BMC Plant Biology

RESEARCH

Open Access

Unveiling the heterosis pattern of modern maize breeding in Southwest China through population structure and genetic diversity analysis

Peng Ma[1,2*†], Hua Zhang[1,2†], Hongxia Shui[1,2†], Xuecai Zhang[3], Xiuquan Wang[1,2], Shibin Gao[4,5,6], Haiying Zhang[5,6], Zhi Nie[7], Chunyan Qing[1,2], Tingqi Lu[1,2], Qihua Pang[1,2], Wenzheng Pei[1,2], Hongmei Chen[1,2], Chenyan He[1,2], Bowen Luo[4,5,6*] and Dan He[1,2*]

Abstract

Maize (Zea mays L.) is an important food crop throughout the world and is also one of the earliest crops to use heterosis. In this study, we evaluated the genetic diversity, population structure, and selective sweep of 100 elite inbred maize lines collected from the current breeding program in Sichuan province, Southwest China, using 5,261,175 high-quality single nucleotide polymorphisms (SNPs). We discovered an abundance of genetic diversities and classified them into four groups. By combining kinship relationships, these groups were further divided into Tropic-local A, Improved-tropic, Tropic-local B, and Improved-local. Genomic differentiation was assessed using F_{st} values (0.21–0.44) as well as genetic diversity ($\pi = 6.07 \times 10^{-4} - 6.61 \times 10^{-4}$). We generated 900 (90 × 10) hybrids using 90 and 10 inbred maize lines from 100 diverse maize germplasms. All hybrids were evaluated for 10 traits in three replicate tests across two locations. We found that the patterns of G1 × G3, G1 × G4, G2 × G3, and G3 × G4 exhibited significant heterosis in yield-related traits and have been used in commercial breeding. In addition, we also explored the relationship between 10 traits of hybrid offspring and the number of heterozygous SNP. Under most heterosis modes, the best linear unbiased estimation (BLUE) value of the trait was highly consistent with the trend of deleterious SNPs, but there was a deviation in the G1 × G3 mode. Taken together, the results provide insight into the utilization of the current maize germplasm in Sichuan province to improve hybrid breeding.

Keywords Maize, Whole-genome sequencing, Single-nucleotide polymorphism, Combining ability, Heterosis pattern, Deleterious mutation

†Peng Ma, Hua Zhang and Hongxia Shui have contributed equally to this article.

*Correspondence:
Peng Ma
mapeng561@163.com
Bowen Luo
bowen_maize@sicau.edu.cn
Dan He
myhd095@163.com
Full list of author information is available at the end of the article

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Introduction

Maize stands as a crucial global food crop with significant implications for food security worldwide. According to the National Bureau of Statistics of China, maize ranks as the country's largest grain crop, cultivated across 43.324 million hectares (https://www.stats.gov.cn/). In the mountainous regions of Southwest China, maize productivity faces substantial challenges due to complex topography, dynamic ecological conditions, and low soil fertility [27, 36], particularly evident in Sichuan Province where cultivation occupied 1.849 million hectares in 2021 (http://www.chinawestagr.com/). As both a model organism for genomic studies [1, 2] and a pioneer crop for heterosis utilization [5], maize holds dual significance in agricultural and biological research. Heterosis, manifested as enhanced stress tolerance, adaptability, and yield in F1 hybrids relative to parental lines [12, 15, 21], is typically exploited through heterosis breeding populations comprising genetically diverse inbred lines with superior combining ability. These populations, shaped by successive rounds of natural and artificial selection, exhibit extensive genetic recombination and broad adaptability to various ecological conditions [41]. Among established heterosis patterns, the Reid ×Lancaster model remains predominant in temperate regions due to its proven effectiveness [42, 43].

Recent advances in bioinformatics and genome sequencing have transformed heterosis research. Whole-genome resequencing analyses of 1,604 maize lines revealed a positive correlation between heterozygosity and heterotic performance [39]. Subsequent studies demonstrated that structural variations between parental lines significantly influence hybrid vigor [35], while large-scale genetic analyses of 5,360 hybrid progenies identified 628 loci governing yield-related traits [24]. These findings collectively support the prevailing hypothesis that heterosis arises from complementation of recessive deleterious alleles through hybridization [3], a phenomenon extensively documented in plant genomes [8, 46].

Despite progress in molecular marker-assisted classification of heterotic groups [38, 41] and pedigree analysis [14, 25, 31] research gaps persist regarding deleterious mutation dynamics in Southwest China's maize breeding programs. Our study addresses this limitation through comprehensive genomic and phenotypic analyses of 100 core inbred lines from Southwest China ecological region. We established a 90 ×10 incomplete diallel cross population and evaluated 10 yield-related traits: ear length, ear diameter, cob diameter, ear tip barrenness, row number per ear, kernel number per row, ear weight, cob weight, hundred-kernel weight, and kernel thickness. Whole-genome sequencing of parental lines enabled identification of deleterious mutations, followed by comparative analysis of allelic frequencies across heterotic subgroups to determine optimal combinations for the mountainous regions of Southwest China.

Materials and methods

Plant material

A panel of 100 maize inbred lines used in this study for genome-wide sequencing. The panel was consisted of 52 lines obtained from the Mianyang Academy of Agricultural Sciences (MAAS) and 48 lines from the current breeding program of Sichuan province in Southwest China. Based on the precise delineation of population structure and thorough observations of the field performance of inbred lines, we selected 90 inbred lines from 100 maize inbred lines as female parents and 10 inbred lines with excellent performance as male parents. The 900 (90 ×10) NC II hybrids generated between the 100 inbred lines mentioned above that are applied for sequencing.

DNA extraction and qualification

All 100 lines of maize were cultivated at the MAAS in Mianyang, China in 2022. The extraction of maize DNA for illumina sequencing was carried out using the cetyltrimethylammonium bromide (CTAB) method from young, freshly harvested leaves. The quality of the genomic DNA was assessed by monitoring DNA degradation and contamination using 1% agarose gels. The concentration of DNA was determined using the Qubit® DNA Assay Kit with a Qubit® 2.0 Fluorometer (Life Technologies, CA, USA).

Library preparation, clustering and sequencing

A genomic DNA sample was obtained and then enzymatically fragmented to a size of 350 bp. Afterwards, the DNA fragments were subjected to end polishing, A-tailing, and ligation with the full-length adapter for Illumina sequencing. These libraries were then analyzed for size distribution using agarose gels and quantified through real-time PCR. The index-coded samples were clustered using a cBot Cluster Generation System with the Novaseq6000 S4 Reagent Kit (Illumina) following the manufacturer's instructions. Once clustering was complete, the DNA libraries were sequenced on the Illumina NovaSeq 6000 platform, generating 150 bp paired-end reads.

To eliminate reads that possess artificial bias, such as low-quality paired reads resulting from base-calling duplicates and adaptor contamination, we removed the following types of reads: (i) reads with 3nt unidentified nucleotides (N); (ii) reads containing the adaptor; (iii) reads with ≥20% of bases having phred quality ≤5. All subsequent bioinformatics analyses were conducted

using the high quality clean data, which were retained after performing these steps.

Reads mapping

The paired-end clean reads were aligned to the B73 RefGen_V4 genome were downloaded directly from genome website (http://www.gramene.org/) using BWA [17]. When one or both reads from a pair mapped to multiple positions, BWA selected the most probable placement. If there were multiple equally likely placements, BWA randomly chose one. Samtools [18] and Picard-tools (http://broadinstitute.github.io/picard/) were subsequently utilized to sort BAM files and mark duplicates. This was done to calculate the sequence depth of the entire exome, as well as the coverage and depth of each individual chromosome.

Variation detection

In this study, we utilized Picard-tools(http://broadinstitute.github.io/picard/) and samtools [18] to organize, identify duplicate reads, and reorganize the bam alignment results for each sample. Subsequently, SNPs and Indels were detected using the GATK software [6] and annotated using ANNOVAR software [37].

Clustering and population structure analyses

The population's genetic structure was analyzed using the ADMIXTURE (v1.3.0) software, and multiple K levels (K = 2 to 6) were calculated to determine the optimal number of subpopulations based on the cross validation (CV) error. Ultimately, K = 4 was identified as a reasonable number for grouping the population. To confirm the rationality of the 4 subgroups, a principal component analysis (PCA) was performed using filtered SNPs and the GCTA software. Initially, the genetic relationship matrix was generated using the parameter'make-grm'and subsequently, the top three principal components were estimated using the parameter'PCA3'. Additionally, an individual-based neighbor-joining tree was estimated based on the p-distance using the TreeBest software (v.1.9.2) with 1000 bootstrap replications.

Analysis of linkage disequilibrium

To estimate the linkage disequilibrium (LD) for all samples, we calculated the squared correlation coefficient (r^2) between pairwise SNPs using the software PopLDdecay. The program parameters were set as MaxDist = 1000, MAF = 0.05, Miss = 0.2 to calculate the average r^2 between two SNPs in 1000 kb windows.

Prediction of deleterious mutations

The SIFT4G [34] software was used to predict the putatively deleterious SNPs. For this analyses, when a SNP was normalized probability is < 0.05, this SNP was considered as a deleterious SNP (dSNP), and an SNP was considered as tolerated SNP (tSNP) if it show the SIFT score ≥ 0.05.

Genome diversity and selection sweeps

Fst and π in 20 kb sliding windows with a step size of 10 kb to quantify genomic differentiation using VCFtools. Sliding windows with the top 5% *Fst* and π values were considered as selected region. Cross population composite likelihood ratio (XP-CLR) score in 20 kb sliding windows with a step size of 10 kb to quantify genomic differentiation using XP-CLR python script. The windows with the top 5% of XP-CLR scores were considered as candidate sweeps region.

Field experiments and data collection

In 2023, the NC II hybrids were phenotyped in Mianyang (MY) and Chongzhou (CZ), located in Sichuan Province. Here, all hybrids were cultured in a completely random design in three independent replications completed in agricultural fields with uniform soil distribution. In each replication, each hybrid was planted with one 3-m rows with 25 cm between plants and 50 cm between rows, and the field management applied during the experiment was similar to the management practiced by farmers. At harvesting stage, 10 yield-related traits were collected, including ear length (EL, cm), ear diameter (ED, cm), cob diameter (CD, cm), ear tip barrenness (ETP, cm), row number per ear (RNPE), kernel number per row (KNPR), ear weight (EW, g), cob weight (CW, g), hundred-kernel weight (HKW, g), and kernel thickness (KT, cm). KT was measured using a vernier caliper, measuring the total thickness of ten kernels at the middle position. HKW was measured with five samples of 100 kernels randomly selected from the total kernels. Five similar ears out of ten within each hybrid were sampled, to eliminate possible outliers within hybrids, and all traits' means were taken.

Statistical analysis

Phenotypic descriptive statistics was analyzed using R software. The general combing ability (GCA) and specific combining ability (SCA) were performed by DPS software [32]. Best linear unbiased estimator (BLUE) was used to calculate phenotypic traits across multiple environments based on a linear model and evaluated by R software. Broad-sense heritability (H^2) were calculated as follows:

$$\text{(I)} \quad V_g = \frac{\sigma_{gi}^2 + \sigma_{gj}^2}{\sigma_{gi}^2 + \sigma_{gj}^2 + \sigma_{sij}^2} \times 100$$

$$\text{(II)} \quad V_s = \frac{\sigma_{sij}^2}{\sigma_{gi}^2 + \sigma_{gj}^2 + \sigma_{sij}^2} \times 100$$

$$\text{(III)} \quad y_{ijk} = \mu + G_i + E_j + R(E)_{jk} + E \times G_{ij} + \varepsilon_{ijk}$$

$$\text{(IV)} \quad H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{e} + \frac{\sigma_e^2}{er}}$$

where, (I) and (II) σ_{gi}^2 and σ_{gj}^2 represented genetic factor of the i th and j th inbred line, respectively. σ_{sij}^2 represented the interaction variance between genetic factor between gi and gj. (III) μ is the mean of each trait, G_i is the genetic factor of the i th inbred line, E_j is environmental factor of the j th environment, $E{\times}G_{ij}$ is the interaction factor between environmental and genetic factors, $R(E)_{jk}$ is factor of k th replication within environments, and ε_{ijk} is the residual error. (IV) σ_g^2, σ_e^2, and σ_{ge}^2 were genetic variance, residual error variance and the variance of genetic × environmental interaction, respectively [22].

Results

Genomic characterization of 100 maize inbred lines

The 100 maize inbred lines was performed whole-genome sequencing against the B73 reference genome, achieving an average sequencing depth of 12.7 ×(range: 10.4–20.4 ×) with 98.68% genome coverage (97.36%–99.20%) after data filtering (Table S1). Analysis of aligned clean reads identified 124,726,264 genome-wide SNPs, comprising 114,077,235 (91.46%) intergenic SNPs, 3,821,037 (3.06%) intronic SNPs, and 1,456,186 (1.17%) coding-region SNPs. Among coding region SNPs, 609,089 (41.83%) were synonymous substitutions while 809,093 (55.57%) caused amino acid changes (Fig. S1 A). Transition mutations (C > T and G > A) predominated the SNP spectrum (Fig. S1B). This comprehensive genomic variation profile establishes a critical foundation for subsequent genetic analyses.

SNP-based phylogenetic analysis reveals distinct groups and genetic divergence

A phylogenetic tree constructed from 5,261,175 high-quality SNPs revealed substantial genetic divergence among the 100 inbred lines. Genetic clustering analysis identified four distinct groups (Fig. 1A, Table S2). Group A comprised 24 ZNC442-derived improved lines, including 0151 NY65, 18 T- 1603–1, 20 T- 1908–1, IND, Mian705, and Mian722. Group B contained 31 lines predominantly from Southwest-adapted germplasm (Y9614, SCML0849, RP125, R08) combined with foreign lines (PH6 WC, LH224, PHR31), along with tropical/subtropical-derived materials such as Chang 7–2, C8210, and 18 T- 1657–1. Group C consisted of 27 lines developed from local germplasm, while Group D included 18 lines introduced by the International Maize and Wheat Improvement Center (CIMMYT). This result reflects genomic variations under different breeding backgrounds, geographical origins, or selection pressures, providing a genetic basis for the classification and utilization of maize germplasm resources.

The genetic landscape of 100 maize inbred lines using ADMIXTURE

Population structure analysis of 100 maize inbred lines using filtered SNPs (Fig. 1A-C) revealed distinct genetic clustering patterns, while the ADMIXTURE results were further visualized through PCA plots at K = 2, 3, and 4 (Fig. 2). Linkage disequilibrium analysis showed Group 3 exhibited the fast decay rate, contrasting with the slow decay pattern in Group 2. The coefficient of variation error rate was minimized at K = 4 (Fig. S2), the division boundary of the population became notably distinct, providing a clearer representation of the genetic structure.

Genetic structure clustering (Table S2) revealed that Group 1 (G1) consisted of 15 inbred lines. Half of these originated from both the Sichuan Agricultural University and the MAAS. Their genetic backgrounds were not uniform and exhibited several differences, such as ZNC442, SCML7275, S0906, and Mian755. Group 2 (G2) consisted of 21 inbred lines, and nearly half were sourced from the CIMMYT. These inbred lines exhibited tropical and subtropical genetic backgrounds and included S37, S273, YA8201, and Mian1604, which were cultivated from tropical or subtropical germplasms. Group 3 (G3) was the largest and featured 48 inbred lines, the majority of which originated from the current breeding program in Sichuan province, followed by the MAAS, and a few lines from the Pioneer. Similar to G2, the majority of the lines in the third group belonged to tropical and subtropical germplasms. Group 4 (G4) consisted of 16 inbred lines, which were mostly improved by the MAAS. These lines were primarily based on tropical germplasms, were combined with elite local germplasms, underwent multiple back- and self-crosses with the introduced tropical germplasm from foreign lines, and showed combining ability differences in their heterotic patterns.

Based on population structure and known pedigree information, we established four breeding groups: Local-tropic A (G1) and Local-tropic B (G3) representing tropical introgressed lines (CIMMYT/Pioneer/Suwan derivatives), Improve-tropic (G2) for tropical-adapted lines, and Improve-local (G4) highlighting locally enhanced germplasm. This classification elucidates the genetic improvement pathways via introgression of tropical germplasm and processes of local adaptation in maize germplasm from Southwest China.

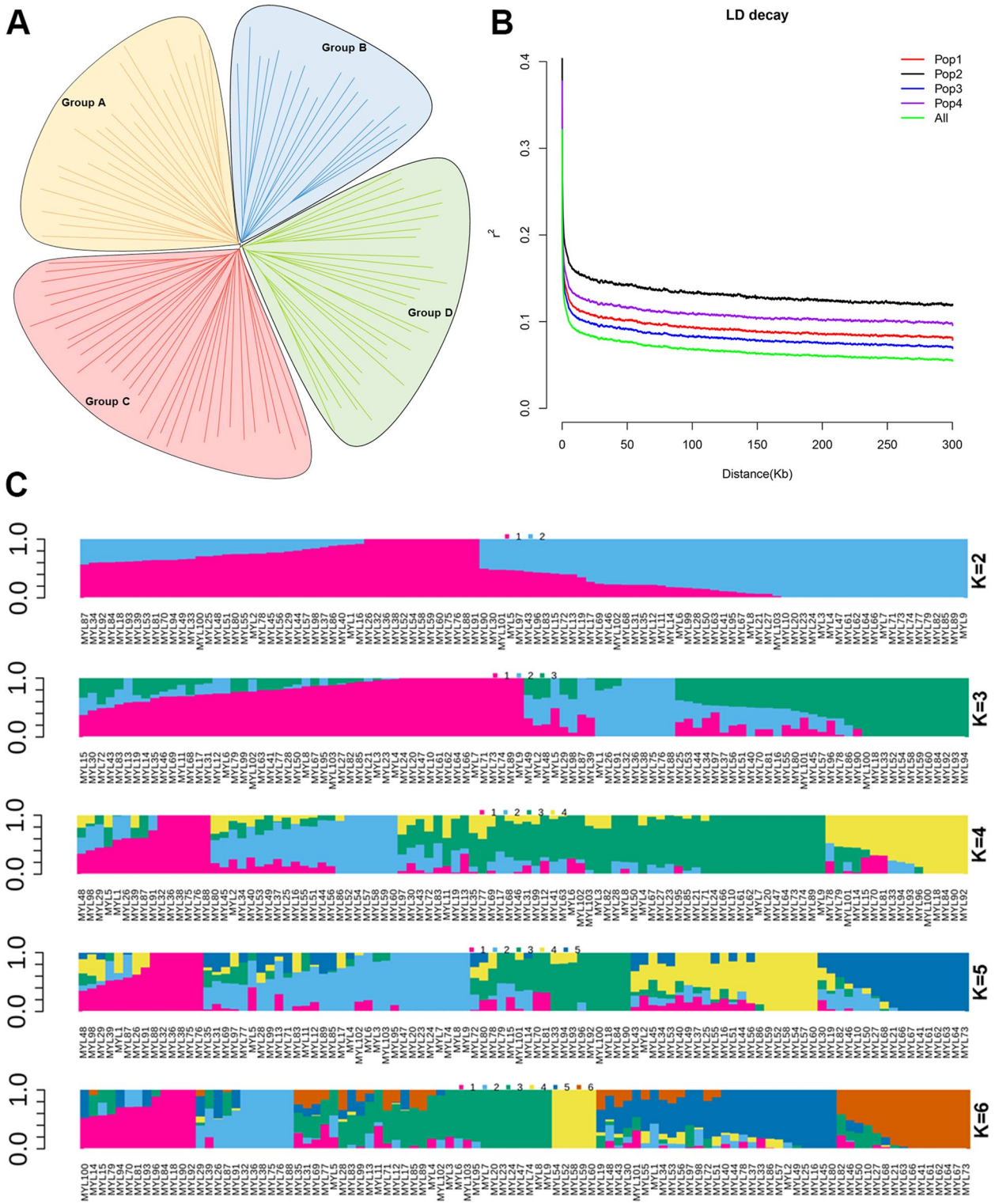

Fig. 1 **A** The neighbor-joining tree of 100 maize inbred lines based on the pairwise distance matrix using filtered SNPs. **B** The LD decay of SNPs. **C** The subgroup assignment by ADMIXTURE software for 100 maize inbred lines when K = 2,3,4,5,6

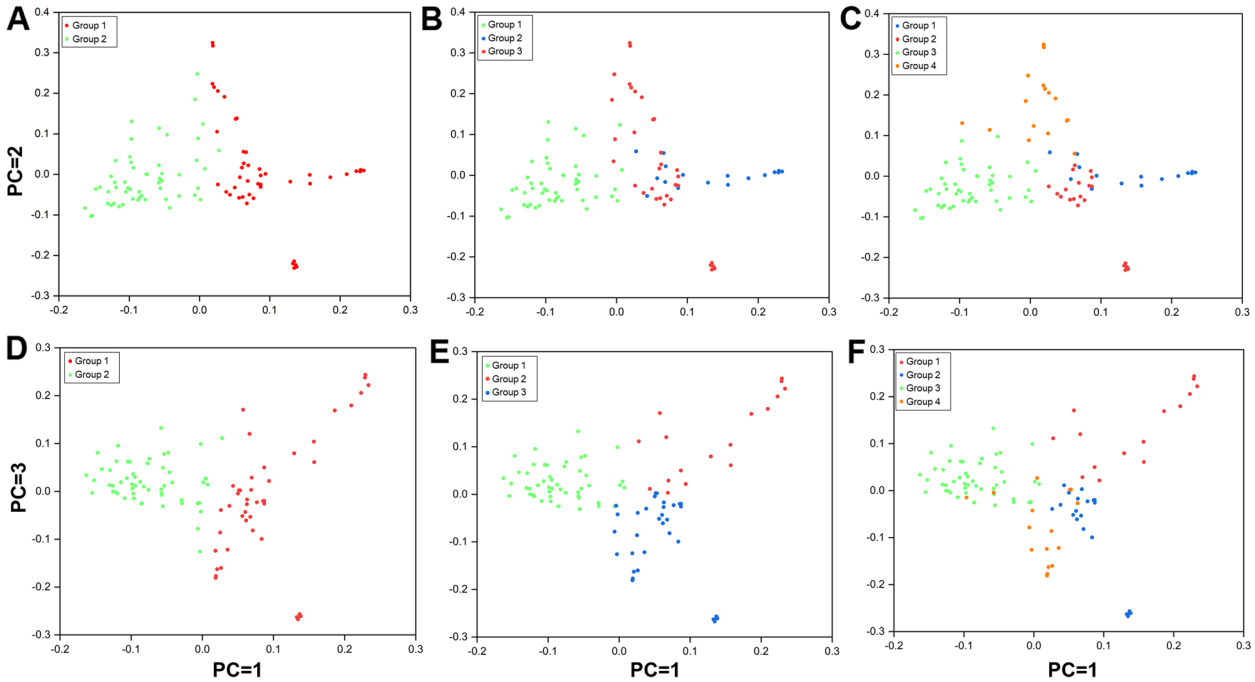

Fig. 2 The 2-D PCA plots of the ADMIKXTURE results for 100 maize inbred lines created by filtered SNPs. (**A**, **D**) K = 2; (**B**, **E**) K = 3; (**C**, **F**) K = 4

Genetic differentiation and diversity among inbred maize line groups

To evaluate genetic differentiation among the groups and genetic diversity within groups of each line, we calculated pairwise fixation indices (F_{st}) and π values among the groups of different lines, respectively. As shown in Fig. 3, the F_{st} value was greater than 0.2, indicating that there was a significant genetic difference among the inbred line groups, particularly between the Improved-tropic and Improved-local groups (F_{st} = 0.33), whereas there was significant genetic differentiation. In addition, we observed siginificant genetic diversity within each inbred line group, as indicated by the π value. The SNP diversity of Improve-local ($\pi = 6.61 \times 10^{-4}$) was greater compared with that of Tropic-local A ($\pi = 6.47 \times 10^{-4}$), Tropic-local B ($\pi = 6.07 \times 10^{-4}$) and Improve-tropic ($\pi = 6.09 \times 10^{-4}$). The results suggest that during the process of genetic improvement, the other three subgroups were mixed with the local germplasm genetic background in the Improve-local group, which resulted in the highest SNP diversity.

Selective sweep detection in 100 inbred lines using XP-CLR

Population structure-based selective sweep analysis of 100 inbred lines identified distinct selection patterns across four subgroups (Fig. 4, Table S3). Comparative XP-CLR analysis revealed the Tropic-local A/B subgroups contained 3,278 selected genomic intervals harboring 2,045 annotated genes with the highest among six comparison groups. In contrast, the Improve-tropic/local subgroups showed 2,968 selection intervals with only 1,523 genes, representing the lowest gene density. Functional enrichment of candidate genes (Fig. S3 and S4) demonstrated differential adaptation strategies, tropical subpopulations (Tropic-local A/B) exhibited strong enrichment in jasmonate signaling and immune activation, cell wall organization, and water stress response. Improved subgroups (Improve-tropic/local) displayed predominant selection signatures in developmental regulation. Notably, the tropical subgroups'genomic architecture showed intensive selection on biotic stress resistance genes (jasmonate-mediated defense pathways) and abiotic adaptation genes (cell wall biosynthesis). The improved subgroups'selection patterns emphasized light signaling and carbohydrate metabolism. These differential selection characteristics between tropical and improved subgroups could reflect distinct evolutionary pressures, with the former maintaining broad genetic diversity for environmental adaptation, while the latter underwent intensive selection for agronomic optimization.

Heterosis pattern evaluation for yield-related traits

Descriptive statistics and broad-sense heritability (H^2) analysis of 10 yield-related traits across Mianyang (MY)

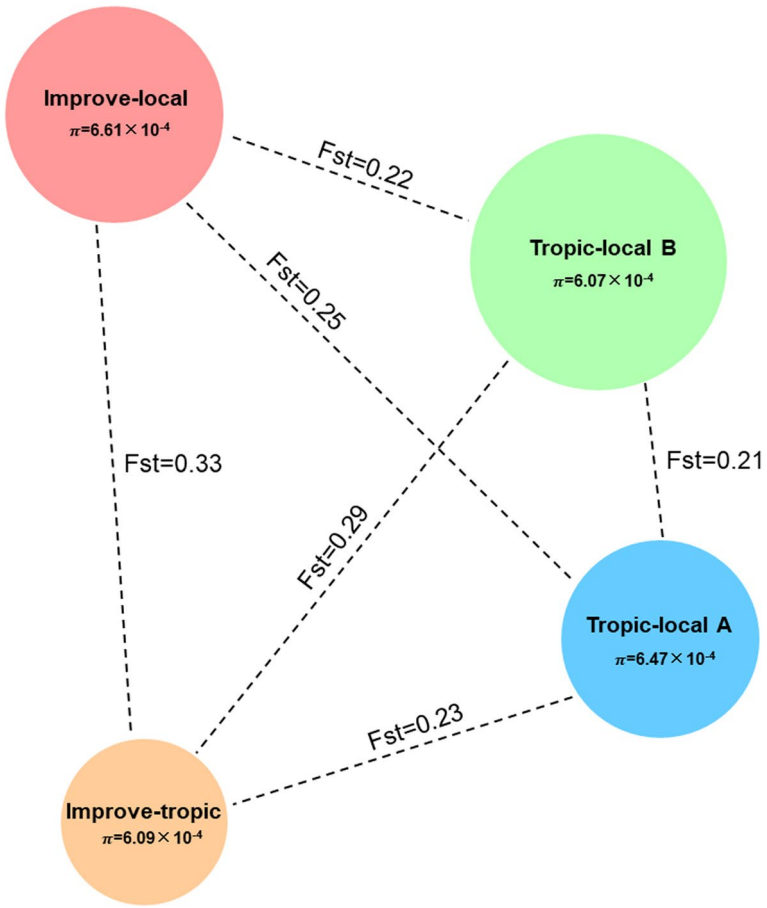

Fig. 3 Diversity and the genetic differentiation among subgroups

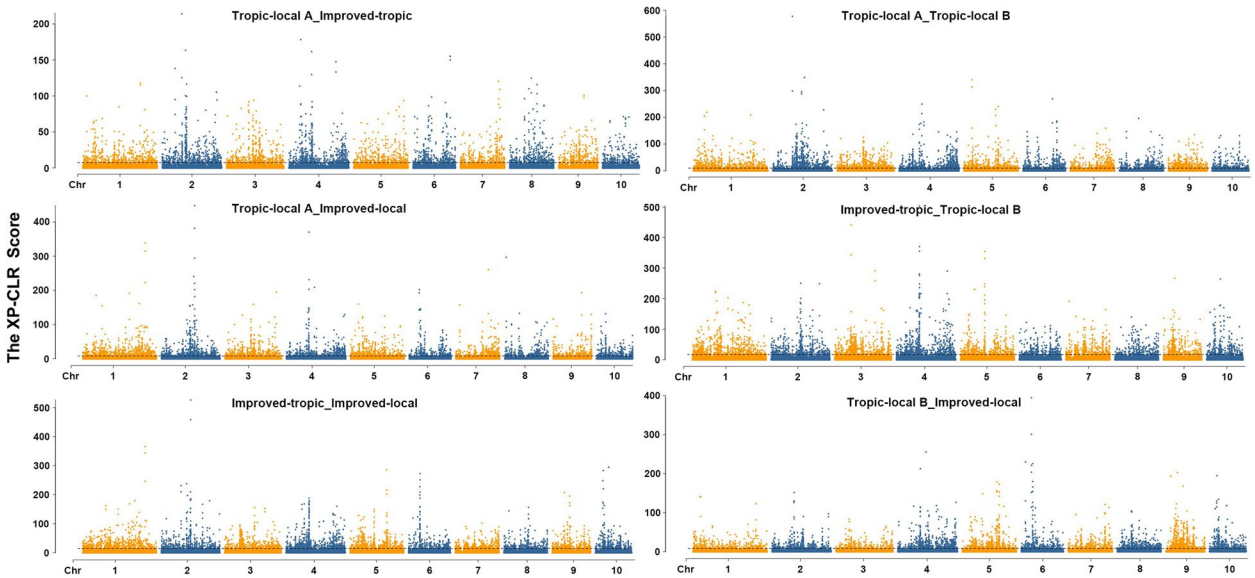

Fig. 4 Genome-wide selection sweep between among subgroups. Manhattan plot of the genome-wide distribution of XP-CLR scores among different subgroups by a 20 kb window size and a 10 kb step size (**A**, **B**, **C**, **D**, **E**, **F**)

and Chongzhou (CZ) environments demonstrated substantial heritable variation (H^2 = 0.62–0.68) for most traits (Table 1). Variance analysis revealed significant genotype × environment interactions for all traits except row number per ear (RNPE), indicating environmental sensitivity of yield components. The correlation analysis of 10 traits carried out for MY and CZ is shown in Table S6. ADMIXTURE-based classification of 100 inbred lines into four genetic groups (excluding mixed group) generated nine heterosis patterns through diallel combinations. Best linear unbiased estimates (BLUE) analysis of hybrid performance across patterns revealed significant trait-specific differentiation (Fig. 5). The ear diameter (ED) trait demonstrated the highest proportion of significant differences among all patterns, which accounted for approximately 72% of the total. The row number (RN) trait followed closely with a significant difference of 61%. In contrast, kernel number per row (KNPR) and kernel thickness (KT) exhibited the lowest proportion of the significant differences. We propose the potential assessment of hybrids created through diverse heterosis patterns, which may be effectively done by focusing on various yield-related traits. For example, the G2 × G3 pattern significantly outperforms the G1 × G4

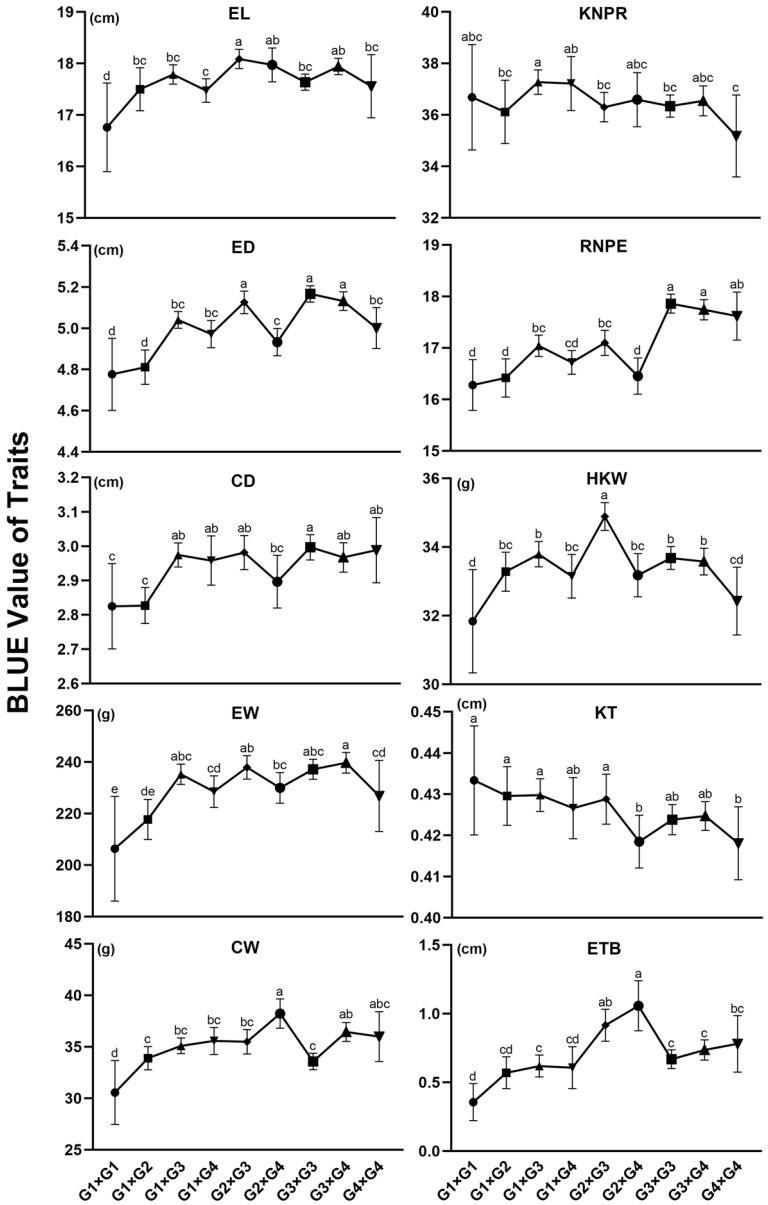

Fig. 5 The BLUE value of traits of hybrids with different heterosis patterns. Same letters are not significant at $P < 0.01$

Table 1 The descriptive statics and broad heritability estimates (H^2) of 10 yield-related traits in Mianyang (MY) and Chongzhou (CZ)

Trait	Environment	Mean	Sd	Min	Max	Skew	Kurtosis	H^2	F-value (Environment)
EL	CZ	17.99	1.272	10.67	29.06	0.04873	0.3188	0.66	1.099**
	MY	17.6	1.334	12.2	34.2	0.05311	2.805		
ED	CZ	5.075	0.2918	4.286	6.075	0.01117	-0.04947	0.68	81.01**
	MY	4.990	2.627	38.5	57.7	0.1046	-0.07348		
CD	CZ	2.981	0.2568	2.329	3.756	0.009834	0.2267	0.66	52.33**
	MY	2.874	1.858	22.8	35.5	0.07395	0.04381		
ETB	CZ	0.789	0.6273	0	7.333	0.02402	2.433	0.22	1.780**
	MY	0.7021	0.4702	0	3.1	0.01872	0.8736		
RNPE	CZ	17.29	1.381	11.56	21.33	0.05287	0.2237	0.62	1.055ns
	MY	17.45	1.344	14	20.9	0.0535	0.04541		
KNPR	CZ	37.57	2.571	22.33	45.17	0.09846	-0.8405	0.67	2.912**
	MY	35.8	4.388	17.3	44.8	0.1747	-0.5803		
KT	CZ	0.439	0.02714	0.3372	0.5325	0.001039	-0.03736	0.67	1.743**
	MY	0.4081	0.03583	0.3	1	0.001426	8.345		
HKW	CZ	33.39	2.313	18.94	38.9	0.08856	-0.9099	0.66	1.326**
	MY	33.69	2.663	21.6	41.4	0.106	-0.4557		
EW	CZ	233.2	26.84	75.82	307.7	1.028	-0.7994	0.64	1.007**
	MY	235.4	26.93	130.7	311.7	1.072	-0.4073		
CW	CZ	36.38	5.884	11.93	61.23	0.2253	0.3313	0.62	1.261**
	MY	33.42	5.241	18.6	53	0.2086	0.2453		

**Significant at $P < 0.01$; ns Not significant

pattern in the ED, indicating that hybrids derived from the G2 × G3 pattern likely exhibit superior performance for this trait compared with those from the G1 × G4 pattern. These results provide valuable insight for the selection of appropriate heterosis patterns in the creation of hybrid maize varieties.

Genetic potential and heterosis patterns in maize hybrids

Based on the heterosis patterns classification, nine distinct heterosis patterns were identified through crosses within the four groups. The general combining ability (GCA) and specific combining ability (SCA) of these hybrids were systematically calculated by using DPS software (Tables S7 and S8), revealing highly significant GCA and SCA effects across all hybrid traits. Most traits exhibited high broad-sense heritability ranging from 0.62 to 0.68 (Table 1), except for TBN (0.22), indicating substantial genetic contributions to phenotypic variation in maize grain quality and yield-related traits. Furthermore, there were significant differences in GCA among the various traits (Table S7). For G1, five inbred lines exhibited negative GCA effects for both HKW and ET, pointing to the genetic potential of these lines to reduce maize yield-related traits. Moreover, three inbred lines (P041, P052, and P089) exhibited positive GCA effects for all 10 measured traits. Notably, inbred line P032 demonstrated a negative GCA effect on the EW, but showed positive GCA effects on the EL and ED. For G2, in which most inbred lines are from CIMMYT, they exhibited strong GCA effects for the EL, ETB, and CW traits. G3 was primarily composed of American germplasm and their improved lines, and exhibited superior GCA effects for the EW, HKW, KT, KNPR, and RNPE traits. Finally, in G4, many inbred lines demonstrated positive GCA effects for most traits, such as P038 and P080, which exhibited high GCA effects and may be widely used by breeders for hybrid breeding.

Heterotic patterns contribute to the advantages of trait characteristics in heterosis

The strength of heterosis is primarily determined by the self-crossing compatibility of maize. For maize breeding, the successful cultivation of significant heterosis in practice depends on the creation of superior self-crossing compatibility. Among the hybrids derived from various heterosis groups, the G3 × G3 hybrids exhibited the lowest SCA effect in several traits (Fig. 6 and Table S8). Conversely, the G2 × G3 hybrids exhibited the highest SCA effect for all traits, except for EL and HKW. Notably, the G1 × G3, G1 × G4, G2 × G3, and G3 × G4 hybrids showed favorable SCA effects in all traits compared with other heterosis patterns, except for ETB. After further

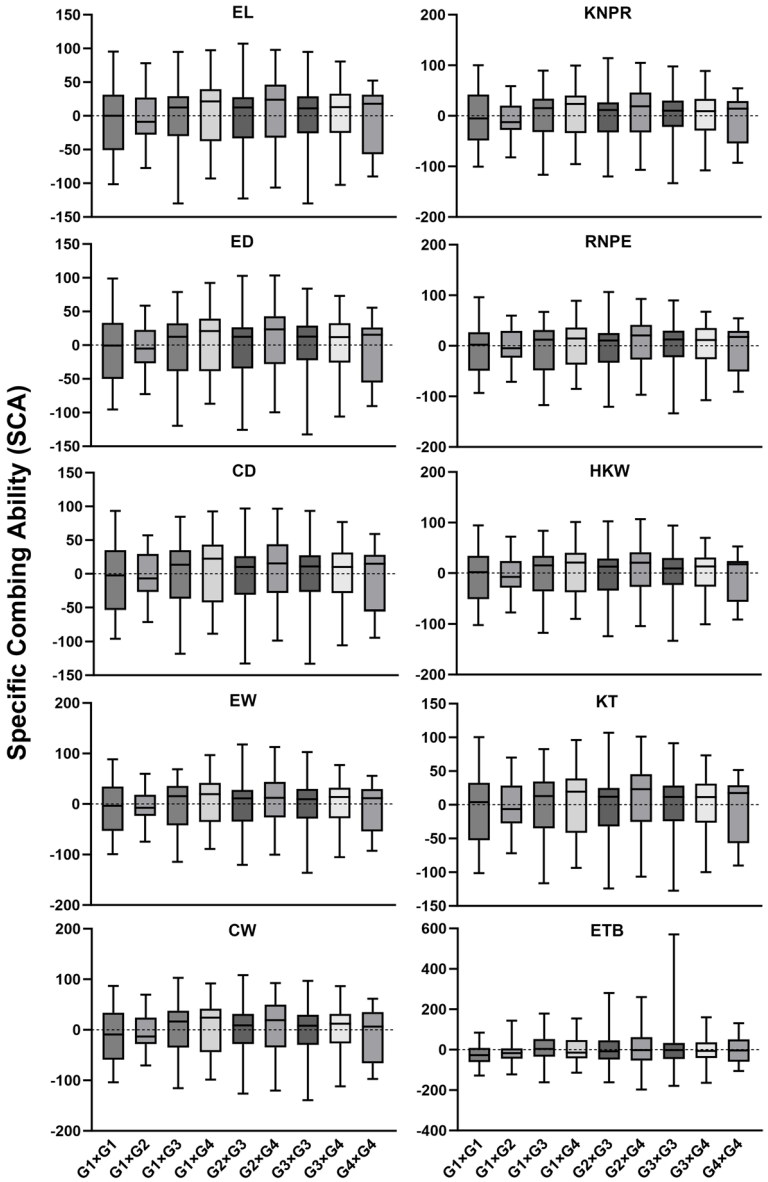

Fig. 6 The specific combing ability of traits of hybrids with different heterosis patterns

analysis of the parents of the officially approved hybrids (Table 2) and their corresponding heterosis pattern classification based on ADMIXTURE analysis, we discovered that these hybrids were specifically bred through the four heterosis patterns: G1 ×G3, G1 ×G4, G2 ×G3, and G3 ×G4. This result demonstrates that dividing the population into four subgroups using ADMIXTURE analysis of SNPs from a population of 100 inbred maize lines is reasonable. Furthermore, it also suggests the applicability of the heterosis group division in the Southwestern maize ecological region, thus providing useful information for future maize breeding.

The impact of deleterious mutations on heterosis in maize

Deleterious mutations can affect the phenotype of crops and their adaptability to the environment and are also one of the critical perspectives for explaining the phenomenon of heterosis [11, 46]. We used SIFT4G software to conduct a deleterious mutation analysis of high-quality SNPs across 100 inbred lines. A total of 6746 deleterious SNPs (SNPs) were identified (Table S9). Of these, non-synonymous dSNPs were the most prominent, totaling 4490, followed by 40 start-lost dSNPs. Conversely, synonymous dSNPs were the least abundant, with only six detected. In addition,

Table 2 The list of modern commercial hybrids which were bred from inbred lines cross in our study

Hybrids	Female parent	Male parent	Officially approved number	Heterotic pattern in this study
Chuandan 428	R08	YA8201	2007001	Group 3 (♀) × Group 2 (♂)
Yunrui 4	YR4M	YR4 F	200602	Group 3 (♀) × Group 2 (♂)
Miandan 232	Mian7897	S0906	20180331	Group 4 (♀) × Group 1 (♂)
Huiyu 366	HF19 - 1	HF09B	2019085	Group 3 (♀) × Group 2 (♂)
Rongyufengzan	Y9614	ZNC442	20200422	Group 3 (♀) × Group 1(♂)
Xieyu 901	CTN3297	CTN2622	20206236	Group 4 (♀) × Group 3 (♂)
Miandan 53	S6338	CTN2622	20212054	Group 4 (♀) × Group 3 (♂)
Miandan 23	Mian787	S6338	20220003	Group 3 (♀) × Group 4(♂)
Chuandan 99	ZNC442	SCML0849	20220504	Group 1 (♀) × Group 3 (♂)
Miandan 903	Mian7247	S6338	20230009	Group 3 (♀) × Group 4 (♂)

within the category of tolerant SNPs (tSNPs), a total of 14,380 such SNPs were detected. Subsequently, we analyzed the degree and quantity of heterozygosity among the identified dSNPs within the nine heterosis patterns. In terms of quantity and frequency of heterozygosity, the dSNPs in the heterosis patterns among the same subgroups (G1 × G1, G3 × G3, G4 × G4) exhibited a relatively dispersed distribution, suggesting a lower degree of heterozygosity for these SNPs (Fig. 7). In contrast, heterosis patterns between different subgroups exhibited pronounced SNP clustering. These findings strongly suggest that hybrids generated from the same subgroup generally exhibit lower dSNP heterozygosity levels, a characteristic that may have non-negligible implications on individual phenotypes and genetic mechanisms.

We delved deeper into the correlation between ten key phenotypic values related to the ear yield of hybrids for nine heterosis patterns and the number of heterozygous SNPs. For most heterosis patterns, the BLUE value of the phenotypes and the trend of heterozygous dSNPs exhibited a high degree of consistency, except for the G1 × G3 patterns, in which a deviation was observed. Notably, under the G2 × G4 heterosis patterns, the number of heterozygous dSNPs exhibited a consistent trend with specific phenotypic traits, including HKW, EL, ETB, KNPR, and CW, whereas this phenomenon was absent in other phenotypes. Furthermore, when focusing on the SCA exhibited by different heterosis combinations, a striking pattern emerged. The trends in SCA for all phenotypes were highly consistent with the heterozygous frequency of dSNPs. This discovery not only enhances our understanding of the interplay between heterosis and genetic complexity, but also provides critical theoretical grounds for subsequent genetic improvements aimed at enhancing crop disease resistance and yield.

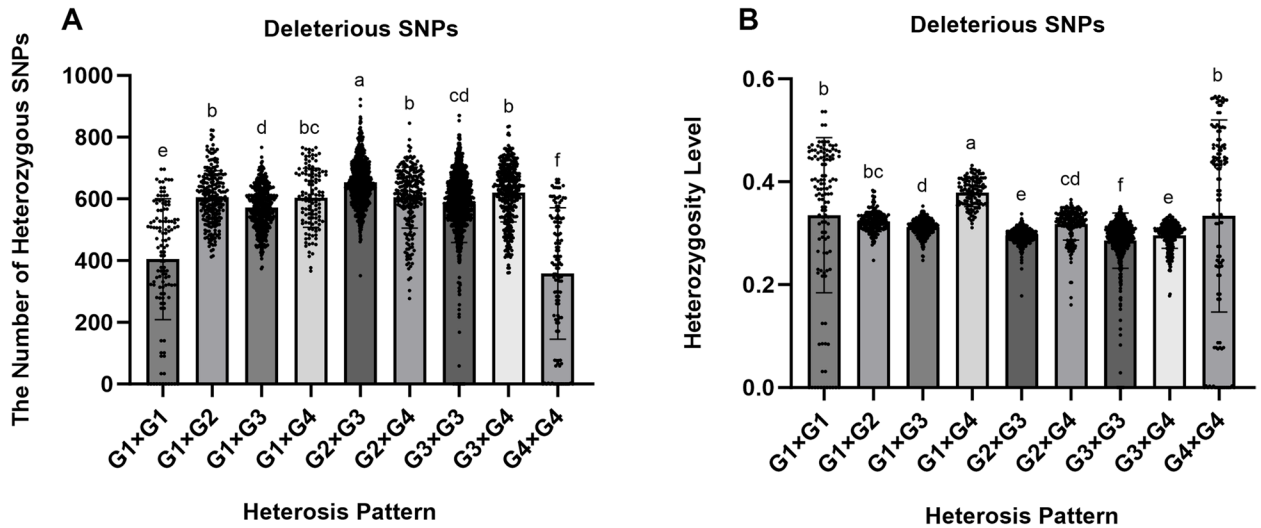

Fig. 7 The number and level of heterozygosity of dSNPs in hybrids with different heterosis patterns. Same letters are not significant at $P < 0.01$

Discussion

Exploring genetic diversity of SNP loci polymorphisms

The significant advantages of high-throughput sequencing technology have been demonstrated in previous studies, accelerating our understanding and exploration of the maize genome [3, 11, 12, 23, 25, 46]. In this study, we identified 5,261,175 high-quality SNPs in 100 maize inbred lines through high-throughput sequencing. These SNPs were evenly distributed across 10 chromosomes with a reference genome mapping rate exceeding 98%. This scale significantly surpasses the 955,120 SNP markers using GBS technology [38], demonstrating the accuracy of genetic variation analysis in this study. Compared with the population structure analysis in specific Chinese ecological regions [45], our genome-wide SNP profiling and discovery of recurrent CG to TA mutation patterns enhance genetic diversity characterization while revealing putative functional mutations underlying maize environmental adaptation. Additionally, the LD decay rate ($r^2 = 0.44$ at 400 kb) observed here was faster than that reported by previous stduy in heterosis pattern analysis [14]. This discrepancy may stem from differences in population genetic backgrounds or improved marker density. This study further demonstrates the feasibility and significance of genome-wide SNP marker analysis in marker-assisted breeding.

Population structure and heterotic group assignment assessed by SNPs

Traditional breeding relies on pedigree records to infer genetic relationships. However, this study found that only 18% of inbred lines had clear pedigrees, and pedigree-based relationships failed to capture genomic-level changes in gene flow caused by selection pressure. To address this, SNP-based ADMIXTURE and PCA analyses classified 100 southwestern maize inbred lines into four genetically distinct subgroups (*Fst* = 0.21–0.33), overcoming pedigree limitations. This classification aligns with Wright's [40] *Fst* criteria for population differentiation and reveals clear subgroup boundaries with minimal admixture. Unlike traditional heterotic group classification based on phenotypic performance, this study quantified genetic divergence using SNPs, establishing a genomic standard for predicting heterosis. For example, crosses between subgroups with high *Fst* values (G2 ×G3) may exhibit stronger dominant complementary effects, while low-divergence combinations (G3 ×G3) show reduced SCA due to genetic redundancy. Therefore, combining molecular markers with traditional pedigree breeding techniques provides an effective approach to analyzing genetic relationships within populations and to classify heterosis groups, thus improving genetics in modern maize breeding.

The difference between population structure and phylogenetic clustering

There are distinct boundaries among different groups and few overlapping regions indicating that ADMIXTURE could be the optimal method in cluster classification. By combining the findings of the genetic correlation and population structure analysis, we discovered that only 73 inbred lines were consistent with the results of the phylogenetic cluster. The remaining 27 inbred lines exhibited differences in population structure and phylogenetic cluster. The results of population structure revealed that 26 inbred lines were classified into three groups based on a phylogenetic cluster, with 11, 5, and 11 lines assigned to Group 1, Group 2, and Group 4, respectively. Notably, Group 3 did not include inbred lines from the phylogenetic cluster, but 33% of the differences between cluster analysis and structure were attributed to Group 3. These results suggest that genetic exchange is common among inbred lines in Group 3, thereby enhancing the genetic diversity of this group. The phylogenetic clustering effectively resolved genetic backgrounds of unspecified lines (1599 A/1719 A grouping with RP125 in Group3; Ye478 clustering with Mian7897) (Table S2). Structure-derived clusters demonstrated inter-group communication patterns. Notably, commercial hybrids (Table 2) derived from single crosses between lines from distinct heterotic groups, highlighting the importance of genetic characterization for optimizing heterotic group assignment and hybrid selection with superior combining ability.

Functional enrichment analysis of genes in selected regions

Through selection sweep analysis, mutations occurring in the genome of this population during genetic improvement were identified [9, 19]. By detecting genomic regions with reduced polymorphism due to strong selection pressure, researchers can locate favorable mutations fixed during evolution. For example, in this study, XP-CLR analysis revealed that genes in tropical subgroups (Tropic-local A/B) were significantly enriched in key functional categories related to disease resistance and environmental adaptation, such as jasmonic acid signaling pathways, immune activation, and water stress response. These genes may play critical roles in helping maize cope with biotic and abiotic stresses. Additionally, selective sweep analysis showed that genes in improved subgroups (Improve-tropic/local) were concentrated in developmental regulation, light signaling, and carbohydrate metabolism, reflecting artificial selection preferences for high-yield and quality traits. However, long-term selection pressures may have reduced genetic diversity in these improved varieties, leading to fewer genes related to disease resistance and environmental

adaptability. Therefore, integrating selective sweep analysis results allows breeders to strategically introduce disease- and stress-resistant genes from tropical germplasm into improved varieties, enhancing hybrid performance in complex ecological environments.

The potential use of heterotic populations in hybrid breeding

Modern maize breeding relies on distinct heterotic groups (maternal groups like SS and PA, and paternal groups like NSS, Iodent, SPT, and PB), which exhibit complementary genetic divergence that enhances heterosis. Maternal groups (SS and PA) are typically selected for traits like early maturity, reduced tassel branching, and efficient grain dehydration, which align with maternal roles in seed production. Paternal groups (NSS and SPT) often contribute to traits such as increased ear diameter, kernel row number, and cob weight, optimizing pollen dispersal and hybrid yield [15, 21].

The primary pattern of heterosis in the Southwest China maize ecological region is the use of "self-breeding" and "foreign-introduced" lines, including foreign and self-bred germplasms. Initially, the heterotic groups were established based on the origin of maize, with little improvement. In the early 1990 s in Northern China, Lancaster, Reid, SPT, Zi330, and E28 were introduced [33],however, the common heterotic patterns used in Northern China were not suitable for the Southwest China maize ecological region, such as the Sichuan province [16, 19, 20]. Based on the breeding practices and molecular marker research results in the Southwestern maize ecological region, the hybrid varieties and their parental inbred lines extensively planted in the Southwest ecological region may be classified into two patterns: temperate germplasm and tropical germplasm [29]. Moreover, the Southwest ecological region has abundant local germplasms, prompting studies on the heterotic groups and models of the local maize. Because many local varieties have yielded traits that are close to or exceed those of famous heterotic groups, they have great potential for the development and use of local germplasms. As shown in Table S2, there are four subgroups were composed of excellent self-crossed lines that combine local superior germplasms with introduced tropical, temperate, and subtropical germplasm, which have been continuously improved through genetic modification, such as Mian7247 (Mian724 ×Reid), Mian722 (Syngenta tropical germplasm), CTN2622 (local germplasms ×hybrid with American temperate), and SCML7275 (F06 ×hybrid with Lancaster). In practical maize breeding programs, the determination of an inbred line's suitability as maternal and paternal parent is guided by its classification within heterotic groups and the evaluation of key traits, including ear height, ear length, kernel weight, pollen shedding duration, and tassel branch number. For instance,"Chuandan 99,"currently the most extensively cultivated and commercially influential maize hybrid in Southwestern China, exemplifies this strategy [30]. Its maternal parent, ZNC442, exhibits superior maternal characteristics such as large ear size, excellent kernel set, and a high kernel output ratio. In this study, ZNC442 was categorized into Group A based on its heterotic pattern. Conversely, the paternal parent, SCML0849, demonstrates paternal advantages including abundant pollen production and photoperiod-insensitive flowering behavior, and was classified into Group C within the heterotic grouping framework. These findings underscore the necessity to clarify genetic relationships between local germplasms and elite inbred lines, establish heterotic groups, and develop rational hybrid heterotic models.

Combining ability and heterosis in hybrid maize breeding

The observed differences in SCA among heterotic groups provide crucial genetic insights for maize breeding in the Southwest China ecological region. The low SCA effects observed in G3 ×G3 hybrids across multiple traits suggest potential genetic redundancy within subgroup G3, likely caused by reduced allelic diversity due to long-term directional selection. This finding highlights the need to avoid excessive intra-group crosses in practical breeding and instead introduce exotic germplasms (tropical and temperate materials) to break genetic homogeneity within subgroup G3. For example, priority should be given to inbred lines with demonstrated genetic complementarity, such as CTN2622 (local ×temperate) and SCML7275 (Lancaster-derived lines), for hybridization with G3 lines to restore heterotic potential. The SCA performance of the G2 ×G3 combination except for EL and HKW traits, suggests dominant-overdominant synergistic effects between these subpopulations. In breeding practice, G2 population (Mian722 containing tropical germplasm) could compensate for genetic deficiencies in G3 population. Simultaneously, individuals in the G3 group with maximum genetic distance from G2 can be selected through SNP markers to maximize heterosis. The subpopulation classification results from ADMIXTURE analysis enable prediction of SCA potential for hybrids. Priority testing combinations like G1 ×G4 and G3 ×G4 can be proposed while bypassing inefficient G3 ×G3 pairings, thereby shortening the breeding cycle.

Heterosis patterns suitable for the Southwest China maize ecological region

Analyzing heterosis patterns tailored for the Southwest China maize ecological region is critical for maize

breeding because of its complex and diverse climatic and soil conditions. Researchers have achieved remarkable success in identifying heterosis patterns for this region, which integrate the traits of the local germplasms with exotic germplasm disease resistance, stress tolerance, and other beneficial genes to enhance maize yields and quality [10, 14, 16, 28, 44]. In this study, using 526,117,5 high-quality SNPs, we were able to categorize inbred maize lines into distinct subgroups using ADMIXTURE. A thorough analysis of the SCA of hybrids resulting from combinations of different heterosis patterns among subgroups revealed four suitable heterosis patterns (G1 ×G3, G1 ×G4, G2 ×G3, and G3 ×G4) for this ecological region. These patterns exhibit significant advantages with respect to yield and quality, such as EL, ED, EW, and CW. This further confirmed that hybrids derived from these heterosis patterns can maintain stable genetic gains and yield advantages, even in complex ecological environments. This classification is not only based on the similarity of genetic information, but also takes into full account ecological adaptability and phenotypic traits, resulting in higher genetic consistency within each subgroup. By conducting hybrid combinations among different subgroups, we can screen out hybrids with strong heterosis and cultivate new high-yielding, high-quality maize hybrids suitable for the Southwest China maize ecological region. This subgroup classification method based on molecular markers not only improves the efficiency and accuracy of heterosis, but provides further insight into modern maize breeding.

Exploring the important role of deleterious mutations in maize genetics and breeding underlying heterosis

Deleterious non-synonymous SNPs (dSNPs) directly disrupt gene function, altering crop phenotypes and environmental adaptability [26]. Reducing these mutations is critical for improving maize yield and quality during breeding. Hybrids from crosses between different subgroups showed higher dSNP heterozygosity, which may enhance heterosis. This phenomenon likely arises from increased genetic diversity and complementary effects, improving hybrid growth, development, and stress tolerance [7]. A significant correlation between hybrid phenotypic values and heterozygous dSNP counts further supports the role of genetic variation in shaping phenotypes. However, deviations in the G1 ×G3 hybrid pattern suggest that heterosis mechanisms involve interactions between genetic and environmental factors. Therefore, a comprehensive approach that considers multiple factors is necessary when elucidating the heterosis phenomenon.

Optimizing breeding strategies requires prioritizing crosses between genetically distinct subgroups to maximize heterozygosity and reduce deleterious allele transmission. Molecular marker-assisted selection can further enhance hybrid genetic quality by minimizing dSNP inheritance. Future research should explore dSNP effects across diverse crops and genetic backgrounds, including interactions with genome structure and epigenetic modifications. Advances in sequencing technologies will deepen understanding of heterosis mechanisms, enabling precise genetic improvements for complex ecological environments.

Conclusion

This study delved into the genetic landscape of 100 elite maize inbred lines from Sichuan province, China, using 5,261,175 high-density SNP markers. The analysis revealed rich genetic diversity among these lines, categorized into four distinct groups. By examining kinship relationships, genomic differentiation, and *Fst* values, we assessed the population structure and genetic diversity of these maize lines. Additionally, 900 hybrids generated from these inbred lines were evaluated for ten agronomic traits, showing significant heterosis patterns in yield-related traits, particularly in G1 ×G3, G1 ×G4, G2 ×G3, and G3 ×G4 combinations, which have been successfully utilized in commercial breeding. Furthermore, the study explored the correlation between hybrid offspring traits and the number of heterozygous SNPs, finding a general consistency with the trend of deleterious SNPs under most heterosis modes. In conclusion, these findings offer valuable insights into the potential of current maize germplasm in Sichuan province for enhancing hybrid maize breeding programs.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-06498-7.

Additional file 1: Supplement Figure 1. The molecular characteristics of 100 maize inbred lines genomes. (A) The SNPs characteristics. (B) The mutation type and number of SNPs. Supplement Figure 2. The coefficient of variation error rate. Supplement Figure 3. GO functional annotation analysis of selected candidate genes among different subpopulations. (A) Tropic-local A_Improved-tropic. (B) Tropic-local A_Tropic-local B. (C) Tropic-local A_Improved-local. (D) Improved-tropic_Tropic-local B. (E) Improved-tropic_Improved-local. (F) Improved-local_Improved-local. Supplement Figure 4. KEGG analysis of selected candidate genes among different subpopulations. (A) Tropic-local A_Improved-tropic. (B) Tropic-local A_Tropic-local B. (C) Tropic-local A_Improved-local. (D) Improved-tropic_Tropic-local B. (E) Improved-tropic_Improved-local. (F) Improved-local_Improved-local

Additional file 2. The information of sequencing.

Additional file 3. The pedigree of 100 maize inbred line based on neighbor-joining cluster analysis and structure analysis by using ADMIXTURE.

Additional file 4. The selected candidate intervals and candidate genes between different subgroups using the XP-CLR.

Additional file 5. The GO annotation analysis of selected candidate genes

Additional file 6. The KEGG analysis of selected candidate genes.

Additional file 7. The correlation analysis of 10 phenotypes.

Additional file 8. General combining ability for 10 tested traits in a maize multiple-hybrid population with 900 hybrids.

Additional file 9. Special combining ability for 10 tested traits in a maize multiple-hybrid population with 900 hybrids.

Additional file 10. The analysis of deleterious SNPs in 100 maize inbred lines.

Acknowledgements
We thank the team members of the Disruptive Materials and Methods Innovation in Maize Breeding in the Sichuan Science and Technology Support Project for providing the maize inbred lines. We thank Xuecai Zhang (International Maize and Wheat Improvement Center, CIMMYT) for providing the maize inbred lines from CIMMYT. We thank Shibin Gao (Maize Research Institute of Sichuan Agricultural University) for the comments in the manuscript revision and Bowen Luo (Maize Research Institute of Sichuan Agricultural University) for comments in the bioinformatics analysis.We would also like to thank Beijing Berry Genomics Co., Ltd. for assistance with the genome resequencing experiments and data analysis.

Authors' contributions
Peng Ma conducted the experimental design, data analysis and drafted the manuscript. Hua Zhang and Hongxia Shui were responsible for data collection and manuscript revision. Xuecai Zhang contributed some experimental materials from CIMMYT. Xiuquan Wang and Shibin Gao provided guidance on the revision of this manuscript and the design of experiment. Haiying Zhang, Zhi Nie, Tingqi Lu, Chunyan Qing, Qihua Pang, Wenzheng Pei, Hongmei Chen and Chenyan He participated in review and editing the manuscript. Bowen Luo was responsible for the bioinformatics analysis of this manuscript and providing comments on the revision. Dan He was involved in review, editing and project administration. All authors read and approved the final manuscript.

Funding
This research was supported by Sichuan Science and Technology Program (2023 NSFSC1175 and 2024 NSFSC0331). Sichuan Science and Technology Support Project (2021YFYZ0017 and 2021YFYZ0027). China Agricultural Research System (CARS-02).

Data availability
The raw sequence data reported in this study have been deposited in the Genome Sequence Archive [4] in National Genomics Data Center, China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA015527) that are publicly accessible at https://ngdc.cncb.ac.cn/gsa.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Mianyang Academy of Agricultural Sciences, Mianyang, Sichuan, China. [2]Crop Characteristic Resources Creation and Utilization Key Laboratory of Sichuan Province, Mianyang, Sichuan, China. [3]International Maize and Wheat Improvement Center (CIMMYT), State of Mexico, Texcoco, Mexico. [4]State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Chengdu, Sichuan, China. [5]Maize Research Institute, Sichuan Agricultural University, Chengdu, Sichuan, China. [6]Key Laboratory of Biology and Genetic Improvement of Maize in Southwest Region, Ministry of Agriculture, Chengdu, Sichuan, China. [7]Sichuan Academy of Agricultural Sciences, Biotechnology and Nuclear Technology Research Institute, Chengdu, China.

Received: 24 October 2024 Accepted: 1 April 2025
Published online: 15 April 2025

References

1. Bennetzen JL, Chandler VL, Schnable P. National Science Foundation-Sponsored Workshop Report. Maize Genome Sequencing Project Plant Physiology. 2001;127:1572–8.
2. Buckler ES, Gaut BS, McMullen MD. Molecular and functional diversity of maize. Curr Opin Plant Biol. 2006;9:172–6.
3. Charlesworth D, Willis JH. The genetics of inbreeding depression. Nat Rev Genet. 2009;10:783–96.
4. Chen TT, Chen X, Zhang SS, Zhu JW, Tang BX, Wang AK, Dong LL, Zhang ZW, Yu CX, Sun YL, Chi LJ, Chen HX, Zhai S, Sun YB, Lan L, Zhang X, Xiao JF, Bao YM, Wang YQ, Zhang Z, Zhao WM. The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types. Genomics Proteomics Bioinformatics. 2021;19:578–83.
5. Chen ZJ. Genomic and epigenetic insights into the molecular bases of heterosis. Nat Rev Genet. 2013;14:471–82.
6. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43:491–8.
7. Dwivedi SL, Heslop-Harrison P, Spillane C, McKeown PC, Edwards D, Goldman I, Ortiz R. Evolutionary dynamics and adaptive benefits of deleterious mutations in crop gene pools. Trends Plant Sci. 2023;28:685–97.
8. Fay J C, Yang J, Mezmouk S, Baumgarten A, Buckler E S, Guill K E, McMullen M D, Mumm R H, Ross-Ibarra J. Incomplete dominance of deleterious alleles contributes substantially to trait variation and heterosis in maize. PLOS Genetics. 2017,13(9): 1-21.
9. Hawliczek A, Borzęcka E, Tofil K, Alachiotis N, Bolibok L, Gawroński P, Siekmann D, Hackauf B, Dušinský R, Švec M, Bolibok-Brągoszewska H. Selective sweeps identification in distinct groups of cultivated rye (Secale cereale L.) germplasm provides potential candidate genes for crop improvement. BMC Plant Biol. 2023;23,323.
10. Hu Y, Colantonio V, Muller BSF, Leach KA, Nanni A, Finegan C, Wang B, Baseggio M, Newton CJ, Juhl EM, Hislop L, Gonzalez JM, Rios EF, Hannah LC, Swarts K, Gore MA, Hennen-Bierwagen TA, Myers AM, Settles AM, Tracy WF, Resende MFR, Jr. Genome assembly and population genomic analysis provide insights into the evolution of modern sweet corn. Nature Communication. 2021;12:1227.
11. Ji F Y, Ma Q G, Zhang W T, Liu J, Feng Y, Zhao P, Song X B, Chen J X, Zhang J P, Wei X, Zhou Y, Chang Y Y, Zhang P, Huang X H, Qiu J, Pei D. A genome variation map provides insights into the genetics of walnut adaptation and agronomic traits. Genome Biol. 2021;22:300.
12. Jiang SQ, Cheng Q, Yan J, Fu R, Wang XF. Genome optimization for improvement of maize breeding. Theor Appl Genet. 2019;133:1491–502.
13. Jiao YP, Peluso P, Shi JH, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei XH, Chin CS, Guill K, Regulski M, Kumari S, Olson A, Gent J, Schneider KL, Wolfgruber TK, May MR, Springer NM, Antoniou E, McCombie WR, Presting GG, McMullen M, Ross-Ibarra J, Dawe RK, Hastie A, Rank DR, Ware D. Improved maize reference genome with single-molecule technologies. Nature. 2017;546:524–7.
14. Leng Y F, Lv C X, Li L J, Xiang Y, Xia C, Wei R J, Rong T J, Lan H. Heterotic grouping based on genetic variation and population structure of maize inbred lines from current breeding program in Sichuan province, Southwest China using genotyping by sequencing (GBS). Molecular Breeding. 2019;39:38.
15. Li CH, Guan HH, Jing X, Li YY, Wang BB, Li YX, Liu XY, Zhang DF, Liu C, Xie XQ, Zhao HY, Wang YB, Liu JB, Zhang PP, Hu GH, Li GL, Li SY, Sun DQ, Wang XM, Shi YS, Song YS, Jiao CZ, Ross-Ibarra J, Li Y, Wang TY, Wang HY. Genomic insights into historical improvement of heterotic groups during modern hybrid maize breeding. Nature Plants. 2022;8:750–63.
16. Li C S, Xiang X L, Huang Y C, Zhou Y, An D, Dong J Q, Zhao C X, Liu H J, Li Y B, Wang Q, Du C G, Messing J, Larkins B A, Wu Y R, Wang W Q. Long-read sequencing reveals genomic structural variations that underlie creation of quality protein maize. Nat Commun. 2020a;11(1).
17. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26:589–95.
18. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.
19. Li T, Qu J Z, Tian X K, Lao Y H, Wei N N, Wang Y H, Hao Y C, Zhang X H, Xue J Q, Xu S T. Identification of Ear Morphology Genes in Maize (Zea

mays L.) Using Selective Sweeps and Association Mapping. Front Genet. 2020b;11,747.
20. Li YX, Li CH, Yang JP, Yang H, Cheng WD, Wang LM, Li FY, Li HY, Wang YB, Li SH, Hu GH, Liu C, Li Y, Wang TY. Genetic Dissection of Heterosis for Huangzaosi, a Foundation Parental Inbred Line of Maize in China. Scientia Agricultura Sinica. 2020;53:4113–26. (in Chinese).
21. Li Z, Sun J B, Hirsch C. Understanding environmental modulation of heterosis. Plant Breeding Reviews. 2022;46:219-237.
22. Liang T H, Hu Y, Xi N, Zhang M Y, Zou C Y, Ge F, Yuan G S, Gao S B, Zhang S Z, Pan G T, Ma L L, Lübberstedt T, Shen Y O. GWAS across multiple environments and WGCNA suggest the involvement of ZmARF23 in embryonic callus induction from immature maize embryos. Theor Appl Genet. 2023;136: 1-15.
23. Liu H J, Wang X, Q Xiao Y J, Luo J Y, Qiao F, Yang W Y, Zhang R Y, Meng Y J, Sun J M, Yan S J, Peng Y, Niu L Y, Jian L M, Song W, Yan J L, Li C H, Zhao Y X, Liu Y, Warburton M L, Zhao J R, Yan J B. CUBIC: an atlas of genetic architecture promises directed maize improvement. Genome Biol. 2020a;21:1-17.
24. Liu HJ, Wang Q, Chen MJ, Ding YH, Yang XR, Liu J, Li XH, Zhou CC, Tian QL, Lu YQ, Fan DL, Shi JP, Zhang L, Kang CB, Sun MF, Li FY, Wu YJ, Zhang YZ, Liu BS, Zhao XY, Feng Q, Yang JL, Han B, Lai JS, Zhang XS, Huang XH. Genome-wide identification and analysis of heterotic loci in three maize hybrids. Plant Biotechnol J. 2019;18:185–94.
25. Liu ZW, Zhu HL, Zhou JH, Jiang SJ, Wang Y, Kuang J, Ji Q, Peng J, Wang J, Gao L, Bai MZ, Jian JB, Ke WD. Resequencing of 296 cultivated and wild lotus accessions unravels its evolution and breeding history. Plant J. 2020;104:1673–84.
26. Lozano R, Gazave E, dos Santos JPR, Stetter MG, Valluru R, Bandillo N, Fernandes SB, Brown PJ, Shakoor N, Mockler TC, Cooper EA, Taylor Perkins M, Buckler ES, Ross-Ibarra J, Gore MA. Comparative evolutionary genetics of deleterious load in sorghum and maize. Nature Plants. 2021;7:17–24.
27. Luo B W, Zhang G D, Yu T, Zhang C, Yang G H, Luo X F, Zhang S H, Guo J Y, Zhang H Y, Zheng H, Tang Z R, Li Q L, Lan Y Z, Ma P, Nie Z, Zhang X, Liu D, Wu L, Gao D J, Gao S Q, Su S Z, Guo J, Gao S B. Genome-wide association studies dissect low-phosphorus stress response genes underling field and seedling traits in maize. Theor Appl Genet. 2024;137:1-17.
28. Lv M, Deng C, Li X Y, Zhao X D, Li H M, Li Z M, Tian Z Q, Leonard A, Jaqueth J, Li B L, Hao J J, Chang Y X, Ding J Q. Identification and fine-mapping of RppCML496, a major QTL for resistance to Puccinia polysora in maize. Plant Genome. 2020;14:1-7.
29. Pan GT, Yang KC. Facing toward challenges and corresponding strategies for maize breeding in southwestern region of China. Acta Agron Sin. 2013;38:1141–7. (in Chinese).
30. Pan G T, Yang K C, Li L J, Gao S B, Shen Y O, Tan D F, Lu H Z. Breeding and application of an elite inbred line SCML0849 in a new round of maize breeding in Southwest China. J Maize Sci. 2024;32:1-15.
31. Shu GP, Cao GQ, Li NN, Wang AF, Wei F, Li T, Yi L, Xu YB, Wang YB. Genetic variation and population structure in China summer maize germplasm. Sci Rep. 2021;11:8012.
32. Tang QY, Zhang CX. Data Processing System (DPS) software with experimental design, statistical analysis and data mining developed for use in entomological research. Insect Science. 2012;20:254–60.
33. Teng WT, Cao QS, Chen YH, Liu XH, Jing XQ, Zhang FJ, Li JS. Analysis of maize heterotic groups and patterns during past decade in China. Agricultural Sciences in China. 2004;3:481–9.
34. Vaser T, Adusumalli S, Leng S, Sikic M, Ng P. SIFT missense predictions for genomes. Nature protocols. 2015;11:1-9.
35. Wang BB, Hou M, Shi JP, Ku LX, Song W, Li CH, Ning Q, Li X, Li CY, Zhao BB, Zhang RY, Xu H, Bai ZJ, Xia ZC, Wang H, Kong DX, Wei HB, Jing YF, Dai ZY, Wang HHL, Zhu XY, Li CH, Sun X, Wang SS, Yao W, Hou GG, Qi Z, Dai H, Li XM, Zheng HK, Zhang ZX, Li Y, Wang TY, Jiang TJ, Wan ZM, Chen YH, Zhao JR, Lai JS, Wang HY. De novo genome assembly and analyses of 12 founder inbred lines provide insights into maize heterosis. Nat Genet. 2023;55:312–23.
36. Wang BB, Lin ZC, Li X, Zhao YP, Zhao BB, Wu GX, Ma XJ, Wang H, Xie YR, Li QQ, Song GS, Kong DX, Zheng ZG, Wei HB, Shen RX, Wu H, Chen CX, Meng ZD, Wang TY, Li Y, Li XH, Chen YH, Lai JS, Hufford MB, Ross-Ibarra J, He H, Wang HY. Genome-wide selection and genetic improvement during modern maize breeding. Nat Genet. 2020;52:565–71.
37. Wang K, Li MY, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164–e164.
38. Wang N, Yuan YB, Wang H, Yu DS, Liu Y, Zhang A, Gowda M, Nair SK, Hao ZF, Lu YL, San Vicente F, Prasanna BM, Li XH, Zhang XC. Applications of genotyping-by-sequencing (GBS) in maize genetics and breeding. Scirntific Reports. 2020;10:16308.
39. Wei T, van Treuren R, Liu XJ, Zhang ZW, Chen JJ, Liu Y, Dong SS, Sun PN, Yang T, Lan TM, Wang XG, Xiong ZQ, Liu YQ, Wei JP, Lu HR, Han SP, Chen JC, Ni XM, Wang J, Yang HM, Xu X, Kuang HH, van Hintum T, Liu X, Liu H. Whole-genome resequencing of 445 Lactuca accessions reveals the domestication history of cultivated lettuce. Nat Genet. 2021;53:752–60.
40. Wright S. The Genetical Structure of Populations. Ann Eugen. 1951;15:323–54.
41. Xiao Y J, Jiang S Q, Cheng Q, Wang X Q, Yan J, Zhang R Y, Qiao F, Ma C, Luo J Y, Li W Q, Liu H J, Yang W Y, Song W H, Meng Y J, Warburton M L, Zhao J R, Wang X F, Yan J B. The genetic mechanism of heterosis utilization in maize improvement. Genome Biology. 2021;22:1-29.
42. Yi Q, Liu Y H, Hou X B, Zhang X G, Li H, Zhang J J, Liu H M, Hu Y F, Yu G W, Li Y P, Wang Y B, Huang Y B. Genetic dissection of yield-related traits and mid-parent heterosis for those traits in maize (Zea mays L.). BMC Plant Biology. 2019;19:392
43. Yong H J, Zhang F Y, Tang J, Yang Z Y, Zhao X Z, Li M S, Zhang D G, Hao Z F, Weng J F, Li X H. Breeding potential of inbred lines derived from five maize (Zea mays L.) populations. Euphytica. 2018;215:1-12.
44. Zhang X, Ren ZY, Luo BW, Zhong HX, Ma P, Zhang HY, Hu HM, Wang YK, Zhang HK, Liu D, Wu L, Nie Z, Zhu YH, He WZ, Zhang SZ, Su S, Shen YO, Gao SB. Genetic architecture of maize yield traits dissected by QTL mapping and GWAS in maize. The Crop Journal. 2022;10:436–46.
45. Zhang X, Zhang H, Li L J, Lan H, Ren Z Y, Liu D, Wu L, Liu H, Jaqueth J, Li B L, Pan G T, Gao S B. Characterizing the population structure and genetic diversity of maize breeding germplasm in Southwest China using genome-wide SNP markers. BMC Genomics. 2016;17:1-16.
46. Zhu M, Cheng Y Y, Wu S L, Huang X H, Qiu J. Deleterious mutations are characterized by higher genomic heterozygosity than other genic variants in plant genomes. Genomics. 2022;114(2):1-9.

Publisher's Note

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