Article

Structural and regulatory evolution across the genus *Hordeum*

https://doi.org/10.1038/s41586-026-11332-7

Received: 2 October 2025

Accepted: 19 October 2026

Published online: 18 November 2026

Open access

Check for updates

Jia-Wu Feng (冯佳武)[1,23], Congcong Jiang (蒋枞璁)[2,23], Beáta Strejčková[3,23], Marek Szecówka[3,23], Einar Baldvin Haraldsson[4,23], Thomas Lux[5,23], Chaodan An (安朝丹)[2,21], Micha Bayer[6], Yann J. K. Bertrand[7], Marie Chudecká[3], Simón Villanueva Corrales[7], Radim Čegan[3,8], Anne Fiebig[1], Guangqi Gao (高广奇)[2], Reza Haghi[1], Axel Himmelbach[1], Yanlong Jin (金彦龙)[2,22], Karol Krak[7,9], Takao Komatsuda[10], Václav Mahelka[7], Klaus F. X. Mayer[5,11], Joanna Melonek[12], Ales Pecinka[3], Shun Sakuma[13], Hana Šimková[3], Ian Small[14], Manuel Spannagl[5,15], Nils Stein[1,16], Helena Toegelová[3], Robbie Waugh[6,17,18], Thomas Wicker[19], Hailin Zhang (张海琳)[1], Zihao Zhu (竺子皓)[1], Jan Šafář[3]✉, David Kopecký[3]✉, Frank R. Blattner[1]✉, Ping Yang (杨平)[2]✉, Maria von Korff[4,20]✉ & Martin Mascher[1,15,20]✉

The genus *Hordeum* comprises cultivated barley and its wild relatives, an evolutionarily, ecologically and agriculturally important group of grasses[1,2]. Here we generated a chromosome-scale comparative genomic resource comprising all recognized diploid *Hordeum* species together with representative polyploid lineages. Comparative analyses reveal extensive structural variation, including inversions and chromosomal rearrangements that disrupt proximal collinearity and reflect lineage-specific genome remodelling. Single-copy gene phylogenies identify ancient hybridization, segmental exchange and introgression, revealing pervasive reticulate evolution during *Hordeum* diversification. We uncover widespread horizontal gene transfer from Panicoideae grasses into *Hordeum*, with many transferred genes showing evidence of transcription and functional integration. Retrotransposon proliferation explains genome size variation and provides insights into polyploid genome evolution. Whole-genome alignments reveal contrasting evolutionary dynamics: whereas coding sequences remain highly conserved, regulatory elements undergo rapid turnover, providing a rich source of regulatory diversity for phenotypic evolution. Case studies of vernalization and glume development demonstrate how natural variation in regulatory and coding sequences contributes to flowering time and floral morphology. This comprehensive pangenomic framework establishes *Hordeum* as a model for studying genome evolution, structural variation and the genetic basis of adaptive traits, offering new avenues for barley improvement.

The genus *Hordeum* comprises approximately 33 species and is widely distributed across the temperate regions of Eurasia and the Americas[1–3]. *Hordeum* is classified into two subgenera based on cytogenetic evidence: one containing the H and Xu genomes and the other containing the I and Xa genomes[4]. The basic chromosome number is seven, and, although most species are diploid, tetraploid and hexaploid taxa have also been described. Cultivated barley (*Hordeum vulgare*) is the best known and most extensively studied species, as it represents one of the earliest domesticated cereal crops. Today, barley is the world's fourth most important grain, valued for its use in food, animal feed and brewing[5]. Cultivated barley and its wild progenitor (*H. vulgare* subsp. *spontaneum*) constitute a single inter-fertile gene pool, allowing for frequent gene flow between wild and cultivated forms[6]. This primary gene pool has long been exploited by breeders to introduce desirable traits[6]. *Hordeum bulbosum* has contributed valuable disease resistance genes through introgression[7–10] due to its ability to hybridize directly with cultivated barley. The other *Hordeum* species are not domesticated but have important ecological and agricultural roles in ecosystem stability, environmental adaptation and forage use[11–15]. Together, *Hordeum* is both an agronomically important genus and a powerful model for studying crop evolution, environmental adaptation and resistance to biotic and abiotic stresses. Factors such as incomplete lineage sorting, recurrent hybridization and gene flow partly obscure species boundaries, leaving major evolutionary questions unresolved[1,2,4,16,17]. Recent barley pangenomes have substantially advanced our understanding of genomic variation within cultivated barley and its closest relatives[10,18,19]. However, these studies were largely confined to two species (*H. vulgare* and *H. bulbosum*) and therefore could not resolve genome-wide evolutionary processes operating across the genus, including the origin of large-scale structural variation, ancient hybridization, polyploid genome evolution and the diversification of regulatory loci. Extending comparative genomic analyses to encompass the entire genus is therefore essential to capture the full breadth of diversity within the *Hordeum* gene pool and to facilitate both evolutionary studies and the effective

A list of affiliations appears at the end of the paper.

utilization of wild germplasm for barley improvement. Pangenomics now offers comprehensive resources to address such complexity. In crops such as tomato, potato, maize, grapevine, citrus and cotton, genus-level pangenomes have uncovered extensive structural variations, gene presence–absence and evolutionary dynamics[20–24]. Investigating structural variations helps to explain evolutionary processes and can inform crop breeding strategies[25]. However, the full extent of structural variation within *Hordeum* remains largely unexplored. To address this gap, we generated chromosome-scale genome assemblies representing all recognized diploid *Hordeum* species together with representative polyploid lineages. This comparative genomic framework provides a view of structural and regulatory genome evolution across the genus, yielding new insights into genome evolution and establishing a foundation for the utilization of wild diversity in barley improvement.

Hordeum species assembly and pangenome

To maximize phylogenetic representation across the genus, we selected 27 accessions representing 24 *Hordeum* species, encompassing all recognized diploid genome groups[1] together with representative tetraploid and hexaploid lineages (Extended Data Fig. 1a, Supplementary Table 1 and Supplementary Figs. 1–25). The dataset comprises 22 diploid accessions, 4 tetraploid accessions and 1 hexaploid accession. Twenty-four genome sequences were assembled using PacBio HiFi accurate long reads and Hi-C data (Supplementary Tables 2 and 3 and Supplementary Figs. 26–28), using the TRITEX pipeline[26,27]. The sequence assemblies of *H. vulgare* and *H. bulbosum* were sourced from our previous work[10,28]. The previously published chromosome-level assembly of *Hordeum marinum* H559 (ref. 29) was updated using the TRITEX pipeline, and previously published transcriptome data, Bionano optical maps and Hi-C data of *Hordeum erectifolium* NGB6816 were used[30]. The average contig N50 of the assemblies is 39.24 Mb. Assembly sizes range from 2.90 Gb to 4.48 Gb (mean, 3.92 Gb) in diploid species and from 7.49 Gb to 8.56 Gb (mean, 8.16 Gb) in tetraploid species, whereas the single hexaploid genome has an assembly size of 10.90 Gb. Chloroplast genome assemblies range from 136,335 bp to 137,079 bp (mean, 136,840 bp) (Supplementary Table 1). High assembly quality is supported by an average long terminal repeat (LTR) assembly index (LAI) of 14.79, a consensus quality value (QV) of 67.30 and *k*-mer completeness of 97.22% (Supplementary Table 4). For the heterozygous diploid species *H. bulbosum*, the two phased haplotypes were designated as Hap1 and Hap2. For polyploid species, independently assembled chromosome sets derived from different ancestral genomes were defined as subgenomes and denoted as Sub1, Sub2 and Sub3 according to their phylogenetic relationships with diploid progenitor lineages (Fig. 1a). To assess the gene content in the *Hordeum* pangenome, we annotated all genomes using transcriptome data (Methods and Supplementary Tables 5 and 6). A gene consolidation method was applied to ensure high comparability of gene predictions, including those for *Hordeum parodii* (6x) and *Hordeum gussoneanum* (4x). The number of high-confidence gene models identified ranges from 35,573 for diploid *Hordeum murinum* to 94,943 for hexaploid *H. parodii* (Supplementary Table 7). We assessed the quality of the *Hordeum* gene annotations using BUSCO (v5.8.0) with the poales_odb12 dataset, comprising 6,282 near-universal single-copy orthologues from Poales. On average, 97.35% of BUSCO genes were identified as complete, 0.69% as fragmented and 1.84% as missing (Supplementary Table 7). Independent benchmarking with OMArk and PSAURON further confirmed the high completeness and overall quality of the gene annotations (Supplementary Table 7). Transposable element (TE) annotation using the panEDTA pipeline[31] revealed that transposons constituted on average 86.12% of the genome, ranging from 83.27% to 90.35% (Supplementary Table 8).

Based on the annotated gene sets of all species, we constructed a gene-based *Hordeum* pangenome. Overall, 42.4% of genes were classified as core genes, 56.5% as shell genes and 1.1% as cloud genes (Extended Data Fig. 1b,c). Gene ontology enrichment analysis showed that phosphorylation- and kinase-related functions were enriched among core genes, ATP and RNA binding were enriched among shell genes, and defence response and oxidative stress response were enriched among cloud genes (Supplementary Fig. 29). Nucleotide-binding leucine-rich repeat (NLR) genes are key components of the plant immune system and have essential roles in resistance to fungal and other pathogens[32,33]. A map of their diversity is critical for advancing our understanding of disease resistance mechanisms in crops. Analysis of NLR genes revealed that I genome species contain significantly more NLR genes than the other *Hordeum* genome groups (Welch's *t*-test, $P = 1.8 \times 10^{-6}$) (Extended Data Fig. 2a,b). To further characterize this diversity, we constructed an NLR pangenome across *Hordeum* haploid genome equivalents (haplomes), comprising 800 orthogroups, of which 13.75% were classified as core, 27.75% as shell and 58.50% as cloud gene families (Extended Data Fig. 2c). In addition, NLR gene clusters were significantly less abundant on chromosomes 4H and 5H than on the remaining chromosomes (Wilcoxon test, $P < 2.2 \times 10^{-16}$) (Extended Data Fig. 2d). Similarly, genome-wide analyses of pentatricopeptide repeat (PPR) proteins and mitochondrial transcription termination factors (mTERFs) revealed significant expansions of restorer-of-fertility-like (RFL) subgroups in the I and Xa genome groups relative to the H and Xu genome groups. Because RFL proteins are involved in fertility restoration in plants carrying cytoplasmic male sterility (CMS)[34], these expansions suggest that the smaller RFL repertoires of the H and Xu genome groups are derived rather than ancestral (Supplementary Fig. 30a–e and Supplementary Table 9). Expansion of RFL-PPRs was most pronounced in the I genome group, whereas RFL-mTERFs were significantly expanded in both the I and Xa genome groups. Notably, RFL-mTERF genes clustered on the short arm of chromosome 6H exhibited extensive copy-number variation between Xa–I and H–Xu genome groups (*t*-test, $P = 5.2 \times 10^{-6}$) (Supplementary Fig. 30f). This region includes the *Rfm3* locus, which partially restores fertility in plants carrying the *msm1* cytoplasm used in hybrid barley breeding[35] (Supplementary Fig. 30g).

To assess the feasibility of graph-based pangenome analyses across highly diverged *Hordeum* genomes, we constructed a pangenome graph using Progressive Cactus and hal2vg, comprising 2.53 billion nodes and 3.46 billion edges (68.7 Gb). Approximately 68% of nodes represented single-base variants, illustrating the extensive sequence divergence captured by the graph. Although computationally intensive tasks such as indexing remained infeasible (Methods), we successfully performed global analyses of core, shell and cloud sequence content, graph saturation, sequence conservation and variant diversity (Supplementary Figs. 31–34). These results demonstrate both the potential of graph-based approaches for genus-wide comparative genomics and the need for more scalable methods for graph exploration and interpretation.

Large structural variations reshape proximal region

Previous studies have inferred the evolutionary relationships among *Hordeum* species from plastome and partial gene sequences[1,17]. However, the relationships among some species remained weakly supported. In this study, we reconstructed a maximum-likelihood tree using a genome-wide single-copy gene set. Unlike previous phylogenies[1,17], our species tree resolved all major evolutionary relationships with maximum bootstrap support, particularly among the recently diverged I-genome species (Supplementary Fig. 35). To further evaluate the robustness of these relationships, we reconstructed an independent coalescent-based species tree using ASTRAL-Pro3 (Supplementary Fig. 36), which integrates individual maximum-likelihood gene trees, and a genome-wide phylogeny inferred by CASTER from whole-genome multiple sequence alignments (MSAs) (Supplementary Figs. 37 and 38). We also estimated the divergence times of these species using fourfold degenerate third-codon transversion (4DTV) sites, excluding genomic

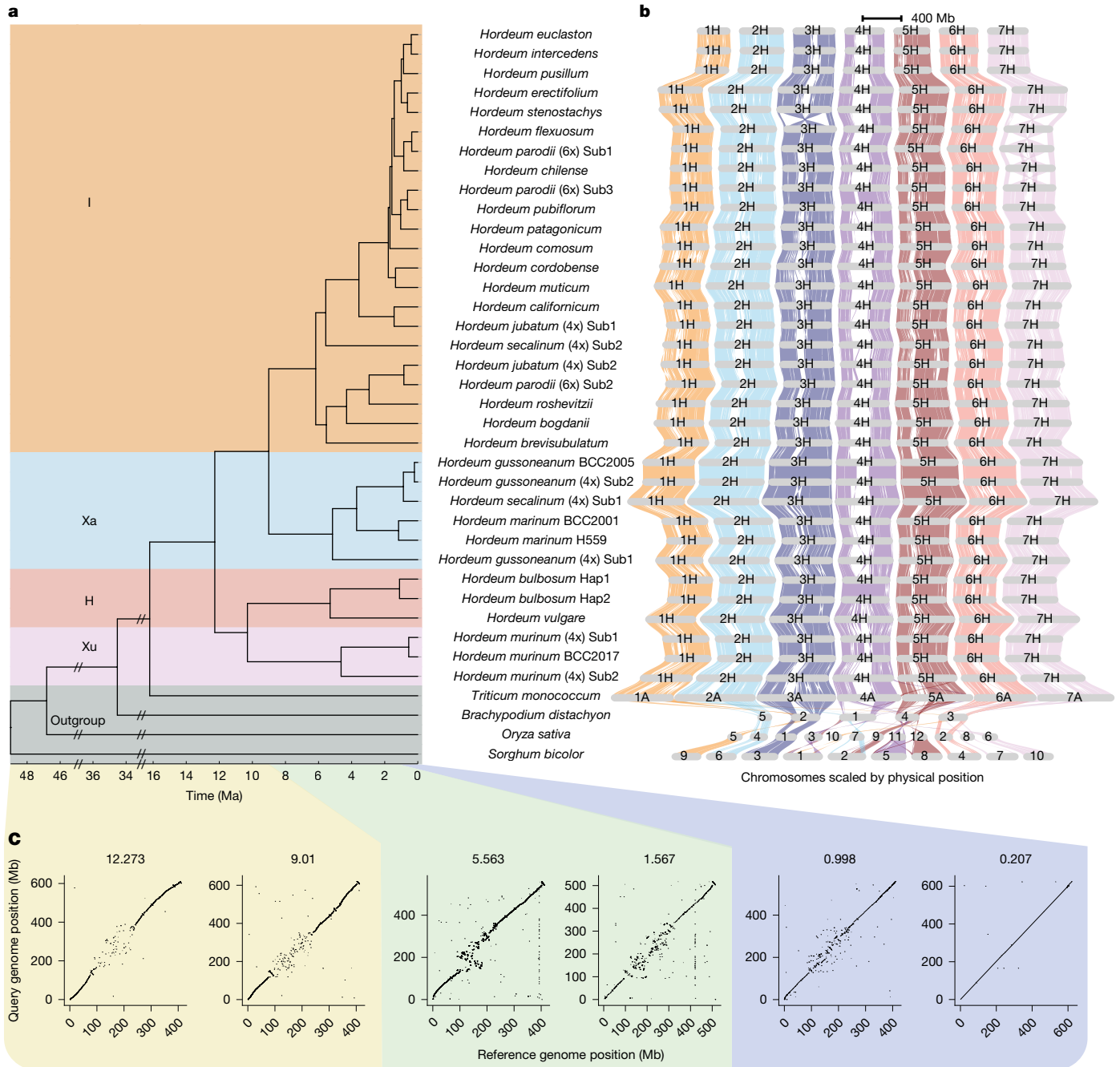


Fig. 1 | Evolution of *Hordeum* species and rapid change in proximal regions. a, Phylogenetic tree of *Hordeum* species. *Sorghum bicolor*, *Oryza sativa*, *Brachypodium distachyon* and *Triticum monococcum* were used as outgroups. Divergence times were inferred using 4DTV sites from genome-wide single-copy genes after excluding genomic regions with strong phylogenetic discordance (nRF > 0.21) identified in Extended Data Fig. 6c. Coloured rectangles represent different genome groups. Hap1 and Hap2 represent phased haplotypes of heterozygous diploid *H. bulbosum*, and Sub1, Sub2 and Sub3 indicate distinct ancestral subgenomes in polyploid species. The background colours represent the genome groups of the species in the following order from top to bottom: I, Xa, H, Xu and outgroup species. **b**, GENESPACE synteny plot among 34 *Hordeum* haplomes and four outgroup genomes. Grey bars represent chromosomes arranged by genomic position. For visualization purposes, the chromosomes of the smaller outgroup genomes (*S. bicolor*, *O. sativa* and *B. distachyon*) were proportionally enlarged by a factor of five. **c**, Dot plot comparisons of chromosome 4H across genomes diverged at different times. Two species were randomly selected from each phylogenetic node for sequence alignment (Supplementary Table 11). Six representative comparisons are shown, with the remaining comparisons presented in Extended Data Fig. 3a. The number displayed at the top of each alignment represents the estimated divergence time between the species shown on the *x* axis and those shown on the *y* axis.

regions identified as putatively derived from ancient hybridization (Supplementary Table 13). We estimated the divergence time between H and Xu at 10.3 million years ago (Ma), whereas I and Xa diverged around 9 Ma. The crown-group age of *Hordeum* dates back approximately 12.27 Ma (Fig. 1a and Supplementary Table 10). The chromosomes of Triticeae species exhibit high collinearity at their distal chromosome regions[36], which is also the case in *Hordeum* (Fig. 1b). However, when aligned, the proximal regions show substantial loss of collinearity and highly fragmented alignments. Whole-genome comparisons across *Hordeum* species revealed that chromosomal rearrangements accumulated progressively with evolutionary divergence, contributing to the loss of collinearity and increasingly fragmented alignments in

proximal chromosomal regions (Fig.1c, Extended Data Fig. 3a, Supplementary Figs. 39–44 and Supplementary Table 11). Quantification of reverse-orientation alignments across all *Hordeum* haplomes further demonstrated that putative inversions accumulated significantly more rapidly in proximal than in distal chromosome regions during *Hordeum* evolution (Supplementary Fig. 45). Among the seven chromosomes, chromosome 4H exhibited the most extensive loss of proximal collinearity (Supplementary Table 12). Recently diverged species retained strong chromosome-scale collinearity despite the presence of numerous structural variants, whereas comparisons between the deeply diverged H–Xu and Xa–I genome groups (more than 9 Ma) showed almost complete fragmentation of proximal collinearity (Fig. 1c, Extended Data Fig. 3a and Supplementary Figs. 39–44). To estimate the size of non-collinear regions, we applied normalized coverage to identify these regions in species that diverged more than 9 Ma (Extended Data Fig. 3c,d, Supplementary Figs. 46–50 and Methods). In total, we determined the genome-wide distribution of non-collinear regions across chromosomes in *Hordeum*, with a cumulative size of 855 Mb identified in the *H. vulgare* genome (Supplementary Table 12). Given the central role of TEs in genome evolution, we compared TE composition between proximal and distal chromosome regions. Proximal regions contained a significantly higher proportion of TEs than distal regions (one-sided paired Wilcoxon test, $P = 9.31 \times 10^{-10}$), primarily due to an increased abundance of LTR/Gypsy retrotransposons (one-sided paired Wilcoxon test, $P = 9.31 \times 10^{-10}$) (Extended Data Fig. 4a,b and Supplementary Fig. 51). Together with the accumulation of large chromosomal rearrangements, these observations motivated a 'stretch and twist' model describing two complementary processes underlying the progressive loss of collinearity. The stretch component represents alignment displacement caused by TE amplification, whereas the twist component represents large-scale chromosomal rearrangements driven by structural variation (Extended Data Fig. 4c).

Large structural variations reveal hybridization

In contrast to the dynamic structural variations found in proximal regions, the distal regions displayed more conserved, shared structural variations. This highlights that the extensive loss of collinearity in proximal regions precludes reliable identification of structural variants across more distantly related *Hordeum* species. Notably, large structural variations in the distal regions were predominantly inversions. On chromosomes 3H and 7H, we identified two large inversions spanning nearly the entire chromosome and present in multiple species (Extended Data Fig. 4d). In total, 1,407 inversions larger than 2 Mb were identified across *Hordeum* species (Extended Data Fig. 4d). The construction of a comprehensive inversion map revealed six inversions on chromosome 1H that were unexpectedly shared across divergent phylogenetic lineages (Fig. 2a and Extended Data Fig. 5). We hypothesized that these large-scale structural variations might have originated from introgression events involving extensive chromosomal segments. We constructed a phylogenetic tree using the single-copy gene set from these regions to test this hypothesis. The resulting tree, resolved with maximum bootstrap support, confirmed that these structural variations are shared between one clade of the I genome group and the H–Xu genome groups (Fig. 2b). Moreover, 10 extant diploid *Hordeum* species diversified within 1.68 million years following the inferred hybridization event, whereas no comparable radiation occurred in other phylogenetic lineages during the same period (Figs. 1a and 2b and Extended Data Fig. 5). Notably, the alternative phylogenetic topology extends across contiguous chromosomal regions of approximately 200 Mb rather than isolated loci, indicating the inheritance of chromosome-scale haplotypes that are more consistent with ancient introgression than with stochastic gene tree discordance or incomplete lineage sorting.

To identify additional signatures of ancient hybridization across the genus, we developed a genome-wide scanning approach based on local phylogenetic discordance among single-copy genes. The single-copy gene set was partitioned into sliding windows, and normalized Robinson–Foulds (nRF) distance was calculated between local phylogenetic trees and the genome-wide species tree (Extended Data Fig. 6a). Peaks in nRF values indicate genomic regions with phylogenetic histories that deviate from the species tree and therefore represent candidate introgression events. Here, we define hybridization footprints as large chromosomal segments shared among multiple descendant species that consistently exhibit an alternative phylogenetic relationship relative to the genome-wide species tree. Using an empirical threshold of 0.21 derived from the genome-wide nRF distribution to distinguish discordance peaks from background variation (Extended Data Fig. 6b and Supplementary Fig. 52), we recovered the chromosome 1H hybridization signal, validating the genome-wide scanning approach. Additionally, we detected more than ten consecutive windows with strong phylogenetic discordance signals at multiple regions (Extended Data Fig. 6c and Supplementary Fig. 52). To further explore the underlying causes of these inconsistencies, we reconstructed local phylogenetic trees using the single-copy gene sets from these regions (Extended Data Fig. 6e,f, Supplementary Figs. 53–58 and Supplementary Table 13). Regions in which local phylogenies consistently supported a specific alternative relationship across a large chromosomal segment were interpreted as evidence of ancient hybridization. By comparing the regions that exhibited phylogenetic discordance with the whole-genome tree, we identified two large-scale segmental exchanges occurring within *Hordeum jubatum*, with maximum bootstrap support (Extended Data Fig. 6e,f and Supplementary Table 13). Surprisingly, this large-scale segmental exchange explains the major duplication and deletion pattern previously identified on chromosome 5H relative to the *H. vulgare* reference genome (Supplementary Figs. 12, 56, 59 and 60). The exchanged regions in the two *H. jubatum* subgenomes are each more closely related to different ancestral genomes, indicating that the apparent duplication and deletion resulted from ancestral segmental exchange rather than independent copy-number changes. (Extended Data Fig. 6d). In the allotetraploid *H. gussoneanum*, extensive segmental exchanges were detected between the two subgenomes on chromosomes 2, 5 and 7 (Extended Data Fig. 6e, Supplementary Figs. 53, 55 and 57 and Supplementary Table 13). In the allohexaploid *H. parodii*, subgenome 1 carried a large chromosomal segment on chromosome 7 derived from *Hordeum stenostachys*, whereas subgenomes 2 and 3 exhibited an additional segmental exchange on chromosome 4 (Extended Data Fig. 6f, Supplementary Figs. 54, 55 and 58 and Supplementary Table 13). Together, these megabase-scale structural variants preserve signatures of ancient interspecific hybridization and large-scale chromosomal exchange that are not evident from genome-wide phylogenies alone, providing new insights into the complex evolutionary history of *Hordeum*.

Horizontal gene transfer

Genetic variation is essential for species evolution. Beyond sexual reproduction, horizontal gene transfer (HGT) serves as an evolutionary shortcut for acquiring novel genes. Traditionally considered rare in vascular plants, genomic advances have revealed widespread occurrences of HGT, challenging prior assumptions[37,38]. Our previous research demonstrated extensive HGT in *Hordeum*. Notably, species within section *stenostachys* (I genome; 16 out of 22 diploid species) possess ribosomal DNA (rDNA) acquired from grasses in the subfamily Panicoideae (for example, *Panicum*, *Paspalum* and *Setaria*)[39,40]. The oldest transfers (from *Panicum*), estimated to have occurred between 5 and 1.7 Ma, resulted in large genomic insertions on chromosome 5H, some spanning over 400 kb and containing TEs and protein-coding genes[41]. Here, we identified Panicoideae-derived nuclear genes across 20 diploid *Hordeum* species, with *H. vulgare* serving as a negative control[39,41]. Our analysis revealed 611 foreign genes that corresponded to 126 orthogroups, predominantly matching *Paspalum* (63%) and *Panicum*

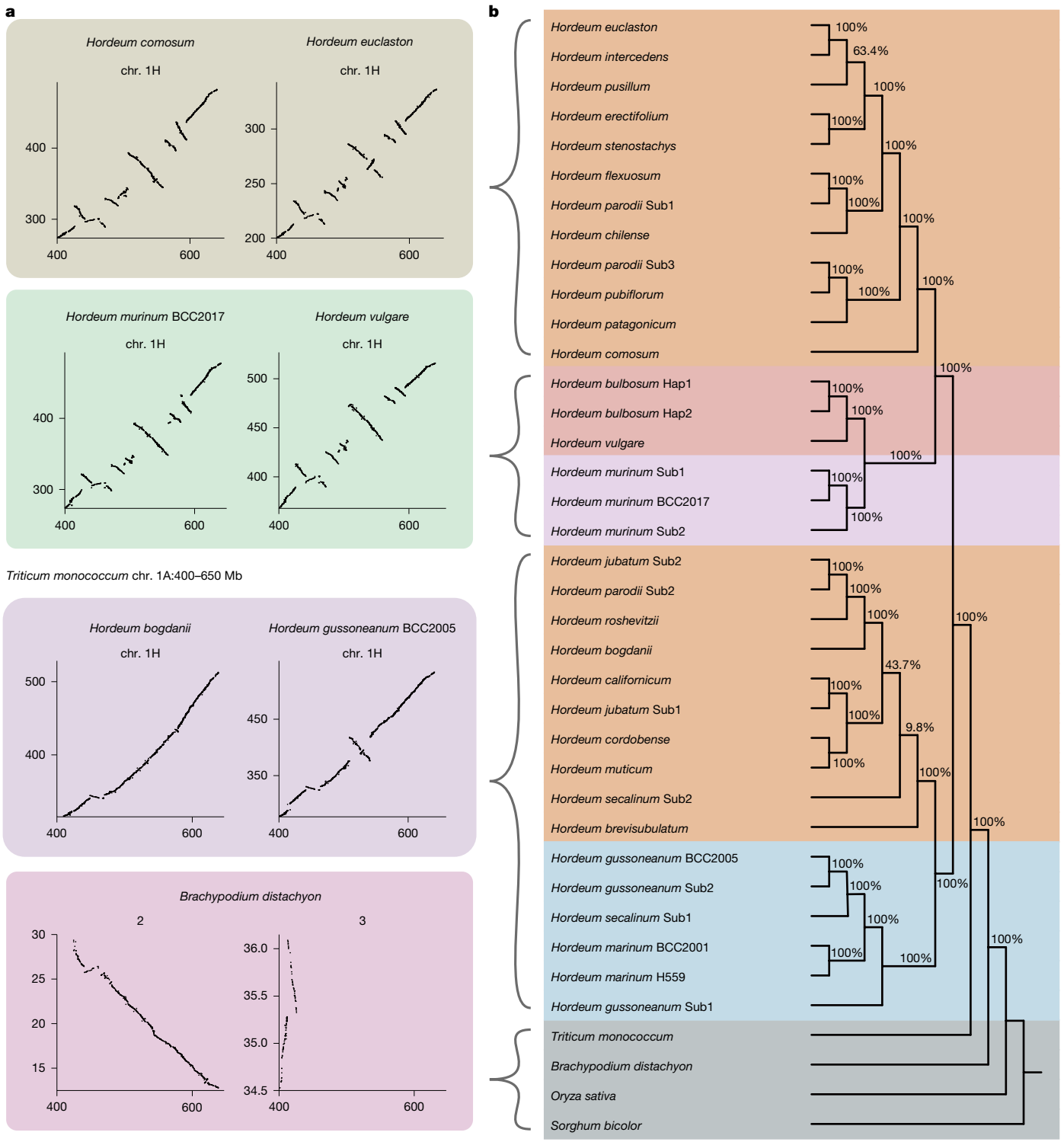


Fig. 2 | Structural variation reveals past introgression in *Hordeum* species. a, Dot plot comparisons of the chr. 1A:400–650 Mb region between *T. monococcum* and seven representative genomes selected for clarity. The remaining genome comparisons are shown in Extended Data Fig. 5. Colours indicate different clusters and are consistent between this panel and Extended Data Fig. 5. **b**, Phylogenetic tree based on single-copy genes within the chr. 1A:400–650 Mb region of *T. monococcum*. Coloured rectangles represent different genome groups. Bootstrap values are shown below each node. Approximately unbiased (AU) test based on concatenated local single-copy genes supports the local tree (p-AU = 1.0) and rejects the reference tree (p-AU = 1.4×10^{-42}).

(22%) (Extended Data Fig. 7a and Supplementary Tables 14 and 15). The extent of HGT varied, with *Hordeum flexuosum* carrying the highest number of foreign genes (68 genes corresponding to 43 orthogroups) (Extended Data Fig. 7a). These findings align with studies indicating that grasses frequently exchange genetic material[42,43]. HGT frequency correlates with *Hordeum* phylogeny, with Mediterranean and West Asian species exhibiting fewer transferred genes than their South American counterparts. This supports the hypothesis that HGT has a cumulative

impact on eukaryotic evolution[44]. The presence of a single panicoid gene in *H. bulbosum* and *H. murinum* can be explained by the introgression of chromosomal segment 1H from I genome species, rather than by an early HGT event predating the diversification of *Hordeum* (Fig. 2b and Extended Data Fig. 6c). Chromosomal mapping indicates that horizontally transferred genes are distributed across all seven *Hordeum* chromosomes, with chromosome 5H carrying the largest number (Extended Data Fig. 7b). These genes are significantly enriched in the distal regions of chromosome arms. Although distal regions generally contain higher local gene densities, negative binomial regression analysis showed that gene density explains only a small fraction of the observed variation in HGT abundance (pseudo-R^2 = 0.034; Supplementary Fig. 61), indicating that higher gene density alone cannot account for the distal enrichment of HGT genes. Although co-transfer of genes remains uncertain, previous research suggests that large fragments of foreign DNA may be transferred together[41,43]. A substantial fraction (82 out of 126) of Panicoideae-derived orthogroups lack native *Hordeum* homologues, indicating novel gene acquisition. Functional annotation suggests that these genes may have roles in key physiological processes such as auxin transport, light response and transcription regulation. RNA sequencing (RNA-seq) analysis detected at least one expressed HGT gene in 73 out of 126 orthogroups (Extended Data Fig. 7c and Supplementary Table 16). The potential adaptive significance of these genes warrants further investigation. Our genus-wide analysis establishes *Hordeum* as a model system for studying the evolutionary impact of HGT.

TE evolution in *Hordeum*

The genome assembly sizes of *Hordeum* species range from 2.90 Gb to 10.9 Gb and are highly correlated with previous flow-cytometric estimates[45] (Fig. 3a). After accounting for the direct effects of polyploidization, the haplomes exhibit pseudomolecule sizes ranging from 2.85 Gb to 4.68 Gb. Apart from a 19 Mb duplication detected in the subgenome of *H. jubatum*, no large duplication events were identified (Supplementary Figs. 59–60). A strong correlation was observed between total TE content and haplome size (Fig. 3b), confirming that TE accumulation has significantly contributed to genome size variation in *Hordeum*. LTR retrotransposons were identified as the primary drivers of these changes, although DNA transposons also had a role (Fig. 3c). Correlation analyses of individual transposon superfamilies showed that multiple TE families contributed to genome size variation, although overall TE abundance remained the strongest predictor of haplome size (Supplementary Fig. 62a). We selected transposons annotated in the TREP database with an average content above 20 Mb and examined their variation across different *Hordeum* haplomes. These TE families clearly differentiated the major evolutionary lineages of *Hordeum* (Supplementary Fig. 62b). To investigate lineage-specific TE dynamics in greater detail, we focused on three LTR retrotransposon families (BARE1, Sabrina and BAGY2) and used TEpop[46] to reconstruct their diversification histories. TE subfamilies correspond closely to the major *Hordeum* clades (Fig. 3d and Extended Data Fig. 8a,b), demonstrating lineage-specific amplification histories that contributed to genome size evolution.

To investigate TE evolution following polyploidization, we compared the TE landscapes of three independently formed tetraploid genomes with those of their diploid progenitors. The three polyploid species exhibited distinct evolutionary trajectories. In *H. murinum*, TE expansion was minimal (Extended Data Fig. 8c and Supplementary Table 17). In *H. jubatum*, both subgenomes were smaller than their diploid progenitors (Fig. 3e and Supplementary Table 17). By contrast, *Hordeum secalinum* experienced extensive TE proliferation, resulting in substantial genome expansion (Extended Data Fig. 8d and Supplementary Table 17). Notably, LTR/Gypsy elements showed the greatest expansion relative to the diploid progenitors and therefore contributed most to genome expansion in *H. secalinum* (Supplementary Table 17). Additionally, the *Vagabond* TE family showed renewed activity following tetraploidization, with significantly higher activity in the tetraploid genome than in either its diploid progenitors or the alternative subgenome (Extended Data Fig. 8e and Supplementary Figs. 63 and 64). To avoid bias introduced by large segmental exchanges between subgenomes, we excluded genomic regions inferred to originate from ancient hybridization before estimating TE insertion ages (Supplementary Table 13 and Supplementary Fig. 65). Although subgenome differentiation alone does not permit precise dating of tetraploid formation, post-polyploid TE dynamics can be reconstructed from TE insertions. Specifically, we inferred the timing of subgenome fusion from the insertion ages of full-length TEs shared exclusively between subgenomes but absent from their respective diploid progenitors (Fig. 3f, Extended Data Fig. 8f,g and Supplementary Figs. 65 and 66). These shared insertions must have occurred after polyploid formation, thereby providing temporal information on the origin of the polyploid genomes and subsequent TE activity. Our findings underscore that TE content is the primary determinant of genome size variation in *Hordeum*. Tetraploidization does not always lead to genome expansion relative to diploid ancestors; we also observed genome contraction. Nevertheless, post-polyploid LTR insertions provided an independent estimate of the timing of polyploid formation and new insights into the evolutionary history of tetraploid and hexaploid *Hordeum* genomes.

Sequence conservation across species

Whole-genome MSAs provide a framework for investigating sequence conservation and divergence across species. We therefore generated a whole-genome MSA of *Hordeum* haplomes using Progressive Cactus[47]. In total, we identified 170.4 Mb of shared aligned sequences across the *Hordeum* genomes (Extended Data Fig. 9a). As expected, rapidly evolving pericentromeric regions showed fewer alignments, whereas chromosome distal ends exhibited higher alignment densities (Extended Data Fig. 9b). The rapid contraction of these alignments is probably due to substantial reshaping of *Hordeum* genomes by TE activity over the past 12 million years. To further investigate sequence conservation across different functional genomic regions, we analysed open chromatin regions identified by assay for transposase-accessible chromatin with sequencing (ATAC–seq) together with annotated genomic features in eight *Hordeum* haplomes (Supplementary Table 18). Untranslated region sequences were the most highly conserved non-coding regions and were lost at a slower rate before the divergence of barley and wheat. By contrast, open chromatin regions exhibited a loss pattern similar to introns, whereas promoter regions underwent more rapid turnover (Extended Data Fig. 10a). Additionally, we used phastCons to identify 74.7 Mb of conserved elements across *Hordeum* genomes (Extended Data Fig. 10b,c). Among all conserved bases, 69.0% overlapped annotated genes and 6.0% overlapped open chromatin regions. Conversely, 44.7% of annotated gene bases and 31.9% of open chromatin bases overlapped conserved elements. Conserved elements were significantly enriched in both annotated genes (odds ratio = 138.0) and open chromatin regions (odds ratio = 27.4), with substantially stronger enrichment in genes, indicating stronger evolutionary constraint on coding sequences than on regulatory regions (Extended Data Fig. 10c).

Overall, our findings indicate that regulatory elements undergo substantially faster evolutionary turnover than coding sequences, generating extensive regulatory variation that may provide a valuable source of alleles for barley improvement. By contrast, the high conservation of coding genes provides opportunities for comparative functional analyses and cross-species transfer of biological knowledge.

Natural variation in vernalization

In cultivated barley, spring and winter growth habits are largely determined by the vernalization genes *VERNALIZATION 1* (*VRN1*), a MADS-box

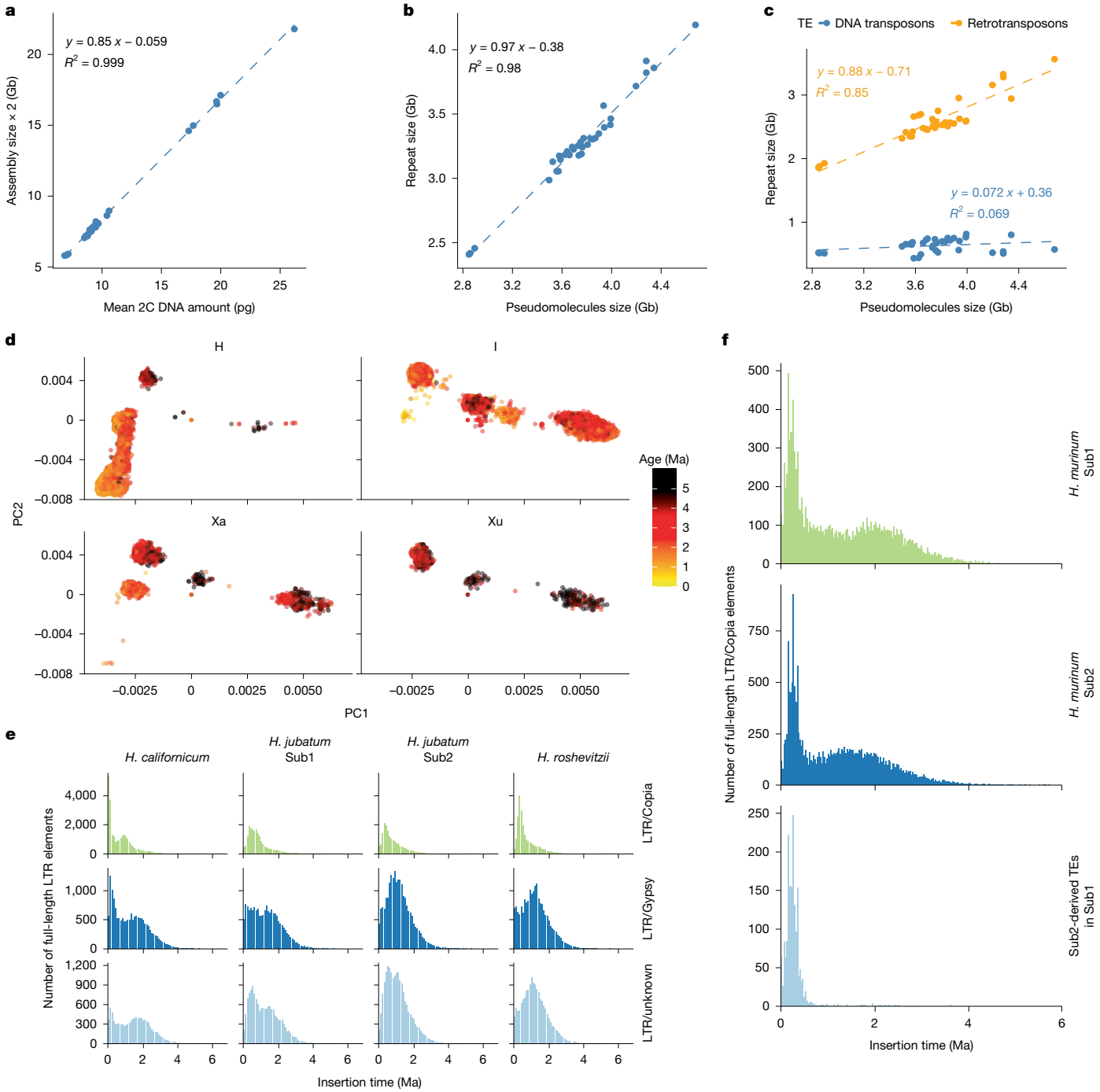


Fig. 3 | TEs reveal genome size variation and tetraploid formation. **a**, Collinearity between assembly size and DNA content ($n = 25$, two-sided linear regression, $P < 2.2 \times 10^{-16}$). **b**, Collinearity between pseudomolecule size and TE content ($n = 34$, two-sided linear regression, $P < 2.2 \times 10^{-16}$). **c**, Correlation between pseudomolecule size and retrotransposon ($n = 34$, two-sided linear regression, $P = 1.24 \times 10^{-14}$) and DNA transposon content ($n = 34$, two-sided linear regression, $P = 0.134$). **d**, Principal component analysis (PCA) of 68,707 full-length *RLG_BAGY2* retrotransposon sequences across *Hordeum* species. Each dot represents one retrotransposon; colours indicate estimated insertion times (Ma). **e**, Distribution of insertion times of full-length LTR elements in *H. jubatum*, *H. californicum* and *Hordeum roshevitzii*. *H. californicum* and *H. roshevitzii* are the respective diploid progenitors of subgenomes 1 and 2 in *H. jubatum*[1]. **f**, Inference of post-allopolyploidization TE insertion dynamics in tetraploid *H. murinum* subsp. *murinum* (BCC 2009) based on full-length LTR/Copia elements shared between subgenomes. Subgenome-derived TEs represent TE insertions detected in one subgenome that are shared with the other subgenome but absent from the corresponding diploid progenitor, reflecting TE accumulation after subgenome fusion. The onset of the insertion peak indicates the estimated timing of allopolyploid formation. *H. murinum* subsp. *glaucum* (BCC 2017) is the diploid progenitor of subgenome 1 in BCC 2009 (ref. 1).

transcription factor, and *VERNALIZATION2* (*VRN2*), which is encoded by a cluster of zinc-finger/CCT (ZCCT) genes. In winter barley, prolonged cold induces VRN1 expression, which represses VRN2 and thereby permits flowering[48,49]. Flowering of winter varieties requires repression of *VRN2* by cold-induced *VRN1* (ref. 49). By contrast, spring barleys often carry deletions in the first intron of *VRN1* and lack *VRN2*, eliminating their vernalization requirement[50–52]. Intermediate facultative types combine the vernalization-sensitive *VRN1* with a non-functional *VRN2* locus. It remains largely unexplored how wild *Hordeum* species regulate flowering across diverse climates.

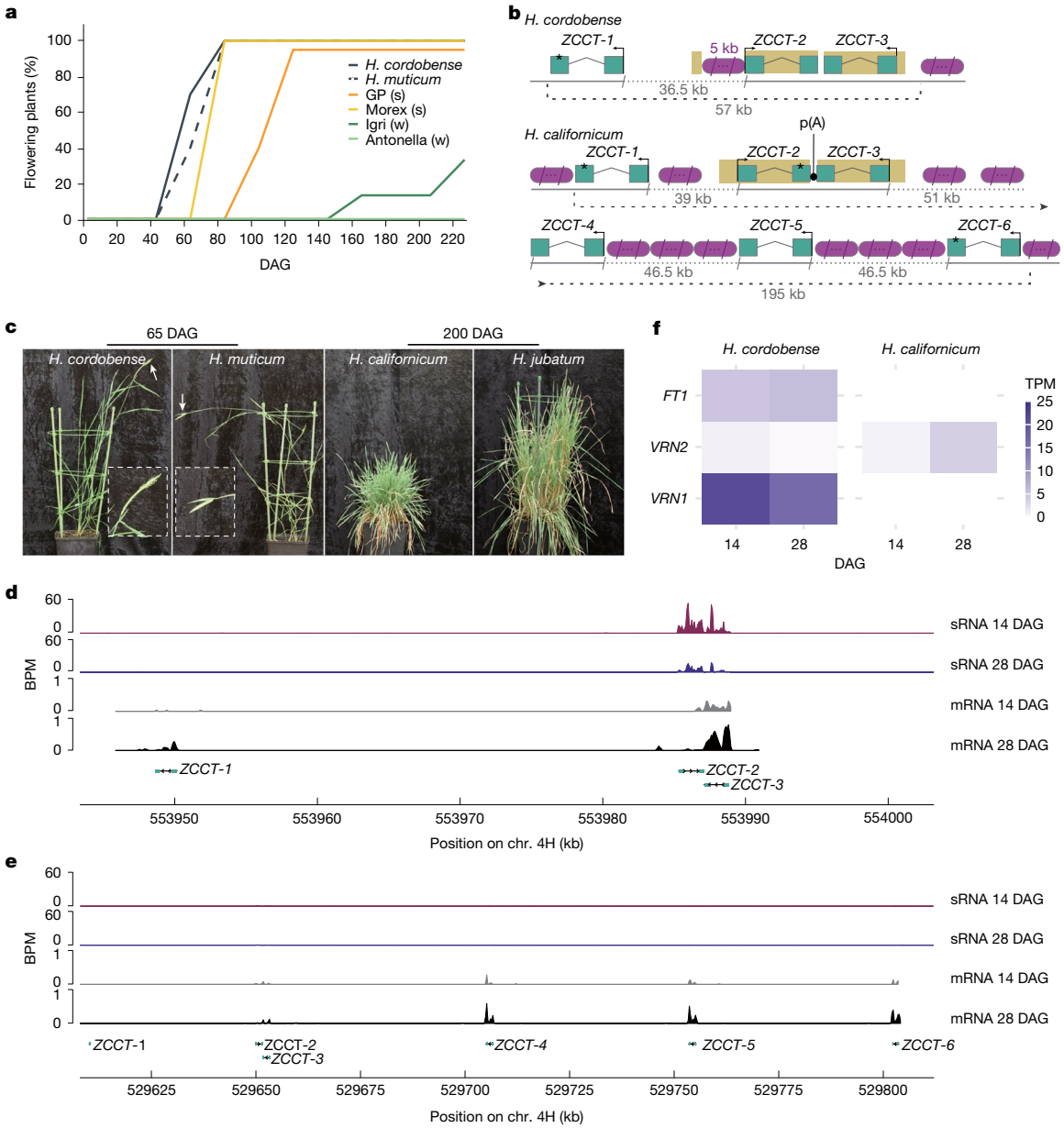


Fig. 4 | Natural variation in vernalization requirement of *Hordeum* species. **a**, Percentage of *H. cordobense* and *H. muticum* plants ($n = 10$ per species) that flower under non-vernalizing conditions. The cultivars Golden Promise (GP) and Morex serve as spring-type (s) controls and cultivars Igri and Antonella serve as winter-type (w) controls. DAG, days after germination. **b**, Structure of the *VRN2* locus in *H. cordobense* with three ZCCT genes, and in *H. californicum* with six ZCCT genes. The scheme is not to scale. Teal rectangles represent exons, asterisks indicate stop codons, purple ovals are TEs, the black dots indicate a polyadenylation (p(A)) signal and sand-coloured regions are inverted repeats. **c**, Phenotypes of four *Hordeum* species with contrasting *VRN2* locus structures at 65 and 200 DAG. Arrows mark visible spikes. **d**,**e**, Mean distribution of RNA-seq and small RNA (sRNA)-sequencing reads over three biological replicates normalized to bins per million (BPM) mapped reads in *H. cordobense* (**d**) and *H. californicum* (**e**) across the *VRN2* locus at 14 and 28 DAG. Read mapping at the *H. cordobense ZCCT-3* locus suggests the presence of improperly spliced, probably aberrant transcripts. **f**, Transcripts per million (TPM) values at 14 and 28 DAG in *H. cordobense* and *H. californicum*.

Examining *VRN1* in *Hordeum* genomes revealed extensive variation in the promoter and the regulatory first intron (5.7 kb to 19.5 kb) caused by TEs (Supplementary Table 19). To assess the functional effect of this variation, we recorded flowering time of 14 wild *Hordeum* species and 4 barley controls under non-vernalizing conditions with a 14 h light:10 h dark photoperiod (Fig. 4a and Extended Data Fig. 11a). Notably, *Hordeum cordobense* and *Hordeum muticum* flowered rapidly (approximately 61 days), surpassing spring barley cultivars (Fig. 4a). Both species share a unique 79-bp deletion in *VRN1* intron 1 critical region[51], suggesting a role in reducing the vernalization requirement (Extended Data Fig. 11b). All species contained the *VRN2* locus with one to seven ZCCT copies (Supplementary Table 19). A key finding was an inverted repeat of two ZCCT genes in *H. cordobense*, *Hordeum californicum*, *H. muticum* and *H. jubatum* (Supplementary Fig. 67). Because inverted repeats may trigger RNA interference (RNAi)[53], we expected silencing of *VRN2* and a spring-type phenotype. Surprisingly, only *H. cordobense* and *H. muticum* flowered early (Fig. 4a,c and Extended Data Fig. 11a). Detailed analysis of *H. cordobense* revealed a non-functional *ZCCT-1* owing to a premature stop codon, and the absence of a polyadenylation signal in the *ZCCT-2*/*ZCCT-3* inverted repeat linker, enabling read-through transcription (detected as an aberrant mRNA transcript), hairpin RNA formation and small RNA production (Fig. 4b,d, Extended Data Fig. 11c and Supplementary Fig. 68). Winter-type

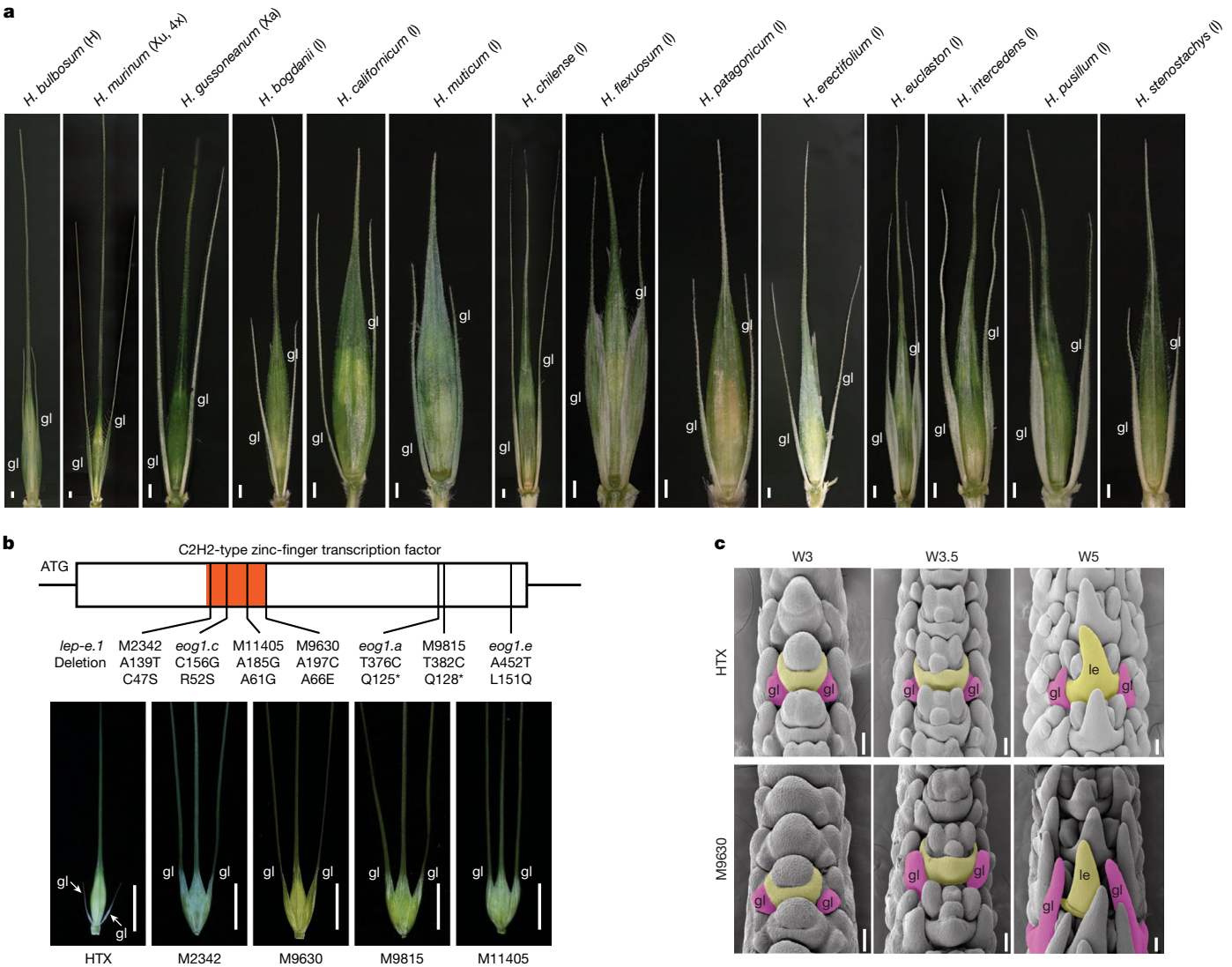


Fig. 5 | Inhibition of glume development by a C2H2-type zinc-finger transcription factor. a, Central spikelet from representative *Hordeum* species exhibiting glume morphology diversity. gl, glume. Scale bars, 500 µm. **b**, Allelic variations of a C2H2-type zinc-finger transcription factor gene (*HORVU.MOREX.r3.2HG0170820.1*) in cultivated barley (*H. vulgare*) lead to glume elongation. The rectangle denotes the exon or full-length coding sequence, with the red rectangle indicating the C2H2-type zinc-finger domain. The asterisk marks a premature stop codon. HTX (Hatiexie) represents the wild type, which exhibits suppressed glume development, whereas the mutants exhibit elongation in glume growth. The mutant allele *lep-e.1* (ref. 59) carries a deletion allele of the *EOG1* gene. Scale bars, 1 cm. **c**, Differentiation on the spikelet morphogenesis along Waddington stage between the wild-type HTX and the induced mutant M9630. W3, glume primordial present; W3.5, floret primordial present; W5, carpel extending. le, lemma. Scale bars, 100 µm.

H. californicum carried additional functional ZCCT copies and a polyadenylation signal in the inverted repeat linker that probably prevented RNAi, thus maintaining the VRN2 repressive function (Fig. 4b,e). This aligns with the expression of *VRN1*, *VRN2* and their target *FLOWERING LOCUS T1* (*FT1*) in the analysed species (Fig. 4f). These findings highlight that genetic diversity shapes vernalization in *Hordeum*. The studied spring-like accessions carry the 79-bp *VRN1* deletion alongside disrupted or silenced *VRN2*, whereas winter-like genotypes retain functional ZCCT copies. Together, our results indicate that adaptation to spring growth habit in wild *Hordeum* species has targeted the same genes and variant types as in cultivated barley, pointing to conserved mechanisms of flowering-time adaptation.

Conserved control of glume development

The glume is a distinctive floral organ in the grass family, which lacks a direct counterpart in dicotyledonous plants[54]. It appears as one or multiple pairs of bract-like structures subtending single or multiple fertile florets. Glume size and shape vary widely, from large lemma-like structures in *Triticum* species to rudimentary axillary buds in *Oryza*. In wild grasses, glume morphology influences seed dispersal, reproductive success and defence against biotic stress[55]. In cultivated cereals, domestication has favoured reduced glume size and roughness, facilitating easier threshing. In *Hordeum*, each floret possesses a pair of glumes positioned on the periaxial side, with considerable variation in length, width and awn extension across species (Fig. 5a). Transcriptomic profiling of five diploid species carrying the I genome (*H. erectifolium*, *Hordeum pusillum*, *Hordeum intercedens H. stenostachys* and *Hordeum euclaston*) identified spike-specific genes expressed in young spikes at early (Waddington stages 3–6 (W3–6)) and late (W9–10) developmental stages[56] (Supplementary Table 21). Genes were classified as spike-specific if their transcript abundance in the remaining 11 out of 13 analysed tissues was less than 5% of their combined expression in the 2 spike samples. Five genes were consistently expressed across species, suggesting conserved regulatory mechanisms (Extended Data Fig. 12a and Supplementary Table 21). These genes encode a C2H2-type

zinc-finger transcription factor, an auxin response factor, a TCP transcription factor, a subtilisin protease (SBT) and a MADS-box transcription factor. Three were predominantly expressed during early spike development, one at later stages and one at both stages (Extended Data Fig. 12b). In rice, loss-of-function mutations in *NONSTOP GLUMES1* (*NSG1*), the orthologue of the C2H2-type zinc-finger transcription factor, result in uncontrolled glume growth[57]. Amino acid variations in this gene across *Hordeum* species along with glume variations suggested a relevance in regulating glume development (Fig. 5a and Supplementary Fig. 69). Mutations in its cultivated barley orthologue (*EOG1*) lead to the elongated outer glume (*eog1*) phenotype (Fig. 5b). These mutants develop enlarged glumes on the central fertile spikelet, with glume awns extending to lemma-awn length, giving the spike a brushy appearance (Extended Data Fig. 12c,d). Scanning electron microscopy (SEM) revealed developmental divergence between wild-type and *eog1* mutants as early as W3.5 (Fig. 5c). By integrating comparative genomics across *Hordeum* with tissue-specific expression data and cultivated barley mutant resources, we identify EOG1 as a conserved regulator of glume development.

Discussion

The chromosome-scale genome assemblies presented here provide a foundation for investigating the evolutionary history of the genus *Hordeum*. Chromosome-scale resolution enables us to investigate how structural variations have shaped genome evolution, particularly in disrupting chromosomal collinearity. Traditionally, grass genomes were thought to possess relatively stable chromosome structures[58]; however, our findings challenge this assumption. Extensive structural variations and chromosomal rearrangements have contributed substantially to genomic divergence, ultimately resulting in the near-complete loss of collinearity in proximal regions. These rearrangements not only mark evolutionary divergence but also provide insights into the processes underlying speciation within *Hordeum*. Structural variations accumulated at markedly different rates across chromosome regions. Whereas proximal regions exhibit high variability, distal regions evolve more slowly. This uneven evolutionary rate has preserved footprints of ancient hybridization events across the genus, offering a historical record of interspecific gene flow. These findings provide empirical evidence for ancient hybridization and genome fusion during *Hordeum* evolution and establish a framework for identifying compatible genomic segments in future interspecific breeding efforts.

Nevertheless, whether the large inversions identified in this study contribute to adaptive evolution, reproductive isolation or other phenotypic consequences remains unknown and warrants future functional investigation. Whole-genome MSAs enabled us to compare genome-wide conservation across *Hordeum* species. Compared with animal genomes, which often display strong interspecies conservation, plant genomes—especially within rapidly evolving genera such as *Hordeum*—undergo dynamic changes driven by short generation times and rapid transposon expansion. These processes fuel the ongoing remodelling of plant genomes, resulting in high variability even among closely related taxa. Despite this rapid evolution, many genes remain conserved, whereas regulatory elements exhibit considerable turnover. This contrast offers unique opportunities: lineage-specific regulatory innovation can be harnessed for crop improvement, while core gene functions are retained. For example, regulatory elements associated with vernalization responses were found to vary across *Hordeum* species, suggesting that novel regulatory variants could be introduced into cultivated barley to fine-tune flowering behaviour. Beyond expanding genomic resources for *Hordeum*, this work provides a framework for investigating structural genome evolution, reconstructing complex evolutionary histories and harnessing genetic and regulatory diversity for barley improvement.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-026-11332-7.

1. 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 https://doi.org/10.1093/sysbio/syv035 (2015).
2. Blattner, F. R., Pleines, T. & Jakob, S. S. Rapid radiation in the barley genus *Hordeum* (Poaceae) during the Pleistocene in the Americas. In *Evolution in Action: Case Studies in Adaptive Radiation, Speciation and the Origin of Biodiversity* (ed. Glaubrecht, M.) 17–33 (Springer, 2010).
3. von Bothmer, R. v. et al. *An Ecogeographical Study of the Genus Hordeum* (IBPGR, 1991).
4. Blattner, F. R. Progress in phylogenetic analysis and a new infrageneric classification of the barley genus *Hordeum* (Poaceae: Triticeae). *Breeding Sci.* **59**, 471–480 https://doi.org/10.1270/jsbbs.59.471 (2009).
5. Jiang, C. et al. Barley2035: a decadal vision for barley research and breeding. *Mol. Plant* **18**, 195–218 https://doi.org/10.1016/j.molp.2024.12.009 (2025).
6. Dawson, I. K. et al. Barley: a translational model for adaptation to climate change. *New Phytol.* **206**, 913–931 https://doi.org/10.1111/nph.13266 (2015).
7. Hoseinzadeh, P., Ruge-Wehling, B., Schweizer, P., Stein, N. & Pidon, H. High resolution mapping of a *Hordeum bulbosum*-derived powdery mildew resistance locus in barley using distinct homologous introgression lines. *Front. Plant Sci.* **11**, 225 https://doi.org/10.3389/fpls.2020.00225 (2020).
8. Wendler, N. et al. Bulbosum to Go: a toolbox to utilize *Hordeum vulgare/bulbosum* introgressions for breeding and beyond. *Mol. Plant* **8**, 1507–1519 https://doi.org/10.1016/j.molp.2015.05.004 (2015).
9. Haas, M. & Mascher, M. Use of the secondary gene pool of barley in breeding improved varieties. *Burl. Dodds Agr. Sci.* **60**, 55–74 https://doi.org/10.19103/As.2019.0051.02 (2019).
10. Feng, J. W. et al. A haplotype-resolved pangenome of the barley wild relative *Hordeum bulbosum*. *Nature* **645**, 429–438 https://doi.org/10.1038/s41586-025-09270-x (2025).
11. Iannone, L. J., Irisarri, J. G. N., Mc Cargo, P. D., Perez, L. I. & Gundel, P. E. Occurrence of *Epichloë* fungal endophytes in the sheep-preferred grass *Hordeum comosum* from Patagonia. *J. Arid. Environ.* **115**, 19–26 https://doi.org/10.1016/j.jaridenv.2014.12.008 (2015).
12. Potthoff, M. et al. Soil biological and chemical properties in restored perennial grassland in California. *Restor. Ecol.* **13**, 61–73 https://doi.org/10.1111/j.1526-100X.2005.00008.x (2005).
13. Janousek, C. N. & Folger, C. L. Inter-specific variation in salinity effects on germination in Pacific Northwest tidal wetland plants. *Aquat. Bot.* **111**, 104–111 https://doi.org/10.1016/j.aquabot.2013.06.009 (2013).
14. Yuan, J., Li, H. & Yang, Y. The compensatory tillering in the forage grass *Hordeum brevisubulatum* after simulated grazing of different severity. *Front. Plant Sci.* **11**, 792 https://doi.org/10.3389/fpls.2020.00792 (2020).
15. Israelsen, K. R., Ransom, C. V. & Waldron, B. L. Salinity tolerance of foxtail barley (*Hordeum jubatum*) and desirable pasture grasses. *Weed Sci.* **59**, 500–505 https://doi.org/10.1614/Ws-D-10-00164.1 (2011).
16. Jin, Y. L., Du, X., Jiang, C. C., Ji, W. Q. & Yang, P. Disentangling sources of gene tree discordance for species via target-enriched sequencing assays. *Mol. Phylogenet. Evol.* https://doi.org/10.1016/j.ympev.2024.108160 (2024).
17. Jakob, S. S. & Blattner, F. R. A chloroplast genealogy of *Hordeum* (Poaceae): long-term persisting haplotypes, incomplete lineage sorting, regional extinction, and the consequences for phylogenetic inference. *Mol. Biol. Evol.* **23**, 1602–1612 https://doi.org/10.1093/molbev/msl018 (2006).
18. Jayakodi, M. et al. The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature* https://doi.org/10.1038/s41586-020-2947-8 (2020).
19. Jayakodi, M. et al. Structural variation in the pangenome of wild and domesticated barley. *Nature* **636**, 654–662 https://doi.org/10.1038/s41586-024-08187-1 (2024).
20. Tang, D. et al. Genome evolution and diversity of wild and cultivated potatoes. *Nature* **606**, 535–541 https://doi.org/10.1038/s41586-022-04822-x (2022).
21. Benoit, M. et al. Pan-genetics reveals paralogues as contingencies in crop engineering. *Nature* https://doi.org/10.1038/s41586-025-08619-6 (2025).
22. Chen, L. et al. Genome sequencing reveals evidence of adaptive variation in the genus *Zea*. *Nat. Genet.* **54**, 1736–1745 https://doi.org/10.1038/s41588-022-01184-y (2022).
23. Wang, M. et al. Genomic innovation and regulatory rewiring during evolution of the cotton genus *Gossypium*. *Nat. Genet.* **54**, 1959–1971 https://doi.org/10.1038/s41588-022-01237-2 (2022).
24. Guo, L. et al. Super pangenome of *Vitis* empowers identification of downy mildew resistance genes for grapevine improvement. *Nat. Genet.* **57**, 741–753 https://doi.org/10.1038/s41588-025-02111-7 (2025).
25. Zhou, Y. et al. Graph pangenome captures missing heritability and empowers tomato breeding. *Nature* **606**, 527–534 https://doi.org/10.1038/s41586-022-04808-9 (2022).
26. 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 https://doi.org/10.1186/s13007-022-00964-1 (2022).
27. Monat, C. et al. TRITEX: chromosome-scale sequence assembly of Triticeae genomes with open-source tools. *Genome Biol.* **20**, 284 https://doi.org/10.1186/s13059-019-1899-5 (2019).
28. Mascher, M. et al. Long-read sequence assembly: a technical evaluation in barley. *Plant Cell* **33**, 1888–1906 https://doi.org/10.1093/plcell/koab077 (2021).

29. Kuang, L. H. et al. The genome and gene editing system of sea barleygrass provide a novel platform for cereal domestication and stress tolerance studies. *Plant Commun*. https://doi.org/10.1016/j.xplc.2022.100333 (2022).
30. Haraldsson, E. B. et al. A chromosome-scale genome assembly of *Hordeum erectifolium*: genomic, transcriptomic and anatomical adaptations to drought in a wild barley relative. *New Phytol*. **250**, 2652–2669 https://doi.org/10.1111/nph.71091 (2026).
31. Ou, S. et al. Differences in activity and stability drive transposable element variation in tropical and temperate maize. *Genome Res*. **34**, 1140–1153 https://doi.org/10.1101/gr.278131.123 (2024).
32. Jones, J. D. G., Staskawicz, B. J. & Dangl, J. L. The plant immune system: from discovery to deployment. *Cell* **187**, 2095–2116 https://doi.org/10.1016/j.cell.2024.03.045 (2024).
33. van Wersch, S., Tian, L., Hoy, R. & Li, X. Plant NLRs: the whistleblowers of plant immunity. *Plant Commun*. **1**, 100016 https://doi.org/10.1016/j.xplc.2019.100016 (2020).
34. Melonek, J. & Small, I. Triticeae genome sequences reveal huge expansions of gene families implicated in fertility restoration. *Curr. Opin. Plant Biol*. **66**, 102166 https://doi.org/10.1016/j.pbi.2021.102166 (2022).
35. Bernhard, T., Koch, M., Snowdon, R. J., Friedt, W. & Wittkop, B. Undesired fertility restoration in *msm1* barley associates with two mTERF genes. *Theor. Appl. Genet*. **132**, 1335–1350 https://doi.org/10.1007/s00122-019-03281-9 (2019).
36. Rabanus-Wallace, M. T. et al. Chromosome-scale genome assembly provides insights into rye biology, evolution and agronomic potential. *Nat. Genet*. **53**, 564 https://doi.org/10.1038/s41588-021-00807-0 (2021).
37. Wickell, D. A. & Li, F. W. On the evolutionary significance of horizontal gene transfers in plants. *New Phytol*. **225**, 113–117 https://doi.org/10.1111/nph.16022 (2020).
38. Chen, R. et al. Adaptive innovation of green plants by horizontal gene transfer. *Biotechnol. Adv*. **46**, 107671 https://doi.org/10.1016/j.biotechadv.2020.107671 (2021).
39. Mahelka, V. et al. Multiple horizontal transfers of nuclear ribosomal genes between phylogenetically distinct grass lineages. *Proc. Natl Acad. Sci. USA* **114**, 1726–1731 https://doi.org/10.1073/pnas.1613375114 (2017).
40. Krak, K., Caklová, P, Kopecky, D., Blattner, F. R. & Mahelka, V. Horizontally acquired nrDNAs persist in low amounts in host *Hordeum genomes* and evolve independently of native nrDNA. *Front. Plant Sci*. https://doi.org/10.3389/fpls.2021.672879 (2021).
41. 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 https://doi.org/10.1111/tpj.15167 (2021).
42. Hibdige, S. G. S., Raimondeau, P., Christin, P. A. & Dunning, L. T. Widespread lateral gene transfer among grasses. *New Phytol*. **230**, 2474–2486 https://doi.org/10.1111/nph.17328 (2021).
43. Dunning, L. T. et al. Lateral transfers of large DNA fragments spread functional genes among grasses. *Proc. Natl Acad. Sci. USA* **116**, 4416–4425 https://doi.org/10.1073/pnas.1810031116 (2019).
44. Martin, W. F. Too much eukaryote LGT. *Bioessays* https://doi.org/10.1002/bies.201700115 (2017).
45. Jakob, S. S., Meister, A. & Blattner, F. R. The considerable genome size variation of *Hordeum* species (poaceae) is linked to phylogeny, life form, ecology, and speciation rates. *Mol. Biol. Evol*. **21**, 860–869 https://doi.org/10.1093/molbev/msh092 (2004).
46. 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*. **3**, 2100022 https://doi.org/10.1002/ggn2.202100022 (2022).
47. Armstrong, J. et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246 https://doi.org/10.1038/s41586-020-2871-y (2020).
48. Trevaskis, B., Hemming, M. N., Peacock, W. J. & Dennis, E. S. *HvVRN2* responds to daylength, whereas *HvVRN1* is regulated by vernalization and developmental status. *Plant Physiol*. **140**, 1397–1405 https://doi.org/10.1104/pp.105.073486 (2006).
49. Deng, W. W. Direct links between the vernalization response and other key traits of cereal crops. *Nat. Commun*. https://doi.org/10.1038/ncomms6882 (2015).
50. Fu, D. L. et al. Large deletions within the first intron in *VRN-1* are associated with spring growth habit in barley and wheat. *Mol. Genet. Genomics* **273**, 54–65 https://doi.org/10.1007/s00438-004-1095-4 (2005).
51. von Zitzewitz, J. et al. Molecular and structural characterization of barley vernalization genes. *Plant Mol. Biol*. **59**, 449–467 https://doi.org/10.1007/s11103-005-0351-2 (2005).
52. Dubcovsky, J., Chen, C. L. & Yan, L. L. Molecular characterization of the allelic variation at the *VRN-H2* vernalization locus in barley. *Mol. Breeding* **15**, 395–407 https://doi.org/10.1007/s11032-005-0084-6 (2005).
53. Hung, Y. H. & Slotkin, R. K. The initiation of RNA interference (RNAi) in plants. *Curr. Opin. Plant Biol*. https://doi.org/10.1016/j.pbi.2021.102014 (2021).
54. Ciaffi, M., Paolacci, A. R., Tanzarella, O. A. & Porceddu, E. Molecular aspects of flower development in grasses. *Sex. Plant Reprod*. **24**, 247–282 https://doi.org/10.1007/s00497-011-0175-y (2011).
55. Xie, P., Wu, Y. R. & Xie, Q. Evolution of cereal floral architecture and threshability. *Trends Plant Sci*. **28**, 1438–1450 https://doi.org/10.1016/j.tplants.2023.08.003 (2023).
56. Waddington, S. R., Cartwright, P. M. & Wall, P. C. A quantitative scale of spike initial and pistil development in barley and wheat. *Ann. Bot*. **51**, 119–130 https://doi.org/10.1093/oxfordjournals.aob.a086434 (1983).
57. Zhuang, H. et al. *NONSTOP GLUMES1* encodes a C2H2 zinc finger protein that regulates spikelet development in rice. *Plant Cell* **32**, 392–413 https://doi.org/10.1105/tpc.19.00682 (2020).
58. Moore, G., Devos, K. M., Wang, Z. & Gale, M. D. Cereal genome evolution: grasses, line up and form a circle. *Curr. Biol*. **5**, 737–739 https://doi.org/10.1016/S0960-9822(95)00148-5 (1995).
59. Hansson, M. et al. A guide to barley mutants. *Hereditas* **161**, 11 https://doi.org/10.1186/s41065-023-00304-w (2024).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026

[1]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Seeland, Germany. [2]State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China. [3]Centre of Plant Structural and Functional Genomics, Institute of Experimental Botany, Czech Academy of Sciences, Olomouc, Czech Republic. [4]Institute of Plant Genetics, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany. [5]Plant Genome and Systems Biology (PGSB), Helmholtz Center Munich–German Research Center for Environmental Health, Neuherberg, Germany. [6]International Barley Hub (IBH), James Hutton Institute (JHI), Dundee, UK. [7]Institute of Botany, Czech Academy of Sciences, Průhonice, Czech Republic. [8]Department of Plant Developmental Genetics, Institute of Biophysics, Czech Academy of Sciences, Brno, Czech Republic. [9]Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic. [10]Shandong Academy of Agricultural Sciences (SAAS), Jinan, China. [11]School of Life Sciences, Technical University Munich, Freising, Germany. [12]Division of Plant Sciences, Research School of Biology, The Australian National University, Acton, Australian Capital Territory, Australia. [13]Faculty of Agriculture, Tottori University, Tottori, Japan. [14]School of Molecular Sciences, The University of Western Australia, Perth, Western Australia, Australia. [15]Centre for Crop and Food Innovation, Food Futures Institute, Murdoch University, Murdoch, Western Australia, Australia. [16]Institute of Agricultural and Nutritional Sciences, Martin Luther University Halle-Wittenberg, Halle, Germany. [17]School of Agriculture, Food and Wine, University of Adelaide, Urrbrae, South Australia, Australia. [18]School of Life Sciences, University of Dundee, Dundee, UK. [19]Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland. [20]Cluster of Excellence on Plant Sciences–SMART Plants for Tomorrow's Needs, Düsseldorf, Germany. [21]Present address: Shiping County Government Offices Administration Bureau, Shiping, China. [22]Present address: Key Laboratory of Oasis Eco-Agriculture, College of Agriculture, Shihezi University, Shihezi, China. [23]These authors contributed equally Jia-Wu Feng, Congcong Jiang, Beáta Strejčková, Marek Szecówka, Einar Baldvin Haraldsson, Thomas Lux. ✉e-mail: Safar@ueb.cas.cz; kopecky@ueb.cas.cz; blattner@ipk-gatersleben.de; yangping@caas.cn; Maria.Korff.Schmising@uni-duesseldorf.de; mascher@ipk-gatersleben.de

Methods

Plant material

Plant material was obtained from genebanks at Institut für Pflanzgenetik und Kulturpflanzenforschung (IPK) Gatersleben, the US 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 high molecular weight (HMW) DNA was isolated from young leaf tissue according to published protocols[10,60]. Construction of HiFi libraries and Circular Consensus Sequencing using the PacBio Sequel IIe instrument (Pacific Biosciences) were essentially as described previously[19]. HiFi sequencing was done at IPK Gatersleben, Genomics and Transcriptomics Laboratory, Heinrich Heine University, Berry Genomics Co., Novogene and the University of Wisconsin Biotechnology Center DNA Sequencing Facility.

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. Libraries were prepared as described[61]. Samples were quantified[62] and sequenced with standard protocols from Illumina using Illumina NovaSeq 6000 devices (Illumina) at IPK Gatersleben and Novogene.

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.[63]. A total of 525 ng of HMW DNA was directly labelled at DLE-1 recognition sites using the standard Bionano Prep Direct Label and Stain protocol (Bionano) and analysed on the Bionano Saphyr platform. The resulting dataset, comprising 1.5 Tb 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 (v3.6.1_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 5) using the RNAeasy Plant Mini Kit (Qiagen, Hilden, Germany) or with Trizol reagent (Invitrogen) according to the manufacturers' protocols. RNA-seq was performed on replicated bulk sample representing a specific developmental stage or tissue type. RNA-seq of individual samples and PacBio IsoSeq of RNA pools was performed as described previously[10]. Library construction and sequencing on Illumina NovaSeq 6000 devices (Illumina RNA-seq) and PacBio Sequel IIe (instruments) was done at IPK Gatersleben, Biomarker Technologies (BMK), the Genomics and Transcriptomics Laboratory at Heinrich Heine University Düsseldorf, Kazusa DNA Research Institute or Berry Genomics. The PacBio IsoSeq sequencing data was processed with the isoseq3 (v3.8.2) 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 reads were mapped to the genome with pbmm2 (v1.10.0) 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)[64]. Pseudomolecules were constructed with the TRITEX pipeline[26,27], 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[10]. The origin of the subgenomes and their potential ancestral contributors were previously investigated in detail by Brassac et al.[1] To classify subgenomes in the polyploid genome, Mash (v2.3)[65] was used 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)[64] 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[30]. Chloroplast genomes were assembled following previously published protocols[10]. HiFi reads were aligned to the barley cv. Morex chloroplast genome (GenBank accession EF115541) using minimap2 (v2.24-r1122)[66] 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)[67] and assembled with Canu (v2.1.1)[68] 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)[69] and annotated with GeSeq[70]. Fully annotated assemblies were aligned to the Morex chloroplast genome using MUMmer4 (v4.0.0beta2)[71], and start positions and orientations of the SSC region were adjusted to match the reference using seqkit (v0.9.1)[72]. *k*-mer frequencies were calculated using Jellyfish(v2.2.6)[73] and analysed with findGSE(v1.94)[74] to estimate genome characteristics. Assembly quality, including QV and *k*-mer completeness, was evaluated using Merqury(v1.3)[75].

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.[28]. As evidences we used RNA-seq and IsoSeq data as described under above and Poales-filtered protein sequences downloaded from UniProt[76] (July 2023). An additional consolidation step was carried out using the liftoff[77] tool. In brief, all predicted gene models from one species were lifted over to all other genomes. Using Mikado[78], 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 (https://trep-db.uzh.ch). 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 (v5.8.0)[79] (dataset: poales_odb12), OMARK (omamer verision 2.1.0/omark v0.3.1)[80] and psauron (v1.0.6)[81].

Construction of the gene-based *Hordeum* pangenome

Orthogroups based on the primary protein sequences from 27 annotated *Hordeum* accessions were calculated using Orthofinder (v2.5.5)[82]. The scripts for the calculation of core, shell and cloud genes are deposited in the repository at https://github.com/PGSB-HMGU/BPGv2. Core

genes contain at least 1 gene model from all 27 compared accessions. Shell genes contain gene models from at least 2 *Hordeum* accessions and at most 26. 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 or cloud categories, we aligned all cloud genes for *H. vulgare* cv. Morex and *H. bulbosum* back to all other genomes and corrected for any missing genes. Gene Ontology-term enrichments were conducted with goatools (v1.4.4)[83]. Gene-based collinearity plots were drawn with GENESPACE[84].

Pan-genome analysis of nucleotide-binding resistance genes

The RGAugury (v2.2)[85] 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)[69]. NLR genes were then grouped into clusters using bedtools (v2.30.0)[69] 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 identified using OrthoFinder (v2.5.5)[82] across all NLR genes. Core orthogroups are defined as orthogroups that contain at least one gene model from each of the compared *Hordeum* accessions. Shell orthogroups include gene models from at least two and at most 26*Hordeum* accessions. Gene models that were not assigned to any orthogroup (singletons) or were grouped only with genes from the same genotype were classified as cloud genes.

Pan-genome analysis of PPR and mTERF genes

Open reading frames longer than 100 codons were searched for the presence of PPR and mTERF motifs using hmmsearch from the HMMER suite (http://hmmer.org) using hidden Markov models as defined in Cheng et al.[86]. for the PPR family and PF02536 from the Pfam 32.0 database (http://pfam.xfam.org)[87] for the mTERF motif. Downstream processing of the hmmsearch results for the PPR proteins followed the pipeline described in Gutmann et al.[88]. The RfCTD domain recently identified as a unique feature of RFL-PPRs[89] was included in the gene annotations. P-class 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)[90].

Graph-based pangenome construction and analysis

The analyses described here used a number of tools from the toolkits vg (v1.61.0)[91], odgi (v0.8.3-26-gbc7742ed)[92], bcftools (v1.21)[67] and seqkit (v2.9.0)[72]. All scripts used for graph construction and analysis are available at https://github.com/cropgeeks/panHordeum. Owing to computational limitations, we built an alignment of *Hordeum* whole genomes with Cactus[47]. We selected *H. marinum* BCC 2001 over *H. marinum* H559 on the basis of its higher assembly quality. Genome haplomes were soft-masked with bedtools[69] using their respective repeat annotations. The guide tree required for the Cactus alignment was taken from the Orthofinder[93] output generated as part of this study. Apart from setting appropriate resource parameters (that is, threads, memory and 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[92]. 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 *H. 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[94]. 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 (vcflib v1.0.2)[95], and finally normalizing the resulting variants with bcftools norm[67]. We then added variant length information (SVLEN) to the VCF files with a custom script to allow extraction of structural variants based on the length difference between reference and alternate allele (>50 bp) and classified structural variations 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)[96]. 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[97], 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)[96].

Whole-genome MSA

Single-copy genes were identified using BUSCO (v5.4.6)[79] in *B. distachyon* (v3.0)[98], *O. sativa* (IRGSP-1.0)[99], *S. bicolor* (NCBIv3)[99], *T. monococcum*[100], *H. vulgare* FT11 (ref. 19) 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)[101]. Well-aligned regions were filtered using Gblocks (v0.91b)[102] with parameters -t=extractedp -b4 = 5 -b5=h. A maximum-likelihood phylogenetic tree was then constructed using RAxML (v8.2.12)[103] 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 (v4.1.1; 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)[47] to generate whole-genome MSAs. 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.

Phylogeny and divergence time estimation

To precisely construct phylogenetic trees and estimate divergence times within the genus *Hordeum*, we included four additional outgroup genomes: *B. distachyon* (v3.0)[98], *O. sativa* (IRGSP-1.0)[99], *S. bicolor* (NCBIv3)[104] (all downloaded from Ensembl Plants; https://plants.ensembl.org/index.html) and *T. monococcum*[100] (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 34 *Hordeum* haplomes and four outgroup genomes using OrthoFinder (v2.5.5)[82], yielding 2,198 single-copy genes. For the concatenation-based approach, the CDS sequences of these genes were concatenated and aligned with MAFFT (v7.490)[101], and well-aligned regions were retained using Gblocks (v0.91b)[102] with the parameters -t=extractedp, -b4 = 5, -b5=h; phylogenetic inference was then performed in IQ-TREE (v2.2.2.6)[105]. For the coalescence-based approach, the CDS sequences of each single-copy gene were aligned separately using MAFFT (v7.490)[101], high-quality regions were selected with Gblocks (v0.91b)[102] using the same parameters, individual gene trees were reconstructed with IQ-TREE (v2.2.2.6)[105], and a species tree was inferred from these gene trees using ASTRAL-Pro3 (CASTLES-Pro, v1.23.3.6)[106]. The remaining two methods were based on whole-genome MSAs generated with Cactus. In the CASTER-pair (CASTLES-Pro, v1.23.3.6)[107] approach, the whole-genome MSA was used to infer a phylogenetic tree rooted with *S. bicolor*, while in the CASTER-site (CASTLES-Pro, v1.23.3.6)[107] approach, the whole-genome MSA was analysed similarly but using site-based inference, also rooted with *S. bicolor*.

To estimate divergence times, genomic regions showing strong phylogenetic discordance (nRF > 0.21) were first excluded based on the genome-wide discordance analysis. 4DTV sites were extracted from MSAs of single-copy genes. Divergence times between barley and outgroup species were obtained from the TimeTree database (https://timetree.org/)[108]. Finally, species divergence times within the genus *Hordeum* were estimated using the MCMCtree program in PAML (v4.4)[109].

Identify the boundary of the proximal regions

We excluded multiple haplomes from the same species and retained a final set of 32 *Hordeum* haplomes for comparative analysis. All haplomes pairs were aligned against each other using minimap2 (v2.24-r1122)[66] with the parameters -2 -K 5 G -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 Ma, a bimodal distribution of normalized coverage was observed, allowing a threshold of 0.5 to distinguish distal and proximal regions (Extended Data Fig. 3b,c). 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. Reverse-orientation alignments were extracted as indicators of putative inversion events. Two metrics were calculated for each pairwise comparison: (1) the number of reverse-orientation alignments per Mb and (2) the proportion of bases covered by reverse-orientation alignments. These metrics were separately calculated for proximal and distal chromosomal regions to compare the rates of inversion accumulation during species divergence. Linear regression models including divergence time, chromosomal region type (proximal or distal) and their interaction term were used to evaluate differences in the accumulation rates of reverse-orientation alignments between regions.

Identification of large inversions

All other *Hordeum* genomes were aligned to the MorexV3 (ref. 28) 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 (v1.5.5)[110] 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 2,198 single-copy orthologues identified in the section 'Phylogeny and divergence time estimation' and using the genome-wide concatenated single-copy gene tree as a reference, we conducted sliding-window gene tree analyses (Extended Data Fig. 6a). All single-copy genes were grouped into non-overlapping windows of 25 genes with a step size of 2 genes. Within each window, protein sequences were aligned using MAFFT (v7.490)[101], and well-aligned regions were retained using Gblocks (v0.91b)[102] with the parameters -t=extractedp -b4 = 5 -b5=h. Local window trees were built with FastTree (v2.1.11)[111] (default GTR + CAT, 20 rate categories), starting from neighbour-joining trees and optimized with minimum-evolution and nearest-neighbour interchange moves; branch support was calculated using Shimodaira–Hasegawa-like values. Each resulting tree was rooted with *S. bicolor* as the outgroup using ETE3 (v3.1.2)[112], and the 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 0.21 as a threshold to separate background from discordant signal peaks (Extended Data Fig. 6b). To evaluate the robustness of discordant region detection, we repeated the sliding-window analyses using alternative parameter settings (30 genes with step sizes of 1 and 5 genes). The major discordance peaks were consistently recovered across different window sizes and step lengths, indicating that peak detection was not sensitive to parameter selection (Supplementary Fig. 52). Genomic regions containing at least 10 overlap windows with nRF ≥ 0.21 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 AU test[113] implemented in IQ-TREE (v2.2.2.6)[105], was used to compare local reference trees and local phylogenetic trees. A *P* value (p-AU) lower than 0.05 indicated that the tree was significantly rejected.

Identification of HGT candidates

Genomes of 20 diploid *Hordeum* species were searched for potential HGT genes from grasses of the subfamily Panicoideae. Regions of interest were identified in the genomes using BLAST (v2.12.0+)[114] 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 HGT genes from the Panicoideae subfamily[39,41,42]. 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)[69], 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 (that is, Panicoideae) origin of the candidate HGT genes identified as described above, phylogenetic trees were constructed for each candidate. Homologues 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 10^{-10}). The retrieved sequences were aligned with the *Hordeum* HGT gene candidates using MUSCLE (v3.8)[115]. The alignments were trimmed using trimAl (v1.2) with an automated trimming method[116]. Phylogenetic trees were constructed using PhyML (v3.1)[117], 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[118]. A candidate gene was considered a true HGT gene if the gene sequence was nested within the PACMAD clade. TE-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)[102] for MSA, followed by the *cons* tool from EMBOSS (v6.5.7.0)[119] 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 neighbour 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.

Evolutionary history of HGT

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 model[120], applying the maximum-likelihood approach described by Pagel[121] and implemented in the phytools package[122]. 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 of HGT candidates

Genes from all orthogroups were annotated using eggNOG-mapper web server (v2.1.12)[123,124]. Gene Ontology terms were extracted for each group and enrichment analysis was performed using the clusterProfiler package (v4.14.3), with a significance threshold of an adjusted *P* value of 0.05 (ref. 125).

Analysis of HGT gene expression

RNA-seq data for inflorescence, leaf and root tissues previously generated for 19 *Hordeum* species were used in this study (RNA-seq data for *H. bulbosum* were not available). Read trimming on quality (Q30) and removal of sequencing adapters were done using Trimmomatic (v0.32)[126]. Cleaned reads from each library were then mapped to the corresponding reference genome using the STAR aligner (v2.7.7a)[127]. Quantification of mapped reads was done using the subread package v1.5.2 with featureCounts (v2.0.1)[128] followed by normalization using the trimmed mean of M-values (TMM) method implemented in the edgeR package (v4.2.1)[129]. Genes were considered to be expressed if TMM ≥ 1 in at least one tissue.

Annotation and analysis of TEs

Structurally intact and fragmented TEs were annotated for each genome using the panEDTA (v0.1) pipeline[31]. In addition to the curated TREP database (https://trep-db.uzh.ch/)[130], TE classification followed the EDTA ontology (https://github.com/oushujun/EDTA/blob/master/bin/TE_Sequence_Ontology.txt), with manual curation of TE categories. High-confidence gene models from *H. 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 used the TEpop framework (https://github.com/wicker314/TEpop) to treat each TE as an individual sample for population structure analysis[46]. PCA was then performed using FlashPCA2 (ref. 131) 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 (v4.8.1)[132] with the parameters -n 5 -c 0.9 -d 0 -aL 0.9 -aS 0.9 -g 1 based on sequence similarity. To avoid potential bias caused by large-scale segmental exchanges between subgenomes, genomic regions identified as exchanged segments were excluded from this analysis (Supplementary Table 13). TEs that were shared between polyploid subgenomes but absent from their corresponding diploid progenitors were defined as post-allopolyploidization insertions. The insertion time distributions of these elements were used to reconstruct TE dynamics after subgenome fusion, and the onset of the insertion peak was used to estimate the timing of polyploid 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 (v2.6.5). Genome-wide MSAs 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 MSAs. A neutral model of evolution was fitted to the *Hordeum* phylogeny using fourfold degenerate sites from chromosome 3H of barley, using phyloFit from the PHAST package (v1.4)[133]. 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 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 $MgCl_2$, 1% Trion X-100, 5 mM β-Mercaptoethanol) supplemented with 1× Halt Protease Inhibitor Cocktail (Thermo Fisher Scientific). Following isolation,

nuclei were quantified via flow cytometry and 75,000 nuclei were subjected to tagmentation using the Tagment DNA Enzyme (TDE1, Illumina) at 37 °C for 30 min. Transposed DNA was purified using the MinElute PCR Purification Kit (Qiagen), libraries were amplified with NEBNext High-Fidelity 2x PCR Master Mix (New England Biolabs), and the final library purification was performed using VAHTSTM DNA Clean Beads (Vazyme) according to standard protocols from the manufacturers. Libraries were quantified as described previously[62] and sequenced (paired-end, 2× 151 cycles, Illumina NovaSeq 6000, IPK Gatersleben) following protocols provided by the manufacturer (Illumina).

Adapter trimming and quality filtering were performed using Trim Galore (v0.6.4_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)[134]. The resulting SAM files were converted to BAM format and sorted using SAMtools (v1.10)[67]. 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 (v2.2.9.1)[135]. 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)[114] with *VRN1* MorexV3 (NC_058522.1:c528157887–528147799) and NCBI GenBank *ZCCTa/b/c* queries (AY485977, AY485978 and AY687931). Corresponding regions were extracted from genome assemblies by bedtools (v2.26.0)[69] and manually annotated and analysed 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)[136]. The 79-bp deletion in *H. cordobense* and *H. muticum VRN1* gene was detected by MSA 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[101]. The critical region of cultivar *Igri* was identified in a previous study[51].

Small RNA and mRNA isolation and sequencing

Three biological replicates were analysed 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-seq libraries and small RNA-seq libraries were constructed according to Illumina instructions and sequenced on a NovaSeq 6000 platform at Novogene. 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)[137]. Only reads of length 21–24 nt were used for subsequent analyses. ShortStack (v4)[138] 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)[67] and deepTools (v3.5.6)[139] with parameters bigwigAverage, bamCoverage–binSize 10,–normalizeUsing BPM. Visualization of small RNA and mRNA bigwig files were done with pyGenomeTracks (v3.9)[140].

RNA-seq data analysis

RNA-seq read trimming on quality (Q30) and sequencing adaptor removal were done with Trimmomatic (v0.32)[126]. Resulting high-quality reads from each library were mapped onto the corresponding reference genome using STAR (v2.7.7a)[127], with default parameters. Mapped reads quantification was done by rsem (v1.3.3)[141].

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[59], 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. Several mutants, namely M2342, M9630, M9815 and M11405, were identified from independent M_2 lines of the previously established barley HTX mutagenesis population[142]. SEM observations were conducted using a Hitachi SEM SU8081 following a standard protocol[143]. Young inflorescences were sampled and examined. Allelism test was performed by crossing M9630 with GSHO 1891 and GSHO 1892, respectively. All F_1 progeny exhibited elongated glumes on the central fertile spikelet (Extended Data Fig. 12d). 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 genes

Clean reads (Supplementary Table 20) were obtained by removing the adapters, reads containing *N* bases and low-quality reads using fastp (v0.20.1) with default parameters[144]. Clean reads were mapped to their respective reference genomes via Hisat2 v2.1.0 with default parameters[145]. Read counts were determined from the mapping files via featureCounts (v2.0.1) using standard parameters[128], and the TPM values of each gene were calculated. Spike-specific expressed genes were defined as having less than 5% TPM in other tissues[146]. 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

(v6.1.0; 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 (v3.8)[115], and the resulting MSA was visualized with the Simple MSA Viewer function of TBtools II[147].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All sequence data generated in this study have been deposited at the European Nucleotide Archive (ENA) under umbrella BioProject PRJEB128330. Accession codes for individual accessions are listed in Supplementary Table 3 (genome sequencing) and Supplementary Table 6 (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, MSAs and coordinates of conserved regions have been deposited at the Plant Genomics and Phenomics Research Data Repository[148] (https://doi.org/10.5447/ipk/2026/11). The pangenome graph in GFA format has been deposited in the Plant Genomics and Phenomics Research Data Repository[148] (https://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 and https://github.com/cropgeeks/panHordeum).

60. Siadjeu, C., Pucker, B., Viehöver, P., Albach, D. & Weisshaar, B. High contiguity de novo genome sequence assembly of trifoliate yam (*Dioscorea dumetorum*) using long read sequencing. *Genes* https://doi.org/10.3390/genes11030274 (2020).
61. 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 https://doi.org/10.1007/978-1-4939-9045-0_28 (2019).
62. Mascher, M. et al. Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J.* **76**, 494–505 https://doi.org/10.1111/tpj.12294 (2013).
63. Šimková, H., Tulpová, Z. & Cápal, P. Flow sorting-assisted optical mapping. *Methods Mol. Biol.* **2672**, 465–483 https://doi.org/10.1007/978-1-0716-3226-0_28 (2023).
64. 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 https://doi.org/10.1038/s41592-020-01056-5 (2021).
65. Ondov, B. D. et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 https://doi.org/10.1186/s13059-016-0997-x (2016).
66. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 https://doi.org/10.1093/bioinformatics/bty191 (2018).
67. Danacek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* https://doi.org/10.1093/gigascience/giab008 (2021).
68. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 https://doi.org/10.1101/gr.215087.116 (2017).
69. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 https://doi.org/10.1093/bioinformatics/btq033 (2010).
70. Tillich, M. et al. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **45**, W6–W11 https://doi.org/10.1093/nar/gkx391 (2017).
71. Marçais, G. et al. MUMmer4: afast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944 https://doi.org/10.1371/journal.pcbi.1005944 (2018).
72. Shen, W., Le, S., Li, Y. & Hu, F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS ONE* **11**, e0163962 https://doi.org/10.1371/journal.pone.0163962 (2016).
73. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 https://doi.org/10.1093/bioinformatics/btr011 (2011).
74. Sun, H., Ding, J., Piednoël, M. & Schneeberger, K. findGSE: estimating genome size variation within human and *Arabidopsis* using k-mer frequencies. *Bioinformatics* **34**, 550–557 https://doi.org/10.1093/bioinformatics/btx637 (2018).
75. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 https://doi.org/10.1186/s13059-020-02134-9 (2020).
76. Consortium, T. U. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res.* **53**, D609–D617 https://doi.org/10.1093/nar/gkae1010 (2025).
77. Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 https://doi.org/10.1093/bioinformatics/btaa1016 (2021).
78. Venturini, L., Caim, S., Kaithakottil, G. G., Mapleson, D. L. & Swarbreck, D. Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *Gigascience* https://doi.org/10.1093/gigascience/giy093 (2018).
79. Kollmar, M. Preface. In *Gene Prediction: Methods and Protocols* (ed. Kollmar, M.) https://doi.org/10.1007/978-1-4939-9173-0 (2019).
80. Nevers, Y. et al. Quality assessment of gene repertoire annotations with OMArk. *Nat. Biotechnol.* **43**, 124–133 https://doi.org/10.1038/s41587-024-02147-w (2025).
81. Sommer, M. J., Zimin, A. V. & Salzberg, S. L. PSAURON: a tool for assessing protein annotation across a broad range of species. *NAR Genomics Bioinformatics* **7**, lqae189 https://doi.org/10.1093/nargab/lqae189 (2025).
82. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 https://doi.org/10.1186/s13059-019-1832-y (2019).
83. Klopfenstein, D. V. et al. GOATOOLS: a Python library for Gene Ontology analyses. *Sci. Rep.* **8**, 10872 https://doi.org/10.1038/s41598-018-28948-z (2018).
84. Lovell, J. T. et al. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife* https://doi.org/10.7554/eLife.78526 (2022).
85. Li, P. et al. RGAugury: a pipeline for genome-wide prediction of resistance gene analogs (RGAs) in plants. *BMC Genomics* **17**, 852 https://doi.org/10.1186/s12864-016-3197-x (2016).
86. Cheng, S. F. et al. Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J.* **85**, 532–547 https://doi.org/10.1111/tpj.13121 (2016).
87. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 https://doi.org/10.1093/nar/gky995 (2019).
88. Gutmann, B. et al. The expansion and diversification of pentatricopeptide repeat RNA-editing factors in plants. *Mol. Plant* **13**, 215–230 https://doi.org/10.1016/j.molp.2019.11.002 (2020).
89. 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 Phytol.* **240**, 830–845 https://doi.org/10.1111/nph.19166 (2023).
90. Appels, R. et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* https://doi.org/10.1126/science.aar7191 (2018).
91. Garrison, E. et al. Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nat. Biotechnol.* **36**, 875–879 (2018).
92. Guarracino, A., Heumos, S., Nahnsen, S., Prins, P. & Garrison, E. ODGI: understanding pangenome graphs. *Bioinformatics* **38**, 3319–3326 https://doi.org/10.1093/bioinformatics/btac308 (2022).
93. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 https://doi.org/10.1186/s13059-015-0721-2 (2015).
94. Liao, W.-W. et al. A draft human pangenome reference. *Nature* **617**, 312–324 https://doi.org/10.1038/s41586-023-05896-x (2023).
95. 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 Comput. Biol.* **18**, e1009123 https://doi.org/10.1371/journal.pcbi.1009123 (2022).
96. Wickham, H. *ggplot2* (Springer, 2016).
97. Zytnicki, M. Assessing genome conservation on pangenome graphs with PanSel. *Bioinformatics Adv.* **5**, vfab018 (2025).
98. Sreedasyam, A. et al. JGI Plant Gene Atlas: an updateable transcriptome resource to improve functional gene descriptions across the plant kingdom. *Nucleic Acids Res.* **51**, 8383–8401 https://doi.org/10.1093/nar/gkad616 (2023).
99. Kawahara, Y. et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**, 4 https://doi.org/10.1186/1939-8433-6-4 (2013).
100. Ahmed, H. I. et al. Einkorn genomics sheds light on history of the oldest domesticated wheat. *Nature* **620**, 830–838 https://doi.org/10.1038/s41586-023-06389-7 (2023).
101. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 https://doi.org/10.1093/molbev/mst010 (2013).
102. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 https://doi.org/10.1093/oxfordjournals.molbev.a026334 (2000).
103. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 https://doi.org/10.1093/bioinformatics/btu033 (2014).
104. 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 https://doi.org/10.1111/tpj.13781 (2018).
105. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 https://doi.org/10.1093/molbev/msaa015 (2020).
106. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* https://doi.org/10.1186/s12859-018-2129-y (2018).
107. Zhang, C., Nielsen, R. & Mirarab, S. CASTER: direct species tree inference from whole-genome alignments. *Science* https://doi.org/10.1126/science.adk9688 (2025).
108. Kumar, S. et al. TimeTree 5: an expanded resource for species divergence times. *Mol. Biol. Evol.* https://doi.org/10.1093/molbev/msac174 (2022).
109. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 https://doi.org/10.1093/molbev/msm088 (2007).
110. 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 https://doi.org/10.1186/s13059-019-1911-0 (2019).

111. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 https://doi.org/10.1093/molbev/msp077 (2009).
112. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 https://doi.org/10.1093/molbev/msw046 (2016).
113. Shimodaira, H. An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**, 492–508 https://doi.org/10.1080/10635150290069913 (2002).
114. Camacho, C. et al. BLAST plus: architecture and applications. *BMC Bioinformatics* https://doi.org/10.1186/1471-2105-10-421 (2009).
115. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 https://doi.org/10.1093/nar/gkh340 (2004).
116. 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 https://doi.org/10.1093/bioinformatics/btp348 (2009).
117. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 https://doi.org/10.1093/sysbio/syq010 (2010).
118. Junier, T. & Zdobnov, E. M. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* **26**, 1669–1670 https://doi.org/10.1093/bioinformatics/btq243 (2010).
119. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 https://doi.org/10.1016/s0168-9525(00)02024-2 (2000).
120. Lewis, P. O. A likelihood approach to estimating phylogeny from discrete morphological character data. *Syst. Biol.* **50**, 913–925 https://doi.org/10.1080/106351501753462876 (2001).
121. Pagel, M. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Syst. Biol.* **48**, 612–622 https://doi.org/10.1080/106351599260184 (1999).
122. Revell, L. J. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *PeerJ* **12**, e16505 https://doi.org/10.7717/peerj.16505 (2024).
123. 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. *Mol. Biol. Evol.* **38**, 5825–5829 https://doi.org/10.1093/molbev/msab293 (2021).
124. 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 https://doi.org/10.1093/nar/gky1085 (2019).
125. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 https://doi.org/10.1016/j.xinn.2021.100141 (2021).
126. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 https://doi.org/10.1093/bioinformatics/btu170 (2014).
127. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 https://doi.org/10.1093/bioinformatics/bts635 (2013).
128. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 https://doi.org/10.1093/bioinformatics/btt656 (2014).
129. 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 https://doi.org/10.1093/nar/gks042 (2012).
130. Wicker, T., Matthews, D. E. & Keller, B. TREP: a database for Triticeae repetitive elements. *Trends Plant Sci.* **7**, 561–562 https://doi.org/10.1016/S1360-1385(02)02372-5 (2002).
131. Abraham, G., Qiu, Y. & Inouye, M. FlashPCA2: principal component analysis of biobank-scale genotype datasets. *Bioinformatics* **33**, 2776–2778 https://doi.org/10.1093/bioinformatics/btx299 (2017).
132. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 https://doi.org/10.1093/bioinformatics/bts565 (2012).
133. Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinformatics* **12**, 41–51 https://doi.org/10.1093/bib/bbq072 (2011).
134. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://doi.org/10.48550/arXiv.1303.3997 (2013).
135. Zhang, Y. et al. Model-based analysis of ChIP–seq (MACS). *Genome Biol.* **9**, R137 https://doi.org/10.1186/gb-2008-9-9-r137 (2008).
136. Seibt, K. M., Schmidt, T. & Heitkam, T. FlexiDot: highly customizable, ambiguity-aware dotplots for visual sequence analyses. *Bioinformatics* **34**, 3575–3577 https://doi.org/10.1093/bioinformatics/bty395 (2018).
137. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
138. Johnson, N. R., Yeoh, J. M., Coruh, C. & Axtell, M. J. Improved placement of multi-mapping small RNAs. *G3* **6**, 2103–2111 https://doi.org/10.1534/g3.116.030452 (2016).
139. 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 https://doi.org/10.1093/nar/gku365 (2014).
140. Lopez-Delisle, L. et al. pyGenomeTracks: reproducible plots for multivariate genomic datasets. *Bioinformatics* **37**, 422–423 https://doi.org/10.1093/bioinformatics/btaa692 (2021).
141. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 https://doi.org/10.1186/1471-2105-12-323 (2011).
142. Jiang, C. et al. A reference-guided TILLING by amplicon-sequencing platform supports forward and reverse genetics in barley. *Plant Commun.* **3**, 100317 https://doi.org/10.1016/j.xplc.2022.100317 (2022).
143. 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 Biotechnol. J.* **20**, 75–88 https://doi.org/10.1111/pbi.13696 (2022).
144. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 https://doi.org/10.1093/bioinformatics/bty560 (2018).
145. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 https://doi.org/10.1038/nmeth.3317 (2015).
146. Kovacik, M. et al. The transcriptome landscape of developing barley seeds. *Plant Cell* **36**, 2512–2530 https://doi.org/10.1093/plcell/koae095 (2024).
147. Chen, C. J. et al. TBtools-II: a "one for all, all for one" bioinformatics platform for biological big-data mining. *Mol. Plant* **16**, 1733–1742 https://doi.org/10.1016/j.molp.2023.09.010 (2023).
148. Arend, D. et al. PGP repository: a plant phenomics and genomics data publication infrastructure. *Database* **2016**, baw033 (2016).
149. Percival-Alwyn, L. et al. UKCropDiversity-HPC: a collaborative high-performance computing resource approach for sustainable agriculture and biodiversity conservation. *Plants People Planet* https://doi.org/10.1002/ppp3.10607 (2024).

Acknowledgements We thank M. Knauft, J. Pohl and P. Jaroschinsky for expert technical work during library construction and DNA sequencing; A. Houben and Y.-T. Kuo for support in validating sequence assemblies. The authors acknowledge Research Computing at the James Hutton Institute for providing computational resources and technical support for the UK's Crop Diversity Bioinformatics HPC (BBSRC grants BB/S019669/1 and BB/X019683/1), use of which has contributed to the results reported within this paper. The authors utilized the University of Wisconsin–Madison Biotechnology Center's DNA Sequencing Facility (Research Resource Identifier; RRID:SCR_017759) to generate and sequence PacBio libraries. M.v.K. and E.B.H. utilized the DFG Research Infrastructure NGS_CC (West German Genome Center (WGGC), project 407493903) as part of the Next Generation Sequencing Competence Network (project 423957469) to generate and sequence PacBio libraries. M.v.K. and E.B.H. acknowledge the computational infrastructure provided by the Centre for Information and Media Technology at Heinrich Heine University Düsseldorf.

Author contributions Conception and coordination of the study: M.M. Selection of *Hordeum* accessions for sequencing: F.R.B., M.M., M.v.K., P.Y., S.S. and T.K. Coordination of genome and transcriptome sequencing: M.M., N.S., F.R.B., M.v.K., E.B.H., P.Y., C.J., S.S., D.K., J.Š. and R.W. Genome sequencing: A.H., E.B.H. and Y.J. Transcriptome sequencing: R.H., A.H., E.B.H. and S.S. Morphological analysis: T.K. ATAC–seq: Z.Z., R.H. and A.H. Sequence data management and submission: A.F. Optical map construction: H.T. and H.Š. Chromosome-scale genome assembly and comparison: J.-W.F. and E.B.H. Gene annotation, orthology inference and gene-based pangenome analysis: T.L., K.F.X.M. and M. Spannagl. Annotation and analysis of the PPR and mTERF gene families: J.M. and I.S. Annotation and analysis of TEs: J.-W.F., E.B.H., H.Z. and T.W. Phylogenomic analyses including structural variation calling, construction of phylogenetic trees and analysis of evolutionary conservation: J.-W.F. Cloning and characterization of *EOG1*: P.Y., C.J., C.A. and G.G. Pan-genome graph construction and analysis: M.B., S.V.C. and Y.J.K.B. Analysis and interpretation of HGT: M. Szecówka, M.C., K.K., V.M. and D.K. Analysis of diversity in VRN genes: B.S., R.Č., A.P. and J.Š. Manuscript preparation: J-W.F. and M.M. All authors read and edited the final manuscript.

Funding This work was supported by the European Research Council (ERC) under the European Union's Horizon Europe research and innovation programme (TRANSFER, 949873 to M.M. and PERLIFE, 101002085 to M.v.K.). M.M. and P.Y. were supported by the Mobility Program of the Sino-German Center for Research Promotion (grant M-0440). P.Y. was supported by the Agricultural Science and Technology Innovation Program of CAAS and State Key Laboratory of Crop Gene Resources and Breeding. Work in the von Korff laboratory was funded by the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy–EXC-2048/2–Project ID: 390686111, the Collaborative Research Centre/Transregio (TRR 341, Project ID: 456082119) and the GRK 2064: Water use efficiency and drought stress responses: from *Arabidopsis* to barley, Project ID: 252965955. J.Š., A.P., B.S. and R.Č. received funding from the Czech Science Foundation (grant 26-20415S). J.Š., A.P., B.S., R.Č., M.C., M. Szecówka and D.K. were supported by the MEYS project TowArds Next GENeration Crops (TANGENC) (CZ.02.01.01/00/22_008/0004581). M. Szecówka, M.C., V.M., K.K., S.V.C., Y.J.K.B. and D.K. were supported by the Czech Science Foundation (project 22-02469S), and V.M., K.K., S.V.C. and Y.J.K.B. were supported by the Czech Academy of Sciences (long-term research development project RVO 67985939). S.S. was supported by the Research Support Project for the Next Generation at Tottori University. J.M. and I.S. received funding from the Australian Research Council (ARC), Linkage grant LP200100547. J.M. was also supported by an ARC Future Fellowship (FT220100792). T.K. received funding from the National Science Foundation of China (NSFC) (grant W2411025), the Shandong Province Key R&D Program for Shandong and the China–Israel Cooperation Program (2023KJHZ003). M. Spannagl and K.F.X.M. acknowledge support from the German Federal Ministry of Education and Research (BMBF) under project number 031B0190 (de.NBI). M.M. and N.S. were supported by BMBF project Epic-p-epBAR (031B1224). Computational resources were partially provided by the e-INFRA CZ project (ID: 90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic, and by the ELIXIR-CZ project (ID: 90255), part of the international ELIXIR infrastructure. The UK's Crop Diversity Bioinformatics HPC[149] was supported by BBSRC grants BB/S019669/1 and BB/X019683/1. M.v.K. and E.B.H. utilized the DFG-funded Research Infrastructure NGS_CC (West German Genome Center (WGGC), project 407493903), established as part of the Next Generation Sequencing Competence Network (project 423957469). Open access funding provided by Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK).

Competing interests The authors declare no competing interests.

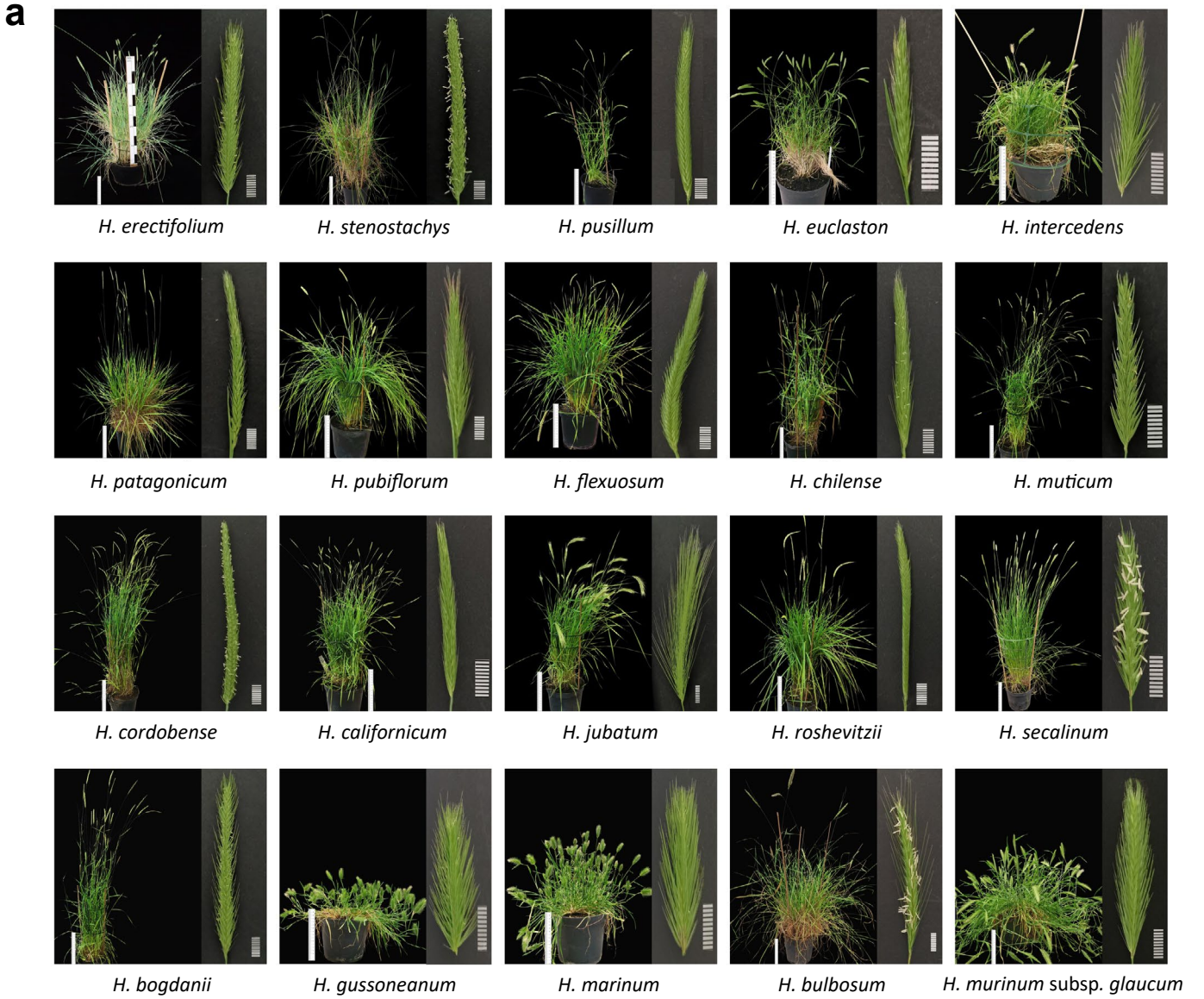
Additional information

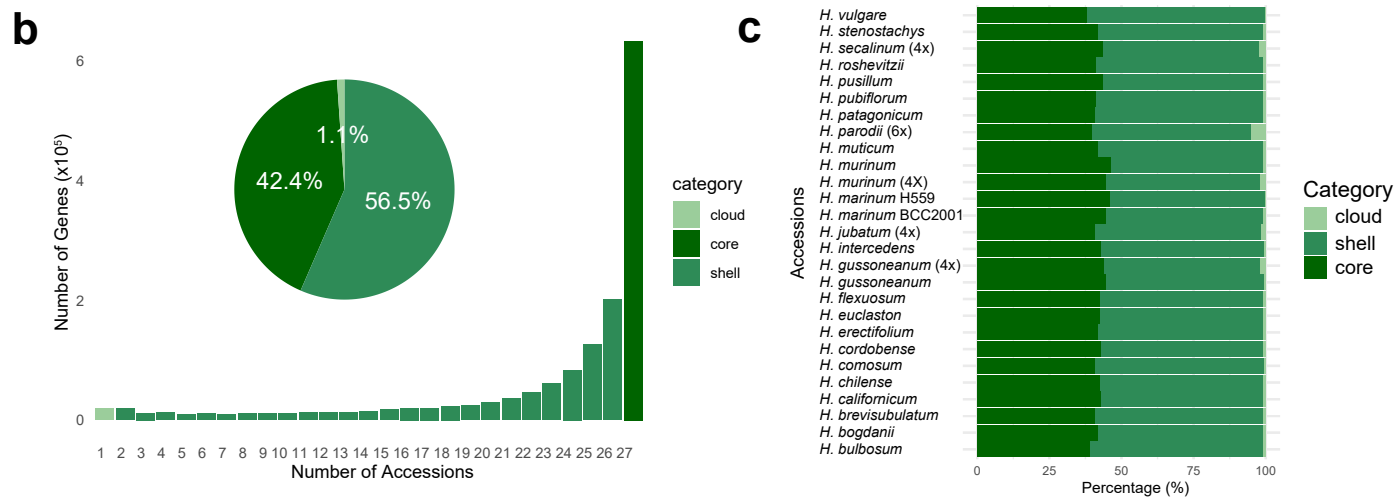
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-026-11332-7.

Correspondence and requests for materials should be addressed to Jan Šafář, David Kopecký, Frank R. Blattner, Ping Yang, Maria von Korff or Martin Mascher.

Peer review information *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

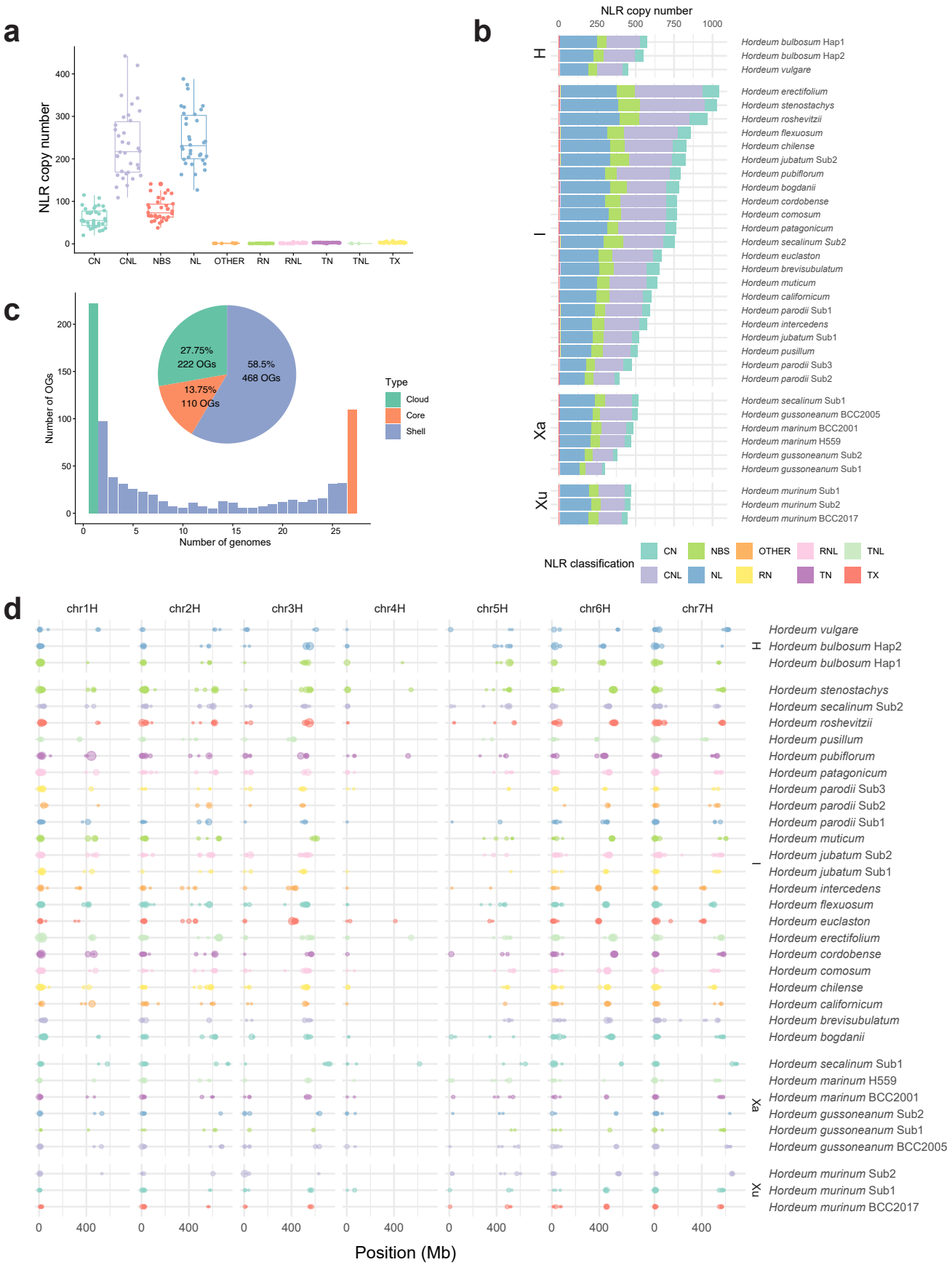
Reprints and permissions information is available at http://www.nature.com/reprints.





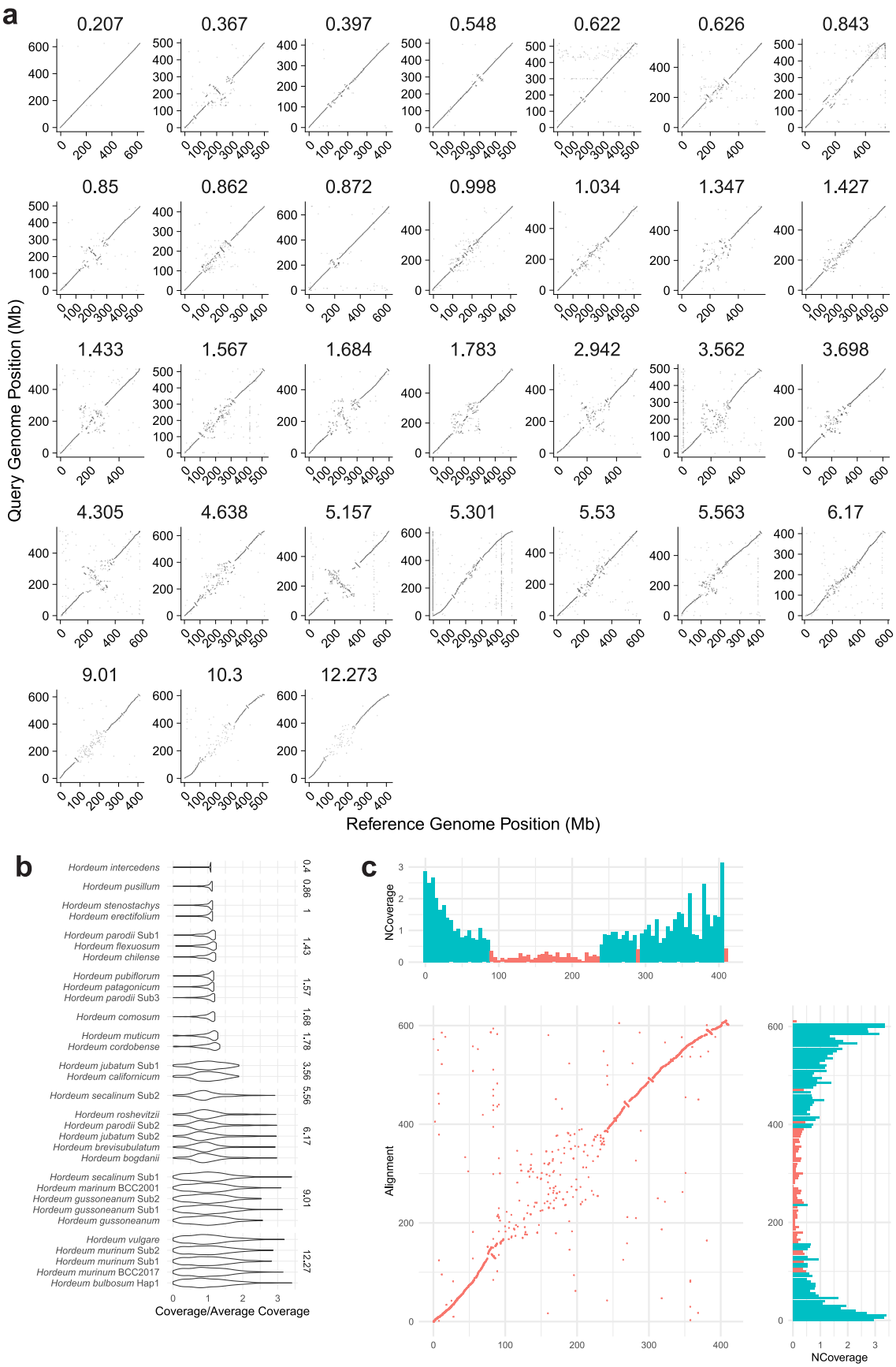
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Morphological and genomics diversity across *Hordeum* species. (**a**) Representative images showing whole plant architecture and spike morphology for 20 species in the genus *Hordeum*. Species shown (left to right, top to bottom): *H. erectifolium*, *H. stenostachys*, *H. pusillum*, *H. euclaston*, *H. intercedens*, *H. patagonicum*, *H. pubiflorum*, *H. flexuosum*, *H. chilense*, *H. muticum*, *H. cordobense*, *H. californicum*, *H. jubatum*, *H. roshevitzii*, *H. secalinum*, *H. bogdanii*, *H. gussoneanum*, *H. marinum*, *H. bulbosum*, and *H. murinum* subsp. *glaucum*. Each panel displays a whole plant (left) and a close-up of the spike (right), highlighting variation in growth habit and inflorescence morphology across the genus. Note that the scale bars differ: the scale bar for whole plant images represents 16 cm, while the scale bar for spike images represents 1 cm. Images show representative plants of the indicated species/accessions; no independent experimental replication was performed for this descriptive panel. (**b**) Pie-chart providing ratios of conserved- and variable-genes for all *Hordeum* pangenome accessions. Bar chart illustrating the number of genes shared between different combinations of accessions within the *Hordeum* pangenome. Genes shared by all 27 accessions constitute the core genome while genes private to a single species are referred to as the cloud genome. (**c**) Bar chart illustrating the proportion of genes contained in the core-, shell-, and cloud- gene portions by individual accessions.



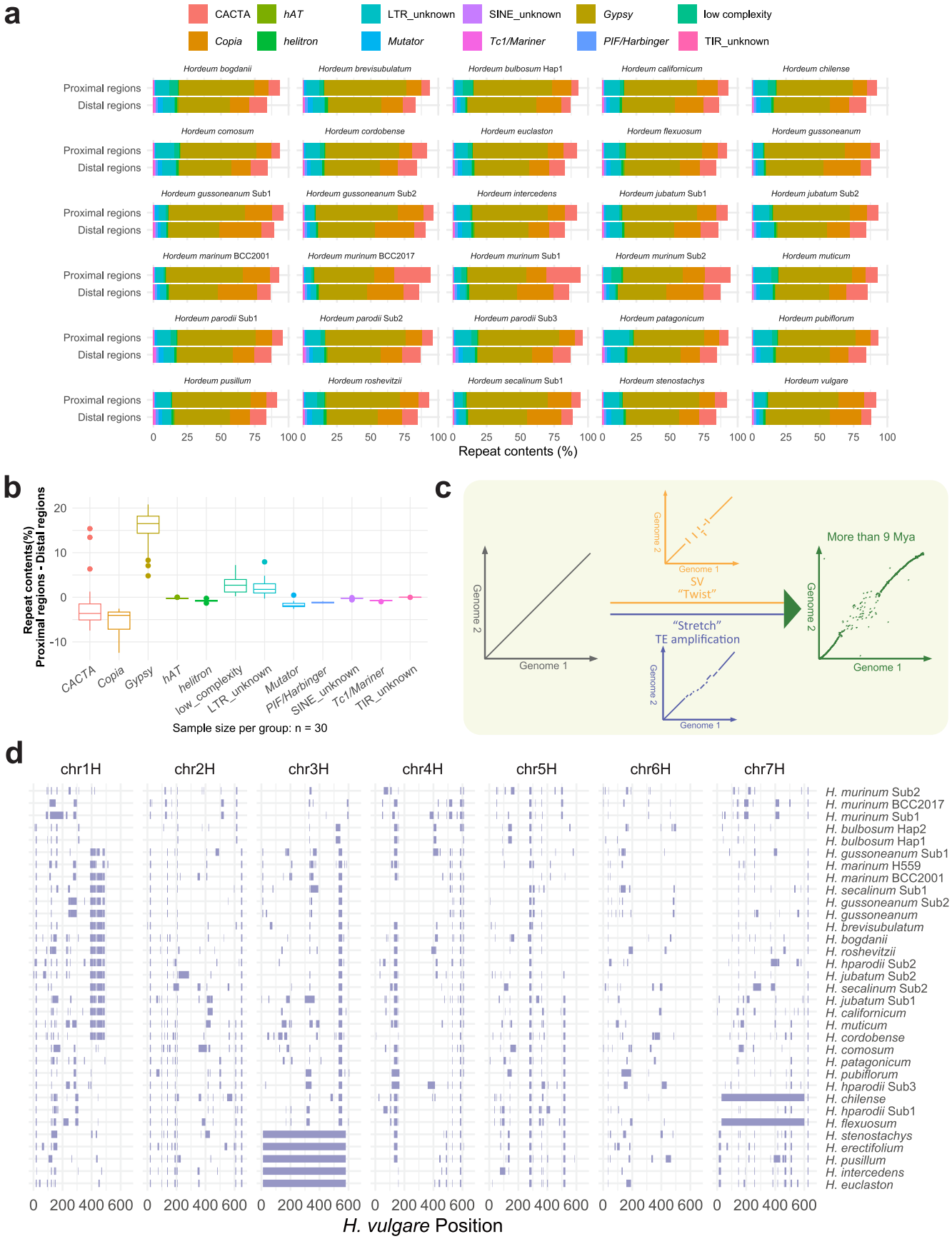
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Landscape of NLR genes in the genus *Hordeum*. (**a**) NLR gene copy numbers across 34 *Hordeum* haplomes. The central line in each boxplot represents the median; the box spans the interquartile range (IQR), and whiskers extend to 1.5 × IQR. 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. NBS-only: Contain only the NBS domain. TIR = Toll/Interleukin-1 Receptor-like; NBS = Nucleotide-Binding Site; LRR = Leucine-Rich Repeat; CC = Coiled-Coil; RPW8 = Resistance to Powdery Mildew 8-like domain. (**b**) Comparison of NLR copy numbers among *Hordeum* genomes. Welch's t-test revealed a highly significant difference between the I genome group and other genome groups (two-sided Welch's t-test, $P = 1.8 \times 10^{-6}$). (**c**) Bar chart showing the number of accessions represented in each NLR orthogroup (OG). The x-axis indicates accession count per OG. The pie chart summarizes conserved vs. variable genes across 27 accessions. (**d**) Distribution of NLR gene clusters (≥3 genes) across 34 haplomes. Dot size represents the number of genes per cluster. The *chr4H/chr5H group* showed significantly different values compared to other chromosomes (two-sided Wilcoxon rank-sum test, $P < 2.2 \times 10^{-16}$).



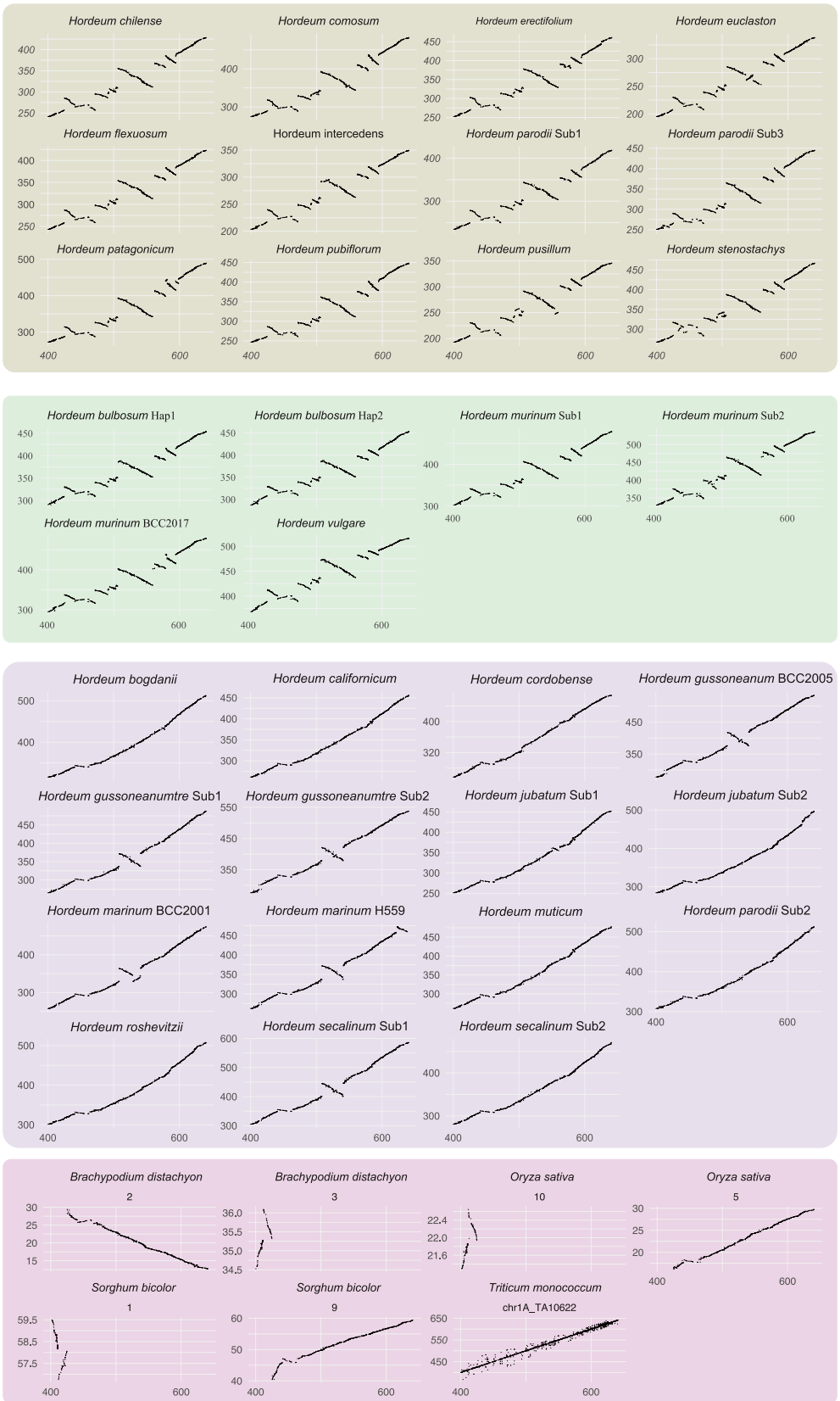
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Normalized alignment coverage defines proximal chromosomal regions. (**a**) Complete set of chromosome 4H dotplot comparisons between *Hordeum* genomes spanning different divergence times, complementing the representative comparisons shown in Fig. 1c. Two species were randomly selected from each phylogenetic node for sequence alignment (Supplementary Table 11). Numbers above the dotplots indicate the estimated divergence times (Ma) between the species shown on the x- and y-axes. (**b**) Normalized coverage density plot in *H. euclaston*. The right shows species ordered by divergence time (Ma); two distinct peaks appear among species that diverged over 9 million years ago. (**c**) Comparison of chromosome 4H between *H. vulgare* and *H. euclaston*. The bar plots along the x- and y-axes represent normalized coverage in 5 Mb windows. Coverage was calculated based on whole-genome alignment between *H. vulgare* and *H. euclaston*. Proximal and distal regions were defined using a normalized coverage threshold of 0.5.



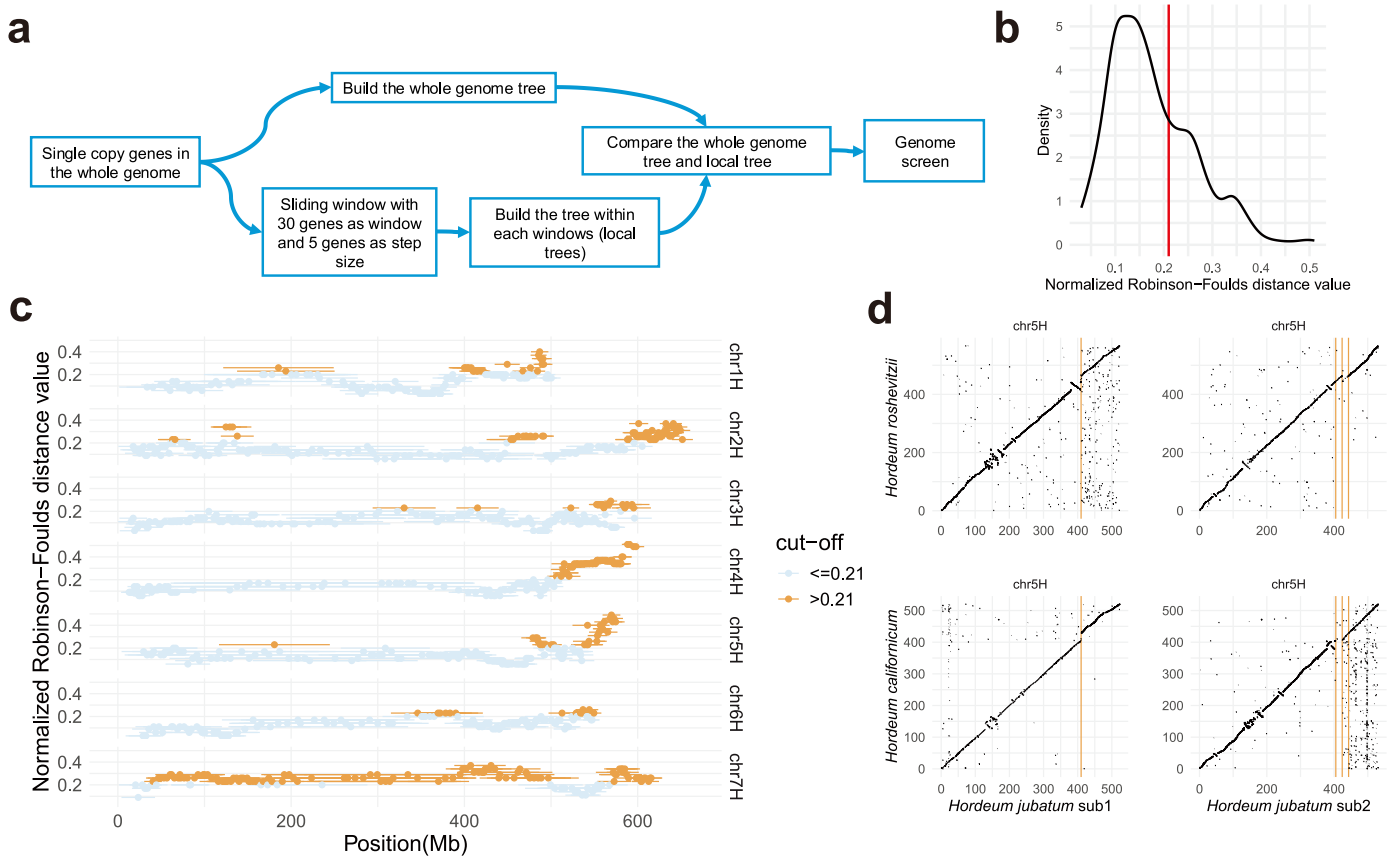
Extended Data Fig. 4 | See next page for caption.

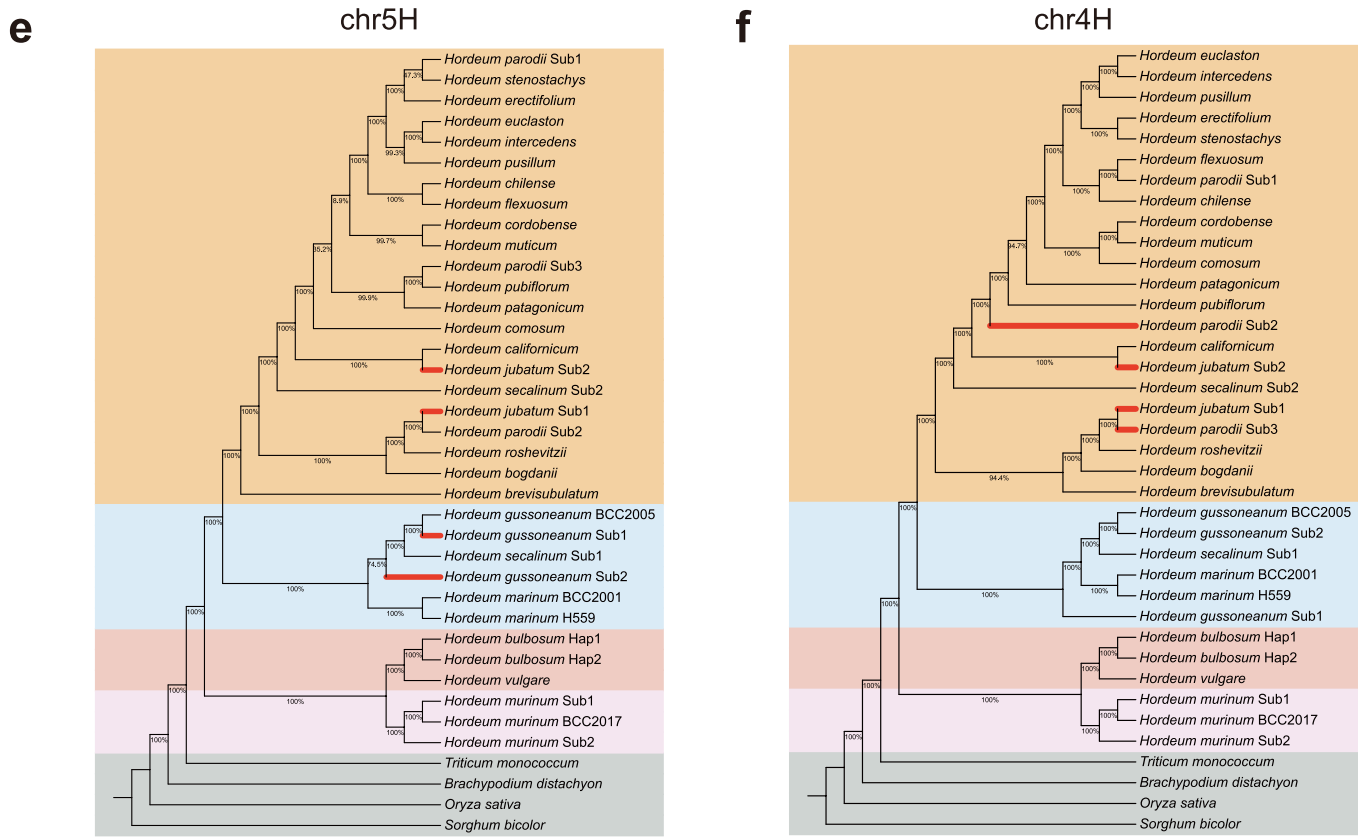
Extended Data Fig. 4 | "Stretch and twist" model of proximal region evolution and inversion map. (**a**) Proportion of repeat types in proximal and distal regions. (**b**) Boxplot showing the difference in repeat composition between distal and proximal regions. *LTR/Gypsy*, *LTR/unknown* and low_complexity showed significant proximal enrichment (one-sided paired Wilcoxon test, $P = 9.31 \times 10^{-10}$, $P = 1.86 \times 10^{-9}$ and $P = 9.31 \times 10^{-10}$, respectively). For the box plots, the center line represents the median, the box bounds indicate the 25th and 75th percentiles, and the whiskers extend to the most extreme values within 1.5× the interquartile range (IQR) from the box boundaries; points beyond the whiskers are shown as outliers. (**c**) A schematic diagram of the "stretch-and-twist" model. (**d**) Genome-wide distribution of distal inversions (≥2 Mb) across *Hordeum* haplomes.



Triticum monococcum chr1A:400Mb-650Mb

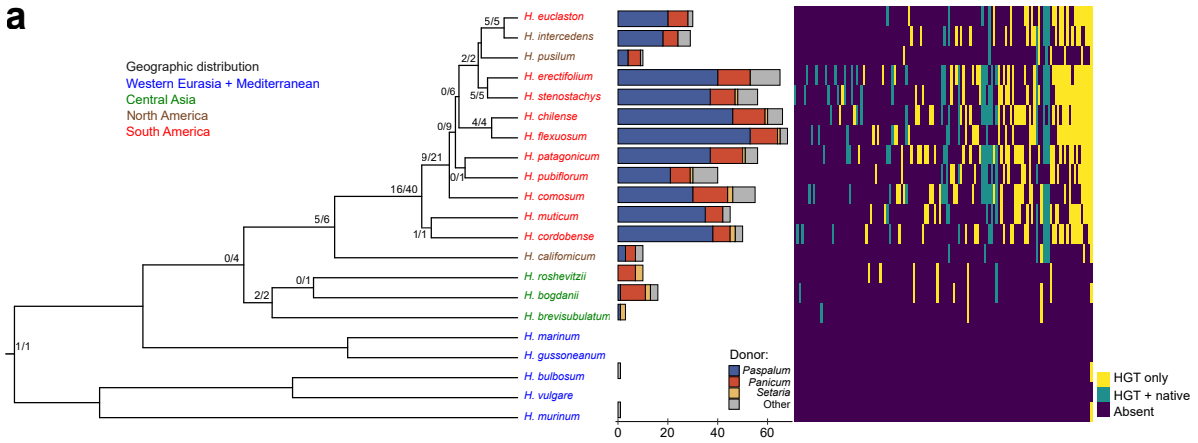
Extended Data Fig. 5 | Six shared inversions support the hybridization region. Dotplot comparisons of the chr1A: 400–650 Mb region between *T. monococcum* and the remaining genomes, complementing the representative comparisons shown in Fig. 2a. Colors indicate different clusters and are consistent with those used in Fig. 2a.

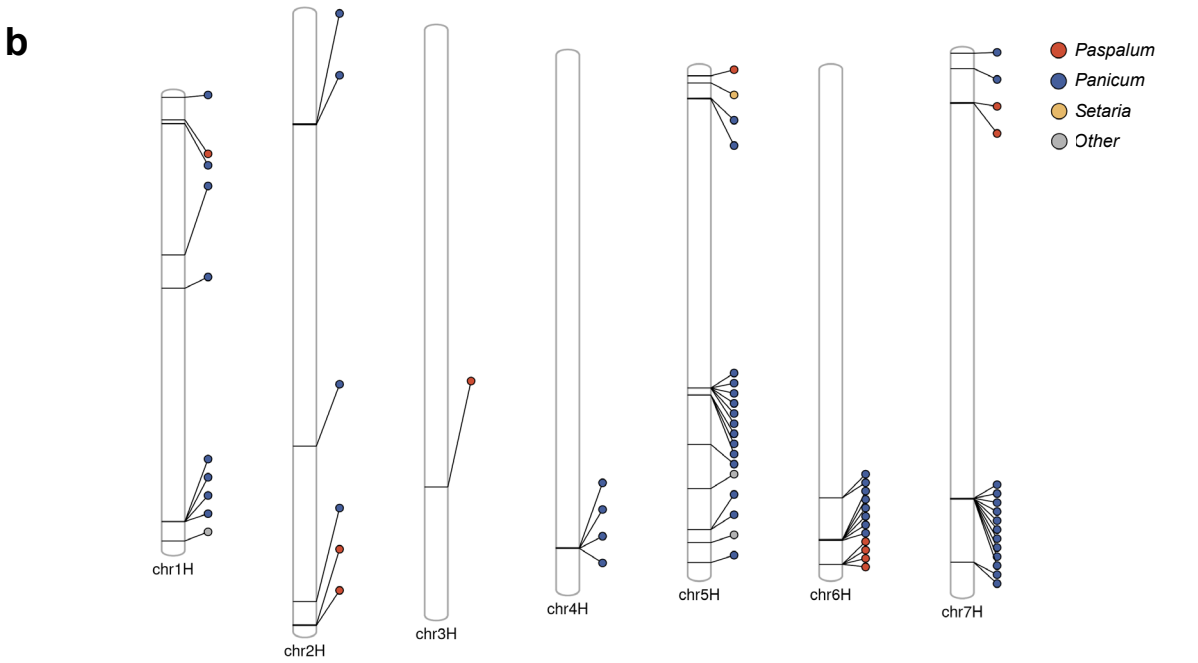


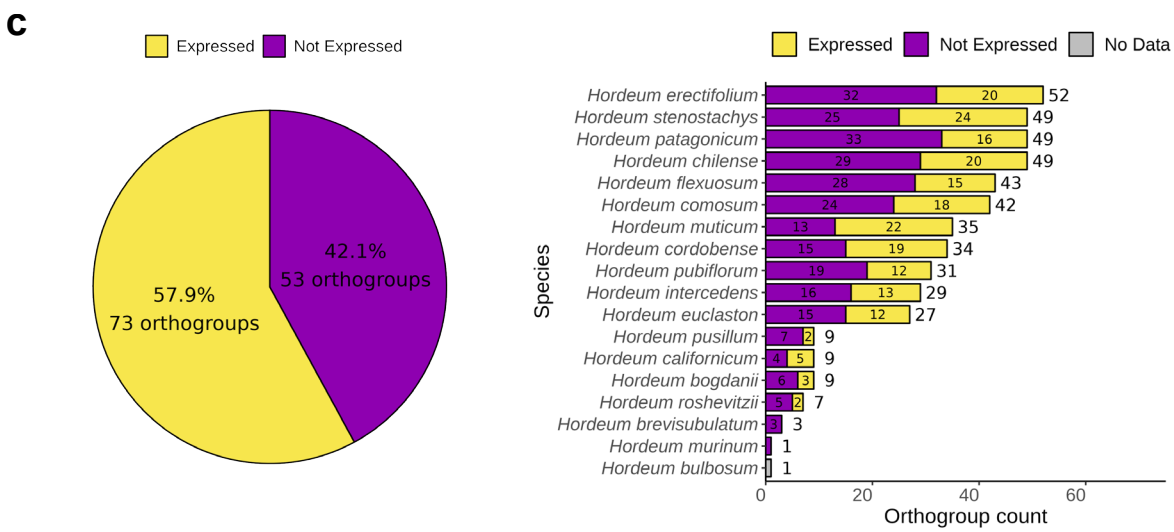


Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Detection of ancient hybridization events and segmental exchanges. (**a**) Schematic overview of the hybridization scanning method. (**b**) Density distribution of normalized Robinson–Foulds (nRF) distances. The red line (nRF = 0.21) represents the empirical threshold separating the two peaks corresponding to background phylogenetic variation and regions with elevated phylogenetic discordance. (**c**) Distribution of normalized Robinson–Foulds (nRF) distances between local gene trees and the species-wide reference tree across the genome. Horizontal line segments indicate the genomic intervals spanned by genes included in each local tree, whereas points mark the midpoint of each interval. The y-axis shows the nRF distance between each local tree and the reference tree, quantifying their topological dissimilarity. Orange indicates nRF values > 0.21, whereas light blue indicates values ≤ 0.21. (**d**) Comparative analysis of chromosome 5H between *H. jubatum* and its ancestral diploid genomes. Orange lines indicate regions with duplications and deletions. A total of n = 128, 178, 164, 142, 144, 118 and 142 local trees were analyzed for chr1H-chr7H, respectively. (**e**) Local phylogenetic analysis of the segmental exchange region on chromosome 5H (523,129,800–584,948,519 bp). The phylogenetic tree was reconstructed using concatenated single-copy genes located within this region, with bootstrap values shown at each node. Red branches indicate topological changes relative to the genome-wide reference species tree. The local phylogeny revealed segmental exchanges between the two subgenomes of the allotetraploid *H. gussoneanum* BCC2005 and between the two subgenomes of the allotetraploid *H. jubatum*. An Approximately Unbiased (AU) test supported the local phylogenetic topology (p-AU = 1.0) and rejected the genome-wide reference topology (p-AU = 2.11×10^{-84}). (**f**) Local phylogenetic analysis of the segmental exchange region on chromosome 4H (495,575,039–607,499,705 bp), performed using the same approach as in (**e**). The local phylogeny revealed segmental exchanges between the two subgenomes of the allotetraploid *H. jubatum* and between subgenomes 2 and 3 of the allohexaploid *H. parodii*. An AU test supported the local phylogenetic topology (p-AU = 1.0) and rejected the genome-wide reference topology (p-AU = 6.55×10^{-57}).

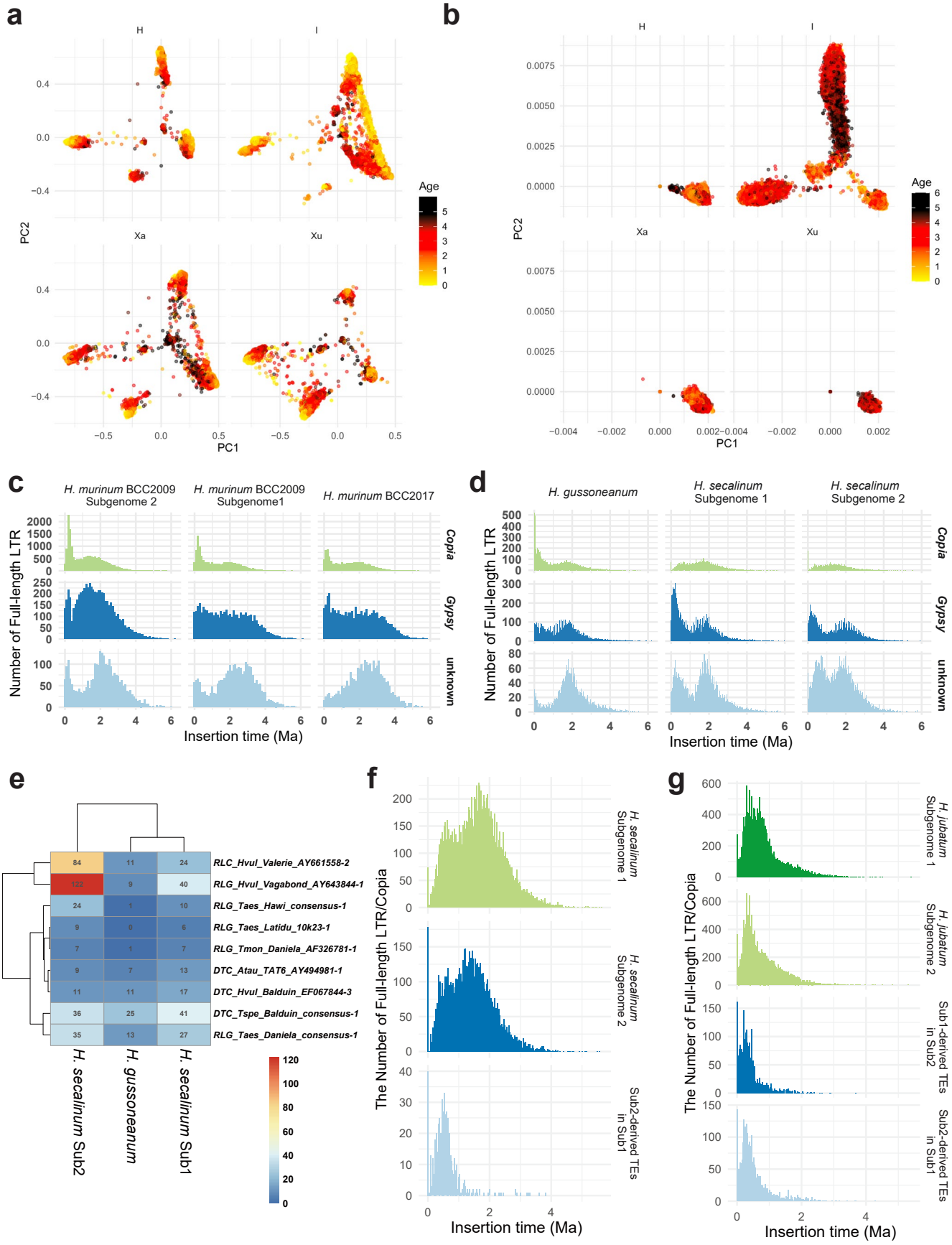






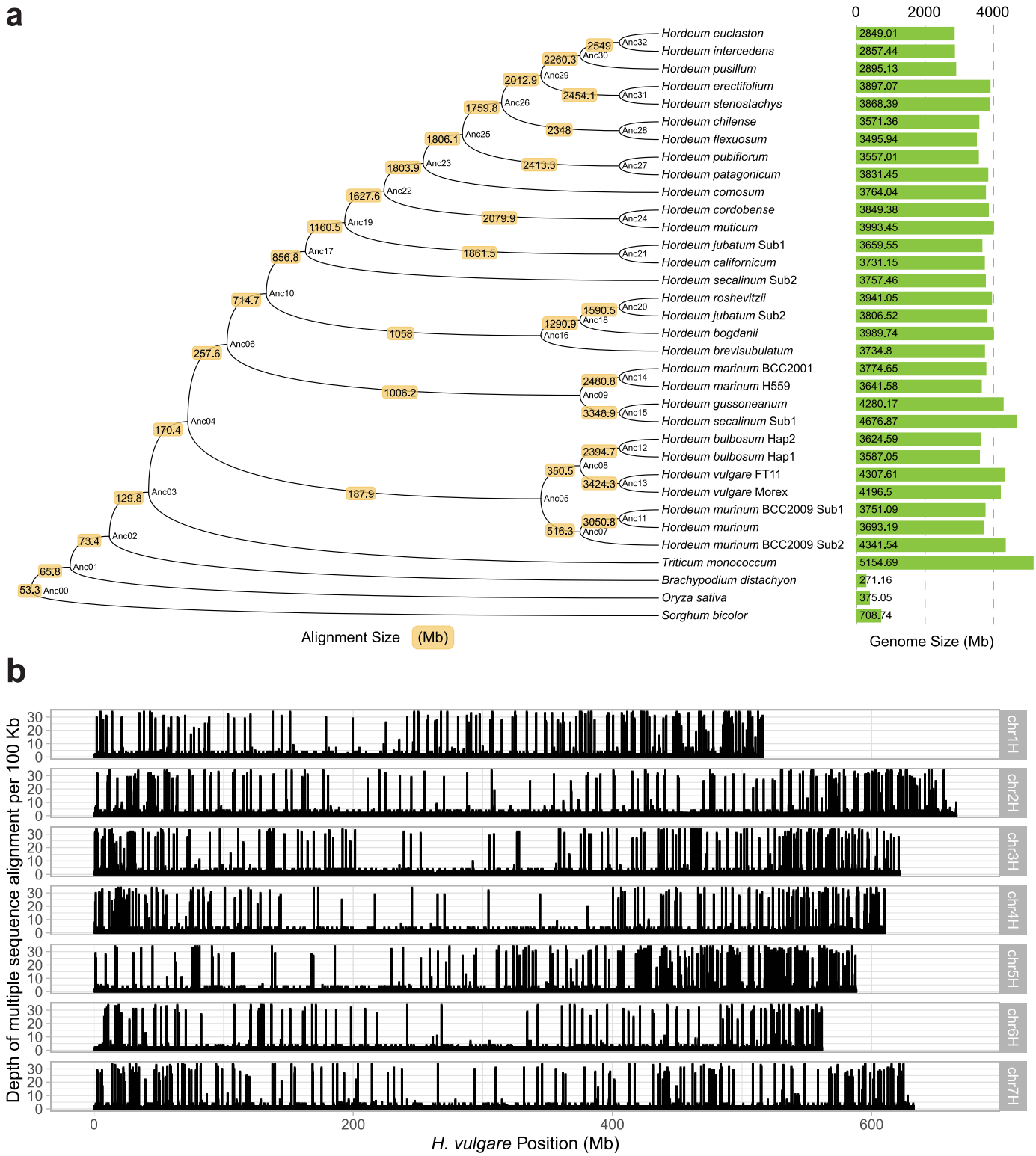
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Horizontal gene transfer (HGT) from Panicoideae into *Hordeum* species. (**a**) Overview of HGT presence across diploid species of *Hordeum* genus. Left panel: Phylogenetic tree of the genus. Numbers adjacent to the nodes indicate the number of HGT in the respective ancestors: the first value reflects ancestral state reconstruction analysis, the second value represents manual correction (see Online methods). Geographical distribution adapted from Brassac & Blattner[1]. Central panel: Number of panicoid genes acquired by HGT in *Hordeum* species, categorized by donor genus: *Paspalum* (blue), *Panicum* (red), *Setaria* (yellow), and other Panicoideae grasses (grey). Right panel: Distribution of panicoid HGT orthogroups across *Hordeum* species: absent (purple), present (yellow), and present with a native copy in the *Hordeum* pan-genome (green). (**b**) Chromosomal distribution of foreign genes in *H. flexuosum*. Dots represent individual HGT genes from donor genus: *Paspalum* (blue), *Panicum* (red), *Setaria* (yellow), and other Panicoideae grasses (grey). (**c**) Expression of HGT. Left panel: Percentage of orthogroups with at least one expressed HGT gene (TMM ≥ 1) in at least one tissue (yellow); no expression in all tissues and species (purple). Right panel: Expression in individual species. The number of orthogroups with at least one expressed HGT gene (TMM ≥ 1) in at least one tissue (yellow); no expression in all tissues (purple); species with no RNA-seq data available (grey).

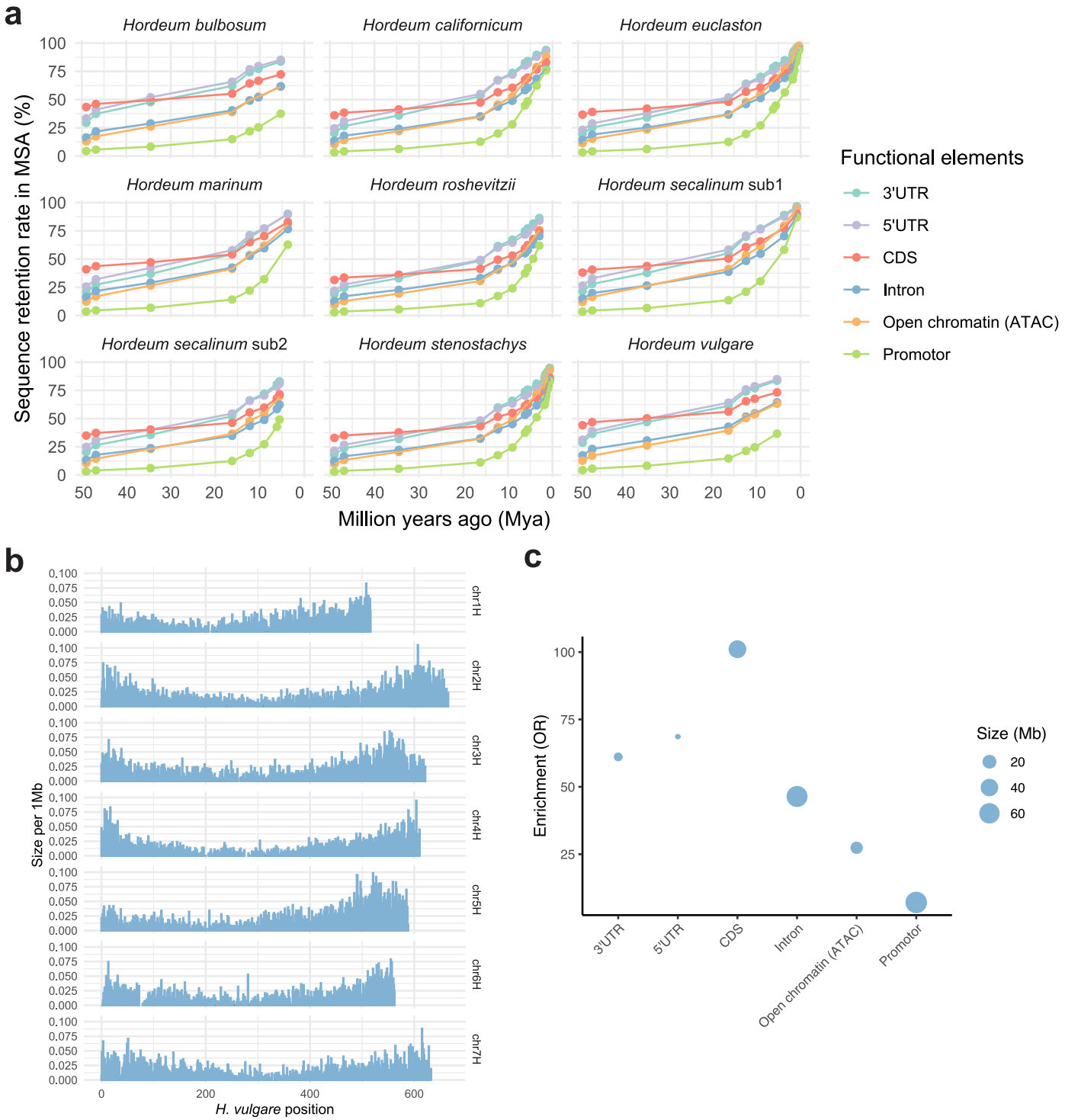


Extended Data Fig. 8 | See next page for caption.

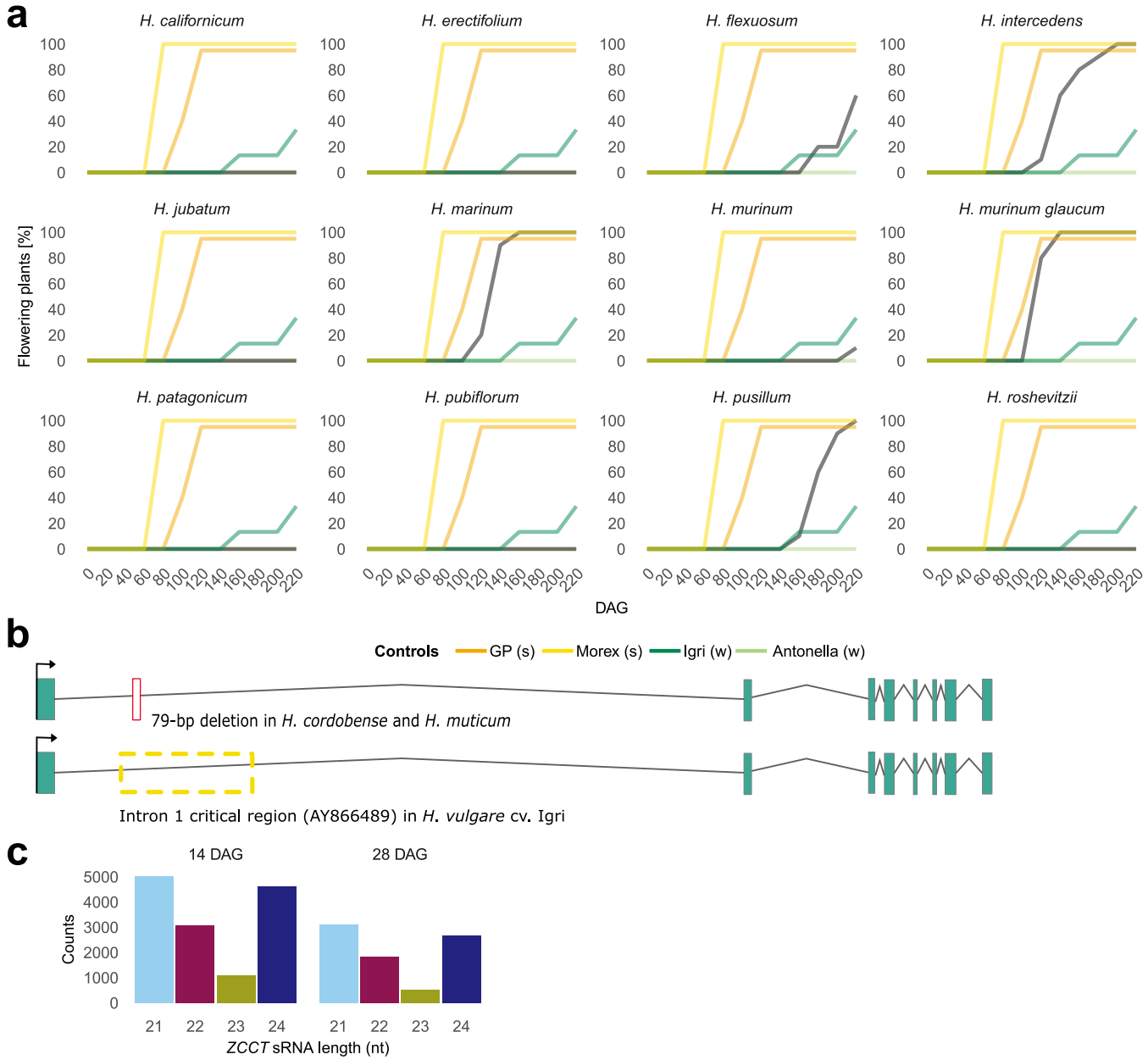
Extended Data Fig. 8 | Transposon dynamics underlie genome expansion in *Hordeum*. (**a**) PCA of 334,142 full-length *RLC_BARE1* elements; each dot represents one element, colored by insertion time (Ma). (**b**) PCA of 254,001 full-length *RLG_Sabrina* elements. (**c**) Insertion time distributions of LTR elements in *H. murinum* subsp. *murinum* (BCC2009) and *H. murinum* subsp. *glaucum* (BCC2017). *H. murinum* subsp. *glaucum* (BCC2017) is the diploid progenitor of subgenome 1 in BCC2009 (ref. 1). (**d**) Insertion time distributions of full-length LTR elements in *H. secalinum* and *H. gussoneanum*. *H. gussoneanum* is the diploid progenitor of *H. secalinum*[1]. (**e**) Heatmap showing TE families expanded in *H. secalinum* subgenome 1 relative to *H. gussoneanum*. (**f**,**g**) Inference of post-allopolyploidization TE insertion dynamics based on full-length *LTR/Copia* elements shared between subgenomes. Subgenome-derived TEs represent TE insertions detected in one subgenome that are shared with the other subgenome but absent from the corresponding diploid progenitor, reflecting TE accumulation after subgenome fusion. The onset of the insertion peak indicates the estimated timing of allopolyploid formation. (**f**) *H. secalinum*. (**g**) *H. jubatum*, *H. californicum* and *H. roshevitzii* are the respective diploid progenitors of subgenomes 1 and 2 in *H. jubatum*[1].



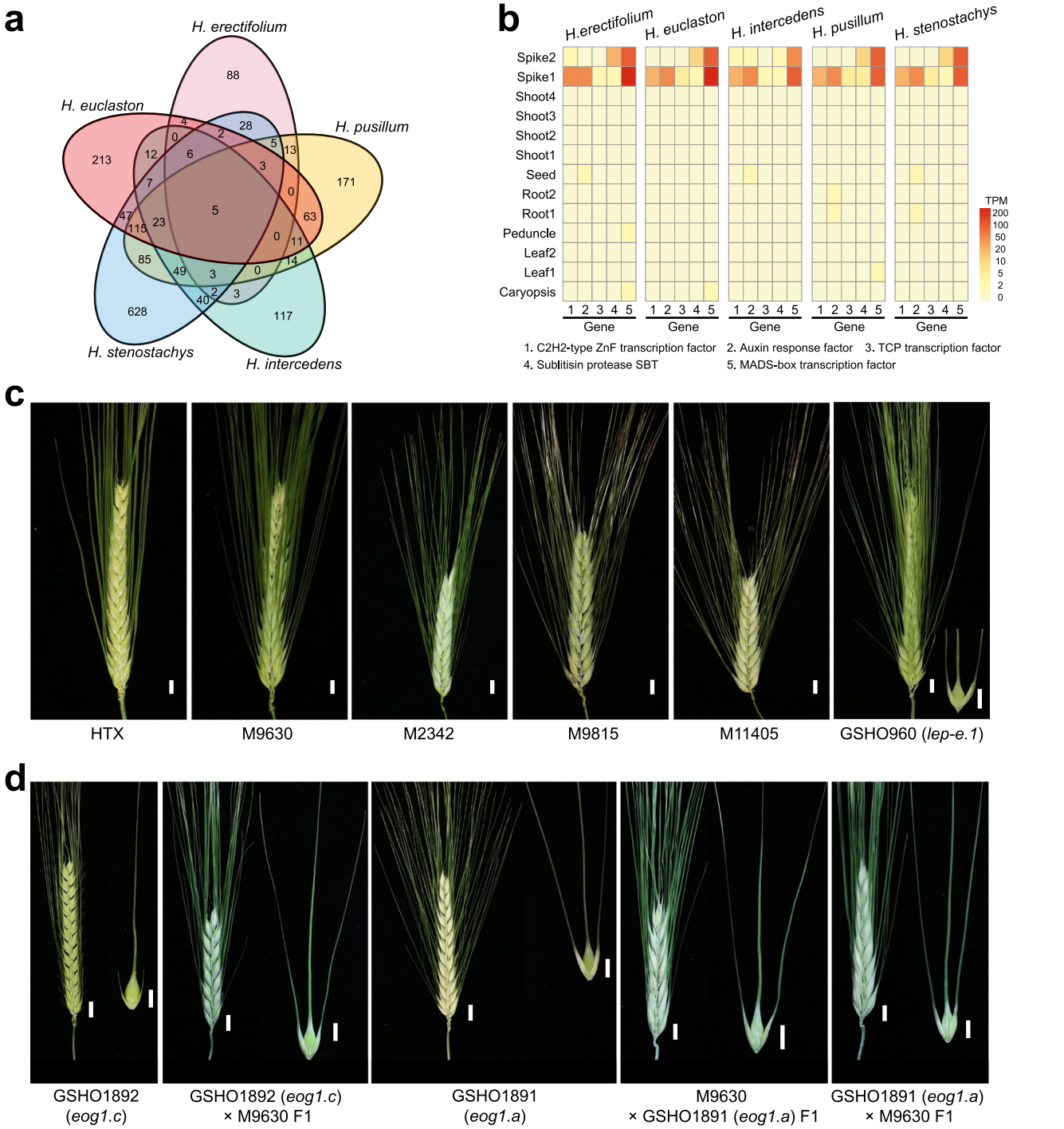
Extended Data Fig. 9 | Whole-genome alignment of 30 *Hordeum* and 4 outgroup haplomes. (**a**) Summary of genome-wide alignments among 30 *Hordeum* and 4 outgroup haplomes. Left: species tree with cumulative alignment length at each node; right: bar plots showing aligned genome size for each species. (**b**) Distribution of alignment depth per 100-kb window based on *H. vulgare*.



Extended Data Fig. 10 | Sequence conservation and divergence across *Hordeum*. (**a**) Genome-wide sequence retention of functional elements along the phylogeny based on multiple sequence alignments (MSAs) of *Hordeum* species and 4 outgroup taxa (as shown in Extended Data Fig. 9). (**b**) Distribution of conserved sequences identified by PhastCons across 30 *Hordeum* haplomes. A y-axis value of 0.1 corresponds to 100 kb of conserved sequence per 1 Mb bin. (**c**) Enrichment of conserved elements in different functional categories; dot size indicates abundance.



Extended Data Fig. 11 | Flowering behavior and molecular basis of spring growth habit in wild *Hordeum*. (**a**) Percentage of 12 wild *Hordeum* species plants flowering under non-vernalizing conditions. The cultivars Golden Promise (GP) and Morex versus cultivars Igri and Antonella serve as the spring-type (s) and winter-type (w) controls, respectively. DAG – days after germination. (**b**) Schematic representation of the *VRN1* gene. The 79 bp deletion present in the spring-type *H. cordobense* and *H. muticum* is outlined by a red rectangle and falls within the *VRN1* intron 1 critical region as defined in winter cultivar Igri (AY866489)[51]. (**c**) Size distribution of sRNAs mapped to *H. cordobense ZCCT-2* and *-3* genes at 14 and 28 DAG.



Extended Data Fig. 12 | Conserved spike-specific genes associated with glume elongation. (**a**) Venn diagram representing genes with spike-specific expression across five *Hordeum* species carrying the I genome. Spike-specific expression was defined as its transcripts per million (TPM) value in 11 of 13 analyzed tissues was less than 5% of its combined expression in the two spike tissues (spike1, W3–6; spike2, W9-10). (**b**) Heatmap of five conserved spike-specific genes among the five species. Functional annotations of these five genes are provided. Scale bar indicates the TPM value. Descriptions of the analyzed tissues and transcriptional profiles of the five genes are given in Supplementary Table 17. (**c**) Glume elongation phenotype on a spike. (**d**) Glume phenotype in F_1 plants through crossing M9630 with other *eog1* mutants. Scale bar = 5 mm.

nature portfolio

Corresponding author(s): Martin Mascher

Last updated by author(s): Oct 12, 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data collection used proprietary software of PacBio and Illumina for raw data processing.
Data analysis	Multiple published software packages were used in the analysis including: bcftools (v1.21), BEDTools (v2.26.0 or v2.30.0), Bionano Solve software (version 3.6.1_11162020), BLAST (v2.12.0+), BUSCO (v5.4.6 or v5.8.0), BWA (v0.7.17), Cactus (v2.6.5), Canu (v2.1.1), CASTLES-Pro (v1.23.3.6), CD-HIT (version 4.8.1), clusterProfiler package (v.4.14.3), cutadapt (v3.4), deepTools (v3.5.6), edgeR (v.4.2.1), eggNOG-mapper web server (v.2.1.12), EMBOSS (v6.5.7.0), ETE3 (version 3.1.2), eulerr package (v.6.1.0), fastp (v0.20.1), FeatureCounts (v2.0.1) , FlashPCA2, flexidot (v1.06), Gblocks (version 0.91b), Geneious Prime software (v2024.0.4, v2024.0.5), GENESPACE, GeSeq, ggplot2 (v3.5.1), goatools (v1.4.4), hifiasm (v0.13-r308 or v0.15.5-r350 or v0.16.1-r375), HMMER , IQ-TREE (version 2.2.2.6), isoseq3 (version 3.8.2), liftoff, MACS2 (version 2.2.9.1, MAFFT (v7.490 or v7.520), MarkDuplicates (v2.21.9, Mash (version 2.3), minimap2 (v2.24-r1122), MUMmer4 (v4.0.0beta2), MUSCLE (v3.8), odgi (v0.8.3-26-gbc7742ed), OMARK (omamer verision 2.1.0/omark version 0.3.1), Orthofinder (version 2.5.5), PAML (version 4.4), panEDTA (v0.1), Pansel, pbmm2 (version 1.10.0) , PHAST package (v1.4), pheatmap , PhyML (v.3.1), phytools, psauron (v. 1.0.6), pyGenomeTracks (v3.9), RAxML (version 8.2.12), RepeatMasker (version 4.1.1), RGAugury (v.2.2), rsem (v1.3.3), SAMtools (v1.10, v1.11 or v1.16.1), seqkit (v.2.9.0, seqkit (v0.9.1 or v.2.9.0), ShortStack (v4), STAR (v.2.7.7a), subread (v.1.5.2), SyRI (version 1.5.5), TBtools II, Trim Galore (version 0.6.4_dev, trimAl (v.1.2) , Trimmomatic (v.0.32), TRITEX pipeline, vcfbub (https://github.com/pangenome/vcfbub), vcfwave (vcflib v1.0.2), vg (v1.61.0), wfmash (v0.9.1-3), Jellyfish(v2.2.6), findGSE(v1.94), Merqury(v1.3)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

AAll sequence data generated in this study have been deposited at the European Nucleotide Archive (ENA) under umbrella BioProject PRJEB128330. Accession codes for individual accessions are listed in Supplementary Table 3 (genome sequencing) and Supplementary Table 6 (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 Repository148 (http://dx.doi.org/10.5447/ipk/2026/11). The pangenome graph in GFA format has been deposited in the Plant Genomics and Phenomics Research Data Repository148 under DOI: http://doi.org/10.5447/ipk/2025/12.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	NA
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We selected 27 accessions representing 25 Hordeum species to cover all recognized diploid genome groups and representative tetraploid and hexaploid lineages. Taxon sampling was guided by the phylogenetic framework of Blattner and Brassac (2015). The sample size was determined by the aim of achieving broad phylogenetic representation across the genus rather than by statistical power analysis.
Data exclusions	No samples were excluded from the comparative genomic analyses. Sequencing datasets that failed quality control were excluded prior to data deposition.
Replication	HiFi and Hi-C sequencing were performed on the same individual plant for each accession. Genome assemblies were not biologically replicated. RNA-seq experiments for differential gene expression analysis included >= 3 biological replicates per condition.
Randomization	Randomization was not applicable to the comparative genome assembly and phylogenomic analyses, as accessions were selected to represent the phylogenetic diversity of Hordeum.
Blinding	Blinding was not applied to the comparative genomic analyses because species identities were required for genome assembly, annotation and phylogenetic comparisons.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☐	☒ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	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.
Novel plant genotypes	No novel seed stocks were generated.
Authentication	Species identities were assessed by comparing phylogenetic trees inferred from the genome assemblies with previously published phylogenies. Genome assembly quality was evaluated using k-mer-based analyses.