

A genus-wide pangenome of Hordeum

Martin Mascher

`mascher@ipk-gatersleben.de`

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben <https://orcid.org/0000-0001-6373-6013>

Jia-Wu Feng

Leibniz Institute of Plant Genetics and Crop Plant Research <https://orcid.org/0000-0003-3737-4850>

Axel Himmelbach

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) <https://orcid.org/0000-0001-7338-0946>

Andreas Houben

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) <https://orcid.org/0000-0003-3419-239X>

Nils Stein

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) <https://orcid.org/0000-0003-3011-8731>

Yi-Tzu Kuo

IPK Gatersleben

Reza Haghi

IPK Gatersleben

Frank Blattner

Institute of Plant Genetics and Crop Plant Research <https://orcid.org/0000-0001-7731-7741>

Anne Fiebig

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) <https://orcid.org/0000-0003-3159-3593>

Hailin Zhang

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben

Zihao Zhu

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) <https://orcid.org/0000-0002-2296-2724>

Jan Safar

Institute of Experimental Botany

Hana Simkova

Centre of the Region Haná for Biotechnological and Agricultural Research <https://orcid.org/0000-0003-4159-7619>

Marie Chudecká

Institute of Experimental Botany <https://orcid.org/0009-0000-8508-2073>

David Kopecky

Institute of Experimental Botany

Marek Szecowka

Institute of Experimental Botany

Helena Toegelová

Institute of Experimental Botany of the Czech Academy of Sciences

Ales Pecinka

Institute of Experimental Botany of the Czech Academy of Sciences <https://orcid.org/0000-0001-9277-1766>

Beata Strejckova

Institute of Experimental Botany

Ping Yang

Institute of Crop Sciences, Chinese Academy of Agricultural Sciences <https://orcid.org/0000-0002-6977-8449>

Congcong Jiang

Institute of Crop Science CAAS <https://orcid.org/0000-0002-9201-4391>

Guangqi Gao

Institute of Crop Science CAAS

Yanlong Jin

Institute of Crop Science CAAS

Chaodan An

Institute of Crop Science CAAS

Manuel Spannagl

Helmholtz Helmholtz-Center Munich, German Research Center for Environmental Health

Klaus Mayer

Helmholtz Center Munich <https://orcid.org/0000-0001-6484-1077>

Thomas Lux

Helmholtz Zentrum München <https://orcid.org/0000-0002-5543-1911>

Maria von Korff

Heinrich Heine University

Einar Haraldsson

University of Düsseldorf

Thomas Wicker

University of Zurich <https://orcid.org/0000-0002-6777-7135>

Shun Sakuma

<https://orcid.org/0000-0003-1622-5346>

Takao Komatsuda

Shandong Academy of Agricultural Sciences

Karol Krak

Institute of Botany CAS

Simon Corrales Villanueva

Institute of Botany CAS <https://orcid.org/0000-0002-0311-878X>

Vaclav Mahelka

Institute of Botany, The Czech Academy of Sciences

Yann Bertrand

Institute of Botany CAS

Radim Cegan

Institute of Biophysics, Czech Academy of Sciences

Ian Small

University of Western Australia <https://orcid.org/0000-0001-5300-1216>

Joanna Melonek

Australian National University <https://orcid.org/0000-0003-4471-2520>

Robbie Waugh

James Hutton Institute <https://orcid.org/0000-0003-1045-3065>

Micha Bayer

The James Hutton Institute <https://orcid.org/0000-0003-0041-3115>

Biological Sciences - Article

Keywords:

Posted Date: October 28th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-7766969/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

A genus-wide pangenome of *Hordeum*

Jia-Wu Feng^{1,22}, Congcong Jiang^{2,22}, Beáta Strejčková^{3,22}, Marek Szećówka^{3,22}, Einar Baldvin Haraldsson^{4,22}, Thomas Lux^{5,22}, Chaodan An², Micha Bayer⁶, Yann J.K. Bertrand⁷, Marie Chudecká³, Simón Villanueva Corrales⁷, Radim Čegan^{3,8}, Anne Fiebig¹, Guangqi Gao², Reza Haghi¹, Axel Himmelbach¹, Andreas Houben¹, Yanlong Jin², Karol Krak^{7,9}, Takao Komatsuda¹⁰, Yi-Tzu Kuo¹, Václav Mahelka⁷, Klaus F.X. Mayer^{5,11}, Joanna Melonek¹², Ales Pecinka³, Shun Sakuma¹³, Hana Šimková³, Ian Small¹⁴, Manuel Spannagl^{5,15}, Nils Stein^{1,16}, Helena Toegelová³, Robbie Waugh^{6,17,18}, Thomas Wicker¹⁹, Hailin Zhang¹, Zihao Zhu¹, Jan Šafář³, David Kopecký³, Frank R. Blattner¹, Ping Yang², Maria von Korff^{4,20}, Martin Mascher^{1,15,21}

¹Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Seeland, Germany

²State Key Laboratory of Crop Gene Resources and Breeding, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

³Institute of Experimental Botany, Czech Academy of Sciences, Centre of Plant Structural and Functional Genomics, Olomouc, Czech Republic

⁴Institute of Plant Genetics, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

⁵PGSB – Plant Genome and Systems Biology, Helmholtz Center Munich – German Research Center for Environmental Health, Neuherberg, Germany

⁶International Barley Hub (IBH), James Hutton Institute (JHI), Dundee, Scotland,

⁷Institute of Botany, Czech Academy of Sciences, Průhonice, Czech Republic

⁹Institute of Biophysics, Czech Academy of Sciences, Department of Plant Developmental Genetics, Brno, Czech Republic

⁹Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic

¹⁰Shandong Academy of Agricultural Sciences (SAAS), Licheng District, Jinan, Shandong, China

¹¹School of Life Sciences, Technical University Munich, Freising, Germany

¹²Division of Plant Sciences, Research School of Biology, The Australian National University

¹³Faculty of Agriculture, Tottori University, Tottori, Japan

¹⁴School of Molecular Sciences, The University of Western Australia

¹⁵Centre for Crop & Food Innovation, Food Futures Institute, Murdoch University, Murdoch, Western Australia, Australia

¹⁶Institute of Agricultural and Nutritional Sciences, Martin Luther University Halle-Wittenberg, Halle, Germany

¹⁷School of Agriculture, Food and Wine, University of Adelaide, Waite Campus, Urrbrae, South Australia, Australia

¹⁸School of Life Sciences, University of Dundee, Dundee, UK

¹⁹Department of Plant and Microbial Biology, University of Zurich, Zurich, Switzerland

²⁰Cluster of Excellence on Plant Sciences “SMART Plants for Tomorrow’s Needs”, Düsseldorf, Germany

²¹German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

²²These authors contributed equally to this work.

✉ Correspondence should be addressed to Martin Mascher (mascher@ipk-gatersleben.de, Maria von Korff (Maria.Korff.Schmising@uni-duesseldorf.de), Ping Yang (yangping@caas.cn), Frank R. Blattner (blattner@ipk-gatersleben.de), David Kopecký (kopecky@ueb.cas.cz) or Jan Šafář (Safar@ueb.cas.cz).

Current address: C.A. – Shiping County Government Offices Administration Bureau, Shiping, China; Y.J. – Key Laboratory of Oasis Eco-Agriculture, College of Agriculture, Shihezi University, Shihezi, China

The genus *Hordeum*, which includes cultivated barley and its wild relatives, represents a diverse group of grasses with agricultural, ecological, and evolutionary importance. Here, we present a genus-wide pangenome based on 25 chromosome-scale genome sequences encompassing all known diploid *Hordeum* species and selected tetraploids. Comparative analysis reveals extensive structural variations, including large-scale inversions and chromosomal rearrangements, which disrupt proximal chromosomal collinearity and reflect lineage-specific genome remodeling. Using single-copy gene phylogenies, we identify multiple deep interspecific hybridization events, segmental exchanges, and introgressions, particularly in regions enriched for conserved inversions. We also uncover widespread horizontal gene transfer from Panicoideae grasses into *Hordeum*, with functional integration of transferred genes. Transposable element dynamics, especially retrotransposon proliferation, account for much of the observed genome size variation and offer insights into polyploid genome evolution. Whole-genome alignments reveal contrasting evolutionary patterns: while coding genes remain highly conserved, regulatory elements show rapid turnover, providing a substrate for phenotypic diversification. 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 Americas¹⁻³. *Hordeum* is classified into two subgenera, harbouring the cytogenetically defined genomes H and Xu, and I and Xa, respectively⁴. The basic chromosome number is seven, and while most species are diploid, tetraploid and hexaploid taxa have also been described. Cultivated barley (*H. 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 significant grain, valued for its use in food, animal feed, and brewing⁵. Cultivated barley and its wild progenitor (*H. vulgare* ssp. *spontaneum*) constitute a single interfertile gene pool, allowing for frequent gene flow between wild and cultivated forms⁶. It defines barley's primary gene pool, which breeders have long exploited to introduce desirable traits⁷. *H. bulbosum* has contributed valuable disease resistance genes through introgression⁸⁻¹¹ due to its ability to hybridize directly with cultivated barley. The other *Hordeum* species are not domesticated but play important ecological and agricultural roles in ecosystem stability, environmental adaptation, and forage use¹²⁻¹⁶. Taken 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^{2-4,17,18}. 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 (SVs), gene presence-absence, and evolutionary dynamics¹⁹⁻²³. Investigating SVs helps to elucidate evolutionary processes and can inform crop breeding strategies²⁴. However, the full extent of structural variation within *Hordeum* remains largely unexplored. To address this gap, we generated and analysed chromosome-scale genomes from 23 *Hordeum* species spanning all major evolutionary lineages. This genus-wide pangenome provides the first comprehensive view of SV across *Hordeum*, offering new insights into *Hordeum* evolutionary history and a genomic foundation for future pre-breeding efforts.

Hordeum species assembly, annotation and pangenome

We selected 25 accessions from 23 different *Hordeum* species for chromosome-level genome assembly (**Extended Data Fig. 1a, Supplementary Table 1 and Supplementary Fig. 1-23**). This includes 22 diploid and three tetraploid (sub-)species (*H. jubatum*, *H. secalinum* and *H. murinum* subsp. *murinum*). The diploid species comprise all known diploid species in the *Hordeum* genus². Twenty-two genome sequences were assembled using PacBio HiFi accurate long reads and Hi-C data (**Supplementary Table 2**), employing the TRITEX pipeline^{25,26}. The sequence assemblies of *H. vulgare* and *H. bulbosum* were sourced from our previous work²⁷. The previously published chromosome-level assembly of *H. marinum* H559²⁸ was updated using the TRITEX pipeline. The average N50 of our assemblies is 39.96 Mb. Assembly sizes range from 2.90 to 4.48 Gb (average 3.92 Gb) in diploids and from 7.49 to 8.56 Gb (average 8.1 Gb) in tetraploids. The chloroplast genome sizes range from 136,335 bp to 137,079 bp (average: 136,837 bp) (**Supplementary Table 1**). An average LTR Assembly Index (LAI) of 14.83, an average consensus quality value (QV) of 67.13, and an average *k*-mer completeness of 97.10% attest to the quality of our assemblies (**Supplementary Table 3**). To assess the gene content in the *Hordeum* pangenome, we *de novo* annotated all genomes using transcriptome data (**Online Methods, Supplementary Table 4**). A gene consolidation method was applied to ensure high comparability of gene predictions. The number of high-confidence gene models identified ranges from 35,573 for diploid *H. murinum* to 89,407 for tetraploid *H. jubatum* (**Supplementary Table 5**). We benchmarked the quality of the *Hordeum* gene predictions against BUSCO (v.5.8.0) with the poales_odb12 lineage dataset²⁹, representing 6,282 Poales near-universal single-copy orthologs. On average, we obtained 98.20% as complete copies, 0.54% as fragmented copies, and 1.26% as missing genes (**Supplementary Table 5**). Additional benchmarking with OMARK and psauron³⁰ confirms the overall high accuracy and completeness of our gene annotation approach (**Supplementary Table 5**). Transposable element annotation using the panEDTA pipeline³¹ revealed that transposons constituted on average 86% of the genome, ranging from 83.27% to 88.93% (**Supplementary Table 6**).

Based on the gene sets of all species, we constructed a gene-based pangenome for *Hordeum*, with 44.02% of genes classified as core genes, 55.03% as shell genes, and 0.95% as cloud genes (**Extended Data Fig. 1b and c**). We find GO terms related to phosphorylation and kinase activity are enriched in the set of core genes, RNA binding and DNA helicase activity are enriched in the set of shell genes, whilst defense response and response to oxidative stress are enriched in the

set of cloud genes (**Supplementary Fig. 24**). Nucleotide-binding Leucine-rich Repeat (NLR) genes are key components of the plant immune system, playing 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 the identified NLR genes revealed that I group genomes contain significantly more NLR genes (**Extended Data Fig. 2a and b**). To investigate this further, we constructed an NLR pangenome across *Hordeum* haploid genome equivalents (or “haplomes”). This pangenome comprises 799 orthologous groups (OGs), categorized as 13.77% core, 54.82% shell, and 31.41% cloud gene families (**Extended Data Figure 2c**). Moreover, examination of NLR gene clusters showed that chromosomes 4H and 5H harbor significantly fewer clusters compared to the other chromosomes (**Extended Data Figure 2d**). Similarly, genome-wide analyses of genes encoding pentatricopeptide repeat (PPR) proteins and mitochondrial transcription termination factors (mTERFs), which include members involved in fertility restoration in plants carrying cytoplasmic male sterility (CMS)³⁴, reveal expansions in the I genome group (**Supplementary Fig. 25a, b and Supplementary Table 7**). In particular, mTERF genes from the *Restorer of fertility like* (RFL) clade, clustered on the short arm of chromosome 6H, show remarkable copy number variation (**Supplementary Fig. 25c, d**). This genomic region includes the *Rfm3* locus, which partially restores fertility in plants carrying the *male sterility maternal* (*msm1*) cytoplasm used in hybrid breeding in barley³⁵ (**Supplementary Fig. 25d**). To evaluate graph-based analysis for highly diverged accessions, we built a pangenome graph using progressive cactus³⁶ and hal2vg, resulting in 2.53 billion nodes and 3.46 billion edges (68.7 Gbp). Around 68% of nodes were single-base, highlighting strong base-level divergence. Although tasks like indexing were computationally prohibitive (**see Online Methods**), we successfully performed several global analyses, including core/shell/cloud analysis, saturation, conservation, and variant surveys (**Supplementary Fig. 26-29**). Our experience demonstrates that scalable methods for graph exploration and interpretation remain a major challenge.

Large SVs rapidly reshape proximal region

Previous studies have inferred the evolutionary relationships among *Hordeum* species from plastome and partial gene sequences^{2,18}. However, the relationships among some species remained weakly supported. In this study, we reconstructed a maximum likelihood (ML) tree using a genome-wide single-copy gene set. Unlike previous phylogenies, the tree fully resolved all species relationships with maximum bootstrap support values, especially for the recent species divergence within the I genome group (**Supplementary Fig. 30**). To further assess the phylogenetic relationships, we additionally constructed a coalescence-based species tree using ASTRAL-Pro3 (**Supplementary Fig. 31**), which integrates individual ML gene trees, as well as a genome-wide phylogenetic tree inferred by CASTER based on whole-genome multiple sequence alignments (MSA) (**Supplementary Fig. 32 and 33**). We also estimated the divergence times of these species using fourfold synonymous third-codon transversion (4Dtv) sites (**Fig. 1a**). We estimated the divergence time between H and Xu at 10.3 million years ago (Mya), while I and Xa diverged 9.2 Mya. The crown-group age of *Hordeum* dates back approximately 12.4 Mya, with the most recent species divergence event being *H. euclaston*/*H. intercedens* around 0.3 Mya (**Fig. 1a and Supplementary Table 8**). The chromosomes of Triticeae species exhibit high collinearity at their distal chromosome regions³⁷, which is also the case in *Hordeum* (**Fig. 1b**). However, when aligned, the proximal regions show substantial loss of collinearity and highly fragmented

alignments. Interestingly, whole-genome comparisons within *Hordeum* species revealed that the extent of chromosomal rearrangements increases with species divergence time, suggesting that such rearrangements are a primary driver of collinearity loss and fragmented alignments in proximal regions (**Fig. 1c and Supplementary Fig. 34-39**). Among all chromosomes, chromosome 4H exhibited the most extensive proximal collinearity loss. Strong collinearity was observed among recently diverged species, along with numerous SVs detected in species that diverged less than 1 million years ago. However, in more distantly related genome groups, H-Xu and Xa-I (divergence time >9 Mya), proximal collinearity became entirely fragmented (**Fig. 1c**). To estimate the size of non-collinear regions, we applied normalized coverage to identify these regions in species that diverged more than 9 mya (**Extended Data Fig. 3a and b; Supplementary Fig. 40-44; 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 9**). Given the importance of transposable elements (TEs) in shaping genomes, we examined TE content in proximal and distal regions. Proximal regions harbored a significantly higher proportion of TEs, with LTR/Gypsy elements contributing more substantially compared to distal regions (**Extended Data Fig. 3c and d, and Supplementary Fig. 45**). Based on these findings, we propose a “Stretch and Twist” model to explain the rapid loss of collinearity and alignment fragmentation. The “Stretch” component refers to alignment displacement or loss due to TE amplification, while the “Twist” describes large-scale chromosomal rearrangements driven by SVs (**Extended Data Fig. 3e**).

Large SVs reveal hybridization

In contrast to the dynamic SVs found in proximal regions, the distal regions displayed more conserved, shared SVs. This highlights the limitation of using *H. vulgare* alone as a reference for identifying SVs across more distantly related *Hordeum* species. Notably, large SVs 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. 4a**). In total, 1,203 inversions larger than 2 Mb were identified across *Hordeum* species (**Extended Data Fig. 4a**). The construction of a comprehensive inversion map revealed six inversions on chromosome 1H that were unexpectedly shared across divergent phylogenetic lineages (**Fig. 2a**). We hypothesized that these large-scale SVs 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, fully resolved with maximum bootstrap support, confirmed that these SVs were specifically shared between a subset of South American I and H/Xu genome groups (**Fig. 2b**). Additionally, we observed that rapid diversification of species occurred within this lineage in the 2 million years following the hybridization event — a pattern absent in other phylogenetic branches (**Fig. 1a**). These findings suggest that large structural variations in the distal chromosome regions preserve signatures of deep hybridization events shared among multiple species.

To further investigate whether other large chromosomal segments exhibit similar hybridization signatures within the phylogenetic tree, we developed a genome-wide scanning approach based on the conservation of single-copy genes. In this approach, the single-copy gene set was divided into multiple sliding windows, and Normalized Robinson-Foulds (nRF) distance was used to assess the topological differences between local phylogenetic trees and the whole-genome

phylogenetic tree (**Extended Data Fig. 4b**). Our initial analysis revealed multiple peaks in nRF values, indicating potential introgression events. We established an nRF threshold of 0.18 to distinguish primary peaks from regions with significant phylogenetic discordance (**Extended Data Fig. 4c**). As expected, we identified a strong hybridization signal on chromosome 1H, validating the effectiveness of our scanning approach. Additionally, we detected more than 10 consecutive windows with strong phylogenetic discordance signals at the terminal regions of chromosomes 4 and 5 (**Fig. 2c**). To further explore the underlying causes of these inconsistencies, we reconstructed local phylogenetic trees using the single-copy gene sets from these regions. By comparing the regions that exhibited phylogenetic discordance with the whole-genome tree, we identified two large-scale segmental exchanges occurring within *H. jubatum*, with maximum bootstrap support (**Extended Data Fig. 4d and e**). Surprisingly, these large-scale segmental exchanges were directly associated with major duplication and deletion events on chromosome 5H previously identified in *H. jubatum* (**Fig. 2d, and Supplementary Fig. 12, 46, 47 and 48**). Our findings demonstrate that large structural variations discovered in genus-wide pangenomes serve as evidence of multiple cross-genome hybridization events, thus shedding light on the complex evolutionary history of *Hordeum* species.

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 HGT occurrences, challenging prior assumptions^{38,39}. Our previous research demonstrated extensive HGT in *Hordeum*. Notably, species within section *Stenostachys* (I genome; 16 of 22 diploid species) possess ribosomal DNA (rDNA) acquired from grasses in the subfamily Panicoideae (e.g., *Panicum*, *Paspalum*, *Setaria*)^{40,41}. The oldest transfers (from *Panicum*), estimated to have occurred between 5 and 1.7 Mya, resulted in large genomic insertions on chromosome 5H, some spanning over 400 kb and containing transposable elements and protein-coding genes⁴². Here, we identified Panicoideae-derived nuclear genes across 20 diploid *Hordeum* species, with *H. vulgare* serving as a negative control^{40,42}. Our analysis revealed 611 foreign genes that corresponded to 126 orthogroups, predominantly matching *Paspalum* (63%) and *Panicum* (22%) (**Extended Data Fig. 5a and Supplementary Table 10 and 11**). The extent of HGT varied, with *H. flexuosum* harboring the highest number of foreign genes (68 genes corresponding to 43 orthogroups) (**Extended Data Fig. 5a**). These findings align with studies indicating that grasses frequently exchange genetic material^{43,44}. 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 evolution⁴⁵. 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. 2c**). Chromosomal mapping indicates widespread distribution of foreign genes across all seven *Hordeum* chromosomes, with chromosome 5H harboring the highest number. These genes are preferentially located in the distal regions of chromosome arms, suggesting selection pressure or integration bias (**Extended Data Fig. 5b**). While co-transfer of genes remains uncertain, previous

research suggests that large fragments of foreign DNA may be transferred together^{42,44}. A substantial fraction (82 out of 126) of Panicoideae-derived orthogroups lack native *Hordeum* homologs, indicating novel gene acquisition. Functional annotation suggests that these genes may play roles in key physiological processes such as auxin transport, light response, and transcription regulation. RNA-seq analysis detected at least one expressed HGT gene in 73 out of 126 orthogroups (**Extended Data Fig. 5c and Supplementary Table 12**). The potential adaptive significance of these genes warrants further investigation. By providing the most comprehensive genus-wide analysis of HGT in plants to date, *Hordeum* emerges as a model system for studying the evolutionary impact of horizontal gene transfer.

TEs drive genome size variation

The genome assembly sizes of *Hordeum* species range from 2.90 Gb to 8.56 Gb and are highly correlated with previous flow-cytometric estimates⁴⁶ (**Fig. 3a**). After accounting for the direct doubling effects of tetraploidization, 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 Fig. 47 and 48**). A strong correlation was observed between total TE content and haplome size (**Fig. 3b**), suggesting that TE accumulation has significantly contributed to genome size variation in *Hordeum*. Long terminal repeat (LTR) retrotransposons were identified as the primary drivers of these changes, though DNA transposons also played a role (**Fig. 3c**). By analyzing correlations between transposon superfamilies and genome size, we found that different TE superfamilies contributed to genome size change, but overall TE abundance was the most direct determinant of genome size variation (**Extended Data Fig. 6a**). We selected transposons annotated in the TREP database with an average content above 10 Mb and examined their variation across different *Hordeum* haplomes. The results showed that these TEs effectively differentiate distinct evolutionary lineages (**Extended Data Fig. 6b**). To further investigate lineage-specific TE dynamics, we focused on three LTR retrotransposon families—BARE1, Sabrina, and BAGY2—and employed TEpop analysis to track their diversification within *Hordeum* haplomes. The findings revealed that the formation of subfamilies within these TE families aligns with different evolutionary clades, making them useful markers for distinguishing genetic lineages (**Fig. 3d, Extended Data Fig. 6c and d**). The clade-specific proliferation and timing of these TE subfamilies collectively contributed to genome size variation.

To assess the impact of polyploidization on genome size variation in tetraploid subgenomes, we compared TE changes in three tetraploid genomes with their diploid progenitors. The results revealed diverse patterns: In *H. murinum*, TE expansion was minimal (**Extended Data Fig. 6e, Supplementary Table 13**); in *H. jubatum*, both subgenomes were smaller than their diploid ancestors (**Fig. 3e, Supplementary Table 13**); in *H. secalinum*, rapid TE proliferation significantly increased genome size, altering the transposon expansion pattern (**Extended Data Fig. 6f, Supplementary Table 13**). Interestingly, in *H. secalinum*, the LTR/Gypsy elements were the main contributors to genome expansion (**Supplementary Table 13**). Additionally, we identified a resurgence of the Vagabond TE family after tetraploidization, where it was significantly more active in the tetraploid genome compared to its diploid progenitors and the alternative subgenome (**Extended Data Fig. 6g–i**). While subgenome differentiation alone does not allow to

date the formation of tetraploid genomes, intersubgenomic TE insertions can provide insight into the timing of genome mergers. We inferred the fusion timeline of tetraploid subgenomes by identifying full-length transposons shared exclusively between subgenomes but absent in their respective diploid ancestors (**Fig. 3f–h**). 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, genome mergers consistently leave detectable TE signatures, providing valuable insights into polyploid evolution.

Sequence conservation across species

Whole-genome MSA provide insights into genomic relationships and help us understand sequence conservation and loss across species. To achieve this, we performed a MSA of *Hordeum* haplomes using progressive Cactus³⁶. In total, we identified 170.4 Mb of shared aligned sequences across the *Hordeum* haplomes (**Extended Data Fig. 7a**). As expected, rapidly evolving pericentromeric regions showed fewer alignments, whereas chromosome distal ends exhibited higher alignment densities (**Extended Data Fig. 7b**). The rapid contraction of these alignments is likely due to substantial reshaping of *Hordeum* genomes by TE activity over the past 12 million years. To further investigate sequence loss across different genomic regions, we analyzed open chromatin regions that were captured by Assay of Transposase-Accessible Chromatin sequencing (ATAC-seq) and annotated regions in eight *Hordeum* haplomes (**Supplementary Table 14**). We found that UTR sequences were lost at a slower rate before the divergence of barley and wheat. In contrast, open chromatin regions showed a loss pattern similar to introns, while promoter regions underwent more rapid turnover (**Extended Data Fig. 8a**). Additionally, we used phastCons to identify 74.7 Mb of conserved elements across *Hordeum* genomes. These elements were primarily enriched within genic regions (**Extended Data Fig. 8b and c**). Our results showed that 69% of conserved elements overlapped with genes, while only 6% were associated with open chromatin regions. Overall, our findings indicate that regulatory elements exhibit a high degree of turnover among *Hordeum* species, highlighting the rich diversity of regulatory sequences that could be leveraged for barley breeding. Meanwhile, the high conservation of genes provides opportunities to explore cross-species gene function and regulatory element substitution, potentially linking genetic changes to phenotypic variation.

Natural Variation Underpins Vernalization Flexibility

In cultivated barley, spring and winter varieties differ in vernalization requirements controlled by the VERNALIZATION 1 (*VRN1*) MADS-box transcription factor and VERNALIZATION 2 (*VRN2*) encoded by the cluster of zinc-finger/CCT domain transcription factor (*ZCCT*) genes^{47,48}. Flowering of winter varieties requires repression of *VRN2* by cold-induced *VRN1*⁴⁸. On the contrary, spring barleys often carry deletions in *VRN1*'s first intron and lack *VRN2*, eliminating their vernalization requirement^{49–51}. 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.

Examining *VRN1* in *Hordeum* genomes revealed extensive variation in the promoter and the regulatory first intron (5.7 to 19.5 kb) caused by transposable elements (**Supplementary Table**

15). To assess the functional impact of this variation, flowering time of 14 wild *Hordeum* species and 4 barley controls was recorded under non-vernalizing conditions (**Fig. 4a and Extended Data Fig. 9a**). Notably, *H. cordobense* and *H. muticum* flowered rapidly (~61 days), surpassing spring barley cultivars (**Fig. 4a**). Both species share a unique 79 bp deletion in *VRN1* intron 1 critical region⁵⁰, suggesting a role in reducing the vernalization requirement (**Extended Data Fig. 9b**). All species contained the *VRN2* locus with one to seven *ZCCT* copies (**Supplementary Table 15**). A key finding was an inverted repeat (IR) of two *ZCCT*s in *H. cordobense*, *H. californicum*, *H. muticum*, and *H. jubatum*. Because IRs may trigger RNA interference (RNAi)⁵², we expected silencing of *VRN2* and a spring-type phenotype. Surprisingly, only *H. cordobense* and *H. muticum* flowered early (**Fig. 4a and c, and Extended Data Fig. 9a**). Detailed analysis of *H. cordobense* revealed a non-functional *ZCCT-1* due to a premature stop codon, and the absence of a polyadenylation signal in the *ZCCT-2/ZCCT-3* IR linker, enabling read-through transcription (detected as an aberrant mRNA transcript), hairpin RNA formation, and small RNA production (**Fig. 4b and d, and Extended Data Fig. 9c**). On the contrary, winter-type *H. californicum* carried additional functional *ZCCT* copies and a polyadenylation signal in the IR linker that likely prevented RNAi, thus maintaining the *VRN2* repressive function (**Fig. 4b and e**). This explains with the expression of *VRN1*, *VRN2* and their target *FLOWERING LOCUS T (FT)* in the analyzed species (**Fig. 4f**). These findings highlight genetic and epigenetic diversity shaping 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. Altogether, our results indicate that adaptation to diverse vernalization regimes in wild *Hordeum* species has targeted the same genes and variant types as in cultivated barley, pointing to conserved mechanisms of flowering-time adaptation to winter.

A Conserved Mechanism Underlying Glume Development

The glume is a distinctive floral organ in the grass family, lacking a direct counterpart in dicotyledonous plants⁵³. 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 defense against biotic stress⁵⁴. In cultivated cereals, domestication has favored 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*, *H. pusillum*, *H. intercedens*, *H. stenostachys*, and *H. euclaston*) identified spike-specific genes expressed in young spikes at early (Waddington stages W3-6) and late (W9-10) developmental stages⁵⁵ (**Supplementary Table 16**). Five genes were consistently expressed across species, suggesting conserved regulatory mechanisms (**Fig. 5b and Supplementary Table 17**). 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 (**Fig. 5c**). In rice, loss-of-function mutations in the *NONSTOP GLUMES1 (NSG1)*, the ortholog of the C2H2-type zinc finger transcription factor result in uncontrolled glume growth⁵⁶. Mutations in its

cultivated barley ortholog (*EOG1*) lead to the elongated outer glume (*eog1*) phenotype (**Fig. 5d**). 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. 10a and b**). Scanning electron microscopy (SEM) revealed developmental divergence between wild-type and *eog1* mutants as early as Waddington stage W3.5 (**Fig. 5e**). By integrating *Hordeum* pan-genomic data with tissue-specific expression data and mutant resources of cultivated barley, we demonstrate that EOG1 functions as a conserved transcriptional repressor controlling glume development.

Discussion

High-quality genome assemblies across the *Hordeum* genus lay the groundwork for exploring its complex evolutionary history. Chromosome-scale resolution allows us to investigate how SVs have shaped genome evolution, particularly in disrupting chromosomal collinearity. Traditionally, grass genomes were thought to possess relatively stable chromosome structures⁵⁷; however, our findings challenge this assumption. Extensive SVs and chromosomal rearrangements have contributed significantly to genomic divergence, ultimately resulting in the near-complete loss of collinearity in proximal regions. These rearrangements are not only markers of divergence but also critical to understanding the speciation events within *Hordeum*. Interestingly, SVs accumulate at different rates across chromosomes. While proximal regions exhibit high variability, distal regions evolve more slowly. This uneven evolutionary rate has preserved footprints of deep hybridization events across the genus, offering a historical record of interspecific gene flow. These insights offer critical empirical evidence for understanding hybrid speciation and genome fusion events and could prove essential for designing novel interspecies crosses in crop improvement strategies. Whole-genome multiple sequence alignments enabled us to compare genome-wide conservation across *Hordeum* species. Unlike animal genomes, which often display strong interspecies conservation, plant genomes, especially within rapidly evolving genera like *Hordeum*, undergo dynamic changes driven by short generation times and rapid transposon expansion. These processes fuel the ongoing remodeling of plant genomes, resulting in high variability even among closely related taxa. Despite this rapid evolution, many genes remain conserved, while regulatory elements exhibit considerable turnover. This contrast offers unique opportunities: while core gene functions are retained, lineage-specific regulatory innovation can be harnessed for crop improvement. 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 behavior. Ultimately, the expansion of *Hordeum* genomic resources does more than increase our understanding of gene content. It opens new avenues for exploring regulatory potential, identifying SV-driven divergence, and reconstructing species relationships. These advances not only deepen our understanding of *Hordeum* evolution but also enhance the genomic toolkit for future barley breeding efforts.

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.Himmelbach, E.B.H., and Y.J. Transcriptome sequencing: R.H., A.Himmelbach, E.B.H. and S.S. Morphological analysis: T.K. ATAC-seq: Z.Z., R.H. and A.Himmelbach. 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. Genome assembly validation: J.-W.F., Y.-T.K. and A.Houben. 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 transposable elements: 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. Pangenome graph construction and analysis: M.B., S.V.C, and Y.J.K.B. Analysis and interpretation of horizontal gene transfer: 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.

Acknowledgements

We acknowledge Manuela Knauff, Jacqueline Pohl and Pascal Jaroschinsky for their expert technical work during library construction and DNA sequencing. This work was supported by the European Research Council (ERC) under the European Union's Horizon Europe research and innovation programme (TRANSFER, No. 949873 to M.M. and PERLIFE, No. 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 lab was funded by the Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy—EXC-2048/1—Project ID: 390686111 and the GRK 2064: Water use efficiency and drought stress responses: From Arabidopsis to Barley, Project ID: 252965955. J. Š., A.P., B.S., R.Č., M.C., M.Szecówka and D.K. received funding from Czech Science Foundation, grant no. 22-00204S MEYS project “TowArds Next GENeration Crops (TANGENC)” (No. 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) grants #W2411025, Shandong Province Key R&D Program for Shandong and the China-Israel Cooperation Program (2023KJHZ003). M.Spannagl and K.F.X.M. acknowledge support by 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 authors acknowledge Research Computing at the James Hutton Institute for providing computational resources and technical support for the “UK’s Crop

Diversity Bioinformatics HPC”⁵⁸ (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.

Figure Legends

Figure 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 Fourfold Degenerate Third-Codon Transversion (4DTv) sites from genome-wide single-copy genes. Colored rectangles represent different genome groups. Star-marked nodes correspond to those shown in panel (c). (b) GENESPACE synteny plot between 29 *Hordeum* haplomes and 4 outgroup genomes. Grey bars represent chromosomes, scaled according to the number of genes with syntenic alignments to other genomes. (c) Dotplot comparisons of chromosome 4H across genomes diverged at different times. Yellow indicates nodes diverged within 1 million years; blue indicates 1–9 million years; and magenta indicates nodes older than 9 million years. Two species were randomly selected from each node for whole-genome comparison (**Supplementary Table 20**).

Figure 2: Structural variation reveals past introgression in *Hordeum* species. (a) Dotplot of all genomes aligned to *T. monococcum* in the chr1A: 400–650 Mb region. (b) Phylogenetic tree based on single-copy genes within the chr1A: 400–650 Mb region of *T. monococcum*. Colored 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 = 2.03×10^{-5}). (c) Distribution of normalized Robinson–Foulds (nRF) distances between gene trees in genome windows and the species-wide tree. Vertical lines represent genomic regions containing single-copy genes; points mark the center of each window. Orange indicates RF values >0.18, while light blue indicates values ≤0.18. (d) Comparative analysis of chromosome 5H between *H. jubatum* and its ancestral diploid genomes. Orange lines indicate regions with duplications and deletions.

Figure 3: Transposable elements reveal genome size variation and tetraploid formation. (a) Correlation between assembly size and DNA content. (b) Correlation between pseudomolecule size and transposable element (TE) content. (c) Correlation between pseudomolecule size and retrotransposon and DNA transposon content. (d) Principal component analysis (PCA) of 68,707 full-length *RLG_BAGY2* retrotransposon sequences across *Hordeum* species. Each dot represents one retrotransposon; colors indicate estimated insertion times. (e) Distribution of insertion times of full-length LTR elements in *H. jubatum*, *H. californicum*, and *H. roshevitzii*. *H. californicum* and

H. roshevitzii are the respective diploid progenitors of subgenomes 1 and 2 in *H. jubatum*². **(f-h)** Insertion time distribution of full-length *LTR/copia* elements in tetraploid species. Elements were selected using CD-HIT clustering from one subgenome of the tetraploid, which are absent in the ancestral diploids but present in the other subgenome. The peak indicates the timing of subgenome fusion. **(f)** *H. secalinum*; *H. gussoneanum* is the diploid progenitor of subgenome 1 in *H. secalinum*². **(g)** *H. murinum* (BCC2009), *H. murinum* (BCC2017) is the diploid progenitor of subgenome 1 in *H. murinum* (BCC2009)². **(h)** *H. jubatum*.

Figure 4: Natural variation in vernalization requirement of *Hordeum* species. **(a)** Percentage of *H. cordobense* and *H. muticum* plants (n = 10 per species) 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)** 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: exons, *: stop codon, purple ovals: transposable elements, black dot with p(A): polyadenylation signal, sand-colored region: inverted repeat. **(c)** Phenotypes of four *Hordeum* species carrying *ZCCT* inverted repeat at 65 and 200 DAG. Arrows mark visible spikes. **(d)** Transcripts per million (TPM) values at 14 and 28 DAG in *H. cordobense* and *H. californicum*. **(e-f)** Mean distribution of RNA-seq and sRNA-seq reads over three biological replicates normalized to bins per million (BPM) mapped reads in *H. cordobense* **(e)** and *H. californicum* **(f)** across the *VRN2* locus at 14 and 28 DAG. Read mapping in *H. cordobense* *ZCCT*-3 suggests an improperly spliced, likely aberrant transcripts.

Figure 5: Inhibition of glume development by a C2H2-type zinc finger transcription factor. **(a)** Central spikelet from representative *Hordeum* species. gl, glume; Scale bar = 500 μ m. **(b)** Venn diagram representing genes with spike-specific expression across five *Hordeum* species. Spike-specific expression was defined by TPM values less than 5% in other tissues within each species. **(c)** Heatmap of five conserved spike-specific genes among the five species. Fold change in transcriptional profiles relative to the mean value are shown. The scale bar represents the magnitude of gene expression fold change. Genes 1 to 5 are annotated with C2H2-type zinc finger transcription factor, auxin response factor, TCP transcription factor, subtilisin protease SBT, and MADS-box transcription factor, respectively. Descriptions of the analyzed tissues (y-axis) and transcriptional profiles of the five genes are given in **Supplementary Table 17**. **(d)** Allelic variations of a C2H2-type zinc finger transcription factor gene (*HORVU.MOREX.r3.2HG0170820.1*) in *H. vulgare* lead to glume elongation. The rectangle denotes the exon/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 with suppressed glume development, while mutants exhibit elongation in glume growth. The mutant allele *lep-e.1* (ref. ⁵⁹) carries a deletion allele of *EOG1* gene. **(e)** Differentiation on the spikelet morphogenesis along Waddington stage between the wild type 'HTX' and the mutant 'M9630'. W3, glume primordial present; W3.5, floret primordial present; W5, carpel extending; gl, glume; le, lemma. Scale bar = 1 cm.

Extended Data Figure 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. **(b)** Pie-chart providing ratios of conserved- and variable- genes for all *Hordeum* pangenome species. Bar chart illustrating the number of genes shared between different combinations of species within the *Hordeum* pangenome. Genes shared by all 25 species 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 species.

Extended Data Figure 2: Landscape of NLR genes in the genus *Hordeum*. **(a)** NLR gene copy numbers across 29 *Hordeum* haplomes. The central line in each boxplot represents the median; the box spans the interquartile range (IQR), and whiskers extend to $1.5 \times \text{IQR}$. 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 ($p \text{ value} = 5.06 \times 10^{-7}$). **(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 25 accessions. **(d)** Distribution of NLR gene clusters (≥ 3 genes) across 29 haplomes. Dot size represents the number of genes per cluster. The *chr4H/chr5H* group showed significantly different values compared to other chromosomes (Wilcoxon test, $p \text{ value} < 2.2\text{e-}16$).

Extended Data Fig. 3: “Stretch and twist” model of proximal region evolution. **(a)** Normalized coverage density plot in *H. euclaston*. The right shows species ordered by divergence time; two distinct peaks appear among species that diverged over 9 million years ago. **(b)** 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. **(c)** Proportion of repeat types in proximal and distal regions. **(d)** Boxplot showing the difference in repeat composition between distal and proximal regions. **(e)** A schematic diagram of the “stretch-and-twist” model.

Extended Data Fig. 4: Detection of deep hybridization events and segmental exchanges. **(a)** Genome-wide distribution of distal inversions (≥ 2 Mb) across *Hordeum* haplomes. **(b)** Schematic overview of the hybridization scanning method. **(c)** Density distribution of normalized Robinson–

Foulds (RF) distances; red line (0.18) separates two peaks. **(d)** Phylogenetic tree constructed from single-copy genes located in a segmental exchange region on chromosome 5H; bootstrap values shown at each node. The green branches show where the changes occurred compared to the reference tree. 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 = 6.11×10^{-12}). **(e)** Analysis of the segmental exchange region on chromosome 4H, performed using the same approach. 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 = 1.11×10^{-37}).

Extended Data Fig. 5: 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 & Blattner² 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. 6: Transposon dynamics underlie genome expansion in *Hordeum*. (a)

Pearson correlation coefficients between TE superfamilies and genome size. **(b)** Heatmap showing abundance of 55 TREP TEs (>10 Mb average cumulative length). The numbers in each cell indicate the cumulative size in Mb of all elements assigned to a family. **(c)** PCA of 334,142 full-length *RLC_BARE1* elements; each dot represents one element, colored by insertion time. **(d)** PCA of 254,001 full-length *RLG_Sabrina* elements. **(e)** Insertion time distributions of full-length LTR elements in *H. secalinum* and *H. gussoneanum*; the latter is the diploid progenitor of the former. **(f)** Insertion time distributions of LTR elements in *H. murinum* BCC2009 (subgenome 1) and its diploid progenitor BCC2017. **(g)** Heatmap showing TE families expanded in *H. secalinum* subgenome 1 relative to *H. gussoneanum*. **(h)** PCA of 1,041 full-length *RLG_Sabrina* elements; colors represent genomes. **(i)** PCA colored by estimated insertion time.

Extended Data Fig. 7: 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. 8: 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. 7**). (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. 9: 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)⁵⁰. (c) Size distribution of sRNAs mapped to *H. cordobense* *ZCCT-2* and *-3* genes at 14 and 28 DAG.

Extended Data Fig. 10: Spike phenotype and test for allelism between M9630 and *eog1* *H. vulgare* mutants. (a) Glume elongation phenotype on a spike. (b) Glume phenotype in F₁ plants through crossing M9630 with other *eog1* mutants. Scale bar = 5 mm.

List of supplementary items

Supplementary Figure 1: *H. bogdanii* H240 Hi-C contact matrix.
 Supplementary Figure 2: *H. brevisubulatum* PI401390 Hi-C contact matrix.
 Supplementary Figure 3: *H. californicum* BCC2058 Hi-C contact matrix.
 Supplementary Figure 4: *H. chilense* GRA1000 Hi-C contact matrix.
 Supplementary Figure 5: *H. comosum* NGB18417 Hi-C contact matrix.
 Supplementary Figure 6: *H. cordobense* BCC2067 Hi-C contact matrix.
 Supplementary Figure 7: *H. erectifolium* NGB6816 Hi-C contact matrix.
 Supplementary Figure 8: *H. euclaston* NGB90233 Hi-C contact matrix.
 Supplementary Figure 9: *H. flexuosum* BCC2023 Hi-C contact matrix.
 Supplementary Figure 10: *H. gussoneanum* BCC2005 Hi-C contact matrix.
 Supplementary Figure 11: *H. intercedens* NGB15676 Hi-C contact matrix.
 Supplementary Figure 12: *H. jubatum* BCC2055 Hi-C contact matrix.
 Supplementary Figure 13: *H. marinum* BCC2001 Hi-C contact matrix.
 Supplementary Figure 14: *H. murinum* BCC2009 Hi-C contact matrix.
 Supplementary Figure 15: *H. murinum* BCC2017 Hi-C contact matrix.
 Supplementary Figure 16: *H. muticum* NGB90062 Hi-C contact matrix.
 Supplementary Figure 17: *H. patagonicum* BCC2065 Hi-C contact matrix.
 Supplementary Figure 18: *H. pubiflorum* BCC2028 Hi-C contact matrix.
 Supplementary Figure 19: *H. pusillum* NGB6888 Hi-C contact matrix.
 Supplementary Figure 20: *H. roshevitzii* BCC2069 Hi-C contact matrix.
 Supplementary Figure 21: *H. secalinum* BCC2004 Hi-C contact matrix.
 Supplementary Figure 22: *H. stenostachys* NGB90152 Hi-C contact matrix.

683 Supplementary Figure 23: *H. marinum* H559 Hi-C contact matrix.

684 Supplementary Figure 24: GO-term enrichment for core, shell and cloud genes visualised by
685 bubble plots.

686 Supplementary Figure 25. Pangenome analyses of the PPR and mTERF gene families in *Hordeum*.

687 Supplementary Figure 26: Percentages of graph assigned to core (present in all accessions), shell
688 (present in two or more but not all accessions) and cloud (present in a single accession only).

689 Supplementary Figure 27: Odgi heaps pan-genome saturation plot.

690 Supplementary Figure 28: Variant density as a function of position on the chromosome.

691 Supplementary Figure 29: Conservation analysis with Pansel. Jaccard index values were
692 computed directly from the graph using Pansel on default settings, using a window size of 1
693 kilobasepairs.

694 Supplementary Figure 30: Phylogenetic tree constructed by concatenated genome-wide single-
695 copy genes.

696 Supplementary Figure 31: Coalescence based species tree constructed by single-copy genes.

697 Supplementary Figure 32. Phylogenetic tree constructed using CASTER-pair based on whole-
698 genome alignments generated by Progressive Cactus.

699 Supplementary Figure 33. Phylogenetic tree constructed using CASTER-site based on whole-
700 genome alignments generated by Progressive Cactus.

701 Supplementary Figure 34: Dotplot comparisons of chromosome 1H across genomes diverged at
702 different times.

703 Supplementary Figure 35: Dotplot comparisons of chromosome 2H across genomes diverged at
704 different times.

705 Supplementary Figure 36: Dotplot comparisons of chromosome 3H across genomes diverged at
706 different times.

707 Supplementary Figure 37: Dotplot comparisons of chromosome 5H across genomes diverged at
708 different times.

709 Supplementary Figure 38: Dotplot comparisons of chromosome 6H across genomes diverged at
710 different times.

711 Supplementary Figure 39: Dotplot comparisons of chromosome 7H across genomes diverged at
712 different times.

713 Supplementary Figure 40: Normalized coverage density plot in *H. vulgare*, *H. murinum* BBC2017
714 and *H. gussoneanum*.

715 Supplementary Figure 41: Distribution plot of genome-wide normalized coverage in *H. euclaston*.

716 Supplementary Figure 42: Distribution plot of genome-wide normalized coverage in *H. vulgare*.

717 Supplementary Figure 43: Distribution plot of genome-wide normalized coverage in *H. murinum*
718 BBC2017.

719 Supplementary Figure 44: Distribution plot of genome-wide normalized coverage in *H.*
720 *gussoneanum*.

721 Supplementary Figure 45: Boxplot of TE family abundance differences between proximal and
722 distal regions.

723 Supplementary Figure 46: Comparative distribution of matched bases between *H. jubatum*
724 subgenomes and their diploid ancestors.

725 Supplementary Figure 47: Dotplot comparisons between *T. monococcum* and *H. jubatum*
726 BCC2055 Subgenome1.

Supplementary Figure 48: Dotplot comparisons between *T. monococcum* and *H. jubatum* BCC2055 Subgenome2.

Supplementary Table 1: Summary statistics of genome assemblies.

Supplementary Table 2: Summary statistics and accession codes of HiFi and Hi-C data.

Supplementary Table 3: Summary of genome assembly evaluation.

Supplementary Table 4: Summary statistics and accession codes of RNA-Seq and Iso-Seq data.

Supplementary Table 5: Summary of gene annotation.

Supplementary Table 6: Summary of the transposable element annotation.

Supplementary Table 7: Number of identified PPR, RFL-PPR, mTERF and RFL-mTERF genes in the *Hordeum* pangenome.

Supplementary Table 8: Summary of *Hordeum* species divergence time.

Supplementary Table 9: Overview of proximal region boundaries in *Hordeum*.

Supplementary Table 10: Genome sources – list of genomes used in phylogenetic validation of HGT candidates.

Supplementary Table 11: HGT genes – HGT genes, their respective orthogroups, presumed donor based on phylogenetic analysis, position and annotation.

Supplementary Table 12: Expression – binary represented (expressed=1/not expressed=0) expression of orthogroups in studied species.

Supplementary Table 13: Transposable element composition of tetraploids and their ancestral genomes.

Supplementary Table 14: Summary statistics and accession codes of ATAC-seq.

Supplementary Table 15: *VRN1* and *VRN2* genes in the wild *Hordeum* species.

Supplementary Table 16: Spike-specific expression genes in five *Hordeum* species.

Supplementary Table 17: Spike-specific expression genes that are commonly detected across five *Hordeum* species.

Supplementary Table 18: Accession codes for genome sequences.

Supplementary Table 19: Accession codes for transcriptome sequences.

Supplementary Table 20: Randomly chosen species and divergence times from all nodes for alignment illustration.

References

- Bothmer, R. v. & International Board for Plant Genetic Resources. *An Ecogeographical study of the genus Hordeum*. (IBPGR, 1991).
- Brassac, J. & Blattner, F. R. Species-Level Phylogeny and Polyploid Relationships in *Hordeum* (Poaceae) Inferred by Next-Generation Sequencing and In Silico Cloning of Multiple Nuclear Loci. *Syst Biol* **64**, 792-808 (2015). <https://doi.org/10.1093/sysbio/syv035>
- Blattner, F. R., Pleines, T. & Jakob, S. S. Rapid Radiation in the Barley Genus *Hordeum* (Poaceae) During the Pleistocene in the Americas. *Evolution in Action: Case Studies in Adaptive Radiation, Speciation and the Origin of Biodiversity*, 17-33 (2010). https://doi.org/10.1007/978-3-642-12425-9_2

770 4 Blattner, F. R. Progress in phylogenetic analysis and a new infrageneric
771 classification of the barley genus (Poaceae: Triticeae). *Breeding Science* **59**, 471-
772 480 (2009). <https://doi.org/DOI> 10.1270/jsbbs.59.471

773 5 Jiang, C. *et al.* Barley2035: A decadal vision for barley research and breeding. *Mol*
774 *Plant* **18**, 195-218 (2025). <https://doi.org:10.1016/j.molp.2024.12.009>

775 6 Dawson, I. K. *et al.* Barley: a translational model for adaptation to climate change.
776 *New Phytol* **206**, 913-931 (2015). <https://doi.org:10.1111/nph.13266>

777 7 Dawson, I. K. *et al.* Barley: a translational model for adaptation to climate change.
778 *New Phytol* **206**, 913-931 (2015). <https://doi.org:10.1111/nph.13266>

779 8 Hoseinzadeh, P., Ruge-Wehling, B., Schweizer, P., Stein, N. & Pidon, H. High
780 Resolution Mapping of a *Hordeum bulbosum*-Derived Powdery Mildew Resistance
781 Locus in Barley Using Distinct Homologous Introgression Lines. *Front Plant Sci* **11**,
782 225 (2020). <https://doi.org:10.3389/fpls.2020.00225>

783 9 Wendler, N. *et al.* Bulbosum to Go: A Toolbox to Utilize *Hordeum vulgare*/bulbosum
784 Introgressions for Breeding and Beyond. *Mol Plant* **8**, 1507-1519 (2015).
785 <https://doi.org:10.1016/j.molp.2015.05.004>

786 10 Haas, M. & Mascher, M. Use of the secondary gene pool of barley in breeding
787 improved varieties. *Burl Dodds Agr Sci* **60**, 55-74 (2019).
788 <https://doi.org:10.19103/As.2019.0051.02>

789 11 Feng, J. W. *et al.* A haplotype-resolved pangenome of the barley wild relative. *Nature*
790 (2025). <https://doi.org:10.1038/s41586-025-09270-x>

791 12 Iannone, L. J., Irisarri, J. G. N., Mc Cargo, P. D., Perez, L. I. & Gundel, P. E.
792 Occurrence of fungal endophytes in the sheep-preferred grass from Patagonia. *J*
793 *Arid Environ* **115**, 19-26 (2015). <https://doi.org:10.1016/j.jaridenv.2014.12.008>

794 13 Potthoff, M. *et al.* Soil biological and chemical properties in restored perennial
795 grassland in California. *Restor Ecol* **13**, 61-73 (2005). <https://doi.org:DOI>
796 10.1111/j.1526-100X.2005.00008.x

797 14 Janousek, C. N. & Folger, C. L. Inter-specific variation in salinity effects on
798 germination in Pacific Northwest tidal wetland plants. *Aquat Bot* **111**, 104-111
799 (2013). <https://doi.org:10.1016/j.aquabot.2013.06.009>

800 15 Yuan, J., Li, H. & Yang, Y. The Compensatory Tillering in the Forage Grass *Hordeum*
801 *brevisubulatum* After Simulated Grazing of Different Severity. *Front Plant Sci* **11**, 792
802 (2020). <https://doi.org:10.3389/fpls.2020.00792>

803 16 Israelsen, K. R., Ransom, C. V. & Waldron, B. L. Salinity Tolerance of Foxtail Barley
804 (*Hordeum jubatum*) and Desirable Pasture Grasses. *Weed Sci* **59**, 500-505 (2011).
805 <https://doi.org:10.1614/Ws-D-10-00164.1>

806 17 Jin, Y. L., Du, X., Jiang, C. C., Ji, W. Q. & Yang, P. Disentangling sources of gene tree
807 discordance for species via target-enriched sequencing assays. *Molecular*
808 *Phylogenetics and Evolution* **199** (2024).
809 <https://doi.org:10.1016/j.ympev.2024.108160>

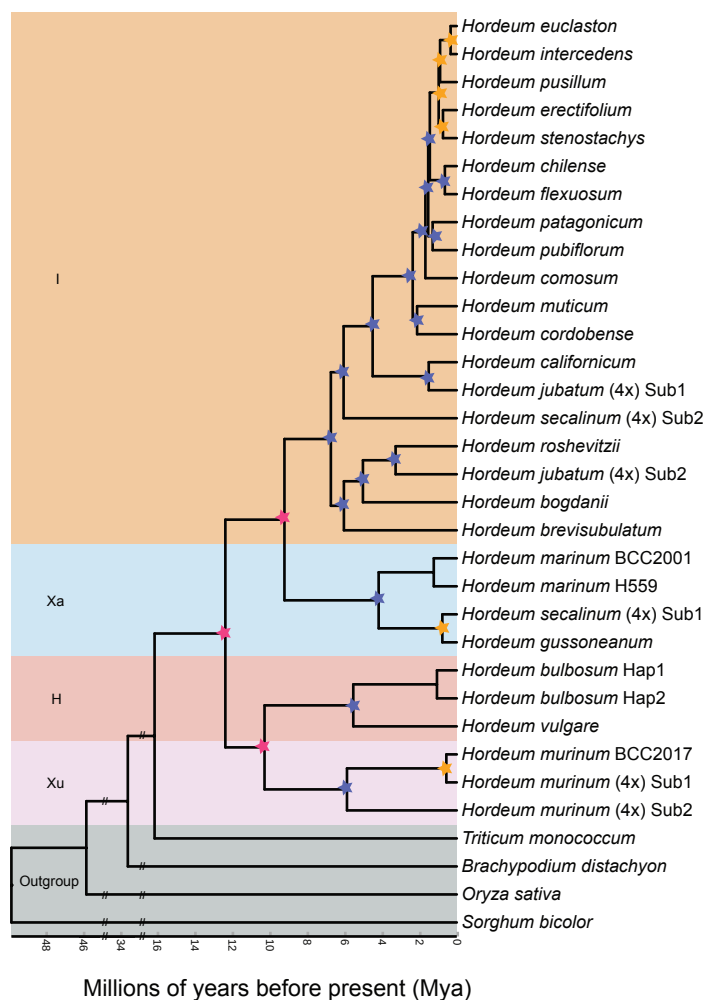
810 18 Jakob, S. S. & Blattner, F. R. A chloroplast genealogy of (Poaceae): Long-term
811 persisting haplotypes, incomplete lineage sorting, regional extinction, and the
812 consequences for phylogenetic inference. *Molecular Biology and Evolution* **23**,
813 1602-1612 (2006). <https://doi.org:10.1093/molbev/msl018>

- 19 Tang, D. *et al.* Genome evolution and diversity of wild and cultivated potatoes. *Nature* **606**, 535-541 (2022). <https://doi.org:10.1038/s41586-022-04822-x>
- 20 Benoit, M. *et al.* pan-genetics reveals paralogues as contingencies in crop engineering. *Nature* **640** (2025). <https://doi.org:10.1038/s41586-025-08619-6>
- 21 Chen, L. *et al.* Genome sequencing reveals evidence of adaptive variation in the genus *Zea*. *Nat Genet* **54**, 1736-1745 (2022). <https://doi.org:10.1038/s41588-022-01184-y>
- 22 Wang, M. *et al.* Genomic innovation and regulatory rewiring during evolution of the cotton genus *Gossypium*. *Nat Genet* **54**, 1959-1971 (2022). <https://doi.org:10.1038/s41588-022-01237-2>
- 23 Guo, L. *et al.* Super pangenome of *Vitis* empowers identification of downy mildew resistance genes for grapevine improvement. *Nat Genet* **57**, 741-753 (2025). <https://doi.org:10.1038/s41588-025-02111-7>
- 24 Zhou, Y. *et al.* Graph pangenome captures missing heritability and empowers tomato breeding. *Nature* **606**, 527-534 (2022). <https://doi.org:10.1038/s41586-022-04808-9>
- 25 Marone, M. P., Singh, H. C., Pozniak, C. J. & Mascher, M. A technical guide to TRITEX, a computational pipeline for chromosome-scale sequence assembly of plant genomes. *Plant Methods* **18**, 128 (2022). <https://doi.org:10.1186/s13007-022-00964-1>
- 26 Monat, C. *et al.* TRITEX: chromosome-scale sequence assembly of Triticeae genomes with open-source tools. *Genome Biol* **20**, 284 (2019). <https://doi.org:10.1186/s13059-019-1899-5>
- 27 Mascher, M. *et al.* Long-read sequence assembly: a technical evaluation in barley. *Plant Cell* **33**, 1888-1906 (2021). <https://doi.org:10.1093/plcell/koab077>
- 28 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 Communications* **3** (2022). <https://doi.org:10.1016/j.xplc.2022.100333>
- 29 Manni, M., Berkeley, M. R., Seppey, M., Simao, F. A. & Zdobnov, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol Biol Evol* **38**, 4647-4654 (2021). <https://doi.org:10.1093/molbev/msab199>
- 30 Sommer, M. J., Zimin, A. V. & Salzberg, S. L. PSAURON: a tool for assessing protein annotation across a broad range of species. *NAR Genom Bioinform* **7**, lqae189 (2025). <https://doi.org:10.1093/nargab/lqae189>
- 31 Ou, S. J. *et al.* Differences in activity and stability drive transposable element variation in tropical and temperate maize. *Genome Research* **34**, 1140-1153 (2024). <https://doi.org:10.1101/gr.278131.123>
- 32 Jones, J. D. G., Staskawicz, B. J. & Dangl, J. L. The plant immune system: From discovery to deployment. *Cell* **187**, 2095-2116 (2024). <https://doi.org:10.1016/j.cell.2024.03.045>
- 33 van Wersch, S., Tian, L., Hoy, R. & Li, X. Plant NLRs: The Whistleblowers of Plant Immunity. *Plant Commun* **1**, 100016 (2020). <https://doi.org:10.1016/j.xplc.2019.100016>

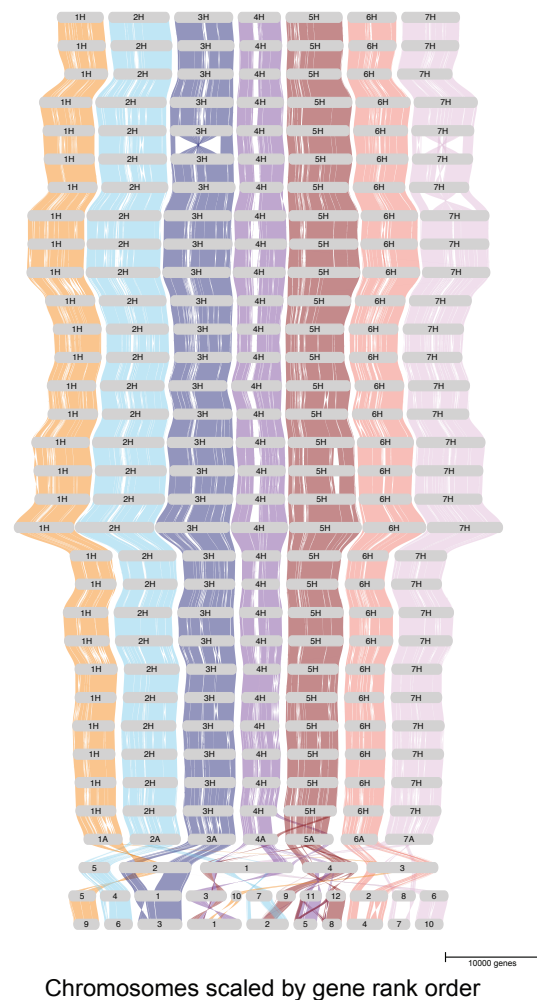
- 34 Melonek, J. & Small, I. Triticeae genome sequences reveal huge expansions of gene families implicated in fertility restoration. *Curr Opin Plant Biol* **66**, 102166 (2022). <https://doi.org/10.1016/j.pbi.2021.102166>
- 35 Bernhard, T., Koch, M., Snowdon, R. J., Friedt, W. & Wittkop, B. Undesired fertility restoration in msml barley associates with two mTERF genes. *Theor Appl Genet* **132**, 1335-1350 (2019). <https://doi.org/10.1007/s00122-019-03281-9>
- 36 Armstrong, J. et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246-+ (2020). <https://doi.org/10.1038/s41586-020-2871-y>
- 37 Rabanus-Wallace, M. T. et al. Chromosome-scale genome assembly provides insights into rye biology, evolution and agronomic potential. *Nature Genetics* **53**, 564-+ (2021). <https://doi.org/10.1038/s41588-021-00807-0>
- 38 Wickell, D. A. & Li, F. W. On the evolutionary significance of horizontal gene transfers in plants. *New Phytologist* **225**, 113-117 (2020). <https://doi.org/10.1111/nph.16022>
- 39 Chen, R. et al. Adaptive innovation of green plants by horizontal gene transfer. *Biotechnol Adv* **46**, 107671 (2021). <https://doi.org/10.1016/j.biotechadv.2020.107671>
- 40 Mahelka, V. et al. Multiple horizontal transfers of nuclear ribosomal genes between phylogenetically distinct grass lineages. *P Natl Acad Sci USA* **114**, 1726-1731 (2017). <https://doi.org/10.1073/pnas.1613375114>
- 41 Krak, K., Caklová, P., Kopecky, D., Blattner, F. R. & Mahelka, V. Horizontally Acquired nrDNAs Persist in Low Amounts in Host Genomes and Evolve Independently of Native nrDNA. *Frontiers in Plant Science* **12** (2021). <https://doi.org/10.3389/fpls.2021.672879>
- 42 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 Journal* **105**, 1141-1164 (2021). <https://doi.org/10.1111/tpj.15167>
- 43 Hibdige, S. G. S., Raimondeau, P., Christin, P. A. & Dunning, L. T. Widespread lateral gene transfer among grasses. *New Phytologist* **230**, 2474-2486 (2021). <https://doi.org/10.1111/nph.17328>
- 44 Dunning, L. T. et al. Lateral transfers of large DNA fragments spread functional genes among grasses. *P Natl Acad Sci USA* **116**, 4416-4425 (2019). <https://doi.org/10.1073/pnas.1810031116>
- 45 Martin, W. F. Too Much Eukaryote LGT. *Bioessays* **39** (2017). <https://doi.org/10.1002/bies.201700115>
- 46 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 (2004). <https://doi.org/10.1093/molbev/msh092>
- 47 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 Physiology* **140**, 1397-1405 (2006). <https://doi.org/10.1104/pp.105.073486>

- 48 Deng, W. W. *et al.* Direct links between the vernalization response and other key traits of cereal crops. *Nature Communications* **6** (2015).
<https://doi.org:10.1038/ncomms6882>
- 49 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 (2005).
<https://doi.org:10.1007/s00438-004-1095-4>
- 50 von Zitzewitz, J. *et al.* Molecular and structural characterization of barley vernalization genes. *Plant Molecular Biology* **59**, 449-467 (2005).
<https://doi.org:10.1007/s11103-005-0351-2>
- 51 Dubcovsky, J., Chen, C. L. & Yan, L. L. Molecular characterization of the allelic variation at the VRN-H2 vernalization locus in barley. *Molecular Breeding* **15**, 395-407 (2005). <https://doi.org:10.1007/s11032-005-0084-6>
- 52 Hung, Y. H. & Slotkin, R. K. The initiation of RNA interference (RNAi) in plants. *Current Opinion in Plant Biology* **61** (2021).
<https://doi.org:10.1016/j.pbi.2021.102014>
- 53 Ciaffi, M., Paolacci, A. R., Tanzarella, O. A. & Porceddu, E. Molecular aspects of flower development in grasses. *Sex Plant Reprod* **24**, 247-282 (2011).
<https://doi.org:10.1007/s00497-011-0175-y>
- 54 Xie, P., Wu, Y. R. & Xie, Q. Evolution of cereal floral architecture and threshability. *Trends in Plant Science* **28**, 1438-1450 (2023).
<https://doi.org:10.1016/j.tplants.2023.08.003>
- 55 WADDINGTON, S. R., CARTWRIGHT, P. M. & WALL, P. C. A Quantitative Scale of Spike Initial and Pistil Development in Barley and Wheat. *Annals of Botany* **51**, 119-130 (1983). <https://doi.org:10.1093/oxfordjournals.aob.a086434>
- 56 Zhuang, H. *et al.* Encodes a C2H2 Zinc Finger Protein That Regulates Spikelet Development in Rice. *Plant Cell* **32**, 392-413 (2020).
<https://doi.org:10.1105/tpc.19.00682>
- 57 Moore, G., Devos, K. M., Wang, Z. & Gale, M. D. Cereal Genome Evolution: Grasses, line up and form a circle. *Current Biology* **5**, 737-739 (1995).
[https://doi.org:https://doi.org/10.1016/S0960-9822\(95\)00148-5](https://doi.org:https://doi.org/10.1016/S0960-9822(95)00148-5)
- 58 Percival-Alwyn, L. *et al.* UKCropDiversity-HPC: A collaborative high-performance computing resource approach for sustainable agriculture and biodiversity conservation. *PLANTS, PEOPLE, PLANET* (2024).
<https://doi.org:10.1002/ppp3.10607>
- 59 Hansson, M. *et al.* A guide to barley mutants. *Hereditas* **161**, 11 (2024).
<https://doi.org:10.1186/s41065-023-00304-w>

a



b



c

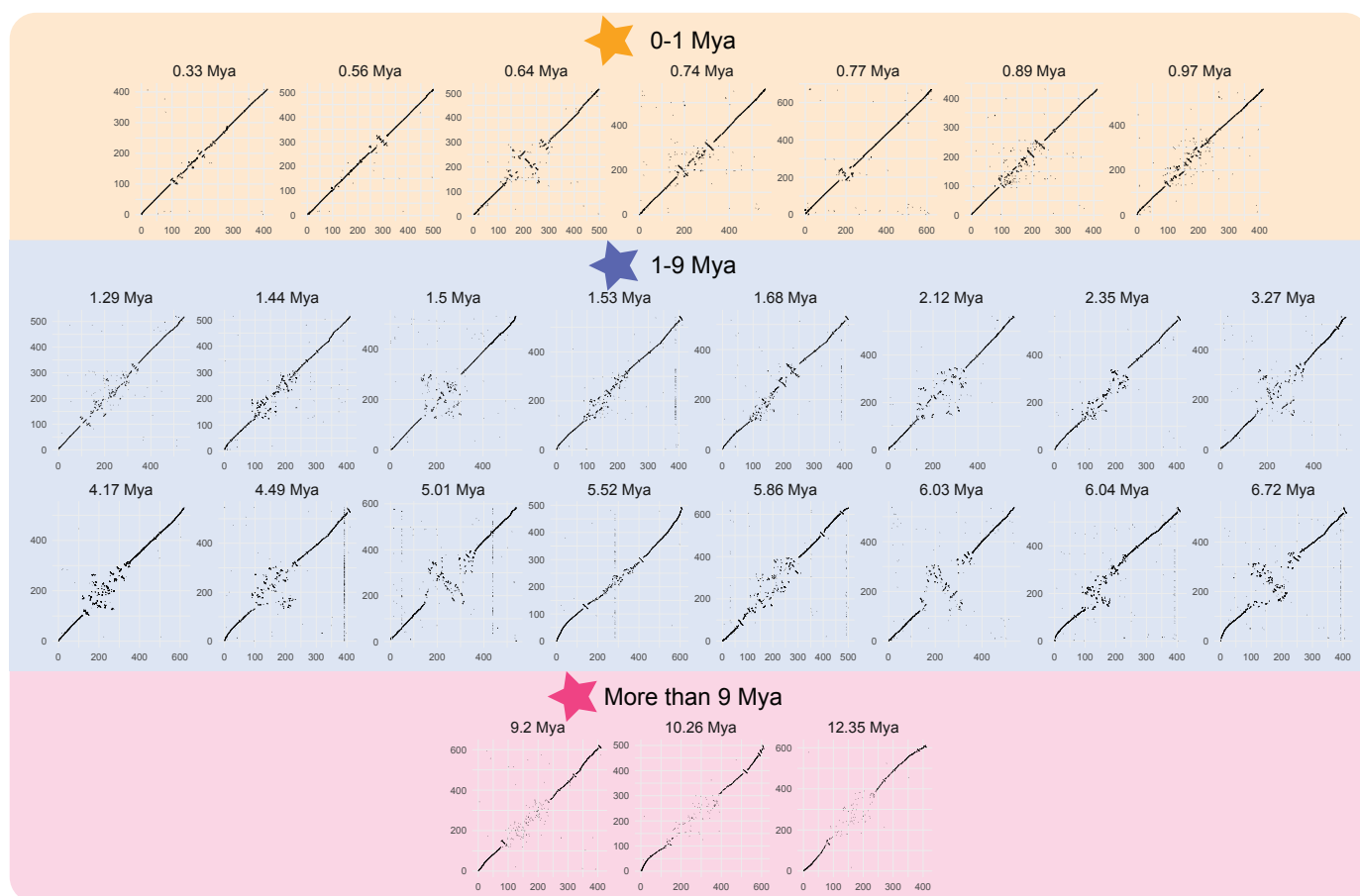


Figure 1: Evolution of *Hordeum* species and rapid change in proximal regions. (a) Phylogenetic tree of *Hordeum* species. *Sorghum bicolor*, *Oryza sativa*, *Brachypodium distachyon*, and *Triticum monococcum* were used as outgroups. Divergence times were inferred using Fourfold Degenerate Third-Codon Transversion (4DTV) sites from genome-wide single-copy genes. Colored rectangles represent different genome groups. Star-marked nodes correspond to those shown in panel (c). (b) GENESPACE synteny plot between 29 *Hordeum* haplotypes and 4 outgroup genomes. Grey bars represent chromosomes, scaled according to the number of genes with syntenic alignments to other genomes. (c) Dotplot comparisons of chromosome 4H across genomes diverged at different times. Yellow indicates nodes diverged within 1 million years; blue indicates 1–9 million years; and magenta indicates nodes older than 9 million years. Two species were randomly selected from each node for whole-genome comparison (Supplementary Table 20).

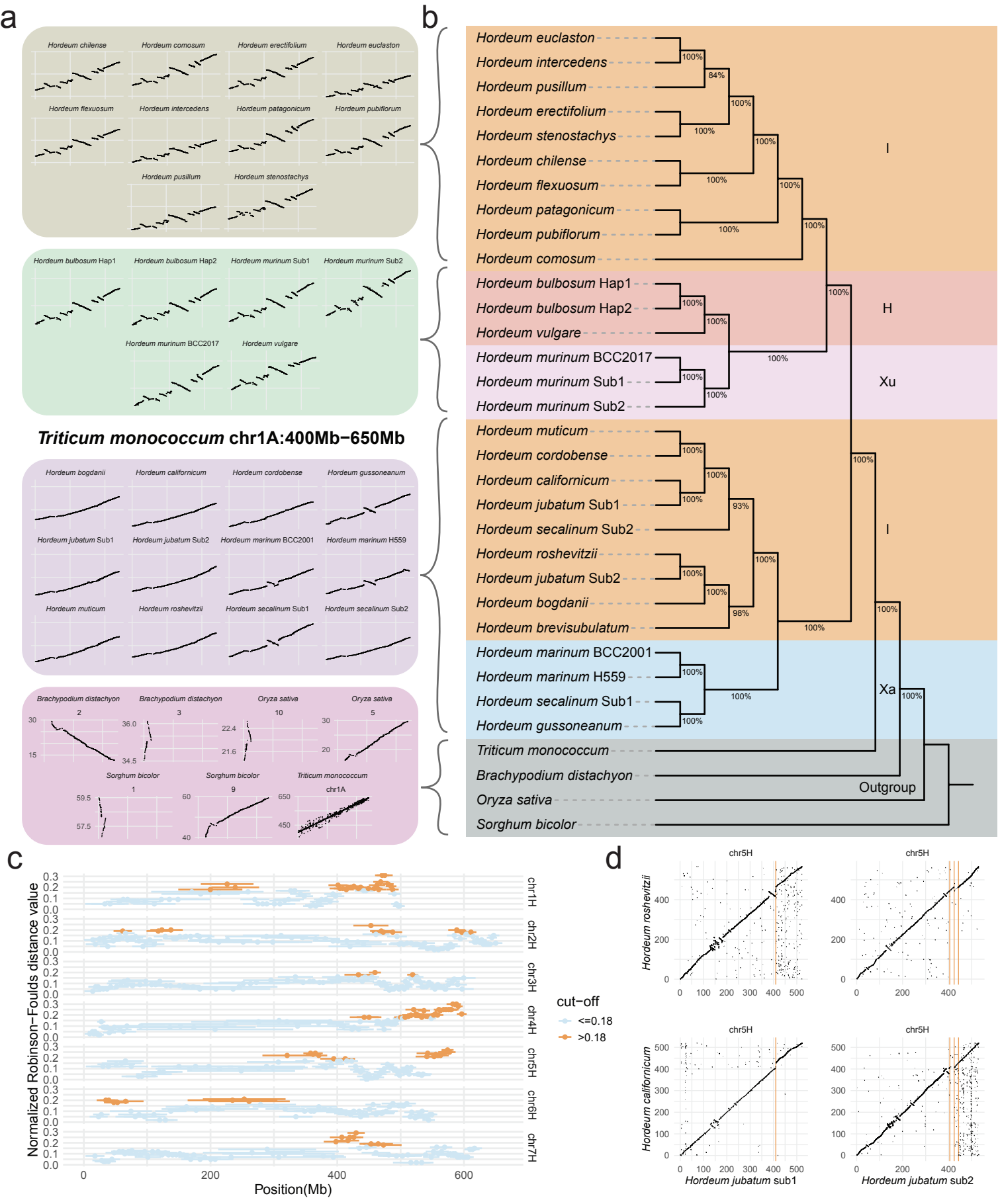


Figure 2: Structural variation reveals past introgression in *Hordeum* species. (a) Dotplot of all genomes aligned to *T. monococcum* in the chr1A: 400–650 Mb region. **(b)** Phylogenetic tree based on single-copy genes within the chr1A: 400–650 Mb region of *T. monococcum*. Colored rectangles represent 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 ($p\text{-AU} = 1.0$) and rejects the reference tree ($p\text{-AU} = 2.03 \times 10^{-5}$). **(c)** Distribution of normalized Robinson–Foulds (nRF) distances between gene trees in genome windows and the species-wide tree. Vertical lines represent genomic regions containing single-copy genes; points mark the center of each window. Orange indicates RF values >0.18, while light blue indicates values ≤0.18. **(d)** Comparative analysis of chromosome 5H between *H. jubatum* and its ancestral diploid genomes. Orange lines indicate regions with duplications and deletions.

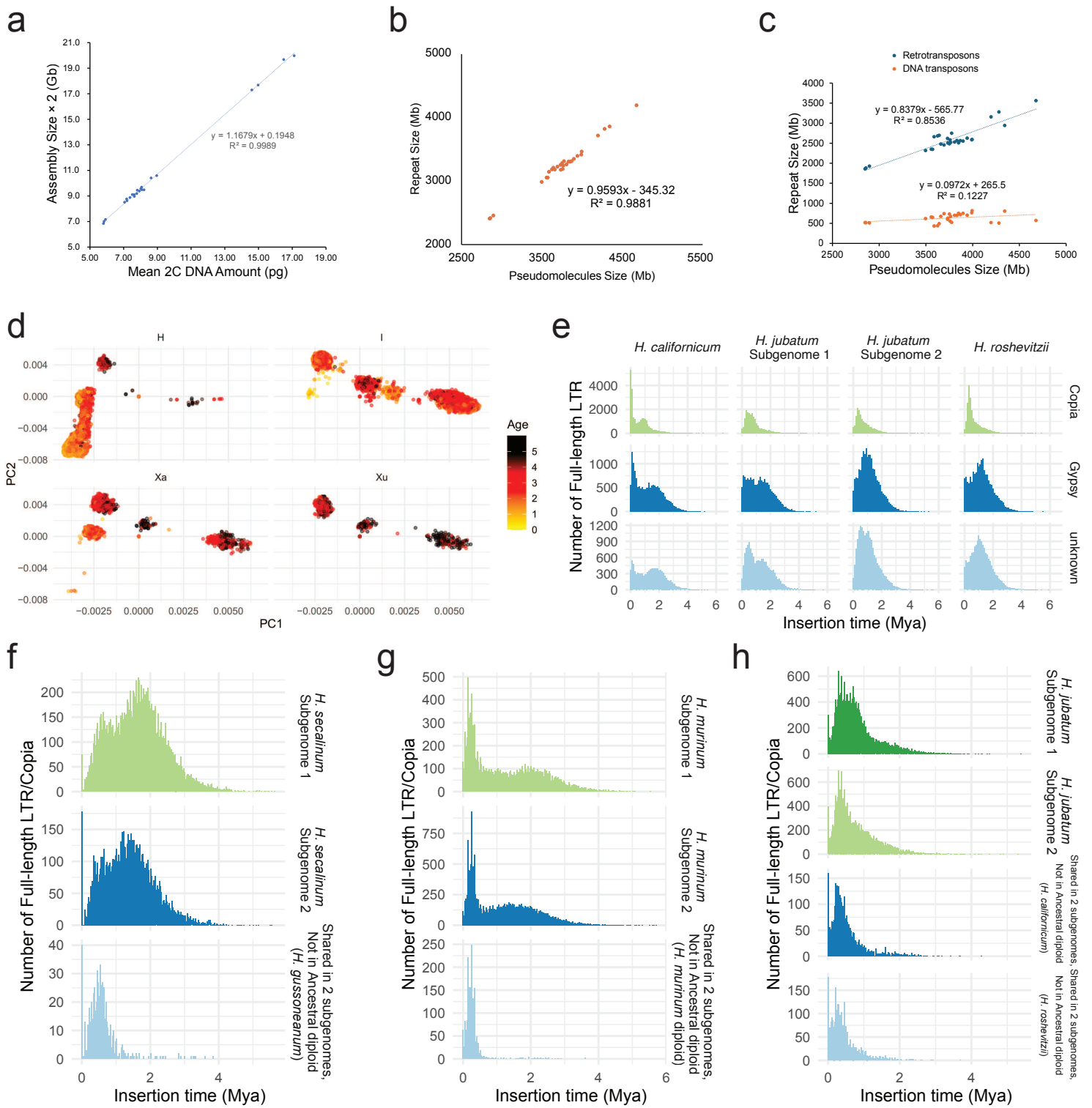


Figure 3: Transposable elements reveal genome size variation and tetraploid formation. (a) Correlation between assembly size and DNA content. (b) Correlation between pseudomolecule size and transposable element (TE) content. (c) Correlation between pseudomolecule size and retrotransposon and DNA transposon content. (d) Principal component analysis (PCA) of 68,707 full-length *RLG_BAGY2* retrotransposon sequences across *Hordeum* species. Each dot represents one retrotransposon; colors indicate estimated insertion times. (e) Distribution of insertion times of full-length LTR elements in *H. jubatum*, *H. californicum*, and *H. roshevitzii*. *H. californicum* and *H. roshevitzii* are the respective diploid progenitors of subgenomes 1 and 2 in *H. jubatum*². (f-h) Insertion time distribution of full-length LTR/Copia elements in tetraploid species. Elements were selected using CD-HIT clustering from one subgenome of the tetraploid, which are absent in the ancestral diploids but present in the other subgenome. The peak indicates the timing of subgenome fusion. (f) *H. secalinum*; *H. gussoneanum* is the diploid progenitor of subgenome 1 in *H. secalinum*². (g) *H. murinum* (BCC2009), *H. murinum* (BCC2017) is the diploid progenitor of subgenome 1 in *H. murinum* (BCC2009)². (h) *H. jubatum*.

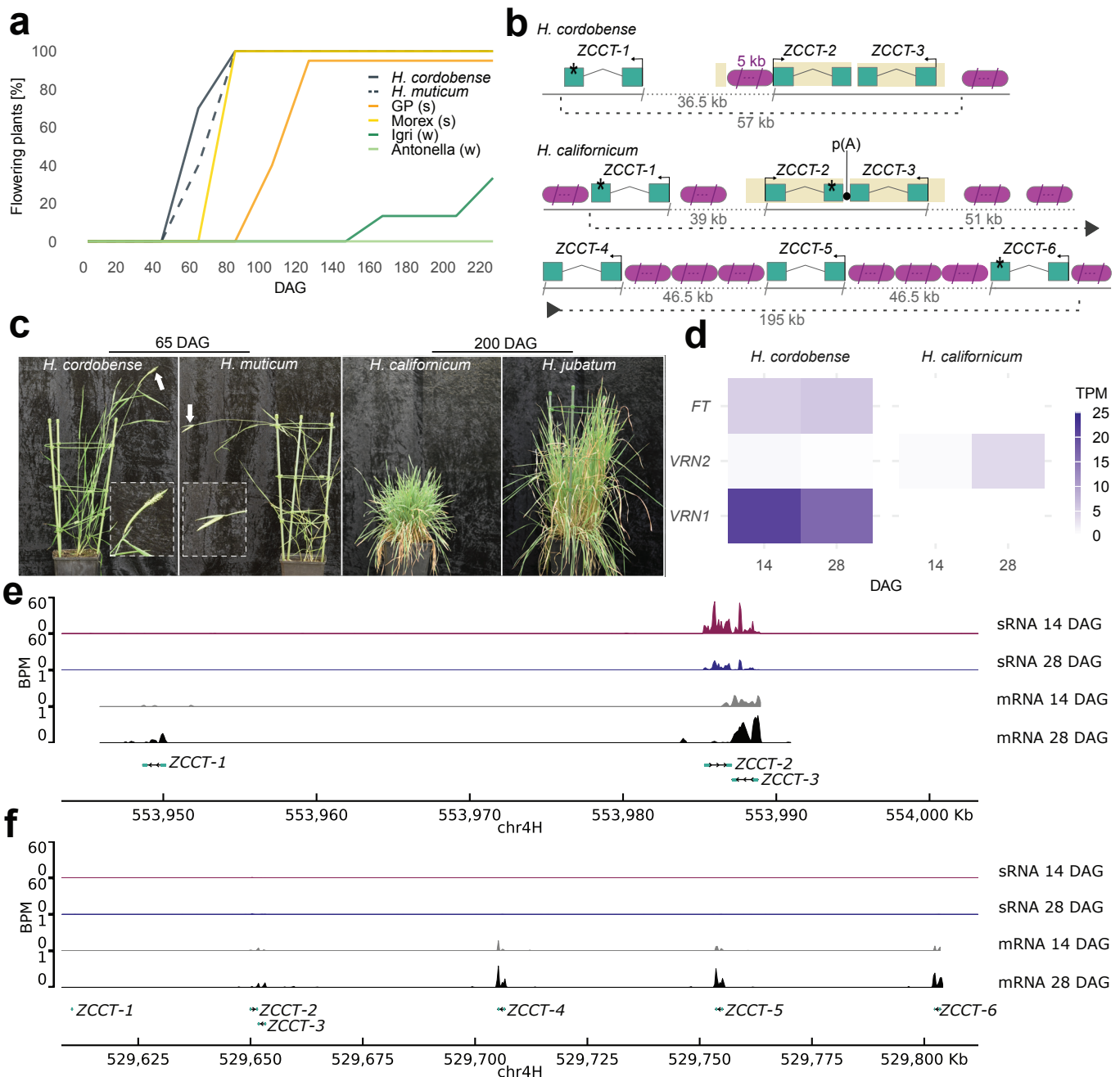


Figure 4: Natural variation in vernalization requirement of *Hordeum* species. (a) Percentage of *H. cordobense* and *H. muticum* plants ($n = 10$ per species) flowering under non-vernalizing conditions. The cultivars Golden Promise (GP) and Morex versus cultivars Igri and Antonella serve as the spring-type (s) and winter-type (w) controls, respectively. DAG – days after germination. (b) 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: exons, *: stop codon, purple ovals: transposable elements, black dot with p(A): polyadenylation signal, sand-colored region: inverted repeat. (c) Phenotypes of four *Hordeum* species carrying *ZCCT* inverted repeat at 65 and 200 DAG. Arrows mark visible spikes. (d) Transcripts per million (TPM) values at 14 and 28 DAG in *H. cordobense* and *H. californicum*. (e-f) Mean distribution of RNA-seq and sRNA-seq reads over three biological replicates normalized to bins per million (BPM) mapped reads in *H. cordobense* (e) and *H. californicum* (f) across the *VRN2* locus at 14 and 28 DAG. Read mapping in *H. cordobense* *ZCCT-3* suggests an improperly spliced, likely aberrant transcripts.

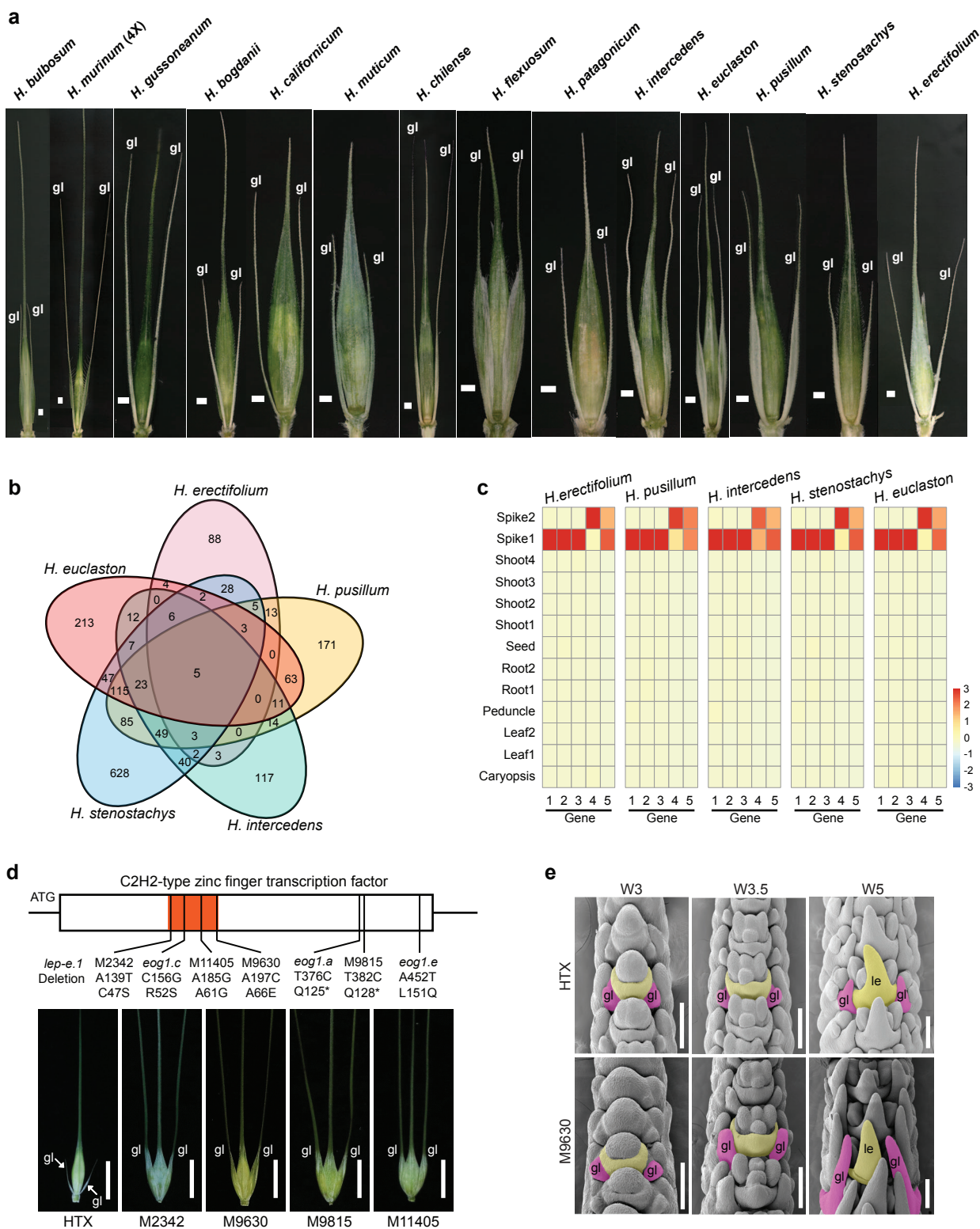


Figure 5: Inhibition of glume development by a C2H2-type zinc finger transcription factor. (a) Central spikelet from representative *Hordeum* species. gl, glume; Scale bar = 500 μ m. **(b)** Venn diagram representing genes with spike-specific expression across five *Hordeum* species. Spike-specific expression was defined by TPM values less than 5% in other tissues within each species. **(c)** Heatmap of five conserved spike-specific genes among the five species. Fold change in transcriptional profiles relative to the mean value are shown. The scale bar represents the magnitude of gene expression fold change. Genes 1 to 5 are annotated with C2H2-type zinc finger transcription factor, auxin response factor, TCP transcription factor, subtilisin protease SBT, and MADS-box transcription factor, respectively. Descriptions of the analyzed tissues (y-axis) and transcriptional profiles of the five genes are given in **Supplementary Table 17**. **(d)** Allelic variations of a C2H2-type zinc finger transcription factor gene (*HORVU.MOREX.r3.2HG0170820.1*) in *H. vulgare* lead to glume elongation. The rectangle denotes the exon/full-length coding sequence, with the red rectangle indicating the C2H2-type zinc finger domain. The asterisk (*) marks a premature stop codon. HTX (Hatixie) represents the wild type with suppressed glume development, while mutants exhibit elongation in glume growth. The mutant allele *lep-e.1* (ref. ⁵⁹) carries a deletion allele of *EOG1* gene. **(e)** Differentiation on the spikelet morphogenesis along Waddington stage between the wild type 'HTX' and the mutant 'M9630'. W3, glume primordial present; W3.5, floret primordial present; W5, carpel extending; gl, glume; le, lemma. Scale bar = 1 cm.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarytables.xlsx](#)
- [onlinemethods.pdf](#)
- [extendeddatafigure.pdf](#)
- [251001Supplementaryfigurescompressed.pdf](#)