Online Methods

Plant material

Plant material was obtained from genebanks at IPK Gatersleben, the United States Department of Agriculture and the Nordic Genetic Resource Center. Accession numbers are listed in **Supplementary Table 1**.

HMW DNA extraction and HiFi sequencing

Plants were grown under greenhouse conditions and HMW DNA was of isolated from young leaf tissue according to published protocols[1,2]. Construction of HiFi libraries and Circular Consensus Sequencing using the PacBio Sequel IIe instrument (Pacific Biosciences, CA, USA) were essentially as described previously[3]. HiFi sequencing was done at IPK Gatersleben, Genomics & Transcriptomics Laboratory, Heinrich Heine University (Düsseldorf, Germany), Berry Genomis Co. (Beijing, China), Novogene Co. (Beijing, China) and the University of Wisconsin Biotechnology Center DNA Sequencing Facility (Madison, WI, USA).

Hi-C sequencing

In situ Hi-C libraries were prepared from young seedlings according to the previously published protocol, using *Dpn*II for the digestion of crosslinked chromatin[2]. Libraries were prepared as described[4]. Samples were quantified[5] and sequenced with standard protocols from Illumina using Illumina NovaSeq6000 devices (Illumina, San Diego, California, USA) at IPK Gatersleben and Novogene Co. (Beijing, China).

Optical genome mapping and hybrid scaffolding

A clone of the same *H. erectifolium* individual plant that was sequenced using PacBio HiFi was used for the construction of the optical genome map (OGM). A total of 2.5 million nuclei, purified from young leaves by flow cytometry, were embedded in agarose miniplugs and treated with proteinase K, following the protocol of Šimková et al[6]. A total of 525 ng of HMW DNA was directly labeled at DLE-1 recognition sites using the standard Bionano Prep Direct Label and Stain (DLS) protocol (Bionano, San Diego, USA) and analyzed on the Bionano Saphyr platform. The resulting dataset, comprising 1.5 Tbp of single-molecule data with an N50 of 247 kb, provided approximately 341× coverage of the *H. erectifolium* genome This dataset was used to generate a *de novo* OGM assembly using Bionano Solve software (version 3.6.1_11162020) with the standard configuration file "optArguments_nonhaplotype_noES_noCut_DLE1_saphyr.xml."

To improve the contiguity of the PacBio HiFi assembly, the automatic hybrid scaffold pipeline integrated in Bionano Solve was run using the OGM assembly. The default DLE-1 Hybrid Scaffold configuration file was applied with the "Resolve conflict" option for conflict resolution. Conflicts between the sequence assembly and the OGM were manually curated and the pipeline was subsequently re-run using a modified *conflict_cut_status.txt* file.

Transcriptome sequencing

Plants were grown under greenhouse conditions and total RNA was isolated from different tissues and developmental stages (**Supplementary Table 4**) using the RNAeasy Plant Mini Kit (Qiagen, Hilden, Germany) or with Trizol reagent (Invitrogen, CA, US) according to the manufacturers' protocols. RNAseq of individual samples and PacBio IsoSeq of RNA pools was performed as described previously[1]. Library construction and sequencing on Illumina

Novaseq6000 devices (Illumina, San Diego, California, USA; RNAseq) and PacBio Sequel IIe (instruments) was done at IPK Gatersleben, Biomarker Technologies (BMK) GmbH (Münster, Germany), the Genomics & Transcriptomics Laboratory at Heinrich Heine Univeristry Düsseldorf (Germany), Kazusa DNA Research Institute (Japan) or Berry Genomis Co. (Beijing, China). The PacBio IsoSeq sequencing data was processed with the isoseq3 (version 3.8.2) pipeline (PacBio - IsoSeq3, 2018/2024). The ploy-A tails were trimmed and concatenated transcripts removed with "refine -require-polya -min-polya-length 12 teloprime.fasta". To generate non-redundant transcripts, we mapped the full-length non-concatenated (FLNC) reads were mapped to the genome with pbmm2 (version 1.10.0) "align -preset ISOSEQ -sort -bam-index CSI" and redundant transcripts collapsed with isoseq3 collapse " -do-not-collapse-extra-5exons".

Genome assembly

PacBio HiFi reads were assembled using hifiasm (v0.13-r308 or v0.15.5-r350 or v0.16.1-r375)[7]. Pseudomolecules were constructed with the TRITEX pipeline[8,9], and chimeric contigs or orientation errors were corrected based on manual inspection of Hi-C contact matrices. For allotetraploid species, subgenomes were phased using a previously developed haplotype-resolved genome assembly strategy[1]. The origin of the subgenomes and their potential ancestral contributors were previously investigated in detail by Brassac et al.[10] To classify subgenomes in the polyploid genome, Mash (version 2.3)[11] was employed to calculate pairwise chromosomal similarities between polyploid chromosomes and candidate ancestral genomes. The resulting distance matrix was used to infer subgenome origin and relationships. In species with high heterozygosity or outcrossing (*H. brevisubulatum* and *H. murinum* BBC2017), Hi-C data were incorporated during hifiasm (v0.15.5-r350)[7] assembly to collapse haplotypes and assemble a representative haploid genome. Except for *H. erectifolium*, whose primary contigs were first hybrid-scaffolded using a Bionano optical map, followed by pseudomolecule construction with Hi-C data and the TRITEX pipeline. Chloroplast genomes were assembled following previously published protocols (ref. HB). HiFi reads were aligned to the barley cv. Morex chloroplast genome (GenBank accession EF115541) using minimap2 (v2.24-r1122)[12] with the parameters "-ax map-hifi --secondary=no --sam-hit-only". Alignments with mapping quality <30 were discarded. Aligned reads were converted to FASTQ format using SAMtools (v1.16.1)[13] and assembled with Canu (v2.1.1)[14] in PacBio-HiFi mode with genome sizes from 120 to 400 kb in 10 kb increments. Circular sequences were identified and trimmed using BEDTools (v2.30.0)[15] and annotated with GeSeq[16]. Fully annotated assemblies were aligned to the Morex chloroplast genome using MUMmer4 (v4.0.0beta2)[17], and start positions and orientations of the SSC region were adjusted to match the reference using seqkit (v0.9.1)[18].

De-novo Gene Annotation of the Hordeum *pangenome*

We performed *de novo* structural gene prediction, confidence classification, and functional annotation, following the protocol described in Mascher, *et al.*[19]. As evidences we used RNAseq and ISOseq data as described under above and Poales-filtered protein sequences downloaded from Uniprot (July 2023). An additional consolidation step was carried out using the liftoff[20] tool. In a nutshell, all predicted gene models from one species were lifted over to all other genomes. Using Mikado we selected the best mapped gene models and unified all gene predictions accordingly. We classified predicted proteins into high and low confidence categories. High-confidence proteins had over 80% coverage and were complete, either with

significant hits in UniMag or UniPoa but not PTREP. Low-confidence proteins were incomplete with hits in UniMag or UniPoa but not PTREP, or were complete without hits in any database. We benchmarked the annotations with BUSCO (v. 5.8.0)[21] (dataset: poales_odb12), OMARK (omamer verision 2.1.0/omark version 0.3.1)[22] and psauron (v. 1.0.6)[23].

Construction of the gene-based Hordeum *pangenome*

Orthogroups (OGs) based on the primary protein sequences from 25 annotated *Hordeum* species were calculated using Orthofinder (version 2.5.5)[24]. The scripts for the calculation of core/shell and cloud genes are deposited in the repository https://github.com/PGSB-HMGU/BPGv2. Core genes contain at least one gene model from all 25 compared species. Shell genes contain gene models from at least two *Hordeum* species and at most 24. Genes clustered with genes only from the same genotype, were defined as cloud genes. Annotations of *H. vulgare* Morex and *H. bulbosum* were not included during the consolidation step. Therefore, to counteract any potential over- or under-presentation of these two species in any of the core/shell/cloud categories, we aligned all could genes for Morex and Bulbosum back to all other genomes and corrected for any missing genes. GO-term enrichments were conducted with goatools (v1.4.4)[25]. Gene-based collinearity plots were drawn with GENESPACE[26].

Pangenome analysis of nucleotide-binding resistance genes

The RGAugury (v.2.2)[27] pipeline was used to identify from our annotation and classify nucleotide-binding site (NB-ARC) domain-encoding genes into different subgroups, based on domain and motif structures: TNL contain TIR, NBS, and LRR domains; CNL contain CC, NBS, and LRR domains; RNL contain RPW8, NBS, and LRR domains; RN contain RPW8 and NBS domains only; NL contain NBS and LRR domains but lack TIR, CC, and RPW8; TN contain TIR and NBS domains only; CN contain CC and NBS domains only; and NBS-only genes contain only the NBS domain. TIR, NBS, LRR, CC, and RPW8 refer to the Toll/Interleukin-1 Receptor-like, Nucleotide-Binding Site, Leucine-Rich Repeat, Coiled-Coil, and Resistance to Powdery Mildew 8-like domains, respectively. An NLR gene cluster is defined based on the genomic proximity of NLR genes. Specifically, any two NLR genes located within 500 Kb of each other are grouped into the same cluster by bedtools (v2.30.0)[15]. NLR genes were then grouped into clusters using bedtools (v2.30.0)[15] merge with a maximum intergenic distance of 250 kb. Only clusters comprising no fewer than three NLR genes were retained. Pan-NLR analysis was performed based on orthogroups (OGs) identified using OrthoFinder (version 2.5.5)[24] across all NLR genes. Core OGs are defined as orthogroups that contain at least one gene model from each of the compared *Hordeum* accessions. Shell OGs include gene models from at least two and at most 24 *Hordeum* accessions. Gene models that were not assigned to any OG ("singletons") or were grouped only with genes from the same genotype were classified as cloud genes.

Pangenome analysis of PPR and mTERF genes

Open reading frames (ORFs) longer than 100 codons were searched for the presence of PPR and mTERF motifs using *hmmsearch* from the HMMER suite (http://hmmer.org) employing hidden Markov models (HMMs) as defined in Cheng et al[28]. for the PPR family and PF02536 from the Pfam 32.0 database (http://pfam.xfam.org)[29] for the mTERF motif. Downstream processing of the *hmmsearch* results for the PPR proteins followed the pipeline described in Gutmann et al[30]. The RfCTD domain recently identified as a unique feature of Restorer-of-

fertility-like PPR proteins (RFL-PPRs)[31] was included in the gene annotations. P- and PLS-class genes with scores below 100 and 240, respectively, were excluded from the annotation as they are unlikely to represent functional PPR genes, as defined previously (IWGSC 2018)[32].

Graph based pangenome construction and analysis

The analyses described here used a number of tools from the toolkits *vg* (v1.61.0)[33], *odgi* (v0.8.3-26-gbc7742ed)[34], *bcftools* (v1.21)[35] and seqkit (v.2.9.0)[18]. All scripts used for graph construction and analysis are available at https://github.com/cropgeeks/panHordeum. Due to computational limitations, we built an alignment of *Hordeum* whole genomes with *cactus* [36]. We selected *H. marinum* BCC2001 over *H. marinum* H559 on the basis of its higher assembly quality. Genome haplotypes were soft-masked with *bedtools* [37] using their respective repeat annotations. The guide tree required for the *cactus* alignment was taken from the *Orthofinder*[38] output generated as part of this study. Apart from setting appropriate resource parameters (*i.e.* threads, memory, disk), we ran *cactus* with default parameters. We transformed the resulting whole genome alignment into a super-pangenome graph using *hal2vg* (https://github.com/ComparativeGenomicsToolkit/hal2vg), with the "--noAncestors" flag, and then further processed it with *vg convert* to get the final pangenome graph in GFA format (https://github.com/GFA-spec/GFA-spec). We computed basic graph statistics using *odgi stats* with the "--summarize" option, which included graph size, number of nodes, edges, paths and steps. To characterize the conservation of the different segments of the graph, we classified nodes as core, shell, or cloud. A node was considered "core" if it was traversed by all samples' paths, "cloud" if it was traversed by a single sample's path, and "shell" otherwise. The same classification was done at the base level. Both computations were performed using the *odgi toolkit* (implementation details in class_stats.sh). To evaluate the agreement of haplotype sequences with their representation in the graph, we first extracted the linear sequences of all haplotypes as encoded in the graph using *odgi paths* with the "--fasta" option. These sequences were subsequently partitioned by source genome using *seqkit grep,* creating direct pairwise comparisons between input genomes and their graph-encoded counterparts. For each genome pair, we performed exact sequence matching, confirming whether each input sequence (or its reverse complement) was precisely preserved in the graph-encoded version (implementation details available in script fidelity.py). We computed pangenome graph growth curves based on the full graph using *odgi heaps* with 100 permutations of the input genomes. Results were plotted using script heaps_fit.R distributed with the *odgi* toolkit[34]. Whole-graph analysis turned out to be computationally prohibitive (job runtime limits and/or memory requirements) for several of the required tools from the *vg* and *odgi* toolkits, given a graph of this size and complexity. We therefore extracted subgraphs equating to individual chromosome paths for *Hordeum vulgare* cv. Morex, which we used as the basis for all variant analyses. We used *vg chunk* for subgraph extraction. Variants were called from the chromosome-level subgraphs directly, using *vg deconstruct*. This generated VCF output containing the full range of variant sizes. To prepare structural variants from the graph for further analysis we broadly followed the variant decomposition approach used by the Human Pangenome Consortium[39]. This involved the removal of top-level variants of length > 100,000 bp with *vcfbub* (https://github.com/pangenome/vcfbub), realigning alternate alleles against the reference for each variant using *vcfwave*[40], and finally normalising the resulting variants with *bcftools norm*[35]. We then added variant length information (SVLEN) to the VCF files with a custom script to allow extraction of structural variants (SVs) based on the length difference between reference and alternate allele (> 50 bp) and classified SVs as

deletions or insertions based on the directionality of the difference (insertion: string length REF < string length ALT; deletion: string length REF > string length ALT). The resulting VCF files were converted to BED format with *bcftools query*. Variant density plots were computed with custom R scripts using *ggplot2* (v3.5.1)[41]. To explore the possibility of using the graph for read mapping and genotyping, we attempted to index the graph using *vg gbwt/vg index/vg minimizer*. This turned out to be computationally prohibitive in terms of job runtime limits on two separate compute facilities used, and thus we were unable to use the graph for the valuation of read mapping. We conducted graph-based conservation analysis with *Pansel*[42], a tool used for calculation of diversity/conservation metrics directly from pan-genome graphs. The default bin size of 1kbp was used, along with all other parameters on default settings. Results were plotted using ggplot2 (v3.5.1)[41].

Whole-genome multiple sequence alignment

Single-copy genes were identified using BUSCO (v5.4.6)[21] in *Brachypodium distachyon* (v3.0)[43], *Oryza sativa* (IRGSP-1.0)[44], *Sorghum bicolor* (IRGSP-1.0)[44], *Triticum monococcum*[45], *Hordeum vulgare* FT11[46], and the 29 *Hordeum* haplomes collected in this study. Protein sequences of BUSCO single-copy genes present in all 34 haplomes were concatenated and aligned using MAFFT (v7.490)[47]. Well-aligned regions were filtered using Gblocks (version 0.91b)[48] with parameters "-t=extractedp -b4=5 -b5=h". A maximum-likelihood phylogenetic tree was then constructed using RAxML (version 8.2.12)[49] with the PROTGAMMAWAG substitution model and the rapid bootstrap algorithm with 100 replicates. To facilitate genome-wide alignment, all 34 haplomes were masked using RepeatMasker (version 4.1.1, http://www.repeatmasker.org) with a pan-genome TE library generated by panEDTA. The masked genomes and the inferred species tree were then used as input for Cactus (v2.6.5)[50] to generate whole-genome multiple sequence alignments. The loss rate of genomic elements was defined as the size of the reconstructed ancestral sequence at the corresponding genomic position in the multiple genome alignment.

Phylogenetic tree construction and Estimation of the divergence time

To precisely construct phylogenetic tree and estimate divergence times within the genus *Hordeum*, we included four additional outgroup genomes: *Brachypodium distachyon* (v3.0)[43], *Oryza sativa* (IRGSP-1.0)[44], *Sorghum bicolo*r (NCBIv3)[51] (all downloaded from Ensembl Plants: https://plants.ensembl.org/index.html), and *Triticum monococcum*[45] (https://wheat.pw.usda.gov/GG3/pangenome). We constructed phylogenetic trees using four approaches. The first two methods were based on whole genome single-copy genes. Orthologous genes were identified from the genomes of our 29 Hordeum haplomes using OrthoFinder (version 2.5.5)[24], yielding 2,969 single-copy genes. For the concatenation-based approach, the CDS sequences of these genes were concatenated and aligned with MAFFT (v7.490)[47], and well-aligned regions were retained using Gblocks (version 0.91b)[48] with the parameters "-t=extractedp, -b4=5, -b5=h"; phylogenetic inference was then performed in RAxML (version 8.2.12)[49] under the GTRGAMMA substitution model using the rapid bootstrap algorithm with 100 replicates. For the coalescence-based approach, the CDS sequences of each single-copy gene were aligned separately using MAFFT (v7.490)[47], high-quality regions were selected with Gblocks (version 0.91b)[48] using the same parameters, individual gene trees were reconstructed with IQ-TREE (version 2.2.2.6)[52], and a species tree was inferred from these gene trees using ASTRAL-Pro3 (CASTLES-Pro, v1.23.3.6)[53]. The remaining two methods were based on whole-genome multiple sequence alignments generated with Cactus. In the

CASTER-pair (CASTLES-Pro, v1.23.3.6)[54] approach, the whole-genome MSA was used to infer a phylogenetic tree rooted with Sorghum bicolor, while in the CASTER-site (CASTLES-Pro, v1.23.3.6)[54] approach, the whole-genome MSA was analyzed similarly but using site-based inference, also rooted with Sorghum bicolor.

To estimate divergence times, fourfold degenerate transversion (4DTV) sites were extracted from the single-copy gene alignments. A phylogenetic tree was reconstructed using RAxML under the same model and bootstrapping scheme. Fourfold degenerate transversion (4DTV) sites were extracted from multiple sequence alignments of single-copy genes. Divergence times between barley and outgroup species were obtained from the TimeTree database (https://timetree.org/)[55]. Finally, species divergence times within the genus *Hordeum* were estimated using the MCMCtree program in PAML (version 4.4)[56].

Identify the boundary of the proximal regions

We excluded multiple haplomes from the same species and retained a final set of 27 *Hordeum* haplomes for comparative analysis. All haplomes pairs were aligned against each other using minimap2 (v2.24-r1122)[12] with the parameters "-2 -K 5G -f 0.005 -x asm10", and the resulting alignments were stored in PAF format for visualization via dot plots. For each reference haplomes, we calculated the alignment coverage in non-overlapping 5 Mb windows, considering only alignments with a mapping quality greater than 20. Coverage values were normalized across the haplomes by the average coverage to produce normalized coverage profiles. In species that diverged more than 9 million years ago, a bimodal distribution of normalized coverage was observed, allowing a threshold of 0.5 to distinguish distal and proximal regions **(Extended Data Fig. 3a and b)**. Finally, genome-wide alignment and normalized coverage profiles were used to manually define the size and boundaries of proximal and distal regions across all species.

Identification of large Inversions

All other *Hordeum* genomes were aligned to the MorexV3[19] reference genome using wfmash (v0.9.1-3, https://github.com/waveygang/wfmash) with the parameters "-p 90 -l 0". Inversions were then identified using SyRI (version 1.5.5)[57] with parameters "-F P -f --nosnp". Only inversions larger than 2 Mb and located outside of the proximal regions were retained for further analysis.

Detection of introgressed regions using gene tree discordance.

Based on the 2,969 single-copy orthologs identified in the section "Phylogenetic tree construction and estimation of divergence time," and using the genome-wide concatenated single-copy gene tree as a reference, we conducted sliding-window gene tree analyses **(Extended Data Fig. 4b)**. All single-copy genes were grouped into non-overlapping windows of 30 genes with a step size of 5 genes. Within each window, protein sequences were aligned using MAFFT (v7.490)[47], and well-aligned regions were retained using Gblocks (version 0.91b)[48] with the parameters "-t=extractedp -b4=5 -b5=h". Local window trees were built with FastTree (version 2.1.11)[58] (default GTR+CAT, 20 rate categories), starting from neighbor-joining trees and optimized with minimum-evolution and nearest-neighbor interchange moves; branch support was calculated using Shimodaira–Hasegawa-like values. Each resulting tree was rooted with *Sorghum bicolor* as the outgroup using ETE3 (version 3.1.2)[59], and the normalized Robinson–Foulds (nRF) distance was calculated against the reference tree using the parameter "--min_support_src 90". The nRF metric quantifies topological dissimilarity

between trees, with values ranging from 0 (identical topology) to 1 (completely different topologies). Based on the nRF distribution, we identified 0.18 as a threshold to separate background from discordant signal peaks **(Extended Data Fig. 4c)**. Genomic regions containing at least 10 overlap windows with nRF ≥ 0.18 were defined as candidate introgressed regions. For each candidate region, we reconstructed local phylogenetic trees using the same procedure as for the genome-wide tree. The topological comparison of these local trees with the reference tree enabled the inference of introgression origins. The Approximately Unbiased (AU) test[60] implemented in IQ-TREE (version 2.2.2.6)[52], was used to compare local reference trees and local phylogenetic trees. A p-value (p-AU) less than 0.05 indicated that the tree was significantly rejected.

Identification of HGT candidates

Genomes of 20 diploid *Hordeum* species were searched for potential horizontal gene transfers from grasses of the subfamily Panicoideae. Regions of interest were identified in the genomes using BLAST (v2.12.0+)[61] against four potential donor species: *Panicum hallii* (v3.2), *Setaria italica* (v2.2), *Paspalum vaginatum* (v3.1), and *Paspalum notatum* (GCA_036689595.1) (**Supplementary Table 10**). BLAST results were filtered to retain hits with a sequence identity of ≥95% and an alignment length of ≥150 bp. The same process was applied to *H. vulgare* cv. Morex (v3), which served as a negative control. *H. vulgare* is widely recognized as a species that lacks horizontally transferred genes (HGTs) from the Panicoideae subfamily[62-64]. Hits from potential donors that were found in both the *Hordeum* recipient genomes and *H. vulgare* were excluded from further analysis. This filtering was conducted using the intersect tool from BEDTools (v2.26.0)[15], with an overlap threshold of 20%. Finally, the identified regions were compared with newly available annotations, and corresponding candidate genes were extracted. In this step, we identified 5,169 candidate genes.

Phylogenetic validation of HGT candidates

To verify the foreign (i.e., Panicoideae) origin of the candidate HGT genes identified as described above, phylogenetic trees were constructed for each candidate. Homologs were retrieved from a set of representative Poaceae species and two outgroups (*Ananas comosus* and *Musa acuminata*) (**Supplementary Table 10**) using BLASTn (e-value cutoff of $1e^{-10}$). The retrieved sequences were aligned with the Hordeum HGT candidates using MUSCLE (v3.8)[65]. The alignments were trimmed using trimAl (v.1.2) with an automated trimming method[66]. Phylogenetic trees were constructed using PhyML (v.3.1)[67], with 1,000 bootstrap iterations and the GTR substitution model. The resulting trees were reviewed semi-automatically with the assistance of the Newick Utilities suite[68]. A candidate gene was considered a true HGT if the gene sequence was nested within the PACMAD clade. Transposable element-related genes were filtered out and only high-confidence, protein-coding genes were retained for further analysis. Out of 5,169 candidate genes, 4,393 did not pass through the primary phylogenetic validation and 776 genes remained. Candidate genes that passed the first step of verification were cross-referenced with orthogroups previously created during pan-genome construction. For groups containing more than one candidate gene, a consensus sequence was generated using MAFFT (v7.520)[47] for multiple sequence alignment, followed by the *cons* tool from EMBOSS (v6.5.7.0)[69] to produce the consensus. These consensus sequences were subsequently used for BLASTn searches. Hits from at least five species, including minimally one from Panicoideae, were required for tree construction. Phylogenetic trees were then constructed as described above, using MAFFT as the aligner. Alignments and trees were

reviewed manually: orthogroups supporting the HGT scenario were retained, and genes in these groups were classified as either 'foreign' or 'native', based on their position on the tree. The neighbor of transferred genes was marked as presumed donor if the branch represented one genus or species, otherwise the donor was marked as unknown. The entire orthogroup was considered to have a native copy if it contained at least one gene marked 'native', regardless of which species that gene was present in.

Mapping the acquisition of foreign genetic material during the evolutionary history of Hordeum

To determine the transfer times of genes during the evolutionary history of the lineage, an ancestral state reconstruction was conducted for each gene independently. The marginal ancestral state was inferred using the equal rate (ER) model[70], applying the maximum likelihood approach described by Pagel[71] and implemented in the phytools package[72]. To ensure accuracy, the results were verified and adjusted for cases where genes were found in multiple tips but not in their common ancestor. In such instances, the presence of the gene was manually introduced at the ancestor node. For each node in the phylogenetic tree, both the originally reconstructed values and the manually adjusted values were visualized.

Gene ontology analysis

Genes from all orthogroups were annotated using eggNOG-mapper web server (v.2.1.12)[73,74]. GO terms were extracted for each group and enrichment analysis was performed using the clusterProfiler package (v.4.14.3), with a significance threshold of an adjusted p-value of 0.05[75].

Analysis of HGT expression

RNA-Seq data for inflorescence, leaf and root tissues previously generated for 19 *Hordeum* species were used in this study (RNAseq data for *H. bulbosum* were not available). Read trimming on quality (Q30) and removal of sequencing adapters were done using Trimmomatic (v.0.32)[76]. Cleaned reads from each library were then mapped to the *corresponding* reference genome using the STAR aligner (v.2.7.7a)[77]. Quantification of mapped reads was done using the subread package v.1.5.2 with featureCounts[78] followed by normalization using the Trimmed Mean of M-values (TMM) method implemented in the edgeR package (v.4.2.1)[79]. Genes were considered to be expressed if TMM ≥ 1 in at least one tissue.

Annotation and analysis of transposable elements

Structurally intact and fragmented TEs were annotated for each genome using the panEDTA (v0.1) pipeline[80]. In addition to the curated TREP database (https://trep-db.uzh.ch/)[81], TE classification followed the EDTA ontology (https://github.com/oushujun/EDTA/blob/master/util/TE_Sequence_Ontology.txt), with manual curation of TE categories. High-confidence gene models from *Hordeum vulgare* cv. MorexV3 was supplied via the "-cds" parameter to exclude genic sequences from TE annotation. For each genome, the length, classification, insertion time, and divergence metrics of intact LTR retrotransposons were extracted from the "intact.gff3" and "pass.list" output files. To investigate TE family expansion and intergenomic relationships, we employed the TEpop framework (https://github.com/wicker314/TEpop) to treat each transposable element (TE) as an individual sample for population structure analysis[82]. Principal component analysis (PCA) was then performed using FlashPCA2[83] to explore the major patterns of TE

variation across *Hordeum* species. To infer the formation time of tetraploid genomes from TE bursts, full-length LTR/Copia elements were extracted from tetraploid assemblies and clustered using CD-HIT (version 4.8.1)[84] with the parameters "-n 5 -c 0.9 -d 0 -aL 0.9 -aS 0.9 -g 1" based on sequence similarity. TEs that were shared between tetraploid subgenomes but absent from their direct diploid progenitors were defined as post-allopolyploidization insertions. Insertion times of these elements were then used to estimate the timing of tetraploid genome formation.

Identification of Hordeum conserved elements

To identify conserved elements across *Hordeum* genomes, outgroup species were removed from the HAL alignment file using halRemoveGenome in Cactus. Genome-wide multiple sequence alignments were then extracted in MAF format using cactus-hal2maf with the parameters "--chunkSize 1000000 --noAncestors --onlyOrthologs --noDupes", using MorexV3 as the reference genome. Alignment blocks containing fewer than 10 sequences were filtered out in multiple sequence alignments. A neutral model of evolution was fitted to the *Hordeum* phylogeny using fourfold degenerate sites from chromosome 3H of barley, employing phyloFit from the PHAST package (v1.4)[85]. Based on the genome-wide multiple alignments and the trained neutral model, conserved elements were identified using PhastCons with the "--most-conserved" flag. The conserved and non-conserved models were trained with a target genome coverage of 0.45, and the expected length of conserved elements was set to 8 bp.

ATAC-seq Sequencing and Data Analysis

ATAC-seq libraries were prepared from fresh leaf tissue at 3-5 leaf stage. Nuclei were isolated by fine chopping the tissue in isolation buffer (0.25 M sucrose, 10 mM Tris-HCl pH 8.0, 10 mM MgCl2, 1% Trion X-100, 5 mM β-Mercaptoethanol) supplemented with 1x Halt™ Protease Inhibitor Cocktail (Thermo Fisher Scientific, MA, USA). Following isolation, nuclei were quantified via flow cytometry and 75,000 nuclei were subjected to tagmentation using the Tagment DNA Enzyme (TDE1, Illumina, San Diego, CA, USA) at 37°C for 30 min. Transposed DNA was purified using the MinElute PCR Purification Kit (Qiagen, Germany), libraries were amplified with NEBNext® High-Fidelity 2x PCR Master Mix (New England Biolabs, MA, USA), and the final library purification was performed using VAHTSTM DNA Clean Beads (Vazyme, China) according to standard protocols from the manufacturers. Libraries were quantified as described previously (Mascher et al., 2013) and sequenced (paired-end, 2 x 151 cycles, Illumina NovaSeq6000, IPK-Gatersleben) following protocols provided by the manufacturer (Illumina, San Diego, California, USA).

Adapter trimming and quality filtering were performed using Trim Galore (version 0.6.4_dev, https://github.com/FelixKrueger/TrimGalore.git) with the following parameters "-q 25 --phred33 --stringency 3 --length 60 --paired -j 10". Only read pairs with both ends retained were used in subsequent analyses. Trimmed reads were aligned to the reference genome using BWA (v0.7.17)[86]. The resulting SAM files were converted to BAM format and sorted using SAMtools (v1.10)[13]. Only properly paired reads with mapping quality ≥ 30 were retained (-f 2 -F 1804 -q 30), and further sorted by read name to enable mate information correction using samtools fixmate. Duplicate reads were marked using Picard MarkDuplicates (v2.21.9, https://broadinstitute.github.io/picard/) with VALIDATION_STRINGENCY=LENIENT and REMOVE_DUPLICATES=false. After marking, only properly paired, non-duplicate reads were retained and re-sorted by coordinate to produce final cleaned BAM files for downstream analysis. BAM indices were generated using samtools index. Peaks were called using MACS2

(version 2.2.9.1)[87]. The effective genome size (-g) was set to 4.2 Gb, approximating the size of the barley genome. Peaks were called with automatic threshold evaluation (--cutoff-analysis), and all duplicate reads were retained during the analysis to preserve biological signal.

Identification of VRN1 and ZCCT genes

VRN1 and *ZCCT* genes were identified in all assembled wild barley genomes using blastn searches (blast-plus/2.12.0)[61] with *VRN1* MorexV3 (NC_058522.1:c528157887-528147799) and NCBI GenBank *ZCCTa/b/c* queries (AY485977, AY485978, AY687931, respectively). Corresponding regions were extracted from genome assemblies by bedtools (v2.26.0)[15] and manually annotated and analyzed by Geneious Prime software (v2024.0.5; https://www.geneious.com). Sequence similarity visualization of the 2 kb region upstream of the transcription start site and the first intron of *VRN1* genes was done by flexidot (v1.06)[88]. The 79-bp deletion in *H. cordobense* and *H. muticum VRN1* gene was detected by multiple sequence alignment of the extracted *VRN1* sequences from wild *Hordeum* species and winter cultivar Igri in Geneious Prime software (v2023.0.4; https://www.geneious.com) with MAFFT[47]. The critical region of cultivar *Igri* was identified in a previous study[89].

Small RNA and mediator RNA isolation and sequencing

Three biological replicates were analyzed for each stage (14 DAG, 28 DAG) of *H. cordobense* and *H. californicum* grown under controlled non-vernalizing conditions (21 °C day/18 °C night, 14 h light/10 h dark photoperiod). Total RNA was extracted from frozen leaf tissue using a Quick-RNA Miniprep Kit (Zymo Research). RNA quality and concentration were evaluated with a Bioanalyzer 2100 (Agilent Technologies) on RNA 6000 Nano chips (Agilent Technologies). *H. cordobense* 28 DAG RNA sequencing (RNA-seq) libraries and small RNA-seq libraries were constructed according to Illumina instructions and sequenced on a Novaseq 6000 platform at Novogene Co. *Ltd.* Other sequencing libraries (*H. cordobense* 14 DAG and *H. californicum* 14 and 28 DAG; small RNA and RNA-seq) were constructed at the Institute of Experimental Botany (Olomouc) using NEBNext® Multiplex Small RNA Library Prep Set for Illumina® and NEBNext Ultra RNA Library Prep Kit for Illumina with poly-A selection and sequenced on Illumina Novaseq 6000 platform, 150 PE mode.

Small RNA-seq data analysis

Small RNA-seq read trimming and sequencing adaptor removal were done with cutadapt (v3.4)[90]. Only the reads of length 21 – 24 nt were used for subsequent analyses. ShortStack (v4)[91] was used to map small RNA reads to the corresponding genome with default parameters. Bigwig files were created from bam files in a combination of samtools (v1.11)[13] and deepTools (v3.5.6)[92] with parameters 'bigwigAverage, bamCoverage --binSize 10, --normalizeUsing BPM'. Visualization of small RNA and mRNA bigwig files were done with pyGenomeTracks (v3.9)[93].

RNA-seq data analysis

RNA-seq read trimming on quality (Q30) and sequencing adaptor removal were done with Trimmomatic(version 0.32)[76]. Resulting high quality reads from each library were mapped onto the *corresponding* reference genome using STAR (version 2.7.7a)[77], with default parameters. Mapped reads quantification was done by rsem (v1.3.3)[94].

Analysis of flowering time

Ten seeds per each of 14 selected *Hordeum* species accessions (BCC 2067 *H. cordobense*, BCC 2058 *H. californicum*, BCC 2014 *H. muticum*, BCC 2055 *H. jubatum*, BCC 2001 *H. marinum*, BCC 2009 *H. murinum* subsp. *murinum*, BCC 2017 *H. murinum* subsp. *glaucum*, BCC 2023 *H. flexuosum*, BCC 2026 *H. erectifolium*, BCC 2028 *H. pubiflorum*, BCC 2043 *H. pusillum*, BCC 2059 *H. intercedens*, BCC 2065 *H. patagonicum*, and BCC 2069 *H. roshevitzii*) were used. Spring cultivated barley genotypes Morex and Golden Promise and winter genotypes Igri and Antonella were included as controls and were treated identically to wild species. Seeds were stratified at 4 °C for 4 days, germinated in Petri dishes on moist filter paper, and then transferred to soil. Plants were grown under controlled non-vernalizing conditions (21 °C day/18 °C night, 14 h light/10 h dark photoperiod) in a growth chamber and cultivated until senescence. Days to flowering were scored in three-day intervals, and the plants were considered flowering at the first emergence of the awn tips. Plants were imaged using a Nikon D5600 digital camera.

Identification of eog1 in cultivated barley and test for allelism

The *elongated outer glume 1* (*eog1*) mutants from cultivated barley[95], including *eog1.a* ("GSHO 1891"), *eog1.c* ("GSHO 1892"), *eog1.e* ("GSHO 285"), and *macrolepis* (*lep-e.1*; "GSHO 960"), were obtained from the Barley Genetic Stock Collection at Aberdeen, Idaho, USA. Several mutants, namely "M2342", "M9630", "M9815", and "M11405", were identified from independent M_2 lines of the previously established barley "Hatiexi" (HTX) mutagenesis population[96]. Scanning electron microscopy (SEM) observations were conducted using a Hitachi SEM SU8081 (Tokyo, Japan) following a standard protocol[97]. Young inflorescences were sampled and examined. Allelism test was performed by crossing "M9630" with "GSHO 1891" and "GSHO 1892," respectively. All F1 progeny exhibited elongated glumes on the central fertile spikelet (**Extended Data Fig. 10b**). Sequence-specific amplification of the C2H2-zinc finger transcription factor gene was performed using primers "T-70820-F" (5'-TAGGACCGCATATTTCCAGG-3') and "T-70820-R" (5'-GAACAGTTCCAAGGCGAAAG-3'). The resulting amplicons were sequenced using Sanger sequencing.

Identification of spike-specific expressed genes and local alignment

Clean reads were obtained by removing the adapters, reads containing N bases, and low-quality reads using fastp (v0.20.1) with default parameters[98]. Clean reads were mapped to their respective reference genomes via Hisat2 v2.1.0 with default parameters[99]. Read counts were determined from the mapping files via FeatureCounts (v2.0.1) using standard parameters[78], and the transcripts per million (TPM) values of each gene were calculated. Spike-specific expressed genes were defined as having less than 5% TPM in other tissues[100]. The heatmap was generated using the pheatmap R package (https://CRAN.R-project.org/package=pheatmap) based on the normalized TPM values. Venn diagrams were drawn using the eulerr package (v.6.1.0, https://CRAN.R-project.org/package=eulerr). A local BLAST search was performed using the deduced amino acids of HORVU.MOREX.r3.2HG0170820.1 (*EOG1*) as the query against the 24 Hordeum genome sequences to identify homologous proteins. Amino acid sequence alignment was carried out using MUSCLE (v3)[65], and the resulting multiple sequence alignment was visualized with the 'Simple MSA Viewer' function of TBtools II[101].

Data availability

All sequence data collected in this study have been deposited at the European Nucleotide Archive (ENA). Accession codes for individual accessions are listed in **Supplementary Table 18** (genome sequencing) and **Supplementary Table 19** (transcriptome sequencing). The assemblies and annotations are available for download from the IPK Galaxy server (https://galaxy-web.ipk-gatersleben.de/libraries). Gene and TE annotations, multiple sequence alignments and coordinates of conserved regions have been deposited at the Plant Genomics and Phenomics Research Data Repository[102] under DOI (temporary DOI https://doi.ipk-gatersleben.de/DOI/ea92c0f5-a14a-464e-92b9-8ff5b6845ed5/845e60e3-6302-4870-acfa-b120431df22a/2/1847940088 for review purposes). The pangenome graph in GFA format has been deposited in the Plant Genomics and Phenomics Research Data Repository[102] under DOI: http://doi.org/10.5447/ipk/2025/12.

Code availability

Code for gene annotation, pangenomic analyses and graph construction is available at GitHub.
https://github.com/PGSB-HMGU/BPGv2
https://github.com/jia-wu-feng/Pan_Hordeum
https://github.com/cropgeeks/panHordeum

Reference for the Online Methods

1 Feng, J. W. *et al.* A haplotype-resolved pangenome of the barley wild relative. *Nature* (2025). https://doi.org:10.1038/s41586-025-09270-x

2 Siadjeu, C., Pucker, B., Viehöver, P., Albach, D. C. & Weisshaar, B. High Contiguity De Novo Genome Sequence Assembly of Trifoliate Yam (Dioscorea dumetorum) Using Long Read Sequencing. *Genes (Basel)* **11** (2020). https://doi.org:10.3390/genes11030274

3 Jayakodi, M. *et al.* Structural variation in the pangenome of wild and domesticated barley. *Nature* **636**, 654-662 (2024). https://doi.org:10.1038/s41586-024-08187-1

4 Padmarasu, S., Himmelbach, A., Mascher, M. & Stein, N. In Situ Hi-C for Plants: An Improved Method to Detect Long-Range Chromatin Interactions. *Methods Mol Biol* **1933**, 441-472 (2019). https://doi.org:10.1007/978-1-4939-9045-0_28

5 Mascher, M. *et al.* Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond. *Plant J* **76**, 494-505 (2013). https://doi.org:10.1111/tpj.12294

6 Šimková, H., Tulpová, Z. & Cápal, P. Flow Sorting-Assisted Optical Mapping. *Methods Mol Biol* **2672**, 465-483 (2023). https://doi.org:10.1007/978-1-0716-3226-0_28

7 Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170-175 (2021). https://doi.org:10.1038/s41592-020-01056-5

8 Marone, M. P., Singh, H. C., Pozniak, C. J. & Mascher, M. A technical guide to TRITEX, a computational pipeline for chromosome-scale sequence assembly of plant genomes. *Plant Methods* **18**, 128 (2022). https://doi.org:10.1186/s13007-022-00964-1

9 Monat, C. *et al.* TRITEX: chromosome-scale sequence assembly of Triticeae genomes with open-source tools. *Genome Biol* **20**, 284 (2019). https://doi.org:10.1186/s13059-019-1899-5

10 Brassac, J. & Blattner, F. R. Species-Level Phylogeny and Polyploid Relationships in Hordeum (Poaceae) Inferred by Next-Generation Sequencing and In Silico Cloning of Multiple Nuclear Loci. *Syst Biol* **64**, 792-808 (2015). https://doi.org:10.1093/sysbio/syv035

11 Ondov, B. D. *et al.* Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* **17**, 132 (2016). https://doi.org:10.1186/s13059-016-0997-x

12 Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094-3100 (2018). https://doi.org:10.1093/bioinformatics/bty191

13 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021). https://doi.org:10.1093/gigascience/giab008

14 Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H. & Phillippy, A. M. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**, 722-736 (2017). https://doi.org:10.1101/gr.215087.116

15 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010). https://doi.org:10.1093/bioinformatics/btq033

16 Tillich, M. *et al.* GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* **45**, W6-W11 (2017). https://doi.org:10.1093/nar/gkx391

17 Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L. & Zimin, A. MUMmer4: A fast and versatile genome alignment system. *PLoS Comput Biol* **14**, e1005944 (2018). https://doi.org:10.1371/journal.pcbi.1005944

18 Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLoS One* **11**, e0163962 (2016). https://doi.org:10.1371/journal.pone.0163962

19 Mascher, M. *et al.* Long-read sequence assembly: a technical evaluation in barley. *Plant Cell* **33**, 1888-1906 (2021). https://doi.org:10.1093/plcell/koab077

20 Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639-1643 (2021). https://doi.org:10.1093/bioinformatics/btaa1016

21 Kollmar, M. Gene Prediction Methods and Protocols Preface. *Gene Prediction: Methods and Protocols* **1962**, V-V (2019). https://doi.org:Book_Doi 10.1007/978-1-4939-9173-0

22 Nevers, Y. *et al.* Quality assessment of gene repertoire annotations with OMArk. *Nat Biotechnol* **43**, 124-133 (2025). https://doi.org:10.1038/s41587-024-02147-w

23 Sommer, M. J., Zimin, A. V. & Salzberg, S. L. PSAURON: a tool for assessing protein annotation across a broad range of species. *NAR Genom Bioinform* **7**, lqae189 (2025). https://doi.org:10.1093/nargab/lqae189

24 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238 (2019). https://doi.org:10.1186/s13059-019-1832-y

25 Klopfenstein, D. V. *et al.* GOATOOLS: A Python library for Gene Ontology analyses. *Sci Rep* **8**, 10872 (2018). https://doi.org:10.1038/s41598-018-28948-z

26 Lovell, J. T. *et al.* GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *Elife* **11** (2022). https://doi.org:10.7554/eLife.78526

27 Li, P., Quan, X., Jia, G., Xiao, J., Cloutier, S. & You, F. M. RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. *BMC Genomics* **17**, 852 (2016). https://doi.org:10.1186/s12864-016-3197-x

28 Cheng, S. F. *et al.* Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant Journal* **85**, 532-547 (2016). https://doi.org:10.1111/tpj.13121

29 El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Research* **47**, D427-D432 (2019). https://doi.org:10.1093/nar/gky995

30 Gutmann, B. *et al.* The Expansion and Diversification of Pentatricopeptide Repeat RNA-Editing Factors in Plants. *Molecular Plant* **13**, 215-230 (2020). https://doi.org:10.1016/j.molp.2019.11.002

31 Huynh, S. D., Melonek, J., des Francs-Small, C. C., Bond, C. S. & Small, I. A unique C-terminal domain contributes to the molecular function of Restorer-of-fertility proteins in plant mitochondria. *New Phytologist* **240**, 830-845 (2023). https://doi.org:10.1111/nph.19166

32 Appels, R. *et al.* Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **361**, 661-+ (2018). https://doi.org:10.1126/science.aar7191

33 Garrison, E. *et al.* Sequence variation aware genome references and read mapping with the variation graph toolkit. *bioRxiv* (2017). https://doi.org:10.1101/234856

34 Guarracino, A., Heumos, S., Nahnsen, S., Prins, P. & Garrison, E. ODGI: understanding pangenome graphs. *Bioinformatics* **38**, 3319-3326 (2022). https://doi.org:10.1093/bioinformatics/btac308

35 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021). https://doi.org:10.1093/gigascience/giab008

36 Armstrong, J. *et al.* Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246-251 (2020). https://doi.org:10.1038/s41586-020-2871-y

37 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010). https://doi.org:10.1093/bioinformatics/btq033

38 Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology* **16**, 157 (2015). https://doi.org:10.1186/s13059-015-0721-2

39 Liao, W.-W. *et al.* A draft human pangenome reference. *Nature* **617**, 312-324 (2023). https://doi.org:10.1038/s41586-023-05896-x

40 Garrison, E., Kronenberg, Z. N., Dawson, E. T., Pedersen, B. S. & Prins, P. A spectrum of free software tools for processing the VCF variant call format: vcflib, bio-vcf, cyvcf2, hts-nim and slivar. *PLOS Computational Biology* **18**, e1009123 (2022). https://doi.org:10.1371/journal.pcbi.1009123

41 Wickham, H. *ggplot2*. (Springer Cham, 2016).

42 Zytnicki, M. Assessing genome conservation on pangenome graphs with PanSel. *bioRxiv*, 2024.2004.2026.591236 (2024). https://doi.org:10.1101/2024.04.26.591236

43 Sreedasyam, A. *et al.* JGI Plant Gene Atlas: an updateable transcriptome resource to improve functional gene descriptions across the plant kingdom. *Nucleic Acids Research* **51**, 8383-8401 (2023). https://doi.org:10.1093/nar/gkad616

44 Kawahara, Y. *et al.* Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. *Rice (N Y)* **6**, 4 (2013). https://doi.org:10.1186/1939-8433-6-4

45 Ahmed, H. I. *et al.* Einkorn genomics sheds light on history of the oldest domesticated wheat. *Nature* **620**, 830-838 (2023). https://doi.org:10.1038/s41586-023-06389-7

46 Jayakodi, M. *et al.* Structural variation in the pangenome of wild and domesticated barley. *Nature* **636** (2024). https://doi.org:10.1038/s41586-024-08187-1

47 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780 (2013). https://doi.org:10.1093/molbev/mst010

48 Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**, 540-552 (2000). https://doi.org:DOI 10.1093/oxfordjournals.molbev.a026334

49 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014). https://doi.org:10.1093/bioinformatics/btu033

50 Armstrong, J. *et al.* Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246-+ (2020). https://doi.org:10.1038/s41586-020-2871-y

51 McCormick, R. F. *et al.* The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J* **93**, 338-354 (2018). https://doi.org:10.1111/tpj.13781

52 Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530-1534 (2020). https://doi.org:10.1093/molbev/msaa015

53 Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *Bmc Bioinformatics* **19** (2018). https://doi.org:10.1186/s12859-018-2129-y

54 Zhang, C., Nielsen, R. & Mirarab, S. CASTER: Direct species tree inference from whole-genome alignments. *Science* **387** (2025). https://doi.org:10.1126/science.adk9688

55 Kumar, S. *et al.* TimeTree 5: An Expanded Resource for Species Divergence Times. *Mol Biol Evol* **39** (2022). https://doi.org:10.1093/molbev/msac174

56 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591 (2007). https://doi.org:10.1093/molbev/msm088

57 Goel, M., Sun, H., Jiao, W. B. & Schneeberger, K. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol* **20**, 277 (2019). https://doi.org:10.1186/s13059-019-1911-0

58 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution* **26**, 1641-1650 (2009). https://doi.org:10.1093/molbev/msp077

59 Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol Biol Evol* **33**, 1635-1638 (2016). https://doi.org:10.1093/molbev/msw046

60 Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* **51**, 492-508 (2002). https://doi.org:10.1080/10635150290069913

61 Camacho, C. *et al.* BLAST plus : architecture and applications. *Bmc Bioinformatics* **10** (2009). https://doi.org:10.1186/1471-2105-10-421

62 Mahelka, V. *et al.* Multiple horizontal transfers of nuclear ribosomal genes between phylogenetically distinct grass lineages. *P Natl Acad Sci USA* **114**, 1726-1731 (2017). https://doi.org:10.1073/pnas.1613375114

63 Mahelka, V. *et al.* A Panicum-derived chromosomal segment captured by Hordeum a few million years ago preserves a set of stress-related genes. *Plant J* **105**, 1141-1164 (2021). https://doi.org:10.1111/tpj.15167

64 Hibdige, S. G. S., Raimondeau, P., Christin, P. A. & Dunning, L. T. Widespread lateral gene transfer among grasses. *New Phytologist* **230**, 2474-2486 (2021). https://doi.org:10.1111/nph.17328

65 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004). https://doi.org:10.1093/nar/gkh340

66 Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009). https://doi.org:10.1093/bioinformatics/btp348

67 Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. & Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321 (2010). https://doi.org:10.1093/sysbio/syq010

68 Junier, T. & Zdobnov, E. M. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* **26**, 1669-1670 (2010). https://doi.org:10.1093/bioinformatics/btq243

69 Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**, 276-277 (2000). https://doi.org:10.1016/s0168-9525(00)02024-2

70 Lewis, P. O. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst Biol* **50**, 913-925 (2001). https://doi.org:10.1080/106351501753462876

71 Pagel, M. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology* **48**, 612-622 (1999). https://doi.org:Doi 10.1080/106351599260184

72 Revell, L. J. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* **12**, e16505 (2024). https://doi.org:10.7717/peerj.16505

73 Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* **38**, 5825-5829 (2021). https://doi.org:10.1093/molbev/msab293

74 Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* **47**, D309-D314 (2019). https://doi.org:10.1093/nar/gky1085

75 Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021). https://doi.org:10.1016/j.xinn.2021.100141

76 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014). https://doi.org:10.1093/bioinformatics/btu170

77 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013). https://doi.org:10.1093/bioinformatics/bts635
78 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014). https://doi.org:10.1093/bioinformatics/btt656
79 McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288-4297 (2012). https://doi.org:10.1093/nar/gks042
80 Ou, S. *et al.* Differences in activity and stability drive transposable element variation in tropical and temperate maize. *Genome Res* **34**, 1140-1153 (2024). https://doi.org:10.1101/gr.278131.123
81 Wicker, T., Matthews, D. E. & Keller, B. TREP: a database for Triticeae repetitive elements. *Trends Plant Sci* **7**, 561-562 (2002). https://doi.org:10.1016/S1360-1385(02)02372-5
82 Wicker, T. *et al.* Transposable Element Populations Shed Light on the Evolutionary History of Wheat and the Complex Co-Evolution of Autonomous and Non-Autonomous Retrotransposons. *Adv Genet (Hoboken)* **3**, 2100022 (2022). https://doi.org:10.1002/ggn2.202100022
83 Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* **33**, 2776-2778 (2017). https://doi.org:10.1093/bioinformatics/btx299
84 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152 (2012). https://doi.org:10.1093/bioinformatics/bts565
85 Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief Bioinform* **12**, 41-51 (2011). https://doi.org:10.1093/bib/bbq072
86 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997* (2013).
87 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008). https://doi.org:10.1186/gb-2008-9-9-r137
88 Seibt, K. M., Schmidt, T. & Heitkam, T. FlexiDot: highly customizable, ambiguity-aware dotplots for visual sequence analyses. *Bioinformatics* **34**, 3575-3577 (2018). https://doi.org:10.1093/bioinformatics/bty395
89 von Zitzewitz, J. *et al.* Molecular and structural characterization of barley vernalization genes. *Plant Molecular Biology* **59**, 449-467 (2005). https://doi.org:10.1007/s11103-005-0351-2
90 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, 10-12 (2011).
91 Johnson, N. R., Yeoh, J. M., Coruh, C. & Axtell, M. J. Improved Placement of Multi-mapping Small RNAs. *G3 (Bethesda)* **6**, 2103-2111 (2016). https://doi.org:10.1534/g3.116.030452
92 Ramirez, F., Dundar, F., Diehl, S., Gruning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, W187-191 (2014). https://doi.org:10.1093/nar/gku365

93 Lopez-Delisle, L. *et al.* pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**, 422-423 (2021). https://doi.org:10.1093/bioinformatics/btaa692

94 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011). https://doi.org:10.1186/1471-2105-12-323

95 Hansson, M. *et al.* A guide to barley mutants. *Hereditas* **161**, 11 (2024). https://doi.org:10.1186/s41065-023-00304-w

96 Jiang, C. *et al.* A reference-guided TILLING by amplicon-sequencing platform supports forward and reverse genetics in barley. *Plant Commun* **3**, 100317 (2022). https://doi.org:10.1016/j.xplc.2022.100317

97 Kong, X. C. *et al.* The wheat AGL6-like MADS-box gene is a master regulator for floral organ identity and a target for spikelet meristem development manipulation. *Plant Biotechnology Journal* **20**, 75-88 (2022). https://doi.org:10.1111/pbi.13696

98 Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884-i890 (2018). https://doi.org:10.1093/bioinformatics/bty560

99 Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-360 (2015). https://doi.org:10.1038/nmeth.3317

100 Kovacik, M. *et al.* The transcriptome landscape of developing barley seeds. *Plant Cell* **36**, 2512-2530 (2024). https://doi.org:10.1093/plcell/koae095

101 Chen, C. J. *et al.* TBtools-II: A "one for all, all for one"bioinformatics platform for biological big-data mining. *Molecular Plant* **16**, 1733-1742 (2023). https://doi.org:10.1016/j.molp.2023.09.010

102 Arend, D., Junker, A., Scholz, U., Schüler, D., Wylie, J. & Lange, M. PGP repository: a plant phenomics and genomics data publication infrastructure. *Database* **2016** (2016).