Heterosis-related genes under different planting densities in maize (*Zea mays* L.)

Juan Ma[1], Dengfeng Zhang[2], Yanyong Cao[1], Lifeng Wang[1], Jingjing Li[1], Thomas Lübberstedt[3], Tianyu Wang[2], Yu Li[2], Huiyong Li [1, *]

[1] Institute of Cereal Crops, Henan Academy of Agricultural Sciences, Zhengzhou, Henan 450002, P.R.China

[2] Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, 100081, P.R.China

[3] Department of Agronomy, Iowa State University, Ames, IA 50010, USA

*** Correspondence: lihuiyong1977@126.com**

Telephone number: 86371 65758930

Running title: Heterosis-related genes in various maize planting densities

Highlight: DEG analysis and WGCNA were used to find heterosis-related genes response to planting densities. Four genes were found for maize heterosis and high planting density.

Juan Ma: majuanjuan85@126.com

Dengfeng Zhang: zhangdengfeng@caas.cn

Yanyong Cao: yanyongcao@126.com

Lifeng Wang: wanglifeng625@126.com

Jingjing Li: jingjingli206@126.com

Thomas Lübberstedt: thomasl@iastate.edu

Tianyu Wang: wangtianyu@caas.cn

Yu Li: liyu03@caas.cn

Date of submission: 6/22/2018

None tables and 4 figures (zero figure in colour in print)

The word count: 6,481

Supplementary data: 5 tables and 16 figures

Abstract

Heterosis and increasing planting density have contributed to improving maize grain yield (GY) for several decades. As planting densities increase, the GY per plot also increases whereas the contribution of heterosis to GY decreases. There are trade-offs between heterosis and planting density, and the transcriptional characterization of heterosis may explain the mechanism involved. In this study, 48 transcriptome libraries were sequenced from four inbred Chinese maize lines and their F_1 hybrids. They were planted at densities of 45,000 plants/ha and 67,500 plants/ha. Maternal-effect differentially expressed genes (DEGs) played important roles in processes related to photosynthesis and carbohydrate biosynthesis and metabolism. Paternal-effect DEGs participated in abiotic/biotic stress response and plant hormone production under high planting density. Weighted gene co-expression network analysis revealed that high planting-density induced heterosis-related genes regulating abiotic/biotic stress response, plant hormone biosynthesis, and ubiquitin-mediated proteolysis but repressed other genes regulating energy formation. Under high planting density, maternal genes were mainly enriched in the photosynthesis reaction center, while paternal genes were mostly concentrated in the peripheral antenna system. Four important genes were identified in maize heterosis and high planting density, all with functions in photosynthesis, starch biosynthesis, auxin metabolism, gene silencing, and RNA interference.

Keywords: differentially expressed gene, heterosis, maize, planting density, transcriptome, weighted gene co-expression network analysis

Introduction

Heterosis or hybrid vigor is an important biological phenomenon that has significantly improved human food supply because agronomically important traits and performance of hybrids are superior to those of their parents.

Plant breeders have utilized heterosis to develop higher-yielding, better-performing hybrids in many crop species. Heterosis in maize (*Zea mays* L.), which was independently proposed by East and Shull in the early 1900s (Shull, 1952), became the primary reason for its successful commercialization (Stuber et al., 1992). The first known attempts at heterosis were made by nineteenth-century farmers who hybridized maize landraces (Anderson and Brown, 1952) that represented a major turning point in developing hybrids with significantly increased yield. Maize landraces (open-pollinated varieties) were developed to adapt to various environmental conditions by long-term domestication and natural and artificial selection, resulting in varieties well adapted to local cultivation conditions and highly resistant to diseases. Maize single-cross hybrids have been produced and cultivated since the late 1950s. In China before 1950, the two main landraces, Lvda Red Cob (LRC) and Tang Si Ping Tou (SPT), had been grown extensively but with low yield. By the mid to late 1970s, progress was achieved by developing significantly improved maize single-cross hybrids with high stability and yield.

Based on classical genetics, several hypotheses have been proposed to explain heterosis in single-cross hybrids. These included dominance, pseudo-overdominance, overdominance, and epistasis (Shull, 1908; Powers, 1944; Birchler *et al*., 2003, 2006). Recent transcriptome analyses revealed that nonadditive gene expression was prevalent (Guo *et al*., 2003; Auger *et al*., 2005; Uzarowska *et al*., 2007; Ma *et al*., 2016), while additive expression was predominant in other studies (Stupar and Springer, 2006; Swanson-Wagner *et*

al., 2006; Meyer *et al*., 2007; Paschold *et al*., 2012). This discrepancy might be explained by differences in the genotypes, plant tissues, experimental designs, and statistics used in these studies (Hochholdinger and Hoecker, 2007).

In addition to heterosis, increasing the planting density also contributed to improving the maize grain yield (GY). Planting density and GY per unit area have been increasingly used as maize productivity metrics since 1930 (Mansfield and Mumm, 2014). Several studies reported that microarray analysis disclosed transcriptional changes in maize in response to planting density stress. For example, in 12-day-old maize seedlings from five inbred lines, St. Pierre *et al.* (2011) found 35 differentially expressed genes (DEGs) in response to plant density stress. In maize hybrid seedlings at the fourth-leaf stage, Guo *et al.* (2004) found two genes presenting allelic differences in transcript accumulation under high and low planting densities. In immature ear tissues of hybrid maize, the percentage of genes exhibiting mid-parent expression decreased with increasing planting density. Nevertheless, the proportion of additively expressed alleles was positively correlated with hybrid yield and heterosis (Guo *et al.,* 2006). Therefore, differential allele regulation may play important roles in hybrid yield and heterosis (Guo *et al.,* 2004, 2006).

In developed countries, increases in GY have been attributed to genetic improvement, advanced crop management practices, and high tolerance to biotic and abiotic stresses, including high planting densities (Tollenaar *et al*., 1997; Tollenaar and Wu, 1999). As planting densities increase, the GY per plot increases whereas the contribution of heterosis to GY decreases. Therefore, a balance between heterosis and planting density is necessary for GY improvement. In this study, four inbred Chinese maize lines (Zheng58, Chang7-2, Ye478, and Huangzaosi) and their cross combinations were subjected to RNA sequencing analysis. These four inbred lines belong to popular heterotic groups used in China that have played important roles in the

genetic improvement of Chinese maize. Zheng58 and Ye478 belong to the Reid heterotic group and were used as maternal parents. Zheng58 was developed from Ye478 and an unknown inbred line. Chang7-2 and Huangzaosi are important inbred lines of the SPT heterotic group that were used as paternal parents. Chang7-2 was developed from Huangzaosi and two other inbred lines. Zheng58 and Chang7-2 are the parents of the commercial hybrid Zhengdan958 which is currently the most widely grown variety in China. Ye478 and Chang7-2 are the parents of the commercial hybrid Anyu5 which was extensively cultivated in the 1990s. This study used transcriptome analysis for examining maternal and paternal effects on heterosis, identifying heterosis-related genes in maize, and elucidating the molecular mechanism of heterosis in response to planting density stress.

Materials and methods

Plant materials and field experiment design

Zheng58, Chang7-2, Ye478, and Huangzaosi, and their four hybrids, Zheng58/Chang7-2 (Zhengdan958), Ye478/Chang7-2 (Anyu5), Zheng58/Huangzaosi, and Ye478/Huangzaosi, were grown at 45,000 plants/ha and at 67,500 plants/ha in a randomized block design (four rows per block) with three replicates. The trials took place in the summer of 2014 and 2015 on the experimental fields of the Henan Academy of Agricultural Sciences, China. For convenience, Zheng58, Ye478, Chang7-2, and Huangzaosi are designated as AA, BB, CC, and DD, respectively; AC, BC, AD, and BD are used to represent Zhengdan958, Anyu5, Zheng58/Huangzaosi, and Ye478/Huangzaosi, respectively.

Agronomic trait measurements and statistical analyses

At maturity, all maize ears within each block were harvested. Twenty ears per plot, replicate, and material were randomly selected from the 45,000 plants/ha. At the 67,500 plants/ha density, 30 ears per plot, replicate, and material were randomly selected. Grain moisture content was measured in two replicates using a PM-8188-A moisture meter (Kett Electric Laboratory, Tokyo, Japan). After threshing, the GY was determined and then adjusted to 14% moisture. The GY per plant (GYP) was calculated by dividing GY by 20 or 30, according to planting density. Kernel row number (KRN), kernel number per row (KNR), kernel number per ear (KNE), ear length (EL), ear diameter (ED), plant height (PH), and ear height (EH) were measured in five maize ears. Four parent lines are inbred with a heterosis index (HI) of one. The HI of the four hybrids was calculated for all these traits according to:

$$\mathrm{HI} = \frac{F_1}{(P_1+P_2)/2} \times 100\% \qquad [1]$$

The difference in GY and other traits between the 45,000 plants/ha and 67,500 plants/ha planting densities were evaluated using Student's *t*-tests. Analysis of variance (ANOVA) of GY and GY HI was conducted considering varieties (eight materials), three replications, two planting densities, and variety × density interactions. Duncan's multiple comparison test was used to evaluate significant differences between means when a significant effect was found by ANOVA. Student's *t*-tests, ANOVA and Duncan's multiple comparison tests were performed with the t.test, aov and duncan.test functions in R (https://www.r-project.org/), respectively.

RNA extraction and sequencing

Photosynthetic efficiency and ear leaf area are significantly and positively correlated with GY in maize (Agrama *et al*., 1999). Therefore, 15 d after pollination, 25 ear leaves were collected from plants in the two middle rows of each replicate within each material and planting density. The three plants in the front and back of each row were excluded. Total RNA was extracted from the leaves using TRIzol® Reagent (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions. Oligo (dT) magnetic beads (Illumina, San Diego, CA, USA) were used to obtain purified mRNA. The cDNA libraries were constructed using a TruSeq Stranded mRNA LT Sample Prep Kit (Illumina). Transcriptome sequencing was performed on the Illumina HiSeq X Ten platform according to the manufacturer's protocol.

Identification and functional analysis of differentially expressed genes

After filtering, clean sequence reads were aligned to the B73 reference genome (ftp://ftp.ensemblgenomes.org/pub/release-29/plants/fasta/zea_mays/dna/Zea_mays.AGPv3.29.dna.toplevel.fa.gz) using TopHat v.2.0.10 (http://ccb.jhu.edu/software/tophat/index.shtml). Only uniquely matched reads

were selected and used. Gene expression was calculated and normalized to fragments per kb per million reads (FPKM) using Cufflinks v.2.1.1 (http://cole-trapnell-lab.github.io/cufflinks/). The expression cutoff was defined as FPKM > 0.5. To graphically interpret the relatedness among the eight materials, cluster analysis was performed using the heatmap.2 function in gplots (http://cran.r-project.org/web/packages/gplots/index.html) with default settings. The average FPKM value of three replicates was used for the clustering.

Zhengdan958 (AC) and Anyu5 (BC) or Zheng58/Huangzaosi (AD) and Ye478/Huangzaosi (BD) have the same paternal lines but different maternal lines. Therefore, differences in the transcription profiling between Zhengdan958 and Anyu5 or between Zheng58/Huangzaosi and Ye478/Huangzaosi might be explained by the maternal lines Zheng58 (AA) and Ye478 (BB). Genes presenting such profiles have the same paternal parents but different maternal lines and are therefore called maternal-effect genes. We found DEGs among AC, BC and CC transcriptomes, and among AD, BD and DD transcriptomes, and common maternal-effect DEGs between them. Genes with the same maternal parents but different paternal lines are called paternal-effect genes. Paternal-effect DEGs were found by comparing the FPKM values of DEGs in AC, BC, AD, and BD with that in their maternal lines (AA and BB). In addition, DEGs between 45,000 plants/ha (CK) and 67,500 plants/ha were identified for the eight materials.

Read counts were obtained by HTSeq (http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html). Negative binomial distribution of DESeq (http://www.bioconductor.org/packages/release/bioc/html/DESeq.html) was used to test significant differences of read counts. Base mean (mean of normalized counts) was used to represent the expression value for differential

expression analysis. The significance threshold was $P < 0.05$. The Log2 fold change threshold was not used in this analysis. Both additive and nonadditive DEG expression patterns were found in the four hybrids. Additive expression means that the hybrid expression level is equal to the mid-parent expression level. Nonadditive expression occurs when the hybrid expression level deviates from the mid-parent expression value. The Audic-Claverie statistic was used to detect significant levels of additive and nonadditive gene expression (Audic and Claverie, 1977).

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) annotations were performed for DEGs. The hypergeometric distribution test was used in GO enrichment analysis and in KEGG enrichment pathway analysis. The GO terms within the category “biological process” (BP) were obtained from http://geneontology.org/. A GO term was considered significantly enriched if the Benjamini and Hochberg (BH) false discovery rate (FDR) cutoff was < 0.05. The KEGG pathways were assigned using the KEGG software package (http://www.kegg.jp/) and considered significant if $P < 0.05$.

Identification of gene co-expression modules

Gene co-expression modules were assigned using the weighted gene co-expression network analysis (WGCNA) protocol (Zhang and Horvath, 2005; Langfelder and Horvath, 2007, 2008), based on the Log2 of the FPKM data of all expressed genes. The Dynamic Tree Cut algorithm was used to cut the hierarchal clustering. The minimum module size was 50 genes and the soft threshold power β was set to 11. Significant module-trait associations were identified by correlating the module eigengenes with the GY HI. The eigengenes represented the gene expression pattern within a module (Langfelder and Horvath, 2007). Cytoscape v.3.0.2 (Shannon *et al.,* 2003) was used to display the co-expression network. A module was considered significant if the absolute gene significance value was substantially > 0.9 and

the absolute correlation coefficient value was > 0.5 and significant ($P < 0.05$). The functions of co-expressed genes were also classified by GO and KEGG enrichment analyses, and their significance tests and threshold were the same as those used for DEG analysis. The FPKM values of genes found by WGCNA were analyzed as the phenotypic data in the ANOVA framework. HI of the genes of the four hybrids was calculated according to the formula described previously [1]. Parental HI of transcriptome was one as they are inbred lines, which was the same as that of GY HI. ANOVA of genes also contained varieties (eight materials), three replications, two planting densities, and variety × density interactions.

Quantitative real-time PCR validation

Four hybrids were grown at Hainan, China, in December 2015, and selfed to obtain F_2 progeny. In the summer of 2016, the four F_2 populations derived from the hybrids were grown in six rows at 90,000 plants/ha on an experimental field at the Henan Academy of Agricultural Sciences. After open pollination, leaves from plants in the four middle rows of each F_2 population were collected. The three plants in the front and back of each row were excluded. The ears corresponding to the excised leaves were harvested, and GYP for each F_2 population was calculated. As it followed a normal distribution, the best and worst six ears were chosen from each F_2 population based on GYP. The four inbred parents were grown at 67,500 plants/ha and their leaves were collected for gene expression validation.

Gene expression quantification was obtained from reverse transcription (RT) PCR. For the RT, 0.5 μg RNA and 2 μL 4× gDNA wiper mix were combined, and nuclease-free H_2O was added to bring the total volume to 8 μL. The reactions were run in a GeneAmp® PCR System 9700 (Applied Biosystems, Foster City, CA, USA) for 2 min at 42 ºC. After adding 2 μL 5× HiScript II Q RT SuperMix IIa (Vazyme, Nanjing, China) to the reaction mixture,

amplifications were run in a GeneAmp® PCR System 9700 (Applied Biosystems) for 10 min at 25 ºC, 30 min at 50 ºC, and 5 min at 85 ºC. The 10 µL RT reaction mix was then diluted 10× in nuclease-free water and stored at -20 ºC. Quantitative real-time PCR (qRT-PCR) was performed in the LightCycler® 480 II Real-time PCR Instrument (F. Hoffmann-La Roche AG, Basel, Switzerland) using 10 µL PCR reaction mixtures including 1 µL cDNA, 5 µL 2× QuantiFast® SYBR® Green PCR Master Mix (Qiagen, Hilden, Germany), 0.2 µL forward primer, 0.2 µL reverse primer, and 3.6 µL nuclease-free water. Reactions were incubated in a 384-well optical plate (F. Hoffmann-La Roche AG) at 95 ºC for 5 min, followed by 40 cycles of 95 ºC for 10 s and 60 ºC for 30 s. Each sample was run in triplicate. At the end of the PCR cycles, a melting curve analysis was performed to validate the expected quantity of the PCR product. The primers for qRT-PCR were designed using mRNA sequences from the National Center for Biotechnology Information (NCBI) database and synthesized by Generay Biotech (Beijing, China). The ratio of target genes to reference gene (Actin) was used to describe the relative target gene expression.

Ten of the genes identified by WGCNA were validated. The FPKM values of genes expressed by plants grown at the 45,000 plants/ha and 67,500 plants/ha densities were correlated with the qRT-PCR expression values (67,500 plants/ha) obtained for the four parents. A regression analysis was performed on the GYP of each of the 12 progenies from the four F_2 populations on the expression values determined by qRT-PCR. The lm function in R was used to perform this regression analysis. The additive and nonadditive target gene patterns were classified for each F_2 population according to the expression found in the 12 F_2 individuals relative to that found in their parents.

Results

Phenotypic performance of maize hybrids and their parents at 45,000 plants/ha and 67,500 plants/ha

Both GY and GYP were significantly and positively correlated with all other traits except KRN (r = 0.7969–0.9676, $P < 0.05$) at both planting densities (Tables S1 and S2). The GY of all materials was significantly higher at 67,500 plants/ha than at 45,000 plants/ha (Supplementary Fig. S1A). There was no significant difference between high and low planting densities for all other traits (Supplementary Fig. S1A). The GY or their heterosis indices for the eight materials, two planting densities, and variety × density interactions were significant ($P < 0.01$, Tables S3 and S4). The GY and their heterosis indices were significantly higher in the four hybrids than in their four parents. Differences in GY between Zhengdan958 (Zheng58/Chang7-2) and Anyu5 (Ye478/Chang7-2), between Anyu5 and Zheng58/Huangzaosi, and between Zheng58/Huangzaosi and Ye478/Huangzaosi were significant at 67,500 plants/ha (Fig. 1A). However, none of the four hybrids significantly differed in terms of GY at 45,000 plants/ha (Fig. 1A). The four hybrids did not significantly differ in terms of GY HI at the high planting density (Fig. 1B), but they significantly differed at the low planting density (Fig. 1B). When the planting density increased from 45,000 plants/ha to 67,500 plants/ha, the GY per plot of all four hybrids increased but their GY HI decreased. Therefore, the contribution of heterosis to GY decreased with increasing planting density.

Global gene expression in maize hybrids and their parents at 45,000 plants/ha and 67,500 plants/ha

To identify heterosis-related genes affected by planting density, 48 maize RNA libraries were constructed and sequenced. Raw sequence data are available at the Sequence Read Archive of NCBI (accession nº SRP136913). Genes with FPKM < 0.5 were filtered out and excluded, resulting in 15,496 expressed genes (Supplementary Table S5). Gene expression correlations were high

among the three biological replicates. The average correlation coefficient range was 0.98–0.99 for both planting densities (Fig. 2A). The maternal lines, Zheng58 and Ye478, and the paternal lines, Chang7-2 and Huangzaosi, were clustered in one group each (Fig. 2B and 2C). The two maternal lines formed a cluster with all four F_1 hybrids whereas the paternal lines did not cluster with any hybrid (Fig. 2B and 2C), indicating that the four hybrids had a closer relationship with their maternal parents than with their paternal parents.

DEGs between both planting densities were identified for parental lines and their hybrids. The 45,000 plants/ha planting density was considered the control condition. For Zheng58 and Ye478, the number of downregulated DEGs between the two planting densities was higher than that of the upregulated DEGs; 77.42% of 1,045 genes and 64.57% of 525 genes were downregulated in Zheng58 and Ye478, respectively (Supplementary Fig. S2A). For Huangzaosi and Chang7-2, the number of upregulated DEGs between the two planting densities was higher than that of the downregulated DEGs; 69.32% of 502 genes and 80.34% of 880 genes were upregulated in Huangzaosi and Chang7-2, respectively (Supplementary Fig. S2A). Therefore, as planting density increased, most maternal genes in the maize ear leaves were repressed while most paternal genes were induced.

Maternal-effect and paternal-effect DEGs have significant functions in response to high planting density stress

To analyze maternal and paternal effects on heterosis in response to planting density, the DEGs in AC, BC, and CC transcriptomes, AD, BD, and DD transcriptomes, AC, AD, and AA transcriptomes, and BC, BD, and BB transcriptomes were examined. We found common DEGs (maternal-effect) between AC, BC and CC transcriptomes and AD, BD and DD transcriptomes. Common paternal-effect DEGs were found between AC, AD and AA transcriptomes, and BC, BD and BB transcriptomes.

At 45,000 plants/ha, there were nine DEGs each in the AC>BC>CC and AD>BD>DD transcriptomes, and in the AC<BC<CC and AD<BD<DD transcriptomes (Supplementary Fig. S2B). All maternal-effect DEGs participated in protein phosphorylation, carbohydrate biosynthesis regulation, lignin biosynthesis, unidimensional cell growth, transcription, and translation. For the paternal-effect group, 16 DEGs each were found in the AC>AD>AA and BC>BD>BB transcriptomes, and in the AC<AD<AA and BC<BD<BB transcriptomes. Paternal-effect DEGs were mostly involved in protein phosphorylation, transcription, and transport.

The number of detected DEGs increased with planting density. This correlation was consistent with the phenotypic variance of GY. Among the 400 DEGs identified in the transcriptomes of AC, AD, BC, BD, CC, and DD (Fig. 2D), 252 DEGs were found in the AC>BC>CC and BC>BD>DD transcriptomes and 148 DEGs were found in the AC<BC<CC and BC<BD<DD transcriptomes. In the transcriptomes of AC, AD, BC, BD, AA, and BB, 318 DEGs were identified (Fig. 2D). Among them, there were 208 DEGs from the AC>AD>AA and BC>BD>BB transcriptomes, and 110 DEGs from the AC<AD<AA and BC<BD<BB transcriptomes (Fig. 2D). Regarding GO annotations in these DEGs, there were more GO terms and DEGs associated with carbohydrate biosynthesis and metabolism and photosynthesis in the AC>BC>CC transcriptomes (34 DEGs) and AD>BD>DD transcriptomes (28 DEGs) than in the AC>AD>AA transcriptomes (seven DEGs) and BC>BD>BB transcriptome (seven DEGs) (Supplementary Fig. 3A). Therefore, maternal-effect genes mainly participated in the synthesis of energy storage materials. The total number of GO terms and DEGs associated with biotic and abiotic stress responses and plant hormone production was higher in AC>AD>AA transcriptomes (24 DEGs) and BC>BD>BB transcriptomes (22 DEGs) than in AC>BC>CC transcriptomes (25 DEGs) and AD>BD>DD transcriptomes (14

DEGs) (Supplementary Figs. S3B and S3C). Therefore, paternal-effect genes were primarily involved in adaptation to environmental stress.

At a density of 67,500 plants/ha, 34 paternal-effect genes common to transcriptomes AC>AD>AA and BC>BD>BB, and four paternal-effect genes common to transcriptomes AC<AD<AA and BC<BD<BB were found (Fig. 2D). There were 28 maternal-effect genes overlapping transcriptomes AC>BC>CC and AD>BD>DD, and four maternal-effect genes overlapping transcriptomes AC<BC<CC and AD<BD<DD (Fig. 2D). These maternal- and paternal-effect genes were classified into additive and nonadditive depending on their relative expression levels in hybrids and their parents. Functional additive and nonadditive gene annotations were also conducted. Most maternal-effect genes were additively expressed in the four hybrid lines (89.29–96.43% of the 28 genes; Fig. 3A). The majority was involved in photosynthesis and in carbohydrate biosynthesis and metabolism (Fig. 3B). Nearly all paternal-effect genes were additively expressed in the four hybrids lines (97.06–100%; Fig. 3A), and they played important roles in abiotic and biotic stress responses (Fig. 3C). In general, both maternal- and paternal-effect DEGs had important functions in planting density stress response.

High planting density caused an increase in heterosis-related genes improving fitness rather than in those producing energy

Network analysis of transcriptome data have been used for omics studies in many plant species, and have been successful in Arabidopsis and tomato (Higashi and Saito, 2013). A WGCNA was constructed to identify networks of co-expressed genes correlated with GY HI. The WGCNA was performed on both maternal and paternal genes to identify them and their functions in response to planting density stress. The four hybrids and their maternal parents were used to find maternal genes correlated with GY HI; the four

hybrids and their paternal parents were used to construct the paternal WGCNA.

All expressed genes (15,496) were used for WGCNA and assigned to 8–17 modules for maternal and paternal WGCNA at the two planting densities. The modules were represented by different colors. They were selected according to gene significance and the module-trait relationship. The eigengene value of each module was correlated with GY HI. In the maternal WGCNA, three modules were very highly correlated with GY HI and all other traits except KNR. The correlation coefficients ranged from -0.87 to 0.80 at 45,000 plants/ha (Fig. 4A and Supplementary Fig. S4A-C). At 67,500 plants/ha, four significant modules were identified (r ranged from -0.93 to 0.70; $P < 0.05$) (Fig. 4A and Supplementary Fig. S4D-G). For the paternal WGCNA, modules Greenyellow (r = 0.77; P = 2e-04), Magenta (r = -0.72; P = 7e-04), Turquoise (r = -0.91; P = 1e-07), and Pink (r = -0.52; P = 0.03) were found at the lower planting density (Fig. 4C and Supplementary Fig. S5A-D). At the higher planting density, three modules were found with module-trait correlation coefficients ranging from -0.93 to 0.70 ($P < 0.01$) (Fig. 4C and Supplementary Fig. S5E-G). Their eigengenes were significantly correlated ($P < 0.01$) with GY HI. Therefore, despite the fact that some modules were negatively correlated, these modules were highly relevant to maize heterosis.

The GO annotations of heterosis-related genes in these modules were analyzed. The number of genes regulating carbohydrate biosynthesis and metabolism increased with planting density for the maternal WGCNA (60 and 72 genes at 45,000 and 67,500 plants/ha, respectively) (Fig. 4B). In contrast, the number of genes associated with photosynthesis decreased with increasing planting density (128 and 74 genes at 45,000 and 67,500 plants/ha, respectively) (Fig. 4B). For the paternal WGCNA, the number of genes associated with energy production varied with planting density differently from

those in the maternal WGCNA (Fig. 4D). The total number of maternal/paternal genes associated with carbohydrate biosynthesis and metabolism decreased with increasing planting density (333 and 312 genes at 45,000 and 67,500 plants/ha, respectively). The numbers of maternal or paternal genes associated with stress response and phytohormone production increased with planting density (31 maternal and 52 paternal genes at 45,000 plants/ha, and 42 maternal and 79 paternal genes at 67,500 plants/ha) (Fig. 4B and 4D). Overall, the heterosis-related genes associated with energy formation and metabolism were mainly expressed at low planting density. However, when planting density increased, the numbers of genes associated with abiotic and biotic stress response and plant hormone biosynthesis increased and participated in fitness.

Heterosis-related maternal and paternal genes have different roles in the photosystem in response to planting density stress

Photosynthesis plays a crucial role in plant heterosis. Photosystems are composed of a peripheral antenna and a reaction center. The maternal and paternal genes have different responses to planting density stress in the maize photosystem. For the maternal WGCNA, the co-expression network showed that the photosynthesis-antenna protein pathway (path:zma00196) was significantly enriched in the Midnightblue (P = 0.002301) and Turquoise (P = 0.01404) modules at 45,000 plants/ha and in the Blue module (P = 0.003643) at 67,500 plants/ha (Supplementary Fig. S6). The photosystem antenna proteins Lhca1, Lhca3, Lhcb1, Lhcb3, and Lhcb4 were associated with the GY HI at both planting densities (Supplementary Fig. S7). Two rarely expressed photosystem antenna genes, *Lhcb6* and *Lhcb7*, were only found at low planting density. Twelve genes participated in the photosystem I (PSI) complexes (PsaD, PsaG, PsaH, and PsaK), in the PSII complexes (PsbO and PsbY), and in the photosynthetic electron transport complexes (PetE, PetF, and PetH) (Supplementary Fig. S8A). These were localized in the Blue modules at 67,500 plants/ha. None of these was found at low planting density.

Therefore, under high planting density stress, maternal genes were mainly enriched in the photosynthesis reaction center but not in the peripheral antenna system.

In the paternal WGCNA, Lhca2, Lhcb1, and Lhcb6 were identified at low planting density (Supplementary Fig. S9). When planting density increased, Lhca1, Lhca2, Lhca3, and Lhca4, and the major PSII antenna protein Lhcb3, positively correlated with maize heterosis (Supplementary Fig. S9). Regarding the photosynthetic pathway, 24 genes for the reaction center were significantly associated with the Turquoise module at 45,000 plants/ha (P = 0.02) and 20 reaction center genes were significantly associated with the Turquoise module at 67,500 plants/ha (P = 0.04274) (Supplementary Fig. S8B and S8C). Except for GRMZM2G059083, all genes identified at the high planting density were also found at the low planting density. Genes PsbY, PsaC, and PetE were found only at low planting density. Therefore, under high planting density stress, paternal heterosis-related genes were mainly enriched in the peripheral antenna system but not in the reaction center.

Heterosis-related genes coding for ubiquitin-mediated proteolysis are mainly enriched at a high planting density

Maternal WGCNA also revealed that heterosis-related genes were enriched in ubiquitin-mediated proteolysis (path:zma04120) in the Midnightblue module (P = 0.008579) at 45,000 plants/ha and in the Turquoise module (P = 0.01808) at 67,500 plants/ha (Supplementary Figs. S6 and S10). At 45,000 plants/ha, the ubiquitin-conjugating enzyme E2-17 kDa, E2-25, and the E3 enzymes COP1 and APC10 were involved in ubiquitin-mediated proteolysis. When the planting density was increased to 67,500 plants/ha, 20 genes were significantly enriched in this pathway. These included the E3 enzymes COP1, UPL6, and RCHY1, an E4-type ligase (E4 SUMO-protein ligase PIAL2), eight E2 enzymes (E2-22, E2-J2, E2-21 kDa 1, E2-2, E2-17 kDa-like, E2-25 and E2-38, and NEDD8-conjugating enzyme Ubc12-like), cullin2, cullin-3A, ubiquitin carrier

protein 7, FZR1, pre-mRNA-splicing factor 19, STUB1, and light-mediated development protein DET1. Therefore, high planting density significantly increased the number of genes associated with the ubiquitination pathway, including E2 and E3 enzymes.

Validation of the heterosis-related genes found by WGCNA

The qRT-PCR validated 10 genes participating in photosynthetic pathways or processes, carbohydrate-related processes, light response processes, and circadian rhythms. All these genes were found by WGCNA. GRMZM2G038519, GRMZM2G162233, and GRMZM2G039455 were also identified by DEG analysis at high planting density.

For the four parent lines, the FPKM values determined by transcriptome analysis at 45,000 plants/ha and 67,500 plants/ha were correlated with the expression values determined by qRT-PCR at 67,500 plants/ha. The correlation coefficient range was 0.6900–0.9469 for GRMZM2G162233, GRMZM2G099334, GRMZM2G162748, GRMZM2G134130, and AC190623.3_FG001 at 45,000 plants/ha (Supplementary Fig. S11). For GRMZM2G038519, GRMZM2G162233, GRMZM2G115914, GRMZM2G099334, GRMZM2G377855, GRMZM2G162748, and GRMZM2G056424, correlation coefficient range was 0.4348–0.9974 at 67,500 plants/ha (Supplementary Fig. S11).

For each F_2 population, a regression analysis was performed for the GYP of the 12 progenies on the expression values determined by qRT-PCR. The linear regressions were significant and showed that the expression values of the 10 genes increased with GYP for all four F_2 populations. The R^2 ranged from 0.3697 to 0.8646, from 0.2636 to 0.6256, from 0.3728 to 0.7333, and from 0.4098 to 0.8933 in the F_2 progenies derived from Zhengdan958, Anyu5, Zheng58/Huangzaosi, and Ye478/Huangzaosi, respectively (Supplementary Fig. S11-S15). Transcriptome analysis showed that most of the 10 genes expressed additively in the four hybrid lines. The qRT-PCR validation in the F_2

population showed the same additive gene expression pattern indicated by transcriptome analysis (Supplementary Fig. S16). The validated density was 90,000 plants/ha for each F_2 progeny. When the density was increased to 90,000 plants/ha, the genes found by WGCNA were still detected and their expression values positively correlated with GYP. Therefore, these genes might play important roles in maize heterosis.

Discussion

Heterosis and planting density contribute to GY improvement. In this study, we investigated maternal and paternal effects on heterosis as well as the trade-off between heterosis and planting density for GY improvement. We tested four popular Chinese maize inbred lines, their hybrids, and the corresponding heterotic pattern. Because Reid × SPT is the most popular heterotic pattern in China, the present study contributes to a better understanding of the most popular heterotic pattern used in maize GY improvement.

Maternal- and paternal-effect DEGs had significant functions in response to planting density stress: maternal-effect DEGs participate in carbohydrate biosynthesis and metabolism and in photosynthesis, and paternal-effect DEGs play roles in abiotic/biotic stress responses and phytohormone production at high planting densities. Li *et al*. (2014) showed that, in hexaploid wheat, paternal-effect DEGs were involved in stress response whereas maternal-effect DEGs participated in plant development. In maize breeding, the maternal parent is selected for high yield. Several studies showed that the maternal parent controls seed size and weight (Kesavan *et al*., 2013; Xin *et al*., 2013; Li and Li, 2015; Zhang *et al*., 2016; Singh *et al*., 2017). In addition, the maternal parent contributes to the offspring seed phenotype. It provides photosynthate and nutrients to support progeny development, coordinates the timing of development, and imprints the maternal gametes (Zhang *et al*., 2016). Similarities in expression patterns between reciprocal crosses and their maternal parents provide valuable insights into the regulatory role of the maternal transcriptome in terms of phenotypic seed size and weight variations (Zhang *et al*., 2016). Recent studies have revealed that the ubiquitin pathway is involved in maternal seed size control in Arabidopsis and rice (Li and Li, 2015). In our study, the ubiquitin-mediated proteolysis comprised 24 genes that were significantly enriched in the maternal WGCNA. Unlike maternal

parents, which can affect seed size in several ways, paternal parents can only influence the seed set by affecting the endosperm and, possibly, the embryo (Jong and Scott, 2007). Paternal parents only contribute to 10% of the variation in seed size (House *et al.,* 2010). In maize breeding, paternal parents are selected for high environmental fitness with respect to abiotic/biotic stress resistance. Therefore, differences in maternal and paternal effects may be the result of different parental genotypes and selection objectives. Moreover, pollen may influence the maternal and paternal effects. However, in maize and soybean, significant reciprocal differences were caused by substantial maternal effects on seed weight and size (Zhang *et al*., 2016; Singh *et al*., 2017). Quantitative changes in gene expression may be the result of *cis* or *trans* variations in gene regulation (Wittkopp *et al*., 2004; Stupar and Springer, 2006). Thus, differences in the gene classes regulated in hybrids might be explained by the various *cis* or *trans* regulatory motifs present in the maternal and paternal lines.

Most maternal- and paternal-effect DEGs were additive. In addition, the gene expression pattern found by WGCNA was predominated by additive expression in the four hybrid lines. It was found that 93.74–97.48% and 90.46–97.41% of the genes were additively expressed in the maternal and paternal groups, respectively. The differential expression profiles in younger vegetative and reproductive stages are mostly nonadditive whereas that in mature vegetative stages are mostly additive (Baranwal *et al*., 2012). The predominance of additive expression in the mature ear leaves of the four hybrid lines corroborated these findings. The significant correlation between the gene modules and HI suggests that additive expression is important for maize hybrids and heterosis.

Under high planting density stress, heterosis-related genes increased to improve or optimize fitness in response to abiotic/biotic stress and to enhance

phytohormone production and ubiquitin-mediated proteolysis. In contrast, the genes regulating photosynthesis and carbohydrate biosynthesis and metabolism decreased. Planting density is recognized as a major factor determining the degree of interplant competition for resources like incident radiation, soil moisture, and nutrients. Resource competition increases almost linearly with stand density (Fasoula and Tollenaar, 2005). In the field, the effect of high plant density is similar to that of shading (Gerakis and Papakosta-Tasopoulou, 1980). Reed *et al*., (1988) reported that photosynthesis decreased and kernel abortion increased when plants were shaded during flowering. The apparent photosynthetic rate in ear leaves significantly decreased with increasing plant density and shading (Singh and Jenner, 1984). Our data showed that high planting density stress might induce the genes regulating environmental fitness at the cost of repressing the genes encoding photosynthesis and carbohydrate-related processes.

An important objective of the present study was to identify heterosis-related genes responding to planting density stress in maize. We identified 70 overlapped DEGs by comparing the transcriptomes of AC, AD, BC, and BD with AA and BB, and those of AC, AD, BC, and BD with CC and DD, all of which found at high planting density. The maternal genes GRMZM2G038519 and GRMZM2G162233 were associated with photosynthesis and carbohydrate biosynthesis and metabolism. Six paternal genes were associated with abiotic/biotic stress response and plant hormone biosynthesis. Moreover, WGCNA indicated that the maternal genes and two of the paternal genes (GRMZM2G436986 and GRMZM2G039455) were significantly correlated with the GY HI. Therefore, these genes might play roles in maize heterosis at high planting density.

GRMZM2G038519, GRMZM2G162233, and GRMZM2G436986 participate mainly in photosynthesis (Fig. 3B and 3C). GRMZM2G038519, or

Lhca1, is involved in the antenna protein pathway, light harvesting, and photosynthesis (Fig. 3B). GRMZM2G162233 and GRMZM2G436986 are significantly involved in photorespiration, photosynthetic electron transport, maltose metabolism, and starch biosynthesis (Fig. 3B and 3C). The ortholog of GRMZM2G162233 is NAD(P)H dehydrogenase (NDH)-dependent flow 6, which is essential for NDH activity and encodes a thylakoid membrane protein specific to terrestrial plants (Ishikawa *et al*., 2008). NDH mediates the cyclic and chlororespiratory electron transport chains in PSI by interacting with multiple copies of the PSI complex via Lhca5 and Lhca6 (Peng *et al*., 2009, 2011). The ortholog of GRMZM2G436986 is the photosynthetic NDH subunit of lumenal location 3, which, like the PsbQ-like protein 1, is a novel NDH subunit that plays an important role in the chloroplast NDH complex (Suorsa *et al*., 2010; Yabuta *et al*., 2010). GRMZM2G039455 participates in auxin metabolism, embryo development, seed dormancy, innate immune response, gene silencing, and RNA interference (Fig. 3C). This gene was named $ZmAGO_{1a}$ in maize (Xu *et al.,* 2016). Previous studies revealed that argonaute proteins are key players in gene silencing mediated by small RNAs (Hutvagner and Simard, 2008; Vaucheret, 2008; Kuan *et al.,* 2016; Xu et al., 2016). The modulation of miRNA levels regulates plant responses to abiotic stressors like drought, cold, salinity, and nutrient deficiency, as well as biotic stress (Baldrich and San, 2016; Kuan *et al.,* 2016; Xu *et al.,* 2016).

For GRMZM2G038519, GRMZM2G162233, and GRMZM2G039455, correlation of qRT-PCR expression values with GYP showed positive linear relationships in all four F_2 populations, and R^2 ranged from 0.4601 to 0.8933 ($P < 0.05$), again showing they might play roles in maize heterosis. Because F_2 is a segregating population with residual heterosis, each F_2 progeny has its own distinct genotype and phenotype. Linear correlation of 10 heterosis-related genes in the four F_2 population suggested that F_2 could be used to validate heterosis.

In addition, the 1,461 genes found by WGCNA were analyzed as phenotypic data in the ANOVA framework. All 1,461 genes were significantly enriched in carbohydrate biosynthesis and metabolism, photosynthesis, stress response and phytohormone production processes, photosynthesis-antenna protein, photosynthesis pathway, and ubiquitin-mediated proteolysis pathway. The effect of variety × density interactions of 52 and 93 genes was significant for FPKM and transcriptome HI (Supplementary Table S6 and S7). Although the percentage of genes with significant variety × density interaction effect is low, the results obtained here support that heterosis varied with planting densities.

Heterosis is a complex quantitative trait. The candidate heterosis-related genes found here will be used to develop DNA microarrays (chips) and validated by Reid and SPT heterotic groups, aiming to predict heterosis by the presence of these candidate genes.

Supplementary data

Table S1 Correlation coefficient among GY/GYP and seven other traits at 45,000 plants/ha.

Table S2 Correlation coefficient among GY/GYP and seven other traits at 67,500 plants/ha.

Table S3 ANOVA of GY.

Table S4 ANOVA of GY HI.

Table S5 FPKM values of expressed genes in DEG analysis and WGCNA.

Table S6 ANOVA of FPKM values of 52 genes.

Table S7 ANOVA of transcriptome HI in 93 genes.

Fig. S1. Student's *t*-test analysis of GY and other traits between 45,000 plants/ha and 67,500 plants/ha.

Fig. S2. Number of DEGs among the planting densities and the transcriptomes of hybrids versus their parents at 45,000 plants/ha.

Fig. S3. Number of DEGs significantly involved in important biological processes for transcriptomes of AC>AD>AA, BC>BD>BB, AC>BC>CC, and AD>BD>DD at 67,500 plants/ha.

Fig. S4. Scatterplots of gene significance for GY HI versus module membership in significant modules for maternal WGCNA at 45,000 plants/ha and 67,500 plants/ha.

Fig. S5. Scatterplots of gene significance for GY HI versus module membership in significant modules for paternal WGCNA at 45,000 plants/ha and 67,500 plants/ha.

Fig. S6. Significant KEGG pathways and number of genes used for maternal and paternal WGCNA at 45,000 plants/ha and 67,500 plants/ha.

Fig. S7. Photosynthesis-antenna protein pathway in maternal significant modules at 45,000 plants/ha and 67,500 plants/ha.

Fig. S8. Photosynthesis pathway in maternal and paternal significant modules at 45,000 plants/ha and 67,500 plants/ha.

Fig. S9. Photosynthesis-antenna protein pathway in paternal significant modules at 45,000 plants/ha and 67,500 plants/ha.

Fig. S10. Ubiquitin-mediated proteolysis pathway in maternal significant modules at 45,000 plants/ha and 67,500 plants/ha.

Fig. S11. Line charts of expression values by transcriptome at 45,000 plants/ha and 67,500 plants/ha, and by qRT-PCR at 67,500 plants/ha of 10 genes in four parental lines.

Fig. S12. Linear regression of GYP in F_2 population derived from Zhengdan958 on expression values using qRT-PCR in 10 genes.

Fig. S13. Linear regression of GYP in F_2 population derived from Anyu5 on expression values using qRT-PCR in 10 genes.

Fig. S14. Linear regression of GYP in F_2 population derived from Zheng58/Huangzaosi on expression values using qRT-PCR in 10 genes.

Fig. S15. Linear regression of GYP in F_2 population derived from Ye478/Huangzaosi on expression values using qRT-PCR in 10 genes.

Fig. S16. Gene expression pattern of 10 genes using qRT-PCR.

Acknowledgements

This work was funded by the National Key Research and Development Program of China (2016YFD100103), the National Natural Science Foundation of China (31471566), and the Major Science and Technology Projects in Henan Province, China (161100110500). We would like to thank Oebiotech company (Shanghai, China) for RNA sequencing and related data analyses.