Plant Mol Biol (2006) 62:579–591
DOI 10.1007/s11103-006-9040-z

Heterosis and polymorphisms of gene expression in an elite rice hybrid as revealed by a microarray analysis of 9198 unique ESTs

Yi Huang · Lida Zhang · Jianwei Zhang · Dejun Yuan · Caiguo Xu · Xianghua Li · Daoxiu Zhou · Shiping Wang · Qifa Zhang

Received: 16 February 2006 / Accepted: 27 June 2006 / Published online: 29 August 2006

Abstract Despite the significant contributions of utilizing heterosis to crop productivity worldwide, the biological mechanisms of heterosis remained largely uncharacterized. In this study, we analyzed gene expression profiles of an elite rice hybrid and the parents at three stages of young panicle development, using a cDNA microarray consisting of 9198 expressed sequence tags (ESTs), with the objective to reveal patterns of gene expression that may be associated with heterosis in yield. A total of 8422 sequences showed hybridization signals in all three genotypes in at least one stage and 5771 sequences produced detectable signals in all slides. Significant differences in expression level were detected for 438 sequences among the three genotypes in at least one of the three stages, as determined by ANOVA validated with 100 permutations at $P < 0.05$. Significant mid-parent heterosis was detected for 141 sequences, which demonstrated the following features: a much larger number of sequences showed negative heterosis than ones showing positive heterosis; genes functioning in DNA replication and repair tended to show positive heterosis; genes functioning in carbohydrate metabolism, lipid metabolism, energy metabolism, translation, protein degradation, and cellular information processing showed negative heterosis; both positive and negative heterosis were observed for genes in amino acid metabolism, transcription, signal transduction, plant defense and transportation. The results are indicative of the biochemical and physiological activities taking place in the hybrid relative to the parents. Identification of genes showing expression polymorphisms among different genotypes and heterotic expression in the hybrid may provide new avenues for exploring the biological mechanisms underlying heterosis.

Keywords cDNA microarray · Expression polymorphism · Expression profile · Heterosis · Hybrid rice · Young panicle

Introduction

Heterosis refers to the phenomenon that hybrids perform better than the parents in an array of characteristics such as yield, biomass, stress tolerance, and reproducibility. Utilization of heterosis has become a major practice for increasing productivity of plants and animals, which has contributed significantly to the great increase of agricultural products worldwide in the last several decades, including the most important food crops such as maize and rice (Stuber 1994; Yuan 1998).

A considerable amount of efforts has been invested in unraveling the genetic basis of heterosis in crop plants. Two hypotheses, i.e., the dominance hypothesis (Davenport 1908) and the overdominance hypothesis (East 1908) were proposed early last century to explain the genetic basis of heterosis. The dominance hypothesis states that deleterious alleles at different loci in the parental genomes are complemented in the F_1 hybrid thus producing superior phenotype. The overdominance

Electronic Supplementary Material Supplementary material is available to authorised users in the online version of this article at http://dx.doi.org/10.1007/s11103-006-9040-z.

Y. Huang · L. Zhang · J. Zhang · D. Yuan · C. Xu · X. Li · D. Zhou · S. Wang · Q. Zhang (✉)
National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, China
e-mail: qifazh@mail.hzau.edu.cn

hypothesis asserts that the improved performance of an F_1 hybrid relative to its inbred parents is a consequence of favorable allelic interactions at heterozygous loci that outperform either homozygous state. Although many investigators favored one hypothesis over the other (Allard 1960), data allowing for critical assessment of the hypotheses remained largely unavailable until very recently with the advent of molecular marker technology and high-density molecular linkage maps. Variable results have been obtained from genetic analyses of heterosis mostly in maize and rice based on molecular marker linkage maps (Hua et al. 2002, 2003; Li et al. 2001; Luo et al. 2001; Stuber et al. 1992; Xiao et al. 1995; Yu et al. 1997), depending on the materials and the experimental designs employed. These results suggested that genetic approach alone is unable to fully characterize the biological mechanisms of heterosis.

Studies have also been performed in attempt to find relationships between gene expression and heterosis. Using the differential display technique, Xiong et al. (1998) studied the relationship between banding patterns of differentially displayed gene expression and the level of heterosis in a set of diallel cross involving eight elite parents of hybrid rice, and showed that dominant type of differential gene expression in flag leaf tissue did not seem to be correlated with heterosis in yield traits, while differential inhibition of gene expression in the hybrids appeared to be significantly correlated with heterosis. In wheat, Sun et al. (1999) observed quite different patterns of gene expression in the hybrids compared with their parents, as well as between heterotic and non-heterotic hybrids. Wu et al. (2003a) cloned cDNAs for a number of genes differentially expressed in hybrids and found that they belonged to functionally diverse categories. In maize, Tian and Dai (2003) observed association between heterosis and inhibition of gene expression in the hybrids. However, the information obtained from these studies was of limited use because of the scale of the experiments due to technical limitations.

Microarray technique has become a useful tool for exploring genome-scale gene expression (Schena et al. 1995). It has been widely applied in gene expression profiling of important organisms (Stoughton 2005). In plants, microarray analysis has been employed to investigate gene expression related to responses to hormones such as abscisic acid and gibberellin (De Paepe et al. 2004; Leonhardt et al. 2004; Yang et al. 2004; Zhong and Burns 2003), circadian clock (Harmer et al. 2000; Schaffer et al. 2001), light (Kim and von Arnim 2006; Ma et al. 2002; Tepperman et al. 2001), low phosphate or nitrogen nutrient conditions (Hammond et al. 2003; Lian et al. 2006; Wang et al. 2003; Wu et al. 2003b), cold, high salinity and drought stresses (Kawasaki et al. 2001; Rabbani et al. 2003; Seki et al. 2002), pathogen infections (Frick and Schaller 2002; Maleck et al. 2000; Marathe et al. 2004; Schenk et al. 2000), and insect attack (Halitschke et al. 2003; Park et al. 2006). There have also been reported studies on gene expression profiles of seed formation (Firnhaber et al. 2005; Girke et al. 2000; Gregersen et al. 2005), anther development (Wang et al. 2005), pollination and fertilization (Lan et al. 2004), and panicle differentiation (Furutani et al. 2006). To our knowledge, however, there has been no reported microarray study addressing heterosis at certain development stages, while data from such studies may open new avenues for unraveling the biological mechanisms governing heterosis.

Recently, there have been efforts in using model system, such as Arabidopsis (Meyer et al. 2004; Mitchell-Olds 1995; Syed and Chen 2005), for investigating heterosis, which may bring new hopes for complete characterization of the biological mechanisms of heterosis. Rice, in addition to being a major crop, has now become a model system for genomic research of monocot species, because of its small genome size, and the near completion of genome sequencing project (International Rice Genome Sequencing Project 2005). Moreover, the efforts in the development of hybrid rice in the last three decades not only demonstrated the existence of strong heterosis in this predominantly self-pollinating species, but also resulted in a large number of highly heterotic hybrids thus providing the material basis for research. Therefore rice can also serve as an excellent system for heterosis studies. In rice, heterosis has been observed for a range of characters, with many of the studies focusing on grain yield (Hua et al. 2002, 2003; Li et al. 2001; Luo et al. 2001; Xiao et al. 1995; Yu et al. 1997; Zhang et al. 1994, 1996). Although yield heterosis is the primary target for increasing productivity, the biological complexity of yield as a trait frequently makes it difficult to draw meaningful conclusions of the data in order to track individual causal elements involved in heterosis. In agronomic terms, yield is composed of three sub-traits (or components): number of panicles (or tillers) per plant, number of grains per panicle, and grain weight. Many studies have established that number of tillers per plant usually has low heritability and grain weight generally features low heterosis. In contrast, number of grains per panicle often shows high heritability as well as strong heterosis in heterotic crosses, and thus heterosis in yield is largely dependent on heterosis displayed by number of grains per panicle. Panicle size (or number of spikelets

per panicle) is formed at the stage of young panicle development, and is a function of the duration and rate of panicle differentiation. Thus, the developing young panicle is the tissue critical for determining heterosis and therefore should be the target for discovering the biological processes leading to heterosis in yield.

In the study reported in this paper, we analyzed the expression profiles of genes in an elite rice hybrid in comparison with the parents at three stages of young panicle development, using a cDNA microarray consisting of 9198 expressed sequence tags (ESTs), likely representing 9198 unique genes. The objectives were to reveal possible patterns of gene expression during young panicle development, discern processes likely to be associated with heterosis in this hybrid, and identify potential candidate genes that are involved in young panicle development and heterosis by comparing with data accumulated in previous studies.

Materials and methods

Rice genotypes and growth conditions

Shanyou 63, the most widely cultivated rice hybrid in China during the last two decades, and its parents Zhenshan 97 (female) and Minghui 63 (male) were used in this study. In the rice growing seasons of 2002 and 2003, germinated seeds were sown, allowed to grow in three plastic boxes for approximately 30 days, and transplanted in the field of the experimental farm of Huazhong Agricultural University, Wuhan, China. Each genotype was planted with two replications in plots consisting of 15 rows with 15 plants per row. The planting density was 16.5 cm between plants within a row, and the rows were 26.4 cm apart. The same planting of the three materials was repeated one more time 7 days after the first planting. The field management followed essentially the normal agricultural practice.

From early internode-elongation stage, the status of the young panicle development of the main culm was monitored using a dissection microscope by examining five plants per genotype sampled every three days. The stages of the young panicle development were determined as described (Xu and Xu 1984), by which young panicle development was considered to start when the first bracket primordium became visible, and end at the day of heading. The period from the first bracket primordium to heading was considered as the duration of young panicle development. The main culms were examined for a number of traits at 25 days after flowering, including number of primary branches, number of secondary branches, number of spikelets on primary branches, number of spikelets on secondary branches, and the total number of spikelets per panicle. The rate of differentiation was estimated as the ratio of number of spikelets to duration of differentiation. The measurements were averaged over the two replications for calculating mid-parent heterosis and high-parent heterosis.

Collection of young panicle tissues

Young panicles at the secondary branch primordium stage (stage III) and pistil/stamen primordium differentiation stage (stage IV) were collected under a dissection microscope, and put into a 1.5 ml microfuge tube with 1.0 ml Trizol regent (GIBCO/BRL) on ice, and stored at –70°C. Panicles at the pollen-mother cell formation stage (stage V) were harvested, placed immediately in liquid N_2 and stored at –70°C. To decrease the experiment error, tissues of the same stage of each genotype were mixed at the time of use before grinding. Isolation of total RNA was performed on three genotypes simultaneously.

Rice cDNA libraries

Chu et al. (2003) described a normalized cDNA library prepared from 15 different tissues of the rice cultivar Minghui 63. However, that library did not specifically include the young panicle tissue, thus genes expressed in young panicles might be under-represented in that library. To avoid such a possibility, we collected cDNA clones from a library prepared from young panicle tissues. For doing so, total RNA was isolated using Trizol reagent (GIBCO/BRL) according to the manufactures' instructions from an equi-mix of panicle tissues harvested at secondary branch primordium differentiation, pistil/stamen primordium differentiation stage and pollen mother cell formation stage. mRNA was purified using Dynabeads oligo$(dT)_{25}$ (Dynal A.S., Oslo, Norway). About 6 μg of mRNA was used for cDNA synthesis with the Superscript kit (GIBCO/BRL), according to the manufacture's instructions. Approximately 1800 randomly selected clones were sequenced, and the sequences were compared with those from the normalized cDNA library (Chu et al. 2003), resulting in 553 new sequences, which were added to the EST set for chip preparation.

Fabrication of rice microarray

A total of 9198 unique EST sequences were selected. cDNA clones were isolated and used as templates for PCR amplification in 96-well plates using T7 and Sp6

primers in 50-μl reaction volume. The amplification products were checked for quality and quantity by electrophoresis in 1% agarose gels. Amplified cDNAs were purified and suspended in 40 μl of 50% dimethylsulfoxide. The purified cDNAs were arrayed on glass slides (GAPS II Coated Slides, Corning, USA) using a GMS 417 arrayer (Genetic MicroSystems, Worburn, Massachusetts, USA). A clone of a rice actin gene (GenBank Acc. No. EI077C02) was also arrayed and repeated 10 times as the positive control, and a clone of a porcine glyceraldehyde 3-phosphate dehydrogenase gene (GenBank Acc. No. AF017079) was repeated 8 times per slide as the negative control. Post-arraying treatment of the slides proceeded by placing at room temperature for 48 h and incubating at 80°C for 4 h and stored in a jar containing silica gel till use. The design and the cDNA sequences of the arrays can be found on the website (http://redb.ricefgchina.org/mged/hy/).

Preparation of hybridization probe

Tissues were ground to fine powder in liquid nitrogen. Total RNA was extracted with Trizol regent (GIBCO/BRL) according to the manufacture's protocol. RNA samples were treated with DNase I for 10 min at 37°C before reverse transcription. The reverse transcription reaction was conducted in a volume of 40 μl, containing 60 μg of total RNA, 6.0 μg Oligo $dT_{(20)}$, 1× first-strand buffer (Life Technologies, Grand Island, New York, USA), 400 U Superscript II reverse transcriptase, 10 mM DTT, 40 U RNase inhibitor (Promega, Madison, Wisconsin, USA), 500 μM of dATP, dCTP and dGTP, 200 μM of dTTP and aa-dUTP (Sigma, St. Louis, Missouri, USA). After incubation at 42°C for 1 h, another 400 U Superscript II reverse transcriptase were added and incubated for another 2 h at 42°C. At the end of incubation, the reaction mixture was heated to 94°C for 3 min to stop the reaction. After cooled to room temperature, 2 U RNase H was added to the reaction mixture and incubated at 37°C for 10 min, followed by adding 1.0 μg RNase A and incubating at 37°C for 10 min. The reaction mixture was extracted with phenol/chloroform, purified with a Microcon YM-30 filter (Millipore, USA), and then coupled to Cy5-dye (Amersham Pharmacia, USA) as previously described (Ma et al. 2002). The probe was then purified with a Microcon YM-30 filter and the volume was adjusted to 7.0 μl with filtrated de-ionized water, and then 1.0 μl 20× SSPE, 1.0 μl blocking solution and 16.0 μl hybridization buffer were added. The mixture was heated to 90°C for 3 min and cooled to room temperature. After spun at 13,000 g for 3 min, the supernatant was ready for use.

Hybridization

About 25.0 μl of pre-hybridization buffer was placed onto the centre of a cover slip, and then the centre of the arrayed slide was placed over the coverslip (Corning, USA) to avoid forming bubbles. The slide was inverted and placed on a glass plate with the coverslip edges sealed by paper bond (Takara, Kyoto, Japan). The glass plate was transferred into a moist plastic box which was pre-incubated in an air bath at 76°C; the slide was kept in the air bath for 5 min then 50°C for 30 min. After pre-hybridization, the slide was put into double de-ionized water to remove the coverslip, washed in fresh double de-ionized water for 2 min, 70% ethanol for 2 min, absolute ethanol for 2 min, and finally spun at no more than 1200 rpm for 3 min for drying. The labeled probe was applied to the slide just like adding the pre-hybridization solution. The plastic box was enveloped with a piece of foil paper and incubated for 16–18 h in the air bath set at 42°C. After hybridization, slides were washed at room temperature with 2× SSC and 0.1% SDS for 10 min (twice), 0.2× SSC and 0.1% SDS for 10 min (twice), 0.2× SSC for 10 min (twice), and 0.02× SSC for 10 min (once). Washed slides were dried by centrifugation at no more than 1200 rpm for 3 min.

Data processing and analysis

The slides were scanned using a GMS 418 array-scanner (Genetic MicroSystems, Worburn, Massachusetts, USA), and the images were analyzed using the ImaGene 4.2 software (BioDiscovery, Los Angeles, CA), by which the scanning parameters were selected to make the sums of signal intensity equal among the slides.

The signal of each spot was subjected to spot filtering and normalization. First, spots flagged Bad by ImaGene 4.2 software were excluded from the analysis. Second, only those spots with fluorescent intensity levels at least two times above the local background were selected for further analysis. Third, in each tissue, only sequences that had detectable signals in all four independent hybridizations of each probe were accepted. The subtract method of background-correction was applied to the accepted sequences and the quantile normalization method was used to the corrected raw signal intensities (Bolstad et al. 2003). To assess statistically the differences of expression levels among the three genotypes in each tissue, the normalized signals were subjected to two-way ANOVA using genotype and year as the variables. A sequence that was identified as showing significant expression

difference among the genotypes at $P < 0.05$ was further tested by performing the ANOVA in the same manner using 100 randomization permutations of the original data of the same sequence without replacement. Then the original F value was compared against the F values calculated from random permutations. If fewer than 5% of the random F values were larger than the original F value, the difference was considered to be statistically significant at the 0.05 probability level. The expression levels of the three genotypes for a sequence that showed significant difference by ANOVA were ranked using the least significant difference (LSD) at the 0.05 probability level. All the data processing and analyses were performed using the R programming language through the limma package of Bioconductor (http://www.bioconductor.org/ and http://www.rproject.org/).

Mid-parent heterosis was calculated as: $(F_1 - MP)/MP$, in which F_1 is the performance of the hybrid and MP is the average performance of the two parents. To test whether mid-parent heterosis of gene expression was significantly different from zero, an h-statistic was devised: $h=[F_1-(P_1+P_2)/2]\Big/\sqrt{V_{F_1}+(V_{P_1}+V_{P_2})\Big/4}$, which is similar to the t-statistic with the degrees of freedom contributed by the observed values of the three genotypes, F_1, P_1 and P_2. This h-statistic was used to screen the heterosis values that were likely to be statistically significant at a pre-set probability level, which was $P < 0.05$ in this analysis. For a sequence that was identified by the h-statistic as showing significant heterosis at $P < 0.05$, the h-statistic was recalculated with 100 random permutations of the original data. If fewer than 5% of the random h values were larger than that based on the original data, the heterosis value was considered to be statistically significant at $P < 0.05$.

Results

Performance of the hybrid relative to the parents in panicle traits

Panicles of the main culms at 25 days after flowering from the hybrid Shanyou 63 were much larger than those from either of the parents (Fig. 1). Shanyou 63 had higher values in all the attributes of the panicle size examined based on samples of 10 plants each, including, on a per panicle basis, numbers of primary branches, secondary branches, spikelets on primary branches, spikelets on secondary branches, and total spikelets (Table 1). Biologically, number of spikelets per panicle was determined by the duration and rate of differentiation; the hybrid had both higher rate and longer duration of differentiation (Table 1).

Fig. 1 Panicles of Zhenshan 97 (left), Minghui 63 (right), and the hybrid Shanyou 63 (middle)

Mid-parent heterosis and high-parent heterosis were calculated for the four traits. There was strong heterosis for all the attributes of panicle size (Table 1). The mid-parent heterosis varied from 19.3% in differentiation rate to 54.6% in spikelets on primary branches. There was also substantial high-parent heterosis for all the attributes of the panicle size.

General features of the hybridization signals

For ease of description, we will refer to each of the 9198 ESTs as a ''sequence''. According to the criteria that were applied for the data processing, a total of 8422 sequences produced hybridization signals in all four replicates, including two biological replicates (2 years) and two technical replicates (repeats within years), in all three genotypes at least in one of the three stages. The number of sequences that produced detectable signals in all three genotypes varied slightly from one stage to another, ranging from 7176 for stage III, to 7223 for stage IV and 7481 for stage V, with 5771 sequences producing detectable signals in all slides. Correlations between technical repeats ranged from

Table 1 Means and heterosis of panicle traits of the parents and hybrid

Trait	Zhenshan 97	Minghui 63	Shanyou 63	Mid-parent heterosis (%)	High-parent heterosis (%)
Primary branches (PB)/panicle	13.1 ± 0.2	11.2 ± 0.1	16.7 ± 1.4	37.8	28.0
Secondary branches (SB)/panicle	22.5 ± 2.9	21.7 ± 0.8	27.1 ± 2.7	22.6	20.5
Spikelets on PB/panicle	50.0 ± 0.7	58.8 ± 1.3	84.0 ± 1.4	54.6	43.1
Spikelets on SB/panicle	56.2 ± 3.8	70.7 ± 1.8	86.4 ± 10.4	36.2	22.2
Spikelets/panicle	106.2 ± 3.1	129.5 ± 3.1	170.5 ± 11.7	44.7	31.7
Differentiation duration[a]	23	30	32	20.8	6.7
Differentiation rate[b]	4.6	4.3	5.3	19.3	15.5

[a] Number of days from meristem elongation to the day before heading

[b] Number of spikelets divided by differentiation duration

0.85 to 0.99, and correlations between biological repeats varied from 0.54 to 0.72.

Expression polymorphisms

On a sequence-by-sequence basis, significant differences in the expression levels were detected for approximately 5.2% (438) of the 8422 sequences among the three genotypes in at least one of the three stages (Table 2), using the screening by the two-way ANOVA at $P < 0.05$, followed by validation with 100 random permutations. We referred to the significant differences so identified as expression polymorphisms. Overall, a larger number of sequences showed expression polymorphisms between the parental lines than between the hybrid and either of the parental lines at all three stages. Comparison between Zhenshan 97 and the hybrid revealed a larger number of sequences showing expression polymorphisms than did the comparison between Minghui 63 and the hybrid. It was also clear from Table 2 that a greater number of sequences showed expression polymorphisms at stage V than stages III and IV.

The details of the differences of the expression levels among the three genotypes ranked using the LSD at $P < 0.05$ are given in Table S2.

Table 2 Number of sequences showing significant differences among the three genotypes at the same stage detected by ANOVA at $p < 0.05$ and validated using 100 random permutations

	Stage III[a]	Stage IV	Stage V	Total
Shanyou 63/Zhenshan 97	60	28	134	217
Shanyou 63/Minghui 63	48	49	65	159
Zhenshan 97/Minghui 63	79	83	163	318
Total	121	100	229	438

[a]III, IV and V represent the secondary branch primordium differentiation, pistil/stamen primordium differentiation and pollen mother-cell formation stages respectively

Functional classification of the sequences showing polymorphic expression

The 438 differentially expressed sequences could be classified into putative major functional categories according to Gene Ontology (GO) and Non-redundancy (NR) databases of NCBI with a BLASTx at *E*-value $\leq 10^{-5}$ or BLASTn sore ≥100: (1) metabolism, (2) genetic information processing, (3) cellular information processing, (4) environment information processing, (5) unclassified, and (6) unknown function (Table 3). These were further classified into sub-categories according to their functions. The largest category consisted of 215 sequences with unknown functions. Among the sequences with homology to genes of known functions, the largest number of sequences was in the category of genetic information processing (87), followed by metabolism (72), and environmental information processing (49). The category of cellular processing had the smallest number of sequences (14).

To test whether the occurrence of sequences with expression polymorphism was differentially associated with any of the functional categories, a χ^2-test was performed to assess goodness-of-fit between the observed numbers of sequences and the expectations based on frequencies of the total 8422 sequences falling in the various categories (Table 3). The results of the test showed that overall there was no significant discrepancy between the expected and the observed numbers ($\chi^2 = 17.1$, $P > 0.58$), indicating no differential occurrence of sequences in the various functional categories.

Mid-parent heterosis and functional categories of the heterotically expressed sequences

Mid-parent heterosis of the expression value (also referred to as heterotic expression) was calculated using the normalized signal for each of the 438 sequences showing expression polymorphism at least in

Table 3 Functional categories of the sequences showing significant differential expression among the three genotypes in at least one of the three stages of young panicle development

Functional categories	Stage			Total	
	III	IV	V	Observed	Expected[a]
Metabolism	20	23	33	72	72.9
Amino acid metabolism	4	5	9	17	12.6
Carbohydrate metabolism	6	9	12	25	28.7
Energy metabolism	3	3	3	8	8.0
Lipid metabolism	2	4	4	10	10.9
Nucleotide metabolism	3	2	2	7	5.9
Metabolism of others	1	0	1	2	2.6
Secondary metabolism	1	0	1	2	4.2
Genetic information processing	26	19	42	87	72.5
DNA replication and repair	0	2	6	8	7.1
Transcription factors	6	5	16	27	24.2
Transcription	3	5	6	14	11.1
Translation	14	3	7	24	14.9
Protein degradation	3	4	7	14	15.3
Cellular processing	5	2	7	14	14.9
Environment information processing	17	8	25	49	58.0
Plant defense	4	3	6	12	8.3
Signal transduction	5	3	10	18	24.7
Transport	8	2	9	19	25.0
Unclassified	0	0	1	1	1.9
Unknown function	53	48	121	215	217.8
Hypothetical	22	16	51	87	85.9
No significant homology	20	16	50	82	89.2
No hits found	11	16	21	47	42.6
Total	121	100	229	438	

[a]Expected numbers based on the frequencies of the total 8422 sequences falling in the various categories

one of the three stages. The significance test of mid-parent heterosis with the *h*-statistic at $P < 0.05$ followed by validation with 100 random permutations identified a total of 141 sequences as showing significant heterotic expression (Table 4). Only one sequence showed significant negative heterosis simultaneously at two stages, and 140 sequences exhibited significant heterosis at only one of the three stages.

When the numbers of sequences that occurred in each of the functional categories were compared with the expected numbers according to the frequencies of the total 8422 sequences falling in the various categories (Table 5), it was shown that overall there was no obvious discrepancy between the observed and expected numbers ($\chi^2 = 17.8$, $P > 0.50$).

Patterns of heterotic expression and differential occurrence of the sequences in various functional categories

As described above, 141 sequences were identified as showing heterotic expression in at least one of the three stages, among them 118 sequences had detectable hybridization signals in the three genotypes at all three stages, and the remaining 23 sequences had detectable hybridization signals for all the three genotypes at only one or two stages (Table S3). Three major patterns could be identified according to the directions and levels of the heterosis displayed by the 118 sequences: (1) positive heterosis at all three stages, (2) positive heterosis at one or two stages but negative heterosis at other stage(s), and (3) negative heterosis at all three stages (Table S3).

Inspection of the data presented in Table S3 clearly revealed that a greater number of sequences showing negative heterosis than ones showing positive heterosis. Only 7 sequences displayed the first pattern in which positive heterosis was detected at all three stages, compared with 50 sequences showing the third pattern in which negative heterosis was observed at all three

Table 4 Number of sequences showing significant mid-parent heterosis at three stages of young panicle development identified using the h-statistic and validated with 100 random permutations.

Range of heterosis (%)	Stage III	Stage IV	Stage V
>100	1	1	1
50–100	3	1	4
25–50	2	3	7
0–25	4	0	7
–25 to 0	4	5	25
–50 to –25	23	11	40
Total	37	21	84

Table 5 Observed number of sequences showing significant heterosis compared with the expected number (based on the classification of the total 8422 sequences) in each of the functional categories

Functional categories	Observed	Expected
Metabolism	23	23.4
Amino acid metabolism	8	4.1
Carbohydrate metabolism	7	9.2
Energy metabolism	3	2.6
Lipid metabolism	2	3.5
Nucleotide metabolism	1	1.9
Metabolism of others	2	1.4
Secondary metabolism	0	0.8
Genetic information processing	28	23.4
DNA replication and repair	4	2.3
Transcription factors	10	7.8
Transcription	4	3.6
Translation	5	4.8
Protein degradation	5	4.9
Cellular processing	4	4.8
Environment information processing	22	18.7
Plant defense	6	2.7
Signal transduction	9	8.0
Transport	7	8.0
Unclassified	1	0.6
Unknown function	63	70.1
Hypothetical	23	27.7
No significant homology	31	28.7
No hits found	9	13.7
Total	141	

stages. There were also similar trend in the sequences showing the second pattern, in which a greater number of sequences showed negative heterosis than ones showing positive heterosis.

It is also clear from Table S3 that the functional categories appeared to occur differentially among the sequences showing positive and negative heterosis. For example, 3 of the 7 sequences showing the first pattern in which positive heterosis occurred at all three stages had homology with genes of known functions, of which one sequence was related to amino acid metabolism, two having roles in transcription. Whereas among the sequences displaying the third pattern in which negative heterosis occurred at all three stages, 26 had homology with genes of known functions. Six of the 26 sequences had roles in metabolisms (2 in carbohydrate metabolism, 2 in lipid metabolism, one in energy metabolism and one in metabolism of others), 13 were involved in genetic information processing (7 in transcription, 3 in translation and 3 in protein degradation), 2 in cellular information processing, 5 in environment information processing (3 in plant defense and 2 in transportation). Thus, sequences showing negative heterosis were involved in more diverse functional categories than ones showing positive heterosis. Again the same trend could be observed among the sequences showing the second pattern in which positive heterosis occurred at one or two stages and negative heterosis at other stage(s).

Further inspection of the results presented in Table S3 indicated that there were associations between the occurrence of heterosis and the functional categories. In the general category of metabolism, sequences involved in amino acid metabolism showed both positive and negative heterosis, while carbohydrate metabolism, energy metabolism and lipid metabolism were associated with negative heterosis. In the category of genetic information processing, DNA replication and repair seemed to be associated with positive heterosis, and translation and protein degradation associated with negative heterosis, while both positive and negative heterosis occurred in sequences involved in transcription. In the category of cellular information processing, all the sequences appeared to be associated with negative heterosis. Whereas the sequences in the category of environmental information processing, which included plant defense, signal transduction and transport, were split between positive heterosis and negative heterosis.

Locations of the significant differentially expressed sequences in QTL containing regions

Previous studies identified a large number of QTLs for various panicle traits using populations derived from a cross between Zhenshan 97 and Minghui 63, including $F_{2:3}$, recombinant inbred lines, and ''immortalized F_2'' populations (Cui et al. 2003; Hua et al. 2002, 2003; Xing et al. 2002; Li et al. 2000; Yu et al. 1997). By homology search ($E < 10^{-5}$), 436 of the 438 sequences that showed significant expression polymorphisms identified above could be mapped to the 12 chromosomes. Of them, 64 sequences were localized to 26 intervals spanning 282.0 cM of the genome where QTLs were previously identified for panicle traits, including number of primary branches, number of secondary branches, number of spikelets on primary branches, number of spikelets on secondary branches and number of spikelets per panicle. Twenty-four of the 64 sequences showed significant heterotic expression as identified above and their correspondence with the QTLs were also indicated in Table 6. The lengths of the QTL intervals as defined by their flanking markers varied from approximately 1.7 cM for *SSBp2-1* for spikelets on secondary branches and *SP6a* for number of spikelets per panicle, to approximately 35 cM for *Sp6b* for number of spikelets per panicle. Most of the intervals were of about 10 cM or less, and

in a number of cases, the intervals were less than 5.0 cM. Table 6 also showed that, for all the traits listed, there were cases in which multiple sequences were located within short QTL intervals.

It can also be seen from Table 6 that there were several intervals each harboring two or more QTLs, due to either close linkage or pleiotropic effects. For example, QTLs for number of secondary branches, number of spikelets on secondary branches and number of spikelets per panicle were detected in the interval C2340-C86 (5.8 cM) on chromosome 1, five sequences with expression polymorphisms including one showing heterotic expression were resolved to this interval. These sequences included a receptor-like kinase protein, sorting nexin 1 protein and an AtPH1-like protein, in addition to a hypothetical and an unknown protein. Three QTLs were detected in the interval R712-RZ324 (4.1 cM) on chromosome 2 for three different traits, number of secondary branches, number of spikelets on secondary numbers and number of spikelets per panicle. One sequence was located in this interval encoding a proline-rich family protein. Similarly, three of the differentially expressed sequences were located in the interval RM239-C1633 (23.3 cM) on chromosome 10 where QTLs for number of primary branches and number of spikelets on primary branches were resolved. Two of the three sequences had homology with a hypothetical protein, and the other in the category of no hits found. Additionally, 8 differentially expressed sequences including 6 showing heterotic expression involving two transcription factors were mapped to the interval C1023-R1440 (8.0 cM) on chromosome 7 where a QTL for number of spikelets per panicle resided.

Discussion

The analysis of cDNA microarray containing 9198 sequences, with 8422 sequences having detectable hybridization signals in all three genotypes at least in one of the three stages, and 5771 sequences showing detectable hybridization signals at all three stages of young panicle development, provided a considerable

Table 6 Number of significant differentially expressed sequences located in each of the QTL regions

Trait	QTL	Chr.	Position (cM)	Interval	Number of sequences	
					Expression polymorphism	Heterotic expression
No. primary branches	NPB6-1	6	75.1–82.9	R2147-RZ667	1	0
	NPB10-1	10	21.1–44.4	RM239-C1633	3	0
	Total				4	0
No. secondary branches	NSB1-1	1	213.4–219.2	C2340-C86	5	1
	NSB1-2	1	28.8–37.3	G359-RG532	2	0
	NSB2-1	2	71.4–75.5	R712-RZ234	1	1
	Total				8	2
No. spikelets on primary branches	SPBp1-1	1	219.2–245.3	C86-RG236	5	1
	SPBp6-1	6	82.9–114.1	RZ667-RG424	12	5
	SPBp6-2	6	0–2.2	C474-R2869	2	0
	SPBp10-2	10	21.1–44.4	RM239-C1633	3	0
	SPBp6-3	6	8.4–12.9	C952-Waxy	1	1
	Total				23	7
No. spikelets on secondary branches	SSBp1-1	1	213.4–219.2	C2340-C86	5	1
	SSBp1-2	1	28.8–37.3	G359-RG532	2	0
	SSBp2-1	2	87.5–89.2	R1843-C777	7	2
	SSBp2-2	2	71.4–75.5	R712-RZ324	1	1
	SSBp3-1	3	117–128.5	C746-C944	1	1
	SSBp6-2	6	18.5–21.8	R1952a-R2749	1	0
	Total				17	5
No. spikelets per panicle	SP1a	1	28.8–37.3	G359-RG532	2	0
	SP1b	1	213.4–219.2	C2340-C86	5	1
	TS2-1	2	71.4–75.5	R712-RZ324	1	1
	SP2	2	7.7–19.4	R2510-RM211	1	0
	SP3	3	88.1–94.4	C1087-RZ403	3	2
	SP6b	6	131.9–166.8	R2549-C962	5	2
	SP7a	7	51.5–59.5	C1023-R1440	8	6
	SP7b	7	94.4–102.6	RM70-R1245	2	0
	SP7c	7	119.1–129	RM234-R1789	8	3
	SP11	11	102.7–115.5	C1003B-G4001	1	0
	Total				36	15

amount of information for assessing the patterns of gene expression that may be relevant to heterosis.

Expression polymorphisms

An important finding of this study is the significant differences in the expression levels of the same genes among the three genotypes at the same developmental stages for a considerable proportion of the sequences, which were referred to as expression polymorphisms. Such polymorphisms occurred in several forms as displayed by the large number of genes included in this study: the genes may be highly expressed in one of the parents but low or little expression in the other parent and hybrid, or highly expressed in both parents but little expressed in the hybrid, or highly expressed in the hybrid but low in both parents. In addition, there were also extensive quantitative differences in the expression levels among the three genotypes that did not reach the level of detection by *F* statistic and LSD at the 0.05 probability level. It should be noted that ESTs arrayed on the chips may hybridize better to the probes prepared from Minghui 63 than those from the other two genotypes in case of DNA polymorphism between the parents that seems to occur frequently even within the indica subspecies (Zhang et al. 2005), which may cause bias in the data analysis.

The expression polymorphism may have important implications in understanding the regulation of gene expression. For long, it has been believed that gene expression is spatially and temporally regulated in the process of growth and development of the organisms, thus resulting in stage or tissue specific expression that can be unambiguously defined. The results from this study clearly revealed that gene expression level at given tissues or stages are also dependent on the genotypes such that a gene may be strongly expressed in one genotype at a given tissue but very weakly expressed in another genotype in the same tissue, resulting in large differences, or polymorphisms, in gene expression level among the genotypes.

The expression polymorphisms, also referred to as allele-specific gene expression (Knight 2004), have received substantial attention recently in humans, animals and plants (Pastinen and Hudson 2004). Allelic variation in gene expressions were reported to be common in mouse (Cowles et al. 2002) and human (Lo et al. 2003), and allele specific accumulations of transcripts were also observed in maize (Guo et al. 2004). Interestingly, the results of recent studies showed that quantitative variations of gene expression levels were controlled genetically, and regulated by both *cis*- and *trans*-acting loci, with trans-acting elements playing important roles (Schadt et al. 2003). Based on current knowledge, *cis*-acting regulatory polymorphisms mostly refer to variations in the promoter regions that alter the level of gene transcript (Pastinen and Hudson 2004). Moreover, polymorphisms in coding regions are also known to affect the level of expression (Pastinen and Hudson 2004). It is also possible that different genotypes may produce different cellular signals for activating the gene expression. Although more sequence information and detailed characterization for each of the differentially expressed genes are necessary in order to resolve the causes of the expression polymorphism as observed in the present study, characterization of the mechanisms for regulating the expression polymorphisms will certainly advance the fundamental understanding of biological processes underlying growth and development leading to phenotypic diversity, and also enhance the understanding of mechanism for heterosis.

Heterotic expression

Approximately 32.2% (141 of 438) of the significant differentially expressed sequences detected in this study showed various levels of heterotic expression, expressed as mid-parent heterosis. It was found, though surprising, that a much larger number of sequences showed negative heterosis than positive heterosis, an observation that appears to be consistent with the finding reported by Xiong et al. (1998) who found negative correlation between F_1-specific expression and heterosis. Such results seemed to suggest that an overall down-regulation of the gene expression may be related to heterosis, which is in contrast to the higher rate of growth and differentiation of the panicles in the hybrid than the parents, a phenomenon that is difficult to understand at present.

Another important feature observed among the heterotically expressed genes was the preferential association of positive or negative heterosis with the functional categories, which could be summarized as the following: (1) genes functioning in DNA replication and repair tended to show positive heterosis, (2) genes functioning in carbohydrate metabolism, lipid metabolism, energy metabolism, translation, protein degradation, and cellular information processing appeared to show negative heterosis, and (3) some of the genes in the categories of transportation, signal transduction, plant defense, amino acid metabolism and transcription showed positive heterosis while others in the same categories showed negative heterosis. Although the information is not conclusive due to the limited representation of the sequences arrayed on the chip, the results are indicative of the biochemical

and physiological activities taking place in the hybrid relative to the parents. However, extensive investigation is needed to determine whether such observed up- and down-regulations provided the molecular basis of heterosis for panicle development.

It should be pointed out that, based on a single locus model, heterosis and dominance describe different attributes of the same phenomenon. Detection of heterosis indicates dominance of various degrees ranging from partial dominance, full dominance to overdominance. Among the 141 heterotically expressed sequences identified in this study, 24.1% (34) showed higher expression levels in the hybrid than the mid-parent, and 74.5% (105) exhibited lower expression levels in the hybrid. Interestingly, most of the sequences with detectable mid-parent heterosis (either positive or negative) also showed high-parent heterosis (Table S3), in which the expression level of the hybrid is higher than the high parent (or lower than the low parent), thus signifying overdominance (or underdominance). In heterosis study of rice, overdominance was reported to be relatively frequent in the trait of grain yield, but rare in yield component traits (Hua et al. 2002, 2003; Yu et al. 1997; Zhang et al. 1994). Such observation was ascribed to a multiplicative effect among the component traits: number of tillers per plant, number of grains per panicle and grain weight, such that small dominance effects of the component traits would multiplicatively amplify each other to produce overdominance in the composite trait (Zhang et al. 1994). Whereas, insofar as the expression level of individual gene was concerned, the above explanation obviously could not hold since there is no such obvious means to resolve the level of the gene expression into individual components.

However, the heterotic expression (either positive or negative) detected for a large number of sequences in the hybrid compared to the parents clearly suggested that the regulatory mechanisms in the hybrid were different from those in the parents. Thus such expression heterosis may be viewed as special cases of expression polymorphisms as discussed above, which warrant intensive investigation in future studies. In addition, the expression heterosis may be used as a model for understanding heterosis at the level of trait phenotype, and provide insight into the causes of heterosis of agronomic traits, such as panicle size targeted in this study.

Correspondence of the significant differentially expressed sequences with QTLs of panicle traits

This analysis also intended to establish correspondence between the differentially (including heterotically) expressed sequences and the QTLs that were previously identified for panicle traits using populations derived from a cross between the parents used in this study. This effort resolved 436 of 438 sequences that showed significant expression polymorphism to the 12 chromosomes, of which, 64 were localized to 26 QTL intervals spanning 282.0 cM region covering approximately 23.8% of the total genome. Hitzemann et al. (2003) suggested that gene expression data together with QTL analysis may provide an avenue for identifying candidate genes for the traits of interest. The practice of this study demonstrated that although such an analysis could not offer precise information about the candidate genes for the traits due to the statistical nature of the QTLs, the results nonetheless suggested a large number of likely candidates for the genes that warrant further studies, especially with the current international efforts in functional genomics aiming at systematic characterization of all the genes in the rice genome.

It should be pointed out that the current studies only included a total of 9198 unique gene like sequences in the chips, which made up of less than 17% of the total genes as predicted (http://www.tigr.org). There has been recent effort to place the entire set of rice genes on the chip (Stolc et al. 2005). Expression profile of the entire set of the rice genes at the stages as did in this study will certainly identify all the genes that are relevant for heterosis study of the panicle traits.

Acknowledgements This work was supported by grants from the National Program on the Development of Basic Research, the National Special Key Project of Functional Genomics and Biochips, and the National Natural Science Foundation of China.

References

Allard RW (1960) Principles of plant breeding. John Wiley and Sons, New York

Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19:185–193

Chu ZH, Peng KM, Zhang LD, Zhou B, Wei J, Wang SP (2003) Construction and characterization of a normalized whole-life-cycle cDNA library of rice. Chinese Sci Bull 48:229–235

Cowles CR, Hirschhorn JN, Altshuler D, Lander ES (2002) Detection of regulatory variation in mouse genes. Nat Genet 32:432–437

Cui KH, Peng SB, Xing YZ, Yu SB, Xu CG, Zhang Q (2003) Molecular dissection of the genetic relationships of source, sink and transport tissue with yield traits in rice. Theor Appl Genet 106:649–658

Davenport CB (1908) Degeneration, albinism and inbreeding. Science 28: 454–455

De Paepe A, Vuylsteke M, Van Hummelen P, Zabeau M, Van Der Straeten D (2004) Transcriptional profiling by cDNA-AFLP and microarray analysis reveals novel insights into

the early response to ethylene in *Arabidopsis*. Plant J 39:537–559

East EM (1908) Inbreeding in corn. In: Report of the Connecticut Agricultural Experiment Station for Years 1907–1908, New Haven, CT, pp 419–428

Firnhaber C, Puhler A, Kuster H (2005) EST sequencing and time course microarray hybridizations identify more than 700 *Medicago truncatula* genes with developmental expression regulation in flowers and pods. Planta 222:269–283

Frick UB, Schaller A (2002) cDNA microarray analysis of fusicoccin-induced changes in gene expression in tomato plants. Planta 216:83–94

Furutani I, Sukegawa S, Kyozuka J (2006) Genome-wide analysis of spatial and temporal gene expression in rice panicle development. Plant J 46:503–511

Girke T, Todd J, Ruuska S, White J, Benning C, Ohlrogge J (2000) Microarray analysis of developing *Arabidopsis* seeds. Plant Physiol 124:1570–1581

Gregersen PL, Brinch-Pedersen H, Holm PB (2005) A microarray-based comparative analysis of gene expression profiles during grain development in transgenic and wild type wheat. Transgenic Res 14:887–905

Guo M, Rupe MA, Zinselmeier C, Habben J, Bowen BA, Smith OS (2004) Allelic variation of gene expression in maize hybrids. Plant Cell 16:1707–1716

Halitschke R, Gase K, Hui D, Schmidt DD, Baldwin IT (2003) Molecular interactions between the specialist herbivore *Manduca sexta* (lepidoptera, sphingidae) and its natural host *Nicotiana attenuata*. VI. Microarray analysis reveals that most herbivore-specific transcriptional changes are mediated by fatty acid-amino acid conjugates. Plant Physiol 131:1894–1902

Hammond JP, Bennett MJ, Bowen HC, Broadley MR, Eastwood DC, May ST, Rahn C, Swarup R, Woolaway KE, White PJ (2003) Changes in gene expression in *Arabidopsis* shoots during phosphate starvation and the potential for developing smart plants. Plant Physiol 132:578–596

Harmer SL, Hogenesch JB, Straume M, Chang HS, Han B, Zhu T, Wang X, Kreps JA, Kay SA (2000) Orchestrated transcription of key pathways in *Arabidopsis* by the circadian clock. Science 290:2110–2113

Hitzemann R, Malmanger B, Reed C, Lawler M, Hitzemann B, Coulombe S, Buck K, Rademacher B, Walter N, Polyakov Y, Sikela J, Gensler B, Burgers S, Williams RW, Manly K, Flint J, Talbot C (2003) A strategy for the integration of QTL, gene expression, and sequence analyses. Mamm Genome 14:733–747

Hua J, Xing Y, Wu W, Xu C, Sun X, Yu S, Zhang Q (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 USA 100:2574–2579

Hua JP, Xing YZ, Xu CG, Sun XL, Yu SB, Zhang Q (2002) Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. Genetics 162:1885–1895

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Kawasaki S, Borchert C, Deyholos M, Wang H, Brazille S, Kawai K, Galbraith D, Bohnert HJ (2001) Gene expression profiles during the initial phase of salt stress in rice. Plant Cell 13:889–905

Kim BH, von Arnim AG (2006) The early dark-response in *Arabidopsis thaliana* revealed by cDNA microarray analysis. Plant Mol Biol 60:321–342

Knight JC (2004) Allele-specific gene expression uncovered. Trends Genet 20:113–116

Lan L, Chen W, Lai Y, Suo J, Kong Z, Li C, Lu Y, Zhang Y, Zhao X, Zhang X, Zhang Y, Han B, Cheng J, Xue Y (2004) Monitoring of gene expression profiles and isolation of candidate genes involved in pollination and fertilization in rice (*Oryza sativa* L.) with a 10K cDNA microarray. Plant Mol Biol 54:471–487

Leonhardt N, Kwak JM, Robert N, Waner D, Leonhardt G, Schroeder JI (2004) Microarray expression analyses of Arabidopsis guard cells and isolation of a recessive abscisic acid hypersensitive protein phosphatase 2C mutant. Plant Cell 16:596–615

Li JX, Yu SB, Xu CG, Tan YF, Gao YJ, Li XH, Zhang Q (2000) Analyzing quantitative trait loci for yield using a vegetatively replicated F2 population from a cross between the parents of an elite rice hybrid. Theor Appl Genet 101:248–254

Li ZK, Luo LJ, Mei HW, Wang DL, Shu QY, Tabien R, Zhong DB, Ying CS, Stansel JW, Khush GS, Paterson AH (2001) Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. Genetics 158:1737–1753

Lian X, Wang S, Zhang J, Feng Q, Zhang L, Fan D, Li X, Yuan D, Han B, Zhang Q (2006) Expression profiles of 10,422 genes at early stage of low nitrogen stress in rice assayed using a cDNA microarray. Plant Mol Biol 60:617–631

Lo HS, Wang Z, Hu Y, Yang HH, Gere S, Buetow KH, Lee MP (2003) Allelic variation in gene expression is common in the human genome. Genome Res 13:1855–1862

Luo LJ, Li ZK, Mei HW, Shu QY, Tabien R, Zhong DB, Ying CS, Stansel JW, Khush GS, Paterson AH (2001) Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. Genetics 158:1755–1771

Ma L, Gao Y, Qu L, Chen Z, Li J, Zhao H, Deng XW (2002) Genomic evidence for COP1 as a repressor of light-regulated gene expression and development in *Arabidopsis*. Plant Cell 14:2383–2398

Maleck K, Levine A, Eulgem T, Morgan A, Schmid J, Lawton KA, Dangl JL, Dietrich RA (2000) The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. Nat Genet 26:403–410

Marathe R, Guan Z, Anandalakshmi R, Zhao H, Dinesh-Kumar SP (2004) Study of *Arabidopsis thaliana* resistome in response to cucumber mosaic virus infection using whole genome microarray. Plant Mol Biol 55:501–520

Meyer RC, Torjek O, Becher M, Altmann T (2004) Heterosis of biomass production in *Arabidopsis*. Establishment during early development. Plant Physiol 134:1813–1823

Mitchell-Olds T (1995) Interval mapping of viability loci causing heterosis in *Arabidopsis*. Genetics 140:1105–1109

Park SJ, Huang Y, Ayoubi P (2006) Identification of expression profiles of sorghum genes in response to greenbug phloem-feeding using cDNA subtraction and microarray analysis. Planta 223:932–947

Pastinen T, Hudson TJ (2004) Cis-acting regulatory variation in the human genome. Science 306:647–650

Rabbani MA, Maruyama K, Abe H, Khan MA, Katsura K, Ito Y, Yoshiwara K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Monitoring expression profiles of rice genes under cold, drought, and high-salinity stresses and abscisic acid application using cDNA microarray and RNA gel-blot analyses. Plant Physiol 133:1755–1767

Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G, Linsley PS, Mao

M, Stoughton RB, Friend SH (2003) Genetics of gene expression surveyed in maize, mouse and man. Nature 422:297–302

Schaffer R, Landgraf J, Accerbi M, Simon V, Larson M, Wisman E (2001) Microarray analysis of diurnal and circadian-regulated genes in *Arabidopsis*. Plant Cell 13:113–123

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470

Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Somerville SC, Manners JM (2000) Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. Proc Natl Acad Sci USA 97:11655–11660

Seki M, Narusaka M, Ishida J, Nanjo T, Fujita M, Oono Y, Kamiya A, Nakajima M, Enju A, Sakurai T, Satou M, Akiyama K, Taji T, Yamaguchi-Shinozaki K, Carninci P, Kawai J, Hayashizaki Y, Shinozaki K (2002) Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. Plant J 31:279–292

Stolc V, Li L, Wang X, Li X, Su N, Tongprasit W, Han B, Xue Y, Li J, Snyder M, Gerstein M, Wang J, Deng XW (2005) A pilot study of transcription unit analysis in rice using oligonucleotide tiling-path microarray. Plant Mol Biol 59:137–149

Stoughton RB (2005) Applications of DNA microarrays in biology. Annu Rev Biochem 74:53–82

Stuber CW (1994) Heterosis in plant breeding. Plant Breed Rev 12:227–251

Stuber CW, Lincoln SE, Wolff DW, Helentjaris T, Lander ES (1992) Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823–839

Sun QX, Ni ZF, Liu ZY (1999) Differential gene expression between wheat hybrids and their parental inbreds in seedling leaves. Euphytica 106:11–17, 123

Syed NH, Chen ZJ (2005) Molecular marker genotypes, heterozygosity and genetic interactions explain heterosis in *Arabidopsis thaliana*. Heredity 94:295–304

Tepperman JM, Zhu T, Chang HS, Wang X, Quail PH (2001) Multiple transcription-factor genes are early targets of phytochrome A signaling. Proc Natl Acad Sci USA 98:9437–9442

Tian ZY, Dai JR (2003) Relationship between differential gene expression patterns in functional leaves of maize inbreds and hybrids at spikelet differentiation stage and heterosis. Acta Genetica Sinica 30:154–162

Wang R, Okamoto M, Xing X, Crawford NM (2003). Microarray analysis of the nitrate response in *Arabidopsis* roots and shoots reveals over 1,000 rapidly responding genes and new linkages to glucose, trehalose-6-phosphate, iron, and sulfate metabolism. Plant Physiol 132:556–567

Wang Z, Liang Y, Li C, Xu Y, Lan L, Zhao D, Chen C, Xu Z, Xue Y, Chong K (2005) Microarray analysis of gene expression involved in anther development in rice (*Oryza sativa* L.). Plant Mol Biol 58:721–737

Wu LM, Ni ZF, Meng FR, Lin Z, Sun QX (2003a) Cloning and characterization of leaf cDNAs that are differentially expressed between wheat hybrids and their parents. Mol Genet Genomics 270:281–286

Wu P, Ma L, Hou X, Wang M, Wu Y, Liu F, Deng XW (2003b) Phosphate starvation triggers distinct alterations of genome expression in *Arabidopsis* roots and leaves. Plant Physiol 132:1260–1271

Xiao J, Li J, Yuan L, Tanksley SD (1995) Dominance is the major genetic basis of heterosis in rice as revealed by QTL analysis using molecular markers. Genetics 140:745–754

Xing YZ, Tan YF, Hua JP, Sun XL, Xu CG, Zhang Q (2002) Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. Theor Appl Genet 105:248–257

Xiong LZ, Yang GP, Xu CG, Zhang Q, Saghai Maroof MA (1998) Relationships of differential gene expression in leaves with heterosis and heterozygosity in a rice diallel cross. Molecular Breeding 4:129–136

Xu SX, Xu XB (1984) Morphologic and anatomic analysis on rice: young panicle development. China Agricultural Press, Beijing (in Chinese)

Yang GX, Jan A, Shen SH, Yazaki J, Ishikawa M, Shimatani Z, Kishimoto N, Kikuchi S, Matsumoto H, Komatsu S (2004) Microarray analysis of brassinosteroids- and gibberellin-regulated gene expression in rice seedlings. Mol Genet Genomics 271:468–478

Yu SB, Li JX, Xu CG, Tan YF, Gao YJ, Li XH, Zhang Q, Maroof MA (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc Natl Acad Sci USA 94:9226–9231

Yuan LP (1998) Hybrid rice breeding in china. In: Virmani SS, Siddiq EA, Muralidharan K (eds) Advances in Hybrid Rice Technology. International Rice Research Institute, Los Banos, pp 27–33

Zhang J, Feng Q, Jin C, Qiu D, Zhang L, Xie K, Yuan D, Han B, Zhang Q, Wang S (2005) Features of the expressed sequences revealed by a large-scale analysis of ESTs from a normalized cDNA library of the elite indica rice cultivar Minghui 63. Plant J 42:772–780

Zhang Q, Gao YJ, Yang S, Ragab R, Saghai Maroof MA, Li ZB (1994) A diallel analysis of heterosis in elite hybrid rice based on RFLPs and microsatellites. Theor Appl Genet 89:185–192

Zhang Q, Zhou ZQ, Yang GP, Xu CG, Liu KD, Saghai Maroof MA (1996) Molecular marker heterozygosity and hybrid performance in indica and japonica rice. Theor Appl Genet 93:1218–1224

Zhong GV, Burns JK (2003) Profiling ethylene-regulated gene expression in *Arabidopsis thaliana* by microarray analysis. Plant Mol Biol 53:117–131