Portrait of a genus: genome sequencing reveals evidence of adaptive variation in *Zea*

Lu Chen[1†‡], Jingyun Luo[1†], Minliang Jin[1†], Ning Yang[1,2†*], Xiangguo Liu[3], Yong Peng[1], Wenqiang Li[1], Alyssa Philips[4], Brenda Cameron[5], Julio Bernal[6], Rubén Rellán-Álvarez[7], Ruairidh JH Sawers[8], Liu Qing[3], Yuejia Yin[3], Xinnan Ye[3], Jiali Yan[1], Qinghua Zhang[1], Xiaoting Zhang[1], Shenshen Wu[1], Songtao Gui[1], Wenjie Wei[1], Yuebin Wang[1], Yun Luo[1], Chengling Jiang[1], Min Deng[1], Min Jin[1], Liumei Jian[1], Yanhui Yu[1], Maolin Zhang[1], Xiaohong Yang[9], Matthew B. Hufford[10], Alisdair R. Fernie[11], Marilyn L. Warburton[12], Jeffrey Ross-Ibarra[4*], Jianbing Yan[1,2*]

[1] National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

[2] Hubei Hongshan Laboratory, Wuhan 430070, China

[3] Biotechnology Research Centre, Jilin Academy of Agricultural Sciences, Changchun 130033, China

[4] Department of Evolution and Ecology, Center for Population Biology, Genome Center, University of California Davis, Davis, California 95616, USA

[5] Department of Evolution and Ecology, University of California Davis, Davis, California 95616, USA

[6] Department of Entomology, Texas A&M University, College Station, TX, 77843-2475, USA

[7] Department of Molecular and Structural Biochemistry, North Carolina State University, Raleigh, NC

[8] Department of Plant Science, The Pennsylvania State University, State College, PA

16802, USA

[9] National Maize Improvement Center of China, Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing 100193, China

[10] Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, Iowa 50011, USA

[11] Department of Molecular Physiology, Max-Planck-Institute of Molecular Plant Physiology, Am Mühlenberg 1, 14476 Potsdam-Golm, Germany

[12] United States Department of Agriculture-Agricultural Research Service: Western Regional Plant Introduction Station, 59 Johnson Hall, Washington State University, Pullman, WA, 99614, USA

[†] These authors contributed equally to this work

[‡] Present 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 100101, China

[*] Correspondence should be addressed to Jianbing Yan (yjianbing@mail.hzau.edu.cn), Jeffrey Ross-Ibarra (rossibarra@ucdavis.edu) and Ning Yang (yangningyingji@126.com)

Abstract

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 the 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 (SNPs) and nearly 9 million Insertion/Deletion (InDel) polymorphisms. The variation map reveals evidence of selection within taxa displaying novel adaptations to traits such as waterlogging, perenniality and regrowth. We focus in detail on adaptive alleles in highland teosinte and temperate maize and highlight the key role of flowering time related pathways in highland and high latitude adaptation. To show how this data can identify useful genetic variants, we generated and characterized novel mutant alleles for two flowering time candidate genes. This work provides the most extensive sampling to date of the genetic diversity of the genus *Zea*, resolving questions on evolution and identifying adaptive variants for direct use in modern breeding.

Introduction

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 problem[3], and projected shifts in temperature and precipitation will lead to further declines in productivity for many major crops[4]. New varieties displaying both higher yield and the better adaptation to diverse environments are thus urgently needed to increase crop productivity under changing climate scenarios[5,6].

Maize (*Zea mays* subsp. *mays*) is one of the world's most widely grown crops, with an annual global production of over 1.1 billion tons in 2018 (FAOSTAT, 2020).

Native American peoples domesticated from the wild grass *Zea mays* subsp. *parviglumis* (hereafter *parviglumis*) approximately 9,000 years ago in the southwest of Mexico[7,8]. Population genetic 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 years[13]. 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 use of the genetic diversity at their disposal, looking beyond modern elite lines to traditional cultivated varieties and locally adapted wild relatives[16].

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 absent in modern maize and humid high elevations in central Western Mexico and the Huehuetenango region of Guatemala[17-19], providing a wealth of genetic diversity that could be utilized in modern breeding. A recent example is a large-effect allele for leaf angle identified in teosinte that was lost during maize domestication[20]. CRISPR-Cas9 editing of maize to mimic the teosinte allele resulted in a ~20% yield increase in modern hybrids grown at high density. Other studies have used genetic mapping to capitalize on teosinte alleles for nutrition[21,22], adaptation to extreme environments[23,24], and disease resistance[25-27]. Population genetic evidence suggests that diverse alleles from the teosinte *Zea mays* subsp. *mexicana* (hereafter *mexicana*) played an important role in allowing maize to adapt to arid

highland conditions[28-30].

Despite the potential for teosinte to contribute to 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 relationship of several taxa is 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 teosinte and maize landraces is the result of local adaptation[48,49]. Low density genotyping or pooled sequencing approaches in *parviglumis* 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 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 described species and subspecies in the genus *Zea* (Fig. 1a, b) to an average depth of 22x, and combined these data with genome resequencing data from 507 cultivated maize inbred lines

representing both temperate and tropical regions[52] (Supplementary Table 1). To ensure the quality of this new *Zea* diversity map, we used a set of strict filtering conditions (Methods) and identified a final set more than 70M SNPs and nearly 9M insertion/deletions (InDels) (Supplementary Table 2), with nearly 80% of SNPs segregating as rare variants (MAF<0.05) (Supplementary Fig. 1). Both classes of variants appeared enriched in genic and regulatory regions (30% of SNPs and 45% of InDels in 14% of the genome), likely reflecting difficulties in read mapping in repetitive regions of the genome. We validated a subset of genic SNPs using Sanger sequencing, with median concordance between datasets >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 greater than 60% ancestry in a single group were clustered into *parviglumis* (*n*=70), *mexicana* (*n*=81), *Zea mays* subsp. *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*), 210 tropical maize and 280 temperate maize (Supplementary Fig. 2a,b and Supplementary Table 1). Principal component analysis of these lines was in strong concordance with population structure results (Supplementary Fig. 2c).

We inferred phylogenetic relationships for the genus *Zea* under the multispecies coalescent model[53] (Fig. 1c); maximum likelihood phylogenies[54] produced largely congruent results (Supplementary Fig. 3, 4). 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 >10M years ago[55]. Within the genus, our results suggest that

144 *nicaraguensis* likely represents a subspecies of *luxurians*, with divergence times similar
145 to those among subspecies of *Zea mays*. The phylogeny supports earlier analysis[34]
146 suggesting that divergence among *Zea mays*, *luxurians*, and *diploperennis* was nearly
147 contemporaneous, occurring ~120,000 years ago (95% highest posterior density (HPD)
148 interval for *luxurians* divergence from other taxa: 125,967-127,200; Fig. 1c and
149 Supplementary Table 4). We further estimate that *perennis* split from its diploid
150 progenitor *diploperennis* only ~48,000 years ago (95% HPD: 38,033-119,100). Tree
151 topologies and divergence times also support earlier analyses[56] showing that
152 *huehuetenangensis* is a subspecies of *Zea mays*, diverging from other annual subspecies
153 ~68,000 years ago (95% HPD: 60,133-106,467), followed by the divergence of
154 highland *mexicana* and lowland *parviglumis* ~30,000 years ago (95% HPD: 26,733-
155 34,500). Our phylogeny estimates the divergence of maize from *parviglumis* at ~12,000
156 years, only slightly older than the earliest archaeological evidence[8] and likely due to
157 population structure within *parviglumis*[37,46]. Independent estimates of divergence times
158 taken from rates of cross-coalescence[57] between taxa are strikingly consistent (Fig. 1d).
159 Population genetic analysis of diversity further reveals changes in demography
160 among taxa in *Zea*. Coalescent estimates of effective population size (N_e) over time
161 reveal the well-established bottleneck associated with maize domestication but also a
162 continued decline in population size for the annual subsepecies *parviglumis* and
163 *mexicana* since their divergence (Supplementary Fig. 5). All other taxa in the genus
164 show parallel trends, with steady declines in population size until about 10,000 years
165 ago, with more recent increases for *luxurians* and *diploperennis*. Patterns of shared
166 derived alleles and sequence divergence both suggest a history of introgression among
167 taxa (Fig. 1e, Supplementary Fig. 6 and Supplementary Table 5), including
168 bidirectional admixture between *parviglumis/huehuetenangensis* and

nicaraguensis/luxurians, and unidirectional introgression from *huehuetenangensis/mexicana* into domesticated maize, highlighting the important role of gene flow in crop adaptation[58].

Novel diversity in *Zea*

SNP data highlight the impressive genetic diversity present in teosinte. 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 of maize lines, even among teosinte with limited geographic ranges (Supplementary Table 2 and Supplementary Fig. 7). Nearly a 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. 2a), this tendency remains the same after choosing comparable samples in each taxon (Supplementary Fig. 8). In teosintes, a substantial proportion of taxon-specific SNPs and InDels are located in genic and regulatory regions (promoter and cis-regulatory elements[59]; Supplementary Fig. 9), suggesting the presence of biologically functional alleles with potential for improving modern maize. Differentiation (F_{ST}) between teosinte taxa is often lower than that found between inbred maize and teosinte (Supplementary Fig. 10a), consistent with the historical reduction of diversity that occurred during modern maize breeding[60]. The annual subspecies of *Zea mays* show much faster decay of linkage disequilibrium than our diverse panel of maize inbreds (10-50Kb compared to ~200Kb; Supplementary Fig. 10b), but historical recombination in other teosintes appears to be even more limited (>500Kb).

Short-read mapping approaches pose challenges in characterizing genetic diversity, including difficulty with repetitive sequences and reference bias. In order to

194 circumvent some of these obstacles, we used a reference-free k-mer approach to
195 characterize each taxon (Methods). Consistent with the reference mapping bias (~8%
196 unmapped reads in average), most taxa showed a substantial proportion of unique k-
197 mers (Supplementary Fig. 11a, b Supplementary Table 2). Since non *Zea mays* species
198 have diverged from *Zea mays* more than ~120,000 years (Fig. 1c), the higher number
199 of unique k-mers were exhibited in their genomes as expected (Fig. 2b and
200 Supplementary Fig. 11c, d). These results not only highlight the novel genetic diversity
201 present in teosinte but also likely point to the ongoing importance of evolutionary
202 processes in generating and filtering diversity in traditional maize populations in
203 Mexico[61].

204 We next investigated the diversity and abundance of transposons and inversion
205 polymorphisms in *Zea*. Transposable elements (TEs) are an important driver of shaping
206 the structure and evolution of the genome[62], and over 85% of the maize genome is
207 repetitive sequence[63]. Clustering repeats from our short-read data accounted for ~74%
208 of sequence across the genus (Supplementary Table 6), with the vast majority (60-70%)
209 coming from LTR retrotransposon. Mapping reads from individual genomes to these
210 clusters revealed broadly similar patterns across species, consistent with previous
211 comparisons of *Zea mays* and *luxurians*[64]. Nonethless, we do identify a notable
212 decrease in the percentage of Ty3 retrotransposons in *Zea mays* compared to other
213 species, and an increase abundance of DNA transposons in *diploperennis* and *perennis*
214 (Fig. 2c and Supplementary Fig 12).

215 Inversions are known to play important roles in adaptation and speciation[65,66],
216 and previous work has highlighted the evolutionary relevance of several large
217 inversions in *Zea*[23,45,46,67], including *Inv9e* in *mexicana* adaptation[46,50,51].
218 Multidimensional scaling of SNP diversity across the genome[68] allowed us to identify

eight large genomic regions (> 1 Mb) indicative of inversion polymorphism (Supplementary Fig. 13, Supplementary Table 7). Six of these are newly identified in the present study, and show clustering patterns delineating the three genotypes (standard; heterozygous inversion and homozygous inversion; Fig. 2d, Supplementary Fig. 14 and Supplementary Table 8).

Given previous evidence suggesting the association between inversions and soil characteristics[46], we performed genome-wide association with nine representative soil traits (Methods) from a rich database of more than 200 soil properties[69] (Supplementary Fig. 15a and Supplementary Table 9). *Inv9e* was significantly associated with gypsum content (0.829-1.383m) which is a representative of 29 soil properties (Supplementary Fig. 15b and Supplementary Table 9). We merged nearby significant SNPs located in *Inv9e* into two QTLs (chr9:127,017,047-127,356,295 and chr9:138,354,955-139,846,464; Supplementary Fig. 16, Supplementary Table 10). These contain 15 genes that have been functionally validated in rice or *Arabidopsis* (Supplementary Table 11) including two (*Zm00001d047667* and *Zm00001d047694*) with orthologs that have been confirmed to affect root development in rice[70,71] and may provide clues to further explore the function of *Inv9e* in adaptation. Given that many inversions found segregating at appreciable frequency are likely adaptive in some environments[72,73], these data argue that improved assembly and characterization of structural variants in teosinte would be a promising avenue for 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 adaptation[17]. To identify potential targets of selection, we calculated F_{ST} between each teosinte taxa and cultivated maize in 5-kb sliding windows (Methods). Here, we found a high proportion of outlier windows shared between the

closely related taxa (56% overlapped between *nicaraguensis* and *luxurians*; 54% overlapped between *diploperennis* and *perennis*; Supplementary Table 12, Supplementary Fig. 17). Shared genes (5,706; Supplementary Table 13) in *nicaraguensis* and *luxurians* comparisons were enriched in core cell component and reproductive system developmental processes (GO:0061458; *P*-value = 1.15E-04; FDR = 6.87E-03; Supplementary Table 14). Candidate adaptive genes (4,659; Supplementary Table 15) in *diploperennis* and *perennis* comparisons were enriched in some basic biological process and core cellular components such as nucleus (GO:0005634; *P*-value = 1.25E-12; FDR = 2.89E -10) (Supplementary Fig. 18, Supplementary Table 16).

We also identify a number of genes related with known pathways involved in meiosis[74], QTLs in regrowth[75] and waterlogging[76-78] (Supplementary Table 17). These include *Zm00001d002945*, an ortholog of the *Arabidopsis* gene *AtNAC082* involved in the regulation of leaf senescence[79], which shows high F_{ST} in *diploperennis* - maize and *perennis* - maize comparisons and is located in a QTL region controlling regrowth[75]. In *nicaraguensis* - maize, *luxurians* - maize comparisons, we find genes potentially involved in the response to waterlogging not only by regulating the content of ethylene and wax, but also the photosynthetic efficiency to adapt to the wetter climate in Guatemala[17]. These include *Zm00001d015637*, the maize ortholog of *AtOSP1* in *Arabidopsis*, a GDSL lipase that is required for wax biosynthesis and stomatal formation[80]. 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 worth further exploration.

In addition to identifying differences among species, our extensive sampling of *parviglumis* (*n*=70), *mexicana* (*n*=81), and both temperate (*n*=280) and tropical maize

(*n*=210) accessions allowed investigation of more recent adaptation to highlands and high latitudes. Both the high elevation and high latitude reflects a climate of lower temperate and longer light period, and previous work identified evidence of convergent selection between temperate maize and its broadly-distributed temperate relative *Tripsacum*[81]. Here, we extended this comparison to investigate convergence between temperate maize and high elevation adapted teosinte (*mexicana*). We applied a composite likelihood genome-scan (see Methods) for selection between *mexicana* vs *parviglumis* and temperate vs tropical maize (Fig. 3a, b and Supplementary Table 18, 19). We found significant overlap in selected windows (P = 0.047; 14.7% higher than permutations; Supplementary Fig. 19a), but less overlap than expected in candidate genes (P = 0.97; 27% less than permutation). Notably, however, ~90% of selected windows in both comparisons were found in noncoding regions of the genome, suggesting adaptation may predominantly have targeted regulatory regions. To test for convergence in regulatory adaptation, used RNA-seq from the shoot base of *parviglumis*, *mexicana* and tropical and temperate maize to search for changes in gene expression. We identified 595 genes differentially expressed between *mexicana* and *parviglumis* (Supplementary Table 20) and 437 genes differentially expressed between tropical and temperate maize (Supplementary Table 21), with significant overlap between the two lists (P = 0.006; 102% higher than permutations; Supplementary Fig. 19b). Those results may point to the importance of convergent regulatory evolution in maize and teosinte 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 range[82], and the alleles involved in flowering time are also a major target of highland landrace adaptation[83]. Experimental data in maize[84] and from orthologs[85] in other

294 species shows that at least 51 genes associated with highland and 61 genes associated
295 with high latitude adaptation were involved in flowering time pathways
296 (Supplementary Fig. 20 and Supplementary Table 22), respectively. For example, the
297 genes *GI* and *PRR7*, both known to participate in the circadian clock pathway in
298 *Arabidopsis* and rice[86-89], show evidence of selection both in *mexicana* and temperate
299 maize. Tracking the flowering time pathway, we found temperate maize has more genes
300 under selection in the photoperiod pathway (eight in temperate maize, five in *mexicana*;
301 Supplementary Table 22), which may be a signal of adaptation to changing latitude.

302 To validate the utility of the selection scan approach, we tested the function of
303 *ZmPRR7* (*Zm00001d047761*), which shows convergent patterns in maize and teosinte,
304 and the maize-specific candidate *ZmCOL9* (*Zm00001d051684*) that is involved in the
305 photoperiod pathway. Mutants of these two genes were obtained from a CRISPR/Cas9-
306 based high-throughput targeted mutagenesis library[90]. The mutant allele of *ZmPRR7* is
307 a 5.8-Kb deletion in the gene region that leads to the total loss of protein function. Plants
308 harboring the mutant allele exhibit significantly earlier flowering than the wildtype in
309 both tropical and temperate environments (Fig. 3c, d and Supplementary Fig. 21). The
310 loss-of-function allele of *ZmCOL9* includes a 5 bp deletion/1bp insertion in the intron
311 and a 2 bp deletion/4 bp deletion in the 3rd exon (Supplementary Fig. 22a, d) that result
312 in premature translation termination. In a tropical environment (Hainan; China; E109°,
313 N18°), *ZmCOL9* knockout mutants showed no difference in flowering time compared
314 to the wild type (Fig. 3e and Supplementary Fig. 23b, e) but overexpression plants
315 exhibit a later flowering phenotype (Fig. 3f and Supplementary Fig. 23a, b). In contrast,
316 when planted in a temperate environment (Jilin; China; E125°, N44°), the *ZmCOL9*
317 knockout mutants flowered earlier (Fig. 3e and Supplementary Fig. 22c, f) and the
318 overexpression lines flowered later than the wild type (Fig. 3f, g and Supplementary

319 Fig. 23c, d). These results confirmed the key roles for both *ZmPRR7* and *ZmCOL9* in
320 regulating flowering time and contributing to the adaptation of highland teosinte and
321 modern maize.

322 **Discussion**

323 The twin projections of increasing human population and decreasing suitable farmland
324 highlight the challenge breeders face in producing high crop yields, and this has
325 motivated an increasing interest in crop wild relatives as sources of genetic diversity
326 for improvment[91,92]. Here, we present a high-resolution genetic variation map that
327 greatly expands the publicly available genetic sequence information for the genus *Zea*.

328 We provide the first complete picture of the phylogeny and demography of the
329 genus *Zea* using genome-wide data, including both divergence times and effective
330 population sizes of *Zea* species. We reaffirm several aspects of the phylogeny of *Zea,*
331 but our data identify a number of new features, including the likely subspecies status of
332 *nicaraguensis*, the short divergence times between the perennial taxa, and the relatively
333 young age of the genus. We caution that our divergence estimate for *Tripsacum* may be
334 underestimated because of the difficulty of mapping short reads from divergent
335 genomes, however, and that high-quality *Tripsacum* and teosinte reference genomes
336 will be essential to better answer this question[93].

337 Our broad sampling of the genus allows us to take advantage of population genetic
338 tools to identify candidate genes involved in adaptation across both long and short time
339 scales. We find evidence of convergent adaptation of highland teosinte and high-
340 latitude maize, exemplifying the utility of studying variation in wild relatives to identify
341 genes important in crops. Finally, we validate these approaches by using genome
342 editing to knock out two candidate flowering time genes. All data and results of this
343 work have been integrated into the ZEAMAP database[94] for easy query and retrieval.

344 It is particularly noteworthy that our work identifies a vast trove of genetic
345 variation absent in cultivated maize and even in its closest wild relative *parviglumis*.
346 Our functional analysis of candidate adaptation genes clarifies the great potential in the
347 utilization of the wild relatives of maize in identifying novel alleles or highlighting
348 potential genes for subsequent editing, potentially accelerating modern genetic
349 improvements[95]. The data and discoveries presented in this study provide the
350 foundation for the use of crop wild relative resources for breeding in the face of
351 increasing human populations and decreasing farmland.

352 **Reference**

353 1. Ray, D. K., Mueller, N. D., West, P. C. & Foley, J. A. Yield trends are insufficient
354 to double global crop production by 2050. *PLoS One* **8**, e66428 (2013).
355 2. Bailey-Serres, J., Parker, J. E., Ainsworth, E. A., Oldroyd, G. E. D. & Schroeder,
356 J. I. Genetic strategies for improving crop yields. *Nature* **575**, 109-118 (2019).
357 3. Lesk, C., Rowhani, P. & Ramankutty, N. Influence of extreme weather disasters
358 on global crop production. *Nature* **529**, 84-87 (2016).
359 4. Challinor, A. J. et al. A meta-analysis of crop yield under climate change and
360 adaptation. *Nat Clim Change* **4**, 287-291 (2014).
361 5. Tigchelaar, M., Battisti, D. S., Naylor, R. L. & Ray, D. K. Future warming
362 increases probability of globally synchronized maize production shocks. *Proc. Natl.*
363 *Acad. Sci. USA* **115**, 6644-6649 (2018).
364 6. Li, Q. & Yan, J. Sustainable agriculture in the era of omics: knowledge-driven crop
365 breeding. *Genome Biol* **21**, 154 (2020).
366 7. Matsuoka, Y. et al. A single domestication for maize shown by multilocus
367 microsatellite genotyping. *Proc. Natl. Acad. Sci. USA* **99**, 6080-6084 (2002).
368 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. Hans-O, P. et al. Climate Change 2022: Impacts, Adaptation, and Vulnerability. (Cambridge Univ. Press, In Press).
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. Tian, J. et al. Teosinte ligule allele narrows plant architecture and enhances high-density maize yields. *Science* **365**, 658-664 (2019).
21. 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).
22. 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).
23. 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).
24. 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).
25. de Lange, E. S., Balmer, D., Brigitte, Mauch-Mani. & Turlings, T. C. J. Insect and pathogen attack and resistance in maize and its wild ancestors, the teosintes. *New Phytol* **204**, 329-341 (2014).
26. 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).
27. 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).
28. Hufford, M. B. et al. The genomic signature of crop-wild introgression in maize. *PLoS Genet* **9**, e1003477 (2013).

29. Calfee, E. et al. Selective sorting of ancestral introgression in maize and teosinte along an elevational cline. *PLoS Genet* **17**, e1009810 (2021).

30. Rodríguez-Zapata. F. et al. Teosinte introgression modulates phosphatidylcholine levels and induces early maize flowering time. Preprint at https://www.biorxiv.org/content/10.1101/2021.01.25.426574v2 (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. Doebley, J. F. Genetic diversity and population structure of teosinte. *Genetics* **169**, 2241-2254 (2005).

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).

443 40. Tenaillon, M. I., Hufford, M. B., Gaut, B. S. & Ross-Ibarra, J. Genome size and
444 transposable element content as determined by high-throughput sequencing in
445 maize and *Zea luxurians*. *Genome Biol Evol* **3**, 219-229 (2011).
446 41. Chia, J. M. et al. Maize HapMap2 identifies extant variation from a genome in flux.
447 *Nat Genet* **44**, 803-807 (2012).
448 42. Ting, Y. C. Spontaneous chromosome inversions of Guatemalan teosintes (*Zea*
449 *mexicana*). *Genetica* **36**, 229-242 (1965).
450 43. Ting, Y.C. Common inversion in maize and teosinte. *Am Nat* **101**, 87-89 (1967).
451 44. Ting Y. C. Chromosome polymorphism of teosinte. *Genetics* **83**, 737-742 (1976).
452 45. Fang, Z. et al. Megabase-scale inversion polymorphism in the wild ancestor of
453 maize. *Genetics* **191**, 883-894 (2012).
454 46. Pyhäjärvi, T., Hufford, M. B., Mezmouk, S., & Ross-Ibarra, J. Complex patterns
455 of local adaptation in teosinte. *Genome Biol Evol* **5**, 1594-1609 (2013).
456 47. Yang, N. et al. Contributions of *Zea mays* subspecies *mexicana* haplotypes to
457 modern maize. *Nat. Commun*. **8**, 1874 (2017).
458 48. Fustier, M. A. et al. Common gardens in teosintes reveal the establishment of a
459 syndrome of adaptation to altitude. *PLoS Genet* **15**, e1008512 (2019).
460 49. Janzen, G. M. et al. Demonstration of local adaptation of maize landraces by
461 reciprocal transplantation. Preprint at
462 https://www.biorxiv.org/content/10.1101/2021.03.25.437076v1 (2021).
463 50. Fustier M. A. et al. Signatures of local adaptation in lowland and highland teosintes
464 from whole-genome sequencing of pooled samples. *Mol Ecol* **26**, 2738-2756
465 (2017).

466 51. Aguirre-Liguori, J. A. et al. Divergence with gene flow is driven by local
467 adaptation to temperature and soil phosphorus concentration in teosinte subspecies
468 (*Zea mays parviglumis* and *Zea mays mexicana*). *Mol Ecol* **28**, 2814-2830 (2019).
469 52. Yang, X. et al. Characterization of a global germplasm collection and its potential
470 utilization for analysis of complex quantitative traits in maize. *Mol Breed.* **28**, 511-
471 526 (2011).
472 53. Flouri, T., Jiao, X., Rannala, B. & Yang, Z. Species tree inference with BPP using
473 genomic sequences and the multispecies coalescent. *Mol Biol Evol* **35**, 2585-2593
474 (2018).
475 54. Lee, T. H., Guo, H., Wang, X., Kim, C. & Paterson, A. H. SNPhylo: a pipeline to
476 construct a phylogenetic tree from huge SNP data. *BMC Genomics* **15**, 162 (2014).
477 55. Wang, X. et al. Genome alignment spanning major Poaceae lineages reveals
478 heterogeneous evolutionary rates and alters inferred dates for key evolutionary
479 events. *Mol Plant* **8**, 885-898 (2015).
480 56. Buckler, E. S^{4th}. & Holtsford, T. P. *Zea* systematics: ribosomal ITS evidence. *Mol*
481 *Biol Evol* **13**, 612-622 (1996).
482 57. Schiffels, S. & Wang, K. MSMC and MSMC2: The Multiple Sequentially
483 Markovian Coalescent. *Methods Mol Biol* **2090**, 147-166 (2020).
484 58. Janzen, G. M., Wang, L. & Hufford, M. B. The extent of adaptive wild
485 introgression in crops. *New Phytol* **221**, 1279-1288 (2019).
486 59. Marand, A. P., Chen, Z., Gallavotti, A. & Schmitz, R. J. A cis-regulatory atlas in
487 maize at single-cell resolution. *Cell* **184**, 3041-3055 (2021).
488 60. van Heerwaarden, J., Hufford, M. B. & Ross-Ibarra, J. Historical genomics of
489 North American maize. *Proc. Natl. Acad. Sci. USA* **109**, 12420-12425 (2012).

490 61. Bellon, M. R. et al. Evolutionary and food supply implications of ongoing maize
491 domestication by Mexican *campesinos*. *Proc Biol Sci* **285**, 20181049 (2018).
492 62. Finnegan, D. J. Eukaryotic transposable elements and genome evolution. *Trends*
493 *Genet* **5**, 103-107 (1989).
494 63. Hufford, M. B. et al. De novo assembly, annotation, and comparative analysis of
495 26 diverse maize genomes. *Science* **373**, 655-662 (2021).
496 64. Tenaillon, M. I., Hufford, M. B., Gaut, B. S. & Ross-Ibarra, J. Genome size and
497 transposable element content as determined by high-throughput sequencing in
498 maize and *Zea luxurians*. *Genome Biol Evol* **3**, 219-229 (2011).
499 65. Faria, R., Johannesson, K., Butlin, R. K. & Westram, A. M. Evolving Inversions.
500 *Trends Ecol Evol* **34**, 239-248 (2019).
501 66. Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and
502 speciation. *Genetics* **173**, 419-434 (2006).
503 67. Crow, T. et al. Gene regulatory effects of a large chromosomal inversion in
504 highland maize. *PLoS Genet* **16**, e1009213 (2020).
505 68. Li, H. & Ralph, P. Local PCA shows how the effect of population structure differs
506 along the genome. *Genetics* **211**, 289-304 (2019).
507 69. Wei S., Dai Y., Duan Q., Liu B. & Hua Y. A global soil data set for earth system
508 modeling. *J Adv Model Earth Syst* **6**, 249-263 (2014).
509 70. Silva, R. et al. *Gluconacetobacter diazotrophicus* changes the molecular
510 mechanisms of root development in *Oryza sativa L.* growing under water stress.
511 *Int J Mol Sci* **21**, 333 (2020).
512 71. Wang, C. et al. Mutation in xyloglucan 6-xylosytransferase results in abnormal root
513 hair development in *Oryza sativa*. *J Exp Bot* **65**, 4149-4157 (2014).

514 72. Corbett-Detig, R. B. & Hartl, D. L. Population genomics of inversion
515 polymorphisms in *Drosophila melanogaster*. *PLoS Genet* **8**, e1003056 (2012).
516 73. Lowry, D. B. & Willis, J. H. A widespread chromosomal inversion polymorphism
517 contributes to a major life-history transition, local adaptation, and reproductive
518 isolation. *PLoS Biol* **8**, e1000500 (2010).
519 74. Yant, L. et al. Meiotic adaptation to genome duplication in *Arabidopsis arenosa*.
520 *Curr Biol* **23**, 2151-2156 (2013).
521 75. Ma, A. et al. The genetics and genome-wide screening of regrowth loci, a key
522 component of perennialism in *Zea diploperennis*. *G3 (Bethesda)* **9**, 1393-1403
523 (2019).
524 76. Guo, Z. et al. Identification of major QTL for waterlogging tolerance in maize
525 using genome-wide association study and bulked sample analysis. *J Appl Genet* **62**,
526 405-418 (2021).
527 77. Yu, F. et al. Dissecting the genetic architecture of waterlogging stress-related traits
528 uncovers a key waterlogging tolerance gene in maize. *Theor Appl Genet* **131**, 2299-
529 2310 (2018).
530 78. Osman, K. A. et al. Dynamic QTL analysis and candidate gene mapping for
531 waterlogging tolerance at maize seedling stage. *PLoS One* **8**, e79305 (2013).
532 79. Kim, H. J. et al. Time-evolving genetic networks reveal a NAC troika that
533 negatively regulates leaf senescence in Arabidopsis. *Proc. Natl. Acad. Sci. USA*
534 **115**, E4930-E4939 (2018).
535 80. Tang, J. et al. GDSL lipase occluded stomatal pore 1 is required for wax
536 biosynthesis and stomatal cuticular ledge formation. *New Phytol* **228**, 1880-1896
537 (2020).

538 81. Yan, L et al. Parallels between natural selection in the cold-adapted crop-wild
539 relative *Tripsacum dactyloides* and artificial selection in temperate adapted maize.
540 *Plant J* **99**, 965-977 (2019).
541 82. Buckler, E. S. et al. The genetic architecture of maize flowering time. *Science* **325**,
542 714-718 (2009).
543 83. Wang, L. et al. Molecular parallelism underlies convergent highland adaptation of
544 maize landraces. *Mol Biol Evol* 38, 3567-3580 (2021).
545 84. Li Y. X. et al. Identification of genetic variants associated with maize flowering
546 time using an extremely large multi-genetic background population. *Plant J* **86**,
547 391-402 (2016).
548 85. Bouché, F., Lobet, G., Tocquin, P. & Périlleux, C. FLOR-ID: an interactive
549 database of flowering-time gene networks in *Arabidopsis thaliana*. *Nucleic Acids*
550 *Res* **44**, D1167-1171 (2016).
551 86. Hayama, R., Yokoi, S., Tamaki, S., Yano, M. & Shimamoto, K. Adaptation of
552 photoperiodic control pathways produces short-day flowering in rice. *Nature* **422**,
553 719-722 (2003).
554 87. Seo, E. et al. Crosstalk between cold response and flowering in *Arabidopsis* is
555 mediated through the flowering-time gene *SOC1* and its upstream negative
556 regulator *FLC*. *Plant Cell* **21**, 3185-3197 (2009).
557 88. Liang, L. et al. The transcriptional repressor *OsPRR73* links circadian clock and
558 photoperiod pathway to control heading date in rice. *Plant Cell Environ* **44**, 842-
559 855 (2021).
560 89. Nakamichi, N. et al. *Arabidopsis* clock-associated pseudo-response regulators
561 *PRR9*, *PRR7* and *PRR5* coordinately and positively regulate flowering time

562 through the canonical CONSTANS-dependent photoperiodic pathway. *Plant Cell*
563 *Physiol* **48**, 822-832 (2007).
564 90. Liu, H. et al. High-Throughput CRISPR/Cas9 Mutagenesis Streamlines Trait Gene
565 Identification in Maize. *Plant Cell* **32**, 1397-1413 (2020).
566 91. Hufford, M. B., Martínez-Meyer, E., Gaut, B. S., Eguiarte, L. E. & Tenaillon, M.
567 I. Inferences from the historical distribution of wild and domesticated maize
568 provide ecological and evolutionary insight. *PLoS One* **7**, e47659 (2012).
569 92. Aguirre-Liguori, J. A. et al. Connecting genomic patterns of local adaptation and
570 niche suitability in teosintes. *Mol Ecol* **26**, 4226-4240 (2017).
571 93. Gault, C. M., Kremling, K. A. & Buckler, E. S. *Tripsacum* de novo transcriptome
572 assemblies reveal parallel gene evolution with maize after ancient polyploidy.
573 *Plant Genome* **11**, 1-13 (2018).
574 94. Gui, S. et al. ZEAMAP, a comprehensive database adapted to the maize multi-
575 omics era. *iScience* **23**, 101241 (2020).
576 95. Fernie, A. R., & Yan, J. Targeting key genes to tailor old and new crops for a
577 greener agriculture. *Mol Plant* **13**, 354-356 (2020).

578 **METHODS**

579 **Samples and whole genome resequencing.** A total of 237 teosinte accessions from
580 CIMMYT, USDA and collaborators were obtained, consisting of 90 *mexicana*, 79
581 *parviglumis*, 20 *diploperennis*, 15 *perennis*, 15 *luxurians*, 13 *nicaraguensis*, five
582 *huehuetenangensis* according to morphological classification (Supplementary Table 1,
583 2). Two *Tripsacum dactyloides* were obtained from Dr. Fajia Chen's lab (Henan
584 Agricultural University, China). Young leaves from one individual of each accession
585 were used for DNA extraction for sequencing using the Illumina HiSeq3000 platform
586 (150-bp paired-end), conducted by BGI (Shenzhen, China) and NovaSeq6000 platform

(150-bp paired-end), conducted by Novogene (Sacramento, USA). DNA sequencing data of 507 cultivated maize were downloaded from the NCBI SRA database (PRJNA531553; Supplementary Table 1).

Read mapping and SNP calling. Raw reads of teosinte were first processed using FastQC (v0.11.3; http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimmomatic[96] (v0.33; HiSeq3000 platform; LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) and fastp[97] (v0.19.4; NovaSeq6000 platform; -g -l 36) were used to remove poor-quality base calls and adaptors. Reads of teosinte and maize were then aligned to the B73 reference genome[98] (v4) using Bowtie2[99] (v2.1.0; --very-fast). Unique mapped reads were sorted and indexed using Picard (v1.119; http://broadinstitute.github.io/picard/). SAMtools[100] (v1.3.1) and UnifiedGenotyper from GATK (v3.5; https://software.broadinstitute.org/gatk/) were used to estimate the variant calling file for each individual. Hard filtering of the individual SNP calls was carried out with mapping quality (MQ $\leq$ 20.0), and thresholds set by sequencing coverage based on minimum coverage (DP $\leq$ 5) and maximum coverage (DP $\geq$ 200). Then, variants from the 237 teosinte and 507 maize were combined by GATK CombineVariants to a single variant calling file. To confirm if unknown variants were discarded reference genotypes in individual calls, we recalled these sites and replaced them with reference genotypes if they had supported reads. Finally, sites with a missing rate higher than 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 classification, principal component analysis and phylogenetic tree construction. We evaluated patterns of population structure using a set of SNPs filtered to remove multi-allelic loci and SNPs with a minor allele frequency

612 < 0.05 (--maf 0.05 -biallelic-only) using PLINK[101] (v1.9). We then ran admixture[102] for
613 different values of the number of clusters (K) from 2 to 20 (--cv = 10; v1.3.0). Each
614 individual with admixture components < 0.6 was classified as 'Teosinte (mix)' or
615 'maize (mix)'. We performed PCA analysis using this same set of SNPs with GCTA[103]
616 (v1.26) recording the first 10 components (--pca 10). We annotated SNPs with a
617 missing data rate less than 0.7 in teosinte and maize with SnpEff (v4.1g;
618 http://snpeff.sourceforge.net/index.html) using the first transcript of B73 v4 genes. We
619 then used synonymous and noncoding SNPs to construct a simple phylogenetic tree
620 with SNPhylo[54] (v20140701) using default parameters and visualized the tree with
621 iTOL[104].

622 **Species tree analysis.** Species delimitation and species trees were inferred using BPP[53]
623 (model A11; v4.1.4). We used the following samples in BPP: three tropical maize, three
624 *parviglumis*, three *mexicana*, three *nicaraguensis*, three *diploperennis*, three *perennis*,
625 three *luxurians*, two *huehuetenangensis* and two *Tripsacum dactyloides*
626 (Supplementary Table 1). Low-quality base calls and adaptors from raw reads of
627 *Tripsacum dactyloides* were removed using Trimmomatic, and the remaining
628 sequences were aligned to the B73 v4 reference genome with Bowtie2 as described
629 above. The consensus base was estimated from the uniquely mapped reads using
630 ANGSD[105] (v0.930). Using the B73 annotation, we randomly selected 2,000 coding
631 sequence genes to estimate the species delimitation and species tree. The prior
632 distribution of ancestral population size (θ) and divergence time from the root (τ)
633 followed an inverse-gamma (IG) prior with means of 0.005 IG (3 0.01)) and 0.75 (IG
634 (3 1.5)), respectively. The consensus of A11 species trees was visualized using
635 DensiTree[106] (v2.2.6).

636 **Imputation and demographic estimation.** SNPs in the 237 teosinte and 507 maize

were imputed with BEAGLE[107] (v4.0), respectively. Divergence times within teosinte and the effective population size of each teosinte were estimated using BPP (A00 model) and MSMC2[57] (v2.1.1). The topological tree in BPP (A00 model) was fixed as the species tree with 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 priors as above. In MSMC2, four haplotype models were applied (Supplementary Table 1). The mutation rate used in BPP (A00 model) and MSMC2 was 3E-08[108].

ABBA-BABA and divergence-based introgression polarization test. We used Patterson's D statistic[109,110] to test for introgression between teosinte. Assuming *Tripsacum dactyloides* as the outgroup (O), we assessed D statistics for the tree (((P1, P2), P3), O), P1/P2/P3 representing different taxa in *Zea* (autotetraploid *perennis* was excluded). The number of ABBA and BABA 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. To estimate the directions of introgression, consensus base was estimated from the uniquely mapped reads using ANGSD for representing 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 suggested in Supplementary Fig. 6 and divided into a 5,000bp windows, which were used as the input of DIP[111].

Linkage disequilibrium, nucleotide diversity and F_{ST} calculation. Linkage disequilibrium (r^2) of *nicaraguensis* (14), *luxurians* (14), *diploperennis* (20), *perennis* (15), *huehuetenangensis* (five), *mexicana* (81), *parviglumis* (70), and maize (507) were

662 estimated for all bi-allelelic SNPs within 500Kb (--geno 0.5 --maf 0.05 --biallelic-only
663 --snps-only) using PLINK. Nucleotide diversity of *nicaraguensis* (14), *luxurians* (14),
664 *diploperennis* (20), *perennis* (15), *huehuetenangensis* (five), *mexicana* (81),
665 *parviglumis* (70) and maize (randomly selected 110 individuals) was calculated using
666 ANGSD (v0.930, -doMaf 1 -doMajorMinor 1 -uniqueOnly 1 -minMapQ 30 -minQ 20
667 -GL 2 -fold 1 -win 5000 -step 5000). Differentiation (F_{ST}) between maize and teosinte
668 with five randomly selected samples was estimated in VCFTools[112] (v0.1.16; --fst-
669 window-size 5000).

670 **Taxon-specific SNPs, InDels and k-mer analysis.** SNPs and InDels found only in one
671 specific taxon in *Zea* in at least two individuals were regarded as taxon-specific SNPs.
672 The longest transcripts of each gene in the B73 annotation and a recent atlas of cis-
673 regulatory elements[59] were used to annotate variants. K-mers of teosinte and maize
674 were counted using Jellyfish[113] (v2.3.0; -m 31). K-mers unique to each taxon that
675 appeared at least two times were obtained with sourmash[114] (v3.2.0; --scaled 1000).

676 **Transposon element analysis.** RepeatExplore2[115] was used to identify repeat clusters
677 of each taxa of *Zea* (two samples were randomly selected from each taxon). Clusters
678 were further annotated by applying RepeatMasker (http://www.repeatmasker.org/;
679 v4.1.0; -species maize). Reads were mapped to the above repeat clusters by using
680 BWA-MEM[116] (v0.7.10), and the number of mapped reads in each repeat clusters were
681 calculated with SAMTools. Abundance of the repeat elements between samples were
682 normalized by their sequenced library size.

683 **Inversion calling**. Localized heterogeneity across chromosomes was identified using
684 lostruct[68] in windows containing 10,000 SNPs. The most related 5% of windows in
685 each chromosome around one of the four outliers (maximum, minimum MDS1 or
686 MDS2) were regarded as candidate inversions and were genotyped using invClust[117]

(v1.0) with B73 as the reference state. Genotypes of the candidates were confirmed via PCA of the SNPs in the corresponding region. Only taxa 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 locations from entire in the NAM population[63].

Genome-wide association analysis. SNPs from *mexicana* were obtained from the imputed teosinte panel according to the name of samples, and then population structure was calculated with admixture (v1.3.0; --cv=10; K=1, 2, 3, 4, 5). The K value with the lowest CV (K=2) was used in downstream analysis. Estimation of the kinship matrix and association analyses using the compressed MLM were performed using TASSEL3[118] (v3.0.174), with a *P*-value cut off set to 1/N (N = the number of tested SNPs). Latitude and longitude information was obtained from Dr.Suketoshi Taba's lab. Global soil properties used as phenotypes for the GWAS were extracted using the R package ncdf4 (v1.16; http://cirrus.ucsd.edu/~pierce/ncdf/) from the Global Soil Dataset for Earth System Modeling[69], a comprehensive database with eight layers to the depth of 2.3m (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). Soil properties were clustered using the R package clValid[119], which tested hierarchical, k-means and k-medoide in combination with 2-40 clusters to find the best method and cluster numbers. GWAS were performed on a subset of nine features identified by hierarchical cluster analysis (Supplementary Fig. 16).

Identification of adaptive regions in non *Zea mays* taxa. Whole genome adaptive genetic variation between different non *Zea mays* taxa and maize were estimated by calculating their F_{ST} value in VCFTools (--fst-window-size 5000). Under each comparison, all available teosinte and maize samples were used. We then Z-

712 transformed the F_{ST} in each window, windows with ZF_{ST} values exceeding the 95th
713 percentile of the whole genome were declared as candidate adaptive regions. GO
714 enrichment analysis was conducted using PANTHER with default parameters[120,121] and
715 visualized with GlueGo[122].

716 **Selective sweeps in teosinte and maize.** Whole genome scanning for regions of
717 teosinte elevation adaptation and maize temperate adaptation was implemented by a
718 mixed method. First, two genetic maps were obtained from a B73 x Teosinte
719 population[123] and a maize B73 x By804 population[124], and the physical locations were
720 converted to coordinates of the B73 v4 reference sequence using CrossMap[125] (v0.2.9).
721 The genetic distance between SNPs in *mexicana* and *parviglumis* were then calculated
722 based on the B73 x Teosinte genetic map, while the distance in temperate maize and
723 tropical maize were calculated based on the B73 x By804 genetic map. Genetic
724 distances between SNPs located between the genetic markers were assigned based on
725 their physical distance. The likelihood of multi-locus allele frequency differentiation
726 between two tested populations was modeled using XP-CLR[126] (v1.0; -w1 0.005 100
727 1000 -p0 0.7) in both the teosinte group (*mexicana*, with *parviglumis* as the reference)
728 and the maize group (temperate maize, with tropical maize as the reference). Finally,
729 we applied a spline-window method (GenWin[127] v0.1; smoothness = 100) to smooth
730 the results. The top 5% of genomic region with the highest W statistic in *parviglumis*
731 and *mexicana* were regarded as candidate teosinte altitude adaptation regions and the
732 top 5% of the W statistic regions in temperate and tropical maize were regarded as
733 candidate maize temperate adaptation regions. Enrichment analysis between candidate
734 teosinte altitude adaptation regions and maize temperate adaptation was conducted
735 using the shuffle function (-excl -noOverlapping) in BEDTools[128] (v2.25.0). Genes,
736 including the promoter and 2kb upstream, that overlapped with the regions identified

737 above were regarded as candidate adaptive genes.

738 **RNA-seq sampling, library construction and data analysis**. The base tissues of V5
739 stage shoots (1-2 cm) of maize (five tropical maize; five temperate maize) and teosinte
740 (three *parviglumis*; three *mexicana*) were sampled for mRNA and total RNA extraction.
741 Both mRNA and total RNA samples were used for library preparation according to
742 Illumina strand-specific library construction protocols. Paired-end libraries were
743 sequenced using a mixture of platforms (Hi-Seq3000, x10, NovaSeq) with 150 cycles.
744 Raw reads were filtered to remove the poor-quality base calls and adaptors specifically
745 for each platform (NovaSeq: fastp -g -l 36; x10: fastp -l 36; Hi-Seq3000: Trimmomatic
746 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36). Reads were then
747 aligned to the B73 reference genome (V4) using TopHat2[129] (v2.2.1) and read counts
748 for each gene were calculated using htseq-count[130] (v0.9.1). Finally, differentially
749 expressed genes were identified between tropical and temperate maize, as well as
750 between *parviglumis* and *mexicana,* using DESeq2[131] (v1.10.1) with absolute fold
751 change higher than 1 and *P*-value < 0.05.

752 **Functional validation of *ZmPRR7* and *ZmCOL9*.** Mutants of *ZmPRR7* and *ZmCOL9*
753 were generated from a high-throughput genome-editing design[90]. In brief, line-specific
754 sgRNAs were filtered based on the assembled pseudo-genome of the receptor KN5585,
755 and a double sgRNAs pool (DSP) approach was used to construct vectors. The vectors
756 were transformed into the receptor KN5585, and the targets of each T_0 individual were
757 assigned by barcode-based sequencing. The genotype of gene-editing lines was
758 identified by PCR amplification and Sanger sequencing using target-specific primers
759 (Supplementary Table 23).

760 Transgenic lines generated with DNA fragments of *ZmCOL9* driven by the *ZmUbi*
761 promoter were created using the modified binary vector pCAMBIA3300. Immature

762 zygotic embryos of maize hybrid HiII (B73 x A188) were infected with *A. tumefaciens*
763 strain EHA105 harboring the binary vector based on the published method for
764 *ZmCOL9*[132]. Transgenic plants were identified by qRT-PCR as well as tests for
765 herbicide resistance and the presence of the bar gene. Flowering-time phenotypes of
766 mutants and transgenic plants of *ZmPRR7* and *ZmCOL9* were investigated in Jilin
767 province (E125°, N44°) and Hainan province (E109°, N18°).

768 **Data availability**

769 DNA- and RNA-sequencing reads from this study were deposited in the NCBI
770 Sequence Read Archive with the accession number of PRJNA641489, PRJNA816255,
771 PRJNA816273 and PRJNA645739, respectively. The SNP data can be downloaded
772 from
773 https://ftp.cngb.org/pub/CNSA/data3/CNP0001565/zeamap/02_Variants/PAN_Zea_V
774 ariants/Zea-vardb/.

775 **Code availability**

776 All custom scripts used in this study are available at
777 https://github.com/conniecl/Zea_genus.

778 **Reference**

779 96. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for
780 Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
781 97. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ
782 preprocessor. *Bioinformatics* **34**, i884-i890 (2018).
783 98. Jiao, Y. et al. Improved maize reference genome with single-molecule technologies.
784 *Nature* **546**, 524-527 (2017).
785 99. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat*
786 *Methods* **9**, 357-359 (2012).

787 100.Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*
788 **25**, 2078-2079 (2009).
789 101.Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and
790 richer datasets. *Gigascience* **4**, 7 (2015).
791 102.Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of
792 ancestry in unrelated individuals. *Genome Res*. **19**, 1655-1664 (2009).
793 103.Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. GCTA: a tool for genome-
794 wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
795 104.Letunic, I. & Bork, P. Interactive tree of life (iTOL) v4: recent updates and new
796 developments. *Nucleic Acids Res* **47**, W256-W259 (2019).
797 105.Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next
798 Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
799 106.Bouckaert, R. R. DensiTree: making sense of sets of phylogenetic trees.
800 *Bioinformatics* **26**, 1372-1373 (2010).
801 107.Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and
802 missing-data inference for whole-genome association studies by use of localized
803 haplotype clustering. *Am J Hum Genet* **81**, 1084-1097 (2007).
804 108.Clark, R. M., Tavaré, S. & Doebley, J. Estimating a nucleotide substitution rate for
805 maize from polymorphism at a major domestication locus. *Mol Biol Evol* **22**, 2304-
806 2312 (2005).
807 109.Green, R. E. et al. A draft sequence of the Neandertal genome. *Science* **328**, 710-
808 722 (2010).
809 110.Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture
810 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. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764-770 (2011).

114. Brown, C. T. & Irber, L. sourmash: a library for MinHash sketching of DNA. *J. Open Source Softw* **1**, 27 (2016).

115. Novák, P., Neumann, P. & Macas, J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* **11**, 378 (2010).

116. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **25**, 1754-1760 (2009).

117. 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).

118. Bradbury, P. J. et al. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633-2635 (2007).

119. Brock, G., Pihur, V., Datta, S. & Datta, S. clValid: An R package for cluster validation. *J Stat Softw* **25**, 1-22 (2008).

120. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).

121. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res* **47**, D330-D338 (2019).

834 122.Bindea, G. et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped
835 gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091-1093
836 (2009).
837 123.Liu, Z. et al. Expanding Maize Genetic Resources with Predomestication Alleles:
838 Maize-Teosinte Introgression Populations. *Plant Genome* **9** (2016).
839 124.Pan, Q. et al. Genome-wide recombination dynamics are associated with
840 phenotypic variation in maize. *New Phytol* **210**, 1083-1094 (2016).
841 125.Zhao, H. et al. CrossMap: a versatile tool for coordinate conversion between
842 genome assemblies. *Bioinformatics* **30**, 1006-1007 (2014).
843 126.Chen, H., Patterson, N. & Reich, D. Population differentiation as a test for selective
844 sweeps. *Genome Res* **20**, 393-402 (2010).
845 127.Beissinger, T. M., Rosa, G. J. M., Kaeppler, S. M., Gianola, D. & de Leon, N.
846 Defining window-boundaries for genomic analyses using smoothing spline
847 techniques. *Genet Sel Evol* **47**, 30 (2015).
848 128.Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
849 genomic features. *Bioinformatics*. **26**, 841-842 (2010).
850 129.Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of
851 insertions, deletions and gene fusions. *Genome Biol* **14**, R36 (2013).
852 130.Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-
853 throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).
854 131.Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and
855 dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
856 132.Frame, B. R. et al. Agrobacterium tumefaciens-mediated transformation of maize
857 embryos using a standard binary vector system. *Plant Physiol* **129**, 13-22 (2002).
858

ACKNOWLEDGMENTS

We would like to thank Dr. Suketoshi Taba from CIMMYT for providing teosinte materials. We would like to thank Dr. Jiafa Chen from Henan Agricultural University for providing *Tripsacum dactyloides*. We would like to thank Mr. Hao Liu from the high-throughput computing platform of National Key Laboratory of Crop Genetic Improvement. We would like to thank Andi Kur from North Carolina State University for drawing the picture of teosinte morphological characteristics. We also thank Taylor AuBuchon-Elder and Sally Fabbri for growing and supplying germplasm. This research was supported by the National Key Research and Development Program of China (2020YFE0202300) and National Natural Science Foundation of China (U1901201, 31771879), as well as the US National Science Foundation (grants 1546719,1822330) and USDA Hatch project CADPLS2066H.

AUTHOR CONTRIBUTIONS

J.Y., J.R.-I. and N.Y. designed and supervised this study. Y.P., W.L., A.P., B.C., J.B., R. R.-A., R.S., J.Y., Q.Z., S.W., S.G., Y.W., Y.L., C.J., M.D., M.J., J.L., L.J., Y.Y., M.Z. and X.Y. 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., L.Q., Y.Y., and X.Y performed genetic transformation and mutant validation. L.C., M.J., N.Y., M.H., A.R.F., M.L.W., J.R.-I. and J.Y. prepared the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

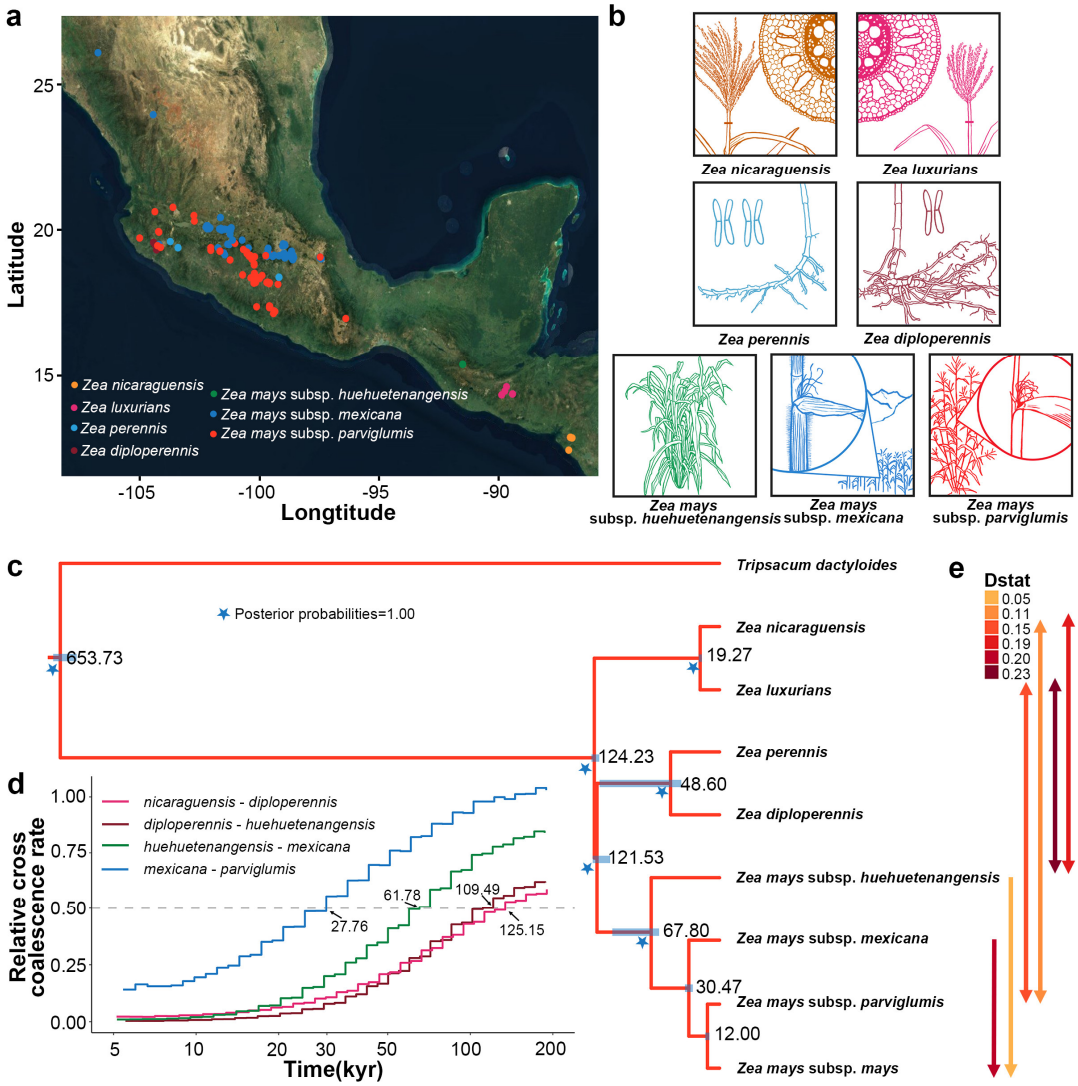

Fig. 1. Phylogeny of *Zea* genus. a, Geographical distribution of collected teosintes, taxa were identified and colored based on morphology. **b**, Morphological characteristics of teosinte (Credit to Dr. Andi Kur). *nicaraguensis* and *luxurians* are distinguished from other teosinte based on aerenchyma in their stems which aerate 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 other teosintes. The Mexican annual teosintes *parviglumis* 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. **c**, Divergence times (in thousands of years before present)

893 estimated from the multispecies coalescent (MSC) model. Blue bars indicate the 95%
894 highest posterior density (HPD) intervals. The star indicates nodes with posterior
895 probability of 1. Edge widths reflect estimates of effective population size
896 (Supplementary Table 4). **d**, Rates of cross-population coalescence among teosinte
897 species. Curves were computed using four phased haplotypes. **e**, Introgression among
898 taxa. Arrows indicate the taxa involved (one-way arrow indicate unidirectional
899 introgression, two-way arrow indicate bidirectional introgression), and arrow color
900 shows the value of Patterson's D-statistic (Supplementary Table 5).
901

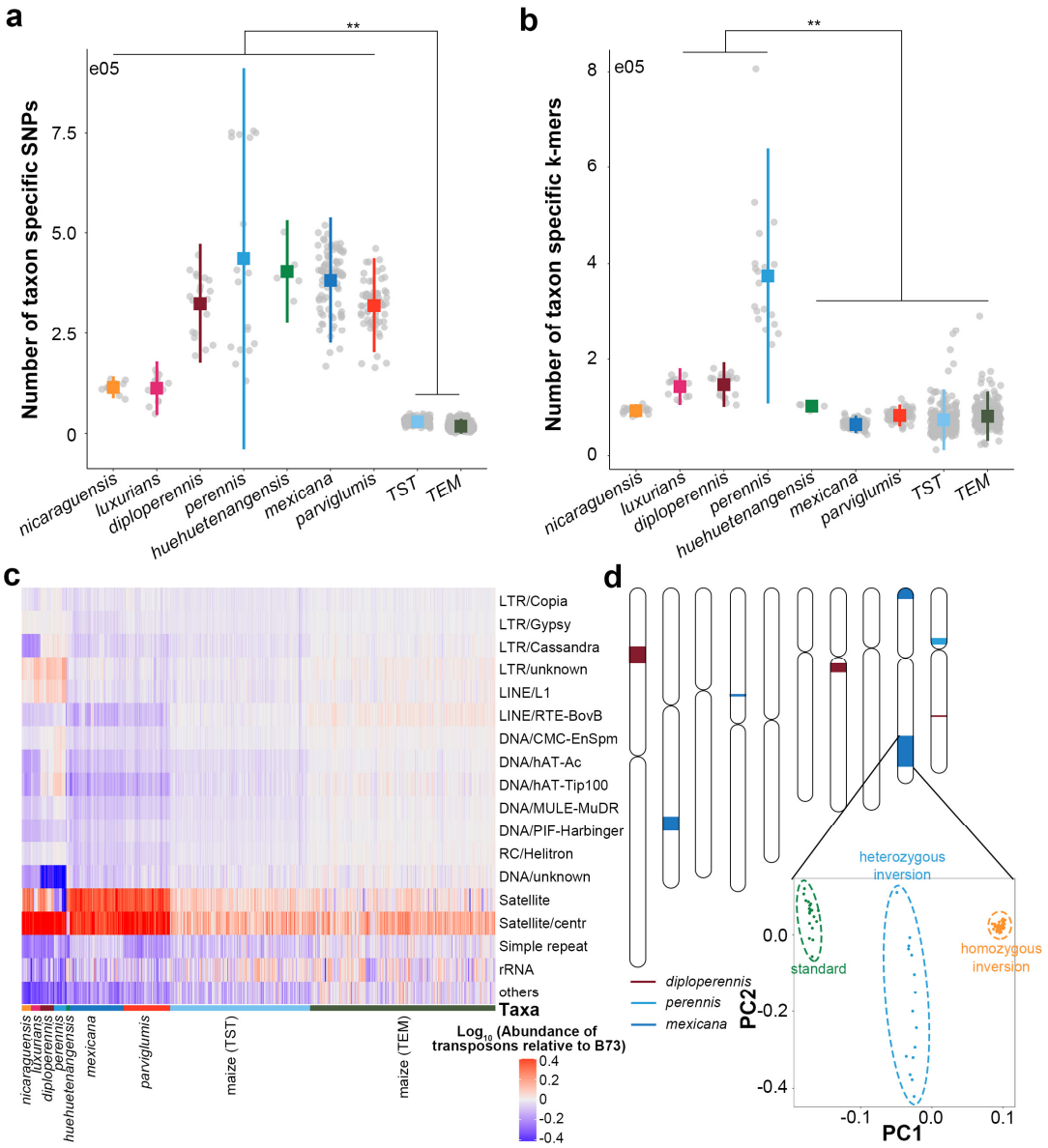

Fig. 2. Variation in the *Zea* genus. a, Taxon specific SNPs and **b**, k-mers (31bp) in *Zea* genus. TST indicates tropical maize, and TEM indicates temperate maize. The significantly lines compare all teosinte to TST and TEM (**a**) or *luxurians*/*diploperennis*/*perennis* to *Zea mays* (**b**). **c**, Abundance of transposon elements relative to B73. Each column represents a sample. **d**, 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*). Inset shows PCA of SNPs data

911 from within *Inv9e*, clearly separating the three genotype classes (left: standard; middle:

912 heterozygous inversion; right: homozygous inversion).

913

914

915

916

917

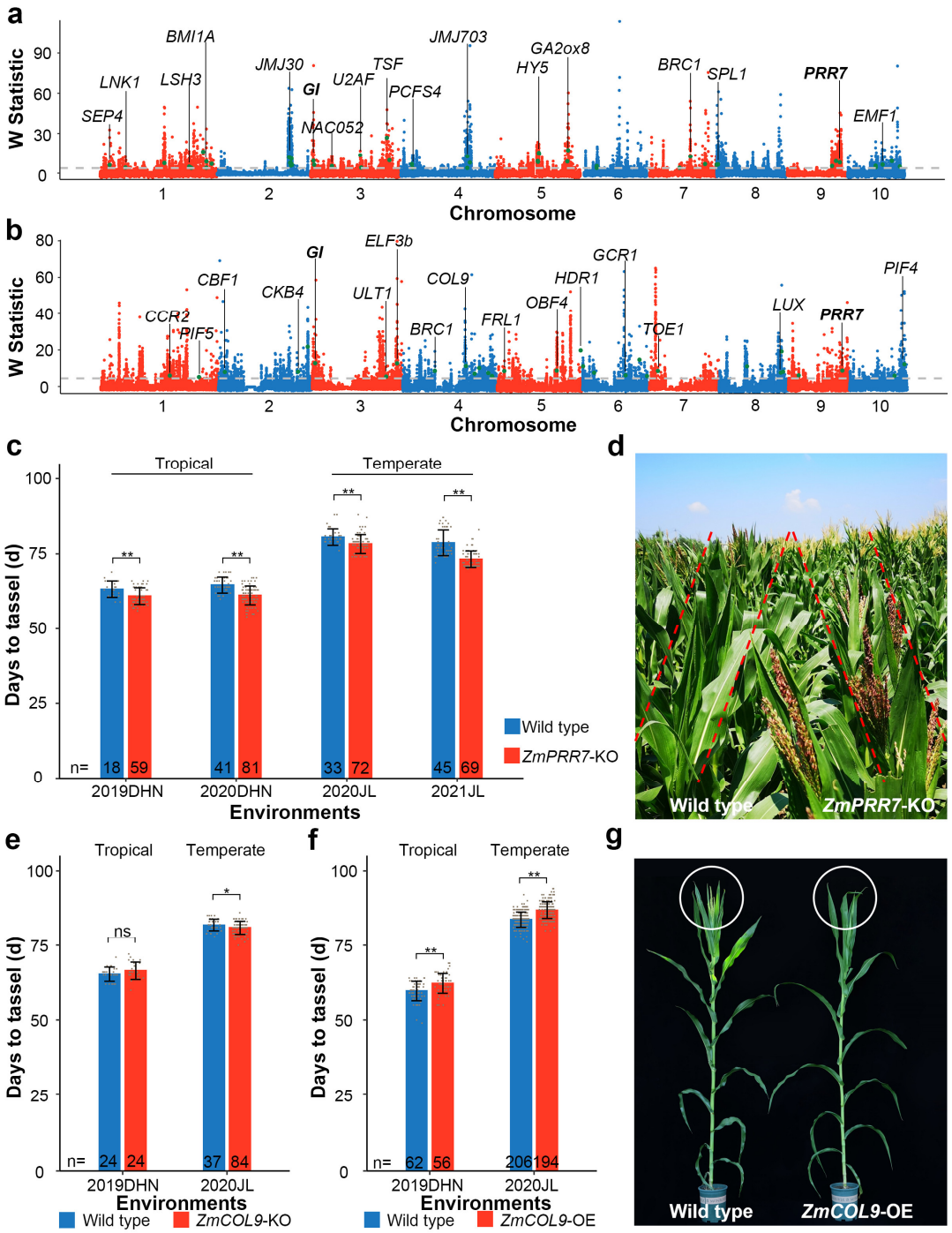

Fig. 3. Local adaptation in teosinte and maize. Genome-wide selection signals (W statistic reflecting smoothed XP-CLR score) between **a**, *mexicana* and *parviglumis*; and **b**, temperate and tropical maize. The horizontal grey dashed line represents the top 5% cutoff. Genes associated with flowering time and floral development in maize, rice and *A. thaliana* are marked with green points. **c**, Days to tassel of wild type and *ZmPRR7* knock out (KO) mutants under tropical (Hainan province 2019 and 2020; China; E109°,

N18°) and temperate (Jilin province 2020 and 2021; China; E125°, N44°) environments. **d**, *ZmPRR7* KO mutants showed earlier flowering relative to wild type. The picture was taken in Jilin province 2020 at 77 d after planting. **e**. Days to tassel of wild type and *ZmCOL9* KO mutants under tropical and temperate environments. **f**. Days to tassel of wild type and *ZmCOL9* over-expression (OE) mutants under tropical and temperate environments. **g**, *ZmCOL9* OE mutants showed later flowering relative to wild type. The picture was taken in Jilin province in 2020 at 78 d after planting. ns indicates non-significant difference between mutants and wild type by two-sided t-test at *P*-value = 0.05, * indicates *P*-value < 0.05, ** indicates *P*-value < 0.01.