

Genome sequencing reveals evidence of adaptive variation in the genus *Zea*

Received: 7 April 2021

Accepted: 10 August 2022

Published online: 20 October 2022

 Check for updates

Lu Chen^{1,14,15}, Jingyun Luo^{1,15}, Minliang Jin^{1,15}, Ning Yang^{1,2,15}✉, Xiangguo Liu³, Yong Peng¹, Wenqiang Li¹, Alyssa Phillips^{4,5}, Brenda Cameron⁵, Julio S. Bernal⁶, Rubén Rellán-Álvarez⁷, Ruairidh J. H. Sawers⁸, Qing Liu³, Yuejia Yin³, Xinnan Ye³, Jiali Yan¹, Qinghua Zhang¹, Xiaoting Zhang¹, Shenshen Wu¹, Songtao Gui¹, Wenjie Wei¹, Yuebin Wang¹, Yun Luo¹, Chenglin Jiang¹, Min Deng¹, Min Jin¹, Liumei Jian¹, Yanhui Yu¹, Maolin Zhang¹, Xiaohong Yang⁹, Matthew B. Hufford¹⁰, Alasdair R. Fernie¹¹, Marilyn L. Warburton¹², Jeffrey Ross-Ibarra¹³✉ and Jianbing Yan^{1,2}✉

Maize is a globally valuable commodity and one of the most extensively studied genetic model organisms. However, we know surprisingly little about the extent and potential utility of the genetic variation found in wild relatives of maize. Here, we characterize a high-density genomic variation map from 744 genomes encompassing maize and all wild taxa of the genus *Zea*, identifying over 70 million single-nucleotide polymorphisms. The variation map reveals evidence of selection within taxa displaying novel adaptations. We focus on adaptive alleles in highland teosinte and temperate maize, highlighting the key role of flowering-time-related pathways in their adaptation. To show the utility of variants in these data, we generate mutant alleles for two flowering-time candidate genes. This work provides an extensive sampling of the genetic diversity of *Zea*, resolving questions on evolution and identifying adaptive variants for direct use in modern breeding.

Global crop production is currently insufficient to meet the anticipated demands of a growing human population^{1,2}. Climate change is affecting crop production in many areas, further exacerbating this problem³, and projected shifts in temperature and precipitation will lead to further

declines in productivity for many major crops⁴. New varieties displaying both higher yield and better adaptation to diverse environments are thus urgently needed to increase crop productivity under changing climate scenarios^{5,6}.

¹National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, China. ²Hubei Hongshan Laboratory, Wuhan, China.

³Institute of Agricultural Biotechnology, Jilin Academy of Agricultural Sciences, Changchun, China. ⁴Center for Population Biology, University of California Davis, Davis, CA, USA. ⁵Department of Evolution and Ecology, University of California Davis, Davis, CA, USA. ⁶Department of Entomology, Texas A&M University, College Station, TX, USA. ⁷Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC, USA.

⁸Department of Plant Science, The Pennsylvania State University, State College, PA, USA. ⁹National Maize Improvement Center of China, Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing, China. ¹⁰Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, USA. ¹¹Department of Molecular Physiology, Max-Planck-Institute of Molecular Plant Physiology, Potsdam-Golm, Germany.

¹²United States Department of Agriculture-Agricultural Research Service: Western Regional Plant Introduction Station, Washington State University, Pullman, WA, USA. ¹³Department of Evolution and Ecology, Center for Population Biology, Genome Center, University of California Davis, Davis, CA, USA.

¹⁴Present address: State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China. ¹⁵These authors contributed equally: Lu Chen, Jingyun Luo, Minliang Jin, Ning Yang.

✉e-mail: ningy@mail.hzau.edu.cn; rossibarra@ucdavis.edu; yjianbing@mail.hzau.edu.cn

Maize (*Zea mays* subspecies *mays*) is one of the world's most widely grown crops. Native American peoples domesticated the wild grass *Z. mays* subspecies *parviflora* (hereafter *parviflora*) approximately 9,000 years ago in the southwest of Mexico^{7,8}. Population genetics analyses largely agree that maize underwent a substantial population bottleneck during domestication^{9–12}, reducing the genetic diversity available for adaptation. Although maize rapidly spread from its center of domestication across a wide range of environments, successful adaptation required hundreds or thousands of years¹³. As global populations increase and climate change accelerates, unprecedented maize yield losses are projected to become commonplace in most maize-producing regions^{5,14,15}. To facilitate adaptation to these new challenges, breeders will need to maximize the use of the genetic diversity at their disposal, looking beyond modern elite lines to traditional cultivated varieties and locally adapted wild relatives¹⁶.

The wild congeners of maize—collectively called teosintes—are annual and perennial grasses native to Mexico and Central America (Fig. 1a). They are adapted to a diverse range of environments, from hot, humid, subtropical regions of Central America to cold, dry, high elevations of the Mexican Central Plateau^{17,18}. Teosintes exhibit biotic and abiotic adaptations that are absent in modern maize^{17–19}, providing a wealth of genetic diversity that could be utilized in modern breeding. Recent examples show that the alleles from teosintes can help maize to adapt to low-temperature and low-phosphorus environments²⁰ and confer resistance to multiple diseases²¹. Other studies have used genetic mapping to capitalize on teosinte alleles for nutrition^{22,23}, adaptation to extreme environments^{24,25} and disease resistance^{26–28}. Population genetics evidence suggests that diverse alleles from the teosinte *Z. mays* subspecies *mexicana* (hereafter *mexicana*) played an important role in allowing maize to adapt to arid highland conditions^{29,30}.

Despite the potential for teosintes to contribute to the breeding and adaptation of cultivated maize, we know relatively little about the genetic diversity and history of these taxa. Estimates of the age of the genus vary substantially^{31–35}, and the phylogenetic relationships of several taxa are debated or unknown^{17,36–38}. Considerable cytological diversity is found within the genus, and transposable element variation^{39–41} and large inversions^{42–47} have been documented as well. Moreover, common garden studies have demonstrated that phenotypic differentiation in both teosintes and maize landraces is the result of local adaptation^{48,49}. Low-density genotyping or pooled sequencing approaches in *parviflora* and *mexicana* have identified a number of candidate loci related to soil, climate and disease resistance, highlighting the importance of inversions^{46,50,51}. However, for most taxa in *Zea*, their potential as sources of useful diversity in maize remains poorly understood.

Here, we present a genus-wide resource of genome-scale genetic diversity in *Zea*. We resequenced 237 teosinte accessions, including all seven taxa of teosinte, and combined these data with sequences from 507 maize inbred lines. Our analyses reveal a detailed phylogeny and demography of the genus *Zea*, identify substantial novel genetic diversity and expand our understanding of adaptation in the genus *Zea*. We predict that these resources will substantially facilitate the efficient use of diverse *Zea* taxa in modern maize breeding and improvement.

Results

The diversity map and phylogeny of the genus *Zea*

We resequenced 237 teosinte accessions encompassing all of the described species and subspecies in the genus *Zea* (Fig. 1a,b) to an average depth of 22× and combined these data with genome resequencing data from 507 cultivated maize inbred lines representing both temperate and tropical regions⁵² (Supplementary Table 1). To ensure the quality of this *Zea* diversity map, we used a set of strict filtering conditions (Methods). We identified a final set more than 70 million single-nucleotide polymorphisms (SNPs) and nearly 9 million insertions/deletions (indels) (Supplementary Table 2), with nearly 80% of SNPs segregating as rare variants (minor allele frequency (MAF) < 0.05)

(Supplementary Fig. 1). Both classes of variants appeared to be enriched in genic and regulatory regions (30% of SNPs and 45% of indels in 14% of the genome), probably reflecting difficulties in read mapping in repetitive regions of the genome. We validated a subset of genic SNPs using Sanger sequencing, with a median concordance between datasets of >95% and reasonable false positive and false negative rates (both ~5% on average) for non-reference alleles (Supplementary Table 3). Based on population structure analysis, samples with >60% ancestry in a single group were clustered into *parviflora* ($n = 70$), *mexicana* ($n = 81$), *Z. mays* subspecies *huehuetenangensis* ($n = 5$; hereafter, *huehuetenangensis*), *Zea diploperennis* ($n = 20$; hereafter, *diploperennis*), *Zea perennis* ($n = 19$; hereafter, *perennis*), *Zea luxurians* ($n = 14$; hereafter, *luxurians*), *Zea nicaraguensis* ($n = 14$; hereafter, *nicaraguensis*), tropical maize ($n = 210$) and temperate maize ($n = 280$) (Extended Data Fig. 1a,b and Supplementary Table 1). Principal component analysis (PCA) of these lines was in strong concordance with population structure results (Extended Data Fig. 1c).

We inferred phylogenetic relationships for the genus *Zea* under the multispecies coalescent model⁵³ (Fig. 1c) and maximum likelihood phylogenies⁵⁴, which produced largely congruent results (Extended Data Fig. 2 and Supplementary Fig. 2). Notably, we estimated a very recent origin for the genus, splitting from its sister genus *Tripsacum* only ~650,000 years ago. This young age is especially striking given the pronounced differences in chromosome structure and sub-genome organization resulting from the two genera's shared polyploidy event >10 million years ago⁵⁵. Within the genus, our results suggest that *nicaraguensis* probably represents a subspecies of *luxurians*, with a divergence time similar to those among subspecies of *Z. mays*. The phylogeny supports earlier analysis³⁴ suggesting that divergence among *Z. mays*, *luxurians* and *diploperennis* was nearly contemporaneous, occurring ~120,000 years ago (95% highest posterior density (HPD) interval for *luxurians* divergence from other taxa = 119,400–127,200 years ago; Fig. 1c and Supplementary Table 4). We further estimate that *perennis* split from its diploid progenitor *diploperennis* only ~48,000 years ago (95% HPD interval = 38,033–119,100 years ago). Tree topologies and divergence times also support earlier analyses⁵⁶ showing that *huehuetenangensis* is a subspecies of *Z. mays*, diverging from other annual subspecies ~68,000 years ago (95% HPD interval = 60,133–106,467 years ago), followed by the divergence of highland *mexicana* and lowland *parviflora* ~30,000 years ago (95% HPD interval = 26,733–34,500 years ago). Our phylogeny estimates the divergence of maize from *parviflora* at ~12,000 years—only slightly older than the earliest archeological evidence⁸ and probably due to population structure within *parviflora*^{37,46}. Independent estimates of divergence times taken from rates of cross-coalescence⁵⁷ between taxa are strikingly consistent (Fig. 1d).

Population genetic analysis of diversity further reveals changes in demography among taxa in *Zea*. Coalescent estimates of the effective population size (N_e) over time reveal the well-established bottleneck associated with maize domestication but also a continued decline in population size for the annual subspecies *parviflora* and *mexicana* since their divergence (Supplementary Fig. 3). All other taxa in the genus show parallel trends, with steady declines in population size until ~10,000 years ago and with more recent increases for *luxurians* and *diploperennis*. Patterns of shared derived alleles and sequence divergence both suggest a history of introgression among taxa (Fig. 1e, Supplementary Fig. 4 and Supplementary Table 5), including bidirectional admixture between *parviflora*/*huehuetenangensis* and *nicaraguensis*/*luxurians* and unidirectional introgression from *huehuetenangensis*/*mexicana* into domesticated maize, highlighting the important role of gene flow in crop adaptation⁵⁸.

Novel diversity in *Zea*

SNP data highlight the impressive genetic diversity present in teosintes. Despite the potential downward bias due to strict filtering parameters and read mapping to a maize reference, heterozygosity and nucleotide diversity are both higher in teosinte taxa than the much larger panel

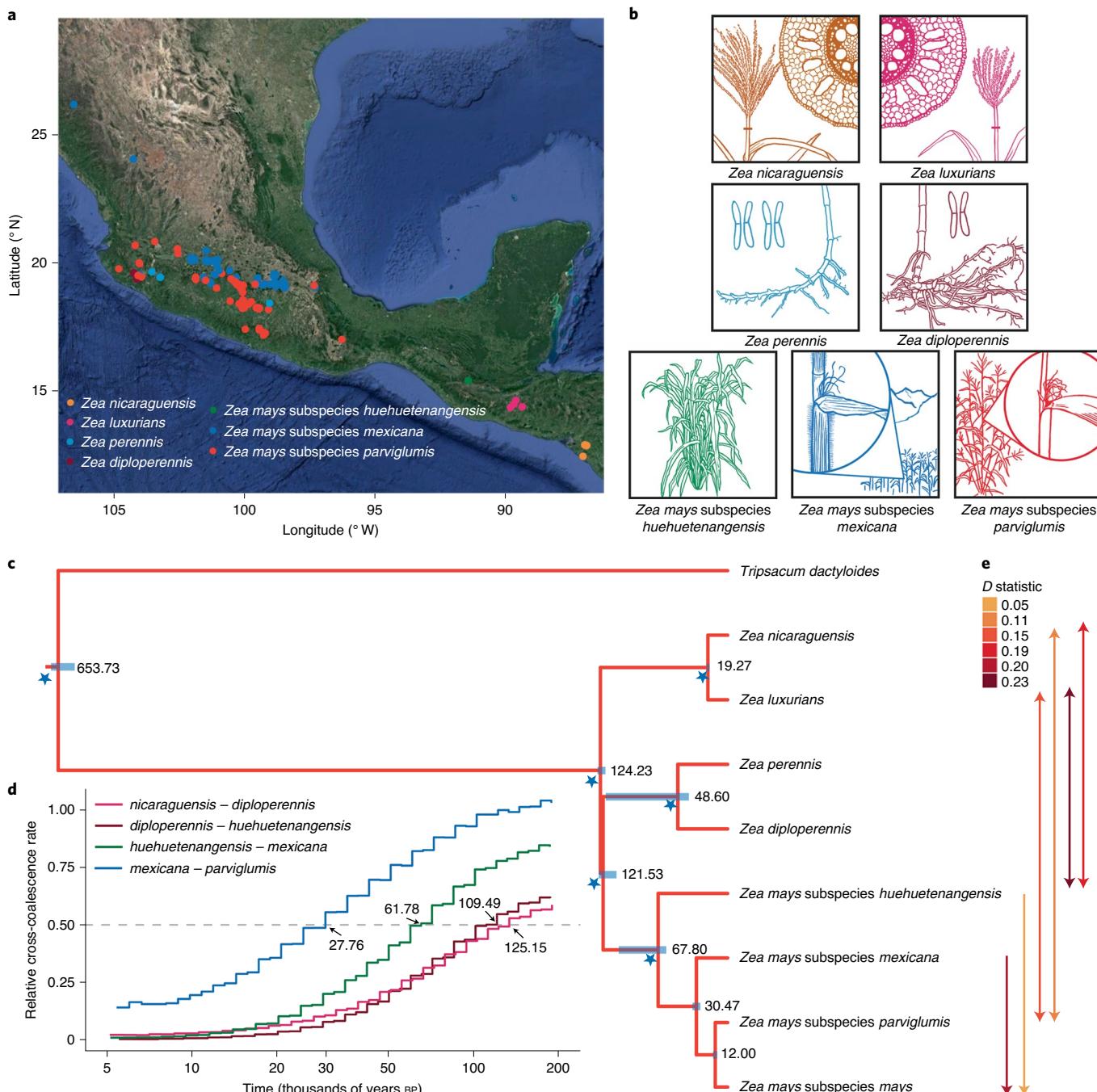


Fig. 1 | Phylogeny of the *Zea* genus. **a**, Geographical distribution of the collected teosintes. The taxa were identified and are colored based on morphology. Adapted from Google Imagery © 2022 TerraMetrics. **b**, Morphological characteristics of teosintes. *nicaraguensis* and *luxurians* are distinguished from the other teosintes based on aerenchyma in their stems (which aerate the roots during submergence), while *nicaraguensis* has a more robust tassel than *luxurians*. *perennis* is a recent autotetraploid of *diploperennis*; the rhizomatous root systems of these perennial taxa distinguish them from the other teosintes. The Mexican annual teosintes *parviflora* and *mexicana* are distinguished from each other based on the presence

of macro-hairs and pigment along their stems—two traits that are linked to highland adaptation. Credit: Andi Kur. **c**, Divergence times estimated from the multispecies coalescent model. The blue bars indicate 95% HPD intervals. Stars indicate nodes with a posterior probability of 1. **d**, Rates of cross-population coalescence among teosinte species. The curves were computed using four phased haplotypes. **e**, Introgression among taxa. The arrows indicate the taxa involved (a one-way arrow indicates unidirectional introgression whereas a two-way arrow indicates bidirectional introgression) and the arrow color shows the value of Patterson's *D* statistic (Supplementary Table 5).

of maize lines, even among teosinte with limited geographic ranges (Fig. 2a,b, Extended Data Fig. 3 and Supplementary Table 2). Differentiation (F_{ST}) between teosinte taxa is often lower than that found between inbred maize and teosintes (Fig. 2a), consistent with the historical reduction of diversity that occurred during modern maize breeding⁵⁹. The annual subspecies of *Z. mays* show much faster decay of linkage

disequilibrium than our diverse panel of maize inbreds (10–50 kilobases (kb) compared with ~200 kb; Fig. 2c), but historical recombination in other teosintes appears to be even more limited (>500 kb). Nearly one-quarter (24%) of the SNPs and 20% of the indels identified across all taxa are taxon specific (Supplementary Table 2) and there are significantly more SNPs specific to each teosinte accession than maize

(Fig. 2d and Supplementary Table 6). This tendency remains the same after choosing comparable samples in each taxon (Supplementary Fig. 5). In teosintes, a substantial proportion of taxon-specific SNPs and indels are located in genic and regulatory regions (promoter and *cis*-regulatory elements⁶⁰; Supplementary Fig. 6), suggesting the presence of biologically functional alleles with the potential for improving modern maize.

Short-read mapping approaches pose challenges in characterizing genetic diversity, including difficulty with repetitive sequences and reference bias. To circumvent some of these obstacles, we used a reference-free *k*-mer approach to characterize the diversity of each taxon (Methods). Consistent with the reference mapping bias (~8% unmapped reads on average), most taxa showed a substantial proportion of unique *k*-mers (Supplementary Fig. 7a,b and Supplementary Table 2), and a higher number of unique *k*-mers were exhibited in the species other than *Z. mays* (Fig. 2e, Supplementary Fig. 7c,d and Supplementary Table 6). These results not only highlight the novel genetic diversity present in teosintes but also probably point to the ongoing importance of evolutionary processes in generating and filtering diversity in traditional maize populations in Mexico⁶¹.

Next, we investigated the diversity and abundance of transposons and inversion polymorphisms in *Zea*. Transposable elements are an important driver shaping the structure and evolution of the genome⁶² and over 85% of the maize genome is repetitive sequence⁶³. Repeat clusters from our short-read data account for an average of ~74% of the genomic sequence across species (Supplementary Table 7), with the vast majority (60–70%) coming from long-terminal repeat retrotransposons. Mapping reads from individual genomes to these clusters revealed broadly similar patterns across species, consistent with previous comparisons of *Z. mays* and *luxurians*⁴⁰. Nonetheless, we identified a notable decrease in the content of Ty3 retrotransposons in *Z. mays* compared with other species, as well as an increased abundance of DNA transposons in *diploperennis* and *perennis* (Fig. 2f and Extended Data Fig. 4).

Inversions are known to play important roles in adaptation and speciation^{64,65}, and previous work has highlighted the evolutionary relevance of several large inversions in *Zea*^{24,45,46,66}, including *Inv9e* in *mexicana* adaptation^{46,50,51}. Multidimensional scaling of SNP diversity across the genome⁶⁷ allowed us to identify eight large genomic regions (>1 megabase) indicative of inversion polymorphism (Supplementary Fig. 8 and Supplementary Table 8), showing the clustering patterns delineating three genotypes: (1) standard, (2) heterozygous inversion and (3) homozygous inversion (Fig. 2g, Extended Data Fig. 5 and Supplementary Table 9).

Given previous evidence suggesting the association between inversions and soil characteristics⁴⁶, we performed genome-wide association analysis with nine representative soil traits (Methods) from a rich database of more than 200 soil properties⁶⁸ (Supplementary Fig. 9a and Supplementary Table 10). *Inv9e* was significantly associated with gypsum content (0.829–1.383 m), which is a representative of 29 soil properties (Supplementary Fig. 9b and Supplementary Table 10).

We merged nearby significant SNPs located in *Inv9e* into two quantitative trait loci (QTLs) on chromosome 9: base pairs 127,017,047–127,356,295 and 138,354,955–139,846,464 (Supplementary Fig. 10 and Supplementary Table 11). These QTLs contain 15 genes that have been functionally validated in rice or *Arabidopsis* (Supplementary Table 12), including two (*Zm00001d047667* and *Zm00001d047694*) with orthologs that have been confirmed to affect root development in rice^{69,70} and may provide clues to further explore the function of *Inv9e* in adaptation. Given that many inversions found segregating at appreciable frequency are probably adaptive in some environments^{71,72}, these data argue that improved assemblies and characterization of structural variants in teosinte would be a promising avenue for the discovery of new functional genetic diversity.

Signals of selection from allele frequency data

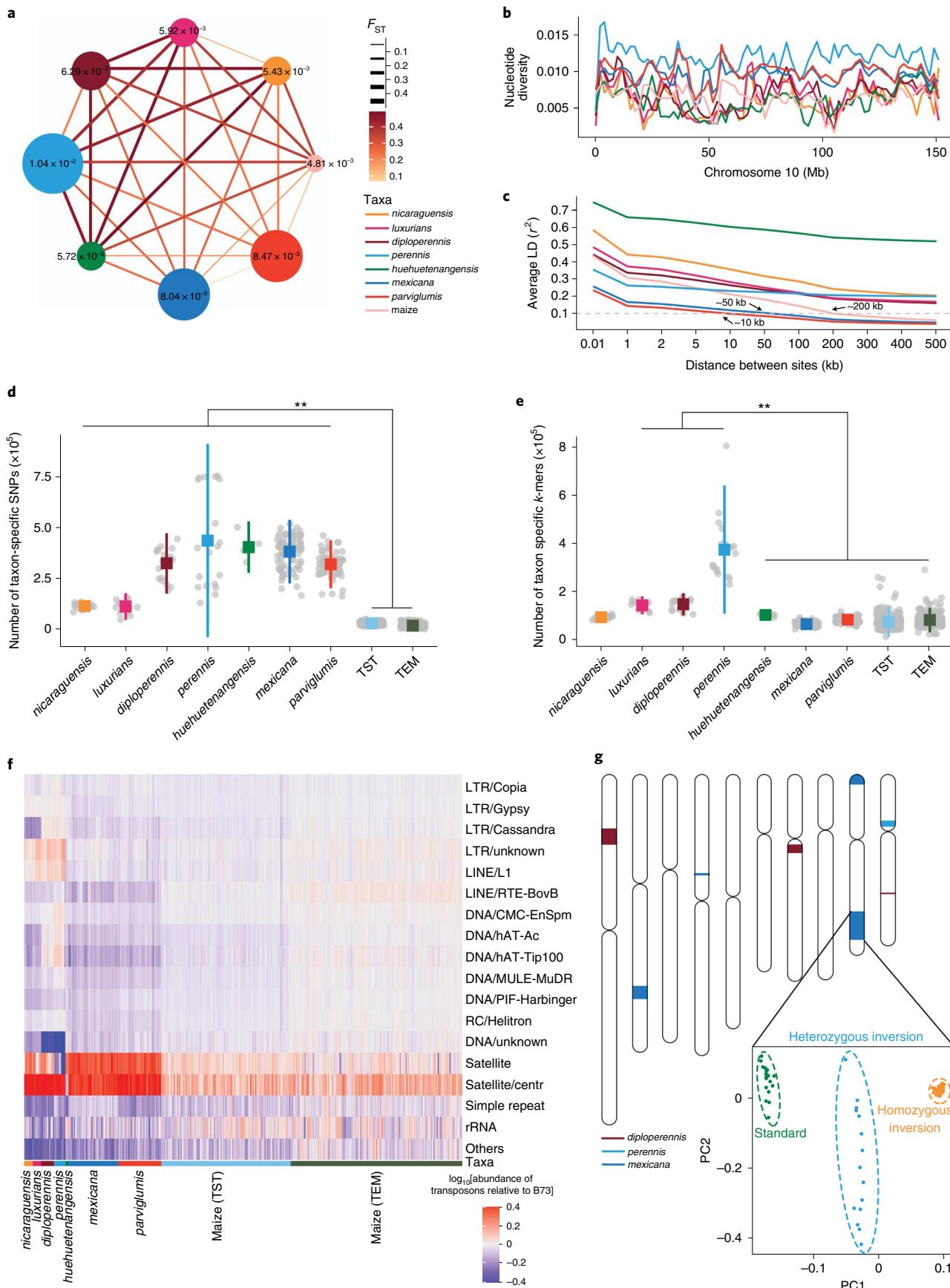
Their genetic, ecological and life history diversity make teosintes an ideal model system for studying adaptation¹⁷. To identify potential targets of selection, we calculated *F_{ST}* between each teosinte taxon and cultivated maize in 5-kb sliding windows (Methods). We found that a high proportion of outlier windows were shared between the closely related taxa (56% overlapped between *nicaraguensis* and *luxurians* and 54% overlapped between *diploperennis* and *perennis*; Supplementary Table 13 and Supplementary Fig. 11). Shared genes (5,706; Supplementary Table 14) in *nicaraguensis* and *luxurians* comparisons were enriched in core cell component and reproductive system developmental processes (GO:0061458; *P* = 1.15 × 10⁻⁴; false discovery rate = 6.87 × 10⁻³; Supplementary Table 15). Candidate adaptive genes (4,659; Supplementary Table 16) in *diploperennis* and *perennis* comparisons were enriched in some basic biological process and core cellular components such as the nucleus (GO:0005634; *P* = 1.25 × 10⁻¹²; false discovery rate = 2.89 × 10⁻¹⁰) (Supplementary Fig. 12 and Supplementary Table 17).

We also identified a number of genes related to meiosis⁷³, QTLs in regrowth⁷⁴ and waterlogging^{75–77} (Supplementary Table 18). These included *Zm00001d002945*—an ortholog of the *Arabidopsis* gene *AtNAC082* involved in the regulation of leaf senescence⁷⁸—which showed high *F_{ST}* in *diploperennis*–maize and *perennis*–maize comparisons and was located in a QTL region controlling regrowth⁷⁴. In *nicaraguensis*–maize and *luxurians*–maize comparisons, we found genes potentially involved in the response to waterlogging not only by regulating the content of ethylene and wax, but also photosynthetic efficiency potentially related to adaptation to wetter climates in Guatemala¹⁷. These include *Zm00001d015637*, the maize ortholog of *AtOSPI* in *Arabidopsis*, which transcribes a GDSL lipase (a subclass of lipolytic enzymes) that is required for wax biosynthesis and stomatal formation⁷⁹. These genes highlight the value of our diversity data in identifying candidate loci of potential adaptive relevance for maize, and present a catalog of genes worthy of further exploration.

In addition to identifying differences among species, our extensive sampling of *parviflumis* (*n* = 70), *mexicana* (*n* = 81) and tropical (*n* = 210) and temperate (*n* = 280) maize accessions allowed investigation of more recent adaptation to highlands and high latitudes. Both high elevation and high latitude reflect a climate of lower temperature

Fig. 2 | Variation in the *Zea* genus. **a**, Mean nucleotide diversity in each taxon (nodes) and mean population differentiation *F_{ST}* between taxa (edges). The size of the nodes and values within the nodes represent the nucleotide diversity, whereas the width and color of the edges represent *F_{ST}*. **b**, Distribution of nucleotide diversity along chromosome 10 in *Zea*. The colors of the taxa are the same as in **a**. Mb, megabases. **c**, Linkage disequilibrium (LD) decay of the *Zea* genus. The labels indicate the distance at which mean *r*² = 0.1. The colors of the taxa are the same as in **a**. **d**, Taxon-specific SNPs. TEM, temperate maize; TST, tropical maize. **e**, 31-bp *k*-mers in *Zea*. The lines show statistical comparisons of all teosintes with TST and TEM (**d**) and of *luxurians/diploperennis/perennis* with *Z. mays* (**e**). Each gray point represents an individual of *nicaraguensis* (*n* = 14), *luxurians* (*n* = 14), *diploperennis* (*n* = 20), *perennis* (*n* = 19), *huehuetenangensis*

(*n* = 5), *mexicana* (*n* = 81), *parviflumis* (*n* = 70), TST (*n* = 210) or TEM (*n* = 280). The squares indicate mean values and the vertical lines show s.d. Statistical significance was determined by two-sided *t*-test for each comparison. ***P* < 0.01 (Pairwise comparisons of *P* values are provided in Supplementary Table 6). **f**, Abundance of transposon elements relative to B73. Each column represents a sample. LTR, long terminal repeat; centr, centromere; rRNA, ribosomal RNA. **g**, Distribution of inversions across the chromosomes. Each colored segment represents an inversion, with colors referring to the population in which the inversion is most prevalent (deep red, *diploperennis*; blue, *perennis*; deep blue, *mexicana*). The inset shows PCA of SNP data from within *Inv9e*, clearly separating the three genotype classes (left, standard; middle, heterozygous inversion; right, homozygous inversion).



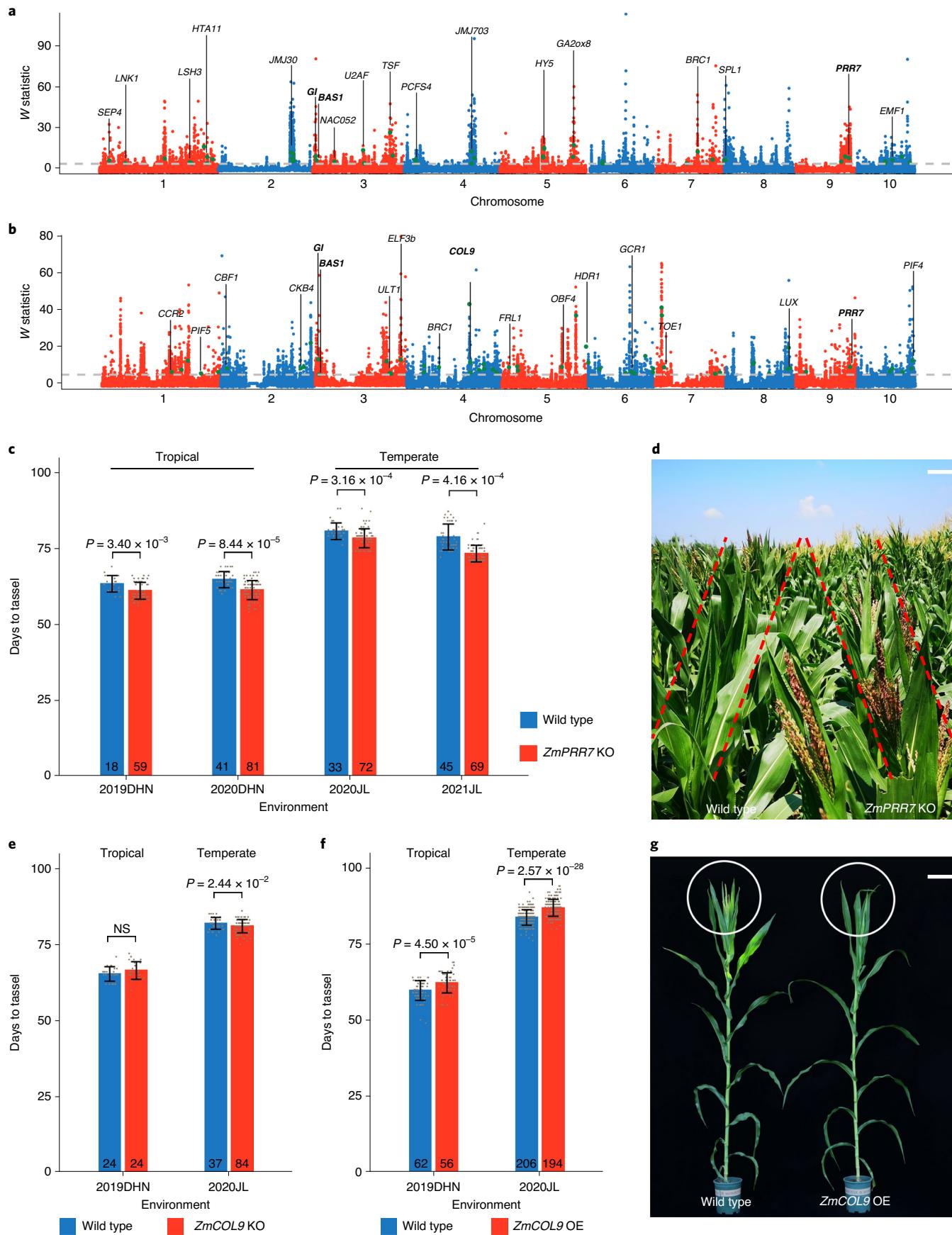


Fig. 3 | Local adaptation in teosintes and maize. **a,b**, Genome-wide selection signals (W statistic reflecting the smoothed cross-population composite likelihood ratio score) between *mexicana* and *parviflora* (**a**) and temperate and tropical maize (**b**). The horizontal gray dashed line represents the top 5% cut off. Genes associated with flowering time and floral development in maize, rice and *Arabidopsis thaliana* are marked with green points. **c**, Days to tassel of wild-type and *ZmPrr7* knockout (KO) mutants under tropical (Hainan Province, China in 2019 and 2020 (2019DHN and 2020DHN, respectively); 109° E, 18° N) and temperate environments (Jilin Province, China in 2020 and 2021 (2020JL and 2021JL, respectively); 125° E, 44° N). **d**, *ZmPrr7* KO mutants showed earlier flowering relative to wild types. The picture was taken in Jilin Province in 2020

77 d after planting (scale bar, 5 cm). Credit: X.L. **e**, Days to tassel of wild-type and *ZmCOL9* KO mutants under tropical and temperate environments. NS, no significant difference between mutants and wild types. **f**, Days to tassel of wild-type and *ZmCOL9* overexpression (OE) mutants under tropical and temperate environments. **g**, *ZmCOL9* OE mutants showed later flowering relative to wild types. The picture was taken in Jilin Province in 2020 78 d after planting (scale bar, 20 cm). Credit: X. Ye. In **c,e** and **f**, the gray points show the number of days to tassel for each individual. The numbers at the bottom of the bars indicate the number of individuals used for phenotyping. The bars represent s.d. Statistical significance was determined by two-sided *t*-test.

and longer light periods, and previous work identified evidence of convergent selection between temperate maize and its broadly distributed temperate relative *Tripsacum*⁸⁰. Here, we extended this comparison to investigate convergence between high-elevation-adapted teosinte (*mexicana*) and temperate maize. We applied a composite likelihood genome scan (Methods) for selection between *mexicana* versus *parviflora* and temperate versus tropical maize (Fig. 3a,b and Supplementary Tables 19 and 20). We found significant overlap in selected windows ($P = 0.047$; 14.7% higher than permutations; Extended Data Fig. 6a), but less overlap than expected in candidate genes ($P = 0.97$; 27% less than permutations). Notably, ~90% of selected windows in both comparisons were found in non-coding regions of the genome, suggesting that adaptation may have predominantly targeted regulatory regions. To test for convergence in regulatory adaptation, we used RNA sequencing datasets from the shoot base of *parviflora*, *mexicana* and tropical and temperate maize to search for changes in gene expression. We identified 595 genes that are differentially expressed between *mexicana* and *parviflora* (Supplementary Table 21) and 437 genes that are differentially expressed between temperate and tropical maize (Supplementary Table 22), with significant overlap between the two lists ($P = 0.006$; 102% higher than permutations; Extended Data Fig. 6b). These results may point to the importance of convergent regulatory evolution in teosinte and maize local adaptation.

Selection for variants that promote early flowering enabled maize to break day-length restrictions and facilitated the spread of maize across a broad geographical range⁸¹. The alleles involved in flowering time are also a major target of highland landrace adaptation⁸². Experimental data in maize⁸³ and from orthologs⁸⁴ in other species show that at least 51 genes associated with highland adaptation and 61 genes associated with high-latitude adaptation were involved in flowering-time pathways (Extended Data Fig. 7 and Supplementary Table 23). For example, the genes *GI*, *BAS1* and *PRR7*—all of which are known to participate in the circadian clock pathway in *Arabidopsis* and rice^{85–89}—show evidence of selection both in *mexicana* and temperate maize. Tracking the flowering-time pathway, we found that temperate maize has more genes under selection in the photoperiod pathway (eight in temperate maize and five in *mexicana*; Supplementary Table 23), which may be a signal of adaptation to changing latitude.

To validate the utility of the selection scan approach, we tested the function of *ZmPrr7* (*Zm00001d047761*), which shows convergent patterns in teosintes and maize, and the maize-specific candidate *ZmCOL9* (*Zm00001d051684*), which is involved in the photoperiod pathway. Mutants of these two genes were obtained from a CRISPR-Cas9-based high-throughput targeted mutagenesis library⁹⁰. The mutant allele of *ZmPrr7* is a 5.8-kb deletion in the gene region that leads to the total loss of protein function. Plants harboring the mutant allele exhibit significantly earlier flowering than the wild type in both tropical and temperate environments (Fig. 3c,d and Extended Data Fig. 8). The loss-of-function allele of *ZmCOL9* includes a 5-bp deletion/1-bp insertion in the intron and a 2-bp deletion/4-bp deletion in the third exon (Extended Data Fig. 9a,d), which result in premature translation termination. In a tropical environment (Hainan, China; 109° E, 18° N), *ZmCOL9* knockout mutants showed no difference in flowering time

compared with the wild type (Fig. 3e and Extended Data Fig. 9b,e), but plants with overexpression exhibited a later flowering phenotype (Fig. 3f and Extended Data Fig. 10a,b). In contrast, when planted in a temperate environment (Jilin, China; 125° E, 44° N), *ZmCOL9* knockout mutants flowered earlier (Fig. 3e and Extended Data Fig. 9c,f) and overexpression lines flowered later than the wild type (Fig. 3f,g and Extended Data Fig. 10c,d). These results confirm key roles for both *ZmPrr7* and *ZmCOL9* in regulating flowering time and contributing to the adaptation of highland teosintes and modern maize.

Discussion

The twin projections of increasing human population and decreasing suitable farmland highlight the challenge breeders face in producing high crop yields and this has motivated an increasing interest in crop wild relatives as sources of genetic diversity for improvement^{91,92}. Here, we present a high-resolution genetic variation map that greatly expands the publicly available genetic sequence information for the genus *Zea*. All of the data and results of this work have been integrated into the ZEAMAP database⁹³ for easy query and retrieval.

We provide a complete picture of the phylogeny and demography of the genus *Zea* using genome-wide data, including both divergence times and effective population sizes of *Zea* species. We reaffirm several aspects of the phylogeny of *Zea*, but our data identify a number of new features, including the likely subspecies status of *nicaraguensis*, the short divergence times between the perennial taxa and the relatively young age of the genus. We caution that our divergence estimate for *Tripsacum* may be underestimated because of the difficulty of mapping short reads from divergent genomes; therefore, high-quality *Tripsacum* and teosinte reference genomes will be essential to better answer this question⁹⁴.

Our broad sampling of the genus allows us to take advantage of population genetics tools to identify candidate genes involved in adaptation across both long and short time scales. We find evidence of convergent adaptation of highland teosintes and high-latitude maize, exemplifying the utility of studying variation in wild relatives to identify genes important in crops. Finally, we validate these approaches using genome editing to knock out two candidate flowering-time genes.

It is particularly noteworthy that our work identifies a vast trove of genetic variation that is absent in cultivated maize and even in its closest wild relative *parviflora*. Our functional analysis of candidate adaptation genes clarifies the great potential in the utilization of the wild relatives of maize in identifying novel alleles or highlighting potential genes for subsequent editing, potentially accelerating modern genetic improvements⁹⁵. The data and discoveries presented in this study provide a foundation for the use of crop wild relative resources for breeding in the face of increasing human populations and decreasing farmland.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41588-022-01184-y>.

References

1. Ray, D. K., Mueller, N. D., West, P. C. & Foley, J. A. Yield trends are insufficient to double global crop production by 2050. *PLoS ONE* **8**, e66428 (2013).
2. Bailey-Serres, J., Parker, J. E., Ainsworth, E. A., Oldroyd, G. E. D. & Schroeder, J. I. Genetic strategies for improving crop yields. *Nature* **575**, 109–118 (2019).
3. Lesk, C., Rowhani, P. & Ramankutty, N. Influence of extreme weather disasters on global crop production. *Nature* **529**, 84–87 (2016).
4. Challinor, A. J. et al. A meta-analysis of crop yield under climate change and adaptation. *Nat. Clim. Change* **4**, 287–291 (2014).
5. Tigchelaar, M., Battisti, D. S., Naylor, R. L. & Ray, D. K. Future warming increases probability of globally synchronized maize production shocks. *Proc. Natl Acad. Sci. USA* **115**, 6644–6649 (2018).
6. Li, Q. & Yan, J. Sustainable agriculture in the era of omics: knowledge-driven crop breeding. *Genome Biol.* **21**, 154 (2020).
7. Matsuoka, Y. et al. A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl Acad. Sci. USA* **99**, 6080–6084 (2002).
8. Piperno, D. R., Ranere, A. J., Holst, I., Iriarte, J. & Dickau, R. Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *Proc. Natl Acad. Sci. USA* **106**, 5019–5024 (2009).
9. Eyre-Walker, A., Gaut, R. L., Hilton, H., Feldman, D. L. & Gaut, B. S. Investigation of the bottleneck leading to the domestication of maize. *Proc. Natl Acad. Sci. USA* **95**, 4441–4446 (1998).
10. Tenaillon, M. I., U'Ren, J., Tenaillon, O. & Gaut, B. S. Selection versus demography: a multilocus investigation of the domestication process in maize. *Mol. Biol. Evol.* **21**, 1214–1225 (2004).
11. Wright, S. I. et al. The effects of artificial selection on the maize genome. *Science* **308**, 1310–1314 (2005).
12. Beissinger, T. M. et al. Recent demography drives changes in linked selection across the maize genome. *Nat. Plants* **2**, 16084 (2016).
13. Swarts, K. et al. Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. *Science* **357**, 512–515 (2017).
14. Zampieri, M. et al. When will current climate extremes affecting maize production become the norm? *Earth's Future* **7**, 113–122 (2019).
15. IPCC Climate Change 2022: Impacts, Adaptation and Vulnerability (eds Pörtner, H.-O. et al.) (Cambridge Univ. Press, 2022).
16. Zhang, H., Li, Y. & Zhu, J. K. Developing naturally stress-resistant crops for a sustainable agriculture. *Nat. Plants* **4**, 989–996 (2018).
17. Hufford, M. B., Bilinski, P., Pyhäjärvi, T. & Ross-Ibarra, J. Teosinte as a model system for population and ecological genomics. *Trends Genet.* **28**, 606–615 (2012).
18. Sánchez González, J. J. et al. Ecogeography of teosinte. *PLoS ONE* **13**, e0192676 (2018).
19. Mammadov, J. et al. Wild relatives of maize, rice, cotton, and soybean: treasure troves for tolerance to biotic and abiotic stresses. *Front Plant Sci.* **9**, 886 (2018).
20. Barnes, A. C. et al. An adaptive teosinte *mexicana* introgression modulates phosphatidylcholine levels and is associated with maize flowering time. *Proc. Natl Acad. Sci. USA* **119**, e2100036119 (2022).
21. Wang, H. et al. A teosinte-derived allele of a MYB transcription repressor confers multiple disease resistance in maize. *Mol. Plant* **14**, 1846–1863 (2021).
22. Karn, A., Gillman, J. D. & Flint-Garcia, S. A. Genetic analysis of teosinte alleles for kernel composition traits in maize. *G3 (Bethesda)* **7**, 1157–1164 (2017).
23. Li, K. et al. Large-scale metabolite quantitative trait locus analysis provides new insights for high-quality maize improvement. *Plant J.* **99**, 216–230 (2019).
24. Mano, Y., Omori, F. & Takeda, K. Construction of intraspecific linkage maps, detection of a chromosome inversion, and mapping of QTL for constitutive root aerenchyma formation in the teosinte *Zea nicaraguensis*. *Mol. Breed.* **29**, 137–146 (2012).
25. Mano, Y. & Omori, F. Flooding tolerance in interspecific introgression lines containing chromosome segments from teosinte (*Zea nicaraguensis*) in maize (*Zea mays* subsp. *mays*). *Ann. Bot.* **112**, 1125–1139 (2013).
26. De Lange, E. S., Balmer, D., Brigitte, M.-M. & Turlings, T. C. J. Insect and pathogen attack and resistance in maize and its wild ancestors, the teosintes. *New Phytol.* **204**, 329–341 (2014).
27. Lennon, J. R., Krakowsky, M., Goodman, M., Flint-Garcia, S. & Balint-Kurti, P. J. Identification of alleles conferring resistance to gray leaf spot in maize derived from its wild progenitor species teosinte. *Crop Sci.* **56**, 222–225 (2015).
28. Lennon, J. R., Krakowsky, M., Goodman, M., Flint-Garcia, S. & Balint-Kurti, P. J. Identification of teosinte alleles for resistance to southern leaf blight in near isogenic maize lines. *Crop Sci.* **57**, 1973–1983 (2017).
29. Hufford, M. B. et al. The genomic signature of crop-wild introgression in maize. *PLoS Genet.* **9**, e1003477 (2013).
30. Calfee, E. et al. Selective sorting of ancestral introgression in maize and teosinte along an elevational cline. *PLoS Genet.* **17**, e1009810 (2021).
31. Gaut, B. S. & Clegg, M. T. Molecular evolution of the *Adh1* locus in the genus *Zea*. *Proc. Natl Acad. Sci. USA* **90**, 5095–5099 (1993).
32. Hilton, H. & Gaut, B. S. Speciation and domestication in maize and its wild relatives: evidence from the globulin-1 gene. *Genetics* **150**, 863–872 (1998).
33. White, S. E. & Doebley, J. F. The molecular evolution of terminal *ear1*, a regulatory gene in the genus *Zea*. *Genetics* **153**, 1455–1462 (1999).
34. Ross-Ibarra, J., Tenaillon, M. & Gaut, B. S. Historical divergence and gene flow in the genus *Zea*. *Genetics* **181**, 1399–1413 (2009).
35. Wang, Q. & Dooner, H. K. Dynamic evolution of *bz* orthologous regions in the *Andropogoneae* and other grasses. *Plant J.* **72**, 212–221 (2012).
36. Buckler, E. S. 4th, Goodman, M. M., Holtsford, T. P., Doebley, J. F. & Sánchez, G. J. Phylogeography of the wild subspecies of *Zea mays*. *Maydica* **51**, 123–134 (2006).
37. Fukunaga, K. et al. Genetic diversity and population structure of teosinte. *Genetics* **169**, 2241–2254 (2005).
38. Sánchez, G. J. J. et al. Three new teosintes (*Zea* spp., Poaceae) from México. *Am. J. Bot.* **98**, 1537–1548 (2011).
39. Lamb, J. C. & Birchler, J. A. Retroelement genome painting: cytological visualization of retroelement expansions in the genera *Zea* and *Tripsacum*. *Genetics* **173**, 1007–1021 (2006).
40. Tenaillon, M. I., Hufford, M. B., Gaut, B. S. & Ross-Ibarra, J. Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zea luxurians*. *Genome Biol. Evol.* **3**, 219–229 (2011).
41. Chia, J. M. et al. Maize HapMap2 identifies extant variation from a genome in flux. *Nat. Genet.* **44**, 803–807 (2012).
42. Ting, Y. C. Spontaneous chromosome inversions of Guatemalan teosintes (*Zea mexicana*). *Genetica* **36**, 229–242 (1965).
43. Ting, Y. C. Common inversion in maize and teosinte. *Am. Nat.* **101**, 87–89 (1967).
44. Ting, Y. C. Chromosome polymorphism of teosinte. *Genetics* **83**, 737–742 (1976).
45. Fang, Z. et al. Megabase-scale inversion polymorphism in the wild ancestor of maize. *Genetics* **191**, 883–894 (2012).

46. Pyhäjärvi, T., Hufford, M. B., Mezmouk, S. & Ross-Ibarra, J. Complex patterns of local adaptation in teosinte. *Genome Biol. Evol.* **5**, 1594–1609 (2013).
47. Yang, N. et al. Contributions of *Zea mays* subspecies *mexicana* haplotypes to modern maize. *Nat. Commun.* **8**, 1874 (2017).
48. Fustier, M. A. et al. Common gardens in teosintes reveal the establishment of a syndrome of adaptation to altitude. *PLoS Genet.* **15**, e1008512 (2019).
49. Janzen, G. M. et al. Demonstration of local adaptation in maize landraces by reciprocal transplantation. *Evol Appl.* **5**, 817–837 (2022).
50. Fustier, M. A. et al. Signatures of local adaptation in lowland and highland teosintes from whole-genome sequencing of pooled samples. *Mol. Ecol.* **26**, 2738–2756 (2017).
51. Aguirre-Liguori, J. A. et al. Divergence with gene flow is driven by local adaptation to temperature and soil phosphorus concentration in teosinte subspecies (*Zea mays parviflora* and *Zea mays mexicana*). *Mol. Ecol.* **28**, 2814–2830 (2019).
52. Yang, X. et al. Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol. Breed.* **28**, 511–526 (2011).
53. Flouri, T., Jiao, X., Rannala, B. & Yang, Z. Species tree inference with BPP using genomic sequences and the multispecies coalescent. *Mol. Biol. Evol.* **35**, 2585–2593 (2018).
54. Lee, T. H., Guo, H., Wang, X., Kim, C. & Paterson, A. H. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* **15**, 162 (2014).
55. Wang, X. et al. Genome alignment spanning major Poaceae lineages reveals heterogeneous evolutionary rates and alters inferred dates for key evolutionary events. *Mol. Plant* **8**, 885–898 (2015).
56. Buckler, E. S. 4th & Holtsford, T. P. Zea systematics: ribosomal ITS evidence. *Mol. Biol. Evol.* **13**, 612–622 (1996).
57. Schiffels, S. & Wang, K. MSMC and MSMC2: the Multiple Sequentially Markovian Coalescent. *Methods Mol. Biol.* **2090**, 147–166 (2020).
58. Janzen, G. M., Wang, L. & Hufford, M. B. The extent of adaptive wild introgression in crops. *New Phytol.* **221**, 1279–1288 (2019).
59. Van Heerwaarden, J., Hufford, M. B. & Ross-Ibarra, J. Historical genomics of North American maize. *Proc. Natl Acad. Sci. USA* **109**, 12420–12425 (2012).
60. 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).
61. Bellon, M. R. et al. Evolutionary and food supply implications of ongoing maize domestication by Mexican campesinos. *Proc. Biol. Sci.* **285**, 20181049 (2018).
62. Finnegan, D. J. Eukaryotic transposable elements and genome evolution. *Trends Genet.* **5**, 103–107 (1989).
63. Hufford, M. B. et al. De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**, 655–662 (2021).
64. Faria, R., Johannesson, K., Butlin, R. K. & Westram, A. M. Evolving inversions. *Trends Ecol. Evol.* **34**, 239–248 (2019).
65. Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434 (2006).
66. Crow, T. et al. Gene regulatory effects of a large chromosomal inversion in highland maize. *PLoS Genet.* **16**, e1009213 (2020).
67. Li, H. & Ralph, P. Local PCA shows how the effect of population structure differs along the genome. *Genetics* **211**, 289–304 (2019).
68. Wei, S., Dai, Y., Duan, Q., Liu, B. & Hua, Y. A global soil data set for earth system modeling. *J. Adv. Model. Earth Syst.* **6**, 249–263 (2014).
69. Silva, R. et al. *Gluconacetobacter diazotrophicus* changes the molecular mechanisms of root development in *Oryza sativa* L. growing under water stress. *Int. J. Mol. Sci.* **21**, 333 (2020).
70. Wang, C. et al. Mutation in xyloglucan 6-xylosyltransferase results in abnormal root hair development in *Oryza sativa*. *J. Exp. Bot.* **65**, 4149–4157 (2014).
71. Corbett-Detig, R. B. & Hartl, D. L. Population genomics of inversion polymorphisms in *Drosophila melanogaster*. *PLoS Genet.* **8**, e1003056 (2012).
72. Lowry, D. B. & Willis, J. H. A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biol.* **8**, e1000500 (2010).
73. Yant, L. et al. Meiotic adaptation to genome duplication in *Arabidopsis arenosa*. *Curr. Biol.* **23**, 2151–2156 (2013).
74. Ma, A. et al. The genetics and genome-wide screening of regrowth loci, a key component of perennialism in *Zea diploperennis*. *G3 (Bethesda)* **9**, 1393–1403 (2019).
75. Guo, Z. et al. Identification of major QTL for waterlogging tolerance in maize using genome-wide association study and bulked sample analysis. *J. Appl. Genet.* **62**, 405–418 (2021).
76. Yu, F. et al. Dissecting the genetic architecture of waterlogging stress-related traits uncovers a key waterlogging tolerance gene in maize. *Theor. Appl. Genet.* **131**, 2299–2310 (2018).
77. Osman, K. A. et al. Dynamic QTL analysis and candidate gene mapping for waterlogging tolerance at maize seedling stage. *PLoS ONE* **8**, e79305 (2013).
78. Kim, H. J. et al. Time-evolving genetic networks reveal a NAC troika that negatively regulates leaf senescence in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **115**, E4930–E4939 (2018).
79. Tang, J. et al. GDSL lipase occluded stomatal pore 1 is required for wax biosynthesis and stomatal cuticular ledge formation. *New Phytol.* **228**, 1880–1896 (2020).
80. Yan, L. et al. Parallels between natural selection in the cold-adapted crop-wild relative *Tripsacum dactyloides* and artificial selection in temperate adapted maize. *Plant J.* **99**, 965–977 (2019).
81. Buckler, E. S. et al. The genetic architecture of maize flowering time. *Science* **325**, 714–718 (2009).
82. Wang, L. et al. Molecular parallelism underlies convergent highland adaptation of maize landraces. *Mol. Biol. Evol.* **38**, 3567–3580 (2021).
83. Li, Y. X. et al. Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. *Plant J.* **86**, 391–402 (2016).
84. Bouché, F., Loyer, G., Tocquin, P. & Périlleux, C. FLOR-ID: an interactive database of flowering-time gene networks in *Arabidopsis thaliana*. *Nucleic Acids Res.* **44**, D1167–D1171 (2016).
85. Hayama, R., Yokoi, S., Tamaki, S., Yano, M. & Shimamoto, K. Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature* **422**, 719–722 (2003).
86. Seo, E. et al. Crosstalk between cold response and flowering in *Arabidopsis* is mediated through the flowering-time gene SOC1 and its upstream negative regulator FLC. *Plant Cell* **21**, 3185–3197 (2009).
87. Peng, H. & Neff, M. M. CIRCADIAN CLOCK ASSOCIATED 1 and ATAF2 differentially suppress cytochrome P450-mediated brassinosteroid inactivation. *J. Exp. Bot.* **71**, 970–985 (2020).
88. Liang, L. et al. The transcriptional repressor OsPRR73 links circadian clock and photoperiod pathway to control heading date in rice. *Plant Cell Environ.* **44**, 842–855 (2021).
89. Nakamichi, N. et al. *Arabidopsis* clock-associated pseudo-response regulators PRR9, PRR7 and PRR5 coordinately and positively regulate flowering time through the canonical CONSTANS-dependent photoperiodic pathway. *Plant Cell Physiol.* **48**, 822–832 (2007).

90. Liu, H. et al. High-throughput CRISPR/Cas9 mutagenesis streamlines trait gene identification in maize. *Plant Cell* **32**, 1397–1413 (2020).
91. Hufford, M. B., Martínez-Meyer, E., Gaut, B. S., Eguiarte, L. E. & Tenaillon, M. I. Inferences from the historical distribution of wild and domesticated maize provide ecological and evolutionary insight. *PLoS ONE* **7**, e47659 (2012).
92. Aguirre-Liguori, J. A. et al. Connecting genomic patterns of local adaptation and niche suitability in teosintes. *Mol. Ecol.* **26**, 4226–4240 (2017).
93. Gui, S. et al. ZEAMAP, a comprehensive database adapted to the maize multi-omics era. *iScience* **23**, 101241 (2020).
94. Gault, C. M., Kremling, K. A. & Buckler, E. S. *Tripsacum* de novo transcriptome assemblies reveal parallel gene evolution with maize after ancient polyploidy. *Plant Genome* **11**, 1–13 (2018).
95. Fernie, A. R. & Yan, J. Targeting key genes to tailor old and new crops for a greener agriculture. *Mol. Plant* **13**, 354–356 (2020).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022

Methods

Samples and whole-genome resequencing

A total of 237 teosinte accessions from CIMMYT, the United States Department of Agriculture and collaborators were obtained, consisting of 90 *mexicana*, 79 *parviflora*, 20 *diploperennis*, 15 *perennis*, 15 *luxurians*, 13 *nicaraguensis* and five *huehuetenangensis* species, according to morphological classification (Supplementary Tables 1 and 2). Two *Tripsacum dactyloides* were obtained from the laboratory of F. Chen (Henan Agricultural University, China). Young leaves were used for DNA extraction for sequencing using the Illumina HiSeq 3000 platform (150-bp paired-end reads; conducted by BGI, Shenzhen, China) and NovaSeq 6000 platform (150-bp paired-end reads; conducted by Novogene, Sacramento, USA). The DNA sequencing data of 507 cultivated maizes were downloaded from the NCBI Sequence Read Archive database ([PRJNA531553](#); Supplementary Table 1).

Read mapping and SNP calling

Raw reads of teosintes were first processed using FastQC (version 0.11.3; <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Trimmomatic⁹⁶ (version 0.33; HiSeq 3000 platform; LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) and fastp⁹⁷ (version 0.19.4; NovaSeq 6000 platform; -g -l 36) were used to remove poor-quality base calls and adapters. Reads of teosintes and maize were then aligned to the B73 reference genome⁹⁸ (version 4) using Bowtie 2 (ref.⁹⁹; version 2.1.0; --very-fast). Unique mapped reads were sorted and indexed using Picard (version 1.119; <http://broadinstitute.github.io/picard/>). SAMtools¹⁰⁰ (version 1.3.1) and the UnifiedGenotyper module from GATK (version 3.5; <https://software.broadinstitute.org/gatk/>) were used to estimate the variants for each individual. Hard filtering of the individual SNP calls was carried out with a mapping quality of ≤ 20.0 , a minimum sequencing coverage of ≤ 5 and a maximum sequencing coverage of ≥ 200 . Then, variants from the 237 teosintes and 507 maizes were combined using GATK CombineVariants to a single variant-calling file. To confirm whether the unknown variants were discarded reference genotypes in each individual call, we recalled these sites and replaced them with reference genotypes if they had supported reads. Finally, sites with a missing rate $> 75\%$ in all samples were excluded. To validate the accuracy of SNPs called from resequencing data, 224 sites in 80 accessions were selected for Sanger sequencing (Supplementary Table 3).

Population structure and PCA

We evaluated patterns of population structure using a set of SNPs that were filtered to remove multi-allelic loci and with a MAF of < 0.05 (--maf 0.05 -biallelic-only) using PLINK¹⁰¹ (version 1.9). We then ran admixture¹⁰² for different values of the number of clusters (K) from 2 to 20 (-cv = 10; version 1.3.0). Each individual with admixture components < 0.6 was classified as teosinte (mix) or maize (mix). We performed PCA using this same set of SNPs with GCTA¹⁰³ (version 1.26) by recording the first ten components (-pca 10).

Phylogenetic tree construction

We annotated SNPs with a missing data rate of < 0.7 in teosinte and maize with SnpEff (version 4.1g; <http://snpEff.sourceforge.net/index.html>) using the first transcript of B73 v4 genes. Then, the synonymous and non-coding SNPs were used to construct a simple phylogenetic tree with SNPhylo⁵⁴ (version 20140701) under default parameters and the tree was visualized with iTOL¹⁰⁴.

Species tree analysis

Species delimitation and species trees were inferred using BPP⁵³ (model A11; version 4.1.4). We used the following samples in BPP: three tropical maizes, three *parviflora*, three *mexicana*, three *nicaraguensis*, three *diploperennis*, three *perennis*, three *luxurians*, two *huehuetenangensis* and two *T. dactyloides* (Supplementary Table 1). Low-quality base calls

and adapters from raw reads of *T. dactyloides* were removed using Trimmomatic, and the remaining sequences were aligned to the B73 version 4 reference genome using Bowtie 2, as described in the section 'Read mapping and SNP calling'. The consensus base was estimated from the uniquely mapped reads using ANGSD¹⁰⁵ (version 0.930). Using the B73 annotation, we randomly selected 2,000 coding sequence genes to estimate the species delimitation and species tree. The prior distribution of ancestral population size (θ) and divergence time from the root (τ) followed an inverse-gamma (IG) prior with means of 0.005 (IG (3, 0.01)) and 0.75 (IG (3, 1.5)), respectively. The consensus of A11 species trees was visualized using DensiTree¹⁰⁶ (version 2.2.6).

Imputation and demographic estimation

SNPs in the 237 teosintes and 507 maizes were imputed with BEAGLE¹⁰⁷ (version 4.0). Divergence times within teosintes and the effective population size of each teosinte were estimated using BPP (A00 model) and MSMC2 (ref.⁵⁷; version 2.1.1). The topological tree in BPP (A00 model) was fixed as the species tree with the highest posterior probability (A11 model) estimated from the above species tree analysis. Sequences used in the A11 model were applied to estimate the effective population size and divergence time using the same priors as above. In MSMC2, four haplotype models were applied (Supplementary Table 1). The mutation rate used in BPP (A00 model) and MSMC2 was 3×10^{-8} (ref.¹⁰⁸).

ABBA–BABA test

We used Patterson's D statistic^{109,110} to test for introgression between teosintes. Assuming *T. dactyloides* to be the outgroup (O), we assessed D statistics for the tree (((P1, P2), P3), O), where P1, P2 and P3 represent different taxa in *Zea* (the autotetraploid *perennis* was excluded). The numbers of ABBA and BABA patterns in each block were calculated in ANGSD (-blockSize 10000). To overcome the problem of non-independence within the sequence, a block-jackknifing procedure was used to test for statistical significance.

Divergence-based introgression polarization test

To estimate the directions of introgression, the consensus base was estimated from the uniquely mapped reads using ANGSD to represent individuals in different taxa of *Zea* and *Tripsacum* (eight taxa in total). The whole-genome consensus files from different taxa were then concatenated into multiple sequence alignment files by different chromosomes. Finally, this eight-taxon alignment was pruned to contain four taxa, according to each test as shown in Supplementary Fig. 4, and divided into 5,000-bp windows, which were used as the input of DIP¹¹¹.

Linkage disequilibrium, nucleotide diversity and F_{ST} calculation

The linkage disequilibria (r^2) of *nicaraguensis* (14), *luxurians* (14), *diploperennis* (20), *perennis* (15), *huehuetenangensis* (five), *mexicana* (81), *parviflora* (70) and maize (507) were estimated for all biallelic SNPs within 500 kb window (--geno 0.5 --maf 0.05 --biallelic-only --snps-only) using PLINK. The nucleotide diversity of *nicaraguensis* (14), *luxurians* (14), *diploperennis* (20), *perennis* (15), *huehuetenangensis* (five), *mexicana* (81), *parviflora* (70) and maize (110 randomly selected individuals) was calculated using ANGSD (version 0.930; -doMaf 1 -doMajorMinor 1 -uniqueOnly 1 -minMapQ 30 -minQ 20 -GL 2 -fold 1 -win 5000 -step 5000). The differentiation (F_{ST}) between maize and teosinte for five randomly selected samples was estimated in VCFTools¹¹² (version 0.1.16; --fst-window-size 5000).

Taxon-specific SNPs, indels and k-mer analysis

SNPs and indels found only in one specific taxon of *Zea* (supported by at least two individuals) were regarded as taxon-specific SNPs. The longest transcripts of each gene in the B73 annotation and a recent atlas of *cis*-regulatory elements⁶⁰ were used to annotate variants. K-mers

unique to each taxon that appeared at least twice were obtained with sourmash¹¹³ (version 3.2.0; --scaled 1000).

Transposon element analysis

RepeatExplore2 (ref. ¹¹⁴; version 0.3.8) was used to identify repeat clusters of each taxon of *Zea* (two samples were randomly selected from each taxon). Clusters were further annotated by applying RepeatMasker (<http://www.repeatmasker.org/>; version 4.1.0; -species maize). Reads were mapped to the above repeat clusters using BWA-MEM¹¹⁵ (version 0.7.10) and the number of mapped reads in each repeat cluster was calculated with SAMTools. The abundance of the repeat elements between samples was normalized by their sequenced library sizes.

Inversion calling

Localized heterogeneity across chromosomes was identified using losstruct⁶⁷ (version 1.0) in windows containing 10,000 SNPs. The most related 5% of windows in each chromosome around one of the four outliers (maximum, minimum, MDS1 or MDS2) were regarded as candidate inversions and were genotyped using invClust¹¹⁶ (version 1.0) with the genotype of B73 as the reference state. Genotypes of the candidates were confirmed via PCA of the SNPs in the corresponding region. Only candidates with three clearly different haplotypes identified by PCA were regarded as true inversions. Candidates near the centromeres were filtered out. Centromere information was obtained by combining the locations from all individuals in the nested association mapping population⁶³.

Genome-wide association analysis

SNPs from *mexicana* were obtained from the imputed teosinte panel according to the name of samples, then the population structure was calculated with admixture (version 1.3.0; --cv = 10; K = 1, 2, 3, 4, 5). The *K* value with the lowest cross validation error (*K* = 2) was used in the downstream analysis. Estimation of the kinship matrix and association analyses using the compressed mixed linear model were performed using TASSEL3 (ref. ¹¹⁷; version 3.0.174) with a *P* value cut off set to 1/*N* (where *N* is the number of tested SNPs; Bonferroni test). Latitude and longitude information was obtained from the laboratory of S. Taba. Global soil properties used as phenotypes for the genome-wide association study were extracted using the R package ncdf4 (version 1.16; <http://cirrus.ucsd.edu/~pierce/ncdf/>) from the Global Soil Dataset for Earth System Modeling⁶⁸—a comprehensive database with eight layers to a depth of 2.3 m (0–0.045, 0.045–0.091, 0.091–0.166, 0.166–0.289, 0.289–0.493, 0.493–0.829, 0.829–1.383 and 1.383–2.296 m). To find the best cluster method and number, all the soil properties were clustered using the R package clValid¹¹⁸ (version 0.7), underling hierarchical, *k*-means and *k*-medoides method in combination with clusters from 2 to 40. Genome-wide association studies were performed on a subset of nine features identified by hierarchical cluster analysis (Supplementary Fig. 10).

Identification of adaptive regions in non-*Z. mays* taxa

Whole-genome adaptive genetic variation between different non-*Z. mays* taxa and maize was estimated by calculating their *F_{ST}* values with VCFTools (-fst-window-size 5000). Under each comparison, all available teosinte and maize samples were used. We then Z-transformed the *F_{ST}* in each window (windows with *ZF_{ST}* values exceeding the 95th percentile of the whole genome were declared as candidate adaptive regions). Gene Ontology enrichment analysis was conducted using PANTHER with default parameters^{119,120} (<https://doi.org/10.5281/zenodo.5725227>; released 16 November 2021) and visualized with GlueGo¹²¹ (version 2.5.9).

Selective sweeps in teosintes and maize

Whole-genome scanning for regions of teosinte elevation adaptation and maize temperate adaptation was implemented using a mixed method. First, two genetic maps were obtained from a B73 × teosinte population¹²²

and a maize B73 × By804 population¹²³ and the physical locations were converted to coordinates of the B73 version 4 reference sequence using CrossMap¹²⁴ (version 0.2.9). The genetic distances between SNPs in *mexicana* and *parviglumis* were then calculated based on the B73 × teosinte genetic map, while the distances in temperate maize and tropical maize were calculated based on the B73 × By804 genetic map. Genetic distances between SNPs located between the genetic markers were assigned based on their physical distances. The likelihood of multi-locus allele frequency differentiation between two tested populations was modeled using XP-CLR¹²⁵ (version 1.0; -w1 0.005 100 1000 -p0 0.7) in both the teosinte group (*mexicana*, with *parviglumis* as the reference) and the maize group (temperate maize, with tropical maize as the reference). Finally, we applied a spline window method (GenWin¹²⁶ version 0.1; smoothness = 100) to smooth the results. The top 5% of genomic regions with the highest *W* statistic in *mexicana* and *parviglumis* were regarded as candidate teosinte altitude adaptation regions and the top 5% of *W* statistic regions in temperate and tropical maize were regarded as candidate maize temperate adaptation regions. Enrichment analysis between candidate teosinte altitude adaptation regions and maize temperate adaptation was conducted using the shuffle function (-excl -noOverlapping) in BEDTools¹²⁷ (version 2.25.0). Genes, including the promoter (2 kb away the transcription start site), that overlapped with the regions identified above were regarded as candidate adaptive genes.

RNA sequencing sampling, library construction and data analysis

The base tissues of V5-stage shoots (1–2 cm) of maize (five tropical maize and five temperate maize) and teosintes (three *parviglumis* and three *mexicana*) were sampled for messenger RNA and total RNA extraction using the Quick RNA Isolation Kit (Huayueyang Biotechnology). Both messenger RNA and total RNA samples were used for library preparation according to Illumina strand-specific library construction protocols. Paired-end libraries were sequenced using a mixture of platforms (HiSeq 3000, X10 and NovaSeq) with 150 cycles. Raw reads were filtered to remove the poor-quality base calls and adapters specifically for each platform (NovaSeq: fastp -g -l 36; X10: fastp -l 36; HiSeq 3000: Trimmomatic LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36). Reads were then aligned to the B73 reference genome (v4) using TopHat2 (ref. ¹²⁸; version 2.2.1) and read counts for each gene were calculated using htseq-count¹²⁹ (version 0.9.1). Finally, differentially expressed genes were identified between *mexicana* and *parviglumis*, as well as between temperate and tropical maize, using DESeq2 (ref. ¹³⁰; version 1.10.1), with an absolute fold change of >1 and *P* < 0.05.

Functional validation of *ZmPRR7* and *ZmCOL9*

Mutants of *ZmPRR7* and *ZmCOL9* were generated from a high-throughput genome editing design⁹⁰. In brief, line-specific single guide RNAs were filtered based on the assembled pseudo-genome of the receptor KN5585, then a double single-guide-RNA-pool approach was used to construct vectors. The vectors (CPB-ZmUbi-hspCas9) were transformed into the receptor KN5585 and the targets of each *T₀* individual were assigned by barcode-based sequencing. The genotype of gene-editing lines was identified by PCR amplification and Sanger sequencing using target-specific primers (Supplementary Table 24).

Transgenic lines generated with DNA fragments of *ZmCOL9* driven by the *ZmUbi* promoter were created using the modified binary vector pCAMBIA3300. Immature zygotic embryos of maize hybrid Hill (B73 × A188) were infected with *Agrobacterium tumefaciens* strain EHA105 harboring the binary vector, based on the published method for *ZmCOL9* (ref. ¹³¹). Transgenic plants were identified by real-time quantitative reverse transcription PCR as well as tests for herbicide resistance and the presence of the bar gene. The flowering-time phenotypes of transgenic plants with mutations in *ZmPRR7* and *ZmCOL9* were investigated in Jilin Province (125° E, 44° N) and Hainan Province (109° E, 18° N).

Reporting summary

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

Data availability

DNA and RNA sequencing reads from this study were deposited in the NCBI Sequence Read Archive with the accession codes [PRJNA641489](#), [PRJNA816255](#), [PRJNA816273](#) and [PRJNA645739](#). The SNP data can be downloaded from https://ftp.cngb.org/pub/CNSA/data3/CNP0001565/zeemap/02_Variants/PAN_Zea_Variants/Zea-vardb/. Source data are provided with this paper.

Code availability

All of the custom scripts used in this study are available at <https://doi.org/10.5281/zenodo.6818334> (ref.¹³²).

References

96. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
97. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
98. Jiao, Y. et al. Improved maize reference genome with single-molecule technologies. *Nature* **546**, 524–527 (2017).
99. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
100. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
101. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
102. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
103. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
104. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
105. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics* **15**, 356 (2014).
106. Bouckaert, R. R. DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* **26**, 1372–1373 (2010).
107. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
108. Clark, R. M., Tavaré, S. & Doebley, J. Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. *Mol. Biol. Evol.* **22**, 2304–2312 (2005).
109. Green, R. E. et al. A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
110. Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).
111. Forsythe, E. S., Sloan, D. B. & Beilstein, M. A. Divergence-based introgression polarization. *Genome Biol. Evol.* **12**, 463–478 (2020).
112. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
113. Brown, C. T. & Irber, L. sourmash: a library for MinHash sketching of DNA. *J. Open Source Softw.* **1**, 27 (2016).
114. Novák, P., Neumann, P. & Macas, J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* **11**, 378 (2010).
115. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
116. Cáceres, A. & González, J. R. Following the footprints of polymorphic inversions on SNP data: from detection to association tests. *Nucleic Acids Res.* **43**, e53 (2015).
117. Bradbury, P. J. et al. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635 (2007).
118. Brock, G., Pihur, V., Datta, S. & Datta, S. clValid: an R package for cluster validation. *J. Stat. Softw.* **25**, 1–22 (2008).
119. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
120. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).
121. Bindea, G. et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped Gene Ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
122. Liu, Z. et al. Expanding maize genetic resources with predomestication alleles: maize-teosinte introgression populations. *Plant Genome* <https://doi.org/10.3835/plantgenome2015.07.0053> (2016).
123. Pan, Q. et al. Genome-wide recombination dynamics are associated with phenotypic variation in maize. *New Phytol.* **210**, 1083–1094 (2016).
124. Zhao, H. et al. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006–1007 (2014).
125. Chen, H., Patterson, N. & Reich, D. Population differentiation as a test for selective sweeps. *Genome Res.* **20**, 393–402 (2010).
126. Beissinger, T. M., Rosa, G. J. M., Kaepller, S. M., Gianola, D. & de Leon, N. Defining window-boundaries for genomic analyses using smoothing spline techniques. *Genet. Sel. Evol.* **47**, 30 (2015).
127. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
128. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
129. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
130. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
131. Frame, B. R. et al. Agrobacterium tumefaciens-mediated transformation of maize embryos using a standard binary vector system. *Plant Physiol.* **129**, 13–22 (2002).
132. Chen, L. Zea_genus. Zenodo <https://doi.org/10.5281/zenodo.6818334> (2022).

Acknowledgements

We thank S. Taba from CIMMYT for providing the teosinte materials, J. Chen from Henan Agricultural University for providing the DNA of *T. dactyloides*, A. Kur from North Carolina State University for drawing the picture of teosinte morphological characteristics and T. AuBuchon-Elder and S. Fabbri for growing and supplying germplasm. Computation resources were provided by the high-throughput computing platform of the National Key Laboratory of Crop Genetic Improvement at Huazhong Agricultural University and supported by H. Liu. This research was supported by the National Key Research and Development Program of China (2020YFE0202300), National Natural Science Foundation of China (U1901201 and 31771879) and Science and Technology major program of Hubei Province (2021ABA011) (all to J.Y.), the Young Elite

Scientist Sponsorship Program of the China Association for Science and Technology (2019QNRC001 to N.Y), and the US National Science Foundation (grants 1546719 and 1822330) and United States Department of Agriculture Hatch project CADPLS2066H (both to J.R.-I.).

Author contributions

Jianbing Yan, J.R.-I. and N.Y. designed and supervised the study. Y.P., W.L., A.P., B.C., J.S.B., R.R.-A., R.J.H.S., Jiali Yan., Q.Z., S.W., S.G., Y.W., Y.L., C.J., M.D., M.J., J.L., L.J., Y.Yu., M.Z. and X.Yang. prepared the materials. X.Z. provided the variant calling pipeline. J.L. performed the Sanger validation of SNPs. W.W. uploaded the SNPs and indels to the database. L.C. and J.L. analyzed the data. M.J., X.L., Q.L., Y.Yin. and X.Ye. performed the genetic transformation and mutant validation. L.C. led the population genomics analyses. L.C., M.J., N.Y., M.B.H., A.R.F., M.L.W., J.R.-I. and Jianbing Yan. prepared the manuscript.

Competing interests

The authors M.J., X.L., J.X., H.L., Y.Yin., L.H., J.G., B.H. and Jianbing Yan. (Chinese patent application number 202110085348.5; “Methods and genes of changing the flowering time of maize”) and X.L., Q.L., M.J., X.L., J.X., H.L., Y.Yin., J.G., Y.L., L.H., B.H., Y.L., Jianbing Yan. and D.H.

(Chinese patent application number 202011480563.7; “Application of a maize plant height and ear height controlling gene ZmCOL14”) have filed patent applications on technology related to the processes described in this article. The other authors declare no competing interests.

Additional information

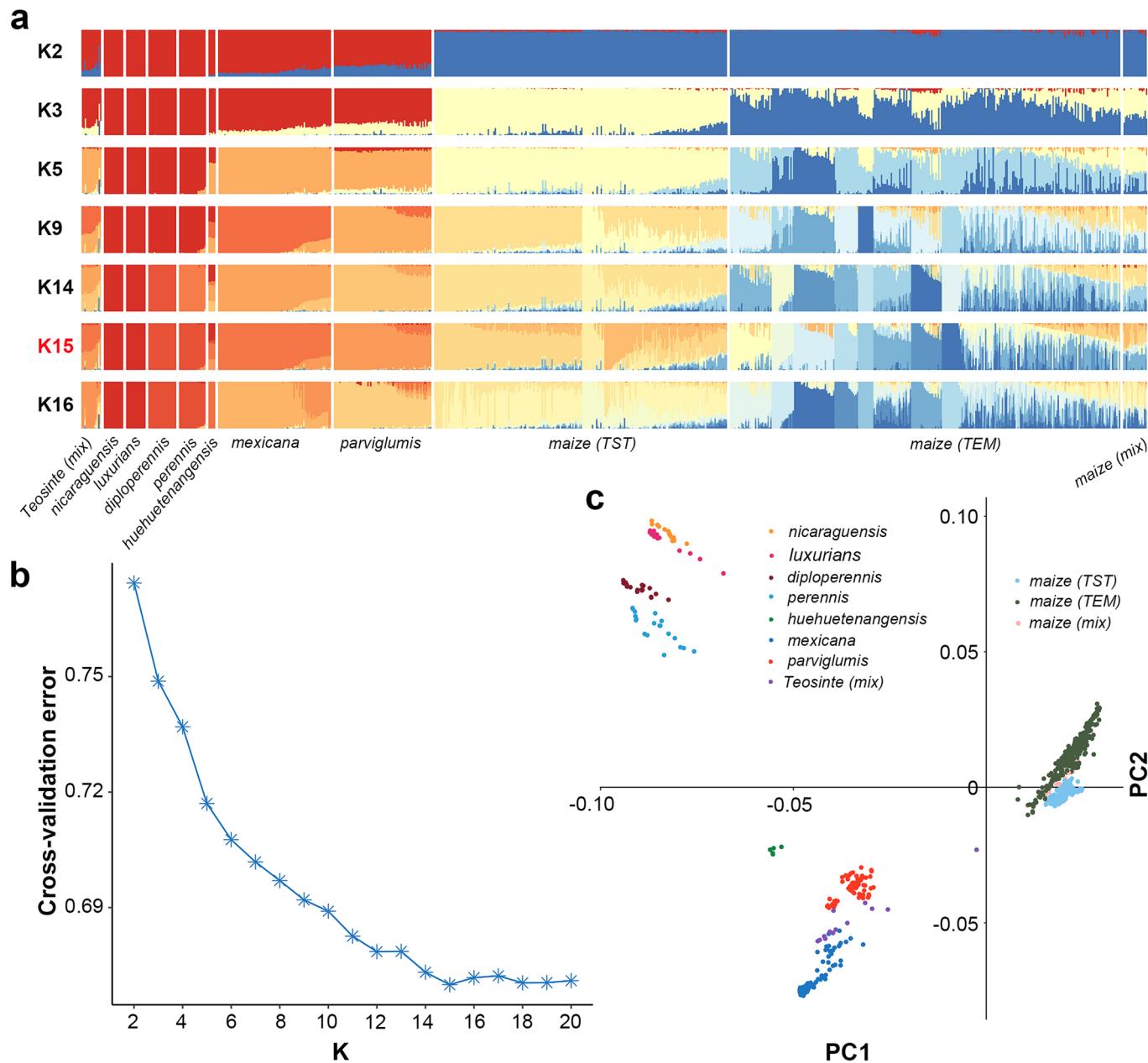
Extended data is available for this paper at
<https://doi.org/10.1038/s41588-022-01184-y>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41588-022-01184-y>.

Correspondence and requests for materials should be addressed to Ning Yang, Jeffrey Ross-Ibarra or Jianbing Yan.

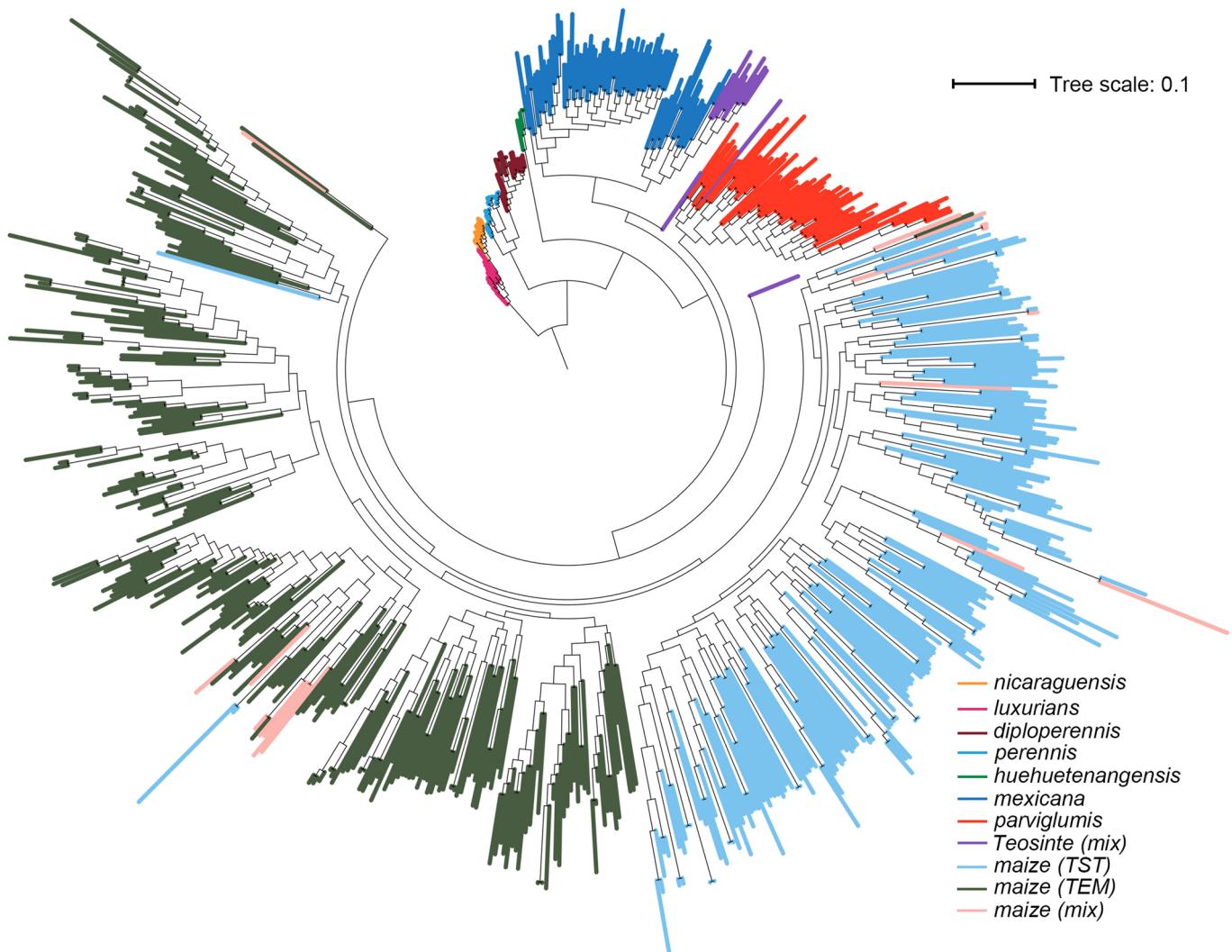
Peer review information *Nature Genetics* thanks Klaus Mayer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

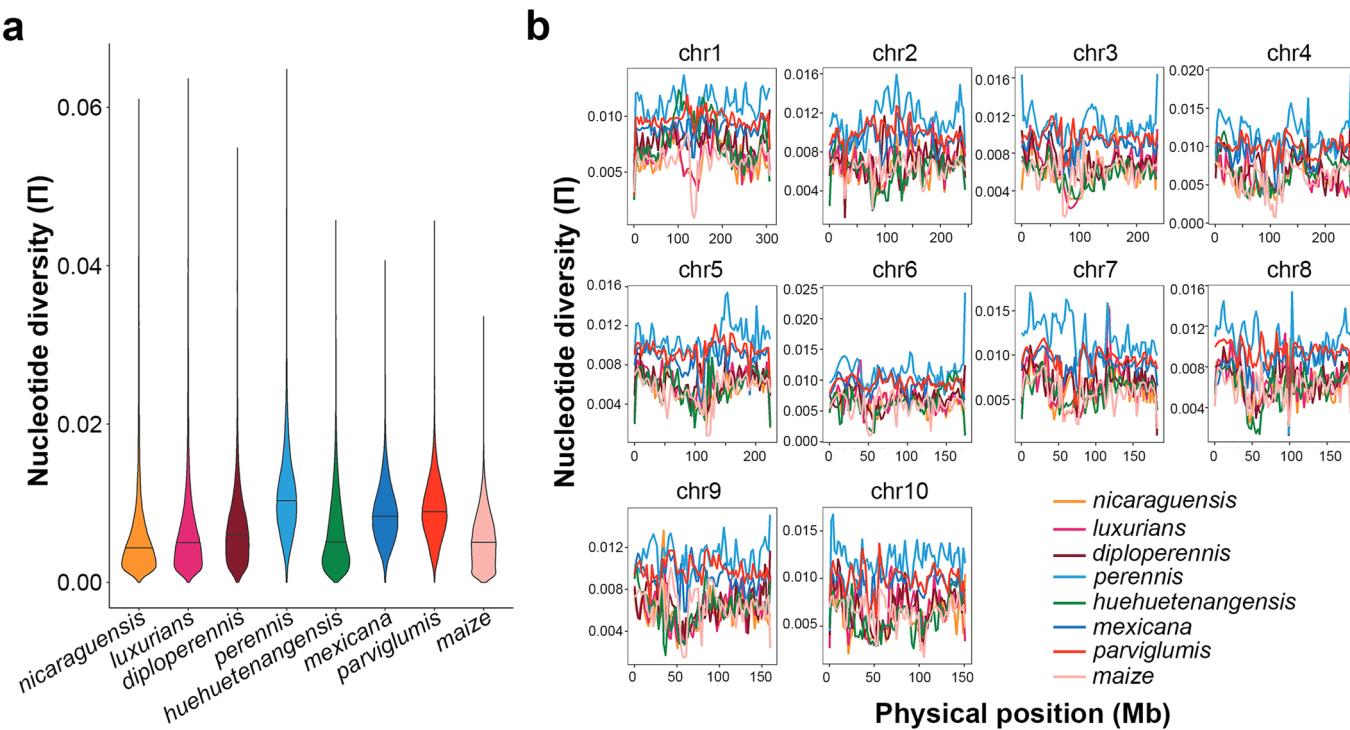


Extended Data Fig. 1 | STRUCTURE and principle component analysis (PCA) of 507 maize and 237 teosinte. **a**, Population structure for $K = 2\text{--}16$. maize (TEM) indicates temperate maize (including 8 components), maize (TST) indicates tropical maize (including 3 components). Samples with assignment to maize (TEM) or maize (TST) lower than 0.6 were classified as maize (mix), and *mexicana*

or *parviglumis* samples with assignment lower than 0.6 to any teosinte were classified as teosinte (mix). **b**, Cross-validation error for $K = 2\text{--}20$ showing $K = 15$ with the lowest cross validation error. **c**, PCA of maize and teosinte; points are colored according to the admixture result ($K = 15$).

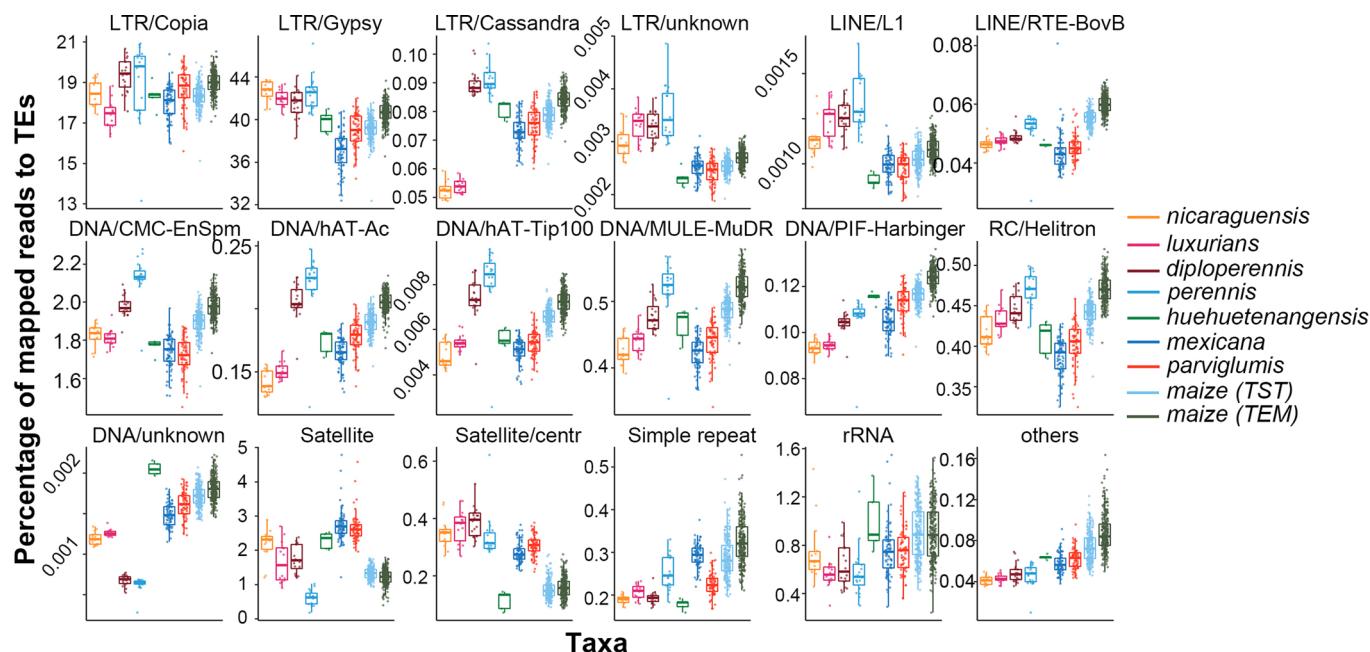


Extended Data Fig. 2 | Phylogenetic tree of *Zea* genus. The maximum likelihood tree was estimated with SNPhylo⁵⁴. Populations are colored based on the admixture result for K = 15.



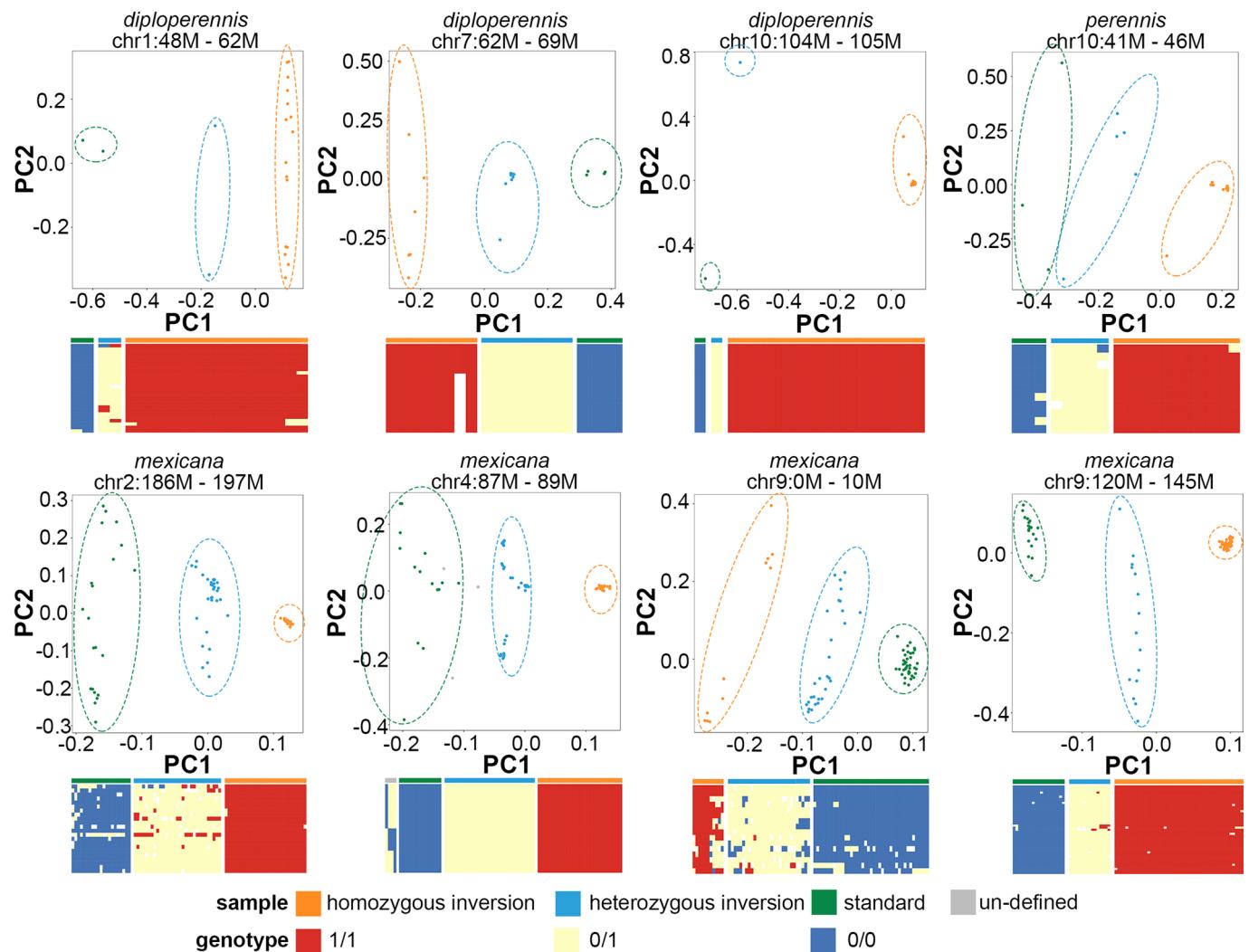
Extended Data Fig. 3 | The distribution of nucleotide diversity. **a**, Violin plot for nucleotide diversity analysis (calculated with 5 kb non-overlap window; windows less than 500 covered sites were excluded to filter out regions of poor alignment). Windows used in *nicaraguensis* (n = 60,778), *luxurians* (n = 51,956), *diploperennis* (n = 52,656), *perennis* (n = 51,306), *huehuetenangensis* (n = 107,481),

mexicana (n = 60,247), *parviflumis* (n = 61,014) and maize (n = 76,375). Cultivated maize was down-sampled to 110 randomly selected individuals. Horizontal lines in each violin plot represent the median value of nucleotide diversity. **b**, Line plot of nucleotide diversity analysis along the chromosomes.



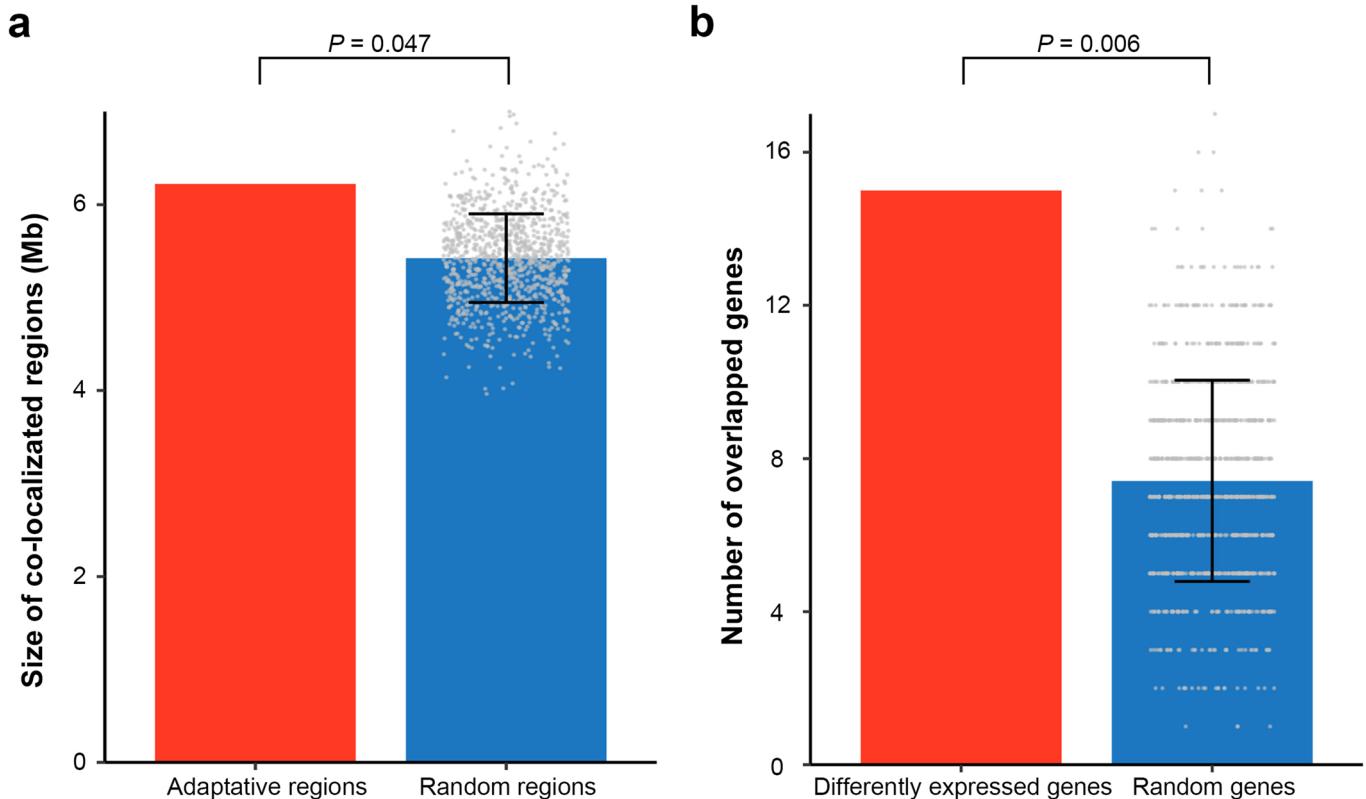
Extended Data Fig. 4 | Repeat comparison of *Zea*. Boxplot of reads mapped to different repeat classes across the samples in different taxa. Center lines in the boxplot indicate the median, edges represent the 25th and 75th percentiles, whiskers further extend by ± 1.5 times the interquartile range from the limits

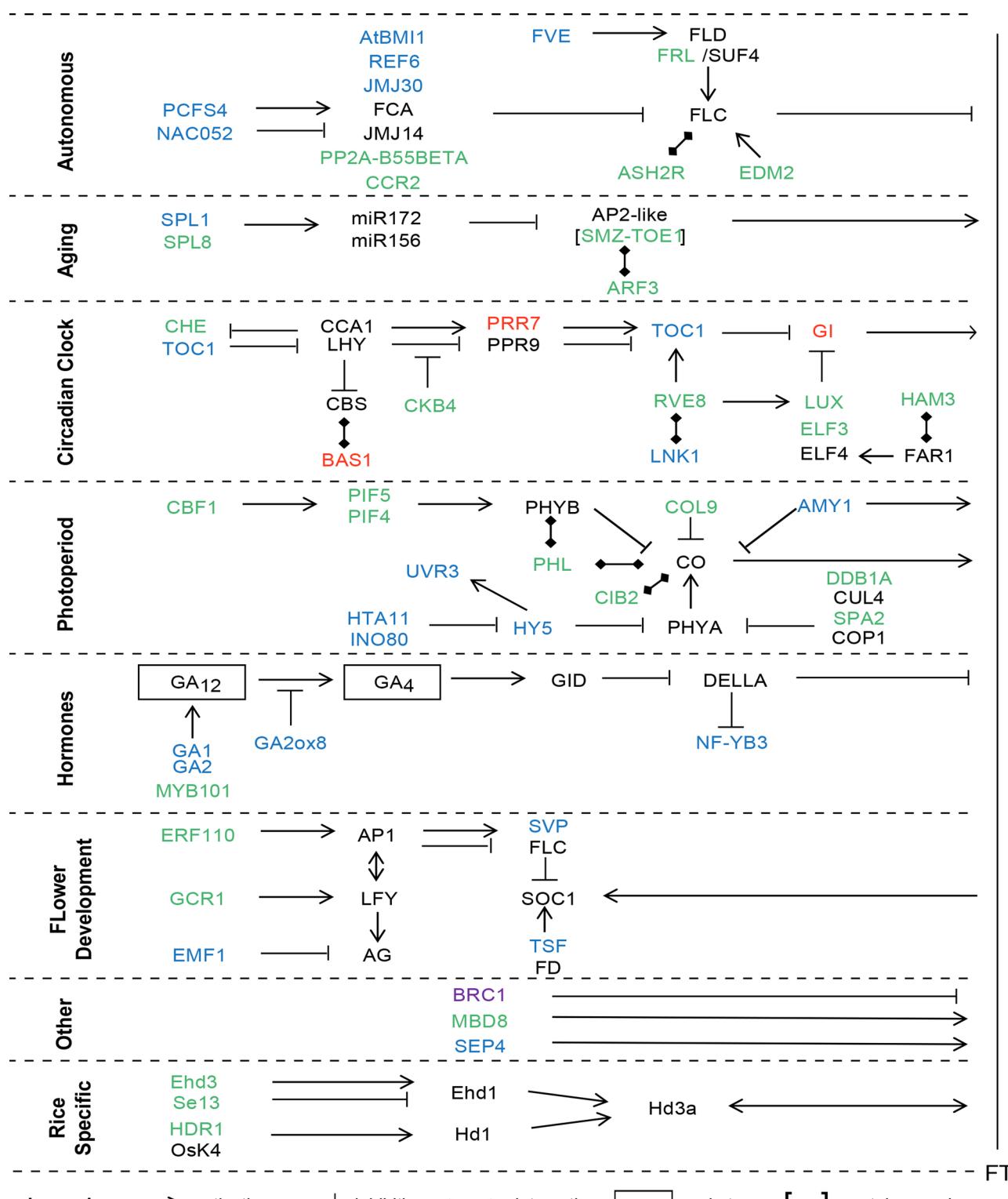
of each box. Each point shows the percentage of mapped reads to different transposon classes in each individual. *nicaraguensis* ($n=14$), *luxurians* ($n=14$), *diploperennis* ($n=20$), *perennis* ($n=19$), *huehuetenangensis* ($n=5$), *mexicana* ($n=81$), *parviflumis* ($n=70$), TST ($n=210$) and TEM ($n=280$).



Extended Data Fig. 5 | Principal component analysis and haplotype of eight large inversions. Each group of plots contains a dotplot showing the principal component analysis of SNPs in the inversion region and a heatmap showing genotypes at the inversion identified by invClust¹⁶. Each point in the dotplot

represents a sample and the color represents the inversion haplotypes. Each row in the heatmap represents the genotyped window used in invClust (500 kb) and each column represents a sample. The bar above the heatmap also shows the inversion cluster identified by PCA.





Blue font Highland adaptive genes

Green font High latitude adaptive genes

Red font Convergent selection genes in highland and high latitude adaptation

Purple font Homologous genes pairs with one copy under selection in the highlands and the other in high latitude

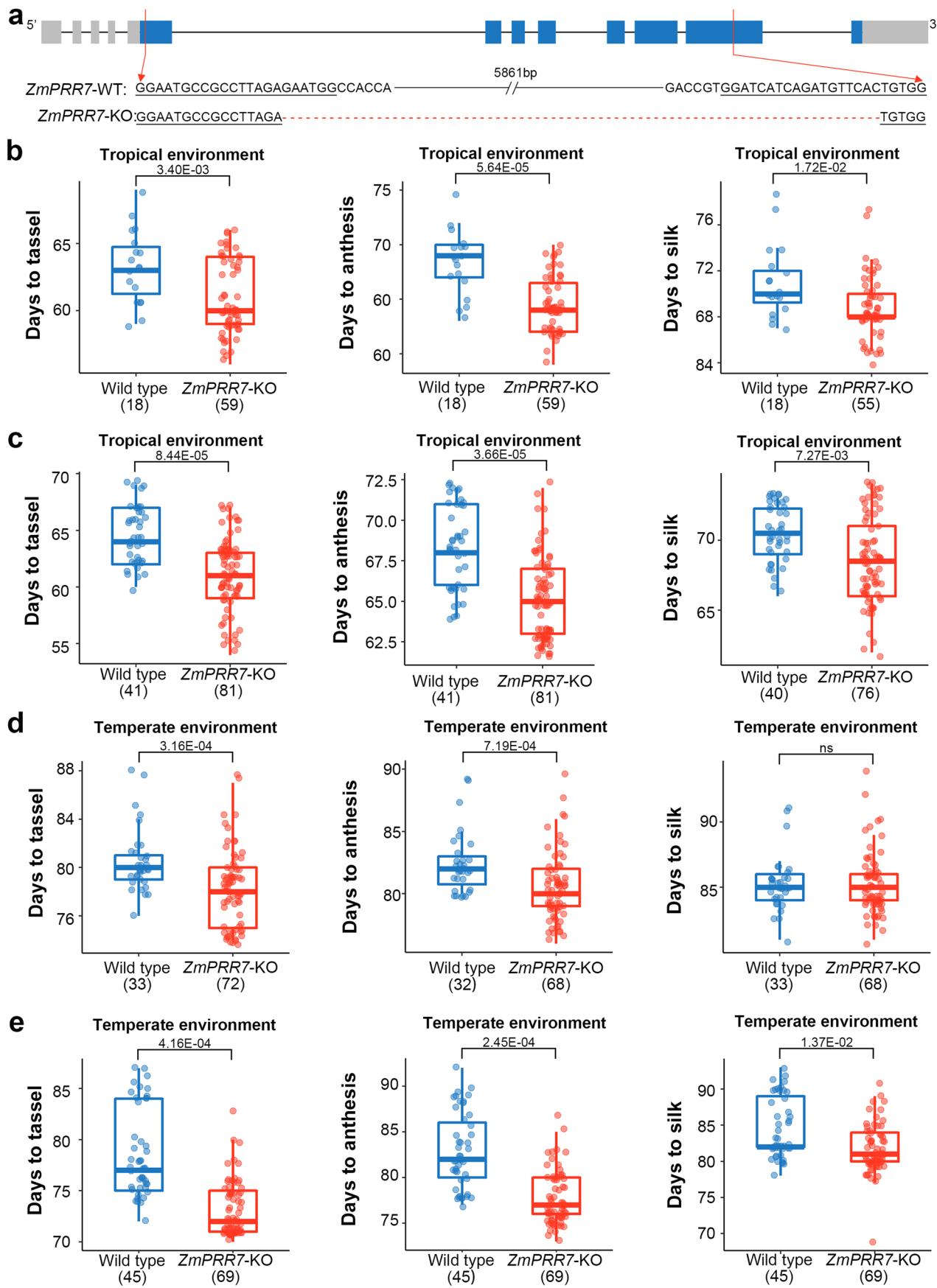
Black font Normal font is other gene and bold font is pathway name

Extended Data Fig. 7 | Flowering time related pathway in highland and high

latitude adaptation. Pathway were integrated from *Arabidopsis* and rice.

Genes colored with 'blue' represent highland adaptive genes, 'green': high

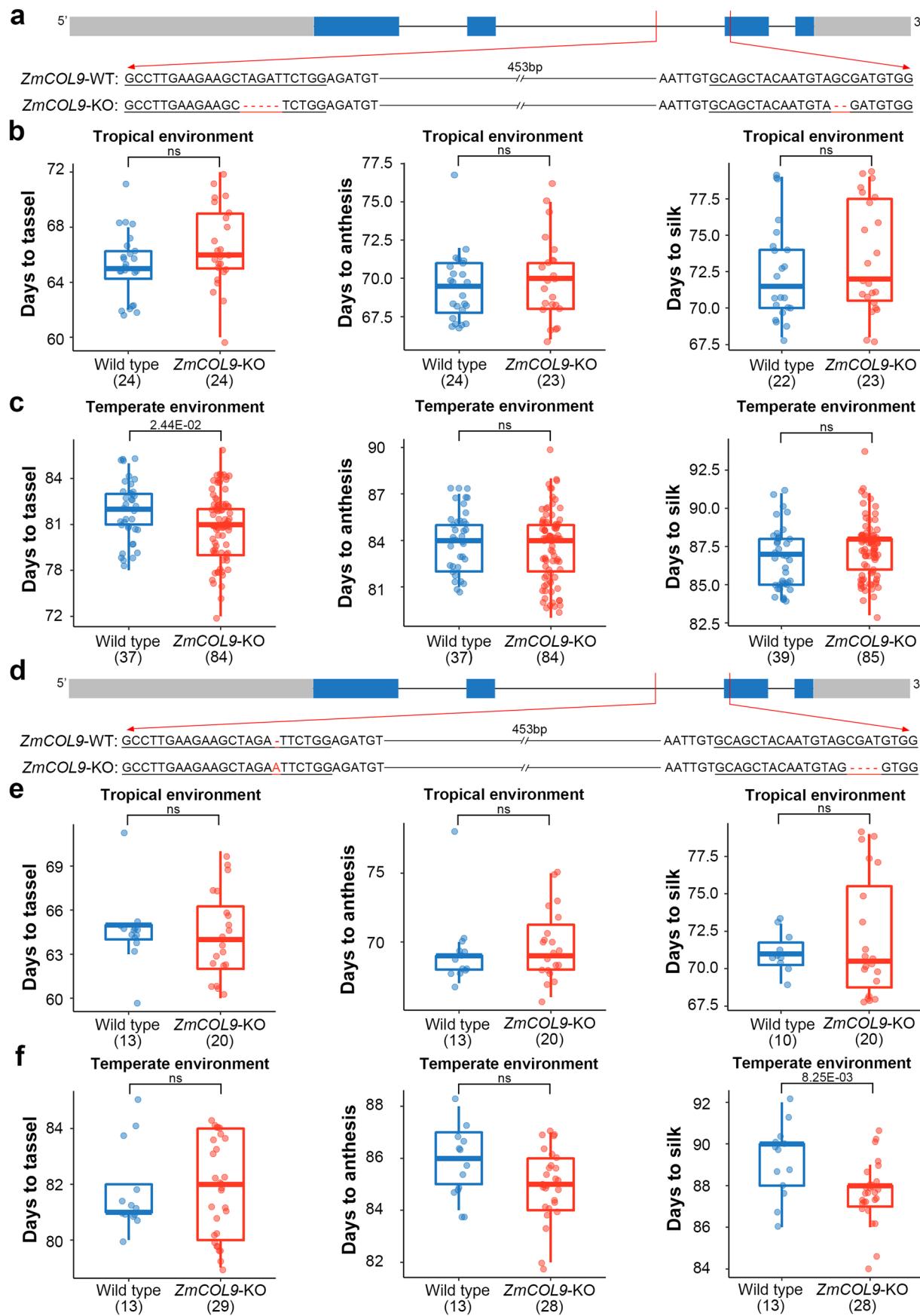
latitude adaptation genes, 'purple': homologous genes pairs with one copy under selection in the highlands and the other in high latitude, 'red': convergent selection genes in highland and high latitude adaptation.



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

Extended Data Fig. 8 | Phenotype analysis of CRISPR/Cas9 mutation for *ZmPRR7* in different environments. **a**, Gene structure and sequences of *ZmPRR7* target regions in wild type, *ZmPRR7* CRISPR/Cas9 knockout mutants. **b** and **c**, Statics of days to tassel, days to anthesis, days to silk in Hainan province (2019 and 2020; China; E109°, N18°; tropical environment). **d** and **e**, Statics of days to tassel, days to anthesis, days to silk in Jilin province (2020 and 2021;

China; E125°, N44°; temperate environment). ns: no significance, two-sided t-test *P*-value shown. Numbers in the blank in x-axis represents the number of individuals. Each point shows the statistics of traits for each individual. Center lines in the boxplot indicate the median, edges represent the 25th and 75th percentiles, whiskers further extend by ± 1.5 times the interquartile range from the limits of each box.

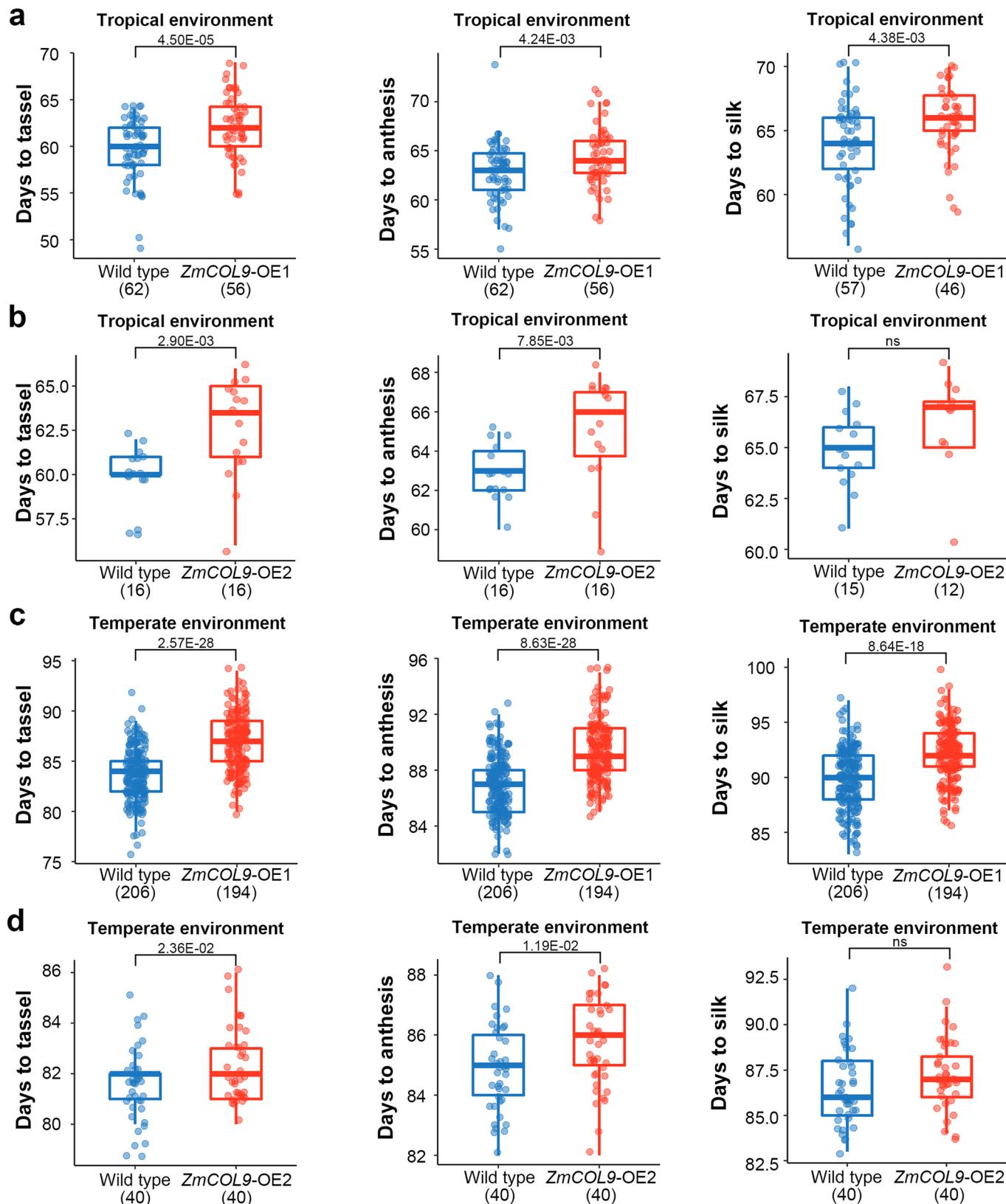


Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Phenotype analysis of different CRISPR/Cas9 mutation

ZmCOL9 in different environments. **a**, Gene structure and sequences of *ZmCOL9* target regions in wild type, *ZmCOL9* CRISPR/Cas9 knockout mutants 1. **b**, Statics of days to tassel, days to anthesis, days to silk in Hainan province (2019; China; E109°, N18°; CRISPR/Cas9 mutation). **c**, Statics of days to tassel, days to anthesis, days to silk in Jilin province (2020; China; E125°, N44°; CRISPR/Cas9 mutation). **d**, Gene structure and sequences of *ZmCOL9* target regions in wild type, *ZmCOL9* CRISPR/Cas9 knockout mutants 2. **e**, Statics of days to tassel, days

to anthesis, days to silk in Hainan province (2019; China; E109°, N18°; CRISPR/Cas9 mutation). **f**, Statics of days to tassel, days to anthesis, days to silk in Jilin province (2020; China; E125°, N44°; CRISPR/Cas9 mutation). ns: no significance, two-sided t-test *P*-value shown. Numbers in the blank in x-axis represents the number of individuals. Each point shows the statistics of traits in each individual. Center lines in the boxplot indicate the median, edges represent the 25th and 75th percentiles, whiskers further extend by ± 1.5 times the interquartile range from the limits of each box.



Extended Data Fig. 10 | Phenotype analysis of different overexpression lines for *ZmCOL9* in different environments. **a**, Statistics of days to tassel, days to anthesis, days to silk in Hainan province (2019; China; E109°, N18°; overexpression line 1). **b**, Statistics of days to tassel, days to anthesis, days to silk in Hainan province (2019; China; E109°, N18°; overexpression line 2). **c**, Statistics of days to tassel, days to anthesis, days to silk in Jilin province (2020; China; E125°, N44°; overexpression line 1). **d**, Statistics of days to tassel, days to anthesis,

days to silk in Jilin province (2020; China; E125°, N44°; overexpression line 2). ns: no significance, two-sided t-test P-value shown. Numbers in the blank in x-axis represents the number of individuals. Each point shows the statistics of traits in each individual. Center lines in the boxplot indicate the median, edges represent the 25th and 75th percentiles, whiskers further extend by ± 1.5 times the interquartile range from the limits of each box.

Reporting Summary

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

Statistics

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

n/a Confirmed

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

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection Soil properties were collected by using R package ncdf4 (v1.16).

Data analysis Published software: FastQC (v0.11.3; <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), Trimmomatic (v0.33), Bowtie2 (v2.1.0), Picard (v1.119; <http://broadinstitute.github.io/picard/>), SAMtools (v1.3.1), GATK (v3.5; <https://software.broadinstitute.org/gatk/>), PLINK (v1.9), admixture (v1.3.0), GCTA (v1.26), SnpEff (v4.1g; <http://snpeff.sourceforge.net/index.html>), SNPhylo (v20140701), BPP (v4.1.4), ANGSD (v0.930), DensiTree (v2.2.6), BEAGLE (v4.0), DIP (v1.0, <https://github.com/EvanForsythe/DIP>), MSMC2 (v2.1.1), VCFTools (v0.1.16), losstruct (v1.0; https://github.com/petrelharp/local_pca), sourmash (v3.2.0), invClust (v1.0), RepeatExplore2 (v0.3.8; https://bitbucket.org/petrnovak/repx_tarean/src/devel/), RepeatMasker (<http://www.repeatmasker.org/>; v4.1.0), BWA-MEM (v0.7.10), TASSEL3 (v3.0.174), CrossMap (v0.2.9), XP-CLR (v1.0), GenWin (v0.1), BEDTools (v2.25.0), fastp (v0.19.4), TopHat2 (v2.2.1), htseq-count (v0.9.1), DESeq2 (v1.10.1), clValid (v0.7), PANTHER (10.5281/zenodo.5725227 Released 2021-11-16), GlueGo (V2.5.9).
Custom scripts are available at: <https://doi.org/10.5281/zenodo.6818334>

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

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

DNA- and RNA-sequencing reads from this study were deposited in the NCBI Sequence Read Archive with the accession number of PRJNA641489, PRJNA816255, PRJNA816273 and PRJNA645739, respectively. The SNP data can be downloaded from https://ftp.ncbi.nlm.nih.gov/pub/CNSA/data3/CNP0001565/zeemap/02_Variants/PAN_Zea_Variants/Zea-vardb/.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

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

Sample size	Samples used in DNA sequencing were gathered as much as we can if Genebanks available. Sample size used in RNA-seq were selected to meet the requirement that at least three biological replications. Sample size used in species tree, divergence time analysis were selected to ensure a balance tree structure of each taxon. Sample size (one sample) of ABBA-BABA and divergence-based introgression polarization analysis were defined by the requirement of the analysis tools. Nucleotide diversity analysis of maize sample (n=110) were selected to fit the handleable sample of the analysis tool. Other studies in this research use all available samples.
Data exclusions	Raw reads and genotype with low quality were excluded as described in the methods.
Replication	The phenotype of the over-expression and knock out mutants used in this study were validated by at least two years field experiments.
Randomization	Randomly selected sample in differently expression gene analysis, species tree, divergence time analysis, ABBA-BABA, divergence-based introgression polarization analysis, nucleotide diversity analysis were preformed by using the function "rand" in PERL. Random shuffled region were used in enrichment analysis between highland and high latitude adaptation regions by using BEDTools. Random genes used in enrichment analysis between teosinte and maize differently expression genes also randomly selected by using the function "rand" in PERL.
Blinding	No blinding for this study. Phenotype were collected without knowledge of genotype information.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging