermediate level tissues was calculated by dividing the individual gene TPM by $1\text{E}+06$. These relative contributions to the total transcript abundance were ranked from the highest to the lowest value

within each of the 22 intermediate tissues. The details of the number of genes at 5% increments is presented in Table S2.

Differential expression (Azhurnaya time course)

We identified genes which were differentially expressed between the 22 intermediate level tissues available for the cultivar Azhurnaya. We selected HC genes which were expressed in at least three samples at >0.5 TPM for differential expression analysis, which was carried out using the R package DESeq2 v1.14.1 (66) (using counts instead of TPM). Pairwise comparisons were made between each of the 22 intermediate level tissues. Genes were considered differentially expressed if up- or down-regulated >2 fold with an FDR (74) adjusted *P* value <0.001 (Fig. S3).

eFP browser

We stored the RNA-Seq TPM values for each homoeolog in an SQL database on the Bio-Analytic Resource for Plant Biology at <http://bar.utoronto.ca>. An image depicting the approximate appearance (growth stage, plant organ) of the samples used for RNA extraction was adapted from online resources (e.g. WheatBP; (75)) and drawn using Inkscape version 0.92.1. A configuration file was created to link up the image with the sample names, and the eFP Browser software (19) was slightly modified to be able to accept wheat gene identifiers. The Wheat eFP Browser is available at http://bar.utoronto.ca/efp_wheat/cgi-bin/efpWeb.cgi and a screenshot is shown in Fig. S5.

Inference of gene families and homoeologous groups

Gene families and homoeologous gene groups were inferred using a phylogenomic approach established recently (10). In brief, orthologous, outparalogous, homoeologous and inparalogous gene relationships were detected using species-tree reconciliation based on gene trees of families inferred from the predicted protein-coding genes of 14 Viridiplantae species that in addition to bread wheat contained nine grasses as well as *Arabidopsis thaliana*, *Selaginella moellendorffii*, *Physcomitrella patens* and *Chlamydomonas reinhardtii* as non-grass outgroups. The nine grasses comprised: *Oryza sativa*, *Sorghum bicolor*, *Zea mays*, *Brachypodium distachyon*, *Hordeum vulgare* (cultivar Morex and variety nudum), *Secale cereale*, *Aegilops tauschii* (D progenitor), *Triticum urartu* (A progenitor). In this procedure homoeologs are defined as orthologs among the sub-genomes which were treated as distinct taxa. Deviating from the procedures described by IWGSC (10), we defined syntenic homoeologous gene groups as homoeologs inferred by the reconciliation approach that were part of a conserved colinear block among the sub-genomes identified by MCScanX (76), i-ADHoRe (77) or DAGchainer (78). A syntenic triad is defined as a triad where at least one of the gene pairs is syntenic.

Ontology term annotation and TF classification

The pipeline also contained a step annotating the domain architectures of the gene family members. The inferred domain architectures were utilized to identify gene families belonging to super-families of transcription factors, transcriptional and post-transcriptional regulators using a HMM-domain rule set established previously (79). The orthologous relationships were utilized to establish Gene Ontology (GO), Plant Ontology (PO) and Plant Trait Ontology (TO) term annotations for bread wheat by homology annotation transfer (10). This pipeline explicitly discarded ontologies related to biotic or abiotic stress. Therefore, to complement the functional annotation, the gene models were aligned to the Arabidopsis proteome (tair10) with blastx. Matches were called with a cut-off e-value <e-10 and GO terms were transferred from the GO assignment of the matching tair10 Arabidopsis annotation. We identified the Arabidopsis proteins with GO terms relating to biotic and abiotic stress, by using the following Plant GO slim (<http://geneontology.org/page/go-slim-and-subset-guide>) terms: GO:0006950: response to stress; GO:0009607: response to biotic stimulus and; GO:0009628: response

to abiotic stimulus. Wheat genes homologous to Arabidopsis proteins with these GO slim terms were extracted from the blastx output and these functional annotations were added to the original IWGSC annotation (10). The GO release was the monthly freeze of 01/01/2017.

Independent measures of D genome homoeolog expression bias

In addition to the analysis in the main text, we analyzed the expression data for possible homoeolog bias in four independent ways.

1. Inference of gene expression level and breadth categories and sub-genome expression bias: TPM abundances for each locus (including both HC and LC genes) were condensed using the median values. The condensed TPM values ≤ 0.5 were placed into the expression level category E0 (no - very weak expression), while TPM > 0.5 were clustered into ascending expression level categories via kmeans using $k=4$ (E1: very weak – weak expression; E2: weak expression – medium expression; E3: medium – strong expression; E4: strong – very strong expression). Relative expression breadth was determined as ratio of samples with TPM signal > 0.5 . Relative expression breadth was categorized using the R built-in *cut* function into five breadth level categories: very few (0 - 0.2), few (0.2 – 0.4), medium (0.4 – 0.6), many (0.6 – 0.8) and most (0.8 – 1).

These data were used to test for genomic expression level bias among the sub-genomes by looking at the numbers of expressed genes. Comparing the complements of expressed genes (E1-E4) reveals a subtle, but significant bias towards $D > A > B$: 52.2% (18,286/35,021) of D, 49.5% (17,953/36,302) of A and 47.3% (17,374/36,738) of B genes are expressed.

2. Genome-wide comparison of expression levels using principal component analysis (PCA): We performed PCA of transcript wise expression levels using the combined TPM data from the 850 RNA-Seq datasets using the R package FactoMineR (80). In this analysis, each sample was used as an independent variable for the transcripts and the subgenomic origin (A, B or D) of each encoding locus was used as a supplementary variable (that is not used to infer the rotated coordinates, but is projected in the resulting components).

In the resulting PCA (Fig. S7), about 62% of the total variation of transcript levels is explained by the first principal component (PC1). The second principal component only accounts for about 2% of the overall variance. Subsequent secondary PCA using PC1-PC10 coordinates of the samples and the sample categories as supplementary variables, indicates that the components $\geq PC2$ represent tissue identity and treatment conditions of the underlying samples. As indicated by the directionality of the variable eigenvectors (Fig. S7, left), the expression strength, i.e. the TPM value in each sample, is the major positive discriminant of the first dimension (PC1, x-axis in Fig. S7, left). Thus, overall high TPM values correspond to large positive PC1 coordinates and low TPM correspond to negative PC1 values. Plotting the PCA projection of the supplementary variable sub-genome origin (Fig. S7, right) reveals a pattern that suggests that overall expression levels of genes located on the D genome are higher than those located on A and B. Genome identity is significantly correlated with PC1 ($p=0$). The overall pattern along PC1 suggests genomic expression levels with $D > A > B$.

3. Comparison of homoeolog expression levels using one-way analysis of variance (ANOVA): For each group of homoeologous genes, locus-wise TPM counts were log-transformed and used to test for significant deviations of expression levels among the homoeologs from the three sub-

genomes using one-way ANOVA (implemented in R using the built-in aov function). Subsequently, we assessed pairwise contrasts using the R built-in TukeyHSD post-hoc test. We corrected resulting *P* values from the post-hoc test using the R p.adjust function employing the method “fdr”. From the results, we tabulated and compared the occurrences of significant deviations in expression levels among the sub-genomes (*P* value cutoff for F-test and post-hoc test: <0.05) using a χ^2 test (R/chisq.test).

54.1% (14,302/26,430) of homoeolog groups had expression levels biased towards D>A, 54.8% (14,699/26,805) towards D>B and 49.0% (13,237/26,996) were biased towards B>A (Fig. S8, left). Assuming equal distribution of biased expression patterns as a null hypothesis, observed frequencies deviate significantly (*P* <1e-16). These numbers support the subtle expression bias towards D homoeologs in the wheat genome.

4. Assessment of family-wise gene expression bias using phylogenetic comparative ANOVA: To assess potential genomic expression bias on the level of gene families, we performed phylogenetic comparative ANOVA of wheat gene expression levels in 5,473 gene families. For this analysis, only families with at least three members and expression data were selected. Phylogenetic comparative ANOVA corrects for the phylogenetic non-independence using branch lengths from the gene trees for each family as weights. As current phylogenetic comparative approaches do not yet support replicated testing, we used the gene-wise coordinates from the first principal component (PC1) from the PCA of the expression levels described above as a proxy variable for gene expression level for each locus. We utilized the function phyANOVA from the R phangorn package (81) to perform a phylogenetic comparative one-way ANOVA with subsequent post-hoc testing and FDR-correction. Subsequently, we tabulated significant deviation in expression levels in gene family members stemming from two sub-genomes and compared the resulting frequencies using the χ^2 test.

The results mirror the pattern observed at the level of homoeologous gene groups; 51.2% (2,544/4,973) of the gene families were biased towards D>A, 51.8% (2,489/4,806) displayed D>B and 48.6% (2,269/4,665) were biased towards B>A (Fig. S8, right). In line with the conclusions from the genome-wide PCA and the homoeolog ANOVA, the results from the family-wise phyloANOVA also point to a mild, albeit significant expression dominance of the D genome.

Genomic compartments

Triads were assigned to genomic compartments based on published criteria (10).

Gene Ontology and Plant Ontology term enrichment

The R package goseq was used to calculate GO and PO enrichments, whilst correcting for the length bias inherent in RNA-seq data. The Benjamini & Yekutieli method (BY, (67)) was used to correct for multiple testing using the R function p.adjust().

CDS, protein and promoter analysis for triads

Protein and CDS sequences were extracted from the RefSeq v1.0 genome (10) using gffread v0.9.8, using the command line:

```
gffread -g <genome.fa> -x CDS.fasta -y Proteins.fasta IWGSCv1.0_UTR_ALL.gff
```


To obtain the sequences of the promoters, we developed a script leveraging PyFaidx (82) and the library comprised in the Mikado program (83). The script takes as input an indexed GFF (produced by Mikado compare) and a list of genes, and subsequently extracts from the genome FASTA file a specified amount of sequence upstream of either the transcription or the translation start site. For this study, we extracted the 5,000 bps upstream of the translation start site of all HC genes included in the RefSeq v1.0 annotation (10). The command line is as follows:

```
extract_promoter_regions.py -nn -eu -d 5000 -z -o promoters <genome>  
IWGSCv1.0_UTR_ALL.gff <list of genes>
```

The version of the script used for this study is present in the Mikado git repository, at the following static address:

https://github.com/lucventurini/mikado/blob/f47aa63/util/extract_promoter_regions.py

A pairwise blast alignment was used on each triad (A vs B, A vs D, and B vs D) for their CDS, translated protein and promoter sequence. The output was produced in XML and parsed with the Ruby package ‘bio-blastxmlparser’ (84). For the promoter regions, the alignments were done for the 1.5kbp, 3kbp, and 4.5kbp upstream; promoters including one or more N’s were discarded from further analyses. The promoter identity was calculated as the number of identical bases over length of the longest HSP (source: https://github.com/TGAC/bioruby-polyploid-tools/blob/master/bin/blast_triads_promoters.rb). For the CDS and protein alignment the identity was on the sum of HSPs of the longest HIT (source: https://github.com/TGAC/bioruby-polyploid-tools/blob/master/bin/blast_triads.rb). Detailed results from these analyses are in Table S21.

Transposable Element (TE) modeling using CLARITE

Annotation of TEs was described elsewhere (10). Briefly, TE modelling was achieved through a similarity search approach based on the ClariTeRep curated databank of repeated elements (github.com/jdaron/CLARI-TE), developed specifically for the wheat genome, and with the CLARITE program that was developed to model TEs and reconstruct their nested structure (85). This generated a GFF file with TE coordinates.

TFBS comparison between homoeologs

As the functional relevance of TFBS copy number in individual promoters is difficult to predict, we considered the presence or absence of unique TFBS within each promoter. Across all genes, we identified a total of 1,031 unique TFBS with a median of 242 TFBS identified in each promoter. For each triad, we categorized each unique TFBS based on the presence/absence of the TFBS in each homoeolog:

- “all_same”: motif present in the A, B and D homoeolog
- “A_diff”: motif only present in the A homoeolog OR motif present in B and D homoeolog and not the A homoeolog
- “B_diff”: motif only present in the B homoeolog OR only present in A and D homoeolog
- “D_diff”: motif only present in the D homoeolog OR only present in A and B homoeolog

We then compared the distribution of each TFBS category between triad expression categories. Kruskal-Wallis and pairwise Mann-Whitney tests with *P* values adjusted using the Benjamini-Hochberg method were used to test for significant differences between distributions.

Ka/Ks analysis

The canonical transcripts of each syntenic and non-syntenic triads were aligned by peptide sequence using MAFFT v7.310 (86) (options -maxiterate 1000 and -localpair) and the binder provided in BioRuby 1.5.1 (84). Source code is available at https://github.com/TGAC/bioruby-polyploid-tools/blob/master/bin/mafft_triads.rb. The alignments were then converted to nucleotide sequence using the CDS sequences and the R function “reverse.align” (R version 3.3.3). From these nucleotide alignments, Ka and Ks values for each pairwise comparison between homoeologs were calculated using the function “kaks” from the R library “seqinr” (R version 3.3.3, seqinr version 3.4-5 (87)). The Ka and Ks values for rates of non-synonymous and synonymous mutations, respectively, are calculated based on the methodology from Li (88). From these values, we obtained the Ka/Ks values reported in the main text. Instances where the Ka/Ks value was infinite (due to a Ks value of 0, and a non-zero Ka) were changed to 10 as a comparatively large, non-infinite number, while negative and NaN values of Ka/Ks (due to missing information or zero Ka and Ks values, respectively) were considered as “NA” values and excluded from further analysis. Comparison of Ka/Ks values between subsets of triads was based on the allocation of triads to stable/dynamic or synteny categories as detailed in the text and in Table S22. The significance of the differences in Ka/Ks values between subsets was calculated using the Mann-Whitney-Wilcoxon test for non-normal distributions in R (function “wilcox.test”, alternative = “two.sided”, R v. 3.3.3).

MADSII phylogenetic tree construction

The three homoeologous MADSII transcription factors in root module 61 were TraesCS2A01G337900.1, TraesCS2B01G344000.1 and TraesCS2D01G325000.1. Since the homoeologues were highly similar, the protein sequence of TraesCS2D01G325000.1 used for finding orthologues using blastp on the EnsemblPlants website (<http://plants.ensembl.org/index.html>) against genomes for *Oryza sativa* ssp. *Japonica* IRGSP 1.0, *Hordeum vulgare* Hv_IBSC_PGSEB_v2, *Glycine max* v1.0, *Arabidopsis thaliana* tair10, *Brachypodium distachyon* v1.0 and *Zea mays* AGPv4. The protein sequence for the canonical transcript was downloaded from EnsemblPlants. Only hits with a percentage ID >60 % were kept for *Oryza sativa*, *Hordeum vulgare*, *Brachypodium distachyon* and *Zea mays*. Hits with percentage ID >50 % were kept for *Glycine max* and *Arabidopsis thaliana*. Wheat genes to include in the tree were identified by blastp against the IWGSC RefSeqv1.0 peptide annotation, keeping the top 30 hits which equated to 26 genes. Peptide sequences for the canonical transcripts for these genes were then extracted. The sequences for all seven species were aligned using clustal omega online (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) (clustal-omegav1.2.4) and the multiple sequence alignment was generated by clustalW2 (v2.1) using default parameters (Neighbour-joining clustering). An excerpt of this tree focusing on the proteins most closely related to the three homoeologous MADSII transcription factors is shown in Figure 4C. The full tree is available at <http://itol.embl.de/shared/borrillp> in the “Ramirez-Gonzalez et al., 2018” project. The section of the tree presented in Figure 4C is highlighted in red, some branches have been rotated in the Figure 4C to present the results more clearly.

Statistical analysis

The statistical tests, sample sizes and the corresponding corrections for multiple testing are listed throughout the main text, supplemental figures, and supplemental tables.

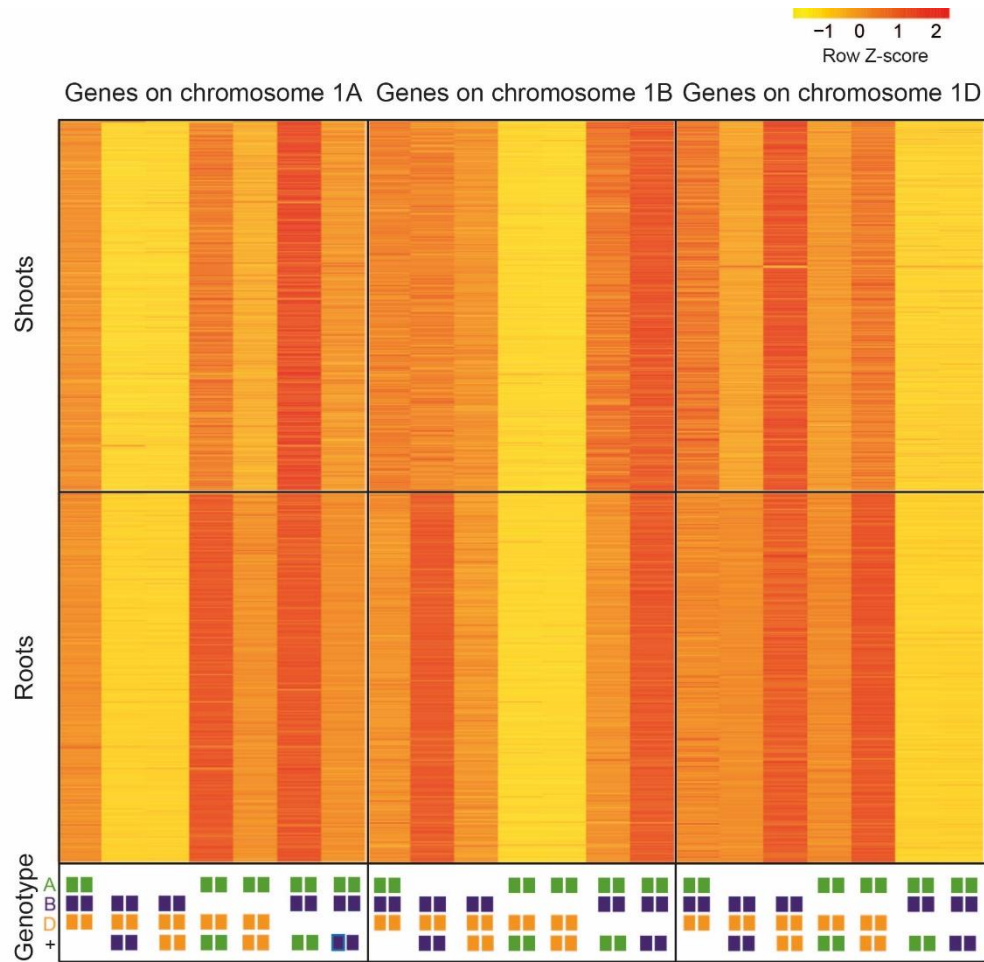


Fig. S1. Expression of genes on chromosome 1 in nulli-tetrasomic wheat lines in shoots and roots. Genotypes for chromosome 1 are indicated in colored squares: A genome in green, B in purple, and D in orange. Squares listed in the bottom row (+) indicate extra copies (tetra), absence of squares indicates deletion (nulli) of entire chromosomes.

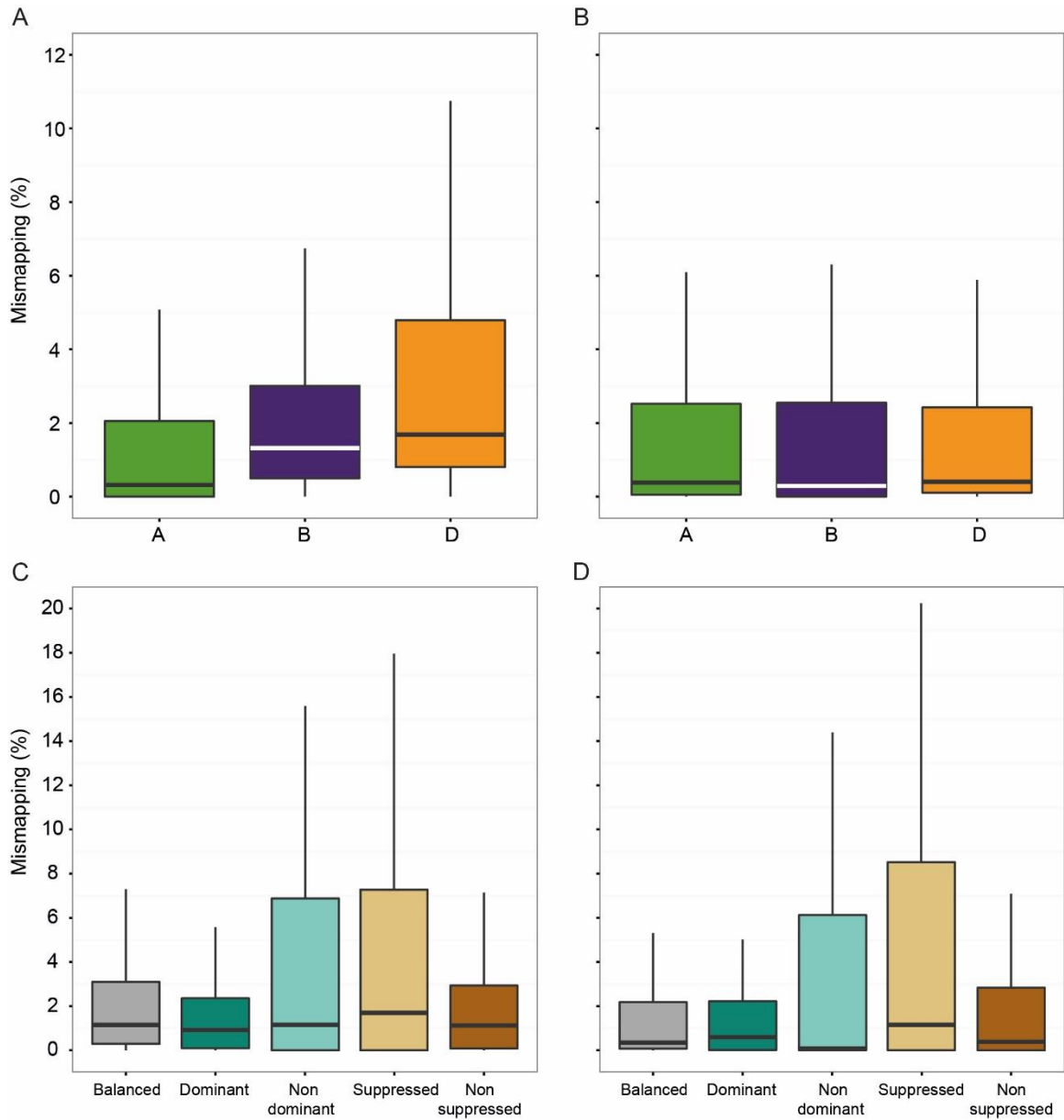


Fig. S2. Mismatching rates estimated using chromosome 1 nulli-tetrasomic lines. Mismatching was calculated for each triad as the TPM of the genome which is missing in the nulli-tetrasomic line (e.g. A genome in the nulli 1A line) divided by the TPM of that genome in Chinese Spring control lines, multiplied by 100. The percentage of mismatching in nulli-tetrasomic lines missing either chromosome 1A, 1B, or 1D are shown for leaf (A) and root (B). Mismatching for triads in different dominance categories is shown for leaf (C) and root (D). Mismappings are shown for triads categorized as balanced (mismatching for all three genomes A, B, and D), dominant (mismatching in the genome of the triad which is dominant, e.g. A for A dominant triad), non-dominant (mismatching in a dominant triad, in the genomes which are not dominant, e.g. B and D genomes for an A dominant triad), suppressed (mismatching in the genome of the triad which is suppressed, e.g. A for A suppressed triad), and non-suppressed (mismatching in a suppressed triad, in the genomes which are not suppressed, e.g. B and D genomes for an A suppressed triad).

		Genes downregulated in tissues (< 0.5 fold padj < 0.001)																					
		anther	awns	embryo	endosperm	flag leaf blade	flag leaf sheath	glumes	grain hard dough and ripening	grain milk and soft dough	internode	leaf blades excl flag	leaf ligule	leaf sheaths excl flag	peduncle	root apical meristem	roots	seedling aerial tissues	shoot apical meristem	shoot axis	spike	spikelets	stigma & ovary
		-	11304	14116	10371	14409	13915	12445	10709	10754	12474	13493	9464	11124	12845	13422	12068	11954	12638	12581	10508	7298	10156
Genes upregulated in tissues (> 2 fold padj < 0.001)	anther																						
	awns	13249	-	15870	13995	1848	1253	736	14015	14349	2461	2124	848	5577	4179	15681	12694	5743	12713	11887	10391	4322	12415
	embryo	10619	9291	-	2603	13137	12018	10476	129	5621	9528	11574	8089	7223	9946	8932	8662	9446	5882	6150	7192	6242	6159
	endosperm	7629	7931	4332	-	11966	10783	9103	817	1936	8407	11001	7314	7697	8895	9166	8263	9212	7603	7571	6799	5067	5468
	flag leaf blade	15336	1273	19258	17136	-	2970	8714	18137	18132	8895	666	1401	10048	8646	18742	16176	8678	16334	15938	15422	8815	16378
	flag leaf sheath	14725	734	17820	15684	1599	-	2866	16432	16423	2592	2133	704	6339	6973	17260	14197	8149	14435	13703	12678	6756	14068
	glumes	13831	1067	17176	14810	8203	4159	-	15502	15301	2844	7835	1341	6164	5537	17407	14299	8178	13832	13426	10267	3715	12135
	grain hard dough and ripening	10171	9842	2699	2191	15856	14286	11760	-	5372	10783	14243	8485	9001	11740	11790	10486	11476	8887	8834	8712	6423	7290
	grain milk and soft dough	9696	9661	9057	4008	15506	13719	11447	4563	-	10407	12814	6413	6726	9729	10769	8634	8093	7258	7496	4250	2629	2819
	internode	13587	2526	15750	13696	7079	3782	3182	14012	13311	-	6047	1083	2817	3022	15188	10818	7069	10576	8607	9434	4025	11412
	leaf blades excl flag	15532	2147	18751	16997	3013	5207	10281	18122	17757	9631	-	1336	9275	9263	18263	15597	7937	15482	15120	14747	8392	15659
	leaf ligule	10030	1003	12283	11213	1805	803	1147	10282	8824	902	1502	-	886	1679	11993	8455	1809	7816	6980	4348	1147	7747
	leaf sheaths excl flag	16934	11069	18472	17061	17930	15375	13948	18930	15267	10914	15685	5176	-	9899	12823	10691	2929	7111	7812	5738	2206	10178
	peduncle	14440	4303	16955	15092	9597	8573	7303	15781	14410	5736	7955	2534	3938	-	15262	12689	4191	11855	11392	8171	1804	12629
	root apical meristem	13581	15289	13707	12486	20311	18779	17657	13585	11919	15284	18494	11755	5257	14635	-	177	8907	2490	3326	8619	6025	8900
	roots	15518	15602	18140	15547	23723	20806	18965	17648	15354	15497	21724	10316	8216	15673	4311	-	10532	6903	7262	12291	7152	12140
	seedling aerial tissues	15509	10072	18639	17120	15657	15514	14696	18741	15860	13994	12836	4960	3570	7780	14039	12643	-	10789	12027	9308	2943	12958
	shoot apical meristem	13998	13180	10725	11380	18889	17168	15368	11869	8857	12363	16531	8905	1815	12130	4049	3415	7246	-	57	4069	2711	5365
	shoot axis	14317	12556	12621	12177	19920	17316	15179	13045	10232	10880	17583	7417	3012	12075	6622	4149	8214	159	-	6606	3631	7481
	spike	12411	9449	13148	11821	16313	14430	11031	12940	6292	10483	13180	4499	2915	7720	11295	10192	4413	5835	7268	-	199	2881
	spikelets	9219	4544	11710	9915	10237	8993	5509	9706	5101	5732	7043	1992	2107	3690	10046	7883	2913	5528	6087	119	-	3325
	stigma & ovary	8289	6752	6796	5626	11569	9896	7465	5188	849	7407	8689	3996	2728	6443	6395	4943	4318	2459	2933	398	227	-

Fig. S3. Differentially expressed genes between the 22 tissues across the Azhurnaya developmental time course. Numbers on top of the matrix diagonal represent genes down-regulated in tissues (< 0.5 fold and $P_{adj} < 0.001$), whereas numbers below the matrix diagonal are up-regulated (> 2 fold and $P_{adj} < 0.001$) in the corresponding tissue.

Fig. S4. Hierarchical clustering of Azhurnaya developmental time course samples separated into A (green), B (purple), and D (orange) genome. Gene expression from the 17,481 expressed 1:1:1 syntenic triads in Azhurnaya were used for this analysis. The high-level tissue for each sample is indicated by the colored bar at the left with samples originating from grain (beige), leaf (pale green), spike (dark green), and roots (brown). The red and green numbers show the significance of the hierarchical clustering determined via multiscale bootstrapping resampling (approximated unbiased P value, au, red) and normal bootstrapping (bootstrap probability, bp, green). If the hypothesis that genome of origin has a strong influence on gene expression was correct, the clustering of the samples would predominantly be determined by the genome of origin (green, purple, and orange). This hypothesis was disproven with the most basal clustering being due to tissue type, rather than genome of origin. On a small-scale level clustering according to genome of origin was observed; however overall the clustering pattern could not be explained by the genome of origin as shown by the green, purple and orange text being interspersed rather than in three large sections. Instead the clustering pattern was more related to tissue type, for example, with the first branch point being between anthers compared to all other tissues, regardless of their genome of origin.

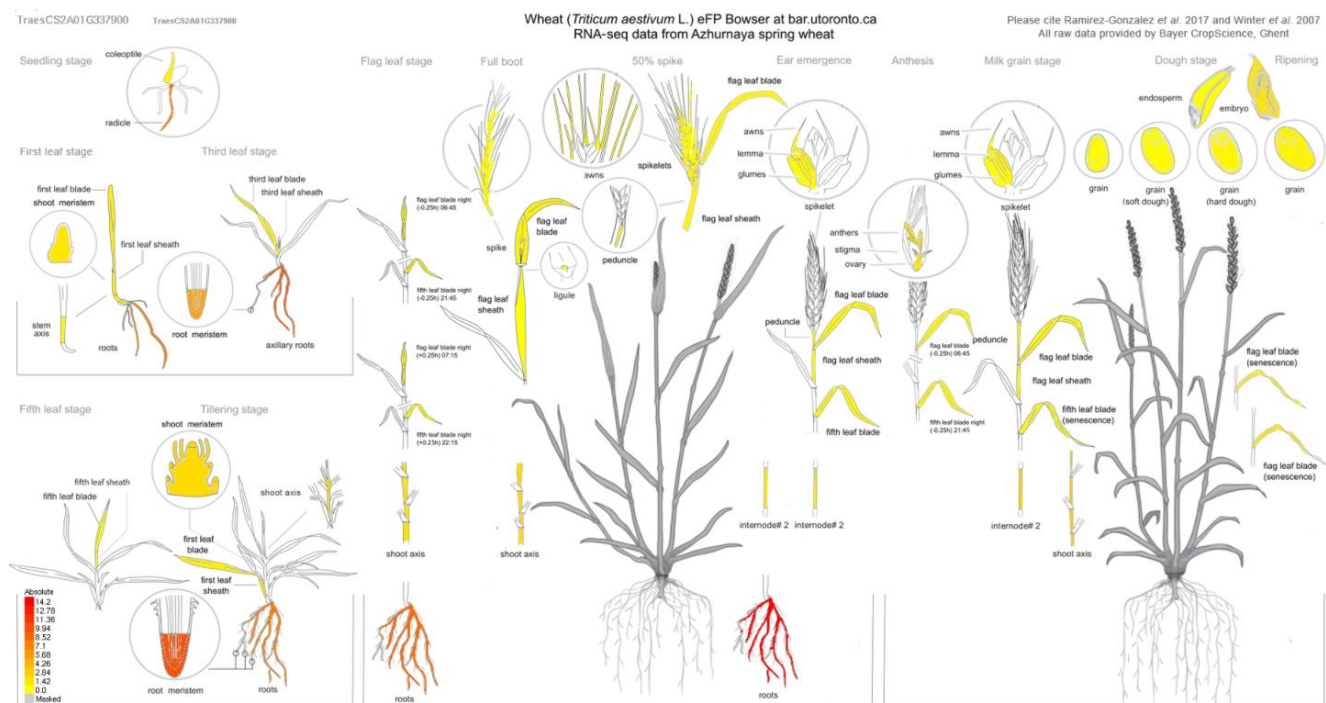


Fig. S5. Wheat eFP browser. The example gene shown is TraesCS2A01G337900, a MADS_II transcription factor expressed specifically in the roots. Color scale is indicated in the bottom left hand of the figure. The wheat eFP browser is available through http://bar.utoronto.ca/efp_wheat/cgi-bin/efpWeb.cgi.

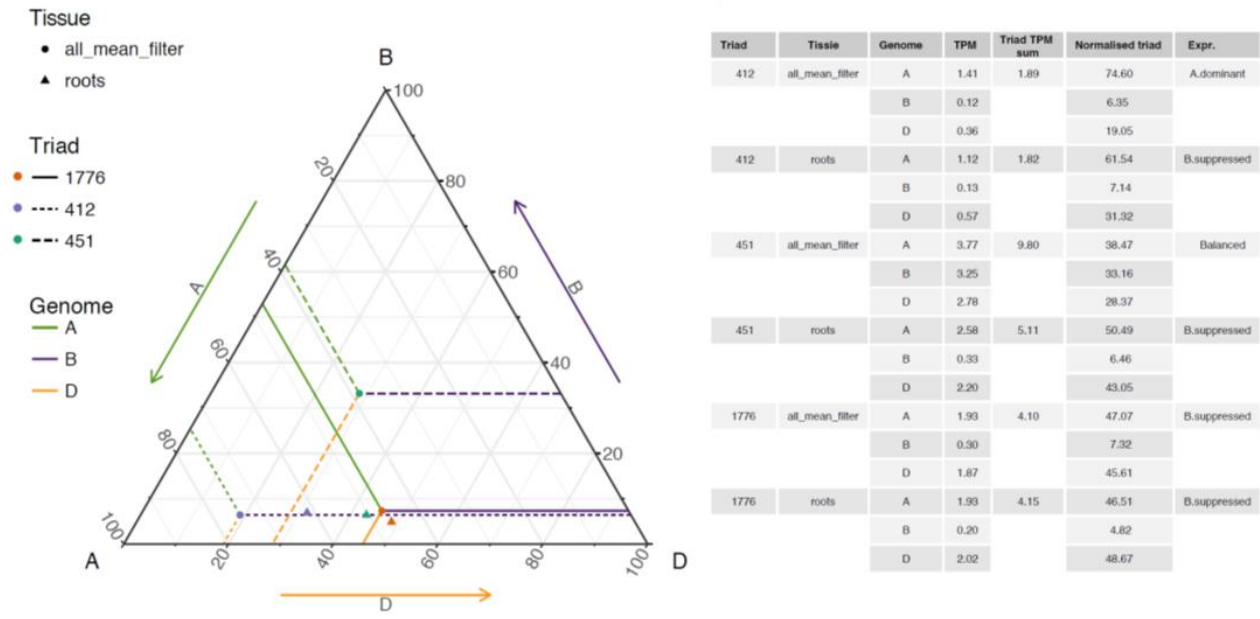


Fig. S6. Normalization of relative expression levels of the A, B, and D genome homoeologs across triads. Table shows the actual absolute expression values in TPM for three gene triads in the root and the combined analysis (all_means_filter) value. First, the expression of the three homoeologs was added (Triad TPM sum) and this was used as the denominator to obtain the normalized triad expression value for each gene/triad. These normalized triad values were used to plot each triad within the ternary plot and to assign the triad to the corresponding homoeolog expression category (Expr. in table). The ternary plot includes the value for the root (small triangle) and for the combined analysis (all_means_filter; circle) with lines being drawn from the latter to each side of the ternary plot to denote the relative contribution of the A, B, and D genome homoeologs.

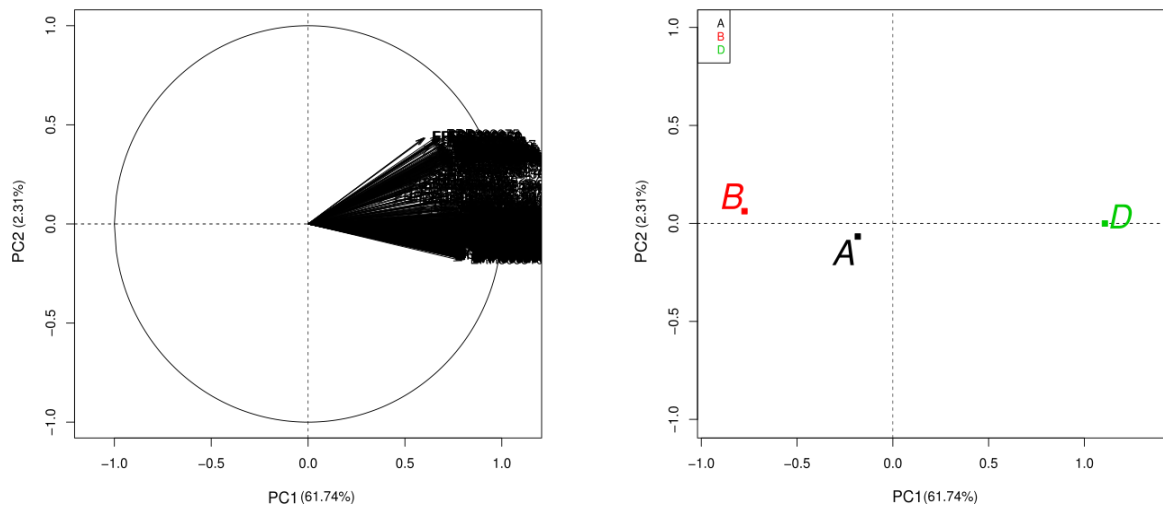


Fig. S7. Principal component analysis (PCA) of gene expression levels. Sub-genome identity was used as supplementary qualitative variable in variable factor (left) or individual factor (right) PCA.

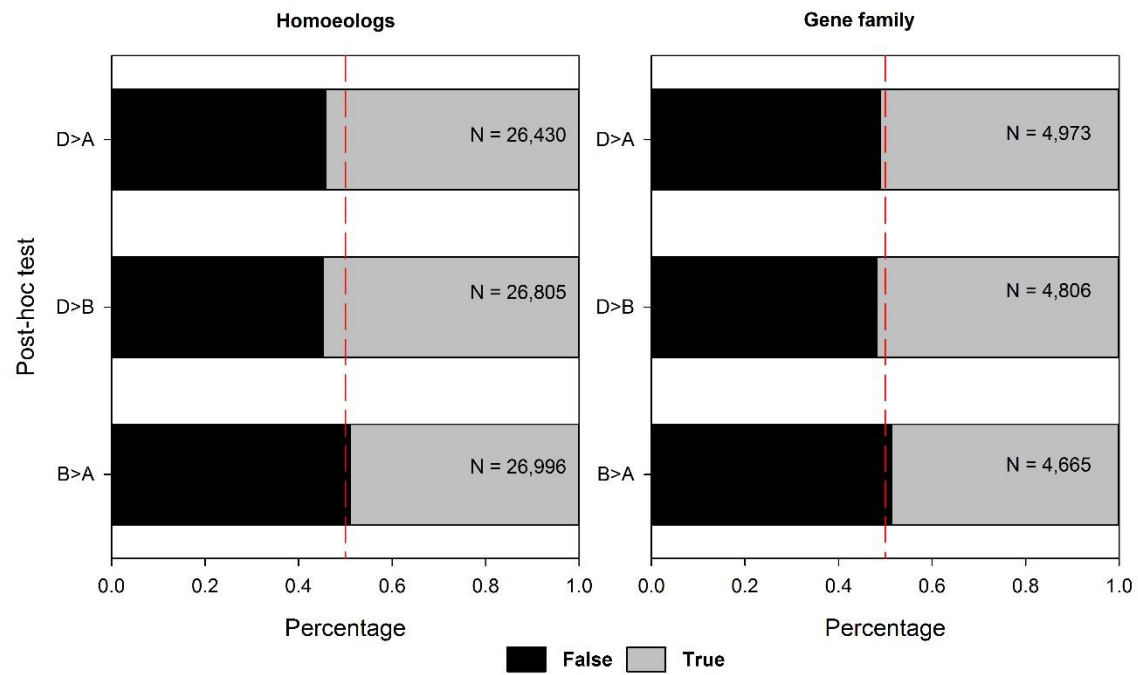


Fig. S8. Significant post-hoc tests for biased expression levels among homoeologs (left) and gene family (right) members. Total number of comparisons (N) are indicated within each bar.

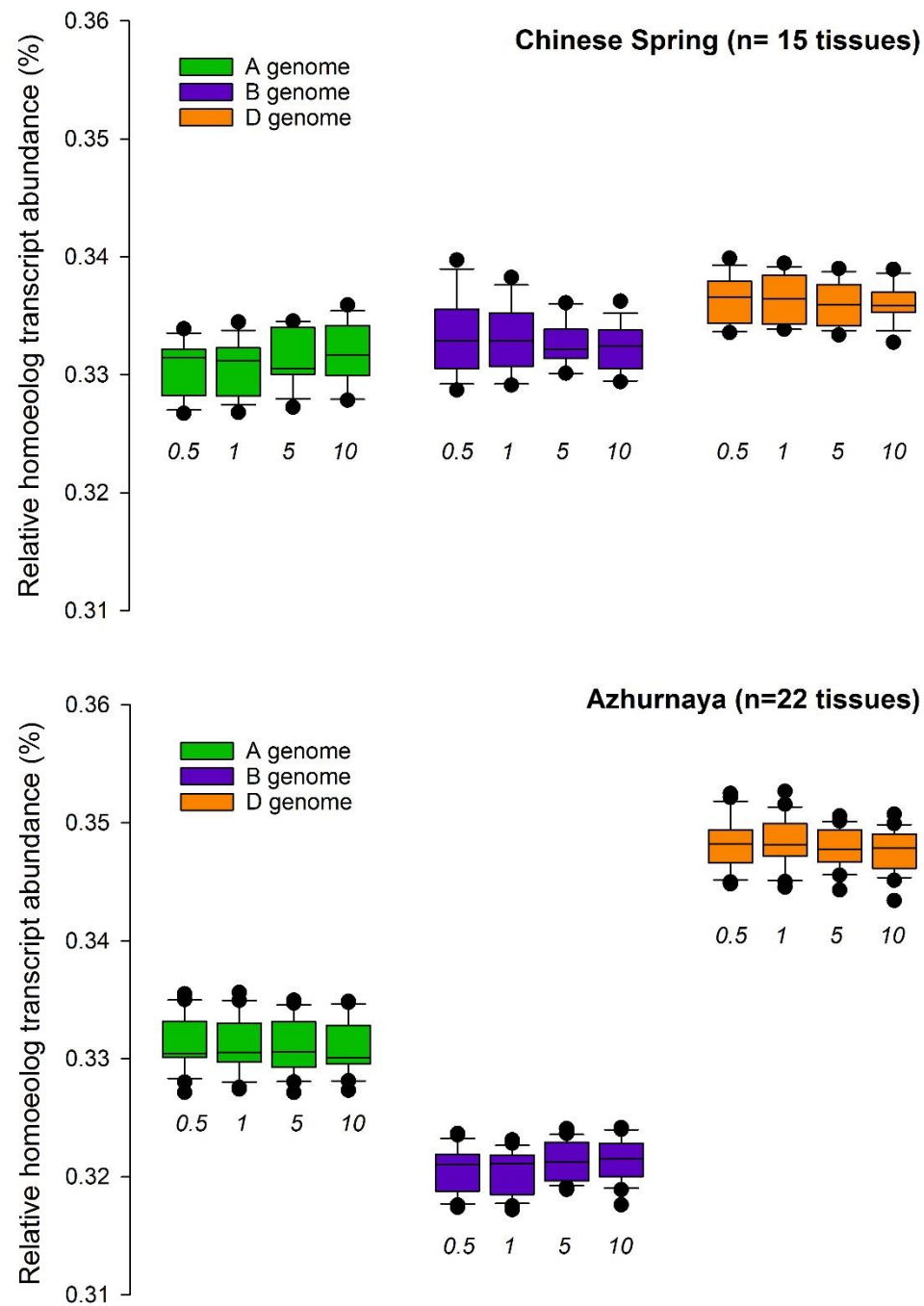


Fig. S9. D genome homoeolog expression bias in CS and Azhurnaya. Boxplots of relative expression abundance of A, B, and D genome homoeologs for syntenic triads across 15 (CS, top) or 22 (Azhurnaya, bottom) tissues at four different minimum TPM cut off values (indicated below each boxplot). Further details in Table S5.



Fig. S10. Homoeolog expression bias assignment across intermediate tissues. A random sampling of 1,000 triads are plotted for each of the 15 tissues indicated in in Fig. 2B. Triads are colored based on their category assignment in the global analysis (Fig. 2A). Only 1,000 triads are plotted since the graphs saturate when all values are shown.

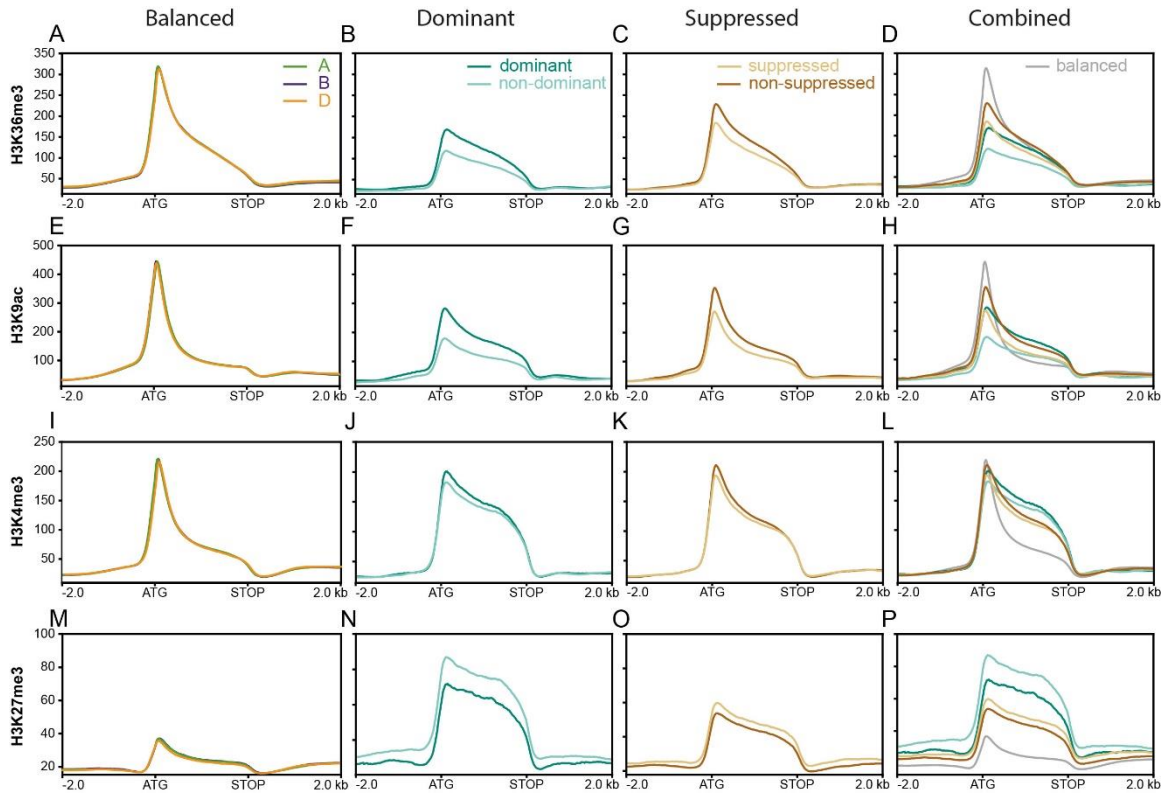


Fig. S11. Metagenome profile for H3K36me3 (A), H3K9ac (B), H3K4me3 (C), and H3K27me3 (D). Histone mark data from -2 kb upstream of the ATG until +2 kb downstream of the stop codon were normalized for gene length and metagenome profiles were plotted for each histone mark. Triads were categorized into balanced, suppressed, and dominant triads. Balanced triads (A, E, I, and M) are shown separated into the A (green), B (purple) and D (orange) genome. Dominant triads (B, F, J, and N) were separated into the more highly expressed dominant homoeolog (teal) and the lower expressed non-dominant homoeologs (pale blue). Suppressed triads (C, G, K, and O) were separated into the more highly expressed non-suppressed homoeologs (brown) and the lower expressed suppressed homoeolog (tan). In panels D, H, L, and P the average of the A, B, and D genomes is shown for balanced triads (grey), dominant triads were separated into dominant (teal) and non-dominant (pale blue) homoeologs, suppressed triads were separated into suppressed (tan) and non-suppressed (brown) homoeologs.

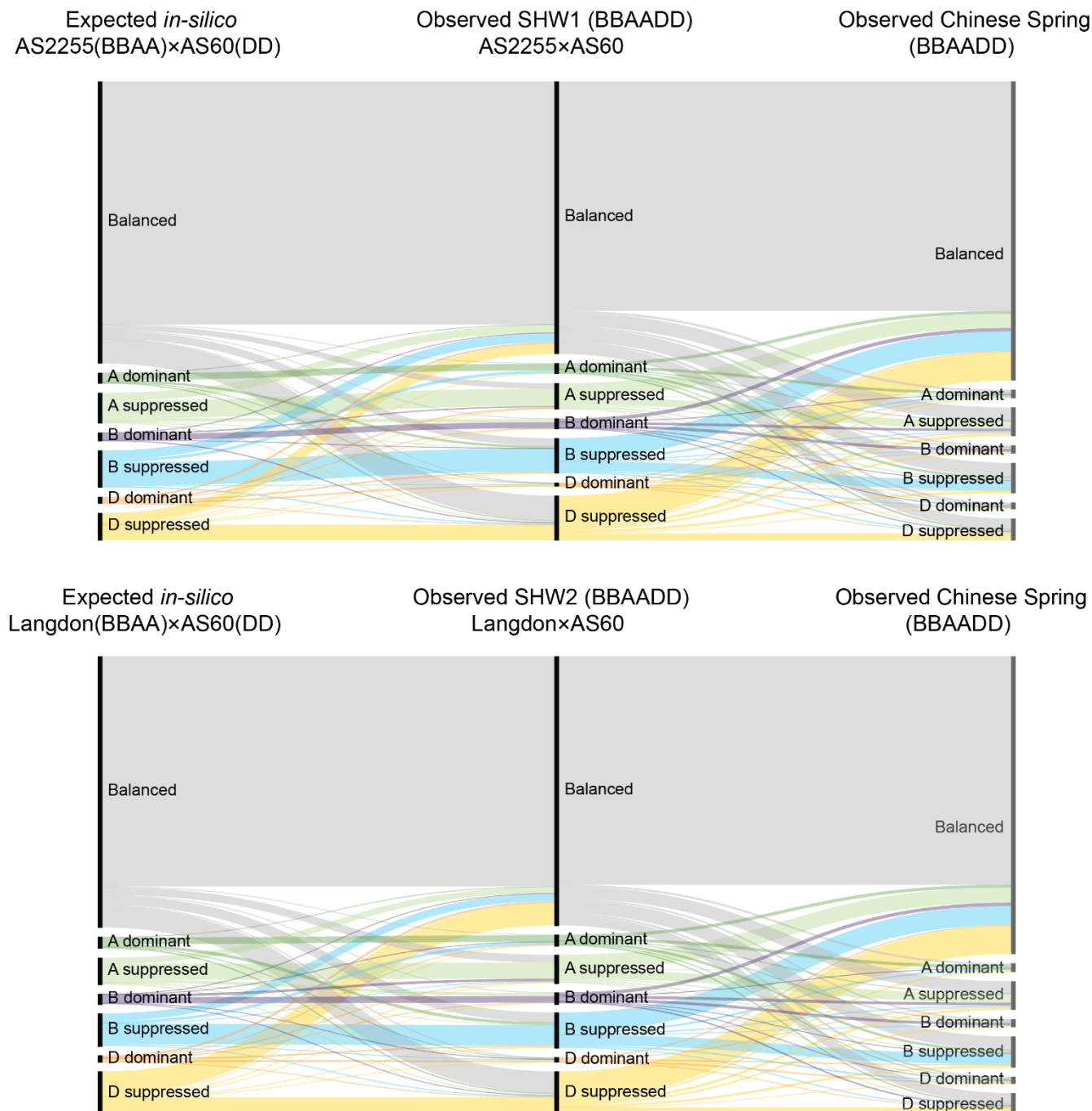


Fig. S12. Alluvial plot of classification of triads across the *in-silico*, synthetic hexaploid, and modern-day Chinese Spring hexaploid wheat. The two graphs represent the comparisons between SHW1 (top; AS2255 x AS60) and Chinese Spring and SHW2 (bottom; Langdon x AS60) and Chinese Spring. The expected *in-silico* data are that which is associated with each SHW.

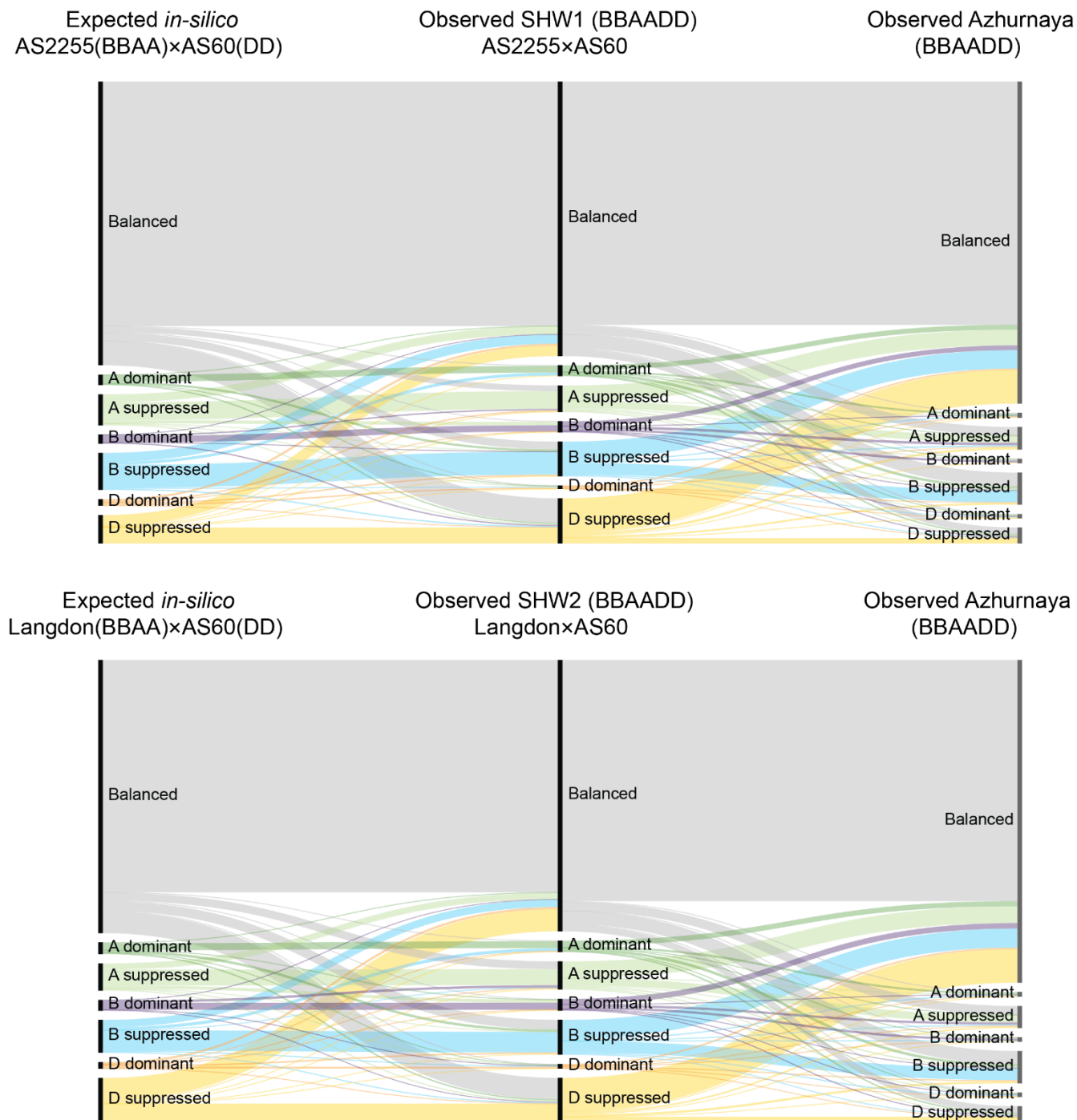
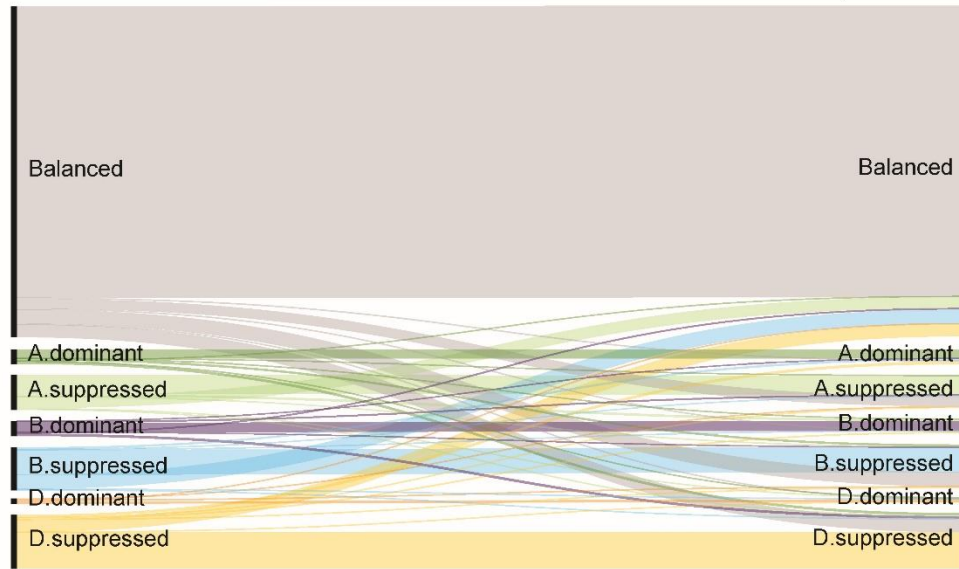


Fig. S13: Alluvial plot of classification of triads across the *in-silico*, synthetic hexaploid, and modern-day Azhurnaya hexaploid wheat cultivar. The two graphs represent the comparisons between SHW1 (top; AS2255 x AS60) and Azhurnaya and SHW2 (bottom; Langdon x AS60) and Azhurnaya. The expected *in-silico* data are that which is associated with each SHW.

Langdon(BBAA)×AS60(DD)
Expected *in-silico*

AS2255(BBAA)×AS60(DD)
Expected *in-silico*



Observed SHW2 (BBAADD)
Langdon×AS60

Observed SHW1 (BBAADD)
AS2255×AS60

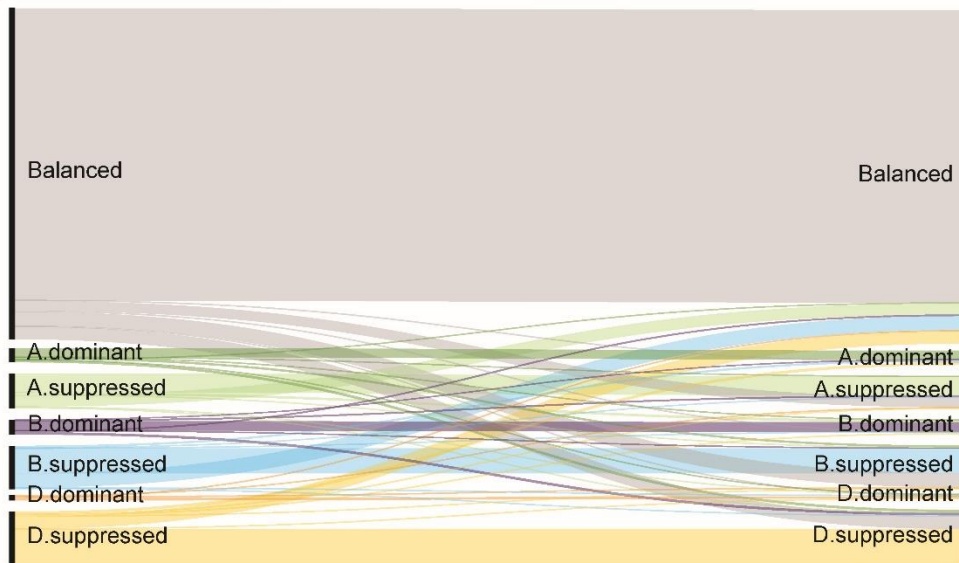


Fig. S14: Alluvial plot of classification of triads comparing the expected *in-silico* (top) and the observed SHW (bottom) datasets.

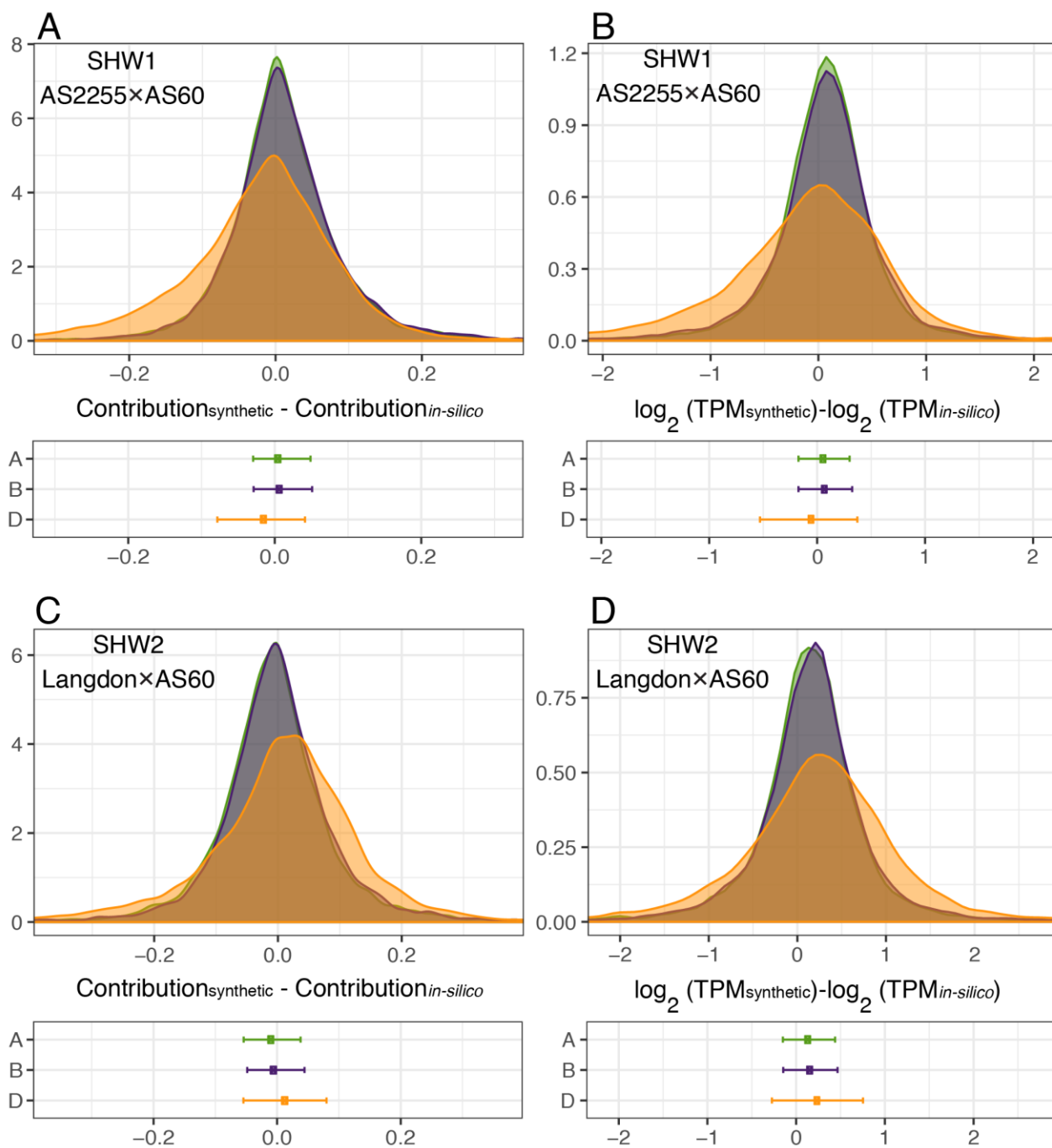


Fig. S15. Differences in the observed and expected relative contributions of each homoeolog between progenitor species and synthetic hexaploid wheat (SHW). Distributions are based on the differences in the relative contribution of each homoeolog to its triad (A, C) and the log₂(TPM) change

(B, D) between the observed values in SHW1 (A, B) and SHW2 (C, D) and the *expected in-silico* datasets. Plots below distributions correspond to distance between the 25% and 75% of the distribution, with the median (50%) shown by the filled square. Positive values indicate that the observed relative contribution was higher in the SHW than would be expected from the *in-silico* dataset from the progenitor species. Likewise, negative values indicate that the observed contribution was lower than expected.

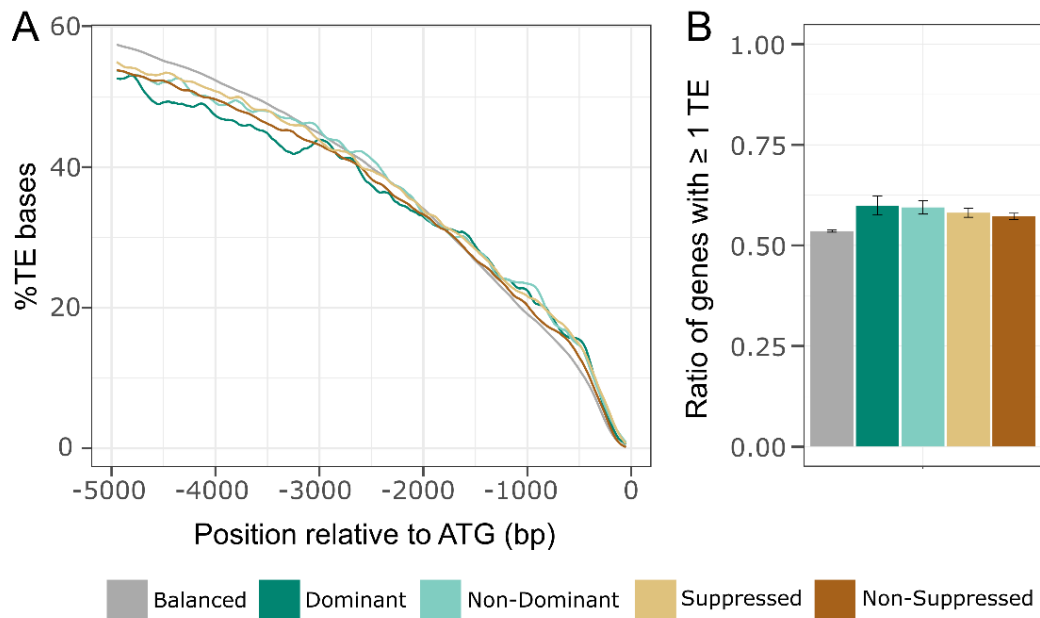


Fig. S16. Transposable element (TE) density and presence across homoeolog expression bias categories. Triads were classified according to their homoeolog expression bias category. Dominant and suppressed triads were further separated into dominant and non-dominant and suppressed and non-suppressed homoeologs based on their relative expression within the triad. (A) TE density was calculated using a sliding window approach over 100 bp intervals for each of the five dominance categories across the 5 kb promoter region. (B) TE presence is the ratio of genes in each category which have at least one TE within the 1.5 kb promoter to all genes in that category. Pairwise comparisons for TE presence between the balanced triads and the other four categories were significant (Mann-Whitney $P < 0.006$ for all but Balanced-Dominant with $P < 0.07$). Error bars show the standard error.

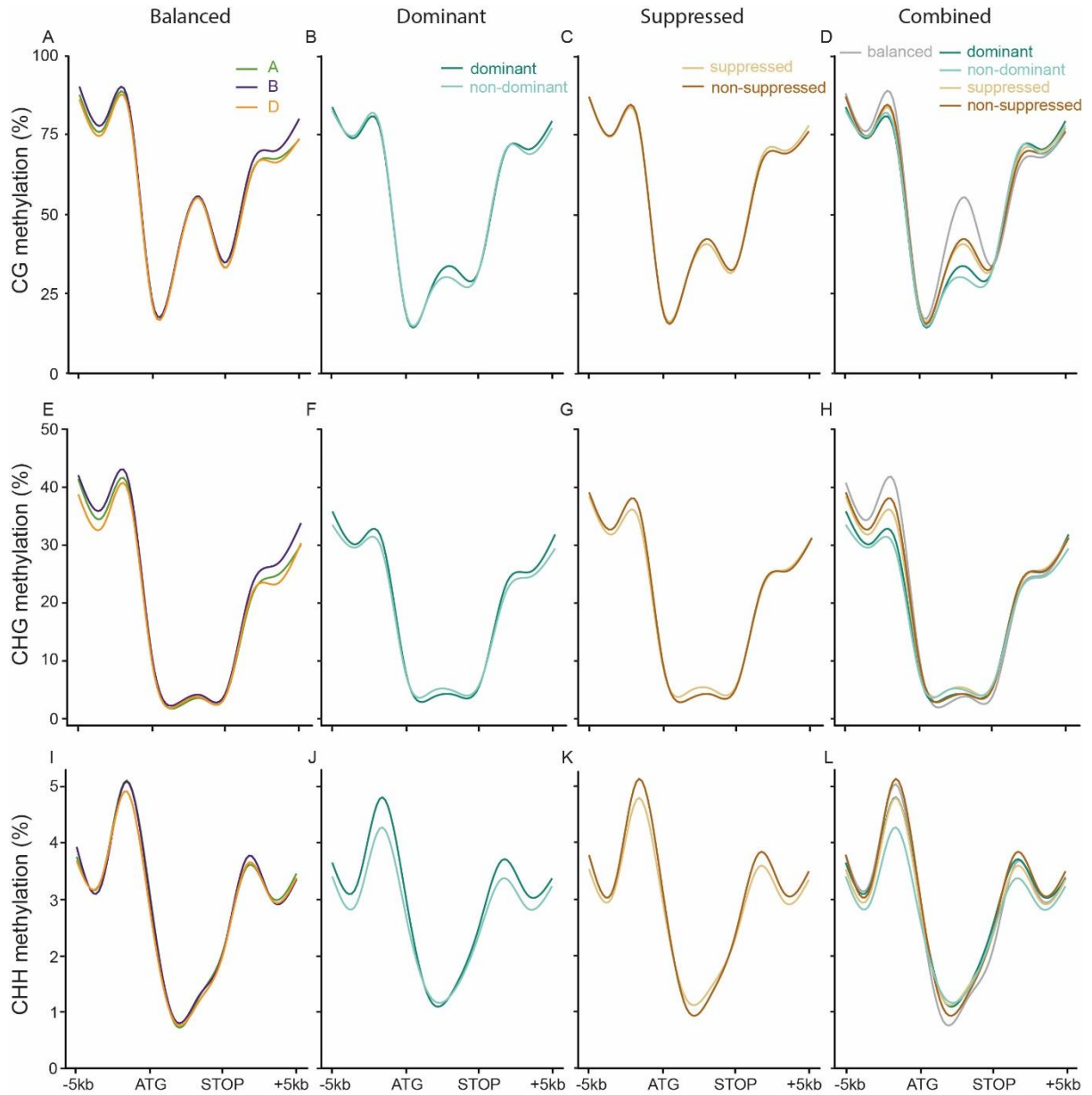


Fig. S17. Metagene profile for CG (A-D), CHG (E-H) and CHH (I-L) methylation. DNA methylation data from -5 kb upstream of the ATG until +5 kb downstream of the stop codon were normalized for gene length and metagene profiles were plotted for each DNA methylation context. Triads were categorized into balanced, suppressed, and dominant triads. Balanced triads (A, E, and I) are shown separated into the A (green), B (purple), and D (orange) genome. Dominant triads (B, F, and J) were separated into the more highly expressed dominant homoeolog (teal) and the lower expressed non-dominant homoeologs (pale blue). Suppressed triads (C, G and K) were separated into the more highly expressed non-suppressed homoeologs (brown) and the lower expressed suppressed homoeolog (tan). In panels D, H, and L the average of the A, B, and D genomes is shown for balanced triads (grey), dominant triads were separated into dominant (teal) and non-dominant (pale blue) homoeologs, and suppressed triads were separated into suppressed (tan) and non-suppressed (brown) homoeologs.

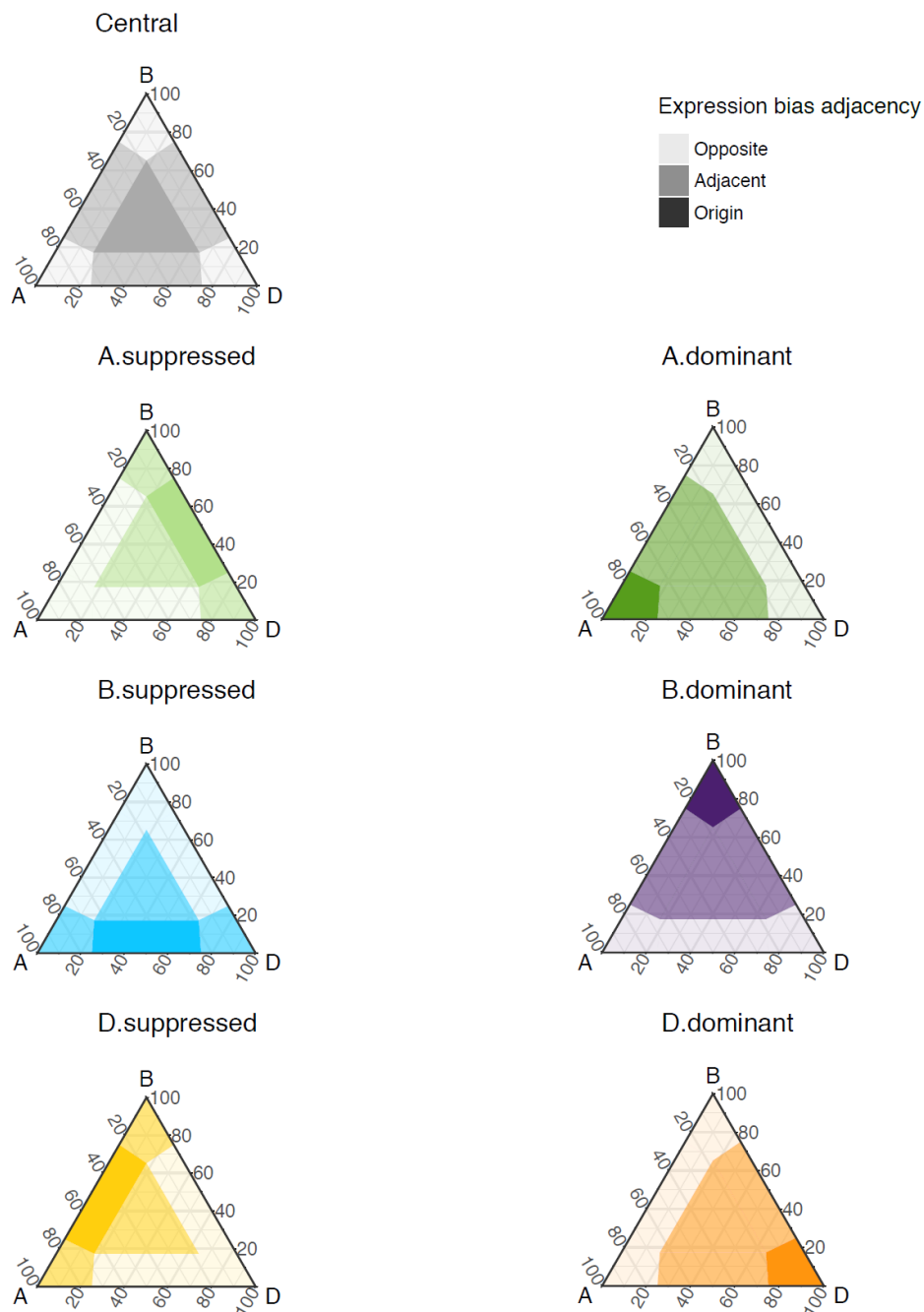


Fig. S18. Schematic of classification for adjacent and opposite categories. Ternary plots are shaded to indicate which homoeolog expression bias categories are considered adjacent (medium shade) and opposite (light shade) for each of the seven origin (dark shade) categories. In the case of the balanced triads, homoeolog-suppressed categories were considered adjacent as they only require variation in a single homoeolog for this classification, whereas dominant categories were considered opposite as they usually require variation in two homoeologs for this classification.

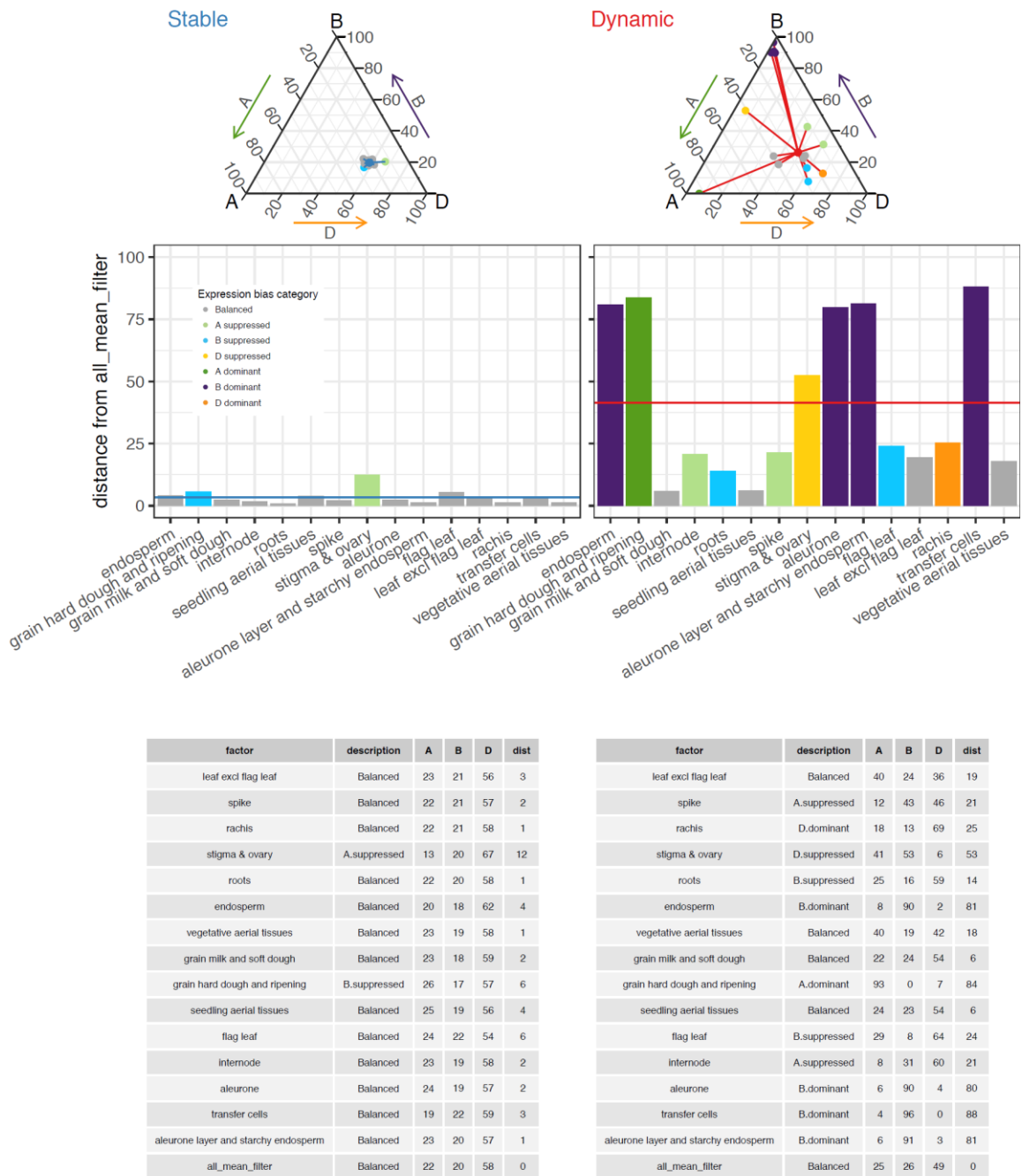


Fig. S19. Variation in triad category assignment across the 15 tissues. Top panel shows the stable and dynamic triad example as in Fig. 3C-D. In the ternary plots, each dot represents an individual tissue in which the triad is expressed and the blue (stable) and red (dynamic) lines illustrate the distance between the 15 tissues and the global average (all_means_filter). Actual values for the ternary plots are indicated in the bottom tables. The distance between each tissue and the global average (all_means_filter) are plotted as bar graphs (center panels) and are also shown as “dist” in the tables. The triad mean distance value which was used to define stable and dynamic triads is indicated by the blue and red horizontal lines in each bar graph. Bars are colored according to their category assignment in each intermediate tissue.

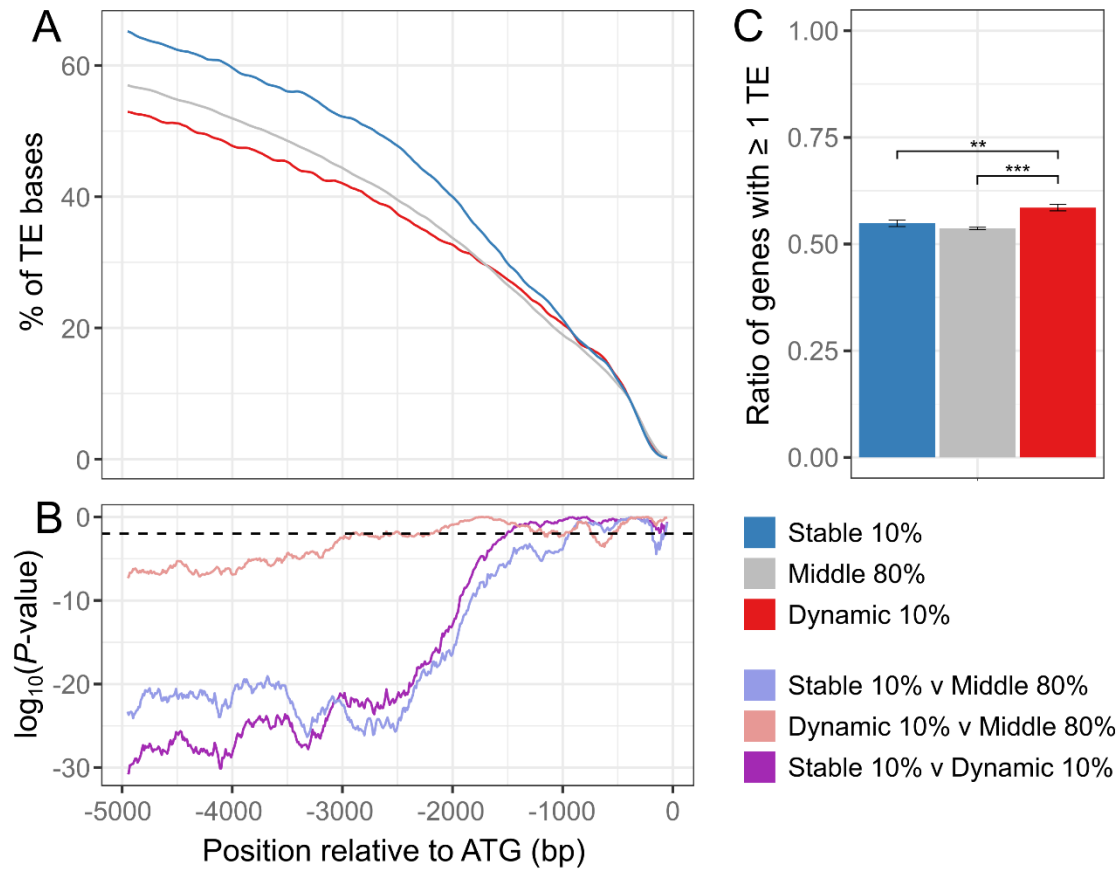


Fig. S20. Transposable element (TE) density and presence across stable, middle and dynamic triads. Triads are classified based on their homoeolog expression bias variation categories into the top 10% most dynamic triads, low 10% most stable triads, and the middle 80% of triads. (A) TE density across the 5 kb promoter region in 100 bp sliding windows is shown for each category. (B) The $\log_{10}(P\text{-value})$ of pairwise Mann-Whitney comparisons between triads shown in panel A. Significantly different comparisons are those below the threshold of -2 (i.e. $P < 0.01$; dashed line). (C) The ratio of genes in each category containing at least one TE within 1.5 kb relative to all genes in the category. Error bars are the standard error; ** Mann-Whitney $P < 0.01$, *** Mann-Whitney $P < 0.001$.

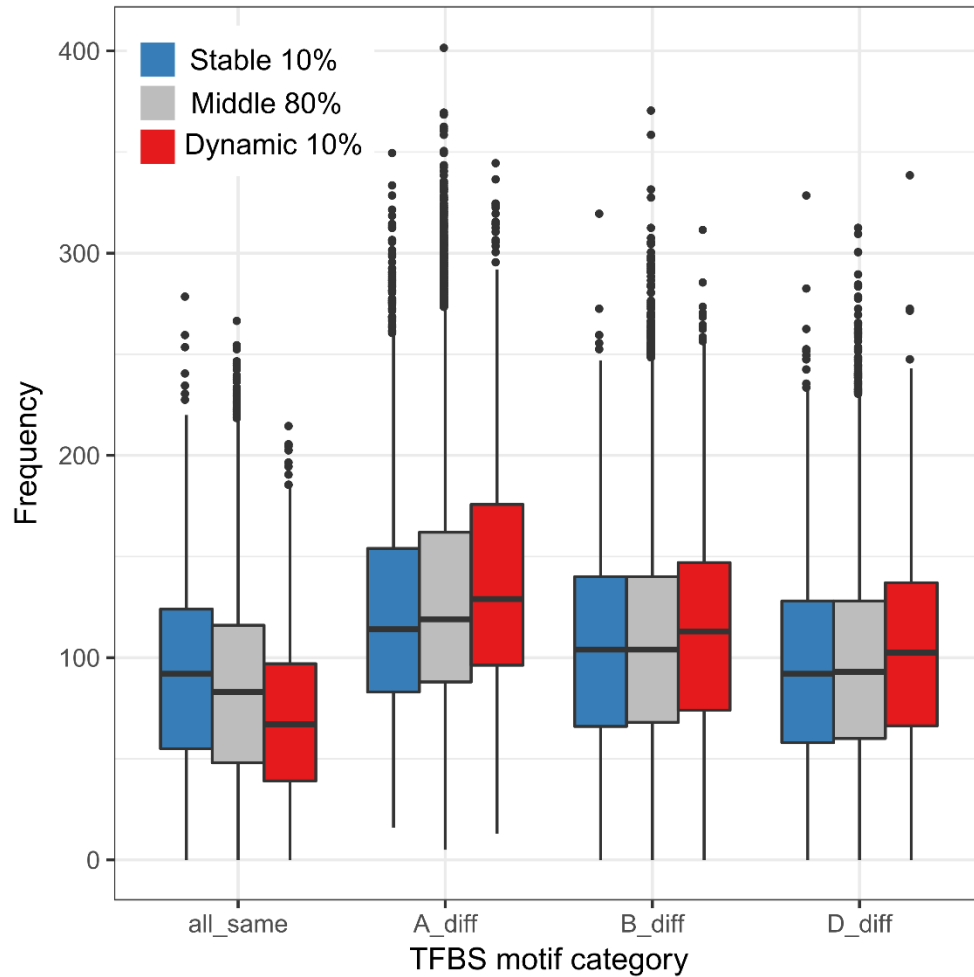


Fig. S21. Transcription factor binding site (TFBS) motif categories across stable, middle and dynamic triads. For each triad, each TFBS motif was assigned to one of four categories based on its presence/absence in the 1.5 kb upstream sequence of each homoeolog: “all_same” = motif present in all three homoeologs; “A_diff” = motif only present in the A homoeolog OR only present in B and D homoeologs; “B_diff” = motif only present in the B homoeolog OR only present in A and D homoeologs; “D_diff” = motif only present in the D homoeolog OR only present in A and B homoeologs. Triads were then grouped according to the stable 10%, middle 80% and dynamic 10% classification. Dynamic triads have significantly fewer motifs shared between all three homoeologs (“all_same”) and significantly more motifs that differed between homoeologs. Within each TFBS motif category, all comparisons were significant (Mann Whitney $P < 0.001$), excluding Stable 10% - Middle 80% in the “B_diff” and “D_diff” categories.

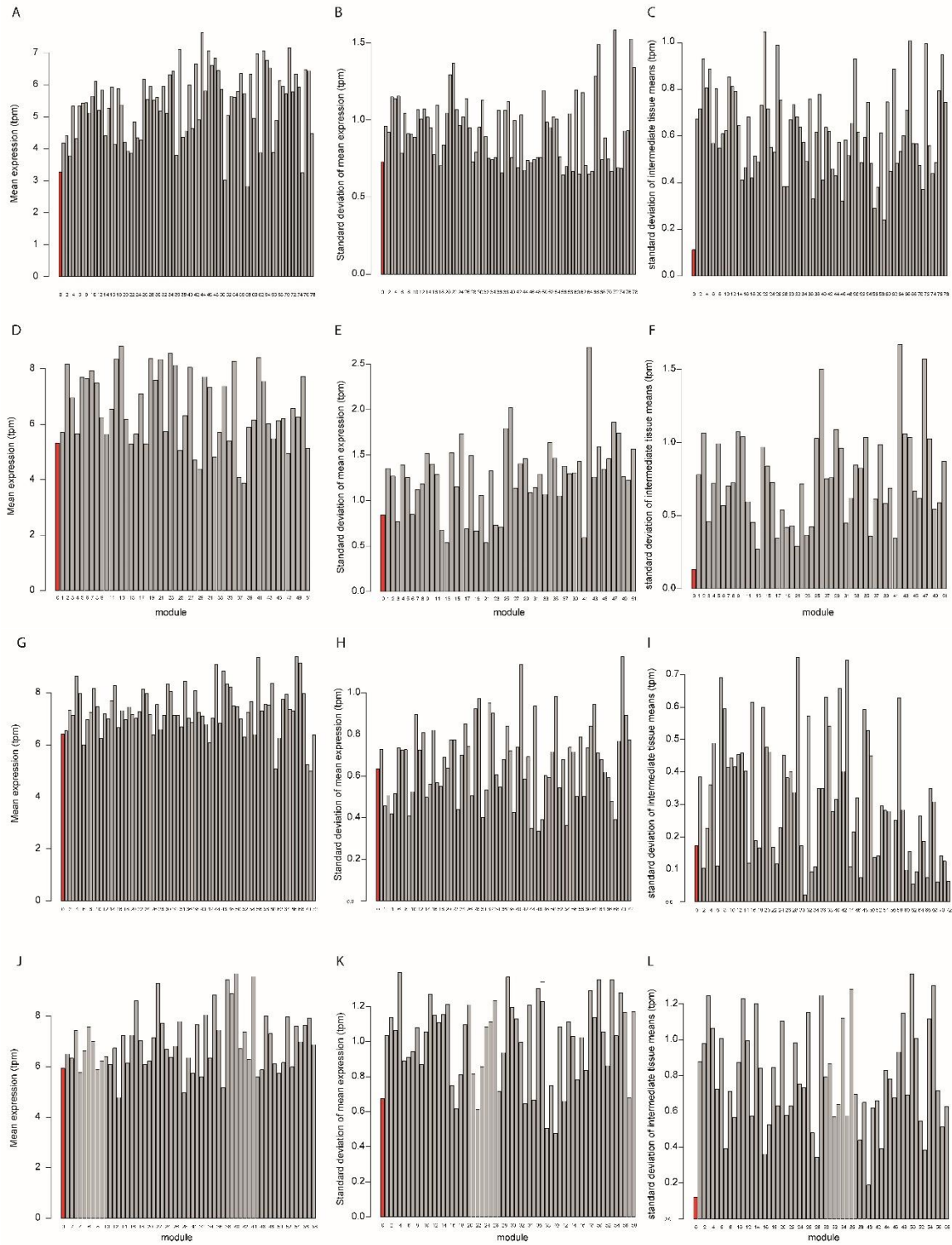


Fig. S22. Features of module 0 in tissue networks. Mean expression (A, D, G, J), standard deviation of mean expression (B, E, H, K), and standard deviation of intermediate tissue means (C, F, I, L) for grain (A-C), leaf (D-F), root (G-I), and spike (J-L) specific networks.

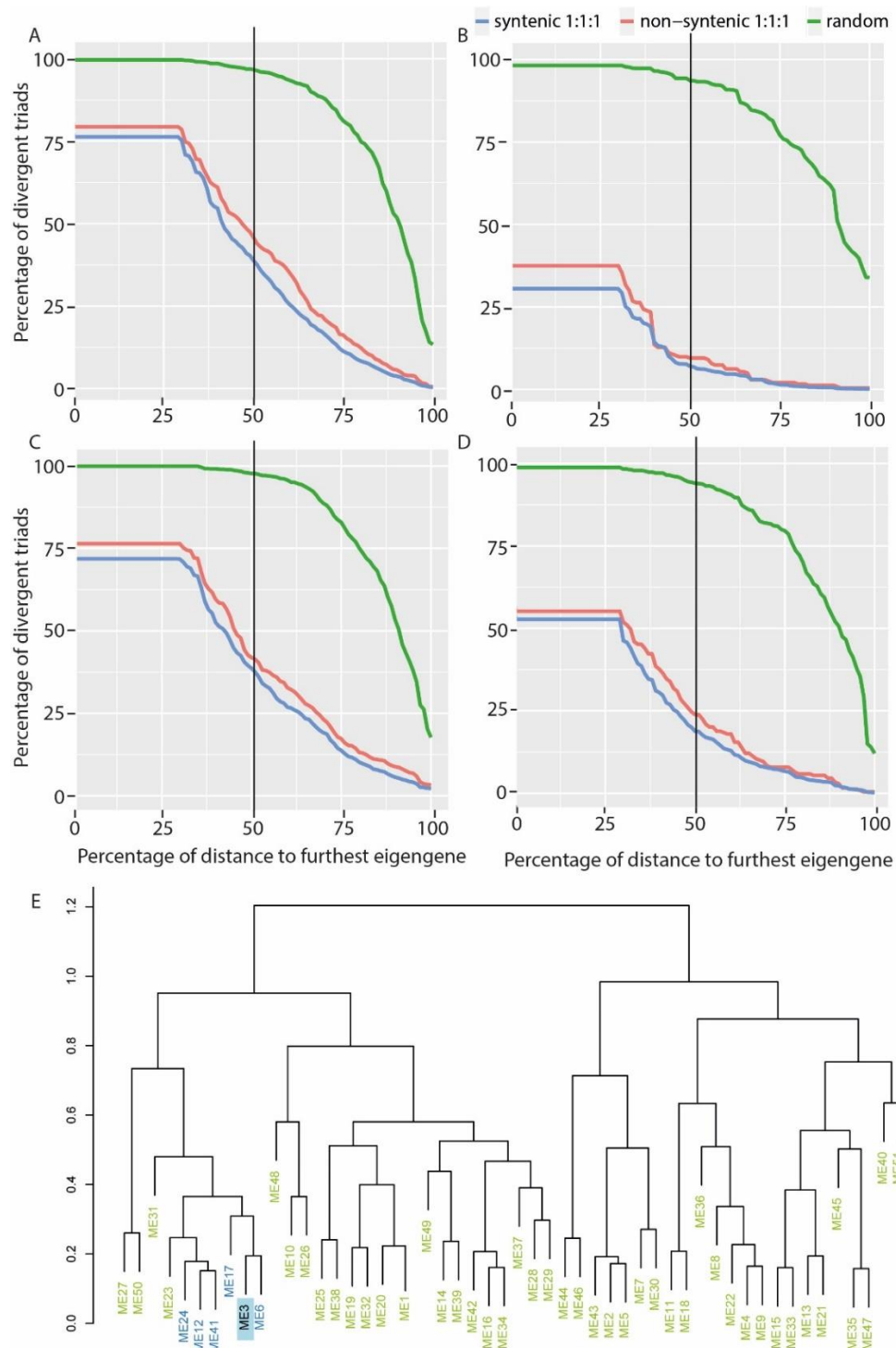


Fig. S23. Determining homoeolog co-expression in networks. Percentage of divergent triads in tissue networks across different thresholds for A) grain, B) leaf, C) root, D) spike. Black line at 50% indicates threshold discussed in main text. E) Dendrogram of eigengene (ME) relatedness in leaf network illustrating an example in relation to module eigengene 2 (ME2, pale blue). Modules which are similar (below 50% threshold) are shown in dark blue, those which are divergent (over 50% threshold) are shown in green.

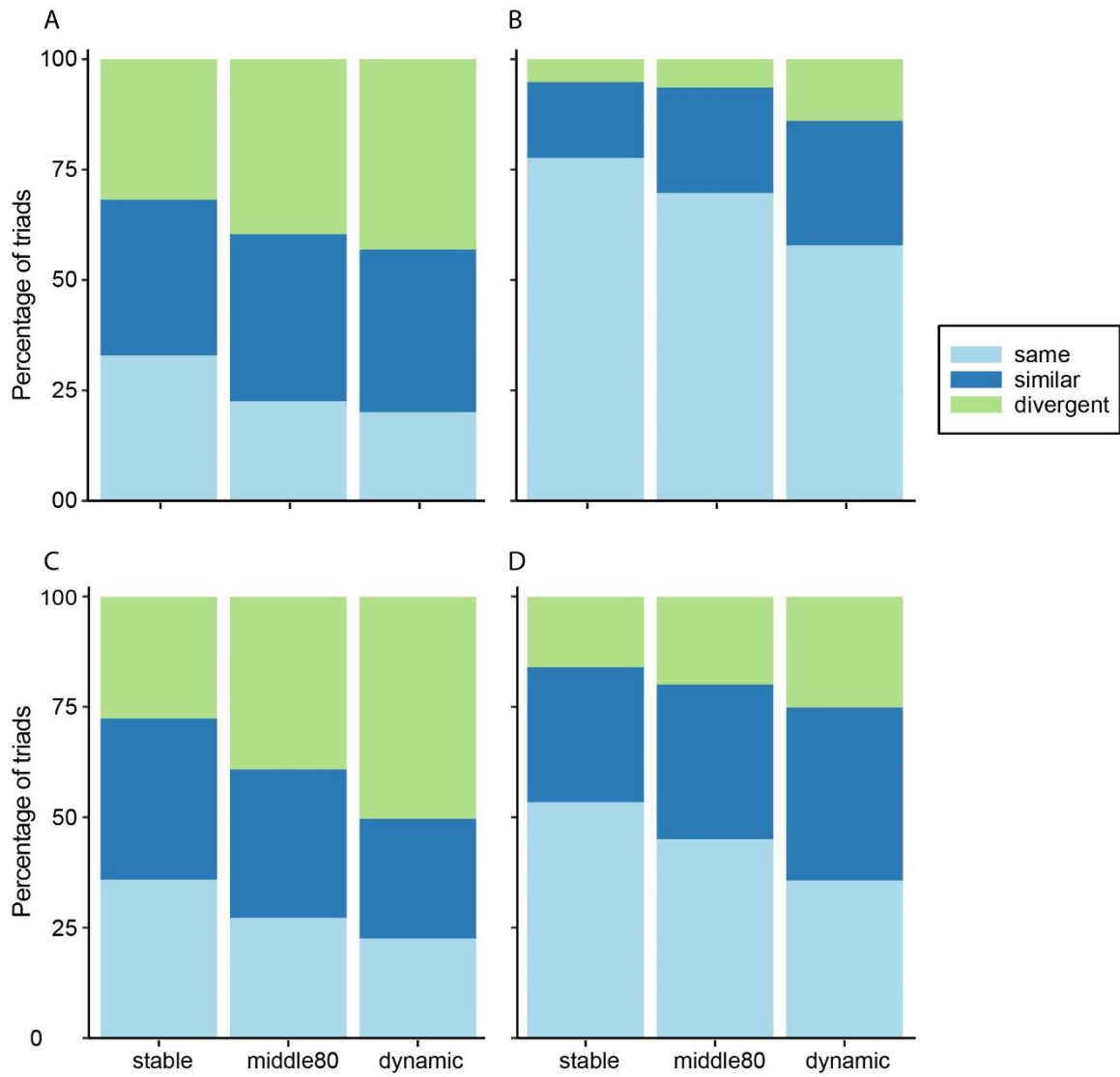


Fig. S24. Stable 10%, middle 80% and dynamic 10% triad assignment to same, similar, and divergent modules in the grain (A), leaf (B), root (C), and spike (D) networks. The root panel (C) is presented in the main text Figure 4B.

Table S1. Metadata for 850 RNA-Seq samples. Details for each sample including variety, tissue, age, stress conditions and original publication. Subsets of samples used for each analysis are indicated by the final eleven columns.

(included in separate Excel file)

Table S2. RNA-Seq complexity in the developmental time course. Expression complexity was calculated for each of the 22 intermediate tissues from Azhurnaya. Values corresponds to the number of genes from which a given percent of transcripts is derived.

(included in separate Excel file)

Table S3. List of 1:1:1 High Confidence syntenic and non-syntenic triads. Includes both expressed (TRUE) and non-expressed (FALSE) triads based on 123 Chinese Spring no stress samples.

(included in separate Excel file)

Table S4. Percentage of triads expressed over min0.5-10tpm cut-offs. The percentage of triads expressed over each min_tpm cut-off are indicated for single tissues and across all tissues in the combined analysis (100% = 16,746 triads/50,238 genes in Chinese Spring and 16,844 triads/50,532 genes in Azhurnaya).

Dataset	Tissue	min0.5tpm	min1tpm	min5tpm	min10tpm
Chinese Spring no stress	aleurone	64.8%	59.8%	42.0%	30.5%
	aleurone layer and starchy endosperm	61.0%	54.6%	32.1%	20.3%
	endosperm	59.8%	52.1%	27.1%	15.9%
	flag leaf	76.6%	71.4%	54.3%	41.9%
	grain hard dough and ripening	69.2%	62.8%	43.8%	32.9%
	grain milk and soft dough	81.6%	75.6%	52.5%	37.5%
	internode	85.7%	80.9%	63.1%	50.6%
	leaf excl flag leaf	82.5%	76.7%	56.5%	42.7%
	rachis	82.2%	77.0%	59.6%	47.4%
	roots	88.3%	83.8%	65.4%	51.4%
	seedling aerial tissues	83.7%	77.6%	57.1%	43.5%
	spike	89.7%	85.1%	66.0%	52.4%
	stigma & ovary	75.3%	69.5%	50.4%	38.0%
	transfer cells	63.5%	55.6%	30.8%	19.1%
	vegetative aerial tissues	81.6%	75.9%	54.8%	40.3%
	<i>Average single tissues</i>	76.4%	70.6%	50.4%	37.6%
	<i>Combined analysis (all tissues)</i>	100.0%	97.9%	85.5%	74.6%
Azhurnaya Development	anther	76.4%	69.2%	46.7%	34.8%
	awns	76.9%	71.3%	54.2%	41.6%
	embryo	73.0%	67.2%	48.6%	37.7%
	endosperm	68.6%	61.4%	39.2%	26.3%
	flag leaf blade	75.7%	69.8%	52.1%	39.8%
	flag leaf sheath	76.0%	70.5%	53.6%	41.9%
	glumes	79.9%	74.5%	57.9%	46.3%
	grain hard dough and ripening	78.0%	71.1%	48.8%	36.1%
	grain milk and soft dough	80.0%	73.9%	53.1%	39.8%
	internode	79.6%	74.2%	58.2%	46.7%
	leaf blades excl flag	77.5%	71.3%	51.8%	39.3%
	leaf ligule	78.4%	73.3%	56.5%	44.4%
	leaf sheaths excl flag	85.3%	80.8%	62.7%	49.6%
	peduncle	79.0%	73.6%	55.8%	43.6%
	root apical meristem	84.5%	79.2%	59.1%	45.7%
	roots	88.6%	84.0%	64.9%	51.0%
	seedling aerial tissues	83.8%	78.5%	59.8%	45.8%
	shoot apical meristem	85.1%	80.4%	60.5%	47.4%
	shoot axis	86.0%	81.0%	61.1%	48.0%
	spike	86.9%	82.1%	63.4%	51.2%
	spikelets	86.5%	81.6%	63.0%	50.2%
	stigma & ovary	81.9%	77.0%	59.8%	49.2%
	<i>Average single tissues</i>	80.3%	74.8%	56.0%	43.5%
	<i>Combined analysis (all tissues)</i>	100.0%	98.5%	87.9%	78.1%

Table S5. Relative contribution of the three genomes to the overall transcript abundance of triads across min_tpm cut-offs. n is the number of genes in each tissue or in the combined analysis. P values for Kruskal-Wallis (K-W) Tukey multiple comparison tests are shown below each min_tpm category. The significantly highest genome in each tissue is highlighted in orange.

(included in separate Excel file)

Table S6. Percentage of syntenic triads assigned to the seven homoeolog expression bias categories. Four different min_tpm cut-offs are shown, across 15 (Chinese Spring no stress) or 22 (Azhurnaya Developmental time course) tissues. The average corresponds to the mean of the 15 or 22 tissues, whereas the combined value corresponds to assignment of each triad when all tissues are combined. Kruskal-Wallis P values for comparisons among genomes are presented in Table S7. Orange highlight shows that D genome is always least suppressed.

(included in separate Excel file)

Table S7. Statistics comparing syntenic triad homoeolog expression bias categories. *P* values of Kruskal-Wallis (K-W) analysis of variance, followed by Tukey multiple comparison test when significant. For the suppression category, the relative frequency of triads suppressed among the A, B and D genome (A suppressed, B suppressed, D suppressed; Table S6) was compared within each min_tpm value. Likewise, for the dominant category, the relative frequency of dominant triads (A dominant, B dominant, D dominant) was compared at each min_tpm.

		Dominance				Suppression			
		K-W	Tukey			K-W	Tukey		
			A vs D	A vs B	B vs D		A vs D	A vs B	B vs D
Chinese Spring no stress	min0.5tpm	0.880	-	-	-	0.001	0.002	0.843	0.012
	min1tpm	0.903	-	-	-	<0.001	<0.001	0.814	0.006
	min5tpm	0.782	-	-	-	<0.001	0.001	0.970	0.003
	min10tpm	0.775	-	-	-	0.001	0.004	0.993	0.005
Azhurnaya Development	min0.5tpm	0.047	0.750	0.210	0.042	<0.001	<0.001	<0.001	<0.001
	min1tpm	0.038	0.748	0.182	0.034	<0.001	<0.001	<0.001	<0.001
	min5tpm	0.203	-	-	-	<0.001	<0.001	<0.001	<0.001
	min10tpm	0.162	-	-	-	<0.001	<0.001	<0.001	<0.001

Table S8. Percentage of non-syntenic triads assigned to the seven homoeolog expression bias categories. The min0.5tpm cut-off was applied across 15 (Chinese Spring no stress) or 22 (Azhurnaya Developmental time course) tissues. The average corresponds to the mean of the 15 or 22 tissues, whereas the combined value corresponds to assignment of each triad when all tissues are combined. Orange highlight shows that D genome is always least suppressed. For the χ^2 test, the values for the seven categories in the non-syntenic triads (observed) were compared against the syntenic triads (expected; Table S6).

(included in separate Excel file)

Table S9. Characteristics of syntenic triads based on homoeolog expression bias assignment. Characteristics were calculated for the Chinese Spring no stress dataset (15 tissues) and the Azhurnaya Developmental time course (22 tissues).

(included in separate Excel file)

Table S10. Statistics comparing genome-specific expression levels in homoeolog expression bias categories. FDR-corrected P values for two-sample Kolmogorov–Smirnov test to determine if the dominant and suppressed TPM abundance distributions for each genome are lower than the TPM abundance distribution of the same genome in the balanced category (i.e. A.dominant compared to A.balanced; B.dominant compared to B.balanced, etc).

(included in separate Excel file)

Table S11. Triads in observed and expected homoeolog expression bias categories between modern-day wheat and synthetic hexaploid wheat. Triads were classified as either meeting the expectation (“As expected”) or not (“Different category”) between the modern-day wheat accessions Chinese Spring/Azhurnaya and the combined SHW datasets. Both absolute values of triads and the percentages are shown for Chinese Spring and Azhurnaya with respect to the two SHW datasets. Triads assigned to the Non-Balanced categories are also summarized based on the Dominant and Suppressed category (independent of genome) and the genome bias (i.e. A corresponds to both A-Dominant and A-Suppressed triads). Bottom section shows the average of both comparisons.

	Observed Chinese Spring			Non-Balanced				
	Total	Balanced	Non-Balanced	Dominant	Suppressed	A	B	D
As expected	6524	5711	813	170	644	276	320	218
Different Category	3554	1735	1820	427	1393	656	655	509
As expected	64.7%	76.7%	30.9%	28.4%	31.6%	29.6%	32.8%	29.9%
Different Category	35.3%	23.3%	69.1%	71.6%	68.4%	70.4%	67.2%	70.1%

	Observed Azhurnaya			Non-Balanced				
	Total	Balanced	Non-Balanced	Dominant	Suppressed	A	B	D
As expected	6783	6055	729	121	608	226	352	151
Different Category	3381	2005	1376	225	1151	454	568	355
As expected	66.7%	75.1%	34.6%	34.9%	34.6%	33.2%	38.3%	29.8%
Different Category	33.3%	24.9%	65.4%	65.1%	65.4%	66.8%	61.7%	70.2%

	Observed modern-day 6x			Non-Balanced				
	Total	Balanced	Non-Balanced	Dominant	Suppressed	A	B	D
As expected	6653	5883	771	145	626	251	336	184
Different Category	3468	1870	1598	326	1272	555	611	432
As expected	65.7%	75.9%	32.5%	30.8%	33.0%	31.1%	35.5%	29.9%
Different Category	34.3%	24.1%	67.5%	69.2%	67.0%	68.9%	64.5%	70.1%

Table S12. Triads in observed and expected homoeolog expression bias categories following polyploidization. Triads were classified as either meeting the expectation (“As expected”) or not (“Different category”) between the observed synthetic hexaploid wheat (SHW) data and the expected *in-silico* dataset. Both absolute values of triads and the percentages are shown for SHW1 (AS2255 x AS60) and SHW2 (Langdon x AS60) with respect to their corresponding *in-silico* datasets. Triads assigned to the Non-Balanced categories are also summarized based on the Dominant and Suppressed category (independent of genome) and the genome bias (i.e. A corresponds to both A-Dominant and A-Suppressed triads). Bottom section shows the average of both experiments.

	Observed SHW1 (AS2255 x AS60)			Non-Balanced				
	Total	Balanced	Non-Balanced	Dominant	Suppressed	A	B	D
As expected	7775	6018	1757	358	1399	601	727	429
Different Category	2233	737	1496	271	1225	316	406	774
As expected	77.7%	89.1%	54.0%	57.0%	53.3%	65.5%	64.2%	35.7%
Different Category	22.3%	10.9%	46.0%	43.0%	46.7%	34.5%	35.8%	64.3%

	Observed SHW2 (Langdon x AS60)			Non-Balanced				
	Total	Balanced	Non-Balanced	Dominant	Suppressed	A	B	D
As expected	7606	5840	1766	386	1381	576	685	506
Different Category	2629	987	1642	309	1334	424	491	728
As expected	74.3%	85.5%	51.8%	55.5%	50.9%	57.6%	58.2%	41.0%
Different Category	25.7%	14.5%	48.2%	44.5%	49.1%	42.4%	41.8%	59.0%

	Observed SHW (BBAA x AS60)			Non-Balanced				
	Total	Balanced	Non-Balanced	Dominant	Suppressed	A	B	D
As expected	7690	5929	1761	372	1390	588	706	467
Different Category	2431	862	1569	290	1279	370	448	751
As expected	76.0%	87.3%	52.9%	56.2%	52.1%	61.4%	61.2%	38.4%
Different Category	24.0%	12.7%	47.1%	43.8%	47.9%	38.6%	38.8%	61.6%

Table S13. Characterization of transposable elements within triad promoters. Transposable elements (TE) were identified and triads classified according to the homoeolog expression bias category across combined “all tissues” (top) and on their variation across 15 tissues (bottom). Promoter lengths were defined as 1.5 kb and 5 kb upstream of the ATG start-site. χ^2 tests were carried out for the number of genes and number of triads with at least one TE; Kruskal-Wallis tests were carried out for the median size and distance of TEs. Significant *P* values are highlighted in green.

Category	Number of Genes	Distance to nearest TE (bp)	1.5 kb			5 kb		
			Genes with ≥ 1 TE (%)	Triads with ≥ 1 TE (%)	Median Size of TE (bp)	Genes with ≥ 1 TE (%)	Triads with ≥ 1 TE (%)	Median Size of TE (bp)
Balanced	35502	1243	53.5%	80.6%	233	93.1%	98.9%	304
Non-Dominant	858	1003	59.9%	88.8%	234	92.9%	98.8%	266
Dominant	429	958	59.4%	88.8%	211	93.5%	98.8%	282
Suppressed	1995	1025	58.1%	86.4%	235	93.3%	99.2%	287
Non-Suppressed	3990	1122	57.2%	86.4%	226	92.9%	99.2%	275
<i>P</i> value		< 2.2E-16	1.1E-09	< 2.2E-16	0.37	0.90	0.16	1.5E-13

Category	Number of Genes	Distance to nearest TE (bp)	1.5 kb			5 kb		
			Genes with ≥ 1 TE (%)	Triads with ≥ 1 TE (%)	Median Size of TE (bp)	Genes with ≥ 1 TE (%)	Triads with ≥ 1 TE (%)	Median Size of TE (bp)
Dynamic 10%	4275	1113	58.5%	88.3%	220	92.8%	99.6%	272
Middle 80%	34221	1242	53.7%	81.1%	230	91.9%	98.9%	302
Stable 10%	4278	1234	54.9%	79.2%	259	93.2%	98.8%	429
<i>P</i> value		3.0E-06	1.1E-08	< 2.2E-16	1.5E-10	1.9E-03	9.3E-06	< 2.2E-16

Table S14. Relative chromosome position of balanced, dominant and suppressed triads. The Chinese Spring no stress and Azhurnaya Developmental time course were analyzed separately. Observed and expected values are shown as percentages, whereas χ^2 tests were performed on the absolute expected and observed values.

(included in separate Excel file)

Table S15. Inter-cultivar comparison of homoeolog expression bias based on genomic compartment. Percentage of genes which remain in the same homoeolog expression bias category between Chinese Spring and Azhurnaya cultivars based on their genomic compartment. The combined analysis for all tissues was compared as well as the nine common tissues between the two datasets to account for any tissue bias in the combined analysis. *P* values for χ^2 test comparing R1/R3 vs R2/C.

Tissues	Genomic compartment					Average		<i>P</i> value
	R1	R2A	C	R2B	R3	R1/R3	R2/C	
endosperm	65.6%	77.4%	75.9%	76.8%	68.3%	67.6%	76.9%	5.0E-19
flag leaf	68.1%	78.7%	81.3%	79.0%	66.6%	67.0%	79.2%	2.0E-38
grain hard dough and ripening	61.3%	72.8%	72.4%	71.2%	63.4%	62.9%	71.8%	1.4E-21
grain milk and soft dough	68.0%	79.6%	80.9%	79.6%	69.2%	68.9%	79.7%	1.7E-32
internode	64.4%	75.9%	77.6%	76.8%	67.0%	66.3%	76.6%	3.2E-30
roots	71.7%	84.9%	85.5%	84.7%	73.0%	72.6%	84.8%	5.5E-43
seedling aerial tissues	71.5%	82.9%	83.9%	84.1%	72.6%	72.3%	83.7%	2.9E-35
spike	71.7%	81.8%	81.9%	83.0%	72.0%	71.9%	82.5%	1.4E-32
stigma & ovary	58.8%	65.9%	67.7%	67.0%	58.9%	58.9%	66.7%	2.3E-19
Combined global analysis	71.0%	84.0%	84.6%	84.8%	72.7%	72.2%	84.5%	1.1E-50

Table S16. Comparison of homoeolog expression bias classification across tissues. Comparisons were made between individual tissues and the global analysis homoeolog expression bias category for Chinese Spring no stress and Azhurnaya developmental time course. See Fig. S18 for visualization of adjacent and opposite categories.

Dataset	Tissue types	Movement across tissues	Homoeolog Expression Bias Category (combined global analysis)		
			Balanced	Dominant	Suppressed
Chinese Spring no stress	15	Invariable	83.6%	73.4%	62.2%
		Variable	16.4%	26.6%	37.8%
		adjacent ¹	13.3%	24.9%	35.0%
		opposite ²	3.1%	1.6%	2.9%
Azhurnaya Development	22	Invariable	84.8%	78.4%	69.8%
		Variable	15.2%	21.6%	30.2%
		adjacent ¹	12.3%	20.5%	28.0%
		opposite ²	2.9%	1.1%	2.2%

¹ For Balanced triads, suppressed categories were considered adjacent as they only require variation in a single homoeolog for this classification

² For Balanced triads, dominant categories were considered opposite as they usually require variation in two homoeologs for this classification

Table S17. Characteristics of stable and dynamic triads. Stable (Low10, Low25) and dynamic (Top25, Top10) triads are shown for Chinese Spring no stress and Azhurnaya Developmental time course. *P* values corresponds to Mann Whitney test.

(included in separate Excel file)

Table S18. GO slim enrichment of stable and dynamic triads. GO slim enrichment is shown for stable (Low10, Low25) and dynamic (Top25, Top10) triads for Chinese Spring no stress and Azhurnaya developmental time course. GO terms in red are those which are unique to a single dataset, whereas all other GO terms are common between datasets. Only enrichments with *P* value significance below E-10 are shown.

(included in separate Excel file)

Table S19. Homoeolog expression bias category assignment for stable and dynamic triads. Stable (Low10, Low25) and dynamic (Top25, Top10) triads are shown for Chinese Spring no stress and Azhurnaya developmental time course compared to expected values across all triads in each dataset. *P* values are reported for χ^2 tests between balanced:dominant:suppressed categories between observed and expected.

(included in separate Excel file)

Table S20. Relative chromosome position of stable, middle and dynamic triads. Two different sets of stable, middle and dynamic triads are shown: Top10, Middle80, Low10 (10-80-10) and Top25, Middle50, Low25 (25-50-25) based on the mean distance across tissues. Chinese Spring no stress and Azhurnaya Developmental time course were analyzed separately. Observed and expected values are shown in percentages, whereas χ^2 tests were performed on the absolute expected and observed values.

(included in separate Excel file)

Table S21. Coding sequence and promoter conservation between stable, middle and dynamic triads. Coding sequence (nucleotide/protein identity and protein similarity) and promoter (1.5 kb) conservation between triads based on most stable, middle and dynamic triads as defined by the mean distance across tissues. The analysis was done at three cutoffs (5, 10, and 25% of the distribution) which defined the middle 90, 80, or 50%, respectively. Values are based on the average of the three pairwise comparisons between homoeologs (A to B, A to D, and B to D). Only promoters without N's in the 1.5 kb 5' upstream and alignments over 200 bp were considered for the analysis. *P* values in bottom section are based on Kruskal-Wallis ANOVA on ranks, followed by Dunn's test for the corresponding comparison. For promoter id, Mann Whitney tests were also performed for the Stable vs dynamic comparison at 10-80-10.

(included in separate Excel file)

Table S22. Ka/Ks ratio values for stable, middle and dynamic triads and syntenic/non-syntenic triads. Values shown are the average Ka/Ks ratios of the three pairwise comparisons between homoeologs (A to B, A to D, and B to D). *P* values of the difference in Ka/Ks ratios between the subsets is also shown, from the Mann Whitney test.

		Chinese Spring no stress			Azhurnaya Development		
Category	Subset	Mean ($\pm$ SE)	Median	N	Mean ($\pm$ SE)	Median	N
5-90-5	Stable 5%	0.23 $\pm$ 0.016	0.16	2136	0.23 $\pm$ 0.016	0.16	2253
	Middle 80%	0.26 $\pm$ 0.0029	0.20	38499	0.26 $\pm$ 0.011	0.20	40563
	Dynamic 5%	0.32 $\pm$ 0.011	0.26	2139	0.32 $\pm$ 0.0029	0.26	2256
10-80-10	Stable 10%	0.21 $\pm$ 0.0094	0.15	4275	0.21 $\pm$ 0.0094	0.15	4506
	Middle 80%	0.26 $\pm$ 0.0031	0.20	34221	0.26 $\pm$ 0.0030	0.20	36057
	Dynamic 10%	0.33 $\pm$ 0.011	0.26	4278	0.33 $\pm$ 0.010	0.26	4509
25-50-25	Stable 25%	0.22 $\pm$ 0.0055	0.16	10692	0.22 $\pm$ 0.0052	0.16	11265
	Middle 50%	0.26 $\pm$ 0.0042	0.20	21387	0.26 $\pm$ 0.0039	0.20	22539
	Dynamic 25%	0.30 $\pm$ 0.0057	0.25	10695	0.31 $\pm$ 0.0062	0.25	11268
Non-Syntenic	All	0.39 $\pm$ 0.014	0.32	2490	0.39 $\pm$ 0.013	0.32	2706
		Chinese Spring no stress			Azhurnaya Development		
Category	Comparison	5-90-5	10-80-10	25-50-25	5-90-5	10-80-10	25-50-25
Triad Movement	Stable vs Dynamic	<2.2E-16	<2.2E-16	<2.2E-16	<2.2E-16	<2.2E-16	<2.2E-16
	Stable vs Middle	1.03E-12	<2.2E-16	<2.2E-16	7.11E-16	<2.2E-16	<2.2E-16
	Dynamic vs Middle	<2.2E-16	<2.2E-16	<2.2E-16	<2.2E-16	<2.2E-16	<2.2E-16
Syteny	Comparison to Syntenic	Stable 10%	Middle 90%	Dynamic 10%	Stable 10%	Middle 90%	Dynamic 10%
	Non-Syntenic	<2.2E-16	<2.2E-16	7.95E-07	<2.2E-16	<2.2E-16	1.38E-07

Table S23. Features of four tissue specific WGCNA networks

Tissue	Number of genes expressed	Percentage of genes assigned to modules	Number of modules
Grain	72,370	77.9	78
Leaf	81,025	42.3	51
Root	73,232	88.0	72
Spike	84,699	53.2	58

Table S24. Triad co-expression based on module assignment of homoeologs across tissue networks.

	Network	Category (count)			Category (%)		
		same	similar	divergent	same	similar	divergent
Syntenic Triads	grain	2160	3417	3542	23.7%	37.5%	38.8%
	leaf	3568	1208	359	69.5%	23.5%	7.0%
	root	2948	3482	4052	28.1%	33.2%	38.7%
	spike	2803	2016	1115	47.2%	34.0%	18.8%
	<i>Weighted average across networks</i>				37.4%	33.0%	29.6%
Non-syntenic Triads	grain	99	161	220	20.6%	33.5%	45.8%
	leaf	152	68	23	62.6%	28.0%	9.5%
	root	131	192	233	23.6%	34.5%	41.9%
	spike	107	75	57	44.8%	31.4%	23.8%
	<i>Weighted average across networks</i>				32.2%	32.7%	35.1%
Random	grain	3	30	967	0.3%	3.0%	96.7%
	leaf	18	46	936	1.8%	4.6%	93.6%
	root	1	22	977	0.1%	2.2%	97.7%
	spike	12	48	940	1.2%	4.8%	94.0%
	<i>Weighted average across networks</i>				0.9%	3.7%	95.5%

Table S25. Plant Ontology (PO) terms enriched in root module 61.

PO term	description	P value
PO:0004545	shoot-borne shoot system	1.87E-10
PO:0006307	root procambium	2.27E-08
PO:0003021	central root cap of primary root	4.69E-08
PO:0006081	primary root apical meristem	2.59E-06
PO:0000026	primary root tip	6.69E-06
PO:0020123	root cap	7.16E-06
PO:0025181	root elongation zone	1.75E-05
PO:0005059	root endodermis	2.04E-05
PO:0006036	root epidermis	2.68E-05
PO:0020124	root stele	4.09E-05
PO:0006504	leaf trichome	8.46E-05

Table S26. TFs in root module 61. Orthologs involved in root development are noted.
(included in separate Excel file)

Table S27. GO term enrichment of target genes of MADS_II TFs in root module 61. *P* values were calculated using the classic Fisher's test in topGO.

GO.ID	Term	<i>P</i> value for target genes of		
		TraesCS2A01G337900	TraesCS2B01G344000	TraesCS2D01G325000
GO:0071554	cell wall organization or biogenesis	1.3E-30	1.8E-20	< 1e-30
GO:0071555	cell wall organization	3.5E-28	9.5E-21	< 1e-30
GO:0009808	lignin metabolic process	6E-27	3.8E-22	< 1e-30
GO:0009664	plant-type cell wall organization	4.6E-25	1.9E-18	< 1e-30
GO:0071669	plant-type cell wall organization or biogenesis	3.4E-23	4.9E-14	< 1e-30
GO:0009809	lignin biosynthetic process	3.3E-19	9E-20	2.8E-29

Table S28. Top three enriched GO terms of TF targets from genie3 network. GO enrichment was calculated using topGO classic Fisher's test.
(included in separate Excel file)

Table S29. Correlation between abiotic network modules and abiotic stresses. *P* values are the Student asymptotic *P* value of the correlations corrected for multiple testing using the Benjamini & Yekutieli method.
(included in separate Excel file)

Table S30. Correlation between disease network modules and disease stresses. *P* values are the Student asymptotic *P* value of the correlations corrected for multiple testing using the Benjamini & Yekutieli method.
(included in separate Excel file)

Table S31. Genes associated with abiotic and/or disease stress and their modules. *P* values are the Student asymptotic *P* value of the correlations corrected for multiple testing using the Benjamini & Yekutieli method.
(included in separate Excel file)

Table S32. GO slim enrichment of genes in disease module 12 and abiotic module 2.

ontology	category	description	P value disease module12	P value abiotic module2
Biological Process	GO:0009607	response to biotic stimulus	2.2E-63	4.0E-06
	GO:0007165	signal transduction	1.0E-57	8.3E-08
	GO:0008219	cell death	9.7E-57	9.0E-04
	GO:0007154	cell communication	2.5E-55	2.2E-06
	GO:0009605	response to external stimulus	1.6E-51	6.3E-05
	GO:0006950	response to stress	6.8E-42	1.2E-04
	GO:0009719	response to endogenous stimulus	1.4E-38	-
		cellular protein modification		
	GO:0006464	process	1.9E-36	-
	GO:0006810	transport	2.2E-21	-
	GO:0019538	protein metabolic process	7.1E-16	-
	GO:0009991	response to extracellular stimulus	2.2E-15	-
	GO:0009856	pollination	3.0E-14	-
	GO:0009875	pollen-pistil interaction	1.1E-12	-
	GO:0008150	biological_process	8.1E-07	-
	GO:0019748	secondary metabolic process	1.6E-05	-
	GO:0009838	abscission	1.9E-05	-
	GO:0009056	catabolic process	0.0E+00	8.6E-04
Cellular Component	GO:0016020	membrane	4.8E-33	1.1E-04
	GO:0005768	endosome	2.1E-29	4.7E-05
	GO:0005618	cell wall	2.7E-16	-
	GO:0030312	external encapsulating structure	9.0E-12	-
	GO:0005783	endoplasmic reticulum	7.6E-09	-
	GO:0005794	Golgi apparatus	1.5E-08	-
	GO:0005886	plasma membrane	1.8E-03	-
Molecular Function	GO:0004872	receptor activity	3.0E-60	-
	GO:0004871	signal transducer activity	1.1E-56	-
	GO:0016301	kinase activity	4.9E-55	-
	GO:0016740	transferase activity	9.4E-30	-
	GO:0030246	carbohydrate binding	2.4E-20	-
	GO:0000166	nucleotide binding	2.6E-20	-
	GO:0005515	protein binding	8.7E-08	-
	GO:0005215	transporter activity	8.5E-07	-
	GO:0003824	catalytic activity	2.4E-05	-
	GO:0003674	molecular_function	7.9E-05	-
	GO:0005102	receptor binding	1.2E-04	-

Table S33. Sixteen TFs shared in abiotic module 2 and disease module 12. The functions of rice and Arabidopsis orthologs are indicated where known.

(included in separate Excel file)

Table S34. Top 10 hub genes in disease module 12. Calculated as intramodular connectivity.

Gene	Intramodular connectivity (eigengene correlation)	P value	TF family
TraesCS1A01G350400	0.961	0	HSF
TraesCS4D01G305100	0.950	0	NA
TraesCS5D01G226400	0.948	0	NA
TraesCS4A01G106400	0.948	0	NA
TraesCS6B01G342800	0.947	0	NA
TraesCS4B01G190100	0.945	0	NA
TraesCS5A01G237900	0.943	0	HSF
TraesCS2A01G188600	0.943	0	NA
TraesCS5B01G236400	0.941	0	HSF
TraesCS3D01G262000	0.940	0	NA

Table S35. Predicted target genes of HSF TFs. The strength of edge (edge weight) connecting the TF to the target gene is shown, along with the module in which the target gene was allocated in the abiotic and disease networks.

(included in separate Excel file)

Table S36: Expression of a dynamic triad across six tissues and in the combined analysis. The expression of the individual homoeologous genes which constitute the triad are shown, alongside the triad sum, normalized expression and homoeolog expression category. This triad was considered expressed across the six tissues shown since the triad_sum > 0.5 TPM for all six tissues. For individual genes, however, these were considered expressed >0.5 TPM on a gene basis in fewer tissues. For example, the A genome homoeolog TraesCS7A01G524500 was considered expressed only in roots, seedling aerial tissues and internodes (3 tissues) given that its expression was lower than 0.5 TPM in the spike, rachis and vegetative aerial tissues. This *gene* would therefore be considered expressed in three tissues, whereas the *triad* was considered expressed in six tissues.

(included in separate Excel file)

Table S37: Definition of homoeolog expression bias categories. A, B, and D represent the relative expression levels of the A, B, and D genome homoeologs across an individual triad.

Category	A	B	D
Balanced	0.33	0.33	0.33
A suppressed	0	0.5	0.5
B suppressed	0.5	0	0.5
D suppressed	0.5	0.5	0
A dominant	1	0	0
B dominant	0	1	0
D dominant	0	0	1

International Wheat Genome Sequencing Consortium (IWGSC) Collaborator List

Name	Email	Affiliation
Abraham Korol	korol@research.haifa.ac.il	26
Andrew G. Sharpe	andrew.sharpe@gifs.ca	43
Angéla Juhász	A.Juhasz@murdoch.edu.au	36,37
Antje Rohde	antje.rohde@bayer.com	72
Arnaud Bellec	arnaud.bellec@inra.fr	20
Assaf Distelfeld	adistel@tauex.tau.ac.il	25
Bala Ani Akpınar	aniakpinar@gmail.com	14
Beat Keller	bkeller@botinst.uzh.ch	41
Benoit Darrier	benoit.darrier@inra.fr	7
Bikram Gill	bsgill@ksu.edu	28
Boulos Chalhoub	boulos.chalhoub@yahoo.com	62,63
Burkhard Steuernagel	burkhard.steuernagel@jic.ac.uk	10
Catherine Feuillet	feuillet@bayer.com	17
Chanderkant Chaudhary	ckryptone@gmail.com	71
Cristobal Uauy	cristobal.uauy@jic.ac.uk	10
Curtis Pozniak	curtis.pozniak@usask.ca	11
Danara Ormanbekova	danara.ormanbekova2@unibo.it	9,48
Daoquan Xiang	daoquan.xiang@nrc-cnrc.gc.ca	22
David Latrasse	david.latrasse@u-psud.fr	15
David Swarbreck	david.swarbreck@earlham.ac.uk	50
Delfina Barabaschi	delfina.barabaschi@crea.gov.it	16
Dina Raats	dina.raats@earlham.ac.uk	50
Ekaterina Sergeeva	sergeeva@bionet.nsc.ru	47
Elena Salina	salina@bionet.nsc.ru	47
Etienne Paux	etienne.paux@inra.fr	7
Federica Cattonaro	cattonaro@igatechnology.com	58
Frédéric Choulet	frederic.choulet@inra.fr	7
Fuminori Kobayashi	kobafumi@affrc.go.jp	31
Gabriel Keeble-Gagnere	gabriel.keeble-gagnere@ecodev.vic.gov.au	1
Gaganpreet Kaur	gaganchahal@gmail.com	28
Gary Muehlbauer	muehl003@umn.edu	30
George Kettleborough	kettleg@gmail.com	50
Guotai Yu	guotai.yu@jic.ac.uk	10
Hana Šimková	simkovah@ueb.cas.cz	8
Heidrun Gundlach	h.gundlach@helmholtz-muenchen.de	9
Hélène Berges	helene.berges@inra.fr	20
Hélène Rimbart	helene.rimbart@inra.fr	7
Hikmet Budak	hikmet.budak@montana.edu	14

Hirokazu Handa	hirokazu@affrc.go.jp	31
Ian Small	ian.small@uwa.edu.au	45
Jan Bartoš	bartos@ueb.cas.cz	8
Jane Rogers	janerogersh@gmail.com	6
Jaroslav Doležel	dolezel@ueb.cas.cz	8
Jens Keilwagen	jens.keilwagen@julius-kuehn.de	40
Jesse Poland	jpoland@ksu.edu	28
Joanna Melonek	joanna.melonek@uwa.edu.au	45
John Jacobs	j.jacobs@bayer.com	18
Jon Wright	Jon.Wright@earlham.ac.uk	50
Jonathan D. G. Jones	jonathan.jones@sainsbury-laboratory.ac.uk	35
Juan Gutierrez-Gonzalez	jgutier@umn.edu	30
Kellye Eversole	eversole@eversoleassociates.com	2,3
Kirby Nilsen	kirby.nilsen@usask.ca	11
Klaus F.X. Mayer	k.mayer@helmholtz-muenchen.de	9,44
Kostya Kanyuka	kostya.kanyuka@rothamsted.ac.uk	38
Kuldeep Singh	kuldeep35@pau.edu	65
Liangliang Gao	lianggao@ksu.edu	28
Lorenzo Concia	lorenzo.concia@u-psud.fr	15
Luca Venturini	Luca.Venturini@earlham.ac.uk	50
Luigi Cattivelli	luigi.cattivelli@crea.gov.it	16
Manuel Spannagl	manuel.spannagl@helmholtz-muenchen.de	9
Martin Mascher	mascher@ipk-gatersleben.de	4,67
Matthew Hayden	matthew.hayden@ecodev.vic.gov.au	1
Michael Abrouk	abrouk@ueb.cas.cz	8,19
Michael Alaux	michael.alaux@inra.fr	13
Mingcheng Luo	mcluo@ucdavis.edu	34
Miroslav Valárik	valarik@ueb.cas.cz	8
Moussa Benhamed	moussa.benhamed@u-psud.fr	15
Nagendra K. Singh	nksingh4@gmail.com	70
Naveen Sharma	naveenlalosharma@gmail.com	71
Nicolas Guilhot	nicolas.guilhot@inra.fr	7
Nikolai Ravin	nravin@biengi.ac.ru	23, 51
Nils Stein	stein@ipk-gatersleben.de	4,5
Odd-Arne Olsen	odd-arne.olsen@nmbu.no	56
Om Prakash Gupta	opgupta@pau.edu	65
Paramjit Khurana	param@genomeindia.org	71
Parveen Chhuneja	pchhuneja@pau.edu	65
Philipp E. Bayer	philipp.bayer@uwa.edu.au	24
Philippa Borrill	Philippa.Borrill@jic.ac.uk	10

Philippe Leroy	philippe.leroy.2@inra.fr	7
Philippe Rigault	prigault@gydle.com	39
Pierre Sourdille	pierre.sourdille@inra.fr	7
Pilar Hernandez	phernandez@ias.csic.es	33
Raphael Flores	raphael.flores@inra.fr	13
Ricardo H. Ramirez-Gonzalez	Ricardo.Ramirez-Gonzalez@jic.ac.uk	10
Robert King	robert.king@rothamsted.ac.uk	42
Ron Knox	ron.knox@agr.gc.ca	21
Rudi Appels	rudi.appels@unimelb.edu.au	1,36
Ruonan Zhou	zhou@ipk-gatersleben.de	4
Sean Walkowiak	sean.walkowiak@usask.ca	11
Sergio Galvez	galvez@uma.es	27
Sezgi Biyiklioglu	sezgi.biyiklioglu@montana.edu	14
Shuhei Nasuda	nasushu@kais.kyoto-u.ac.jp	46
Simen Sandve	simen.sandve@nmbu.no	57
Smahane Chalabi	smahane.chalabi@gmail.com	63
Song Weining	sweining2002@yahoo.com	66
Sunish Sehgal	sunish.sehgal@sdsstate.edu	53
Suruchi Jindal	suruchi-coasab@pau.edu	65
Tatiana Belova	tatiana.belova@nmbu.no	56
Thomas Letellier	thomas.letellier@inra.fr	13
Thomas Wicker	wicker@botinst.uzh.ch	41
Tsuyoshi Tanaka	tstanaka@affrc.go.jp	31
Tzion Fahima	fahima@research.haifa.ac.il	26
Valérie Barbe	vbarbe@genoscope.cns.fr	61
Vijay Tiwari	vktywari@umd.edu	54
Vinod Kumar	kumar.vinod81@gmail.com	70
Yifang Tan	yifang.tan@nrc-cnrc.gc.ca	22

¹AgriBio, Centre for AgriBioscience, Department of Economic Development, Jobs, Transport and Resources, 5 Ring Rd, La Trobe University, Bundoora, Victoria 3083 Australia.

²International Wheat Genome Sequencing Consortium (IWGSC), 5207 Wyoming Road, Bethesda, Maryland, 20816, United States.

³Eversole Associates, 5207 Wyoming Road, Bethesda, Maryland, 20816, United States.

⁴Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Genebank, Corrensstr. 3, 06466 Stadt Seeland, Germany.

⁵The University of Western Australia (UWA), School of Agriculture and Environment, 35 Stirling Highway, Crawley WA 6009, Australia.

⁷GDEC (Genetics, Diversity and Ecophysiology of Cereals), INRA, Université Clermont Auvergne (UCA), 5 chemin de Beaulieu, 63039 Clermont-Ferrand, France.

⁸Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelů 31, CZ-78371, Olomouc, Czech Republic.

- ⁹Helmholtz Center Munich, Plant Genome and Systems Biology (PGSB), Ingolstaedter Landstr. 1 85764 Neuherberg, Germany.
- ¹⁰John Innes Centre, Crop Genetics, Norwich Research Park, Norwich NR4 7UH, United Kingdom.
- ¹¹University of Saskatchewan, Crop Development Centre, Agriculture Building, 51 Campus Drive, Saskatoon SK, S7N 5A8, Canada.
- ¹³URGI, INRA, Université Paris-Saclay, 78026 Versailles, France.
- ¹⁴Montana State University, Plant Sciences and Plant Pathology, Cereal Genomics Lab, 412 Leon Johnson Hall, Bozeman, MT 59717, USA.
- ¹⁵Institute of Plant Sciences - Paris-Saclay, Biology Department, Bâtiment 630, rue de Noetzelin, Plateau du Moulon, CS80004, 91192 - Gif-sur-Yvette Cedex, France.
- ¹⁶Council for Agricultural Research and Economics (CREA), Research Centre for Genomics & Bioinformatics, via S. Protaso, 302, I -29017 Fiorenzuola d'Arda, Italy.
- ¹⁷Bayer CropScience, Crop Science Division, Research & Development, Innovation Centre, 3500 Paramount Parkway, Morrisville, NC 27560, United States.
- ¹⁸Bayer CropScience, Trait Research, Innovation Center, Technologiepark 38, 9052, Gent, Belgium.
- ¹⁹King Abdullah University of Science and Technology, Biological and Environmental Science & Engineering Division, Thuwal, 23955-6900, Kingdom of Saudi Arabia.
- ²⁰INRA, CNRGV, Chemin de Borde Rouge CS 52627 31326 Castanet Tolosan cedex, France.
- ²¹Agriculture and Agri-Food Canada, Swift Current Research and Development Centre, Box 1030, Swift Current, SK S9H 3X2, Canada.
- ²²National Research Council Canada, Aquatic and Crop Resource Development, 110 Gymnasium Place, Saskatoon SK S7N 0W9, Canada.
- ²³Research Center of Biotechnology of the Russian Academy of Sciences, Institute of Bioengineering, Leninsky Ave. 33, bld 2, Moscow 119071, Russia.
- ²⁴University of Western Australia, School of Biological Sciences and Institute of Agriculture, University of Western Australia, Perth, 6009 Australia.
- ²⁵School of Plant Sciences and Food Security, Tel Aviv University, Ramat Aviv 69978, Israel.
- ²⁶University of Haifa, Institute of Evolution and the Department of Evolutionary and Environmental Biology, 199 Abba-Hushi Avenue, Mount Carmel, Haifa 3498838, Israel.
- ²⁷Universidad de Málaga, Lenguajes y Ciencias de la Computación, Campus de Teatinos, 29071 Málaga, Spain.
- ²⁸Kansas State University, Plant Pathology, Throckmorton Hall, Kansas State University, Manhattan KS, 66506, United States.
- ³⁰University of Minnesota, Department of Agronomy and Plant Genetics, 411 Borlaug Hall, St. Paul, MN 55108.
- ³¹Institute of Crop Science, NARO (former NIAS), 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8518, Japan.
- ³³Instituto de Agricultura Sostenible (IAS-CSIC), Consejo Superior de Investigaciones Científicas, Alameda del Obispo s/n, 14004 Córdoba, Spain.
- ³⁴University of California, Davis, Department of Plant Sciences, One Shield Avenue, Davis, CA 95617, United States.
- ³⁵The Sainsbury Laboratory, Norwich Research Park, NR4 7UH, Norwich, United Kingdom.

- ³⁶Murdoch University, Australia China Centre for Wheat Improvement, School of Veterinary and Life Sciences, 90 South Street, Murdoch WA 6150, Australia.
- ³⁷Agricultural Institute, MTA Centre for Agricultural Research, Applied Genomics Department, 2 Brunszvik Street, Martonvásár H 2462, Hungary.
- ³⁸Rothamsted Research, Biointeractions and Crop Protection, West Common, Harpenden, AL5 2JQ, United Kingdom.
- ³⁹GYDLE, Suite 220, 1135 Grande Allée, Ouest, Suite 220, Québec, QC G1S 1E7, Canada.
- ⁴⁰Julius Kühn-Institut, Institute for Biosafety in Plant Biotechnology, Erwin-Baur-Str. 27 06484 Quedlinburg, Germany.
- ⁴¹University of Zurich, Department of Plant and Microbial Biology, Zollikerstrasse 107, 8008 Zurich, Switzerland.
- ⁴²Rothamsted Research, Computational and Analytical Sciences, West Common, Harpenden, AL5 2JQ, United Kingdom.
- ⁴³University of Saskatchewan, Global Institute for Food Security, 110 Gymnasium Place Saskatoon SK S7N 4J8, Canada.
- ⁴⁴Technical University of Munich, School of Life Sciences, Weihenstephan, Germany.
- ⁴⁵The University of Western Australia, School of Molecular Sciences, ARC Centre of Excellence in Plant Energy Biology, 35 Stirling Highway, Crawley WA 6009, Australia.
- ⁴⁶Kyoto University, Graduate School of Agriculture, Kitashirakawaoiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan.
- ⁴⁷The Federal Research Center Institute of Cytology and Genetics, SB RAS, pr. Lavrentyeva 10, Novosibirsk 630090, Russia.
- ⁴⁸University of Bologna, Department of Agricultural Sciences, Viale Fanin, 44 40127 Bologna, Italy.
- ⁴⁹Palacký University, Centre of the Region Haná for Biotechnological and Agricultural Research, Department of Molecular Biology, Šlechtitelů 27, CZ-78371 Olomouc, Czech Republic.
- ⁵⁰Earlham Institute, Core Bioinformatics, Norwich, NR4 7UZ, United Kingdom.
- ⁵¹Moscow State University, Faculty of Biology, Leninskie Gory, 1, Moscow, 119991, Russia.
- ⁵³South Dakota State University, Agronomy Horticulture and Plant Science, 2108 Jackrabbit Dr, Brookings, SD 57006, United States.
- ⁵⁴University of Maryland, Plant Science and Landscape Architecture, 4291 Fieldhouse Road, 2102 Plant Sciences Building College Park, MD 20742, United States.
- ⁵⁶Norwegian University of Life Sciences, Faculty of Bioscience, Department of Plant Science, Arboretveien 6, 1433 Ås, Norway.
- ⁵⁷Norwegian University of Life Sciences, Faculty of Bioscience, Department of Animal and Aquacultural Sciences, Arboretveien 6, 1433 Ås, Norway.
- ⁵⁸Istituto di Genomica Applicata, Via J. Linussio 51, Udine, 33100, Italy.
- ⁶¹CEA - Institut de Biologie François-Jacob, Genoscope, 2 Rue Gaston Cremieux 91057 Evry Cedex, France.
- ⁶²Monsanto SAS, 28000 Boissay, France.
- ⁶³Institut National de la Recherche Agronomique (INRA), 2 rue Gaston Crémieux, 9057 Evry, France.
- ⁶⁵Punjab Agricultural University, Ludhiana, School of Agricultural Biotechnology, ICAR-National Bureau of Plant Genetic Resources, Dev Prakash Shastri Marg, New Delhi 110012, India.

⁶⁶Northwest A&F University, State Key Laboratory of Crop Stress Biology in Arid Areas, College of Agronomy, Northwest A&F University, Yangling 712101, Shaanxi, China.

⁶⁷German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany.

⁷⁰ICAR-National Research Centre on Plant Biotechnology, LBS Building, Pusa Campus, New Delhi 110012, India.

⁷¹University of Delhi South Campus, Interdisciplinary Center for Plant Genomics & Department of Plant Molecular Biology, Benito Juarez Road, New Delhi-110021, India.

⁷²Bayer CropScience, Breeding & Trait Development, Technologiepark 38, 9052, Gent, Belgium.

Additional Data File S1 (separate excel file)

Assignment of genes to modules within the six different WGCNA co-expression networks. Each network is presented in a separate tab within the Excel file.

Additional Data File S2 (separate excel file)

Comparison of modules across networks. The modules within each network (grain, leaf, root and spike) were compared to modules in all other tissue networks. Modules which did not have a significant overlap ($p_{adj} < 0.05$) to any module in the other network are listed as "no overlap", "-" means there was an overlap to at least one module in the other network. Additional tabs give details of pairwise comparisons between networks.

References and Notes

1. W. Albertin, P. Marullo, Polyploidy in fungi: Evolution after whole-genome duplication. *Proc. Biol. Sci.* **279**, 2497–2509 (2012). [doi:10.1098/rspb.2012.0434](https://doi.org/10.1098/rspb.2012.0434) [Medline](#)
2. S. P. Otto, J. Whitton, Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**, 401–437 (2000). [doi:10.1146/annurev.genet.34.1.401](https://doi.org/10.1146/annurev.genet.34.1.401) [Medline](#)
3. A. Salman-Minkov, N. Sabath, I. Mayrose, Whole-genome duplication as a key factor in crop domestication. *Nat. Plants* **2**, 16115 (2016). [doi:10.1038/nplants.2016.115](https://doi.org/10.1038/nplants.2016.115) [Medline](#)
4. S. Renny-Byfield, J. F. Wendel, Doubling down on genomes: Polyploidy and crop plants. *Am. J. Bot.* **101**, 1711–1725 (2014). [doi:10.3732/ajb.1400119](https://doi.org/10.3732/ajb.1400119) [Medline](#)
5. M. Feldman, A. A. Levy, T. Fahima, A. Korol, Genomic asymmetry in allopolyploid plants: Wheat as a model. *J. Exp. Bot.* **63**, 5045–5059 (2012). [doi:10.1093/jxb/ers192](https://doi.org/10.1093/jxb/ers192) [Medline](#)
6. Y. Van de Peer, E. Mizrachi, K. Marchal, The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424 (2017). [doi:10.1038/nrg.2017.26](https://doi.org/10.1038/nrg.2017.26) [Medline](#)
7. S. Ohno, *Evolution by Gene Duplication* (Springer-Verlag, 1970).
8. F. A. Kondrashov, Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc. Biol. Sci.* **279**, 5048–5057 (2012). [doi:10.1098/rspb.2012.1108](https://doi.org/10.1098/rspb.2012.1108) [Medline](#)
9. K. D. Makova, W.-H. Li, Divergence in the spatial pattern of gene expression between human duplicate genes. *Genome Res.* **13**, 1638–1645 (2003). [doi:10.1101/gr.1133803](https://doi.org/10.1101/gr.1133803) [Medline](#)
10. International Wheat Genome Sequencing Consortium, Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**, eaar7191 (2018). [doi: 10.1126/science.aar7191](https://doi.org/10.1126/science.aar7191)
11. E. Martinez-Perez, P. Shaw, G. Moore, The *Ph1* locus is needed to ensure specific somatic and meiotic centromere association. *Nature* **411**, 204–207 (2001). [doi:10.1038/35075597](https://doi.org/10.1038/35075597) [Medline](#)
12. G. Moore, K. M. Devos, Z. Wang, M. D. Gale, Cereal genome evolution. Grasses, line up and form a circle. *Curr. Biol.* **5**, 737–739 (1995). [doi:10.1016/S0960-9822\(95\)00148-5](https://doi.org/10.1016/S0960-9822(95)00148-5) [Medline](#)
13. Additional materials and methods are available as supplementary materials.
14. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016). [doi:10.1038/nbt.3519](https://doi.org/10.1038/nbt.3519) [Medline](#)
15. P. Borrill, R. Ramirez-Gonzalez, C. Uauy, expVIP: A customisable RNA-seq data analysis and visualization platform. *Plant Physiol.* **170**, 2172–2186 (2016). [doi:10.1104/pp.15.01667](https://doi.org/10.1104/pp.15.01667) [Medline](#)
16. M. Melé, P. G. Ferreira, F. Reverter, D. S. DeLuca, J. Monlong, M. Sammeth, T. R. Young, J. M. Goldmann, D. D. Pervouchine, T. J. Sullivan, R. Johnson, A. V. Segrè, S. Djebali, A. Niarchou, GTEx Consortium, F. A. Wright, T. Lappalainen, M. Calvo, G. Getz, E. T. Dermitzakis, K. G. Ardlie, R. Guigó, The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015). [doi:10.1126/science.aaa0355](https://doi.org/10.1126/science.aaa0355) [Medline](#)

17. J. W. Walley, R. C. Sartor, Z. Shen, R. J. Schmitz, K. J. Wu, M. A. Urich, J. R. Nery, L. G. Smith, J. C. Schnable, J. R. Ecker, S. P. Briggs, Integration of omic networks in a developmental atlas of maize. *Science* **353**, 814–818 (2016).
[doi:10.1126/science.aag1125](https://doi.org/10.1126/science.aag1125) [Medline](#)
18. M. Pfeifer, K. G. Kugler, S. R. Sandve, B. Zhan, H. Rudi, T. R. Hvidsten, International Wheat Genome Sequencing Consortium, K. F. X. Mayer, O.-A. Olsen, Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* **345**, 1250091 (2014).
[doi:10.1126/science.1250091](https://doi.org/10.1126/science.1250091) [Medline](#)
19. D. Winter, B. Vinegar, H. Nahal, R. Ammar, G. V. Wilson, N. J. Provart, An “Electronic Fluorescent Pictograph” browser for exploring and analyzing large-scale biological data sets. *PLOS ONE* **2**, e718 (2007). [doi:10.1371/journal.pone.0000718](https://doi.org/10.1371/journal.pone.0000718) [Medline](#)
20. P. Borrill, N. Adamski, C. Uauy, Genomics as the key to unlocking the polyploid potential of wheat. *New Phytol.* **208**, 1008–1022 (2015). [doi:10.1111/nph.13533](https://doi.org/10.1111/nph.13533) [Medline](#)
21. R. Avni, R. Zhao, S. Pearce, Y. Jun, C. Uauy, F. Tabbita, T. Fahima, A. Slade, J. Dubcovsky, A. Distelfeld, Functional characterization of *GPC-I* genes in hexaploid wheat. *Planta* **239**, 313–324 (2014). [doi:10.1007/s00425-013-1977-y](https://doi.org/10.1007/s00425-013-1977-y) [Medline](#)
22. K. J. Simons, J. P. Fellers, H. N. Trick, Z. Zhang, Y. S. Tai, B. S. Gill, J. D. Faris, Molecular characterization of the major wheat domestication gene *Q*. *Genetics* **172**, 547–555 (2006). [doi:10.1534/genetics.105.044727](https://doi.org/10.1534/genetics.105.044727) [Medline](#)
23. C. E. Grover, J. P. Gallagher, E. P. Szadkowski, M. J. Yoo, L. E. Flagel, J. F. Wendel, Homoeolog expression bias and expression level dominance in allopolyploids. *New Phytol.* **196**, 966–971 (2012). [doi:10.1111/j.1469-8137.2012.04365.x](https://doi.org/10.1111/j.1469-8137.2012.04365.x) [Medline](#)
24. S. Renny-Byfield, J. P. Gallagher, C. E. Grover, E. Szadkowski, J. T. Page, J. A. Udall, X. Wang, A. H. Paterson, J. F. Wendel, Ancient gene duplicates in *Gossypium* (cotton) exhibit near-complete expression divergence. *Genome Biol. Evol.* **6**, 559–571 (2014).
[doi:10.1093/gbe/evu037](https://doi.org/10.1093/gbe/evu037) [Medline](#)
25. M. Hao, A. Li, T. Shi, J. Luo, L. Zhang, X. Zhang, S. Ning, Z. Yuan, D. Zeng, X. Kong, X. Li, H. Zheng, X. Lan, H. Zhang, Y. Zheng, L. Mao, D. Liu, The abundance of homoeologue transcripts is disrupted by hybridization and is partially restored by genome doubling in synthetic hexaploid wheat. *BMC Genomics* **18**, 149 (2017).
[doi:10.1186/s12864-017-3558-0](https://doi.org/10.1186/s12864-017-3558-0) [Medline](#)
26. M. J. Yoo, E. Szadkowski, J. F. Wendel, Homoeolog expression bias and expression level dominance in allopolyploid cotton. *Heredity* **110**, 171–180 (2013).
[doi:10.1038/hdy.2012.94](https://doi.org/10.1038/hdy.2012.94) [Medline](#)
27. P. P. Edger, R. Smith, M. R. McKain, A. M. Cooley, M. Vallejo-Marin, Y. Yuan, A. J. Bewick, L. Ji, A. E. Platts, M. J. Bowman, K. L. Childs, J. D. Washburn, R. J. Schmitz, G. D. Smith, J. C. Pires, J. R. Puzey, Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. *Plant Cell* **29**, 2150–2167 (2017). [doi:10.1105/tpc.17.00010](https://doi.org/10.1105/tpc.17.00010) [Medline](#)
28. A. J. Bewick, R. J. Schmitz, Gene body DNA methylation in plants. *Curr. Opin. Plant Biol.* **36**, 103–110 (2017). [doi:10.1016/j.pbi.2016.12.007](https://doi.org/10.1016/j.pbi.2016.12.007) [Medline](#)

29. D. Zilberman, An evolutionary case for functional gene body methylation in plants and animals. *Genome Biol.* **18**, 87 (2017). [doi:10.1186/s13059-017-1230-2](https://doi.org/10.1186/s13059-017-1230-2) [Medline](#)
30. A. J. Bewick, L. Ji, C. E. Niederhuth, E.-M. Willing, B. T. Hofmeister, X. Shi, L. Wang, Z. Lu, N. A. Rohr, B. Hartwig, C. Kiefer, R. B. Deal, J. Schmutz, J. Grimwood, H. Stroud, S. E. Jacobsen, K. Schneeberger, X. Zhang, R. J. Schmitz, On the origin and evolutionary consequences of gene body DNA methylation. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9111–9116 (2016). [doi:10.1073/pnas.1604666113](https://doi.org/10.1073/pnas.1604666113) [Medline](#)
31. X. Zhang, O. Clarenz, S. Cokus, Y. V. Bernatavichute, M. Pellegrini, J. Goodrich, S. E. Jacobsen, Whole-genome analysis of histone H3 lysine 27 trimethylation in Arabidopsis. *PLOS Biol.* **5**, e129 (2007). [doi:10.1371/journal.pbio.0050129](https://doi.org/10.1371/journal.pbio.0050129) [Medline](#)
32. E. D. Akhunov, A. W. Goodyear, S. Geng, L. L. Qi, B. Echaliier, B. S. Gill, Miftahudin, J. P. Gustafson, G. Lazo, S. Chao, O. D. Anderson, A. M. Linkiewicz, J. Dubcovsky, M. La Rota, M. E. Sorrells, D. Zhang, H. T. Nguyen, V. Kalavacharla, K. Hossain, S. F. Kianian, J. Peng, N. L. Lapitan, J. L. Gonzalez-Hernandez, J. A. Anderson, D. W. Choi, T. J. Close, M. Dilbirligi, K. S. Gill, M. K. Walker-Simmons, C. Steber, P. E. McGuire, C. O. Qualset, J. Dvorak, The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. *Genome Res.* **13**, 753–763 (2003). [doi:10.1101/gr.808603](https://doi.org/10.1101/gr.808603) [Medline](#)
33. R. J. A. Buggs, N. M. Elliott, L. Zhang, J. Koh, L. F. Viccini, D. E. Soltis, P. S. Soltis, Tissue-specific silencing of homoeologs in natural populations of the recent allopolyploid *Tragopogon mirus*. *New Phytol.* **186**, 175–183 (2010). [doi:10.1111/j.1469-8137.2010.03205.x](https://doi.org/10.1111/j.1469-8137.2010.03205.x) [Medline](#)
34. H. Zhao, W. Zhang, L. Chen, L. Wang, A. P. Marand, Y. Wu, J. Jiang, Proliferation of regulatory DNA elements derived from transposable elements in the maize genome. *Plant Physiol.* **176**, 2789–2803 (2018). [doi:10.1104/pp.17.01467](https://doi.org/10.1104/pp.17.01467) [Medline](#)
35. C. D. Hirsch, N. M. Springer, Transposable element influences on gene expression in plants. *BBA Gene Regul. Mech.* **1860**, 157–165 (2017).
36. A. Roulin, P. L. Auer, M. Libault, J. Schlueter, A. Farmer, G. May, G. Stacey, R. W. Doerge, S. A. Jackson, The fate of duplicated genes in a polyploid plant genome. *Plant J.* **73**, 143–153 (2013). [doi:10.1111/tjp.12026](https://doi.org/10.1111/tjp.12026) [Medline](#)
37. J.-T. Li, G.-Y. Hou, X.-F. Kong, C.-Y. Li, J.-M. Zeng, H.-D. Li, G.-B. Xiao, X.-M. Li, X.-W. Sun, The fate of recent duplicated genes following a fourth-round whole genome duplication in a tetraploid fish, common carp (*Cyprinus carpio*). *Sci. Rep.* **5**, 8199 (2015). [doi:10.1038/srep08199](https://doi.org/10.1038/srep08199) [Medline](#)
38. C. Yu, Y. Liu, A. Zhang, S. Su, A. Yan, L. Huang, I. Ali, Y. Liu, B. G. Forde, Y. Gan, MADS-box transcription factor *OsMADS25* regulates root development through affection of nitrate accumulation in rice. *PLOS ONE* **10**, e0135196 (2015). [doi:10.1371/journal.pone.0135196](https://doi.org/10.1371/journal.pone.0135196) [Medline](#)
39. L.-H. Yu, Z.-Q. Miao, G.-F. Qi, J. Wu, X.-T. Cai, J.-L. Mao, C.-B. Xiang, MADS-box transcription factor *AGL21* regulates lateral root development and responds to multiple external and physiological signals. *Mol. Plant* **7**, 1653–1669 (2014). [doi:10.1093/mp/ssu088](https://doi.org/10.1093/mp/ssu088) [Medline](#)

40. S. Guo, Y. Xu, H. Liu, Z. Mao, C. Zhang, Y. Ma, Q. Zhang, Z. Meng, K. Chong, The interaction between OsMADS57 and OsTB1 modulates rice tillering via *DWARF14*. *Nat. Commun.* **4**, 1566 (2013). [doi:10.1038/ncomms2542](https://doi.org/10.1038/ncomms2542) [Medline](#)
41. A. Ghazalpour, S. Doss, B. Zhang, S. Wang, C. Plaisier, R. Castellanos, A. Brozell, E. E. Schadt, T. A. Drake, A. J. Lusis, S. Horvath, Integrating genetic and network analysis to characterize genes related to mouse weight. *PLOS Genet.* **2**, e130 (2006). [doi:10.1371/journal.pgen.0020130](https://doi.org/10.1371/journal.pgen.0020130) [Medline](#)
42. M. Kumar, W. Busch, H. Birke, B. Kemmerling, T. Nürnberger, F. Schöffl, Heat shock factors HsfB1 and HsfB2b are involved in the regulation of *Pdfl.2* expression and pathogen resistance in *Arabidopsis*. *Mol. Plant* **2**, 152–165 (2009). [doi:10.1093/mp/ssn095](https://doi.org/10.1093/mp/ssn095) [Medline](#)
43. K. M. Pajerowska-Mukhtar, W. Wang, Y. Tada, N. Oka, C. L. Tucker, J. P. Fonseca, X. Dong, The HSF-like transcription factor TBF1 is a major molecular switch for plant growth-to-defense transition. *Curr. Biol.* **22**, 103–112 (2012). [doi:10.1016/j.cub.2011.12.015](https://doi.org/10.1016/j.cub.2011.12.015) [Medline](#)
44. M. Ikeda, N. Mitsuda, M. Ohme-Takagi, Arabidopsis HsfB1 and HsfB2b act as repressors of the expression of heat-inducible Hsfs but positively regulate the acquired thermotolerance. *Plant Physiol.* **157**, 1243–1254 (2011). [doi:10.1104/pp.111.179036](https://doi.org/10.1104/pp.111.179036) [Medline](#)
45. G. Xu, M. Yuan, C. Ai, L. Liu, E. Zhuang, S. Karapetyan, S. Wang, X. Dong, uORF-mediated translation allows engineered plant disease resistance without fitness costs. *Nature* **545**, 491–494 (2017). [doi:10.1038/nature22372](https://doi.org/10.1038/nature22372) [Medline](#)
46. K. V. Krasileva, H. A. Vasquez-Gross, T. Howell, P. Bailey, F. Paraiso, L. Clissold, J. Simmonds, R. H. Ramirez-Gonzalez, X. Wang, P. Borrill, C. Fosker, S. Ayling, A. L. Phillips, C. Uauy, J. Dubcovsky, Uncovering hidden variation in polyploid wheat. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E913–E921 (2017). [doi:10.1073/pnas.1619268114](https://doi.org/10.1073/pnas.1619268114) [Medline](#)
47. Y. Zhang, Z. Liang, Y. Zong, Y. Wang, J. Liu, K. Chen, J.-L. Qiu, C. Gao, Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nat. Commun.* **7**, 12617 (2016). [doi:10.1038/ncomms12617](https://doi.org/10.1038/ncomms12617) [Medline](#)
48. FAO, www.fao.org/faostat/.
49. L. J. Leach, E. J. Belfield, C. Jiang, C. Brown, A. Mithani, N. P. Harberd, Patterns of homoeologous gene expression shown by RNA sequencing in hexaploid bread wheat. *BMC Genomics* **15**, 276 (2014). [doi:10.1186/1471-2164-15-276](https://doi.org/10.1186/1471-2164-15-276) [Medline](#)
50. N. Hamilton, ggtern: An extension to 'ggplot2', for the creation of ternary diagrams. R package version xxxx (2016); <https://CRAN.R-project.org/package=ggtern>.
51. H.-B. Zhang, X. Zhao, X. Ding, A. H. Paterson, R. A. Wing, Preparation of megabase-size DNA from plant nuclei. *Plant J.* **7**, 175–184 (1995). [doi:10.1046/j.1365-313X.1995.07010175.x](https://doi.org/10.1046/j.1365-313X.1995.07010175.x)
52. S. Lindgreen, AdapterRemoval: Easy cleaning of next-generation sequencing reads. *BMC Res. Notes* **5**, 337 (2012). [doi:10.1186/1756-0500-5-337](https://doi.org/10.1186/1756-0500-5-337) [Medline](#)

53. F. Krueger, S. R. Andrews, Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011). [doi:10.1093/bioinformatics/btr167](https://doi.org/10.1093/bioinformatics/btr167) [Medline](#)
54. R Core Team, R: A language and environment for statistical computing (R Foundation for Statistical Computing, 2013); www.R-project.org/.
55. M. Dowle, M. Srinivasan, data.table: Extension of 'data.frame'. R package version 1.10.4-3 (2017); <https://CRAN.R-project.org/package=data.table>.
56. A. Akalin, M. Kormaksson, S. Li, F. E. Garrett-Bakelman, M. E. Figueroa, A. Melnick, C. E. Mason, methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol.* **13**, R87 (2012). [doi:10.1186/gb-2012-13-10-r87](https://doi.org/10.1186/gb-2012-13-10-r87) [Medline](#)
57. A. Akalin, V. Franke, K. Vlahoviček, C. E. Mason, D. Schübeler, Genomation: A toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics* **31**, 1127–1129 (2015). [doi:10.1093/bioinformatics/btu775](https://doi.org/10.1093/bioinformatics/btu775) [Medline](#)
58. H. Wickham, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2009).
59. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014). [doi:10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170) [Medline](#)
60. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012). [doi:10.1038/nmeth.1923](https://doi.org/10.1038/nmeth.1923) [Medline](#)
61. F. Ramírez, D. P. Ryan, B. Grüning, V. Bhardwaj, F. Kilpert, A. S. Richter, S. Heyne, F. Dündar, T. Manke, deepTools2: A next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016). [doi:10.1093/nar/gkw257](https://doi.org/10.1093/nar/gkw257) [Medline](#)
62. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010). [doi:10.1093/bioinformatics/btq033](https://doi.org/10.1093/bioinformatics/btq033) [Medline](#)
63. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011). [doi:10.1093/bioinformatics/btr064](https://doi.org/10.1093/bioinformatics/btr064) [Medline](#)
64. C.-N. Chow, H.-Q. Zheng, N.-Y. Wu, C.-H. Chien, H.-D. Huang, T.-Y. Lee, Y.-F. Chiang-Hsieh, P.-F. Hou, T.-Y. Yang, W.-C. Chang, PlantPAN 2.0: An update of plant promoter analysis navigator for reconstructing transcriptional regulatory networks in plants. *Nucleic Acids Res.* **44**, D1154–D1160 (2016). [doi:10.1093/nar/gkv1035](https://doi.org/10.1093/nar/gkv1035) [Medline](#)
65. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). [doi:10.1186/1471-2105-9-559](https://doi.org/10.1186/1471-2105-9-559) [Medline](#)
66. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014). [doi:10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8) [Medline](#)
67. Y. Benjamini, D. Yekutieli, The Control of the False Discovery Rate in Multiple Testing under Dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
68. V. A. Huynh-Thu, A. Irrthum, L. Wehenkel, P. Geurts, Inferring regulatory networks from expression data using tree-based methods. *PLOS ONE* **5**, e12776 (2010). [doi:10.1371/journal.pone.0012776](https://doi.org/10.1371/journal.pone.0012776) [Medline](#)

69. A. Alexa, J. Rahnenfuhrer, topGO: Enrichment analysis for gene ontology. R package version 2.30.0 (2016).
70. H. J. Schoonbeek, H.-H. Wang, F. L. Stefanato, M. Craze, S. Bowden, E. Wallington, C. Zipfel, C. J. Ridout, Arabidopsis EF-Tu receptor enhances bacterial disease resistance in transgenic wheat. *New Phytol.* **206**, 606–613 (2015). [doi:10.1111/nph.13356](https://doi.org/10.1111/nph.13356) [Medline](#)
71. H. Webster, thesis, Murdoch University, Perth, Australia (2014).
72. V. Chagué, J. Just, I. Mestiri, S. Balzergue, A.-M. Tanguy, C. Huneau, V. Huteau, H. Belcram, O. Coriton, J. Jahier, B. Chalhoub, Genome-wide gene expression changes in genetically stable synthetic and natural wheat allohexaploids. *New Phytol.* **187**, 1181–1194 (2010). [doi:10.1111/j.1469-8137.2010.03339.x](https://doi.org/10.1111/j.1469-8137.2010.03339.x) [Medline](#)
73. R. Suzuki, H. Shimodaira, Pvcust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006). [doi:10.1093/bioinformatics/btl117](https://doi.org/10.1093/bioinformatics/btl117) [Medline](#)
74. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. A Stat. Soc.* **57**, 289–300 (1995).
75. P. A. Wilkinson, M. O. Winfield, G. L. A. Barker, S. Tyrrell, X. Bian, A. M. Allen, A. Burridge, J. A. Coghill, C. Waterfall, M. Caccamo, R. P. Davey, K. J. Edwards, CerealsDB 3.0: Expansion of resources and data integration. *BMC Bioinformatics* **17**, 256 (2016). [doi:10.1186/s12859-016-1139-x](https://doi.org/10.1186/s12859-016-1139-x) [Medline](#)
76. Y. Wang, H. Tang, J. D. Debarry, X. Tan, J. Li, X. Wang, T. H. Lee, H. Jin, B. Marler, H. Guo, J. C. Kissinger, A. H. Paterson, MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49–e49 (2012). [doi:10.1093/nar/gkr1293](https://doi.org/10.1093/nar/gkr1293) [Medline](#)
77. S. Proost, J. Fostier, D. De Witte, B. Dhoedt, P. Demeester, Y. Van de Peer, K. Vandepoele, i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Res.* **40**, e11–e11 (2012). [doi:10.1093/nar/gkr955](https://doi.org/10.1093/nar/gkr955) [Medline](#)
78. B. J. Haas, A. L. Delcher, J. R. Wortman, S. L. Salzberg, DAGchainer: A tool for mining segmental genome duplications and synteny. *Bioinformatics* **20**, 3643–3646 (2004). [doi:10.1093/bioinformatics/bth397](https://doi.org/10.1093/bioinformatics/bth397) [Medline](#)
79. D. Lang, B. Weiche, G. Timmerhaus, S. Richardt, D. M. Riaño-Pachón, L. G. G. Corrêa, R. Reski, B. Mueller-Roeber, S. A. Rensing, Genome-wide phylogenetic comparative analysis of plant transcriptional regulation: A timeline of loss, gain, expansion, and correlation with complexity. *Genome Biol. Evol.* **2**, 488–503 (2010). [doi:10.1093/gbe/evq032](https://doi.org/10.1093/gbe/evq032) [Medline](#)
80. S. Lê, J. Josse, F. Husson, FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Softw.* 10.18637/jss.v025.i01 (2008).
81. K. P. Schliep, phangorn: Phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011). [doi:10.1093/bioinformatics/btq706](https://doi.org/10.1093/bioinformatics/btq706) [Medline](#)
82. M. D. Shirley, Z. Ma, B. S. Pedersen, S. J. Wheelan, Efficient "pythonic" access to FASTA files using pyfaidx. *PeerJ Prepr.* **3**, e970v971 (2015).

83. L. Venturini, S. Caim, G. Kaithakottil, D. L. Mapleson, D. Swarbreck, Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *BioRxiv* 216994 [Preprint]. 9 November 2017. <https://doi.org/10.1101/216994>.
84. N. Goto, P. Prins, M. Nakao, R. Bonnal, J. Aerts, T. Katayama, BioRuby: Bioinformatics software for the Ruby programming language. *Bioinformatics* **26**, 2617–2619 (2010). [doi:10.1093/bioinformatics/btq475](https://doi.org/10.1093/bioinformatics/btq475) [Medline](#)
85. J. Daron, N. Glover, L. Pingault, S. Theil, V. Jamilloux, E. Paux, V. Barbe, S. Mangenot, A. Alberti, P. Wincker, H. Quesneville, C. Feuillet, F. Choulet, Organization and evolution of transposable elements along the bread wheat chromosome 3B. *Genome Biol.* **15**, 546 (2014). [doi:10.1186/s13059-014-0546-4](https://doi.org/10.1186/s13059-014-0546-4) [Medline](#)
86. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002). [doi:10.1093/nar/gkf436](https://doi.org/10.1093/nar/gkf436) [Medline](#)
87. D. Charif, J. R. Lobry, in *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*, U. Bastolla, M. Porto, H. E. Roman, M. Vendruscolo, Eds. (Springer Berlin Heidelberg, 2007), pp. 207–232.
88. W.-H. Li, Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.* **36**, 96–99 (1993). [doi:10.1007/BF02407308](https://doi.org/10.1007/BF02407308) [Medline](#)
89. T. Chujo, R. Takai, C. Akimoto-Tomiyama, S. Ando, E. Minami, Y. Nagamura, H. Kaku, N. Shibuya, M. Yasuda, H. Nakashita, K. Umemura, A. Okada, K. Okada, H. Nojiri, H. Yamane, Involvement of the elicitor-induced gene OsWRKY53 in the expression of defense-related genes in rice. *BBA Gene Struct. Expression* **1769**, 497–505 (2007). [doi:10.1016/j.bbaexp.2007.04.006](https://doi.org/10.1016/j.bbaexp.2007.04.006) [Medline](#)
90. L. Hu, M. Ye, R. Li, Y. Lou, OsWRKY53, a versatile switch in regulating herbivore-induced defense responses in rice. *Plant Signal. Behav.* **11**, e1169357 (2016). [doi:10.1080/15592324.2016.1169357](https://doi.org/10.1080/15592324.2016.1169357) [Medline](#)
91. L. Hu, M. Ye, R. Li, T. Zhang, G. Zhou, Q. Wang, J. Lu, Y. Lou, The Rice Transcription Factor WRKY53 Suppresses Herbivore-Induced Defenses by Acting as a Negative Feedback Modulator of Mitogen-Activated Protein Kinase Activity. *Plant Physiol.* **169**, 2907–2921 (2015). [Medline](#)
92. Y. Jiang, M. K. Deyholos, Functional characterization of Arabidopsis NaCl-inducible *WRKY25* and *WRKY33* transcription factors in abiotic stresses. *Plant Mol. Biol.* **69**, 91–105 (2009). [doi:10.1007/s11103-008-9408-3](https://doi.org/10.1007/s11103-008-9408-3) [Medline](#)
93. Z. Zheng, S. A. Qamar, Z. Chen, T. Mengiste, Arabidopsis WRKY33 transcription factor is required for resistance to necrotrophic fungal pathogens. *Plant J.* **48**, 592–605 (2006). [doi:10.1111/j.1365-313X.2006.02901.x](https://doi.org/10.1111/j.1365-313X.2006.02901.x) [Medline](#)
94. S. Matsukura, J. Mizoi, T. Yoshida, D. Todaka, Y. Ito, K. Maruyama, K. Shinozaki, K. Yamaguchi-Shinozaki, Comprehensive analysis of rice *DREB2*-type genes that encode transcription factors involved in the expression of abiotic stress-responsive genes. *Mol. Genet. Genomics* **283**, 185–196 (2010). [doi:10.1007/s00438-009-0506-y](https://doi.org/10.1007/s00438-009-0506-y) [Medline](#)

95. D. Guo, G. Qin, EXB1/WRKY71 transcription factor regulates both shoot branching and responses to abiotic stresses. *Plant Signal. Behav.* **11**, e1150404 (2016).
[doi:10.1080/15592324.2016.1150404](https://doi.org/10.1080/15592324.2016.1150404) [Medline](#)
96. Y.-W. Yang, H.-C. Chen, W.-F. Jen, L.-Y. Liu, M.-C. Chang, Comparative Transcriptome Analysis of Shoots and Roots of TNG67 and TCN1 Rice Seedlings under Cold Stress and Following Subsequent Recovery: Insights into Metabolic Pathways, Phytohormones, and Transcription Factors. *PLOS ONE* **10**, e0131391 (2015).
[doi:10.1371/journal.pone.0131391](https://doi.org/10.1371/journal.pone.0131391) [Medline](#)
97. T. Pick, M. Jaskiewicz, C. Peterhansel, U. Conrath, Heat shock factor HsfB1 primes gene transcription and systemic acquired resistance in Arabidopsis. *Plant Physiol.* **159**, 52–55 (2012). [doi:10.1104/pp.111.191841](https://doi.org/10.1104/pp.111.191841) [Medline](#)