Epistasis Activation Contributes Substantially to Heterosis in Temperate by Tropical Maize Hybrids

Zhiqin Sang[1,2]*, Hui Wang[1,3,4], Yuxin Yang[1], Zhanqin Zhang[2], Xiaogang Liu[1], Zhiwei Li[1] and Yunbi Xu[1,5]*

[1] Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, [2] Xinjiang Academy of Agricultural and Reclamation Science, Shihezi, China, [3] Crop Research Institute, Shandong Academy of Agricultural Sciences, Jinan, China, [4] National Engineering Research Center of Wheat and Maize, Shandong Technology Innovation Center of Wheat, Shandong Academy of Agricultural Sciences, Jinan, China, [5] International Maize and Wheat Improvement Center, Texcoco, Mexico

Epistasis strongly affects the performance of superior maize hybrids. In this study, a multiple-hybrid population, consisting of three hybrid maize sets with varied interparental divergence, was generated by crossing 28 temperate and 23 tropical inbred lines with diverse genetic backgrounds. We obtained 1,154 tested hybrids. Among these tested hybrids, heterosis increased steadily as the heterotic genetic distance increased. Mid-parent heterosis was significantly higher in the temperate by tropical hybrids than in the temperate by temperate hybrids. Genome-wide prediction and association mapping was performed for grain weight per plant (GWPP) and days to silking (DTS) using 20K high-quality SNPs, showing that epistatic effects played a more prominent role than dominance effects in temperate by tropical maize hybrids. A total of 33 and 420 epistatic QTL were identified for GWPP and DTS, respectively, in the temperate by tropical hybrids. Protein–protein interaction network and gene-set enrichment analyses showed that epistatic genes were involved in protein interactions, which play an important role in photosynthesis, biological transcription pathways, and protein synthesis. We showed that the interaction of many minor-effect genes in the hybrids could activate the transcription activators of epistatic genes, resulting in a cascade of amplified yield heterosis. The multiple-hybrid population design enhanced our understanding of heterosis in maize, providing an insight into the acceleration of hybrid maize breeding by activating epistatic effects.

Keywords: maize, heterosis, GWAS, epistatic effects, protein–protein interaction, multiple-hybrid population

INTRODUCTION

Hybrid vigor or heterosis refers to the phenotypic superiority of F_1 hybrid plants over their parents. The mechanism of heterosis can be determined according to three primary hypotheses based on classical genetics. The dominance hypothesis explains heterosis by the action of superior dominant alleles from both parents at multiple loci, which complement corresponding unfavorable alleles leading to the enhancement of hybrid vigor. Such complementation might allow hybrids to be similar to or better than the superior parent (Jones, 1917; Birchler, 2015). According to

the single locus overdominance hypothesis, various alleles interact to perform a function better than that performed by homozygous alleles. Therefore, the increase in vigor is proportional to the degree of heterozygosity (East, 1936; Lamkey and Edwards, 1999). Dominance and overdominance hypotheses are based on the action of a single gene, but most heterosis-related traits are quantitatively inherited and involve multiple genes with different effects. The epistasis hypothesis emphasizes the role of inter-allele interactions among genetic loci and associated pathways, which might include all possible forms of molecular interactions (Powers, 1944; Jinks and Jones, 1958; Jiang et al., 2017). Many researchers suggested that partial or complete dominance, rather than super dominance, account for the inheritance of heterosis in maize (Hallauer et al., 2010). Although some researchers have focused on single-gene models, several studies have suggested that heterosis is generally the result of the action of multiple loci, which affect heterosis of different traits and hybrids (Bauman, 1959; Doebley et al., 1995; Ma et al., 2007; Hallauer et al., 2010).

As East (1936) stated, "the problem of heterosis is the problem of the inheritance of quantitative characters," and quantitative traits are typically affected by multiple genes (Flint-Garcia et al., 2009). The genotype of a population has a "net-like" structure; hence, different loci might affect the variation in several characters (Wright, 1984). Additionally, substituting one gene might affect several characters (Yu et al., 1997; Xiao et al., 2021). Based on this perspective, epistasis is one of the most important genetic components in the inheritance of quantitative characters. Epistasis might also contribute to the genetic basis of heterosis (Hua et al., 2003; Hochholdinger and Baldauf, 2018). Epistasis not only shapes which loci can express heterosis but can also mimic overdominance (Fiévet et al., 2010). Several studies have investigated the epistatic effects using quantitative and biparental population genetic approaches (Doebley et al., 1995; Culverhouse et al., 2004). Biparental populations have a narrow genetic basis, which further restricts the effective detection of epistasis (Jiang et al., 2017; Xiao et al., 2021). Due to the lack of tailored quantitative genetic approaches to determine the function of epistasis in hybrid breeding populations, epistatic effects in the populations developed by crossing a panel of diverse breeding materials could not be assessed (Boeven et al., 2020). A quantitative genetic framework was developed to determine the relative contributions of dominance and epistatic effects to heterosis (Jiang et al., 2017), allowing the integration of epistasis in hybrid populations to derive from different parents.

In line with the quantitative genetic hypothesis, hybrid vigor is determined by the interparental genetic distance (Wei and Zhang, 2018). Assuming that all quantitative trait loci (QTL) contribute to heterosis, the genetic distance can be estimated by the squared difference of the interparental allele frequency (Frankham, 1996; Boeven et al., 2020). When heterosis mainly results from dominance and overdominance effects, it is positively correlated with genetic distance (Lamkey and Edwards, 1998). The genetic distance between parents in maize was found to be moderately or highly correlated with middle parent heterosis (Laude and Carena, 2015), which was contrary to several other reports from the tropical region

(Badu-Apraku et al., 2013; Oyekunle et al., 2015). Most artificial selection techniques involved reshaping gene networks rather than single genes (Doust et al., 2014). Thus, due to thousands of years of artificial and natural selection, temperate germplasm exhibits significantly lower genetic diversity compared to tropical germplasm, as shown by the diversity of haplotypes and SNP markers (Lu et al., 2009, 2011). Therefore, the genetic distance between temperate and tropical maize is greater than that within temperate maize. Ignoring the genetic architecture of heterosis within and between alleles might show incorrect relationships between heterosis and genetic distance, especially for the temperate by tropical hybrids with significant genetic differentiation (Boeven et al., 2020).

To understand the genetic mechanism of heterosis in different types of maize, three different hybrid maize panels were developed in this study. The temperate by temperate panel comprised 377 temperate maize hybrids, the temperate by tropical panel comprised 641 temperate and tropical hybrids, and the tropical by tropical panel comprised 136 tropical hybrids. In this study, we found that the heterosis of grain yield and flowering stage increased with heterotic genetic distance. Genome-wide association studies on the heterosis of grain yield and days to silking revealed changes in the genetic architecture of the hybrids from the temperate by temperate to temperate by tropical panels, indicating novel epistatic effects underlying the heterosis in the temperate by tropical panel. The results of the protein–protein interaction (PPI) network analysis and gene-set enrichment analysis (GSEA) demonstrated that many minor loci interacting with the gene networks might improve the performance of the hybrids.

MATERIALS AND METHODS

Plant Materials and Phenotyping

The multiple-hybrid population used in this study was derived from the cross between 23 tropical and 28 temperate inbred lines, representing high diversity in temperate and tropical regions. These lines were grouped into seven heterotic pools according to their pedigrees, original regions, and genetic population structure inferred from molecular markers (Wang et al., 2017). The temperate diallel contained 377 hybrids derived from Griffing IV (temperate by temperate) with 28 temperate maize inbred lines (13 U.S. and 15 Chinese lines), representing different heterotic groups. The NC II (temperate by tropical) hybrid panel contained 641 hybrids generated from 23 tropical and 28 temperate inbred lines. The tropical diallel contained 136 hybrids derived from Griffing IV (tropical by tropical) with 17 tropical inbred lines as parents, many of which were developed at the International Maize and Wheat Improvement Center (CIMMYT) (**Supplementary Figure 1**). For 325 temperate diallel crosses and 263 NC II crosses, enough hybrid seeds were generated to conduct field trials for phenotyping in 2013, 2014, and 2015 at Xinxiang, Henan (35.1°N, 113.8°E) and Shunyi, Beijing (40.2°N, 116.6°E) using randomized block design with two duplicates for each experiment (Wang et al., 2017; Yu et al., 2020). Tropical diallel hybrids were phenotyped in

Jinghong, Yunnan (22.0°N, 100.8°E) in 2014 and Sanya, Hainan (18.4°N, 109.2°E) in 2015 using the same experimental design (Wang et al., 2017). For 377 temperate diallel crosses and 641 NC II crosses, enough hybrid seeds were generated to conduct field trials for 2 years (2017–2018) in Shihezi, Xinjiang (45.2°N, 84.68°E), using alpha designs with two replicates (**Supplementary Table 1**). For each replicate, during field trials, all the hybrids and their 51 parents were split into two adjacent trials. Herbicides, insecticides, and fertilizers were applied following the farmer's practices in intensive maize production.

Days to silking (DTS) and days to anthesis (DTA) were recorded as the number of days from planting to when 50% of the plants in a plot had shed pollen and extruded silks, respectively; anthesis–silking interval (ASI) was defined as the time interval between DTS and DTA. After female flowering, the PH values were recorded as the averaged value of the five plants from the center of the plot. The grain yield per plant (GWPP) was estimated based on the mean value of 10 plants. The grain number per row (GNPR), the row number (RN), and the ear barren tips (TIP) were measured as the averaged value of 10 ears from each plot. Harvesting was performed manually, and the harvest was adjusted to a moisture content of 140 g H_2O kg^{-1}.

Phenotypic Data Analysis

A two-stage method was adopted for phenotypic data analysis (Möhring and Piepho, 2009), where effects were modeled and estimated for each environment first, and then the means of genotypes across environments were calculated. In the first step, we used a mixed model approach for analyzing the modeling effects of individual environments for genotypes, replications, and the blocks within replications, based on the statistical model

$$y_{ijkl} = g_{ij} + r_k + b_{lk} + e_{ijkl} \qquad (1)$$

where y_{ijk} represents the phenotypic performance of the ijth genotype (hybrid $j \neq i$ or parental line $i = j$) within the lth incomplete block for the kth replication, g_{ij} represents the genetic effect of the ijth genotype, r_k represents the effect of the kth replication, b_{lk} represents the effect of the lth incomplete block within the kth replication, and e_{ijkl} represents the residual. Except for the effect of g_{ij}, the remaining effects were random.

In the second step, to analyze the phenotypes in various environments, we used a linear mixed model based on the adjusted best linear unbiased estimates (BLUEs) of individual environments:

$$y_{ijn} = g_{ij} + l_n + (g_{ij}l) + e_{ijn} \qquad (2)$$

where g_{ij} and l_n denote the genotypes and individual environments, respectively. Except for the effect of g_{ij}, all other effects were random.

The fixed genotypic effects were used to obtain the BLUEs for parental and hybrid genotypic values. Next, we used the BLUEs for calculating mid-parent heterosis ($_{MPH}$) of all hybrids by $MPH = F_1 - MP$, where, F_1 represents the hybrid performance, and MP represents the mid-parent value for two parents, P_1 and P_2. Then, we determined the relative $MPH(\%)$ for all hybrids using the formula $MPH(\%) = (\frac{MPH}{MP}) \times 100$. Next, we

determined better-parent heterosis ($_{BPH}$) using the formula $BPH = F_1 - P_{\text{Better}}$, where P_{Better} represents the performance of the better-performing parental line. We measured relative BPH using the formula $BPH(\%) = (\frac{BPH}{P_{\text{Better}}}) \times 100$. Pearson's product–moment correlation coefficients were calculated to test the BLUEs-based correlations. Student's t-tests were performed to compare the BLUEs among diverse genotypes.

The mixed linear model approach was used to estimate variance components, and except for the group effect, the remaining effects were random. Then, we decomposed the total variance of the temperate by temperate and temperate by tropical set to the variances resulting from the effects of the general combining ability (GCA) of female and male subjects, as well as the variance resulting from the specific combining ability (SCA) of the hybrids (Zhao et al., 2015):

$$y_{ijkl} = a + l_n + lr_{nk} + g_i + g_j + g_{ij}$$
$$+g_i : l_n + g_j : l_n + g_{ij} : l_n + e_{ijn} \qquad (3)$$

where y_{ijkl} represents the phenotypic performance of the ijth entry (hybrid $i \neq j$, or line $i = j$) in the nth environment, a represents hybrid and line group effects, l_n indicates the effect of the nth environment, g_i and g_j indicate the genetic effects of the parental lines, g_{ij} represents the SCA effect of the crosses between i and j lines, $g_i : l_n$ and $g_j : l_n$ represents the interplay effect between the i/jth parental lines and the nth environment, $g_{ij} : l_n$ represents the interplay effect between the SCA and the environment, whereas, e_{ijn} represents the residual. Heritability was estimated by the genotypic-to-phenotypic variance ratio, $H^2 = \frac{\delta_G^2}{\delta_G^2 + \delta_G^2/E + \delta_e^2/(E \times R)}$, where, E represents environment number, R represents the mean replication number in each entry in one location, and δ_e^2 represents the combined error variance. The MPH was estimated according to the block-corrected values of the hybrids in each environment and their parental BLUEs obtained across all environments. These MPH values calculated for each environment were used to construct the linear mixed model (Equation 3) and estimate heritability. The ASReml-R 4.0 software package was used for performing statistical analysis in the R environment (Butler et al., 2009).

Genotypic Data Analysis

In a previous study, we genotyped the 51 parental lines with the Maize55K chip (Wang et al., 2017; Xu et al., 2017). From this SNP dataset, quality control of minor allele frequency (MAF > 5%), missing data (<5%), and SNP were filtered based on linkage disequilibrium (LD) (eliminating one in each SNP pair when LD was greater than 0.5 among 50 SNPs). We retained 22,510 high-quality SNP markers for further analyses. For some SNPs with less than 5% MAF in the temperate hybrids, 20,555 high-quality SNPs were finally selected and used for genome-wide prediction and association mapping in the temperate by temperate set. The marker profiles for each hybrid were deduced from the corresponding parental lines. In the genome-wide prediction and association mapping of the additive model, the minor homozygote was coded as "2," the major homozygote was coded as "0," and the heterozygote was coded as "1,"

The F_{st} statistic and SNP nucleotide divergence (π) between the temperate and tropical inbred lines were calculated using VCFtools (Danecek et al., 2011) and visualized in R (R Core Team, 2018). For each 1,000-kb window, we calculated the F_{st} statistics and sequence diversity statistics (π) with 100-kb steps in the maize genome. We also determined π values for temperate and tropical inbred lines, and the ratio of π ($\pi_{temperate}/\pi_{tropical}$) was used to detect the genetic-improvement sweeps. For parental lines and hybrids, the LD decomposition, based on the genetic map distance, was evaluated by fitting the natural smoothing splines to r^2 values using the software package PLINK (Purcell et al., 2007). Principal component analysis (PCA) was performed to analyze the population structures of the hybrids and parental lines using the software TASSEL (Bradbury et al., 2007), and MEGA version 7.0.26 was used for clustering based on modified Nei's genetic distance (Tamura et al., 2013). Rogers' distance (RD) was used to measure the genetic distance (Boggs and Rogers, 1990). We conducted genome-wide prediction for the hybrid phenotype to obtain dominance effects for GWPP and DTS (Zhao et al., 2013; Alves et al., 2019); then, we used them to weigh the marker loci. Furthermore, we determined the heterotic genetic distance developed by Boeven et al. (2020), which was expressed depending on Rogers' distance by including the predicted dominance effects for SNP, as shown in Equation 4:

$$\smallint_{RD}(X, Y) = \frac{1}{L} \sum_{u=1}^{L} w_u \sqrt{\frac{\sum_{j=1}^{n_u}(X_{uj} - Y_{uj})^2}{2}} \qquad (4)$$

where X and Y represent two genotypes, X_{uj} and Y_{uj} represent the frequency of the jth allele at the uth locus, n_u represents the allele number in the uth locus, L represents the locus number, and w_u indicates the dominance weight at the uth locus. Bayes method was used to predict the dominance effects of SNPs (Zhao et al., 2013; Alves et al., 2019). Additionally, five-fold cross-validation was conducted with 100 iterations for predicting dominance effects and assessing the relationship of heterosis with $\smallint_{RD}$. Locally weighted linear regression was performed to determine the relationship between heterosis and genetic distance.

Partitioning of Genetic Variance Components for Mid-Parent Heterosis

We fitted the extended genomic BLUE model that included the digenic epistatic and dominance effects to estimate the genetic variance components of MPH (Zhou et al., 2012; Jiang et al., 2017; Boeven et al., 2020). The model is shown below:

$$y = g_d + g_{aa} + g_{ad} + g_{dd} + e \qquad (5)$$

Here, y represents the MPH vector of every hybrid, g_d, g_{dd}, g_{ad}, and g_{aa} indicate genetic values of the vectors related to dominance, dominance-by-dominance, additive-by-dominance, and additive-by-additive effects, and e represents the residual. It is assumed that $g_d \sim N(0, K_d\delta_d^2)$, $g_{aa} \sim N(0, K_{aa}\delta_{aa}^2)$, $g_{ad} \sim N(0, K_{ad}\delta_d^2)$, $g_{dd} \sim N(0, K_{dd}\delta_{dd}^2)$, and $e \sim N(0, TT'\delta_e^2)$ in the formula, where K_d, K_{aa}, K_{ad}, and K_{dd} indicate the kinship matrices derived from the markers of different genetic effects, T indicates the linear transformation $r \times (r + s)$ matrix from the

original trait vectors to the MPH vector, r indicates the hybrid number, and s represents the parental line number. Next, the F metric was used for calculating kinship matrices derived from the markers, resulting in model non-orthogonal parametrization (Jiang et al., 2017). The multi-kernel approach was adopted to estimate the variance components δ_d^2, δ_{aa}^2, δ_{ad}^2, and δ_{dd}^2 using the R package BGLR (Pérez and De Los Campos, 2014), following previously published settings (Jiang et al., 2017).

Estimation of Heterotic Effects

For a locus, its heterotic effect represents its genetic contribution to MPH and also the combination of the self-dominance effect and epistatic interaction effect with the entire genetic background (Jiang et al., 2017). Specifically, Q was assumed to be the entire QTL set for each phenotypic trait. QTL was assigned a value of 0, 1, or 2, based on the selected allele number at each locus. For each hybrid, R_{kl} (k, $l = 0$ or 2) was denoted as the locus subset, where the male and female parents had the genotypes l and k, respectively. For $i, j \in Q$ and $i \neq j$, d_i is assumed as the dominance effect of the ith QTL, while aa_{ij}, ad_{ij}, and dd_{ij} represent the additive-by-additive, additive-by-dominance, and dominance-by-dominance epistatic effects between ith and jth QTL. For the ith locus, its heterotic effect is shown below:

$$h_i = \begin{cases} d_i - \frac{1}{2}\sum_{j \in R_{20}} aa_{ij} + \frac{1}{2}\sum_{j \in R_{02}} aa_{ij} + \frac{1}{2}\sum_{j \in R_{22}} ad_{ij} \\ \quad - \frac{1}{2}\sum_{j \in R_{00}} aad_{ij} + \frac{1}{2}\sum_{j \in R_{20} \cup R_{02}} dd_{ij} \text{ if } i \in R_{20} \\ d_i - \frac{1}{2}\sum_{j \in R_{02}} aa_{ij} + \frac{1}{2}\sum_{j \in R_{20}} aa_{ij} + \frac{1}{2}\sum_{j \in R_{22}} ad_{ij} \\ \quad - \frac{1}{2}\sum_{j \in R_{00}} aad_{ij} + \frac{1}{2}\sum_{j \in R_{20} \cup R_{02}} dd_{ij} \text{ if } i \in R_{02} \\ \quad \frac{1}{2}\sum_{j \in R_{20} \cup R_{02}} ad_{ij} \text{ if } i \in R_{22} \\ \quad \frac{1}{2}\sum_{j \in R_{20} \cup R_{02}} dd_{ij} \text{ if } i \in R_{00} \end{cases}$$
$$(6)$$

According to the definition, the MPH value of the hybrid represents the summation of all heterotic effects among polymorphic loci.

Genome-Wide Scanning for Significant Heterotic Effects

A three-step process proposed by Jiang et al. (2017) was used to detect significant heterotic effects. First, we conducted genome-wide association mapping to identify significant component effects. Additionally, a standard linear mixed model was constructed along with the kinship matrix derived from the markers to control the structure of various polygenic background effects and relatedness levels (Yu et al., 2005). Assuming the presence of epistasis, a model that controls polygenic background effects, including epistatic and main effects, needs to be constructed (Xu, 2013; Jiang et al., 2017). This model is shown below:

$$y = m\alpha + g_d + g_{aa} + g_{ad} + g_{dd} + e \qquad (7)$$

Here, y, g_d, g_{aa}, g_{ad}, g_{dd}, and e represent the same parameters as those presented in Equation 5. Particularly, α represents the dominance effect of a marker or the epistatic interaction effect of a pair of markers, and m represents the relevant coefficient. We assumed α to be an unknown fixed parameter, while the rest were assumed to be the same as those in Equation 5. We

converted this model into a standard linear regression model for calculating efficiency, where only the residual terms were random. The converted model was similar to the original one as long as the effect of different parameters on the estimation of the variance components was negligible (Lippert et al., 2011; Xu, 2013). Next, we performed the F test to assess the significance of the effect of α (Jiang et al., 2017).

Second, we integrated the significant component effects into the heterotic effects based on Equation 6. We set each non-significant effect as zero.

Third, for all loci, we analyzed their heterotic effect h_i using the permutation test. Specifically, we predicted the hybrid absolute MPH values through their respective heterotic effects. Then, we determined Pearson's correlation coefficient between the actual and estimated MPH values and conducted permutation tests for the correlation coefficients.

In steps one and three, we examined the genome-wide threshold for the p. For the temperate by tropical panel, heterotic and dominance effects were tested by Bonferroni corrected threshold at $p = 0.001/n$. In contrast, the additive-by-additive, additive-by-dominance, and dominance-by-dominance epistatic effects were tested at $p < 0.001/[n(n-1)]$, where n denotes the tested number of SNPs (Holm, 1979; Jiang et al., 2017). The small population size restricted the power to detect significant effects for the temperate by temperate panel. Thus, the threshold for the heterotic and dominance effects in GWPP was set at $p < 0.05/n$, whereas the epistatic effects were tested at $p < 0.05/[n(n-1)]$.

Omics Network Analysis and Gene-Set Enrichment

To further interpret the genetic structure and identify the important candidate epistatic genes for the two tested traits, the genes adjacent to the significant SNPs less than 55 kb were considered to be the candidate genes, while the length of linkage disequilibrium (LD) decay was 245 kb at $r^2 = 0.1$ (Jung et al., 2004). In this study, the MaizeGDB database[1] was used to convert candidate gene names from AGPv3 to AGPv4, and the candidate genes were annotated based on the data from the MaizeGDB database (Harper et al., 2016). A seed indicated the input gene used to query protein–protein interactions (PPI), and the interactome comprised its direct interactor genes. The first layer network comprised the seeds; the second layer network comprised all interactors that interacted with the seeds; the specific network was composed of the first + second layer interactomes. The candidate genes were assigned to PPI using the Maize Interactome Platform (MIP) developed in another study (Han et al., 2020), and the first layer network was constructed based on the interaction of the candidate genes with each other. Each gene in the first layer network acted as seeds, which were used to query the direct interactor genes in the MIP Network Creation tool. Then, all of their genes were selected with the MIP Slim-interactive Omics Network tool to classify them into corresponding modules. The direct interactor genes were used to compose the second layer network. The gene network obtained from MIP was imported into Cytoscape version 3.7.2 for further

analysis and displayed. The genes in the different modules were then assigned to perform Gene Ontology under the categories of biological process (GO-BP) to identify biological functions with a false discovery rate (FDR) less than 0.05, based on the online AgriGO Singular Enrichment Analysis tool[2] (Tian et al., 2017).

RESULTS

Broad Breeding Germplasm From Temperate and Tropical Regions to Determine Heterosis

We produced a multiple-hybrid population of maize with 1,154 hybrids from three subpopulations. The largest genetic distance was identified between the temperate and tropical lines ($p < 0.001$; 0.38), followed by the groups of the tropical (0.36) and temperate lines (0.35). The genetic diversity in the intergroup was also determined, and a faster LD decay was found in the temperate by tropical hybrid set than in the other two hybrid sets. The PCA confirmed the differences among the three hybrid sets, with three final patterns being distributed in different spaces. Using an NJ phylogenetic tree, cluster analysis based on GD estimates identified 49 groups, and the hybrids derived from a parent were mostly clustered into the same group (**Figure 1**). The F_{ST} and $θπ$ for the entire maize genome were observed from the volcano plot. We discovered 125 candidate selective-sweep areas, which covered 0.61% of the maize genome.

Higher Level of Heterosis Identified in Temperate by Tropical Hybrids

We evaluated the heterosis of DTS and GWPP for the temperate by temperate and temperate by tropical hybrids from eight agro-ecological sites and the tropical by tropical hybrids in two agro-ecological sites. The heterosis for both GWPP and DTS were normally distributed (**Supplementary Figures 2, 3**). A moderate-to-high heritability was found, with a value of 0.79 for the temperate by temperate set and a value of 0.84 for the temperate by tropical set for GWPP (**Supplementary Figure 4**). The broad-sense heritability for GWPP was 0.63 for the temperate by tropical hybrids and 0.77 for the temperate by temperate hybrids. The broad-sense heritability for MPH of DTS was 0.88 for the temperate by temperate hybrids and 0.93 for the temperate by tropical hybrids. The MPH heritability estimates for both traits were higher for the temperate by tropical hybrids (**Supplementary Figure 4**).

For GWPP, the mean relative MPH for the temperate by temperate, temperate by tropical, and tropical by tropical hybrids were 128.92, 166.24, and 185.88%, respectively (**Supplementary Figure 2; Table 1**). The average absolute MPH decreased gradually from the temperate by tropical hybrids to the tropical by tropical hybrids, and finally, to the temperate by temperate hybrids. Hybrids in the temperate by tropical panel had significantly higher absolute MPH than in the temperate by the temperate panel ($p < 0.01$; 84.15 g plant^{-1} vs. 78.48 g plant^{-1};

---

[1]http://www.maizegdb.org

[2]http://systemsbiology.cau.edu.cn/agriGOv2

FIGURE 1 | Genetic diversity identified among the hybrids in multiple hybrid populations. Population structure and LD were analyzed for 28 temperate parents, 23 tropical parents, and 1,154 hybrids from the multiple hybrid populations. **(A)** LD decay (R^2) is plotted as a function of the genetic map distance based on pairwise correlations of the LD phase. Horizontal dotted lines represent LD decay lower than 0.1. **(B)** Principal component analysis (PCA) based on the maximum likelihood method. The numbers in the parentheses are genotypic variance proportions explained with the PC1, PC2, and PC3. **(C)** Genome-wide selective sweep analysis on tropical and temperate maize groups. Vertical and horizontal solid lines represent threshold lines for the top 5% of $\theta\pi$ ratios and F_{ST} values, respectively. Red dots in the right upper sector indicate selective signatures for temperate maize lines. The portions of histograms in red and blue colors indicate the $\theta\pi$ ratio (upper) and the F_{ST} (right) values above the thresholds. **(D)** Neighbor-joining tree based on genetic distances.

Figures 2A,B). Lower yield in tropical parents might lead to a higher level mid-parent heterosis in the temperate by tropical hybrids, considering that the temperate parents had 46.49% higher GWPP than tropical parents. However, the temperate by tropical hybrids had 4.7% higher absolute BPH than the temperate by temperate hybrids. The average absolute MPH for DTS decreased significantly from the temperate by tropical hybrids (−4.46 days) to the temperate by temperate hybrids (−5.07 days), and finally, to the tropical by tropical hybrids (−7.00 days) (**Figures 2C,D**). The extent of heterosis showed high specificity for every hybrid set, which could be interpreted either by different interparental genetic distances or the genetic mechanisms for the heterosis between hybrid panels.

Linear Relationship Between Heterosis and Heterotic Genetic Distance

Based on the dominance model, heterosis for GWPP showed a monotonic increase as the interparental heterotic genetic distance increased, explaining 70.78% of the variation (**Figure 3A**;

Supplementary Figures 5, 6). The unknown part of the variation probably resulted from epistatic effects or the noise. The former was obtained from heterosis variance partitioning with various components, and the results should be interpreted with caution (Huang and Mackay, 2016). For GWPP and DTS, epistasis accounted for 62 and 79% of the total genetic variation in the temperate by tropical hybrids and 57 and 60% in the temperate by temperate hybrids, respectively (**Figure 4**). The contribution of the dominant effect to the total genetic variance of heterosis was relatively low, and the close relationship between heterotic genetic distance and heterosis was not contradictory. On the contrary, the estimated dominant effect also captured different types of epistatic interactions, as shown by the correlation between the marker-derived dominant kinship matrix and the three types of digenic epistatic effects (**Supplementary Tables 2, 3**). The correlation between genetic distances and heterosis was weak when the dominant effects were ignored while estimating the genetic distances (**Figures 3B,D**; **Supplementary Figures 5, 6**).

TABLE 1 | Summary statistics for tested traits (GWPP and DTS) and their heterosis.

Source	DTS (days)	GWPP (g plant^{-1})
Temperate line (Min; Max)	61.75 (56.03; 68.61)	62.45 (36.91; 94.32)
Tropical line (Min; Max)	68.23 (59.76; 77.30)	42.63 (29.07; 63.08)
Temperate by temperate hybrids (n = 377)		
Average (Min; Max)	56.67 (51.67; 65.48)	140.99 (57.26; 217.01)
Average MPH (Min; Max)	−5.07 (−10.06; 2.23)	78.48 (8.36; 166.97)
Average MPH% (Min; Max)	−8.20 (−15.89; 3.53)	128.92 (16.5; 333.64)
Average BPH (Min; Max)	−6.90 (−15.88; 1.83)	70.15 (2.83; 161.59)
Average BPH% (Min; Max)	−10.79 (2.87; −23.15)	103.09 (5.20; 291.53)
Temperate by tropical hybrids (n = 641)		
Average (Min; Max)	63.48 (52.48; 81.7)	136.73 (35.47; 205.06)
Average MPH (Min; Max)	−4.46 (−11.84; 12.09)	84.15 (−29.69; 157.24)
Average MPH% (Min; Max)	−6.61 (−17.02; 17.46)	166.24 (−45.56; 422.31)
Average BPH (Min; Max)	−10.68 (−22.02; 6.95)	73.45 (−46.86; 149.99)
Average BPH% (Min; Max)	−14.28 (−27.81; 9.58)	125.21 (−56.92; 355.62)
Tropical by tropical hybrids (n = 136)		
Average (Min; Max)	67.56 (60.77; 76.27)	126.93 (87.26; 165.67)
Average MPH (Min; Max)	−7.00 (−13.05; −2.09)	79.71 (28.7; 124.25)
Average MPH% (Min; Max)	−9.37 (−16.87; −2.78)	185.88 (38.62; 377.9)
Average BPH (Min; Max)	−9.85 (−17.35; −2.88)	70.70 (−15.59; 123.08)
Average BPH% (Min; Max)	−12.67 (−21.65; −3.64)	152.75 (−13.15; 375.15)

DTS, days to silking; GWPP, grain weight per plant; BPH%, better parent heterosis percentage; BPH, absolute better-parent heterosis; MPH%, mid-parent heterosis percentage; MPH, absolute mid-parent heterosis.

Heterosis continuously increased with heterotic genetic distance, and the mean heterosis increased considerably in the temperate by tropical hybrids than in the temperate by temperate hybrids. The regression line of heterosis with heterotic genetic distance for the temperate by tropical hybrids was lower than that for the temperate by temperate hybrids at the beginning, and then the regression line of the temperate by tropical hybrids exceeded that of the temperate by temperate hybrids, increasing almost in parallel with the increase in the heterotic genetic distance (**Figure 3A**). However, for a given heterotic genetic distance, heterosis performance for GWPP was lower in the temperate by temperate hybrids compared to the temperate by tropical hybrids, but for DTS, it showed an opposite trend (**Figure 3C**), which might be due to the violation of the assumption that genetic effects should be similar among different hybrid sets when heterotic genetic distances are estimated.
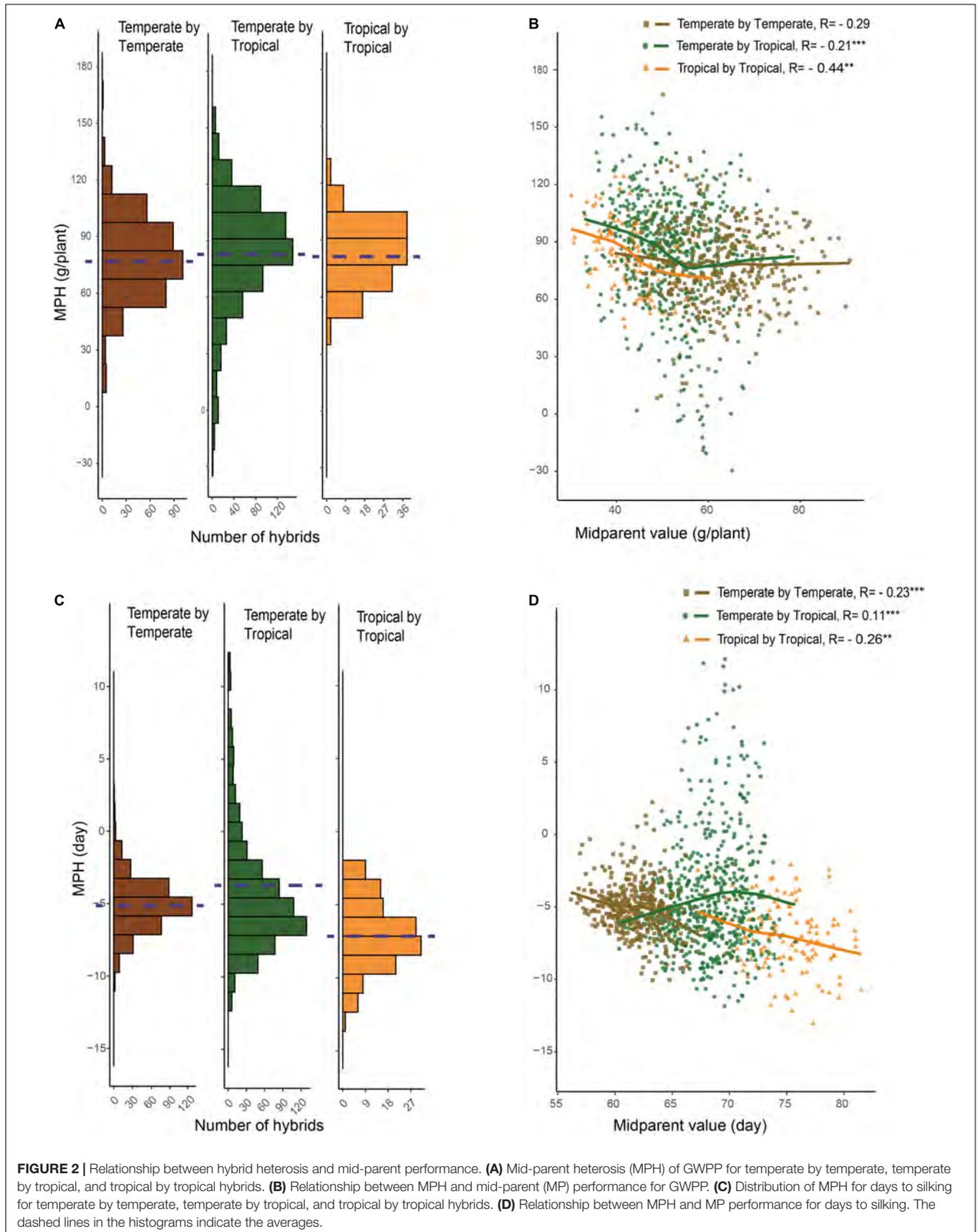
Epistatic Effects Make Greater Contributions to Heterosis in the Temperate by Tropical Hybrids
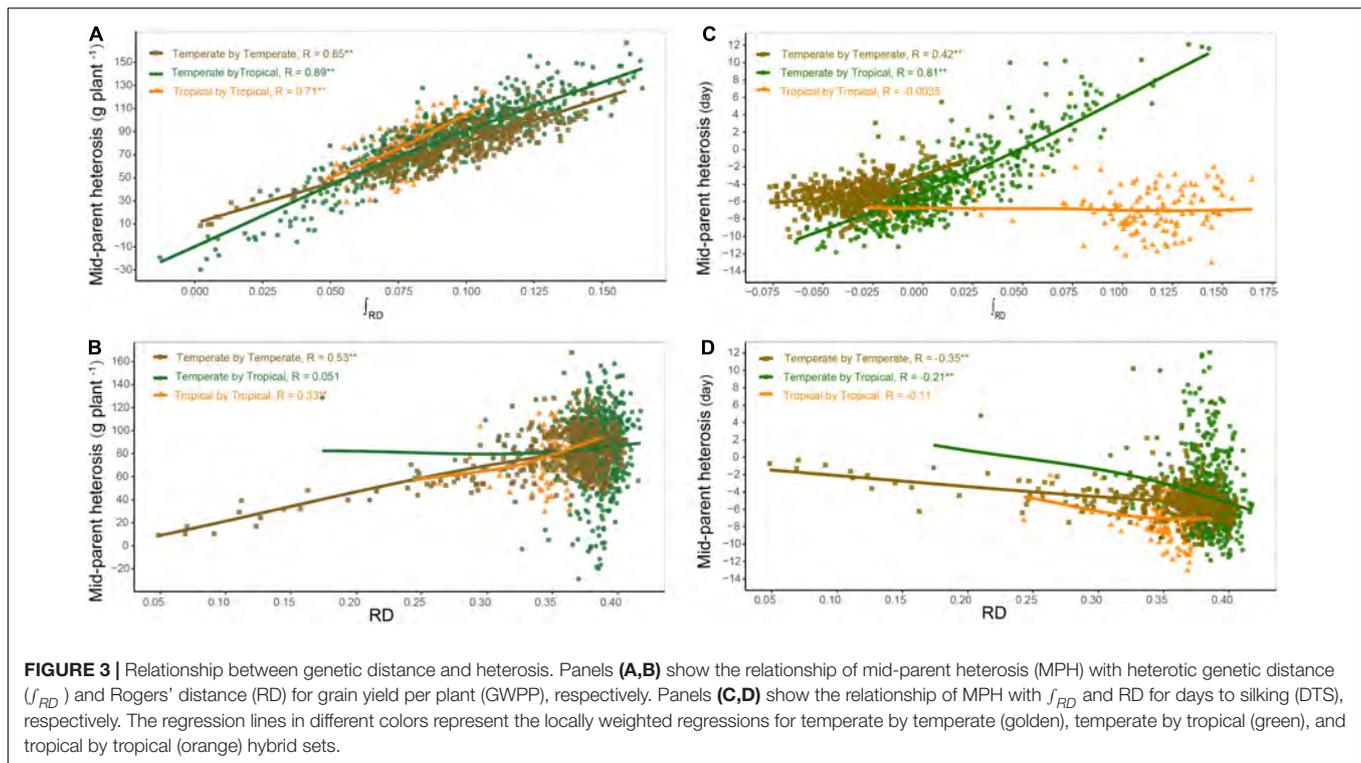
Based on the framework proposed by Jiang et al. (2017), we conducted whole-genome prediction of heterosis for GWPP using 644 temperate by tropical hybrids and 377 temperate by temperate hybrids by modeling digenic epistatic and dominance effects. As revealed by five-fold cross-validation with 100 runs, the model predicted 68.43 and 78.56% of heterosis-related genetic variances for GWPP, and 65.93 and 80.76% for DTS, in the temperate by tropical and the temperate by temperate hybrid sets, respectively (**Supplementary Table 4**). The additive-by-additive

model performed better than the dominance effects, with an increase in the genome-wide prediction accuracy by 6.5 and 4.5% in the temperate by tropical hybrids for GWPP and DTS, respectively. However, the combination of the effects in the two models of digenic epistatic and dominance did not affect prediction accuracy, which was probably caused by the high correlation between epistasis and dominance kinship matrices derived from the markers (**Supplementary Table 2** and **Supplementary Figure 7**). Additionally, whole-genome prediction accuracies for GWPP based on the dominance effects outperformed the additive-by-additive effects by 1.46% in the temperate by temperate hybrids, and combining both types of effects did not improve the prediction accuracy, mainly because the dominance model explained the major portion of genetic variance (**Figure 4**; **Supplementary Figure 7**). Partitioning the total genetic variance of heterosis into its components described a particularly important role in epistasis (**Figure 4**), and the additive-by-dominance and additive-by-additive epistasis largely contributed to heterosis in the temperate by tropical hybrids.

Genome-wide association mapping for GWPP discovered four SNP loci with significant dominance effects in the temperate by temperate hybrids and 33 pairs of markers with significant epistatic effects (**Figures 5A–F** and **Supplementary Data Sheets 2, 3**) in the temperate by tropical hybrids, including 11 additive-by-additive, 13 additive-by-dominance, and nine dominance-by-dominance interactions (**Figures 5A,C**). The absence of significant dominance effects in the temperate by tropical hybrids could be partly interpreted by the low contribution (38%) of the dominance effects to the grain-yield heterosis-related genetic variance. The dominance effects in the temperate by temperate hybrids were positive, while 31 of the 33 (93.9%) epistatic effects in the temperate by tropical hybrids were negative. Whether there were significant heterotic QTL effects was determined by the association of the detected heterosis with the estimated contribution of single heterosis QTL, with 14 and 3 heterotic QTL identified in the temperate by tropical and the temperate by temperate hybrid sets, explaining 32.67 and 31.16% of the mean phenotypic variance, respectively (**Figures 5Ac,Dc**).

Genome-wide association mapping for DTS detected 81 pairs of markers with significant epistatic effects in the temperate by temperate hybrids (**Supplementary Data Sheets 4, 5**), including 34 additive-by-additive and 46 additive-by-dominance interactions (**Figures 6Da,E,F**), and 420 marker pairs with significant epistatic effects in the temperate by tropical hybrids, including 93 additive-by-additive, 145 additive-by-dominance, and 182 dominance-by-dominance interactions (**Figures 6A–C**). The absence of significant dominance effects could partly be interpreted by the low contribution of the dominance effects (21 and 40%) to the DTS heterosis-related genetic variance. Of the 420 epistatic effects identified in the temperate by tropical hybrids, 68 (16.43%) were negative (**Supplementary Data Sheet 4**), while 33 of the 238 (13.87%) epistatic effects in the temperate by temperate hybrids were negative. A total of 139 and 37 heterotic QTL were identified in the temperate by tropical and the temperate by temperate hybrid sets (**Figure 8A**; **Supplementary Data Sheet 4**), respectively, which explained

FIGURE 2 | Relationship between hybrid heterosis and mid-parent performance. **(A)** Mid-parent heterosis (MPH) of GWPP for temperate by temperate, temperate by tropical, and tropical by tropical hybrids. **(B)** Relationship between MPH and mid-parent (MP) performance for GWPP. **(C)** Distribution of MPH for days to silking for temperate by temperate, temperate by tropical, and tropical by tropical hybrids. **(D)** Relationship between MPH and MP performance for days to silking. The dashed lines in the histograms indicate the averages.

FIGURE 3 | Relationship between genetic distance and heterosis. Panels **(A,B)** show the relationship of mid-parent heterosis (MPH) with heterotic genetic distance ($\int_{RD}$) and Rogers' distance (RD) for grain yield per plant (GWPP), respectively. Panels **(C,D)** show the relationship of MPH with $\int_{RD}$ and RD for days to silking (DTS), respectively. The regression lines in different colors represent the locally weighted regressions for temperate by temperate (golden), temperate by tropical (green), and tropical by tropical (orange) hybrid sets.

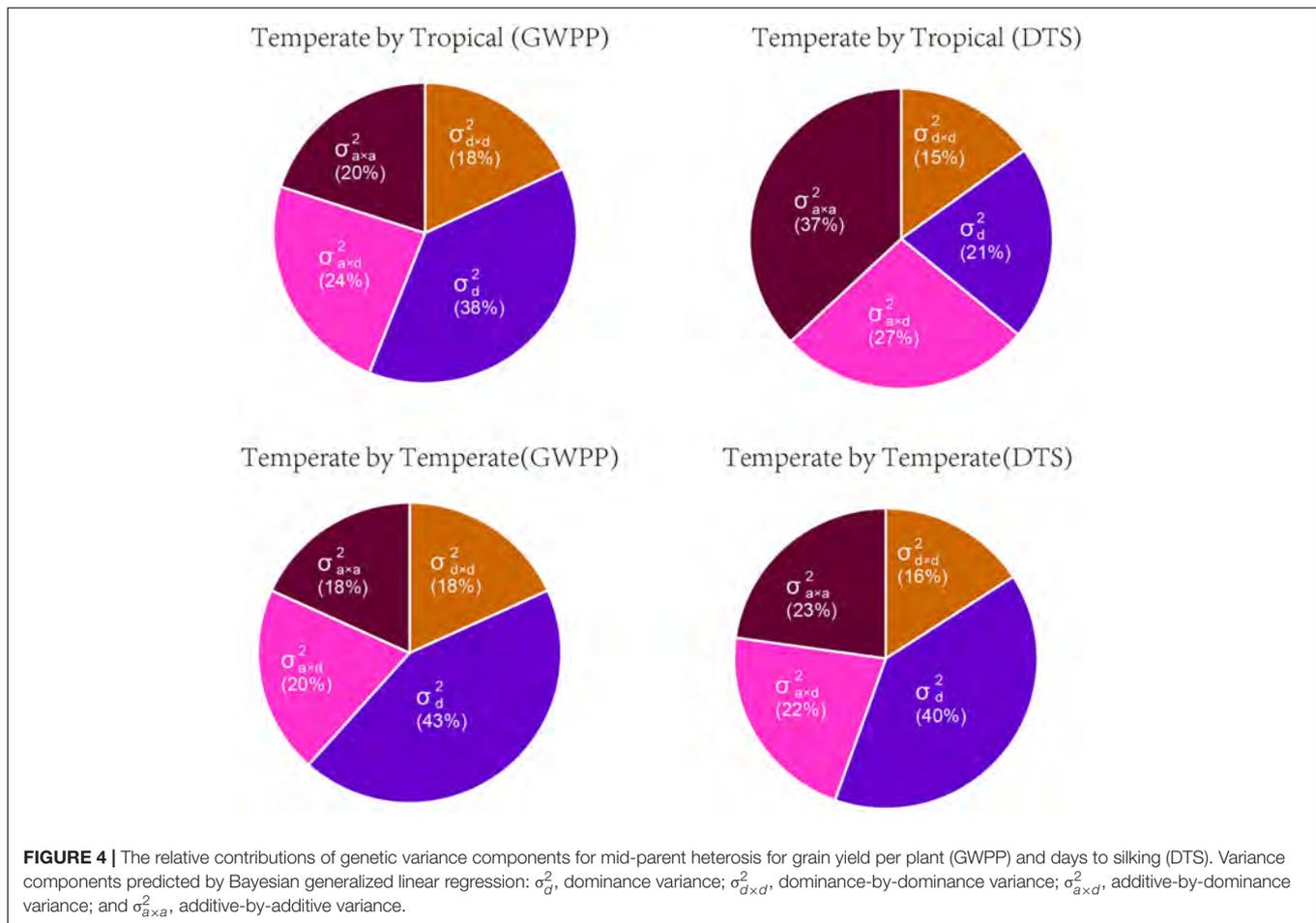31.9 and 50.71% of the phenotypic variance on average (**Figures 6Ac,Dc**).

Candidate Genes and Epistasis Identified in the Temperate by Tropical Hybrids

The interactomes associated with digenic epistasis showed that gene interactions were involved in various biological processes. A total of nine million interactions spanning all levels of genetic information flow across the entire maize lifecycle have been elucidated using the Maize Interactome platform (Han et al., 2020). For GWPP in the temperate by tropical hybrids, a significant locus (Chr2.47114632) was analyzed, along with the other five epistatic loci. A total of 16 candidate genes within 55 kb widows of the epistatic loci were identified using the online Maize Interactome platform PPI Network tool, and five core genes were found with interactions (**Figure 7A**). The five genes were used to build a protein network using the Network Creation instrument, with 90 interacted genes identified. The 90 genes were then analyzed using the online Maize Interactome platform Slim-interactive Omics Network tool, which could be classified into two modules. The first module included 25 genes enriched with GO-BP terms, which could be classified into three functional categories, photosynthesis system (photosystem II assembly, photosynthesis, light reaction, and photosynthesis), positive regulation of transcription (nucleic acid templated transcription modulation, transcription modulation, and RNA biosynthetic process modulation), and protein synthesis (protein complex assembly, protein complex biosynthesis, ribonucleoprotein complex biogenesis, protein complex subunit organization, and ribosome biogenesis) (**Figure 7B**).

For DTS in the temperate by tropical hybrids, a significant locus (Chr6.47114632) with the other nine epistatic interaction loci were analyzed. A total of 40 candidate genes within 55 kb widows of their epistatic loci were analyzed; 12 core genes showed interactions (**Figure 7C**). The 12 core genes were used to build a protein network, with 296 interacted genes and four modules identified. The first module with 64 genes was enriched with GO-BP terms, which were involved in various growth and development processes, such as "cell differentiation," "single-organism development process," "multicellular organism development," "cell development process," "cell response to stress," "reproductive process," and "cell response to stimulus" (**Figure 7D**). The third module with 31 genes was enriched with GO-BP terms, which supported various biosynthesis processes, including the "cellular carbohydrate biosynthesis process," "glucan metabolic process," and " beta-glucan biosynthesis process," and the abiotic stress responses, including the "salt stress response" and "osmotic stress response" (**Figure 7E**).

Epistasis Activation Contributes to Greater Heterosis in the Temperate by Tropical Hybrids

In the parental population, the homozygous background was assumed to suppress the expression of Gene 1 (**Figure 8A**). In the F_1 hybrids, however, such suppression was relieved due to the complementation of heterozygosity, which resulted in the activation of epistasis between Gene 1 and Gene 2. Thus, Gene 1 and Gene 2 might be considered to be epistatically controlled QTL (**Figure 8A**). Using the half-sib hybrids generated between

FIGURE 4 | The relative contributions of genetic variance components for mid-parent heterosis for grain yield per plant (GWPP) and days to silking (DTS). Variance components predicted by Bayesian generalized linear regression: σ^2_d, dominance variance; $\sigma^2_{d \times d}$, dominance-by-dominance variance; $\sigma^2_{a \times d}$, additive-by-dominance variance; and $\sigma^2_{a \times a}$, additive-by-additive variance.

the temperate and tropical maize, the genetic mechanisms of heterosis for single-crosses might be explained. The alleles at Gene 1 had additive or dominant effects, but they were undetectable by GWAS in the temperate population due to the absence of the weak activating allele and no phenotypic difference. However, a phenotypic difference was found in the temperate by tropical hybrids, as shown by GWPP. One possible mechanism for the difference identified using the two sets of hybrids might involve transcriptional regulation and PPI, that is, the upstream PPI network in the hybrids released the inhibition of background Gene 1, and the product of Gene 1 further activated the expression of Gene 2 (**Figure 8B**).

The molecular interpretation of the above-mentioned epistasis can be exemplified by the QTL chr2.47114632 and the epistatic QTL chr6.119876370. We hypothesized that this epistatic interaction agreed with the scenario illustrated in **Figure 8B**, because the genotype chr2.47114632-AA was detected in only three tropical parents. Genetic analysis of the two loci, chr2.47114632 and chr6.119876370, indicated that the homozygous genotype at chr2.47114632 was probably inhibited by the background genotype in the parent lines. In the F_1 hybrids, however, a specific of PPI occurred when two parental genomes combined, as the proteome of each parent might provide novel interacting partners (Li et al., 2020). As a

result, the repression at chr2.47114632 was removed, which allowed the expression of the chr2.47114632 allele. The full expression of the genotype chr2.47114632-TT activated the chr6.119876370 allele to fully express GWPP, and the genotype chr2.47114632-AT activated the chr6.119876370 allele to weakly express GWPP (**Figure 9A**). The three tropical lines carrying the AA genotypes at chr2.47114632 exhibited lower GWPP in the parent lines, and in the F_1 hybrids, the genotype chr2.47114632-AT also had a relatively low GWPP, indicating that the genotype chr2.47114632-AA was unfavorable to GWPP. Similar interaction occurred between chr2.47114632 and three other loci (**Figures 9B–D**), indicating that chr2.47114632 might host an important regulatory gene involved in multiple biological processes in the formation of GWPP.

DISCUSSION

The hybrids generated from inbred lines in the earlier generation showed higher levels of fitness and heterosis (Rhode and Cruzan, 2005). Various models have been proposed to explain heterosis, including dominance, overdominance, and epistasis with complex allelic, intragenomic, and intergenomic interactions (Birchler et al., 2010; Kaeppler, 2012). According
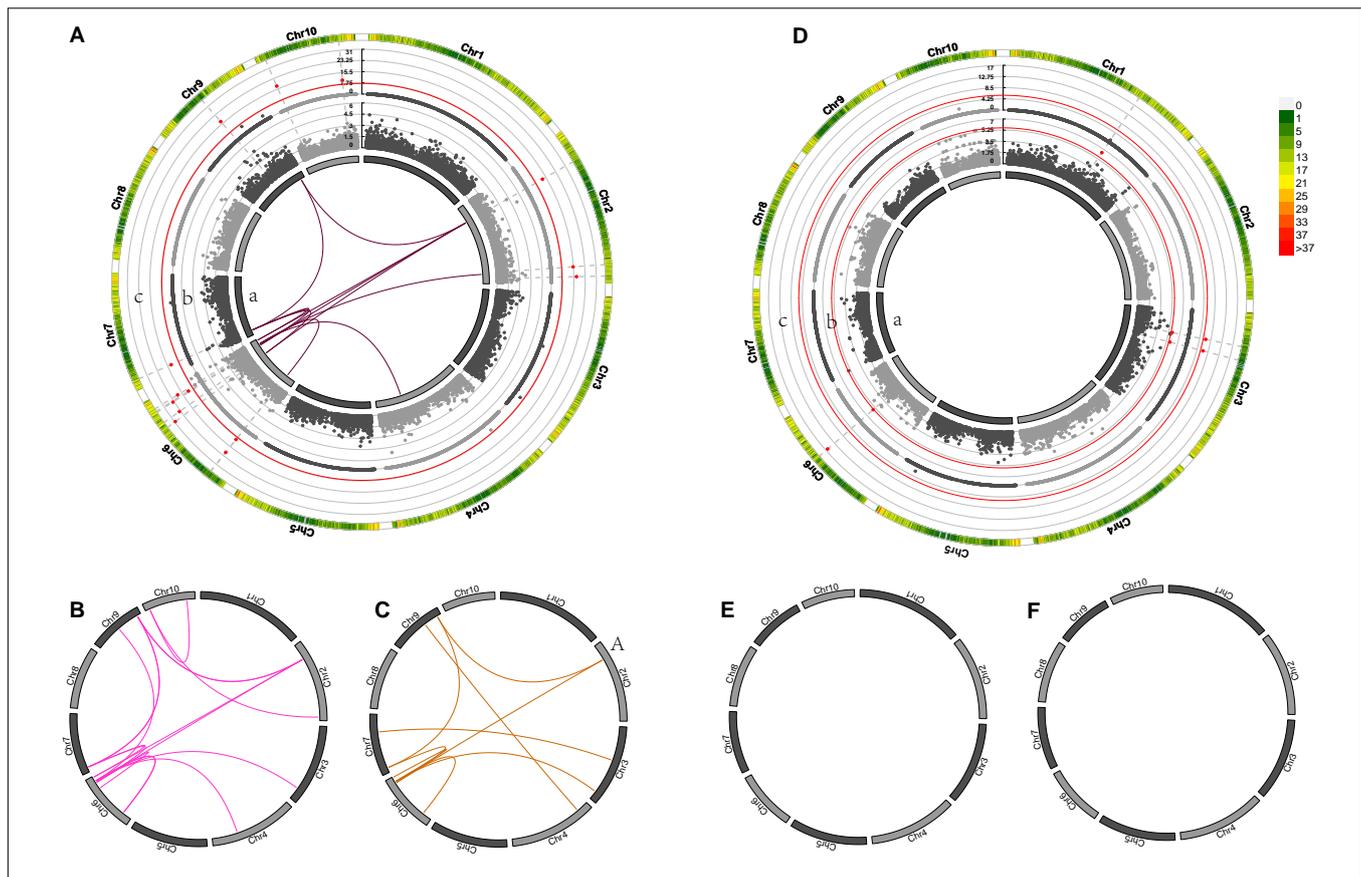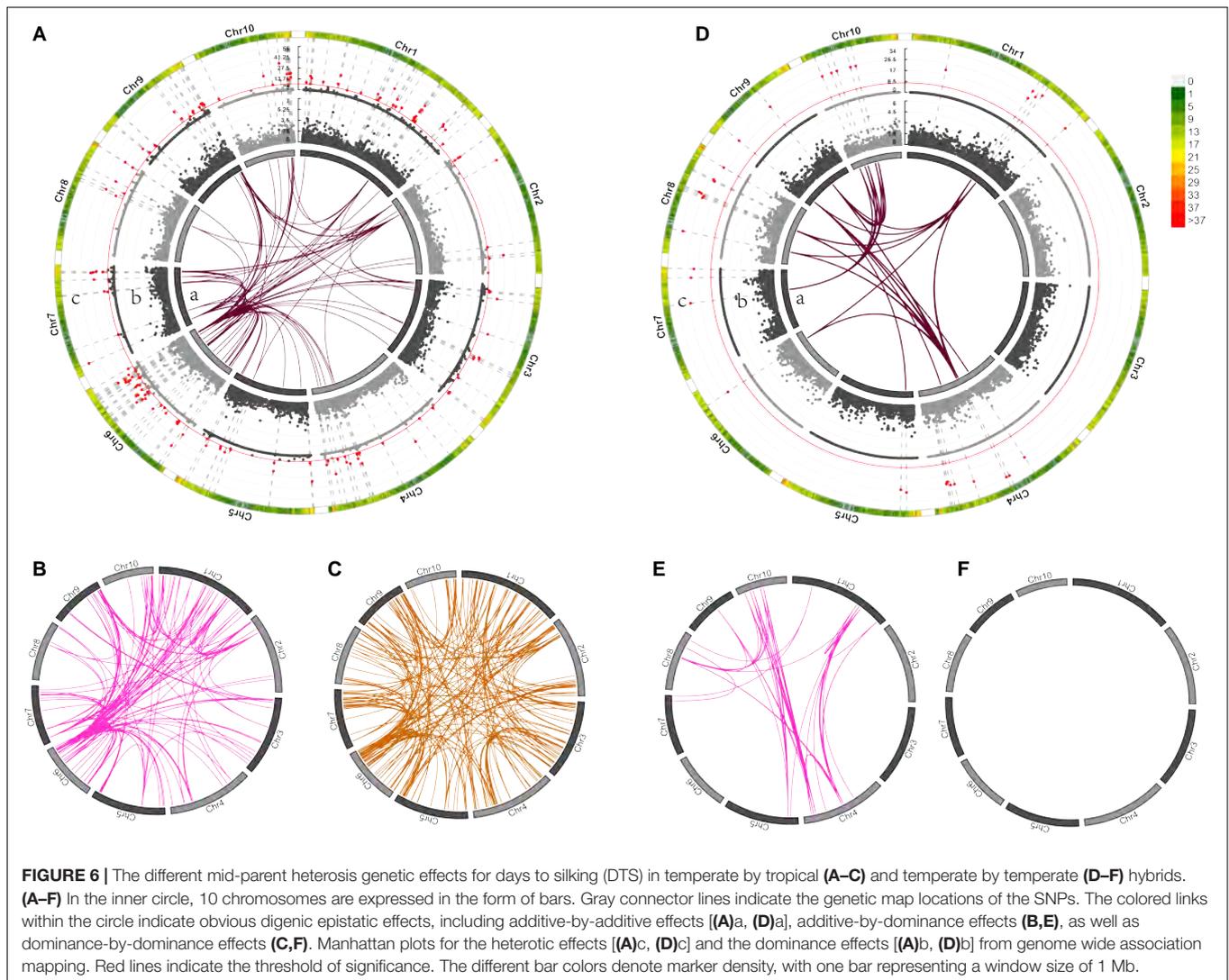
FIGURE 5 | Genetic effects of mid-parent heterosis for maize GWPP in temperate by tropical **(A–C)** and temperate by temperate **(D–F)** hybrids. **(A–F)** In the inner circle, 10 chromosomes are indicated by bars. Gray lines indicate the genetic map locations of SNPs. Color links within the circle indicate significant digenic epistatic interactions, including additive-by-additive [**(A)**a, **(D)**a], additive-by-dominance **(B,E)**, and dominance-by-dominance **(C,F)** interactions. Manhattan plots for the heterotic effects [**(A)**c, **(D)**c] and the dominance effects [**(A)**b, **(D)**b] from genome wide association mapping are presented using Manhattan plots from GWAS. Red lines indicate the threshold of significance. Bar colors denote marker density, with one bar representing a window size of 1 Mb.

to recent studies on rice and maize (Huang et al., 2015; Yang et al., 2017), most heterotic genes exhibit incomplete or partial dominance. However, the relationship between multiple-locus epistasis and heterosis has been under-studied due to the complexity and difficulty in profiling a bonafide and comprehensive interactome in F$_1$ hybrids (Li et al., 2020). Additionally, the overall contribution of individual small-scale mutations to heterosis was mostly weaker than the effects arising from genetic variations in major alleles, but the cumulative effects from dozens to hundreds of small-scale mutations might contribute to the undiscovered component of heterosis. The unique design of multi-hybrid populations derived from the temperate and tropical lines provides a chance to understand the genetic mechanism of heterosis. Furthermore, tropical maize with favorable alleles for abiotic and biotic stress resistance should be used to increase genetic diversity and accumulate favorable QTL with minor effects. With such a design, the alleles in the tropical lines with minor positive or negative effects on fitness might have been retained, resulting in the accumulation of many minor favorable or deleterious mutations. Therefore, the founder lines used in our study can be used to detect such rare

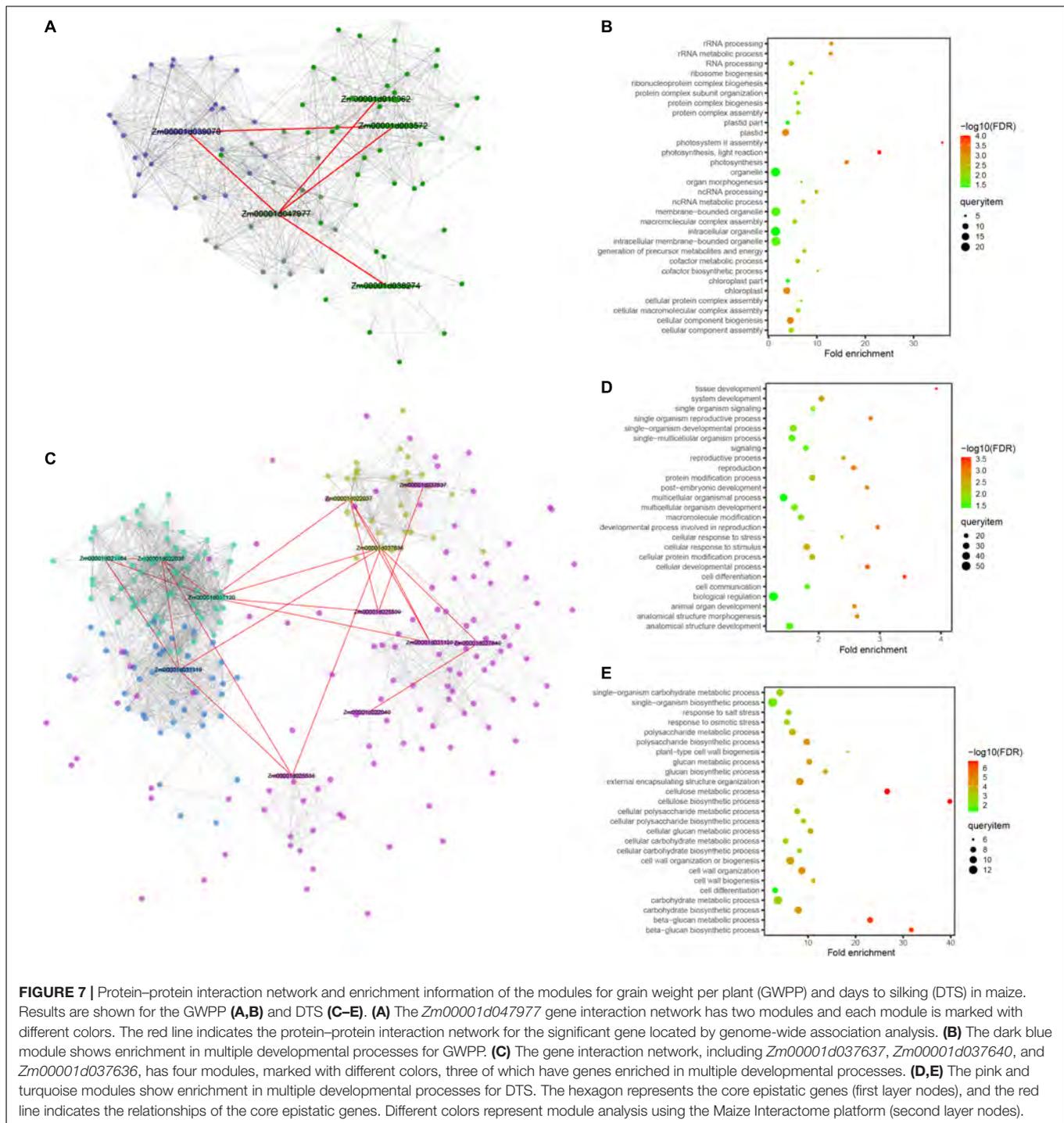mutations to increase the understanding of the contribution of minor gene interactions to heterosis.

Hybrid breeding largely depends on heterosis, which can be determined by the genetic distance between parental lines (Tian et al., 2016). The level of heterosis is generally related to parental genetic diversity (Tian et al., 2019). However, some studies have found a weak or no relationship between marker diversity and yield (Oyekunle et al., 2015; Boeven et al., 2020). Therefore, genetic diversity is important but is not enough to give rise to desirable heterosis performance (Fu et al., 2014). Heterosis in the temperate by tropical and temperate by temperate hybrid sets increased with the heterotic genetic distance, and the temperate by tropical hybrids showed a relatively higher level of heterosis for grain yield. This indicated that intensive selection by breeding, possible genetic drift, or both had created a divergence between the two maize groups, resulting in an increase in heterosis associated with a non-dominant effect. This is supported by the fact that GWPP showed a weak correlation with plant height ($r = 0.07$), grain number per row (0.25), ASI ($r = 0.17$), and hundred-grain weight (0.34) for tropical parents, due to the negative correlation between mid-parent performance and

FIGURE 6 | The different mid-parent heterosis genetic effects for days to silking (DTS) in temperate by tropical **(A–C)** and temperate by temperate **(D–F)** hybrids. **(A–F)** In the inner circle, 10 chromosomes are expressed in the form of bars. Gray connector lines indicate the genetic map locations of the SNPs. The colored links within the circle indicate obvious digenic epistatic effects, including additive-by-additive effects [**(A)**a, **(D)**a], additive-by-dominance effects **(B,E)**, as well as dominance-by-dominance effects **(C,F)**. Manhattan plots for the heterotic effects [**(A)**c, **(D)**c] and the dominance effects [**(A)**b, **(D)**b] from genome wide association mapping. Red lines indicate the threshold of significance. The different bar colors denote marker density, with one bar representing a window size of 1 Mb.

heterosis (**Supplementary Figure 8**). Genetic divergence occurs in isolated populations, which decreases the hybrid fitness, but there are many instances where hybrids show a higher degree of fitness than their parents (Dagilis et al., 2019). Thus, in this study, we adopted the quantitative genetic framework proposed by Jiang et al. (2017) to understand possible genetic causes for heterosis and identified epistatic effects that contributed to stronger heterosis in the temperate by tropical hybrids. Additionally, our results indicated that the most prevalent heterosis was controlled by epistatic genes in the tropical by temperate hybrids, and the prevalence of multiple-locus epistatic interactions might explain the genetic control of hybrid vigor in general. This is analogous to Fisher's geometric model where heterosis is involved in the crosses between distantly related inbred lines, and the fitness values of the hybrids include epistatic effects among many loci (Simon et al., 2018; Dagilis et al., 2019).
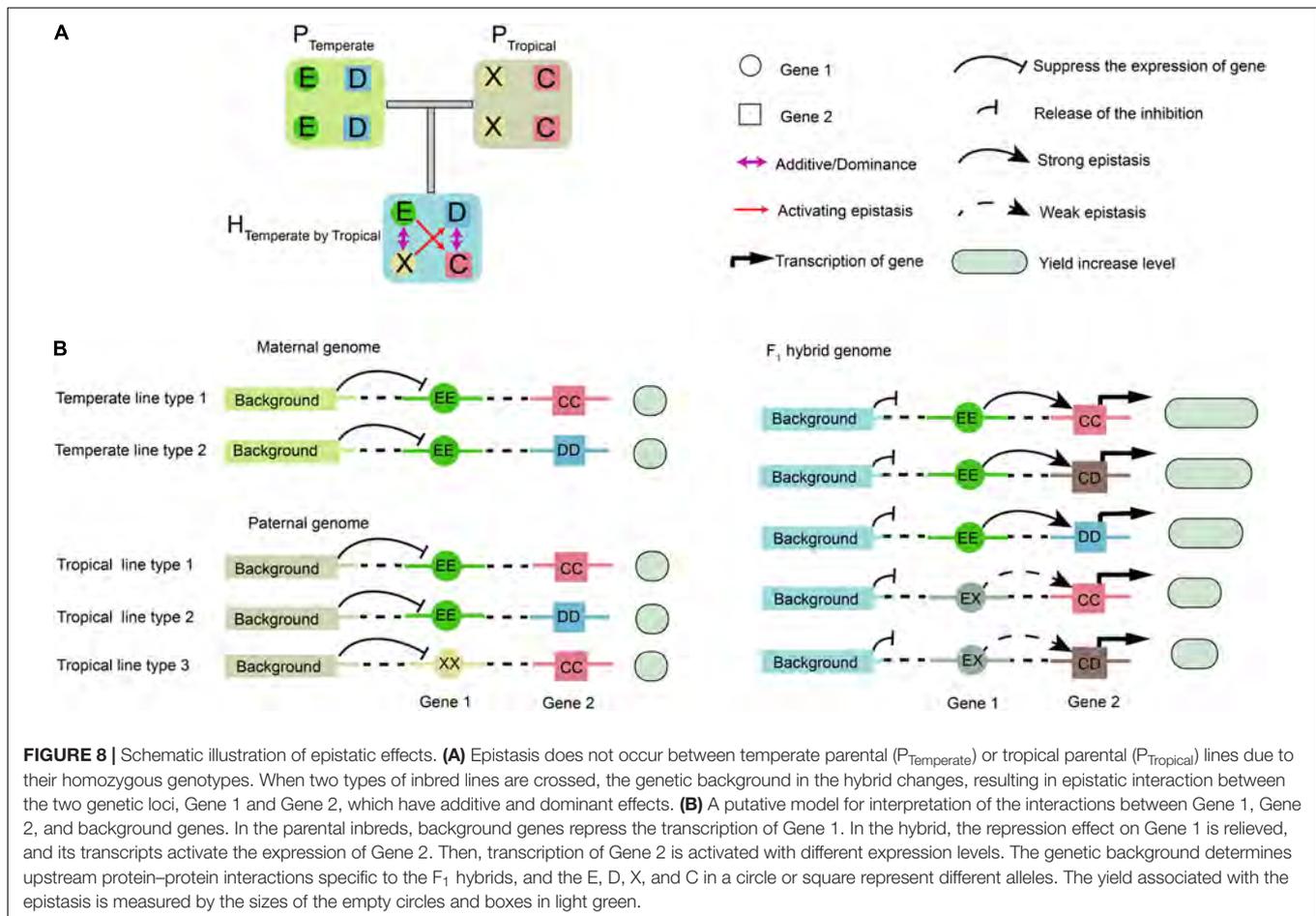
The complexity of the digenic epistatic network identified in this study suggests that many rare genes with minor effects were modulated by the core genes for yield heterosis. The signals of

selection identified between the temperate and tropical hybrids indicated the presence of polygenic heterogeneity along the whole genome (**Figure 1C**), and in plants, GWPP and DTS are some examples of traits with a polygenic basis. Combining GWAS hits with the PPI networks, we found that the epistatic genes identified by GWAS interacted with many undetected background genes, indicating that genomic variations might cause the expression of many minor differential genes and molecular interactions between the two parents. Additionally, the PPI genes for GWPP were mainly related to photosynthesis, regulation of transcription, and protein complex assembly, suggesting that enhanced photosynthetic or biological pathways during development might be associated with hybrid vigor. The maize yield heterosis results from multiple QTL effects accumulated during the development of the hybrid plants (Xiao et al., 2021). In our study, many epistatic QTL were discovered simultaneously within network pathways for different traits, with nine heterosis QTL identified in common for GWPP and DTS in the temperate by tropical hybrids (**Supplementary Figure 9**). The

FIGURE 7 | Protein–protein interaction network and enrichment information of the modules for grain weight per plant (GWPP) and days to silking (DTS) in maize. Results are shown for the GWPP **(A,B)** and DTS **(C–E)**. **(A)** The *Zm00001d047977* gene interaction network has two modules and each module is marked with different colors. The red line indicates the protein–protein interaction network for the significant gene located by genome-wide association analysis. **(B)** The dark blue module shows enrichment in multiple developmental processes for GWPP. **(C)** The gene interaction network, including *Zm00001d037637*, *Zm00001d037640*, and *Zm00001d037636*, has four modules, marked with different colors, three of which have genes enriched in multiple developmental processes. **(D,E)** The pink and turquoise modules show enrichment in multiple developmental processes for DTS. The hexagon represents the core epistatic genes (first layer nodes), and the red line indicates the relationships of the core epistatic genes. Different colors represent module analysis using the Maize Interactome platform (second layer nodes).

floral transition might be a key stage in the formation of heterosis, where epistatic QTL are activated by parental contributions of alleles that counteract the recessive deleterious maternal alleles (Xiao et al., 2021). The correlation of heterosis between yield and many other traits suggests that yield heterosis reflects both the cumulative influence of heterosis with minor effects for many traits and the interaction through various molecular mechanisms (Flint-Garcia et al., 2009). Although our data and
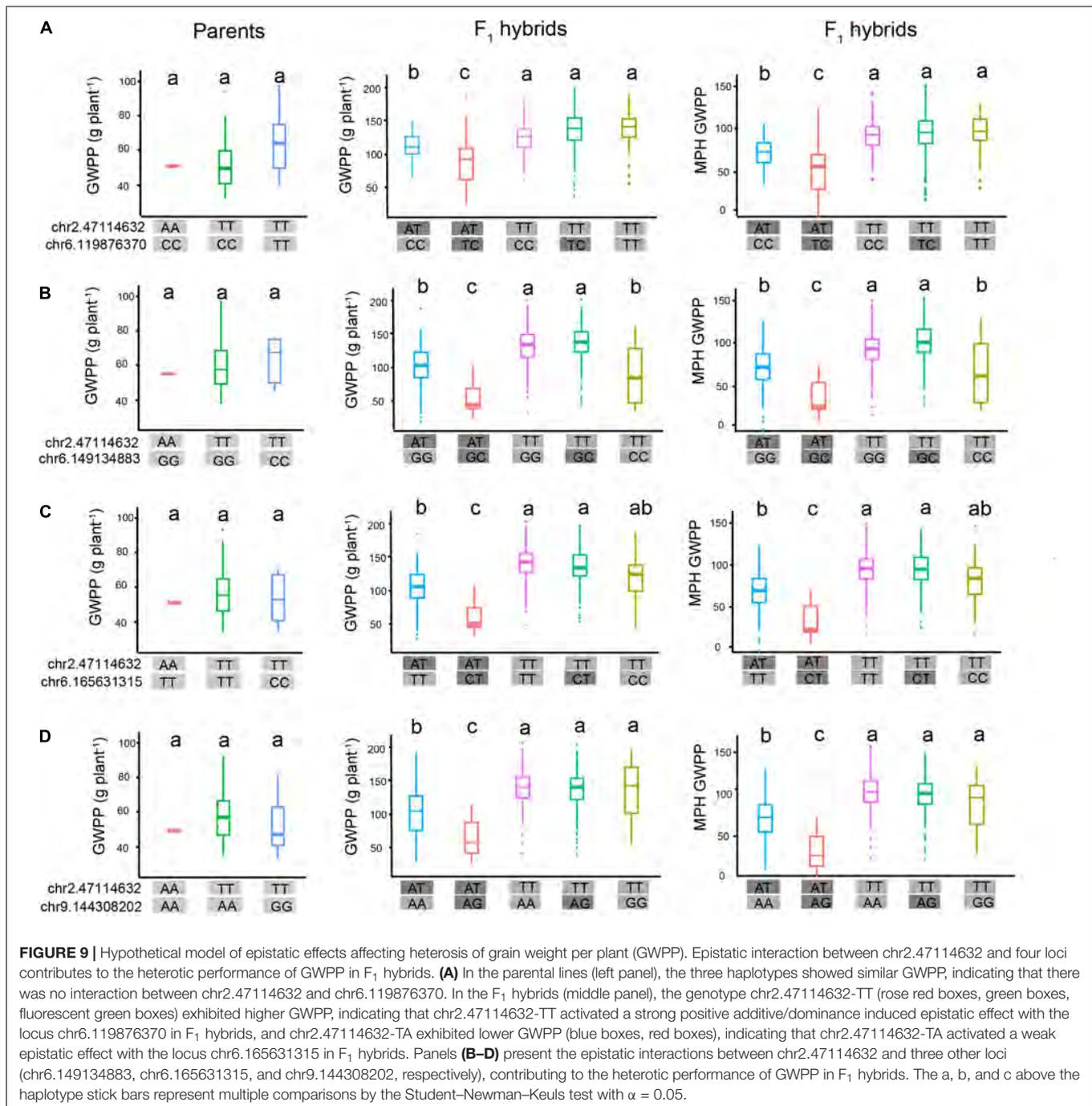
results were compelling, our study had some limitations. We obtained GWAS hits from the experimental sets of different sample sizes, which could have introduced bias. However, we mitigated this effect by taking the different significance thresholds for each experimental set. We suggest that future studies should use other methods to validate our candidate core genes, by deep sequencing to identify rare variants and wet-lab experiments to validate yield relevance.

FIGURE 8 | Schematic illustration of epistatic effects. **(A)** Epistasis does not occur between temperate parental (P_Temperate) or tropical parental (P_Tropical) lines due to their homozygous genotypes. When two types of inbred lines are crossed, the genetic background in the hybrid changes, resulting in epistatic interaction between the two genetic loci, Gene 1 and Gene 2, which have additive and dominant effects. **(B)** A putative model for interpretation of the interactions between Gene 1, Gene 2, and background genes. In the parental inbreds, background genes repress the transcription of Gene 1. In the hybrid, the repression effect on Gene 1 is relieved, and its transcripts activate the expression of Gene 2. Then, transcription of Gene 2 is activated with different expression levels. The genetic background determines upstream protein–protein interactions specific to the F₁ hybrids, and the E, D, X, and C in a circle or square represent different alleles. The yield associated with the epistasis is measured by the sizes of the empty circles and boxes in light green.

Our GWAS and haplotype analyses indicate that epistasis contributes to the greater heterosis identified in the tropical by temperate hybrids. Epistasis might be related to various molecular interactions, and single or combined alterations of hybrid biological networks might contribute to heterosis to different degrees (Li et al., 2020). The complementary dominant gene expression of hybrids between transcriptional regulatory networks involved in biological pathways across developmental stages contributes to heterosis (Liu et al., 2021; Li et al., 2022). Many minor complementary dominant genes activate epistatic interactions with their PPI genes, resulting in a cascade of amplified phenotypic effects in the hybrids. Multiple alleles at Chr2.47114632 had different types of epistatic effects, producing a series of amplified effects on upstream and downstream gene regulation. This might explain why we detected many epistatic QTL through GWAS with a very strict significance threshold, although the overall contribution of complementary harmful alleles to heterosis was also found. The gene GRMZM2G147158 encoding calmodulin-binding protein 60 C (CBP60C) located at chr2.47114632 belongs to a plant-specific protein family that plays an important role in plant growth/development and biotic/abiotic stress responses, and CBP60 is a central transcriptional activator of immunity in *Arabidopsis* and positively regulates salicylic acid and abscisic

acid biosynthesis (Wan et al., 2012; Li et al., 2021; Yu et al., 2021). Our results suggest that the concurrent detection of chr2.47114632 with the other four loci arose from different epistatic effects, and thus, the epistatic genes might be a central transcriptional activator to regulate downstream gene expression.

Designing optimal genotypic combinations and increasing the favorable alleles among parental lines might further enhance hybrid performance. In this study, we found that increasing parental heterotic genetic distance is necessary for maximizing heterosis. The mid-parent values only contributed to a fraction of hybrid performance in maize, while the lines with high GCA for yield were more suitable as parents for hybrid breeding (**Supplementary Figure 10**). Conversely, the mid-parent values accounted for a large proportion of hybrid performance in wheat hybrids (Zhao et al., 2013; Jiang et al., 2017; Boeven et al., 2020). In this study, mid-parent values were relatively stable across all cross combinations, and the increase in the performance of the hybrids was mainly contributed by mid-parent heterosis (**Supplementary Figures 10A,B**). With the increase in the mid-parent values, heterosis decreased at the beginning and then remained stable. Thus, whether hybrid yield can be improved with a further increase in the parental genetic distance while maintaining a certain level of parental

FIGURE 9 | Hypothetical model of epistatic effects affecting heterosis of grain weight per plant (GWPP). Epistatic interaction between chr2.47114632 and four loci contributes to the heterotic performance of GWPP in F_1 hybrids. **(A)** In the parental lines (left panel), the three haplotypes showed similar GWPP, indicating that there was no interaction between chr2.47114632 and chr6.119876370. In the F_1 hybrids (middle panel), the genotype chr2.47114632-TT (rose red boxes, green boxes, fluorescent green boxes) exhibited higher GWPP, indicating that chr2.47114632-TT activated a strong positive additive/dominance induced epistatic effect with the locus chr6.119876370 in F_1 hybrids, and chr2.47114632-TA exhibited lower GWPP (blue boxes, red boxes), indicating that chr2.47114632-TA activated a weak epistatic effect with the locus chr6.165631315 in F_1 hybrids. Panels **(B–D)** present the epistatic interactions between chr2.47114632 and three other loci (chr6.149134883, chr6.165631315, and chr9.144308202, respectively), contributing to the heterotic performance of GWPP in F_1 hybrids. The a, b, and c above the haplotype stick bars represent multiple comparisons by the Student–Newman–Keuls test with α = 0.05.

performance is uncertain. With known heterotic groups, maize breeders usually use pedigree breeding for breeding inbred lines, by which dominant genes between parental lines from two heterotic groups can complement each other. However, early-generation selection lines are characterized by a high level of heterozygosity, making it impossible to select such dominant loci because of lower favorable allele frequencies and the complexity of epistasis. Thus, in the early generations, abiotic and biotic stress resistance, plant maturity, plant height, and other traits are prioritized for selection. In later generations, a large effective population is crucial for selecting desirable combinations of genotypes for grain yield.

DATA AVAILABILITY STATEMENT

The original contributions presented in this study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

YX and ZS conceived and designed the experiments and wrote the manuscript. ZS, HW, and ZZ performed the experiments. ZS, YY, XL, and ZL analyzed the data. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.921608/full#supplementary-material

REFERENCES

Alves, F. C., Granato, ÍS. C., Galli, G., Lyra, D. H., Fritsche-Neto, R., and De Los Campos, G. (2019). Bayesian analysis and prediction of hybrid performance. *Plant Methods* 15:14.

Badu-Apraku, B., Oyekunle, M., Akinwale, R., and Aderounmu, M. (2013). Combining ability and genetic diversity of extra-early white maize inbreds under stress and nonstress environments. *Crop Sci.* 53, 9–26. doi: 10.2135/cropsci2012.06.0381

Bauman, L. F. (1959). Evidence of Non-Allelic Gene Interaction in Determining Yield, Ear Height, and Kernel Row Number in Corn 1. *Agron. J.* 51, 531–534. doi: 10.2134/agronj1959.00021962005100090007x

Birchler, J. A. (2015). Heterosis: the genetic basis of hybrid vigour. *Nat. Plants* 1:15020. doi: 10.1038/nplants.2015.20

Birchler, J. A., Yao, H., Chudalayandi, S., Vaiman, D., and Veitia, R. A. (2010). Heterosis. *Plant Cell* 22, 2105–2112. doi: 10.1105/tpc.110.076133

Boeven, P. H., Zhao, Y., Thorwarth, P., Liu, F., Maurer, H. P., Gils, M., et al. (2020). Negative dominance and dominance-by-dominance epistatic effects reduce grain-yield heterosis in wide crosses in wheat. *Sci. Adv.* 6:eaay4897. doi: 10.1126/sciadv.aay4897

Boggs, P. T., and Rogers, J. E. (1990). Orthogonal distance regression. *Contemp. Math.* 112, 183–194.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Butler, D. G., Cullis, B. R., Gilmour, A. R., and Gogel, B. J. (2009). *ASREMLQ23 R Reference Manual*. Release 3.0. Technical Report, Queensland Department of Primary Industries, Australia. Available online at: http://www.vsni.co.uk/downloads/asreml/release2/doc/asreml-R.pdf (accessed June 27, 2021).

Culverhouse, R., Klein, T., and Shannon, W. (2004). Detecting epistatic interactions contributing to quantitative traits. *Gen. Epidemiol.* 27, 141–152. doi: 10.1002/gepi.20006

Dagilis, A. J., Kirkpatrick, M., and Bolnick, D. I. (2019). The evolution of hybrid fitness during speciation. *PLoS Genet.* 15:e1008125. doi: 10.1371/journal.pgen.1008125

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

Doebley, J., Stec, A., and Gustus, C. (1995). teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* 141, 333–346. doi: 10.1093/genetics/141.1.333

Doust, A. N., Lukens, L., Olsen, K. M., Mauro-Herrera, M., Meyer, A., and Rogers, K. (2014). Beyond the single gene: how epistasis and gene-by-environment effects influence crop domestication. *Proc. Natl. Acad. Sci. U.S.A* .111, 6178–6183. doi: 10.1073/pnas.1308940110

East, E. M. (1936). Heterosis. *Genetics* 21, 375–397. doi: 10.1093/genetics/21.4.375

Fiévet, J. B., Dillmann, C., and De Vienne, D. (2010). Systemic properties of metabolic networks lead to an epistasis-based model for heterosis. *Theor. Appl. Genet.* 120, 463–473. doi: 10.1007/s00122-009-1203-2

Flint-Garcia, S. A., Buckler, E. S., Tiffin, P., Ersoz, E., and Springer, N. M. (2009). Heterosis is prevalent for multiple traits in diverse maize germplasm. *PLoS One* 4:e7433. doi: 10.1371/journal.pone.0007433

Frankham, R. (1996). *Introduction to Quantitative Genetics*, 4th Edn. London: Longman.

Fu, D., Xiao, M., Hayward, A., Fu, Y., Liu, G., Jiang, G., et al. (2014). Utilization of crop heterosis: a review. *Euphytica* 197, 161–173.

Hallauer, A. R., Carena, M. J., and Filho, J. D. (2010). *Quantitative Genetics in Maize Breeding*. London: Springer Science & Business Media.

Han, L., Zhong, W., Qian, J., Jin, M., Tian, P., Zhu, W., et al. (2020). An Interactome Map of Maize (Zea mays L.). *Biol. Sci.* [preprint]. doi: 10.21203/rs.3.rs-88694/v2

Harper, L., Gardiner, J., Andorf, C., and Lawrence, C. J. (2016). "MaizeGDB: the maize genetics and genomics database," in *Plant Bioinformatics* (New York: Springer), 187–202.

Hochholdinger, F., and Baldauf, J. A. (2018). Heterosis in plants. *Curr. Biol.* 28, R1089–R1092. doi: 10.1016/j.cub.2018.06.041

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Statist.* 6, 65–70.

Hua, J., Xing, Y., Wu, W., Xu, C., Sun, X., Yu, S., et al. (2003). Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. U.S.A.* 100, 2574–2579. doi: 10.1073/pnas.0437907100

Huang, W., and Mackay, T. F. (2016). The genetic architecture of quantitative traits cannot be inferred from variance component analysis. *PLoS Genet.* 12:e1006421. doi: 10.1371/journal.pgen.1006421

Huang, X., Yang, S., Gong, J., Zhao, Y., Feng, Q., Gong, H., et al. (2015). Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat. Commun.* 6:6258. doi: 10.1038/ncomms7258

Jiang, Y., Schmidt, R. H., Zhao, Y., and Reif, J. C. (2017). A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. *Nat. Genet.* 49, 1741–1746. doi: 10.1038/ng.3974

Jinks, J., and Jones, R. M. (1958). Estimation of the components of heterosis. *Genetics* 43, 223–234. doi: 10.1093/genetics/43.2.223

Jones, D. F. (1917). Dominance of linked factors as a means of accounting for heterosis. *Genetics* 2, 466–479. doi: 10.1093/genetics/2.5.466

Jung, M., Ching, A., Bhattramakki, D., Dolan, M., Tingey, S., Morgante, M., et al. (2004). Linkage disequilibrium and sequence diversity in a 500-kbp region around the adh1 locus in elite maize germplasm. *Theor. Appl. Genet.* 109, 681–689.

Kaeppler, S. (2012). Heterosis: many genes, many mechanisms—end the search for an undiscovered unifying theory. *ISRN Bot.* 2012:682824. doi: 10.5402/2012/682824

Lamkey, K. R., and Edwards, J. W. (1998). "Heterosis: theory and estimation," in *Proceedings of the 34th 1998 Illinois Corn Breeders' School*, Urbana, IL: University of Illinois 62–77.

Lamkey, K. R., and Edwards, J. W. (1999). "Quantitative genetics of heterosis," in *Proceedings of the International Symposium on the Genetics and Exploitation of Heterosis in Crops*, eds J. G. Coors and S. Pandey (Mexico: CIMMYT), 31–48.

Laude, T., and Carena, M. (2015). Genetic diversity and heterotic grouping of tropical and temperate maize populations adapted to the northern US Corn Belt. *Euphytica* 204, 661–677.

Li, D., Lu, X., Zhu, Y., Pan, J., Zhou, S., Zhang, X., et al. (2022). The multi-omics basis of potato heterosis. *J. Integr. Plant Biol.* 64, 671–687. doi: 10.1111/jipb.13211

Li, H., Jiang, S., Li, C., Liu, L., Lin, Z., He, H., et al. (2020). The hybrid protein interactome contributes to rice heterosis as epistatic effects. *Plant J.* 102, 116–128. doi: 10.1111/tpj.14616

Li, L. S., Ying, J., Li, E., Ma, T., Li, M., Gong, L. M., et al. (2021). Arabidopsis CBP60b is a central transcriptional activator of immunity. *Plant Physiol.* 186, 1645–1659. doi: 10.1093/plphys/kiab164

Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I., and Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. *Nat. Methods* 8, 833–835. doi: 10.1038/nmeth.1681

Liu, W., He, G., and Deng, X. W. (2021). Biological pathway expression complementation contributes to biomass heterosis in Arabidopsis. *Proc. Natl. Acad. Sci.* 118:e2023278118. doi: 10.1073/pnas.2023278118

Lu, Y., Shah, T., Hao, Z., Taba, S., Zhang, S., Gao, S., et al. (2011). Comparative SNP and haplotype analysis reveals a higher genetic diversity and rapider LD decay in tropical than temperate germplasm in maize. *PLoS One* 6:e24861. doi: 10.1371/journal.pone.0024861

Lu, Y., Yan, J., Guimaraes, C. T., Taba, S., Hao, Z., Gao, S., et al. (2009). Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. *Theor. Appl. Genet.* 120, 93–115.

Ma, X., Tang, J., Teng, W., Yan, J., Meng, Y., and Li, J. (2007). Epistatic interaction is an important genetic basis of grain yield and its components in maize. *Mol. Breed.* 20, 41–51.

Möhring, J., and Piepho, H. P. (2009). Comparison of weighting in two-stage analysis of plant breeding trials. *Crop Sci.* 49, 1977–1988. doi: 10.2135/cropsci2009.02.0083

Oyekunle, M., Badu-Apraku, B., Hearne, S., and Franco, J. (2015). Genetic diversity of tropical early-maturing maize inbreds and their performance in hybrid combinations under drought and optimum growing conditions. *Field Crops Res.* 170, 55–65. doi: 10.1016/j.fcr.2014.10.005

Pérez, P., and De Los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198, 483–495. doi: 10.1534/genetics.114.164442

Powers, L. (1944). An expansion of Jones's theory for the explanation of heterosis. *Am. Natural.* 78, 275–280. doi: 10.1086/281199

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rhode, J. M., and Cruzan, M. B. (2005). Contributions of heterosis and epistasis to hybrid fitness. *Am. Nat.* 166, E124–E139. doi: 10.1086/491798

Simon, A., Bierne, N., and Welch, J. J. (2018). Coadapted genomes and selection on hybrids: fisher's geometric model explains a variety of empirical patterns. *Evol. Lett.* 2, 472–498. doi: 10.1002/evl3.66

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197

Tian, H. Y., Channa, S. A., and Hu, S. W. (2016). Relationships between genetic distance, combining ability and heterosis in rapeseed (*Brassica napus* L.). *Euphytica* 213, 1–11.

Tian, S., Xu, X., Zhu, X., Wang, F., Song, X., and Zhang, T. (2019). Overdominance is the major genetic basis of lint yield heterosis in interspecific hybrids between *G. hirsutum* and *G. barbadense*. *Heredity* 123, 384–394.

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2. 0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* 45, W122–W129. doi: 10.1093/nar/gkx382

Wan, D., Li, R., Zou, B., Zhang, X., Cong, J., Wang, R., et al. (2012). Calmodulin-binding protein CBP60 g is a positive regulator of both disease resistance and drought tolerance in Arabidopsis. *Plant Cell Rep.* 31, 1269–1281.

Wang, H., Xu, C., Liu, X., Guo, Z., Xu, X., Wang, S., et al. (2017). Development of a multiple-hybrid population for genome-wide association studies: theoretical consideration and genetic mapping of flowering traits in maize. *Sci. Rep.* 7:40239. doi: 10.1038/srep40239

Wei, X., and Zhang, J. (2018). The optimal mating distance resulting from heterosis and genetic incompatibility. *Sci. Adv.* 4:eaau5518. doi: 10.1126/sciadv.aau5518

Wright, S. (1984). *Evolution and The Genetics of Populations, Volume 1: Genetic and Biometric Foundations*. Chicago, IL: University of Chicago press.

Xiao, Y., Jiang, S., Cheng, Q., Wang, X., Yan, J., Zhang, R., et al. (2021). The genetic mechanism of heterosis utilization in maize improvement. *Genome Biol.* 22:148 doi: 10.1186/s13059-021-02370-7

Xu, C., Ren, Y., Jian, Y., Guo, Z., Zhang, Y., Xie, C., et al. (2017). Development of a maize 55 K SNP array with improved genome coverage for molecular breeding. *Mol. Breed.* 37:20.

Xu, S. (2013). Mapping quantitative trait loci by controlling polygenic background effects. *Genetics* 195, 1209–1222. doi: 10.1534/genetics.113.157032

Yang, J., Mezmouk, S., Baumgarten, A., Buckler, E. S., Guill, K. E., McMullen, M. D., et al. (2017). Incomplete dominance of deleterious alleles contributes substantially to trait variation and heterosis in maize. *PLoS Genet.* 13:e1007019. doi: 10.1371/journal.pgen.1007019

Yu, J., Pressoir, G., Briggs, W. H., Bi, I., Yamasaki, M., Doebley, J. F., et al. (2005). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702

Yu, K., Wang, H., Liu, X., Xu, C., Li, Z., Xu, X., et al. (2020). Large-scale analysis of combining ability and heterosis for development of hybrid maize breeding strategies using diverse germplasm resources. *Front. Plant Sci.* 11:660. doi: 10.3389/fpls.2020.00660

Yu, Q., Liu, Y. L., Sun, G. Z., Liu, Y. X., Chen, J., Zhou, Y. B., et al. (2021). Genome-Wide Analysis of the Soybean Calmodulin-Binding Protein 60 Family and Identification of GmCBP60A-1 Responses to Drought and Salt Stresses. *Int. J. Mol. Sci.* 22:13501. doi: 10.3390/ijms222413501

Yu, S., Li, J., Xu, C., Tan, Y., Gao, Y., Li, X., et al. (1997). Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. U.S.A.* 94, 9226–9231. doi: 10.1073/pnas.94.17.9226

Zhao, Y., Li, Z., Liu, G., Jiang, Y., Maurer, H. P., Würschum, T., et al. (2015). Genome-based establishment of a high-yielding heterotic pattern for hybrid wheat breeding. *Proc. Natl. Acad. Sci. U.S.A.* 112, 15624–15629. doi: 10.1073/pnas.1514547112

Zhao, Y., Zeng, J., Fernando, R., and Reif, J. C. (2013). Genomic prediction of hybrid wheat performance. *Crop Sci.* 53, 802–810. doi: 10.2135/cropsci2012.08.0463

Zhou, G., Chen, Y., Yao, W., Zhang, C., Xie, W., Hua, J., et al. (2012). Genetic composition of yield heterosis in an elite rice hybrid. *Proc. Natl. Acad. Sci. U.S.A.* 109, 15847–15852. doi: 10.1073/pnas.1214141109

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.