nature communications

Article

https://doi.org/10.1038/s41467-025-57932-1

# Constraint of accessible chromatins maps regulatory loci involved in maize speciation and domestication

Received: 10 September 2024

Accepted: 7 March 2025

Published online: 12 March 2025

Check for updates

Yuting Liu[1,6], Xiang Gao[1,5,6], Hongjun Liu[2], Xuerong Yang[2], Xiao Liu[3], Fang Xu[3], Yuzhi Zhu[1], Qingyun Li[1], Liangliang Huang[4], Fang Yang[1], Jinsheng Lai[4] & Junpeng Shi[1]

Comparative genomic studies can identify genes under evolutionary constraint or specialized for trait innovation. Growing evidence suggests that evolutionary constraint also acts on non-coding regulatory sequences, exerting significant impacts on fitness-related traits, although it has yet to be thoroughly explored in plants. Using the assay for transposase-accessible chromatin by sequencing (ATAC-seq), we profile over 80,000 maize accessible chromatin regions (ACRs), revealing that ACRs evolve faster than coding genes, with about one-third being maize-specific and regulating genes associated with speciation. We highlight the role of transposable elements (TEs) in driving intraspecific innovation of ACRs and identify hundreds of candidate ACRs potentially involved in transcriptional rewiring during maize domestication. Additionally, we demonstrate the importance of accessible chromatin in maintaining subgenome dominance and controlling complex trait variations. This study establishes a framework for analyzing the evolutionary trajectory of plant regulatory sequences and offers candidate loci for downstream exploration and application in maize breeding.

The intricate regulation of gene expression is crucial for executing complex biological processes in higher organisms. Maize, a classical genetic model in gene regulation, has been extensively studied since Dr. Barbara McClintock's groundbreaking discovery of transposable elements (TEs), commonly known as jumping genes, which cause the mosaic pigmentation seen on maize kernels[1]. Recent assembly and analysis of maize genomes have unveiled a highly sophisticated and hierarchical manner of gene regulation in maize, characterized by the prevalence of distal *cis*-regulatory elements (CREs)[2], particularly enhancers, that are likely attributed to the proliferated maize genome (~2.3 Gb) with over 85% DNAs composed of TEs. Meanwhile, genetic mapping efforts in maize unveiled an enrichment of quantitative trait loci (QTLs) within non-genic regulatory regions[3], including well-known loci such as *tb1*[4,5], *vgt1*[6], *KRN4*[7], *ZmCCT9*[8], *ZmCCT10*[9], and *UPA1*[10]. On a higher-dimensional scale, the widespread distal CREs in maize can orchestrate with transcription factors (TFs) and other trans-acting elements while simultaneously interacting with promoters or other gene-proximal regions to regulate gene expression through chromatin interactions[2]. These findings underscore the crucial role of regulatory sequences in controlling maize development and trait variations.

[1]State Key Laboratory of Biocontrol, Guangdong Provincial Key Laboratory of Plant Stress Biology, School of Agriculture and Biotechnology, The Shenzhen Campus of Sun Yat-sen University, Sun Yat-sen University, Shenzhen 518107, China. [2]State Key Laboratory of Wheat Improvement, College of Life Science, Shandong Agricultural University, Tai'an 271018, China. [3]The Key Laboratory of Plant Development and Environmental Adaption Biology, Ministry of Education, School of Life Sciences, Shandong University, Qingdao 266237, China. [4]College of Biotechnology and Agronomy, China Agricultural University, Beijing 100193, China. [5]Present address: Centre for Brain Connectome and Behavior, the Brain Cognition and Brain Disease Institute, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China. [6]These authors contributed equally: Yuting Liu, Xiang Gao.
e-mail: shijp6@mail.sysu.edu.cn

The progress of high-throughput chromatin profiling assays, such as ATAC-seq, CUT&Tag and DAP-seq technologies, has revolutionized the genome-wide identification of CREs in plants at base-pair resolution[11]. ATAC-seq utilizes engineered hyper-sensitive Tn5 transposons preloaded with sequencing adaptors to selectively target DNAs within accessible chromatin regions (ACRs)[12]. ACRs provide biomacromolecules physical access to DNAs and are usually colocalized with CREs and transcription factor binding sites (TFBSs), thus playing a critical role in transcriptional regulation[13]. In maize, a series of studies have employed ATAC-seq to profile ACRs, which typically encompass tens of megabases (Mbs) across major maize tissues including seedlings[14,15], roots[16], leaf[2,17], tassel[18] and ear[2,18]. Apart from these progresses in bulk tissues, ATAC-seq has also been continually optimized to work with the increasingly smaller sample input, even down to a single cell. A recent study utilized single-cell ATAC-seq to investigate the chromatin dynamics associated with cellular heterogeneity in a few maize tissues[19]. However, as a model species with substantially high tissue complexity, the existing ATAC-seq data in maize is undoubtedly unsaturated, and an in-depth characterization of the accessible chromatins in maize is still awaited.

In comparative genomics, it is well established that highly constrained DNA sequences shared among related species are essential for organismal survival and may have undergone stronger purifying selection. However, those that specifically exist in a single or a small group of species are usually under positive selection and responsible for innovative traits[20,21]. In the past few years, increasing evidence indicates that a considerable proportion of genome sequences including not only the coding regions but also massive CREs in the non-coding regions can be evolutionary constrained[20–23]. Since CREs are typically associated with epigenetic features such as accessible chromatin, active histone modifications, low DNA methylation, and unique binding of TFs[11], an emerging research field of evolutionary epigenomics has combined both epigenomic and comparative genomic approaches to study the constraint of DNAs associated with CREs[20,21,24]. In mammals, the Zoonomia project estimated that approximately half of the human CREs are evolutionarily constrained in 240 other mammal genomes[22]. Another work reported that approximately 80% and 50% of the evolutionary constrained DNA bases in primates are located outside protein-coding exons and have no annotation in human ENCODE data[21]. Relevant scenarios are also found in plants, where more than half of the highly constrained DNAs between diploid potato and other Solanaceae genomes occur in non-coding regions[23]. Similarly, several published works have utilized evolutionary constraint to identify conserved non-coding sequences (CNSs) in maize and other major crops[25–27], however they usually lacked enough ATAC-seq or other similar data to further classify potential CREs within these CNSs. With the technical progress of CRE annotation methods and the availability of multiple high-quality genome sequences in diverse Poaceae species, we can now generate a comprehensive map of regulatory loci in maize[2,19,28–33] and trace their evolutionary trajectory during approximately 100 million years of Poaceae genome evolution[34], revealing new insights into maize trait innovations compared to its wild ancestor and other grass species.

Here, we build a multi-tissue map of accessible chromatin in over ten major maize tissues and explore their interspecific and intraspecific constraint in multiple Poaceae and maize genomes. Unlike in primates, our analysis reveals strikingly low constraint (~1%) of ACRs in Poaceae, and highlights the major role of TEs in generating maize specific ACRs primarily occurring in regions distal to genes. By comparison, approximately half of ACRs are constrained in diverse maize genomes, exhibiting higher enrichment in trait-associated SNPs compared to other highly variable ACRs. We also uncover a set of ACRs likely associated with maize domestication, demonstrating their entanglement with protein-coding genes that either are artificially selected or differentially expressed between maize and wild teosintes. Our work establishes a framework for the evolutionary analysis of massive plant regulatory elements, and unveils some potentially important regulatory loci involved in maize speciation and domestication.

## Results

### A multi-tissue map of accessible chromatins in maize

We profiled the accessible chromatins for 12 major tissues or cell types in maize (Supplementary Data 1). Extending from the earlier data of ear[2,18], tassel[18], leaf[2], seedling[14,15], root[16] and mesophyll[17], we generated the ATAC-seq data for the embryo, endosperm, pericarp, shoot apical meristem (SAM), silk and stem (Fig. 1a). Each data had two biological replicates that exhibited high reproducibility, canonical enrichment near the transcription start sites (TSSs) of genes, and clustered well in both principal component (PCA) and heatmap analysis (Supplementary Fig. 1 **and** Supplementary Fig. 2). Therefore, the data from two biological replicates were combined for subsequent analysis. Around 26,743 to 39,561 ACRs (Fig. 1b) were identified in each tissue, covering approximately 0.67% to 1.0% of the B73 reference genome (Fig. 1b). As expected, ACRs were mainly distributed in euchromatin regions (Fig. 1c), aggregating into some hotspots conserved or that specifically existed in different tissues (Supplementary Fig. 3). We further classified ACRs into three categories according to their intersection with protein-coding genes[2]: genic ACRs (gACRs, overlapping ≥1 bp with genes, 36% to 52%), proximal ACR (pACRs, overlapping ≥1 bp with 2 kb regions flanking genes, 19% to 25%) and distal ACRs (dACRs, those other than gACRs and pACRs, 29% to 46%) (Fig. 1d). The proliferation of the maize genome has led to a significant number of dACR located in "gene deserts" regions (100 kb away from the nearest gene, Supplementary Fig. 4a). Interestingly, these ultra-distal ACRs showed high conservation across the tissues we profiled and were independently verified by maize single-cell ATAC-seq data[19] and ChIP-seq data[2] of various active histone modifications (Supplementary Fig. 4a and 4b). They were also found to be entwined with chromatin interactions, with some pinpointed to genes in 3D genomic data[2] (Supplementary Fig. 4b), suggesting the *cis*-regulatory role of maize ACRs in both local and distal manners.

In contrast to the maize genome that primarily composed of TEs (~85%), only ~25% of ACRs overlapped with TEs (Fig. 1e), suggesting the enrichment of ACRs for low-copy sequences in maize. Specifically, dACRs (~40%) had higher TE composition over pACRs (~23%) and gACRs (~15%). With respect to TE superfamilies, *Ty3* (*Gypsy*) accounted for over half of TEs in dACRs, while *Helitrons* enhanced substantially in pACRs and gACRs which was consistent with their strong preference for gene-rich regions in maize[35] (Supplementary Fig. 5a). It was worth noting that the insertion time of LTRs associated with ACRs seemed to be more ancient than the remaining non-association LTRs (Supplementary Fig. 5b). Unlike TEs, simple repeats have been reported to be enriched in CREs in humans[36] and also found to be overrepresented in maize ACRs (Supplementary Fig. 5c). Specifically, they exhibited a notable valley in the core region of ACRs while being elevated in the flanking regions, which was consistent with their presumed role in assisting the binding of TFs[36].

By integrating the data of PlantTFDB[37], JASPAR[38] and DAP-seq[2,39], the DNA binding motifs of 306 maize TFs from 36 families were curated (Supplementary Data 2). We discovered the motifs of 243 TFs within 50 bp sequences flanking the summit of ACRs, with each tissue having 157 to 191 TFs (Supplementary Data 3). Similar to a previous report in cotton[40], the motifs of BRR-BPC, C2H2, Dof, ERF and TCP families, which were fundamentally important for plant growth and development, were highly enriched in all profiled tissues (Fig. 1f and Supplementary Fig. 6). Specifically, the enrichment of SBP families was high in the ACRs of ear, tassel and SAM, consistent with one of their major roles in establishing meristem boundaries within inflorescence tissues[41]. Additionally, we found motifs from some WRKY family genes

https://doi.org/10.1038/s41467-025-57932-1

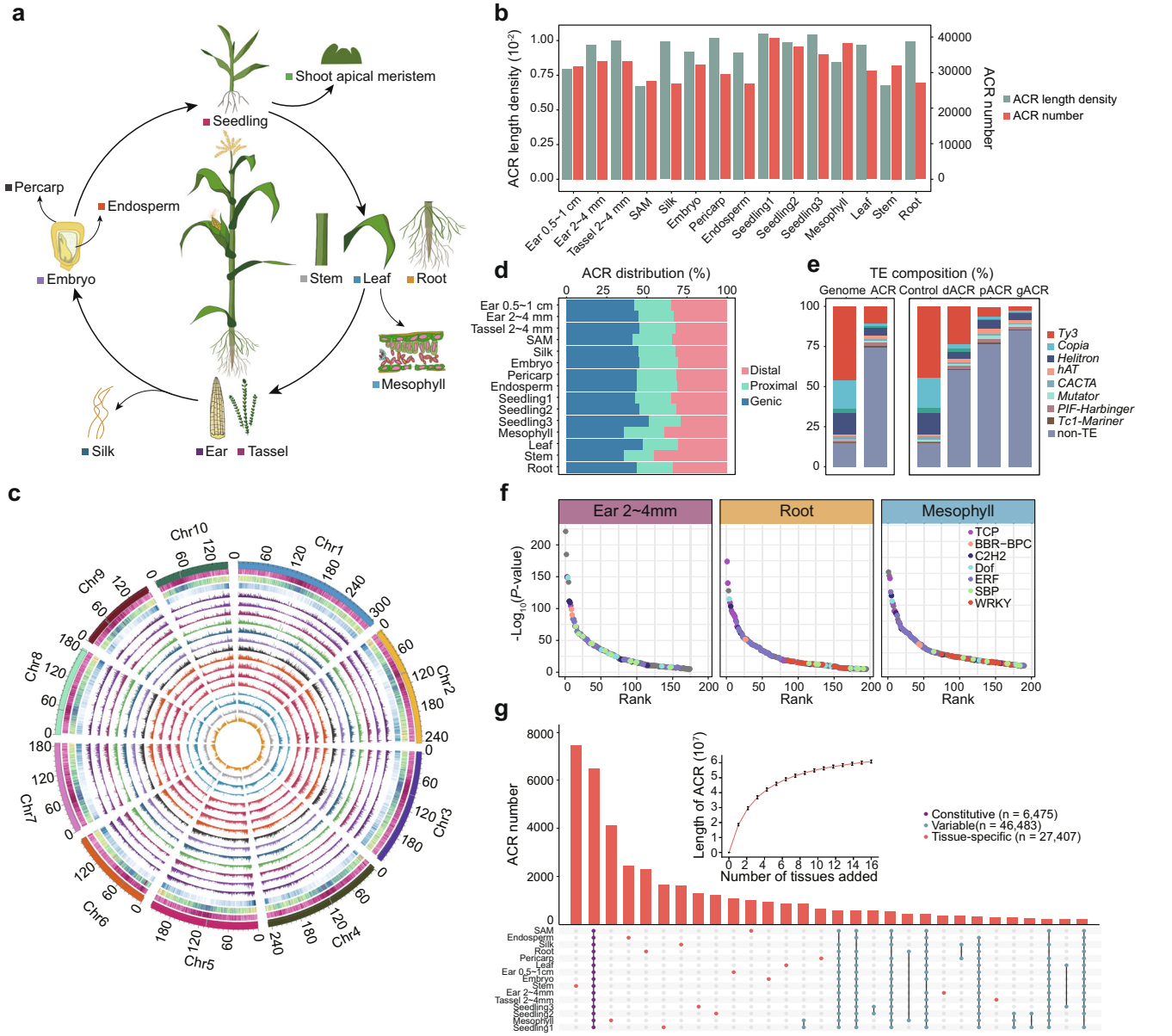


**Fig. 1 | A multi-tissue map of accessible chromatins in maize. a** Overview of the 12 major maize tissues we profiled in this study. Except for the ear and seedling, which had two and three ATAC-seq data, the other tissues had a single ATAC-seq data. We hereafter used sample to represent each of the 15 ATAC-seq data. **b** The number and length of ACRs identified in each sample (seedling1, seedling2 and seedling3 represent three different seedling samples). **c** Circos plot showing the chromosomal distribution of ACRs in each sample. Tracks from outer to the inner represented chromosome ideograms, retrotransposons, DNA transposons, gene density and ACR density of 15 samples, respectively. **d** The distribution of ACRs on different genomic features (genic, proximal and distal). **e** Bar plot showing the composition of different TE superfamilies in ACRs and random control regions. **f** The ranking of enriched motifs in ear (2 ~ 4 mm), root, and mesophyll. Different TF families are indicated by colored dots. The *P*-value of each motif was estimated using one-sided Fisher's exact test. **g** Upset plot showing the ACRs that conserved or specifically existed across 15 samples. A plot of saturation analysis by fitting the Michaelis-Menten equation was attached and the error bars represent the average value ± standard deviation (SD). Source data are provided as a Source Data file.

that were overrepresented in roots and mesophyll, a phenomenon consistent with their significantly enhanced transcription in similar tissues[42,43].

We further merged the ACRs of all tissues into 80,365 multi-tissue ACRs (~61.4 Mb, ~2.9% of the maize genome), accounting for 2.8 to 4.4-fold of the ACR length in a single tissue. Our estimation approached the maize single-cell ATAC-seq assays (~4%), which integrated data from 56,573 nuclei across six major tissues[19]. However, this ACR map was apparently not saturated, representing approximately 85% saturation according to the Michaelis-Menten equation analysis (Fig. 1g), indicating a need to further incorporate more specialized tissues or cell-types. ACRs were further classified into constitutive, variable and tissue-specific ACRs (Fig. 1g). The contrasting numbers of variable (n = 46,483) and constitutive ACRs (n = 6475) again suggested the high tissue specificity of accessible chromatins in maize. To further reveal the selective pressure on ACRs, we estimated the SNP density within and flanking ACRs using the data obtained from maize HapMap3[44]. As expected, the constitutive ACRs had the lowest SNP density in their central region (Supplementary Fig. 7), followed by variable and tissue-specific ACRs. This observation was in accordance with their conservation degree. Interestingly, the flanking regions of constitutive and variable ACRs showed a significantly higher SNP density than tissue-specific ACRs. In summary, we generated a map of over 80,000 ACRs in multiple

https://doi.org/10.1038/s41467-025-57932-1

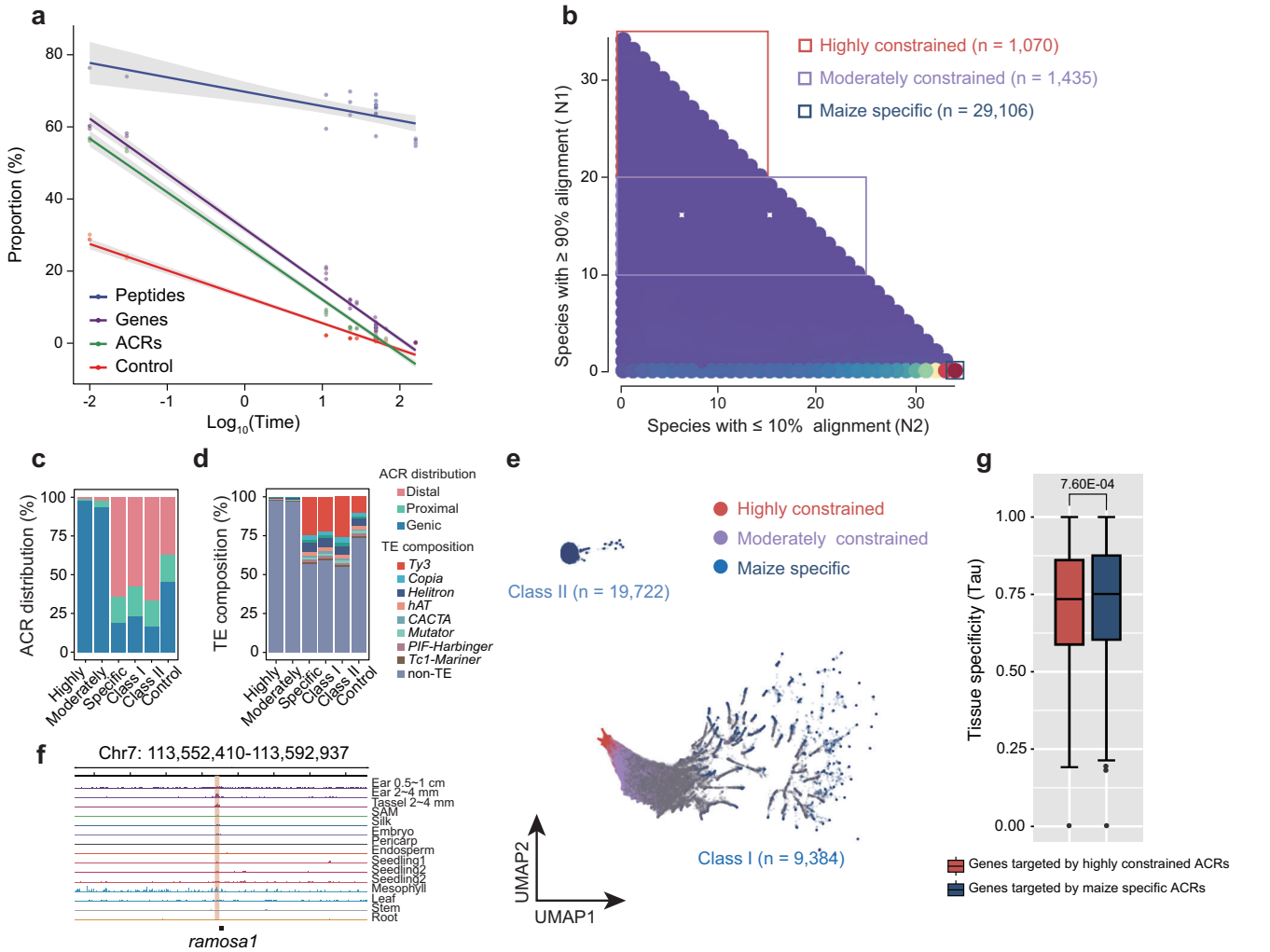


**Fig. 2 | Interspecific constraint of maize ACRs in grasses. a** The proportion of constrained peptides (blue), genes (purple), ACRs (green) and randomly selected sequences (red) with the increasing of pairwise divergence time (Mya) between maize and other non-maize genomes. The error bands indicate 95% confidence intervals. **b** Distribution of maize ACRs according to their constraint in other grass genomes. N1 denotes the number of grass species in which ≥ 90% of maize ACRs can be aligned. N2 denotes the number of grass species in which ≤ 10% of maize ACRs can be aligned. Three distinct groups (highly constrained, moderately constrained and maize specific) are highlighted in squares. **c** The distribution of genomic features of six groups including highly constrained, moderately constrained, maize specific, class I maize specific, class II maize specific ACRs and randomly selected ACRs as control. **d** Barplot showing the composition of different TE superfamilies in different categories of ACRs as shown in (**c**). **e** UMAP projection of all 80,365 ACRs based on their constraint level in each of the non-maize genomes. Each point is one ACR and the colors represent ACR groups. **f** The chromatin accessibility of *ra1* (*Zm00001d020430*) and its ~40 kb flanking region. The red region indicates a maize specific ACR. The structure of genes is shown in the bottom track. **g** Tissues specificity of genes targeted by highly constrained ACRs and maize-specific ACRs across 12 tissues. The *P*-values were calculated by two-sided Wilcoxon test. Tau was used as the index of tissue specificity. The index varied from 0 (consistently expressed genes) to 1 (strongest tissue-specificity of gene expression). The box denotes the maxima, 25th, median, 75th percentiles and minima, and the whiskers indicate the 1.5× interquartile range. Source data are provided as a Source Data file.

maize tissues, which allowed their evolutionary characterization in subsequent sections.

## Evolutionary constraint of maize ACRs in other grass genomes

Comparative genomic studies in plants have commonly characterized the conservation or innovation of protein-coding genes, yet the evolutionary constraint of more widespread regulatory sequences remains less investigated. Based on the multi-tissue ACRs, we employed a Cactus-based comparative approach (Supplementary Fig. 8) to trace their evolutionary trajectory in 34 Poaceae species and 4 outgroups (Supplementary Fig. 9 and Supplementary Data 4). To demonstrate the robustness of our method, we identified the regions in sorghum genome that were highly constrained (alignment coverage ≥ 90%) with maize ACRs, which could successfully reconstruct the two maize subgenomes (Supplementary Fig. 8b). When expanding to other Poaceae species, the number of highly constrained ACRs negatively correlated with their pairwise divergence (Fig. 2a). For instance, 7117, 3592 and 1421 ACRs were highly constrained in sorghum (diverged ~11.1 Mya), foxtail millet (~22.8 Mya) and rice (~49.1 Mya) respectively, exhibiting good consistency with their overall phylogeny (Fig. 2a and Supplementary Data 4). In general, the constraint of ACRs was situated between protein-coding sequences and randomly selected DNA sequences (Fig. 2a), suggesting an accelerated evolution of ACRs over protein-coding sequences in crops.

We then projected ACRs into a planar map (N1-N2, Fig. 2b) according to the numbers of highly constrained species (N1) and the number of species without constraint (N2, alignment coverage ≤ 10%). It was worth noting that only 1% ($n = 1,070$) of the maize ACRs were highly constrained in most Poaceae species (N1 ≥ 20) (Fig. 2b), in contrast to the high conservation in humans that approximately half of

CREs were constrained in primates. In addition, a substantial number of ACRs (~36.2%) specifically existed in maize (N2 ≥ 34) (Fig. 2b). These data implied an unprecedented high genetic diversity of regulatory sequences in grasses, which further prompted us to explore the driving forces behind maize specific ACRs. We found that a profound proportion (~65%) of maize specific ACRs were located in distal regions (Fig. 2c), while both highly and moderately constrained ACRs were exclusively located (> 90%) in genic regions. Meanwhile, nearly half of the sequences in maize specific ACRs were annotated as TEs (Fig. 2d), particularly *Ty3* elements, contrasting with the rare TE composition in highly and moderately constrained ACRs. These data highlighted the major role of TEs, mostly residing in regions distal to genes, which were involved in the innovation of maize specific ACRs.

To further characterize the interspecific constraint of ACRs, we applied the Uniform Manifold Approximation and Projection (UMAP) algorithm to all ACRs based on their alignment coverage in each of the non-maize genomes (Fig. 2e). The constrained (N1 ≥ 10) ACRs were largely clustered on the left panel, while the maize specific ACRs formed two distinct groups, one (class I, $n = 9384$) scattered on the right panel, and the other (class II, $n = 19{,}722$) forming an isolated island. We found almost no sequence homology of class II ACRs in other Poaceae genomes, compared with sparse homology of class I ACRs. Although class II ACRs exhibited higher affinity with TEs than class I ACRs, more than half of both class II and class I ACRs did not originate from TEs (Fig. 2d). These non-TE maize specific ACRs might result from sequence mutations that had accumulated due to genome redundancy after the whole genome duplication of maize.

Genes regulated by highly constrained or species-specific ACRs are likely to encode for traits of conservation or innovation, respectively[22]. Using our previously generated Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PETs) data[45], we identified these two different sets of genes ($n = 2583$ and 11,735), many of which could not be identified by the simple principle of proximity (Supplementary Fig. 10). Genes targeted by highly constrained ACRs were found to be enriched in photosynthesis pathways including NADPH: quinone reductase activity and response to light stimulus (Supplementary Fig. 11a). They also participated in the transmembrane transport of some polyols (Supplementary Fig. 11a), including mannitol, which served as a major product of photosynthesis[46], as well as xylose which could play a crucial role during the formation of plant cell walls[47,48] to mediate plant response to abiotic stresses (Supplementary Fig. 11a). On the other hand, genes ($n = 11{,}735$) regulated by maize specific ACRs were primarily enriched in plant immune or defense response pathways as was commonly observed for PAV genes[49,50] (Supplementary Fig. 11b). They include jasmonic acid and salicylic acid metabolic processes (Supplementary Fig. 11b), which potentially regulate plant resistance against pathogen invasion[51] during the continuous evolution of plant immune system[52]. Meanwhile, genes regulated by maize specific ACRs were also enriched in flower development and morphogenesis pathways (Supplementary Fig. 11b). One special example was *ramosa1* (*ra1*), a pivotal gene that regulates maize inflorescence branching architecture[53] and specifically expressed in immature tassel and ear in our profiled tissues (Supplementary Fig. 12). Notably, only one ACR was identified in both the gene body and 20-kb flanking regions of *ra1* (Fig. 2f), which was likely a decisive regulatory locus of *ra1*. This ACR was maize-specific since it had no constraint in sorghum and other Poaceae species. Therefore, it might function through *ra1* to contribute to the substantial changes in the inflorescence architecture of maize compared to other grasses.

We further investigated the expression profiles of the aforementioned two sets of genes to find more clues about their roles in maize speciation. Although genes regulated by highly constrained ACRs exhibited an overall higher expression level (Supplementary Fig. 13a), those regulated by maize specific ACRs, by contrast, exhibited significantly higher tissue specificity (Fig. 2g). This result was consistent with previous findings that tissue-specific expressed genes usually evolved faster[54] and were more likely to be affected by speciation and gene duplication[55]. We further utilized an independent dataset that applied the phylogenetic approach to estimate the age of genes in Poaceae[56], and found a higher proportion of young genes among the genes regulated by maize specific ACRs (Supplementary Fig. 13b). Collectively, we demonstrated the massive number of maize specific ACRs that primarily derived from the activity of TEs, and that genes they regulated might be involved in the innovation of some maize specific traits, such as inflorescence development and immune response.

## Intraspecific constraint of ACRs and their association with maize domestication

The recent release of diverse maize reference genomes[14,57,58] offers us an opportunity to further explore the intraspecific constraint of maize ACRs. By adopting the same comparative approach, we identified highly constrained ACRs in a total of 53 non-B73 maize genomes (Fig. 3a and Supplementary Data 5). We found that approximately half of the ACRs were fixed in nearly all maize genomes (Supplementary Fig. 14a). This ratio was comparable to the estimation of the interspecific constraint of CREs (~50%) in primates[21,22], and accorded with a previous estimation of genetic diversity in maize, which was comparable to or even higher than the divergence between human and chimpanzee[59]. On the other hand, only ~7% ($n = 5466$) of ACRs were constrained in fewer than 10 maize genomes (Fig. 3a), among which 337 were specifically found in B73. It seemed that the conservation of ACRs we estimated here was stronger than genes, since a pan-genome analysis of the maize NAM population had reported a higher degree of gene variability[14] (~70% were dispensable or private genes). To clarify, we simultaneously estimated the constraint of ACRs and genes in our panel. We found apparently stronger constraint of genes over ACRs, and their constraint were highly correlated ($r = 0.98$) with only a few exceptions (e.g., LH244 and PH207) that the constraint of ACRs was slightly stronger than genes (Supplementary Fig. 14b). Interestingly, ACRs that constrained intraspecifically exhibited quite different genomic properties compared to interspecifically constrained ACRs. For instance, TEs were common in intraspecifically constrained ACRs (Fig. 3c), while rare in interspecifically constrained ones. In addition, a considerable proportion of intraspecifically constrained ACRs were found in distal regions (Fig. 3b), while almost no distal ACRs was found in interspecifically constrained ones. These findings underscored the importance of historical TEs shared by the common maize ancestors, which might reside in regulatory regions and subsequently be fixed after maize diversification.

Genome comparison between wild teosinte and maize had uncovered hundreds to thousands of candidate genes associated with maize domestication[5,7]. Still, little is known regarding the regulatory sequences involved in this process. Based on the classical case of *tb1*, which was a gain-of-function *Hopscotch* insertion that was fixed in maize while being absent in teosintes[5], we hypothesized that ACRs that were highly constrained in most maize genomes but were absent in the majority of wild teosintes[60] (≥ 3 of 5 analyzed teosinte genomes (3 *Zea mays* ssp. *parviglumis*[61,62] and 2 *Z. mays* ssp. *mexicana*[62])), should play a pivotal role in rewiring transcriptional networks during maize domestication. Under this scenario, we could successfully validate the ACR that colocalized with *tb1*[63] (Supplementary Fig. 15a). On a genome-wide scale, we identified 856 ACRs potentially involved in maize domestication. Interestingly, compared with all highly constrained ACRs in maize, this small group of highly-constrained and domestication-associated ACRs showed uncanonically higher TE composition (Fig. 3c), as well as a profound ratio within distal non-genic regions. These results once again highlighted the prevalence and importance of TEs, particularly those located distal to genes during maize domestication. We were interested in the TFs involved in

https://doi.org/10.1038/s41467-025-57932-1

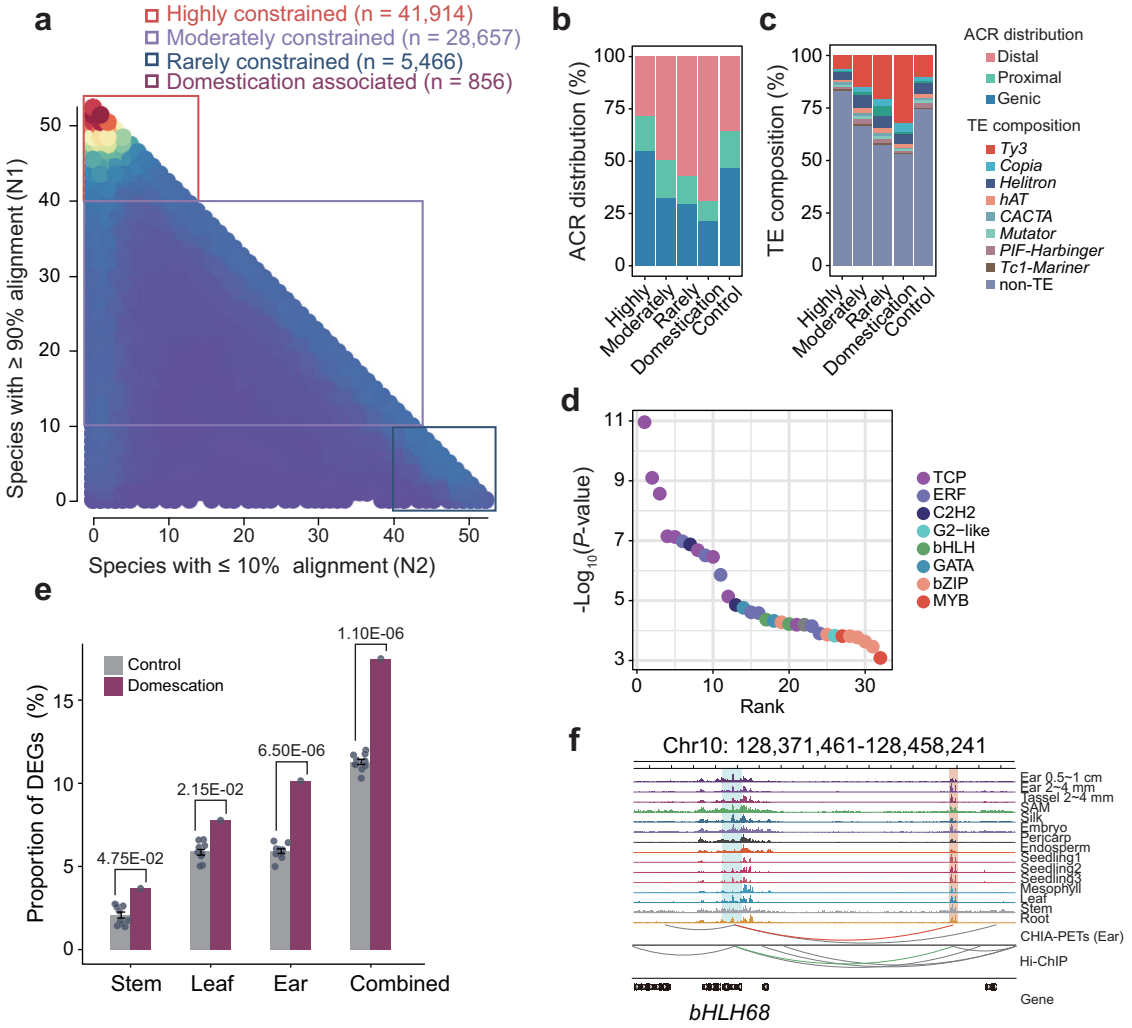


**Fig. 3 | The characteristics of intraspecifically constrained ACRs and domestication-associated ACRs. a** Distribution of ACRs in B73 according to their constraint in other non-B73 maize genomes. **b** The distribution of genomic features of five ACR groups including highly constrained, moderately constrained, weakly constrained, domestication-associated ACRs and randomly selected ACRs as control. **c** Barplot exhibiting the composition of different TE superfamilies in different ACR categories as shown in (**b**). **d** The ranking of enriched motifs in domestication-associated ACRs. Different TF families are depicted by colored dots. The *P*-value of each motif was estimated using one-sided Fisher's exact test. **e** The proportions of DEGs among genes targeted by domestication-associated ACRs and randomly selected control ACRs ($n = 856$). The *P*-values were calculated by two-sided Wilcoxon test. The error bars represent the average value ± SD. **f** The chromatin accessibility of *bHLH68* (*Zm00001d025752*) and its ~90 kb flanking region. The red region indicates domestication-associated ACR; the blue region indicates the target gene; the red and blue lines represent chromatin interactions between domestication-associated ACR and *bHLH68* identified by ChIA-PETs and HiChIP, respectively. The structure of genes is shown in the bottom track. Source data are provided as a Source Data file.

regulating these domestication associated ACRs and further predicted their TFBSs. We found significant motif enrichment of TCP family (Fig. 3d), which was accorded with the identity of master regulator gene *tb1* of maize domestication. In addition, we also discovered the enriched motifs of C2H2 family including *ramosa1* gene, as well as ERF and bZIP families that were fundamentally important during plant development.

To offer more functional clues into these domestication-associated ACRs, we again utilized the ChIA-PETs data to find a total of 463 genes that entangled with these ACRs through chromatin interaction. Notably, we noticed significant enrichment of these genes in another independent domestication gene list (Supplementary Fig. 15b), which was identified by scanning SNPs between teosinte and maize populations[64]. Moreover, among the differentially expressed genes (DEGs) identified in three major tissues between teosintes and maize[65], we found a significantly higher ratio (17.5%, Supplementary Data 6) of DEGs within these interacting genes over randomly selected genes without interactions (12.4%, Fig. 3e). As an example, we identified an ACR located ~51 kb upstream of *bHLH68*, also known as *phytochrome-interacting factor1* or *ZmPIF1*, which could play versatile roles in regulating the growth, development and drought tolerance of maize[66]. Chromatin interactions between this ACR and *bHLH68* can be found not only in our ChIA-PETs data but also in other 3D genomic datasets (Fig. 3f), and this ACR was fixed in multiple maize lines while absent in teosintes, which correlated well with its up-regulation in maize over teosintes (Supplementary Fig. 12). A similar case was also found for a domestication-associated ACR and its interacting gene *cmu2* (Supplementary Fig. 15c), which encoded a chorismate mutase with elevated expression in maize ear and leaf and might be involved in plant pathogen infection resistance[67] (Supplementary Fig. 12). In summary, our results demonstrated the great promise of analyzing evolutionary constraint of ACRs to identify candidate regulatory loci associated with crop domestication and subsequent adaptation processes.

### Constraint and dominance of ACRs between two maize subgenomes

It is well documented that genes in one maize subgenome always dominated expression over genes in the other[68], a phenomenon known as subgenome dominance. However, it has been less investigated for chromatin accessibility. Using 3719 high-quality genes pairs with exact 2:1 synteny to sorghum genes (Supplementary Fig. 16), we reconstructed two maize subgenomes (m1, 1,181 Mb; m2, 726 Mb) and classified ACRs into m1 (n = 47,666) and m2 (n = 28,442), respectively (Supplementary Data 7). Notably, the numbers of highly constrained ACRs in m1 (n = 4647, 9.7%) and m2 (n = 2326, 8.1%) when compared to sorghum were higher than the highly constrained ACRs when mutually compared between m1 and m2 (-1200, Fig. 4a, b), a phenomenon accorded with the well-known reciprocal subgenome fractionation observed for genes[68]. We then characterized key metrics of ACRs between the two subgenomes, and observed that both the number and density of ACRs in m1 were higher than in m2, although their overall differences were subtle (Fig. 4c and Supplementary Fig. 17a). Accordingly, the transcriptional differentiation was also weak when profiled across the two subgenomes (Supplementary Fig. 18a). However, when we focused on these 3,719 pairs of syntenic genes, the dominance of both chromatin accessibility (Fig. 4d and Supplementary Fig. 17b-c) and gene transcription was significantly enhanced (Supplementary Fig. 18b). Meanwhile, the chromatin accessibility of these conserved regions was also much stronger than the remaining less-conserved regions (Fig. 4c, d). These results suggested the importance of chromatin accessibility in differentiating duplicated gene loci to introduce subsequent differentiation of gene expression[69].

To further explore accessible chromatin in specific duplicated loci, we focused on *teosinte branched1* (*tb1*, m1) and its m2 paralog *tcp6*, both of which encoded the TCP family transcription factors. Although gene density and chromatin accessibility were fairly high in up- and downstream regions of *tcp6*, no ACRs were identified in its gene body and 10-kb flanking regions (Fig. 4f), which was consistent with its silenced expression in major maize tissues (Supplementary Fig. 19a). In contrast, *tb1* was exclusively activated in ear, which might be attributed to the functional *Hopscotch* insertion that showed colocalization with two independent ACRs in our data (Fig. 4f). We also investigated other two syntenic gene pairs, *ub2* (m1) and *ub3* (m2), *gt1* (m1) and *Zm00001d048172* (m2), both were crucial in regulating plant architecture[70,71]. Cases of *ub3* and *gt1* were similar to *tb1* that their causative sequence insertions, named *KRN4*[7] and *prol1.1*[71], exhibited accessible chromatin in most of the profiled tissues (Supplementary Fig. 19b and 19c). However, no parallel functional insertion was found near *ub2* and *Zm00001d048172*, and consequently their transcription was weakened or completely silenced (Supplementary Fig. 19a). These data suggested that the insertion of subgenome-specific functional sequences, mostly TEs, might be one of the driving forces underlying the subgenome differentiation of chromatin accessibility and gene expression.

We next sought to decipher the divergence of TFBSs between the two maize subgenomes. Unlike genes and ACRs that exhibited strong fractionation[68] (Fig. 4a), approximately 80% of the TFBSs (69.1% to 86.3% across the profiled tissues) could be identified in both subgenomes (Fig. 4e and Supplementary Fig. 20a). However, the divergence of TFBSs enhanced in both promoter and distal ACRs (Fig. 4e and Supplementary Fig. 20b and 20c), e.g., only one-third of the TFBSs in distal ACRs of seedling were shared between two subgenomes. Therefore, the strong conservation of TFBSs between two subgenomes might be attributed to the complementation of ACRs from different functional regions. Additionally, we found a higher number of subgenome-specific TFBS in m1 than m2, which was accorded with the general principal of m1 dominance over m2 (Fig. 4e and Supplementary Fig. 20a-c). For instance, the TFBSs of the GLK family, which regulates the earliest photosynthesis steps by activating genes encoding chloroplast-localized and photosynthesis-related proteins[72], were exclusively found in m1 (Supplementary Fig. 21). Interestingly, the genes that encoded these GLK TFs also existed in m1, indicating their regulatory functions that preferentially exerted within the same subgenome. These results indicated that the regulomes of the two maize subgenome were largely shared except for a few TF families that specifically functioned within a single subgenome to precisely regulate the development of particular tissues.

### Trait-associated SNPs are enriched in ACRs that highly constrained in maize

In species with complex genomes, such as maize[29] and humans[73], trait-associated SNPs (TASs) identified by genome-wide association studies (GWASs) often fall into non-coding regulatory regions[73]. We anticipated a better annotation of maize TASs using our ACR map. Therefore, we first performed a GWAS of 15 complex traits in a maize-associated mapping panel ($n = 260$) that was genotyped across HapMap3 SNPs, resulting in 110 to 363 TASs (minor allele frequency > 0.05, $P$-value < $10^{-5}$) for each trait (Supplementary Fig. 22 and Supplementary Table 1). We then integrated all TASs ($n = 2984$) to improve their statistical power when assessed with ACRs. As expected, TASs were overrepresented in ACRs with their density being 1.64 (cob-weight) to 6.49-fold (leaf-length) higher over the remaining non-ACR regions (Fig. 5a). In addition, the overall density of TASs in m1 ACRs was higher than in m2 ACRs (Fig. 5b), indicating a predominant role of the m1 subgenome in the genetic regulation of maize complex traits. Interestingly, although all the traits we investigated were believed to be classical quantitative traits, six out of them (for example, upper-leaf-angle and days-to-tassel) exhibited non-canonical results with higher TAS densities in the ACRs of the m2 subgenome (Fig. 5b). The existence of hotspots with an excessive number of TASs (Supplementary Fig. 23a and **b**), which usually entangled with a strong LD (linkage disequilibrium) and can be mapped within the m1 or m2 subgenomes for different traits, might provide one of the appropriate explanations for these observations.

Research in humans has revealed that TASs of disease-related or other complex traits are more likely to be enriched in interspecific constrained CREs[21,22]. However, due to low and limited number of interspecifically constrained ACRs in Poaceae (Fig. 2b), the power to investigate maize TASs in ACRs with a varying degree of interspecific constraint was statistically weak. However, we observed that the TASs were overrepresented in ACRs with higher intraspecific constraint (Fig. 5c). Specifically, the density of TASs in ACRs that highly constrained in most maize genomes exhibited 1.4-fold increase compared with ACRs with weak constraint in maize. We further verified this result (Supplementary Fig. 24) by using another independent TASs dataset (n = 2360) that also mapped on maize B73 V4 genome and tested in an ultra-large maize population with more than 1000 lines[74]. These findings suggested that the complex traits in maize are primarily driven by the genetic variants in ACRs, especially for those fixed in diverse maize populations.

We finally attempted to annotate some GWAS loci that resided in ACRs. As an illustration, we identified a lead SNP (chr3: 215,791,183) that significantly associated with both days-to-silk and days-to-tassel, which fell into a non-coding region that was ~5.3 kb away from the nearest gene. Notably, this lead SNP colocalized with an intraspecifically highly constrained ACR accessible in three reproductive tissues (ear, tassel, and SAM) (Fig. 5d), suggesting its potential role in regulating the genes involved in flowering. Further incorporation of 3D chromatin interaction data, such as the HiChIP data (Fig. 5d), could further pinpoint to a list of candidate genes likely interacting with this ACR, although the exact function of these GWAS loci and interacting genes need further experimental validations. To end up our work, we integrated the multi-tissue ATAC-seq data with our recently established HEMU online database (https://shijunpenglab.com/HEMUdb/

https://doi.org/10.1038/s41467-025-57932-1

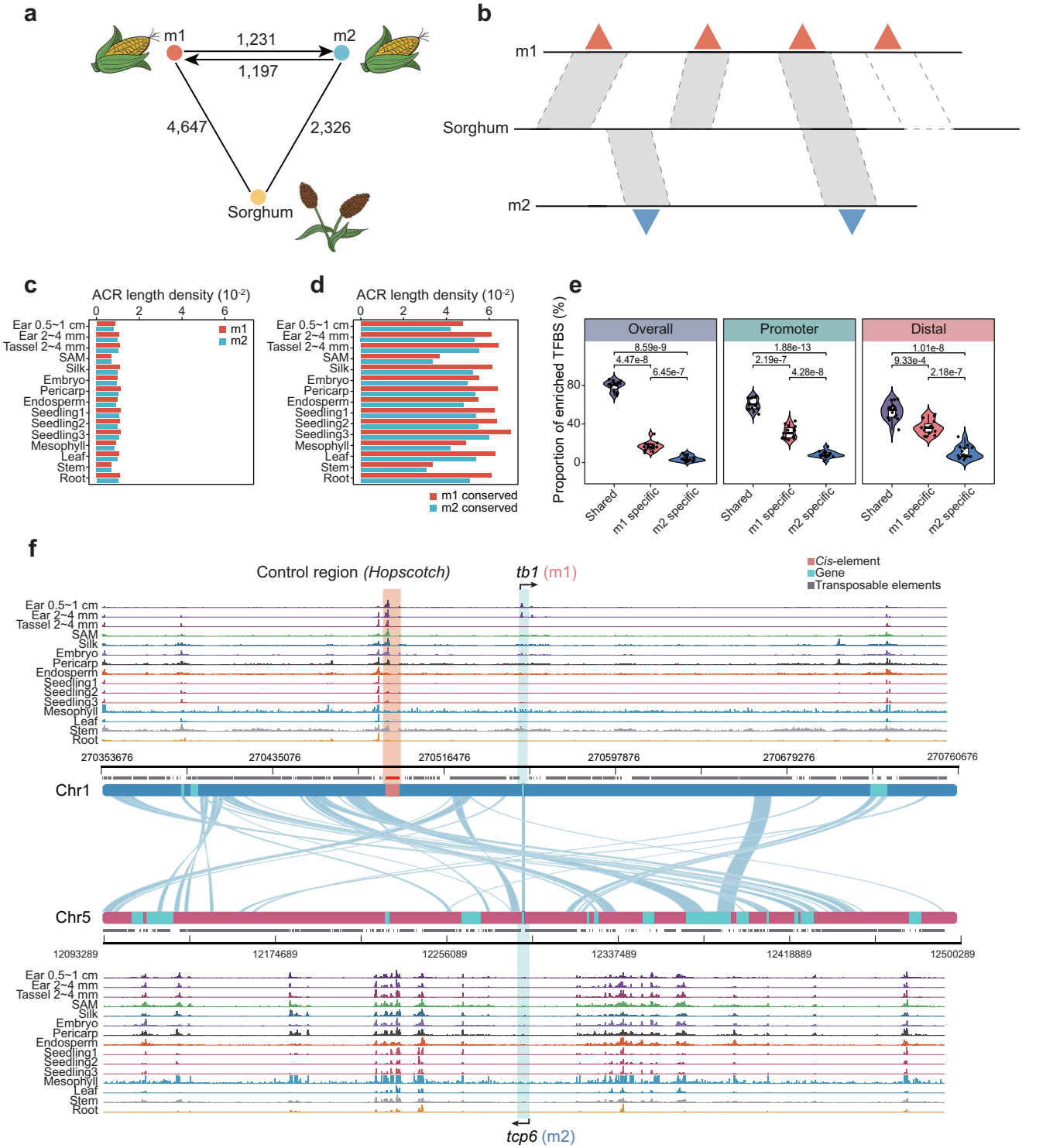


**Fig. 4 | Characteristics of ACRs between two maize subgenomes. a** The number of constrained ACR on m1 and m2 when compared to sorghum and when mutually compared. **b** A schematic of reciprocal subgenome sequence loss happened on ACRs. Straight lines represent genome sequences; gray dashed boxes represent regions of collinearity between genomes; and triangles represent ACRs on genomes. **c** ACR length density normalized by the lengths of m1 (red) and m2 (blue) across 15 samples. **d** ACR length density of highly conserved region normalized by the lengths of highly conserved regions of m1 (red) and m2 (blue) across 15 samples. **e** The proportion of enriched TFBSs that shared between two subgenomes and specifically existed in m1 or m2 subgenome ($n = 15$). Except for the TFBSs that annotated across the two subgenomes, we also independently annotated TFBSs in promoter and distal ACRs, respectively. The box denotes the maxima, 25th, median, 75th percentiles and minima, and the whiskers indicate the 1.5× interquartile range. The *P*-values were calculated by two-sided Wilcoxon test. **f** An overview of ACRs across a ~400 kb region between *tb1* and its m2 syntenic gene *tcp6*. Red region indicated the reported functional *Hopscotch* region; the blue box indicates the genes; the grey bar indicates the transposable elements. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-025-57932-1

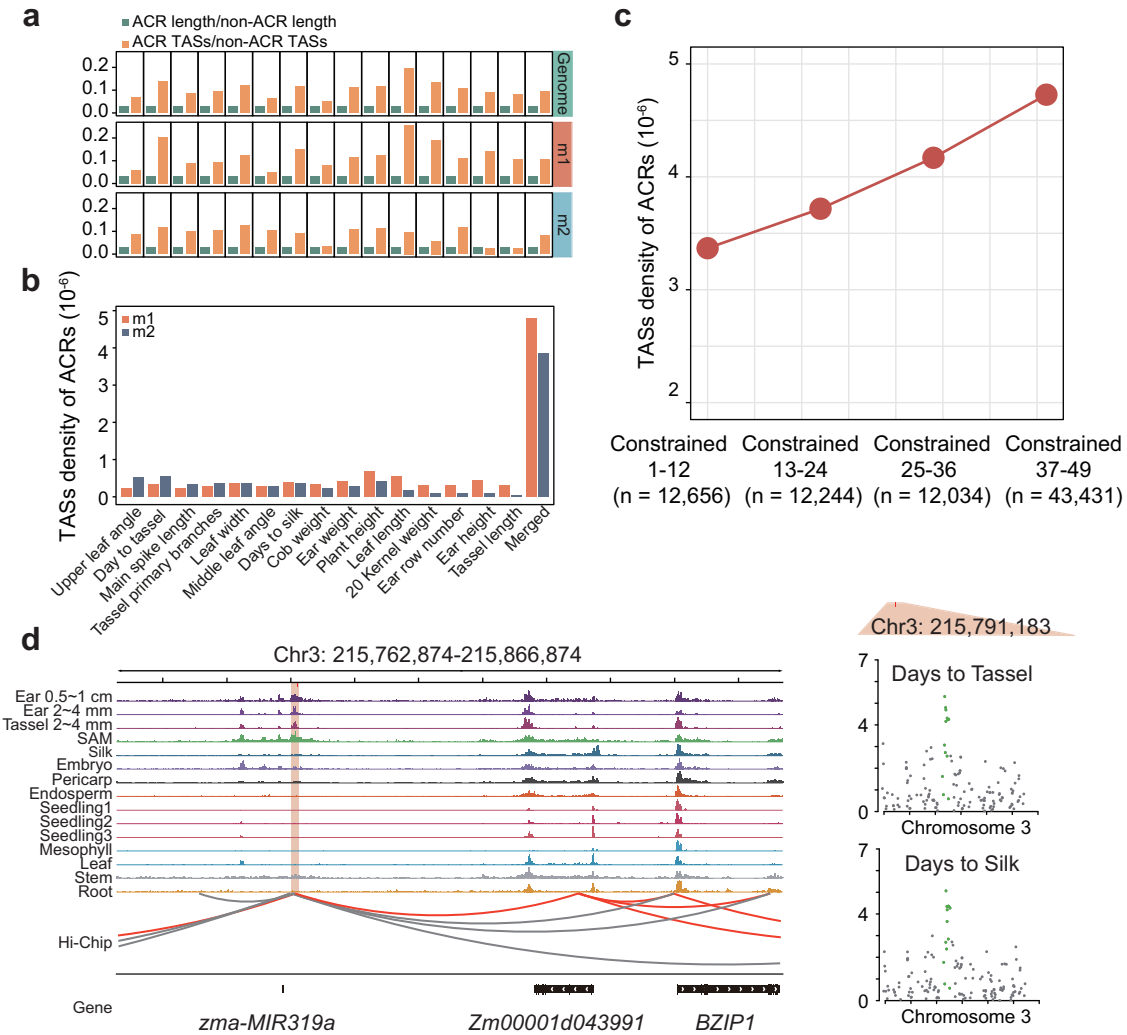


**Fig. 5 | The enrichment of TASs in ACRs. a** The ratio of TASs_in_ACRs/TAS_in_non-ACRs and ACR length/non-ACR length across the whole genome, m1 and m2 subgenomes for 15 complex traits. **b** The TASs density in ACRs of in m1 and m2 subgenome across 15 traits and those that merged into a non-redundant TASs. **c** The TASs density in 4 different degrees of intraspecifically constrained ACRs. **d** An overview of ATAC-seq signals and HiChIP interactions (red indicates H3K4me3 and grey indicates H3K27m3) of a non-coding GWAS locus. The position of lead SNP is indicated by light red shading. Regional Manhattan plots of this significant SNP associations for Days_to_tassel and Days_to_silk traits are shown in the right panel. Source data are provided as a Source Data file.

databrowse)[75] to facilitate easy visualization and analysis of any loci that recorded by B73 V4 genome coordinates.

## Discussion

With the rapid advancement in plant genomics, we are on a highway to enter the post-genomic era, with primary research goals progressively shifting from gene-centric studies to encompass a wider spectrum of genomic regulatory sequences. In this work, we generated a comprehensive map of accessible chromatin by integrating the ATAC-seq data of 12 major maize tissues. Although ATAC-seq has been commonly used by many studies on maize to explore the tissue-specificity of accessible chromatins[2,14,16–19], most of them are carried out on a single or relatively few tissues. They include the recently published single-cell atlas of ACRs across six maize tissues, highlighting the strong association between GWAS SNPs and tissue/cell-type specificity of ACRs. To our knowledge, the map in our study might be one of the most comprehensive ACR maps in plants to date[76], although it is still far from saturation as we await the incorporation of more ATAC-seq data from distinct tissues, cell types and other non-B73 maize lines[14,77]. We anticipate the growth of ATAC-seq data in maize and other plant species in the forthcoming years[32,33], especially through the application of the single-cell ATAC-seq technology that has shown its unique advantages in decoding cell heterogeneity and the flexibility in obtaining chromatin accessibility of bulk tissues when data from tens of thousands of cells are integrated. Meanwhile, the recently developed long-read Fiber-seq is another promising approach[78] for profiling chromatin accessibility in maize and other crops with considerably high ratios of TEs. We also anticipate the continuous development of cutting-edge artificial intelligence algorithms to predict functional CREs in complex crop genomes with the integration of multi-omics data[79–81], and leverage these epigenetic information to dissect complex traits and improve breeding design.

We performed large-scale comparative genomic analysis to estimate the inter- and intraspecific evolutionary constraint of maize ACRs. Although it is well-documented that highly constrained DNA sequences in a group of related species are usually functional, there are also many circumstances where loss-of-function mutations can occur in these highly constrained regions, for example, large-effect SNPs that can result in pseudogenes. Therefore, ATAC-seq data of other crop species are required to unveil the true extent of chromatin accessibility for any DNA sequence constrained within an ACR in maize. In addition, by analyzing the intraspecific constraint, we also identified several

hundreds of candidate ACRs that are fixed in most maize lines while absent in teosintes, which were likely associated with maize domestication. However, compared with these gain-of-function regulatory elements, the effect of the genetic bottleneck happened during crop domestication are more usually to result in loss of beneficial alleles in their ancestors. Therefore, generating similar regulatory maps in wild teosintes[82] will help shed new light on these untapped regulatory loci that should also play an important role in the maize domestication process. For high-resolution TF binding motifs or CREs, which typically span tens of base pairs within ACRs, it requires more meticulously designed methods to investigate their evolutionary constraint in order to minimize inherent alignment bias. We also demonstrated the power of utilizing ATAC-seq data to annotate the genetic loci discovered by GWAS in maize, although experimental tools, e.g., transient transcription activation assay or genome editing is required for downstream verification. In summary, our study expands our knowledge of plant evolutionary epigenomics, and offers a wealthy data resource for the excavation and utilization to enhance the genetic improvement of maize.

## Methods

### Plant materials

The seeds of maize B73 inbred line were planted at the experimental station of Shandong Agricultural University, Tai'an, China in 2021. After 12 days after self-pollination, the embryo, endosperm and pericarp from roughly 50 seeds were separated, then freezen quickly in liquid nitrogen. SAM tissues were dissected from 62-week-old field-grown plants and immediately frozen in liquid nitrogen.

### Library construction and ATAC sequencing

Plant tissues were ground into powder in liquid nitrogen, then incubated with lysis buffer at 4 °C for 10 min. After filtration with 40 μm cell strainer, cell solution was transferred to the top 60% percoll solution and centrifuged at 1800 xg for 20 min. The intermediate phase of the centrifuged liquid was extracted and then diluted with 10 mL RSB buffer. Cell nuclei pellets were harvested by centrifuge at 1,000 xg for 10 min and washed with RSB buffer once time. Approximate 50,000 nucleus were added to transposition reaction solution to perform tagmentation. Tn5 transposed DNA were purified by AMPure DNA magnetic beads, then amplified by PCR. The final qualified ATAC-seq libraries were sequenced on the Illumina Novo-seq platform (San Diego, CA, United States) with PE150 mode.

### ATAC-seq analysis

ATAC-seq raw data was processed using ATAC-pipe pipeline[83]. The quality of mapping reads was evaluated using the "--MappingQC" module in ATAC-pipe. After trimming adapter sequences using the script in ATAC-pipe, reads were mapped to *Z. mays* AGPv4 reference genome using Bowtie2 (v.2.3.0)[84]. The PCR duplicates were removed using the MarkDuplicates function from Picard tools (v.2.9.2; http://broadinstitute.github.io/picard/). The mapped reads were then shifted +4/-5bp depending on the strand of the reads, so that the first base of each mapped read can represent the Tn5 cleavage position. All mapped reads were further extended to 50 bp centered by the cleavage position. Reads mapping to highly repeated regions, chromosome M and P, and unanchored contigs were removed. Two biological replicates from the same sample were merged using the "--Merge" module in the ATAC-pipe.

### Sample clustering and PCA

The clustering among all samples was performed by the "dba.plotHeatmap" function of DiffBind (v.3.4)[85] with default parameters. The PCA of all samples was performed using the "dba.plotPCA" function of DiffBind (v.3.4) with default parameters.

### Transcriptome data analysis

Firstly, FASTP (v.0.20.1)[86] was used to trim adaptors and filter out low quality reads. Trimmed reads were further mapped to the *Z. mays* AGPv4 reference genome by HISAT2 (v.2.2.1)[87] with default parameters. Alignments were subsequently sorted and converted into bam files using Samtools (v.1.12)[88]. Gene expression was estimated by StringTie (v.2.1.5)[89] in the forms of both fragments per kilobase of transcript per million mapped reads (FPKM) and transcripts per kilobase million reads (TPM). Genes with FPKM or TPM > 1 were considered to be expressed in the corresponding samples.

### ACR identification

MACS2[90] was used to call peaks by the "--PeakCalling" module in ATAC-pipe with the parameters of "macs2 callpeak -f BED -g 2.1e + 9 -q 0.01 --nomodel --shift 0". Tn5 integration site density (insertions per kb) was estimated for each peak, as well as the matched and randomized control regions (established using bedtools (v.2.26.0)[91], shuffled specifically excluding ACRs from the randomized selection space). ACRs with Tn5 integration site density less than the 99.9% upper quantile of randomized control regions were removed according to the published method[14].

### Annotation of TEs and their association with ACRs

Repetitive elements were identified using the Extensive de novo TE annotation (EDTA, v.1.9.6)[92] pipeline with the parameters "--species Maize --anno 1". Next, the unclassified transposons from EDTA was re-annotated using DeepTEs[93]. Then, TE composition in ACRs was calculated by intersectBed (v.2.29.2)[91] with the parameters "intersectBed -wo".

The insertion time of intact LTRs associated with ACRs (50% TE sequence overlap with ACR) was calculated. Firstly, intact LTRs were extracted by two long terminal repeat sequences and mutually aligned by Muscle[94]. The distance *K* with Kimura Two-Parameter approach between LTRs was calculated by distmat program in EMBOSS (v6.6.0)[95]. The activity of each LTR (*T*) was calculated by the following formular:

$$T = K/(2 \times r) \quad (1)$$

where "*r*" refers to a general substitution rate of $1.3 \times 10^{-8}$ per site per year in the grass family[96]. RepeatMasker was used to annotate simple repeats in the ACR with the parameters "-no_is -norna".

### TF annotation and motif analysis

TF genes (n = 2,361) in the B73 V4 genome were annotated using Transcription Factor Prediction function of PlantTFDB[37] (http://planttfdb.gao-lab.org/). A total of 259 and 45 TF motif models were downloaded from PlantTFDB and JASPAR, respectively. Additional 12 TF motif models were detected via de novo motif identification using the discriminative motif discovery workflow of MEME-ChIP (v5.0.2) with default settings based on DAP-seq data[2,39]. The motifs of 306 TF genes from 36 families were curated. Finally, the enriched motifs in the interest region of ACRs were identified using the AME v.5.0.2 tool of MEME Suite with the parameters; "--scoring avg --method fisher --hit-lo-fraction 0.25 --evalue-report-threshold 10.0 --control".

### Evolutionary constraint analysis of ACRs

To measure the sequence constraint of ACRs, Progressive Cactus (v2.6.0)[97] was used to perform pairwise alignment of maize with other genomes, with the parameters "cactus./js_grass./B73.othergenome.txt./ B73.othergenome.hal --consMemory 320 G --binariesMode local --stats". After obtaining the hal file, the hal-to-chains function of Cactus was utilized to convert into chain file with the parameters "cactus-hal2chains./js_grass./ B73.othergenome..hal chain/ --refGenome othergenome". Then, we followed the published code to

employed the LiftOver[98] software to determine the conservation of ACR sequences from B73 genome in other genomes[22]. Alignment ≥ 90% was considered as constrained ACRs (see Supplementary Fig. 7a). Phylogenetic relationships among the 34 Poaceae species and 4 outgroups were obtained from TimeTree[99].

### UMAP analysis on ACRs

For each ACRs, the LiftOver software was used to get the percentages of aligned positions in the other 34 genomes. The resulting matrix of ACRs by 34 genomes was then used as input to run UMAP[100] with default parameters.

### GO analysis

GO enrichment analysis was performed using GOATOOLS[101] with *Z. mays* AGPv4 GO annotation and go-basic.obo (http://current.geneontology.org/ontology/go-basic.obo). The module "find_enrichment.py" with Fisher's exact test was used.

### Analysis of differentially expressed genes

The published RNA-seq data between 9 teosintes and 6 maize inbred lines (generated by John Doebley's lab[65], NCBI SRA accession number of SRP047528) from ear, leaf and stem tissues were downloaded. For the identification of differentially expressed genes between teosintes and maize inbred lines, we treated the bam files from teosintes (n = 18, 2 replicates from each of 9 lines) and maize inbred lines (n = 12, 2 replicates from each of 6 lines) as replicates, then used Cuffdiff[102] to calculate the FPKM in teosinte and maize. Genes with the fold-change of FPKM > 2 and $q$-value < 0.05 were identified as DEGs between teosintes and maize.

### Maize two subgenome identification

The two subgenome of *Z. mays* AGPv4 were identified according to our published pipeline[103]. Briefly, we used the Synmap pipeline in CoGe (https://genomevolution.org/CoGe/SynMap.pl). We used last to blast[104] the CDS sequences between maize and sorghum, then detected syntenic blocks with DAGchainer[105] with options -D 20 -A 5. QuotaAlign[106] was further used to merge adjacent syntenic blocks. The syntenic depth was set to 2:1 for maize and sorghum comparison, and the overlapped distance was set to 40 to permit overlapped syntenic regions. Fractionation bias was applied to determine subgenome organization in maize compared with sorghum. For each pair of chromosomes, the copy with a greater number of unique genes was assigned to the maize1 subgenome, whereas the pair with fewer uniquely retained genes was assigned to the maize2 subgenome. The CodeML utility in the PAML[107] software package was used to calculate the non-synonymous (*Ka*) and synonymous (*Ks*) rates between orthologous genes.

### High-quality syntenic gene pairs between the two maize subgenomes

According to the above COGE output file of QuotaAlign results, the gene pairs in maize having 2:1 relationship with sorghum genes were extracted, and among those with $Ks > 2$ were further removed (potential pre-grass homeologs). Finally, remaining gene pairs were assigned to the m1 and m2 subgenomes based on the result of subgenome classification.

### Genome wide association studies (GWAS)

Plink[108] was used to convert vcf files to bed files. Then the sites with missrate > 0.5 or MAF < 0.05 were filter out. GCTA[109] was used to perform PCA. The kinship matrix was obtained using emmax-kin for calculating genetic relatedness. EMMAX[110] was used to conduct GWAS analysis and correcte for population structure by including the top 20 principal components (to account for population structure) and the kinship matrix (to correct for genetic relatedness and other fixed effects).

### Reporting summary

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

## Data availability

All the previously published ATAC-seq data that used in this study were summarized in the Supplementary Data 1. All the newly generated ATAC-seq data in this study were deposited into NCBI SRA under BioProject ID PRJNA1048297. Source data are provided with this paper.

## Code availability

All codes related to analysis is available at Github [https://github.com/ShiJunpenglab/Inter-and-intraspecific-constraint-of-maize-ACR].

## References


1. Wessler, S. R. Phenotypic diversity mediated by the maize transposable elements Ac and Spm. *Science* **242**, 399–405 (1988).
2. Ricci, W. A. et al. Widespread long-range cis-regulatory elements in the maize genome (vol 5, pg 1237, 2019). *Nat. Plants* **6**, 328–328 (2020).
3. Liang, Y., Liu, H. J., Yan, J. & Tian, F. Natural variation in crops: realized understanding, continuing promise. *Annu Rev. Plant Biol.* **72**, 357–385 (2021).
4. Clark, R. M., Wagler, T. N., Quijada, P. & Doebley, J. A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat. Genet.* **38**, 594–597 (2006).
5. Studer, A., Zhao, Q., Ross-Ibarra, J. & Doebley, J. Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat. Genet.* **43**, 1160–1163 (2011).
6. Castelletti, S., Tuberosa, R., Pindo, M. & Salvi, S. A MITE transposon insertion is associated with differential methylation at the maize flowering time QTL Vgt1. *G3 (Bethesda)* **4**, 805–812 (2014).
7. Liu, L. et al. KRN4 Controls quantitative variation in maize kernel row number. *PLoS Genet* **11**, e1005670 (2015).
8. Huang, C. et al. ZmCCT9 enhances maize adaptation to higher latitudes. *Proc. Natl Acad. Sci. USA* **115**, E334–e341 (2018).
9. Yang, Q. et al. CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the post-domestication spread of maize. *Proc. Natl Acad. Sci. USA* **110**, 16969–16974 (2013).
10. Tian, J. et al. Teosinte ligule allele narrows plant architecture and enhances high-density maize yields. *Science* **365**, 658–664 (2019).
11. Marand, A. P., Eveland, A. L., Kaufmann, K. & Springer, N. M. cis-Regulatory elements in plant development, adaptation, and evolution. *Annu. Rev. Plant Biol.* **22**,111–137 (2022).
12. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213 (2013).
13. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
14. Hufford, M. B. et al. De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**, 655 (2021).
15. Tu, X. et al. Reconstructing the maize leaf regulatory network using ChIP-seq data of 104 transcription factors. *Nat. Commun.* **11**, 5089 (2020).
16. Crisp, P. A. et al. Stable unmethylated DNA demarcates expressed genes and their cis-regulatory space in plant genomes. *Proc. Natl Acad. Sci. USA* **117**, 23991–24000 (2020).
17. Dong, P. et al. 3D chromatin architecture of large plant genomes determined by local A/B compartments. *Mol. Plant* **10**, 1497–1509 (2017).

18. Sun, Y. et al. 3D genome architecture coordinates trans and cis regulation of differentially expressed ear and tassel genes in maize. *Genome Biol.* **21**, 143 (2020).
19. Marand, A. P., Chen, Z., Gallavotti, A. & Schmitz, R. J. A cis-regulatory atlas in maize at single-cell resolution. *Cell* **184,** 3041–3055 (2021).
20. Wirthlin, M. E. et al. Vocal learning-associated convergent evolution in mammalian proteins and regulatory elements. *Science*, **383**, eabn3263 (2024).
21. Kuderna, L. F. K. et al. Identification of constrained sequence elements across 239 primate genomes. *Nature* **625**, 735–742 (2024).
22. Andrews, G. et al. Mammalian evolution of human cis-regulatory elements and transcription factor binding sites. *Science* **380**, eabn7930 (2023).
23. Wu, Y. et al. Phylogenomic discovery of deleterious mutations facilitates hybrid potato breeding. *Cell* **186**, 2313–2328.e15 (2023).
24. Hu, Y. et al. Comparative analysis reveals epigenomic evolution related to species traits and genomic imprinting in mammals. *Innov. (Camb.)* **4**, 100434 (2023).
25. Song, B. et al. Conserved noncoding sequences provide insights into regulatory sequence and loss of gene expression in maize. *Genome Res* **31**, 1245–1257 (2021).
26. Chow, C.-N. et al. PlantPAN 4.0: updated database for identifying conserved non-coding sequences and exploring dynamic transcriptional regulation in plant promoters. *Nucleic Acids Res.* **52**, D1569–D1578 (2024).
27. Xin, H. et al. Identification and functional characterization of conserved cis-regulatory elements responsible for early fruit development in cucurbit crops. *Plant Cell* **36**, 2272–2288 (2024).
28. Savadel, S. D. et al. The native cistrome and sequence motif families of the maize ear. *PLoS Genet* **17**, e1009689 (2021).
29. Rodgers-Melnick, E., Vera, D. L., Bass, H. W. & Buckler, E. S. Open chromatin reveals the functional maize genome. *Proc. Natl Acad. Sci. USA* **113**, E3177–E3184 (2016).
30. Engelhorn, J. et al. Genetic variation at transcription factor binding sites largely explains phenotypic heritability in maize. *bioRxiv*, https://doi.org/10.1101/2023.08.08.551183 (2024).
31. Oka, R. et al. Genome-wide mapping of transcriptional enhancer candidates using DNA and chromatin features in maize. *Genome Biol.* **18**, 137 (2017).
32. Yan, H. et al. Evolution of plant cell-type-specific cis-regulatory elements. *bioRxiv* https://doi.org/10.1101/2024.01.08.574753 (2024).
33. Marand, A. P. et al. The genetic architecture of cell-type-specific *cis*-regulation. *bioRxiv*, (2024). 2024.08.17.608383.
34. Huang, W. et al. A well-supported nuclear phylogeny of Poaceae and implications for the evolution of C(4) photosynthesis. *Mol. Plant* **15**, 755–777 (2022).
35. Morgante, M. et al. Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nat. Genet.* **37**, 997–1002 (2005).
36. Horton, C. A. et al. Short tandem repeats bind transcription factors to tune eukaryotic gene expression. *Science* **381**, eadd1250 (2023).
37. Jin, J. et al. PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* **45**, D1040–D1045 (2017).
38. Castro-Mondragon, J. A. et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **50**, D165–d173 (2022).
39. Galli, M. et al. The DNA binding landscape of the maize AUXIN RESPONSE FACTOR family. *Nat. Commun.* **9**, 4526 (2018).
40. Han, J. et al. Genome-wide chromatin accessibility analysis unveils open chromatin convergent evolution during polyploidization in cotton. *Proc. Natl Acad. Sci.* **119**, e2209743119 (2022).
41. Chuck, G., Whipple, C., Jackson, D. & Hake, S. The maize SBP-box transcription factor encoded by tasselsheath4 regulates bract development and the establishment of meristem boundaries. *Development* **137**, 1243–1250 (2010).
42. Zhang, C., Li, X., Wang, Z., Zhang, Z. & Wu, Z. Identifying key regulatory genes of maize root growth and development by RNA sequencing. *Genomics* **112**, 5157–5169 (2020).
43. Tao, S. et al. Single-cell transcriptome and network analyses unveil key transcription factors regulating mesophyll cell development in maize. *Genes* **13**, 374 (2022).
44. Bukowski, R. et al. Construction of the third-generation *Zea mays* haplotype map. *Gigascience* **7**, 1–12 (2018).
45. Li, E. et al. Long-range interactions between proximal and distal regulatory regions in maize. *Nat. Commun.* **10**, 2633 (2019).
46. Rumpho, M. E., Edwards, G. E. & Loescher, W. H. A pathway for photosynthetic carbon flow to mannitol in celery leaves: activity and localization of key enzymes. *Plant Physiol.* **73**, 869–873 (1983).
47. Hatfield, R. D., Rancour, D. M. & Marita, J. M. Grass cell walls: a story of cross-linking. *Front Plant Sci.* **7**, 2056 (2016).
48. Faik, A. Xylan biosynthesis: news from the grass. *Plant Physiol.* **153**, 396–402 (2010).
49. Lai, J. et al. Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat. Genet.* **42**, 1027–1030 (2010).
50. Gabur, I. et al. Gene presence-absence variation associates with quantitative Verticillium longisporum disease resistance in Brassica napus. *Sci. Rep.* **10**, 4131 (2020).
51. Ye, J., Zhang, L., Zhang, X., Wu, X. & Fang, R. Plant defense networks against insect-borne pathogens. *Trends Plant Sci.* **26**, 272–287 (2021).
52. Han, G. Z. Origin and evolution of the plant immune system. *N. Phytol.* **222**, 70–83 (2019).
53. Sigmon, B. & Vollbrecht, E. Evidence of selection at the ramosa1 locus during maize domestication. *Mol. Ecol.* **19**, 1296–1311 (2010).
54. Liao, B. Y. & Zhang, J. Low rates of expression profile divergence in highly expressed genes and tissue-specific genes during mammalian evolution. *Mol. Biol. Evol.* **23**, 1119–1128 (2006).
55. Mantica, F. et al. Evolution of tissue-specific expression of ancestral genes across vertebrates and insects. *Nat. Ecol. Evolution* **8**, 1140–1153 (2024).
56. Tong, Y.-B. et al. GenOrigin: A comprehensive protein-coding gene origination database on the evolutionary timescale of life. *J. Genet. Genomics* **48**, 1122–1129 (2021).
57. Wang, B. et al. De novo genome assembly and analyses of 12 founder inbred lines provide insights into maize heterosis. *Nat. Genet.* **55**, 312–323 (2023).
58. Chen, J. et al. A complete telomere-to-telomere assembly of the maize genome. *Nat. Genet.* **55**, 1221–1231 (2023).
59. Buckler, E. S. et al. The Genetic Architecture of Maize Flowering Time. *Science* **325**, 714–718 (2009).
60. Yang, N. et al. Two teosintes made modern maize. *Science* **382**, eadg8940 (2023).
61. Huang, Y. et al. THP9 enhances seed protein content and nitrogen-use efficiency in maize. *Nature* **612**, 292–300 (2022).
62. Stitzer, M. C. et al. Extensive genome evolution distinguishes maize within a stable tribe of grasses. *bioRxiv*, https://doi.org/10.1101/2025.01.22.633974 (2025).
63. Liu, J., Fernie, A. R. & Yan, J. The past, present, and future of maize improvement: domestication, genomics, and functional genomic routes toward crop enhancement. *Plant Commun.* **1**, 100010 (2020).
64. Chen, W. et al. Convergent selection of a WD40 protein that enhances grain yield in maize and rice. *Science* **375**, eabg7985 (2022).
65. Lemmon, Z. H., Bukowski, R., Sun, Q. & Doebley, J. F. The role of cis regulatory evolution in maize domestication. *PLoS Genet* **10**, e1004745 (2014).

66. Gao, Y. et al. A maize phytochrome-interacting factors protein ZmPIF1 enhances drought tolerance by inducing stomatal closure and improves grain yield in Oryza sativa. *Plant Biotechnol. J.* **16**, 1375–1387 (2018).
67. Jan, R. et al. Over-expression of Chorismate mutase enhances the accumulation of salicylic acid, lignin, and antioxidants in response to the white-backed planthopper in rice plants. *Antioxidants (Basel)* **10,** 1680 (2021).
68. Schnable, J. C., Springer, N. M. & Freeling, M. Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. *Proc. Natl. Acad. Sci. USA* **108**, 4069–74 (2011).
69. Fang, C. et al. Dynamics of cis-regulatory sequences and transcriptional divergence of duplicated genes in soybean. *Proc. Natl Acad. Sci. USA* **120**, e2303836120 (2023).
70. Kong, D. et al. UB2/UB3/TSH4-anchored transcriptional networks regulate early maize inflorescence development in response to simulated shade. *Plant Cell* (2022).
71. Wills, D. M. et al. From many, one: genetic control of prolificacy during maize domestication. *PLoS Genet* **9**, e1003604 (2013).
72. Li, X. et al. Maize GOLDEN2-LIKE genes enhance biomass and grain yields in rice by improving photosynthesis and reducing photoinhibition. *Commun. Biol.* **3**, 151 (2020).
73. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
74. Li, C. et al. Genomic insights into historical improvement of heterotic groups during modern hybrid maize breeding. *Nat. Plants* **8**, 750–763 (2022).
75. Zhu, Y. et al. HEMU: an integrated comparative genomics database and analysis platform for Andropogoneae grasses. *Plant Communications*, **5**, 100786 (2023).
76. Zhu, T. et al. Comprehensive mapping and modelling of the rice regulome landscape unveils the regulatory architecture underlying complex traits. *Nat. Commun.* **15**, 6562 (2024).
77. Zhu, Y. et al. Pan-cistrome analysis of the leaf accessible chromatin regions of 214 maize inbred lines. *bioRxiv* https://doi.org/10.1101/2024.10.14.618191 (2024).
78. Bubb, K. L. et al. The regulatory potential of transposable elements in maize. *bioRxiv* https://doi.org/10.1101/2024.07.10.602892 (2024).
79. Li, T. et al. Modeling 0.6 million genes for the rational design of functional cis-regulatory variants and de novo design of cis-regulatory sequences. *Proc. Natl Acad. Sci. USA* **121**, e2319811121 (2024).
80. Zhou, X., Ruan, Z., Zhang, C., Kaufmann, K. & Chen, D. Deep learning chromatin profiles reveal the cis-regulatory sequence code of the rice genome. *J. Genet. Genomics* https://doi.org/10.1016/j.jgg.2024.12.007(2024).
81. Zhang, D., Gan, Y., Le, L. & Pu, L. Epigenetic variation in maize agronomical traits for breeding and trait improvement. *J. Genet. Genomics* https://doi.org/10.1016/j.jgg.2024.01.006 (2024).
82. Cahn, J. et al. MaizeCODE reveals bi-directionally expressed enhancers that harbor molecular signatures of maize domestication. *Nat. Commun.* **15**, 10854 (2024).
83. Zuo, Z. et al. ATAC-pipe: general analysis of genome-wide chromatin accessibility. *Brief. Bioinform* **20**, 1934–1943 (2019).
84. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
85. Stark, R. & Brown, G. *DiffBind differential binding analysis of ChIP-Seq peak data*. *In R package version* **100** (2011).
86. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
87. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
88. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
89. Kovaka, S. et al. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, 278 (2019).
90. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
91. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
92. Ou, S. et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **20**, 275 (2019).
93. Yan, H., Bombarely, A. & Li, S. DeepTE: a computational method for de novo classification of transposons with convolutional neural network. *Bioinformatics* **36**, 4269–4275 (2020).
94. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).
95. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European molecular biology open software suite. *Trends Genet* **16**, 276–277 (2000).
96. Ma, J. & Bennetzen, J. L. Rapid recent growth and divergence of rice nuclear genomes. *Proc. Natl Acad. Sci. USA* **101**, 12404–12410 (2004).
97. Armstrong, J. et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).
98. Hinrichs, A. S. et al. The UCSC genome browser database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
99. Hedges, S. B., Marin, J., Suleski, M., Paymer, M. & Kumar, S. Tree of life reveals clock-like speciation and diversification. *Mol. Biol. Evolution* **32**, 835–845 (2015).
100. Mcinnes, L. & Healy, J. UMAP: uniform manifold approximation and projection for dimension reduction. *J. Open Source Softw.* **3**, 861 (2018).
101. Klopfenstein, D. V. et al. GOATOOLS: a python library for gene ontology analyses. *Sci. Rep.* **8**, 10872 (2018).
102. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
103. Sun, S. et al. Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nat. Genet* **50**, 1289–1295 (2018).
104. Kiełbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res* **21**, 487–493 (2011).
105. Haas, B. J., Delcher, A. L., Wortman, J. R. & Salzberg, S. L. DAGchainer: a tool for mining segmental genome duplications and synteny. *Bioinformatics* **20**, 3643–3646 (2004).
106. Tang, H. et al. Screening synteny blocks in pairwise genome comparisons through integer programming. *BMC Bioinforma.* **12**, 102 (2011).
107. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
108. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet* **81**, 559–575 (2007).
109. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. *Methods Mol. Biol.* **1019**, 215–236 (2013).
110. Kang, H. M. et al. Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet* **42**, 348–354 (2010).


## Acknowledgements

We want to thank the colleagues from Frasergen company in Wuhan, China for their assistance in preparing the ATAC-seq sequencing libraries. We also appreciate the assistance and suggestions from Dr. Zhong, Silin from The Chinese University of Hong Kong. We also want to

thank Dr. Yaoyao Wu from Nanjing Agricultural University for her suggestions on genome constraint analysis, as well as The High-performance Computing Public Platform (Shenzhen Campus) of SUN YAT-SEN UNIVERSITY for providing computing resources. This work was supported by grants from the National Natural Science Foundation of China (32172014 and 32372115), the Shenzhen Outstanding Youth Science Fund project (RCYX20231211090225038), the Young Elite Scientists Sponsorship Program by CAST (2021QNRC001), and the Fundamental Research Funds for the Central Universities, Sun Yat-sen University (23lgbj016) to J.S.

## Author contributions

J.S. and X.G. conceived this research project. Y.L. completed most of the data analysis. X.G., Y.Z., L.H., Q.L., H.L., X.Y., F.X., F.Y., and J.L. participated in data analysis. X.G., H.L., X.L., F.X., and X.Y. involved in sample collection and sequencing. J.S., X.G. and Y.L. wrote and finalized this manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-57932-1.

**Correspondence** and requests for materials should be addressed to Junpeng Shi.

**Peer review information** *Nature Communications* thanks Lin Li, Jeffrey Ross-Ibarra and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.