nature plants

Supplementary information

https://doi.org/10.1038/s41477-025-02128-0

Striking convergent selection history of wheat and barley and its potential for breeding

In the format provided by the authors and unedited

Additional online files

Supplementary online DataSets: https://entrepot.recherche.data.gouv.fr/dataset.xhtml?persistentId=doi%3A10.57745%2FMX18MU&version=DRAFT

DataSet S1. List of orthogroups between Triticeae genomes
Dataset S2. Passport information of the barley and wheat accessions
DataSet S3. Vcf of non-imputed and imputed SNPs
DataSet S4. Matrices of orthoSNPs for barley, tetraploid and hexaploid wheats
DataSet S5. List of candidate genes and functional validation information

Supplementary Tables

Supplementary information

Comparative genomics and reconstruction of Triticeae Ancestor.

To reconstruct the ancestral karyotypes of Triticeae, hereafter referred to as ATK, the gene sets from barley and wheat were compared to identify paraloguous and orthologus genes. Triticeae genomes (*Hordeum vulgare*, Hvulgare_462_r1; *Triticum urartu*, TuG1812 IGDB; *Aegilops tauchii*, Aet_MR_1.0; *Triticum dicoccoides* Zavitan, WEWSeq v.1.0; *Triticum durum* cv. Svevo, svevo v1; *Triticum aestivum*, v1.0 iwgsc) have been compared (BlastP between annotated proteomes) to identify orthologous genes using OrthoFinder (Emms and Kelly 2019) defining families of conserved genes referenced to as OrthoGroups (OGs, Supplementary Table 1). Ancestral Triticeae karyotype (ATK) has been reconstructed according to a two-stage procedure. While the ancestral karyotype is reconstructed in the first stage, the ancestral gene content of such karyotypes, and the gene order, are inferred in the second stage. In Stage 1, the ancestral karyotype (*i.e.* the ancestral chromosome number) is determined based on genome alignments (using cumulative identity percentage and cumulative alignment length percentage blast parameters (Murat et al. 2017). Conserved genes (*i.e.* putative protogenes or pPG) are identified from these alignments and pPGs conserved in all investigated species (*i.e.* core protogenes or core-pPG) are then extracted. These core-pPGs are used to identify synteny blocks (SBs) removing groups of fewer than five genes. SBs are then merged into ancestral protochromosomes (also referred to as contiguous ancestral regions, CARs), corresponding to independent sets of blocks sharing paralogous and/or orthologous relationships in modern species. Stage 2 of the procedure is aiming at ordering protogenes on the previously defined protochromosomes. Genes (pPGs) conserved between pairs of species but not constituting core-pPGs (used in stage 1) are mapped onto the protochromosomes, delivering an exhaustive set of ordered protogenes (oPGs) consisting of 32,833 ancestral genes positioned on 7 ancestral chromosomes (Dataset S1, Supplementary Table 2). This consists in 11,925 genes 'perfectly' conserved ancestral genes showing complete (*i.e.* core protogenes or core-pPG with a ratio of 1:2:3 respectively in barley, tetraploid wheat and hexaploidy wheat) orthologous relationships between the investigated crops, and 7,328 'partially' conserved genes (pPGs), for which the ortholog ratio deviates from 1:2:3 due to deletions, local duplications, or assembly and annotation inconsistency between species. From ATK the proposed evolutionary scenario refines the understanding of the emergence of ATK from the known AGK, as well as the complex history of chromosome fusion, fission and translocation events that shaped the modern Triticeae genomes over millions of years of evolution (Supplementary Figure 1).

SNPs, conserved orthoSNPs and phylogenomics analysis in wheat and barley.

SNP calling. Our general strategy for the comparative analyses of wheat and barley genetic diversity was based on independent production of arrays of nucleotide diversity (in VCF format) for the three species - diploid barley, tetraploid wheat and hexaploid wheat, and subsequent integration of this diversity data through the concept of orthogenes and orthoSNPs. The 679 barley (*Hordeum vulgare*, *H. spontaneum*) and 672 wheat accessions (252 tetraploid and 420 hexaploids) representing a broad sample of cultivar, landrace and wild genotypes across their eco-geographical ranges have been sequenced for exome diversity analysis as described in Pont et al. 2019 (Dataset S2). The three datasets were mapped on their respective reference genome (*i.e.* cv. Chinese Spring for hexaploidy wheat, cv. Svevo for tetraploid wheat, cv Morex for Barley). The 679 barley lines used here are a subset of those analyzed in our previous works (Bustos-Korts et al. 2019, Chen et al. 2022). The 252 tetraploid wheat genotypes considered herein have been carefully selected from a wide collaborative collection referred to as the Global Durum Genomic Resource, including the Tetraploid Germplasm Collection (TGC, Maccaferri et al. 2019) of 1,856 wild and domesticated emmer and *T. turgidum* landrace genotypes and the Global Durum Panel (GDO, Mazzucotelli et al. 2020) of 1,000 genotypes, mainly cultivated materials and landraces. Based on the Illumina 90K wheat SNP array as a common genotyping platform, genomic data from both collections were subjected to phylogenetic and population structure analysis and a reduced number of genotypes were short-listed and exome-captured to provide an accurate representation of major non-admixed, pure haplotypes present in main populations and subpopulations of the GDGR (Maccaferri et al. personal communication). A detailed description of the variant calling workflow is available in the original publication (Pont et al. 2019, Maccaferri et al. 2019), but a brief summary is presented here. The paired end Illumina data was mapped to the reference sequence of cultivar Morex (Mascher et al. 2017, Monat et al. 2019) using BWA-MEM (Li 2013), and further pre-processing of the resulting BAM files was carried out using GATK v.3.4.0 in accordance with the GATK Best Practices pipeline (Van der Auwera et al. 2013). Variants were then called with the GATK's HaplotypeCaller and the join GenotypeCaller tools (Poplin et al. 2018). The tetraploid and hexaploid wheat datasets were prepared following similar pipelines. The pre-processed paired reads of the hexaploid and tetraploid samples were mapped onto the Chinese Spring v1.0 (IWGSC 2018) and Svevo v2 (Maccaferri et al. 2019) reference genomes, respectively, using bwa mem with default settings (Li, 2013). In order to avoid exceeding the maximum sequence length for sam/bam indexing, the reference chromosomes had been split into halves, approximately resembling chromosomal arms. Initial mapping files were post-processed using picard-tools to mark duplicates, fix mates and sort the output. The resulting bam files were then used to jointly call variants per-chromosome using samtools mpileup (version 1.3) with parameters --min-MQ 20, --excl-flags UNMAP,SECONDARY,QCFAIL,DUP (samtools) and bcftools call (version 1.3) with parameters -vcO z -f GQ. In

the resulting vcf files, heterozygous variant calls and those with depth <3 and genotype quality <10 were masked as missing data. From this point onwards, the vcf files of the three species (barley, tetraploid and hexaploid wheats) were treated with identical procedures. Only biallelic SNPs with minor allele frequency >0.005 located within annotated genes and 1kb up-/down-stream regions were retained. These basic vcf files were used in preliminary Principal Component Analysis (PCA) in order to identify samples with uncertain genetic identity (e.g. suspected interploidy hybrids and introgressed lines in wheats, wild-domesticated hybrids etc.). These suspected hybrids were removed, together with samples with high proportions of missing data (outliers defined as 3rd quartile +1.5*inter-quartile range). The currently delivered repertoire of 1351 accessions were finally retained for further genetic analyses. Only sites with minor allele counts ≥4 were kept. Subsequently, two sets of vcf files were prepared, one for the analysis of the population structure, and another for the detection of the selection signatures. In the population structure datasets, sites with a high proportion of missing calls (>0.333) and sites with outlier sum depth values were removed, yielding 536,669, 799,883 and 236,760 SNPs for barley, tetraploid and hexaploid wheat, respectively (Dataset S3). In the datasets intended for the analysis of selection signatures, sites with >10% of heterozygous calls (based on the raw vcf files before masking) were removed, and missing genotype calls were imputed with Beagle5.3 (Browning et al. 2021). This final 'imputed' dataset consists of 1,850,997, 1,537,720 and 485,900 SNPs for barley, tetraploid and hexaploid wheat (Dataset S3, Supplementary Table 3). ***Population structure*.** Genetic populations were defined separately in each of the species (barley, tetraploid and hexaploid wheat) on the basis of PCA, sNMF ancestry coefficients (Frichot et al. 2014), Maximum-Likelihood (ML) tree topology and passport data. PCA was performed with smartpca from the EIGENSOFT package (Patterson et al. 2006) with the least-square PCA projection and no outlier removal. sNMF was run for K2-15 with five iterations using various values of the regularization parameter (10-1000). The ML phylogeny was inferred with iqtree (Nguyen et al. 2015), using Modelfinder (Kalyaanamoorthy et al. 2017) to identify the optimal substitution model (TVM+F+ASC+R4 for both wheat datasets and TVM+F+ASC+R3 for barley). In the barley dataset, wild accessions were excluded from the population structure analysis, since the wild diversity strongly confounded the structure of the domesticated groups. Number of populations was identified empirically by comparing the sNMF ancestry coefficients obtained for various K-levels to the ML-topology and PCA clusters. The K-level of sNMF best matching the ML-topology was used for the primary population assignment (in barley and tetraploid wheat), and samples with inconsistent positions across the two methods were designated as admixed (referred as ADM). Since the hexaploid wheat dataset produced a poorly-supported ML tree, likely due to low inter-population divergence and frequent gene flow, the primary population assignment was based on the comparison of the sNMF and PCA outputs. Final adjustments were performed on the basis of known ancestry conflicts (from passport data). Genetic diversity structure defined subpopulations of barley (9 domesticated and 1 wild group designated as Hv_pop1-9 and Hv_pop10, respectively). Tetraploid wheat genetic diversity defines BA-types with 4 groups of domesticated emmer wheat (DEW, labelled as DEW_pop3-6), 4 groups of durum wheat (DW, labelled as DW_pop7-10), as well as one group of wild BA-type (*Triticum dicoccoides*) and one group of GA-type wheat (*T. araraticum* and *T. timopheevii*). Finally, the hexaploid bread genetic diversity defines 6 populations (labelled BW_pop1-6) wheats (Supplementary Table 4, Supplementary Figure 2). Geographical coordinates were used to place samples on the map at their coordinates using the map_data function from the R package ggplot2 (Wickham 2016). When known, precise GPS positions of the samples were used. For all other samples, GPS coordinates of the country's capital from which the sample came from were used instead. Average latitudinal and longitudinal coordinates across all the individuals per population were used to place the geographical midpoint of each population on the map. ***OrthoSNPs identification***. From the catalog of 19,253 conserved gene families (or orthogroups, OG) and the total number of 3,923,023 original non-imputed SNPs identified separately on the investigated *Triticeae* species*,* we performed a multi-sequence alignment (Bawono et al. 2017) between conserved genes in order to characterize orthoSNPs, *i.e.* SNPs located at the same nucleotidic position between conserved genes in barley, tetraploid and hexaploid wheat. A total number of 418,791 orthoSNPs on 16,937 conserved genes have been identified. After imputation (see previous section), a total number of 375,901 orthoSNPs were recovered. Most of identified orthoSNPs, 336,016 (~89%) are species-specific (177,206 for barley, 139,691 for tetraploid wheat and 19,119 for hexaploid wheat). A catalog of 39,032 (~11%) orthoSNPs are shared between hexaploid, tetraploid wheats and barley (Supplementary Table 5, Dataset S4). ***Phylogeny inference***. Phylogenomic analysis was performed through RRHS trees (Lischer et al. 2014) in the form of conventional consensus topologies and densitree visualizations to infer taxonomic clades. We analyzed the evolutionary distances among the tips of the 1,000 trees using the minimum spanning tree (MST) algorithm in Python. The MST graphs were subsequently combined into a weighted, phylogenetic consensus network whose nodes were clustered into clades using the Girvan–Newman Edge-Betweenness algorithm in Cytoscape (Kucera et al. 2016). The clustered network topology was plotted considering edge-betweenness in Cytoscape and taxonomic clades were inferred by intersection of community clusters with taxon information that was annotated using the AutoAnnotate plugin. The relative number of RRHS trees for which a respective edge was selected by the MST algorithm were used as edge weights and were interpreted similar to bootstrap support values in the consensus tree topologies. Phylogenetic tree between wheat accessions (738 in total)

was done using orthoSNPs datasets between tetraploids and hexaploids (*i.e.* 33,905 shared SNP positions) converted to pseudo vcf format. The resulted vcfs were then used as inputs to create a matrix for phylogenetic analysis (in PHYLIP format) as described in Ortiz, (2019). The final tree (newick format) was drawn using iqtree2 (Minh et al. 2020) for either A (13,430 shared positions) and B (20,029 shared positions) subgenomes separately, or the full dataset combing the two subgenomes. Visualization and plots were realized with the ggtree R package (Xu et al. 2022). We considered AGW (GA cultivars) populations as outgroup to root the trees (Supplementary Figure 3).

Characterization of conserved sweep signals (OrthoSweeps)

Detection of selection signatures. Signatures of selection (that jointly characterize genomic loci under selection called 'sweeps') were obtained per population for every single position of the three 'imputed' vcf datasets. A variety of methods were tested for their ability to detect convergent signals of selection across the tree datasets. These included SweeD (Pavlidis et al. 2013), OmegaPlus (Alachiotis et al. 2012), iHS (Voight et al. 2006), F_{ST}, Tajima's D and various measures of genetic diversity. Our preliminary tests focused on several attributes that were considered important for the application of the selective sweep analysis in a comparative genomics framework: (i) The statistics should not be consistently associated with the distance from the centromere. Due to strong syntenic relationships among the Triticeae genomes, correlations with centromere distance could lead to false signals of sweep sharing across species. (ii) Assuming that true targets of selection exhibit multiple genomic signatures associated with selection, the statistics are expected to correlate with each other within species (although this is often not observed (Pavlidis et al. 2013)). (iii) The statistics are expected to correlate among the populations of the same species (since they share some selection history by descent), and to a lesser extent, if at all, among populations of different species. The eventual correlation of the statistics across species would indicate the existence of convergent selection. Outputs of several methods were deemed substandard and not included in the inter-species comparisons. SweeD failed to detect strong selection signals, as reported previously (Russell et al. 2016, Civan et al. 2021), probably due to a bug in calculations involving folded site-frequency spectra. The iHS outputs contained large amounts of missing values, especially in bread wheat populations, likely due to low SNP densities. Common measures of diversity and the shifts in site-frequency spectra (*Pi* and Tajima's D, respectively) were not used due to flaws in their estimation from vcf files (Korunes and Samuk, 2021), but also due to a significant correlation of *Pi* with the centromere distance (r = 0.27, averaged across all populations). The statistics chosen for the comparative analysis of selection were omega, F_{ST} and a customized measure of nucleotide diversity ϑ_W*MAF. Omega, which detects a particular pattern of linkage disequilibrium associated with selection, was calculated with OmegaPlus (Alachiotis et al. 2012) for each population and each position of the vcf file, setting the -minwin and -maxwin parameters to 200 and 5,000, respectively. The output omega values were log-transformed. Weir and Cockerham's F_{ST} measuring the level of population differentiation was calculated for each population and the remainder of the species (e.g., DEW_pop3 compared to DEW_pop4-6, etc…) in vcftools (Danecek et al. 2011), using windows of 400 SNPs and a sliding step of 1 SNP. This was achieved by changing the SNP coordinates in the vcf files to consecutive numbers starting at 1 for each chromosome. In order to estimate nucleotide diversity along the chromosomes, we calculated Watterson's Theta (ϑ_W) for each population as the number of segregating sites (polymorphisms) divided by the (n-1)th harmonic number, where n is the population size. Theta was calculated for each gene as the average of 100 consecutive genes (not limited to orthogenes but including all annotated high- and low-confidence genes) centered around the gene of interest. Theta was then multiplied by average MAF (minor allele frequency) calculated in the same genomic windows. Consequently, ϑ_W*MAF is a measure of diversity that reflects both the density of polymorphisms and the shifts of the site-frequency spectrum***. *Comparative analysis of selection signatures***. The three metrics (omega, F_{ST} and ϑ_W*MAF) were firstly included in a region-based analysis for the identification of sweeps in each population, comparing them between the investigated species, using a percentile-based thresholds. For omega and F_{ST}, in each population, signals above the population-based 95th or the 99th percentile were scored as soft-and hard-sweep, respectively, while for ϑ_W*MAF, 5th and 1st percentile were used as thresholds. Population sweep blocks were defined as contiguous regions including two or more adjacent scored polymorphisms. For each species, distal, pericentromeric and centromeric chromosome compartment coordinates were retrieved from Mascher et al. (2017), Maccaferri et al. (2019) and IWGSC et al. (2018) and intersected with sweep block coordinates. This revealed that while strong F_{ST} and omega signals are, on average, uniformly distributed along chromosome subregions, ϑ_W*MAF blocks are preferentially located on recombination-depleted centromeric regions. Since large parts of *Triticeae* genomes often show characteristics of selective sweeps (which can also arise from neutral processes; Supplementary Figure 4), it is difficult to assess whether sweep overlaps across different species occur by chance or represent true cases of parallel selection. For this reason, we did not quantify parallel selection between species using intersections of region-based sweep contents. Instead, we regarded each gene from the 'perfect OGs' as an independent selection unit associated with specific values of the three statistics, E, allowing us to quantify enrichment of signal sharing (see below). Each orthogene was assigned selection values using the mean (for F_{ST} and ϑ_W*MAF) or maximum (for omega) of the data points intersecting the orthogene's genomic position. In cases of no intersection, the closest data point was used. Subsequently, the three metrics were standardized as Z-scores, using the mean and standard

deviation of the variables across orthogenes in all populations of the same species. As seen from the chromosomal distribution of region-based sweeps, ϑ_W*MAF shows positive correlation with the centromere distance consistently across all populations (r = 0.41 on average). This is a consequence of the general pattern of nucleotide diversity along chromosomes (low diversity in the pericentromeric regions, high diversity in the subtelomeric regions; Supplementary Figure 4). Since this pattern can introduce bias in the assessment of convergent selection (see attribute i above), we used local (corresponding to 500 ordered orthogenes) rather than global mean and standard deviation when calculating ϑ_W*MAF z-scores. This adjustment reduced the correlation with the centromere distance to 0.13. Global means and standard deviations were used in the standardization of omega and F_{ST}, as their correlations with the centromere distance were not consistent (both positive and negative, depending on the population; Supplementary Figure 5a). The standardized scores were analyzed separately, but also in combination as $z_c = z_{omega} + z_{FST} - z_{\vartheta*MAF}$. In respect to the attribute (ii) (see above), omega is strongly negatively correlated with ϑ_W*MAF (Supplementary Figure 5b), supporting the expectation that selective sweeps with high omega values are characterized with low nucleotide diversity. Our F_{ST} estimates are not strongly correlated with either omega or nucleotide diversity (Supplementary Figure 5b), which is not surprising, as regions with low differentiation in a pair of populations can have either high or low nucleotide diversity, and the same is true for regions with high population differentiation. Omega and ϑ_W*MAF profiles are clearly correlated among populations of the same species, and to a lesser extent among populations of different species (Supplementary Figure 7a,b), meeting the expectation of the attribute (iii). This is not the case for the profiles of F_{ST} (Supplementary Figure 7c), where correlation between populations of the same species is not expected and generally low. As a consequence, the linear combination of the three metrics shows only a weak positive correlation for most population combinations (Supplementary Figure 7d), indicating that some convergent selection could be at play. ***Enrichment of signal sharing.*** Neither of the three statistics detects signatures that are exclusively attributable to selection. Genomic signatures indicative of selection can also arise through neutral (stochastic) processes, such as lineage sorting, bottlenecks and drift. Such signal noise is difficult to separate from the true signals of selection, potentially leading to false conclusions. It is therefore critical to have a robust strategy to distinguish real selection signals from the neutral background. In our detection of selected loci, we relied on empirical outliers, *i.e.* we did not compare the obtained statistics against a chosen demographic model. However, we took advantage of the comparative nature of our sweep detection objective. Using thresholds defined by percentiles of the chosen statistics, it is possible to calculate sharing expected purely by chance, *i.e.* the number of OGs that are expected to be shared among two or more randomly drawn sets of OGs. Subsequently, we can compare an observed number of OGs shared by two unrelated populations at a given percentile (og_o) to the number of OGs expected to be shared by chance (og_e), and calculate the 'signal enrichment' (og_o/og_e). Ratios below or close to 1 indicate that either the chosen statistic failed to detect real selection signals (low sensitivity, high false-positive rate), or that the populations being compared have not experienced convergent selection. On the other hand, signal enrichment substantially over 1 indicates convergent selection. We have therefore calculated signal enrichment for an array of alpha percentiles (from 0.6 to 0.99 with 0.01 increment, and for 0.999) for each of the three statistics and their linear combination z_c. The expected probabilities were calculated for the cases when the signal is shared by all homeologs of polyploid wheat (eq. 1), and for the cases when the signal is shared by any homeolog of polyploid wheat (eq. 2). (eq. 1) $og_e = 11925 * (1-a)^x$, where a is the alpha percentile and x is the number of genes within an OG required to be above the given percentile. For example, we can expect that just by chance, ~1.5 out of 11,925 OGs will be above the 95th percentile simultaneously in a particular population of barley and both subgenomes of a tetraploid population($og_e = 11925 * 0.05^3$). (eq. 2) $og_e = 11925 * (1-a^y) * (1-a^z)$, where a is the alpha percentile, y is the number of homoeologs in the first population, and z is the number of homeologs in the second population. For example, we can expect that simply by chance, ~3.5 OGs will be above the 99th percentile in a particular population of barley and any of the three subgenomes of a hexaploid population $\left(og_e = 11925 * (1-0.99^1) * (1-0.99^3)\right)$. When enrichment of signal sharing >20x was detected at any of the tested percentile levels (i.e., when the number of observed orthogenes with selection signals of a particular strength in two populations of different crops simultaneously was 20x higher that that number expected by chance), all of the implicated orthogenes were treated as affected by selection, either directly or indirectly (by linkage), and listed in Supplementary Table 6-9. We compared these results to the analysis of region-based sweeps (see above) in order to assess whether our orthogene-based approach corresponds to the traditional analysis of sweeps defined as linkage blocks under selection (Supplementary Figure 6-8). When 95th percentile is used to define sweep blocks in the region-based analysis, the majority of the orthogenes with >20x signal sharing are found within the sweep blocks, with matching metric and populations. In particular, 386 (out of 451) orthogenes implied in >20x enrichment are found in the sweep blocks defined by omega, 81 (out of 106) for the sweep blocks defined by F_{ST}, and 2 (out of 56) for the sweep blocks defined by ϑ_W*MAF. The low consistency in ϑ_W*MAF estimates are due to the correction of the centromere-distance dependency that was used in the orthogene-based z-score calculation, but not in the region-based sweep detection. ***Investigation of known domestication genes***. We checked the performance of the three metrics on four genes (*Btr1, Btr2, Vrn1* and *Vrn3*) known to be involved in domestication and adaptation. Mutations disrupting

the function of the *Btr* gene products are responsible for the non-brittle rachis of cultivated forms in both barley and wheat, which is the crucial transition during the domestication of these crops. In barley, mutations in either of the two *Btr* genes, which are separated by only 90 kb, bring about the non-brittle phenotype (Porkheirandish et al. 2015). Since each barley population contains both non-brittle types (*btr1-Btr2* and *Btr1-btr2*) in various proportions, the diversity at these two loci does not indicate selection. Nonetheless, six out of nine cultivated barley populations have omega values above the 90th percentile at the two loci, suggesting selection. In wheats, the situation at the *Btr* loci is different. Based on studies on domesticated einkorn and emmer (Pourkheirandish et al. 2018, Avni et al. 2017), it appears that mutations on *Btr1* (both its homeologs in the case of tetraploids) are sufficient for the transition to non-brittleness, and the function of *Btr2* has not been disrupted. The situation is likely similar in bread wheat; however, population diversity at the *Btr* loci has not been assessed in polyploid wheats. In our data, the *Btr1-Btr2* region has low levels of nucleotide diversity, variable F_{ST} and high values of omega. Both *Btr1-Btr2* loci on chromosomes 3A and 3B are located within a >1Mb region with virtually no nucleotide variation in domesticated tetraploids (though the loci are at the distal ends of those swept regions on both homoeologous chromosomes). The uniformity of these regions across all tetraploid populations is not reflected by F_{ST} because our calculation was based on sliding windows with constant number of SNPs (400), spanning >3Mb and >5Mb in the case of the A and B homeologs, respectively. Overall, the profiles are consistent with selection acting on the *Btr1* homeologs, with the selection signal extending to the linked *Btr2* locus. An exception is the domesticated emmer from the Horn of Africa, where omega is below the median. In contrast to the *Btr* genes, the *Vrn* genes are characterized by large allelic variation in both wheat and barley (Fernandez-Calleja et al. 2021), which is confirmed in our data. The vernalization genes *Vrn1* and *Vrn3* are major determinants of the flowering time and drivers of phenological adaptation (another vernalization gene *Vrn2* was not examined due to its deletion in the barley reference genome and consequent absence of data in the vcf matrix). *Vrn1* is the central regulator of vernalization-induced flowering, modulating the expression of *Vrn3* after being induced by prolonged exposure to cold. *Vrn3* is the main integrator of photoperiod and vernalization signals, and its expression leads to the transition from the vegetative to the reproductive state. The complexity of the vernalization-photoperiod pathway and its interactions with the environment (Fernandez-Calleja et al. 2021) allows a continuous gradation of vernalization responses and the promotion of flowering. A phenotype advantageous in a given environment can be achieved *via* multiple polymorphisms, some of which have pleiotropic effects affecting other traits like frost tolerance and yield (Fernandez-Calleja et al. 2021). It is therefore impossible to predict which of the *Vrn* genes should be under selection in which population. Nonetheless, since the winter growth habit (*i.e.* the vernalization requirement) is the ancestral 'wild-type' state, and the spring-habit mutations allowed the expansion of cultivation into areas where a strong cold requirement cannot be satisfied, it can be expected that one or more *Vrn* genes would be under selection in e.g. subtropical regions and also in temperate regions but under spring sowing (e.g. central to north EU). In accordance with this expectation, both *Vrn1* and *Vrn3* are associated with high omega (95th percentile) signals of selection in Ethiopian barley_pop6. This is consistent with the observation that Ethiopian barleys are uniform in their 'winter' allele at *Vrn3*, but have lost the essential vernalization-responsive region of *Vrn1* intron 1 (Tsehaye et al. 2012), likely contributing to their spring habit. Additional signals of selection (omega or F_{ST} above 95th percentile) are associated with the vernalization genes in multiple barley and wheat populations. ***Genes associated with orthoSweeps***. Our analysis identified 22 orthoSweeps containing 451 orthogenes (with 1 to 231 orthogenes per orthoSweep), see details in Supplementary Table 9. The strongest signal of convergent selection in our analyses lies on chromosome 5 (orthoSweep5.1) for a region of ~160 kb 0.98 percentile of omega, and 0.96 percentile of $-\vartheta_W$*MAF) in the Hv_pop5 (Eastern Fertile Crescent) and both subgenomes of the DEW_pop5 (Transcaucasia/Iran). The omega signal is also present in the Hv_pop3 (Transcaucasia), where it contributes to an extreme peak of z_c (0.999 percentile) shared with the B subgenome of the DW_pop10 (Mediterranean). The core region contains three OGs (OG0013770, OG0013769 and OG0019192), one of which encodes a subunit of polycomb repressive complex (OG0013770, Strejčková et al. 2020). Polycomb repressive complexes play important roles in epigenetic gene regulation by post-translationally modifying specific histone residues. The same gene was identified as a candidate for the vigor of temperature response, which affects the timing of stem elongation and final height in bread wheat (Kronenberg et al. 2021). This region also includes a cytokin glucosyl transferase, a key regulator of cytokinin homeostasis with potential impacts on crop productivity (Chen et al. 2021). Besides the strongest signal of convergent selection, chromosome 5 contains additional sweeps shared by multiple crops. One region (OrthoSweep5.2) is above the 0.999 $-\theta_W$*MAF percentile on the A subgenome of DW_pop9 (Transcaucasia) and the B subgenome of BW_pop5 (Tunisia and Levant) simultaneously, and contains a gene implicated in the resistance response to *Rhizoctonia cerealis* (OG0014053, Geng et al. 2022). Another region (OrthoSweep5.3) is above the 0.98 $-\theta_W$*MAF percentile in Hv_pop4 (Horn of Africa) and both subgenomes of DEW_pop4 (Continental Europe, Balkans) and carries OGs associated with kernel traits (OG0013987, Wang et al. 2022a) and responsive to drought stress (OG0013986, Wang et al. 2022b). Chromosome 1 contains several strong signals of convergent selection. OrthoSweep1.1 is above the 0.999 omega percentile in Hv_pop1 (European I) and the B subgenome of DEW_pop6 (Horn of Africa). This locus contains a gene for a RING-type E3 ligase, which was identified as a candidate

regulating high amylose starch biosynthesis in wheat (OG0008850, Parveen et al. 2021). OrthoSweep1.2 is above the 0.999 F_{ST} percentile in geographically overlapping populations Hv_pop1 (European I) and the A subgenome of BW_pop2 (Continental Europe and Scandinavia), containing a gene associated with root-related traits (OG0016778, Soriano and Alvaro 2019). A locus on chromosome 1 is above the 0.97 -ϑ_W*MAF percentile in both subgenomes of the emmer population 5 (Transcaucasia/Iran) and the durum population 10 (Mediterranean). It contains OG0021046 encoding a protein with an NHL domain, which is also present in a rice protein regulating kernel weight (Qu et al. 2021, Chen et al. 2015). The largest shared sweeps encompass over 46 Mb on chromosome 2 (orthoSweep2.1). The locus is above the 0.99 omega percentile simultaneously in the B subgenome of DW_pop7 (*T.t.* subsp. *carthlicum* and Horn of Africa) and the D subgenome of BW_pop4 (Mediterranean). It is also weakly selected (above 0.78 omega percentile) in both subgenomes of DW_pop7 and all three subgenomes of BW_pop6 (Western Europe). The locus contains 99 OGs, many of which are involved in processes related to adaptation and productivity. These include genes associated with resistance to pathogens (OG0009795, OG0009836, OG0009890, OG0017381, OG0017397), abiotic stresses (OG0017381, OG0009787, OG0009812, OG0009815, OG0009821, OG0009831, OG0009833, OG0009835), sugar and nitrate transporter genes implicated in growth, development and stress tolerance (OG0009825, OG0009826, OG0009827, OG0009828, OG0020478, OG0009776, OG0009780), cytokinin-related genes (OG0020471, OG0017391), genes associated with glycogen (OG0006475, OG0009778), and candidate genes controlling grain weight (OG0009792) and Fe and Zn content (OG0006485, OG0009777). Another region on the chromosome 2 (OrthoSweep2.2, Supplementary Table 9) is above the 0.999 F_{ST} percentile in all the three species grown in hot and dry areas: Hv_pop4 (Horn of Africa) and the B subgenome of both DEW_pop3 (Southern Levant - Europe) and BW_pop5 (Tunisia and Levant). In wheat populations, this sweep extended to include a gene encoding a histone-modifying enzyme that gets phosphorylated in response to heat (OG0020460, Vu et al. 2018), and members of a Receptor-Like Protein Kinase 1 gene family associated with drought tolerance (OG0009591, Rahim et al. 2022). Additionally, chromosome 2 contains an orthoSweep2.3 (Supplementary Table 9) above the 0.999 omega percentile simultaneously in the B subgenome of DW_pop7 (Persian and Horn of Africa) and the D subgenome of BW_pop5 (Tunisia and Levant). It contains three OGs implicated in kernel size-related traits (OG0009047, OG0009048, OG0009052, Ma et al. 2022), a gene involved in pathways important for drought tolerance (OG0017115, Abdi et al. 2023), and a bHLH47 homolog involved in Fe homeostasis (OG0006358, Wang et al. 2020, Kumar et al. 2022), but this orthosweep also encompasses the key gene of the photoperiod response, *Ppd,* on both 2B and 2D, whose ancestral photoperiod-sensitive alleles have been fully lost partially or completely in most of the modern genotypes as a consequence of human-driven selection for indel variants mostly occurred in promoters and introns while the photoperiod sensitivity allele is still present in wheat landraces. Thus, the *Ppd* gene could be the true selection target of this orthosweep, although not detectable based on the applied SNP capture-based approach. OrthoSweep3.1 on the chromosome 3 is above the 0.999 omega percentile on the B subgenome of DEW_pop6 (Horn of Africa) and the subgenome D of BW_pop4 (Mediterranean). It carries a gene involved in snRNA import into nucleus, identified as contributing to seed morphology in bread wheat (Rabieyan et al. 2022). Chromosome 7 contains orthoSweep7.1 above the 0.98 F_{ST} percentile in Hv_pop7 (Mediterranean) and both subgenomes of DEW_pop3 (Southern Levant and Europe). The locus contains a serotonin N-acetyltransferase (SNAT) gene involved in salt tolerance (OG0020134, Zhan et al. 2019), as well as a gene implicated in the grain Zn and Fe content (OG0020136, Singh et al. 2022). Finally, a large orthoSweep7.2 (chromosome 7) is above the 0.85 F_{ST} percentile in all subgenomes of BW_pop4 (Mediterranean) and DW_pop9 and 10 (Transcaucasia and Mediterranean, respectively). It contains genes involved in flower development (OG0016146, Szeliga et al. 2023) and flowering time (OG0016138, Feng et al. 2020) known to be under positive selection in other crops; a gene controlling awn elongation (OG0016148, Huang et al. 2020), genes implicated in seed germination (OG0020214, Cheng et al. 2022), cold stress (OG0022929, Konstantinov et al. 2021) and pathogen resistance (OG0016153, Pradhan et al. 2020).

Convergent selection from conserved differentiated OrthoSNPs (CD-OrthoSNPs)

Differentiated OrthoSNPs. From the catalog of 39,032 orthoSNPs shared (by position and polymorphism) between hexaploid, tetraploid wheats and barley (see previous 'OrthoSNPs identification' section), we characterized differentially (*i.e* swept) orthoSNPs between populations, focusing on the reference allele frequency and applied a cut-off of ≦0.1 and ≧0.9 (*i.e* an orthoSNP is considered differentially fixed when at least two populations show contrasted 0.1/0.9 allele frequencies). When comparing domesticated populations separately for each species, we identify 18,872, 20,756, 6,381 and 2,807 differentiated orthoSNPs in Hv, DEW, DW and BW, respectively. This delivered 6,055, 6,551, 2,424 and 1,138 genes hosting differentiated orthoSNPs between domesticated populations in Hv, DEW, DW and BW (Supplementary Figure 9, Supplementary Table 10 (complete list), Supplementary Table 11 (percentage per populations/subgenome)). GO analysis showed enrichment in different biological functions according to the targeted species (Supplementary Figure 10). In Hv, the identified genes are enriched in cytokines-activated-signaling pathway. For DEW, genes are enriched in functions related to RNA processing, DNA repair, chromatin remodeling, gene silencing, protein transport (Golgi), embryo development, cell division, etc… For DW, similar functions have been identified like chromatin remodeling, cell division, DNA repair, chloroplast organization, etc. while for BW, functions related to

microtubule-based movement, DNA rewinding and cell cycle were found. Beside GO analysis the identified genes include known genes involved in grain weight or yield (*GSN1:* OG0007903, *OsARF4:* OG0018196, *GL7:* OG0006415*, AN1:* OG0009628*, FLO2:* OG0017642*, SSG4:* OG0010640*, GSD1:* OG0010240, *NOG1:* OG0011265) and in panicle development, spikelet architecture and tillering (*OsARG:* OG0017084, *FUWA:* OG0014551, *DEP3:* OG0007705*, EP2:* OG0009268, *NAL1:* OG0006537, *D11:* OG0009866*, SGL1:* OG0015020, *OsGATA28:* OG0012048, *IAA6:* OG0011212). When comparing wild to domesticated population separately for each species, a total of 8,918, 7,781 and 6,228 orthoSNPs were differentiated between the wild and at least one domesticated population in Hv, DEW and DW, respectively. Most of these orthoSNPs are specific to each crop (8,906 for Hv, 5,677 for DEW and 4,126 for DW) and only 8, 182 and 446 orthoSNPs had variant frequencies differentiated between the wild and all domesticated populations in Hv, DW and DEW respectively (Supplementary Table 14). For Hv, *SPL7 (*OG0023164) gene involved in the control of flowering were identified. In DEW, known genes in seed germination, growth and development like *BUI1* (OG0022644), *G11a* (OG0015515), *OsbZIP60* (OG0009185) were found. The vast majority of candidate genes were identified from the comparison between the wild emmer and DW, including genes involved in flowering (*TAF15b:* OG0017772, *S-locus:* OG0012984, *OsHAK26:* OG0020050, *Osnop:* OG0016082, *OsNTL5:* OG0020092, *OsSPL15:* OG0015632), grain yield (*HD3b:* OG0006323, *ARE1:* OG0013620, *OsMED15_2:* OG0015760, *XIAO:* OG0017507), panicle development (*ES1:* OG0010723, *OsGAMYBL1:* OG0016076, *OsGAMYBL2:* OG0015307, *OsSEU:* OG0012065, *OsSEU2:* OG0012065, *OsSEU3:* OG0016326), starch synthesis (*CDP3.2:* OG0012290), seed germination and plant development (*Gas2:* OG0010873, *SRD2:* OG0011644*, Sweet1a:* OG0011625). ***CD-orthoSNPs***. Conserved and non-conserved differentiated OrthoSNPs (CD-OrthoSNPs and not CD-OrthoSNPs) between different subpopulations of the three investigated species (*i.e* diploid barley, 2 tetraploid wheats with durum wheat and domesticated emmer and the hexaploid wheat) were identified in order to assess convergent signatures of selection (when comparing domesticated genotypes) and domestication (when comparing wild *vs* domesticated genotypes. When looking for convergent signatures of selection between the domesticated populations of the crops (*i.e.* signatures of modern selection and improvement), a total of 1,582 CD-OrthoSNPs were identified, corresponding to 948 non-redundant orthogenes, with 1-16 CD-OrthoSNPs per OG (Supplementary Table 12). Such genes are enriched in function related to protection from non-homologous end joining at telomere, and includes also known genes such as *EP2* (OG0009268) and *FUWA* (OG0014551) for panicle architecture and grain size, *SSG4* (OG0010640) for regulation of starch grain size, *IAA6* (OG0011212) for control of tiller outgrowth, *DRO1* (OG0019052) for root system architecture and drought avoidance, *GI* (OG0010632) for photoperiodic control of flowering and osmotic stress response, *osDR11* (OG0014598) for positive regulation of disease resistance, *osGRAS23* (OG0019340) for drought stress response, *RSS1* (OG0014622) for maintenance of meristematic activity under stress condition, *OsLCBK2* OG0010047 for regulation of disease resistance response and programmed cell death, Dataset S5. Homeologous bias has been investigated in characterizing CD-orthoSNPs between the domesticated populations of the crops, either mapped on single homeologous gene copies or multiple homeologs (Supplementary Table 13). When looking for convergent signatures of domestication between wild and (at least one) domesticated populations of the crops, only few CD-OrthoSNPs are identified (4 between Hv:DEW:DW, 3 between Hv:DW, 5 between Hv:DEW and 2,095 between DEW:DW), and 119 CD-OrthoSNPs are identified between wild and all DEW:DW domesticated populations (Supplementary Table 14, Supplementary Figure 13-15). The 12 domestication CD-orthoSNPs common to barley and various wheats included *RNP-1* (OG0009752) for RNA-binding protein involved in cold adaptation, *DCD* (OG0019930) for development and cell death involved in immune responses and *OMA1* (OG0014948) for mitochondrial proteases involved in heat stress responses. CD-OrthoSNPs identified between tetraploid and hexaploidy wheat could either reflect (independent) convergent selection or (shared) ancestry. In order to differentiate sign of convergence vs ancestry, OrthoSNPs and reference sequences for tetra- and hexaploid wheats were used to reconstruct individual haplotypes of all orthogenes containing one or more CD-orthoSNPs. These haplotypes were then used to construct multi-species gene-trees and median-joining networks (Bandelt et al. 1999), which were visually checked for cases of homoplasy. Homoplasies found in multiple species can be regarded as evolutionary convergence on a nucleotide level (if they are non-neutral), Supplementary Figure 12.

Characterization of the *FUWA* and *DRO1* genes in wheat.

Functional validation of TdFUWA (Supplementary Figure 11a)***.*** *Gene IDs.* TdFUWA-A: TRITD6Av1G100490 - chr6A: 270,171,952-270,190,995. Based on the sequence similarity with the *OsFUWA* gene we referred to the isoform TRITD6Av1G100490.2 and TdFUWA-B: TRITD6Bv1G115800 - chr6B: 373,410,600-373,420,244. Selected isoform: TRITD6Bv1G115800.3. *Construct preparation***.** To edit the TdFUWA homeologous genes, a target region of about 1Kb was identified using the Cas-Designer tool (Park et al. 2015). sgRNAs with an out-of-frame score higher than 66 have been further analyzed for potential off-targets using CasOT (Xiao et al. 2014). Finally, four sgRNAs have been selected for plant transformation and gene editing. Each guide was individually inserted into plasmid pU6-gRNA as for the protocol reported by Shan et al. 2014. The four FUWA pU6-gRNAs were co-bombarded along two more plasmids: pFH66, containing an optimized *Cas9* for wheat editing and pAH20 containing the BASTA gene (marker selection: resistance to phosphinotricine, PPT) (Hahn et al. 2020). *Durum wheat transformation***.** Durum wheat ears (*Triticum turgidum*

subsp. durum, cv Svevo) were collected from donor plants grown in controlled environment rooms with 18°C day/15°C night temperatures and a 16 h photoperiod. Relative humidity was maintained at 50–70 %. Immature seeds (16–21 days post anthesis) were manually removed from the surrounding panicles and surfaced-sterilize by rinsing in 70 % (v/v) aqueous ethanol for 3–5 min, followed by 10 min in 10 % (v/v) aqueous bleach with occasional gentle shaking. The seeds were then rinsed several times with sterile distilled water and drained. Immature embryos, ranging from 1.5 to 2 mm in length, were isolated from the seed and the axis removed to prevent precocious germination. 30 embryos were placed onto induction medium in a 9 cm Petri dish plates and incubated at 22–23 °C in the dark for 1–2 days' pre-culture prior to bombardment (Sparks and Jones, 2014). A total of 260 embryos were bombarded. Callus Induction Medium (MS) was composed by 4.3 g/L Murashige and Skoog (MS) basal salts, 40 g/L maltose, 0.50 mg/L thiamine-HCl, 150 mg/L aspargine, 3.5 g/L Phytagel, 2 mg/L 2,4- dichlorophenoxyacetic acid (2,4-D). *Wheat Transformation by Particle Bombardment.* Plasmid DNA was delivered in wheat immature embryos using the Bio-Rad PDS-1000/He particle gun according to manufacturer's instructions. The exact quantity of each plasmid to employ for gene delivery was calculated depending on the molecular weight according to He et al. (1999). Microcarriers were prepared by mixing plasmid DNA with appropriate aliquots of previously sterilized 0.6 µm (submicron) gold particles, CaCl2 and spermidine. 650 psi rupture discs were used and the following settings were used as standard for the procedure: 2.5 cm gap (distance between rupture disc and macrocarrier), 5.5 cm target distance (distance between stopping screen and target plate), 0.8 cm stopping plate aperture (distance between macrocarrier and stopping screen), 28–29" Hg vacuum, 5.0 vacuum flow rate, 4.5 vacuum vent rate (Sparks and Jones, 2014; Gadaleta et al. 2008). *Regeneration and selection of transformed calluses.* After the bombardment, embryos were evenly spread across the induction medium and incubated in the dark at 22–23 °C for 1 week. Only good quality calluses, showing a regular round shape and a bright yellow color, were subjected to next in vitro culture phases. Calluses were then transferred to selection medium containing PPT (2.5mg/l) for 4-5 weeks in the dark. Responsive calluses were moved to regeneration medium to initiate regeneration from the somatic embryos and incubate in the light at 22–23 °C for 2 weeks. Calluses showing green spots were then moved to rooting media in the light for 2 weeks, after which, some showed little green shoots and some small roots, so to be transferred and pot up in soil. When the plantlets reached the two leaves stage, a small amount of leaf was sampled for DNA extraction and construct occurrence screening. *Mutants' genotype***.** Genome editing experiments led to the creation of the following alleles: *TdFUWA-A_m1:* Inversion of the genomic region comprised between bp 253 and 1306 downstream the ATG. *TdFUWA-B_m1*: Deletion of 394 bp located 250 bp downstream the ATG and deletion of 27 bp located 2300 bp downstream the ATG. *TdFUWA-B_m2*: Inversion of the genomic region comprised between bp 252 and 1313 downstream the ATG. Mutant line ML1 is homozygous for the *Td-FUWA-A_m1* and *Td-FUWA-B_m1* alleles, mutant line ML2 is homozygous for the *Td-FUWA-A_m1* and *Td-FUWA-B_m2* alleles. *Phenotypic analysis***.** Wild-type durum wheat, cv Svevo, BC1F2-ML1, BC1F2-ML2 and homozygous ML1 and ML2 plants were grown in controlled standard conditions. Days to heading were scored when the spike fully emerged from the flag leaf. Leaf length and width (second, third and flag leaf) was measured at flowering. Phenotypic data have been analysed using R statistical environment. Statistical significance of observed difference was assessed using the ANOVA test. Statistically different groups have been identified using the Tukey post hoc test. The R library ggplot2 has been used to generate graphs. ***Haplotype analysis for FUWA and DRO1*** (Supplementary Figure 11b-c). SNPs in FUWA and DRO1 gene intervals were extracted from 6x and 4x vcf files using vcftools (Danecek et al. 2011), and haplotype network analysis was performed and visualized for each species with the geneHapR R package (Zhang et al. 2023). Consensus haplotype sequences for each allele were then retrieved again from the 6x and 4x vcf files using bcftools consensus (Danecek et al. 2021). Cross-species haplotype allele multiple alignment (after manual masking of specie-specific SNPs) and NJ tree were constructed using Molecular Evolutionary Genetics Analysis (MEGA) 11 software (Tamura et al. 2021). ***Phenotyping of root system architecture for DRO1*** (Supplementary Figure 11c)***.*** Root growth angle at the seedling stage was evaluated in 100 domesticated emmer and durum-turgidum accessions (out of the 232 tetraploid accessions included in the genetic study) selected for representing the diverse sub-populations, with the exclusion of AGW and WEW. Phenotyping was done in seedlings grown for 5 days at 22˚ C under a 16-hour light photoperiod in semi-hydroponic condition as described in Maccaferri et al. 2016 and Rufo et al. 2020.

Impact of convergence on gene function (High impact variant) and traits (GWAS).

High impact (HI) variant. High impact SNPs were identified using SnpEff (Cingolani et al. 2012) with default parameters for all "non-reference" alleles (*i.e.* those that differ from the alleles of the reference genomes used for tetraploid wheat, hexaploid wheat, and barley). Statistics on variant numbers, types, and allele frequencies in populations (VAF for Variant allele Frequency: number of variant alleles / total number of alleles) were computed in R (Supplementary Figure 16). Nonsense mutations (mutation types including a "stop gained" event) showed overall the highest proportions among mutation types classified as "high impact" variants (HI-variants), with barley showing the highest proportion (64%). Barley also showed the highest proportion (8%) of "start lost" variants, while tetraploid and hexaploid wheat had the highest proportion (13%) of variants causing loss of stop codons. Among variants affecting splicing sites, tetraploid and hexaploid wheat presented the highest proportions (27 and 28%, respectively). When comparing VAFs

across the species, in general tetraploid wheat had higher VAFs compared to barley and hexaploid wheat, for the variants identified (Supplementary Table 15-16). We also analyzed HI-variants at the population level and found that, for tetraploid wheat, "stop gained" variants showed high VAF for GA and DEW_pop6_Horn of Africa (70% and 66%, respectively), while DW_pop10_Mediterranean showed the least frequency (21%). For hexaploid wheat, France/UK population had the highest VAF (35%), while Nepal the least (25%). For barley, Eastern Fertile Crescent population showed the highest VAF (40%) and *H. spontaneum* showed the least (12%). Finally, among mutations affecting splicing sites, the highest VAFs were observed for GA and DEW_pop6_HornAfrica for tetraploid wheat (82% and 70%, respectively), France/UK (44%) for hexaploid wheat, and Horn of Africa (46%) for barley. Enrichment analyses for annotated gene ontology (Ashburner et al. 2000) biological processes (GO-BP, GO term size > 10) for genes carrying high-impact variants were performed with gprofiler2 (Kolberg et al. 2020) using threshold=0.05 and Benjamini-Hochberg FDR (False Discovery Rate) for multiple testing correction. To gain functional insights on high-impact variants, we investigated the over-representation of biological processes associated to those genes. Among the over-represented processes, we observed strong enrichment signals for "response to stress" and "defense response", with the latter found significantly enriched in every panel population (Supplementary Table 17). For this reason, we focused on the "defense response" process evaluating functional over-representation at increasing variant allele frequencies (e.g. VAF≥0.1, VAF≥0.2, etc..) for each panel population. Generally, while strong enrichment signals were observed for genes carrying variants at low VAF, functional over-representation could still be found for gene sets harboring monomorphic variants in two barley populations (Horn of Africa, Eastern Fertile Crescent) and nearly fixed (VAF≥0.9) in one tetraploid wheat population (DEW_pop6_Horn of Africa). In the hexaploid wheat panel, enrichment signals were observed for genes harboring variants with VAF≤0.3. To gain additional functional insights into convergent HI-variants across barley and wheats, Intersection between HI-variants and orthoSNPs, deliverd 97 HI-orthoSNPs on 96 genes shared across wheat and barley (Supplementary Table 18). To functionally investigate variant impacts on plant phenotypes at the cross-species level, RARGE II database (Akiyama et al. 2014) was queried to gain information on *Arabidopsis* genes that, when mutated, show a significant change in plant phenotypes. ***GWAS analysis***. Barley and hexaploid wheat SNP variants were associated to phenotypic data collected as previously described (Bustos-Korts et al. 2019, Pont et al. 2019). Briefly, common garden experiments were run in Turkey, Italy, France, Hungary, UK, Scotland following an augmented partially replicated design. The number of days from sowing to heading (HD, Zadoks stage 55), plant height (PH) and thousand grain weight (GW) were scored for 425 barley accessions (240 landraces and 185 cultivars) and 420 bread wheat accessions (189 landraces and old cultivars, 231 cultivars), Supplementary Table 19. Genome-Wide Association Scans were run separately for each trait in each environment, using the imputed SNP datasets filtered for MAF 0.05 (ca. 900,000 and 400,000 SNPs for barley and wheat, respectively), with the statgenGWAS R package (van Rossum and Kruijer 2023; https://github.com/Biometris/statgenGWAS/). To identify loci reproducibly associated with the traits under study in different environments, we utilized the summary statistics from single GWAS as input for a meta-analysis with METAL (Willer et al. 2010), in which a sample size weighted fixed-effects model was implemented. Two different thresholds were considered for declaring significant SNP markers from the meta-analyses: the top 0. 5% of whole genome p-value distribution and a p-value <0.0001 after applying a genomic control correction method. Significant markers were then linked to OrthoGroups (OG) described above, with the final aim of identifying common loci controlling the traits under study in the different species (e.g. barley and hexaploid wheat) or in the different subgenomes within polyploid species (e.g. hexaploid genomes A and B), Supplementary Figure 17, Supplementary Table 21.

Ancient-modern GA diversity analysis.

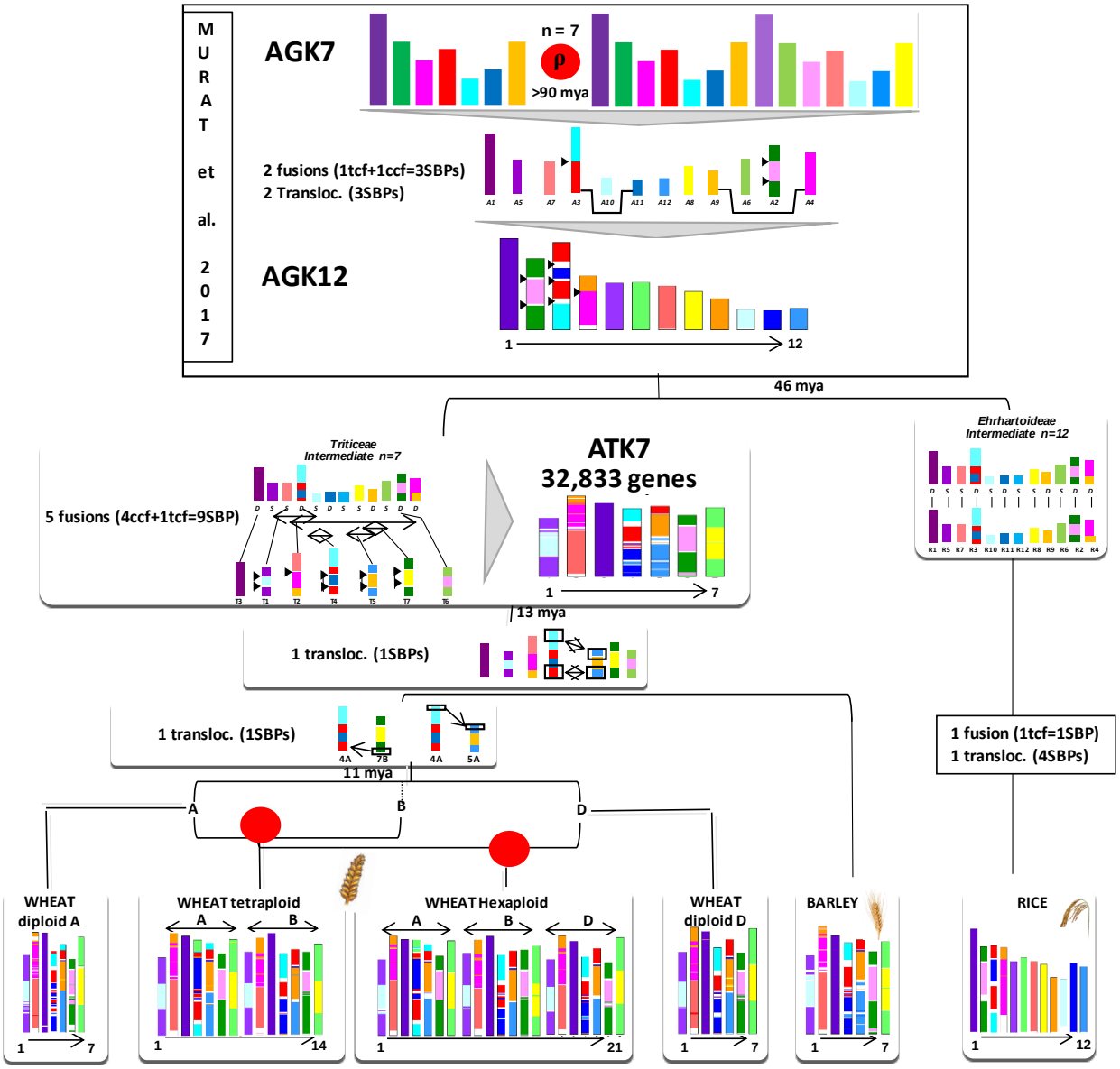
GA diversity analysis. To detect GA introgressions in the populations of BA tetraploids, we used the imputed tetraploid dataset to identify 'GA-specific' variants. We define these as variants fixed in GA wheats (variant frequency 1) and absent in wild *T. dicoccoides* (variant frequency 0). Subsequently, we calculated average allelic frequencies of the GA-specific variants in 5Mb genomic windows along all chromosomes, separately for all DEW and DW populations (Supplementary Figure 18, Supplementary Table 22). ***Ancient wheat samples***. From layer 0, dated between 3200 to 3150 BC (archaeological and dendrochronological inferences), a total of 2074 remains of 62 different taxa were isolated (Supplementary Table 23). ***Morphometry***. We performed an automated morphometry (based on length of abscission *scars and U shape measurements among 10 measured parameters)* of spikelet remains compared to modern spikelet experimentally charred and to carpological references (Supplementary Figure 19). To evaluate the proximity between archaeological NGW and either *T. timopheevii* or *T. dicoccon*, five measurements were taken using photographs of archaeological NGW and *T. dicoccon* spikelets and of spikelets from modern accessions of *T. timopheevii* and *T. dicoccon*. Measurements *b* and *c* are according to Kohler-Schneider (2003). Measurement *d* differs from Kohler-Schneider, who measured the distance between the disarticulation scar and the glume base insertion: the change proved efficient to take measurements from photographs, where absence of depth and varying appearance of the spikelets led to an heterogenous reading. Measurements *e* and *f* were introduced in this study. Measurements on archaeological spikelets were normalized to temper the natural size discrepancies which may exist between ancient and modern individuals. ***aDNA extraction***. Two samples, each composed of very similar morphology of fragments (~50 fragments for #A

and ~300 fragments for #B), were used for DNA extraction and library preparation for shotgun sequencing. Examination of ancient DNA requires a specific analysis environment and protocols to prevent potential contamination (Pont et al. 2019b; Fulton and Shapiro 2019). Appropriate precautions were taken in our DNA extraction method at The Paleogenomics Lab (https://pgl.soe.ucsc.edu/) to ensure the authenticity of the aDNA results (Fulton and Shapiro 2019). Briefly, the steps of DNA extraction (including grinding of grain, DNA extraction and polymerase chain reaction) were carried out in separate rooms dedicated exclusively to the study of aDNA, a mock sample was taken at all stages. All consumables were cleaned with BLEACH-RITE® and irradiated with ultraviolet (UV) light before use. For aDNA extraction, each ancient sample was rinsed in water and air dried under UV irradiation in clean room facilities. After grinding to powder without liquid nitrogen (Retsch MM400), the powder of each sample was collected in a 1.5 ml tube for the DNA extraction and 1 ml extraction buffer containing 1% CTAB together with 20 mM EDTA pH 8.0, 100 mM Tris-HCl pH 8.0, 1.5M NaCl and 3% cross linked PVP (polyvinylpyrrolidone). After suspension by vortexing, samples were incubated at 55°C for 1 hour and rotated at 37°C for 24 hours. Mock sample were done using exactly the same protocol without powder. Following the silica column-based method described in Dabney et al. (2013), extracts were centrifuged at maximum speed (16,100 × g) for 5 minutes, and the supernatant was transferred to a 50 ml tube with a binding apparatus reservoir containing 13 ml of binding buffer (including guanidine hydrochloride, isopropanol, sodium acetate and Tween-20). After centrifugation at 1500 x g for 7 minutes, the extension reservoir of the MinElute column were removed and placed into a clean 2 ml collection tube. The MinElute column was dry-spun at 3,300 × g for 1 minute in a bench-top centrifuge. Two washing steps were performed by adding 750 µl of PE buffer (Qiagen) to the MinElute column, centrifuging at 3,300 × g, and discarding the flow-through. The column was dry-spun at maximum speed (16,100 × g) for 1 minute and then transferred to a fresh 1.5-ml collection tube. For elution, 12.5 µl of TET buffer (1 mM EDTA, 10 mM Tris-HCl, 0.05% Tween20) was pipetted onto the silica membrane and collected by centrifugation at maximum speed for 30 seconds after a 2- to 5-min incubation. This step was repeated for a total of 25 µl of DNA extract. The library was constructed using a protocol for sequencing on the Illumina sequencing platform optimized for ancient and degraded DNA (Kapp et al. 2021) using splinted ligation to simultaneously ligate the Illumina P5 and P7 adapters. ***aDNA authentication***. A minute proportion (N=49,005) provided high-quality alignments against the tetraploid (Svevo) reference genome showing typical signatures of post-mortem DNA damage, showing both an excess of C→T mis-incorporations at alignment termini and an over-representation of purines at those reference positions preceding alignments starts. The former represents the hallmark of post-mortem cytosine deamination (Briggs et al. 2007), while the latter confirm depurination driving post-mortem DNA fragmentation (Stiller et al. 2006). This confirms the presence of ancient DNA fragments in the samples analyzed (Supplementary Figure 20). ***Genetic identification***. 131 580 wheat reads were mapped to the NCBI nucleotide database using BLASTn2.11.0+ (release April 2022) with default parameters and analyzed in MEGAN6.22.2 (Huson et al. 2007) using the Naive LCA (last common ancestor) algorithm for binning. To characterize the taxonomical diversity of the sample CG102, we mapped reads to the NCBI nucleotide database using the NCBI taxonomy summarized in MEGAN software. We obtained 299 taxonomic classes, the two best are *Paraburkholderia aromaticivorans* (isolated from gasoline-contaminated soil according to Lee and Jeon (2018) and *Triticum aestivum*. When considering eukaryote only, we identified Obtectomera (butterflies), Clupeocephala (fishes), Leotiomyceta (filamentous ascomycete fungi) and *Homo sapiens* (less than 0,47%). Most of the other species represented at more than 1% correspond to water microorganisms in line with a sample originating from the shores of lake. Reads where mapped against the wheat genome references (Supplementary Figure 11). ***Read processing and mapping***. The demultiplexed fastq read pairs were processed using PALEOMIX (Schubert et al. 2014) bam_pipeline v1.2.13.2. Sequencing adaptors were trimmed and paired-end reads were collapsed using AdapterRemoval2 (Schubert et al., 2016) v2.3.1, with default parameters. All the resulting reads, *i.e.* those collapsed (and possibly truncated for lower quality ends), and those remaining paired, were aligned against a tetraploid naked wheat (Svevo) reference genome, using bwa (Li & Durbin, 2009) v.0.7.17-r1194-dirty, disabling seeding and using the mapping options -n: 0.1 -o: 2. Alignments showing mapping qualities strictly inferior to 30 were discarded. Read alignment was repeated against the reference genome of an hexaploid naked wheat (IWGSC) reference genome to query the ploidy level of the ancient sample examined (Sample47). PCR duplicates were removed using Picard MarkDuplicates (http://picard. sourceforge.net) and DNA fragmentation and nucleotide misincorporation patterns were visualized using mapDamage2 (Jónsson et al. 2013) v.2.0.8 with default parameters. Per chromosome coverage estimates were computed by paleomix, both at the library and the pan-sample levels, considering read alignments against both reference genomes. The coverage proportion of A, B and/or D genomes provided insights into the genomic makeup of the ancient sample examined, especially whether D genetic material was present. As read alignments against the IWGSC reference did not support the presence of D genetic material, we further only considered the alignments against the Svevo tetraploid reference sequence. The BAM alignment file was further filtered for alignment sizes superior or equal to 32 bp, and processed through PMDtools (Skoglund et al. 2014) v0.60 and the trim_bam.py python script from Librado et al. (2021) to reduce the impact of post-mortem DNA damage on downstream analyses. More specifically, the alignments were binned into two categories according to their likelihood to contain post-mortem damage. Alignments

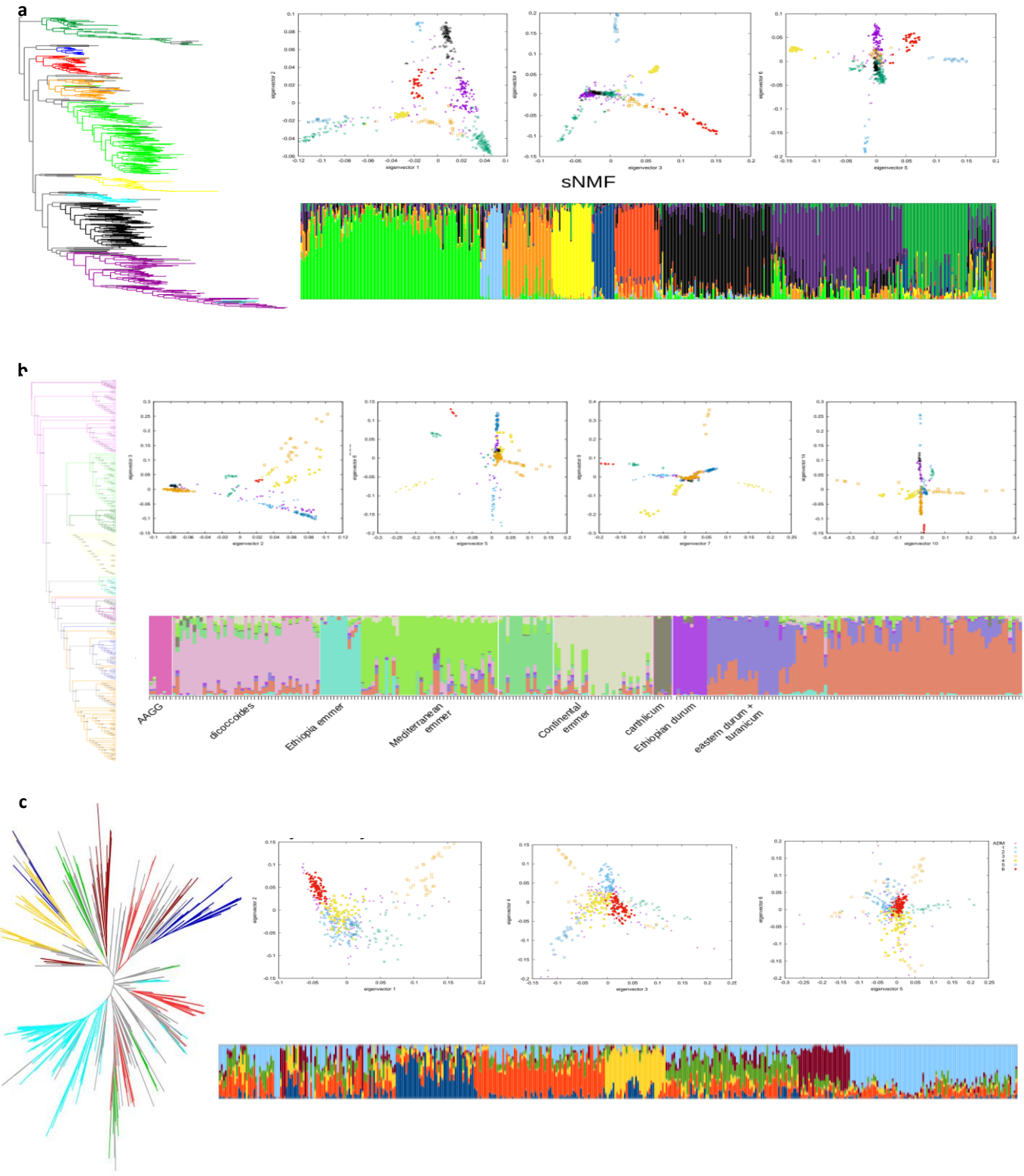
showing PMD scores superior to 1 (--threshold 1) were considered likely to be affected by post-mortem damage, while those remaining (--upperthreshold 1) were considered not affected. The former were trimmed for 3 bases on both termini, whereas the latter were trimmed for a single nucleotide at the 3' end of the alignment. All the resulting alignments were merged into a single BAM alignment before being further pseudo-haploidized, random-sampling one read at positions characterized by one or more alignments. Pseudo-haploid genotypes were called using ANGSD (Korneliussen et al. 2014) v.0.930 (htslib: 1.9), skipping positions and/or reads showing base and/or mapping Phred-quality scores strictly lower than 30 (--doHaploCall 1 -doCounts 1 -minMapQ 30 -minQ 30 -remove_bads 1 -uniqueOnly 1), and restricting calls for those 1,537,720 SNP positions common to the current modern diversity panel. A total of 748 positions that are polymorphic in the current modern diversity panel were also covered in the pseudo-haploid ancient genome. The ancient genotype calls were merged together with the current modern diversity imputed panel, without the hexaploid accessions, using plink v1.9 (Purcell et al. 2007). The whole procedure presented above was applied to the raw sequence data of an ancient Egyptian emmer sample (Scott et al. 2019) for comparison. Population structure. The genetic variation present in our SNP panel was studied using Principal Component Analysis (PCA, using smartpca (Patterson et al. 2006)) and ADMIXTURE (Alexander et al. 2009) v.1.3.0. The PCA leveraged the whole sequence information present in the current modern diversity imputed panel for a total of 242 wheat accessions. These include all accessions, excepting the hexaploid accessions and the 10 that we used as controls, and were selected to be representing important geographic and/or taxonomic groups (LIST). The genotype data available for all 10 control samples as well as the ancient Egyptian emmer sample were down-sampled for 748 random positions, which represents the number of sites from the current modern diversity panel covered in the ancient Sample47 sample. Down-sampling was carried out for a total of 25 replicates to monitor the effect of missingness in the PCA projections. The 25 replicates generated for the 10 control samples plus the ancient Egyptian emmer were projected together with the Sample47 sample against the structure characterizing the 242 tetraploid accessions of the current modern diversity panel, using the "lsqproject: YES" option of smartpca, and considering the autoshrink mode, or not. Both PCA analyses returned similar results. Unsupervised admixture analyses were carried out considering the tetraploid imputed panel accessions supplemented with the two ancient specimens, and SNP positions covered across all samples (plink --geno 0). The genotype matrix was also thinned for linkage disequilibrium following the plink parameters from Scott et al. (2019) (--indep-pairwise 200 10 0.9), leaving a total of 278 SNPs. ADMIXTURE models including K=2 to K=15 genetic ancestry components were considered, and the best model was selected using a cross-validation procedure (--cv=10) only asymptotically declining from K=10. Admixture proportions for the best model (K=10) were estimated from a total of 100 bootstrap pseudo-replicates. A final supervised admixture analysis was run to assess the genomic makeup of the 10 controls and the two ancient samples, pre-defining 10 representative groups from the tetraploid panel accessions (N=242; 532 195 SNPs after pruning with the same parameters as above). ***Principal components analysis***. PCA analysis was performed using the smartPCA tool from the EIGENSOFT package (version 5.0; ttps://github.com/chrchang/eigensoft/blob/master/POPGEN/README), with outlier removal turned off. PCA projections were performed by solving least square equations and not by an orthogonal projection step. Principal Component Analysis, calculated on the tetraploid panel described previously, revealed that the ancient sample from Chalain projected in an intermediate position between *T. timopheevi* (GA karyotypes) and the AABB accessions (regrouping durum, emmer and wild wheat), Supplementary Figure 22. Admixture analyses confirmed the presence of ~25.7±5.2% (unsupervised, Supplementary Figure 23) and ~58.6±6.4% (supervised, Supplementary Figure 24) of an ancestry component characteristic of GA accession. Downsampling demonstrated that both analyses were robust to the amount of missingness characteristic of the ancient sample. Using the same dataset (mapped positions), we computed maximum-likelihood based phylogenies using Iqtree 2.2.0.3 (Nguyen et al. 2015), where the aDNA sample is phylogenetically located between the GA wheats (*T. araraticum*, *T. timopheevii*) and BA *T. dicoccon* groups. This indicates that genetic analysis of the ancient DNA suggests a relatedness to both *T. araraticum - T. timopheevii* group and *T. dicoccon* group. The relevance of the modern wheat diversity panel provided here is illustrated with the analysis of the ancient Egyptian sample previously sequenced and investigated in Scott et al. (2019). Whereas this sample was considered to represent an ancient form of emmer wheat, it clearly clustered together with durum wheat accessions from the Horn of Africa and *T. carthlicum* in our expertise dataset. ***Phylogenetic tree***. The genotype matrix was converted into Fasta format to run IQtree 2.2.0.3 (Chernomor et al. 2016) and compute a phylogenetic tree between samples. The final model used was GTR+F+ASC+R4 as recommended by the tool after an initial run of 1000 replicates. The results were visualized in Dendroscope (Huson et al. 2007). ***Validation of diagnostic markers***. We calculated the frequency of each position to detect diagnostic and common variation between BA and GA tetraploids. Within the vcf file of modern wheat diversity, we identified 167 positions differentially fixed in BA and GA and present in the aDNA sample. Validation was performed using the Fluidigm 96.96 Dynamic ArrayIFCs (Integrated Fluidic Circuits) experiment. Primer pairs were designed for 78 samples consisting of 53 GA samples (8 *T. araraticum* and 45 *T. timopheevii*), 1 GAA sample (*T. zhukovskyi*), 7 *T. aestivum* (BAD), 2 *T. spelta* (BAD), 1 *T. sphaerococcum*, 5 diploids (*T. monococcum*, *T. boeoticum*, *Ae. tauschii*) and 9 tetraploid BA wheats (*T. dicoccon*, *T. dicoccoides*, *T. durum*). Dynamic arrays were performed for samples from the field trials as

described in Fluidigm 96.96 Dynamic Array IFCs (Integrated Fluidic Circuits) (PN 100-1208 B). Chips were run with the Biomark System and data were collected with Fluidigm Real-Time PCR Analysis Software (https://www.standard-bio.com/products-services/software).

<u>Supplementary Figure 1</u>. *Triticeae* genome evolution from reconstructed ATK. ATK derived from the known ancestral grass karyotype (AGK, Murat et al. 2017) reported to consist of 7 chromosomes (AGK7) that were duplicated (ρ paleotetraploidization) to reach an AGK12 intermediate through two fissions and two translocations. Relative to AGK12, ATK underwent 5 fusions and a complex interplay of translocation shaping the modern Triticeae chromosomes 4 and 5. The wheat A subgenome (in tetraploids and hexaploid) shows an additional synteny break point corresponding to the known 4A/7B translocation. The ATK has been then proposed to have derived from *n = 12* AGK (with rice as the modern genome resembling the most to AGK) through 4 centromeric ancestral chromosome fusions (leading to functional monocentric neochromosomes), 1 fission and 2 telomeric ancestral chromosome fusions, involving, with T for ancestral Triticeae chromosomes, the following chromosome relationships: T1 = A10 + A5, T2 = A7 + A4, T3 = A1, T4 = A11 + A3, T5 = A9 + A12, T6 = A2, T7 = A8 + A6. The paleohistory of modern Triticeae genomes from the reconstructed AGK paleokaryotypes involves numerous WGD events (1R-3R, red dots) as well as ancestral chromosome fusion (6) and fission (1) events. The modern barley genome has retained the original ATK chromosome number of 7, making this genome a reference karyotype for comparative genomics investigation within the Triticeae tribe. The modern wheat genomes share the ancestral reciprocal translocation (between A4-A5 on the A subgenome) and underwent two hybridization events (2R, 3R) between A, B and D progenitors as well as an additional lineage-specific translocation (between chromosomes 4A and 7B) to reach the modern 21 chromosomes.

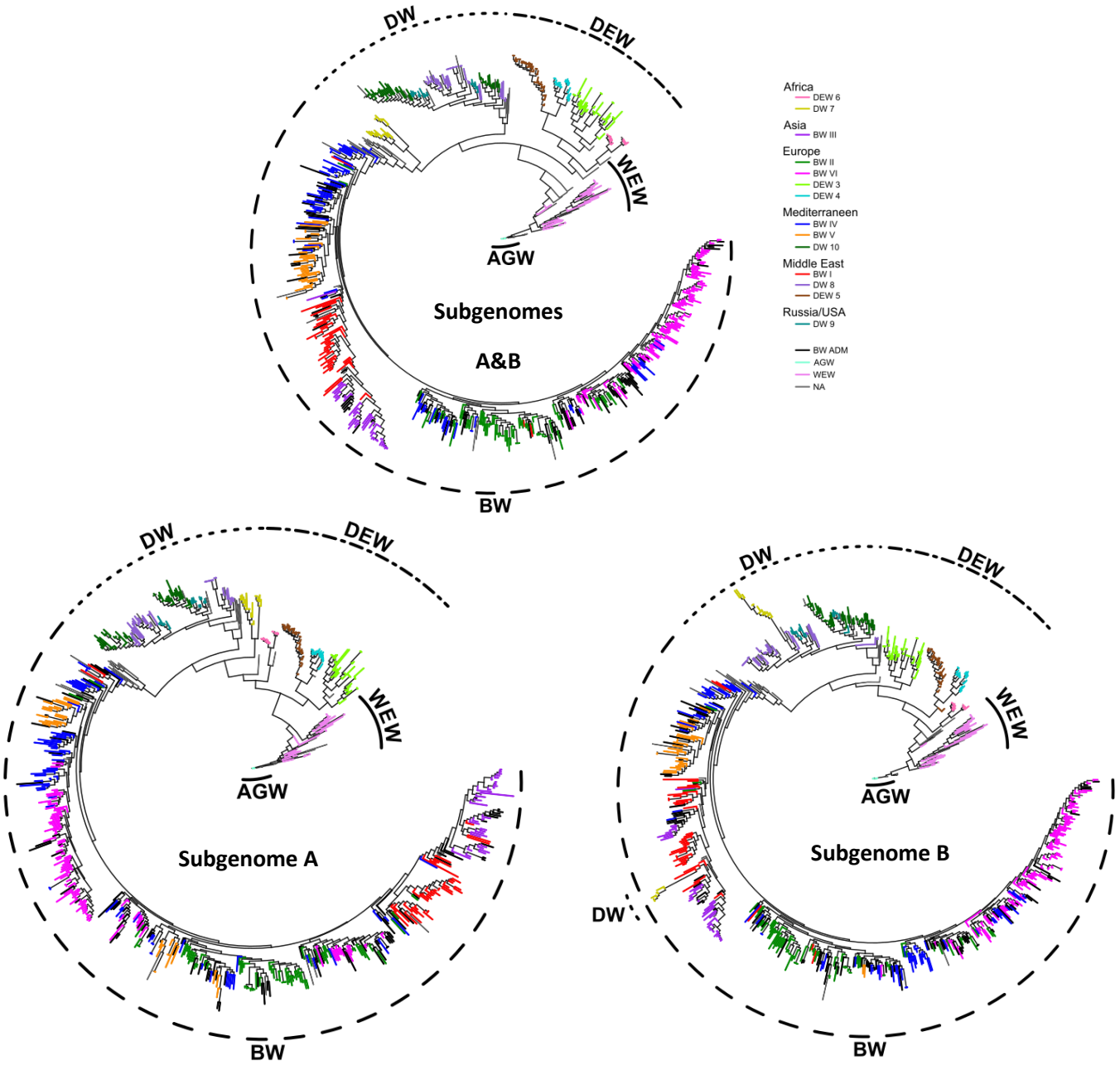


<u>Supplementary Figure 2</u>. Population structures of barley (a), tetraploid (b) and hexaploid (c) wheats, with a color code illustrating subpopulations (as described in the Fig. 2 of the main manuscript) with, for each panel, the phylogenetic tree (left), PCA analysis (top) and structure analysis (bottom).

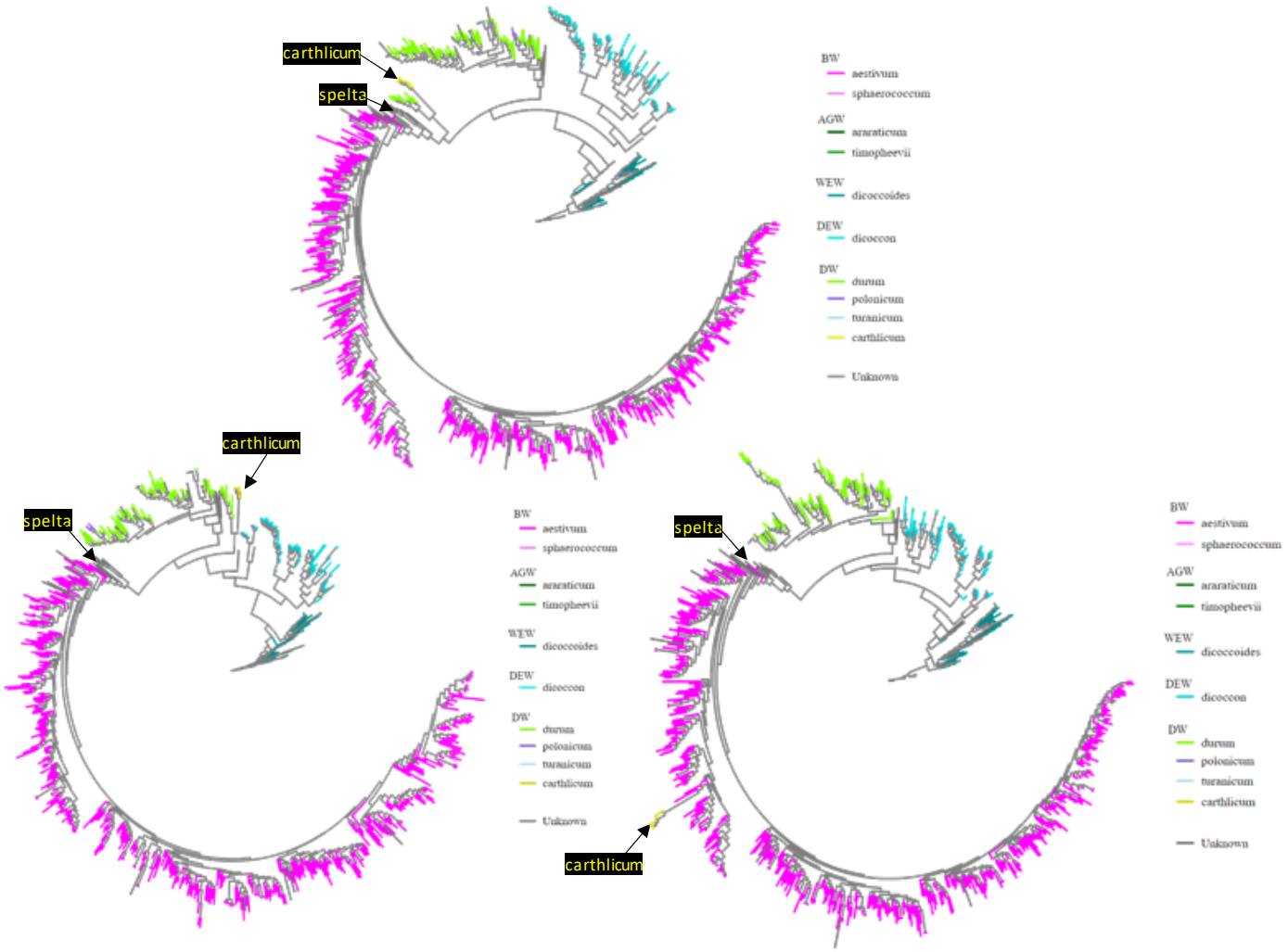


Supplementary Figure 3. Phylogenetic analysis of tetraploid and hexaploid accessions. Phylogenetic tree of the genetic diversity highlighted with a color code (see legend at the right) for **(a)** the population, with (TOP) 738 individuals with 33,905 shared OrthoSNPs between Td (*Triticum durum)* and Ta (*Triticum aestivum)*, (BOTTOM) 738 individuals with 13,430 shared OrthoSNPs between TdA (*Triticum durum* subgenome A) and TaA (*Triticum aestivum* subgenome A) at the left and 737 individuals with 20,029 shared OrthoSNPs between TdB (*Triticum durum* subgenome B) and TaB (*Triticum aestivum* subgenome B) at the right; and (**b**) the taxa, with (TOP) 738 individuals with 33,905 shared OrthoSNPs between Td (*Triticum durum)* and Ta (*Triticum aestivum)*, (BOTTOM) 738 individuals with 13,430 shared OrthoSNPs between TdA (*Triticum durum* subgenome A) and TaA (*Triticum aestivum* subgenome A) at the left and 737 individuals with 20,029 shared OrthoSNPs between TdB (*Triticum durum* subgenome B) and TaB (*Triticum aestivum* subgenome B) at the right.

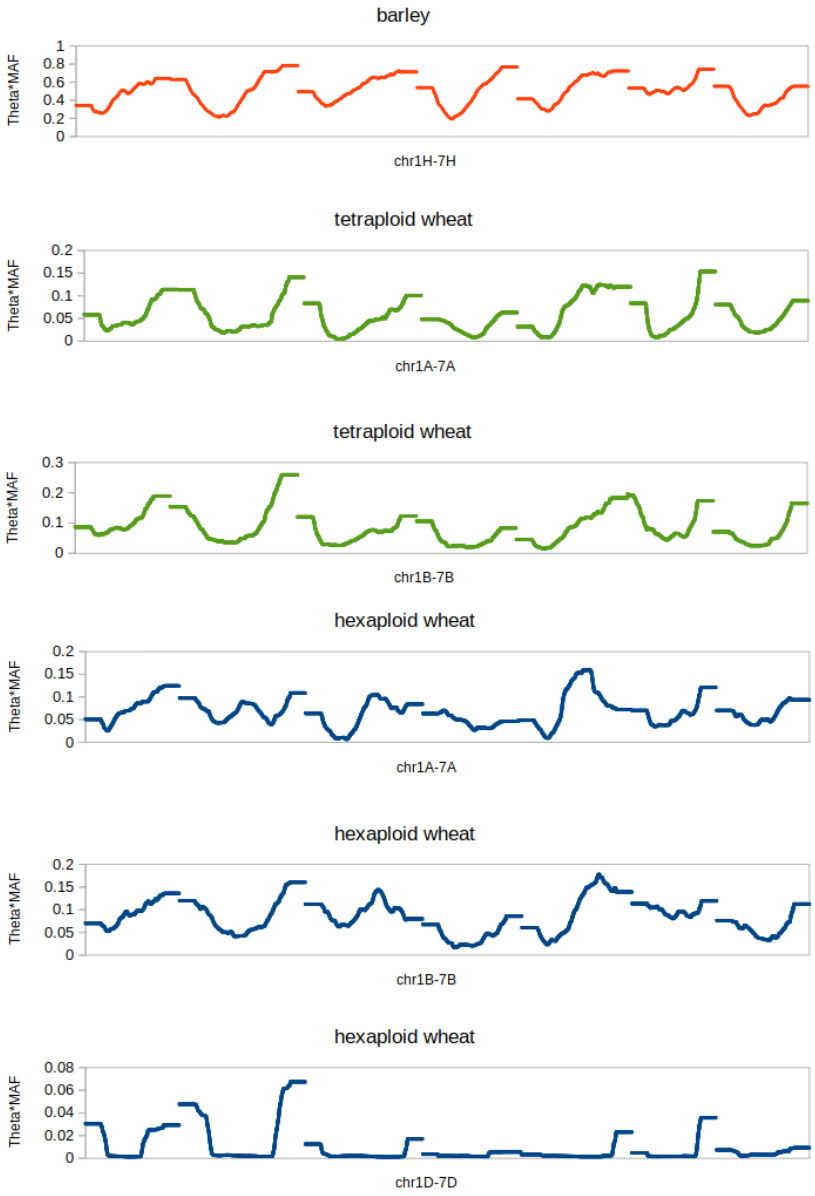
a



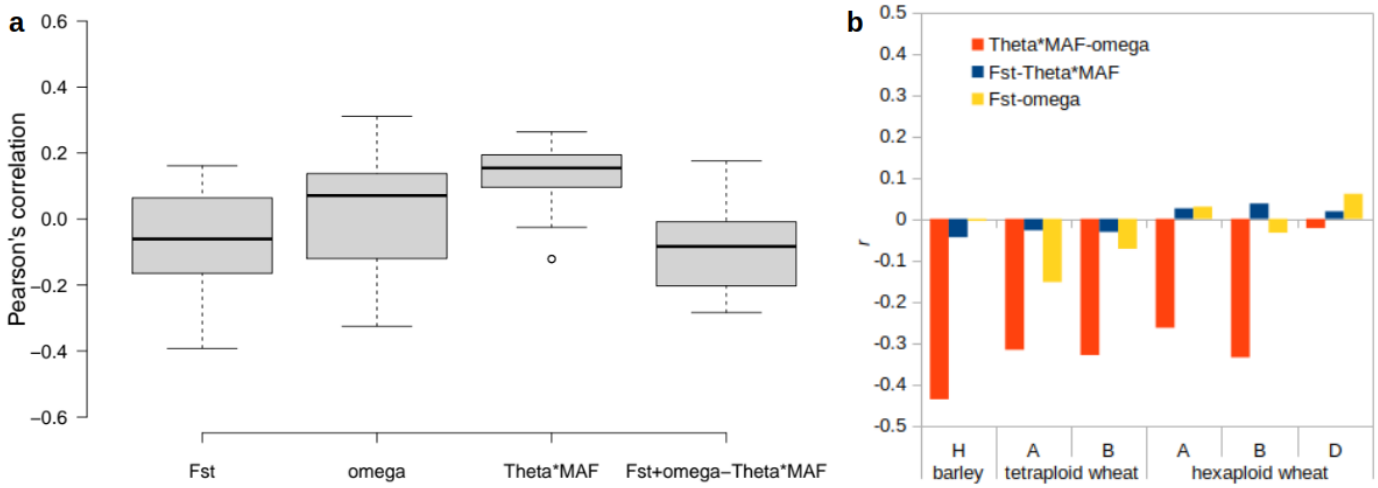
b


carthlicum
spelta
BW
aestivum
sphaerococcum
AGW
araraticum
timopheevii
WEW
dicoccoides
DEW
dicoccon
DW
durum
polonicum
turanicum
carthlicum
Unknown
carthlicum
spelta
BW
aestivum
sphaerococcum
AGW
araraticum
timopheevii
WEW
dicoccoides
DEW
dicoccon
DW
durum
polonicum
turanicum
carthlicum
Unknown
spelta
carthlicum
BW
aestivum
sphaerococcum
AGW
araraticum
timopheevii
WEW
dicoccoides
DEW
dicoccon
DW
durum
polonicum
turanicum
carthlicum
Unknown

<u>Supplementary Figure 4</u>. General pattern of nucleotide diversity measured as ϑ_W*MAF averaged over 500 ordered orthogenes, shown for the seven chromosomes of barley and each of the polyploid wheat subgenomes. These data points were used in the standardisation of ϑ_W*MAF z-scores.

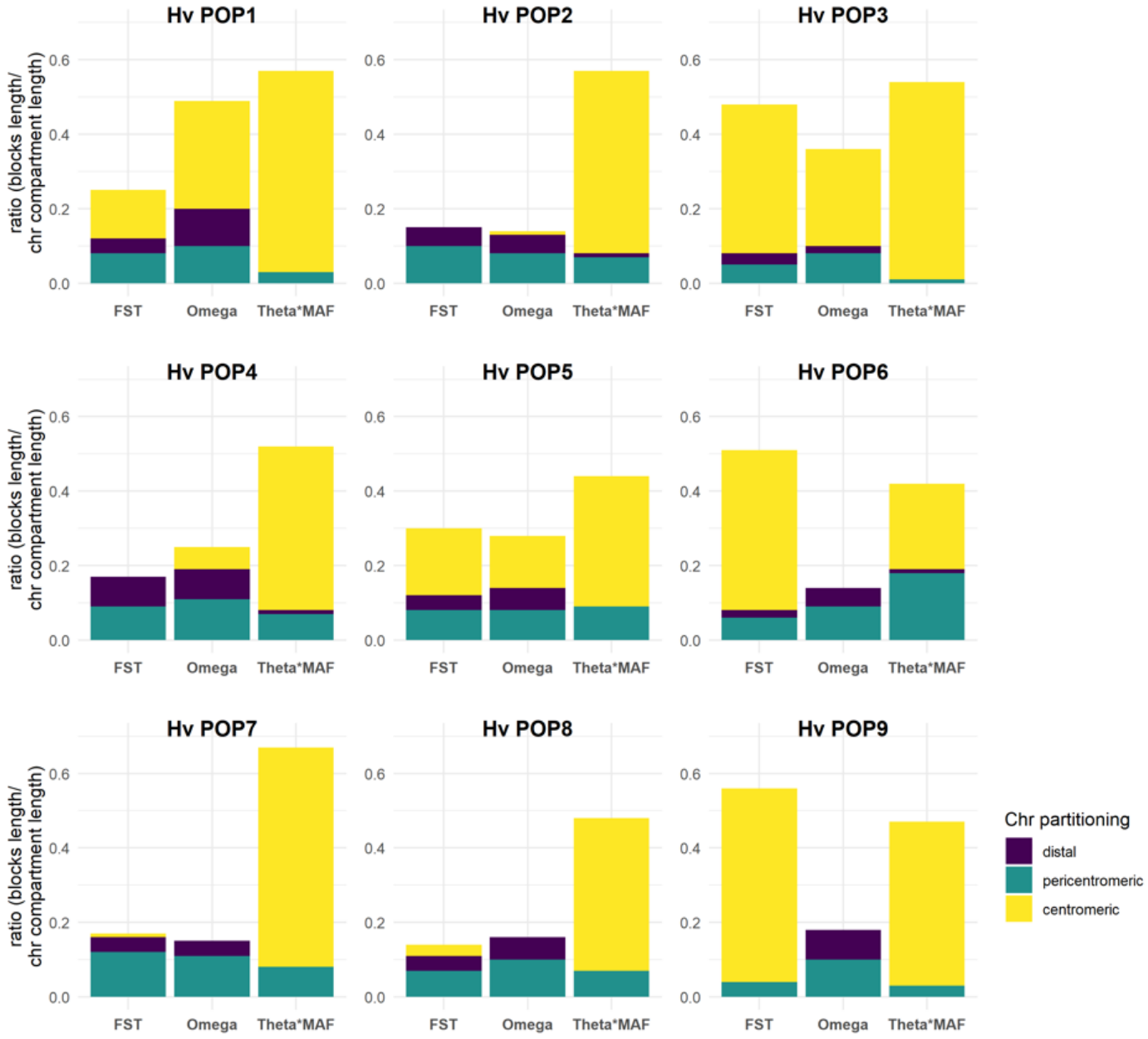


<u>Supplementary Figure 5</u>. Assessment of the chosen selection statistics. (**a**) For each statistic assigned to the 11,925 'perfect' OGs, correlation with the OGs distance from the centromere was calculated. The boxplots show the distribution of the correlation coefficients across all populations and subgenomes. (**b**) Correlations between the selection statistics, showing high negative correlation between omega and ϑ_W*MAF, and inconsistent correlations involving F_{ST} across different species and subgenomes.

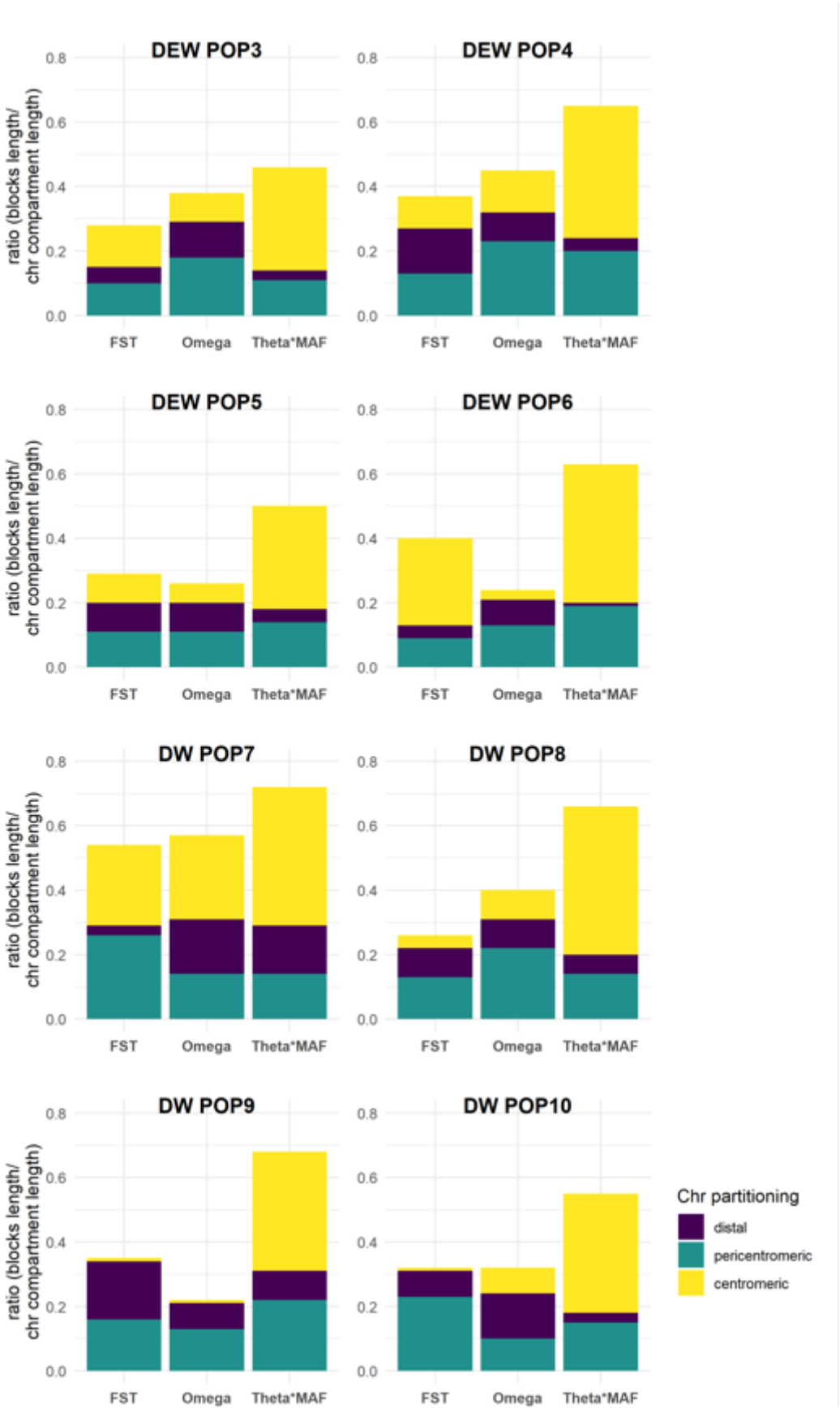


<u>Supplementary Figure 6</u>. Sweep detection between genomic compartments. Bar plots showing the abundance of FST, Omega and ϑ_W*MAF blocks in each barley population in respect to the barley (panel **a**), DW-DEW (panel **b**) and BW (panel **C**) chromosome compartments.

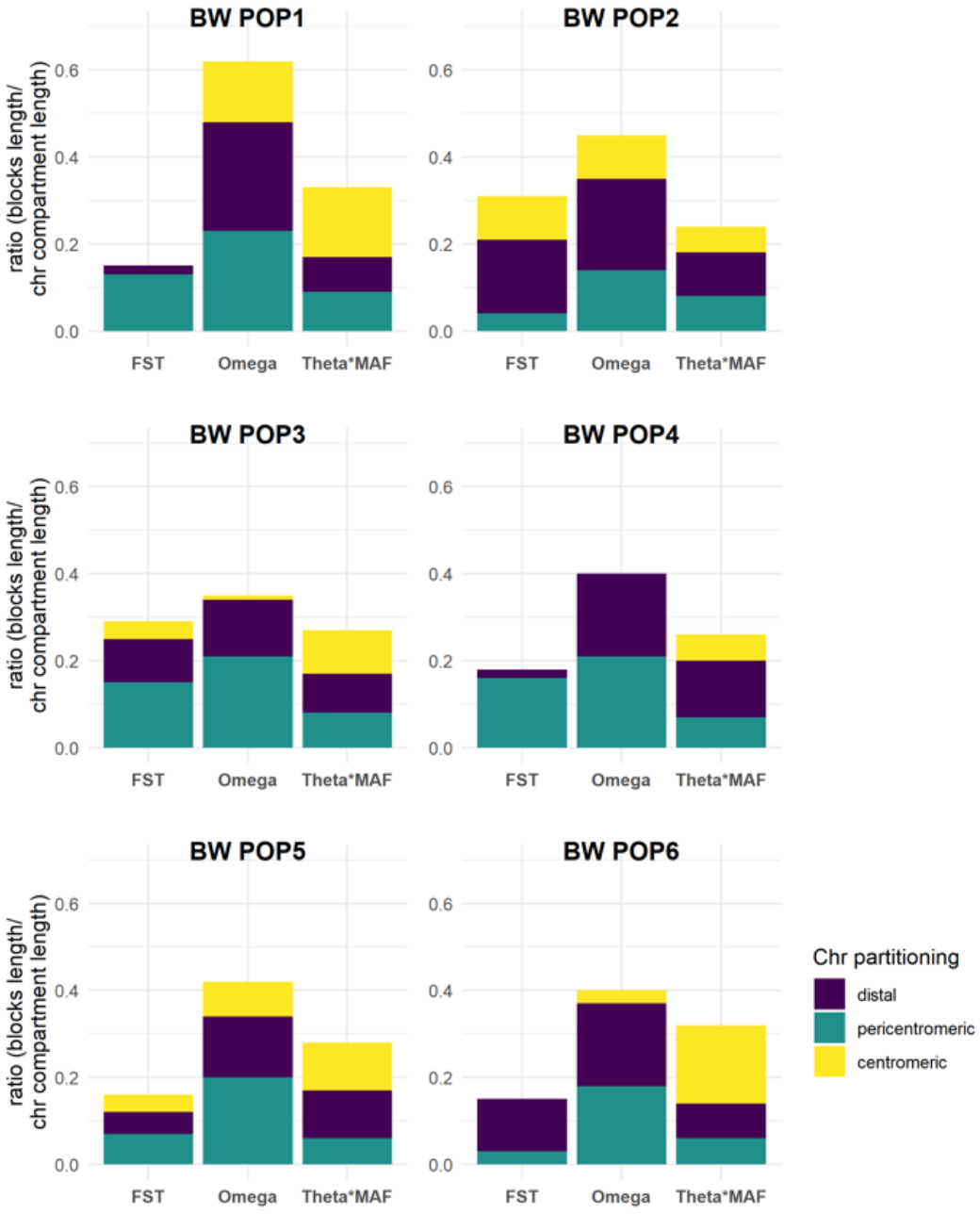
a



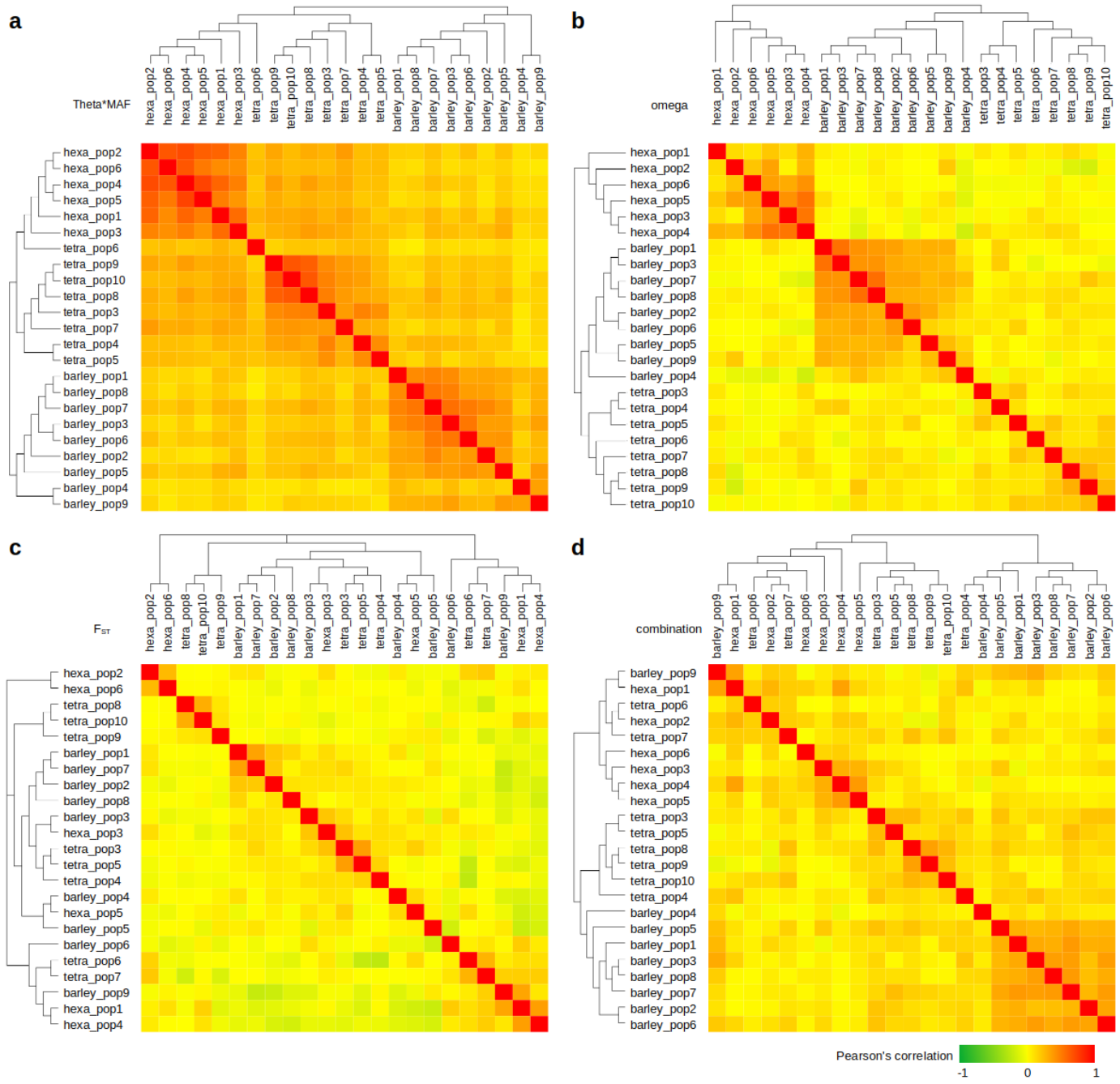
b


DEW POP3
DEW POP4
DEW POP5
DEW POP6
DW POP7
DW POP8
DW POP9
DW POP10
ratio (blocks length/ chr compartment length)
0.0
0.2
0.4
0.6
0.8
FST
Omega
Theta*MAF
Chr partitioning
distal
pericentromeric
centromeric

c

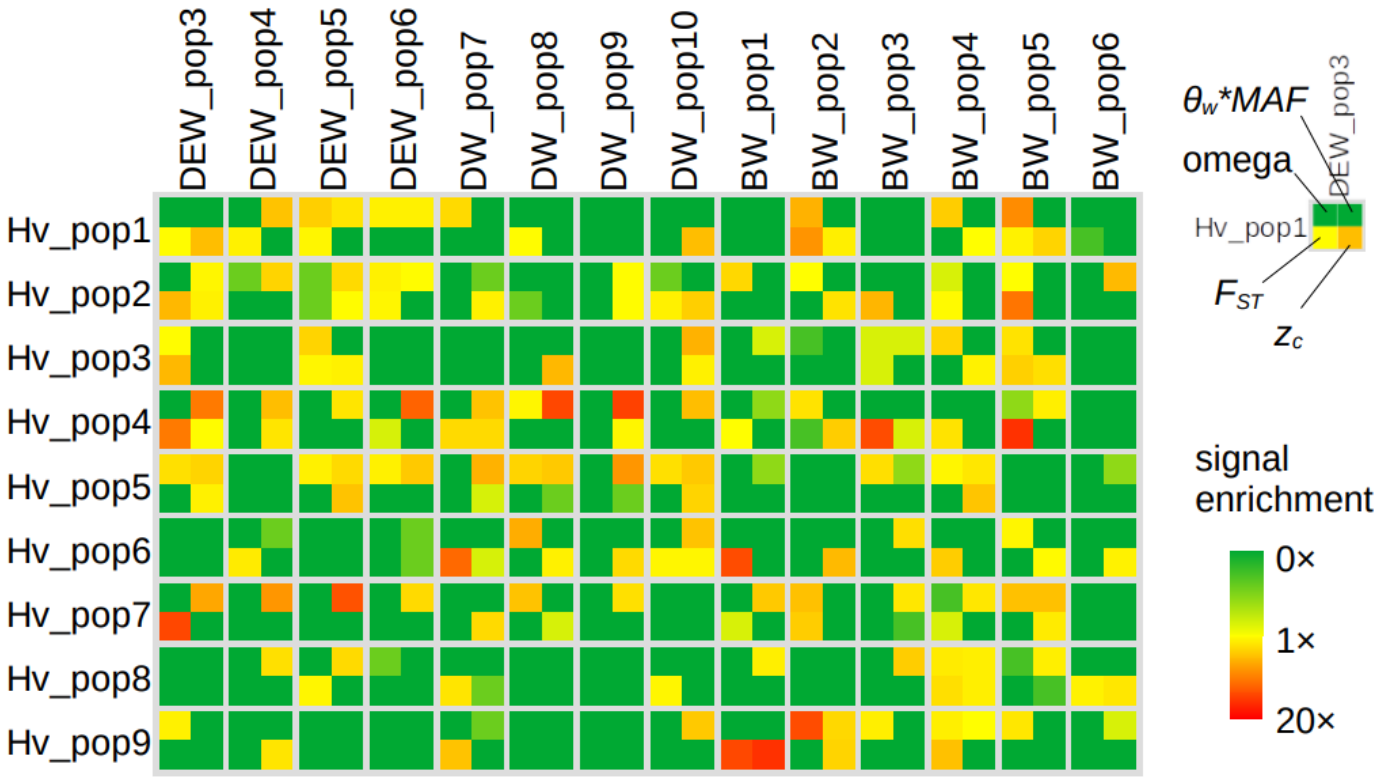

BW POP1
BW POP2
BW POP3
BW POP4
BW POP5
BW POP6
ratio (blocks length/ chr compartment length)
0.0
0.2
0.4
0.6
FST
Omega
Theta*MAF
Chr partitioning
distal
pericentromeric
centromeric

<u>Supplementary Figure 7</u>. Correlation-based clustering of the statistics across all pairwise combinations of populations. The correlation coefficients were calculated across the 11,925 perfect orthogroups, where the maximum of the two/three homeologs was used in the populations of polyploid wheats. (**a**) ϑ_W*MAF; (**b**) omega; (**c**) F_{ST}; (**d**) z_c.

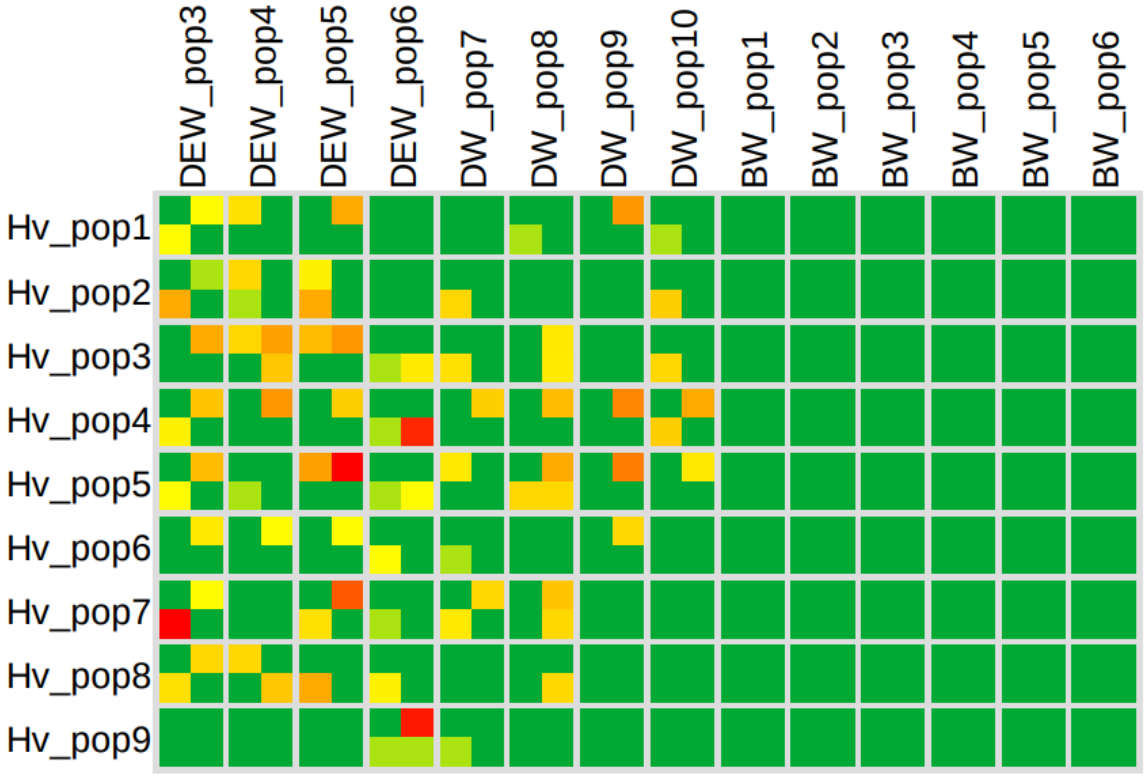


Supplementary Figure 8. OrthoSweep enrichment results. (**a**) Enrichment of the sharing of selection signals between populations of barley and wheat (any subgenome) at the 99th percentile level. Zero signal enrichment (green) means that the two populations do not share any orthogenes with the given metric above the 99th percentile. (**b**) Enrichment of the sharing of selection signals between populations of barley and wheat (all subgenomes) at the 95th percentile level. Note that no OG has a selection signal of this strength on all homeologs of hexaploid wheat and barley simultaneously (any combination of populations, any metric). Enrichment of signal sharing was calculated for each percentile in the range 60-99, and the 99.9th percentile. Genes associated with signal enrichment >20x (at any of the percentile thresholds) are listed in Table S6.

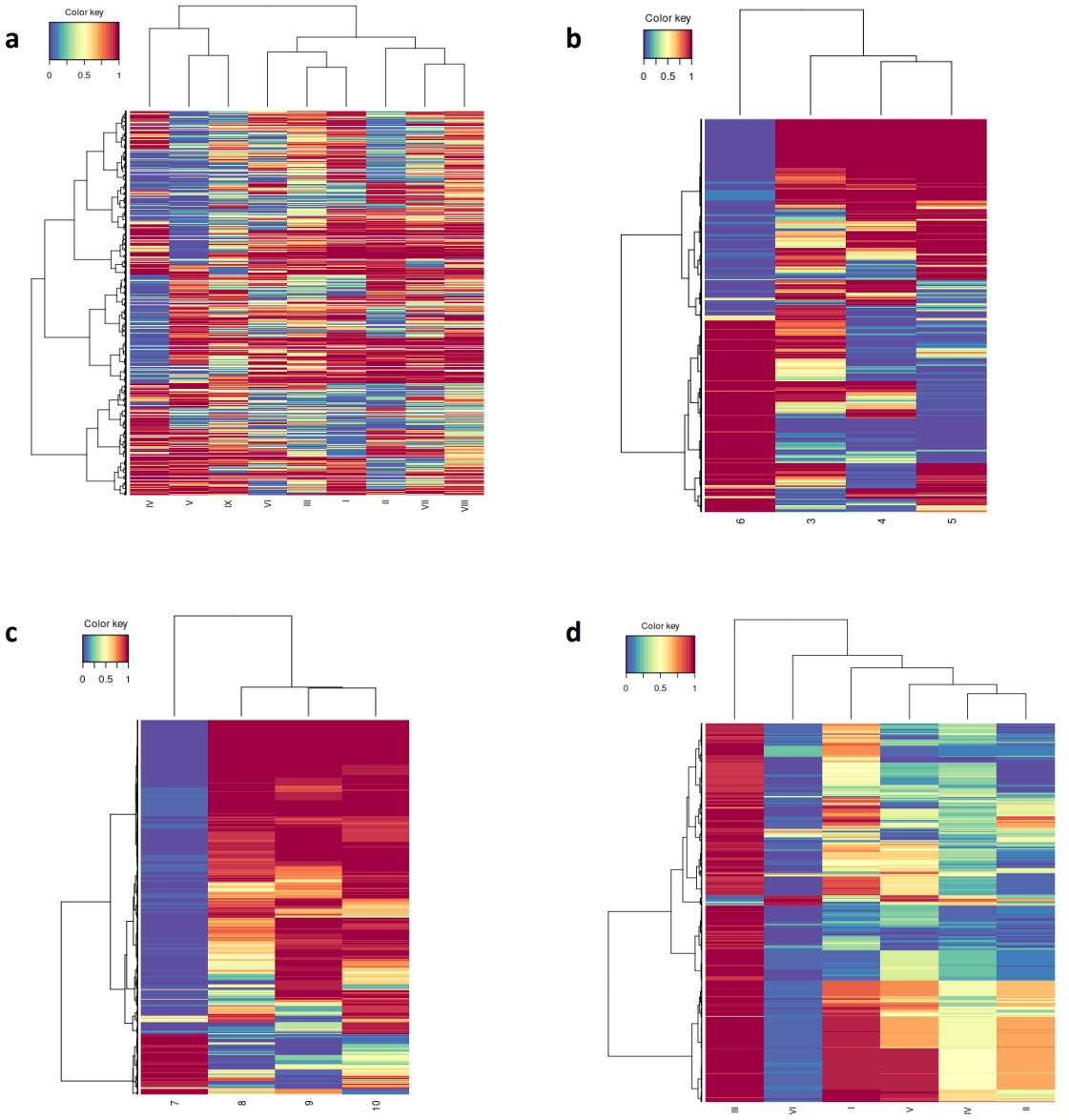
a



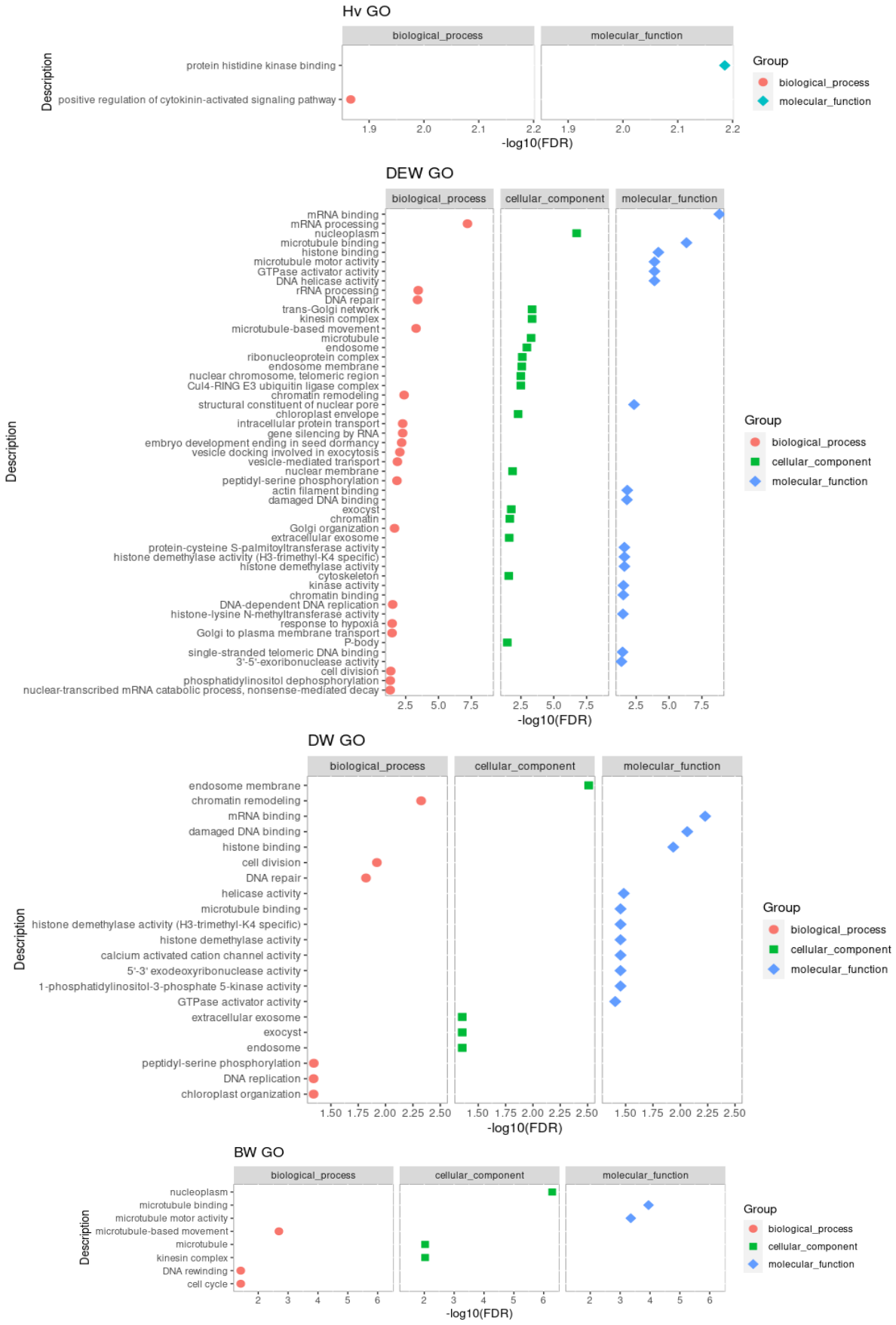
b



<u>Supplementary Figure 9</u>. Differentiated orthoSNPs between domesticated populations in Hv (a), DEW (b), DW (c) and BW (d). Values correspond to REF allele frequency in each crop according to the defined populations. Plots are drawn using mixomics R package with ward hierarchical clustering.



<u>Supplementary Figure 10</u>. Gene ontology enrichment of differentiated orthoSNPs between domesticated populations for biological process and molecular function in Hv (a), DEW (b), DW (c) and BW (d). GO analysis was done using TGT (Triticeae-GeneTribe, Chen et al. 2020). A customized background corresponding to the list of 19250 orthogroups was used and FDR fixed below 0.05.



<u>Supplementary Figure 11.</u> Characterization of *FUWA* and *DRO1* genes in wheat.

a- FUWA mutants produced for A and B homoeologous copies were phenotypically evaluated for ML1 and ML2 plants for flowering time and leave length- width compared to BC1F2 and Svevo plants. **b**- Haplotype network for 4x wheat (panel i) and 6x wheat (panel ii) *FUWA* orthologs on chromosome 6A (*FUWA-A)*. Each circle represents a haplotype (H), with the pie color depicting its diffusion in each sub-population identified within the species. The sizes of the circles indicate the number of accessions carrying the specific haplotype. The dots on line between haplotypes represent number of variants. Cross-species multiple haplotype alignment (panel iii) and NJ tree pointing at the relationship between alleles in the two species. Stars point to the orthoSNPs identified. **c**- Haplotype network for 4x wheat (panel i) and 6x wheat (panel ii) *DRO1* orthologs on chromosome 5B (*DRO1-B)*. Integration of genotype data and phenotypic variation for Root Growth Angle (RGA) in tetraploid wheats indicate that 4x_H002 is associated with significantly lower RGA than other alleles (panel iv). T-test statistical significance *: p <= 0.05, **: p <= 0.01.

a

Mutants' genotype

Genome editing experiments led to the creation of the following alleles:

TdFUWA-A_m1: Inversion of the genomic region comprised between bp 253 and 1306 downstream the ATG.

TdFUWA-B_m1: Deletion of 394 bp located 250 bp downstream the ATG and deletion of 27 bp located 2300 bp downstream the ATG.

TdFUWA-B_m2: Inversion of the genomic region comprised between bp 252 and 1313 downstream the ATG.

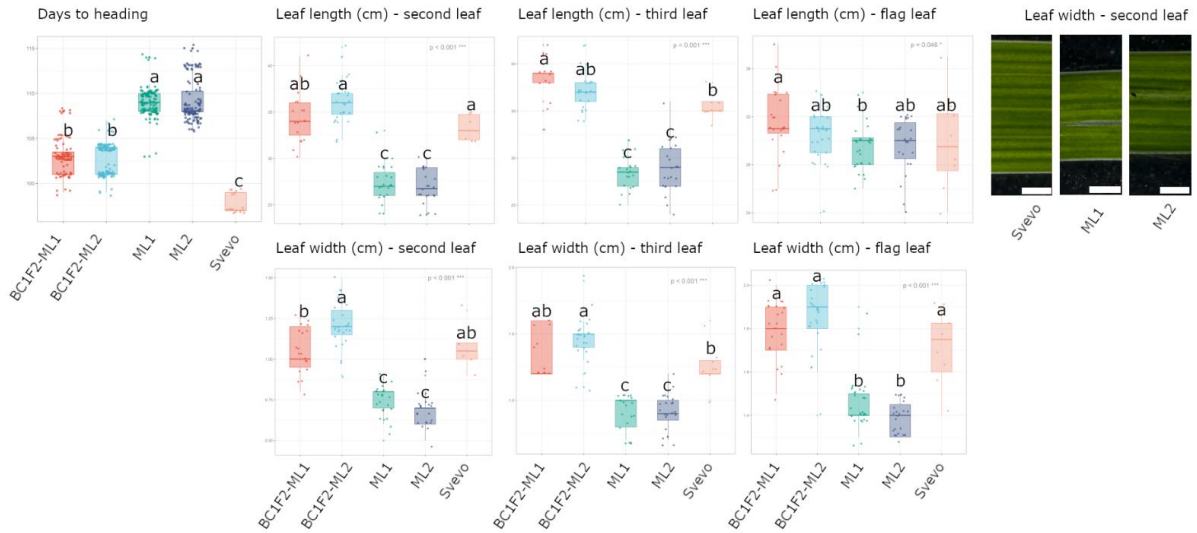
Mutant line ML1 is homozygous for the *Td-FUWA-A_m1* and *Td-FUWA-B_m1* alleles, mutant line ML2 is homozygous for the *Td-FUWA-A_m1* and *Td-FUWA-B_m2* alleles.

gRNAs and primers used in this study

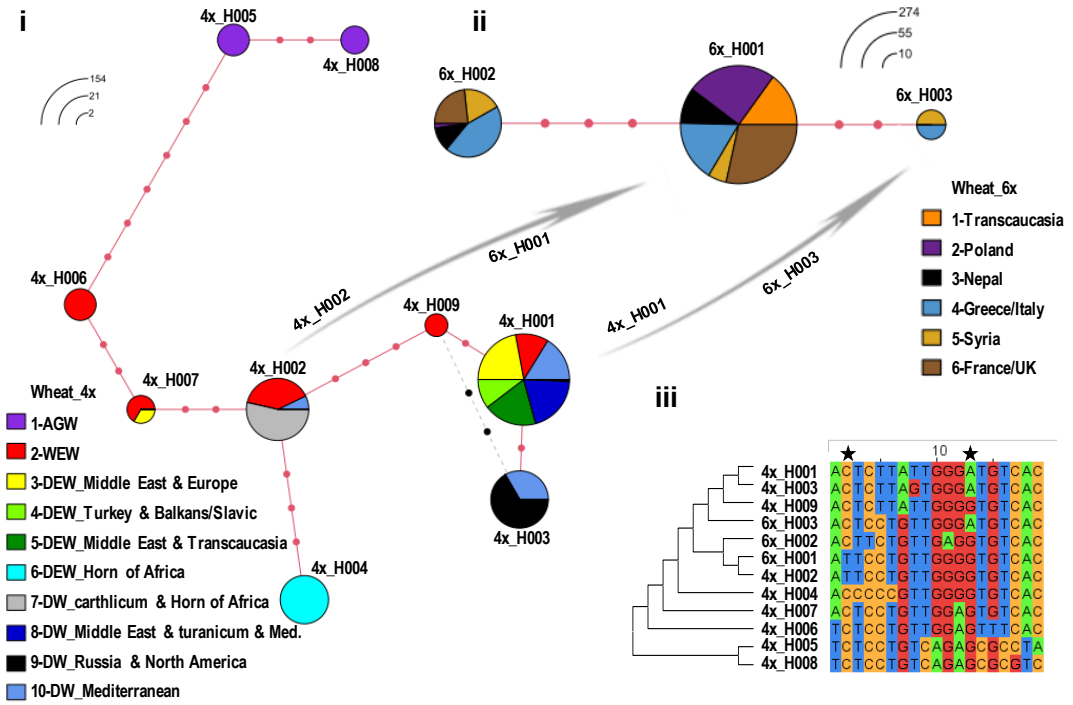
gRNA1	5'-CATCGTGAGGCAGCTATCGT-3'	
gRNA2	5'- GAGTGGGGGGGAGAACCGCG-3'	
gRNA3	5'-GTGTCACGCCCCACACCATT-3'	
gRNA4	5'-AAGCTCTCCACCCTCAACTG-3'	
TdFUWA-A for	5'-CAAACAGTCATCGAAGAATAAAATGATCG-3'	Primers used for genotyping wild-type and mutant alleles
TdFUWA-B for	5'-AGTTTTAGTACCTCAGTATTTTTTTAGGG-3'	
TdFUWA rev	5'-AGGTATTTCGTACTTCCAT-3'	
TdFUWA-A/B_m rev	5'-GTCCTCATTCTAGGCCTGACC-3'	
TdFUWA-B_m1 rev	5'-GAACTGTGGGCAGCAAGAGAAAAAC-3'	

Phenotypic data

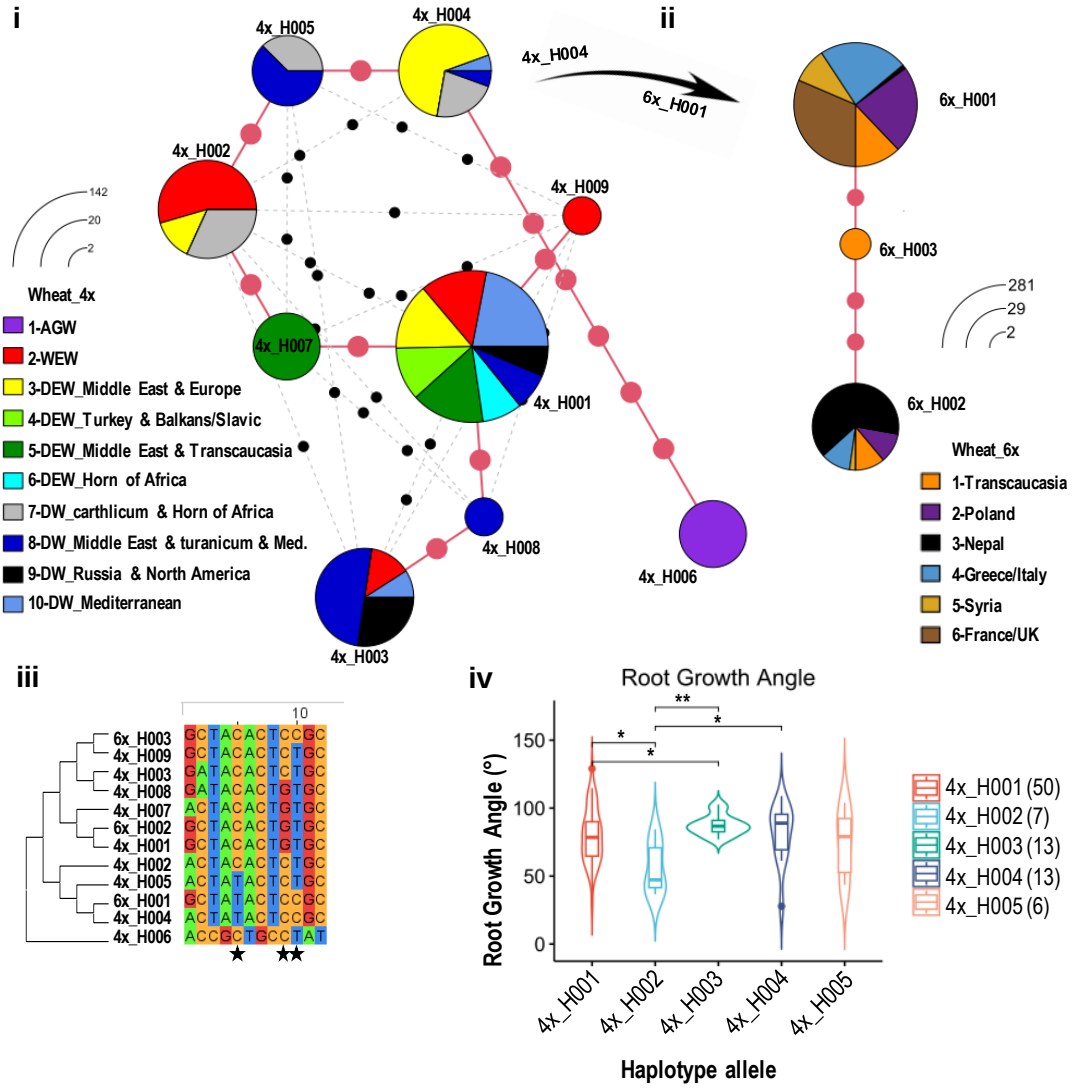
ML1 and ML2 plants were characterized by late flowering time compared to BC1F2 and Svevo plants, shorter and narrower leaves (measures in cm of second and third leaves at flowering are reported). Microscopy image of second leaf width in Svevo, ML1 and ML2 plants. The proximal region of second leaf of 4 weeks-old plants was photographed. Scale bar corresponds to 500 µm.



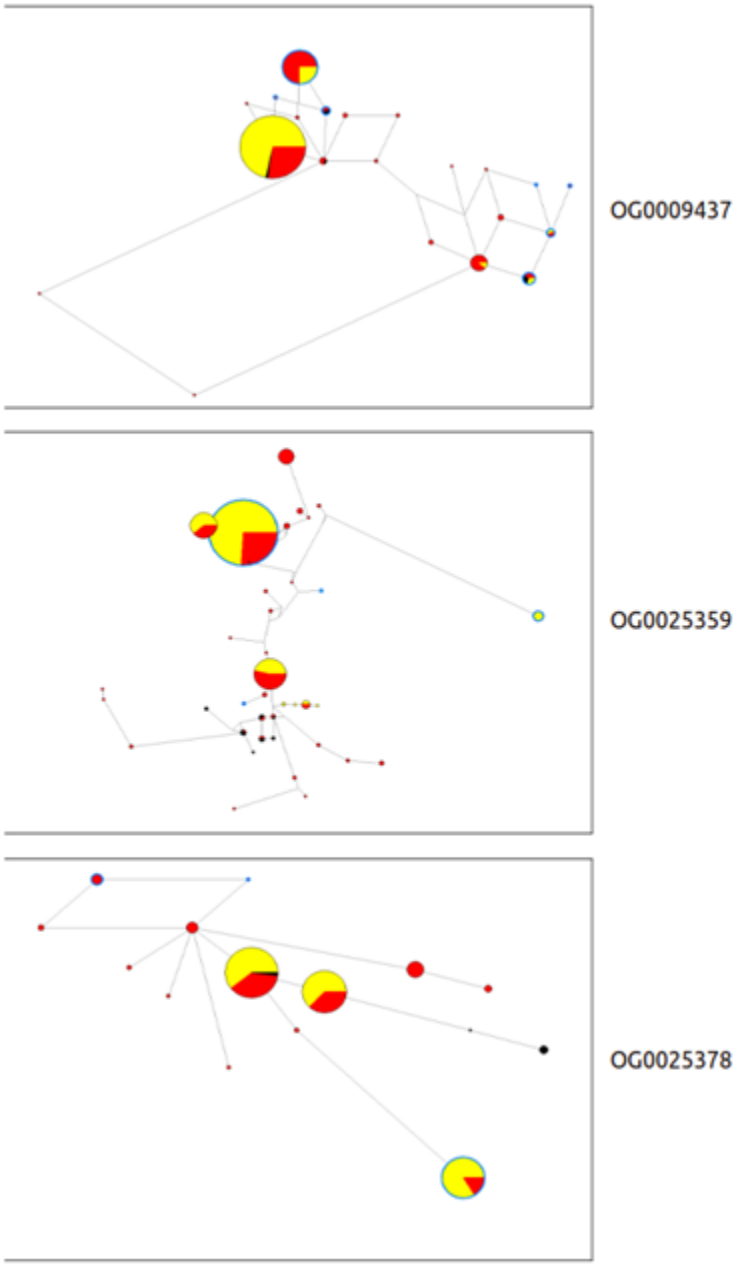
b


i
4x_H005
4x_H008
154
21
2
4x_H006
4x_H007
4x_H002
4x_H009
4x_H001
4x_H003
4x_H004
Wheat_4x
1-AGW
2-WEW
3-DEW_Middle East & Europe
4-DEW_Turkey & Balkans/Slavic
5-DEW_Middle East & Transcaucasia
6-DEW_Horn of Africa
7-DW_carthlicum & Horn of Africa
8-DW_Middle East & turanicum & Med.
9-DW_Russia & North America
10-DW_Mediterranean
ii
6x_H002
6x_H001
6x_H003
274
55
10
4x_H002
6x_H001
4x_H001
6x_H003
Wheat_6x
1-Transcaucasia
2-Poland
3-Nepal
4-Greece/Italy
5-Syria
6-France/UK
iii
10
4x_H001
4x_H003
4x_H009
6x_H003
6x_H002
6x_H001
4x_H002
4x_H004
4x_H007
4x_H006
4x_H005
4x_H008

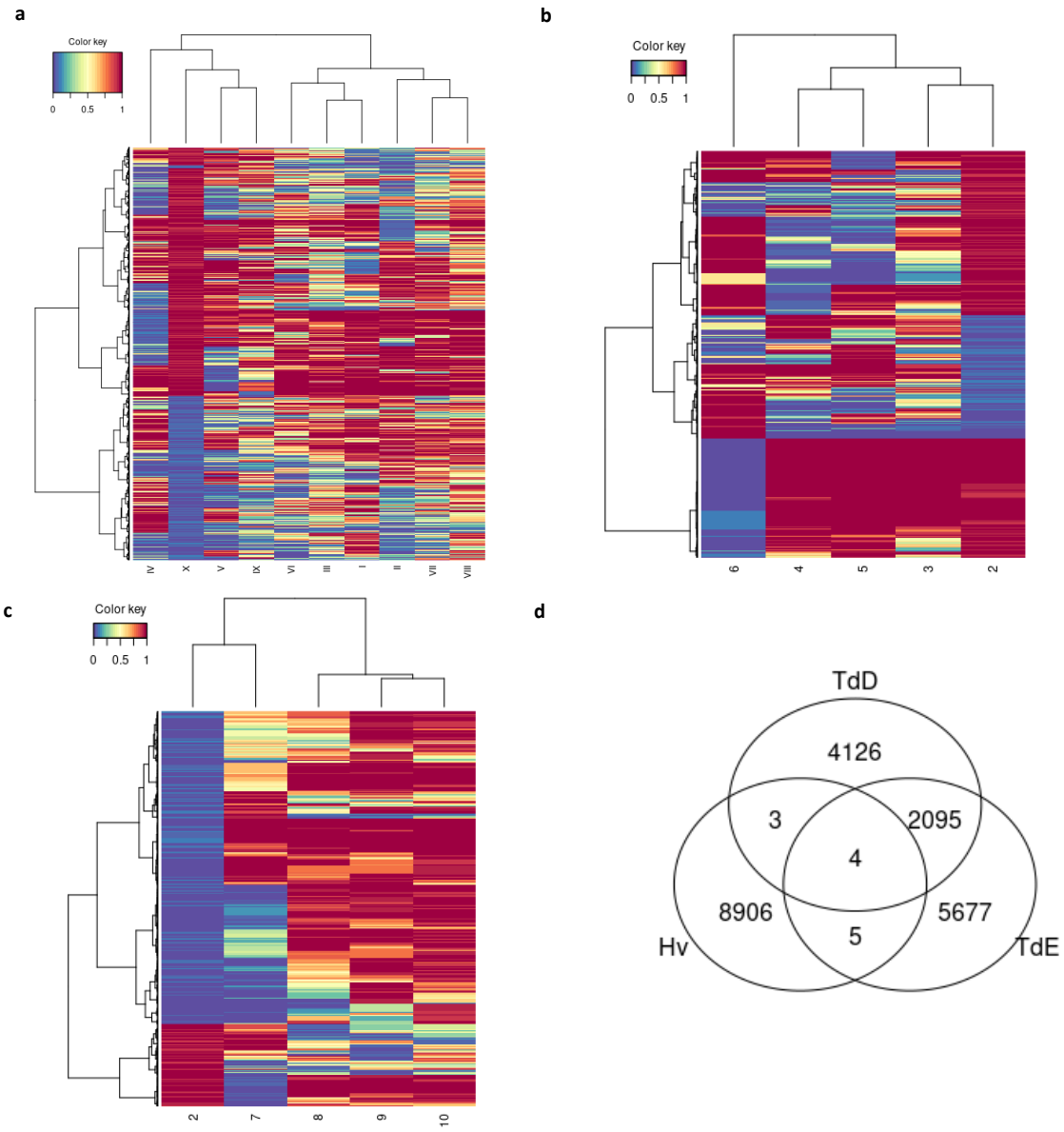
c


i
4x_H005
4x_H004
4x_H002
4x_H009
4x_H007
4x_H001
4x_H008
4x_H006
4x_H003
142
20
2
Wheat_4x
1-AGW
2-WEW
3-DEW_Middle East & Europe
4-DEW_Turkey & Balkans/Slavic
5-DEW_Middle East & Transcaucasia
6-DEW_Horn of Africa
7-DW_carthlicum & Horn of Africa
8-DW_Middle East & turanicum & Med.
9-DW_Russia & North America
10-DW_Mediterranean
ii
4x_H004
6x_H001
6x_H001
6x_H003
281
29
2
6x_H002
Wheat_6x
1-Transcaucasia
2-Poland
3-Nepal
4-Greece/Italy
5-Syria
6-France/UK
iii
10
6x_H003
4x_H009
4x_H003
4x_H008
4x_H007
6x_H002
4x_H001
4x_H002
4x_H005
6x_H001
4x_H004
4x_H006
iv
Root Growth Angle
Root Growth Angle (°)
150
100
50
0
4x_H001
4x_H002
4x_H003
4x_H004
4x_H005
Haplotype allele
4x_H001 (50)
4x_H002 (7)
4x_H003 (13)
4x_H004 (13)
4x_H005 (6)

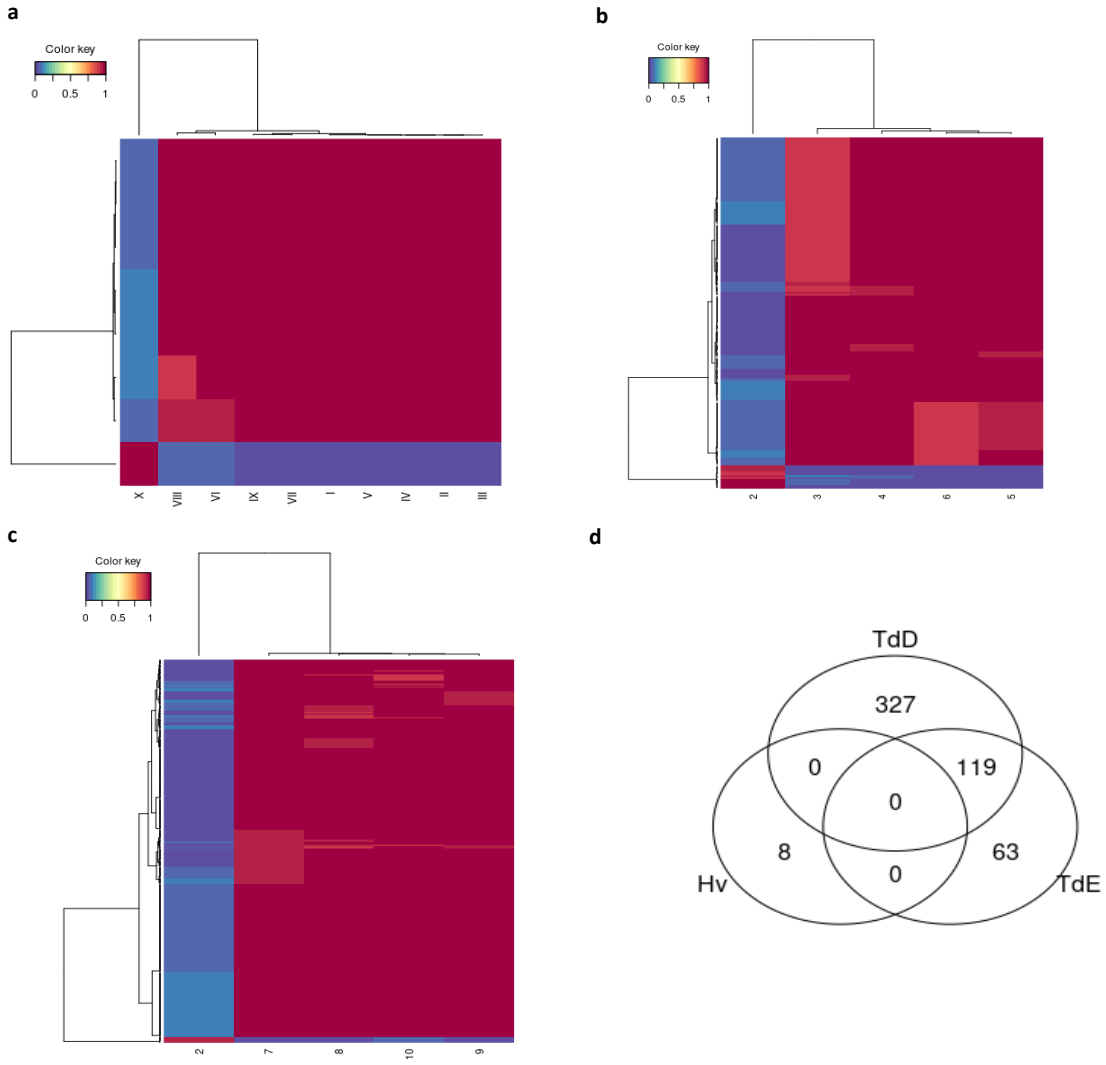
<u>Supplementary Figure 12</u>. Homoplasic concomitantly differentiated (CD-)orthoSNPs between tetra- and hexa-ploid wheats. OrthoSNPs and reference sequences for tetra- and hexa-ploid wheats were used to reconstruct individual haplotypes of all orthogenes containing one or more CD-orthoSNPs. These haplotypes were then used to construct multi-species gene-trees and median-joining networks (Bandelt et al. 1999), which were visually checked for cases of homoplasy. Homoplasies found in multiple species can be regarded as evolutionary convergence on a nucleotide level (if they are non-neutral). Mediant-joining networks show for three OGs, OG0009437 (a), OG0025359 (b) and OG0025378 (c), each of which encodes a putative disease resistance protein. Each network node corresponds to a haplotype, with node sizes proportional to the number of samples and edge lengths proportional to the number of substitutions separating the nodes. Tetraploid and hexaploid wheats are represented by red and yellow color, respectively. Haplotypes containing identical variants implied to be homoplasies by the network topology are indicated with blue circle lines. The haplotype networks do not indicate whether the homoplasies emerged in tetra- and hexaploid wheats separately. The variants could have occurred two or more times in tetraploid wheats, with each haplotype being passed to hexaploid wheats.



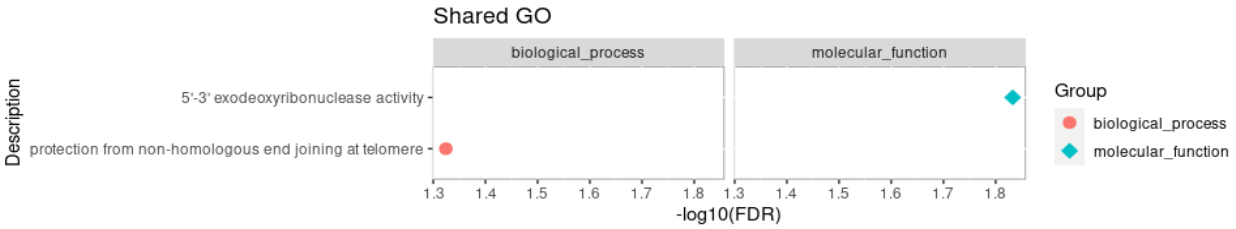
<u>Supplementary Figure 13</u>. Concomitantly differentiated (CD-)orthoSNPs between the wild and at least one domesticated population in Hv (a), DEW (b) and DW (c) and shared between crops (d). Heatmap values correspond to REF allele frequency of genes (OGs, y-axis) in each crop according to the defined populations (x-axis). Heatmaps are drawn using mixomics R package with ward hierarchical clustering (with, 'TdD' = 'DW' and 'TdE' = 'DEW').



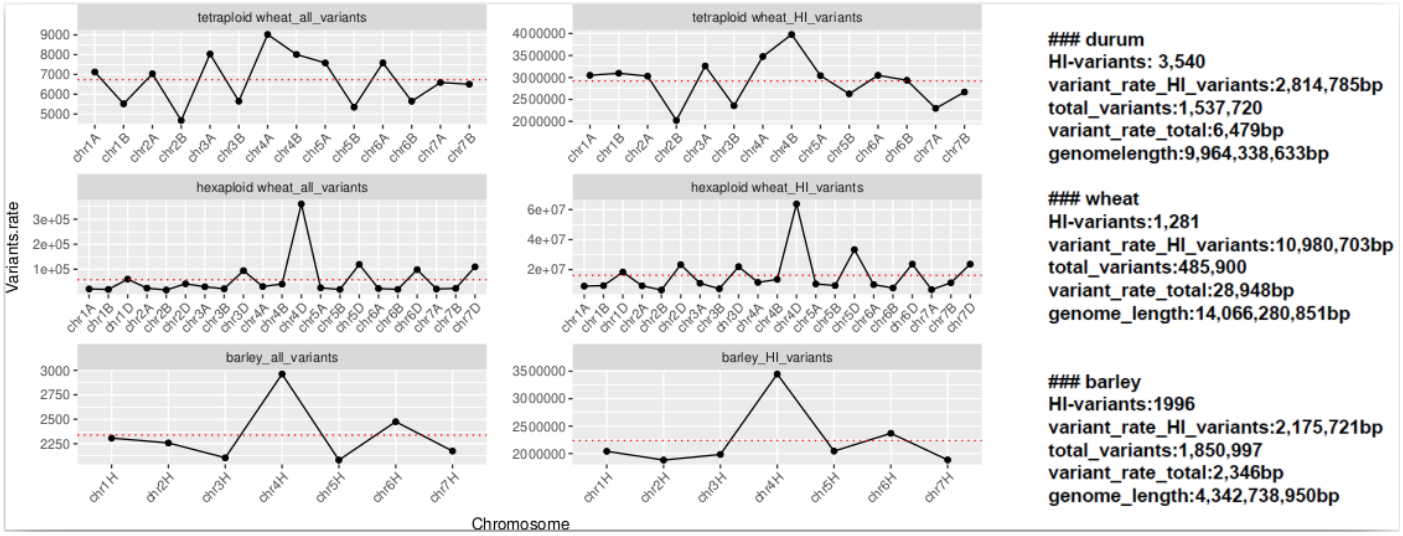
<u>Supplementary Figure 14</u>. Concomitantly differentiated (CD-)orthoSNPs between the wild and all domesticated populations in Hv (a), DEW (b) and DW (c) and shared between crops (d). Heatmap values correspond to REF allele frequency of genes (OGs, y-axis) in each crop according to the defined populations (x-axis). Heatmaps are drawn using mixomics R package with ward hierarchical clustering (with, 'TdD' = 'DW' and 'TdE' = 'DEW').



<u>Supplementary Figure 15</u>. Gene ontology enrichment of concomitantly differentiated (CD-)orthoSNPs between wild and domesticated populations. GO analysis was done using TGT (Triticeae-GeneTribe, Chen et al. 2020). A customized background corresponding to the list of 19250 orthogroups was used and FDR fixed below 0.05.



<u>Supplementary Figure 16</u>. Chromosome distribution of HI-impact variants. Illustration of the number (y-axis) of identified variants (left) and HI-impact variant (right) on chromosomes (x-axis) for tetraploid wheat (top), hexaploid wheat (center) and barley (bottom).

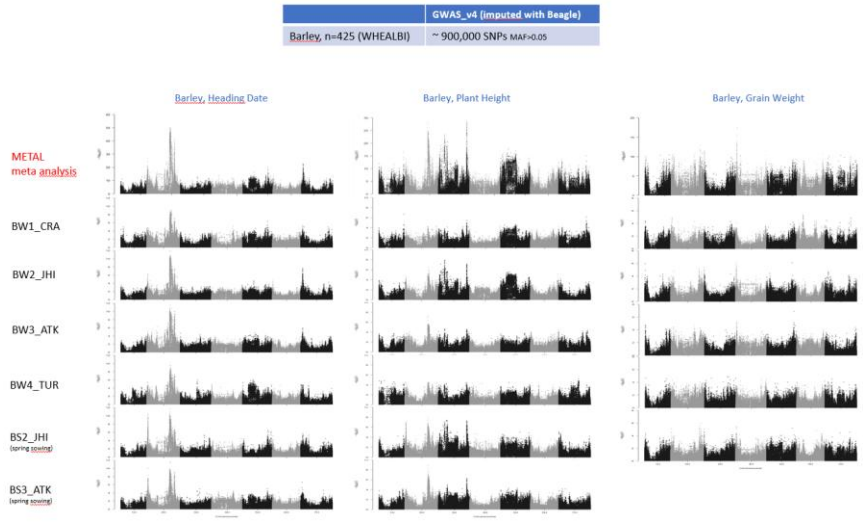


<u>Supplementary Figure 17</u>. Comparative GWAS between barley and bread what. **a**. For each species, GWAS were run for each trait in each environment, followed by a meta-analysis with METAL to find the main genetic determinants across environments (no GxE). The 99.5% distribution of the p-value from the METAL was set as a first threshold for defining significant SNPs (low confidence). METAL results were then subjected to correction (Genomic Control) and the -log10 (P)>4 was set as a second threshold for defining significant SNPs (high confidence). The table below shows a summary of the number of SNPs identified with the two thresholds. Please notice that for HD_Barley the number of high_conf SNPs is higher that low_conf SNPs. **b**. Meta-Analysis of GWA for HD (Heading Date), PH (Plant Height), TWK (Thousand Kernel Weight) in barley. **c**. Meta-Analysis of GWA for HD, PH, TKW in bread wheat.

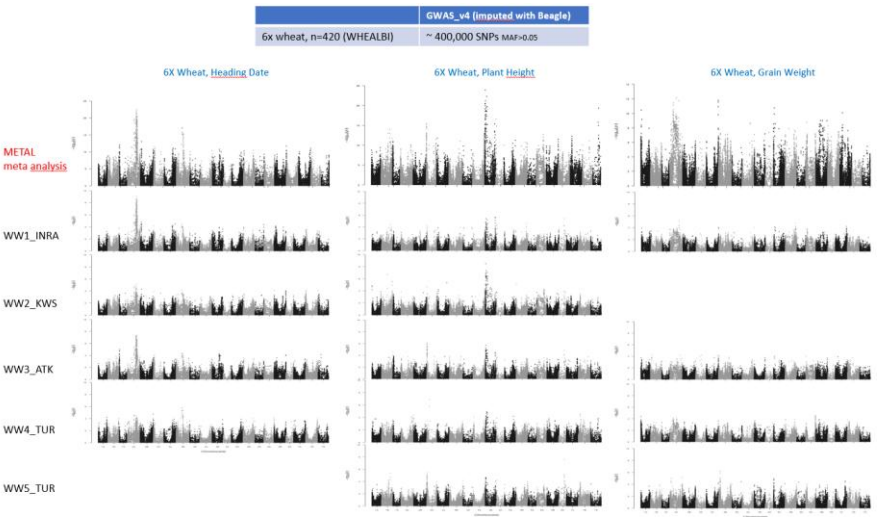
a

		Low_conf: Metal, 0.5% p-value			High_conf: Corrected Metal p-value, LOD>4		
Trait	**crop**	**N° of SNPs**	**MIN -log10(P)**	**MAX -log10(P)**	**N° SNPs**	**MIN -log10(P)**	**MAX -log10(P)**
PH	barley	4734	12,1	28,4	**2296**	4,0	7,5
PH	wheat	2039	8,3	23,9	**89**	4,0	7,0
HD	barley	4589	16,2	50,1	**5076**	4,0	12,7
HD	wheat	2011	9,5	22,4	**1378**	4,0	7,2
TKW	barley	4634	5,9	17,4	**133**	4,0	6,3
TKW	wheat	2107	7,8	12,1	**1214**	4,0	5,8

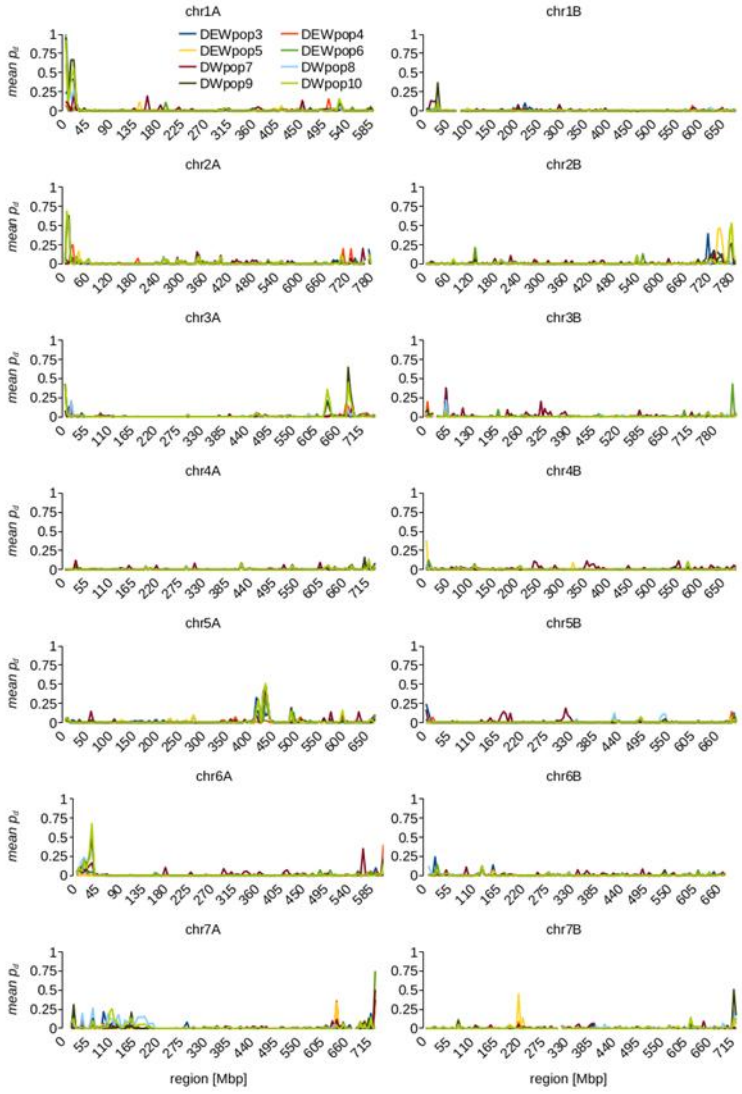
b



c

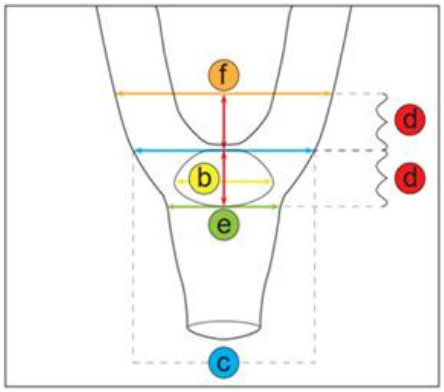


<u>Supplementary Figure 18</u>. GA introgression in DEW and DW populations. 'GA diagnostic' markers corresponding to variants that are absent in WEW (REF allele frequency 0) and fixed in GA (REF allele frequency 1) were tracked in DEW and DW populations on each chromosome. Mean p_d= mean allelic frequencies calculated for 5 Mbp windows. Average allelic frequencies of GA-specific variants (mean *pd*) calculated for each DEW and DW population in 5Mb genomic windows on all 14 chromosomes of tetraploid wheat. True introgressions are expected to result in clustering of the GA-specific variants, which would be reflected in the mean *pd*. A value above the noise threshold but smaller than 1 then indicates that a GA introgression is found only in a part of the given population, or that it is shorter than 5Mb.

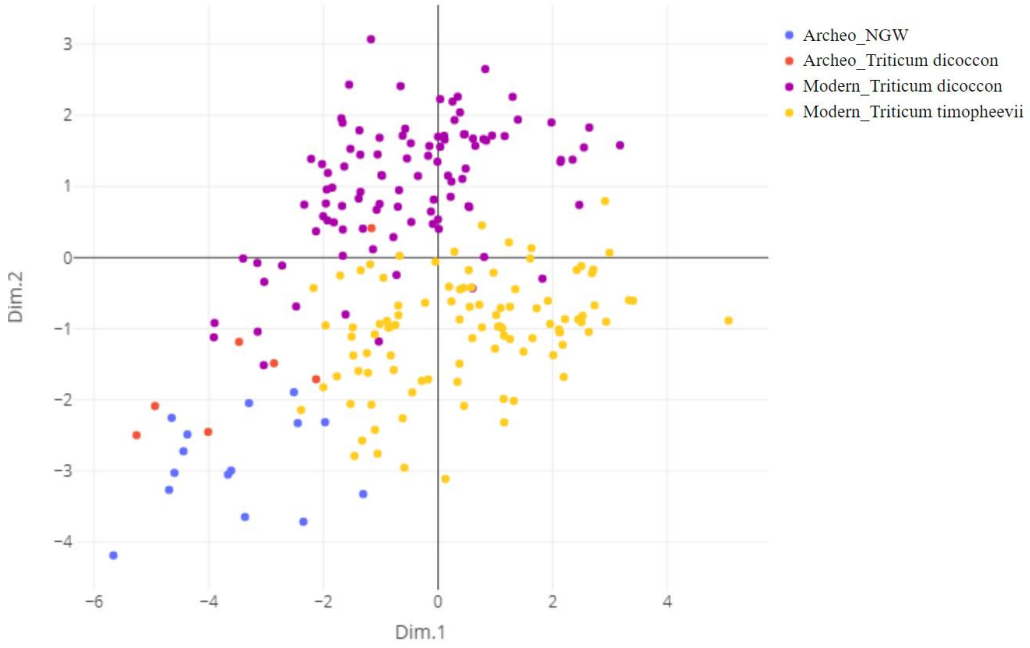


<u>Supplementary Figure 19</u>. **Morphometry analysis of ancient wheat remain. a.** Measurements used to conduct a morphometric analysis on modern and ancient spikelet basis. Measurements performed on the wheat spikelets using Adobe Illustrator™: 'b' width of the disarticulation scar (Kohler-Schneider 2003); 'c' width of the spikelet base (ibid.); 'd' height of the disarticulation scar; 'e' width of the spikelet base, measured along the bottom of the disarticulation scar; 'f' width across both glumes as seen from the ventral surface (height from which f is measured corresponds to length d, replicated above the upper edge of the disarticulation scar). **b.** Principal components analysis (PCA) showcasing the proximity between the four main groups studied, before normalization of the measurements: carbonized archaeological NGW (blue, number of individuals (N) = 15), carbonized archaeological *T. dicoccon* (red, N = 7), modern accessions of *T. dicoccon* (purple, N = 100) and *T. timopheevii* (yellow, N = 99).

a

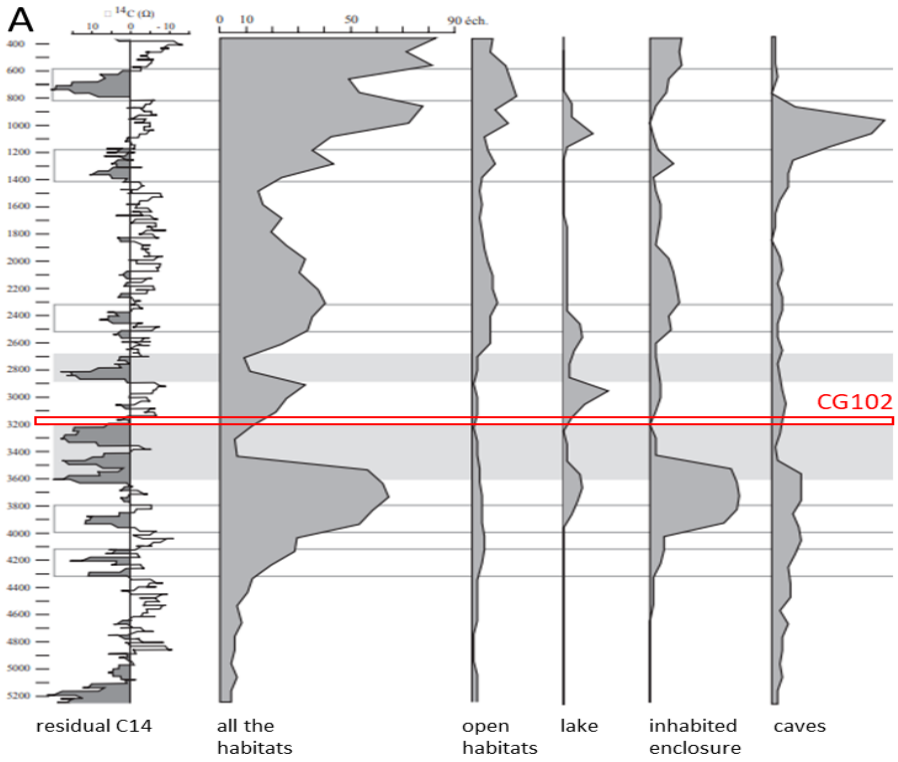


b

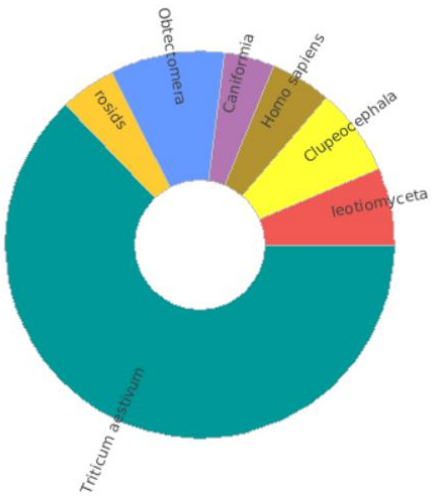


<u>Supplementary Figure 20.</u> aDNA analysis of ancient wheat remain. a Comparison of the residual C14 curve in the atmosphere (indirect indicator of climate variations) and the curve of archaeological settlement indices in the lake of Chalain area, for all the habitats, open habitats, lake habitats, inhabited enclosure and caves. Horizontal red rectangle shows the investigated CG102 layer (adapted from Pétrequin et al. 2005). **b** Taxonomical diversity of the sequenced DNA CG102 by mapping to the NCBI nucleotide database and summarized in MEGAN, showing the distribution of eukaryote species only. **c** Post-mortem DNA damage signature with MapDamage profiles.

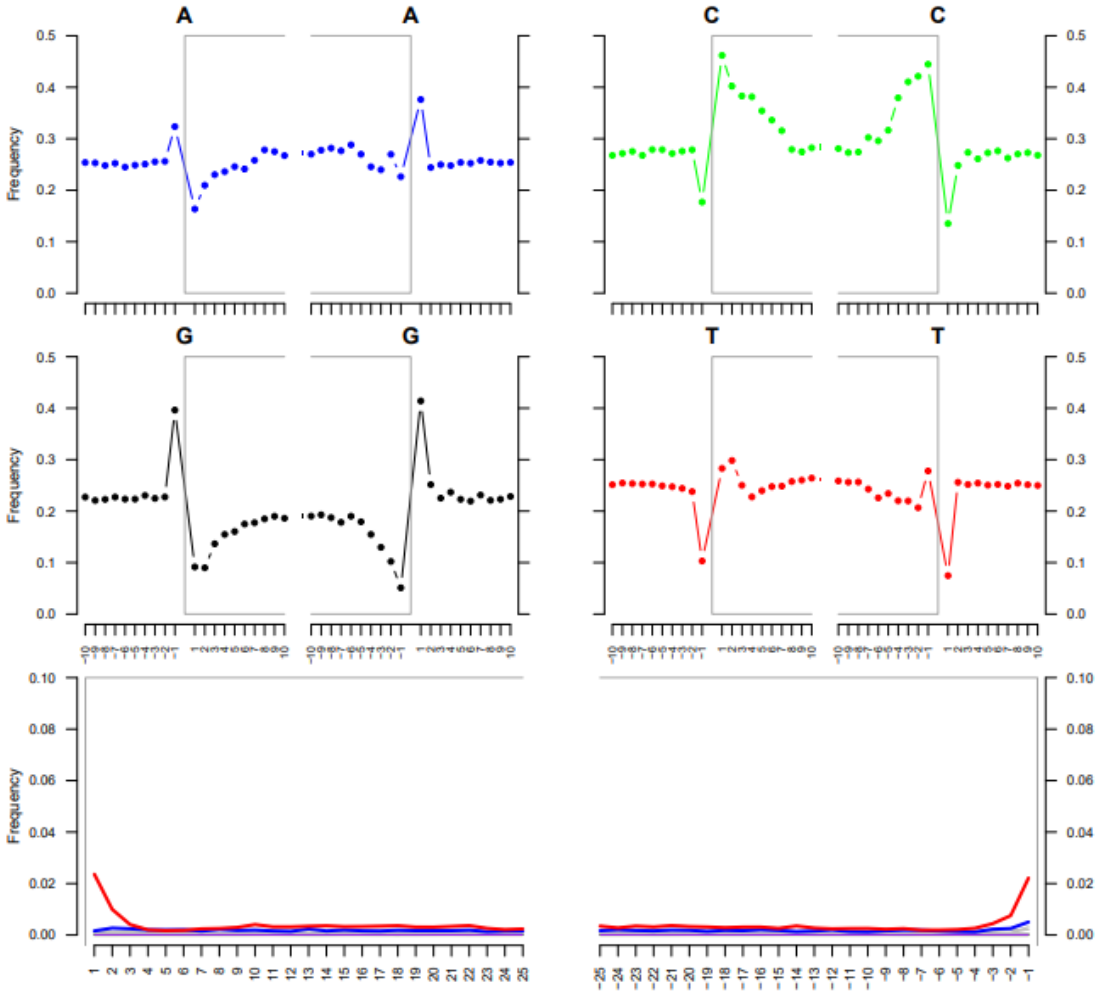
a



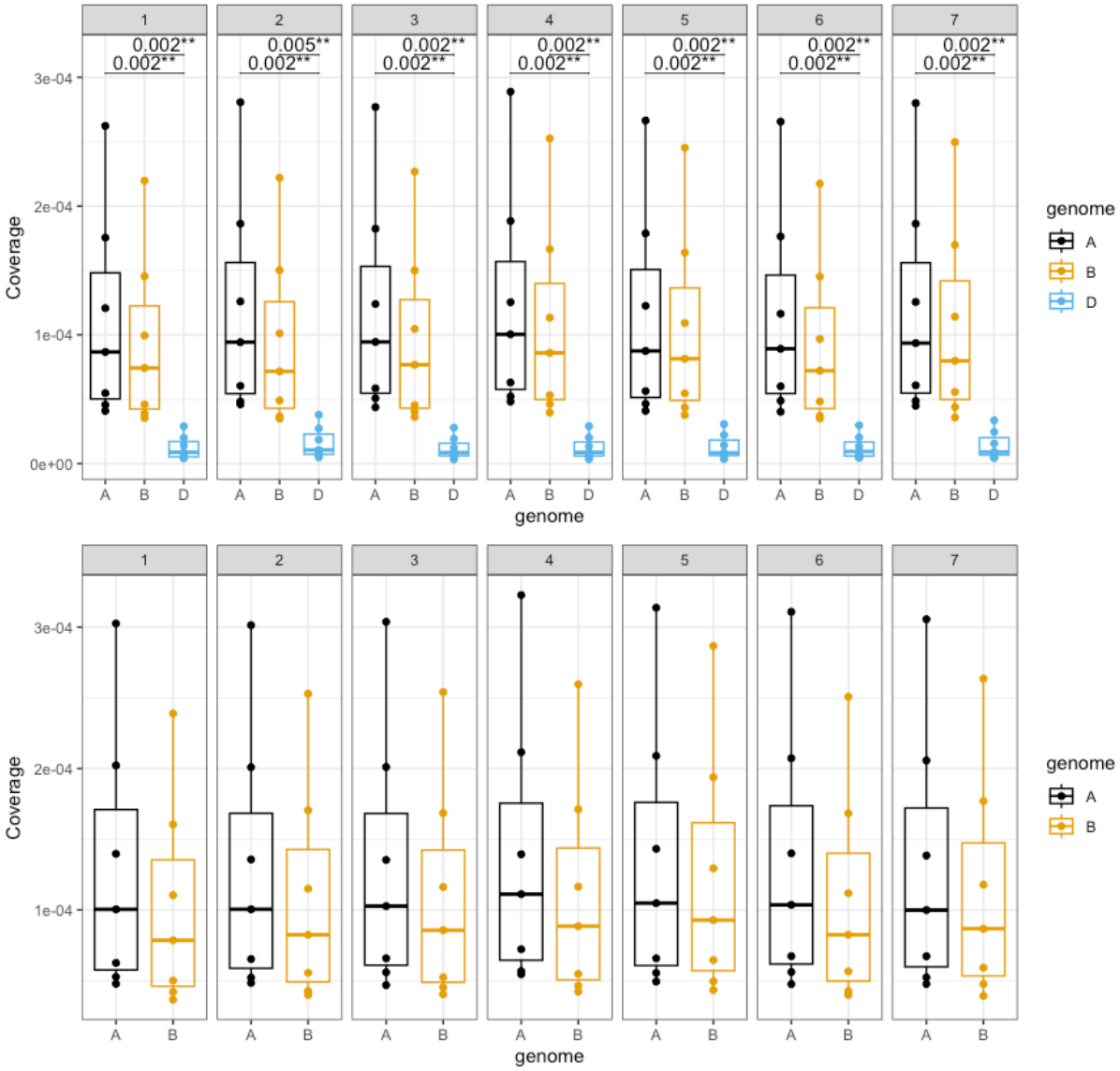
b



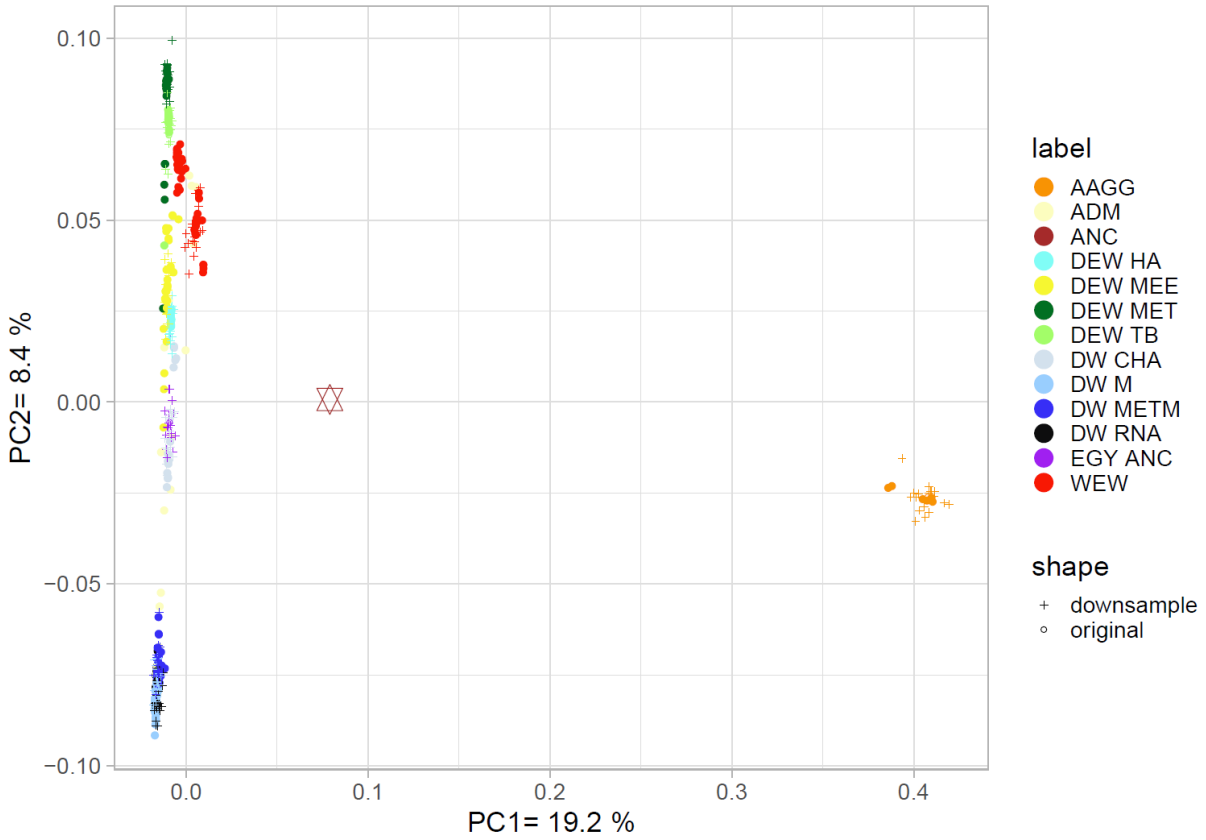
c



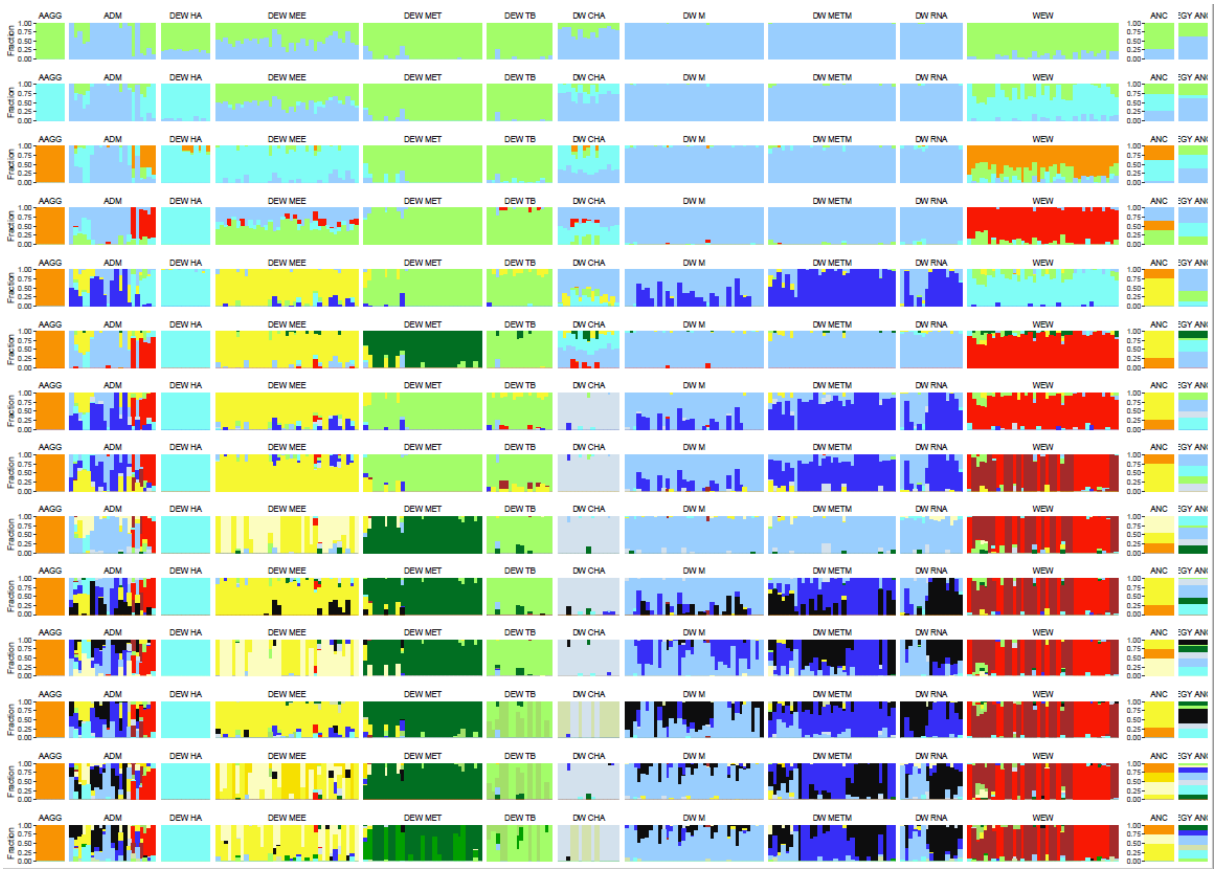
<u>Supplementary Figure 21</u>. Coverage proportions of A, B and/or D genomes from aDNA. Top: high-quality alignments against the AA BB DD hexaploid IWGSC reference genome. Bottom: high-quality alignments against the AA BB tetraploid Svevo reference genome sequence.



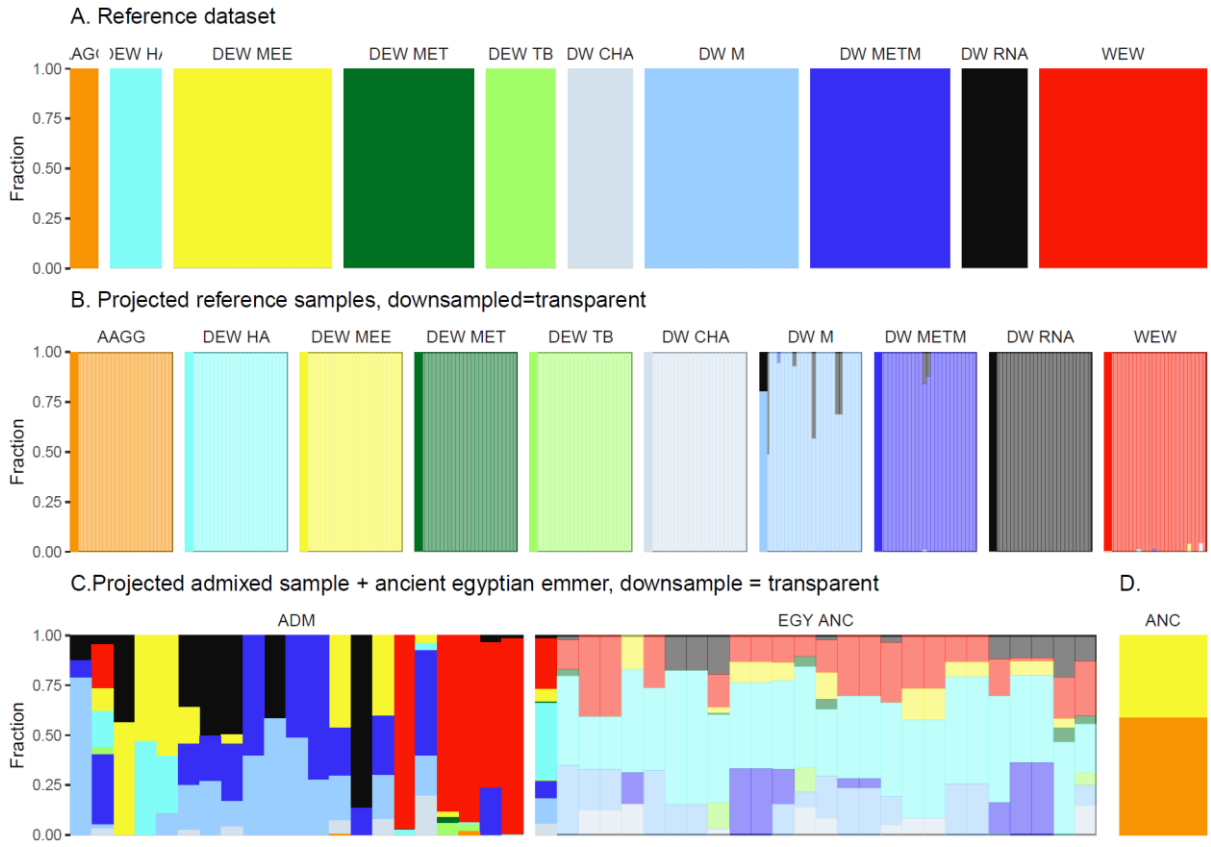
<u>Supplementary Figure 22.</u> Principal Component Analysis of aDNA. The two ancient samples (Sample47 and the ancient Egyptian emmer wheat) were projected together with 10 control modern samples against the genetic structure characteristic of 242 tetraploid imputed accessions within the current modern diversity panel. The genotype data available for the 10 control modern samples and the ancient Egyptian emmer wheat were down-sampled to represent the same amount of SNPs identified in the ancient Sample47. A total of 25 replicates were considered for each of the 10 control samples and the ancient Egyptian emmer wheat.



<u>Supplementary Figure 23</u>. Unsupervised Admixture Analyses of aDNA (for K=2 to K=15, in lines) (for K=10) with the samples from the current modern diversity panel as well as (at the right) ancient Egyptian emmer sample (Scott et al. 2019) and the NGW sample (current study).



<u>Supplementary Figure 24</u>. **Supervised Admixture Analysis of aDNA** (for K=10) with **(a)** the reference dataset, 242 tetraploid imputed accessions from the current modern diversity panel, (**b**) the control samples and their down-sampled replicates, and (**c**) the admixed samples from the current modern diversity panel, ancient Egyptian emmer sample (Scott et al. 2019) and its down-sampled replicates and the NGW sample (current study).



References

1. Abdi, H., Alipour, H., Bernousi, I., Jafarzadeh, J. & Rodrigues, P. C. Identification of novel putative alleles related to important agronomic traits of wheat using robust strategies in GWAS. *Sci Rep* **13**, 9927 (2023).
2. Akiyama, K. *et al.* RARGE II: An Integrated Phenotype Database of Arabidopsis Mutant Traits Using a Controlled Vocabulary. *Plant and Cell Physiology* **55**, e4–e4 (2014).
3. Alachiotis, N., Stamatakis, A. & Pavlidis, P. OmegaPlus: a scalable tool for rapid detection of selective sweeps in whole-genome datasets. *Bioinformatics* **28**, 2274–2275 (2012).
4. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
5. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat Genet* **25**, 25–29 (2000).
6. Avni, R. *et al.* Wild emmer genome architecture and diversity elucidate wheat evolution and domestication. *Science* **357**, 93–97 (2017).
7. Bandelt, HJ. Forster, P. & Röhl, A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 1999 Jan;16(1):37-48.
8. Bawono, P. *et al.* Multiple Sequence Alignment. in *Bioinformatics* (ed. Keith, J. M.) vol. 1525 167–189 (Springer New York, New York, NY, 2017).
9. Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14616–14621 (2007).
10. Browning, B. L., Tian, X., Zhou, Y. & Browning, S. R. Fast two-stage phasing of large-scale sequence data. *The American Journal of Human Genetics* **108**, 1880–1890 (2021).
11. Bustos-Korts, D. *et al.* Exome sequences and multi-environment field trials elucidate the genetic basis of adaptation in barley. *The Plant Journal* **99**, 1172–1191 (2019).
12. Chen, J. *et al*. An evolutionarily conserved gene, FUWA, plays a role in determining panicle architecture, grain shape and grain weight in rice. Plant J. **83**(3), 427-38 (2015).
13. Chen, Y. et al. A Collinearity-Incorporating Homology Inference Strategy for Connecting Emerging Assemblies in the Triticeae Tribe as a Pilot Practice in the Plant Pangenomic Era. Molecular Plant 13, 1694–1708 (2020).
14. Chen, L., Zhao, J., Song, J. & Jameson, P. E. Cytokinin glucosyl transferases, key regulators of cytokinin homeostasis, have potential value for wheat improvement. *Plant Biotechnol J* **19**, 878–896 (2021).
15. Chen, Y. *et al.* The evolutionary patterns of barley pericentromeric chromosome regions, as shaped by linkage disequilibrium and domestication. *The Plant Journal* **111**, 1580–1594 (2022).
16. Cheng, X. *et al.* Characterization of the wheat VQ protein family and expression of candidate genes associated with seed dormancy and germination. *BMC Plant Biol* **22**, 119 (2022).
17. Chernomor, O., von Haeseler, A. & Minh, B. Q. Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices. *Systematic Biology* **65**, 997–1008 (2016).
18. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
19. Civáň, P. *et al.* Episodes of gene flow and selection during the evolutionary history of domesticated barley. *BMC Genomics* **22**, 227 (2021).
20. Crisci, J. L., Poh, Y.-P., Mahajan, S. & Jensen, J. D. The impact of equilibrium assumptions on tests of selection. *Front Genet* **4**, 235 (2013).
21. Dabney, J., Meyer, M. & Paabo, S. Ancient DNA Damage. *Cold Spring Harbor Perspectives in Biology* **5**, a012567–a012567 (2013).
22. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
23. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. Gigascience. **10**(2), giab008 (2021).
24. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238 (2019).
25. Feng, K. *et al.* The improved assembly of 7DL chromosome provides insight into the structure and evolution of bread wheat. *Plant Biotechnology Journal* **18**, 732–742 (2020).
26. Fernández-Calleja, M., Casas, A. M. & Igartua, E. Major flowering time genes of barley: allelic diversity, effects, and comparison with wheat. *Theor Appl Genet* **134**, 1867–1897 (2021).
27. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & François, O. Fast and Efficient Estimation of Individual Ancestry Coefficients. *Genetics* **196**, 973–983 (2014).
28. Fulton, T. L. & Shapiro, B. Setting Up an Ancient DNA Laboratory. in *Ancient DNA* (eds. Shapiro, B. et al.) vol. 1963 1–13 (Springer New York, New York, NY, 2019).
29. Gadaleta, A., Blechl, A. E., Nguyen, S., Cardone, M. F., Ventura, M., Quick, J. S., & Blanco, A. Stably expressed D-genome-derived HMW glutenin subunit genes transformed into different durum wheat genotypes change dough mixing properties. Molecular Breeding **22**, 267-279 (2008).

30. Geng, X. *et al.* Comparative transcriptome analysis of resistant and susceptible wheat in response to Rhizoctonia cerealis. *BMC Plant Biol* **22**, 235 (2022).
31. Hahn, F., Korolev, A., Sanjurjo Loures, L., & Nekrasov, V. A modular cloning toolkit for genome editing in plants. *BMC plant biology* **20**, 1-10 (2020).
32. He, G. Y., Rooke, L., Steele, S., Bekes, F., Gras, P., Tatham, A. S., ... & Lazzeri, P. A. Transformation of pasta wheat (Triticum turgidum L. var. durum) with high-molecular-weight glutenin subunit genes and modification of dough functionality. *Molecular Breeding* **5**, 377-386 (1999).
33. Huang, D. *et al.* Dominant inhibition of awn development by a putative zinc-finger transcriptional repressor expressed at the *B1* locus in wheat. *New Phytologist* **225**, 340–355 (2020).
34. Huson, D. H., Auch, A. F., Qi, J. & Schuster, S. C. MEGAN analysis of metagenomic data. *Genome Res.* **17**, 377–386 (2007).
35. International Wheat Genome Sequencing Consortium (IWGSC). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**, eaar7191 (2018).
36. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
37. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**, 587–589 (2017).
38. Kapp, J. D., Green, R. E. & Shapiro, B. A Fast and Efficient Single-stranded Genomic Library Preparation Method Optimized for Ancient DNA. *Journal of Heredity* **112**, 241–249 (2021).
39. Kohler-Schneider, M. Contents of a storage pit from late Bronze Age Stillfried, Austria: another record of the 'new' glume wheat. *Veget Hist Archaeobot* **12**, 105–111 (2003).
40. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. F1000Res 9, 709 (2020).
41. Konstantinov, D. K., Zubairova, U. S., Ermakov, A. A. & Doroshkov, A. V. Comparative transcriptome profiling of a resistant vs susceptible bread wheat (*Triticum aestivum* L.) cultivar in response to water deficit and cold stress. *PeerJ* **9**, e11428 (2021).
42. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
43. Korunes, K. L. & Samuk, K. PIXY : Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular Ecology Resources* **21**, 1359–1368 (2021).
44. Kronenberg, L. *et al.* Temperature response of wheat affects final height and the timing of stem elongation under field conditions. *J Exp Bot* **72**, 700–717 (2021).
45. Kucera, M., Isserlin, R., Arkhangorodsky, A. & Bader, G. D. AutoAnnotate: A Cytoscape app for summarizing networks with semantic annotations. *F1000Res* **5**, 1717 (2016).
46. Kumar, A. *et al.* Strategies and Bottlenecks in Hexaploid Wheat to Mobilize Soil Iron to Grains. *Front. Plant Sci.* **13**, 863849 (2022).
47. Lee, Y. & Jeon, C. O. Paraburkholderia aromaticivorans sp. nov., an aromatic hydrocarbon-degrading bacterium, isolated from gasoline-contaminated soil. *Int J Syst Evol Microbiol* **68**, 1251–1257 (2018).
48. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://doi.org/10.48550/arXiv.1303.3997 (2013).
49. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
50. Librado, P. *et al.* The origins and spread of domestic horses from the Western Eurasian steppes. *Nature* **598**, 634–640 (2021).
51. Lischer, H. E. L., Excoffier, L. & Heckel, G. Ignoring Heterozygous Sites Biases Phylogenomic Estimates of Divergence Times: Implications for the Evolutionary History of Microtus Voles. *Molecular Biology and Evolution* **31**, 817–831 (2014).
52. Ma, J. *et al.* Identification of quantitative trait loci (QTL) and meta-QTL analysis for kernel size-related traits in wheat (Triticum aestivum L.). *BMC Plant Biol* **22**, 607 (2022).
53. Maccaferri, M. *et al.* Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat Genet* **51**, 885–895 (2019).
54. Maccaferri, M. *et al.* Prioritizing quantitative trait loci for root system architecture in tetraploid wheat. J Exp Bot. **67**(4), 1161-78 (2016).
55. Mascher, M. *et al.* A chromosome conformation capture ordered sequence of the barley genome. *Nature* **544**, 427–433 (2017).
56. Mazzucotelli, E. *et al.* The Global Durum Wheat Panel (GDP): An International Platform to Identify and Exchange Beneficial Alleles. *Front Plant Sci* **11**, 569905 (2020).

57. Minh, BQ. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the *Genomic Era. Mol Biol Evol*. 2020 May 1;37(5):1530-1534. doi: 10.1093/molbev/msaa015. Erratum in: Mol Biol Evol. 2020 Aug 1;37(8):2461.
58. Monat, C. *et al.* TRITEX: chromosome-scale sequence assembly of Triticeae genomes with open-source tools. *Genome Biol* **20**, 284 (2019).
59. Murat, F., Armero, A., Pont, C., Klopp, C. & Salse, J. Reconstructing the genome of the most recent common ancestor of flowering plants. *Nat Genet* **49**, 490–496 (2017).
60. Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**, 268–274 (2015).**54. Ortiz, EM. 2019.** vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. DOI:10.5281/zenodo.2540861
61. Ortiz, E. M. vcf2phylip v2.0: convert a VCF matrix into several matrix formats for phylogenetic analysis. DOI:10.5281/zenodo.2540861 (2019).
62. Park, J., Bae, S., Kim, J.S. Cas-Designer: a web-based tool for choice of CRISPR-Cas9 target sites. *Bioinformatics*, **31**(24):4014-6 (2015).
63. Parveen, A. *et al.* Genome-wide analysis of RING-type E3 ligase family identifies potential candidates regulating high amylose starch biosynthesis in wheat (Triticum aestivum L.). *Sci Rep* **11**, 11461 (2021).
64. Patterson, N., Price, A. L. & Reich, D. Population Structure and Eigenanalysis. *PLoS Genet* **2**, e190 (2006).
65. Pavlidis, P. & Alachiotis, N. A survey of methods and tools to detect recent and strong positive selection. *Journal of Biological Research-Thessaloniki* **24**, 7 (2017).
66. Pavlidis, P., Živkovic, D., Stamatakis, A. & Alachiotis, N. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol Biol Evol* **30**, 2224–2234 (2013).
67. Petrequin, P., Magny, M. and Bailly M., Habitat lacustre, densité de population et climat - L'exemple du Jura français, in : P. Della Casa et M. Trachsel (ed.), Wes'04. Wetland economies and societies, Schweizerisches Landesmuseum Zürich, Zürich, Chronos Verlag : 143-168 (2005).
68. Pont, C. *et al.* Tracing the ancestry of modern bread wheats. *Nat Genet* **51**, 905–911 (2019).
69. Pont, C. *et al.* Paleogenomics: reconstruction of plant evolutionary trajectories from modern and ancient DNA. *Genome Biol* **20**, 29 (2019).
70. Poplin, R. *et al.* Scaling accurate genetic variant discovery to tens of thousands of samples. Preprint at https://doi.org/10.1101/201178 (2018).
71. Pourkheirandish, M. *et al.* Evolution of the Grain Dispersal System in Barley. *Cell* **162**, 527–539 (2015).
72. Pourkheirandish, M. *et al.* Elucidation of the origin of ' *agriocrithon* ' based on domestication genes questions the hypothesis that Tibet is one of the centers of barley domestication. *The Plant Journal* **94**, 525–534 (2018).
73. Pradhan, A. K. *et al.* Identification of QTLs/Defense Genes Effective at Seedling Stage Against Prevailing Races of Wheat Stripe Rust in India. *Front. Genet.* **11**, 572975 (2020).
74. Purcell, S., Daly, M. J. & Sham, P. C. WHAP: haplotype-based association analysis. *Bioinformatics* **23**, 255–256 (2007).
75. Qu, X., Liu, J., Xie, X., Xu, Q., Tang, H., Mu, Y., Pu, Z., Li, Y., Ma, J., Gao, Y., Jiang, Q., Liu, Y., Chen, G., Wang, J., Qi, P., Habib, A., Wei, Y., Zheng, Y., Lan, X., Ma, J. Genetic Mapping and Validation of Loci for Kernel-Related Traits in Wheat (*Triticum aestivum* L.). *Front Plant Sci*. **12**, 667493 (2021).
76. Rabieyan, E., Bihamta, M. R., Moghaddam, M. E., Mohammadi, V. & Alipour, H. Genome-wide association mapping for wheat morphometric seed traits in Iranian landraces and cultivars under rain-fed and well-watered conditions. *Sci Rep* **12**, 17839 (2022).
77. Rahim, A. A. *et al.* Genome-Wide Identification and Characterization of Receptor-Like Protein Kinase 1 (RPK1) Gene Family in Triticum aestivum Under Drought Stress. *Front. Genet.* **13**, 912251 (2022).
78. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
79. Rufo, R., Salvi, S., Royo, C., Soriano, J.M. Exploring the Genetic Architecture of Root-Related Traits in Mediterranean Bread Wheat Landraces by Genome-Wide Association Analysis. *Agronomy* **10***,* 613 (2020).
80. Russell, J. *et al.* Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation. *Nat Genet* **48**, 1024–1030 (2016).
81. Shan, Q., Wang, Y., Li, J., Gao C. Genome editing in rice and wheat using the CRISPR/Cas system. *Nat Protoc*, **9**(10):2395-410 (2014).
82. Schubert, M. *et al.* Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat Protoc* **9**, 1056–1082 (2014).
83. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes* **9**, 88 (2016).

84. Scott, M. F. *et al.* A 3,000-year-old Egyptian emmer wheat genome reveals dispersal and domestication history. *Nat. Plants* **5**, 1120–1128 (2019).
85. Singh, R. *et al.* Meta-QTLs, ortho-MetaQTLs and candidate genes for grain Fe and Zn contents in wheat (Triticum aestivum L.). *Physiol Mol Biol Plants* **28**, 637–650 (2022).
86. Skoglund, P., Sjödin, P., Skoglund, T., Lascoux, M. & Jakobsson, M. Investigating Population History Using Temporal Genetic Differentiation. *Molecular Biology and Evolution* **31**, 2516–2527 (2014).
87. Soriano, J. M. & Alvaro, F. Discovering consensus genomic regions in wheat for root-related traits by QTL meta-analysis. *Sci Rep* **9**, 10537 (2019).
88. Sparks, C. A., & Jones, H. D. Genetic transformation of wheat via particle bombardment. *Cereal genomics: methods and protocols*, 201-218 (2014).
89. Stiller, M. *et al.* Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proceedings of the National Academy of Sciences* **103**, 13578–13584 (2006).
90. Strejčková, B., Čegan, R., Pecinka, A., Milec, Z. & Šafář, J. Identification of polycomb repressive complex 1 and 2 core components in hexaploid bread wheat. *BMC Plant Biol* **20**, 175 (2020).
91. Szeliga, M., Bakera, B., Święcicka, M., Tyrka, M. & Rakoczy-Trojanowska, M. Identification of candidate genes responsible for chasmogamy in wheat. *BMC Genomics* **24**, 170 (2023).
92. Tamura., K, Stecher., G, Kumar., S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. **38**(7), 3022-3027 (2021).
93. Tsehaye, Y., Bjørnstad, Å. & Abay, F. Phenotypic and genotypic variation in flowering time in Ethiopian barleys. *Euphytica* **188**, 309–323 (2012).
94. Van Der Auwera, G. A. *et al.* From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *CP in Bioinformatics* **43**, (2013).
95. Van Rossum, B. J. & Kruijer, W. statgenGWAS: Genome wide association studies. Available at: https://CRAN.R-project.org/package=statgenGWAS (2023).
96. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol* **4**, e72 (2006).
97. Vu, L. D. *et al.* Temperature-induced changes in the wheat phosphoproteome reveal temperature-regulated interconversion of phosphoforms. *Journal of Experimental Botany* **69**, 4609–4624 (2018).
98. Wang, M., Gong, J. & Bhullar, N. K. Iron deficiency triggered transcriptome changes in bread wheat. *Computational and Structural Biotechnology Journal* **18**, 2709–2722 (2020).
99. Wang, P. *et al.* Genome-Wide Association Study of Kernel Traits Using a 35K SNP Array in Bread Wheat (Triticum aestivum L.). *Front. Plant Sci.* **13**, 905660 (2022a).
100. Yang, B., Wen, X., Wen, H., Feng, Y., Zhao, J., Wu, B., Zheng, X., Yang, C., Yang, S., Qiao, L., Zheng, J. Identification of Genetic Loci Affecting Flag Leaf Chlorophyll in Wheat Grown under Different Water Regimes. *Front Genet.* **13**, 832898 (2022b).
101. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer Publishing Company, Incorporated, 2016).
102. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
103. Xiao, A., Cheng, Z., Kong, L., Zhu, Z., Lin, S., Gao, G., Zhang, B. CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics*, **30**(8):1180-1182 (2014).
104. Xu, S. *et al. Ggtree*: A serialized data object for visualization of a phylogenetic tree and annotation data. *Imeta*. **1**(4), e56 (2022).
105. Zhan, H., Nie, X., Zhang, T., Li, S., Wang, X., Du, X., Tong, W., Song, W. Melatonin: A Small Molecule but Important for Salt Stress Tolerance in Plants. *Int J Mol Sci.* **20**(3), 709 (2019).
106. Zhang, R., Jia, G., Diao, X. geneHapR: an R package for gene haplotypic statistics and visualization. BMC Bioinformatics **24**, 199 (2023).