

Multi-omics analyses reveal transposon-associated genomic evolution and regulatory rewiring in barley

Wenhao Yue

ZheJiang Academy of Agricultural Sciences

Kangfeng Cai

ZheJiang Academy of Agricultural Sciences

Lei Liu

ZheJiang Academy of Agricultural Sciences

Yong Li

ZheJiang Academy of Agricultural Sciences

Fangying Ge

ZheJiang Academy of Agricultural Sciences

Qiuyu Wang

ZheJiang Academy of Agricultural Sciences

Shenguan Cai

Zhejiang University

Junmei Wang

wangjunmei@zaas.ac.cn

ZheJiang Academy of Agricultural Sciences

Research Article

Keywords: Transposable element, methylation, domestication, three-dimensional (3D) genome architecture

Posted Date: March 18th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9029460/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background:

Transposable elements (TEs) are major components of plant genomes and possess the capacity to transpose across genomes. Nevertheless, their associations with DNA methylation landscapes and three-dimensional (3D) genome organization in barley (*Hordeum vulgare*) remain poorly understood.

Results:

Here, we performed integrated comparative genomic, methylomic, population genomic and Hi-C analyses to investigate the associations between TEs and barley genome evolution and domestication. We found that continuous proliferation of long terminal repeat (LTR) retrotransposons is strongly associated with the barley genome expansion and with distinct genome-wide DNA methylation landscapes across plant species. Population genomic analyses identified candidate domestication-associated genes enriched in sucrose metabolism pathways and revealed an exceptionally slow decay of linkage disequilibrium in barley. Comparative Hi-C and genomic analyses further uncovered extensive differences in A/B chromatin compartments and divergence of topologically associating domains (TADs) between wild and domesticated barley, and indicated that regulatory rewiring, rather than widespread gene gain or loss, represents a prominent feature during barley domestication. Notably, the enrichment of DNA transposons and Copia retrotransposons at TAD boundaries suggested a potential role for TEs in higher-order chromatin organization.

Conclusions:

Together, our study provides valuable genomic resources for future functional investigations in barley and underscores the potential value of incorporating 3D genome features into molecular breeding strategies.

Introduction

The majority of hereditary information in plants is stored within their genomes. Comparative analyses of plant genomes provide fundamental insights into genomic evolution, speciation and the emergence of species-specific genomic features that underpin environmental adaptability and shape plant diversity [1]. Transposable elements (TEs) are ubiquitous and essential components of plant genomes, and can move from one location in a genome to another through a process called transposition [2]. TEs are broadly classified into two major classes: DNA transposons, where DNA sequences mobilize via a cut-and-paste mechanism, and retrotransposons, which mobilize via a copy-and-paste mechanism requiring RNA intermediates and represent the predominant TE class in plant genomes [2, 3]. The retrotransposons are further divided into long terminal repeat (LTR) retrotransposons, which largely contribute to plant genome size variations, and non-LTR retrotransposons [2, 3]. Although TEs generate raw genetic variations and can facilitate phenotypic diversification during plant domestication and adaptation [4, 5],

they can also result in deleterious even lethal mutations. To maintain genomic stability while preserving the ability of rapid responses to environmental changes, plants employ reversibly epigenetic inheritance to silence TE activity [2]. DNA methylation, also known as 5-methylcytosine (5mC), is an epigenetic modification vital for silencing TEs as well as transcriptional regulations and plant growth and development [6, 7]. DNA methylation occurs in three sequence contexts: CG, CHG and CHH (H denotes A, T or C). DNA methylation influences the transcription of genes and TEs mainly in three patterns. Fully silenced genes and TEs are associated with high levels of methylated CG (mCG), mCHG and mCHH. Gene body CG methylation alone is often linked to active transcription, whereas any type of DNA methylation mostly leads to transcription repression [2, 6]. Widespread variations in DNA methylation exist among plant species, and different DNA methylation patterns reflect the evolution histories of plant species [8].

The plant genomes are spatially organized within the nucleus into complex three-dimensional (3D) chromatin architectures, which modulate gene transcriptions by allowing regulatory elements, far from target genes in the one-dimensional sequence, to approach target genes in spatial structures [9–11]. The architectures of plant chromatin are mainly classified in three levels from low resolution in the megabase scale to high resolution in the kilobase scale: chromatin compartment, topologically associating domain (TAD) and chromatin loop [9–11]. A compartment regions are generally gene-rich and transcriptionally active, whereas B compartment regions are enriched in TEs and repressed epigenetic modifications [9–11]. TADs are relatively local units and the interactions within TADs are significantly stronger than those between different TADs [9–11]. The chromatin loops establish physical contacts between distant genomic regions, such as regulatory elements and promoters, to modulate gene expressions [9–11]. Different chromatin structures contribute to the cell type-specific gene expression patterns [11], and chromatin landscape exhibits domestication-associated remodeling [12, 13] and adaptation to environmental changes [14, 15].

Barley (*Hordeum vulgare*), one of the earliest domesticated crops, possesses a complex evolutionary and domestication history [16, 17], and plays a vital role in the global feed and malting industries [18]. The self-pollinating nature, diploid genome and rich germplasm resources of barley established it as a model species for basic research in the 20th century [18, 19]. Nevertheless, the large genome of barley (approximately 4.5Gb), characterized by more than 80% repeat elements and low recombination [20], posed significant challenges for the whole-genome sequencing and assembly, resulting in the first report of chromosome-scale barley reference genome in 2017 [20] considerably later than model species *Arabidopsis* and rice. The rapid innovations of high-throughput sequencing technologies and decreasing costs advanced the qualities of barley reference genomes in recent years [21–23] and produced many omics data of barley involved in whole-genome resequencing [21, 23], methylation [24] and high-throughput chromosome conformation capture (Hi-C) [21, 25], making it possible to perform multi-omics analyses of barley. Despite extensive omics resources available for barley, a comprehensive understanding of how transposable elements are associated with genome evolution, epigenetic landscapes and 3D chromatin organization during domestication remains lacking. In this study, we integrated the comparative genomic and methylomic analyses across barley and four representative

Poaceae (grass) species, with population genomics and 3D chromatin architecture dynamics of wild and domesticated barley. Our study suggested that TE-associated regulatory rewiring, rather than widespread gene gain or loss, represents a prominent feature during barley domestication, and provides genomic resources for future functional genomics and molecular breeding.

Materials and methods

Genomic, resequencing and RNA-Seq data resources

The reference genomes of *Hordeum vulgare* [23], *Arabidopsis thaliana* [26] and four taxa in the Poaceae (grass) clade, including *Sorghum bicolor* [27], *Zea mays* [28], *Oryza sativa* [29], *Triticum aestivum* [30] were downloaded from public repositories or specific websites (Table S1). Only chromosome-level assemblies were considered for subsequent analysis (TE annotation, methylation level detection, synteny analysis and so on), while scaffold sequences were excluded.

The VCF file containing SNP variations of 291 diverse barley (*Hordeum vulgare*) varieties, including 93 wild and 198 domesticated accessions, were downloaded from previously published study [23].

The raw RNA-Seq data of different developmental tissues [20] and different tissues of germinating seeds [31] were derived, and mapped to reference genome [23] of barley to calculate gene expressions following the previous described methods [32].

Population genetics analysis and SNP annotation

The geographic distribution of downloaded resequencing barley varieties was shown using R package maps (version 3.4.3) and ggplot2 (version 4.0.0).

The downloaded SNPs with minor allele frequency lower than 5% and missing rate larger than 10% in the population were discarded using VCFtools [33] (version 0.1.17) and then the package beagle (version 5.4) with the default parameters were employed for imputation of missing data. The filtered SNPs were used to calculate the linkage disequilibrium (LD) decay distance within wild and domesticated groups using PopLDdecay, and annotated based on the barley reference genome [23] using ANNOVAR (version 2019-10-24).

The VCFtools (version 0.1.17) was employed with 100kb-10kb sliding windows to calculate the pairwise genetic differentiation index F_{ST} values between wild and domesticated populations, and π and Tajima's D values within wild and domesticated populations. The ratio of diversity ($\pi_{wild}/\pi_{domesticated}$) and difference in Tajima's D between wild and domesticated populations were calculated. For each method, the top 1% genomic regions were determined as selective sweeps and the genes overlapped with selective sweeps identified in at least two methods were identified as candidate domesticated genes.

The filtered SNPs were further filtered by PLINK [34] (version 1.90) with the parameter: `--indep-pairwise 50 10 0.2`, and used to carry out phylogenetic tree and Principal Component analysis (PCA) of resequenced accessions. The neighbor-joining phylogeny tree was constructed using PHYLIP software (version 3.697) and shown with iTOL (<https://itol.embl.de/>). The PCA analysis was performed using PLINK (version 1.90) with the parameter: `--pca` set as 291, and the first and second eigenvectors were plotted.

Enrichment analysis, gene expression illustration and haplotype analysis

The R package clusterProfiler (version 4.8.2) was employed to perform the KEGG and GO enrichment analysis and the results were shown using R package ggplot2 (version 3.5.2). The R package ComplexHeatmap (version 2.10.0) was used to generate the heatmaps of gene expression profiles based on the $\log_2(\text{TPM} + 1)$ transformation.

The SNPs within and in the 1 kb flanking regions of candidate genes were extracted to perform haplotype analysis. PLINK [34] (version 1.90) was employed to calculate the pairwise linkage disequilibrium (R^2) of SNPs and the results were displayed with R package LDheatmap (version 1.0–6).

Evolution analysis

The longest translation form of each gene from five taxa (barley, sorghum, rice, maize, wheat) in the grass clade and the outgroup Arabidopsis, downloaded from public repositories or specific websites, were used to perform comparative genomics analysis. The gene family clustering was conducted using OrthoFinder [35] (version 2.5.4). The orthogroups, containing three genes in wheat and only one gene in other plant species, were filtered and we randomly discarded two wheat genes in each orthogroups to generate the super genes for construction of maximum-likelihood (ML) species phylogenetic tree using IQ-Tree (version 1.6.12) with the parameters: `-m MFP -bb 1000`. The gene family expansion and contraction along plant phylogenetic tree was estimated using CAFE5, based on the divergence time of the six plant species obtained from TimeTree (<https://timetree.org/>).

The genome-wide syntenic analysis within or among the six plant species were carried out as described previously [32]. The synteny illustrations among different plant species were generated using python version JCVI (<https://github.com/tanghaibao/jcvi>) of MCScan. The syntenic blocks containing > 200 genes among different plant species were shown with yellow colors. The duplicated genes within a plant species were classified into whole genome /segmental, tandem, proximal or dispersed duplications using duplicate_gene_classifier [36] with default parameters.

The reciprocal best hit orthologs between different plant species were determined using WGD [37], and their protein sequences were aligned and then transformed into nucleotide sequences using ParaAT. The

KaKs_Calculator 2.0 as employed to estimate the synonymous nucleotide substitution per synonymous site (Ks) using Yang-Nielsen model [38]. The syntenic gene pairs in the syntenic blocks within plant species were used to calculate Ks values as described above. The R package ggridges (version 0.5.6) was employed to display the density distribution of Ks.

Transposable element (TE) annotation and LTR insertion time analysis

The TE annotation of a plant genome was first conducted using EDTA [39] (version 2.0.1) with parameters: --step all --sensitive 0 --anno 1. Since a proportion of long terminal repeat retrotransposons (LTRs) in the TE library generated in the original EDTA pipeline could not be classified into a specific superfamily, DeepTE was then used to further classify these unknown LTR. Finally, the improved non-redundant TE library was used as input to produce TE annotation in the plant genome with parameters: --step anno --overwrite 1 --anno 1. The LTR insertion time was estimated using full-length LTRs with neutral mutation rate of 1.3×10^{-8} per bp per year.

Phylogenetic tree of full-length LTR of barley and rice

The full-length LTRs of barley and rice were extracted using BEDtools (version 2.31.1) according to gff3 files generated in EDTA pipeline and classified at clade levels using TESorter [40] with rexdb-plant repeat database [41]. The RT domain larger than 200 amino acids of full-length Copia LTR of barley and rice were extracted and aligned using MUSCLE (version 3.8.31). The phylogenetic tree of Copia elements of barley and rice was constructed using IQ-TREE (version 1.6.12) and shown with iTOL (<https://itol.embl.de/>). The same methods were employed for phylogenetic tree of RT domain (larger than 150 amino acids) of full-length Gypsy LTR of barley and rice. The insertion time of full-length LTR was displayed using the R package ggridges (version 0.5.6).

Circos plots

The density distributions of SNPs, F_{ST} , π and Tajima's D in contiguous 100 kb windows and the density distributions of the intact LTR elements, Copia and Gypsy elements in contiguous 5 Mb windows across barley genome were displayed using Circos (version 0.69-8).

WGBS analysis

Given the whole-genome bisulfite sequencing raw data of barley [24] was from wild accession Rossa166, we also downloaded the resequencing data of Rossa166 to construct a pseudo-genome for methylation level analysis as previously described [24], eliminating the interference of nucleotide variations. The whole-genome bisulfite sequencing raw data of barley accession Rossa166 was filtered using fastp (version 0.12.4) with the default parameters. The high-quality clean reads were mapped to the pseudo

reference genome using Bismark (version 0.24.2) with default parameters. The methylation level for each cysteine site was calculated using BatMeth2 [42] with command: BatMeth2 calmeth -Q 20 --remove_dup --coverage 4 -nC 1. The methylation level within gene or TE and flanking regions were conducted using BatMeth2 [42] with command: BatMeth2 methyGff --coverage 4 -nC 1 --distance 2000 --step 0.01. The visualization of methylation level was conducted using ggplot2 (version 3.5.2).

The whole-genome bisulfite sequencing raw data of Arabidopsis [43], sorghum [44], rice [45], maize [46], wheat [47] were from the same varieties, that was used for reference genome assemblies. The methylation levels of these plant species were conducted as described above without construction of pseudo-genomes.

A/B Compartment identification

The Hi-C sequencing data of barley varieties Morex [20], EC_N1 and EC_S1 [25] were downloaded from previous studies. First, the Hi-C sequencing data were filtered using fastp (version 0.12.4) with the default parameters and mapped to the reference genome of Morex [23] using bwa (version 0.7.17) with parameters: bwa mem -A1 -B4 -E50 -L0. The Hi-C contact matrix was generated using hicBuildMatrix [48] (version 3.7.5) with the parameters: --binSize 10000 --minDistance 500 --maxLibraryInsertSize 1500 --restrictionSequence AAGCTT --danglingSequence AGCT. Next, the hicNormalize [48] (version 3.7.5) were employed to normalize contact matrices among three varieties with parameters: --normalize smallest, and then the matrices resolutions were zoomified using Cooler (version 0.10.2) with iterative correction and eigenvector decomposition (ICE) method. Finally, the 70 kb resolutions, at which 80% of loci showed at least 1000 contacts, were selected to perform compartment analysis using R package Calder [49] with gene density as compartment A/B classification standard.

Hi-C contact matrices in chromosome-levels were displayed using plotMatrix from R package HiContacts (version 1.6.0) at 1 Mb resolution.

TAD boundary identification

The Hi-C reads of three barley varieties were processed as described in the compartment A/B identification, except mapped to their own reference genomes. The 70 kb resolution of contact matrices were used to detect the TAD boundaries using hicFindTADs [48] (version 3.7.5) with the default parameters. The conserved and specific TAD boundaries among the three varieties were identified using Tcbf [50]. The Hi-C matrices and TAD domains were displayed using pyGenomeTracks (version 3.9). The TE counts around TAD boundaries were calculated using Deeptools (version 3.5.5) and displayed using ggplot2 (version 4.0.0).

Statistics and Reproducibility

Statistical analyses were performed in R (version 4.4.0). Pearson's correlation and corresponding statistical significance analyses were assessed using the R package rstatix (version 0.7.2). Differences

between groups were evaluated using the Wilcoxon rank-sum test. All statistical tests were two-sided unless otherwise specified. Exact *P* values are provided in the figures and/or figure legends.

Results

Comparative genomic analyses

To investigate the genomic evolution of barley, we obtained the genome assemblies (Table S1) of barley, Arabidopsis and four other grass species-sorghum, maize, rice and wheat-and performed gene family clustering. A total of 32,207 gene families were identified across the six plant species (Table S2). Among them, 21,024 gene families were detected within barley, fewer than wheat (25,554 gene families) but more than Arabidopsis (12,803), sorghum (19,172), maize (18,936) and rice (19,618) (Fig. 1A and Table S3). A total of 667 gene families comprising 10,447 genes were specific to barley. The Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment of these genes was mainly involved in mismatch repair, spliceosome, amino sugar and nucleotide sugar metabolism and circadian rhythm pathways (Fig. 1B and Table S4). Consistently, the GO enrichment of the barley-specific genes was associated with primary cellular process such as RNA transcription, mismatch repair and chromatin organization (Fig. S1 and Table S5). The gene family expansion and contraction were examined across the six plant species, and we identified 34 expanded gene families in barley, substantially fewer than the closely related species wheat (1268 gene family), whereas 651 gene family contraction in barley, markedly more than wheat (19 gene family) (Fig. 1C).

Species divergence was inferred based on the density distributions of synonymous nucleotide substitution (*K_s*) within syntenic blocks between barley and the five other plant species, where the wheat genome was divided into A, B and D subgenomes. These analyses indicated that barley diverged earliest from Arabidopsis, followed by divergence from the common ancestor of sorghum and maize, subsequently from rice and wheat (Fig. 1D and Fig. S2A). Furthermore, polyploidization event analyses based on intra-genomic syntenic blocks revealed three peaks in barley and two peaks in other plant species in the (Fig. 1E and Fig. S2B; Table S6). The ancient *K_s* density peak (*K_s* ~ 4.0) among all species denoted the ancient whole-genome duplication among angiosperms [51, 52] (Fig. 1E), and the shared peak (*K_s* ~ 1.0) corresponded to the grass-specific rho whole-genome duplication [53] (Fig. 1E). To investigate whether the peak (*K_s* ~ 0.2) in barley represent a whole-genome polyploidization event, we carried out the genomic comparisons between barley with sorghum that have not experienced whole-genome duplication in recent 70 million years [54]. The comparison of genome-wide synteny (Fig. 1F) and syntenic depth analysis (Fig. 1G) indicated that no whole-genome duplication (WGD) or whole-genome triplication (WGT) existed after divergence of barley and sorghum, and the *K_s* peak (*K_s* ~ 0.2) represented the previously reported burst of recent gene duplications in Triticeae [55]. We then mapped the barley gene sequences onto the other plant species to infer their inter-genomic synteny and polyploidization events. More syntenic blocks were observed between barley and species that underwent recent WGD (maize, 866 syntenic blocks, 30 large blocks with gene pairs > 200) or hexaploidy species wheat (1436, 140) than other species such as rice (534, 50), sorghum (523, 47) and Arabidopsis (177, 0)

(Fig. S3-5). Moreover, barley exhibited a 1:1 syntenic ratio with sorghum and rice, a 1:2 ratio with maize and a 1:3 ratio with hexaploid wheat (Fig. 1H, Fig. S6-8), consistent with known polyploidization histories [53]. Furthermore, barley exhibited good synteny with each subgenome of wheat in chromosome levels except a large inversion in chromosome 4A (Fig. S8). These results demonstrated that barley genome did not undergo whole-genome polyploidization event after divergence from sorghum, but instead experienced recent segmental duplications.

Comparative transposable element landscapes

TEs are major components of plant genomes and are closely associated with genomic evolution [56, 57]. To characterize the potential contributions of TEs to genome evolution in barley, we annotated the repeat sequences in the genomes of barley and the five other plant species using a uniform annotation pipeline. Repeat sequences accounted for 88.34% of barley genome (genome size ~ 4.26 Gb), comparable to the proportion in maize (85.43%, 2.11 Gb) and wheat (88.37%, 14.23 Gb), but higher than sorghum (68.91%, 683.65 Mb), rice (51.12%, 373.25 Mb) and Arabidopsis (14.31%, 119.15 Mb) (Table S7), suggesting repeat sequence content increased with genome size. We further focused on DNA transposons and long terminal repeat retrotransposons (LTR-RTs). The LTR-RTs, most of which were Gypsy elements (Fig. 2A), constituted the major fraction of repeat sequences across all grass plant species, accounting for 80.83% proportion in barley, comparable to maize (80.21%) and wheat (84.12%), but higher than sorghum (75.81%) and rice (43.62%) (Fig. 2A and Table S7).

To characterize transposable element amplification dynamics, we identified the full-length LTR-RTs in each genome and estimated their insertion times. In barley, LTR-RT insertions appeared to occur continuously, with distinct insertion peaks at approximately 1.0-1.5 and 1.8–2.2 million years ago (Mya) for Copia and Gypsy elements, respectively (Fig. 2B). These peaks were older than those observed in the other four grass species (Fig. 2B and Fig. S9). The LTR-RT expansions in barley were further supported by the phylogenetic tree of full-length LTR-RTs following divergence from rice (Fig. 2C). These results revealed that extensive LTR-RTs proliferation coincides with, and may contribute to the large genome size of barley.

Comparative DNA methylation analyses

Plants repress TE activity to maintain genome stability through various chromatin modifications, such as cytosine DNA methylation [2]. Given the large variations of TE content among surveyed plant species, we expected pronounced differences in the DNA methylation patterns. We downloaded the leaf whole-genome bisulfite sequencing (WGBS) data of the comparative plant species (Table S1; > 26-fold for all species) and quantified the DNA methylation levels at single-nucleotide resolution. As expected, the genome-wide DNA methylation levels increased with the proportions of TEs in the genome (Fig. S10). Moreover, methylation levels of mCG and mCHG were positively correlated with genomic TE proportions, whereas mCHH exhibited no significantly correlation (Fig. 3A). We next examined the methylation levels of different contexts in the distinct genomic features. For genes, the methylation levels of different contexts decreased at the boundaries of genes across all species relative to flanking regions. The

methylation of mCG within gene body returned to the flanking levels, whereas mCHG and mCHH remained low levels throughout gene body regions (Fig. 3B). Additionally, barley exhibited the highest methylation levels of mCG and lowest methylation levels of mCHH within gene regions among grass species examined (Fig. 3B). For TEs, methylation levels of different contexts increased at the boundaries in all species relative to flanking regions and remained high levels within the TE body regions (Fig. 3B). Similar to gene regions, barley exhibited almost the highest methylation levels of mCG and lowest methylation levels of mCHH in the TE regions among grass species (Fig. 3B). Furthermore, chromosome-scale analyses revealed that methylation levels of mCG and mCHG positively correlated with LTR-RT density and negatively with gene density across all species, whereas mCHH exhibited distinct correlation patterns among different plant species (Fig. 3C and Fig. S11-20). Notably, the methylation levels of mCHH in barley positively correlated with gene density (correlation coefficient $r = 0.72$; $p < 2.2 \times 10^{-16}$) and negatively correlated with LTR-RT density ($r = -0.54$; $p < 2.2 \times 10^{-16}$) (Fig. S20), a pattern also observed in rice and sorghum. In contrast, no strong correlation of mCHH with the gene or LTR-RT density ($r < 0.2$; $p < 0.001$) were detected in maize and wheat (Fig. S20). These results demonstrate that DNA methylation patterns are closely associated with TE content and distribution in the barley genome.

Identification of candidate domestication-associated genes in barley

The selective sweeps or selection genes during barley domestication have been investigated using exome sequencing [58, 59], which targeted approximately 60 Mb of coding sequences of barley genome [60]. To identify the candidate genes under selection during barley domestication at whole-genome scale, single-nucleotide polymorphisms (SNPs) of 291 accessions (Fig. 4A), including 93 wild and 198 domesticated accessions from previous study [23], were employed to detect selective sweeps. The phylogenetic analysis and principal component analysis (PCA) using whole-genome SNPs clearly separated all accessions into wild and domesticated groups (Fig. 4B and 4C). The linkage disequilibrium (LD; measured as r^2) dropped to half of its maximum value at approximately 401 kb for wild and 2.12 Mb for domesticated accessions (Fig. 4D), comparable to wheat [61] but substantially larger than most cultivated crops, such as rice (20kb for wild and 123kb for cultivars) [62] and maize (22kb for wild and 30kb for cultivars) [63].

Population differentiation value (F_{ST}), nucleotide diversity (π) ratios and Tajima's D statistic values between wild and domesticated populations across the whole genome were detected (Fig. 5A), and genomic regions with top 1% values were defined as selective sweeps. A total of 546, 398 and 435 genes overlapped with the selective sweeps identified by F_{ST} , π and Tajima's D methods, respectively, and the 243 genes identified by at least two methods were defined as candidate domestication-associated genes (Fig. 5B and Table S8). The KEGG enrichments of candidate domestication-associated genes were significantly enriched in starch and sucrose metabolism, cyanoamino acid metabolism and base excision repair (Fig. 5C and Table S9).

To further characterize the KEGG-enriched genes, their transcriptional expression patterns of these genes were examined in different developmental [20] and germinating tissues [31] (Fig. 5D). We then focused on a hexokinase-encoding gene Horvu_MOREX_1H01G192600, which exhibited high transcriptional levels in examined tissues (Fig. 5D). Hexokinases play essential roles in glucose sensing and sugar metabolism, thereby impacting plant growth and development [64]. A total of 16 SNPs were identified within Horvu_MOREX_1H01G192600 and its 1 kb flanking regions, including six upstream SNPs, one synonymous and one nonsynonymous SNPs in the exons, four intronic SNPs and four downstream SNPs (Fig. 5E and Table S10). These SNPs defined 14 haplotypes (H1-H14), with the frequency of H1 increasing from 23.7% (22/93) in the wild to 93.9% (186/198) in the domesticated population (Fig. 5E). Furthermore, Horvu_MOREX_1H01G192600 was located within a selective sweep detected by both π and Tajima's D statistic (Fig. 5F). The amino acid corresponding to nonsynonymous mutation of Horvu_MOREX_1H01G192600 was not conserved among grass plant species (Fig. 5G), and its significant divergence of the nonsynonymous mutation between wild (20.4%; 19/93) and domesticated (0%; 5/198) population was not observed (Fig. 5E). However, the variant frequency of downstream SNP (c. A3509C) was decreased from wild (76.3%; 71/93) to domesticated population (3.5%; 7/198), consistent with strong selection signals during domestication (Fig. 5E).

A total of 64 candidate domestication-associated genes were identified by three methods and clustered into three groups based on their transcriptional expressions, with group III showing high expressions in different tissues (Fig. 6A). Based on functional annotation, we further focused on a GTPase-encoding gene Horvu_MOREX_7H01G519400, as GTPases are crucial switches by binding GTP and GDP and regulate diverse cellular processes [65]. A total of 36 SNPs were identified within Horvu_MOREX_7H01G519400 and its the 1 kb flanking regions, including ten upstream SNPs, two synonymous SNPs in the exons, 18 intronic SNPs and six downstream SNPs (Fig. 6B and Table S11). These SNPs defined 43 haplotypes (H1-H43), with the frequency of H1 increased from 0% (0/93) in the wild to 93.4% (185/198) in the domesticated population (Fig. 6B and 6C). Notably, Horvu_MOREX_7H01G519400 was located in a selective sweep consistently identified by all three methods (Fig. 6D). The three downstream SNPs (c. T3560C, T3657C and C3924T) exhibited strong linkage disequilibrium (Fig. 6B), and their variant frequencies were decreased from wild (89.3%, 83/93; 98.9%, 92/93; 98.9%, 92/93) to domesticated population (1.0%, 2/198; 0.5%, 1/198; 0.5%, 1/198), reflecting strong selection during domestication.

Comparative A/B compartment organizations between wild and domesticated barley

Dynamic changes in 3D genome architecture contribute to the regulations of gene expression and are associated with plant growth, development, and environmental responses [9, 10]. To investigate the changes of chromatin organization between wild and domesticated barley, we obtained the reference genome assemblies [23, 25] and seedling leaf Hi-C datasets [20, 25] of the domesticated cultivar Morex and two wild barley accessions (ECN1 and ECS1). The wild barley ECN1 was from north slope of evolution canyon with a cooler climate and higher relative humidity, whereas the wild barley ECS1 was

from south slope with higher daily temperatures [25]. We observed strong contact signals in the diagonals as expected and relatively weak anti-diagonal signals in the intrachromosomal Hi-C contact matrices of all three genomes (Fig. S21), consistent with previous reported Rabl-type configurations in barley[20]. Furthermore, comparative analysis of Hi-C contact signals revealed an enrichment of short-range chromatin contacts in the wild barley accessions relative to Morex (Fig. 7A and Fig. S22A).

Plant chromatin is broadly partitioned into euchromatic A compartments and heterochromatic B compartments [9, 10]. We annotated the compartment regions in the three barley genomes, and found that Morex harbor higher proportion of A compartment compared with ECS1 or ECN1 (Fig. 7B), and approximate 8.9% compartment exhibited status variations among the three genomes (Fig. 7C). In total, 1445 and 2037 genes exhibited B-to-A compartment transitions in the comparisons of ECN1 versus (vs) Morex (Table S12) and ECS1 vs Morex (Table S13), respectively, whereas only 798 and 660 genes showed A-to-B compartment transitions in the corresponding comparisons. KEGG enrichment analysis of genes with B-to-A compartment transitions in ECN1 or ECS1 vs Morex were involved in the phenylpropanoid biosynthesis pathway (Fig. 7D and Fig. S22B; Table S14 and Table S15). The phenylpropanoid metabolism plays crucial roles in plant defense and stress tolerance in plant [66], raising the possibility that selection on genes in this pathway contributed to enhanced environmental adaptation during barley domestication. To further evaluate this hypothesis, we overlapped phenylpropanoid-biosynthesis enriched genes with candidate domestication-associated genes, and identified two shared genes (a cinnamate-f-hydroxylase-encoding gene Horvu_MOREX_3H01G65700 and a peroxidase-encoding gene Horvu_MOREX_3H01G166100) in ECN1 and ECS1 relative to Morex (Fig. 7D and Fig. S22B).

A total of 12 SNPs were observed within Horvu_MOREX_3H01G65700 and its the 1 kb flanking regions, including four upstream SNPs, one nonsynonymous SNP in the exons, three intronic SNPs and four downstream SNPs (Fig. 7E and Table S16). These SNPs defined 14 haplotypes (H1-H14), with the frequency of H1 increased from 0% (0/93) in the wild to 91.4% (181/198) in the domesticated population (Fig. 7E). Furthermore Horvu_MOREX_3H01G65700 was located within a selective sweep detected by π and Tajima's D statics (Fig. 7F). The nonsynonymous (c. T3038G) and the downstream SNPs (c. C4148T) showed strong linkage disequilibrium (Fig. 7E) and their variant frequencies were decreased from wild (100%, 93/93) to domesticated population (8.6%, 17/198), consistent with strong selection signals during domestication (Fig. 7E). However, the amino acid corresponding to nonsynonymous mutation of Horvu_MOREX_3H01G65700 was not conservative among grass plant species (Fig. 7G), suggesting the downstream SNPs (c. G4148A) might be the site under artificial selection. Similarly, 25 SNPs in candidate domestication-associated gene Horvu_MOREX_3H01G166100 defined 19 haplotypes (H1-H19), with the frequency of H1 increased from 0% (0/93) in the wild to 91.4% (181/198) in the domesticated population (Fig. S22C and Table S17). The three downstream SNPs (c. C1743T, C1560G and T1439C) of Horvu_MOREX_3H01G166100 showed strong linkage disequilibrium (Fig. S22C), whose variant mutations decreased from wild (100%, 93/93; 98.9%, 92/93; 98.9%, 92/93) to domesticated population (8.1%, 16/198; 8.1%, 16/198; 7.6%, 15/198), consistent with strong selection signals during

domestication. These results indicated that the dynamic A/B compartment reorganization between wild and domesticated barley represents a prominent chromatin feature accompanying barley domestication.

Comparative TAD structures between wild and domesticated barley

To investigate the evolution of TADs during domestication, we annotated the TAD domains across the three genomes, and identified 4,162, 4,355 and 2,644 TAD domains in ECN1, ECS1 and Morex, covering 99.40%, 99.28% and 99.37% of their respective genomes (Table S18). Morex exhibited significantly larger TAD domains compared with ECN1 and ECS1 (Fig. 8A). To assess the conservation and divergence of the TAD boundaries among the three genomes, we clustered TAD boundaries and found that 1514 boundary clusters were shared across the three genomes (Fig. 8B and Table S19). Notably, a substantial proportion of variety-specific TAD boundary clusters were observed, with 27.7% (1,117/4,030) in ECN1, 28.8% (1,209/4,197) in ECS1 and 21.1% (546/2,587) in Morex (Fig. 8B), indicating extensive divergence of TAD boundaries between wild and domesticated barley. For example, within the syntenic regions on chromosome 7, we identified 14 conserved TAD boundaries between ECN1 and ECS1 along with five ECN1-specific and two ECS1-specific boundaries. In contrast, only eight of the 14 boundaries existed in Morex, accompanied by five Morex-specific boundaries (Fig. 8C). However, gene clustering analysis revealed a high degree of conservation among the three genomes, with 21311 gene families shared, accounting for 81.5% of total gene families in ECN1, 80.9% in ECS1 and 86.7% in Morex, respectively (Fig. S23). These results demonstrated that although gene count and molecular functions are highly conserved between wild and domesticated barley, the regulatory architecture-particularly TAD organizations-has diverged.

Discussion

Barley, which exhibits strong adaptation to diverse environments, ranks as the fourth most important cereal crop worldwide in terms of harvested area and production, and serves as a major resource for livestock feed and brewing industries [18]. The increasing frequency of extreme heat and drought events has imposed substantial negative impacts on barley production and global beer supply [67]. The long-time artificial selection of barley has significantly reduced the genetic diversity of modern cultivars and promoted genetic homogeneity [68]. Therefore, a comprehensive understanding of barley genomic evolution and domestication is essential to develop environmentally resilient and high-performing cultivars.

Our study indicates that the large genome size of barley is primarily attributed to extensive proliferation of TEs (Fig. 2), particularly Copia and Gypsy LTR retrotransposons, a feature commonly observed in plant species with large genomes [3]. A recent segmental duplication ($K_s \sim 0.2$) observed in barley (Fig. 1E) corresponds to previously reported TE-driven gene duplication in Triticeae species [55] and may have provided additional genetic materials for environmental adaption. The enrichment of barley-specific genes in mismatch repair pathway may be associated with adaptation to environments characterized by

elevated UV radiation, such as high-altitude regions including the Qinghai-Tibetan plateau. The absence of WGD or WGT after divergence from sorghum (Fig. 1F and 1G), as well as high collinearity with wheat (Fig. S8) establishes barley as an informative model crop for genomic investigation within the Triticeae.

The long-time domestication has led to a suite of traits that distinguish cultivated barley from its wild progenitors, such as row type of spike and grain disperse [18]. Although barley was first domesticated in the Fertile Crescent before spreading to other regions [16], the mosaic genomic ancestry feature of barley [16], which results in high haplotype differentiation among geographically distinct barley populations, poses challenges for mapping domesticated genes.

Here, we provide valuable candidate domestication-associated genes for further functional investigation. It should be emphasized that candidate domestication-associated genes were identified without explicitly accounting for the geographic structure of wild and domesticated populations, largely due to limited resequencing sample size. With the accumulation of geographically representative resequencing datasets, integrating geographic population structure will enable more precise dissection of domestication signals and reduce confounding effects of population differentiation. Besides genes, TE insertion polymorphisms (TIPs) also contribute to plant diversity [5] and could undergo artificial selection via affecting nearby gene expressions during crop domestication [4]. It will be of interest in future studies to investigate the potential roles of TIPs in barley domestication.

Strong positive correlations of TEs and methylation levels of mCG or mCHH contexts were observed across plant species at genome-wide and chromosome scales (Fig. 3A, 3C and Fig. S20), suggesting repression effects of DNA methylation on TE transposition to preserve genomic stability, consistent with previous reports [2, 6]. The 3D architecture dynamics of plant genomes play essential roles in gene regulations [11], environmental responses [14], and also are involved in crop domestication [13]. High conservation of gene content and substantial divergence of TAD boundaries were observed between wild and cultivated barley (Fig. 8B, 8C and Fig. S23). In animals the TAD boundaries are established by CTCF binding sites, which facilitate the recruitment of CTCF proteins and cohesin [11]. However, plants lack CTCF proteins, and TADs are likely formed through alternative mechanisms, including TE transposition [69]. Therefore, the reduced Gypsy density but elevated Copia and DNA transposon densities at the TAD boundaries of wild and cultivated barley (Fig. 8D) suggested potential roles of Copia retrotransposons and DNA transposons in higher-order chromatin organizations in barley. Together, these findings highlight the essential role of TE in shaping barley genome evolution and driving regulatory rewiring during domestication, particularly through their interactions with DNA methylation and high-order chromatin architecture.

Conclusion

In this study, comparative genomic analyses and TE component profiling indicate contiguous insertions of LTR retrotransposon are closely associated with barley genome expansion. In addition, comparative methylome analyses across plant species reveal that transposable elements exhibit strong

correspondence with genome-wide DNA methylation landscapes. Furthermore, by integrating analyses of transposable elements, A/B compartmentalization and topologically associating domains, we demonstrate that barley domestication is accompanied by extensive reorganization of higher-order chromatin particularly in genomic regions enriched for TEs, largely in the context of conserved gene content. These results suggest that regulatory rewiring, rather than widespread gene gain or loss, represents a prominent feature associated with barley domestication. Together, our study provides valuable genomic resources for future functional investigations in barley and offers a conceptual framework for incorporating 3D genome features into molecular breeding strategies.

Declarations

Authorship contributions

WY conceived the research, performed data analysis and drafted manuscript. KC, LL, YL, FG and QY performed data analysis. SC and JW reviewed and edited the manuscript. All authors reviewed and approved the final manuscript.

Funding

This work was supported by Zhejiang Provincial Talent Cultivation Project (2023R5259), Precision identification, evaluation, and storage of barley germplasm resources in Zhejiang Province (2025JZJD001), China Agriculture Research System (CARS-05-01A-06), and National Natural Science Foundation of China (32301793).

Data availability

The reference genome assembly of *Hordeum vulgare* (cv. Morex) [23] is publicly available via barley pangenome repository at <https://barley-pangenome.ipk-gatersleben.de>. The genome assemblies of two wild barley accessions (ECN1 and ECS1) [25] have been deposited in the NCBI database under BioProject accession PRJNA947680. The reference genomes of *Arabidopsis thaliana*, *Sorghum bicolor*, *Zea mays*, *Oryza sativa* and *Triticum aestivum* were obtained from Phytozome (<https://phytozome-next.jgi.doe.gov/>). The whole-genome bisulfite sequencing (WGBS) data of *Arabidopsis* (SRR13650156 and SRR13650157), maize (SRR12463972 and SRR12463973), barley (SRR13345339), and wheat (SRP133674) are accessible through Sequence Read Archive (SRA, <https://www.ncbi.nlm.nih.gov/sra>). The WGBS data of sorghum (CRR145305) and rice (CRR190552) are available in the Genome Sequence Archive (GSA, <https://ngdc.cnbc.ac.cn/gsa/>). The Hi-C data of Morex (PRJEB14169), ECN1 (PRJNA748178) and ECS1 (PRJNA748178) have been deposited in the NCBI database. The SNP VCF file [23] of diverse barley accessions analyzed in this study were downloaded from <https://doi.ipk-gatersleben.de/DOI/c4d433dc-bf7c-4ad9-9368-69bb77837ca5/1b898e9a-11bf-4eb6-b165-e1fa2167f0ec/0>. The RNA-Seq data of different developmental tissues [20] and different tissues of germinating seeds [31] of barley are available in NCBI database under PRJEB14349 and PRJNA378132, respectively.

The datasets supporting the conclusions of this article are available in the Figshare repository (10.6084/m9.figshare.31563343), or included in the article and its supplementary materials. Additional materials and resources are available from the corresponding authors upon reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

All authors declare that there are no competing interests.

References

1. Shen S, Li N, Wang Y, Zhou R, Sun P, Lin H, Chen W, Yu T, Liu Z, Wang Z *et al*: **High-quality ice plant reference genome analysis provides insights into genome evolution and allows exploration of genes involved in the transition from C3 to CAM pathways**. *Plant Biotechnol J* 2022, **20**(11):2107-2122.
2. Liu P, Cuerda-Gil D, Shahid S, Slotkin RK: **The Epigenetic Control of the Transposable Element Life Cycle in Plant Genomes and Beyond**. *Annu Rev Genet* 2022, **56**:63-87.
3. Novak P, Guignard MS, Neumann P, Kelly LJ, Mlinarec J, Koblizkova A, Dodsworth S, Kovarik A, Pellicer J, Wang W *et al*: **Repeat-sequence turnover shifts fundamentally in species with large genomes**. *Nat Plants* 2020, **6**(11):1325-1329.
4. Cai X, Lin R, Liang J, King GJ, Wu J, Wang X: **Transposable element insertion: a hidden major source of domesticated phenotypic variation in Brassica rapa**. *Plant Biotechnol J* 2022, **20**(7):1298-1310.
5. Dominguez M, Dugas E, Benchouaia M, Leduque B, Jimenez-Gomez JM, Colot V, Quadrana L: **The impact of transposable elements on tomato diversity**. *Nat Commun* 2020, **11**(1):4058.
6. Zhang H, Zhu JK: **Epigenetic gene regulation in plants and its potential applications in crop improvement**. *Nat Rev Mol Cell Biol* 2025, **26**(1):51-67.
7. Lloyd JPB, Lister R: **Epigenome plasticity in plants**. *Nat Rev Genet* 2022, **23**(1):55-68.
8. Niederhuth CE, Bewick AJ, Ji L, Alabady MS, Kim KD, Li Q, Rohr NA, Rambani A, Burke JM, Udall JA *et al*: **Widespread natural variation of DNA methylation within angiosperms**. *Genome Biol* 2016, **17**(1):194.
9. Pei L, Li G, Lindsey K, Zhang X, Wang M: **Plant 3D genomics: the exploration and application of chromatin organization**. *New Phytol* 2021, **230**(5):1772-1786.
10. Domb K, Wang N, Hummel G, Liu C: **Spatial Features and Functional Implications of Plant 3D Genome Organization**. *Annu Rev Plant Biol* 2022, **73**:173-200.

11. Misteli T: **The Self-Organizing Genome: Principles of Genome Architecture and Function.** *Cell* 2020, **183**(1):28-45.
12. Graf C, Winkler TS, Maughan PJ, Stetter MG: **Domestication shaped the chromatin landscape of grain amaranth.** *Nat Commun* 2025, **16**(1):10407.
13. Wang L, Jia G, Jiang X, Cao S, Chen ZJ, Song Q: **Altered chromatin architecture and gene expression during polyploidization and domestication of soybean.** *Plant Cell* 2021, **33**(5):1430-1446.
14. Li Z, Sun L, Xu X, Liu Y, He H, Deng XW: **Light control of three-dimensional chromatin organization in soybean.** *Plant Biotechnol J* 2024, **22**(9):2596-2611.
15. Liang Z, Zhang Q, Ji C, Hu G, Zhang P, Wang Y, Yang L, Gu X: **Reorganization of the 3D chromatin architecture of rice genomes during heat stress.** *BMC Biol* 2021, **19**(1):53.
16. Guo Y, Jayakodi M, Himmelbach A, Ben-Yosef E, Davidovich U, David M, Hartmann-Shenkman A, Kislev M, Fahima T, Schuenemann VJ et al: **A haplotype-based evolutionary history of barley domestication.** *Nature* 2025, **647**(8090):680-688.
17. Civan P, Drosou K, Armisen-Gimenez D, Duchemin W, Salse J, Brown TA, Venezolana AC: **Episodes of gene flow and selection during the evolutionary history of domesticated barley.** *BMC Genomics* 2021, **22**(1):227.
18. Jiang C, Kan J, Gao G, Dockter C, Li C, Wu W, Yang P, Stein N: **Barley2035: A decadal vision for barley research and breeding.** *Mol Plant* 2025, **18**(2):195-218.
19. Lundqvist U: **Scandinavian mutation research in barley - a historical review.** *Hereditas* 2014, **151**(6):123-131.
20. Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T, Radchuk V, Dockter C, Hedley PE, Russell J et al: **A chromosome conformation capture ordered sequence of the barley genome.** *Nature* 2017, **544**(7651):427-433.
21. Jayakodi M, Lu Q, Pidon H, Rabanus-Wallace MT, Bayer M, Lux T, Guo Y, Jaegle B, Badea A, Bekele W et al: **Structural variation in the pangenome of wild and domesticated barley.** *Nature* 2024, **636**(8043):654-662.
22. Mascher M, Wicker T, Jenkins J, Plott C, Lux T, Koh CS, Ens J, Gundlach H, Boston LB, Tulpova Z et al: **Long-read sequence assembly: a technical evaluation in barley.** *Plant Cell* 2021, **33**(6):1888-1906.
23. Jayakodi M, Padmarasu S, Haberer G, Bonthala VS, Gundlach H, Monat C, Lux T, Kamal N, Lang D, Himmelbach A et al: **The barley pan-genome reveals the hidden legacy of mutation breeding.** *Nature* 2020, **588**(7837):284-289.
24. Cai S, Shen Q, Huang Y, Han Z, Wu D, Chen ZH, Nevo E, Zhang G: **Multi-Omics Analysis Reveals the Mechanism Underlying the Edaphic Adaptation in Wild Barley at Evolution Slope (Tabigha).** *Adv Sci* 2021, **8**(20):e2101374.
25. Pan R, Hu H, Xiao Y, Xu L, Xu Y, Ouyang K, Li C, He T, Zhang W: **High-quality wild barley genome assemblies and annotation with Nanopore long reads and Hi-C sequencing data.** *Sci Data* 2023, **10**(1):535.

26. Cheng CY, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD: **Araport11: a complete reannotation of the Arabidopsis thaliana reference genome.** *Plant J* 2017, **89**(4):789-804.
27. McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, Kennedy M, Amirebrahimi M, Weers BD, McKinley B et al: **The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization.** *Plant J* 2018, **93**(2):338-354.
28. Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS et al: **Improved maize reference genome with single-molecule technologies.** *Nature* 2017, **546**(7659):524-527.
29. Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L et al: **The TIGR Rice Genome Annotation Resource: improvements and new features.** *Nucleic Acids Res* 2007, **35**(Database issue):D883-887.
30. Zhu T, Wang L, Rimbert H, Rodriguez JC, Deal KR, De Oliveira R, Choulet F, Keeble-Gagnere G, Tibbits J, Rogers J et al: **Optical maps refine the bread wheat Triticum aestivum cv. Chinese Spring genome assembly.** *Plant J* 2021, **107**(1):303-314.
31. Betts NS, Berkowitz O, Liu R, Collins HM, Skadhauge B, Dockter C, Burton RA, Whelan J, Fincher GB: **Isolation of tissues and preservation of RNA from intact, germinated barley grain.** *Plant J* 2017, **91**(4):754-765.
32. Yue W, Cai K, Xia X, Liu L, Wang J: **Genome-wide identification, expression pattern and genetic variation analysis of SWEET gene family in barley reveal the artificial selection of HvSWEET1a during domestication and improvement.** *Front Plant Sci* 2023, **14**:1137434.
33. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST et al: **The variant call format and VCFtools.** *Bioinformatics* 2011, **27**(15):2156-2158.
34. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ et al: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81**(3):559-575.
35. Emms DM, Kelly S: **OrthoFinder: phylogenetic orthology inference for comparative genomics.** *Genome Biol* 2019, **20**(1):238.
36. Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H et al: **MCSanX: a toolkit for detection and evolutionary analysis of gene syntenic and collinearity.** *Nucleic Acids Res* 2012, **40**(7):e49.
37. Chen H, Zwaenepoel A: **Inference of ancient polyploidy from genomic data.** In: *Polyploidy: Methods and protocols.* Springer; 2023: 3-18.
38. Yang Z, Nielsen R: **Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models.** *Mol Biol Evol* 2000, **17**(1):32-43.
39. Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T et al: **Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline.** *Genome Biol* 2019, **20**(1):275.

40. Zhang RG, Li GY, Wang XL, Dainat J, Wang ZX, Ou S, Ma Y: **TEsorter: an accurate and fast method to classify LTR-retrotransposons in plant genomes.** *Hortic Res* 2022, **9**:uhac017.
41. Neumann P, Novak P, Hostakova N, Macas J: **Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification.** *Mobile DNA* 2019, **10**:1.
42. Zhou Q, Lim JQ, Sung WK, Li G: **An integrated package for bisulfite DNA methylation data analysis with Indel-sensitive mapping.** *BMC Bioinformatics* 2019, **20**(1):47.
43. Tang K, Zhu X, Xie S, Lang Z, Zhu JK: **Transgenerational increases in DNA methylation in Arabidopsis plants defective in active DNA demethylation.** *Proc Natl Acad Sci U S A* 2024, **121**(22):e2320468121.
44. Zhou C, Yuan Z, Ma X, Yang H, Wang P, Zheng L, Zhang Y, Liu X: **Accessible chromatin regions and their functional interrelations with gene transcription and epigenetic modifications in sorghum genome.** *Plant Commun* 2021, **2**(1):100140.
45. Cao S, Chen K, Lu K, Chen S, Zhang X, Shen C, Zhu S, Niu Y, Fan L, Chen ZJ et al: **Asymmetric variation in DNA methylation during domestication and de-domestication of rice.** *Plant Cell* 2023, **35**(9):3429-3443.
46. Liang Z, Anderson SN, Noshay JM, Crisp PA, Enders TA, Springer NM: **Genetic and epigenetic variation in transposable element expression responses to abiotic stress in maize.** *Plant Physiol* 2021, **186**(1):420-433.
47. Ramirez-Gonzalez RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F, Pasha A et al: **The transcriptional landscape of polyploid wheat.** *Science* 2018, **361**(6403):662.
48. Ramirez F, Bhardwaj V, Arrigoni L, Lam KC, Gruning BA, Villaveces J, Habermann B, Akhtar A, Manke T: **High-resolution TADs reveal DNA sequences underlying genome organization in flies.** *Nat Commun* 2018, **9**(1):189.
49. Liu Y, Nanni L, Sungalee S, Zufferey M, Tavernari D, Mina M, Ceri S, Oricchio E, Ciriello G: **Systematic inference and comparison of multi-scale chromatin sub-compartments connects spatial organization to cell phenotypes.** *Nat Commun* 2021, **12**(1):2439.
50. He X, Huang X, Long Y, Liu Z, Chang X, Zhang X, Wang M: **Tcbf: a novel user-friendly tool for pan-3D genome analysis of topologically associating domain in eukaryotic organisms.** *Bioinformatics* 2023, **39**(9):btad576.
51. Clark JW, Donoghue PCJ: **Whole-Genome Duplication and Plant Macroevolution.** *Trends Plant Sci* 2018, **23**(10):933-945.
52. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS et al: **Ancestral polyploidy in seed plants and angiosperms.** *Nature* 2011, **473**(7345):97-100.
53. Zhang T, Huang W, Zhang L, Li DZ, Qi J, Ma H: **Phylogenomic profiles of whole-genome duplications in Poaceae and landscape of differential duplicate retention and losses among major Poaceae lineages.** *Nat Commun* 2024, **15**(1):3305.

54. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A *et al*: **The Sorghum bicolor genome and the diversification of grasses.** *Nature* 2009, **457**(7229):551-556.
55. Wang X, Yan X, Hu Y, Qin L, Wang D, Jia J, Jiao Y: **A recent burst of gene duplications in Triticeae.** *Plant Commun* 2022, **3**(2):100268.
56. Colonna Romano N, Fanti L: **Transposable Elements: Major Players in Shaping Genomic and Evolutionary Patterns.** *Cells* 2022, **11**(6):1048.
57. Pulido M, Casacuberta JM: **Transposable element evolution in plant genome ecosystems.** *Curr Opin Plant Biol* 2023, **75**:102418.
58. Russell J, Mascher M, Dawson IK, Kyriakidis S, Calixto C, Freund F, Bayer M, Milne I, Marshall-Griffiths T, Heinen S *et al*: **Exome sequencing of geographically diverse barley landraces and wild relatives gives insights into environmental adaptation.** *Nat Genet* 2016, **48**(9):1024-1030.
59. Chen YY, Schreiber M, Bayer MM, Dawson IK, Hedley PE, Lei L, Akhunova A, Liu C, Smith KP, Fay JC *et al*: **The evolutionary patterns of barley pericentromeric chromosome regions, as shaped by linkage disequilibrium and domestication.** *Plant J* 2022, **111**(6):1580-1594.
60. Mascher M, Richmond TA, Gerhardt DJ, Himmelbach A, Clissold L, Sampath D, Ayling S, Steuernagel B, Pfeifer M, D'Ascenzo M *et al*: **Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond.** *Plant J* 2013, **76**(3):494-505.
61. Guo J, Guo J, Li L, Bai X, Huo X, Shi W, Gao L, Dai K, Jing R, Hao C: **Combined linkage analysis and association mapping identifies genomic regions associated with yield-related and drought-tolerance traits in wheat (Triticum aestivum L.).** *Theor Appl Genet* 2023, **136**(12):250.
62. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z *et al*: **Genome-wide association studies of 14 agronomic traits in rice landraces.** *Nat Genet* 2010, **42**(11):961-967.
63. Hufford MB, Xu X, van Heerwaarden J, Pyhajarvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler SM *et al*: **Comparative population genomics of maize domestication and improvement.** *Nat Genet* 2012, **44**(7):808-811.
64. Eom SH, Kim E, Hyun TK: **HXK, SnRK1, and TOR signaling in plants: Unraveling mechanisms of stress response and secondary metabolism.** *Sci Prog* 2024, **107**(4):368504241301533.
65. Lu Y, Cheng K, Tang H, Li J, Zhang C, Zhu H: **The role of Rab GTPase in Plant development and stress.** *J Plant Physiol* 2024, **296**:154239.
66. Dong NQ, Lin HX: **Contribution of phenylpropanoid metabolism to plant development and plant-environment interactions.** *J Integr Plant Biol* 2021, **63**(1):180-209.
67. Xie W, Xiong W, Pan J, Ali T, Cui Q, Guan D, Meng J, Mueller ND, Lin E, Davis SJ: **Decreases in global beer supply due to extreme drought and heat.** *Nat Plants* 2018, **4**(11):964-973.
68. Landis JB, Guercio AM, Brown KE, Fiscus CJ, Morrell PL, Koenig D: **Natural selection drives emergent genetic homogeneity in a century-scale experiment with barley.** *Science* 2024, **385**(6705):eadl0038.

69. Wang M, Li J, Wang P, Liu F, Liu Z, Zhao G, Xu Z, Pei L, Grover CE, Wendel JF *et al*: **Comparative Genome Analyses Highlight Transposon-Mediated Genome Expansion and the Evolutionary Architecture of 3D Genomic Folding in Cotton.** *Mol Biol Evol* 2021, **38**(9):3621-3636.
70. Feschotte C, Pritham EJ: **DNA transposons and the evolution of eukaryotic genomes.** *Annu Rev Genet* 2007, **41**:331-368.

Figures

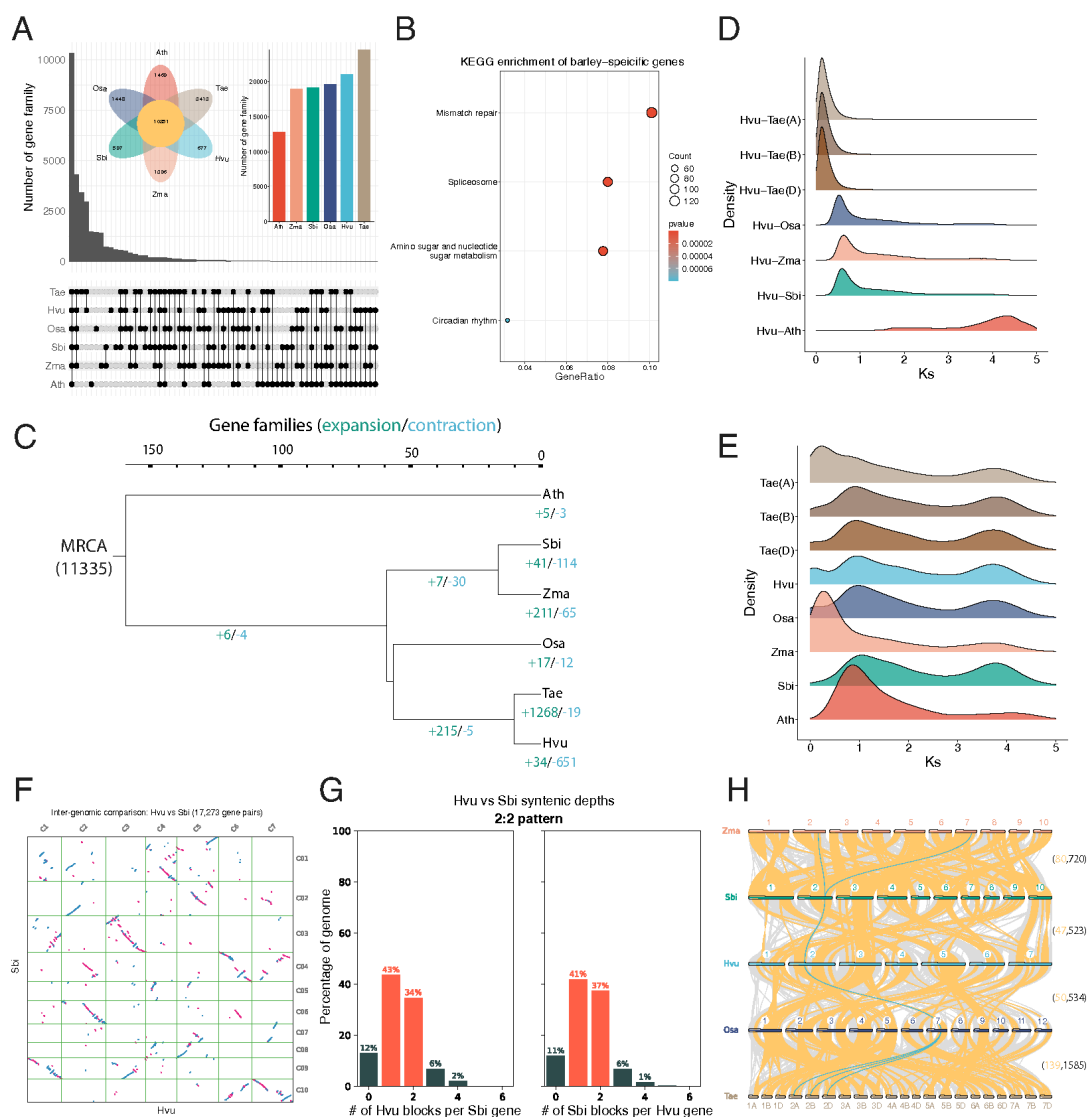


Fig. 1. Comparative genomic analysis between barley, four representative grass species and Arabidopsis. (A) Gene family clustering analysis. Insert images displaying the numbers of the common and species-specific gene families, and the total numbers of gene families in each plant species. (B) KEGG analysis of barley-specific gene families. (C) Gene family expansion and contraction analysis. The phylogenetic tree of plant species was based on the concatenation of 2502 single copy genes. The green and blue colors mark the numbers of expanded and contracted gene families, respectively. MRCA, most recent common ancestor. (D and E) Estimations of species divergences between barley and other plant species (D) and whole-genome duplications within different plant species (E). (F and G) Dot plot analysis (F) and syntenic depth analysis (G) of genome-wide syntenic blocks between barley and sorghum. (H) The genome-wide syntenic analysis among different plant species. The numbers in the parentheses indicate large (> 200 gene pairs, yellow) and small (5-200 gene pairs, black) syntenic blocks. Representative relationships among the plant species are denoted by blue lines. Abbreviation: Ath, *Arabidopsis thaliana*; Osa, *Oryza sativa*; Sbi, *Sorghum bicolor*; Zma, *Zea mays*; Hvu, *Hordeum vulgare*; Tae, *Triticum aestivum*.

Figure 1

See image above for figure legend.

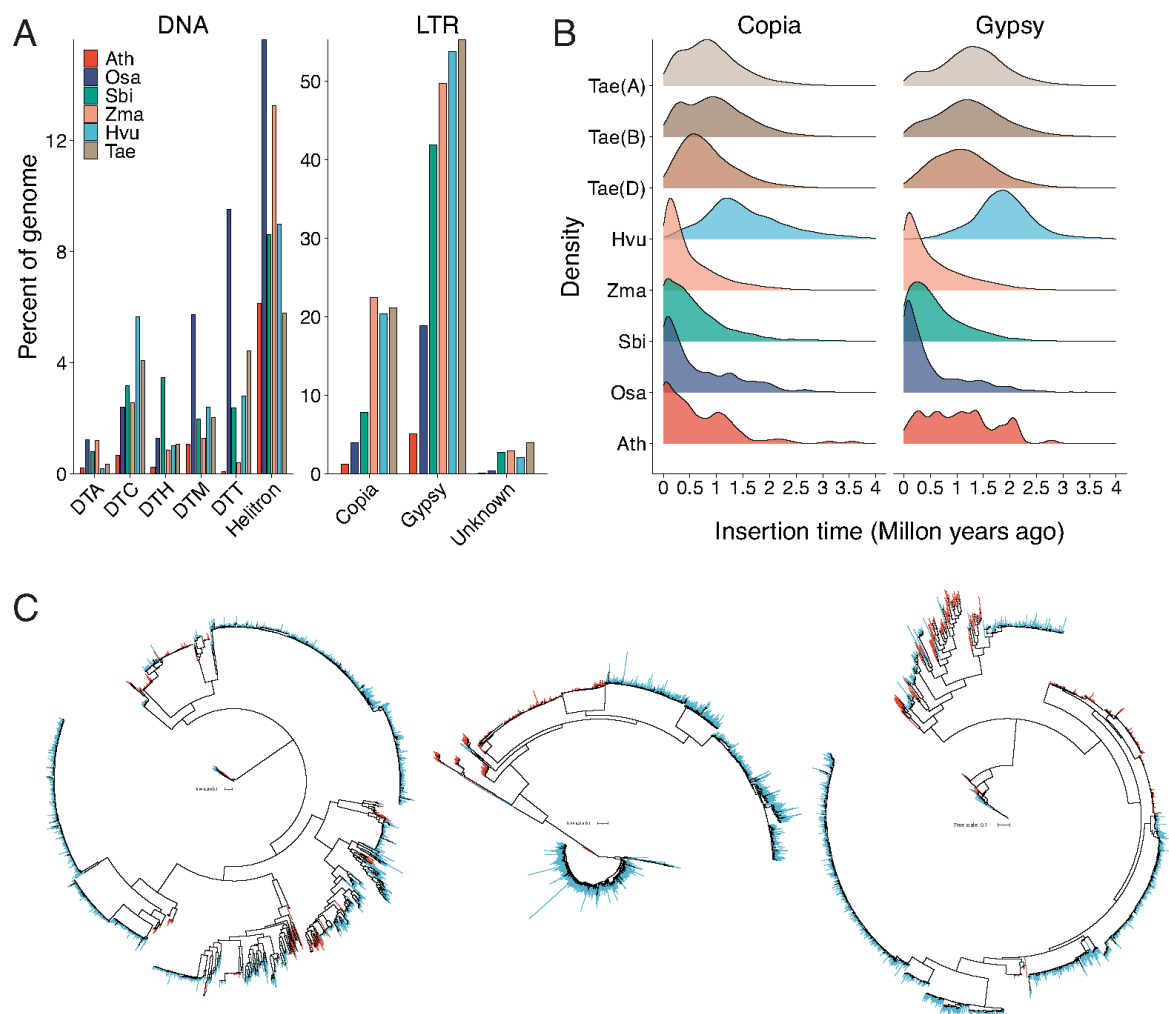


Fig. 2. LTR retrotransposon proliferation contributes to barley genome expansion. (A) The contents of transposable elements. DNA transposons were classified into Helitron and five major superfamilies [70], including DTA (hAT), DTC (CATA), DTH (PIF/Harbinger), DTM (Mutator) and DTT (Tc1/Mariner). (B) Estimation of insertion times of Copia and Gypsy transposons. (C) Phylogenetic tree of intact Copia transposons (left), chromovirus type Gypsy transposons (middle) and non-chromovirus type Gypsy transposons (right) of barley and rice. The phylogenetic trees include 1646 Copia transposons of barley and 318 of rice (left), 947 chromovirus type Gypsy transposons of barley and 414 of rice (middle), and 2360 non-chromovirus type Gypsy transposons of barley and 500 of rice (right). Abbreviation: Ath, *Arabidopsis thaliana*; Osa, *Oryza sativa*; Sbi, *Sorghum bicolor*; Zma, *Zea mays*; Hvu, *Hordeum vulgare*; Tae, *Triticum aestivum*.

Figure 2

See image above for figure legend.

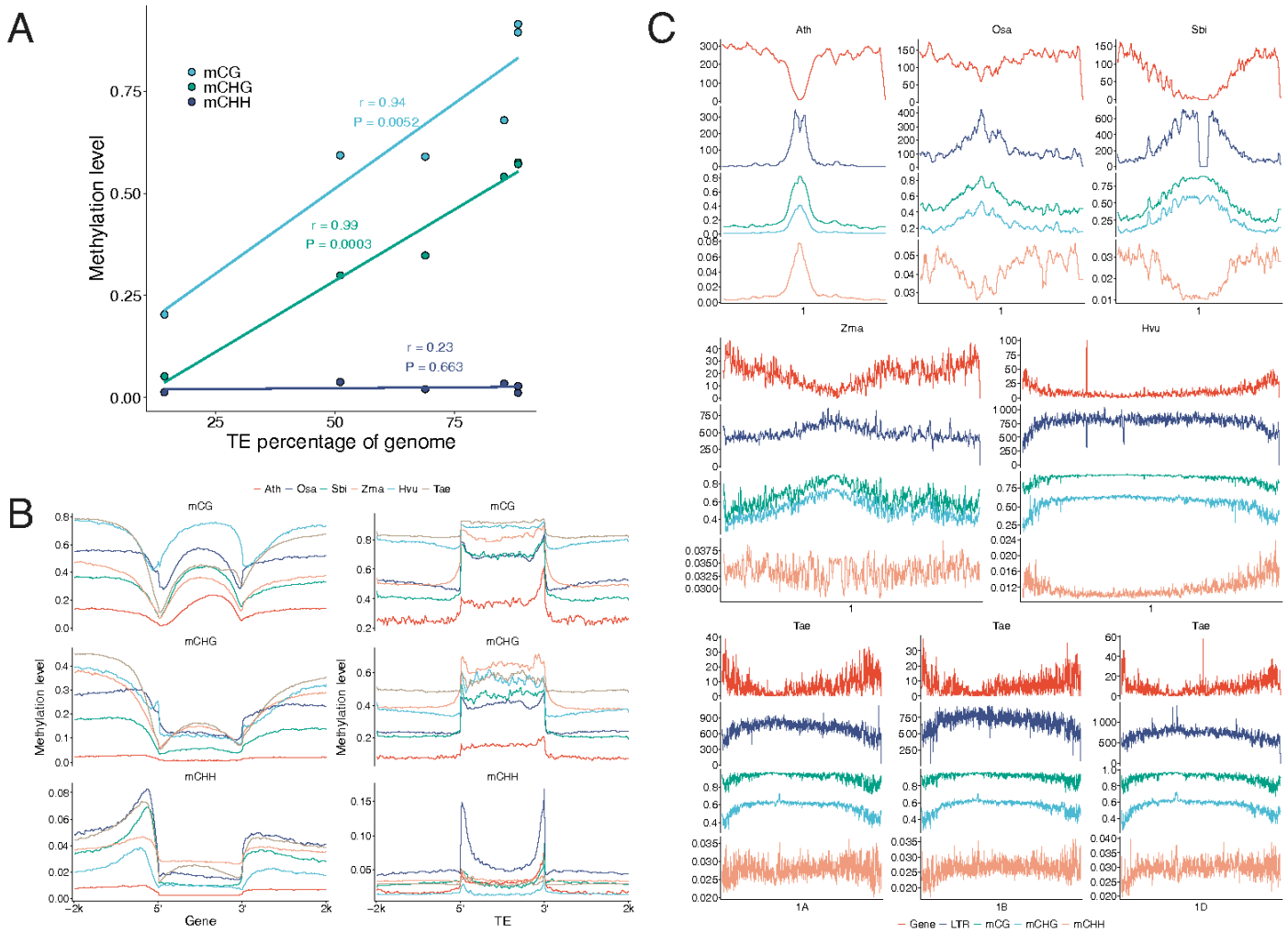


Fig. 3. Distinct DNA methylation patterns among barley, four representative grass species and Arabidopsis. (A) Correlation between DNA methylation levels and TE content. (B) Methylation patterns of mCG, mCHG and mCHH across the genomic regions ranging from 2kb upstream to 2kb downstream of genes (left) and TEs (right). (C) Distribution of genes, LTRs, and DNA methylation levels of mCG, mCHG and mCHH along chromosome 1 in different plant species, using 1Mb sliding windows with 200 kb steps. Abbreviation: Ath, *Arabidopsis thaliana*; Osa, *Oryza sativa*; Sbi, *Sorghum bicolor*; Zma, *Zea mays*; Hv, *Hordeum vulgare*; Tae, *Triticum aestivum*.

Figure 3

See image above for figure legend.

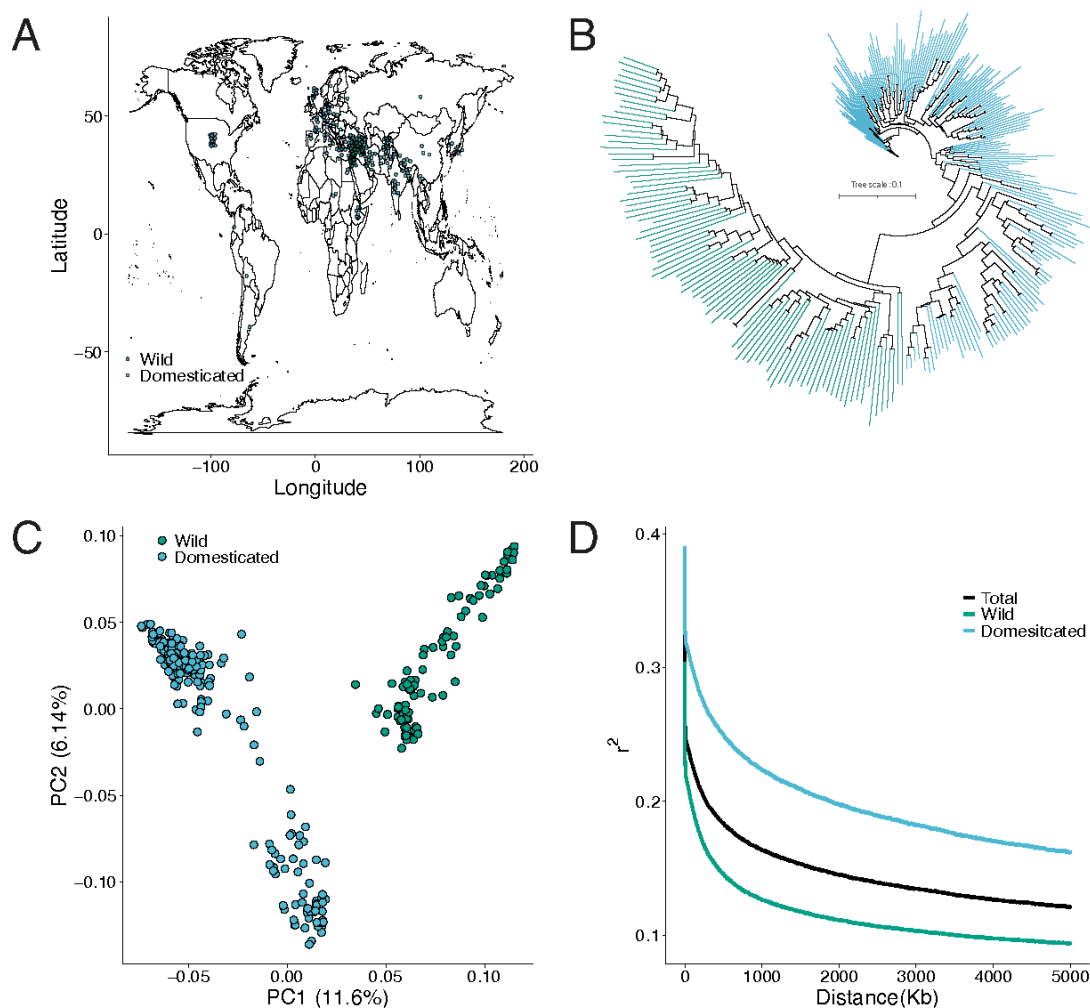


Fig. 4. Geographic distribution, population structure and decay of linkage disequilibrium (LD) of resequenced barley accessions. (A) The global geographic distribution of resequenced accessions. (B) The phylogenetic tree of resequenced accessions based on the whole-genome SNPs. (C) Principal component analysis of resequenced accessions in the first and second dimensions. (D) Decay of LD in whole, wild and domesticated populations. The whole, wild and domesticated populations are marked with black, green and blue colors, respectively.

Figure 4

See image above for figure legend.

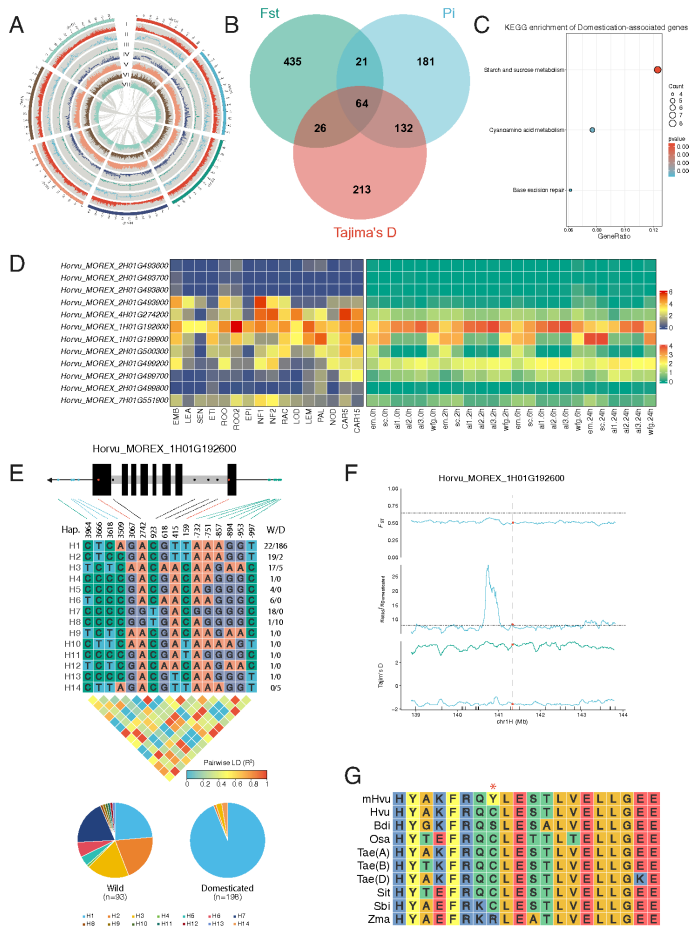


Figure 5

See image above for figure legend.

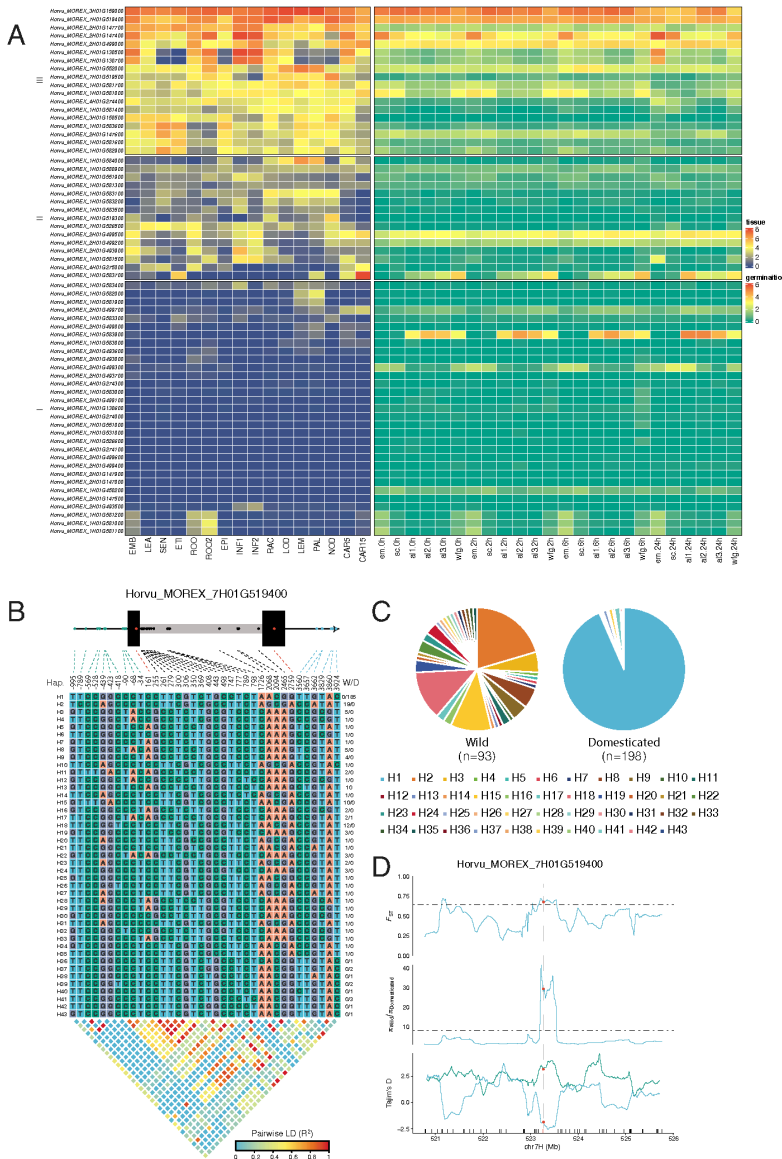
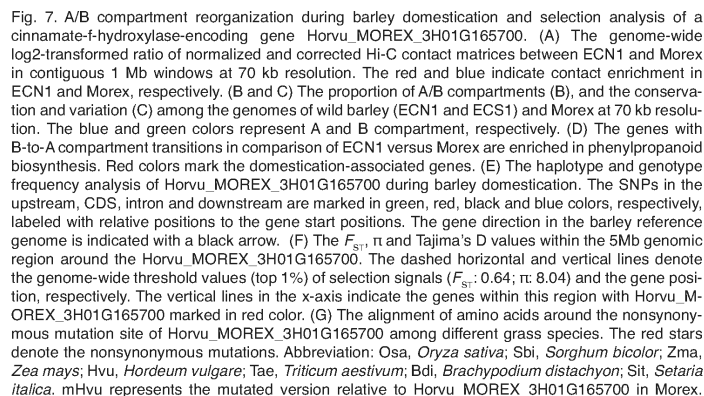


Fig. 6. Transcriptional profiles of the domestication-associated genes simultaneously identified by three methods and selection analysis of a GTPase-encoding gene Horvu_MOREX_7H01G519400. (A) Gene expression profiles. The values of Transcripts Per Million reads (TPM) are transformed by $\log_2(\text{TPM}+1)$. (B and C) The genetic variations (B) and genotype frequency (C) analysis of Horvu_MOREX_7H01G519400 during barley domestication. The SNPs in the upstream, CDS, intron and downstream to the gene start positions. The gene direction in the barley reference genome is indicated with a black arrow. (d) The F_{ST} and Tajima's D values within the 5Mb genomic region around Horvu_MOREX_7H01G519400. The dashed horizontal and vertical lines denote the genome-wide threshold values (top 1%) of selection signals (F_{ST} : 0.64; π : 8.04) and the gene position, respectively. The vertical lines in the x-axis indicate the genes within this region with Horvu_MOREX_7H01G519400 marked in red color.

Figure 6

See image above for figure legend.



See image above for figure legend.

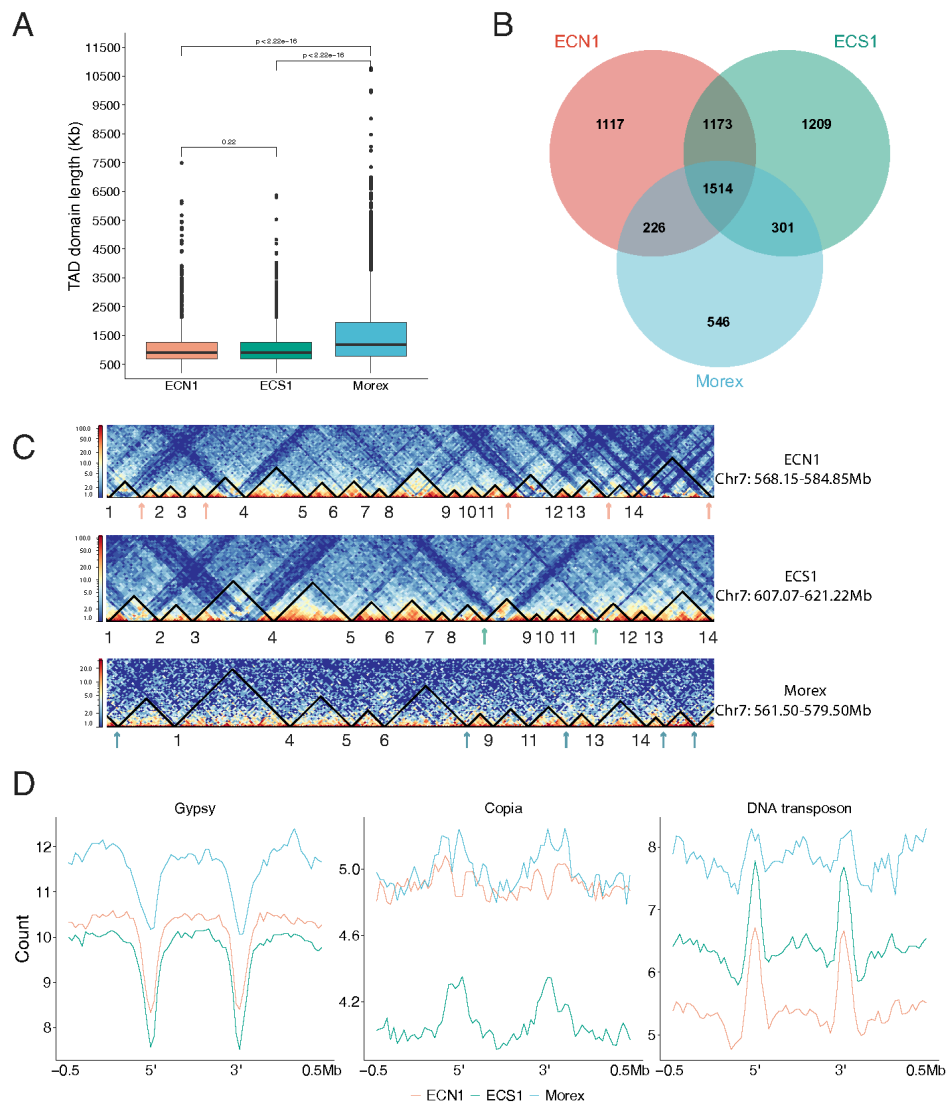


Fig. 8. The divergence of topologically associating domains (TADs) between wild and domesticated barley is associated with TE enrichment. Boxplots of TAD domain lengths (A) and the TAD boundary conservation (B) during barley domestication at 70 kb resolution. P values of Wilcoxon rank-sum test are indicated in (A). (C) The TAD structures in the syntenic blocks on chromosome 7 among the three genomes at 70 kb resolution. The log-transformed Hi-C contact matrices are shown. The conserved TAD boundaries among the three genomes are denoted by black numbers. The colored arrows indicate species-specific TAD boundaries. (D) The transposon distributions around TAD domains among the three barley genomes using 20 kb windows.

Figure 8

See image above for figure legend.

Supplementary Files

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

- [SupplementaryFigure.pdf](#)
- [SupplementaryTable.xlsx](#)