

RESEARCH ARTICLE

Open Access



# Cytoplasmic genome contributions to domestication and improvement of modern maize

Shuai Cao<sup>1,4†</sup>, Huanhuan Zhang<sup>2†</sup>, Yang Liu<sup>1</sup>, Yi Sun<sup>2</sup> and Z. Jeffrey Chen<sup>3\*</sup> 

## Abstract

**Background** Studies on maize evolution and domestication are largely limited to the nuclear genomes, and the contribution of cytoplasmic genomes to selection and domestication of modern maize remains elusive. Maize cytoplasmic genomes have been classified into fertile (NA and NB) and cytoplasmic-nuclear male-sterility (CMS-S, CMS-C, and CMS-T) groups, but their contributions to modern maize breeding have not been systematically investigated.

**Results** Here we report co-selection and convergent evolution between nuclear and cytoplasmic genomes by analyzing whole genome sequencing data of 630 maize accessions modern maize and its relatives, including 24 fully assembled mitochondrial and chloroplast genomes. We show that the NB cytotype is associated with the expansion of modern maize to North America, gradually replaces the fertile NA cytotype probably through unequal division, and predominates in over 90% of modern elite inbred lines. The mode of cytoplasmic evolution is increased nucleotypic diversity among the genes involved in photosynthesis and energy metabolism, which are driven by selection and domestication. Furthermore, genome-wide association study reveals correlation of cytoplasmic nucleotypic variation with key agronomic and reproductive traits accompanied with the diversification of the nuclear genomes.

**Conclusions** Our results indicate convergent evolution between cytoplasmic and nuclear genomes during maize domestication and breeding. These new insights into the important roles of mitochondrial and chloroplast genomes in maize domestication and improvement should help select elite inbred lines to improve yield stability and crop resilience of maize hybrids.

**Keywords** Maize, Hybrid, Domestication, Improvement, Cytoplasmic genomes, Convergent evolution

<sup>†</sup>Shuai Cao and Huanhuan Zhang contributed equally to this work.

\*Correspondence:

Z. Jeffrey Chen  
zjchen@austin.utexas.edu

<sup>1</sup> State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, 1 Weigang Road, Nanjing 210095, China

<sup>2</sup> Shanxi Key Laboratory of Minor Crops Germplasm Innovation and Molecular Breeding, Shanxi Agricultural University, Shanxi, Taiyuan 030031, China

<sup>3</sup> Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX 78712, USA

<sup>4</sup> Temasek Life Sciences Laboratory, 1 Research Link, National University of Singapore, Singapore 117604, Singapore

## Background

Corn or maize (*Zea mays* ssp. *mays*) is the highest yielding crop, accounting for over 30% of cereal production worldwide [1]. Besides serving as a source of food, feed, and fuel, maize is an experimental model for studying plant evolution and crop domestication [2, 3]. Maize was domesticated from the wild ancestor teosinte (*Z. mays* ssp. *parviglumis*) approximately 9000 years ago near the Balsas river [4] through dramatic changes in morphology and agronomic traits, which transform from long branches, multiple small ears, and fruitcase over grain in teosinte to very few branches with a single ear on each



branch and many grains with exposed fruitcase in modern maize [5, 6].

This domestication process may involve one [7] or two teosintes [8] through introgression of several regulatory genes [9]. *teosinte branched1* (*tb1*) encodes a transcription factor of TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL NUCLEAR ANTIGEN FACTOR (TCP) family, whose expression is regulated by an upstream transposable element (TE) in modern maize [10], and can inhibit the outgrowth of axillary buds to control apical dominance [11]. *Teosinte glume architecture1* (*tga1*), encoding a squamosa-promoter binding protein (SBP) transcription factor, is another maize domestication gene. *tga1* affects the binding activities of its protein TGA1, leading to the shift from kernels encased in a hardened fruitcase in teosinte to naked kernels exposed on the ear in modern maize [12]. Domestication can also enhance maize adaptation to higher latitudes through the alteration of some genes involved in photoperiodic flowering, such as *ZmCCT9*, encoding a CCT (CONSTANS, CO-like, and TOC1) domain transcription factor, which regulates photoperiodic sensitivity and flowering [13, 14], and *ZEA CENTRORADIALIS8* (*ZCN8*), the maize florigen gene, which mediates flowering and adaptation [15]. In addition to gain or enhanced traits of adaptation and seed quality, some agronomic traits are decreased or lost during domestication. For example, *TEOSINTE HIGH PROTEIN9* (*THP9*) encodes an asparagine synthetase 4 enzyme that is highly expressed in teosinte, but not in the B73 inbred line. The low expression of *THP9* is associated with a deletion of the tenth intron and decreased seed protein content in modern maize [16].

Maize is a genetically diverse crop with a rapidly evolving nuclear genome, probably due to an abundance of transposable elements (TEs) [17, 18] and intensive domestication and breeding for adaptation to a wide range of environments [19, 20]. Re-sequencing analyses of the maize nuclear genome [21–24] have confirmed the origin of modern maize from teosinte and revealed strong selection during domestication and improvement. The application of modern breeding technology has further improved modern maize lines, including several generations and geographical locations of inbred lines used in commercial hybrid production [13, 25, 26].

However, the impact of cytoplasmic genomes on maize evolution and domestication is poorly understood [27–29]. Cytoplasmic genomes are important to growth and development of plants and animals. Mitochondria and chloroplasts are semi-autonomous organelles with their own genomes, providing cell metabolism and photosynthesis in flowering plants. Mitochondrial and chloroplast genomes, also known as mitogenomes and plastomes,

respectively, are commonly used as molecular markers for evolutionary studies because their rate of sequence changes is rather constant [30], although the evolution rate of mitochondrial genomes can be 10 times faster than the nuclear genome in animals [31]. Moreover, many common diseases including those of paternal origins are associated with mutations in their mitochondrial genomes [32, 33].

Consistent with the mitochondrial role in animal health, mutations in mitogenome can lead to cytoplasmic nuclear male-sterility (CMS) in plants [34], including maize, many of which rapidly evolve through sequence rearrangements and exchanges [35, 36]. In maize, there are three major CMS types: CMS-C, CMS-T, and CMS-S, which are classified according to the fertility by their respective nuclear restorer-of-fertility (*Rf*) genes [37]. The CMS-C type is related to three chimeric mitochondrial open reading frames (orfs) of *atp9-C*, *atp6-C*, and *coxII-C* [38], and its fertility can be fully restored by the nuclear gene *rf4* [39]. CMS-T is caused by a mutation of *T-urf13* through rearrangement of the mitogenome in the Texas cytoplasm and susceptible to Southern corn leaf blight [40], and the nuclear gene *rf2*, encoding a putative aldehyde dehydrogenase, can restore the pollen fertility in the CMS-T lines [41]. The CMS-S is associated with co-transcription of *orf355/orf77* due to rearrangements [42], and *rf3* is a dominant fertility restorer gene for CMS-S [43]. In addition to these CMS types, modern maize accessions are distinguished by the male-fertile cytotypes of NA and NB, which are subject to 16 rearrangements in the mitogenome [36]. The fertile NB type is present in most commercial hybrids [36], while the other fertile NA type, originally identified in the A188 inbred line, is found in most cultivated maize lines in North America [44]. Interestingly, sequence analysis of organellar genomes from a ~5000-year-old archaeological maize sample is related to the fertile NB cytype, suggesting that CMS lines (CMS-S, CMS-T, and CMS-C) evolve after the NB cytoplasm [27], which is different from the conclusion of sequencing eight mitochondrial genomes [45]. This may suggest that the role of mitochondrial and possibly chloroplast genomes during maize domestication has yet to be elucidated.

Here, we compared high-quality whole genome sequencing (WGS) data of 630 maize and relative accessions from China and North America. To investigate the cytoplasmic genome evolution, we analyzed WGS data from representative maize populations from North America, including 61 inbred lines, 23 landraces, and 19 maize relatives from diverse ecogeographic regions in the US and Mexico. We further explored the association of cytoplasmic genome variation during domestication and improvement with key agronomic traits and found

candidate loci for breeding applications. Collectively, these data provide new insights into the evolutionary history of cytoplasmic genomes during maize domestication and will help further improve maize crop yield and resilience.

## Results

### Evolution of mitochondrial genomes driven by domestication and improvement

To investigate the variation of mitochondrial (Mt) genomes or mitogenomes during maize domestication and improvement, we analyzed 24 assembled cytoplasmic genomes (Additional 1: Table S1), including 24 fully assembled genomes, from a representative population of *Zea* genus, consisting of 61 maize elite inbred lines, 23 landraces, and 19 wild relatives from diverse eco-geographic regions in the US, Mexico, and China (Additional 1: Table S2). After quality control, 1249 high-quality SNPs of the mitogenomes were identified to construct the neighbor-joining phylogenetic tree (Additional 1: Table S3). Two mitogenomes of sorghum (black branches) were used as an outgroup. The clade next to sorghum consisted of *Tripsacum dactyloides*, *Zea perennis*, *Zea luxurians*, *Zea mays* subsp. *mexicana* (hereafter, *Z. mexicana*), *Zea mays* subsp. *parviglumis* (*Z. parviglumis*), and *Zea mays* subsp. *mays* (*Z. mays*) (Fig. 1A); the latter included landraces and improved inbred lines. Notably, two mitogenomes from *mexicana* were clustered in a subgroup of *parviglumis*, which indicates its origin from *parviglumis*. Alternatively, there is possible introgression of the latter into the former through naturally outcrossing or human intervention. Interestingly, ~69% (16/23) of landraces and ~26% (16/61) improved inbred lines were also clustered in the *parviglumis* subgroup. This result suggests that the mitogenome of *parviglumis* has largely been retained in the landraces but gradually lost among improved inbred lines probably via modern breeding.

Maize mitogenomes can be classified into fertile (NA and NB) and cytoplasmic-nuclear male-sterility (CMS-S, CMS-C, and CMS-T) cytotypes [29]. The mitogenome of an archaeological maize SM10 resembles the NB cytype [27], indicating that maize NB cytype appeared prior to the domestication of landraces. SM10 is estimated to have existed ~5000 years ago, and landraces began to spread into the temperate zone of the US ~4000 years ago [21]. This is supported by the phylogenetic tree (Fig. 1A), as SM10 is close to the older NA cytype, compared to most of the modern maize. Most landraces, including maize NA cytype (A188) and CMS, were clustered in the *parviglumis*, whereas most modern inbred lines belonged to the NB cytype. Moreover, landraces were distributed across all branches but

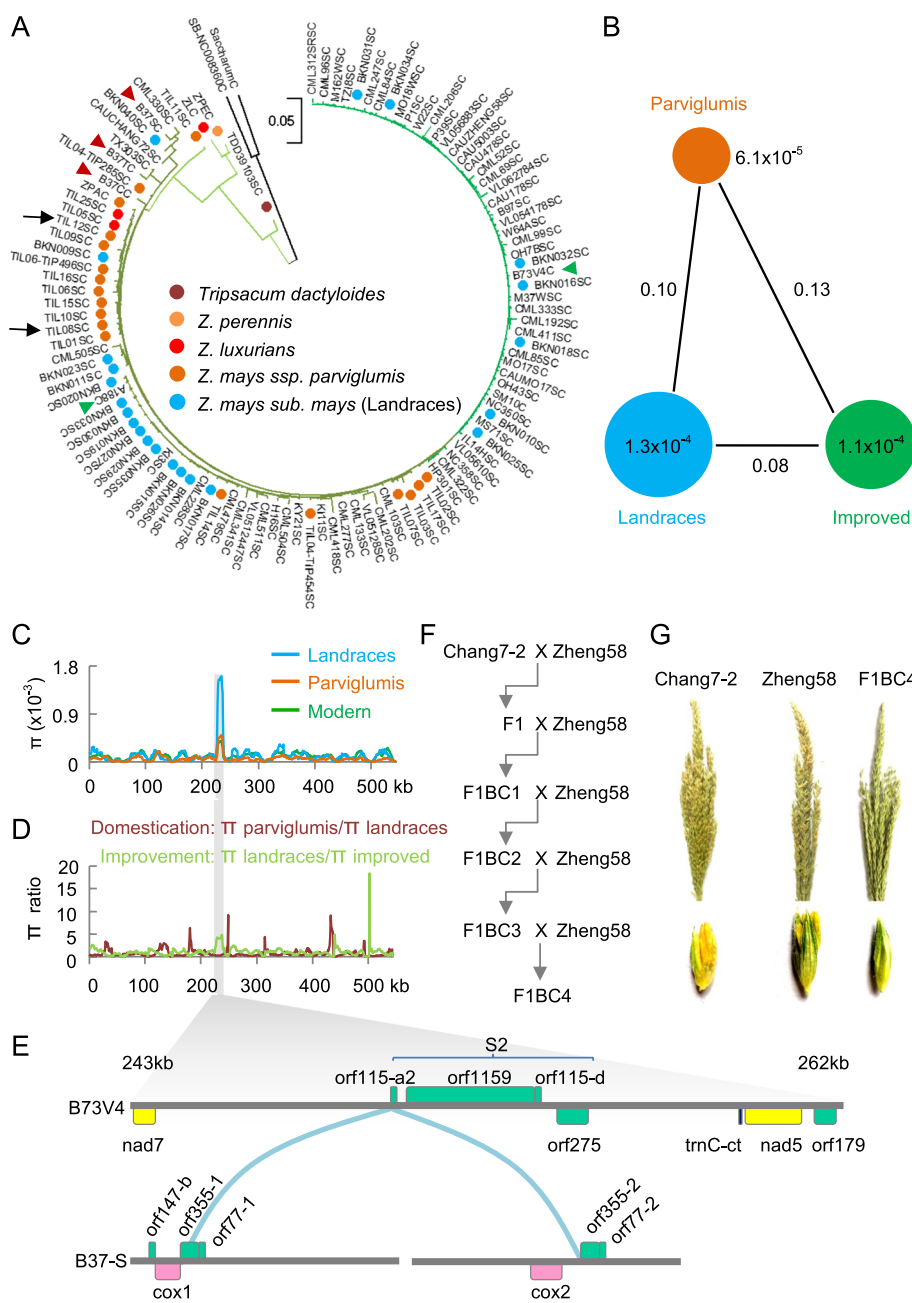
more concentrated in the NA and CMS clusters. These results suggest that landraces were domesticated from *parviglumis* and used for the improvement of modern maize inbred lines [11, 19]. The above conclusion was also supported by the phylogenetic tree constructed from mitochondrial genomes of 175 elite inbred lines in China (Additional 2: Figure S1). The data collectively suggest that the NB cytype of mitogenomes has been selected during modern maize breeding and spread in the temperate zone of the US [21].

### Nucleotypic evolution in mitogenomes

Among diverse cytotypes, analysis of population fixation index ( $F_{ST}$ ) showed the lowest  $F_{ST}$  value (0.08) between landraces and improved populations, followed by 0.10 between *parviglumis* and improved population, and 0.13 between *parviglumis* and landraces (Fig. 1B). The data suggest modern breeding has a strong effect on mitogenomes between improved populations and landraces and can shorten the time for population improvement during modern breeding [46]. At the nucleotide variation ( $\pi$ ) level, mitogenomes showed greater variation in the landraces ( $1.3 \times 10^{-4}$ ) and improved lines ( $1.1 \times 10^{-4}$ ) than among *parviglumis* accessions ( $6.5 \times 10^{-5}$ ) (Fig. 1B). This suggests accelerated variation of mitogenomes during domestication and breeding. The highest level of nucleotide variation ( $\pi$ ) was localized in a region (243–262 kb) among three populations, especially in landraces (Fig. 1C). This region, along with another region (near 500 kb) (Fig. 1D), may represent a selection sweep during modern maize improvement.

The selection sweep region consisted of *tmC-ct* and eight orfs, including *nad7*, *orf115-a2*, *orf1159*, *orf115-d*, *orf275*, *nad5*, and *orf179* (Fig. 1E). Four orfs, including *orf115-a2*, *orf1159*, *orf115-d*, and *orf275*, were derived from the S2 linear plasmid, which were uniquely related to CMS-S [36]. *orf115-a2* in the maize NB cytype was collinear to a region spanning *orf335-1/2* in the fully assembled mitogenome of the CMS-S cytype [36] (Fig. 1E) and another CMS-S cytype (Additional 2: Figure S2), which is known as the CMS-S cytype [42]. Landraces contained all cytoplasmic cytotypes, including *parviglumis*, CMS (CMS-T, CMS-C, and CMS-S), and fertile NA and NB cytotypes. These results suggest the evolution and rearrangement of mitogenomes between CMS-S and fertile NB cytotypes from domestication of landraces to breeding selection of the favorable NB cytype in modern maize.

For example, a common commercial hybrid ZD958 in China was produced by crossing Zheng58 (a maternal parent) with pollen from Chang7-2 (CMS-S cytype) [47]. In the reciprocal backcrossing scheme using



**Fig. 1** Contribution of the mitogenomes to maize domestication and improvement. **A** Neighbor-joining (NJ) phylogenetic tree reconstructed using SNPs of mitogenomes from different species of *Zea* using *Sorghum bicolor* as an outgroup. Two *Z. parviglumis* accessions (black arrows) were clustered in a clade with CMS cytotypes. Also indicated are representative CMS-S (B37S-KP966116), CMS-T (B37T-KP966117), and CMS-C (B37C-KP966115) cytotypes and fertility NA type (A188-KF241980) and NB type (B73V4) (green triangles) cytotypes. Branch length scale = 0.05. **B** Genetic diversity based on SNPs of mitogenomes among improved lines, landraces, and *Z. parviglumis*. **C** Nucleotide diversity ( $\pi$ ) of mitogenomes among improved lines, landraces, and *Z. parviglumis*. **D** Candidate regions of mitogenomes under domestication (brown) and improvement (light green). Dashed line indicates the ratio value of 1. **E** Genomic region of high-nucleotide diversity showing Mt groups of B73V4 (NB) and B37S (CMS-S). **F** Backcross scheme of F1(Chang7-2 x Zheng58) to the recurrent parent Zheng58. **G** Tassel (upper panel) and floret (lower panel) images showing male sterility occurred in the F1BC4

Chang7-2 as a maternal parent and Zheng58 as a paternal parent (Fig. 1F), male sterility was shown clearly in the F1BC4 progeny (Fig. 1G). The data also suggest that the

nuclear fertility restorer (*Rf*) gene in Zheng58 is either absent or non-functional to restore the CMS-S fertility in Chang7-2.

### Evolution of chloroplast genomes during maize domestication and improvement

From chloroplast genomes or plastomes of the same set of genetic materials, we identified 2925 sequence variants, including 2327 SNPs (79.6%) and 598 Indels (20.4%) (Additional 1: Table S4). The phylogenetic neighbor-joining tree revealed that all *parviglumis* and most landraces and maize NA (A188) were clustered in the *parviglumis* clade with one exception (Fig. 2A). One *parviglumis* (accession TIL02) and CMS cytotypes were clustered in another clade, suggesting this *parviglumis* accession may contribute to the CMS phenotype. The same trend was observed among elite inbred lines from China, where improved lines and CMS cytotypes were shown in a separate group (Additional 2: Figure S3). The data indicates that plastomes and mitogenomes are co-selected during maize breeding improvement.

At the population level, landraces and improved populations had the lowest  $F_{ST}$  value (0.07), compared with *parviglumis* and landraces (0.18) and improved populations (0.25), respectively (Fig. 2B). This suggests a close genetic relationship between landraces and improved maize lines, consistent with the conclusion from mitogenomes. The nucleotide variation ( $\pi$ ) was the highest in the improved populations ( $2.3 \times 10^{-4}$ ) (Fig. 2B), indicating a strong human selection that drives plastome evolution. The nucleotide variation ( $\pi$ ) in the 20–80-kb region was higher in the landraces than that in *parviglumis* (Fig. 2C and D), while the nucleotide variation in the 20–45-kb region was increased in the improved populations compared with landraces. These regions encode genes and components related to photosynthetic machinery, including photosystem I and II, ATP synthase, rubisco large subunit, ribosomal proteins (SSU and LSU), and ribosomal RNA (Fig. 2E). This result suggests pervasive changes in the chloroplast genome from *parviglumis* in Mexico to increase adaptation to diverse light and temperate environments in the US during modern maize breeding.

### Contribution of cytoplasmic genomes to modern maize breeding

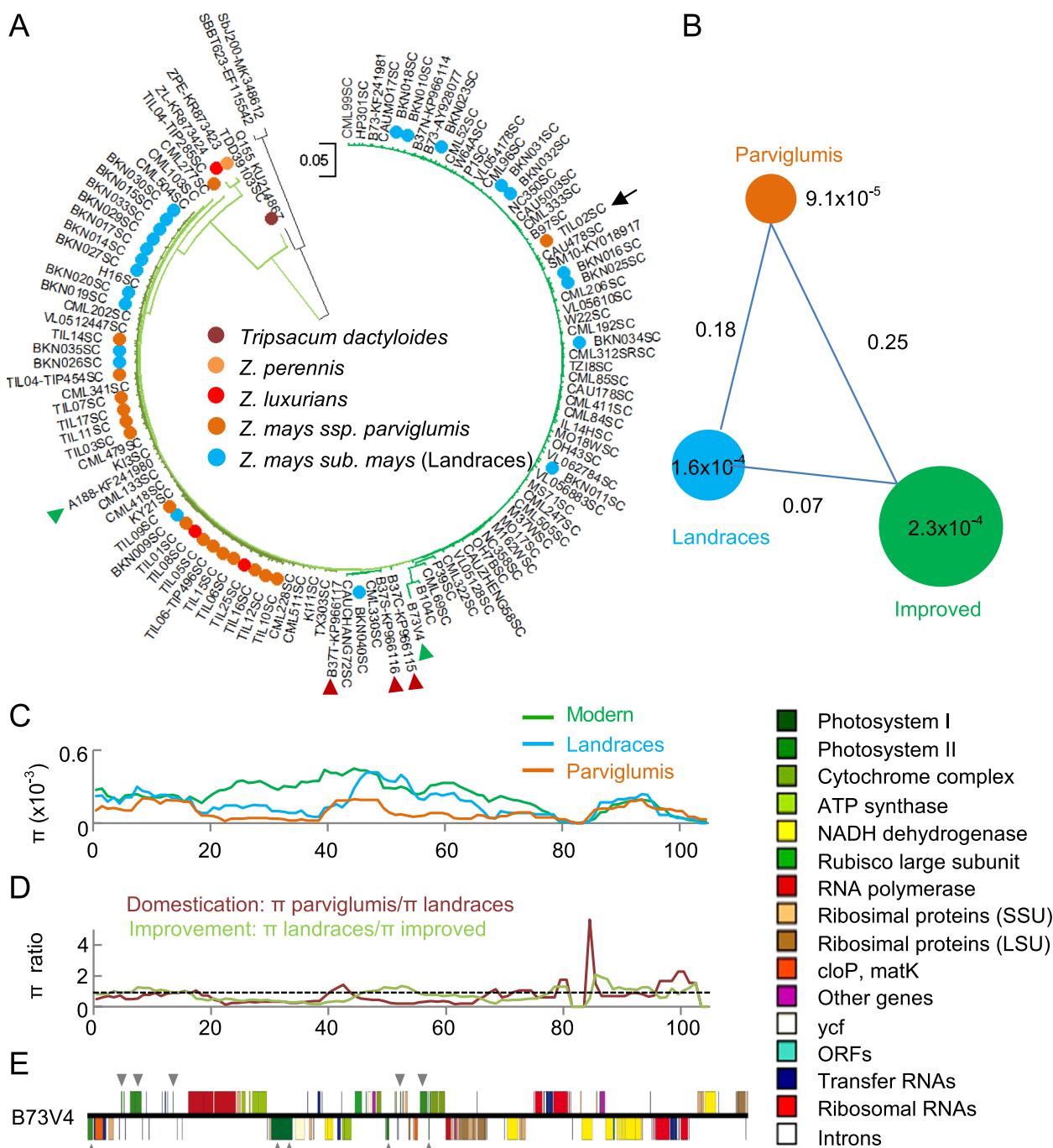
Modern breeding has greatly improved corn yield, especially through successful utilization of single-cross hybrids [48, 49]. Although abundant variation of nuclear genomes has been associated with selection during modern breeding, it cannot fully explain many domestication traits involved in energy production, storage, and yield [46]. To explore a role of cytoplasmic genome in modern breeding, we performed a comprehensive analysis of cytoplasmic genomes using a chronological sampling of 343 elite inbred lines (improved populations) from China and North America. The analysis recorded a total of

11,102 variants including 5668 SNPs and 5431 Indels in mitogenome and 1879 variants including 338 SNPs and 1541 Indels in plastome. After quality control, 2440 SNPs of the mitogenome were used to classify these inbred lines into two groups, including a maize NB group and another group with maize NA cytotype and CMS cytotypes (Fig. 3A and Additional 1: Table S5). These inbred lines from different breeding stages were distributed over all branches of the phylogenetic tree (Fig. 3A), different from the nuclear genome variation [46].

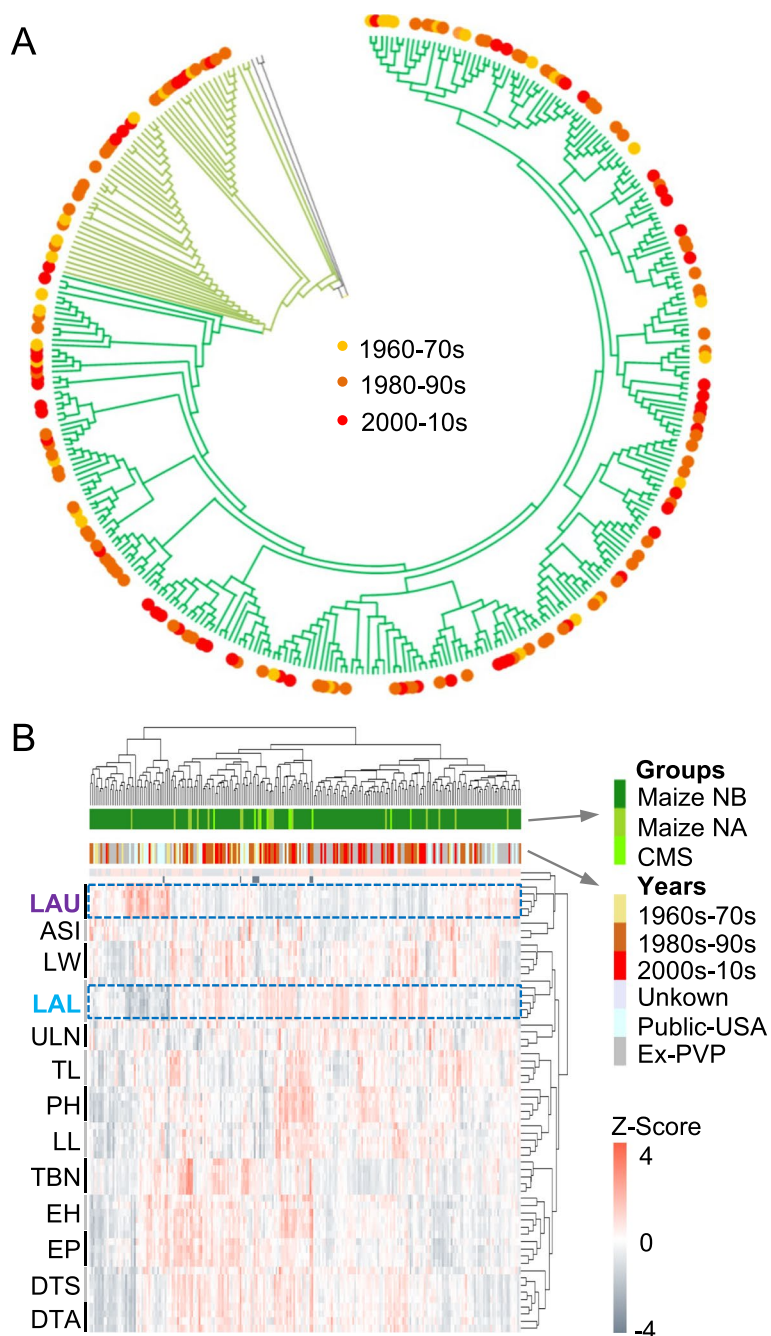
To explore the ideotype of mitogenome for modern inbreeding, we further analyzed 64 individual traits of 15 agronomic trait groups among different locations and years [46] (Fig. 3B). While most traits from the same agronomic group were clustered together, the relationship between upper leaf angle (LAU) and lower leaf angle (LAL) is antagonistic (Fig. 3B), in agreement with the ideotype of modern breeding for large LAL and small LAU. Similar discordance occurred between one subgroup of LAL and anthesis to silking interval (ASI) (Additional 2: Figure S4) and another subgroup of other agronomic traits. Some other traits such as relative ear position (EP) (Additional 2: Figure S5), days to silking (DTS) (Additional 2: Figure S6), days to anthesis (DTA), tassel branch number (TBN), and ear height (EH) were positively correlated. Notably, NA cytotype and CMS cytotypes of modern inbred lines had large LAL and small LAU (Fig. 3B), suggesting the contribution of mitogenomes to the ideotype of maize inbred lines.

To further dissect which variation of mitogenome is related to these traits, we next performed GWAS for all 64 traits using mixed linear model-based association analysis (GCTA-MLMA) [50]. After removing variants with >40% missing calls and minor allele frequency (MAF) of <0.005, we identified 548 high-quality SNPs in the mitogenome and 45 SNPs in the plastome (Additional 1: Table S6). Significance thresholds for trait association were set at  $P$ -values of  $9.124 \times 10^{-5}$  (0.05/548) and  $1.086 \times 10^{-3}$  (0.05/45) for mitogenome and plastomes, respectively. We found three traits, including LAL (Additional 2: Figure S7), TBN (Additional 2: Figure S8), and cob color, stem diameter, and kernel color (Additional 2: Figure S9), were associated with mitogenome diversity at a significant level ( $P < 4.386 \times 10^{-5}$ ), while plastome SNPs were not associated with these traits (Additional 2: Figure S10).

The mitogenome SNPs located at the position 184,809, is related to genes of *NADH dehydrogenase subunit 1* (*nad1*), *NADH dehydrogenase subunit 2* (*nad2*), and *tRNA-Ala*, and associated with LAL\_BLUP ( $P < 2.09 \times 10^{-05}$ ) (Fig. 4A), LAL\_LF2016 ( $P < 6.31 \times 10^{-05}$ ) (Fig. 4B), LAL\_JL2017 ( $P < 1.55 \times 10^{-04}$ ) (Fig. 4C), and LAL\_LF2017 ( $P < 4.24 \times 10^{-05}$ ) (Fig. 4D). There were four major haplotypes, including TT (269 lines), GG



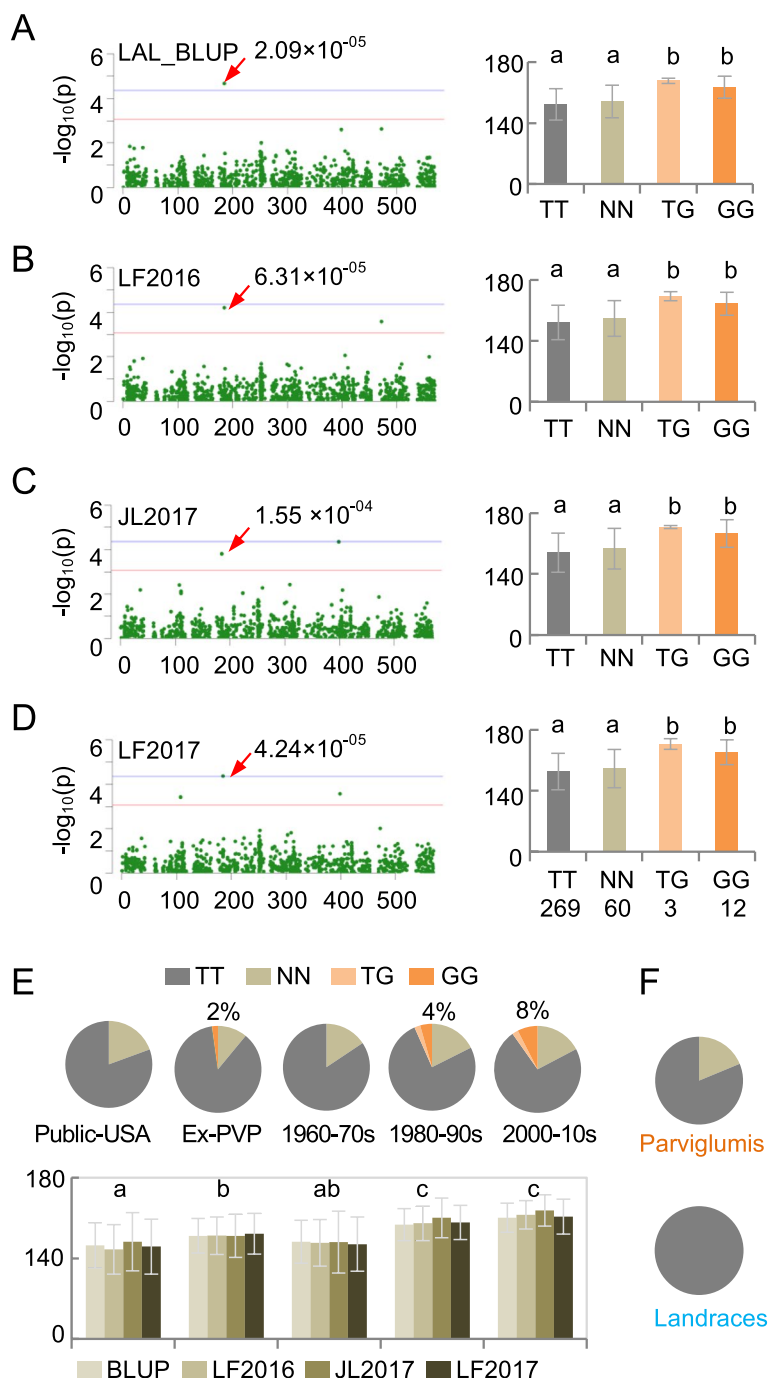
**Fig. 2** Contribution of plastomes to maize domestication and improvement. **A** Neighbor-joining (NJ) phylogenetic tree reconstructed using SNPs of plastomes between different species of *Zea* using *Sorghum bicolor* as an outgroup. Two *Z. parviglumis* accessions (black arrows) were clustered in a clade with CMS cytotypes. Also indicated are representative CMS-S (B375C), CMS-T (B377C), and CMS-C (B377C) (red triangles) cytotypes and fertility NA type (A188C) and NB type (B73V4C) (green triangles) cytotypes. Branch length scale=0.05. **B** Genetic diversity based on SNPs of plastomes among improved lines, landraces, and *Z. parviglumis*. **C** Nucleotide diversity ( $\pi$ ) of plastomes among improved lines, landraces, and *Z. parviglumis*. **D** Candidate regions of Cp genome under domestication (brown) and improvement (light green). Dashed line indicates the ratio value of 1. **E** Maize plastome of the NB cytotypic (Genebank: KP966114) showing the location of photosystem I and II (indicated by triangles) gene clusters (shown on the right with colors)



**Fig. 3** Association of mitogenome diversity with maize breeding lines. **A** Neighbor-joining (NJ) phylogenetic tree reconstructed using SNPs of the mitogenomes between breeding lines in different stages, 1960–1970s (yellow), 1980–1990s (light orange), and 2000–2010s (red). **B** Hierarchical clustering of principal components using 64 agronomic traits among breeding lines in different stages. Abbreviations are LAU (upper leaf angle), ASI (anthesis to silking interval), LW (leaf width), LAL (lower leaf angle), ULN (upper leaf number), TL (tassel length), TBN (tassel branch number), LL (leaf length), PH (plant height), EH (ear height), EP (relative ear position), DTS (days to silking), and DTA (days to anthesis). LAU and LAL are boxed to indicate their antagonistic relationship

(12 lines), GT (3 lines), and NN (60 lines,  $N$ =unknown base). The GT and GG haplotypes were related to increased LAL values, compared with haplotypes of TT and NN (Student's  $t$ -test,  $P < 0.05$ ). Interestingly,

increased frequencies of these haplotypes were coincident with the increased LAL trait during modern maize breeding in both the USA and China (Fig. 4E, Student's  $t$ -test,  $P < 0.05$ ).



**Fig. 4** Association of lower leaf angle with mitogenome diversity during maize breeding. **A–D** Manhattan plot (left) and mitogenome haplotypes (right) showing association (with a red arrow and  $P$  value) of lower leaf angle (LAL) of LAL\_BLUP (best linear unbiased predictor) (**A**), LAL\_LF2016 (**B**), LAL\_JL2017 (**C**), and LAL\_LF2017 (**D**) with the haplotypes TT, NN, TG, and GG, respectively, in the position of 184,809. The LAL were collected in the city of Langfang in Hebei province in 2016 (LF2016) and 2017 (LF2017), Ledong County in Hainan province in 2016 (HN2016), and Gongzhuling in Jilin province in 2017 (JL2017) in China. The arrow in (**A**) indicates the best linear unbiased predictor value for data of LAL phenotyped across four environments in 2 consecutive years. Different letters indicate statistical significance of  $P < 0.05$  (Student's  $t$ -test). **E** Proportion of mitogenome haplotypes at 184,809 (upper panel) with the lower leaf angle trait which is one of four listed in **A–D** (lower panel) among breeding lines in different stages and regions. Public-USA and Ex-PVP indicate public inbred lines in the US and elite commercial lines with expired Plant Variety Protection Act Certificates mostly released after 2003, respectively. **F** Proportion of mitogenome haplotypes in *Z. parviglumis* (upper panel) and landraces (lower panel)

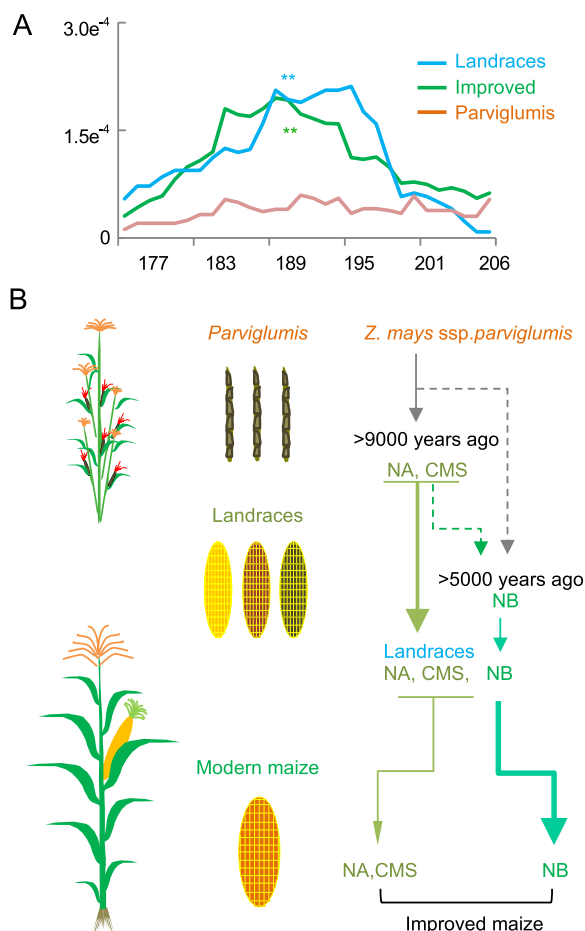


However, the GG haplotype was absent in either *parviglumis* or landraces (Fig. 4F), indicating that the GG haplotype could derive from a natural mutation, leading to a larger LAL during modern maize breeding. This notion is supported by the higher level of nucleotide diversity ( $\pi$ ) in the mitogenome region (177–206 kb) that spans the haplotype in the improved lines than in the landraces and *parviglumis* (Fig. 5A).

#### GWAS of 17 agronomic traits from another maize population

Different population may select different loci for shaping traits, and we analyzed another large population of modern maize. To investigate more beneficial alleles for improving breeding, we performed GWAS of cytoplasmic genomes on their trait data [24, 51]. We randomly selected 124 inbred lines with a high depth of re-sequencing data from published data [24] (Additional 2: Figure S11A) and identified a total of 13,420 primary variants from mitogenomes, including 12,024 SNPs (89.6%) and 1396 Indels (10.4%), and a total of 2793 primary variants from the plastomes, including 2194 SNPs (78.6%) and 599 Indels (21.4%). After removing variants with >40% missing calls and minor allele frequency (MAF) of <0.005, we identified 924 SNPs in the mitogenome and 144 SNPs in the plastome (Additional 1: Table S7). Based on the SNPs of the mitogenome, the population consisted of two cytoplasmic groups with the NB cytotypic being predominant one (Additional 2: Figure S11A). Hierarchical clustering of principal components revealed that the NB cytotypic cytoplasm was associated with a group of reproductive traits (vertical box), including HD (heading date), PS (pollen shed), ST (silking time), LNBE (leaf number above the ear), TBN (tassel branch number), and EH (ear height), which is in contrast to the values of these traits in most CMS accessions (horizontal box) (Additional 2: Figure S11B). This result indicates the role of the NB cytoplasm in the development of reproductive traits.

Using genome-wide complex trait analysis-mixed linear model-based association (GCTA-MLMA) [52], we identified the association of GWAS traits with cytoplasmic genomes, using the  $p$ -values of  $5.411 \times 10^{-5}$  (0.05/924) and  $8.197 \times 10^{-4}$  (0.05/61) for mitogenomes and plastomes, respectively. The trait of kernel number per row was related to two SNPs at mitogenome positions 109,061 ( $P < 4.14 \times 10^{-5}$ ) and 109,669 ( $P < 3.03 \times 10^{-5}$ ), respectively (Additional 2: Figure S11C). Both SNPs are located in the region encoding *nad1*, *nad2*, *tRNA-Ala*, and ribosomal protein S3 (*rps3*), while the kernel width was associated with two other mitogenome loci at 103,594 ( $P < 4.83 \times 10^{-6}$ ) and 445,983 ( $P < 5.01 \times 10^{-10}$ ), respectively (Additional 2: Figure S11D). The SNPs located at 103,594 are associated with the *nad1*, *nad2*,



**Fig. 5** A model for the origin of NA, CMS, and NB cytotypes in maize mitogenomes. **A** Nucleotide diversity ( $\pi$ , y-axis) of mitogenome (x-axis) in the region of 177–206 kb among improved lines, landraces, and *Z. parviglumis*. An asterisk indicates the change between landraces or improved lines and wild maize with statistical significance ( $P < 0.05$ , two-tailed Student's  $t$ -test). **B** Mitogenome of maize was derived from *Z. parviglumis* cytoplasm with the nuclear genomes through modern breeding programs produced mitogenome cytotypes including NA and CMS (S, T, and C). Approximately ~5000 years ago, NB cytotypic of the mitogenome was generated possibly through recombination of NA and CMS (S, T, and C) cytotypes. The combination of NA and CMS cytotypes has been retained through artificial selection, while the NB cytotypic became predominant in the commercial inbred lines associated with maize improvement

*tRNA-Ala* region, and the SNPs located at 445,983 are related to the *nad1* and *nad2* region. In the plastome, the ear leaf width (ELW) was associated with a SNP locus at 52,460 ( $P < 3.44 \times 10^{-4}$ ), which is located at the ~300 bp upstream of *NADH dehydrogenase D3* (*ndhC*) (Additional 2: Figure S11E). These data from two large populations collectively suggest co-evolution between cytoplasmic and nuclear genomes during maize domestication and improvement.

## Discussion

Cytoplasmic genomes (plasmon) of mitogenomes and plastomes with origins of bacterial endosymbionts are essential for life activities of carbon cycle and energy metabolism, including photosynthesis, cellular respiration, and ATP synthesis in flowering plants. In contrast to the nuclear genome, cytoplasmic genomes evolve rapidly through rearrangement to create mutations related to many common diseases of maternal inheritance in mammals [32, 33] and cytoplasmic nuclear male-sterility (CMS) in plants [34].

Maize, the highest-yielding crop, was domesticated from tropical wild ancestor teosinte (*Z. parviglumis*) approximately 9000 years ago near the Balsas river, spreading from the temperate zone to more than the 50th parallel north. This requires wide adaptation to critical limiting factors of light and temperature through adjustment of photosynthesis and energy metabolism during domestication and breeding improvement. We found that 90% or more modern elite inbred lines possess the fertile NB cytotype. Moreover, mitogenomes of *Z. parviglumis* populations were clustered in the subgroups of fertile NA and CMS, suggesting that the fertile NA and CMS cytotypes were generated before landraces of maize were domesticated from *Z. parviglumis* in Mexico [21, 22]. However, the maize NB cytotype has not been found in *Z. parviglumis* populations, which suggests that the NB cytotype has been distinctively produced during modern maize breeding (Fig. 1A). Analysis of ancient DNA from a >5000-year-old maize (SM10) [27] found the fertile NB cytotype of mitogenome, suggesting that the fertile NB cytotype was formed before the maize was brought to the southwestern USA [26]. These data let us to propose the origin of the NB cytotype between 5000 and 9000 years during early domestication (Fig. 5B). In contrast, CMS and fertile NA cytotypes could be formed before maize domestication along with the *Z. parviglumis* population. According to integration analysis of the above archaeological evidence and genetics, The NB cytotype may be derived directly from *Z. parviglumis* or through possible rearrangement(s) of the NA and CMS (S, T, and C) cytotypes.

The fertile NB cytotype of mitogenomes is popular among maize landraces, which has been selected during maize breeding as a predominant (>90%) group in modern elite inbred lines and nearly all of commercial hybrids [36]. The discovery and application CMS-T in modern hybrid breeding was a success but did not last long [40, 53], partly because these CMS (T, C, and S) cytotypes such as CMS-T that is highly susceptible to bacterial blight infection [40]. Furthermore, restoration lines for the CMS-C cytotype are uncommon, and the male sterility of CMS-S cytotype is unstable [54].

Our results provide genomic evidence for the origin of three CMS cytotypes from *Z. parviglumis*, but not for selection during domestication of modern maize. The introduction of natural variation among *parviglumis* populations can break this bottleneck of narrow genetic background among CMS cytotypes.

There are some exceptions. For example, the plastome of one *Z. parviglumis* accession (TIL02) was clustered in the maize group, despite its mitogenome being relatively distant from the fertile NB group (Figs. 1A and 2A). This is probably because *Z. parviglumis* and maize are naturally outcrossing during the early stages of domestication [55], such that *Z. parviglumis* gene flow has contributed to local adaptation during maize improvement [56]. The plastome inheritance may be retained from unequal division of the cytoplasm, independent of mitogenome inheritance [32]. The unequal division and independent inheritance may also lead to the incongruent evolution between nuclear and cytoplasmic genomes. Notably, the increased nucleotypic diversity of cytoplasmic genomes during maize domestication and improvement is contrary to the evolution features of nuclear genomes (Hufford et al., 2012; Wang et al., 2020). This may indicate the adaptive evolution of mitogenomes in mammals [57], as well as positive or relaxed selection, resulting from increased patterns of nucleotide substitutions as reported in the plastomes of Geraniaceae [58]. The increased nucleotypic variation in the photosynthetic system of the plastome may help adapt rapidly to light and temperature conditions during domestication and improvement (Kistler et al., 2018). Moreover, different cytotypes could have been selected for local adaptation through outcrossing and/or introgression of cytoplasmic genomes by selection and breeding. Similarly, large population of agricultural production and extensive breeding for modern maize could have increased the nucleotide diversity of cytoplasmic genomes compared to nuclear genomes. The association of nucleotypic variation in the mitogenomes and plastomes with different traits in two different populations (Fig. 4 and Additional 2: Figure S10) may also indicate a wide range of effects of the cytoplasmic genomes on photosynthesis, energy metabolism, ideal architecture, and reproductive traits, which play an important role in modern maize breeding and improvement. For example, the association of mitogenomic variation with the lower leaf angles (LAL) suggests the gradual accumulation of this trait through direct or indirect selection during modern maize breeding. These new insights learned from the evolution of cytoplasmic genomes should facilitate and accelerate biological breeding applications in maize production.

## Conclusions

After selection and domestication from one or two tropical wild ancestor teosintes (*Z. parviglumis*), maize cultivation has spread from the temperate zone to more than the 50th parallel north. This wide distribution is accompanied by improved adaptation to critical limiting factors such as light and temperature largely through selection and breeding to enhance photosynthesis and energy metabolism. The genomic basis for maize domestication and breeding improvement depends on convergent evolution between cytoplasmic and nuclear genomes. Using whole genome sequencing data of maize populations and phylogenetic analysis of mitochondrial and chloroplast genomes, we found that the fertile NB cytotype of the mitochondrial genome is associated with the expansion of maize to North America, which gradually replaces the fertile NA cytotype and predominates in more than 90% of modern elite inbred lines. Moreover, population genetics studies have uncovered important roles of mitochondrial and chloroplast genomes during modern maize breeding. Both mitogenomes and plastomes have increased levels of nucleotypic diversity among the genes involved in photosynthesis and energy metabolism, and the cytoplasmic nucleotypic variation in those genes is associated with key agronomic and reproductive traits under diversification and selection of nuclear genomes. These new insights into the evolutionary history of cytoplasmic and nuclear genomes during maize domestication and improvement could help select elite inbred lines to improve yield stability and crop resilience of maize hybrids.

## Methods

### Plant materials

Each year, two inbred lines, Chang7-2 and Zheng58, were grown in Taiyuan in summer and Sanya in winter. The Chang7-2 was backcrossed with Zheng58 pollen in summer and winter, respectively. Cytoplasmic nuclear male-sterility lines were identified from backcrossed offspring in each generation.

### Acquisition of WGS data and alignments

We collected a total of 630 whole-genome resequenced accessions of maize and relatives, including 102 accessions from SRA051245 [59], 51 accessions randomly selected from SRA049859 [47], 350 accessions from the PRJNA609577 [46], 124 accessions randomly selected from PRJNA531553 [24], and 3 accessions from PRJNA528290 [55]. The single- or paired-end resequencing reads were downloaded from NCBI Short Reads Archive by the fastq-dump program using the SRA-toolkit (version 2.10.5, <https://ftp-trace.ncbi.nlm.nih.gov/sra/sdk/2.10.5/>). All reads were filtered into clean data using NGSQCtoolkit v2.3 (<http://www.nipgr.res.in/ngsqctoolkit.html>). All cleaned reads were mapped to the cytoplasmic genome sequence of B37NB (KP966114 and AY506529) using BWA (v0.7.15) with four mismatches allowed per read. Only uniquely mapped paired reads were extracted to create new bam files using Perl scripts. Cleaned reads were used for further analysis, after potential PCR duplicates were removed using the “Samtools rmdup” of the Samtools program (version 1.3.1).

gov/sra/sdk/2.10.5/). All reads were filtered into clean data using NGSQCtoolkit v2.3 (<http://www.nipgr.res.in/ngsqctoolkit.html>). All cleaned reads were mapped to the cytoplasmic genome sequence of B37NB (KP966114 and AY506529) using BWA (v0.7.15) with four mismatches allowed per read. Only uniquely mapped paired reads were extracted to create new bam files using Perl scripts. Cleaned reads were used for further analysis, after potential PCR duplicates were removed using the “Samtools rmdup” of the Samtools program (version 1.3.1).

### Alignment of assembled cytoplasmic genomes

Assembled cytoplasmic genomes were downloaded from the genebank of NCBI (see Additional 1: Table S1 for details). The sequence of assembled cytoplasmic genomes was split into 200-bp short fragments with steps of 50-bp using Perl scripts. These short fragments are imitated as reads of WGS data to produce fastq files using Perl scripts. The reads were mapped to the cytoplasmic genome sequence of B73 using BWA (v0.7.15) with four mismatches allowed per read, and only uniquely mapped paired reads were extracted for further analysis.

### Detection of genetic variation

The Genome Analysis Toolkit (GATK, version 3.5.0) and Picard tools (version 2.0.1) were applied to detect sequence variants. “CombineGVCFs” in GATK was applied to merge raw variant calls. The SNPs and Indels were separately classified using “SelectVariants” in GATK. To reduce the variant discovery rate, the SNP calls were filtered according to the following threshold:  $QD < 10.0$  ||  $MQ < 20.0$  ||  $FS > 30.0$  ||  $SOR > 3.0$  ||  $MQRankSum < -2.5$  ||  $ReadPosRankSum < -3.5$  ||  $DP < 5$ . Retained SNPs were used in further analyses.

### Phylogenetic analysis

SNPs from the mitogenomes and plastomes among maize and its relatives were used to construct a neighbor-joining tree using MEGAX with 1000 replicates for bootstrap and confidence analyses.

### Population parameter estimation

The cytoplasmic genome was scanned in 1-kb window size, and population parameters ( $\pi$  and  $F_{ST}$ ) were estimated for each window by VCFtools (<https://github.com/vcftools/vcftools>). Nucleotide diversity ( $\pi$ ) was measured with the defined parameters “-window-pi 1,000 -window-pi-step 100.” The average 1-kb window size  $\pi$  value was estimated as the genetic diversity. For measurement of population differentiation,  $F_{ST}$  was calculated using the setting “-fst-window-size 1,000 -fst-window-step 100.”

### Genome-wide association analyses (GWAS)

For the population of 343 accessions with 15 traits [46], GWAS was performed using 1140 high-quality SNPs in mitogenomes and 59 high-quality SNPs in plastome genomes, after removing variants with >40% missing calls and minor allele frequency (MAF) < 0.005. To control spurious associations, genetic relatedness was estimated using “GCTA-GRM” [52]. GWAS was performed in a mixed linear model-based association analysis (MLMA) in the Genome-wide Complex Trait Analysis (GCTA) program, using effective numbers of 548 and 45 independent SNPs for mitogenomes and plastomes, respectively, and their corresponding *P*-values of  $9.124 \times 10^{-5}$  (0.05/548) and  $1.086 \times 10^{-3}$  (0.05/45). Significant association signals were identified using the thresholds set by GCTA-MLMA [52].

For the population of 73 accessions associated with 17 traits [51], GWAS was performed using a total of 924 high-quality SNPs in mitogenomes and 61 high-quality SNPs in plastomes, after removing variants with >40% missing calls and minor allele frequency (MAF) < 0.005. Genetic relatedness was estimated using “GCTA-GRM.” GCTA was performed in “MLMA” [52], using an effective number of independent SNPs of 924 and 61 for mitogenomes and plastomes, respectively, and their corresponding *P*-values of  $5.411 \times 10^{-05}$  and  $8.197 \times 10^{-04}$ . Significant association signals were identified using the thresholds set by GCTA-MLMA [52].

### Abbreviations

CMS	Cytoplasmic-nuclear male-sterility
TCP	TEOSINTE BRANCHED1/CYCLOIDEA/PROLIFERATING CELL NUCLEAR ANTIGEN FACTOR
tb1	Teosinte branched 1
tga1	Teosinte glume architecture 1
TE	Transposable element
CCT	CONSTANS, CO-like, and TOC1
ZCN8	ZEA CENTRORADIALIS8
THP9	TEOSINTE HIGH PROTEIN9
Rf	Restorer-of-fertility
orf	Open reading frame
mt	Mitochondrion or mitochondrial
cp	Chloroplast
SNP	Single nucleotide polymorphism
MAF	Minor allele frequency
GWAS	Genome-wide association study
GCTA-MLMA	Genome-wide complex trait analysis-mixed linear model-based association
WGS	Whole genome sequencing
LAU	Upper leaf angle
LAL	Lower leaf angle
ASI	Anthesis to silking interval
ELW	Ear leaf width
EP	Ear position
DTS	Days to silking
DTA	Days to anthesis
TBN	Tassel branch number
EH	Ear height

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-024-01859-4>.

**Additional file 1: Table S1.** List of assembled maize cytoplasmic genomes. **Table S2.** Whole genome resequencing of 630 maize accessions. **Table S3.** High-quality (HQ) SNPs of mitogenomes for evolution analysis. **Table S4.** High-quality (HQ) SNPs of plastomes for evolution analysis. **Table S5.** High-quality (HQ) SNPs of mitogenomes for population genetic analysis. **Table S6.** High-quality (HQ) SNPs of mitogenomes and plastomes for GWAS with 64 individual traits of 15 agronomic trait groups. **Table S7.** High-quality (HQ) SNPs of mitogenomes and plastomes for GWAS with 17 agronomic traits.

**Additional file 2: Figure S1.** Neighbor-joining (NJ) phylogenetic tree reconstructed using SNPs of mitogenomes from elite inbred lines of China. **Figure S2.** Multiple sequence alignment (MSA) in the *orf115-a2* region of the mitogenome of NB and CMS-S cytotypes. **Figure S3.** Neighbor-joining (NJ) phylogenetic tree reconstructed using SNPs of plastomes from elite inbred lines of China. **Figure S4.** Manhattan and Q-Q plots showing significant signals for the anthesis-silking interval (ASI) trait using SNPs of mitogenomes. **Figure S5.** Manhattan and Q-Q plots showing significant signals for the relative ear position (EP) trait using SNPs of mitogenomes. **Figure S6.** Manhattan and Q-Q plots showing significant signals for the days to silking (DTS) trait using SNPs of mitogenomes. **Figure S7.** Manhattan and Q-Q plots showing significant signals for the lower leaf angle (LAL) trait using SNPs of mitogenomes. **Figure S8.** Manhattan and Q-Q plots showing significant signals for the tassel branch number (TBN) trait using SNPs of mitogenomes. **Figure S9.** Manhattan and Q-Q plots showing significant signals for the cob color, stem diameter, and kernel color traits using SNPs of mitogenomes. **Figure S10.** Manhattan and Q-Q plots showing significant signals for the relative ear position (EP) using SNPs of plastomes. **Figure S11.** Genome-wide association analysis with 17 agronomic traits.

### Acknowledgements

We thank Dr. Yufeng Wu in the Bioinformatics Center of Nanjing Agricultural University for providing a computing facility for data analysis.

### Authors' contributions

S.C. and Z.J.C. designed the research; S.C., H.Z., and Y.L. performed the research; Y.L. contributed analytical tools; S.C. and Y.L. analyzed the data; Y.S. provided materials and supervision; S.C. and Z.J.C. wrote the paper.

### Funding

This work is supported by grants from the National Science Foundation (IOS1238048 to Z.J.C.), Jiangsu Collaborative Innovation Center for Modern Crop Production, and Shanxi Key Laboratory of Minor Crops Germplasm Innovation and Molecular Breeding (202105D121010-26 to H.Z.).

### Availability of data and materials

All sequencing data are available under NCBI under the BioProject accession numbers SRA051245 [59], SRA049859 [47], PRJNA609577 [46], PRJNA531553 [24], and PRJNA528290 [55] and accession numbers in Additional 1: Table S1.

### Declarations

#### Ethics approval and consent to participate

Ethics approval was not needed for this study.

#### Consent for publication

All authors read and approved the paper for publication.

#### Competing interests

The authors declare that they have no competing interests.

Received: 12 September 2023 Accepted: 28 February 2024  
Published online: 13 March 2024

## References

- Shiferaw B, Prasanna BM, Hellin J, Bänziger M. Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Secur.* 2011;3(3):307–27.
- Bennetzen JL, Hake S. *Maize Handbook*. New York: Springer Science + Business Media, LLC; 2009.
- Hake S, Ross-Ibarra J. Genetic, evolutionary and plant breeding insights from the domestication of maize. *ELife.* 2015;4:e05861.
- Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, 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 B.P. maize from the Central Balsas River Valley, Mexico. *Proc Natl Acad Sci U S A.* 2009;106(13):5019–24.
- Yang CJ, Samayoa LF, Bradbury PJ, Olukolu BA, Xue W, York AM, Tuholski MR, Wang W, Daskalska LL, Neumeyer MA, et al. The genetic architecture of teosinte catalyzed and constrained maize domestication. *Proc Natl Acad Sci U S A.* 2019;116(12):5643–52.
- 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.
- Yang N, Wang Y, Liu X, Jin M, Vallebuena-Estrada M, Calfee E, Chen L, Dilkes BP, Gui S, Fan X, et al. Two teosintes made modern maize. *Science.* 2023;382(6674):eadg8940.
- Stitzer MC, Ross-Ibarra J. Maize domestication and gene interaction. *New Phytol.* 2018;220(2):395–408.
- Li E, Liu H, Huang L, Zhang X, Dong X, Song W, Zhao H, Lai J. Long-range interactions between proximal and distal regulatory regions in maize. *Nat Commun.* 2019;10(1):2633.
- Doebley J, Stec A, Hubbard L. The evolution of apical dominance in maize. *Nature.* 1997;386(6624):485–8.
- Dorweiler J, Stec A, Kermicle J, Doebley J. Teosinte glume architecture 1: a genetic locus controlling a key step in maize evolution. *Science.* 1993;262(5131):233–5.
- Huang C, Sun H, Xu D, Chen Q, Liang Y, Wang X, Xu G, Tian J, Wang C, Li D, et al. ZmCCT9 enhances maize adaptation to higher latitudes. *Proc Natl Acad Sci U S A.* 2018;115(2):E334–41.
- Yang Q, Li Z, Li W, Ku L, Wang C, Ye J, Li K, Yang N, Li Y, Zhong T, 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.
- Guo L, Wang X, Zhao M, Huang C, Li C, Li D, Yang CJ, York AM, Xue W, Xu G, et al. Stepwise cis-regulatory changes in ZCN8 contribute to maize flowering-time adaptation. *Curr Biol.* 2018;28(18):3005–3015 e3004.
- Huang Y, Wang H, Zhu Y, Huang X, Li S, Wu X, Zhao Y, Bao Z, Qin L, Jin Y, et al. THP9 enhances seed protein content and nitrogen-use efficiency in maize. *Nature.* 2022;612(7939):292–300.
- Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS, et al. Improved maize reference genome with single-molecule technologies. *Nature.* 2017;546(7659):524–7.
- Sun S, Zhou Y, Chen J, Shi J, Zhao H, Zhao H, Song W, Zhang M, Cui Y, Dong X, et al. Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nat Genet.* 2018;50(9):1289–95.
- 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):594–7.
- Hufford MB, Xu X, van Heerwaarden J, Pyhajarvi T, Chia JM, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaepler SM, et al. Comparative population genomics of maize domestication and improvement. *Nat Genet.* 2012;44(7):808–11.
- Swarts K, Gutaker RM, Benz B, Blake M, Bukowski R, Holland J, Kruse-Peoples M, Lepak N, Prim L, Romay MC, et al. Genomic estimation of complex traits reveals ancient maize adaptation to temperate North America. *Science.* 2017;357(6350):512–5.
- Kistler L, Maizumi SY, de GregorioSouza J, Przelomska NAS, Malaquias Costa F, Smith O, Loiselle H, Ramos-Madriral J, Wales N, Ribeiro ER, et al. Multiproxy evidence highlights a complex evolutionary legacy of maize in South America. *Science.* 2018;362(6420):1309–13.
- Rojas-Barrera IC, Wegier A, Sanchez Gonzalez JJ, Owens GL, Rieseberg LH, Pinerio D. Contemporary evolution of maize landraces and their wild relatives influenced by gene flow with modern maize varieties. *Proc Natl Acad Sci U S A.* 2019;116(42):21302–11.
- Yang N, Liu J, Gao Q, Gui S, Chen L, Yang L, Huang J, Deng T, Luo J, He L, et al. Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. *Nat Genet.* 2019;51(6):1052–9.
- Tang ZX, Yang ZF, Hu ZQ, Zhang D, Lu X, Jia B, Deng DX, Xu CW. Cytonuclear epistatic quantitative trait locus mapping for plant height and ear height in maize. *Mol Breeding.* 2013;31(1):1–14.
- Romero Navarro JA, Willcox M, Burgueno J, Romay C, Swarts K, Trachsel S, Preciado E, Terron A, Delgado HV, Vidal V, et al. A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat Genet.* 2017;49(3):476–80.
- Perez-Zamorano B, Vallebuena-Estrada M, Martinez Gonzalez J, Garcia Cook A, Montiel R, Vielle-Calzada JP, Delaye L. Organellar genomes from a approximately 5,000-year-old archaeological maize sample are closely related to NB genotype. *Genome Biol Evol.* 2017;9(4):904–15.
- Bosacchi M, Gurdon C, Maliga P. Plastid genotyping reveals the uniformity of cytoplasmic male sterile-T maize cytoplasms. *Plant Physiol.* 2015;169(3):2129–37.
- Clifton SW, Minx P, Fauron CM, Gibson M, Allen JO, Sun H, Thompson M, Barbazuk WB, Kanuganti S, Tayloe C, et al. Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiol.* 2004;136(3):3486–503.
- Lemieux C, Otis C, Turmel M. Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution. *Nature.* 2000;403(6770):649–52.
- Brown WM. Mechanisms of evolution in animal mitochondrial DNA. *Ann N Y Acad Sci.* 1981;361:119–34.
- Luo S, Valencia CA, Zhang J, Lee NC, Slone J, Gui B, Wang X, Li Z, Dell S, Brown J, et al. Biparental inheritance of mitochondrial DNA in humans. *Proc Natl Acad Sci U S A.* 2018;115(51):13039–44.
- Yonova-Doing E, Calabrese C, Gomez-Duran A, Schon K, Wei W, Karthikeyan S, Chinnery PF, Howson JMM. An atlas of mitochondrial DNA genotype-phenotype associations in the UK Biobank. *Nat Genet.* 2021;53(7):982–93.
- Chen L, Liu YG. Male sterility and fertility restoration in crops. *Annu Rev Plant Biol.* 2014;65:579–606.
- Palmer JD, Herbon LA. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J Mol Evol.* 1988;28(1–2):87–97.
- Allen JO, Fauron CM, Minx P, Roark L, Oddiraju S, Lin GN, Meyer L, Sun H, Kim K, Wang CY, et al. Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. *Genetics.* 2007;177(2):1173–92.
- Beckett JB. Classification of male-sterile cytoplasms in maize (*Zea mays* L.). *Crop Sci.* 1971;11(5):724–0.
- Dewey RE, Timothy DH, Levings CS 3rd. Chimeric mitochondrial genes expressed in the C male-sterile cytoplasm of maize. *Curr Genet.* 1991;20(6):475–82.
- Kheyr-Pour A, Gracen VE, Everett HL. Genetics of fertility restoration in the C-group of cytoplasmic male sterility in maize. *Genetics.* 1981;98(2):379–88.
- Levings CS III. The Texas cytoplasm of maize: cytoplasmic male sterility and disease susceptibility. *Science.* 1990;250:942–7.
- Cui X, Wise RP, Schnable PS. The rf2 nuclear restorer gene of male-sterile T-cytoplasm maize. *Science.* 1996;272(5266):1334–6.
- Zabalá G, Gabay-Laughnan S, Laughnan JR. The nuclear gene Rf3 affects the expression of the mitochondrial chimeric sequence R implicated in S-type male sterility in maize. *Genetics.* 1997;147(2):847–60.
- Zhang ZF, Wang Y, Zheng YL. AFLP and PCR-based markers linked to Rf3, a fertility restorer gene for S cytoplasmic male sterility in maize. *Mol Genet Genomics.* 2006;276(2):162–9.
- Fauron CM, Casper M. A second type of normal maize mitochondrial genome: an evolutionary link. *Genetics.* 1994;137(3):875–82.
- Darracq A, Varre JS, Touzet P. A scenario of mitochondrial genome evolution in maize based on rearrangement events. *BMC Genomics.* 2010;11:233.

46. Wang B, Lin Z, Li X, Zhao Y, Zhao B, Wu G, Ma X, Wang H, Xie Y, Li Q, et al. Genome-wide selection and genetic improvement during modern maize breeding. *Nat Genet.* 2020;52(6):565–71.
47. Lai J, Li R, Xu X, Jin W, Xu M, Zhao H, Xiang Z, Song W, Ying K, Zhang M, et al. Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat Genet.* 2010;42(11):1027–30.
48. Andorf C, Beavis WD, Hufford M, Smith S, Suza WP, Wang K, Woodhouse M, Yu J, Lubberstedt T. Technological advances in maize breeding: past, present and future. *Theor Appl Genet.* 2019;132(3):817–49.
49. Kusmec A, Zheng ZH, Archontoulis S, Ganapathysubramanian B, Hu GP, Wang LZ, Yu JM, Schnable PS. Interdisciplinary strategies to enable data-driven plant breeding in a changing climate. *One Earth.* 2021;4(3):372–83.
50. Yang J, Lee SH, Goddard ME, Visscher PM. Genome-wide complex trait analysis (GCTA): methods, data analyses, and interpretations. *Methods Mol Biol.* 2013;1019:215–36.
51. Yang N, Lu Y, Yang X, Huang J, Zhou Y, Ali F, Wen W, Liu J, Li J, Yan J. Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. *PLoS Genet.* 2014;10(9):e1004573.
52. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011;88(1):76–82.
53. Dewey RE, Timothy DH, Levings CS. A mitochondrial protein associated with cytoplasmic male sterility in the T cytoplasm of maize. *Proc Natl Acad Sci U S A.* 1987;84(15):5374–8.
54. Xiao H, Zhang F, Zheng Y. The 5' stem-loop and its role in mRNA stability in maize S cytoplasmic male sterility. *Plant J.* 2006;47(6):864–72.
55. Xue W, Anderson SN, Wang X, Yang L, Crisp PA, Li Q, Noshay J, Albert PS, Birchler JA, Bilinski P, et al. Hybrid decay: a transgenerational epigenetic decline in vigor and viability triggered in backcross populations of teosinte with maize. *Genetics.* 2019;213(1):143–60.
56. van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, de Jesus Sanchez Gonzalez J, Ross-Ibarra J. 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.
57. da Fonseca RR, Johnson WE, O'Brien SJ, Ramos MJ, Antunes A. The adaptive evolution of the mammalian mitochondrial genome. *BMC Genomics.* 2008;9:119.
58. Guisinger MM, Kuehl JV, Boore JL, Jansen RK. Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proc Natl Acad Sci U S A.* 2008;105(47):18424–9.
59. Chia JM, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, et al. Maize HapMap2 identifies extant variation from a genome in flux. *Nat Genet.* 2012;44(7):803–7.

## Publisher's Note

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