Genome-Wide Transcriptional Analysis of Yield and Heterosis-Associated Genes in Maize (Zea mays L.)

ZHANG Ti-fu, LI Bo, ZHANG Deng-feng, JIA Guan-qing, LI Zhi-yong and WANG Shou-cai

National Maize Improvement Center/Maize Breeding Engineering Center, Ministry of Education/Key Laboratory of Crop Genomics and Genetic Improvement, Ministry of Agriculture/College of Agriculture and Biotechnology, China Agricultural University, Beijing 100193, P.R.China

Abstract

Heterosis has contributed greatly to yield in maize, but the nature of its contribution is not completely clear. In this study, two strategies using whole-genome oligonucleotide microarrays were employed to identify differentially expressed genes (DEGs) associated with heterosis and yield. The analysis revealed 1838 heterosis-associated genes (HAGs), 265 yield-associated genes (YAGs), and 85 yield heterosis-associated genes (YHAGs). 37.1% of HAGs and 22.4% of YHAGs expressed additively. The remaining genes expressed non-additively, including those with high/low-parent dominance and over/under dominance, which were prevalent in this research. Pathway enrichment analysis and quantitative trait locus (QTL) co-mapping demonstrated that the metabolic pathways for energy and carbohydrates were the two main enriched pathways influencing heterosis and yield. Therefore, the DEGs participating in energy and carbohydrate metabolism were considered to contribute to heterosis and yield significantly. The investigation of potential groups of HAGs, YAGs, and YHAGs might provide valuable information for exploiting heterosis to improve yield in maize breeding. In addition, our results support the view that heterosis is contributed by multiple, complex molecular mechanisms.

Key words: microarray, heterosis, yield, QTL, metabolism, maize

INTRODUCTION

Heterosis is commonly used to describe the superior performance of heterozygous $F_1$ hybrid plants relative to their homozygous parental inbred lines for a wide range of traits, with yield being the most important (Shull 1908, 1952). Although multiple genetic models have been proposed to explain the phenomenon of heterosis, there is still no strong consensus on the genetic and molecular basis of heterosis due to its complexity.

Two classic genetic hypotheses to explain heterosis are the dominance (or complementation) hypothesis (Davenport 1908; Bruce 1910) and the overdominance hypothesis (East 1908; Shull 1908). Although these two fundamental hypotheses have provided guidance to explain heterosis at the genomic level, another theory suggests that differential gene expression between inbreds and hybrids may be responsible for heterosis at transcriptional level (Song and Messing 2003; Hubner et al. 2005). It has been reported in many studies that gene expression patterns can be divided into either additive or non-additive allelic expression depending on the different levels of transcript accumulation in hybrids compared to their parental lines (Song and Messing 2003; Auger et al. 2005; Vuylsteke et al. 2005; Huang et al. 2006; Stupar and Springer 2006; Swanson-Wagner et al. 2006; Meyer et al. 2007; Uzarowska et al.
2007). In additive expression, the hybrids accumulate levels of transcript equal to the mid-parental level, while in non-additive gene expression, the hybrid transcript levels are significantly different from the mid-parent value. In this case, non-additive gene expression in hybrids includes levels of transcript equal to the high or low parent (high/low-parent dominance), above the high parent (overdominance) or below the low parent (underdominance). However, in general, the proportions of the different expression patterns have not shown a uniform trend in these gene expression studies. Some studies show that additivity is the main mode of gene action (Huang et al. 2006; Stupar and Springer 2006; Swanson-Wagner et al. 2006; Meyer et al. 2007), but in other studies non-additive gene expression was observed for the majority of differentially expressed genes (DEGs) (Guo et al. 2003; Song and Messing 2003; Auger et al. 2005; Uzarowska et al. 2007). In addition to these two trends, the proportion of additive versus non-additive expression was found to be similar in one study (Guo et al. 2006). Different approaches to analysis may underlie the discrepancy of these results, because due to technological limitations, some studies focus on selected genes and others focus on global and transcriptome-wide gene expression. Analysis of only a few, selected genes may not be sufficient to explain the complicated mechanism of heterosis from the genome-wide perspective.

In recent years, the rapid development of genome-wide expression technology, especially microarrays, has facilitated the collection of comprehensive information about gene expression from hybrids and their parental lines at genome-wide level. Many plants have been investigated for their gene expression profiles using microarrays for traits including anther development, panicle development, and response to low nitrogen stress in rice (Wang et al. 2005; Furutani et al. 2006; Lian et al. 2006), grain development in transgenic and wild type wheat (Gregersen et al. 2005), seed formation and response to virus infection or early dark treatment in Arabidopsis (Girke et al. 2000; Marathe et al. 2004; Kim and von Arnim 2006), fusicoccin-induced change in tomato (Frick and Schaller 2002), and the response to greenbug phloem feeding in sorghum (Park et al. 2006).

In maize, several analyses of DEGs as underlying heterosis-associated genes (HAGs) have also been performed using a high-throughput microarray strategy (Guo et al. 2006; Stupar and Springer 2006; Swanson-Wagner et al. 2006; Uzarowska et al. 2007). A variety of different tissues, including young embryos (Stupar and Springer 2006), seedlings (Stupar and Springer 2006; Swanson-Wagner et al. 2006), shoot apical meristems (Uzarowska et al. 2007) and immature ears (Guo et al. 2006; Stupar and Springer 2006) were used to identify HAGs in these studies. Although there are no consensus HAG member sets found among these diverse tissue types, ear tissue is promising for identifying yield and heterosis-associated genes because it is a reproductive organ, and therefore straightforwardly linked to yield. Previous studies demonstrated that early primordium stages are critical in determining the organ developmental fate or pattern (Sachs 1969; Cho et al. 2002). Therefore, analysis of immature ear will likely reveal YAGs.

In this study, we sought to identify maize yield and heterosis-associated genes through gene expression comparison of the immature ears at the spikelet differentiation stage using two different microarray strategies. Strategy I was expected to identify HAGs differentially expressed between a highly heterotic maize hybrid and its parents. Strategy II was designed to investigate YAGs differentially expressed between a high-yield pool (HYP) and a low-yield pool (LYP), which was comprised of five high-yield maize hybrids and five low-yield maize hybrids, respectively. Genes identified using both strategies I and II were considered maize yield heterosis-associated genes (YHAGs) in the strict sense. Furthermore, for the DEGs, we also wanted to investigate their metabolic pathways and related quantitative trait loci (QTLs). Finally, we attempted to establish relationships among genes, pathways, and QTLs for insight into the role of heterosis in yield.

RESULTS

Differential expression analysis

Based on the statistical analysis, there were 1,838 DEGs between the highly heterotic hybrid C8605-2×W1445 and either of its parents in strategy I (Li et al. 2009). These were considered to be HAGs due to the strong F1 heterosis performance. In strategy II, 265 DEGs were detected in the comparison between the HYP and
LYP (Appendix A). These 265 DEGs were classified as YAGs. Among the 265 YAGs, there were two expression trends: 118 (44.5%) were upregulated and 147 (55.5%) were downregulated in the HYP compared to the LYP. Of 265 YAGs, the 85 DEGs which showed significant differences using strategy I as well were considered YHAGs (Fig. 1, the order of the three gene sets is HAGs, YAGs, and YHAGs if they are not otherwise labeled). Of these, 85 YHAGs, 43 (50.6%) were upregulated and 47 (55.3%) downregulated in strategy I (5 YHAGs had the reverse expression orientation in comparison between hybrid and either of its parents) and 49 (57.6%) upregulated, 36 (42.4%) downregulated in strategy II. Only 21 YHAGs showed the same expression orientation both in strategies I and II, with 13 (61.9%) upregulated and 8 (38.1%) downregulated.

Expression patterns of DEGs

Expression patterns, also referred to expression polymorphisms, may provide clues to the regulation of gene expression. For the 1838 HAGs and 85 YHAGs, approximately 682 (37.1%) and 19 (22.4%) DEGs had additive expression levels, respectively. The remaining 1156 (62.9%) and 66 (77.6%) DEGs had non-additive expression levels (Appendix B). According to our calculation of expression levels, non-additive expression can be sorted into five patterns: high or low-parent dominance if the gene expression level in the hybrid is significantly higher/lower than that in one of the parents but is not significantly different from the other; overdominance or underdominance if the gene expression level in the hybrid is significantly higher/lower than both parents; and partial-dominance if the gene expression level in the hybrid is significantly higher than one parent but significantly lower than the other. Therefore, of the 1838 HAGs and 85 YHAGs, 675 (36.7%) and 24 (28.2%) displayed high-parent dominance, 339 (18.4%) and 24 (28.2%) displayed low-parent dominance, 89 (4.8%) and 15 (17.6%) displayed underdominance, 28 (1.5%) and 3 (3.5%) displayed partial-dominance. Only 25 (1.4%) HAG transcripts showed an overdominant expression pattern (Fig. 2 and Appendix B).

Functional analysis of DEGs

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a well-known bioinformatics resource with a large pathway database which facilitates the anchoring of genes of interest directly to known pathways. To further understand the biological functions of the DEGs identified in this research, we carried out pathway anchoring analysis using KEGG orthology based annotation system KOBAS (Mao et al. 2005). We found that among these three gene sets, 336 of 1838 (18.3%) HAGs, 48 of 265 (18.1%) YAGs, and 17 of 85 (20%) YHAGs could be classified into 23 function categories. However, three functional categories contained more DEGs in all three gene sets: (1) carbohydrate metabolism (66 HAGs, 14 YAGs, and 8 YHAGs), (2) energy metabolism (52 HAGs, 7 YAGs, and 1 YHAG), and (3) genetic information processing, including folding, sorting, and degradation (98 HAGs, 20 YAGs, and 10 YHAGs). It is known that enriched pathway identification can be affected by pathway size and background distribution. We applied the set of microarray probes as the background distribution for further analysis. A hypergeometric distribution test and false discovery rate (FDR) correction revealed that two functional categories were significantly enriched in the HAG set, eight functional categories were enriched in YAG set, and seven
functional categories were enriched in YHAG set ($P<0.05$ and $q<0.05$) (Table 1). Energy metabolism with $P<0.05$ and $q<0.3$ in the HAG set was also considered an enriched pathway to avoid that too small enriched pathway produced by relatively stringent cutoff could not be on target in this study (Appendix C).

**Fig. 2** Distribution of expression patterns. A, expression patterns of HAGs. B, expression patterns of YHAGs.

**Table 1** Enrichment analysis of functional classification among three gene sets

<table>
<thead>
<tr>
<th>Functional categories</th>
<th>HAGs</th>
<th>YAGs</th>
<th>YHAGs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Metabolism</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amino acid metabolism</td>
<td>34</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Biosynthesis of other secondary metabolites</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biosynthesis of polyketides and terpenoids</td>
<td>3</td>
<td>2*</td>
<td></td>
</tr>
<tr>
<td>Carbohydrate metabolism</td>
<td>66</td>
<td>14</td>
<td>8**</td>
</tr>
<tr>
<td>Energy metabolism</td>
<td>521)</td>
<td>7*</td>
<td>1</td>
</tr>
<tr>
<td>Glycan biosynthesis and metabolism</td>
<td>10</td>
<td>2**</td>
<td>1*</td>
</tr>
<tr>
<td>Lipid metabolism</td>
<td>21</td>
<td>1*</td>
<td>1</td>
</tr>
<tr>
<td>Metabolism of cofactors and vitamins</td>
<td>9</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Metabolism of other amino acids</td>
<td>14</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Nucleotide metabolism</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Xenobiotics biodegradation and metabolism</td>
<td>11</td>
<td>3*</td>
<td>2**</td>
</tr>
<tr>
<td><strong>Generic information processing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Folding, sorting, and degradation</td>
<td>98**</td>
<td>20**</td>
<td>10**</td>
</tr>
<tr>
<td>Replication and repair</td>
<td>60</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Transcription</td>
<td>16</td>
<td>11**</td>
<td>5**</td>
</tr>
<tr>
<td>Translation</td>
<td>45</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td><strong>Environmental information processing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Membrane transport</td>
<td>5</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Signal transduction</td>
<td>18</td>
<td>6**</td>
<td>4**</td>
</tr>
<tr>
<td>Signaling molecules and interaction</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td><strong>Cellular processes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell communication</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell growth and death</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell motility</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transport and catabolism</td>
<td>18</td>
<td>6*</td>
<td>5**</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unclassified</td>
<td>1502</td>
<td>207</td>
<td>68</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>1838</td>
<td>265</td>
<td>85</td>
</tr>
</tbody>
</table>

*Enrichment cutoff $P<0.05$ and $q<0.3$.

*, enrichment cutoff $P<0.05$ and $q<0.05$; **, enrichment cutoff $P<0.01$ and $q<0.05$. 
Energy metabolism was significantly enriched in both HAGs and YAGs. Carbohydrate metabolism was only enriched in YHAGs. In addition, glycan biosynthesis and metabolism and xenobiotic biodegradation and metabolism were enriched in both YAGs and YHAGs. In the category of genetic bioinformatics processing, folding, sorting, and degradation was the only subcategory significantly enriched in all three gene sets. Another enriched pathway in this category identified in both HAGs and YHAGs was transcription. Compared to metabolism and genetic pathways, signal transduction as an environmental information processing pathway was enriched in HAGs and YHAGs and the category of cellular transport and catabolism was only enriched in YHAGs.

Co-mapping QTLs with DEGs

To investigate the relationships between DEGs in three gene sets and QTLs, we collected 1726 maize QTLs distributed in seven categories from Gramene before April 13, 2010. Co-mapping QTLs and DEGs was subsequently implemented. QTL confidence intervals varied from approximately 12 to 40 cM. Those which were no more than 15 cM in normal maize genetic maps were selected to better represent real QTLs associated with DEGs.

The results demonstrated that 795 of 1838 (43.3%) HAGs, 120 of 265 (45.3%) YAGs and 39 of 85 (45.9%) YHAGs were located within QTL regions containing 218, 153, and 82 QTLs, respectively (Appendix D). The QTL region on chromosome 1 defined by flanking markers umc1318 and umc128 was found to be hot spot containing 52 QTLs in seven QTL categories (abiotic stress, anatomy, biochemical, development, quality, vigor, yield). Another abundant QTL interval flanked by bnlg1712 and umc1171 on chromosome 10 contained 23 QTLs in 5 QTL categories. The QTL category distribution of the three gene sets is shown in Fig. 3. Most of the QTL categories relate to yield, development, biochemical, and abiotic stress. Among these QTL categories, we were most interested in yield. Further analysis revealed that 180 of 795 (22.6%) HAGs, 23 of 120 (19.2%) YAGs and 7 of 39 (17.9%) YHAGs were mapped to 24 yield QTL intervals containing 54 yield QTLs. These yield QTLs on 10 maize chromosomes involved five yield traits: ear number, grain number, grain yield, kernel row number, and seed weight (Fig. 4).

We subsequently extracted candidate genes within yield QTLs with known KEGG pathways to deduce which pathways were related to yield significantly and found that only three metabolic pathways were included: carbohydrate, energy, and amino acid metabolism (Table 2). The pathway enrichment analysis showed carbohydrate and energy metabolisms to be enriched in one or two of the three gene sets. Both results confirm that carbohydrate metabolism and energy metabolism are critical to yield. It is consistent with the physiological process of yield. Photosynthesis (energy metabolism) converts solar energy to chemical energy which accumulates in organic compounds, and carbohydrate metabolism participates in the synthesis of carbohydrate, the primary dry matter of yield. Some of the genes identified in this experiment play key roles in energy and carbohydrate metabolism. MZ00024066, MZ00037162, and MZ00042017 are annotated as photosystem II proteins in photosynthesis (ko00195). MZ00005720 (EC:2.7.7.9), UDP glucose pyrophosphorylase, is a key enzyme in starch and sucrose metabolism, and MZ00009925 (EC:1.1.1.37), malate dehydrogenase, is found in several carbohydrate metabolic pathways, including glyoxylate and dicarboxylate metabolism (ko00630), pyruvate metabolism (ko00620), and the citrate cycle (ko00020).

DISCUSSION

Implications of HAGs, YAGs, and YHAGs

Heterosis is attractive to breeders because of its super-
ority in substantially increasing crop production, based on hybrids that outperform their parents. It has been widely used in crop breeding systems to produce elite hybrids for many years in rice, maize, and sorghum. But the contribution of heterosis to yield in high yielding hybrids has been debated. A previous study indicated that heterosis per se does not appear to be associated with yield improvement in USA maize hybrids (Duvick 1999); their results showed that heterosis has been declining in hybrids during the time that the yield of inbreds and hybrids was increasing from the 1930s to 1990s. However, we assumed that if the genes associated with both heterosis and yield could be identified, then the contribution of heterosis to yield could be explained.

In this study, we identified 1838 HAGs, 265 YAGs, and 85 YHAGs through two microarray strategies. Strategy I was a common microarray design which has been used by other researchers to monitor heterotic gene action between a hybrid and its parents (Swanson-Wagner et al. 2006; Pea et al. 2008; Wei et al. 2009). These 1838 HAGs, differentially expressed between the hybrid and either of its parents, were considered to be responsible for heterosis on the basis of the strong heterotic performance of hybrid F<sub>1</sub> versus its parents. Strategy II used a pooling strategy to compare the HYP and LYP, composed of five maize hybrids at the extremes of yield, in which the yield trait showed significant diversity and the other traits declined reciprocally. To reduce influence of the genetic background on the various hybrids in each pool, we fixed one of the parents in each pool: W1445 in HYP and W245 in LYP. Despite the fact that genomic interference could not be completely excluded, we believe that most of the 265 YAGs are genes associated with yield, with the possibility that other real YAGs were excluded by our criteria. Therefore, the number of genes associated with yield could be greater than what we identified, but our research identified 265 representative YAGs as significant. According to our definition, 85 YHAGs were defined as genes differentially expressed both in strategies I and II. Most of the YHAGs showed reversed gene expression trends in the two strategies. To date, we haven’t observed any implication in the proportion of transcripts with the two expression directions.

The link between heterosis and yield was established by introducing YHAGs at the transcriptional level in maize, and this could be a clue for further exploiting heterosis in yield improvement in hybrid maize breeding programs.

Fig. 4 Distribution of yield QTLs co-mapped with HAGs in maize linkage group.
Energy and carbohydrate metabolism are the two main pathways influencing yield and heterosis.

Pathway- and metabolite-based approaches have received more attention in studies of heterosis and yield in recent years (Schauer et al. 2008; Meyer et al. 2010; Song et al. 2010). To obtain further insight into the functions of the DEGs in three gene sets and their relations with established QTLs, we carried out pathway analysis based on KOBAS and co-mapping analysis with yield-related QTLs.

In the category of metabolism, we observed some
genes involved in primary metabolism, including amino acid, carbohydrate, and energy metabolisms, as well as some important secondary metabolites such as polyketides and terpenoids, cofactors and vitamins, other amino acids, and xenobiotic biodegradation and metabolism in both HAGs and YAGs. Lisee et al. (2008) indicated that primary metabolism, not secondary metabolism, was more closely linked to plant growth and development. Pathway enrichment analysis demonstrated that energy metabolism was statistically significantly enriched in HAGs and YAGs and carbohydrate metabolism was significantly enriched in YHAGs. A recent study reported energy metabolism was enriched in DEGs between a hybrid and its parents, rather than between parents, in rice (Wei et al. 2009). Pathways involved in energy metabolism were mainly photosynthesis and CO₂ fixation. It has been recognized that grain yield increase requires improvement of photosynthetic performance (Horton 2000; Richards 2000). Dry material accumulation in grains, contributed mostly by CO₂ assimilates, was observed to be promoted by higher photosynthetic capacity in wheat (Yang et al. 2007). As we known that photosynthetic yield as the main assimilates could contribute to the grain yield directly, the DEGs participating in the enriched energy metabolism might work together to increase assimilates source for yield improvement. The results illustrated that not all metabolic pathways, but only specific enriched pathways, were responsible for heterosis and yield significantly. It should be noted that pathways involved in carbohydrate metabolism might also contribute to heterosis in spite of not being enriched in HAGs but enriched in YHAGs. For example, differences in carbohydrate metabolism including the tricarboxylic acid (TCA) cycle, starch and sucrose metabolism, and inositol phosphate metabolism were identified as significant in a transcriptional comparison between superhybrid rice LYP9 and its parents (Ge et al. 2008; Wei et al. 2009; Song et al. 2010). Some subcategories of carbohydrate metabolism, glycolysis, pyruvate metabolism, and citrate cycle were the important respiration pathway which could provide material and energy used for growth and development. It might suggest reasonable that the role of DEGs deposited in carbohydrate metabolism be related to yield positively.

The integration of QTLs with DEGs provides new evidence in the search for candidate genes responsible for yield and heterosis and linking genes and phenotypes. In this study, the majority of QTLs co-localized with HAGs, YAGs, and YHAGs were distributed in the four categories of yield, development, biochemical, and abiotic stress. It reflects the close relationship between DEGs in the three gene sets and the corresponding traits. In studies on rice, this approach was also employed to establish relationships between heterosis related genes and yield QTLs (Huang et al. 2006; Song et al. 2010). Duvik (1999) indicated that improvement in tolerance to biotic and abiotic stresses would increase yield in maize hybrids. It is not surprising then, that abiotic stress QTLs contained many of HAGs and YAGs. We observed that QTLs were distributed discontinuously in the genome and there were some QTL-rich regions, also referred to hot spots, suggesting the presence of pleiotropic genes for additional biological functions underlying heterosis. The yield QTL category is our top interest because it involves traits that directly contribute to grain composition. Further analysis revealed five yield QTLs distributed around the maize genome. It is reasonable to hypothesize that the whole genome may be mobilized to favor yield, because it is such a complex trait. An interesting finding for searching candidate genes responsible for yield with known KEGG pathways is that all candidate genes localized to yield QTLs are annotated with roles in carbohydrate, energy, and amino acid metabolisms. The links among DEGs, metabolic pathways, and yield QTLs are expected to demonstrate that the metabolic pathways of carbohydrate, energy, and amino acids have a greater influence on yield compared to other metabolic pathways.

With the limited information in KEGG pathways and QTLs, we could not link all DEGs to annotation and QTL co-localization. Thus, at least to some extent, pathway enrichment analysis and yield QTL co-mapping indicate that energy and carbohydrate metabolisms are the two main metabolic pathways influencing heterosis and yield. The genes within yield QTLs involved in these two metabolic pathways are important candidate genes for heterosis and yield.

**Consideration of molecular mechanism of heterosis**

Despite the great success of using heterosis in crop
production, there is no consensus on its mechanism, from debates of the historical theories of dominance and overdominance, to the action of a few genes or global gene expression on cellular metabolism. Recently, investigations of global gene expression have deepened the understanding of the molecular mechanism of heterosis. Based on the hypothesis that heterosis could be caused by differential gene expression between a hybrid and its parents (Song and Messing 2003; Hubner et al. 2005), gene expression models are focused on deciphering heterosis.

In this research, 37.1% of HAGs and 22.4% of YHAGs exhibited additive expression. The remaining 63.9% of HAGs and 77.6% of YHAGs exhibiting non-additive expression predominated. However, this finding is not consistent with other studies, which demonstrated that additive expression was prevalent at the transcriptional level (Stupar and Springer 2006; Swanson-Wagner et al. 2006; Meyer et al. 2007; Song et al. 2010). The similar result was also observed even at the metabolite level (Marcon et al. 2010). The most likely reason for the difference among studies is the differing degrees of heterosis within the experimental materials used. All possible modes of action were also exhibited in the study which focus on expression of miRNAs and highly expressed small RNAs in rice (Chen et al. 2010). Additive expression is consistent with the dominance hypothesis and might be controlled by cis-elements and/or dosage-dependent trans-factors (Stupar and Springer 2006; Swanson-Wagner et al. 2006). Non-additive expression could be contributed by dominant allelic and nonallelic epistatic control of transcripts (Swanson-Wagner et al. 2006). In particular, although the possibility of overdominant gene action resulting from pseudooverdominance has been debated, the recent discovery of a single overdominant gene for yield in tomato confirms the existence of overdominance genes (Krieger et al. 2010). Combined with previous studies, our findings support the view that multiple molecular mechanisms give rise to heterosis.

MATERIALS AND METHODS

Experimental design and tissue sampling for microarrays

Previous studies on 33 maize hybrids consisting of a 6×6 incomplete diallel cross mating scheme from 12 inbred maize lines have been carried out by the authors (Li et al. 2009). Based on yield analysis, five hybrids were identified as high-yield hybrids for the HYP, and five were identified as low-yield hybrids for the LYP for use in comparison analysis in strategy II (Table 3). Hybrid C8605-2×W1445 was selected for strategy I because of its higher heterosis for yield (Li et al. 2009).

For tissue sampling in the maize growing season, the ten hybrids and two inbred lines (C8605-2, W1445) were grown at the Experiment Station of China Agricultural University (Beijing, China). Field management followed normal agricultural practices. The development stage of the topmost ears was monitored using a dissecting microscope to monitor morphological characters. Immature ears at the spikelet differentiation stage from the 12 lines were collected, immediately frozen in liquid nitrogen and stored at -80°C for RNA extraction. To maximally synchronize the samples of various materials, we collected all samples in the morning, between 8:00 a.m. and 10:00 a.m.

RNA isolation, cDNA synthesis, and labeling

Immature ears of each genotype were ground to a fine powder in liquid nitrogen, and total RNA was isolated with the Trizol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacture’s protocol. RNA samples were treated with the RNeasy Mini Spin Column Kit (Qiagen, Valencia, CA, USA) for purification. The purified RNAs were checked for quality by electrophoresis using a formaldehyde denaturing gel and quantified by spectrophotometry. To label probes, double-stranded complementary DNAs (cDNAs) were synthesized from 5 μg total RNAs and a T7-Oligo (dT) 9 primer with a cDNA Synthesis Kit (TaKaRa, Japan). After the double strand cDNA was purified with the QiAquick PCR Purification Kit (Qiagen, Germany), cRNAs were synthesized with templates of the second-strand cDNAs by in vitro transcription (T7 RiboMAX Express Large Scale RNA Production System, Promega, USA). Reverse transcription was conducted in the following reaction volume: 2 μg purified cRNAs, 300 U SuperScript II (Invitrogen, USA) and 2 μg 9-mer random primers. Following purification of reverse transcription products, 1 μg of product was labeled with Cy3-dCTP (Cy5-dCTP) in a reaction mixture containing 5 μL 5×Klenow buffer, 1 μL Cy3-dCTP (Cy5-dCTP), and 1.2 μL Klenow fragment.

| Table 3 Composition of the HYP and the LYP used in strategy II |
|------------------|------------------|------------------|------------------|------------------|
| HYP(1) | Maternal | Paternal | Maternal | Paternal |
| W1445 | C8605-2 | W245 | C8605-2 |
| W1445 | L87 | W245 | 757 |
| W1445 | K12 | W245 | SCY-4 |
| F349 | 87-1 | W227 | K12 |
| W245 | L87 | F349 | L87 |

(1) C8605-2×W1445 and its parents were used in strategy I.
Hybridization and data analysis

Maize oligomicroarrays (ver. 1.3) were obtained from University of Arizona (Arizona, USA). Each microarray set consisted of two slides (A and B) containing 57452 oligos, 70 nucleotides in length, representing 55000 maize-specific genes and 302 positive and 472 negative controls (complete information about the microarray is available at http://www.maizearray.org). Microarrays were hybridized overnight at 42°C in a mixture of 30 μL hybridization solution and equal amounts of cDNA from the two corresponding genotypes labeled with Cy3 or Cy5. After hybridization, microarrays were washed in 0.2% SDS and 2×SSC at 42°C for 5 min, 0.2% SDS at room temperature for 5 min, and finally dried by centrifugation. High quality hybridization is critical to the validation of the microarray data. Therefore, dye-swapping was used to assess the reproducibility and quality of the microarray hybridizations by alternating Cy3 and Cy5 for each genotype. Including the duplicate samples with reversed dye labeling, there are 8 microarray sets and 16 hybridizations in our experiment (Fig. 5).

A two channel LuxScan scanner (CapitalBio Co., China) was used to scan all microarrays, and GenePix Pro 4.0 software (Axon Instruments Inc., USA) was used to grid arrays. After analysis of the 16-bit TIFF images, the original fluorescence signal intensity values were processed. The extended package maanova v1.4.0 based on the R statistical environment was used for data analysis (Wu et al. 2003). First, flagged spots were eliminated from the analysis manually. Second, each spot’s background intensity was subtracted from signal intensity for further analysis. Before analysis, all background-subtracted signal intensity values were log2 transformed. To minimize the variation generated from spatial and intensity-dependent effects, the LOWESS normalization method was applied to the data (Wu et al. 2003).

To assess statistical differences in expression levels among these genotypes, analysis of variance (ANOVA) was carried out on the normalized signal data. The model of ANOVA is

\[ y_{ijk} = \mu + A_i + D_j + AD_{ij} + G_k + VG_{kg} + DG_{jk} + AG_{ig} + \varepsilon_{ijk} \]

In our analysis, \( VG \) is the most interesting term, because it represents the interaction effect between samples and genes. The value of \( \log \), ratio, the differential gene expression level between two particular samples, can be obtained from the difference in the estimated \( VG \) values (Cui and Churchill 2002). We introduced a statistical approach to determine whether the differential gene expression is statistically significant by adjusting \( P \)-values calculated from the ANOVA in which FDR could be controlled (Benjamini and Hochberg 1995).

In the biological and statistical senses, we defined DEGs as genes with \(|\log \text{ ratio}| \geq 2\), a two-fold cut widely used to make gene expression data comparable and adjusted \( P<0.05 \). In the highly heterotic hybrid C8605-2×W1445, the two parents C8605-2 and W1445 each contribute one dose to the genetic constitution. Therefore, a \( t \)-test was employed to compare gene transcription levels in hybrid C8605-2×W1445 to its mid-parent level \((P_{\text{male}}+P_{\text{female}})/2\). Additive expression in the hybrid would give a mid-parent level \((P_{\text{male}}+P_{\text{female}})/2\) with adjusted \( P>0.05 \), and non-additive expression was considered to be significantly different from the mid-parent level \((P_{\text{male}}+P_{\text{female}})/2\) with adjusted \( P<0.05 \). Non-additive expression was further classified into the following specific significant patterns. DEGs exhibited high-parent (low-parent) dominance if the \( F_1 \) genotype mean was significantly different from one parent but significantly larger (smaller) than the other parent and overdominance (underdominance) if the \( F_1 \) genotype mean was significantly larger (smaller) than both parents. Partial-dominance referred to DEGs with an \( F_1 \) genotype mean significantly larger than one parent but significantly smaller than the other parent.

Metabolic pathway analysis

We performed metabolic pathway analysis of DEGs using KOBAS (Mao et al. 2005). For each DEG sequence, KEGG Orthology (KO) terms were assigned based on similarity to amino acid sequences in the KEGG GENES dataset with default thresholds. It is essential to identify the statistically significantly enriched pathways in a set of sequences compared with a background because some pathways are naturally large and would involve more genes or proteins by chance alone. We also performed the FDR correction to reduce the false positives caused by multiple hypothesis tests in a large number environment (Benjamini and Hochberg 1995). Enriched pathways were determined by using a hypergeometric distribution test with the entire microarray probe
set as background distribution, with $P<0.05$ and $q<0.05$ simultaneously as the rather strict significance threshold.

**Co-mapping QTLs with DEGs**

First, we searched the maize sequence database (http://www.maizesequence.org) using DEG sequences as the query for Blastn to collect the closest markers flanking the sequence. Second, QTLs acquired from the Gramene database (www.gramene.org) were integrated into the high dense maize standard map IBM2 2005 Neighbors (Lee et al. 2002) employing QTL-Finder (Li et al. 2007) to locate the correct physical positions. Co-mapping was finished if a DEG was located within the confidence interval. To better describe the relationships between DEG and QTL, QTL confidence intervals spanning less than 40 cM were filtered which represent ~10 cM because the IBM2 2005 Neighbors Map is four times larger than normal maize genetic maps.

**Acknowledgements**

We are grateful to Dr. Li Lin, China Agricultural University, for his support in the work of QTL co-mapping. This work was supported by the National Basic Research Program of China (2007CB109000 and 2009CB118400).

**Appendix** associated with this paper can be available on http://www.ChinaAgriSci.com/V2/En/appendix.htm

**References**


