

RESEARCH

Open Access



# Population genomics of *Zea* species identifies selection signatures during maize domestication and adaptation

Gen Xu<sup>1,2†</sup>, Xuan Zhang<sup>1,2†</sup>, Wenkang Chen<sup>1,2†</sup>, Renyu Zhang<sup>1,2†</sup>, Zhi Li<sup>1,2</sup>, Weiwei Wen<sup>3</sup>, Marilyn L. Warburton<sup>4</sup>, Jiansheng Li<sup>1</sup>, Huihui Li<sup>5\*</sup> and Xiaohong Yang<sup>1,2\*</sup>

## Abstract

**Background:** Maize (*Zea mays* L. ssp. *mays*) was domesticated from teosinte (*Zea mays* ssp. *parviglumis*) about 9000 years ago in southwestern Mexico and adapted to a range of environments worldwide. Researchers have depicted the maize domestication and adaptation processes over the past two decades, but efforts have been limited either in sample size or genetic diversity. To better understand these processes, we conducted a genome-wide survey of 982 maize inbred lines and 190 teosinte accessions using over 40,000 single-nucleotide polymorphism markers.

**Results:** Population structure, principal component analysis, and phylogenetic trees all confirmed the evolutionary relationship between maize and teosinte, and determined the evolutionary lineage of all species within teosinte. Shared haplotype analysis showed similar levels of ancestral alleles from *Zea mays* ssp. *parviglumis* and *Zea mays* ssp. *mexicana* in maize. Scans for selection signatures identified 394 domestication sweeps by comparing wild and cultivated maize and 360 adaptation sweeps by comparing tropical and temperate maize. Permutation tests revealed that the public association signals for flowering time were highly enriched in the domestication and adaptation sweeps. Genome-wide association study identified 125 loci significantly associated with flowering-time traits, ten of which identified candidate genes that have undergone selection during maize adaptation.

**Conclusions:** In this study, we characterized the history of maize domestication and adaptation at the population genomic level and identified hundreds of domestication and adaptation sweeps. This study extends the molecular mechanism of maize domestication and adaptation, and provides resources for basic research and genetic improvement in maize.

**Keywords:** *Zea*, Evolutionary relationship, Domestication, Adaptation, Genome-wide association study, Flowering time

## Background

Maize (*Zea mays* L. ssp. *mays*) is the most widely planted crop species for food, feed, and industrial materials [1]. Maize, along with its wild relatives, also serves as an excellent model organism for understanding the genetic and functional mechanisms of plant domestication and adaptation. Maize and teosinte make up the genus *Zea*, which consists of five species distributed from northern Mexico through Central America [2–4]. The five species are *Zea nicaraguensis* (hereafter *nicaraguensis*), *Zea*

\*Correspondence: lihuihui@caas.cn; yxiaohong@cau.edu.cn

<sup>†</sup>Gen Xu, Xuan Zhang, Wenkang Chen and Renyu Zhang contributed equally to this work.

<sup>2</sup> Joint International Research Laboratory of Crop Molecular Breeding, China Agricultural University, Beijing 100193, China

<sup>5</sup> Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Full list of author information is available at the end of the article



*luxurians* (hereafter *luxurians*), *Zea diploperennis* (hereafter *diploperennis*), *Zea perennis* (hereafter *perennis*), and *Zea mays*. Of these, *diploperennis* and *perennis* are diploid and tetraploid perennial teosinte, respectively, whereas the others are diploid annual species. The annual species *Zea mays* consists of four subspecies, including the domesticated maize, the lowland adapted *Zea mays* ssp. *parviglumis* (hereafter *parviglumis*), the highland adapted *Zea mays* ssp. *mexicana* (hereafter *mexicana*), and the mid-altitude adapted *Zea mays* ssp. *huehuetenangensis* (hereafter *huehuetenangensis*). A refined understanding of the genetic relationship within the genus *Zea* can help elucidate the trajectories of maize domestication and adaptation.

Previously published genetic and archaeological data clearly reveal that maize was domesticated from *parviglumis* in a single domestication event in southern Mexico ~9000 years ago [5–7]. During this period, maize underwent dramatic phenotypic changes in both morphological and physiological characteristics [8–10]. The genetic basis of the morphological differences between maize and teosinte has been intensely investigated by quantitative trait locus (QTL) mapping using maize-teosinte populations [11–15]. However, only a limited number of domestication QTLs have been mapped to the underlying genes, including *teosinte branched1* (*tb1*) controlling branching [16–18], *teosinte glume architecture1* (*tga1*) controlling the formation of the stony fruit case [19, 20], and *grassy tillers1* (*gt1*) affecting prolificacy [21]. In addition to the cloning of single genes, population genetics comparisons of maize and teosinte have revealed evidence for positive selection in hundreds of genes during maize domestication [3, 22].

After its domestication, maize began to spread from southern Mexico into North and South America, where it adapted to these diverse environmental conditions [4, 5]. One of the most important events in this adaptation process was the divergence between tropical and temperate lines around 3400–6700 years ago [23]. Various environmental differences between temperate and tropical regions, such as temperature and day length, shaped maize diversity and facilitated its movement, and the footprints of this adaptation process were recorded in its genome. As with the study of the domestication process, genome-wide-level genotypic datasets provide an excellent resource for characterizing the genetic basis of adaptation. Adaptation studies involving these datasets linked to many aspects of the maize phenotypes and its metabolic pathways have identified a large number of selected loci, which reveal the complex genetic architecture of adaptation [23–25].

Flowering time is a key component in the adaptation of maize to local conditions as it moved to higher latitudes

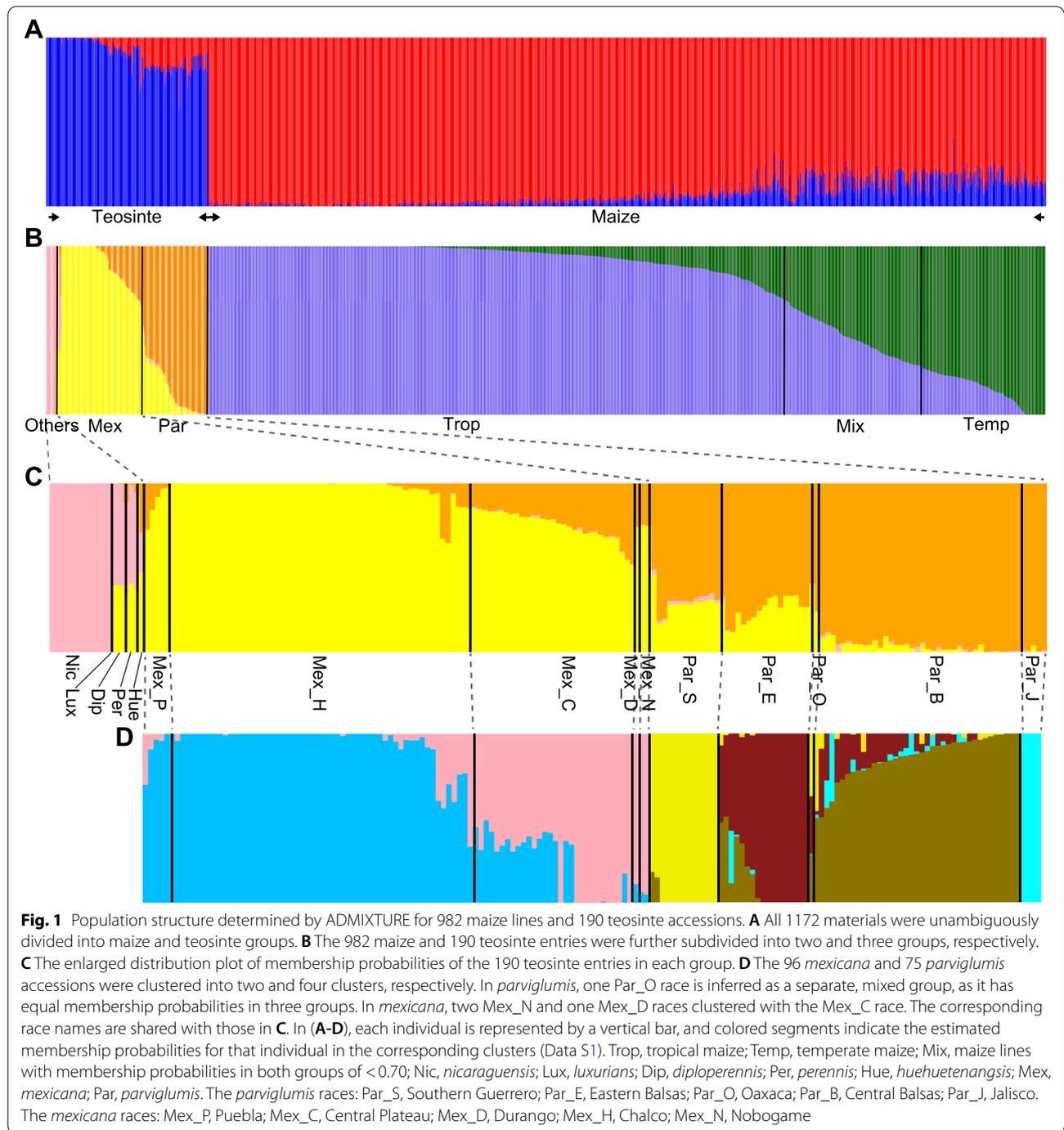
post-domestication. Five *cis*-variants in four genes, including a miniature transposon (MITE) located ~70 kb upstream of *ZmRap2.7* [26], a CACTA-like transposon in the *ZmCCT10* promoter [27], a Harbinger-like transposon located ~57 kb upstream of *ZmCCT9* [28], and SNP-1245 and InDel-2339 in the promoter of *ZCN8* [29], have been identified that contribute to phenotypes that allowed the pre-Columbian spread of maize throughout the Americas. The map of these interacting genes suggests that the SNP-1245A allele of *ZCN8* may have been the first to be selected, whereas the other four early-flowering alleles made specific contributions to northward expansion in North America [29]. These results suggest that the adaptation of maize was a complex process, involving numerous genetic loci that were selected at different evolutionary times for local adaptation [23].

During the past two decades, researchers have depicted the history of maize domestication and adaptation using genetic information from cultivated maize and its wild relatives [3, 5, 7, 22–24], but efforts have been limited either in sample size or geographic range. Here, a collection of 982 maize inbred lines representing global tropical, subtropical, and temperate germplasm and 190 teosinte accessions from Mexico and Central America were genotyped using the Illumina MaizeSNP50 Bead-Chip. We used this resource to determine the evolutionary relationship of the genus *Zea*, and to identify the loci that have undergone selection during maize domestication and adaptation. Subsequently, we performed co-localization analysis of selective sweeps with known selected genes, and associated genes for adaptation traits identified via genome-wide association studies (GWASs). We found that parts of the selected loci were associated with domestication and adaptation traits. This study will provide insights into maize evolutionary history, and the genetic resource should facilitate future maize breeding.

## Results

### Genetic structure within the genus *Zea*

Using 42,204 high-quality single-nucleotide polymorphisms (SNPs), all 1172 materials (982 maize inbred lines and 190 teosinte accessions) were unambiguously assigned to the maize or the teosinte clusters through population structure analysis (Fig. 1A; Data S1). Membership probabilities of each teosinte individual in the maize cluster ( $0 < P < 0.5$ ) reflected the common ancestry between some teosinte accessions and maize. Maize inbred lines were further divided into tropical/subtropical (hereafter tropical; 669 lines) and temperate (157 lines) subgroups (Fig. 1B; Data S1), consistent with the historical separation of these two subgroups [5, 30]. A substantial mixed group (156 lines) also shows the effect of more recent breeding efforts to expand diversity within



each breeding pool by bringing in germplasm from the other. Twelve teosinte accessions from *nicaraguensis* and one accession from *luxurians* clustered into a single subgroup (Fig. 1B, C; Data S1), suggesting the possibility of genetic similarity between these two subspecies. Accessions from *mexicana* and *parviglumis* clustered independently, and each formed a unique subgroup with 96

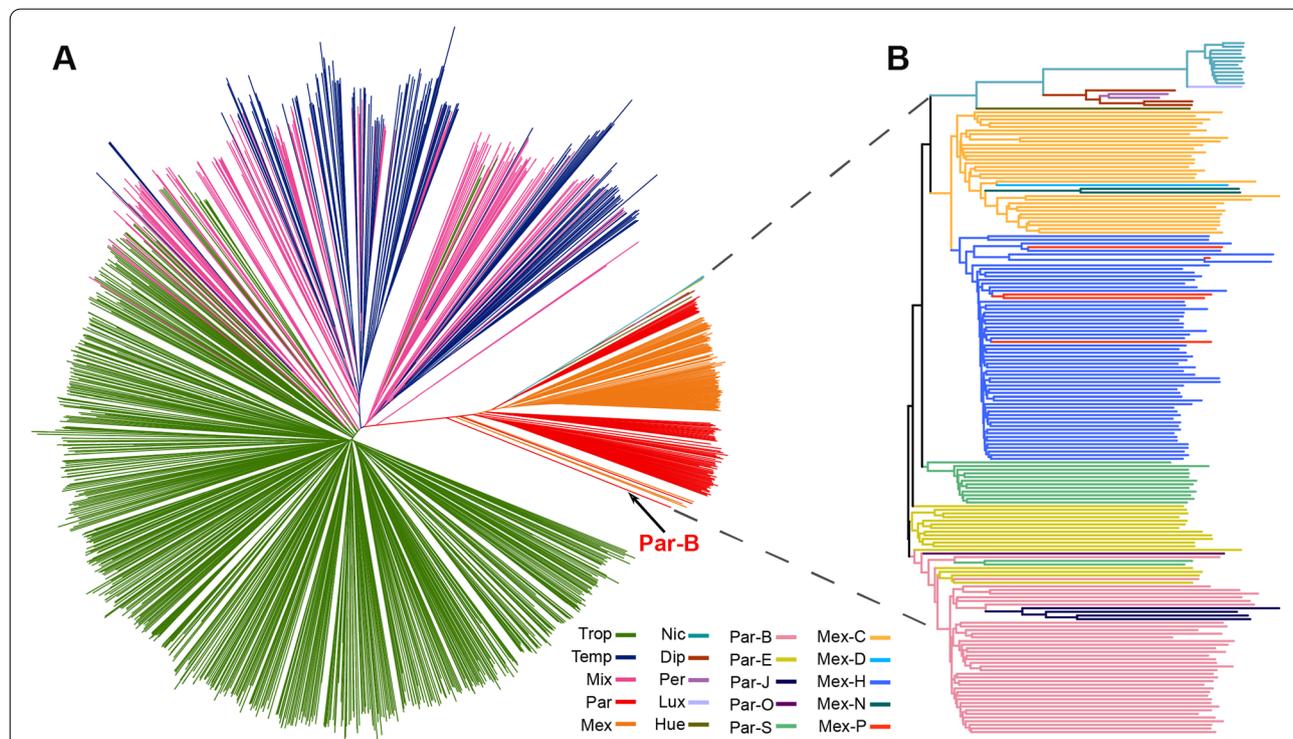
and 75 accessions, respectively (Fig. 1B, C; Data S1). The *diploperennis*, *perennis*, and *huehuetenangsis* accessions clustered into a mixed subgroup, and the membership probabilities of *diploperennis* and *perennis* in *mexicana* and *nicaraguensis* subgroups were similar (Fig. 1C; Data S1). Subsequent differentiation of *mexicana* and *parviglumis* revealed two major subgroups including

two *mexicana* clusters and four *parviglumis* clusters, in agreement with races classified by geographical distribution (Fig. 1D; Fig. S1; Data S1).

In addition to the population structure analysis, we also carried out a principal component analysis (PCA) using the same SNP data set, and found that the PCA results strongly supported the classification of species, subspecies, and races based on the population structure analysis of the genus *Zea* (Fig. S2). Whereas, the PCA plots show that the extreme points in maize represent B73 and Mo17, and that the spread of the maize points is distorted and over-stretched. This phenomenon might be caused by SNP ascertainment bias, especially from the Syngenta SNPs. To evaluate the effect of ascertainment bias caused by Syngenta SNPs, we re-analyzed population structure and principal component using 30,974 non-Syngenta SNPs. The results from ADMIXTURE show the correlation of membership probabilities calculated by all SNPs and non-Syngenta SNPs are pretty high ( $R^2 > 0.99$ ) for each assigned group (Fig. S3A, B). In addition, the PCA plots show similar distribution of maize (Fig. S3C, D).

Furthermore, we calculated the polymorphic information content (PIC) for each SNP, and found that the genetic diversity was quite similar between the results calculated from two different datasets (Fig. S3E). Taken together, these results suggest that the ascertainment bias caused by Syngenta SNPs did not affect the global estimation of genetic relationship and genetic diversity in the genus *Zea* although it indeed affected the genetic distance of maize inbred lines.

To identify the primary sources of maize genetic diversity, we constructed a neighbor-joining phylogenetic tree that included all entries in this study (Fig. 2). In the phylogenetic tree, the accession in the *luxurians* group was closest to *nicaraguensis* (chosen as the root of the tree), followed in order by *diploperennis* and *perennis*, *huehuetenangsis*, *mexicana*, *parviglumis*, and, finally, maize. These groupings reflect the evolutionary lineage of all *Zea* species and subspecies. The monophyletic clade including all maize lines (Fig. 2A) strongly supports a single domestication event in maize. The *parviglumis* accessions from the Central Balsas race were closest to maize



**Fig. 2** Neighbor-joining phylogenetic tree of maize and teosinte lines rooted with one *nicaraguensis* accession. **A** The genetic relationship of all maize plus teosinte. The *parviglumis* accessions from the Central Balsas race (Par\_B) were closest to maize although four *mexicana* accessions from three races (Mex\_H, Mex\_P, Mex\_C) were mixed among Par\_B. **B** The genetic relationship of all teosinte accessions and all races in *parviglumis* and *mexicana*. All materials are marked according to the inferred clusters from the ADMIXTURE analysis. Trop, tropical maize; Temp, temperate maize; Mix, maize lines with membership probabilities in both tropical and temperate groups of < 0.70; Nic, *nicaraguensis*; Lux, *luxurians*; Dip, *diploperennis*; Per, *perennis*; Hue, *huehuetenangsis*; Mex, *mexicana*; Par, *parviglumis*. The *parviglumis* races: Par\_S, Southern Guerrero; Par\_E, Eastern Balsas; Par\_O, Oaxaca; Par\_B, Central Balsas; Par\_J, Jalisco. The *mexicana* races: Mex\_P, Puebla; Mex\_C, Central Plateau; Mex\_D, Durango; Mex\_H, Chalco; Mex\_N, Nobogame

(Fig. 2), favoring the Balsas River valley as the center of maize domestication [5, 6, 31, 32]. In addition, the groups formed by the *mexicana* and *parviglumis* accessions seemed to be interconnected in a manner consistent with their geographical overlap (Fig. 2B; Fig. S1). Collectively, the evolutionary relationship of all *Zea* species and subspecies inferred by three methods is fully consistent with the current taxonomy of the genus *Zea*.

**Shared and unique haplotypes in maize and teosinte**

Because of their proximity to maize, further analyses were focused on *mexicana* and *parviglumis* teosinte, as compared with tropical and temperate maize. These comparisons allowed the determination of genetic variation acquired by maize from teosinte during the domestication period, as compared to variation partitioned during its adaptation from tropical to temperate environments. High pairwise  $F_{ST}$  among these four subgroups ( $0.10 < F_{ST} < 0.21$ ) indicated high population differentiation (Table S1). Furthermore, high pairwise  $F_{ST}$  between teosinte and maize and relatively small  $F_{ST}$  between tropical and temperate maize reflect maize domestication and adaptation history. Whereas, we found the haplotype richness in *parviglumis* was similar with that in tropical maize (Table 1). To exclude the biased estimation of haplotypes caused by sample size, we randomly selected 75 samples in each group with 100 bootstraps except *parviglumis* that had the smallest sample size. As expected, the window-based haplotype number in teosinte was much greater than modern maize, with the order following as *parviglumis* > *mexicana* > tropical maize > temperate maize (Fig. S4). These findings indicate that the genetic diversity in maize, especially temperate maize, was dramatically reduced during maize domestication and adaptation.

Many group-specific haplotypes were also observed in the four subgroups, *parviglumis*, *mexicana*, tropical and temperate maize (Fig. 3; Table 1). The presence of relatively fewer maize-specific haplotypes suggests that

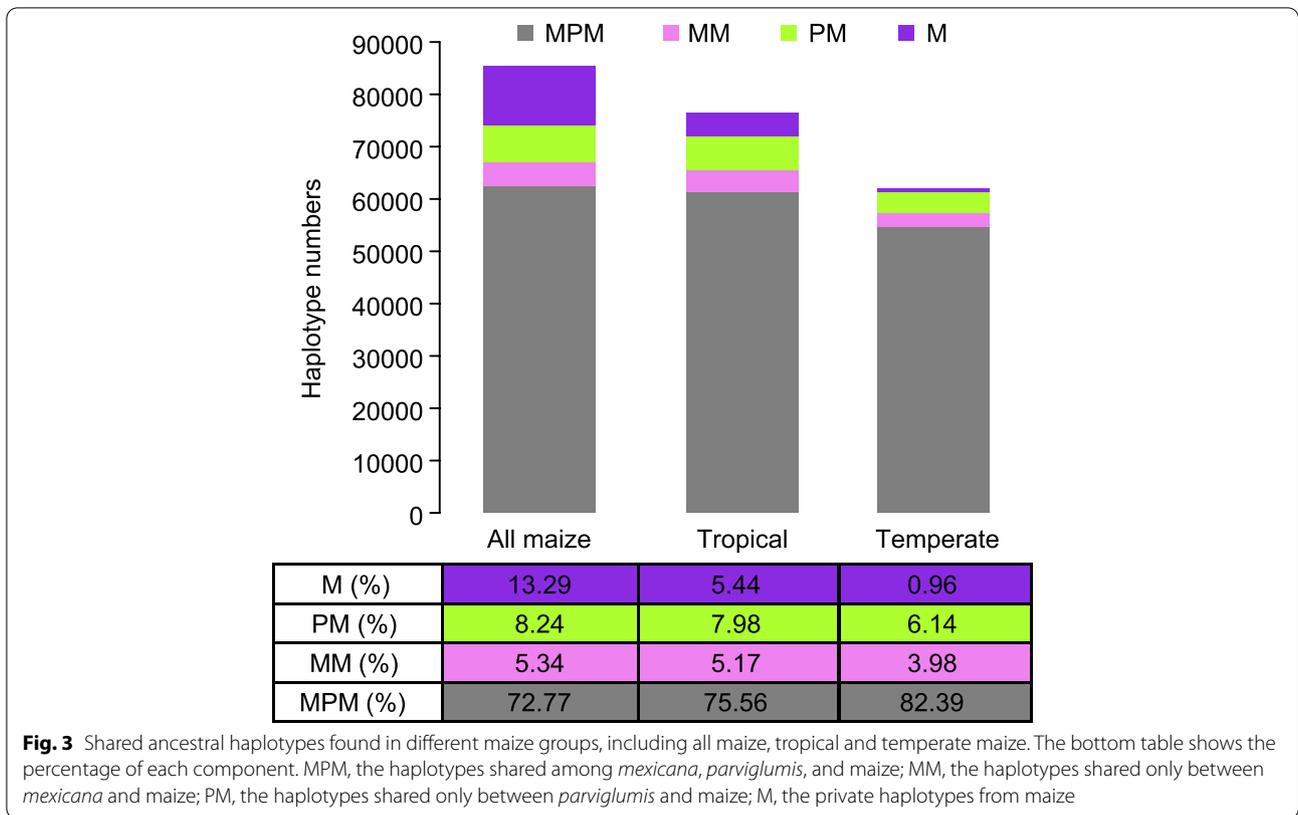
most of the diversity present in the domesticated maize gene pool is contributed by teosinte, and is not due to de novo haplotype creation since domestication. Both tropical and temperate maize had a great proportion of haplotypes in common with *parviglumis* and *mexicana* (Fig. 3), suggesting that both *parviglumis* and *mexicana* contributed to ancestral alleles into domesticated maize. Whereas, the contribution of *parviglumis* to maize during domestication may be overestimated because of the rapid expansion of the initial maize progenitor population.

**Footprints of selection in the genome**

The domestication of maize from its wild progenitor resulted in extreme morphological changes in plant and ear architecture, followed by further changes as a result of selection during crop adaptation [8, 33]. To determine if these changes can be detected as footprints of selection in the maize genome, two between-population comparisons, the calculation of  $F_{ST}$ , and a cross-population composite likelihood ratio (XP-CLR) approach, were implemented for sliding windows between teosinte and maize, and between tropical and temperate maize (Fig. 4; Table 2; Data S2). Based on the top 0.5% of XP-CLR and  $F_{ST}$  values, we identified 141 and 295 regions, respectively, associated with domestication, with 42 regions identified in common by both methods (Fig. 4C; Table 2). We similarly identified 138 and 268 regions, respectively, for adaptation, with 46 regions identified by both methods (Fig. 4D; Table 2). The small portion of overlapping sweeps (~30%) between different methods may be due to the different aspects the two methods focus on.  $F_{ST}$  is based on single marker analysis with large variance of its measurements, while XP-CLR is a model-based extension of  $F_{ST}$  to multiple-loci analysis using linkage disequilibrium (LD) in the reference population to weight SNPs and then to reduce the high ratio of false positives [34]. Collectively, we identified 394 regions with domestication features and 360 regions with adaptation

**Table 1** Summary statistics of window-based haplotypes for all lines, groups, and subgroups in this study

Groups	All entries	Maize			Teosinte		
		Overall	Tropical	Temperate	Overall	<i>mexicana</i>	<i>parviglumis</i>
Sample size	1172	982	671	148	190	96	75
Number of haplotypes	98,177	85,863	81,928	66,359	86,770	74,032	78,605
Number of haplotypes per locus	5.74	5.02	4.79	3.88	5.07	4.33	4.59
Haplotype richness	6.73	5.84	5.55	4.41	5.9	4.98	5.31
Group-specific haplotypes	23,721	11,407	4421	640	12,314	2536	4346
Group-specific haplotypes/line	20.24	11.62	6.59	4.32	64.81	26.42	57.95
Group-specific haplotype (%)	0.24	13.29	5.40	0.96	14.19	3.43	5.53



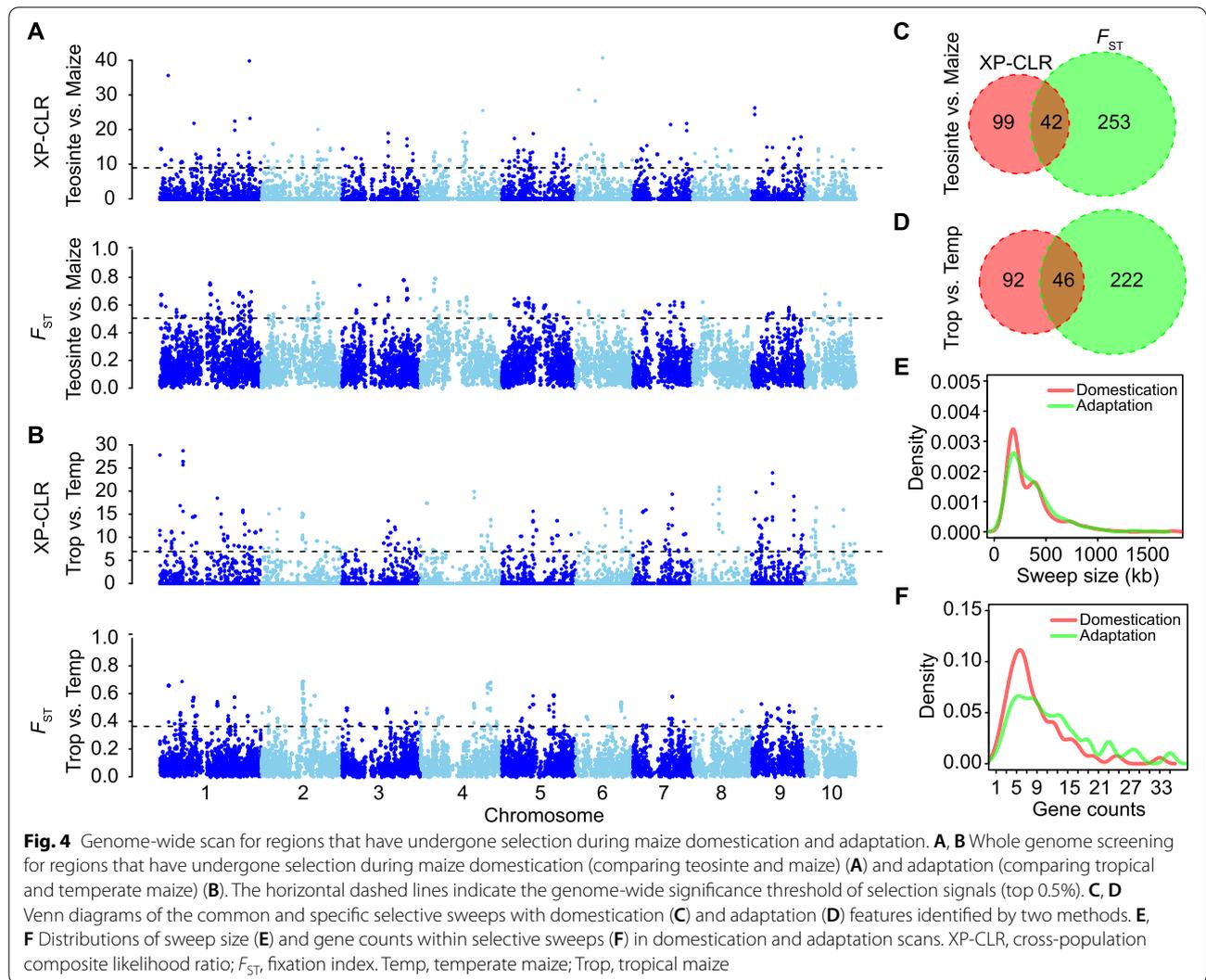
features, covering 5.7% (131 Mb) and 5.5% (127 Mb) of the genome, respectively (Table 2). For domestication, the size of these selection footprint regions ranged from 100 kb to 1.7 Mb, with a mean size of 333 kb, harboring 2218 genes; fewer selection footprint regions with a similar average size (352 kb) were detected during the adaptation process (Fig. 4E, F; Table 2). In addition, 69 of the domestication-related selective sweeps showed evidence of selection during adaptation, indicating that a subset of around 17% of the domestication loci may have also contributed to adaptation related phenotypes (Data S2).

To test if genetic variation within selected regions contributed to phenotypic changes during maize domestication and adaptation, we collected 29 previously reported genes with evidence of selection during domestication and adaptation (Table S2) and performed a co-localization analysis (Fig. 5; Data S2). Of the 29 genes, nine genes fell within the selective sweeps detected in our study, and eight genes which were previously reported to be domestication-related genes were physically located within the domestication-related selective sweeps identified here, i.e. *tb1* [18, 35], *pb1* [36], and *zagl1* [37, 38]. The finding that we didn't identify all the 29 known selected genes may be a consequence of the low marker density or different germplasms. Taken together, our results provide

evidence that some selective sweeps identified here are associated with domestication traits, although the causative genes in most selective sweeps remain unknown.

### Selection footprint regions associated with adaptation traits

To mine more loci or genes under selection during the adaptation process, we are using flowering-time traits as a representative for adaptation traits, since flowering time plays a key role in the process of adaptation that allowed maize to spread so widely [27–29]. We performed an additional co-localization analysis between selection sweeps and genomic regions associated with flowering-time traits as identified by genome-wide association studies (GWASs) [27, 39, 40]. A total of 32 domestication (8.1%) and 39 adaptation (10.8%) sweeps were co-located with GWAS signals for flowering time (Fig. 5; Data S2). Then we carried out a 1000-permutation test using the randomly sampled genomic regions with the same number and size as the selective sweeps compared to these public GWAS hits for flowering time [27, 39, 40]. The results revealed that the GWAS signals for flowering time were highly enriched in the domestication and adaptation sweeps (Permutation test,  $P < 0.001$ ) (Fig. S5). Notably, three reported flowering-time genes,



**Table 2** Summary of selective sweeps identified during maize domestication and adaptation

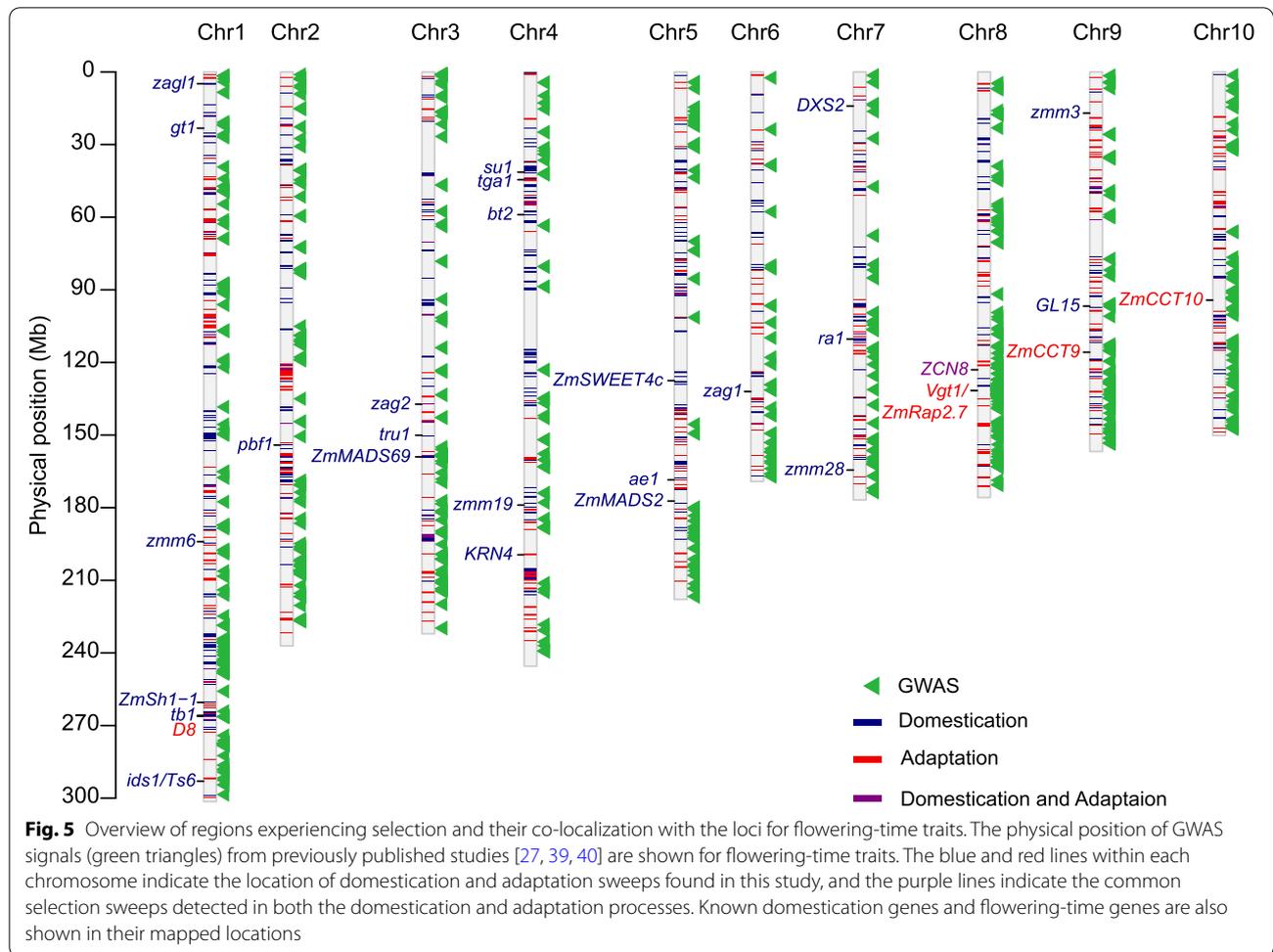
Selection Feature	Method	Selective sweeps			Gene number
		Number	Average length (kb)	Range (kb)	
Domestication	$F_{ST}$	295	386	110–1730	1759
	XP-LCR	141	166	100–420	585
	Total <sup>a</sup>	394	333	100–1730	2218
Adaptation	$F_{ST}$	268	412	120–1530	2181
	XP-LCR	138	171	100–340	622
	Total <sup>a</sup>	360	352	100–1530	2665

<sup>a</sup> shows the characteristics of the selective sweeps identified jointly by two methods

*ZmMADS69* [41], *PhyB1* [42, 43], and *zmm3* [37, 44], were detected within the GWAS signals as well as the selective sweeps. These results suggest that the genes underlying these co-localized regions for flowering-time

traits might have undergone selection during maize domestication and adaptation.

In addition to characterization of selected regions potentially related to flowering time, we compared our



selected regions to a marker-trait association mapping that was done for four flowering-time traits using a set of 508 maize inbred lines with ~1.25 million SNPs [45]. At a  $P$ -value  $\leq 6.05 \times 10^{-6}$  (1/165,202), a total of 10, 6, 11, and 4 loci were significantly associated with days to anthesis (DTA), days to silking (DTS), anthesis photoperiod response (APR), and silking photoperiod response (SPR), respectively (Data S3) when using best linear unbiased prediction (BLUP) values. Comparison of our selective sweeps to this GWAS on flowering-time traits using the set of 508 inbred lines grown at seven locations at diverse latitudes was also instructive; that GWAS identified 188 additional SNPs that resolved to 106 loci, and ten co-located with adaptation-related selective sweeps (Table 3; Data S3). The function of these ten candidate genes for flowering time that underwent selection during maize adaptation were annotated as transcription factors, flavonol synthase, MYB DNA-binding domain superfamily protein, etc (Table 3). Of these loci, association and

adaptation-related selective signals were both noted at the gene *GRMZM2G169293* (Fig. 6A, B), which encodes a ceramide and inositol phosphotransferase. We found that 77% of tropical inbred lines carried the C allele at the SNP (S8\_167550959) that showed the most significant association at the *GRMZM2G169293* locus, and that the percentage of lines with the C allele increased to 99% among temperate inbred lines (Fig. 6C). These contrasting frequency distributions suggest that the C allele of SNP S8\_167550959 might be associated with distinct patterns of geographic dispersal. Interestingly, SNP S8\_167550959 exhibited significant association with flowering time only at high latitudes, and the effects increased with latitude (except within Yunnan, China; Fig. 6D). Although the function of *GRMZM2G169293* affecting flowering time need more solid evidence, i.e., overexpression or mutant analysis, these findings suggested the characterization of genes responsible for adaptation from tropical to temperate regions.

**Table 3** Summary of maize adaptation loci that were significantly associated with flowering-time traits

Candidate Gene <sup>a</sup>	Lead trait <sup>b</sup>	Lead SNP <sup>c</sup>	Position <sup>d</sup>	Allele <sup>e</sup>	MAF <sup>f</sup>	P-value	Other trait <sup>g</sup>	Annotation
GRMZM2G382569	DTS_BJ	S1_62146730	Chr1_62146730	<u>G</u> /T	0.05	6.0 × 10 <sup>-6</sup>		Flavonol synthase/flavanone 3-hydroxylase
GRMZM2G180847	DTA_GX	S2_194077681	Chr2_194077681	I/ <u>C</u>	0.13	1.5 × 10 <sup>-6</sup>	DTS_GX	Probable transcription factor PosF21
GRMZM2G151044	DTA_YN	S4_180919965	Chr4_180919965	<u>C</u> /T	0.10	3.7 × 10 <sup>-6</sup>		Trehalose-6-phosphate phosphatase4
GRMZM2G111537	DTA_HB	S5_31012027	Chr5_31012027	<u>A</u> /C	0.48	5.1 × 10 <sup>-7</sup>	APR_BLUP/DTS_HB	Adenine nucleotide alpha hydrolases-like superfamily protein
GRMZM5G812923	DTA_DHN	S5_81991055	Chr5_81991055	<u>G</u> /A	0.09	1.3 × 10 <sup>-6</sup>	DTS_DHN	Protein SCO1 homolog 1 mitochondrial
GRMZM2G147867	DTA_YN	S6_147732326	Chr6_147732326	<u>G</u> /A	0.07	1.1 × 10 <sup>-6</sup>		NAC domain-containing protein 35
GRMZM2G435034	DTA_BJ	S8_156956875	Chr8_156956875	<u>C</u> /T	0.10	1.1 × 10 <sup>-6</sup>		MACPF domain-containing protein
GRMZM2G169293	DTA_BJ	S8_167550959	Chr8_167550959	<u>C</u> /G	0.10	6.0 × 10 <sup>-7</sup>		Ceramide inositol phosphotransferase 1
GRMZM2G166780	DTA_GX	S9_139587213	Chr9_139587213	<u>C</u> /T	0.15	4.5 × 10 <sup>-6</sup>		THO complex subunit 4B
GRMZM2G425427	DTA_BJ	S10_85333950	Chr10_85333950	<u>C</u> /G	0.08	6.6 × 10 <sup>-9</sup>		Putative MYB DNA-binding domain superfamily protein

<sup>a</sup> A plausible biological candidate gene at the locus or the nearest annotated gene to the lead associated SNP

<sup>b</sup> The trait with the most significant association signals among the tested traits. The trait is named by the abbreviation of flowering-time traits and locations for field trials. APR, anthesis photoperiod response; SPR, silking photoperiod response; DTA, days to anthesis; DTS, days to silking. DHN, Hainan; YN, Yunnan; GX, Guangxi; CQ, Chongqing; HB, Hubei; HN, Henan; BJ, Beijing; BLUP, best linear unbiased prediction

<sup>c</sup> The SNP with the most significant association signals for the lead trait at a given locus, where the major allele is underlined

<sup>d</sup> Chromosome and physical position for SNPs according to B73 RefGenV3 reference genome

<sup>e</sup> The major allele for each SNP is underlined

<sup>f</sup> MAF, minor allele frequency

<sup>g</sup> The remaining associated traits at a given locus

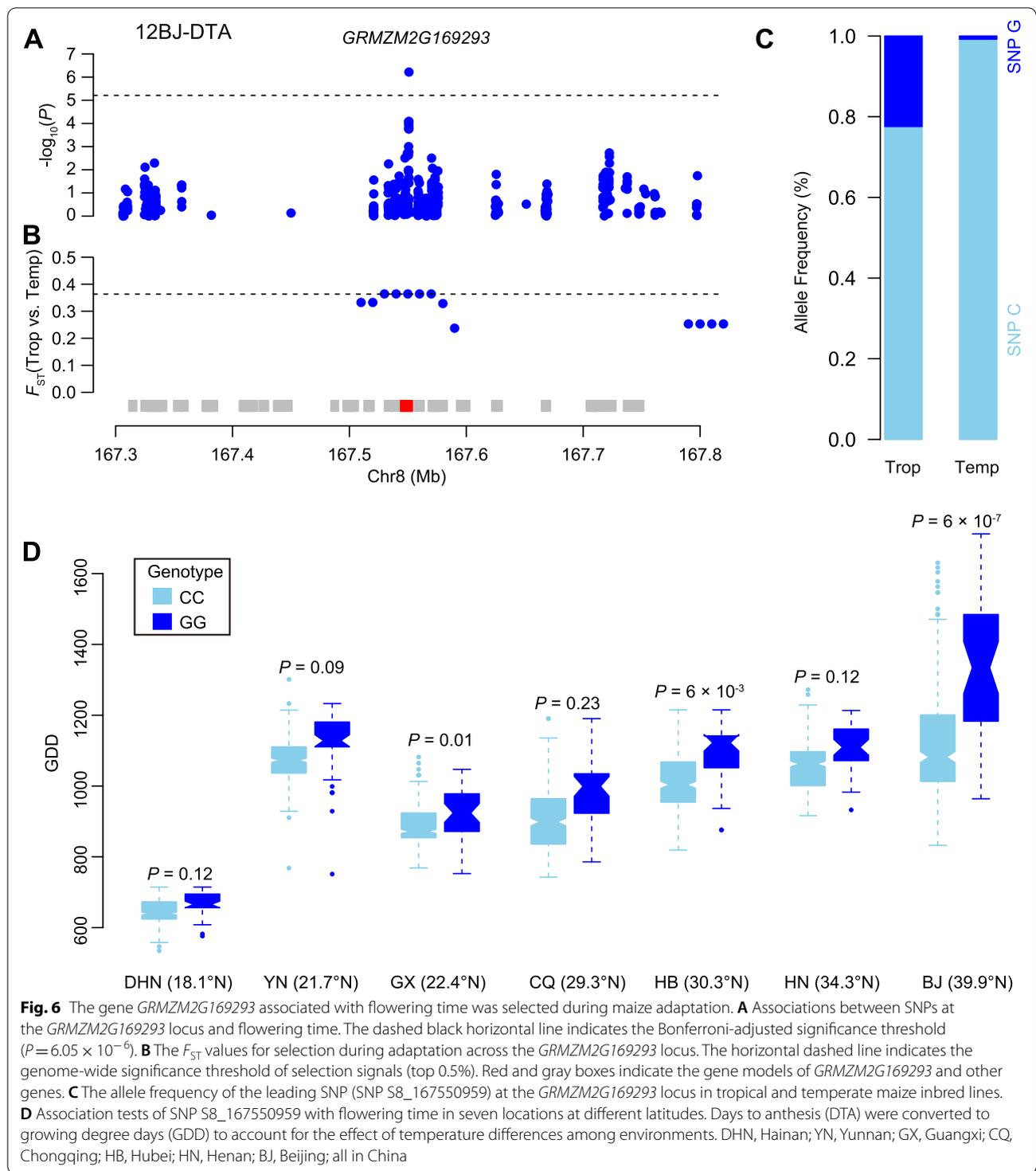
## Discussion

The germplasm analyzed here is comprised of an ecologically diverse collection of species including domesticated maize from tropical and temperate regions, and its close wild relatives. These taxa provide an excellent genetic resource to address multiple questions about speciation and evolution, structural and functional genomics, and utilization of teosinte germplasm in maize breeding. Cultivated maize has experienced a long period of artificial selection for desirable traits such as high yield (e.g., large seeds), nutrient richness (e.g., high levels of starch, oil, carotenoids, etc.), and ease of harvest [8–10, 15, 46]. This productivity-directed selection process generally results in the loss of genetic diversity in maize and an increased vulnerability to biotic and abiotic stresses [9].

Comparison of polymorphism data between maize landraces and teosinte in previous studies report a substantial loss (17%) of diversity during the domestication bottleneck [47]. Following further and more intense artificial selection, modern maize lost even more (18.6%) genetic diversity compared to teosinte [47]. Thus, in comparison with cultivated maize, its wild relative teosinte is a reservoir of genetic variation, and often exhibits

favorable nutritional attributes [15], stress resilience [48, 49], and even agronomic and yield performance [12–14, 50]. Multiple favorable alleles from teosinte have been mined, such as *ZmWAK* for resistance to head smut [51], *ZmNAC111* for drought tolerance in maize seedlings [52], and *UPA2* for leaf angle [50]. Notably, the teosinte *UPA2* allele reducing the leaf angle, which has a pretty low allele frequency (4.4%) in teosinte that has not been used in modern maize, was introgressed into an elite modern maize hybrid, Nongda108, via marker-assisted selection, and finally enhanced the maize yield under dense planting [50]. It is a successful example to incorporate the teosinte germplasm to improve the maize breeding. These findings suggest the potential to identify other beneficial variants useful for maize genetic improvement that may be hidden in teosinte. The five species of teosinte in the genus *Zea*, *parviglumis*, *mexicana*, *huehuetenangensis*, *diploperennis* and *luxurians*, can be hybridized with modern maize [3], enabling the transfer of favorable alleles that currently exist in wild relatives into modern maize breeding pools.

Capitalizing on the development of efficient genotyping technology, teosinte represents an attractive system



for the study of population and ecological genomics of maize domestication, introgressive hybridization, and local adaptation [3, 53]. In our study, different methods including ADMIXTURE analysis, PCA, and phylogenetic tree analysis clearly elucidated the genetic relationship

between maize and its wild relatives based on over 40,000 SNPs across the genome. Consistent with previous studies [2, 5, 6, 31, 32], our results confirm a single domestication event in maize from the Central Balsas *parviglumis* race and favor the Balsas River valley as the center of

maize domestication. Notably, the domestication process inferred from paleogenomic data was both gradual and complex, in which different genetic loci were selected at different time points, and the transformation of teosinte to maize was completed in the last 5000 years [54]. In addition to the evolutionary relationship between maize and teosinte, we also determined the evolutionary lineage of all species within teosinte, namely that *parviglumis* are closest to *mexicana*, followed in order by *huehuetenangsis*, *diploperennis* and *perennis*, *luxurians* and *nicaraguensis*. These findings answer a fundamental question in the taxonomic classification of teosintes, which has been debated during the last five decades [2, 55–59].

Our comparative genomic analysis between wild and modern maize, and between tropical and temperate maize, identified 5.7% of the genome that had been selected during maize domestication, and 5.5% of the genome that had been selected during adaptation. Our data cannot differentiate selective sweeps with domestication features from those with improvement features because we didn't look at maize landraces. In comparisons to previous studies, the size of the selected genomic regions we identified is smaller, and only 24% (95/394) of putative domestication-related selective sweeps overlapped with the results of Hufford et al. [50], and 17% (62/360) of putative adaptation genes overlapped with the results of Liu et al. [23]. These low percentages may result from different genetic germplasms, sample sizes, and SNP densities as well as from differences in the quality of the reference genome (Table S3). Although the SNP density used was relatively low, the larger sample size in our study shows greater genetic diversity (Fig. S6) and could increase the power of detecting selection signals [60]. With newer developments in sequencing technology, re-sequencing our germplasm plus a set of maize landraces will refine what we are able to conclude about maize domestication, improvement, and adaptation.

Maize was subjected to drastic morphological or physiological changes during domestication that now differentiate it from its teosinte progenitor. Given these changes, the selective sweeps identified in this study could be associated with domestication and adaptation traits. These associations were supported by the co-location of the selective sweeps identified here and eight domestication genes (e.g., *tb1* [16–18] and *pb1* [36]) plus a set of GWAS signals for flowering-time traits (Fig. 5). In addition to the known genes and existing GWAS signals reported in previous studies, ten candidate genes were identified that colocalized at both GWAS and selection signals. As an example, *GRMZM2G169293* had a genetic effect on flowering time that was dependent on altitude. Similar trends have been seen in known adaptation genes such as *ZmCCT10* [27] and *ZmCCT9* [28].

Such temperature-related highland adaptation loci could be important for maize breeding in the face of climate change [3]. Therefore, identification of selective sweeps during maize domestication and adaptation will extend our understanding of these processes, and greatly benefit maize breeding if this information is included in the process of maize improvement.

## Conclusions

In summary, we determined the genetic structure reflected the historical evolutionary relationships among *Zea* species and subspecies, namely that maize is closest to *parviglumis*, followed by *mexicana*, *huehuetenangsis*, *diploperennis* and *perennis*, *luxurians* and *nicaraguensis*. Our comparative population genomic studies identified more than 600 domestication and adaptation sweeps, and the existing GWAS hits for flowering time were highly enriched in the selective sweeps. Combining with the GWAS results, we identified ten candidate genes that were significantly associated with adaptation traits and that have undergone selection during maize adaptation. Notably, a candidate gene *GRMZM2G169293* was identified, which located within an adaptation selective sweep and was associated with photoperiod responses. Taken together, our results will provide increasing insights into the evolutionary history of maize and will greatly benefit the maize breeding.

## Materials and methods

### Plant material

A set of 982 maize lines and 190 teosinte accessions were used in this study. The maize lines, representative of tropical, subtropical, and temperate germplasm, were collected from maize breeding programs of the International Maize and Wheat Improvement Center (CIMMYT) ( $n=691$ ), China ( $n=221$ ), the USA ( $n=66$ ), Thailand ( $n=3$ ), and Peru ( $n=1$ ) (Data S4). The teosinte accessions, representative of the entire geographical distribution of teosinte across Mexico and Central America, included 12 *nicaraguensis*, one *luxurians*, three *diploperennis*, two *perennis*, one *huehuetenangsis*, 96 *mexicana*, and 75 *parviglumis* accessions (Data S4). Based on their geographical distribution, the *mexicana* accessions were further divided into five geographical groups from Puebla, Central Plateau, Chalco, Durango and Nobogame, and *parviglumis* accessions were also further divided into five geographical groups from Southern Guerrero, Oaxaca, Eastern Balsas, Central Balsas, and Jalisco (Fig. S1 and Data S4).

### Genotyping and SNP quality control

DNA was extracted from leaves that were obtained from a pool of at least six individuals for each maize line and

one individual per teosinte accession. All maize lines and teosinte accessions were genotyped using the Illumina MaizeSNP50 BeadChip (Illumina Inc., San Diego, CA, USA) containing 56,110 SNPs [61]. SNP genotypes were manually checked as reported previously [62]. A total of 2353 SNPs with poor performance were removed from subsequent analyses. In addition, only the SNPs with probe sequences uniquely mapped to the B73 reference genome (B73 RefGenV3) using the Burrows-Wheeler Aligner (BWA) were retained [63]. A final set of 42,204 polymorphic and single-copy SNPs with <20% missing data across all 1172 accessions was used in the final analyses. The PIC for each SNP was calculated using PowerMarker version 3.25 [64].

### Population structure analysis

Hierarchical population structure of all maize lines and teosinte accessions was estimated with the program ADMIXTURE, which implemented a Structure-like model-based maximum likelihood clustering algorithm [65]. The maize lines and teosinte accessions were subsequently analyzed separately. For maize, lines with membership probabilities  $\geq 0.70$  were assigned to that corresponding group, and lines with a probability of <0.70 for both the temperate and tropical groups were assigned to a mixed group. For teosinte, the entries were assigned to the corresponding subspecies and geographical groups based on their known origins and ADMIXTURE results. ADMIXTURE results showing individual assignments to corresponding groups were graphically displayed using R Version 3.1.1 ([www.R-project.org](http://www.R-project.org)).

### Visualization of relationships

PCA was performed at the individual level using the GCTA software [66]. Subgroups were formed that included all maize and teosinte accessions, maize inbred lines only, teosinte accessions only, and teosinte accessions split into two subgroups. The first three principal components were used to visualize the genetic relatedness among individuals and to investigate the groups. The identity-by-state distance matrix was calculated between each pair of lines with PLINK Version 1.7 [67], and was then imported into the MEGA6 program [68] to construct a neighbor-joining phylogenetic tree. One *nicaraguensis* accession was used as the outgroup.

### Haplotype phasing and visualization

Haplotype phasing was done independently for each chromosome by SHAPEIT Version 2.12 [69, 70] with 2-Mb window size, 20 burn-in iterations, 20 iterations of the pruning stage, and 30 main iterations. Then the genome was divided into 50-kb windows to determine the haplotypes of linked SNPs in each window. If a

window contains more than five SNPs, a random subset of five SNPs was selected for haplotype analysis, and the same randomly selected SNPs were used for all individuals. As a result, the SNP number used for haplotype analysis in each window ranged from one to five. For subsequent analyses, each haplotype window was defined as a locus, and each unique haplotype within the window was defined as an allele. In total, 17,109 loci were visualized for the window-based haplotypes.

### Genome scanning for regions that have undergone selection

To achieve maximum statistical power, XP-CLR hosted on GitHub [34] was implemented along with the population fixation statistic,  $F_{ST}$ , using VCFtools [71] to detect loci that may have undergone selection during maize domestication and adaptation. In the analysis of XP-CLR, we used a 100-kb sliding window and a 10-kb step size. To ensure comparability of the composite likelihood score in each window, we fixed the number of SNPs assayed in each window to five with the setting '--maxsnps 5 --minsnps 5' [34]. Meanwhile, to keep the used genomic windows consistent in the XP-CLR analysis, the weighted  $F_{ST}$  values were estimated in each window that required at least five SNPs with the setting '--fst-window-size 100,000 --fst-window-step 10,000' [71]. Pairwise differentiation between populations ( $F_{ST}$ ) was calculated using the "hierfstat" package of R [72].

Evidence for selection across the genome during the domestication and adaptation processes were evaluated in two separate comparisons: teosinte versus maize for domestication and tropical maize lines versus temperate maize lines for adaptation. For each method, we merged the adjacent windows with top 10% values into a single window, and the top 0.5% outliers were determined to represent putative selection signals. In addition, adjacent sweeps separated by a physical distance of <100 kb were merged into a single selected locus.

### Genome-wide association mapping for flowering-time traits

The 508 diverse inbred lines that made up an association mapping panel [73] were planted in seven environments, including six long-day (>13 h) and one short-day (<13 h) growing-season environments [27]. Flowering time was recorded as DTA and DTS, and these values were then converted into GDDs. APR and SPR were calculated as the difference between GDDs under long- and short-day conditions for pollen shed and silking, respectively. The best linear unbiased prediction (BLUP) values for each trait were used for the marker-trait association analysis. Using ~1.25 million previously reported SNPs with a minor allele

frequency of  $\geq 0.05$  [45], the marker-trait association analyses were performed using a mixed linear model [74] presented in TASSEL 5.2 [75], which accounted for population structure and relative kinship [76]. Because the SNPs used for GWAS are in LD at different levels, we first performed LD pruning for the 1.25 million SNPs (window size 50, step size 50,  $r^2 \geq 0.2$ ) using PLINK [67], and obtained 165,202 independent SNPs. Consequently, the Bonferroni-corrected threshold,  $6.05 \times 10^{-6}$  ( $P < 1/165,202$ ), was used as the whole-genome significance cutoff. Marker-trait associations were also analyzed with this dataset for flowering time in each environment.

### Data availability

SNP data for this study has been uploaded to European Variation Archive and can be retrieved through the project number PRJEB41335 (<http://wwwdev.ebi.ac.uk/eva/?eva-study=PRJEB41335>).

### Abbreviations

QTL: Quantitative trait locus; MITE: Miniature transposon; SNP: Single-nucleotide polymorphism; PCA: Principal component analysis; XP-CLR: Cross-population composite likelihood ratio; LD: Linkage disequilibrium; GWAS: Genome-wide association studies; NAM: Nested-association mapping; DTA: Days to anthesis; DTS: Days to silking; GDD: Growing degree days; APR: Anthesis photoperiod response; SPR: Silking photoperiod response; BLUP: Best linear unbiased prediction; MAF: Minor allele frequency; PIC: Polymorphic information content.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-022-03427-w>.

**Additional file 1: Figure S1.** Geographical distribution of all teosinte accessions. **Figure S2.** Genetic relationships of maize and teosinte assessed by PCA. **Figure S3.** Evaluation of the ascertainment bias caused by Syngenta SNPs. **Figure S4.** Haplotype richness in maize and teosinte groups estimated via window-based methods. **Figure S5.** Co-localization of putative selective sweeps with public GWAS hits for flowering time. **Figure S6.** Genetic relationships of maize and teosinte assessed by PCA using 36,839 common SNPs between this study and Hufford et al's study. **Table S1.** Population divergence among maize and teosinte subgroups estimated by pairwise  $F_{ST}$  values between different groups. **Table S2.** List of known domestication, improvement and adaptation genes in maize. **Table S3.** Comparisons of selective sweeps identified in this study and previous studies, and the factors affecting the identification of selective sweeps.

**Additional file 2: Data S1.** Genetic relationship of maize and teosinte inferred by ADMIXTURE.

**Additional file 3: Data S2.** Summary of selection sweeps with domestication and adaptation features.

**Additional file 4: Data S3.** Summary of SNPs significantly associated with flowering-time traits detected by GWAS.

**Additional file 5: Data S4.** List of plant materials used in this study.

### Acknowledgements

We greatly appreciate Dr. Jianbing Yan at Huazhong Agricultural University and Drs. S. Taba, B.M. Prasanna, S.D. Nicolas and H.H. Kuang of the International Maize and Wheat Improvement Center (CIMMYT) for their critical comments about the early version of this manuscript.

### Authors' contributions

X.Y. and H.L. conceived and designed the research. G.X., X.Z., W.C., R.Z., and H.L. analyzed the data. X.Z., W.C., R.Z., and Z.L. performed the experiments. W.W. provided the geography information for teosinte. G.X., M.L.W., and X.Y. wrote the manuscript. X.Z., W.C., R.Z., J.L., and H.L. edited the manuscript. All authors have read and approved the manuscript.

### Funding

This work was supported by the National Natural Science Foundation of China (91935302, 31722039, 32022064) and Beijing Outstanding Young Scientist Program (BJJWZYJH01201910019026).

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>State Key Laboratory of Plant Physiology and Biochemistry, National Maize Improvement Center of China, MOA Key Laboratory of Maize Biology, China Agricultural University, Beijing 100193, China. <sup>2</sup>Joint International Research Laboratory of Crop Molecular Breeding, China Agricultural University, Beijing 100193, China. <sup>3</sup>Key Laboratory of Horticultural Plant Biology (MOE), College of Horticulture and Forestry Sciences, Huazhong Agricultural University, Wuhan 430070, China. <sup>4</sup>United States of Department of Agriculture, Agricultural Research Service, Corn Host Plant Resistance Research Unit, Box 9555, Mississippi, MS 39762, USA. <sup>5</sup>Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China.

Received: 3 August 2021 Accepted: 5 January 2022

Published online: 18 February 2022

### References

- Ranum P, Pena-Rosas JP, Garcia-Casal MN. Global maize production, utilization, and consumption. *Ann N Y Acad Sci*. 2014;1312(1):105–12.
- Fukunaga K, Hill J, Vigouroux Y, Matsuoka Y, Sanchez J, Liu KJ, et al. Genetic diversity and population structure of teosinte. *Genetics*. 2005;169(4):2241–54.
- Hufford MB, Bilinski P, Pyhajarvi T, Ross-Ibarra J. Teosinte as a model system for population and ecological genomics. *Trends Genet*. 2012;28(12):606–15.
- Manchanda N, Snodgrass SJ, Ross-Ibarra J, Hufford MB. Evolution and adaptation in the maize genome. In: *The Maize Genome*: Springer; 2018. p. 319–32.
- Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez J, Buckler E, Doebley J. A single domestication for maize shown by multilocus microsatellite genotyping. *Proc Natl Acad Sci U S A*. 2002;99(9):6080–4.
- Piperno DR, Ranere AJ, Holst I, Iriarte J, Dickau R. Starch grain and phytolith evidence for early ninth millennium BP maize from the central Balsas River valley. Mexico. *Proc Natl Acad Sci USA*. 2009;106(13):5019–24.
- van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, Gonzalez JDS, et al. Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proc Natl Acad Sci U S A*. 2011;108(3):1088–92.
- Doebley J. The genetics of maize evolution. *Annu Rev Genet*. 2004;38:37–59.

9. Doebley JF, Gaut BS, Smith BD. The molecular genetics of crop domestication. *Cell*. 2006;127(7):1309–21.
10. Flint-Garcia SA. Kernel evolution: from teosinte to maize. In: *Maize Kernel Development*; 2017. p. 1–15.
11. Liu Z, Cook J, Melia-Hancock S, Guill K, Bottoms C, Garcia A, et al. Expanding maize genetic resources with predomestication alleles: Maize-teosinte introgression populations. *Plant Genome*. 2016;9(1):plantgenome2015-07.
12. Chen Q, Yang CJ, York AM, Xue W, Daskalska LL, DeValk CA, et al. TeoNAM: a nested association mapping population for domestication and agronomic trait analysis in maize. *Genetics*. 2019;213(3):1065–78.
13. Fu Y, Xu G, Chen H, Wang X, Chen Q, Huang C, et al. QTL mapping for leaf morphology traits in a large maize-teosinte population. *Mol Breeding*. 2019;39:103.
14. Liu L, Huang J, He L, Liu N, Du Y, Hou R, et al. Dissecting the genetic architecture of important traits that enhance wild germplasm resource usage in modern maize breeding. *Mol Breeding*. 2019;39:157.
15. Fang H, Fu X, Wang Y, Xu J, Feng H, Li W, et al. Genetic basis of kernel nutritional traits during maize domestication and improvement. *Plant J*. 2020;101(2):278–92.
16. Doebley J, Stec A, Gustus C. *Teosinte branched1* and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics*. 1995;141(1):333–46.
17. Doebley J, Stec A, Hubbard L. The evolution of apical dominance in maize. *Nature*. 1997;386(6624):485–8.
18. Studer A, Zhao Q, Ross-Ibarra J, Doebley J. Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat Genet*. 2011;43(11):1160–3.
19. Doebley J, Stec A. Inheritance of the morphological differences between maize and teosinte: comparison of results for two F<sub>2</sub> populations. *Genetics*. 1993;134(2):559–70.
20. Wang H, Nussbaum-Wagler T, Li B, Zhao Q, Vigouroux Y, Faller M, et al. The origin of the naked grains of maize. *Nature*. 2005;436(7051):714–9.
21. Wills DM, Whipple CJ, Takuno S, Kursel LE, Shannon LM, Ross-Ibarra J, et al. From many, one: genetic control of prolificacy during maize domestication. *PLoS Genet*. 2013;9:e1003604.
22. Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, et al. The effects of artificial selection on the maize genome. *Science*. 2005;308(5726):1310–4.
23. Liu H, Wang X, Warburton ML, Wen W, Jin M, Deng M, et al. Genomic, transcriptomic, and phenomic variation reveals the complex adaptation of modern maize breeding. *Mol Plant*. 2015;8(6):871–84.
24. Pyhäjärvi T, Hufford MB, Mezouk S, Ross-Ibarra J. Complex patterns of local adaptation in teosinte. *Genome Biol Evol*. 2013;5(9):1594–609.
25. Takuno S, Ralph P, Swarts K, Elshire RJ, Glaubitz JC, Buckler ES, et al. Independent molecular basis of convergent highland adaptation in maize. *Genetics*. 2015;200:1297–312.
26. Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, et al. Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc Natl Acad Sci U S A*. 2007;104(27):11376–81.
27. Yang Q, Li Z, Li W, Ku L, Wang C, Ye J, et al. CACTA-like transposable element in *ZmCCT* attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc Natl Acad Sci U S A*. 2013;110(42):16969–74.
28. Huang C, Sun H, Xu D, Chen Q, Liang Y, Wang X, et al. *ZmCCT9* enhances maize adaptation to higher latitudes. *Proc Natl Acad Sci U S A*. 2018;115(2):E334–41.
29. Guo L, Wang X, Zhao M, Huang C, Li C, Li D, et al. Stepwise cis-regulatory changes in *ZCN8* contribute to maize flowering-time adaptation. *Curr Biol*. 2018;28(18):3005–15.
30. Vigouroux Y, Glaubitz JC, Matsuoka Y, Goodman MM, Jesus SG, Doebley J. Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. *Am J Bot*. 2008;95(10):1240–53.
31. Doebley J. Molecular evidence and the evolution of maize. *Econ Bot*. 1990;44(3):6–27.
32. Ranere AJ, Piperno DR, Holst I, Dickau R, Iriarte J. The cultural and chronological context of early Holocene maize and squash domestication in the central Balsas River valley, Mexico. *Proc Natl Acad Sci USA*. 2009;106(13):5014–8.
33. Hake S, Ross-Ibarra J. Genetic, evolutionary and plant breeding insights from the domestication of maize. *eLife*. 2015;4:e05861.
34. Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. *Genome Res*. 2010;20(3):393–402.
35. Clark RM, Wagler TN, Quijada P, Doebley J. A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat Genet*. 2006;38(5):594.
36. Lang Z, Wills DM, Lemmon ZH, Shannon LM, Bukowski R, Wu Y, et al. Defining the role of *prolamin-box binding factor1* gene during maize domestication. *J Hered*. 2014;105(4):576–82.
37. Zhao Q, Weber AL, McMullen MD, Guill K, Doebley J. MADS-box genes of maize: frequent targets of selection during domestication. *Genet Res*. 2011;93(1):65–75.
38. Wills DM, Fang Z, York AM, Holland JB, Doebley JF. Defining the role of the MADS-box gene, *Zea Agamous-like1*, a target of selection during maize domestication. *J Hered*. 2018;109(3):333–8.
39. Lu F, Romay MC, Glaubitz JC, Bradbury PJ, Elshire RJ, Wang T, et al. High-resolution genetic mapping of maize pan-genome sequence anchors. *Nat Commun*. 2015;6(1):1–8.
40. Li YX, Li C, Bradbury PJ, Liu X, Lu F, Romay CM, et al. Identification of genetic variants associated with maize flowering time using an extremely large multi-genetic background population. *Plant J*. 2016;86(5):391–402.
41. Liang Y, Liu Q, Wang X, Huang C, Xu G, Hey S, et al. *ZmMADS69* functions as a flowering activator through the *ZmRap2.7-ZCN8* regulatory module and contributes to maize flowering time adaptation. *New Phytol*. 2019;221(4):2335–47.
42. Sheehan MJ, Farmer PR, Brutnell TP. Structure and expression of maize phytochrome family homeologs. *Genetics*. 2004;167(3):1395–405.
43. Sheehan MJ, Kennedy LM, Costich DE, Brutnell TP. Subfunctionalization of *PhyB1* and *PhyB2* in the control of seedling and mature plant traits in maize. *Plant J*. 2007;49(2):338–53.
44. Studer AJ, Wang H, Doebley JF. Selection during maize domestication targeted a gene network controlling plant and inflorescence architecture. *Genetics*. 2017;207(2):755–65.
45. Liu H, Luo X, Niu L, Xiao Y, Chen L, Liu J, et al. Distant eQTLs and non-coding sequences play critical roles in regulating gene expression and quantitative trait variation in maize. *Mol Plant*. 2017;10(3):414–26.
46. Flint-Garcia SA, Bodnar AL, Scott MP. Wide variability in kernel composition, seed characteristics, and zein profiles among diverse maize inbreds, landraces, and teosinte. *Theor Appl Genet*. 2009;119(6):1129–42.
47. Hufford MB, Xu X, van Heerwaarden J, Pyhäjärvi T, Chia JM, Cartwright RA, et al. Comparative population genomics of maize domestication and improvement. *Nat Genet*. 2012;44(7):808–11.
48. Mano Y, Omori F. Breeding for flooding tolerant maize using "teosinte" as a germplasm resource. *Plant Root*. 2007;1:17–21.
49. Wang L, Yang A, He C, Qu M, Zhang J. Creation of new maize germplasm using alien introgression from *Zea mays* ssp. *mexicana*. *Euphytica*. 2008;164(3):789–801.
50. Tian J, Wang C, Xia J, Wu L, Xu G, Wu W, et al. Teosinte ligule allele narrows plant architecture and enhances high-density maize yields. *Science*. 2019;365(6454):658–64.
51. Zuo W, Chao Q, Zhang N, Ye J, Tan G, Li B, et al. A maize wall-associated kinase confers quantitative resistance to head smut. *Nat Genet*. 2015;47(2):151–7.
52. Mao H, Wang H, Liu S, Li Z, Yang X, Yan J, et al. A transposable element in a *NAC* gene is associated with drought tolerance in maize seedlings. *Nat Commun*. 2015;6(1):1–13.
53. Aguirre-Liguori JA, Tenaillon MI, Vazquez-Lobo A, Gaut BS, Jaramillo-Correa JP, Montes-Hernandez S, et al. Connecting genomic patterns of local adaptation and niche suitability in teosintes. *Mol Ecol*. 2017;26(16):4226–40.
54. Ramos-Madriral J, Smith BD, Moreno-Mayar JV, Gopalakrishnan S, Ross-Ibarra J, Gilbert MTP, et al. Genome sequence of a 5,310-year-old maize cob provides insights into the early stages of maize domestication. *Curr Biol*. 2016;26(23):3195–201.
55. Wilkes HG. Teosinte: the closest relative of maize. Teosinte: the closest relative of maize; 1967.
56. Illis HH, Doebley JF, Guzmán R, Pazy B. *Zea diploperennis* (Gramineae): a new teosinte from Mexico. *Science*. 1979;203(4376):186–8.

57. Sanchez JJ, De la Cruz L, Vidal VA, Ron J, Taba S, Santacruz-Ruvalcaba F, et al. Three new teosintes (*Zea* spp., Poaceae) from México. *Am J Bot.* 2011;98(9):1537–48.
58. Pena GT, Larios LD, Gonzales JDS, Corral JAR, Nava JJC, Santacruz-Ruvalcaba F, et al. Relationships among teosinte populations (*Zea* spp.) from Mexico, Guatemala and Nicaragua. *Acta Bot Mex.* 2015;111:17–45.
59. Rivera-Rodriguez DM, Gonzalez JDS, Larios LD, Santacruz-Ruvalcaba F, Corral JAR. Morphological and climatic variability of teosinte (*Zea* spp.) and relationships among taxa. *Syst Bot.* 2019;44(1):41–51.
60. Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, et al. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* 2009;19(5):826–37.
61. Ganai MW, Durstewitz G, Polley A, Berard A, Buckler ES, Charcosset A, et al. A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS One.* 2011;6:e28334.
62. Yan J, Yang X, Shah T, Sanchez-Villeda H, Li J, Warburton M, et al. High-throughput SNP genotyping with the Golden Gate assay in maize. *Mol Breeding.* 2010;25(3):441–51.
63. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2009;25(14):1754–60.
64. Liu K, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics.* 2005;21(9):2128–9.
65. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 2009;19(9):1655–64.
66. Yang JA, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88:76–82.
67. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559–75.
68. Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* 2013;30(12):2725–9.
69. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods.* 2012;9(2):179–81.
70. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods.* 2013;10(1):5–6.
71. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011;27(15):2156–8.
72. Goudet J. HIERFSTAT, a package for R to compute and test hierarchical *F*-statistics. *Mol Ecol Notes.* 2005;5(1):184–6.
73. Yang X, Gao S, Xu S, Zhang Z, Prasanna BM, Li L, et al. Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol Breeding.* 2011;28(4):511–26.
74. Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet.* 2006;38(2):203–8.
75. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics.* 2007;23(19):2633–5.
76. Li Q, Yang X, Xu S, Cai Y, Zhang D, Han Y, et al. Genome-wide association studies identified three independent polymorphisms associated with  $\alpha$ -tocopherol content in maize kernels. *PLoS One.* 2012;7:e36807.

## Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

