Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Molecular clustering and genetic distance do not predict yield combining ability in a tropical × temperate maize testcross panel

Yuanyuan Yan
Hebei Academy of Agriculture and Forestry Sciences
Rui Guo
Hebei Academy of Agriculture and Forestry Sciences
Haihong Fan
Hebei Academy of Agriculture and Forestry Sciences
Xinghua Li
Hebei Academy of Agriculture and Forestry Sciences
Jianfeng Wei
Hebei Academy of Agriculture and Forestry Sciences
Liwei Wang
Hebei Academy of Agriculture and Forestry Sciences
Liang Song
Hebei Academy of Agriculture and Forestry Sciences
Thanda Dhliwayo
Centro Internacional de Mejoramiento de Maíz Y Trigo
Jianghao Wang
Hebei Academy of Agriculture and Forestry Sciences
Wei Song
sw1717@126.com

Hebei Academy of Agriculture and Forestry Sciences

Research Article

Keywords: tropical maize germplasm, combining ability, heterotic group, modified Rogers' distance, SNP marker

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11449732/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Molecular clustering and genetic distance do not predict yield combining ability in a tropical × temperate maize testcross panel

Yuanyuan Yan [a+], Rui Guo [b+], Haihong Fan [b], Xinhua Li [b], Jianfeng Wei [b], Liwei Wang [a], Liang Song [a], Thanda Dhliwayo [c], Jianghao Wang [b*] and Wei Song [b*]

[a] Hebei Key Laboratory of Crop Cultivation Physiology and Green Production, Institute of Cereal and Oil Crops, Hebei Academy of Agriculture and Forestry Sciences, Shijiazhuang, China

[b] Key Laboratory of Crop Genetics and Breeding of Hebei Province, Institute of Cereal and Oil Crops, Hebei Academy of Agriculture and Forestry Sciences, Shijiazhuang, China

[c] International Maize and Wheat Improvement Center (CIMMYT), State of Mexico, Texcoco, Mexico

Correspondence: Jianghao Wang and Wei Song, Institute of Cereal and Oil Crops, Hebei Academy of Agriculture and Forestry Sciences, Shijiazhuang 050035, China.

Email: jhw78@163.com, sw1717@126.com

Tel: +86 0311 8767 0662

Fax: +86 0311 8767 0663

Yuanyuan Yan, Assistant Research Professor, PhD Candidate

Rui Guo, Associate Research Professor, PhD Candidate

Yuanyuan Yan and Rui Guo contributed equally to this work.

Abstract

Maize hybrid breeding relies on reliable estimates of parental combining ability. Tropical germplasm can expand the genetic diversity available for temperate maize improvement, yet the utility of SNP-derived candidate heterotic groups and genome-wide modified Rogers' distance (MRD) for predicting tropical × temperate testcross performance remains elusive. We adopted an NC-II line × tester mating design comprising 46 CIMMYT tropical inbred lines and four temperate tester lines adapted to the Yellow-Huai-Hai Plain. This crossing scheme generated 184 testcross hybrids, which were evaluated across two field sites within a single agro-ecological zone in 2025. We characterized population genetic structure and defined SNP-based candidate heterotic groups and applied mixed-model analyses to examine whether genome-wide average MRD could predict specific combining ability (SCA) for grain yield and F_1 grain yield, while accounting for non-independence among related testcrosses. Genome-wide average MRD showed no significant linear association with yield SCA ($P = 0.742$) or F_1 grain yield ($P = 0.624$). No tester-specific MRD-trait relationships were detected among the four temperate testers. Within this panel and the tested sites, genome-wide SNP-based grouping and average pairwise genetic distance exhibited limited linear predictive capacity for yield combining ability and hybrid performance. These findings offer practical implications for parental selection in tropical × temperate maize crosses. In particular, direct field phenotyping of testcrosses remains essential for evaluating combining ability, whereas SNP markers are valuable for germplasm characterization and diversity management.

Key words: tropical maize germplasm; combining ability; heterotic group; modified Rogers' distance; SNP marker

Introduction

Maize (*Zea mays* L.) is one of the most important cereal crops globally, providing staple food, animal feed and industrial raw materials. Narrowing genetic base is a long-standing bottleneck in temperate maize breeding. Introduction of tropical inbred lines from CIMMYT serves as a promising approach to broaden the genetic diversity of temperate breeding pools [1,2]. The Yellow–Huai–Hai (YHH) Plain is the dominant summer maize production region in China. In this region, high planting density together with increased post-anthesis drought and heat stress has aggravated yield instability [3,4]. Local elite temperate germplasm harbors limited allelic variation for stress tolerance, while tropical and subtropical inbred lines possess abundant favorable alleles for drought escape, heat tolerance, disease and pest resistance [5,6]. However, large-scale direct utilization of tropical germplasm is greatly restricted by inherent photoperiod sensitivity, serious flowering delay, and unclear heterotic patterns when crossed with local YHH testers [7,8]. Photoperiod sensitivity is regulated by CCT family genes, which offer valuable molecular targets for pre-adaptation of tropical germplasm to local photoperiod conditions [9]. Nevertheless, few studies have focused on the combining ability and heterotic relationships between CIMMYT tropical lines and widely used temperate testers of the YHH Plain under local field environments.

Heterotic-group classification and combining-ability evaluation are two cornerstones of hybrid maize breeding [10,11]. General combining ability (GCA) reflects

70 additive genetic effects, whereas specific combining ability (SCA) describes non-
71 additive interactions, including dominance and epistasis, between parental inbred lines
72 [2]. The line × tester (NC-II) incomplete diallel mating design allows simultaneous
73 estimation of GCA for a large set of female inbred lines and a small number of testers,
74 together with SCA for all hybrid combinations. This design has been widely used to
75 evaluate genetic performance in tropical-by-temperate maize crosses [12]. Recent studies
76 have integrated genome-wide SNP genotyping with NC-II mating designs to explore
77 the potential of molecular markers for predicting combining ability [13]. Nevertheless,
78 multi-site NC-II evaluations with large tropical germplasm panels and a limited set of
79 temperate testers are still scarce for the Yellow–Huai–Hai Plain of China. Therefore,
80 this study adopted an NC-II design to phenotype tropical-by-temperate testcross
81 hybrids across two sites, characterize GCA and SCA patterns for major agronomic traits,
82 and evaluate the predictive potential of genome-wide modified Rogers' distance (MRD)
83 for yield SCA.

84 High-throughput SNP genotyping, represented by widely adopted 50 K and 60 K
85 maize SNP arrays, has become a routine tool for dissecting population genetic structure
86 and conducting heterotic-group classification [14,15]. However, several studies have
87 observed only weak agreement between SNP-derived molecular clustering and
88 heterotic grouping defined by combining ability [16-18]. Marker-based genetic distance
89 also often shows inconsistent prediction for heterosis [19]. Genome-wide genetic
90 distance often fails to fully capture non-additive parental interactions, which restricts
91 its capacity to predict heterotic performance. To date, few investigations have examined

whether genome-wide marker-based genetic distance can reliably predict yield SCA in tropical-by-temperate maize testcross populations, particularly for materials targeted to the YHH agroecological zone.

To date, systematic genetic characterization of the CML486–CML615 tropical inbred series remains limited, particularly for breeding applications targeting the YHH Plain. Most phenotypic assessments of these tropical lines have been performed under southern subtropical environments, limiting their direct deployment in temperate breeding programs. Few studies have adopted the four widely used YHH temperate testers (Z58, Chang7-2, Ji1877, and Ji42) within a multi-site NC-II mating design combined with genome-wide SNP genotyping to evaluate tropical germplasm. Although previous reports have documented weak associations between genome-wide molecular distance, heterosis and hybrid performance, few studies have specifically assessed the predictive power of MRD for yield SCA using an NCII panel of CIMMYT tropical lines × Chinese YHH temperate testers. Therefore, this study aimed to estimate GCA and SCA effects for grain yield and sink-related traits across 184 tropical-by-temperate testcross hybrids derived from 46 tropical inbred lines and four temperate testers at two YHH field sites. Mixed linear models were applied to partition additive and non-additive genetic variance components. Based on combining ability estimates, we identified tropical donor lines and hybrid combinations with favorable performance under the tested environments. Using genome-wide SNP markers, we further characterized population genetic structure, defined SNP-derived candidate heterotic groups, and evaluated the predictive potential of MRD for grain yield and yield SCA.

We hypothesized that genome-wide MRD would show limited predictive power for yield SCA, because yield heterosis is largely determined by locus-specific allelic interactions and genotype-by-site interactions. These results provide insights into parental selection for tropical–temperate hybridization and enrich our understanding of combining ability within this inter-ecotype germplasm panel.

Materials and methods

Plant materials

Forty-six tropical maize inbred lines belonging to the CML486–CML615 series and introduced from CIMMYT (Mexico) were designated as female parents (test lines). Four core backbone inbred lines extensively deployed in maize breeding within the YHH region, Zheng58 (Z58), Chang7-2, Ji1877, and Ji42, were used as male testers. The widely planted commercial hybrid Zhengdan 958 (ZD958) was included as the field check for comparison of agronomic performance and grain yield. Origins and ecotypes of the 46 tropical inbred lines are listed in Additional file 1.

Experimental design and field management

An NCII (line × tester) mating design was adopted to develop testcross combinations at the Sanya National Breeding and Multiplication Base, Hainan, during the winter of 2024. In 2025, the testcross progenies were evaluated at two experimental sites in Gaocheng and Luancheng Districts, Shijiazhuang, Hebei Province. The experiment was arranged in a randomized complete block design with two replications. Each plot consisted of two rows with a length of 5 m and row spacing of 0.6 m, resulting in a plot area of 6 m^2. Plants were spaced 0.25 m within rows, yielding a planting density

of 67,500 plants ha^{-1}. The entire plot area was harvested for grain yield assessment. Standard local agronomic practices were applied for field management.

Phenotypic traits and yield measurement

Phenological and plant architecture traits were systematically surveyed in this study. Days to tasseling (DTT), days to anthesis (DTA), and days to silking (DTS) were recorded as days from sowing to the corresponding stage. At physiological maturity, plant height (PH) and ear height (EH) were determined on ten representative plants per plot using a calibrated measuring pole.

For ear morphological traits, five uniform and representative ears were harvested from each plot and air-dried at ambient temperature until a constant weight was achieved. Ear length (EL), ear diameter (ED), kernel row number (KRN), barren ear tip length (BTL), and total kernel number (TKN) were quantified using an automatic maize ear analysis system.

For grain yield evaluation, all ears harvested from each plot were manually shelled, air-dried, and weighed. Grain yield was adjusted to 14% moisture on a wet basis and converted to kilograms per hectare (Mg ha^{-1}) using the harvested plot area.

Missing plant correction followed standard field trial specifications. Plots with a missing plant rate ≤ 5% were used directly without correction. For plots with a missing rate between 5% and 10%, plot yield was corrected based on the average single-plant yield of surviving plants within the same plot. Plots with a missing rate exceeding 10% were excluded from statistical analyses. These thresholds follow standard practice for maize yield trials in this region. Because the vast majority of plots fell within the ≤ 5%

158 category and required no correction, and only a small number of plots required
159 adjustment, the influence of stand correction on the overall yield dataset was limited.
160 All yield values are reported as Mg ha^{-1} throughout the manuscript.

161 **SNP genotyping and population genomic analysis**

162 A total of 46 tropical and four temperate maize inbred lines were grown at the
163 Dishang Experimental Station, Shijiazhuang, Hebei, China. Genotyping was performed
164 as described in our previous study [20]. Briefly, leaf tissue pooled from 15 plants per
165 inbred line was submitted to MolBreeding Biotechnology Co., Ltd. for genotyping
166 using the Maize 10K GBTS BeadChip. SNP variants were called against the B73
167 RefGen_v5 assembly, and 10,005 high-confidence SNPs were retained following
168 quality filtering. Principal component analysis (PCA) was performed on the filtered
169 SNP matrix using TASSEL software to evaluate population stratification, and the first
170 two principal components were applied to visualize genetic differentiation patterns
171 among accessions. A neighbor-joining (NJ) phylogenetic tree was further constructed
172 based on pairwise genetic distances [21].

173 Heterotic group classification and population genetic structure were assessed
174 using STRUCTURE v2.3.4 with the 10 005 high-quality filtered SNPs [22,23]. The
175 correlated-allele-frequency model was used; the burn-in period was set to 10 000,
176 followed by 10 000 post-burn-in MCMC replicates. K was varied from 1 to 10, with
177 five independent replicate runs for each K. Label-switching across independent runs
178 was corrected using CLUMPP. Mean ancestry membership probabilities obtained from
179 aligned replicate runs were used to generate the main STRUCTURE bar plot (Fig. 5).

The optimal K was inferred using the ΔK method [24], which computes the second-order rate of change of log-likelihood values across successive K values. The K with the highest ΔK was considered the optimal number of genetic subgroups [25].

Population-structure analyses were performed at two hierarchical levels. First, all 50 inbred lines (46 tropical and four temperate) were jointly analyzed to detect the primary genetic partition across K = 1–10. The ΔK statistic supported K = 3 as the optimal value. At K = 3, the four temperate testers formed one genetic group, while tropical germplasm was partitioned into two major clusters. Second, to resolve fine-scale genetic structure within tropical germplasm, the 46 tropical inbred lines were analyzed separately across K = 1–10. ΔK identified K = 4 as the optimal cluster number for this tropical subset.

Genomic grouping for the 46 tropical inbred lines was finalized by integrating outputs from STRUCTURE, PCA, and phylogenetic tree analyses. In addition, genomic-based clustering results were cross-validated against phenotypic grouping derived from multi-trait GCA performance.

Pairwise MRD [26] across all parental lines (46 female inbred lines and four male testers) was calculated using genome-wide SNP markers. The resulting MRD matrix was subsequently used to explore the relationships between parental genomic divergence and hybrid performance, including grain yield and yield SCA.

Statistical and genetic analysis

Traditional ANOVA for NC-II mating design

Raw phenotypic data were compiled in Microsoft Excel 2019. Multi-site data from

the NC-II (line × tester) mating design were analyzed using two complementary approaches in R (v4.3.1).

As an exploratory descriptive analysis, a traditional fixed-effects analysis of variance (ANOVA) was performed to partition phenotypic variation for each trait. Site, block nested within site, female line, male tester, line × tester interaction, and their interactions with site were all treated as fixed effects. F-statistics and significance levels were derived to describe the significance of each variation source (Table 2). All formal genetic inference and variance component estimation were based on the REML mixed models described in the next section.

Mixed-model analysis, variance components and genetic parameters

Mixed linear models were fitted via restricted maximum likelihood (REML) using the *sommer* package (v4.3.1) in R (v4.3.1) to estimate variance components and best linear unbiased predictions (BLUPs) of combining ability for the 12 agronomic traits (Tables 3 and 5). The block structure for fixed effects was consistent with that used in the ANOVA analysis to ensure design comparability.

In these models, site and block nested within site were specified as fixed effects, given that the two experimental sites represented our target test sites. Female GCA, male GCA, SCA, and their respective interactions with site were treated as random effects. The full linear mixed model was:

$$y_{ijkl} = \mu + E_i + R_{j(i)} + F_k + M_l + (FM)_{kl} + (EF)_{ik} + (EM)_{il} + (EFM)_{ikl} + e_{ijkl}$$

where y_{ijkl} is the phenotypic observation for the cross derived from the k^{th} female line × l^{th} male tester in the j^{th} block within the i^{th} site; μ is the overall

population mean; E_i is the fixed effect of the i^{th} site; $R_{j(i)}$ is the fixed effect of the j^{th} block nested within the i^{th} site; F_k and M_l are random GCA effects of the female and male parents, respectively; $(FM)_{kl}$ is the random SCA effect; $(EF)_{ik}$, $(EM)_{il}$, and $(EFM)_{ikl}$ represent random interaction effects of site with female GCA, male GCA, and SCA; and e_{ijkl} denotes the residual term.

All random genetic terms were assumed mutually independent and normally distributed with zero means. Residual variances were allowed to differ across sites using the diagonal residual variance structure by environmental group implemented in $sommer$.

Two candidate models with full random effect structures were fitted per trait, differing only in residual variance specification: (i) homogeneous residual model, with constant residual variance across sites; (ii) heterogeneous residual model, with site-specific residual variances. The model with the lowest Akaike Information Criterion (AIC) was retained for downstream analysis.

Both one-tailed Wald tests and likelihood ratio tests (LRT) were applied to assess the statistical significance of each random variance component. BLUPs for GCA and SCA effects were extracted directly from the optimal model solution.

Total genetic variance, total genotype-by-site (G×S) interaction variance, entry-mean phenotypic variance, and broad-sense heritability (H^2) at both entry-mean and plot levels were derived from the REML variance-component estimates. Standard errors for all derived parameters were calculated using the delta method.

Under NCII genetic assumptions, variance components correspond to classical

quantitative genetic expectations. However, SCA variance may also capture epistatic effects, and should be interpreted as total non-additive variance rather than pure dominance variance. Given only four male testers in this experiment, male-associated variance components have lower estimation precision; female and male GCA variances are therefore reported separately, and pooled estimates should be interpreted with caution [27].

Predictive ability of MRD for F_1 grain yield and yield SCA

We used mixed-effects regression to test the linear predictive relationship between parental MRD and hybrid performance, using two response variables: the entry-mean F_1 grain yield (across the two sites) and the estimated SCA effects for grain yield, both extracted from the above-described NC-II mixed-model analysis. Entry-mean phenotypes and SCA point estimates were used to avoid the shrinkage bias associated with genomic prediction, thereby providing a more objective evaluation of the predictive ability of MRD. This mixed-model framework accounts for non-independence among the 184 testcrosses originating from shared female inbred lines and shared male testers. The mixed-effect regression model was:

$$y_{kl} = \mu + \mathrm{MRD}_{kl} + F_k + M_l + e_{kl}$$

where y_{kl} is either the BLUP for F_1 grain yield or the BLUP for yield SCA of the cross between female line k and male tester l; μ is the overall intercept; MRD_{kl} is the pairwise modified Rogers' distance between female line k and male tester l, fitted as a fixed effect; F_k is the random effect of female inbred line k; M_l is the random effect of male tester l; and e_{kl} is the residual term. Because the estimated SCA effects

carry their own sampling variance from the first-stage NC-II model, the residual term ekl the SCA regression encompasses both the prediction error and the SCA estimation error. To evaluate whether the MRD–trait relationship varied across tester backgrounds, the MRD × tester interaction term was further added to the base model.

Models were fitted by REML using the $sommer$ package (v4.3.1) in R (v4.3.1). LRT were used to assess statistical significance for the MRD main effect and the MRD × tester interaction. Variance-partitioning estimates and tester-specific slopes derived from mixed-effects models were exported and summarized in Supplementary Additional file 6.

For descriptive visualization only, separate ordinary least-squares (OLS) regressions were fitted within each tester group to generate regression lines displayed in Additional file 7 (F_1 grain yield) and Additional file 8 (yield SCA). These OLS fits were not used for statistical inference because they ignored the non-independence of testcrosses sharing parental lines. Residual diagnostics including Q–Q plots and residual-versus-fitted-value plots were applied to verify the mixed-model assumptions of normality and homoscedasticity (Additional file 9).

Data visualization

All Figs in this study were generated using Origin 2024 and the ggplot2 package (v3.4.2) in R (v4.3.1). Conventional graphical plots, including bar charts, line charts, and scatter plots, were produced in Origin 2024. Genomic and population genetic graphs, including phylogenetic trees, STRUCTURE bar plots, and PCA scatter plots, were visualized using the ggplot2 package.

Results

Descriptive statistics of agronomic traits

Descriptive statistics were calculated for 12 agronomic traits of the 184 testcross hybrids (Table 1). Wide phenotypic variation was detected for all evaluated traits. Flowering-related traits had the lowest coefficients of variation (CVs), with CV values ranging from 4.7% to 5.1% for DTT, DTA and DTS. These CVs reflect variation among hybrids, not within-hybrid flowering synchrony.

For plant-architecture traits, PH averaged 303.22 cm (CV = 8.64%) and EH averaged 147.91 cm (CV = 15.45%). The CV of EH was approximately 1.8-fold that of PH, indicating greater phenotypic plasticity of EH relative to PH.

Ear-related traits exhibited broad-ranging phenotypic variation. ED showed relatively low variation, while KRN, EL and KPR displayed moderate variation. TKN, BTL and grain yield displayed the highest levels of phenotypic variation, reflecting large differences in yield potential among these testcross combinations.

Overall, the testcross population exhibited substantial variation for agronomic traits, particularly BTL, TKN and grain yield. This variation provides the basis for subsequent combining-ability analysis and hybrid screening.

307 **Table 1 The descriptive statistics of major traits**

Trait	DTT (d)	DTA (d)	DTS (d)	PH (cm)	EH (cm)	KRN (rows)	KPR (kernels)	TKN (kernels)	EL (cm)	ED (cm)	BTL (cm)	Yield (Mg ha^{-1})
Max	62.00	65.00	63.00	397.41	221.47	20.55	38.84	649.67	24.60	7.37	11.89	13.16
Min	47.00	50.00	50.00	171.18	81.19	10.91	15.40	158.11	11.40	3.82	0.70	0.43
Mean	54.17	57.28	57.24	303.22	147.91	15.23	28.92	438.41	18.62	4.46	3.42	5.17
SD	2.77	2.70	2.87	26.20	22.85	1.69	4.16	88.35	2.14	0.45	1.50	2.13
CV (%)	5.10	4.70	5.02	8.64	15.45	11.09	14.37	20.15	11.52	7.78	43.81	41.28

308 Note: DTT, days to tasseling; DTA, days to anthesis; DTS, days to silking; PH, plant height; EH, ear height; KRN, kernel row number; KPR,
309 kernels per row; TKN, total kernel number; EL, ear length; ED, ear diameter; BTL, barren tip length; Yield, grain yield per plot. The same below.

310 **Table 2 Analysis of variance about 12 major traits (*F-values*)**

Source of variation	Df	DTT	DTA	DTS	PH	EH	KRN	KPR	TKN	EL	ED	BTL	Yield
Site	1	518.92 **	206.66 *	161.72 **	497.32 ns	7866.12 **	5.00 ns	23.85 ns	38341.96 ns	58.48 ns	0.0024 ns	0.42 ns	0.45 ns
Block (Site)	2	1.01 ns	2.11 ns	1.15 ns	380.37 ns	21.38 ns	2.14 ns	4.07 *	3763.18 ns	123.81 ***	0.12 ns	21.28 ***	52.72 ***
Genotypes	183	19.59 ***	21.69 ***	25.21 ***	1831.34 ***	1593.57 ***	5.05 *	40.80 **	13324.23 *	7.79 *	0.34 *	4.11 ns	11.48 ***
Female line	45	55.22 ***	64.08 ***	76.05 ***	4725.05 ***	3288.52 ***	6.77 *	49.53 *	15068.55 ns	8.78 ns	0.53 **	5.35 ns	18.38 ***
Male tester	3	119.12 *	118.19 **	71.05 ns	15363.22 *	28668.59 **	42.36 *	195.65 ns	115040.69 ns	34.05 ns	2.29 ns	12.27*	103.05 *
Female × Male	135	5.50 **	5.42 ***	7.25 ***	566.06 ns	426.92 *	3.65 ns	34.45 *	10482.43 ns	6.88 *	0.23 ns	3.51 ns	7.15 *
Genotype × Site (G×S)	183	4.24 ***	3.71 ***	4.11 ***	599.03 ***	328.86 ***	3.59 ***	26.32 ***	9626.90 ***	5.97 ***	0.25 ***	3.30 ***	5.48 ***
Line × Site	45	6.87 ***	5.90 **	4.99 ns	740.92 ns	398.37 ns	3.58 ns	29.88 ns	12108.98 ns	8.64 *	0.22 ns	4.67 *	6.68 ns
Tester × Site	3	8.32 ns	2.03 ns	15.08 **	1048.52 ns	919.86 *	2.08 ns	41.74 ns	23169.18 *	7.29 ns	0.32 ns	1.09 ns	4.83 ns
Line × Tester × Site	135	3.27 ***	3.01 ***	3.57 ***	541.74 ***	292.55 ***	3.63 ***	24.78 ***	8498.60 ***	5.05 ***	0.25 ***	2.90 ***	5.09 ***
Error (Residuals)	366	2.02	1.31	1.46	160	65.38	1.38	1.06	4073.98	1.52	0.11	0.68	0.37

311 Note: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Combined analysis of variance for agronomic traits across two sites

Preliminary fixed-effects analysis of variance across the two sites revealed substantial genotypic variation among hybrids for most measured traits (Table 2). Site effects were significant for DTT, DTA, DTS, and EH, but not for PH, KRN, KPR, TKN, EL, ED, BTL, or grain yield. Block-within-site effects were significant only for KPR, EL, BTL, and grain yield, indicating that the randomized complete block design effectively accounted for spatial heterogeneity in these traits.

The overall genotypic effect was highly significant ($P < 0.001$) for most traits, whereas BTL showed no significant genotypic variation. When partitioned, female GCA effects were significant for nine of the twelve traits, and male GCA effects reached significance for seven traits. The line × tester (SCA) interaction was significant for seven traits, with significance levels ranging from $P < 0.05$ to $P < 0.001$. Note that these F-statistics and *P*-values are reported for exploratory phenotypic screening only; estimations of variance components, BLUPs of combining ability, and calculations of genetic parameters were all performed via REML mixed models (Section 2.5).

Because only four testers were used, F-statistics for male GCA cannot be interpreted as direct measures of additive genetic contributions. Significant line × tester effects indicate tester-dependent hybrid performance, but do not by themselves confirm non-additive genetic mechanisms.

Genotype-by-site (G×S) interactions were highly significant ($P < 0.001$) for all twelve traits, with particularly strong G×S effects for grain yield and TKN, demonstrating that hybrid performance differed markedly between the two

experimental sites. Therefore, both combining ability and cross-site phenotypic stability should be considered when selecting superior hybrids.

Genetic parameter estimation

Variance components and genetic parameters were estimated based on the best-fitting mixed-effect models (Table 3a, 3b). The results revealed contrasting genetic architectures underlying the 12 evaluated agronomic traits. For most traits including DTT, DTA, DTS, PH, EH, and grain yield, variance components for female GCA ($V_{GCA\ (female)}$) and SCA (V_{SCA}) were statistically significant. By contrast, $V_{GCA\ (male)}$ was generally small and frequently non-significant (DTS, PH). Notably, for ED and BTL, only $V_{GCA\ (female)}$ reached significance ($P < 0.01$ and $P < 0.05$, respectively). Several variance-component estimates for TKN exhibited very large standard errors, indicating low estimation precision for this trait.

The SCA-by-site interaction ($V_{SCA \times site}$) was highly significant ($P < 0.001$) for all twelve traits. In comparison, GCA-by-site interactions were mostly non-significant: $V_{GCA \times site\ (female)}$ was significant only for DTT and DTA, whereas $V_{GCA \times site\ (male)}$ was significant only for DTS.

Broad-sense heritability estimated on an entry-mean basis ($H^2_{entry\text{-}mean}$) varied substantially across traits (Table 3b). High heritability ($H^2_{entry\text{-}mean}>0.65$) was observed for DTS, DTA, EH, DTT, and PH, indicating these traits are under strong genetic control. Grain yield exhibited moderate heritability ($H^2_{entry\text{-}mean} = 0.54$). Several traits, namely KRN, ED, KPR, TKN, EL, and BTL, showed relatively low entry-mean broad-sense heritability, which can be attributed to large residual and genotype-by-site variances

356 relative to total genetic variance. Furthermore, plot-level broad-sense heritability
357 ($H^2_{\text{plot-level}}$) followed a similar trait-ranking pattern but yielded lower values overall,
358 ranging from 0.10 (BTL) to 0.66 (DTS). This pattern corresponds to the well-known
359 expectation that heritability decreases when evaluated on individual-plot observations
360 compared with genotype entry means.

Table 3 Variance components and derived genetic parameters for 12 agronomic traits estimated from optimal mixed models.

(a) Variance components

Genetic Parameter	$V_{GCA\ (female)}$	$V_{GCA\ (male)}$	V_{SCA}	$V_{GCA \times site(female)}$	$V_{GCA \times site(male)}$	$V_{SCA \times site}$	$V_{Residual}$
DTT	2.88 ± 0.73 ***	0.59 ± 0.53 *	0.56 ± 0.19 **	0.45 ± 0.19 **	0.05 ± 0.07 ns	0.63 ± 0.21 ***	2.02 ± 0.15
DTA	3.48 ± 0.85 ***	0.61 ± 0.52 **	0.61 ± 0.19 ***	0.36 ± 0.16 **	0.00 ± 0.03 ns	0.84 ± 0.19 ***	1.31 ± 0.10
DTS	4.21 ± 1.01 ***	0.29 ± 0.33 ns	0.92 ± 0.25 ***	0.18 ± 0.14 ns	0.13 ± 0.14 *	1.06 ± 0.22 ***	1.46 ± 0.11
PH	242.50 ± 62.00 ***	76.37 ± 67.18 *	6.17 ± 24.06 ns	22.07 ± 20.83 ns	5.23 ± 9.18 ns	195.19 ± 33.55 ***	159.18 ± 12.34
EH	173.28 ± 44.06 ***	149.45 ± 126.75 *	33.48 ± 15.68 *	13.63 ± 11.45 ns	6.79 ± 8.14 ns	113.05 ± 17.88 ***	65.46 ± 4.92
KRN	0.19 ± 0.11 ns	0.21 ± 0.19 *	0.01 ± 0.16 ns	0.00 ± 0.11 ns	0.00 ± 0.03 ns	1.11 ± 0.22 ***	1.38 ± 0.10
KPR	0.62 ± 0.83 ns	0.79 ± 0.89 ns	2.42 ± 1.29 ns	0.64 ± 0.87 ns	0.19 ± 0.37 ns	11.86 ± 1.51 ***	1.06 ± 0.08
TKN	60.98 ± 274.62 ns	488.52 ± 520.83 ns	495.96 ± 410.63 ns	451.30 ± 344.30 ns	159.46 ± 205.94 ns	2212.31 ± 538.68 ***	4073.98 ± 301.16
EL	0.00 ± 0.19 ns	0.15 ± 0.17 ns	0.44 ± 0.28 ns	0.32 ± 0.23 ns	0.02 ± 0.07 ns	2.21 ± 0.34 ***	1.38 ± 0.14
ED	0.02 ± 0.0082 **	0.01 ± 0.01 ns	0.00 ± 0.01 ns	0.00 ± 0.0073 ns	0.00092 ± 0.0029 ns	0.07 ± 0.01 ***	0.11 ± 0.0081
BTL	0.0014 ± 0.10 ns	0.05 ± 0.06 ns	0.16 ± 0.14 ns	0.23 ± 0.13 ns	0.00 ± 0.03 ns	1.09 ± 0.18 ***	0.68 ± 0.05
Yield	0.59 ± 0.26 *	0.51 ± 0.45 *	0.51 ± 0.26 *	0.19 ± 0.19 ns	0.00 ± 0.05 ns	2.33 ± 0.31 ***	0.37 ± 0.03

(b) Derived genetic parameters

Genetic Parameter	V_G (Total genetic variance)	$V_{G \times site}$ (Total genotype × environment variance)	V_P (Entry-mean phenotypic variance)	H^2 (Broad-sense heritability, entry-mean basis)	V_P (Plot-level phenotypic variance)	H^2 (Broad-sense heritability, Plot-level basis)
DTT	4.03 ± 0.92	1.13 ± 0.25	5.10 ± 0.91	0.79 ± 0.04	7.18 ± 0.92	0.56 ± 0.06
DTA	4.70 ± 1.01	1.21 ± 0.21	5.63 ± 1.00	0.83 ± 0.03	7.22 ± 1.01	0.65 ± 0.05
DTS	5.42 ± 1.07	1.36 ± 0.25	6.47 ± 1.06	0.84 ± 0.03	8.24 ± 1.07	0.66 ± 0.05
PH	325.03 ± 92.89	222.50 ± 32.59	476.07 ± 91.47	0.68 ± 0.07	706.70 ± 93.09	0.46 ± 0.08

EH	356.21 ± 134.62	133.46 ± 18.71	439.31 ± 134.30	0.81 ± 0.06	555.14 ± 134.64	0.64 ± 0.09
KRN	0.42 ± 0.24	1.11 ± 0.19	1.32 ± 0.22	0.32 ± 0.14	2.91 ± 0.25	0.14 ± 0.08
KPR	3.83 ± 1.50	12.68 ± 1.40	10.43 ± 1.33	0.37 ± 0.10	17.57 ± 1.50	0.22 ± 0.08
TKN	1045.46 ± 652.29	2823.07 ± 555.57	3475.49 ± 594.97	0.30 ± 0.14	7942.51 ± 669.45	0.13 ± 0.08
EL	0.59 ± 0.32	2.56 ± 0.34	2.22 ± 0.27	0.27 ± 0.12	4.53 ± 0.32	0.13 ± 0.07
ED	0.03 ± 0.02	0.07 ± 0.01	0.09 ± 0.01	0.33 ± 0.12	0.21 ± 0.02	0.15 ± 0.07
BTL	0.21 ± 0.15	1.32 ± 0.18	1.04 ± 0.12	0.21 ± 0.12	2.21 ± 0.15	0.10 ± 0.07
Yield	1.62 ± 0.55	2.53 ± 0.29	2.97 ± 0.53	0.54 ± 0.09	4.51 ± 0.55	0.36 ± 0.09

Note: $V_{GCA\ (female)}$ and $V_{GCA\ (male)}$, variance components of general combining ability for female lines and male testers, respectively; V_{SCA}, variance component of specific combining ability (line × tester interaction); $V_{GCA \times S(female)}$ and $V_{GCA \times S(male)}$, variance components of GCA-by-site interaction for female and male parents, respectively; $V_{SCA \times site}$, variance component of SCA-by-site interaction; $V_{Residual}$, residual variance; V_G, total genetic variance; $V_{G \times S}$, total genotype-by-site interaction variance; V_p, phenotypic variance; H^2, broad-sense heritability. Values are presented as estimate ± standard error. *, **, *** denote significance at $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively; ns, non-significant.

Analysis of GCA effects

GCA effects of male tester inbred lines

Estimates GCA effects of the four temperate male testers are presented in Table 4. The four testers exhibited distinct GCA patterns across all evaluated agronomic traits. For grain yield, Ji1877 had the largest positive GCA effect, while Chang7-2 showed a negative GCA effect. For plant-architecture traits, Z58 possessed the strongest negative GCA effects for PH and EH, whereas Chang7-2 had the highest positive GCA effect for EH. For phenological traits, Ji1877 showed positive GCA effects for DTT, DTA and DTS, while Z58 and Ji42 exhibited negative GCA effects for these three traits. Among ear-related traits, Ji42 had positive GCA effects for KPR and EL, and Chang7-2 showed positive effects for KRN and TKN.

Table 4 GCA effects of four temperate tester inbred lines for 12 agronomic traits

Inbred line	DTT	DTA	DTS	PH	EH	KRN	KPR	TKN	EL	ED	BTL	Yield
Z58	-0.76	-0.47	-0.13	-12.25	-13.63	-0.31	-1.04	-27.90	-0.05	0.05	**0.18**	-0.11
Chang7-2	0.27	-0.05	0.29	3.59	**15.57**	**0.64**	0.30	**15.44**	-0.25	**0.11**	0.07	-0.86
Ji1877	**0.90**	**1.10**	**0.42**	1.83	-0.06	-0.23	-0.02	1.83	-0.18	-0.10	-0.25	**0.81**
Ji42	-0.40	-0.57	-0.59	**6.84**	-1.88	-0.10	**0.76**	10.64	**0.48**	-0.06	0.00	0.15

Note: Bold values denote the largest positive GCA effect for each trait.

GCA effects of female inbred lines

Estimated GCA effects for 12 agronomic traits of the 46 tropical female inbred lines are presented in Additional file 2. GCA-effect variation differed markedly among traits. PH and EH had the widest ranges of GCA effects, while EL and BTL showed GCA effects close to zero across all lines. The remaining traits showed moderate variation.

For traits with obvious GCA-effect variation, lines with large GCA effects showed

clear trait-specific patterns. For phenological traits (DTT, DTA and DTS), CML505 and CML507 had large negative GCA effects, whereas CML601 had the highest positive GCA effects for these three traits. For plant-architecture traits (PH, EH), CML505 and CML507 also showed the most prominent negative GCA effects. For ear-related traits, CML554 had the largest positive GCA effects for TKN and KRN. For grain yield, CML555, CML554 and CML551 exhibited the largest positive GCA effects.

No single inbred line had consistently positive or negative GCA effects for all traits; most lines carried a mixture of positive and negative GCA effects. To remove scale differences and characterize multi-trait GCA profiles, Z-score standardization was performed on GCA estimates for the ten traits with considerable GCA-effect variation, followed by hierarchical cluster analysis. The tropical germplasm was classified into four major clusters (Fig. 1). GCA-Group 1 (represented by CML505 and CML507) showed strong negative GCA effects for phenological (DTT, DTA and DTS) and plant-architecture (PH, EH) traits, with mostly negative or near-zero effects for ear-related and yield traits. GCA-Group 2 (represented by CML486 and CML506) had near-zero GCA effects for all measured traits. GCA-Group 3 (represented by CML554, CML577 and CML602) exhibited strongly positive GCA effects for ear-related traits and grain yield, with near-neutral effects for phenological and plant-architecture traits. GCA-Group 4 (represented by CML522 and CML567) showed strongly positive GCA effects for PH and EH, together with near-neutral effects for the remaining traits.

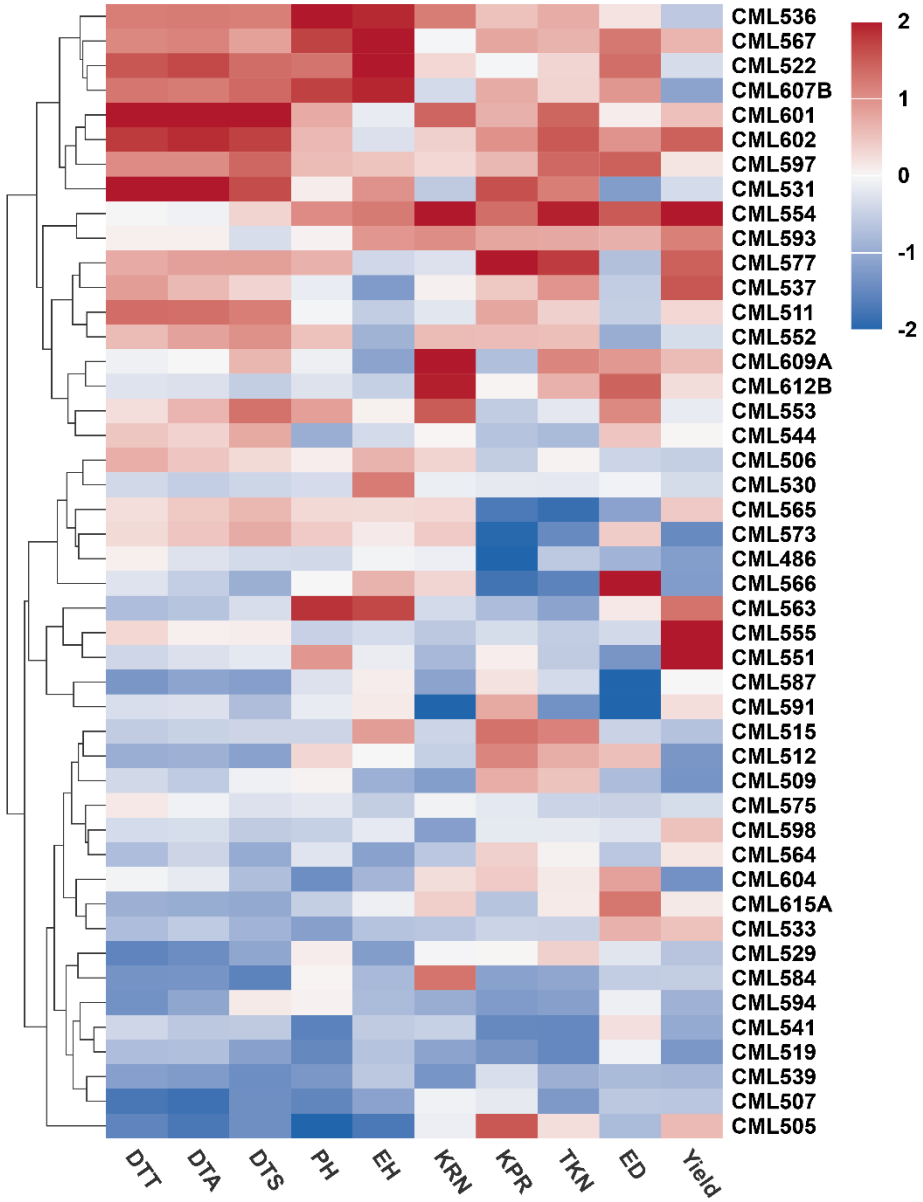


Fig. 1. Bidirectional cluster heatmap of Z-score standardized GCA effects for 10 genetically variable agronomic traits across 46 tropical maternal maize inbred lines. Note: Rows represent individual inbred lines clustered using Euclidean distance and Ward's algorithm; columns indicate evaluated traits after column-wise Z-score normalization. The color bar denotes standardized GCA values ranging from -2 to 2, with red for positive general combining ability effects, blue for negative effects, and white for zero values.

Analysis of SCA effects

The breeding value of maize hybrids relies on both GCA and SCA effects. In this study, SCA effects for 12 agronomic traits of the 184 testcross hybrids are presented in

Additional file 3. These results were used to characterize non-additive genetic variation, describe tester-specific SCA patterns, and screen promising hybrid combinations.

Overall variation patterns of SCA effects across the population

Among the 12 evaluated agronomic traits, ED showed no detectable SCA effects across all 184 test-cross hybrids, indicating an absence of non-additive genetic variation for this trait (Table 5). Although the line × tester term for ED was significant in the fixed-effects ANOVA (Table 2), the corresponding variance component was estimated as zero under the REML mixed model (Table 3), and all SCA values for ED were equal to zero (Table 5). This discrepancy arises from the different statistical assumptions underlying the two analytical approaches (Section 2.5). The remaining 11 traits exhibited SCA variation of varying magnitudes, reflecting substantial trait-to-trait differences in non-additive genetic variation (Table 5). TKN possessed the widest SCA-effect range (−26.08 to 23.61, range = 49.69), followed by EH and KPR, whereas grain yield displayed a moderate SCA range. For most traits, the numbers of hybrids with positive and negative SCA effects were roughly balanced. By contrast, more hybrids exhibited negative SCA effects for KRN, EL, BTL, DTS, and EH.

436 **Table 5** Variation ranges of SCA effects for 12 agronomic traits across 184 hybrids

Trait	SCA Min	SCA Max	SCA Range	Std. Deviation	No. of positive SCA	No. of negative SCA	Positive proportion (%)
DTT	-1.24	1.25	2.49	0.42	94	90	51.09
DTA	-1.25	1.13	2.38	0.45	94	90	51.09
DTS	-1.68	1.59	3.27	0.60	88	96	47.83
PH	-2.40	1.19	3.59	0.45	92	92	50.00
EH	-7.41	7.40	14.81	2.85	88	96	47.83
KRN	-0.04	0.04	0.08	0.01	81	97	44.02
KPR	-3.24	1.75	4.99	0.79	99	85	53.80
TKN	-26.08	23.61	49.69	9.24	96	88	52.17
EL	-0.92	0.91	1.83	0.31	85	99	46.20
ED	0.00	0.00	0	0.00	0	0	0.00
BTL	-0.26	0.72	0.98	0.17	81	103	44.02
Yield	-0.84	0.74	1.58	0.35	98	86	53.26

437 Note: SCA effects were estimated based on the optimal m4 model using REML. SCA
438 range = max - min. Positive and negative SCA percentages indicate the proportion of
439 hybrids with positive or negative SCA effects for each trait.

440 **Tester-dependent SCA variation and identification of promising testcrosses**

441 For a given trait, the dispersion of SCA effects varied among testcross hybrids
442 derived from different testers. Taking TKN and grain yield as examples, the median
443 SCA effects of hybrids within each tester group were close to zero, while obvious
444 differences in dispersion were detected (Fig. 2-A-B). Among the four testers, hybrids
445 from Z58 showed the widest dispersion of SCA effects for TKN, and several hybrids
446 had prominent positive and negative SCA effects. Hybrids from Ji1877 exhibited the
447 broadest SCA dispersion for grain yield. By comparison, hybrids derived from the other
448 two testers showed relatively narrow SCA-effect dispersion.

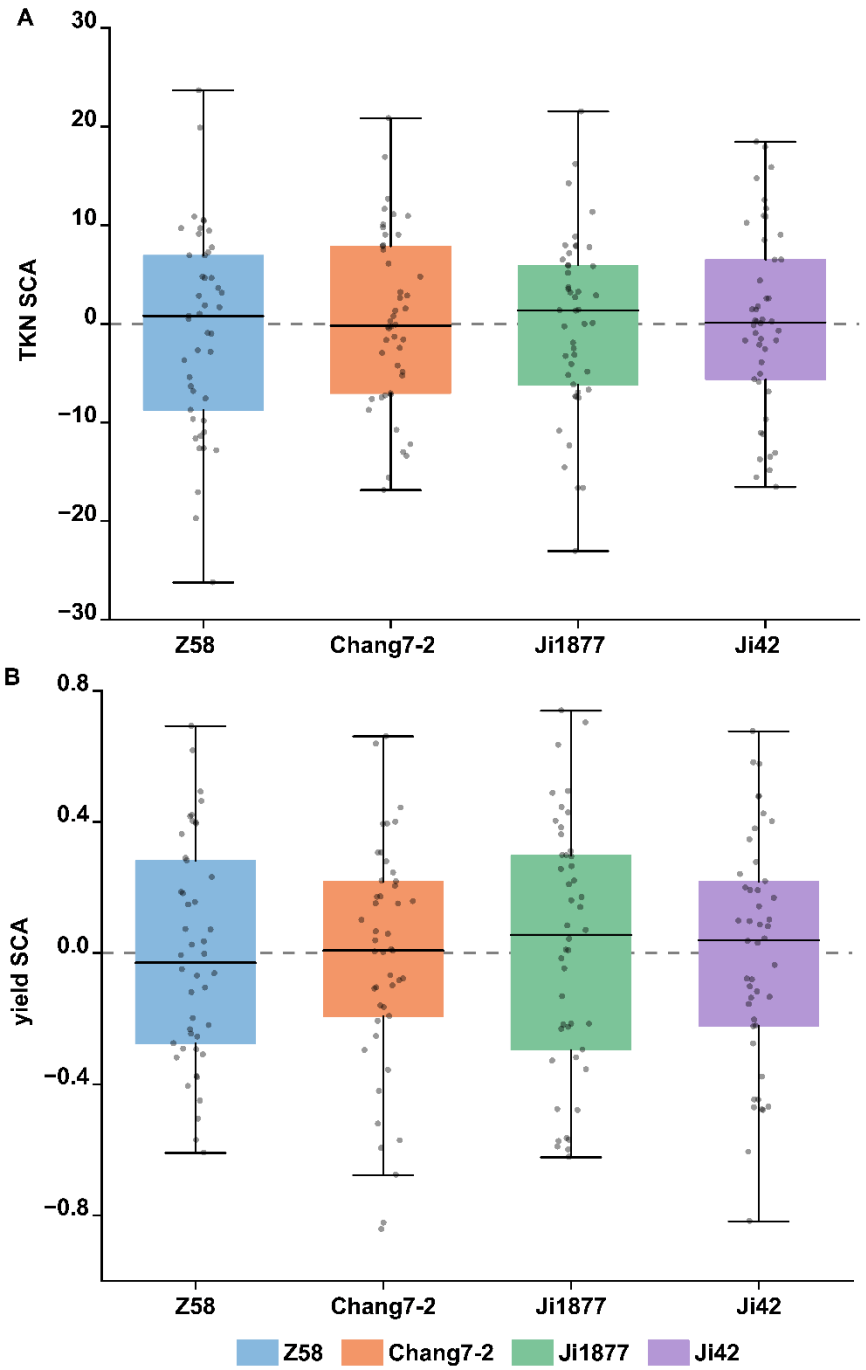


Fig. 2. Box plots of SCA effects for (A) TKN and (B) Yield across four testers. Note: Each gray point represents the SCA estimate for one testcross. Black lines indicate medians, and horizontal gray dashed lines at y = 0 indicate zero SCA. The spread of SCA estimates differed among testers, with Z58-derived testcrosses showing the widest range.

Based on predefined selection criteria prioritizing positive SCA effects for grain yield and sink-related traits (TKN, EL), and avoiding large positive SCA effects for plant-architecture traits (PH, EH) and flowering-time traits (DTA, DTS), six promising

testcross combinations were identified (Table 6). Their adjusted grain-yield means, standard errors and performance relative to the commercial check ZD958 are presented in Table 7.

In terms of adjusted grain-yield means, CML554 × Ji1877 (9.92 Mg ha^{-1}), CML577 × Ji1877 (8.54 Mg ha^{-1}), CML601 × Ji42 (7.36 Mg ha^{-1}) and CML522 × Ji42 (7.10 Mg ha^{-1}) had higher means than the commercial check ZD 958 (5.32 Mg ha^{-1}), with yield advantages ranging from 33.46% to 86.47% (Table 7). Among these crosses, CML522 × Ji42 had the smallest standard error (SE = 0.08), showing the most reliable mean estimate. In contrast, CML554 × Ji1877 had the largest SE (SE = 2.66), indicating greater uncertainty in its mean value despite the highest adjusted grain-yield mean. Because each selected cross involved only one tester, cross performance across different tester backgrounds could not be evaluated. Nevertheless, cross performance across the two experimental sites was used to evaluate phenotypic variation between the tested sites (Table 7).

Mid-parent and better-parent heterosis were not calculated in this study. Therefore, these crosses are treated as promising candidates rather than confirmed superior hybrids. Additional multi-site trials are required to verify their yield advantage over the commercial check and cross-site phenotypic stability.

Table 6 SCA of six promising testcross combinations for key agronomic traits

Hybrids	Yield	TKN	EL	PH	EH	DTA	DTS
CML512 × Z58	0.40	20.06	0.79	-0.03	0.85	-0.50	-0.45
CML522 × Ji42	0.58	11.52	0.56	-0.01	-1.32	-0.34	-0.48
CML601 × Ji42	0.48	10.99	0.31	-0.07	-0.78	0.09	-0.23
CML486 × Chang7-2	0.15	9.91	0.12	-0.50	-1.35	-0.44	-0.80
CML554 × Ji1877	0.70	11.08	0.22	-0.39	-3.20	0.58	0.29
CML577 × Ji1877	0.45	14.18	0.34	-0.93	−6.38	0.25	0.23

Note: SCA effects were estimated by REML under model M4. These six testcrosses are candidates for validation across additional sites and years; they should not be considered superior hybrids until their observed yield, stability, and uncertainty are evaluated.

Table 7 Adjusted grain-yield means, standard errors, and performance relative to the commercial check ZD 958 for six promising testcross combinations across two sites

Combination	Adjusted mean (Mg ha^{-1})	SE	Site 1 (Mg ha^{-1})	Site 2 (Mg ha^{-1})	% of check
CML512 × Z58	5.64	0.44	6.08	5.20	106.0
CML522 × Ji42	7.10	0.08	7.18	7.02	133.5
CML601 × Ji42	7.36	0.7	8.07	6.66	138.4
CML486 × Chang7-2	4.09	0.112	4.74	3.44	76.9
CML554 × Ji1877	9.92	2.66	12.59	7.26	186.5
CML577 × Ji1877	8.54	1.27	9.81	7.28	160.5
ZD958 (CK)	5.32	1.31	4.01	6.63	100.0

Note: Means are in Mg ha^{-1}. SE, standard error of the mean. % of check = (combination mean / ZD958 mean) × 100. ZD958 served as the commercial check. Site stability was evaluated via cross-combination performance across the two experimental sites.

Analysis of population genetic structure and heterotic-group classification based on SNP markers

Population genetic structure

An NJ phylogenetic tree was constructed based on MRD calculated from genome-wide SNP markers to explore genetic relationships among the 50 maize inbred lines. These accessions were classified into three major clusters (Fig. 3). Most tropical inbred lines belonged to Genetic groups 1 and 3, while the four temperate testers (Ji42, Z58, Chang7-2 and Ji1877) formed Genetic group 2 as a separate clade. This result indicated

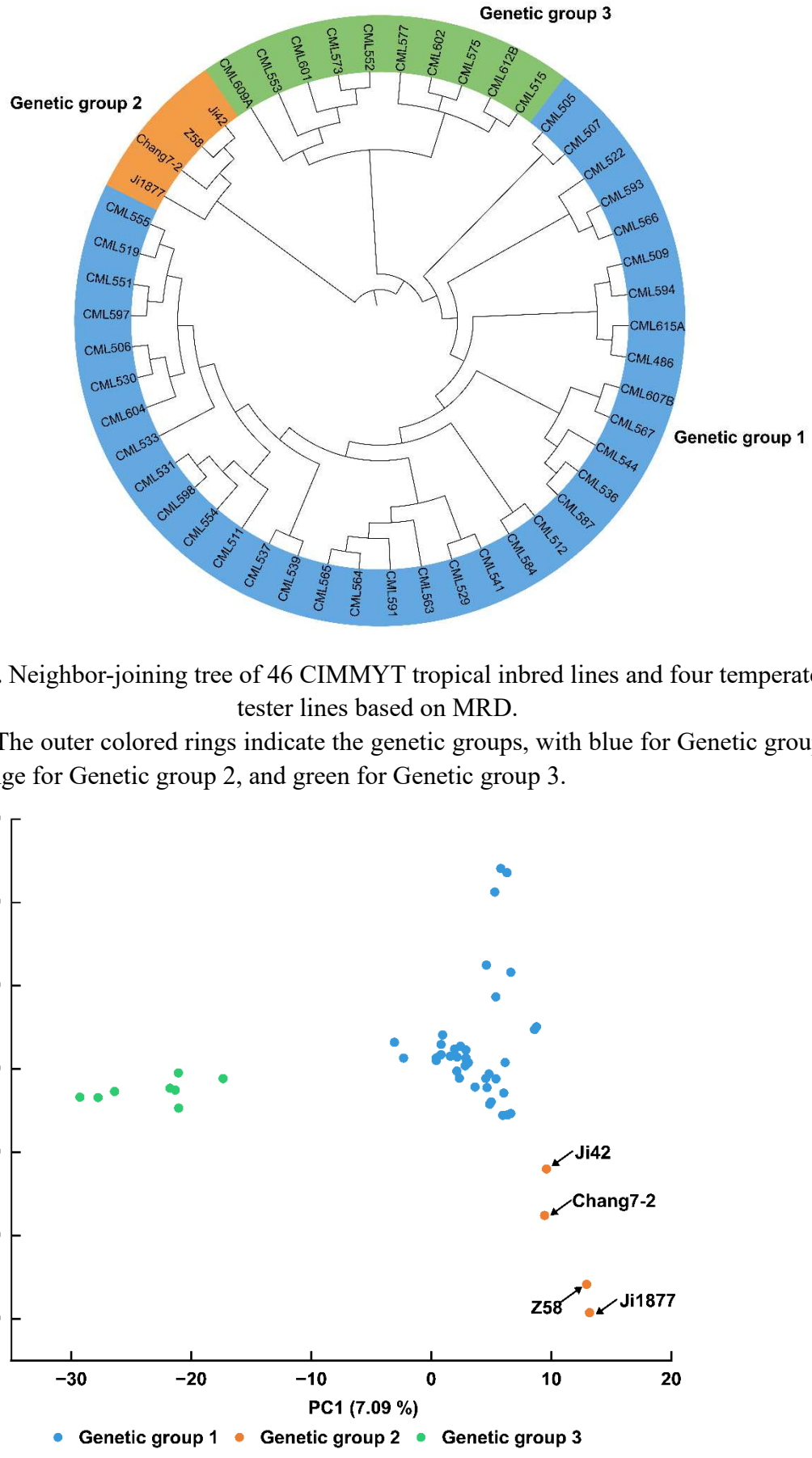
494 clear genetic divergence between temperate testers and tropical parental lines.

495 PCA (Fig. 4) revealed clustering patterns consistent with the NJ tree. PC1 and PC2
496 explained 7.09% and 5.25% of the total genetic variation, respectively. The four
497 temperate testers were clearly separated from most tropical inbred lines along PC1,
498 while PC2 further distinguished subgroups within the tropical germplasm.

499 Bayesian clustering implemented in STRUCTURE was performed at two
500 hierarchical levels. For the full panel of 50 inbred lines, the ΔK method identified K =
501 3 as the optimal number of genetic groups (Additional file 4-A). At K = 3, the four
502 temperate testers were predominantly assigned to one genetic group, and the tropical
503 germplasm was divided into two major clusters (Fig. 5-A). To resolve finer substructure
504 within tropical materials, STRUCTURE analysis was conducted separately using only
505 the 46 tropical inbred lines. The ΔK method supported K = 4 as the optimal partition,
506 which distinguished four genetic subgroups within the tropical germplasm (Additional
507 file 4-B, Fig. 5-B).

508 The STRUCTURE results were consistent with the NJ tree and PCA, clearly
509 separating temperate testers from tropical inbred lines. Ji42 and Z58 showed nearly
510 pure ancestry from the temperate cluster (Genetic group 2). By contrast, Chang7-2 and
511 Ji1877 mainly belonged to Genetic group 2, with small ancestral components from
512 Genetic group 1 (tropical origin).

513 Overall, the marked genetic differentiation between tropical germplasm and
514 temperate testers, together with clearly resolved population substructure, provides a
515 robust basis for developing candidate heterotic-group classification.



516 **Fig. 3.** Neighbor-joining tree of 46 CIMMYT tropical inbred lines and four temperate
517 tester lines based on MRD.
518 Note: The outer colored rings indicate the genetic groups, with blue for Genetic group
519 1, orange for Genetic group 2, and green for Genetic group 3.

520

Fig. 4. Scatter plot of the first two principal components (PC1, PC2) for 46 CIMMYT tropical inbred lines and four temperate tester lines.

Note: Points are colored according to the three genetic groups: blue for Genetic group 1, orange for Genetic group 2, and green for Genetic group 3. The four temperate testers are explicitly labeled.

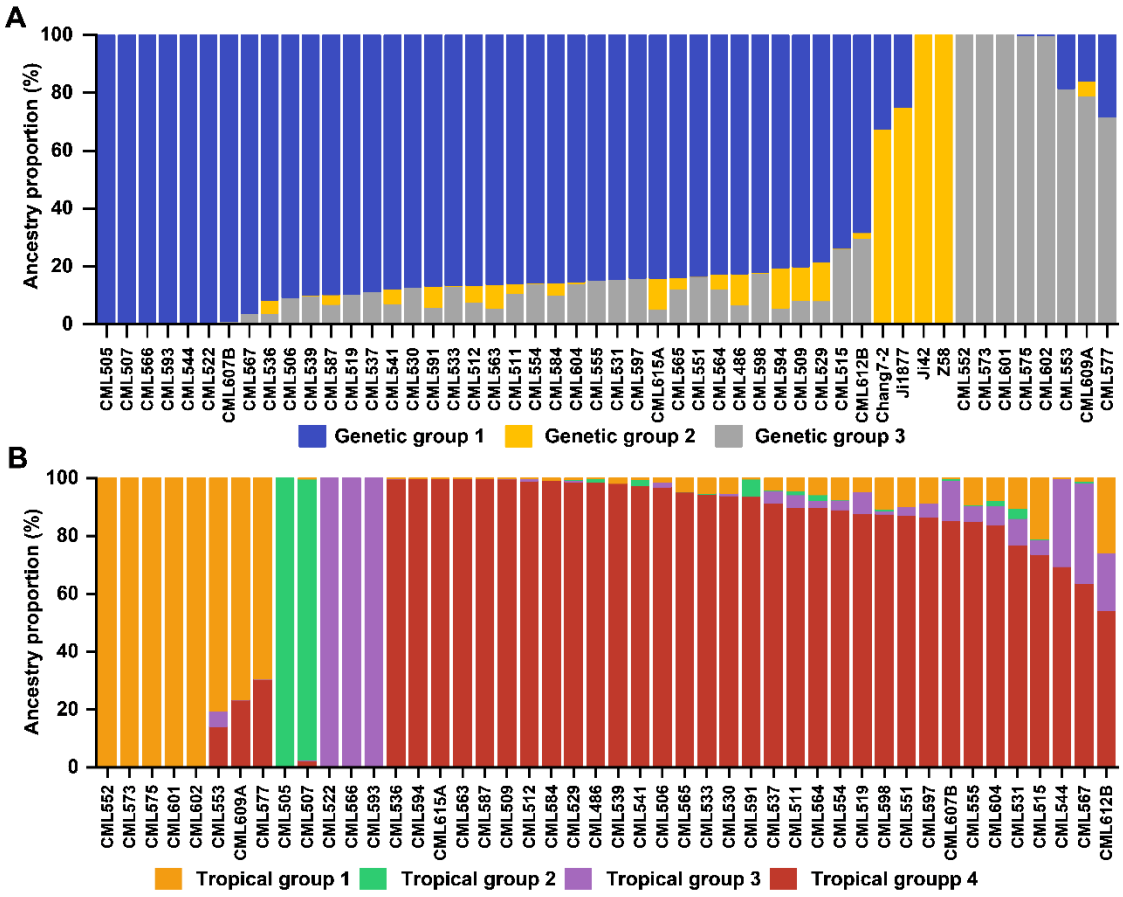


Fig. 5. Population genetic structure of the maize inbred line panel.

Note: (A) Ancestry proportions from STRUCTURE analysis (K = 3) for all 50 inbred lines, identifying three major genetic groups. Each vertical bar denotes one inbred line, and colours show its ancestry fraction for each genetic group. (B) Ancestry proportions from STRUCTURE analysis (K = 4) for the subset of 46 tropical inbred lines, resolving four subgroups within tropical germplasm. Each vertical bar denotes one tropical inbred line, and colours indicate the proportional ancestry for each tropical subgroup.

Molecular genetic groups as candidates for heterotic-group classification

Based on the population genetic structure identified by STRUCTURE, the full panel of 50 inbred lines was divided into three genetic groups at $K = 3$ (Fig. 5-A). The subset of 46 tropical inbred lines was further classified into four molecular genetic groups (Tropical groups 1-4; Fig. 5-B). These molecular groups are regarded as

candidate heterotic group partitions, based on the assumption that genome-wide genetic differentiation may partly reflect heterotic patterns. Nevertheless, heterotic groups are ultimately defined by hybrid performance such as SCA or heterosis. Therefore, these molecular classifications require validation through testcross evaluation.

These molecular groups provide a framework for comparison with combining-ability-based classifications. Although descriptive visual comparison suggested partial overlap between molecular groups and GCA-based breeding types (Section 3.3.2), overall concordance was weak. Therefore, to formally evaluate the predictive value of parental molecular genetic background, we further tested whether the continuous measure of parental MRD had a linear association with grain yield SCA and hybrid grain yield.

Predictive ability of MRD for SCA and yield

Based on the SNP-derived candidate heterotic groups described above, we further evaluated whether parental MRD could predict testcross performance. The analysis accounted for the non-independence of the 184 testcrosses arising from shared female and male parents. Mixed-effects regression models were fitted with MRD as a fixed effect, and female line and male tester as random effects. MRD values of the 184 testcross combinations ranged from 0.34 to 0.43 (mean = 0.37, SD = 0.02), showing an approximately unimodal distribution (Additional file 5).

The global MRD effect was non-significant for F_1 grain yield (slope = −3.47, P = 0.624) and grain yield SCA (slope = −0.38, P = 0.742; Table 8). The MRD × Tester interaction was also non-significant for F_1 grain yield (LRT = 0.78, $P = 0.855$) and grain

yield SCA (LRT = 0.89, P = 0.828), providing no statistical evidence for differing MRD-trait associations across the four tester backgrounds.

Variance partitioning results and tester-specific slopes from the mixed model are presented in Supplementary Additional file 6. Ordinary least-squares slopes fitted for each tester group were used only for visual description rather than statistical inference (Additional file 7-8). Residual diagnostics indicated that the assumptions of normality and homoscedasticity for the mixed model were satisfied (Additional file 9). Overall, MRD explained only a small proportion of the variation in F_1 grain yield and its SCA effect, with non-significant slopes and no tester-specific relationships identified.

Table 8 Mixed-effects regression of grain yield and yield SCA on MRD

Trait	Slope (β)	SE	95% CI	t	P	MRD × Tester (LRT)
Grain yield	-3.47	7.07	-17.32 to 10.38	-0.49	0.62	0.78 (P = 0.86)
yield SCA	-0.38	1.16	-2.66 to 1.89	-0.33	0.74	0.89 (P = 0.83)

Note: MRD, molecular-relatedness distance; SCA, specific combining ability; LRT, likelihood ratio test. Full outputs including tester-specific slopes and variance components are provided in Supplementary Additional file 6.

DISCUSSION

Phenotypic variation and genetic architecture of tropical × temperate testcross hybrids under YHH conditions

Flowering-related traits (DTT, DTA, DTS) had low phenotypic coefficients of variation, while grain yield and ear-component traits such as TKN and BTL displayed much larger phenotypic variation, in agreement with previous tropical × temperate maize testcross studies [28,29]. Narrow flowering-time variation indicated adequate photoperiod adaptation of the testcross panel under long-day conditions in the YHH

region, ensuring pollination synchrony and stable grain yield when tropical germplasm is used in temperate environments [28]. This pattern, however, does not necessarily indicate full long-day adaptation of the tropical parental lines, because F_1 flowering is jointly controlled by tropical female parents and temperate male testers.

Preliminary fixed-effects ANOVA revealed highly significant genotypic effects for most traits, with female GCA significant for nine traits, male GCA for seven traits, and SCA for seven traits. Notably, BTL showed no significant genotypic variation. Furthermore, SCA-by-site interactions were highly significant for all traits, whereas GCA-by-site interactions were mostly non-significant. It should be noted that significance from fixed-effect ANOVA F-tests may differ from LRT of REML variance components. The F-test assesses phenotypic mean differences under fixed-effect assumptions, whereas LRT tests whether variance components exceed zero under random-effect assumptions. Such discrepancies frequently occur in unbalanced line × tester designs, particularly when variance components carry large standard errors. These results demonstrate that additive and non-additive effects jointly determine hybrid performance for most traits. Specifically, the pervasive SCA-by-site interactions indicate that non-additive effects are highly environment-dependent, highlighting the necessity of multi-site trials for reliable selection [2,29].

Trait-specific genetic architectures were observed. Flowering traits (DTT, DTA, DTS) had high entry-mean broad-sense heritability, with significant GCA effects, confirming predominant additive control, in agreement with the consensus that maize flowering time is governed by many small-effect loci [30]. Therefore, phenotypic

603 selection for flowering adaptation is efficient in early generations during inbred
604 development [2].

605 By contrast, ear traits KPR and EL exhibited significant SCA effects, indicating
606 the presence of non-additive genetic variation, consistent with previous reports for
607 tropical × temperate maize [31]. Early-generation selection is inefficient for these traits,
608 and testcross evaluation is required in breeding practice [2]. For BTL, however, the
609 non-significant genotypic variation and low entry-mean broad-sense heritability imply
610 that its genetic variance was largely obscured by environmental error and G×S
611 interactions, limiting direct selection gains. Notably, ED represented an extreme case
612 where no SCA effects were detected (variance component = 0), indicating an absence
613 of non-additive genetic variation for this trait. For TKN, both female GCA and SCA
614 variance components were significant; however, several of its variance-component
615 estimates presented extremely large standard errors, indicating low estimation precision.
616 Consequently, its relatively low entry-mean broad-sense heritability restricts the
617 exploitation of additive variation, prioritizing hybrid screening over inbred
618 improvement.

619 Grain yield exhibited mixed genetic control: the female GCA variance component
620 was significant, pointing to a substantial additive contribution, while the SCA variance
621 component was also significant, reflecting non-additive effects. Significant GCA and
622 SCA effects for grain yield are commonly reported in maize NC-II experiments [2].

623 In summary, introgressing tropical germplasm into YHH environments requires
624 trait-specific selection strategies. Flowering adaptation can be improved via

additive-based selection during inbred-line development, while grain yield and sink-related ear traits need testcross phenotyping to capture non-additive genetic effects and their site-specific interactions.

GCA-based breeding-type classification and identification of superior tropical donor lines and hybrid combinations

Multi-trait GCA estimates from the NC-II mating design enabled evaluation of additive parental effects for tropical female lines and temperate male testers. Using standardized Z-score multi-trait GCA profiles, the 46 tropical female inbred lines were classified into four distinct breeding types following the HGCAMT analytical framework for combining-ability-based germplasm characterization [32,33]. Importantly, these four GCA-derived breeding types are phenotypic categories defined by multi-trait general-combining-ability profiles, rather than canonical heterotic groups inferred from SCA or hybrid performance. This distinction avoids conceptual confusion between additive parental breeding potential and non-additive heterotic differentiation [2,11,17].

GCA Group 1 (represented by CML505 and CML507) exhibited strong negative GCA for phenological and plant-architecture traits, indicating that this subset of tropical lines can act as donors to shorten growth cycle and reduce plant height in tropical × temperate breeding pipelines. GCA Group 3 (CML554, CML577, CML602) showed favorable positive GCA for grain yield and sink-related ear traits while maintaining neutral effects for flowering and plant architecture, making this group the most valuable donor pool for improving yield potential in tropical × temperate hybrid breeding targeting the tested YHH environments. In contrast, GCA Group 2 lines had near-zero

GCA across all evaluated traits and may serve as neutral genetic bridge materials for germplasm introgression. GCA Group 4 lines showed positive GCA for plant height and ear height and should be deployed cautiously to prevent excessive vegetative growth under high-density production in the YHH Plain [15]. Collectively, no individual inbred line exhibited uniformly positive or negative GCA across all measured traits, confirming that CIMMYT tropical CML germplasm harbors complementary favorable alleles distributed across distinct genetic backgrounds. Pyramiding favorable alleles from multiple donor lines will be required to achieve comprehensive multi-trait improvement [2,33].

Distinct GCA profiles were observed among tropical female lines and the four temperate testers, highlighting substantial variation in the capacity of parental lines to transmit favorable alleles to their progeny. Selection of tropical female lines with high positive GCA for grain yield and key ear traits represents a practical route to enrich favorable additive alleles within temperate-adapted breeding pools. Meanwhile, tester-dependent SCA variation revealed that top-performing hybrid combinations arose from complementation between high-GCA female donors and specific male testers. This observation underscores that superior hybrid performance relies on both additive parental merit and non-additive inter-parental complementarity. Although high GCA can serve as a primary criterion for preliminary screening of tropical donor lines, promising hybrid combinations cannot be reliably predicted from GCA alone, because SCA contributes substantially to yield and sink-related ear traits.

Notably, high-yielding testcross combinations may exhibit larger cross-site

669 variation. Among the six promising hybrids, CML554 × Ji1877 had the greatest yield
670 potential but greater uncertainty in cross-site performance, whereas CML522 × Ji42
671 displayed consistent phenotypic performance across the two tested sites with a
672 moderate yield advantage. Therefore, balancing yield potential and phenotypic
673 performance across environments is critical, and multi-site and multi-year testing
674 remains indispensable for candidate hybrid validation before considering further
675 breeding evaluation under YHH field conditions.

676 **Limited concordance between SNP-derived molecular genetic groups and GCA-**
677 **based breeding-type grouping**

678 Population genetic structure analyses (NJ tree, PCA, and STRUCTURE)
679 consistently revealed clear genetic differentiation between the 46 tropical CML inbred
680 lines and four temperate testers. Ji42 and Z58 exhibited nearly pure ancestry from the
681 temperate Genetic group 2. By contrast, Chang7-2 and Ji1877 were predominantly
682 assigned to Genetic group 2, with small ancestral components derived from tropical
683 Genetic group 1. Hierarchical STRUCTURE analysis identified $K = 3$ as optimal for
684 the full panel of 50 lines and $K = 4$ for the subset of tropical lines, resolving fine-scale
685 substructure within tropical germplasm [24]. These SNP-derived genetic groups serve as
686 candidate heterotic-group partitions, based on the assumption that genome-wide
687 divergence partially reflects heterotic patterns. As shown in Section 3.6.2, descriptive
688 comparisons indicated limited concordance between these molecular groups and
689 combining-ability-based classifications.

690 This limited concordance arises because genome-wide neutral SNP variation

captured by molecular clustering does not directly reflect breeding value or combining ability for target agronomic traits; neutral marker divergence cannot be equated to the functional allelic variation governing grain yield and sink-component traits. Multiple mechanisms underlie this pattern [35]. First, widespread ancestral admixture in tropical maize creates mosaic haplotypes, where favorable functional alleles recombine onto diverse neutral genomic backgrounds and weaken the association between marker relatedness and multi-trait breeding performance [36]. Second, GCA is governed by additive and additive-by-additive epistatic effects, whereas SNP clustering mainly captures neutral relatedness and often fails to capture variation at the small set of functional loci controlling combining ability [35,37]. Third, GCA estimates were obtained under the tested YHH field environments; genotype-by-site (G×S) interactions and environmental plasticity further decouple neutral genomic divergence from phenotypic combining-ability classification [16].

From a breeding perspective, molecular population structure provides useful background information for germplasm management but cannot replace empirical testcross-based GCA evaluation for parental selection in tropical × temperate hybrid breeding. Sole reliance on SNP clustering risks discarding valuable donors with favorable yield-related combining ability observed in this trial.

Limited linear predictive power of genome-wide MRD for yield-related SCA and F_1 grain yield performance

Given the limited correspondence between genome-wide genetic background and combining-ability variation, we further evaluated whether pairwise MRD could linearly

predict entry-mean F_1 grain yield and estimated yield SCA across the 184 tropical × temperate testcrosses using mixed-effects regression. The model accounted for non-independence among hybrids by incorporating female line and male tester as random effects, with MRD fitted as a fixed effect. Using entry-mean yield and SCA point estimates avoided the potential underestimation of genetic variation caused by shrinkage effects**.** The global MRD effect was non-significant for F_1 grain yield (slope = −3.47, 95% CI: -17.32 to 10.38, $P = 0.624$) and yield SCA (slope = −0.38, 95% CI: -2.66 to 1.89, $P = 0.742$; Table 8). The MRD × Tester interaction was also non-significant for both traits, indicating no tester-dependent difference in the MRD–trait association. Tester-specific slopes and variance components from the mixed model are provided in Supplementary Table S5, while OLS slopes for each tester group were fitted only for descriptive visualization rather than statistical inference. Residual diagnostics confirmed that mixed-model assumptions were satisfied. It should be noted that the MRD values in this panel spanned a relatively narrow range (0.34–0.43), which may limit statistical power to detect weak linear associations; additionally, the sampling errors inherent in the estimated SCA effects were absorbed into the residual term. Nevertheless, because both the main effect and the interaction were non-significant, our conclusion remains statistically conservative**.** Collectively, genome-wide MRD showed limited linear predictive capacity for hybrid yield and yield SCA within this germplasm panel and tested environments.

This weak predictive performance arises from several genetic features. First, MRD primarily quantifies neutral genome-wide allelic differences, whereas yield SCA is

735 determined by locus-specific dominance, overdominance and epistasis. Neutral SNPs
736 are often imperfectly linked to causal functional loci, decoupling average parental
737 genomic distance from non-additive complementarity [37,38]. Second, all crosses in this
738 panel follow tropical female × temperate male, already possessing substantial inter-
739 ecotype divergence. Within this constrained cross type, incremental changes in
740 genome-wide MRD do not translate into proportional changes in heterosis or SCA [39].
741 Notably, this result only rejects simple linear prediction using average MRD; it does
742 not exclude non-linear relationships or prediction approaches using functional markers,
743 haplotypes or genomic prediction models [37,40].

744 This study has several limitations. First, the 46 CML tropical lines represent only
745 a subset of CIMMYT tropical germplasm; therefore, our results cannot be broadly
746 generalized. Second, phenotypic evaluation was performed at two sites within a single
747 growing season in the YHH Plain. Accordingly, all phenotypic and statistical
748 conclusions are limited to these tested sites, which restricts extrapolation of our findings.
749 Multi-year and multi-ecozone trials are needed to validate these results and improve the
750 robustness of GCA and SCA estimates. Third, only genome-wide average MRD, a
751 population-level genomic dissimilarity metric, was assessed in the present study. Future
752 work could explore alternative genomic prediction approaches such as GBLUP to
753 enhance the prediction of combining-ability performance.

754 For tropical × temperate maize breeding targeting the tested YHH conditions, these
755 results indicate that genome-wide MRD cannot replace empirical testcross phenotyping
756 for screening promising hybrid combinations or evaluating the GCA of tropical donor

lines. Even so, SNP-derived population structure remains valuable for germplasm conservation, diversity monitoring, and the removal of redundant accessions, despite its limited capacity to predict combining ability.

Conclusions

Genome-wide SNP-based population-structure analyses clearly differentiated the 46 tropical CML inbred lines from the four temperate testers. Among the 184 tropical × temperate test-cross hybrids evaluated across two sites in a single growing season, pairwise MRD exhibited no significant linear association with yield SCA or F_1 grain yield. SNP-derived molecular genetic groups showed only partial correspondence with multi-trait GCA-based breeding types. Collectively, these findings demonstrate that simple genome-wide clustering and average parental genomic distance cannot replace empirical test-cross evaluation for parental screening within this germplasm panel. Future efforts involving multi-year and multi-ecozone phenotypic trials, statistical models that account for relatedness among shared parents, and genomic prediction approaches incorporating functional markers or haplotype effects are needed to validate these outcomes and extend inferences to other germplasm panels or growing environments.

Abbreviations

BTL: Barren Tip Length

CV: Coefficient of Variation

DTA: Days to Anthesis

DTS: Days to Silking

779 DTT: Days to Tasseling

780 ED: Ear Diameter

781 EH: Ear Height

782 EL: Ear Length

783 GBLUP: Genomic Best Linear Unbiased Prediction

784 GCA: General Combining Ability

785 G×S: Genotype-by-Site Interaction

786 H^2: Broad-sense Heritability

787 HGCAMT: Heterosis Grouping by Combining Ability of Multiple Traits

788 HSGCA: Heterotic Grouping based on Specific and General Combining Ability

789 KRN: Kernel Row Number

790 MRD: Modified Rogers' Distance

791 NC-II: Line × Tester

792 NJ: Neighbor-Joining

793 PCA: Principal-Component Analysis

794 PH: Plant Height

795 REML: Restricted Maximum Likelihood

796 SCA: Specific Combining Ability

797 SNP: Single Nucleotide Polymorphism

798 TKN: Total Kernel Number

799 V_A: Additive Variance

800 V_D: Dominance Variance

801 $V_{G \times S}$: Total Genotype-by-Site Interaction Variance

802 V_{GCA}: Variance of General Combining Ability

803 $V_{GCA \times site}$: Variance Component of GCA-by-Site Interaction

804 V_P: Phenotypic Variance

805 $V_{Residual}$: Residual Variance

806 V_{SCA}: Estimated Variance of Specific Combining Ability

807 $V_{SCA \times site}$: Variance Component of SCA-by-Site Interaction

808 YHH: Yellow–Huai–Hai Plain

809 ## Supplemental information

810 Additional file 3ontain additional population genetic analysis outputs, phenotypic
811 statistics, heritability estimates, raw GCA/SCA values for the 12 agronomic traits
812 across testcross combinations, and supplementary figures.

813 ## Acknowledgments

814 We thank Dr. Tai An from the Institute of Crop Sciences, Chinese Academy of
815 Agricultural Sciences for assistance with data analysis.

816 ## Author contributions

817 Yuanyuan Yan: Writing – review & editing, Writing – original draft, Visualization,
818 Validation, Supervision, Formal analysis, Data curation, Conceptualization,
819 investigation; methodology; Rui Guo: Writing – review & editing, Formal analysis,
820 Validation, Formal analysis, Data curation, Software. Haihong Fan: Writing – original
821 draft, Data curation. Jianghao Wang: Writing – review & editing, Supervision,
822 Resources, Conceptualization. Wei Song: Writing – review & editing, Supervision,

Resources, Project administration, Funding acquisition, Conceptualization. Liang Song and Liwei Wang: Validation, Formal analysis, Data curation. Jianfeng Wei and Xinghua Li: Validation, Supervision, Project administration.

Funding

This work was supported by the HAAFS International Science and Technology Cooperation Project (2025KJCXZX-LYS-GH01), the Hebei Provincial Science and Technology Plan Project (262N6309D), and the Natural Science Foundation of Hebei Province (C2025301136).

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests

References

1. Morris ML, Edmeades GO, Barker TC, Wilkinson DR. The exploitation of tropical maize germplasm for temperate breeding. Crop Sci. 1999;39:1033-1040. https://doi.org/10.2135/cropsci1999.001118.

2. Hallauer AR, Carena MJ, Miranda Filho JB. Quantitative genetics in maize breeding (3rd ed.). New York: Springer; 2010. ISBN 9781441907653.

https://doi.org/10.1007/978-1-4419-0766-0.

3. Li Q, Wang PJ, Li Y, Tang JX, Li X, Zhang YD, Han DC, Zhao YC, Wang Q, Yang ZQ. Identifying trigger thresholds for heat damage at different growth stages of summer maize: a case study of the Huang-Huai-Hai Plain, China. Eur J Agron. 2025;170:127729. https://doi.org/10.1016/j.eja.2025.127729.

4. Fu ZZ, Liu ZJ, Guo SB, Zhang ZT, Lu XM, Gong JJ, Dong MQ, Yang XG. Impact assessment of heat stress during post-silking period of summer maize in the Huang-Huai-Hai Plain, China. J Agric Sci. 2023;161:318-327. https://doi.org/10.1017/S0021859623000321.

5. Cairns JE, Crossa J, Zaidi PH, Grudloyma P, Sanchez C, Araus JL, Thaitad S, Makumbi D, Magorokosho C, Bänziger M, Menkir A, Hearne S, Atlin GN. Identification of drought, heat, and combined drought and heat tolerant donors in maize. Crop Sci. 2013;53:1335-1346. https://doi.org/10.2135/cropsci2012.09.0545.

6. Mudadirwa D, Olakojo SA, Ndhlela T, Mhike X, Shayanowako AIT, Mukuze C. Understanding genetic diversity and population structure of CIMMYT and IITA early maturing provitamin A maize (Zea mays L.) inbred lines using phenotypic traits and SNP markers for variable tolerance to drought and heat stress. Discover Plants. 2025;2:119. https://doi.org/10.1007/s44372-025-00205-4.

7. Mungoma C, Pollak LM. Photoperiod sensitivity in tropical maize accessions, early inbreds, and their crosses. Crop Sci. 1991;31:415-420. https://doi.org/10.2135/cropsci1991.0011183x.

8. Chen YH, Wang LM, Dai JR. Studies on the heterotic patterns between tropical subtropical and Chinese temperate germplasms in maize. Acta Agron Sin. 2000;26:557-564. https://doi.org/10.3321/j.issn:0496-3490.2000.05.007.

9. Hung HY, Shannon LM, Tian F, Bradbury PJ, Chen C, Flint-Garcia SA, McMullen MD, Ware D, Buckler ES, Doebley JF, Holland JB. ZmCCT and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. Proc Natl Acad Sci U S A. 2012;109:E1913-E1921. https://doi.org/10.1073/pnas.1203189109.

10. Sprague GF, Tatum LA. General vs. specific combining ability in single crosses of corn. J Am Soc Agron. 1942;34:923-932. https://doi.org/10.2134/ja42100923.

11. Griffing B. Concept of general and specific combining ability in relation to diallel crossing systems. Aust J Biol Sci. 1956;9:463-493. https://doi.org/10.1071/BI9560463.

12. Singh RK, Chaudhary BD. Biometrical methods in quantitative genetic analysis. New Delhi: Kalyani Publishers; 1985. ISBN 9781842657102.

13. Ma J, Cao YY. Genome-wide association analysis of yield traits and their specific combining ability in maize hybrid populations. Acta Agron Sin. 2024;50:363-372. https://doi.org/10.3724/SP.J.1006.2024.33022.

14. Tian HL, Yang Y, Yi HM, Xu LW, He H, Fan YM, Wang L, Ge JR, Liu YW, Wang FG, Zhao JR. New resources for genetic studies in maize (Zea mays L.): a genome-wide Maize6H-60K single nucleotide polymorphism array and its application. Plant J. 2021;105:1113-1122. https://doi.org/10.1111/tpj.15089.

15. Xue YJ, Zhao YK, Zhang YL, Wang R, Li XH, Liu ZH, Wang WW, Zhu SX, Fan YM, Xu LW, Zhao W, Zhao JR, Wang FG. Insights into the genomic divergence of maize heterotic groups in China. J Integr Plant Biol. 2025;67:1467-1486. https://doi.org/10.1111/jipb.13884.

16. Akinwale RO, Badu-Apraku B, Vroh I, Fakorede MAB. Heterotic grouping of tropical early-maturing maize inbred lines based on combining ability in Striga-infested and Striga-free environments and the use of SSR markers for genotyping. Field Crops Res. 2014;156:48-62. https://doi.org/10.1016/j.fcr.2013.10.015.

17. Badu-Apraku B, Oyekunle M, Fakorede MAB, Vroh I, Akinwale RO, Aderounmu M. Combining ability, heterotic patterns and genetic diversity of extra-early yellow inbreds under contrasting environments. Euphytica. 2013;192:413-433. https://doi.org/10.1007/s10681-013-0876-4.

18. Wu YS, Vicente FS, Huang KJ, Dhliwayo T, Costich DE, Semagn K, Nair S, Olsen M, Prasanna BM, Zhang XC, Babu R. Molecular characterization of CIMMYT maize inbred lines with genotyping-by-sequencing SNPs. Theor Appl Genet. 2016;129:753-765. https://doi.org/10.1007/s00122-015-2646-8.

19. Suwarno WB. The usefulness of molecular markers approach for developing heterotic groups in maize. J Trop Crop Sci. 2014;1:4-10. https://doi.org/10.29244/jtcs.1.2.4-10.

20. Fan HH, Wang JH, Yan YY, Zhang QG, Wang LW, Song L, Wei JF, Li XH, Zhang DM, Guo JJ, Guo R, Song W. Molecular and genetic characterization of newly

912 released CIMMYT inbred maize lines. Plants. 2025;14:3866.
913 https://doi.org/10.3390/plants14243866.
914 21. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing
915 phylogenetic trees. Mol Biol Evol. 1987;4:406-425.
916 https://doi.org/10.1093/oxfordjournals.molbev.a040454.
917 22. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using
918 multilocus genotype data. Genetics. 2000;155:945-959.
919 https://doi.org/10.1093/genetics/155.2.945.
920 23. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population
921 structure with the assistance of sample group information. Mol Ecol Resour.
922 2009;9:1322-1332. https://doi.org/10.1111/j.1755-0998.2009.02591.x.
923 24. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals
924 using the software STRUCTURE: a simulation study. Mol Ecol.
925 2005;14:2611-2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x.
926 25. Guo R, Li TT, Zhang QG, Wang JH, Guo JJ, Wang LW, Song L, Yan YY, Zhang
927 DM, Wei JF, Li XH, Song W. Genetic and molecular characterization of maize
928 landraces from central China. Agronomy. 2024;14:1278.
929 https://doi.org/10.3390/agronomy14061278.
930 26. Rogers JS. Measures of genetic similarity and genetic distance. In: Studies in
931 Genetics VII. University of Texas Publication 7213; Austin; 1972. p.145-153.
932 27. Bernardo R. Breeding for Quantitative Traits in Plants (3rd ed.). Woodbury, MN:
933 Stemma Press; 2020.

28. Choquette NE, Weldekidan T, Brewer J, Davis SB, Wisser RJ, Holland JB. Enhancing adaptation of tropical maize to temperate environments using genomic selection. G3 (Bethesda). 2023;13:jkad141. https://doi.org/10.1093/g3journal/jkad141.

29. Sang ZQ, Wang H, Yang YX, Zhang ZQ, Liu XG, Li ZW, Xu YB. Epistasis activation contributes substantially to heterosis in temperate by tropical maize hybrids. Front Plant Sci. 2022;13:921608. https://doi.org/10.3389/fpls.2022.921608.

30. Buckler ES, Holland JB, Bradbury PJ, McMullen MD. The genetic architecture of maize flowering time. Science. 2009;325:714-718. https://doi.org/10.1126/science.1174276.

31. Souza Neto IL de, Figueiredo AST, Uhdre RS, Contreras-Soto RI, Scapim CA, Zanotto MD. Combining ability and heterotic pattern in relation to F1 performance of tropical and temperate-adapted sweet corn lines. Bragantia. 2022;81:e3422. https://doi.org/10.1590/1678-4499.20220056.

32. Fan XM, Chen HM, Tan J, Xu CX, Zhang YM, Huang YX, Kang MS. A new maize heterotic pattern between temperate and tropical germplasms. Agron J. 2008;100:917-923. https://doi.org/10.2134/agronj2007.0298.

33. Adu GB, Badu-Apraku B, Akromah R, Awuku FJ. Combining abilities and heterotic patterns among early maturing maize inbred lines under optimal and Striga-infested environments. Genes. 2022;13:2289. https://doi.org/10.3390/genes13122289.

35. Reif JC, Melchinger AE, Frisch M. Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. Crop Sci. 2005;45:1-7. https://doi.org/10.2135/cropsci2005.0001.

36. van Heerwaarden J, Hufford MB, Ross-Ibarra J. Historical genomics of North American maize. Proc Natl Acad Sci U S A. 2012;109:12420-12425. https://doi.org/10.1073/pnas.1209275109.

37. Bernardo R. Breeding for Quantitative Traits in Plants (2nd ed.). Woodbury, MN: Stemma Press; 2010. ISBN 978-0-9720724-1-0.

38. Melchinger AE. Genetic diversity and heterosis. In: Coors JG, Pandey S, editors. Genetics and Exploitation of Heterosis in Crops. Madison, WI: ASA-CSSA-SSSA; 1999. p.99-118. https://doi.org/10.2134/1999.geneticsandexploitation.c10.

39. Wegary D, Vivek B, Labuschagne M. Association of parental genetic distance with heterosis and specific combining ability in quality protein maize. Euphytica. 2013;191:205-216. https://doi.org/10.1007/s10681-012-0757-2.

40. Schrag TA, Möhring J, Maurer HP, Dhillon BS, Melchinger AE, Piepho HP, Sørensen AP, Frisch M. Molecular marker-based prediction of hybrid performance in maize using unbalanced data from multiple experiments with factorial crosses. Theor Appl Genet. 2009;118:741-751. https://doi.org/10.1007/s00122-008-0934-9.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.xlsx
- Additionalfile2.xlsx
- Additionalfile3.xlsx
- Additionalfile49.docx