DOI: 10.1007/s11427-006-2031-0

Comparative analysis of gene expression at early seedling stage between a rice hybrid and its parents using a cDNA microarray of 9198 uni-sequences

HUANG Yi, LI Lihua, CHEN Ying, LI Xianghua, XU Caiguo, WANG Shiping & ZHANG Qifa

National Key Laboratory of Crop Genetic Improvement, National Center of Plant Gene Research (Wuhan), Huazhong Agricultural University, Wuhan 430070, China
Correspondence should be addressed to Zhang Qifa (email: qifazh@mail.hzau.edu.cn)

Received March 17, 2006; accepted June 20, 2006

Abstract Using a cDNA microarray consisting of 9198 expressed sequence tags, we surveyed the gene expression profiles in shoots and roots of a rice hybrid, Liangyoupei 9 and its parents Peiai 64s and 93-11 at 72 h after germination. A total of 8587 sequences had detectable signals in both shoots and roots of the three genotypes. A total of 1571 sequences exhibited significant (P<0.01) expression differences in shoots or roots among the three genotypes, of which 121 showed expression polymorphisms in both shoots and roots, and 870 revealed significant expression differences between the hybrid and one of the parents. The expression polymorphism of the sequences was associated with the functional categories of the sequences. They occurred more frequently in categories of carbohydrate, energy and lipid metabolisms and stress response than expected, while less frequently in categories of amino acid metabolism, transcription and translation regulation, and signal transduction. A total of 214 sequences exhibited significant (P<0.05) mid-parent heterosis in expression, of which 117 had homology to genes with known functions, assigned in the categories of basic metabolism, genetic information processing, cell growth and death, signal transduction, transportation and stress response. The results may provide useful information for exploring the relationship between gene expression polymorphism and phenotypic variation, and for characterizing the molecular mechanism of seedling development and heterosis in rice.

Keywords: ***Oryza sativa*** **shoots, roots, gene expression, microarray, heterosis.**

The completion of the rice genome sequencing project and optimization of various technological platforms have greatly enhanced rice functional genomics research. Such developments are also important for interpreting genetic variation and phenotypic diversity, as well as for understanding the molecular mechanism underlying heterosis[1−3]. Genes in hybrid are inherited from the parents，thus variation in regulations of the genes often leads to variation in the level of gene expression in hybrid, which in turn may alter the phenotype of the hybrids, thus causing heterosis[4]. Therefore differences in gene expression between hybrid and parents may provide important clues for revealing the molecular mechanisms of heterosis.

Germination is the first stage of the life cycle in plants. Investigation of gene expression in the germination stage may provide information for understanding mechanism of rice seedling development and pos-

sible heterosis at the early stage. There have been reports on difference of gene expression between hybrids and their parents. Using DDRT-PCR, Xiong *et al*.[5] and Sun *et al*.[6] investigated mRNA abundance and expression patterns of the hybrids and parents in rice and wheat, and found associations between heterosis and patterns of gene expression. In maize, Tian and Dai[7] observed an association between heterosis and inhibition of gene expression in the hybrids. However, due to technical limitations, the numbers of genes that could be investigated by these methods are limited.

Microarray technology is a powerful tool for analyzing genome-wide gene expression[8]. It has been applied to identifying pathways involved in the process of seed development[9], plant hormone regulation[10,11] and expression profiles of plant genes under environmental stresses[12−14]. Seedling is the starting point of rice life cycle, and also the beginning for building-up heterosis. Matsumura *et al*.[15] showed that most of the highly expressed genes in rice seedlings belonged to the category of housekeeping (genes encoding ribosomal proteins or proteins for metabolism and cell structure) in anaerobically treated rice seedlings. Bao *et al*.[16] surveyed transcriptomes of panicles, leaves and roots of a 'two-line' rice hybrid Liangyoupei 9 and its parents using the serial analysis of gene expression (SAGE) technique, and found that a large number of tags were significantly differentially expressed in hybrid, as compared to the parents, most of which were relevant to basic metabolisms and cell growth. However, there have not been reported studies on gene expression profiles at the early germination stage using the microarray technology.

In this study, we investigated gene expression in shoots and roots of the rice hybrid Liangyoupei 9 in comparison with its parents at 72 h after germination using a cDNA microarray containing 9198 unique sequences. It was expected that the expression data would provide useful information for understanding the molecular mechanism underlying rice seedling development as well as heterosis at the transcriptome level.

1 Materials and methods

1.1 Materials and growth conditions

The rice hybrid Liangyoupei 9 and its parent Peiai 64s (female) and 93-11 (male) were used in this study. Seeds of the parents and hybrid were soaked in tap water at 25℃ for 72 h and then kept in an incubator at 37℃ for 8 h to allow germination. Germinated seeds were planted in wet sand in plastic boxes for seedling development with two replications. Each box was covered with a piece of nylon net. At one-leaf stage, 30 seedlings per genotype were carefully transferred in a greenhouse into a plastic box (62.0 cm × 37.0 cm × 14.3 cm) containing the culture solution[17] and covered with a wood plate with bored holes (1.5 cm of diameter, one plant in each hole). The plants were cultured hydroponically, with the solution changed every three days. Two independent plantings were carried out, each with two replications.

1.2 Phenotype measurements

Five healthy seedlings of each genotype were randomly chosen from each replicate at 72 h after germination, 2-leaf, 3-leaf, 4-leaf and 5-leaf stages. Measurements were taken for the following characteristics: shoot height (cm), length of main root (cm), number of lateral roots, total length of lateral roots and leaf area (length×width, cm^2). In addition, dry weight of the shoots and roots at 4- and 5-leaf stages was also measured. To measure dry weight of shoots and roots, each seedling was put in a paper bag and baked in an oven at 120℃ for 2 h and kept at 80℃ for 30 h. The measurements were averaged over replications and plantings in the analysis. One-way ANOVA and the least significant difference (LSD) test were performed to assess statistical significance of differences of the measurements among the three genotypes. Mid-parent heterosis [= $(F_1\text{-}MP)/\times 100$, in which F_1 is the performance of the hybrid and MP is the average performance of the two parents] was calculated for each trait.

1.3 Sample collection and RNA preparation

Roots and shoots of the three genotypes were separately collected at 72 h after germination, immediately frozen in liquid nitrogen and kept at −70℃until RNA isolation. Total RNA was isolated from pooled tissue samples for each genotype using Trizol reagent (GIBCO/BRL) according to the manufacture's instruction (http://www.invitrogen.com). mRNA was isolated

from the total RNA of each genotype using the Oligo $dT_{(25)}$ magnetic Dynal beads (http://www.dynalbiotech.com/) (Oslo, Norway).

1.4 Microarray fabrication

A collection of unique cDNA sequences from a normalized whole-life-cycle library of Minghui 63[18] was used as the main source of the cDNAs for the chip preparation. Four subtractive cDNA libraries of Minghui 63 prepared using tissues harvested after low-nitrogen, low-phosphate, drought stress and striped stem borer feeding treatments, and a cDNA library of young panicle of Minghui 63 were also randomly sequenced for unique ESTs. In this way, a total of 9198 unique sequences were selected for cDNA microarray printing.

For printing the microarray, plasmids containing cDNA inserts were isolated and used as templates for PCR amplification in 96-well plates using T7 and Sp6 primers. The amplification products were cleaned by precipitation with iso-propanol and two washes with 70% ethanol, and resuspended in 40 μL of 50% dimethylsulfoxide (DMSO). The quality and quantity of the amplified cDNAs were checked by electrophoresis in 1% agarose gel. Only the cDNAs showing single band amplification products were selected and kept in 384-well microtiter plates at –20℃ for microarray preparation.

Purified cDNAs were arrayed from the 384-well microtiter plates onto poly-lysine-coated glass slides (Corning, USA) with a GMS 417 arrayer (Genetic MicroSystems, Worburn, Massachusetts, USA). A cDNA clone of the rice actin gene (GenBank Acc. No. EI077C02) was also arrayed 10 times with random positions on the slide as the positive control, and a clone of a porcine glyceraldehyde 3-phosphate dehydrogenase gene (GenBank Acc. No. AF017079) repeated 8 times per slide as the negative control. Post-arraying treatment of the slides was proceeded by placing at room temperature for 48 h, incubated at 80℃ for 4 h and kept in the dark at room temperature in a jar containing silica gel until use.

1.5 Preparation of hybridization probe

Reverse transcription was conducted in a volume of 40 μL containing 2.0 μg mRNA, 6.0 μg Oligo $dT_{(20)}$, 1×first-strand buffer (Life Technologies, Grand Island, New York, USA), 400 U Superscript II reverse transcriptase, 10 mmol/L DTT, 40 U RNase inhibitor (Promega, Madison, Wisconsin, USA), 500 μmol/L of dATP, dCTP and dGTP, 200 μmol/L of dTTP and aa-dUTP (Sigma, St. Louis, Missouri, USA). After incubation at 42℃ for 3 h, the reaction mixture was heated to 94℃ for 3 min to stop the reaction, and treated with RNase H to remove RNA. The reaction mixture was extracted with phenol/chloroform, purified with a Microcon YM-30 filter (Millipore, USA), and coupled to Cy5- and Cy3-dye (Amersham Pharmacia, USA) as previously described[19]. The probe was 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 denatured at 90℃ for 3 min and cooled to 4℃ for hybridization (the details of the protocol can be found at the website: http://redb.ncpgr.cn/protocols/).

1.6 Hybridization

The procedures for pre-hybridization and hybridization were as described previously[14]. We performed forward and reverse labeling for each hybridization. In the forward hybridization, the mRNA from the hybrid was labeled with Cy3 and the tissue from one of the parents with Cy5, and the labeled probes were mixed in equal amounts and hybridized with the cDNA chip. In the reverse hybridization, the mRNA from the parent was labeled with Cy5 and the hybrid tissue with Cy3.

1.7 Data processing and analysis

Slides were scanned with a GMS418 Array Scanner (Genetic MicroSystems, Worburn, MA) by two separate laser channels for Cy3 and Cy5 emissions. The details of the protocol can be found at the website (http://redb.ncpgr.cn/protocols/). The scanning parameters were set to make the sums of signal intensities of Cy3-dye and Cy5-dye nearly equal, and the sum of signal intensities of slide 1 equal to that of slide 2. Data were extracted from the image using the ImaGene 4.2 software (BioDiscovery, Los Angeles, CA).

The signal of each spot was subjected to spot filtering and normalization as follows. First, spots flagged "Bad" by ImaGene 4.2 software using the criteria adopted were excluded from the analysis. Second, only those spots that exhibited fluorescent intensity levels at least in one channel above two times the local background were included. Third, in each tissue, only sequences that had detectable signals in the two hybridizations of a probe pair were accepted.

A one-way ANOVA and LSD test were conducted to assess the statistical significance of the differences of gene expression levels among the three genotypes in a given tissue. Mid-parent heterosis of the gene expression level was calculated as: (F1-MP)/MP×100, in which F_1 is the expression level of the hybrid and *MP* is the average of the two parents. An *h*-statistic was devised to test statistical significance of mid-parent heterosis of gene expression: $h = [F_1 - (P_1 + P_2)/2] / \sqrt{V_{F_1} + (V_{P_1} + V_{P_2})/4}$, which is similar to the *t*-statistic with the degrees of freedom contributed by the three genotypes, F_1, P_1 and P_2. V_{F1}, V_{P1} and V_{P2} represent the variances of the hybrid and the two parents, respectively.

1.8 Northern blot analysis

Total RNA was prepared as described above, 15.0 μg total RNA of each genotype was denatured by mixing with an equal volume of formamide containing 0.05% bromophenol blue and 0.01% SybrGreen II. The RNA was separated on 1.5% agarose gel running in 1×MOPS buffer, then blotted onto Hybond-N^+ membranes (Amersham Pharmacia, USA) and fixed under UV light. Membranes were hybridized with ^{32}P-labeled probes prepared using individual cDNA clones. The hybridization was performed in Church buffer (7% SDS, 1% Bovine serum albumin, 1.0 mmol/L EDTA, 0.25 mol/L Na_2HPO_4, pH7.2) at 65℃ overnight, then the membrane was washed in 2×SSC+0.1%SDS at 65℃ for 30 min, and in 0.1× SSC+0.1%SDS at 65℃ for another 30 min. Image of individual gel was generated with PhosphorImager SI[20] (Molecular Dynamics, Sunnyvale, CA) and the signal intensity of each band was extracted using the Syngene software (Synoptics Ltd., UK).

2 Results

2.1 Heterosis of the hybrid at the seedling stage

Seedlings of the hybrid and parents at 72 h after germination are shown in Fig. 1, and the measurements of the characteristics taken at 5 stages are given in Table 1. The measurements of seedlings sampled at 72 h after germination, 2-leaf, 3-leaf, 4-leaf and 5-leaf stages showed that the hybrid was significantly higher than Peiai 64s in shoot height at all the stages sampled, and 93-11 was significantly higher than Peiai 64s in shoot height from the 3- to 5-leaf stages. The hybrid and 93-11 had significantly larger values than Peiai 64s in total length and number of roots from the 2- to 5-leaf stages.

The mid-parent heterosis of seedling height increased gradually with seedling development. Negative heterosis was detected for total length and number of roots at 72 h after germination, while the heterosis became positive from the 2- to 5-leaf stages. Additionally, there were various degrees of heterosis in dry weight of roots, shoots and biomass of the seedling at the 4- and 5-leaf stages.

2.2 General features of microarray data

The number of sequences, that had detectable signals as revealed by the microarray analysis using root and shoot tissues harvested at 72 h after germination,

Fig. 1. Phenotypes of Peiai 64s (left), Liangyoupei 9 (middle) and 93-11 (right) at 72 h after germination.

Table 1 Means and heterosis of seedling traits of the parents and hybrid [a)]

Genotypes	72 HAG [b)]	2-leaf stage	3-leaf stage	4-leaf stage	5-leaf stage
	Plant height (cm)				
Peiai 64s	3.1±0.3[b]	18.3±0.5[b]	18.5±1.1[b]	23.9±0.8[b]	25.1±0.9[b]
Liangyoupei 9	3.8±0.3[a]	18.7±0.7[a]	21.3±1.4[a]	27.5±1.2[a]	30.0±1.0[a]
93-11	3.4±0.3[b]	18.0±1.6[b]	22.2±1.3[a]	27.0±1.8[a]	29.8±1.3[a]
Mid-parent heterosis (%)	15.0	2.8	5.0	8.1	9.2
	Number of lateral roots				
Peiai 64s	1.9±0.4[b]	5.7±1.1[b]	8.4±1.3[b]	18.8±3.0[b]	25.5±4.4[b]
Liangyoupei 9	2.1±0.9[b]	7.9±1.4[a]	12.6±2.3[a]	23.5±2.2[a]	36.0±3.7[a]
93-11	3.0±1.1[a]	8.0±3.2[a]	14.9±2.7[a]	23.8±3.1[a]	38.1±5.2[a]
Mid-parent heterosis (%)	–13.4	15.3	8.2	10.3	13.2
	Total length of lateral roots (cm)				
Peiai 64s	0.8±0.3[b]	26.0±8.0[b]	53.6±11.1[b]	146.7±16.3[b]	308.8±11.2[b]
Liangyoupei 9	1.1±0.5[b]	32.8±6.4[a]	79.1±11.4[a]	191.0±17.5[a]	386.9±17.5[a]
93-11	2.4±1.2[a]	35.6±2.2[a]	95.9±13.5[a]	195.1±17.4[a]	418.7±17.8[a]
Mid-parent heterosis (%)	–34.5	6.6	5.8	11.8	6.4

a) a, b showing significant difference at probability level 0.05 between genotypes as determined by the LSD test.
b) HAG, hours after germination.

varied slightly among the three genotypes (Table 2). About 97.3% (8947) and 94.3% (8675) of the sequences were detected in shoots and roots respectively in all three genotypes, and 9035 sequences had reproducible signals in the two hybridization replicates among three genotypes in at least one of the two tissue samples, indicating that majority of the sequences were expressed in both shoots and roots.

Correlations between the dye-swap hybridizations ranged from 0.95 to 0.98. We also selected 7 sequences, representing several expression patterns in the hybrid and parents, to investigate the correspondence between the signal intensity in the microarray and Northern hybridization using rRNA as the reference (Fig. 2). The signal intensity of each Northern blot band was quantified with Syngene software (Synoptics Ltd.). The correlation between microarray and Northern blot of 7 sequences was 0.62, indicating that the data generated in the microarray analysis is reliable.

2.3 Gene expression polymorphism

Results of ANOVA and LSD test revealed a total of 779 and 913 sequences that showed significant ($P<$ 0.01) differences in expression levels among the three genotypes, of which 121 sequences showed significant differences in both shoots and roots. Thus, significant expression differences were detected in either roots or shoots in a total of 1571 sequences (for details of the sequences, see Table S1 in http://redb.ncpgr.cn/ mged/ lyp/). We referred to the significant difference at the expression level so identified as expression polymorphisms.

Table 2 Number of sequences expressed in shoots and roots of three genotypes at 72 h after germination

Genotypes	Roots	Shoots	Common
Peiai 64s	9010	8993	8924
Liangyoupei 9	8675	8947	8587
93-11	8710	9057	8652
Common	8675	8947	8587

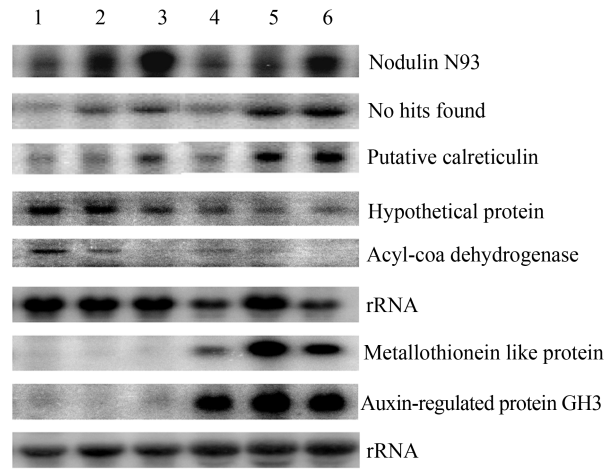

Fig. 2. Seven representative differentially expressed sequences were selected as probes for RNA-gel blot. 1—3, 72-HAG-stage shoots; 4—6, 72-HAG-stage roots, from left to right: 93-11 (left), Liangyoupei 9 (middle) and Peiai 64s (right) in each triad.

Among the 779 sequences showing expression polymorphisms in shoots, 286 showed significant expression differences only between the two parents, 493 showed significantly different expression levels be-

tween the hybrid and one or both parents. Based on the patterns of the expression in the hybrid and the parents, the 493 sequences could be classified into three groups (Table 3): (1) significant expression differences between the hybrid and both parents (76 sequences), (2) significant expression differences between the hybrid and Peiai 64s (163 sequences), and (3) significant expression differences between the hybrid and 93-11 (254 sequences). There were 58, 121, and 221 sequences showing significant expression differences between the two parents in the above three groups.

Table 3 Number of sequences that showed significant differences between hybrid and parents detected by LSD tests at $P<0.01$ probability level

Patterns [a)]	Shoots	Roots	Common
I	76 (15.4%)	111 (27.1%)	0
II	163 (33.1%)	205 (50.0%)	13
III	254 (51.5%)	94 (22.9%)	2
Total	493	410	15

a) I, significant difference between the hybrid and both parents; II, significant difference between the hybrid and Peiai 64s; III, significant difference between the hybrid and 93-11.

Among the 913 sequences showing expression polymorphisms in roots at 72 h after germination, 503 sequences displayed significant differences in expression levels only between the parents, and the remaining 410 sequences revealed significant expression differences between the hybrid and one or both parents. Similarly, the 410 sequences could be divided into three groups according to the expression patterns: (1) significant expression differences between the hybrid and both parents (111 sequences), (2) significant expression differences between the hybrid and Peiai 64s (205 sequences), and (3) significant expression differences between the hybrid and 93-11 (94 sequences) (Table 3). There were 104, 186, and 86 sequences showing significant expression differences between the two parents in the three groups.

In shoots, the number of sequences showing significant expression differences between the hybrid and 93-11 is greater than that between the hybrid and Peiai 64s. The reverse was the case in roots, in which the number of sequences showing significant expression differences between the hybrid and Peiai 64s is greater than that between the hybrid and 93-11.

Taken together, a total of 870 sequences showed significant expression differences between the hybrid and parents in shoots and roots, of which 187 sequences showed significant expression differences between the hybrid and both parents, 683 sequences had expression differences between the hybrid and one of the parents.

2.4 Functional classification of the differentially expressed sequences

Functional classification of the differentially expressed sequences was performed by homology search using the BLAST program[21]. Putative functions for cDNA sequences with *E*-value $\leqslant 10^{-5}$ or BLASTn scores $\geqslant 100$ were assigned on the basis of Gene Ontology (GO) and Non-redundancy (NR) databases of NCBI. The biochemical and physiological pathways were classified according to the KEGG database (http:// www.genome.jp/kegg/).

The 870 sequences showing expression polymorphisms detected above could be classified into 6 major functional categories: (1) metabolism, (2) genetic information processing, (3) cellular information processing, (4) environment information processing, (5) unclassified, and (6) unknown function (Table 4). Sequences were further classified into several sub-categories in each functional category. The largest category consisted of 434 sequences with unknown function (including unclassified). In the remaining 436 sequences, 160 sequences were placed in the category of metabolism, 120 in genetic information processing, 109 in environment information processing, and 36 in cellular information processing. The same procedure was applied to the 493 and 410 sequences showing expression polymorphisms in shoots and roots, which resulted in similar distributions of the sequences (Table 4).

To test whether the occurrence of sequences with expression polymorphisms was differentially associated with any of the functional categories, a χ^2-test was performed to evaluate goodness-of-fit between the observed numbers of sequences and the expectations based on the frequencies of the total 8587 sequences that fell in the 16 sub-categories of known functions in shoots and roots (Table 4). The results showed that there were significant discrepancies between the ex-

Table 4 Functional categories of the sequences showing polymorphic expression among the three genotypes

Functional categories	Shoots		Roots		Total	
	observed	expected [a)]	observed	expected	observed	Expected[a)]
Metabolism	86	81	79	68	158	144
Amino acid metabolism	8	14	8	12	16	26
Carbohydrate metabolism	33	32	38	26	69	56
Energy metabolism	19	9	7	7	25	16
Lipid metabolism	19	12	11	10	28	21
Nucleotide metabolism	5	7	10	5	15	12
Metabolism of others	1	5	1	4	1	8
Secondary metabolism	1	3	4	2	4	5
Genetic information processing	83	81	46	67	120	142
DNA replication and repair	8	8	6	6	12	13
Transcription factors	30	27	12	22	40	48
Transcription	17	12	10	10	25	22
Translation	14	17	6	14	18	29
Protein degradation	14	17	12	14	25	30
Cellular processing	18	17	19	14	37	30
Environment information processing	67	63	49	53	110	112
Plant defense	16	9	15	8	28	16
Signal transduction	20	27	17	22	36	47
Transport	31	27	17	23	46	48
Unclassified	7	9	4	7	11	15
Unknown function	232	242	213	201	434	427
Hypothetical	83	97	88	81	167	172
No significant homology	107	99	79	82	181	174
No hits found	42	46	46	38	86	82
Total	493		410		870	

a) Expected numbers based on the frequencies of the 8587 sequences that occurred in the various functional categories.

pected and observed numbers in both shoots and roots. In shoots, the observed numbers of sequences in the sub-categories of energy and lipid metabolisms and stress response were significantly larger than expected, whereas the observed number of sequences in metabolism of others was smaller than expected (χ^2=34.29, P=0.0031). In roots, the observed numbers of sequences in the sub-categories of carbohydrate, nucleonic acid metabolisms and stress response were greater than expected, while the reverse was the case in the sub-categories of transcription and translation (χ^2=36.24, P=0.0016). Taking roots and shoots together, there were highly significant discrepancies between observed and expected numbers in the various sub-categories (χ^2=41.45, P=0.00027). The observed numbers of sequences in the sub-categories of carbohydrate, energy and stress response were in excess, whereas those in amino acid metabolism, metabolism of others and translation regulation were in deficiency.

2.5 Heterosis of gene expression

To identify significant mid-parent heterosis of gene expression, an h-statistic analysis at P<0.05 probability level was performed with the normalized signal intensity for the 870 sequences with detectable expression polymorphisms between the hybrid and parents. A total of 214 sequences showed significant expression heterosis in shoots or roots (Table 5), of which 66 and 161 sequences were identified as showing significant heterosis in roots and shoots respectively, and 13 sequences showed heterosis in both roots and shoots.

There were obvious differences in roots and shoots in the numbers of heterotic expression sequences and directions of mid-parent heterosis (Table 5). The number of heterotically expressed sequences in shoots (161) was about 2.5 times of the number in roots (66). The number of sequences with positive heterosis (83) in shoots was nearly equal to the ones showing negative heterosis (78), and about 44.5% (69) of the se-

Table 5 Number of sequences showing various degrees of mid-parent heterosis in the two tissues at 72 h after germination

Range of heterosis (%)	Number of sequences	
	roots	shoots
−100—−50	10	2
−50—−25	17	9
−25—0	19	67
0—25	14	14
25—50	5	42
50—100	1	27
Total	66	161

quences showed more than 25% positive heterosis. Whereas in roots, the number of sequences showing positive heterosis (20) was less than the ones showing negative heterosis (69), of which 42.9% (27) of the sequences showed less than −25% mid-parent heterosis.

2.6 Patterns and functional categories of heterotically expressed sequences

Based on the values and directions of mid-parent heterosis displayed by the 214 heterotically expressed sequences, as listed in Table 6, and Table S2 in http://redb.ncpgr.cn/mged/lyp/, four patterns could be identified: (1) positive heterosis in both roots and shoots, (2) positive heterosis in shoots but negative heterosis in roots, (3) positive heterosis in roots but negative heterosis in shoots, (4) negative heterosis in both roots and shoots (Table 6 and Table S2 in http://redb.ncpgr.cn/mged/lyp/).

Table 6 Pattern and number of heterotically expressed sequences in shoots and roots at 72 h after germination

Patterns[a)]	Roots	Shoots	Common	Total
I	9	53	2	64
II	7	26	2	35
III	9	29	0	38
IV	28	40	9	77
Total	53	148	13	214

a) I, positive heterosis both in roots and shoots; II, positive heterosis in shoots and negative heterosis in roots; III, negative heterosis in shoots and positive heterosis in roots; IV, negative heterosis both in roots and shoots.

The first pattern involved 64 sequences, of which 35 showed homology to genes of known function ($E \leqslant 10^{-5}$). Of the 35 sequences, 30 and 4 showed heterotic expression in shoots and roots respectively, and one, having homology to EF-hand Ca^{2+}-binding protein CCD1 (BAD07944.1), showed significant positive heterosis in both shoots and roots. Eleven of the 30 sequences showing heterosis in shoots were relevant to metabolisms, including 4 in energy, 4 in lipid, 2 in carbohydrate and 1 in amino acid metabolisms. Nine of the 30 sequences had homology to genes involved in genetic information processing, 8 of which were relevant to transcription regulation including 7 transcription factors, and 1 involved in translation. Six of the 30 sequences had similarity to genes in environment information processing including 2 in signal transduction, 3 in transportation and 1 in plant defense. The remaining 4 sequences were involved in cellular information processing. Among the 4 sequences heterotically expressed in roots, one was involved in plant defense, one in signal transduction, and the remaining 2 encode products playing roles in protein degradation respectively.

The second pattern was displayed by 35 sequences, of which 15 had homology ($E \leqslant 10^{-5}$) to genes of known functions. Eleven of the 15 sequences were heterotically expressed in shoots, 3 were heterotically expressed in roots, and 1, encoding flavanone 3-hydroxylase (AAN74830.1), had significant heterotic expression in both shoots and roots. Among the 11 sequences heterotically expressed in shoots, 4 were relevant to transportation, 3 encode proteins playing roles in transcription regulation, 2 had homology to genes in carbohydrate and lipid metabolisms, and the remaining 2 sequences were related to plant defense and signal transduction respectively. One of the 3 sequences showing significant negative heterosis in roots had homology to a gene in carbohydrate metabolism, the other 2 sequences were involved in plant defense and protein degradation.

Among the 38 sequences exhibiting the third pattern, 21 had homology ($E \leqslant 10^{-5}$) to genes with known functions. Sixteen of the 21 sequences were heterotically expressed in shoots, of which 6 were relevant to metabolisms including one in carbohydrate, 2 in energy, 2 in lipid and one in nucleic acid metabolism. Seven of the 16 sequences were involved in environment information processing, of which 2 were relevant to transportation, 2 had homology to genes in signal transduction, and the other 3 had roles in plant defense. Two of the 16 sequences played roles in transcription

regulation. The remaining one sequence was related to cell cycle. Of the 5 sequences showing expression heterosis in roots, 2 encode transporting proteins, 2 had homology to genes in amino acid and nucleic acid metabolisms, and one was involved in cellular information processing.

The fourth pattern was displayed by 77 sequences, of which 47 had homology ($E \leqslant 10^{-5}$) to genes with known functions. Seven of the 47 sequences showed negative heterotic expression in shoots and roots respectively, 20 were negatively heterotically expressed in shoots and another 20 in roots. Three of the 7 sequences had homology to plant defense, 2 had similarity to transcription factors, and the remaining 2 encode components of transportation and lipid metabolism respectively. Of the 20 sequences heterotically expressed in shoots, 7 were involved in metabolisms, 7 in genetic information processing, 4 in environment information processing, and 2 in cellular information processing. Among the 20 sequences heterotically expressed in roots, 13 were relevant to metabolisms, 3 were involved in stress response, 2 encode proteins which play roles in protein degradation, and the remaining 2 were relevant to cell growth and transportation.

Together, 6.1% (13) of the 214 heterotically expressed sequences showed significant heterosis both in shoots and roots, and 93.9% (201) of the sequences showed expression heterosis only in one of the two tissues. About 65.9% (141) of the sequences exhibited expression heterosis in the same directions in shoots and roots, while the remaining 34.1% (73) of the sequences showed opposite directions of expression heterosis in shoots and roots.

3 Discussion

In this study, we investigated gene expression profiles in shoots and roots in a rice hybrid and its parents at 72 h after germination using a cDNA microarray containing 9198 unique expressed sequences of rice. It was shown that about 18.3% (1571) of the sequences exhibited expression polymorphisms among the three genotypes at least in one of the two tissues, 10.1% (870) of the sequences exhibited significant expression differences between the hybrid and parents, and 2.5% (214) of the sequences displayed significant expression heterosis.

Gene expression polymorphisms, also referred to as allele-specific gene expression[22] that occurs frequently in nature, have been studied recently in maize[23], animals[24] and humans[25]. In this study, about 8.0% (683) of the sequences showed significant differences in expression levels between the hybrid and one of the parents, and 2.2% (187) of the sequences exhibited significant differences between the hybrid and both parents. The expression levels appeared as partial or full dominance for 7.1% of the sequences, and overdominance or under-dominance for 3.0% (259) of the sequences. These results differ substantially from a previous study with the same genotypes using a SAGE technique by Bao *et al.*, who reported that most of the genes that were differentially expressed in panicles, leaves and roots of hybrid compared with the parents were up-regulated, and only a very small portion of these genes were down-regulated[16]. It is not clear whether the difference in the results of these two studies was due to difference in tissues or the methods used in the two studies.

Our results also showed that the distribution of the sequences showing expression polymorphisms between the hybrid and one or both of the parents was associated with the functional categories. In general, sequences in the sub-categories of carbohydrate metabolism, energy metabolism and stress response occurred more frequently than expected, whereas those in the sub-categories of amino acid metabolism, translation regulation and metabolism occurred less frequently than expected. The results indicated that the hybrid seemed to differ from the parents in activities in a number of biochemical and physiological processes at the seedling stage.

Further analysis identified 214 sequences as showing significant mid-parent heterosis in expression, involving genes that participate in carbohydrate, amino acid, energy and lipid metabolisms, transcription and translation regulation, protein degradation, cell growth and death, transportation, and biotic or abiotic stress responses. It is interesting to note that a sequence encoding the CIP7 protein shows strong positive heterosis in shoots. CIP7 plays a role in mediating the light activation of gene expression as a tran-

scription factor and may serve as one of the targets for downstream COP1 that regulates the expression levels of other light activating genes[26], thus triggering systematically the expression of other genes related to photomorphogenesis. Thus, heterotic expression of CIP7 may accelerate the developing rate of photosynthetic system of hybrid. Similarly, sequences encoding chlorophyll a/b binding proteins and Calvin cycle elements like ribulose 1,5-bisphosphate carboxylase large subunit relevant to CO_2 assimilation also showed positive heterotic expression in shoots. It indicates that photosynthesis in hybrid is more robust, leading to more rapid accumulation of biomass. Additionally, some sequences encoding transcription factors such as the NAC2 protein showed positive heterosis in shoots and roots of hybrid. NAC2 protein is involved in formation of the shoot apical meristem, floral organs and lateral shoots, as well as in plant hormonal control and defense[27]. Heterotic expression of NAC2 indicates that cell division and organization may be accelerated in the hybrid compared with the parents. It is interesting to note that some sequences showing homology to genes that are important elements of glycolysis and stress response exhibited significant negative heterosis in shoots and roots. This seems to indicate that hybrid has strong stress tolerance and high efficiency on energy utilization, thus ensuring the energy supply in growth and development of the seedlings.

This study identified a large number of sequences as showing significant expression polymorphisms among the three genotypes at 72 h after germination and also a large number of genes as showing heterotic expression in the hybrid compared with the parents. Functional classification of these genes provided important information for understanding the processes underlying seedling growth and development, as well as for the causes of heterosis. It should be noted that the 9198 possible uni-genes assembled in the microarray employed in this study made up approximately 17% of the predicted genes in the rice genome (http://www.tigr.org/), which would not be able to provide complete information for the genes expressed at this stage. Future studies should employ the whole genome microarrays to explore the gene expression profiles of multiple developmental stages of even the whole life cycle of rice. Based on genome sequence information and mutant databases of rice, the functional roles of these genes could be unveiled systematically using multiple molecular tools including both forward and backward genetic methods. In this way, the molecular mechanisms underlying rice development processes and heterosis could be characterized at the transcription level.

Acknowledgements We thank Profs. Sibin Yu, Lizhong Xiong and Qian Ma for comments and suggestions, and Dr. Mengliang Cao for providing the seeds. This work was supported by the National Program on the Development of Basic Research (Grant No. 2001CB1088), the National Special Key Project of Functional Genomics and Biochips (Grant No.2002AA2Z1002), and the National Natural Science Foundation of China (Grant No. 30321005).

References

1 Yu J, Hu S, Wang J, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science, 2002, 296: 79－92
2 Goff S A, Ricke D, Lan T H, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science, 2002, 296: 92－100
3 Matsuoka M. New directions of post-rice genome sequencing. Plant Cell Physiol, 2005, 46: 1－2
4 Doebley J, Lukens L. Transcriptional regulators and the evolution of plant form. Plant Cell, 1998, 10: 1075－1082
5 Xiong L Z, Yang G P, Xu C G, et al. Relationships of differential gene expression in leaves with heterosis and heterozygosity in a rice diallel cross. Mol Breeding, 1998, 4: 129－136
6 Sun Q X, Ni Z F, Liu Z Y. Differential gene expression between wheat hybrids and their parental inbreds in seedling leaves. Euphytica, 1999, 106: 11－17, 123
7 Tian, Z Y, Dai, J R. Relationship between differential gene expression patterns in functional leaves of maize inbreds & hybrids at spikelet differentiation stage and heterosis. Acta Genet Sin, 2003, 30: 154－162
8 Schena M, Shalon D, Davis R W, et al. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science, 1995, 270: 467－470
9 Girke T, Todd J, Ruuska S, et al. Microarray analysis of developing Arabidopsis seeds. Plant Physiol, 2000, 124: 1570－1581
10 Yang G X, Jan A, Shen S H, et al. Microarray analysis of brassinosteroids- and gibberellin-regulated gene expression in rice seedlings. Mol Genet Genomics, 2004, 271: 468－478
11 Bao F, Hu Y X, Li J Y. Identification of auxin responsive genes in Arabidopsis by cDNA array. Chin Sci Bull, 2002, 47: 548－552
12 Wu P, Ma L, Hou X, et al. Phosphate starvation triggers distinct alterations of genome expression in Arabidopsis roots and

leaves. Plant Physiol, 2003, 132: 1260－1271

13 Liu R, Zhao J, Xiao Y, et al. Identification of prior candidate genes for Sclerotinia local resistance in *Brassica napus* using Arabidopsis cDNA microarray and Brassica-Arabidopsis comparative mapping. Sci China Ser C-Life Sci, 2005, 48: 460－470

14 Lian X, Wang S, Zhang J, et al. Expression profiles of 10,422 genes at early stage of low nitrogen stress in rice assayed using a cDNA microarray. Plant Mol Biol, 2006, 60: 617－631

15 Matsumura H, Nirasawa S, Terauchi R. Technical advance: transcript profiling in rice (*Oryza sativa* L.) seedlings using serial analysis of gene expression (SAGE). Plant J, 1999, 20: 719－726

16 Bao J, Lee S, Chen C, et al. Serial analysis of gene expression study of a hybrid rice strain (LYP9) and its parental cultivars. Plant Physiol, 2005, 138: 1216－1231

17 Yoshida S, Forno D A, Cook J H, et al. Laboratory Manual for Physiological Studies of Rice. Manila: International Rice Research Institute, 1976. 61－67

18 Chu Z H, Peng K M, Zhang L D, et al. Construction and characterization of a normalized whole-life-cycle cDNA library of rice. Chin Sci Bull, 2003, 48: 229－235

19 Ma L, Li J, Qu L, et al. Light control of Arabidopsis development entails coordinated regulation of genome expression and cellular pathways. Plant Cell, 2001, 13: 2589－2607

20 Potokina E, Sreenivasulu N, Altschmied L, et al. Differential gene expression during seed germination in barley (*Hordeum vulgare* L.). Funct Integr Genomics, 2002, 2: 28－39

21 Altschul S F, Gish W, Miller W, et al. Basic local alignment search tool. J Mol Biol, 1990, 215: 403－410

22 Knight J C. Allele-specific gene expression uncovered. Trends Genet, 2004, 20: 113－116

23 Guo M, Rupe M A, Zinselmeier C, et al. Allelic variation of gene expression in maize hybrids. Plant Cell, 2004, 16: 1707－1716

24 Cowles C R, Hirschhorn J N, Altshuler D, et al. Detection of regulatory variation in mouse genes. Nat Genet, 2002, 32: 432－437

25 Lo H S, Wang Z, Hu Y, et al. Allelic variation in gene expression is common in the human genome. Genome Res, 2003, 13: 1855－1862

26 Yamamoto Y Y, Matsui M, Ang L-H, et al. Role of a COP1 interactive protein in mediating light-regulated gene expression in Arabidopsis. Plant Cell, 1998, 10: 1083－1094

27 He X, Mu R, Cao W, et al. AtNAC2, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development. Plant J, 2005, 44: 903－916