Journal of Advanced Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Advanced Research

journal homepage: www.elsevier.com/locate/jare

Genetic architecture of heterosis in maize NCII breeding populations

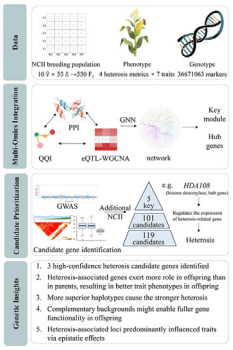
Ying Chen [a], Fei Gao [b], Jingtian Wang [a], Aifang Wang [b], Miaomiao Zhao [a], Yanhong Hu [b], Qiong Zhao [a], Yibo Wang [b], Guoping Shu [b,*], Yuan-Ming Zhang [a,*]

[a] College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China
[b] Beijing Lantron Seed; Yuan Long Ping High-tech Agriculture Co. LTD, Zhengzhou 450001, China

HIGHLIGHTS

- Validates three high-confidence genes (*HDA108*, *MADS3*, *Zm00001eb080510*) in multiple environments.
- Complementary parental genetic backgrounds enable enhanced gene function in hybrids.
- Identifies 101 heterosis-associated genes with robust haplotype effects across two populations, 6 hub genes and one heterosis module.
- Superior haplotype accumulation in hybrids correlates strongly with heterosis strength.
- Epistasis explains up to 89.6% of genetic variance in heterosis.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 18 October 2025
Revised 17 May 2026
Accepted 29 May 2026
Available online xxxx

Keywords:
Heterosis
Genome-wide association study
Epistasis
Candidate gene
Maize

ABSTRACT

Introduction: Despite the extensive utilization of heterosis in modern breeding, its genetic architecture remains incompletely understood.
Objectives: This study aimed to decipher the genetic basis of heterosis in maize.
Methods: Four heterosis indices derived from seven traits in a 2023 maize NCII breeding population were analyzed in association with 36,671,063 genetic markers.
Results: Among the 119 candidate genes identified through multi-omics integration and haplotype analyses, 101 exhibited consistent and significant haplotype–heterosis associations across both the 2019 and 2023 NCII breeding populations. Five high-confidence candidate genes were highlighted, with *HDA108*, *MADS3*, and *Zm00001eb080510* repeatedly detected across multiple environments and populations. eGWAS linked the known heterosis-related genes *ZAR1*, *bZIP29*, and *ACCO2* to one, two, and two candidate genes, respectively. Five candidate genes showed co-localization of eQTLs and heterosis-associated QTNs within linkage disequilibrium windows (2.60 to 8.63 kb). Integration of multi-dimensional networks using a graph neural network framework identified six hub genes, including *HDA108* (degree = 224). Notably, complementary genetic backgrounds of parental lines enabled more complete gene functionality in hybrid offspring. Candidate genes exhibited substantially more significant haplotype–trait associations in F_1 hybrids (98) than in parental lines (35), and 97 candidates showed significant allele frequency divergence (*Fst* > 0.2) between parental groups. The number of superior haplotypes was significantly correlated with both best-parent and mid-parent heterosis across the two breeding populations. These findings demonstrate that heterosis-related genes exert stronger functional effects in hybrids than in inbred

* Corresponding authors.
E-mail addresses: xugp2011@163.com (G. Shu), soyzhang@mail.hzau.edu.cn (Y.-M. Zhang).

https://doi.org/10.1016/j.jare.2026.05.049

Please cite this article as: Y. Chen, F. Gao, J. Wang et al., Genetic architecture of heterosis in maize NCII breeding populations, Journal of Advanced Research, https://doi.org/10.1016/j.jare.2026.05.049

parents, thereby enhancing trait performance in offspring. Furthermore, variation in the number of superior haplotypes directly contributed to heterosis differences among hybrids. All heterosis-associated QTNs primarily influenced trait phenotypes through epistatic interactions, which emerged as the predominant genetic driver of heterosis.

Conclusion: This study has identified some novel candidate genes and elucidated some key genetic mechanisms underlying heterosis, providing some valuable insight and resource for precision hybrid breeding.

Introduction

Over the past century, heterosis has played a crucial role in enhancing crop productivity and ensuring global food security [1]. Maize was among the first crops to harness heterosis, achieved via F_1 hybrids generated from controlled crosses between inbred lines [2]. Despite its widespread application in modern breeding programs, the underlying genetic mechanisms governing heterosis remain elusive [3].

Three classical genetic hypotheses were proposed to explain the heterosis in the 20th century: dominance [4], over-dominance [5], and epistasis [6] hypothesis. The predominant roles of non-additive effects (e.g., epistasis) and environmental dependencies in hybrid vigor formation have been studied through large-scale genome-wide association studies (GWAS) [2,7–10] and multi-environment phenotypic analyses [11]. Genomic studies reveal that heterosis arises from allelic interactions between parental genomes [12], parental genetic complementarity [13], allelic introgression of rare advantageous alleles [14–16], and dynamic allele-specific expression patterns with tissue-specific dominance effects [17]. Although significant advances have been made in understanding the underlying mechanism of heterosis, critical gaps remain and the theoretical basis of heterosis has not been fully cleared.

Although many heterosis loci have been mapped, the number of heterosis-related genes that have undergone functional validation is limited due to the complexity of both heterosis and the genetic background of germplasm, which makes validation difficult [18], e.g., *RH8* [19], *S5* [20] and *OsMADS1* [21] in rice; *ZAR1* and *ACCO2* [13], *ZmCCA1* [22] and *bZIP29* [23] in maize. The limited number of validated heterosis genes restricts our mechanistic understanding and breeding applications, as the polygenic nature of heterosis likely involves the coordinated action of numerous genes, which could benefit from multi-omics data.

To address these questions, 10 maternal lines were crossed with 55 paternal lines to generate 550 F_1 hybrids, constituting a single maize NCII breeding population in 2023 (2023 NCII population). This population was subjected to GWAS and multi-omics analyses to identify candidate genes and decipher the genetic architecture of heterosis. The GWAS incorporated four phenotypic metrics, including best-parent heterosis (BPH), mid-parent heterosis (MPH), general combining ability (GCA), and specific combining ability (SCA), derived from seven agronomic traits. To unravel regulatory networks, we integrated expression quantitative trait locus (eQTL) mapping with weighted gene co-expression network analysis (WGCNA), protein–protein interaction (PPI) prediction, and epistatic genetic network analysis to pinpoint hub genes governing heterosis. Importantly, an additional 89 × 8 NCII breeding population in 2019 (2019 NCII population) was employed to validate the associations between gene haplotypes and heterosis metrics and to validate the key heterosis candidates obtained from GWAS in the 2023 NCII population. Both populations were further used to perform regression analyses of trait phenotypes against the number of elite haplotypes. This integrated strategy aims to uncover novel heterosis-related genes and refine the genetic framework of hybrid breeding, thereby facilitating more precise breeding practices.

Material and methods

Population and heterosis calculation

The 2023 NCⅡ breeding population was composed of 55 paternal lines, 10 maternal lines, and 550 F_1 hybrids. The parental lines were planted in Zhengzhou, and the F_1 hybrids were planted in Zhengzhou, Luohe, Shijiazhuang, and Xuchang in 2023. The 2019 NCII breeding population was composed of 8 paternal lines, 89 maternal lines and 712 F_1 hybrids. The parental lines were planted in Zhengzhou, and the F_1 hybrids were planted in Zhengzhou, Luohe, and Baixiang in 2019. The parental lines of the 2019 and 2023 NCII breeding populations are derived from six elite founder inbred lines, and four of them are derived from the Stiff Stalk (SS) heterotic group and two from the Non-Stiff Stalk (NSS) group (Fig. S1; Table S1). These two NCII populations shared 7 paternal lines, 10 maternal lines, and 70 F_1 hybrids (9.83% of total F_1 offsprings). The measurement method was shown in Supplemental method. The quality control of the trait phenotypes was conducted by LongPing HighTech Maize Innovation Center, including no extreme phenotypic values.

Best linear unbiased prediction (BLUP) values were obtained using *lme4* v1.1.31 [24]. GCA and SCA were calculated by *Sommer* v4.2.0.1 [25]. General combining ability (GCA) and specific combining ability (SCA) were estimated using a mixed linear model implemented in the R package *Sommer* v4.2.0.1. Specifically, a NCII mating design-based model was fitted as:

$$y = X\beta + Z_f f + Z_m m + Z_h h + \varepsilon$$

where y represents the phenotypic values of the F_1 hybrids, β denotes the fixed effects including the overall mean and environmental effects, f and m represent the random effects of maternal and paternal lines corresponding to GCA, respectively, h denotes the random effects of hybrids corresponding to SCA, and ε is residual error. The corresponding variance components as well as BLUP values of GCA and SCA were obtained using the *mmer* function in the *sommer* package.

The maternal and paternal lines and their corresponding F_1 hybrids were simultaneously measured for phenotypes in the same field environment (Zhengzhou). Using the phenotypes in Zhengzhou, MPH and BPH (%) were calculated as:

$$MPH = \frac{F_{ij} - (P_i + P_j) \times 0.5}{(P_i + P_j) \times 0.5} \times 100$$

$$BPH = \frac{F_{ij} - Max(P_i, P_j)}{Max(P_i, P_j)} \times 100$$

where F_{ij} is the F_1 hybrid of P_i and P_j, $Max(P_i, P_j)$ is the maximum of P_i and P_j trait values. To ensure data reliability, field trials were conducted using a randomized complete block design with three replicates, with the commercial hybrid Yufeng 303 included in each

block as a common control. Trait values were calculated as the mean of each trait for each parent or hybrid across replicates.

Genomic data

The marker genotypes of 145 parents in the 2019 and 2023 NCII breeding populations were obtained through high-depth, whole-genome resequencing. The method of SNP variant calling was shown in Supplemental method. The *bcftools* v1.20 [26] was used to integrate the SNPs and indels. *Beagle* v5.2 [27] was used to impute missing markers, while *Plink* v1.90 [28] was employed to filter markers with minor allele frequency (MAF) < 1%. Finally, 36,671,063 high quality markers were used for subsequent analysis. Offspring genotypes are inferred from their parental genotypes [29,30]. For example, parental genotypes AA (P_1) and aa (P_2) were used to infer the F_1 genotype Aa. When parental genotypes were heterozygous, F_1 genotypes were inferred probabilistically based on expected Mendelian segregation ratios.

Genome-wide association study

Fast3VmrMLM [29] was used for GWAS. In GWAS, three principal components (PCs) calculated by *flashpcaR* v2.1 [31] was used to control for population structure, and the kinship matrix was calculated by *Fast3VmrMLM*. The *Q* matrix of population structure was calculated by *Admixture* v1.3.0 [32] (Table S1) to clarify the genetic background of these parents. The Bonferroni correction threshold ($P = 1.36 \times 10^{-9}$) was used to determine significant QTLs, while LOD = 3.0 was used to determine suggested QTLs [29,33]. Marker annotation was conducted using *snpEff* v4.3t [34] based on B73V5 genome. *VCFtools* v0.1.16 [35] was used to calculate fixation index (F_{st}) for comparing paternal and maternal lines.

Dominance ratio

The additive (a) and dominance (d) effects of significant or suggestive loci were estimated using Fast3VmrMLM [29]. The dominance ratio (d/a) was subsequently calculated. According to the locus classification criteria described by Jiang et al. [36], loci with $|d/a| < 0.2$, $0.2 \le |d/a| < 0.8$, $0.8 \le |d/a| < 1.2$, and $|d/a| \ge 1.2$ were categorized as additive, partial dominance, complete dominance, and overdominance, respectively.

Identification of candidate genes

Differential expression analysis was based on transcriptome data from Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo; accession number: GSE115796, GSE124543 and GSE155947) using *DESeq2* v1.38.3 [37] with the criterion of $|\log_2\text{FC}| > 1$ and P value < 0.05. RNA-seq datasets from 14-day-old seedlings of parents B73 and Mo17 and their F_1 hybrid were obtained to compare the gene expression levels between the parents and their F_1 hybrids. Gene annotation was performed using *clusterProfiler* v4.6.2 [38] based on the annotation downloaded from MaizeGDB (https://download.maizegdb.org/Zm-B73-REFERENCE-NAM-5.0). Kyoto Encyclopedia of Genes and Genomes (KEGG) were conducted using *KOBAS-i* v3.0 (https://bioinfo.org/kobas). Known rice genes were obtained from the China Rice Data Centre (https://ricedata.cn/gene/) and previous studies. Maize protein sequences (https://download.maizegdb.org/Zm-B73-REFERENCE-NAM-5.0) and rice protein sequences (https://rice.uga.edu/, version 7.0) were used to find the homologous genes using *Othofinder* v2.3.8 [39]. The regulation map downloaded from PlantRegMap (https://plantregmap.gao-lab.org/) was used to build the relationship between known genes and candidate genes.

Transcriptome analysis of parent lines

Three parent lines with significant GCA differences were sampled for RNA-seq: stems at V6 stage and ears at R2 stage. Each sample was pooled from five plants. Total RNA quality was assessed using NanoDrop 2000 (purity and concentration) and Agilent 2100/4200 (integrity and quantity). RNA-seq libraries were prepared by mRNA purification with polyT beads, fragmentation (300–350 bp), and reverse transcription. After second-strand synthesis, double-strand cDNA ends were blunted, 3′ adenylated, and ligated with sequencing adaptors. Library fragments were purified, PCR-enriched, and purified again. Libraries from different samples were pooled and sequenced on an NGS platform by Berry Genomics Co., Ltd (Beijing, China; https://www.berrygenomics.com/). Raw reads were filtered to remove those with ambiguous bases (N), adaptor contamination, or ≥ 20% bases with Phred quality ≤ 5. Clean reads were aligned to the silva database with bowtie2 to remove rRNA, and the remaining reads were mapped to the B73V5 reference genome using HISAT2. Gene read counts were obtained with featureCounts and subsequently normalized to transcripts per million (TPM). The correlation analysis was conducted by R function cor.test using the GCA of parents and gene expression.

Haplotype analysis

The gene haplotypes were built by Tag SNPs obtained by *Plink* v1.90 [28]. The potential candidate genes that were significant in the haplotype analysis using one-way ANOVA were considered as candidate genes. Superior haplotype was determined by LSD multi-comparison. The effect of the number of gene superior haplotypes on F_1 heterosis was calculated by *lm* function in R.

Datasets in expression genome-wide association study

Transcriptomic and genotypic datasets of expression GWAS (eGWAS) were obtained from Pang et al. (2019) [40], where kernels of 282 inbred lines were collected at 5 days after pollination. To align reference genome versions, *minimap2* v2.17-r941 (https://github.com/lh3/minimap2) was used for the alignment of B75V5 and B73V3 genomes (https://download.maizegdb.org/B73_RefGen_v3/), and then *transanno* v0.4.5 (https://github.com/informationsea/transanno) was used to create chain file from *minimap2* result. *Crossmap* v0.2.6 [41] was used to convert the genotypes to B73V5 genome. *Beagle* v5.2 [27] was used to impute missing markers, while *Plink* v1.90 [28] was employed to filter SNPs based on MAF < 1%. The transcriptome data was normalized to TPM based on the GFF file (https://download.maizegdb.org/B73_RefGen_v3/). Gene expression exceeded 0.1 in at least 5% of individuals in the population was used for eGWAS.

Gene network

The gene network contains gene co-expression network, PPI and genetic network. First, transcriptome data downloaded from MaizeGDB (https://qteller.maizegdb.org/) was used to build the gene co-expression network using *WGCNA* v1.72–1 [42], and then combined with the result of eGWAS to obtain the gene expression network. Then, PPI was predicted by STRING (https://string-db.org) to obtain the PPI network. Finally, all the significant/suggested heterosis quantitative trait nucleotides (QTNs) were used to construct epistasis network using empirical Bayes [43].

$$y = \mu + u_a + u_d + u_{aa} + u_{ad} + u_{da} + u_{dd} + \varepsilon$$

where u_a and u_d is additive and dominant effects, respectively; u_{aa}, u_{ad}, u_{da} and u_{dd} are additive-by-additive, additive-by-

dominant, dominant-by-additive, and dominant-by-dominant interaction effects, respectively; ε is residual error; and μ is total average. Wald test was used to calculate the *P* value. The threshold was set as 0.01. The relationship was built between candidate genes around the significant QTN-by-QTN interactions (QQIs).

Integration of gene networks

We integrated three gene networks (STRING, eGWAS-WGCNA and epistasis) using a graph neural network framework:

(1). Data Integration: Edge features (STRING combines scores, $-\log_{10}P$ value in eGWAS, WGCNA weights, and epistasis *P* value) were merged. The *Node2Vec* function was used to generate 16D embeddings (zero-padded for missing nodes).
(2). Ensemble Graph Model: An ensemble of three EnhancedGATEncoders with multi-head attention ($8 \rightarrow 4 \rightarrow 1$ heads) and skip connections was trained for 300 epochs (Adam optimizer, lr = 0.001). Loss combined edge reconstruction and attention regularization.
(3). Module Detection: Louvain communities were ranked by L2-norm of node embeddings. Results included module importance rankings and Cytoscape files.
(4). Validation: Bootstrap stability (50 resamples, Jaccard similarity) ensured robustness.

Genetic variance estimation using BGLR

Reproducing kernel Hilbert space (RKHS) regression, implemented by *BGLR* v1.1.4, was used to estimate genetic variance components in the below model.

$$y = X\beta + u_{main} + u_{epi} + \varepsilon$$

where *y* is a vector of heterosis indicator, *X* includes intercept and population structure (top three PCs), $u_{main} \sim N(0, \sigma^2_{main}K_{main})$, $u_{epi} \sim N(0, \sigma^2_{epi}K_{epi})$ and $\varepsilon \sim N(0, \sigma^2_e I)$ is residual error. The model was fitted using 12,000 MCMC iterations with a burn-in of 2,000. The main and epitasis variances were estimated as the posterior means of the corresponding components.

Results

Genetic architecture and heterosis-related genes

The seven agronomic traits analyzed in the 2023 NCII population displayed significant heterosis, including plant height (PH), ear height (EH), kernel row number (KRN), kernel number per row (KNR), 100-seed weight (SW), ear thickness (ET), and yield per plot (Yield). Both BPH and MPH exhibited considerable variation among hybrids (Fig. 1A).

Using the Fast3VmrMLM method, four heterosis indicators for each trait were associated with approximately 36 million markers. This analysis detected 671 significant QTNs ($P < 1.36 \times 10^{-9}$) and 28 suggested (LOD score ≥ 3) QTNs [29]. After removing duplicates, 690 unique QTNs were retained, of which 37.24% were located within genic regions (Fig. 1B; Table S2). To further characterize QTNs in potential regulatory regions, promoter regions were defined as the 2 kb sequences upstream of the transcription start site. Accordingly, 29 QTNs (4.20%) were located within promoter regions, suggesting a potential role of transcriptional regulation in heterosis (Table S2).

Notably, QTNs associated with GCA were predominantly characterized by additive effects (dominance ratio ≈ 0), whereas those linked to SCA were enriched with over-dominant alleles ($|d/a| \geq 1.2$), consistent with established heterosis genetic models (Table S2; Fig. 1C). For MPH and BPH QTNs, additive and dominance effects were further quantified. Overdominant QTNs accounted for 55.37% and 53.21% of MPH- and BPH-associated loci, respectively, whereas additive and partial-dominant QTNs contributed 24.76%–29.64% and 9.64%–11.73%, respectively. These results indicate that overdominance represents the predominant genetic basis of heterosis, with additive and partial dominance effects also contributing substantially (Table S2; Fig. 1D). Notably, distinct genetic architectures were observed among traits. For example, additive and partial-dominant QTNs underlying SW-derived MPH (50.00%) and BPH (66.15%) collectively accounted for significantly larger proportions than overdominant QTNs for SW-derived MPH (44.00%) and BPH (21.54%) (Table S2). These findings suggest that heterosis for SW is primarily governed by additive and partial dominance effects rather than overdominance, highlighting pronounced trait-dependent differences in the genetic basis of heterosis.

Population differentiation analysis further revealed strong allele frequency divergence (*F*st > 0.2) at 15.6% (108/690) of the heterosis-related QTNs between the paternal and maternal groups (Fig. 1E). In the genomic regions surrounding the heterosis-associated QTNs, 15 trait-related (non-heterosis) genes have been previously reported, including known candidates such as *D9* [44] and *HDA108* [45] (Table S2; Fig. 2A). To further identify candidate genes potentially involved in heterosis, we performed Gene Ontology (GO) annotation, homology searches against known rice genes, differential expression analysis, and gene haplotype analysis.

Using transcriptomic data from Luo et al. [46], we identified 208 differentially expressed candidate genes between parents and their hybrid offspring ($|\log_2 FC| > 1$, *P* value < 0.05; Fig. 2B), six of which were homologous to previously reported rice genes (Table S4). Linkage disequilibrium (LD) blocks among markers within the candidate gene region were analyzed to support haplotype classification using the tag-SNPs derived from *Plink*, identifying 119 potentially candidate genes with significance at the 0.05 level (Table S5). Among these, 97 potentially candidates exhibited significant allele frequency divergence (*F*st > 0.2) between paternal and maternal groups (Table S6). Functional annotations revealed that these 119 candidate genes were mainly involved in metabolic processes, stress response, ion transport, and cell communication (Fig. 2C). For example, *HDA108* (*Zm00001eb177560*), which is associated with MPH for EH, is known to regulate plant developmental processes including plant height [45]. Its haplotype Hap1 conferred approximately 10% higher MPH compared to Hap2 (*P* value = 1.76×10^{-4}; Fig. 2D). Markers within this gene region also showed pronounced divergence between paternal and maternal lines (Fig. 2D).

In summary, through an integrated approach incorporating transcriptomic profiling, cross-species homology, gene annotation, and haplotype-heterosis association analysis, we identified 119 candidate genes that might be associated with dissecting the genetic basis of heterosis.

Integrated eQTL analysis of candidate genes

To investigate the potential role of candidate heterosis-related genes, expression levels of 113 out of the 119 candidate genes, along with four previously known heterosis-associated genes, were utilized as phenotypic traits in an eGWAS (Fig. 3A). Only genes with expression levels greater than 0.1 in at least 5% of the accessions were included, resulting in 117 genes being analyzed. The analysis revealed five candidate heterosis-associated genes located near eQTLs linked to the expression of established heterosis-related genes *ZAR1*, *bZIP29* and *ACCO2*. Specifically, one candidate gene (*Zm00001eb250560*) was associated with *ZAR1*, two (*Zm00001eb366040* and *Zm00001eb366090*) with *bZIP29*, and two (*Zm00001eb330020* and *Zm00001eb107940*) with *ACCO2*, with *P*

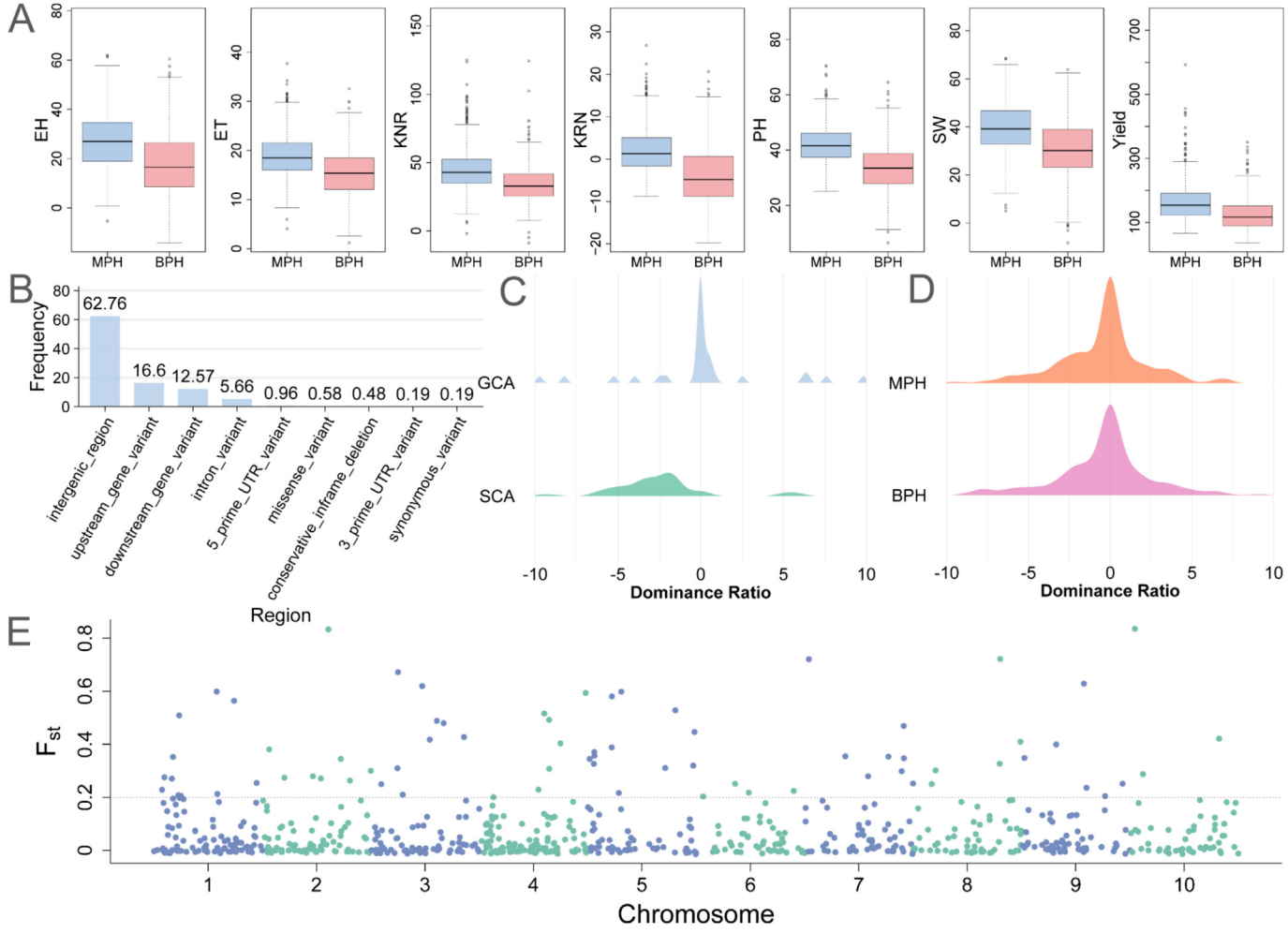


Fig. 1. Characterization of heterosis and heterosis-associated loci. (A) Best-parent and mid-parent heterosis (BPH and MPH) for seven traits in a maize 2023 NCII breeding population: plant height (PH), ear height (EH), kernel row number (KRN), kernel number per row (KNR), seed weight (SW), ear thickness (ET) and yield per plot (Yield). (B) Distribution of heterosis-associated loci in maize genome. (C) Dominant ratio of loci associated with general combining ability (GCA) and specific combining ability (SCA). (D) Dominant ratio of loci associated with BPH and MPH. (E) Genetic diversity of heterosis-associated loci between maternal and paternal lines.

values ranging from 7.49×10^{28} to 1.72×10^{5} (Figs. 3B and S2; Table S7). It should be noted that *Zm00001eb366090* (*ZmGA20ox5*) for BPH from PH is a gibberellin-associated gene [47], while *Zm00001eb061930* (*bZIP29*) is a known heterosis gene [23]. These results revealed potential *trans*-regulatory interactions between candidate and known heterosis-related genes and enhancing the credibility of candidate genes.

Of these 113 candidate genes, five exhibited co-localization of their eQTLs and heterosis-associated QTNs within the same LD windows, with physical distances ranging from 2.60 to 8.63 kb, implying possible shared genetic variants, with *P* values ranging from 1.20×10^{11} to 8.92×10^{121} (Fig. 3C and S3; Table S8). A notable example is *Zm00001eb215710*, for which an eQTL of this gene and a QTN associated with MPH for ET were located within a shared LD window. This gene also showed a significant association between its haplotypes and MPH for ET, supporting its role as a candidate gene influencing this trait. In summary, the co-localization of functional QTNs, combined with significant haplotype-heterosis associations, strengthens the evidence that these candidate genes contribute to heterosis.

Multi-dimensional networks integrated by graph neural network

To identify heterosis-related genes and gene-by-gene interactions, multi-omics network was integrated by three complementary methods (Fig. 3D). First, integrating eGWAS with WGCNA produced a significant co-expression network comprising 1,694 gene pairs. Second, all the detected QTNs associated with heterosis were incorporated into a full epistasis model to detect high-confidence QQIs. Using empirical Bayes estimation and Wald tests ($P < 0.01$), we identified 25 significant epistatic interactions involving 21 candidate genes. Finally, PPI prediction was performed between the 119 candidate genes and genes located near the detected QTNs using the STRING database, resulting in 1,219 predicted PPI pairs.

The three networks described above were integrated using a graph neural network framework. Topological patterns were captured through *Node2Vec* embedding, and edge features were consolidated within an ensemble learning framework implemented in PyTorch Geometric with customized graph attention networks. Using Louvain community detection coupled with L2-based module importance ranking, we reproducibly identified 31 modules, encompassing 1,046 nodes and 2,088 edges (bootstrap stability = 1.000 ± 0.000) (Fig. 3D). All genes in the resulting integrated network were associated with heterosis. The 'Blue' module was identified as the most significant and exhibited strong enrichment in five key biological pathways: alpha-linolenic acid metabolism, glycerolipid metabolism, aminoacyl-tRNA biosynthesis, plant hormone signal transduction, and glycosylphosphatidylinositol (GPI)-anchor biosynthesis (corrected *P* values < 0.05; Fig. S4), high-

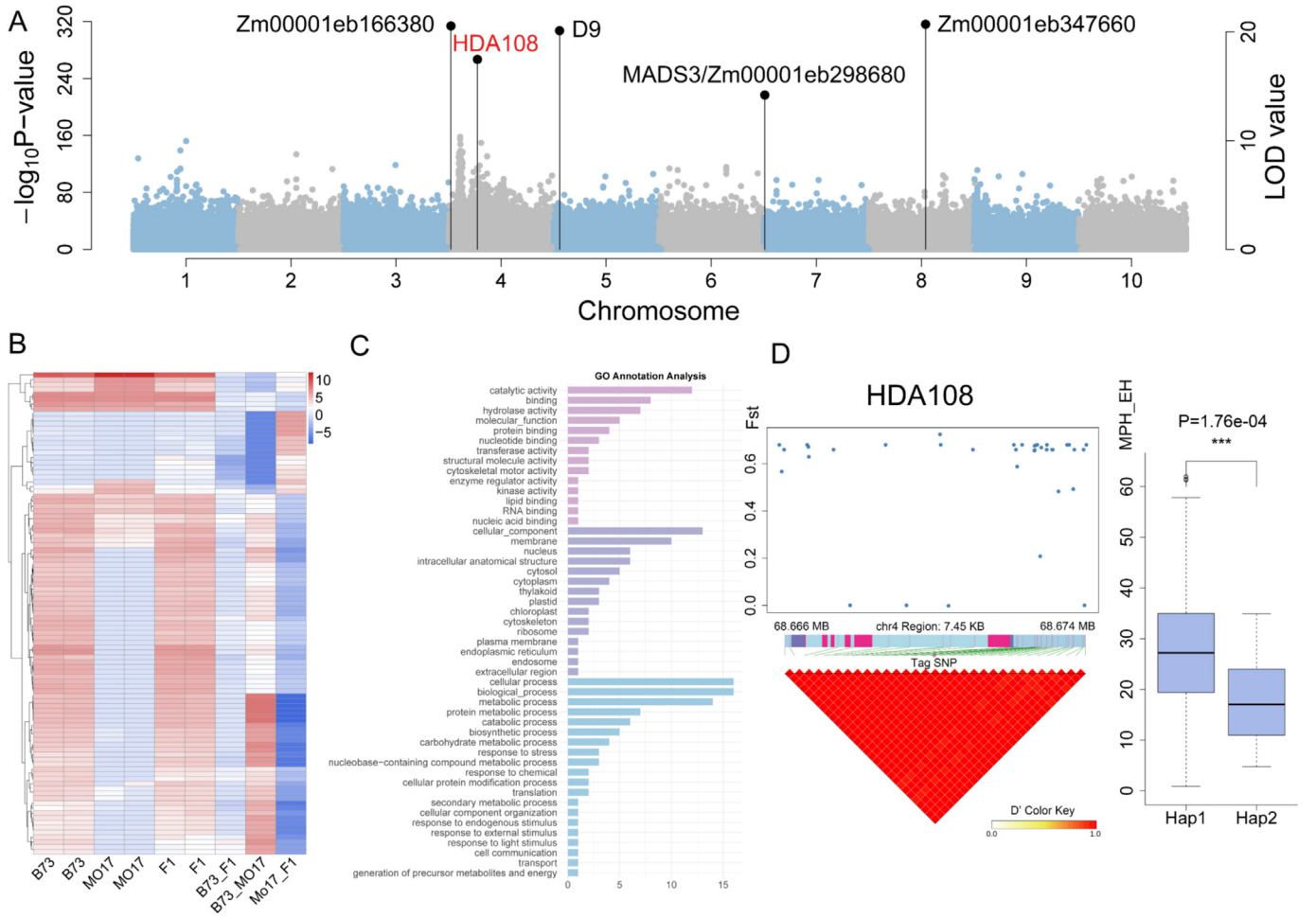


Fig. 2. Identification of candidate genes for heterosis. (A) Manhattan plot of mid-parent heterosis from ear height. *HDA108*, indicated by red color, is a key heterosis-associated candidate gene. (B) The $\log_2$(expression level + 1) in B73, Mo17 and their F_1 hybrid (two replicates) in columns 1 to 6 and their fold change of expression levels in columns 7 to 9 for heterosis-associated candidate genes (C) Gene ontology annotation analysis of heterosis-associated candidate genes. The molecular function, cell component and biological process were indicated by purple, deep purple and blue, respectively. (D) *F*st, linkage disequilibrium and haplotype analysis for heterosis-associated candidate gene *HDA108*. ***: $P < 0.001$. The LD heatmap was generated using the markers within the candidate gene region to construct gene haplotypes based on the tag-SNP (68,669,973 bp) obtained from *plink* software.

lighting their potential importance in heterosis. Furthermore, six hub genes with degrees exceeding 75 were identified, including *HDA108* (*Zm00001eb177560*) (Fig. 3E), suggesting that these genes may play central roles in influencing heterosis within the network.

Support for haplotype-heterosis associations in the 2019 and 2023 NCII populations

To support the haplotype-heterosis associations identified in the 2023 NCII population, we performed haplotype analysis of the 119 candidate heterosis genes in an additional 2019 NCII population comprising 712 hybrids. Results demonstrated that 101 candidate genes exhibited significant associations, and these candidates with an average r^2 of 1.90% (ranging from 0.066% to 17.646%) (Fig. 4 and S6; Tables S9 and S10). Among these, nine candidate genes corresponded to previously reported trait-related genes; 92 were significantly differentially expressed between parents and hybrids; six were homologous to known trait-related genes in rice; the eQTL of candidate gene *Zm00001eb215710* was co-localized with a QTL for MPH of ET; five were identified as hub genes in the integrated network; and nine were predicted to interact with known heterosis genes (Table S10). Overall, the haplotype-heterosis associations of 101 candidate genes were consistently robust across both populations (Figs. 4 and S5; Table S9).

To further elucidate the genetic basis of heterosis, multiple comparisons of heterosis indicators among haplotypes were conducted to identify superior haplotypes for the 101 candidate genes. For each candidate gene, high- and low-heterosis pools were constructed by selecting individuals from the top and bottom 5% of the distribution of the heterosis index associated with that gene. The frequency of each superior haplotype was compared between these pools. Notably, 96 candidate genes showed an average 14.70% higher frequency of superior haplotypes in the high pool compared to the low pool (Fig. 5A; Table S11), underscoring the reliability of these candidates and supporting an association between heterosis and the accumulation of superior haplotypes.

To strengthen this conclusion, linear regression analysis was performed to assess the relationship between BPH / MPH and the number of superior haplotypes per individual. A significantly positive correlation was observed (Fig. 5B). This trend was further validated in eight of the fourteen BPH/MPH phenotypes in the 2019 NCII population (Fig. 5C), despite the existence of genetic differences between the two populations. These consistent results highlight the crucial role of superior haplotype accumulation in influencing heterosis.

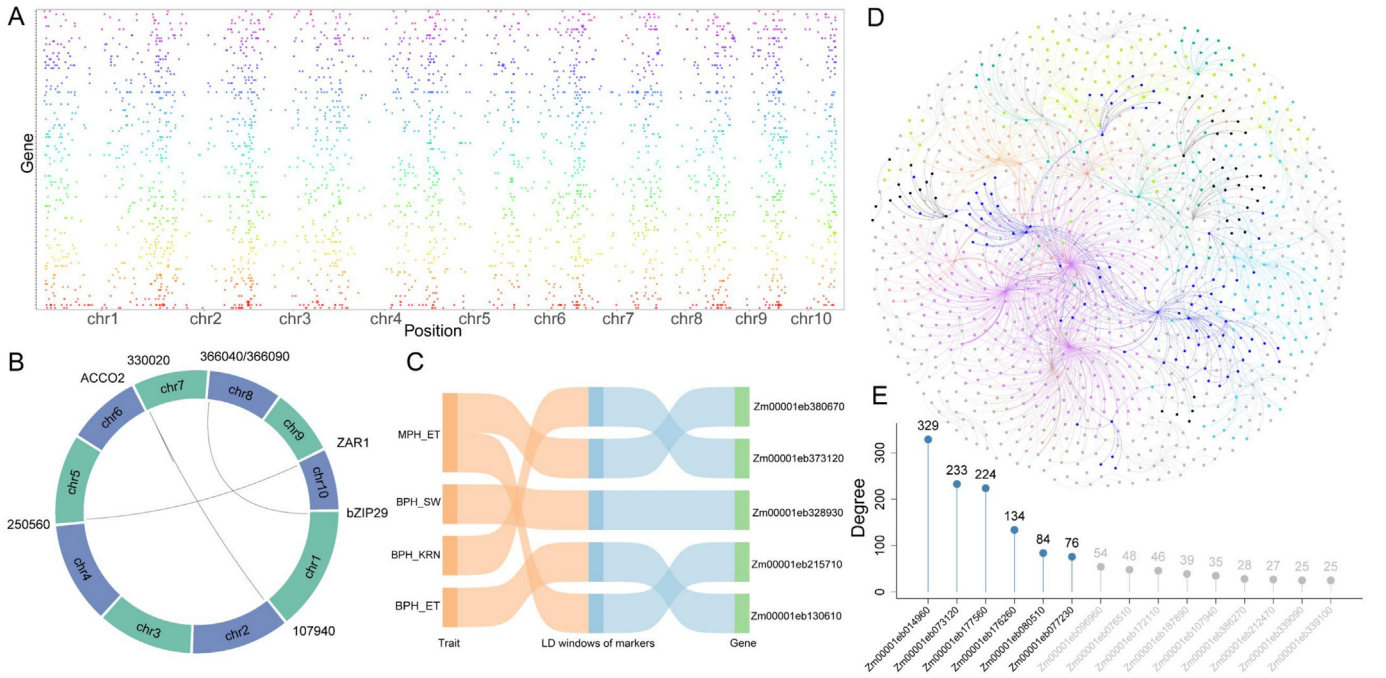


Fig. 3. Integration of multi-dimensional networks. (A) Distribution of eQTLs in a maize genome. (B) Trans-regulatory between heterosis-associated candidate genes and three known genes, in which '*Zm00001eb*' was omitted from all the candidate gene names. (C) Co-localization of eQTLs for heterosis-associated candidate genes and heterosis-associated QTLs. BPH: best-parent heterosis; MPH: mid-parent heterosis; EH: ear height; KRN: kernel row number; SW: seed weight; ET: ear thickness. (D) Multi-dimensional networks integrated by graph neural network framework. Different modules are indicated by different colors. (E) Degrees of the nodes for six hub genes (black) and others (gray) in the integrated network.

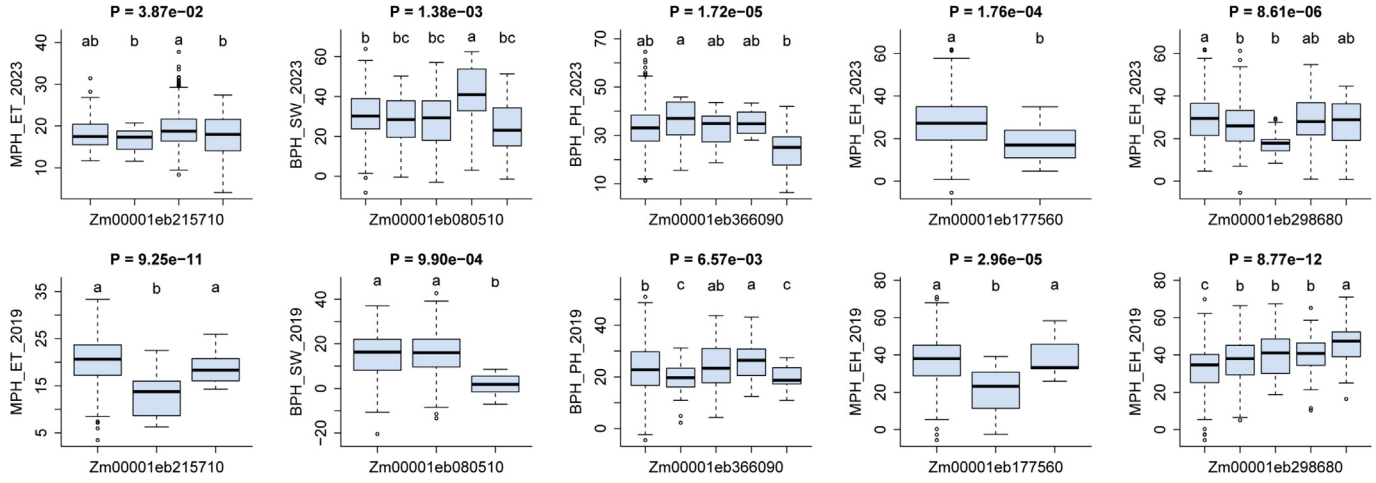


Fig. 4. Haplotype-heterosis association for five key candidate genes in the 2019 and 2023 NCII breeding populations. BPH: best-parent heterosis; MPH: mid-parent heterosis; PH: plant height; EH: ear height; SW: seed weight; ET: ear thickness.

Discussion

In this study, we employed maize NCII breeding populations to dissect the genetic architecture of heterosis, analyzing four heterosis metrics derived from seven agronomic traits. The NCII design was selected due to its ability to provide broader genetic diversity and more practical breeding guidance compared to F_2 populations [48,49]. Our results revealed that additive effects were predominantly enriched in QTNs associated with GCA, whereas over-dominant effects were mainly associated with SCA (Table S2; Fig. 1C), which is consistent with classical genetic models of heterosis.

A total of 119 candidate genes located near heterosis-associated QTNs were identified as potentially contributing to heterosis through an integrated approach that included haplotype analysis, differential expression screening, GO annotation, homology analysis with known rice genes, and literature mining (Tables S3–S5; Fig. 2C). Among these, 97 candidates exhibited significant allele frequency differentiation (*Fst* > 0.2) between parental groups (Table S6), and 101 were further validated via haplotype-heterosis association analysis in an additional 2019 NCII population (Fig. S5; Table S9).

By integrating co-expression, epistatic (QQI), and PPI networks within a graph neural network framework, we identified a highly

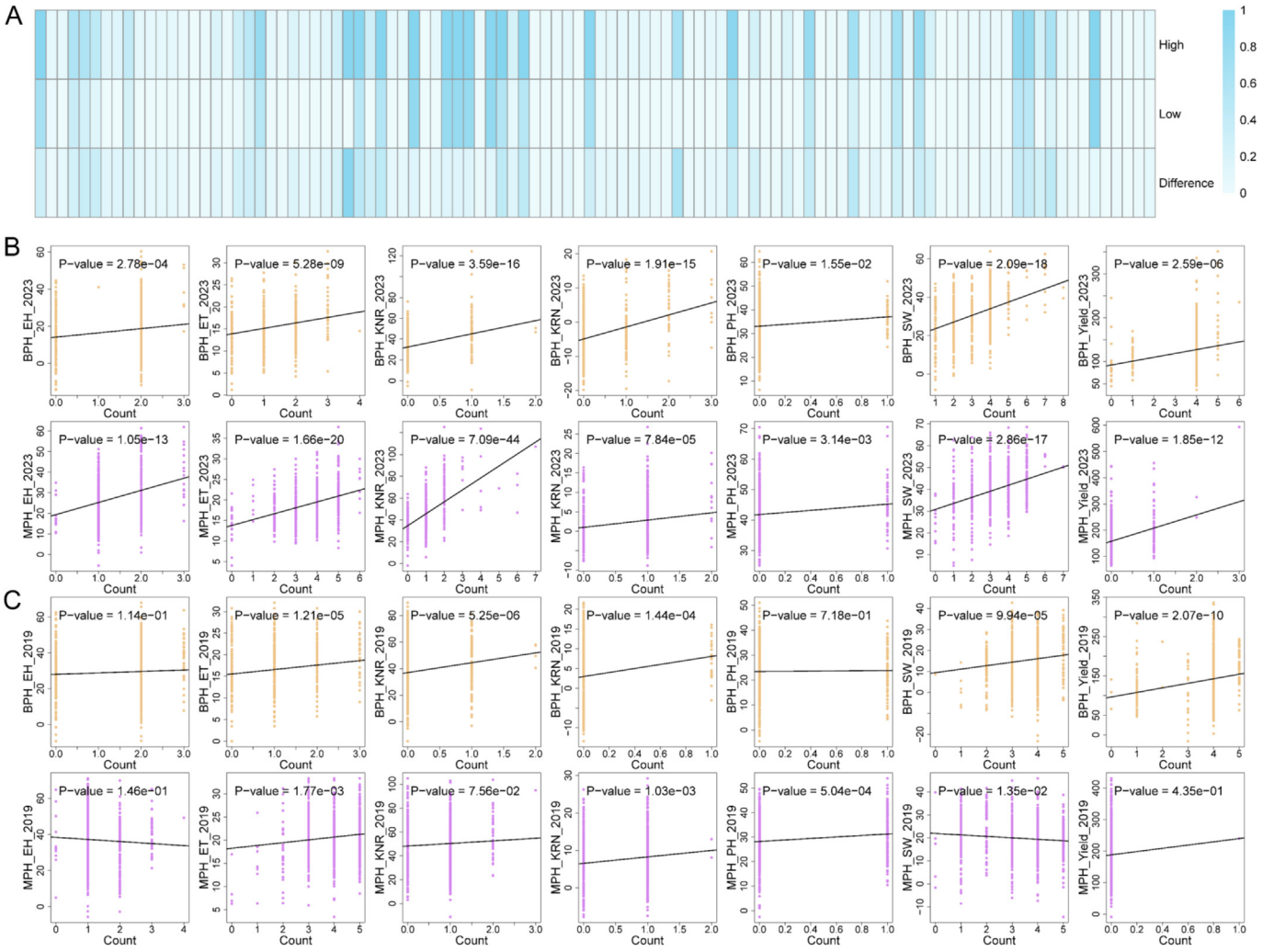


Fig. 5. Effect of superior haplotypes for heterosis-associated candidate genes on heterosis. (A) Frequency of superior haplotype for each candidate gene in high- and low-heterosis pools and their difference between the two extreme pools. (B and C) Regression of individual heterosis on the number of superior haplotypes carried by each individual in the 2019 (C) and 2023 (B) NCII breeding population. Best-parent (BPH) and mid-parent (MPH) heterosis are indicated by orange and purple colors, respectively. PH: plant height; EH: ear height; KRN: kernel row number; KNR: kernel number per row; SW: seed weight; ET: ear thickness; Yield: yield per plot.

significant "Blue" module, six hub genes, including *HDA108*, and nine candidates predicted to interact with four previously reported heterosis-related genes in maize (Fig. 3D and E). Based on multi-evidence support, we prioritized five high-confidence candidate genes for heterosis, evaluated through: (1) functional annotation and prior knowledge, (2) significant differential expression between parents and hybrids, (3) hub status in the integrated network, (4) predicted interactions with known heterosis genes, and (5) robust haplotype-heterosis associations across multiple populations.

Furthermore, our findings underscore the important role of epistasis in heterosis, and heterosis QTNs influence traits mainly through epistasis interactions. Finally, we demonstrate that the accumulation of superior haplotypes within candidate genes significantly contributes to heterosis, reinforcing the utility of superior haplotype in predictive breeding.

Key candidate genes for heterosis

Here, we identify five high-confidence candidate genes associated with heterosis: *Zm00001eb080510*, *Zm00001eb177560* (*HDA108*), *Zm00001eb215710*, *Zm00001eb298680* (*MADS3*), and *Zm00001eb366090* (*ZmGA20ox5*) (Tables 1, S4 and S9; Fig. 4). To enhance the reliability of these candidates, GWAS was performed using each-environment data for the 2023 population and across-environment average data for the 2019 population. Results indicated that four candidates were detected in multiple environments within the 2023 population, and three were detected in the 2019 population (Table S12). Among these, *Zm00001eb080510*, *HDA108*, and *MADS3* were consistently identified in both populations, though not necessarily for the same trait or heterosis metric, marking them as the most promising candidate genes. In detail, *HDA108* was found to be associated with the EH-derived MPH in the 2023 NCII population ($P = 3.43 \times 10^{-18}$) and the KNR- ($P = 2.50 \times 10^{-4}$) and KRN-derived ($P \approx 0$) GCA in the 2019 NCII population. *Zm00001eb080510* was found to be associated with the SW-derived BPH ($P = 4.27 \times 10^{-26}$) in the 2023 NCII population and the KRN-derived GCA ($P \approx 0$) in the 2019 NCII population. *MADS3* was found to be associated with the EH-derived MPH ($P = 6.51 \times 10^{-15}$) in the 2023 NCII population and the ET-derived GCA ($P = 7.69 \times 10^{-7}$) in the 2019 NCII population. To further enhance the reliability of these candidates, we used the G2F dataset comprising 4,423 hybrids [50] to identify their superior haplotypes for four of the five heterosis candidates (due to limited marker density) using LSD multiple comparisons ($\alpha = 0.05$). These genes were *Zm00001eb177560*, *Zm00001eb298680*, *Zm00001eb366090*, and *Zm00001e-*

Table 1
Evidence of five key heterosis-associated candidate genes.

Trait	Marker	*P* value	Gene	Gene name	Description	Differential expression	Rice homology	Co-localization	Degree of hub gene	Interacted with known gene
BPH_SW	2_42683321	4.27E-26	*Zm00001eb080510*	Inapplicable	tryptophan biosynthetic process	Yes	*TDD1*	No	84	No
MPH_EH	4_68813420	3.43E-18	*Zm00001eb177560*	*HDA108*	histone deacetylase	No	––	No	224	*ZmCCA1*, STRING
MPH_ET	5_9890550	3.58E-69	*Zm00001eb215710*	Inapplicable	alpha-tubulin protein	Yes	*SRS5*	Yes	No	No
MPH_EH	7_1817612	6.51E-15	*Zm00001eb298680*	*MADS3*	MADS-box transcription factor	No	––	No	No	*ZmCCA1*, PlantRegMap
BPH_PH	8_171488969	5.33E-136	*Zm00001eb366090*	*ZmGA20ox5*	gibberellin 20-oxidase	No	––	No	No	*bZIP29*, eGWAS

BPH: best-parent heterosis, MPH: mid-parent heterosis, PH: plant height, EH: ear height, SW: seed weight and ET: ear thickness.

b080510. We then calculated the frequencies of these superior haplotypes in extreme pools comprising the top and bottom 1% of hybrids with respect to trait phenotypes. The results showed that the four candidate genes were significant in both the haplotype and frequency difference analyses, with the exception that *Zm00001eb080510* was close to significance in the frequency difference analysis (Table S13). These results indicate the significant influence of these candidate heterosis genes on trait phenotypes between the two extreme hybrid pools.

HDA108 is prioritized as a key candidate influencing EH-derived MPH in four aspects. First, it encodes a histone deacetylase previously implicated in plant development and height regulation [45]. Second, it was identified as a hub gene in the integrated network with a degree of 224 (Table 1; Fig. 3E) and is predicted to interact with the known heterosis-related gene *ZmCCA1* [22] (Table 1; Table S10). Moreover, histone deacetylation is known to dynamically modulate gene expression and has been linked to heterosis in other species; for example, *OsHDT1* regulates expression patterns in hybrids and contributes to heterosis in rice [51–53]. Thus, as a highly connected regulator, *HDA108* may influence EH-derived MPH by modifying the expression of heterosis-related genes through epigenetic mechanisms.

Zm00001eb080510 is a strong candidate for SW-derived BPH. It participates in the tryptophan biosynthetic pathway and is a homolog of rice *TDD1*, which affects auxin biosynthesis via tryptophan metabolism [54]. This gene also acts as a hub in the network (degree = 84; Fig. 3E) and shows differential expression between parents and hybrids. Given that auxin signaling has been associated with heterosis in canola, sorghum, and *Arabidopsis* [55–57], variation in *Zm00001eb080510* expression may lead to differential auxin levels, thereby influencing phenotypic variation between parents and hybrids both directly and through the function of hub gene in the gene network (Table 1; Table S10).

MADS3 is another key candidate contributing to EH-derived MPH. As a MADS-box transcription factor, it has been reported to affect spikelet development and plant height [58]. It is predicted to interact with the known heterosis gene *ZmCCA1* [22] based on PlantRegMap analysis (Table 1; Table S10). Related MADS-box genes in rice, such as *OsMADS18*, which modulates plant architecture, and *OsMADS23*, involved in strigolactone and gibberellin crosstalk [59,60], supporting the functional relevance of this gene family. Notably, *OsMADS1* has been directly implicated in rice heterosis [21]. Thus, *MADS3* may influence plant architecture through hormone signaling pathways and protein interactions with *ZmCCA1* to regulate EH-derived MPH.

Although strong haplotype-heterosis associations were observed in both the 2019 and 2023 NCII populations (Fig. 4), and functional predictions were supported by genetic analyses, network inference, and previous experimental evidence, further validation through molecular and field experiments is required to confirm the roles of these candidate genes. Notably, plant hormone signal transduction was significantly enriched in the most important network module (Fig. S4), and three of the five key candidates are implicated in hormone-related processes (Table S10), underscoring the central role of hormonal regulation in heterosis. Furthermore, the graph neural network framework is a powerful tool for integrating multi-dimensional networks and capturing key modules and hub genes in integrated networks in the big data era.

Parental genomic complementarity enhances gene functionality in hybrids

To analyze the investigate between heterosis-related genes and trait phenotypes, haplotype analysis was conducted on the BLUP trait values across heterosis-related gene haplotypes. Significant associations between haplotypes and trait values were significant for 98 candidate genes (97.03%) in the offspring ($P < 0.05$; Table S14), compared to only 35 genes (34.65%) in the parents ($P < 0.05$; Table S15). This partly reveals the effect of most heterosis-related genes on the genetic architecture of offspring. We found that 97 candidates exhibited genetic diversity between maternal and paternal (Table S6), and that 88.31% of markers had higher heterozygous frequencies in the offspring than in the parents (Table S16). The complementary genetic backgrounds resulting from parental hybridization thus enhance the functional contributions of these genes in the hybrids. These results, obtained from a haplotype analysis of real data, are similar to the HoIIB model derived from simulation studies by Xie et al. [11]. In summary, heterosis-related genes exert more pronounced effects in hybrid offspring than in their inbred parents, leading to improved trait performance. Variation in the number of superior haplotypes among offspring underlies the observed differences (Tables S14 and S15; Fig. 5).

Crucial role of epistasis between heterosis QTNs in the genetic dissection of traits and heterosis

Previous studies have established the predominant contribution of non-additive effects, particularly epistasis, to hybrid vigor in crops such as rice, maize, and wheat [2,7–11,61]. In our inheritance analysis of both main and epistatic effects, epistasis accounted for a greater proportion of genetic variance than main effects across all heterosis metrics, in detail, 89.62% for BPH, 55.03% for MPH, and

56.04% for SCA, compared to 10.38%, 44.97%, and 43.96%, respectively, explained by main effects (Table S17). This highlights the central role of epistasis, especially in BPH. To evaluate how heterosis-associated QTNs influence trait phenotypes, we constructed a genome-wide epistasis model incorporating all detected QTNs and estimated interaction effects using empirical Bayes methods [43]. Notably, the number of significant QQIs substantially exceeded that of significant main-effect QTNs for all seven traits ($P < 0.01$; Table S18), indicating that epistatic interactions constitute the primary mechanism through which heterosis QTNs affect phenotypic variation. This further underscores the importance of genetic background compatibility between parental lines. Variance component analysis reinforced these findings, showing that epistasis, especially additive-by-additive (*aa*) and dominant-by-dominant (*dd*) interactions, explained the largest proportion of genetic variance for all traits (Table S19). GO annotation analysis was subsequently performed using genes located near QQIs. Three GO terms were significantly enriched ($P < 0.05$), including manganese ion binding, cell wall, and apoplast (Table S20). These functional categories are closely associated with photosynthesis, biomass accumulation, and intercellular transport, suggesting that these QQIs in maize are involved in the coordinated regulation of nutrient utilization, cell expansion, and physiological robustness. Thus, epistasis emerges as the most important genetic component underlying heterosis in the maize NCII breeding population.

Integrative framework for robust candidate gene prioritization in heterosis

To further strengthen the reliability of the 101 candidate genes, we referenced three independent transcriptomic maize heterosis datasets. First, Zhou et al. [62] identified allele-specific expressed genes in multiple hybrid combinations. Of these, 52 overlapped with our results and included *Zm00001eb080510* and *Zm00001eb215710*. Second, Shi et al. [63] performed RNA-seq analysis on 2–4 mm immature maize ears to explore the transcriptional basis of heterosis. Of their differentially expressed candidate genes, ten overlapped with our results and included *Zm00001eb080510*, which further supports their regulatory potential. Third, Hu et al. [64] investigated heterosis-related allele-specific expression during floret and spikelet differentiation stages, revealing 44 and 57 overlapping candidate genes, respectively (e.g., *Zm00001eb298680* and *Zm00001eb215710*). In total, 82 of the 101 candidate genes were independently supported by at least one external dataset, and 52 genes were validated by two or more datasets, providing strong cross-study evidence for their involvement in heterosis regulation (Table S21). It should be noted that the tissue types and developmental stages of an eGWAS dataset in Pang et al. [40] were consistent with those for five out of seven traits in this study, except for plant height and ear height. As previously mentioned, independent transcriptomic studies of maize heterosis at corresponding developmental stages [46,62–64] and two NCII breeding populations (in 2019 and 2023) repeatedly identified the aforementioned 101 candidate genes, which may be involved in the conserved regulatory processes associated with heterosis.

More importantly, we analyzed RNA-seq data from stem tissue at the V6 stage and ear tissue at the R2 stage in parental lines exhibiting significant variation in GCA. For plant height and ear height, correlations between gene expression level and GCA were calculated using stem RNA-seq data at the V6 stage, whereas ear RNA-seq data at the R2 stage were used for the remaining traits. Twelve candidate genes showed significant expression–GCA correlations, exemplified by *Zm00001eb177560* ($r = 0.9982$, $P = 0.038$), suggesting their potential regulatory roles in hybrid performance (Table S22).

It should be noted that although differential expression analysis was performed using publicly available RNA-seq data from Luo et al. [46], these transcriptomic results were intended to provide auxiliary support rather than serve as a primary criterion for candidate gene identification. Increasing evidence indicates that heterosis is governed by complex genetic mechanisms, in which heterosis-associated loci do not necessarily exhibit consistent differential expression across tissues or developmental stages, particularly when regulatory variation, epistasis, or post-transcriptional regulation is involved.

The candidate genes were prioritized based on a series of comprehensive analyses conducted in this study. Their identification primarily relied on consistent associations between haplotypes and heterosis observed across two NCII breeding populations. This was further supported by functional annotation of the genes, orthology with rice, analysis of genetic diversity, previous studies, and multi omics integration. Together, these lines of evidence constitute a robust, multi layered framework. Within this strategy, transcriptomic data were incorporated to provide additional biological context rather than to serve as a mandatory filtering criterion.

Conclusion

Of the 119 candidate genes identified in this study, 101 show the robust haplotype-heterosis associations across the 2019 and 2023 NCII breeding populations; five are prioritized with relatively enough evidence, in which three were detected by multiple environments and populations, including *HDA108*, *MADS3* and *Zm00001eb080510*. Heterosis-related genes exert more role in offspring than in parents, resulting in better trait phenotypes in offspring than in parents, and the more superior haplotypes in offspring result in the stronger heterosis, partially validating these heterosis genes and explaining the genetic basis of heterosis. Ninety-seven out of 101 candidates showed genetic diversity ($Fst > 0.2$) between parent groups, together with more haplotype-trait associations in offspring, suggesting that complementary backgrounds enable fuller gene functionality in offspring. Furthermore, heterosis-associated QTNs predominantly influenced trait and heterosis performances via epistasis in seven traits.

Compliance with ethics requirements

This article does not contain any studies with human or animal subjects.

Data availability

Transcriptome data was downloaded from NCBI Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo; accession numbers GSE115796 and GSE155947). The RNA-seq dataset for eGWAS could be accessed through NCBI GEO under accession number GSE110315, and the SNP dataset was download from the European Variation Archive (EVA; https://www.ebi.ac.uk/eva; accession number PRJEB24974). Transcriptome data for the gene co-expression network was downloaded from MaizeGDB (https://qteller.maizegdb.org/). The G2F maize dataset were downloaded from https://doi.org/10.25739/tq5e-ak26. The code for graph neural network framework is available at Supplemental files. The Fast3VmrMLM software can be downloaded from Github (https://github.com/YuanmingZhang65/Fast3VmrMLM). The data support the results can be downloaded from Supplemental files.

CRediT authorship contribution statement

Ying Chen: Formal analysis, Data curation, Writing – original draft, Writing – review & editing. **Fei Gao:** Investigation. **Jingtian**

Wang: Data curation. **Aifang Wang:** Investigation. **Miaomiao Zhao:** Data curation. **Yanhong Hu:** Investigation. **Qiong Zhao:** Data curation. **Yibo Wang:** Supervision, Investigation, Writing – review & editing. **Guoping Shu:** Supervision, Investigation, Writing – review & editing. **Yuan-Ming Zhang:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We are sincerely thankful for the valuable comments from editors and reviewers. This study was supported by the National Natural Science Foundation of China, China (32270673; 32470657) and State Key Laboratory of Crop Gene Resources and Breeding, China (CGRB-2025-02).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jare.2026.05.049.

References

[1] Hochholdinger F, Baldauf JA. Heterosis in plants. Curr Biol 2018;28:R1089–92. doi: https://doi.org/10.1016/j.cub.2018.06.041.
[2] Xiao Y, Jiang S, Cheng Q, Wang X, Yan J, Zhang R, et al. The genetic mechanism of heterosis utilization in maize improvement. Genome Biol 2021;22:148. doi: https://doi.org/10.1186/s13059-021-02371-1.
[3] Liu W, Zhang Y, He H, He G, Deng XW. From hybrid genomes to heterotic trait output: challenges and opportunities. Curr Opin Plant Biol 2022;66:102193. doi: https://doi.org/10.1016/j.pbi.2022.102193.
[4] Bruce AB. The Mendelian theory of heredity and the augmentation of vigor. Science 1910;32:627–8. doi: https://doi.org/10.1126/science.32.827.627-a.
[5] Shull GH. The genotypes of maize. Am Nat 1911;45:234–52. doi: https://doi.org/10.1086/279207.
[6] Powers L. An expansion of Jones's theory for the explanation of heterosis. Am Nat 1944;78:275–80. doi: https://doi.org/10.1086/281199.
[7] Jiang Y, Schmidt RH, Zhao Y, Reif JC. A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. Nat Genet 2017;49:1741–6. doi: https://doi.org/10.1038/ng.3974.
[8] Huang X, Yang S, Gong J, Zhao Q, Feng Q, Zhan Q, et al. Genomic architecture of heterosis for yield traits in rice. Nature 2016;537:629–33. doi: https://doi.org/10.1038/nature19760.
[9] Wei X, Chen M, Zhang Q, Gong J, Liu J, Yong K, et al. Genomic investigation of 18,421 lines reveals the genetic architecture of rice. Science 385 2024: eadm8762. doi: https://doi.org/10.1126/science.adm8762.
[10] M. Zhao, X. Han, A. Zheng, Y. Yan, H. Zhang, X. Ma, et al. Dissecting the genetic foundation of heterosis for rice complex traits using GWAS, TWAS and mGWAS, J. Integr. Agric. in press.
[11] Xie J, Wang W, Yang T, Zhang T, Zhang Q, Zhu Z, et al. Large-scale genomic and transcriptomic profiles of rice hybrids reveal a core mechanism underlying heterosis. Genome Biol 2022;23:264. doi: https://doi.org/10.1186/s13059-022-02822-8.
[12] Chen ZJ. Genomic and epigenetic insights into the molecular bases of heterosis. Nat Rev Genet 2013;14:471–82. doi: https://doi.org/10.1038/nrg3503.
[13] Wang B, Hou M, Shi J, Ku L, Song W, Li C, et al. De novo genome assembly and analyses of 12 founder inbred lines provide insights into maize heterosis. Nat Genet 2023;55:312–23. doi: https://doi.org/10.1038/s41588-022-01283-w.
[14] Huang X, Yang S, Gong J, Zhao Y, Feng Q, Gong H, et al. Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. Nat Commun 2015;6:6258. doi: https://doi.org/10.1038/ncomms7258.
[15] Gu Z, Gong J, Zhu Z, Li Z, Feng Q, Wang C, et al. Structure and function of rice hybrid genomes reveal genetic basis and optimal performance of heterosis. Nat Genet 2023;55:1745–56. doi: https://doi.org/10.1038/s41588-023-01495-8.
[16] Lin Z, Qin P, Zhang X, Fu C, Deng H, Fu X, et al. Divergent selection and genetic introgression shape the genome landscape of heterosis in hybrid rice. PNAS 2020;117:4623–31. doi: https://doi.org/10.1073/pnas.1919086117.
[17] Sun Z, Peng J, Lv Q, Ding J, Chen S, Duan M, et al. Dissecting the genetic basis of heterosis in elite super-hybrid rice. Plant Physiol 2023;192:307–25. doi: https://doi.org/10.1093/plphys/kiad078.
[18] Ouyang Y, Li X, Zhang Q. Understanding the genetic and molecular constitutions of heterosis for developing hybrid rice. J Genet Genomics 2022;49:385–93. doi: https://doi.org/10.1016/j.jgg.2022.02.022.
[19] Li D, Huang Z, Song S, Xin Y, Mao D, Lv Q, et al. Integrated analysis of phenome, genome, and transcriptome of hybrid rice uncovered multiple heterosis-related loci for yield increase. PNAS 2016;113:E6026–35. doi: https://doi.org/10.1073/pnas.1610115113.
[20] Chen J, Ding J, Ouyang Y, Du H, Yang J, Cheng K, et al. A triallelic system of *S5* is a major regulator of the reproductive barrier and compatibility of indica-japonica hybrids in rice. PNAS 2008;105:11436–41. doi: https://doi.org/10.1073/pnas.0804761105.
[21] Wang C, Tang S, Zhan Q, Hou Q, Zhao Y, Zhao Q, et al. Dissecting a heterotic gene through GradedPool-Seq mapping informs a rice-improvement strategy. Nat Commun 2019;10:2982. doi: https://doi.org/10.1038/s41467-019-11017-y.
[22] Ko DK, Rohozinski D, Song Q, Taylor SH, Juenger TE, Harmon FG, et al. Temporal shift of circadian-mediated gene expression and carbon fixation contributes to biomass heterosis in maize hybrids. PLoS Genet 2016;12:e1006197. doi: https://doi.org/10.1371/journal.pgen.1006197.
[23] Zhang J, Gu R, Miao X, Schmidt RH, Xu Z, Lu J, et al. GWAS-based population genetic analysis identifies *bZIP29* as a heterotic gene in maize. Plant Commun 2025;20:101289. doi: https://doi.org/10.1016/j.xplc.2025.101289.
[24] Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw 2015;67:1–48. doi: https://doi.org/10.18637/jss.v067.i01.
[25] Covarrubias-Pazaran G. Genome-assisted prediction of quantitative traits using the R package sommer. PLoS One 2016;11:e0156744. doi: https://doi.org/10.1371/journal.pone.0156744.
[26] Danecek P, McCarthy SA. BCFtools/csq: haplotype-aware variant consequences. Bioinformatics 2017;33:2037–9. doi: https://doi.org/10.1093/bioinformatics/btx100.
[27] Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale sequence data. Am J Hum Genet 2021;108:1880–90. doi: https://doi.org/10.1016/j.ajhg.2021.08.005.
[28] Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 2015;4:7. doi: https://doi.org/10.1186/s13742-015-0047-8.
[29] Wang J, Chen Y, Shu G, Zhao M, Zheng A, Chang X, et al. Fast3VmrMLM: a fast algorithm that integrates genome-wide scanning with machine learning to accelerate gene mining and breeding by design for polygenic traits in large-scale GWAS datasets. Plant Commun 2025;6:101385. doi: https://doi.org/10.1016/j.xplc.2025.101385.
[30] Wang J, Han X, Zhao M, Zhang H, Chen Y, Jiang Q, et al. A fast method for breeding by design via G×E interactions detected in large-scale climatic, phenomic and genomic data. Natl Sci Rev 2026:nwag095. doi: https://doi.org/10.1093/nsr/nwag095.
[31] Abraham G, Inouye M. Fast principal component analysis of large-scale genome-wide data. PLoS One 2014;9:e93766. doi: https://doi.org/10.1371/journal.pone.0093766.
[32] Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res 2009;19:1655–64. doi: https://doi.org/10.1101/gr.094052.109.
[33] Li M, Zhang YW, Zhang ZC, Xiang Y, Liu MH, Zhou YH, et al. A compressed variance component mixed model for detecting QTNs and QTN-by-environment and QTN-by-QTN interactions in genome-wide association studies. Mol Plant 2022;15:630–50. doi: https://doi.org/10.1016/j.molp.2022.02.012.
[34] Cingolani P, Platts A, Wang L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 2012;6:80–92. doi: https://doi.org/10.4161/fly.19695.
[35] Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics 2011;27:2156–8. doi: https://doi.org/10.1093/bioinformatics/btr330.
[36] Jiang L, Ge M, Zhao H, Zhang T. Analysis of heterosis and quantitative trait loci for kernel shape related traits using triple testcross population in maize. PLoS One 2015;10:e0124779. doi: https://doi.org/10.1371/journal.pone.0124779.
[37] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550. doi: https://doi.org/10.1186/s13059-014-0550-8.
[38] Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 2012;16:284–7. doi: https://doi.org/10.1089/omi.2011.0118.
[39] Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 2015;16:157. doi: https://doi.org/10.1186/s13059-015-0721-2.
[40] Pang J, Fu J, Zong N, Wang J, Song D, Zhang X, et al. Kernel size-related genes revealed by an integrated eQTL analysis during early maize kernel development. Plant J 2019;98:19–32. doi: https://doi.org/10.1111/tpj.14193.
[41] Zhao H, Sun Z, Wang J, Huang H, Kocher JP, Wang L. CrossMap: a versatile tool for coordinate conversion between genome assemblies. Bioinformatics 2014;30:1006–7. doi: https://doi.org/10.1093/bioinformatics/btt730.
[42] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinf 2008;9:559. doi: https://doi.org/10.1186/1471-2105-9-559.

[43] Xu S. An expectation-maximization algorithm for the Lasso estimation of quantitative trait locus effects. Heredity 2010;105:483–94. doi: https://doi.org/10.1038/hdy.2009.180.
[44] Winkler RG, Freeling M. Physiological genetics of the dominant gibberellin-nonresponsive maize dwarfs, *Dwarf8* and *Dwarf9*. Planta 1994;193:341–8. doi: https://doi.org/10.1007/BF00201811.
[45] Forestan C, Farinati S, Rouster J, Lassagne H, Lauria M, Dal Ferro N, et al. Control of maize vegetative and reproductive development, fertility, and rRNAs silencing by *HISTONE DEACETYLASE 108*. Genetics 2018;208:1443–66. doi: https://doi.org/10.1534/genetics.117.300625.
[46] Luo JH, Wang M, Jia GF, He Y. Transcriptome-wide analysis of epitranscriptome and translational efficiency associated with heterosis in maize. J Exp Bot 2021;72:2933–46. doi: https://doi.org/10.1093/jxb/erab074.
[47] Paciorek T, Chiapelli BJ, Wang JY, Paciorek M, Yang H, Sant A, et al. Targeted suppression of gibberellin biosynthetic genes *ZmGA20ox3* and *ZmGA20ox5* produces a short stature maize ideotype. Plant Biotechnol J 2022;20:1140–53. doi: https://doi.org/10.1111/pbi.13797.
[48] Chen J, Zhou H, Xie W, Xia D, Gao G, Zhang G, et al. Genome-wide association analyses reveal the genetic basis of combining ability in rice. Plant Biotechnol J 2019;17:2211–22. doi: https://doi.org/10.1111/pbi.13134.
[49] Hu Y, Xiong J, Shalby N, Zhuo C, Jia Y, Yang QY, et al. Comparison of dynamic 3D chromatin architecture uncovers heterosis for leaf size in *Brassica napus*. J Adv Res 2022;42:289–301. doi: https://doi.org/10.1016/j.jare.2022.01.001.
[50] Washburn J, Varela J, Xavier A, Chen Q, Ertl D, Gage J, et al. Global genotype by environment prediction competition reveals that diverse modeling strategies can deliver satisfactory maize yield estimates. Genetics 2025;229:iyae195. doi: https://doi.org/10.1093/genetics/iyae195.
[51] Li C, Huang L, Xu C, Zhao Y, Zhou DX. Altered levels of histone deacetylase *OsHDT1* affect differential gene expression patterns in hybrid rice. PLoS One 2011;6:e21789. doi: https://doi.org/10.1371/journal.pone.0021789.
[52] Ma X, Jia Q, Li S, Chen Z, Ming X, Zhao Y, et al. An enhanced network of energy metabolism, lysine acetylation, and growth-promoting protein accumulation is associated with heterosis in elite hybrid rice. Plant Commun 2023;4:100560. doi: https://doi.org/10.1016/j.xplc.2023.100560.
[53] Chen Y, Guo P, Dong Z. The role of histone acetylation in transcriptional regulation and seed development. Plant Physiol 2024;194:1962–79. doi: https://doi.org/10.1093/plphys/kiad614.
[54] Sazuka T, Kamiya N, Nishimura T, Ohmae K, Sato Y, Imamura K, et al. A rice tryptophan deficient dwarf mutant, *tdd1*, contains a reduced level of indole acetic acid and develops abnormal flowers and organless embryos. Plant J 2009;60:227–41. doi: https://doi.org/10.1111/j.1365-313X.2009.03952.x.
[55] Groszmann M, Gonzalez-Bayon R, Lyons RL, Greaves IK, Kazan K, Peacock WJ, et al. Hormone-regulated defense and stress response networks contribute to heterosis in *Arabidopsis* F1 hybrids. PNAS 2015;112:E6397–406. doi: https://doi.org/10.1073/pnas.1519926112.
[56] Zhu A, Wang A, Zhang Y, Dennis ES, Peacock WJ, Greaves IK. Early establishment of photosynthesis and auxin biosynthesis plays a key role in early biomass heterosis in *Brassica napus* (canola) hybrids. Plant Cell Physiol 2020;61:1134–43. doi: https://doi.org/10.1093/pcp/pcaa038.
[57] Shang J, Pang H, Wang L, Li X, Wang Y, Li Y. Study on the relationship between auxin and sorghum heterosis, Xinjiang. Agric Sci 2024;4:841–6.
[58] Heuer S, Hansen S, Bantin J, Brettschneider R, Kranz E, Lörz H, et al. The maize MADS box gene *ZmMADS3* affects node number and spikelet development and is co-expressed with *ZmMADS1* during flower development, in egg cells, and early embryogenesis. Plant Physiol 2001;127:33–45. doi: https://doi.org/10.1104/pp.127.1.33.
[59] Yin X, Liu X, Xu B, Lu P, Dong T, Yang D, et al. *OsMADS18*, a membrane-bound MADS-box transcription factor, modulates plant architecture and the abscisic acid response in rice. J Exp Bot 2019;70:3895–909. doi: https://doi.org/10.1093/jxb/erz198.
[60] Li X, Xie Z, Qin T, Zhan C, Jin L, Huang J. The *SLR1-OsMADS23-D14* module mediates the crosstalk between strigolactone and gibberellin signaling to control rice tillering. New Phytol 2025;246:2137–54. doi: https://doi.org/10.1111/nph.20331.
[61] Wen J, Zhao X, Wu G, Xiang D, Liu Q, Bu SH, et al. Genetic dissection of heterosis using epistatic association mapping in a partial NCII mating design. Sci Rep 2015;5:18376. doi: https://doi.org/10.1038/srep18376.
[62] Zhou T, Afzal R, Haroon M, Ma Y, Zhang H, Li L. Dominant complementation of biological pathways in maize hybrid lines is associated with heterosis. Planta 2022;256:111. doi: https://doi.org/10.1007/s00425-022-04028-5.
[63] Shi X, Li W, Guo Z, Wu M, Zhang X, Yuan L, et al. Comparative transcriptomic analysis of maize ear heterosis during the inflorescence meristem differentiation stage. BMC Plant Biol 2022;22:348. doi: https://doi.org/10.1186/s12870-022-03695-6.
[64] Hu X, Wang H, Diao X, Liu Z, Li K, Wu Y, et al. Transcriptome profiling and comparison of maize ear heterosis during the spikelet and floret differentiation stages. BMC Genomics 2016;17:959. doi: https://doi.org/10.1186/s12864-016-3296-8.