Molecular Plant • Volume 1 • Number 5 • Pages 720–731 • September 2008 RESEARCH ARTICLE

A Genome-Wide Transcription Analysis Reveals a Close Correlation of Promoter INDEL Polymorphism and Heterotic Gene Expression in Rice Hybrids

Hui-Yong Zhang[a,2], Hang He[a,b,2], Liang-Bi Chen[c], Lei Li[d,e], Man-Zhong Liang[c], Xiang-Feng Wang[a,e], Xi-Gang Liu[a], Guang-Ming He[a], Run-Sheng Chen[b], Li-Geng Ma[a] and **Xing Wang Deng[a,d,e,1]**

a National Institute of Biological Sciences, Zhongguancun Life Science Park, Beijing 102206, China
b Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Beijing 100101, China
c Department of Botany, College of Life Sciences, Hunan Normal University, Changsha 410081, China
d Peking–Yale Joint Research Center of Plant Molecular Genetics and Agrobiotechnology, College of Life Sciences, Peking University, Beijing 100871, China
e Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, Connecticut 06520, USA

ABSTRACT Heterosis, or hybrid vigor, refers to the phenomenon in which hybrid progeny of two inbred varieties exhibits enhanced growth or agronomic performance. Although a century-long history of research has generated several hypotheses regarding the genetic basis of heterosis, the molecular mechanisms underlying heterosis and heterotic gene expression remain elusive. Here, we report a genome-wide gene expression analysis of two heterotic crosses in rice, taking advantage of its fully sequenced genomes. Approximately 7–9% of the genes were differentially expressed in the seedling shoots from two sets of heterotic crosses, including many transcription factor genes, and exhibited multiple modes of gene action. Comparison of the putative promoter regions of the ortholog genes between inbred parents revealed extensive sequence variation, particularly small insertions/deletions (INDELs), many of which result in the formation/disruption of putative *cis*-regulatory elements. Together, these results suggest that a combinatorial interplay between expression of transcription factors and polymorphic promoter *cis*-regulatory elements in the hybrids is one plausible molecular mechanism underlying heterotic gene action and thus heterosis in rice.

INTRODUCTION

Heterosis, or hybrid vigor, refers to the phenomenon in which progeny of two inbred varieties exhibits enhanced agronomic performance such as biomass production, growth rate, and fertility relative to both parents (Shull, 1908). Heterosis has been widely exploited in the breeding of maize, rice, and other crops, which is important to yield growth worldwide. For example, it is estimated that hybrid rice technology for large-scale production has a yield advantage of 15–20% over the elite inbred varieties.

The classical quantitative genetic explanations for heterosis center on a few concepts, including dominance (Davenport, 1908), over-dominance (Shull, 1908; East, 1908), and epistasis (Yu et al., 1997; Li et al., 2001). The dominance model posits that each of the inbred lines contains slightly deleterious alleles that reduce their fitness. The hybrid will benefit from the complementation of these deleterious alleles and will display a superior phenotype. The over-dominance model suggests that the heterozygous combination of alleles at a given locus is phenotypically superior to either of the homozygous combinations for that locus, thereby resulting in a superior hybrid. Both the dominance and over-dominance hypotheses are based on single locus theory, and they may be inadequate to address the molecular mechanism for heterosis (Birchler et al., 2003). Epistasis is classically defined as interactions of superior alleles at different loci from two parents, and the effects may show additivity, dominance or over-dominance (Yu et al., 1997; Li et al., 2001).

Genetic analyses have implied the involvement of multiple genetic loci in bringing about heterosis (Li et al., 2001; Stuber

[1] To whom correspondence should be addressed. E-mail xingwang.deng@yale.edu, fax 1-203-432-5726, tel. 1-203-432-8908.

[2] These authors contributed equally to this work.

doi: 10.1093/mp/ssn022, Advance Access publication 5 August 2008
Received 14 February 2008; accepted 25 March 2008

et al., 1992; Xiao et al., 1995; Hua et al., 2002, 2003; Lu et al., 2003). However, comprehensive assessment of the expression of the loci contributing to heterosis is largely not available due to the limited resolution of classic genetic analysis in tracking multiple loci (Birchler et al., 2003; Lippman and Zamir, 2007). Advances in DNA microarray-based analysis have made it possible to examine transcription of all individual genes in the genome in inbred parents and their F_1 hybrid simultaneously. It has been widely adopted for analyzing the global gene expression of a number of organisms (Stoughton, 2005). Recently, several studies have analyzed the association of gene expression and heterosis in maize, rice, and *Arabidopsis* using microarrays (Swanson-Wagner et al., 2006; Guo et al., 2003, 2006; Stupar and Springer, 2006; Meyer et al., 2007; Vuylsteke et al., 2005; Huang et al., 2006). Multiple modes of gene action, including additivity, high- and low-parent dominance, overdominance, and under-dominance, were found in those studies. However, no consensus as to how those differential expression patterns arose and how they relate to heterosis in the organism has been examined.

Rice is one of the most important stable food crops and an excellent model for grass species. Further, the availability of the complete sequences of the two rice genomes (Goff et al., 2002; Yu et al., 2002, 2005) and the readily traceable phenotypes resulting from heterosis make rice an ideal system in which to study the molecular events accompanying heterosis. As a first step to understanding the molecular basis of heterosis, we used a whole genome microarray to examine gene expression patterns in two heterotic crosses of rice. Analysis of the promoter regions of the differentially expressed genes indicated that INDEL polymorphism is an important contributor to heterotic gene action.

RESULTS

Characterization of the Two Heterotic Crosses Used in this Study

We used a rice whole genome microarray to examine gene expression patterns in two heterotic crosses (see Methods). One of the crosses involves the super hybrid rice commonly planted in China—*Liang-You-Pei-Jiu* (*LYP9*), an F_1 hybrid derived from crossing of the inbred paternal line *93-11* (*Oryza sativa* L. ssp. *Indica*) and maternal line *Pei-Ai* 64s (*PA64s*, with a mixed genetic background of *indica*, *japonica*, and *javanica*). The F_1 hybrid shows taller, stronger tillering activity and higher yield than both parents (Figure 1A, top). The other cross (Figure 1A, bottom) involves the inbred paternal line *93-11* and maternal line *Nipponbare* (*Oryza sativa* L. ssp. *japonica*); both

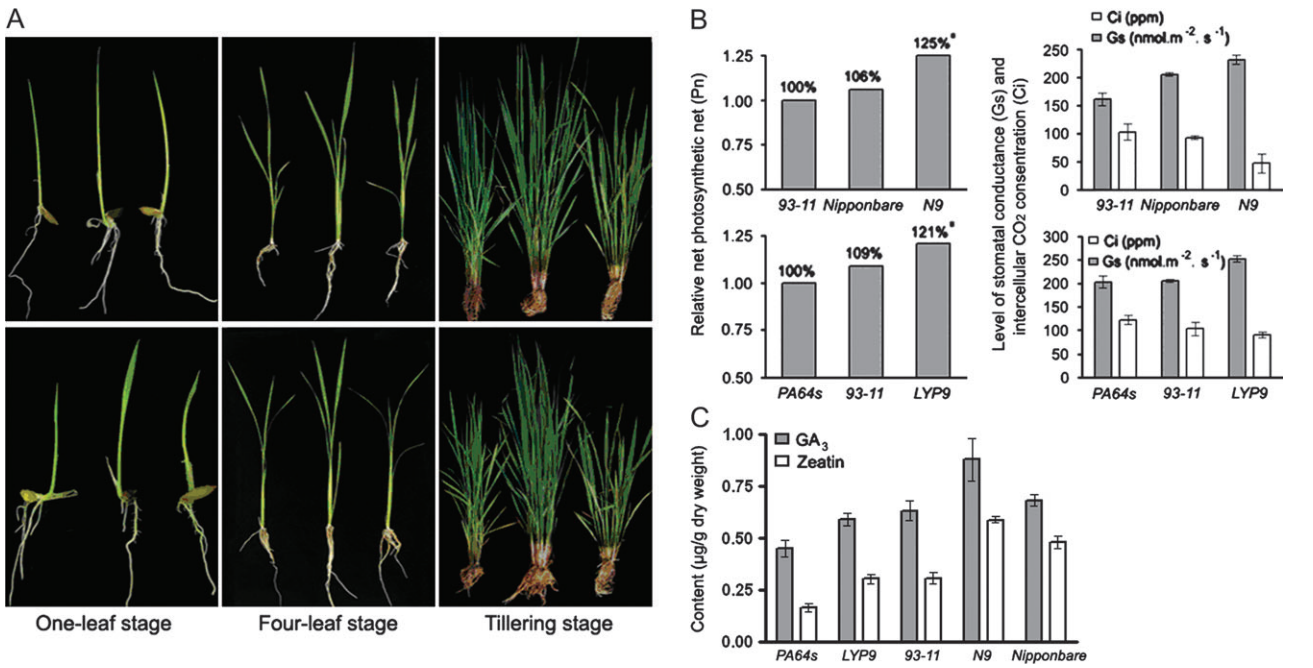


Figure 1. Heterosis in Two Rice Heterotic Crosses.

(A) Upper panel is the cross between *PA64s* (♀) and *93-11* (♂), and the lower panel is the cross between *Nipponbare* (♀) and *93-11* (♂). Left, maternal line; right, paternal line; middle, F_1 hybrid. Rice seedlings at the four-leaf stage are used for microarray analysis.

(B) Comparison of net photosynthetic net (Pn) among inbred parents and their F_1 hybrids. Star indicates that the Pn levels have significant difference between F_1 hybrids and inbred parents ($p < 0.05$). The net photosynthetic rate is positively correlated with stomatal conductance (Gs) and negatively correlated with intercellular CO_2 concentration (Ci).

(C) Measurement of GA3 and zeatin content in the seedling shoots at the four-leaf stage. *N9* represents F_1 hybrid of *Nipponbare* × *93-11* cross; *LYP9*, the F_1 hybrid of *PA64s* × *93-11* cross.

strains have their genome sequenced. The F_1 hybrid of this cross displays a strong hybrid vigor and higher biomass production than both parents. Both F_1 hybrids show significantly increased net photosynthetic rate (Pn) (Figure 1B) and GA_3 and zeatin content (Figure 1C) compared with the respective inbred parents. We selected four-leaf stage seedlings as the target for microarray analyses, as heterosis at this stage is already evident for both crosses (Supplemental Table 1).

Identification and Characterization of Differentially Expressed Genes

The processed microarray data was used to determine differential gene expression between inbred parents or between parental lines and their F_1 hybrids (Figure 2A and 2B). A total of 3488 (9.4%) genes were identified with significant differential expression ($p < 0.05$) in the *PA64s* × *93-11* cross. Among these genes, 66% (2317 of 3488) exhibited an expression pattern that was not distinguishable from additivity (average expression level of the two parental lines), while the other 34% (1171 of 3488) genes showed non-additive expression patterns. Similarly, a total of 2416 (6.5%) differentially expressed genes were detected from the *Nipponbare* × *93-11* cross, with 44% (1055 of 2466) and 56% (1361 of 2416) genes displaying additive and non-additive expression, respectively. The non-additive differentially expressed genes from the two crosses were further classified into four distinct modes based on their deviation from the mid-parent prediction: high-parent dominance, low-parent dominance, over-dominance, and under-dominance (Table 1). A random set of differentially expressed genes selected from both crosses was used to validate the transcriptional level by semi-quantitative RT–PCR (Supplemental Figure 1). The PCR analysis confirmed approximately 95% (41 of 43 for *PA64s* × *93-11* and 33 of 35 for *Nipponbare* × *93-11*) of differentially expressed genes identified from microarray analysis. These findings are consistent with a recent study in maize (Swanson-Wagner et al., 2006) that supports the involvement of multiple modes of gene action in association with heterosis.

We further investigated the profile changes of differentially transcribed genes in individual rice genotype. Using a hierarchical clustering algorithm (Figure 2C), we found that in the *PA64s* × *93-11* hybrid cross, gene expression pattern of the F_1 hybrid (*LYP9*) was more similar to the paternal line (*93-11*) but deviated from that of the maternal line (*PA64s*). On the other hand, gene expression profile in the F_1 hybrid in the *Nipponbare* × *93-11* hybrid cross was more similar to the maternal line (*Nipponbare*) than the paternal line (*93-11*). These results indicate that the hybrid could exhibit a gene expression pattern closer to either the paternal or maternal parent, depending on parental combination of individual crosses. We also compared gene expression profiles in the F_1 hybrids between the reciprocal cross, *93-11* (maternal) × *Nipponbare* (paternal), and found that only a small portion (less than 10%) of differentially expressed genes seemed to be caused by maternal effects (data not shown). Therefore, it appears that the differentially expressed genes identified from the two heterotic crosses were not mainly caused by maternal effects and thus are excellent candidates for studying the molecular events associated with heterosis.

Differential Regulation of Metabolic Processes

An analysis of the distribution among the Gene Ontology (GO) functional categories indicated that in both hybrid crosses, those differentially expressed genes are relatively enriched in pathways for carbohydrate metabolism, metabolism of cofactors and vitamins, amino acid metabolism, and biosynthesis of secondary metabolites (Supplemental Figure 2). We further examined the expression of individual genes encoding enzymatic steps in the standardized AraCyc-defined metabolic pathways (Mueller et al., 2003) between F_1 hybrid and inbred parents (Supplemental Figure 3), which confirmed the involvement of multiple biosynthetic and secondary metabolic pathways in both rice heterotic crosses. Analysis of two representative pathways, namely carbon metabolism (Calvin cycle) and gibberellin biosynthesis, are illustrated in Figure 3. We found that in the F_1 hybrid, genes involved in the Calvin cycle exhibited strong non-additive gene actions of high-parent dominance and over-dominance, which may help to reduce the effects of rate-limiting steps in carbon metabolism (Figure 3A and 3C). This observation is consistent with a significant net photosynthetic rate (Pn) increase in F_1 hybrid plants (Figure 1B). Furthermore, we also found that GA3 and zeatin contents increased in F_1 hybrid plants compared to parental inbred lines at the four-leaf stage (Figure 1C), which is consistent with recent reports that bioactive gibberellins are responsible for elongating shoots and biomass production (Eriksson et al., 2000; Biemelt et al., 2004). This is also consistent with observed changes in up-regulated expression of genes in the gibberellins (Figure 3B) and cytokinin biosynthetic pathways (figure not shown). Interestingly, genes involved in

Table 1. Statistical Analysis of Differentially Expressed Genes ($p < 0.05$).

Hybrid cross	Total	Additivity	Non-additivity	High-parent dominance	Low-parent dominance	Over-dominance	Under-dominance
PA64s × 93-11	3488	2317	1171	323	267	215	83
Nipponbare × 93-11	2416	1055	1361	190	115	245	153

F_1 represents the hybrid lines; P, paternal lines; and M, maternal lines.
Additivity, $F_1 = 1/2$ (P+M); non-additivity, $F_1 > ½$(P+M) or $F_1 < ½$(P+M). High-parent dominance, $F_1 = P > M$ or $F_1 = M > P$; low-parent dominance, $F_1 = P < M$ or $F_1 = M < P$; over-dominance, $F_1 > P$ and $F_1 > M$; under-dominance, $F_1 < P$ and $F_1 < M$.

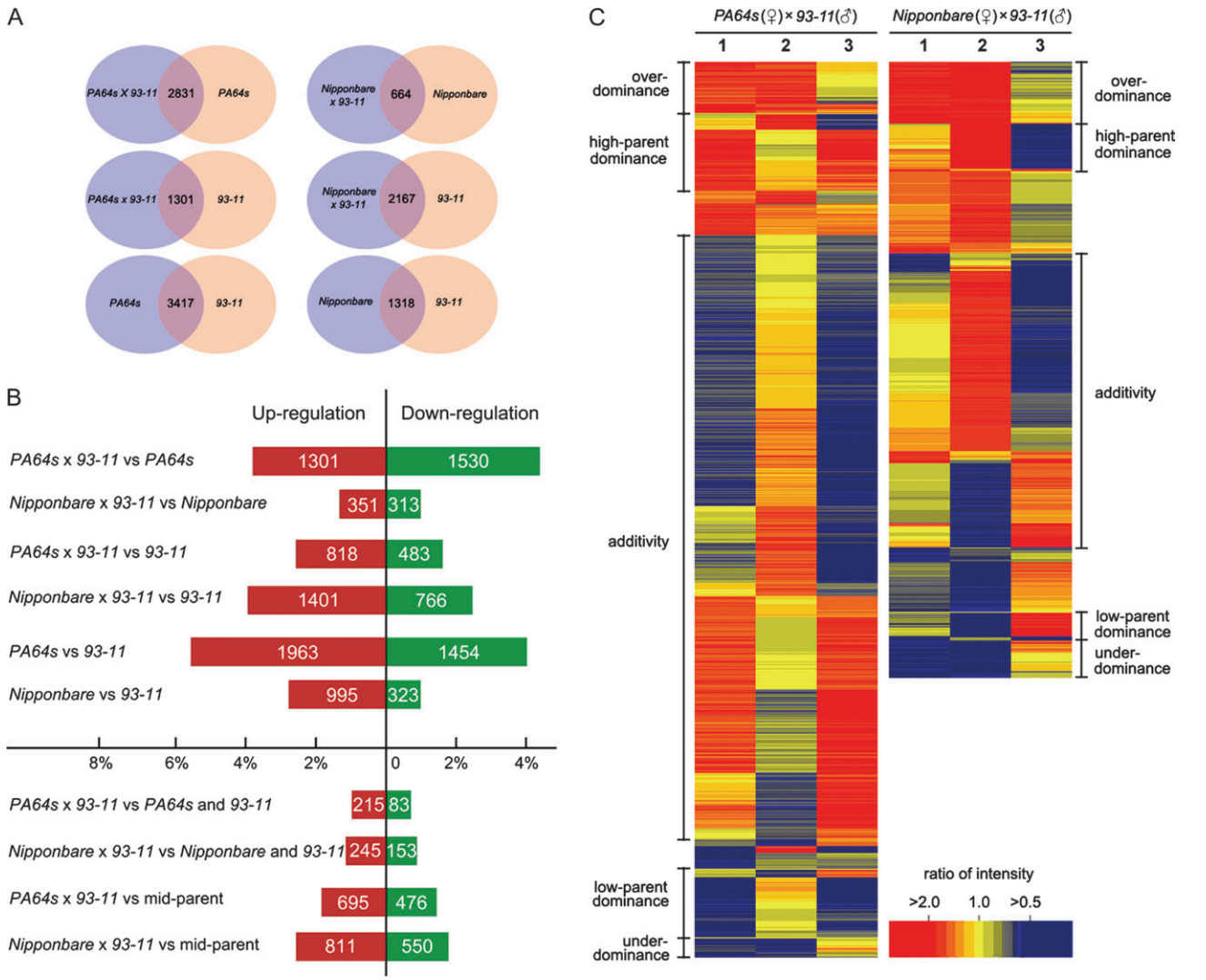


Figure 2. Differentially Expressed Genes in the Two Rice Heterotic Crosses.

Venn diagram **(A)** and statistical analysis **(B)** of differentially expressed genes among inbred parents and their F_1 hybrids.
(C) Hierarchical clustering display of differentially expressed genes with the ratio of intensity according to expression patterns among inbred parents and their F_1 hybrids in the two rice hybrid crosses. The color scale is shown at the bottom and mode of gene action is indicated on the side. Lane 1 represents intensity ratios of F_1 hybrid to female parent; Lane 2, intensity ratios of F_1 hybrid to male parent; Lane 3, intensity ratios of male parent to female parent.

biosynthesis of the hormones exhibited a strong over-dominance/under-dominance mode of gene action (Figure 3B and 3C). Some of the genes involved in other biosynthetic pathways, such as glycolysis and gluconeogenesis (Supplemental Figure 3), were also verified to exhibit multiple expression profiles by semi-quantitative RT–PCR (data not shown). These results suggest that the observed modes of gene action are generally in line with the functionality of the involved genes.

Enrichment of INDEL Polymorphism in Rice Promoter Regions

Because of the availability of the quality genomic sequences, the *Nipponbare* × *93-11* cross was first examined regarding promoter sequence polymorphism between the two parental lines and its relation to differential gene expression among the three genotypes. To this end, the 3-kb upstream region from the start codon (ATG) of orthologous gene pairs between the two parental lines was compared for all sequence polymorphism sites by Smith-Waterman local alignment algorithm (see Methods), after excluding repetitive sequences. As shown in Figure 4A, there are evidently higher sequence variations between the 500 –1500-bp upstream regions, coinciding with the anticipated promoter regions. This enrichment of sequence variations in the anticipated promoter regions is more dramatic for the small insertion/deletion polymophisms (INDEL) than that of the single nucleotide polymorphism

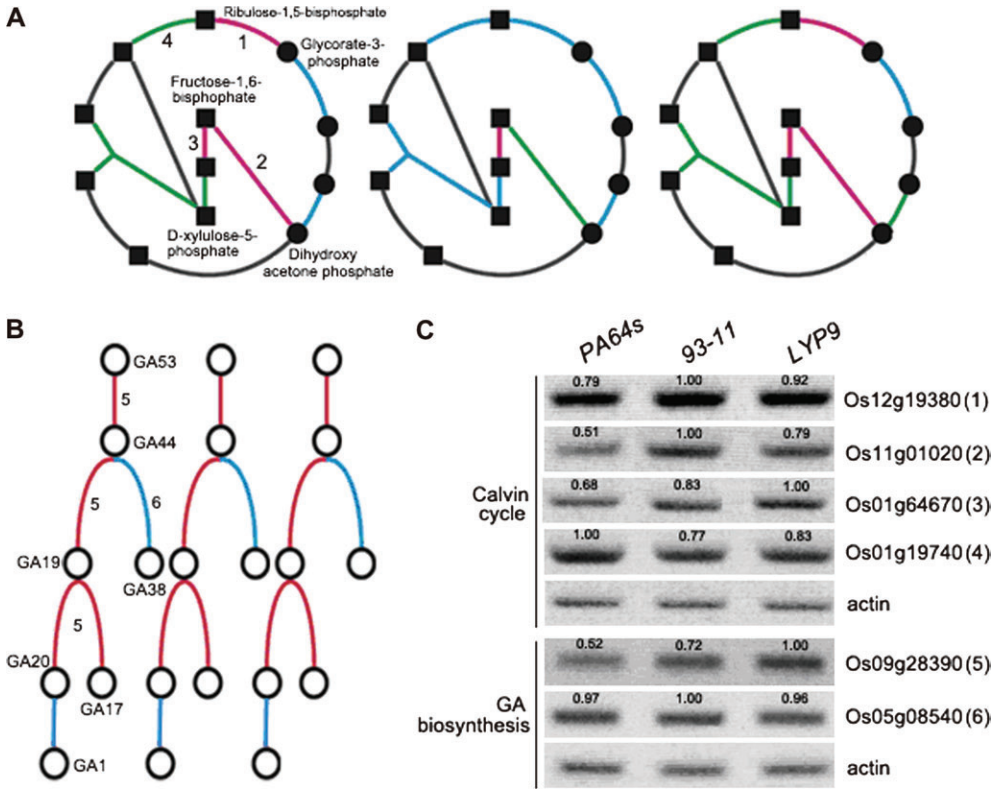


Figure 3. Analysis of Differentially Expressed Genes Involved in Different Metabolic Pathways.

Pathways of the Calvin cycle **(A)** and gibberellin biosynthesis **(B)** are represented with the ratios of expression levels among inbred parents and their F_1 hybrid. In each pathway, left indicates F_1 hybrid versus female parent; middle, F_1 hybrid versus male parent; right, male parent versus female parent. Expression level change of each reaction is shown in a color representing the ratio of two genotypes. Red represents up-expression; green, down-expression; and blue, no detectable expression change.

(C) Semi-quantitative RT–PCR analysis of selected genes from the two representative pathways. Genes representing different steps in the pathway were numbered, (1)–(4) for the Calvin cycle and (5) and (6) for the gibberellin biosynthesis pathway.

(SNP) (Figure 4B). Many putative functional *cis*-elements were found in those promoter INDELs by searching PLACE database (see Methods) (Supplemental Table 2).

Significantly Higher Promoter INDEL Polymorphism among Heterotic Expressed Genes

To further examine a possible relationship between promoter polymorphisms and differential expression of genes, we grouped the non-differentially and differentially expressed genes and calculated their respective promoter sequence polymorphism frequency (Figure 4C). The percentage of promoter regions with INDELs for non-differentially expressed genes was 40.1%, but significantly increased to 49.8% for differentially expressed genes. The percentage of SNP between the two groups of genes was similar. Figure 4E illustrates some specific examples of differentially expressed genes with promoter INDELs between inbred parental *93-11* and *Nipponbare* or *PA64s*. In each of those cases, the promoter INDEL polymorphism coincides with a putative plant promoter *cis*-element (shaded in pink), thus likely affecting the *cis*-element function and gene expression. For example, genes (1), (3), and (4) illustrated in Figure 3A and quite a number of genes involved in the biosynthetic pathways in Supplemental Figure 3 contain promoter variation between the two parental lines (data not shown). To exclude any possible sequence error in our observed promoter sequence variation, we selected 30 pairs of genes with promoter INDELs for PCR-sequencing using *93-11* and *Nipponbare* genomic DNA as templates. This analysis confirmed accuracy of 93% (28 of 30). The same analysis using *PA64s* and *93-11* genomic DNA as templates confirmed all INDELs in the promoter regions between these two genotypes. These results indicate that promoter regions of differential expressed genes in the heterotic crosses are enriched with INDEL polymorphism.

We further investigated the promoter polymorphic variation among distinct modes of differentially expressed genes in the two heterotic crosses. As shown in Figure 4D, in the *Nipponbare* × *93-11* cross, the promoter region of the additive and over-/under-dominant genes all exhibited significant higher frequency of promoter INDELs compared to non-differentially expressed genes. On the contrary, the high-parent dominance gene group showed a significantly lower

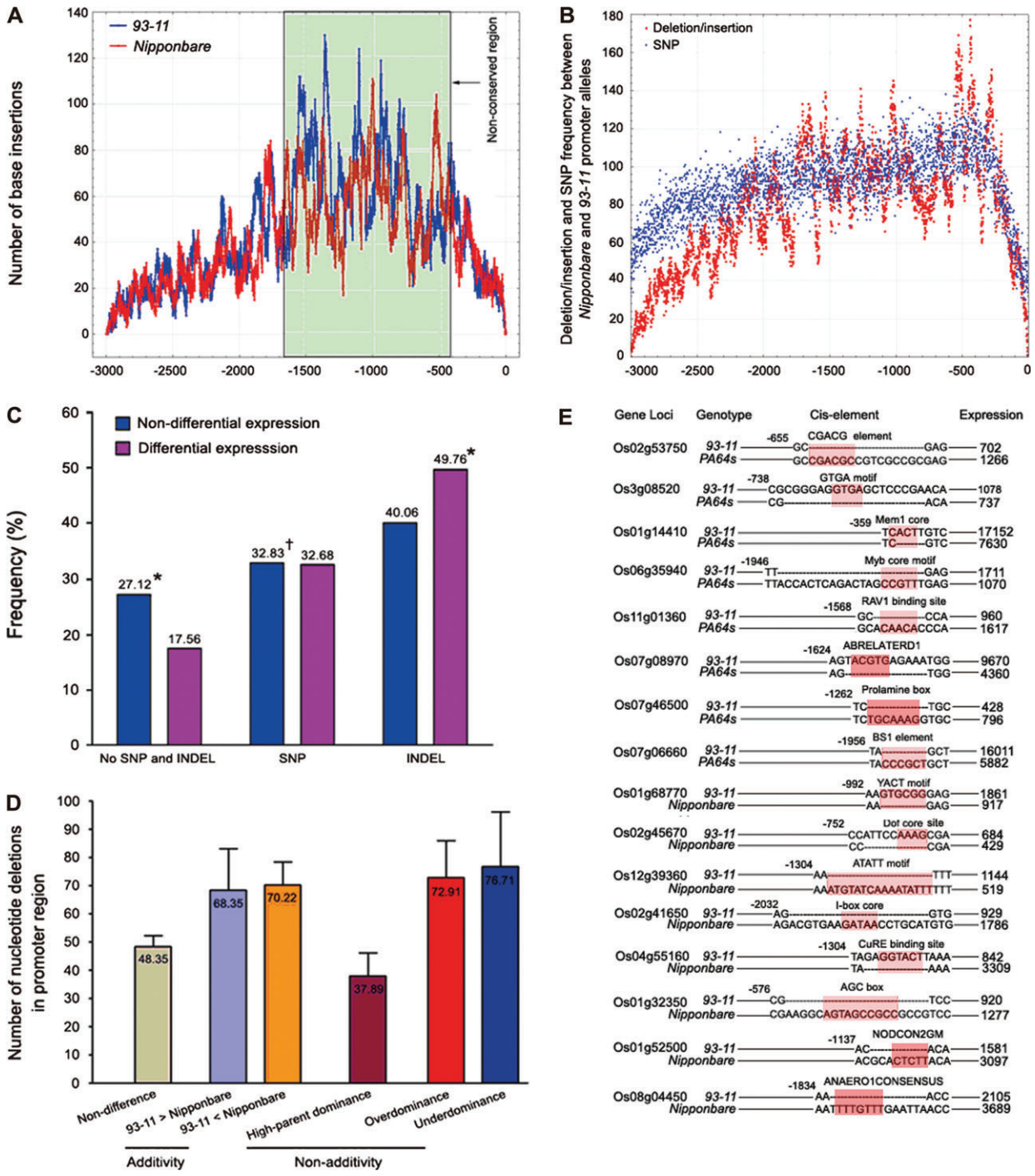


Figure 4. Sequence Variations between Parental Promoter Alleles.

(A) Comparative analysis of base insertions between *Nipponbare* and *93-11* promoter alleles. Aqua box points to the promoter position with higher sequence diversity.

(B) Comparison of sequence insertions/deletions (INDELs) and SNP (single nucleotide polymorphism) frequency between *Nipponbare* and *93-11* promoter alleles corresponding to the best homologous gene pairs (see Methods). INDELs and SNP are in red and blue, respectively.

(C, D) Sequence variation frequency (C) and base deletion frequency (D) of parental promoter alleles between differentially and non-differentially expressed genes from the *Nipponbare* × *93-11* cross. A λ^2 test was used to determine the statistical significance between the two group genes. * $P < 0.01$, † no significant difference.

(E) Examples for sequence deletions/insertions (INDELs) between parental promoter alleles from differentially expressed genes. Promoters of inbred parental alleles are compared to detect inserted/deleted fragments which are subsequently extracted and seared for *cis*-element in the Plant *Cis*-acting DNA Elements database (PLACE). For *Nipponbare* and *93-11* lines, their 3-kb promoter regions were directly extracted

frequency of promoter INDELs compared to non-differentially expressed genes, while the low-parent dominance group lacks sufficient gene number to be statistically significant. Analysis of the *PA64s* × *93-11* cross expression data and the privilegedly access to *PA64s* genomic sequence generated similar results (see Supplemental Figure 4). Therefore, differentially expressed genes with different modes of gene action in heterotic crosses are correlated with distinct patterns of INDEL enrichment in their promoter regions.

Models for Heterotic Gene Expression Patterns

Transcription factors regulate the binding of RNA polymerase to their target genes by binding to specific *cis*-regulatory DNA sequences. Polymorphisms in the *cis*-regulatory elements and differential expression of transcription genes in the parental lines, and their unique combination in the F_1 hybrids, thus could bring about the five possible modes of gene action. In both heterotic crosses, many transcriptional factor genes were found as differentially expressed (Table 2 and Supplemental Table 3). The transcription factor genes could in turn regulate other genes that display distinct modes of heterotic gene action.

Based on the above analysis, we propose a working hypothesis (Figure 5) to explain how promoter INDELs and differential expression of transcription factors acting upon them could affect gene expression and contribute to different modes of gene action in F_1 hybrids. As shown in Figure 5A, hybrid progeny inherits one copy of the gene from each parent. In the case of additive expression, the lack (or low activity) of transcription from one allele is assumed to be caused by the lack of (or weak) binding of a transcriptional activator to the promoter due to a *cis*-element mutation; the F_1 hybrid progeny will thus have an expression level between the two parents. In the case of high-parent dominance, where the promoters are the same in both parents but the transcription factor is missing (or with low activity) in one parent, the progeny would have an expression level similar to the high expression parent. In the case of over-dominance, it is possible that the responsible transcription factor is only present in the parent with a dysfunctional promoter due to INDEL, but missing (weak) in the other parent that has a functional promoter. The hybrid would thus have the transcription factor from one parent and the functional promoter from the other, resulting in higher expression levels than either parent. Similarly, the transcription factor can act to repress transcription and its interaction with promoter INDELs can result in other modes of gene action, such as additivity, low-parent dominance, or under-dominance (Figure 5B).

DISCUSSION

Based on our microarray analysis, approximately 9.4% (3488 of 36 962) and 6.5% (2416 of 36 962) of genes exhibited differential transcription in the *PA64s* × *93-11* and the *Nipponbare* × *93-11* hybrid cross, respectively. These differential genes exhibited additive and non-additive expression patterns; however, their proportion differs in the two heterotic crosses (Figure 2). In the *PA64s* × *93-11* cross, a large portion (66%, 2317 of 3488) exhibited a mode of gene action that could not be distinguished from additivity, which was similar to recent studies in maize (Swanson-Wagner et al., 2006; Stupar and Springer, 2006). On the contrary, several studies reported a high proportion of non-additive expression patterns in F_1 hybrids (Song and Messing, 2003; Gibson et al., 2004; Auger et al., 2005), which was consistent with our observation that a majority of the differentially expressed genes exhibit non-additive gene action (56%, 1361 of 2416) in the *Nipponbare* × *93-11* heterotic cross. Among those non-additive expressed genes, the proportion of genes with clear over-dominant action was 6.5% (215 of 3416) and 10% (245 of 2416) in the *PA64s* × *93-11* and the *Nipponbare* × *93-11* heterotic crosses, respectively, which was similar to results from prior studies (Vuylsteke et al., 2005; Gibson et al., 2004). Together, our results support the notion that multiple molecular mechanisms contribute to heterosis.

Many factors affect plant yield and productivity, such as the capacity of fixing carbon dioxide (CO_2) and converting the fixed carbon into dry matters. Functional analysis indicated that a large number of differentially expressed genes in F_1 hybrids were involved in multiple biosynthesis pathways related to carbon metabolism, such as the glycolysis pathway, the tricarboxylic acid (TCA) cycle, the calvin cycle, and the glyconeogenesis pathway (Figure 3 and Supplemental Figures 2 and 3). Our microarray data also showed that many genes involving in photosynthesis were up-regulated in F_1 hybrids, similarly with findings from previous studies (Bao et al., 2005). Additionally, up-regulated genes for hormone biosynthesis are consistent with a higher content of gibberellin and cytokinin in F_1 hybrids (Figure 1C). Thus, novel changes of biosynthetic processes by differentially expressed genes might result in heterotic phenotypes of F_1 hybrid rice.

It is well established that the proportion and composition of transcribed genes change considerably during the lifecycle of plants, and most regulation of gene expression occurs at the transcriptional level. Comparative analysis of promoter alleles between inbreds showed that sequence polymorphism preferentially occurred in those differentially transcribed genes. In this study, mid-parent (or additive) expression of F_1 hybrid

from public rice genome databases. For the *PA64s* × *93-11* cross, however, the promoter regions were amplified and sequenced from *PA64s* and *93-11* genomic DNAs according to conserved sequences between *Nipponbare* and *93-11* promoter alleles, respectively. Pink boxes show nucleotides composition of *cis*-elements in inserted/deleted promoter regions. Gene ID, names and position of *cis*-elements from the start codon together with microarray expression levels of both parents are all shown.

Table 2. Differentially Expressed Transcription Factors in the Microarray Analysis from the *Nipponbare* × *93-11* Hybrid Cross.

Gene locus	Description (e-value)[a]	Significant pattern[b]	MP fold change[c]	Expression model[d]
Os01g11500	Zinc finger, C3HC4 type ($9e^{-6}$)	P<F≈M	1.88	HPD
Os01g21120	AP2 domain ($4e^{-11}$)	P<F≈M	2.32	HPD
Os01g39330	Helix-loop-helix family ($2e^{-9}$)	M≈F<P	1.88	LPD
Os01g60600	WRKY family ($3e^{-15}$)	P≈F<M	2.44	LPD
Os01g61080	WRKY family ($7e^{-27}$)	M≈F<P	1.66	LPD
Os01g62920	Homeobox domain ($9e^{-9}$)	P≈F<M	2.04	LPD
Os01g64790	AP2 domain ($4e^{-9}$)	P<M<F	1.67	ODO
Os01g67480	Helix-loop-helix domain ($8e^{-8}$)	M≈F<P	1.55	LPD
Os01g68860	Zinc finger, CCCH type ($5e^{-5}$)	F<M≈P	1.19	UDO
Os01g74440	SRF-type ($6e^{-14}$)	P<F≈M	1.82	HPD
Os02g13670	Helix-loop-helix domain ($2e^{-11}$)	P<F≈M	1.92	HPD
Os02g34840	Transcriptional Coactivator p15 ($2e^{-17}$)	M≈P<F	1.15	ODO
Os02g40510	response regulator APRR1 ($5e^{-14}$)	M≈P<F	1.28	ODO
Os02g45450	CRT/DRE binding factor 1 ($3e^{-10}$)	M≈F<P	1.76	LPD
Os02g45670	Myb family-like ($1e^{-81}$)	M≈F<P	1.59	LPD
Os02g47060	WRKY family ($1e^{-24}$)	M≈P<F	1.26	ODO
Os02g52780	bZIP family ($6e^{-6}$)	F<M<P	1.74	UDO
Os02g54520	R2R3 Myb family ($6e^{-9}$)	P<F≈M	1.92	HPD
Os03g12760	helix-loop-helix family ($3e^{-7}$)	P<F≈M	1.72	LPD
Os03g55540	zinc finger ZF1 ($2e^{-28}$)	F<M≈P	1.01	UDO
Os04g31800	MADS GGM9 ($7e^{-15}$)	P<F≈M	1.78	ND
Os04g33950	E2F_TDP ($5e^{-23}$)	P<F≈M	2.44	ND
Os04g51070	bHLH family ($2e^{-34}$)	M≈P<F	1.12	ODO
Os04g59380	Zinc finger, C2H2 type ($1e^{-66}$)	P<F≈M	1.85	HPD
Os05g46020	WRKY family ($7e^{-24}$)	P<F<M	3.03	Additivity
Os05g50340	Myb-like domain ($5e^{-20}$)	P<M<F	2.56	ODO
Os06g44010	WRKY family ($2e^{-22}$)	M≈F<P	2.22	LPD
Os07g07350	AN1-like Zinc finger ($5e^{-28}$)	P≈F<M	2.00	ND
Os07g02800	myb-like domain (0.004)	P<M<F	1.88	ODO
Os07g46500	Ankyrin repeats family ($3e^{-22}$)	P<F≈M	1.85	HPD
Os08g36920	AP2 domain ($3e^{-10}$)	M≈F<P	2.44	LPD
Os09g03680	Ankyrin repeats family ($2e^{-20}$)	M<F<P	2.68	Additivity
Os09g25060	WRKY family ($3e^{-22}$)	M≈F<P	2.26	LPD
Os09g33810	Ankyrin repeats family ($1e^{-20}$)	P<F≈M	2.04	HPD
Os09g34950	TCP family ($1e^{-22}$)	P<F≈M	2.32	HPD
Os10g05970	DUF1210 family ($2e^{-39}$)	P<M<F	5.00	ODO
Os10g05980	DUF1210 family ($6e^{-41}$)	P<M<F	4.54	ODO
Os10g34180	BTF3b-like ($1e^{-13}$)	P<F≈M	1.56	HPD
Os11g08020	uncharacterized	M≈F<P	4.31	LPD
Os12g40920	bZIP family ($4e^{-10}$)	M<F≈P	2.02	HPD

a Conserved protein domains corresponding to individual genes were searched in NCBI protein databases by using BLASTP.
b $P < 0.05$, M, female parent; P, male parent; F, F_1 hybrid.
c Fold change between parent with the highest level of gene expression and another parent is shown as a ratio.
d Gene expression of F_1 hybrid is classified into multiple patterns. HPD indicates high-parent dominance; LPD, low-parent dominance; ODO, over-dominance; UDO, under-dominance; ND, undistinguishable model from additivity and non-additivity.

indicated that the alleles from respective inbred parents did not equally contribute to transcript accumulation for variance of promoter sequence, which was consistent with the previous study (Guo et al., 2004). In the hybrids, two alleles are exposed to the common *trans*-acting factors (such as transcription factors), but the promoter INDEL polymorphism may lead to target genes differentially interacting with the common transcription factors, thus resulting in differential transcription

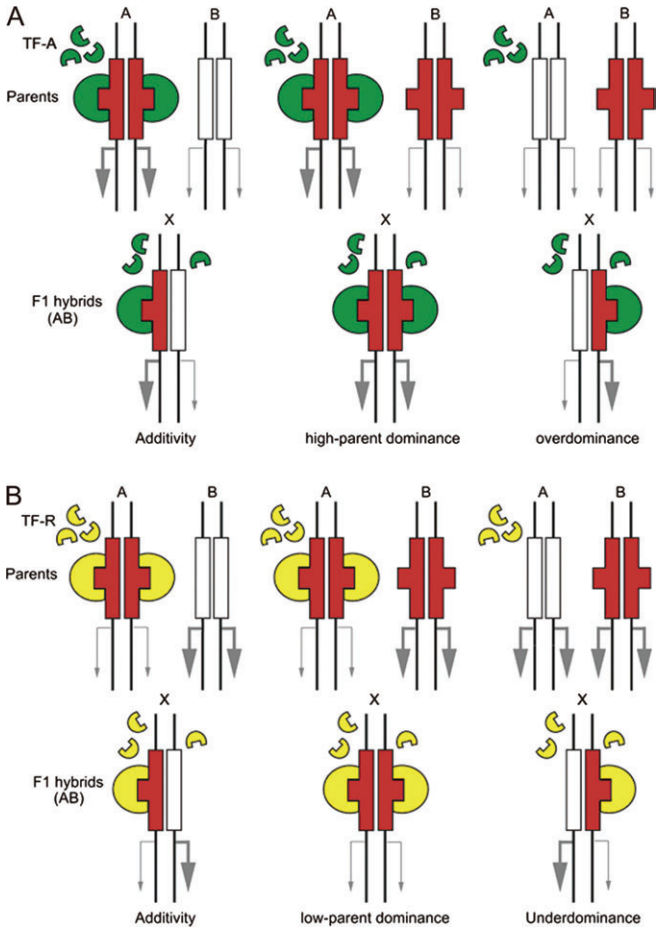


Figure 5. Models for Promoter INDEL-Mediated Gene Action in Heterotic Crosses.

It is assumed that for genes with promoter INDEL, there are two possibilities: either one parent or both contain a certain *cis*-element essential for gene expression.

(A) Differential expression of a transcriptional activator and its action on the *cis*-element could result in gene expression levels in the hybrid in between (additivity) or higher than both parental lines (over-dominance), or similar to the high-parent (high-parent dominance).

(B) Conversely, differential expression of a transcriptional repressor and its action on the *cis*-element could result in gene expression levels in the hybrid in between (additivity) or lower than both parental lines (under-dominance), or similar to the low-parent (low-parent dominance). Red box represents the presence of *cis*-element; white box, absence of *cis*-element; green ellipse, transactivator; yellow ellipse, transrepressor; gray line, gene transcripts.

between alleles tending towards mid-parent or additive levels as described in Figure 5. The majority of genes displaying non-additive profiles are expressed at levels within the range of two inbred parental lines, which might also be explained by the promoter sequence variation. For example, a transactivator or transrepressor in either of the inbred parents could lead to multiple non-additive expression patterns in the F_1 hybrids (Figure 5). The fact of the promoter sequence polymorphism within non-differentially expressed genes might be due to developmental stage-specific expression fashions. In conclusion, the heterotic gene expression in the two rice hybrid crosses was in part or completely due to promoter INDEL polymorphism.

From an evolutionary point of view, the divergence of promoter sequences between inbred lines might be the result of natural selection or domestication through artificial selection for humans' desired traits. Several modes of natural selection on promoter sequences as well as on coding sequences (Wray et al., 2003), such as positive selection, over-dominant selection, and stabilizing selection, have been proposed. Many promoters are organized into functional modules, and each of them produces a discrete effect on the total transcription output, allowing selection to modify this discrete effect independently (Arnone and Davidson, 1997). Alternatively, many promoter alleles may be functionally co-dominant and thus immediately visible to selection, which increases the efficiency of fixing beneficial alleles and eliminating deleterious ones. A number of rice studies have revealed the importance of artificial selection in the establishment of cultivated rice (Cheng et al., 2003; Vaughan et al., 2003; Lu et al., 2006), and a most recent study (Konishi et al., 2006) indicated that the 5′ sequence variation of regulatory region of *qSH1* gene caused loss of seed shattering among *japonica* subspecies of rice, and the sequence variation was a result of artificial selection during rice domestication. Thus, comparison of the promoter–transcription factor combination that results in certain modes of gene action in rice and other grass species should reveal the contribution of natural selection as well as artificial selection in shaping the outcome of heterosis.

The genome scale transcription analysis reported herein revealed that about 7–9% of the genome is differentially expressed, exhibiting all modes of gene action in two heterotic crosses. The number of genes with additive gene action in hybrid F_1 progeny varied over two-fold in the two crosses, while a number of genes with the non-additive gene action mode were quite similar in the two crosses. Enrichment of promoter INDEL polymorphism was highly correlated with differentially expressed genes with additive as well as over- and under-dominance gene action, but not the high- and low-parent dominance. A working model was proposed based on these results that could explain all possible modes of gene action as a result of promoter INDELs and differential expression of transcription factors. These results indicate that the interplay between promoter INDEL and regulated expression of transcription factors is one of the molecular events for heterotic gene expression contributing to heterosis in rice.

METHODS

Plant Materials and Oligonucleotide Microarray

LYP9 (F_1 hybrid), *93-11* (*Oryza sativa* L. ssp *indica* cultivar), *PA64s* (with a mixed genetic background of *indica*, *japonica*, and *javanica*), *Nipponbare* (*Oryza sativa* L. ssp. *japonica*

cultivar), and *Nipponbare* × *93-11* hybrids were cultivated under typical field conditions in south China for physiological characterization and seed propagation of the isogenic strains. For hybrid seed production, *LYP9* was produced in cooler areas, where the *PA64s* strain is male-sterility and can be easily cross-pollinated by *93-11*. The hybrid seeds of *Nipponbare* × *93-11* were produced by manual pollination of the *93-11* pollens into *Nipponbare* stigma. Seedling shoots grown in environmentally controlled growth chambers at 28°C were harvested at the four-leaf stage for microarray analysis. The rice 70-mer oligonucleotide set representing 36 926 unique genes was used in this study as described previously (Jiao et al., 2005; Ma et al., 2005).

RNA Isolation, Probe Labeling, and Microarray Hybridization

Rice seedling shoots were frozen in liquid nitrogen, and total RNA was isolated using RNAwiz reagent (Ambion) and purified by the RNeasy kit (Qiagen) according to the manufacturer's instructions. For each sample, 100 mg of total RNA was labeled with aminoallyl-dUTP (aa-dUTP, Sigma-Aldrich) by reverse transcription. The aminoallyl- dUTP-labeled cDNAs were purified using a Microcon YM-30 filter (Millipore) and resuspended in 0.1 M $NaHCO_3$. The purified cDNAs were further fluorescently labeled by conjugating monofunctional Cy2, Cy3, or Cy5 dye (Amersham) to the aminoallyl functional groups using a loop-like design. Three biological replicates were performed with each hybrid cross, and, in each replicate, alternative RNA pools for three genotypes were labeled with Cy2, Cy3, and Cy5 dye, respectively. After coupling at room temperature for 1 h, the labeling reaction was stopped by ethanolamine. The labeled probes were separated from unincorporated dye using the QIAquick PCR purification kit (Qiagen) and concentrated by a Microcon YM-30 filter, respectively. The following protocols for microarray hybridization, microarray slide washing, and array scanning were the same as described previously (Ma et al., 2001). Hybridized slides were scanned with a GenePix 4000B scanner (Axon), and independent TIFF images for triple Cy2, Cy3, and Cy5 channels were used for subsequent analysis.

Microarray Data Processing

After manual removal of spots with aberrant morphology, microarray spot intensity signals were acquired using the Axon GenePix Pro 5.0 software package without correction for background, and each slide included three troops of intensity data corresponding to Cy5, Cy3, and Cy2 channels. We first remove the dye intensity difference on each slides with the 'LOWESS' normalization method, which was respectively applied to $\log_2$ transformed intensities from each pair of two channels (Cy5/Cy3, Cy3/Cy2, Cy2/Cy5), and three repeats of such processing guaranteed that the dye effects had been removed. Subsequently, quantile normalization was applied to the three LOWESS-normalized microarray data to remove biases among slides. After that, all intensities of three channels can be compared to each other.

For detection of differentially expressed spots among three genotypes, the normalized data was $\log_2$ transformed and fitted into a mixed effect ANOVA model, with the software MAANOVA under R environment. After multiple testing between pairwise comparisons, spots with FDR-corrected *P* values <0.05 were regarded as differentially expressed genes. The same strategy was performed in a linear-in-genotype contrast when F_1 genotype was compared to the average of the two parental lines. Spots with FDR-corrected *P* values <0.05 were regarded as non-additivity, when the spots with FDR-corrected *P* values >0.05 were regarded as no statistically significant difference from additivity. To classify the genes further distinguishably, the high-parent dominant genes and low-parent dominant genes were identified from the non-additive group based on the criterion that F_1 genotype was significantly different from one parent and no significantly different from another parent. From the additive group, the genes were identified as over-dominance or under-dominance when the F_1 genotype was significantly higher or lower than both inbred parents, respectively.

Cluster Analysis

Cluster analysis was applied to all genes showing differential expression among hybrids and inbred parents from both two rice hybrid crosses, respectively. Differential expression was determined as stated above. Hierarchical clustering with the average linkage was performed using the software Cluster and visualized by the Treeview program (Eisen et al., 1998).

Promoter Variation and Motif Search

To detect the promoter variations, homologous gene pairs, with same length, mapping extron position and completely identical coding nucleotide sequence, were selected between two inbred parental genomes in each hybrid cross (*PA64s* × *93-11* and *Nipponbare* × *93-11*). 5084 homologous gene pairs were screened out from the inbred lines *PA64s* and *93-11*, and 5000 from the inbred lines *Nipponbare* and *93-11*. After extracting the 3-kilo bases upstream regions from the start codon in these homologous gene pairs, Smith-Waterman local alignment algorithm was carried out to detect the polymorphism sites. Among those sites, deleted/inserted fragments which meet the following conditions were recorded particularly: (1) the length of inserted/deleted fragments ranges from 4 to 50 bp; and (2) these fragments are not single nucleotide repetitive sequences. Then, these small deleted/inserted fragments were searched in the Plant *Cis*-acting Regulatory DNA Elements database (Higo et al., 1998).

Correlation between Expression Modes and Promoter Deletion Frequency

To examine the relationship between promoter variation and gene expression profiles, the frequency of promoter deletions/insertions was detected in all differential genes. Gene expression patterns include two main groups based on the F_1 expression:

additivity and non-additivity. The former includes those genes exhibiting non-difference and *93-11* higher or lower than *Nipponbare*, and the latter includes those genes exhibiting high-parent, low-parent, over-dominance, and under-dominance. For each kind of distinct gene, 3-kb promoter alleles between inbred parents from the start codon mapped to corresponding genes were compared to each other using BLAT (Kent, 2002). The promoter alleles with identical sequences of more than 2 kb were subjected to BLAT to analyze average nucleotide numbers in the deleted/inserted fragments (gaps) according to each distinct expression profile, and then the Mann-Whitney U test was performed to estimate the different distributions in these differential genes with distinct expression patterns.

Semi-Quantitative RT–PCR

Total RNA was isolated from four-leaf seedling shoots as used in microarray analysis and treated with RNase-free DNase (Promega), and cDNA was synthesized with SuperScript™ II First-Strand cDNA Synthesis kit (Invitrogen). PCR was performed using general standard techniques, and *O. sativa* actin gene was used as control.

Analysis of Net Photosynthesis and Hormones Content

The net photosynthesis was measured using a portable photosynthesis system (CIRAS-1, PP Systems, UK) by an open system in the morning between 09:00 and 11:00 h to avoid potential photoinhibition. Measurements were made by attaching a light source to the leaf chamber window under saturating photosynthetic photon flux densities (1200 μmol m^{-2} s^{-1}). Data were determined at least in five leaves from five different plants of each variety.

The endogenous hormone content was measured using the Agilent 1100 high-performance liquid chromatography (HPLC) system (Agilent, Palo Alto, CA). Fresh leaves at the four-leaf stage were frozen in liquid nitrogen and then cold-dried under vacuum. 0.5 g dried leaves were added to 11 mL 80% aqueous methanol and immediately homogenized on ice. The homogenate was maintained during 15 h at 4°C in darkness with continuous shaking. Afterwards, it was centrifuged for 10 min at 4500 *g* at 4°C, and the clear upper phase was collected and evaporated under vacuum. Dry residue was re-dissolved in 8 mL of ammonium acetic buffer (0.1 mol L^{-1}, pH 9.0). After centrifuging for 20 min at 15 000 rpm, the upper supernatant was purified sequentially through Polyvinylpolypyrrolidone (PVPP) column and DEAE Sephadex A225 column. Before HPLC analysis, the elution with 50% aqueous methanol was concentrated by Sep-Pak C18 column (Waters Chromatography). Standard gibberellin and zeatin were purchased from Fluka Co. (Switzerland). All solvents and buffers were HPLC-quality.

SUPPLEMENTARY DATA

Supplementary Data are available at *Molecular Plant Online*.

FUNDING

This study was supported by the 863 Rice Functional Genomics Program and the National '863' High-Tech Project 2003AA 210070 from the Ministry of Science and Technology of China.

ACKNOWLEDGMENTS

We thank C.M. Lu and X.Y. Chen for suggestions and technical assistance with measurement of the rice photosynthetic characteristics. No conflict of interest declared.

REFERENCES

Arnone, M.I., and Davidson, E.H. (1997). The hardwiring of development: organization and function of genomic regulatory systems. Development. **124,** 1851–1864.

Auger, D.L., Gray, A.D., Ream, T.S., Kato, A., Coe, E.H. Jr, and Birchler, J.A. (2005). Nonadditive gene expression in diploid and triploid hybrids of maize. Genetics. **169,** 389–397.

Bao, J.Y., et al. (2005). Sage analysis of gene expression study of a hybrid rice strain (LYP9) and its parental cultivars. Plant Physiol. **138,** 1216–1231.

Biemelt, S., Tschiersch, H., and Sonnewald, U. (2004). Impact of altered gibberellin metabolism on biomass accumulation, lignin biosynthesis, and photosynthesis in transgenic tobacco plants. Plant Physiol. **135,** 254–265.

Birchler, J.A., Auger, D.L., and Riddle, N.C. (2003). In search of the molecular basis of heterosis. Plant Cell. **15,** 2236–2239.

Cheng, C., Motohashi, R., Tsuchimoto, S., Fukuta, Y., Ohtsubo, H., and Ohtsubo, E. (2003). Polyphyletic origin of cultivated rice: based on the interspersion pattern of SINEs. Mol. Biol. Evol. **20,** 67–75.

Davenport, C.B. (1908). Degeneration, albinism and inbreeding. Science. **28,** 454–455.

East, E.M. (1908). Inbreeding in corn. Connecticut Agric. Exp. Stn. **1907,** 419–428.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl Acad. Sci. U S A. **95,** 14863–14868.

Eriksson, M.E., Israelsson, M., Olsson, O., and Moritz, T. (2000). Increased gibberellin biosynthesis in transgenic trees promotes growth, biomass production and xylem fiber length. Nat. Biotechnol. **18,** 784–788.

Gibson, G., Riley-Berger, R., Harshman, L., Kopp, A., Vacha, S., Nuzhdin, S., and Wayne, M. (2004). Extensive sex-specific nonadditivity of gene expression in Drosophila melanogaster. Genetics. **167,** 1791–1799.

Goff, S.A., et al. (2002). A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science. **296,** 92–100.

Guo, M., Rupe, M.A., Danilevskaya, O.N., Yang, X., and Hu, Z. (2003). Genome-wide mRNA profiling reveals heterochronic allelic variation and a new imprinted gene in hybrid maize endosperm. Plant J. **36,** 30–44.

Guo, M., Rupe, M.A., Yang, X.F., Crasta, O., Zinselmeier, C., Smith, O.S., and Bowen, B. (2006). Genome-wide transcript analysis of maize hybrids: allelic additive gene expression and yield heterosis. Theor. Appl. Genet. **113,** 831–845.

Guo, M., Rupe, M.A., Zinselmeier, C., Habben, J., Bowen, B.A., and Smith, O.S. (2004). Allelic variation of gene expression in maize hybrids. Plant Cell. **16,** 1707–1716.

Higo, K., Ugawa, Y., Iwamoto, M., and Higo, H. (1998). PLACE: a database of plant cis-acting regulatory DNA elements. Nucleic Acids Res. **26,** 358–359.

Hua, J., Xing, Y., Wu, W., Xu, C., Sun, X., Yu, S., and 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. U S A. **100,** 2574–2579.

Hua, J., Xing, Y., Xu, C., Sun, X., Yu, S., and Zhang, Q. (2002). Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. Genetics. **162,** 1885–1895.

Huang, Y., Zhang, L., Zhang, J., Yuan, D., Xu, C., Li, X., Zhou, D., Wang, S., and Zhang, Q. (2006). Heterosis and polymorphisms of gene expression in an elite rice hybrid as revealed by a microarray analysis of 9198 unique ESTs. Plant Mol. Biol. **62,** 579–591.

Jiao, Y., Ma, L., Strickland, E., and Deng, X.W. (2005). Conservation and divergence of light-regulated genome expression patterns during seedling development in rice and Arabidopsis. Plant Cell. **17,** 3239–3256.

Kent, W.J. (2002). BLAT: the BLAST-like alignment tool. Genome Res. **12,** 656–664.

Konishi, S., Izawa, T., Lin, S.Y., Ebana, K., Fukuta, Y., Sasaki, T., and Yano, M. (2006). An SNP caused loss of seed shattering during rice domestication. Science. **312,** 1392–1396.

Li, Z.K., et al. (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.

Lippman, Z.B., and Zamir, D. (2007). Heterosis: revisiting the magic. Trends Genet. **23,** 60–66.

Lu, H., Romero-Severson, J., and Bernardo, R. (2003). Genetic basis of heterosis explored by simple sequence repeat markers in a random-mated maize population. Theor. Appl. Genet. **107,** 494–502.

Lu, J., Tang, T., Tang, H., Huang, J., Shi, S., and Wu, C.I. (2006). The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. Trends Genet. **22,** 126–131.

Ma, L., et al. (2005). A microarray analysis of the rice transcriptome and its comparison to Arabidopsis. Genome Res. **15,** 1274–1283.

Ma, L., Li, J., Qu, L., Hager, J., Chen, Z., Zhao, H., and Deng, X.W. (2001). Light control of Arabidopsis development entails coordinated regulation of genome expression and cellular pathways. Plant Cell. **13,** 2589–2607.

Meyer, S., Pospisil, H., and Scholten, S. (2007). Heterosis associated gene expression in maize embryos 6 days after fertilization exhibits additive, dominant and overdominant pattern. Plant Mol. Biol. **63,** 381–391.

Mueller, L.A., Zhang, P., and Rhee, S.Y. (2003). AraCyc: a biochemical pathway database for Arabidopsis. Plant Physiol. **132,** 453–460.

Shull, G.H. (1908). The composition of a field of maize. Am. Breed Assn. **4,** 296–301.

Song, R., and Messing, J. (2003). Gene expression of a gene family in maize based on noncollinear haplotypes. Proc. Natl Acad. Sci. U S A. **100,** 9055–9060.

Stoughton, R.B. (2005). Applications of DNA microarray in biology. Annu. Rev. Biochem. **74,** 53–82.

Stuber, C.W., Lincoln, S.E., Wolff, D.W., Helentjaris, T., and Lander, E.S. (1992). Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics. **132,** 823–839.

Stupar, R.M., and Springer, N.M. (2006). Cis-transcriptional variation in maize inbred lines B73 and Mo17 leads to additive expression patterns in the F1 hybrid. Genetics. **173,** 2199–2210.

Swanson-Wagner, R.A., Jia, Y., DeCook, R., Borsuk, L.A., Nettleton, D., and Schnable, P.S. (2006). All possible modes of gene action are observed in a global comparison of gene expression in a maize F1 hybrid and its inbred parents. Proc. Natl Acad. Sci. U S A. **103,** 6805–6810.

Vaughan, D.A., Morishima, H., and Kadowaki, K. (2003). Diversity in the Oryza genus. Curr. Opin. Plant Biol. **6,** 139–146.

Vuylsteke, M., Van Eeuwijk, F., Van Hummelen, P., Kuiper, M., and Zabeau, M. (2005). Genetic analysis of variation in gene expression in Arabidopsis thaliana. Genetics. **171,** 1267–1275.

Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., and Romano, L.A. (2003). The evolution of transcriptional regulation in eukaryotes. Mol. Biol. Evol. **20,** 1377–1419.

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

Yu, J., et al. (2002). A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science. **296,** 79–92.

Yu, J., et al. (2005). The genomes of *Oryza sativa*: a history of duplications. PLoS Biol. **3,** e38.

Yu, S.B., Li, J.X., Xu, C.G., Tan, Y.F., Gao, Y.J., Li, X.H., Zhang, Q., and Maroof, M.A. (1997). Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. Proc. Natl Acad. Sci. U S A. **94,** 9226–9231.

Supplementary Material

Supplemental Tables

Supplemental Table 1: Heterosis for rice seedling

Genotypes	Seedling dry weight (g)	Seedling height (cm)	Tillering number*
PA64s x *93-11*	0.145±0.004	31.29±1.15	3.7±0.33
PA64s	0.122±0.007	23.52±1.52	3.0±0.25
93-11	0.120±0.008	27.51±2.52	2.6±0.58
Nipponbare	0.101±0.005	29.75±0.50	2.2±0.20
Nipponbare x *93-11*	0.186±0.020	34.16±2.11	4.1±0.77

*Tillering number was examined at six-leaf stage of the main tiller.

Supplemental Table 2: Functional *cis*-elements discovered from deleted/inserted fragments between *Nipponbare* and *93-11* promoter alleles by PLACE search

Entry ID	Numbers	Factor or site name	Core motif
S000449	637	CACTFTPPCA1	YACT
S000265	514	DOFCOREZM	AAAG
S000454	417	ARR1AT	NGATT
S000028	408	CAATBOX1	CAAT
S000198	370	GT1CONSENSUS	GRWAAW
S000039	364	GATABOX	GATA
S000144	347	EBOXNNAPA	CANNTG
S000407	347	MYCCONSENSUS	CANNTG
S000378	311	GTGANTG10	GTGA
S000447	272	WRKY71OS	TGAC
S000098	270	ROOTPAPOX1	ATATT
S000245	212	POLLENLELAT52	AGAAA
S000415	195	ACGTATERD1	ACGT
S000430	181	GCCORE	GCCGCC

Supplemental Table 3: Differentially expressed transcription factors in the microarray analysis from *PA64s* x *93-11* hybrid cross

Gene locus	Description (e value)[a]	Significant pattern[b]	MP fold change[c]	Expression model[d]
Os01g07390	Zinc finger, C3HC4 type ($3e^{-9}$)	P≈F<M	1.49	ND
Os01g09370	Ankyrin-like protein ($2e^{-14}$)	P<F<M	3.22	ND
Os01g21120	AP2 domain ($2e^{-12}$)	P≈F<M	2.04	ND
Os01g53220	HSF-type family ($6e^{-33}$)	P<F<M	2.63	Additivity
Os01g60600	WRKY family ($3e^{-15}$)	P<M<F	1.51	ODO
Os01g61810	Histone-like transcription factor ($4e^{-20}$)	P≈F<M	2.38	LPD
Os01g61990	Ankyrin repeats ($1e^{-18}$)	P<F<M	1.96	Additivity
Os01g62460	GRAS family ($1e^{-90}$)	M<F≈P	1.42	HPD
Os01g64790	AP2 domain ($4e^{-9}$)	P<F<M	1.82	Additivity
Os01g68860	Zinc finger CCCH type ($5e^{-5}$)	M<F<P	1.83	Additivity
Os02g03340	Transcription factor Tfb4 ($7e^{-35}$)	M≈F<P	1.64	ND
Os02g04640	Myb-like DNA-binding family ($2e^{-4}$)	P<F<M	1.92	Additivity
Os02g06330	ERE binding factor ($1e^{-6}$)	M<F≈P	1.76	HPD
Os02g07930	B-box zinc finger ($2e^{-9}$)	M<F<P	3.29	Additivity
Os02g08440	WRKY family ($4e^{-23}$)	P<F<M	1.88	Additivity
Os02g10860	bZIP family ($1e^{-10}$)	M<F≈P	1.70	HPD
Os02g27060	HMG (high mobility group) box ($1e^{-12}$)	M<F<P	1.84	Additivity
Os02g29550	AP2 domain ($5e^{-8}$)	M<F≈P	1.96	HPD
Os02g32040	AP2 domain ($5e^{-9}$)	P≈F<M	1.69	LPD
Os02g36930	MADS family ($4e^{-22}$)	F<M≈P	1.15	UDO
Os02g38090	AP2 domain ($1e^{-6}$)	M≈F<P	1.48	ND
Os02g44360	GRAS family ($8e^{-51}$)	M<F≈P	1.97	HPD
Os02g47060	WRKY family ($1e^{-24}$)	P<F<M	1.66	Additivity

Os02g47190	myb-like family ($4e^{-5}$)	M<F≈P	1.46	HPD
Os02g56250	GATA zinc finger ($1e^{-9}$)	P<F≈M	1.42	HPD
Os03g06630	HSF-type DNA-binding domain ($1e^{-15}$)	M<F<P	2.23	Additivity
Os03g13790	Myb-like family (0.002)	M<F<P	1.58	Additivity
Os03g16780	Ankyrin repeats ($1e^{-17}$)	P<F<M	2.50	Additivity
Os03g21710	WRKY family ($1e^{-22}$)	P<F<M	2.12	Additivity
Os03g29610	myb-like family ($3e^{-69}$)	M<F<P	5.78	Additivity
Os03g48450	GRAS family ($3e^{-94}$)	M<F≈P	1.41	HPD
Os03g48970	CCAAT-binding ($3e^{-23}$)	M<F<P	1.82	Additivity
Os03g55110	Ankyrin repeats ($2e^{-29}$)	P≈F<M	2.22	LPD
Os03g61030	Transcription antiterminational factor ($5e^{-13}$)	P<F≈M	1.96	ND
Os04g33950	Transcription factor E2F/dimerisation ($5e^{-23}$)	P<F≈M	1.82	ND
Os04g36730	PHD-finger ($1e^{-6}$)	M<F<P	1.90	Additivity
Os04g45810	homeodomain leucine zipper protein ($3e^{-11}$)	M≈F<P	2.02	ND
Os04g55510	Zinc finger, C3HC4 type ($9e^{-9}$)	M<F≈P	1.66	HPD
Os05g02420	Myb-like family ($6e^{-4}$)	M<F<P	1.80	Additivity
Os05g03900	WRKY family ($9e^{-20}$)	M<F≈P	1.81	HPD
Os05g23320	Ankyrin repeats ($1e^{-19}$)	P<F<M	3.84	ND
Os05g45410	HSF-type family ($1e^{-36}$)	P<F<M	2.32	Additivity
Os05g46020	WRKY family ($7e^{-24}$)	P<F<M	2.08	Additivity
Os05g50340	Myb-like family ($5e^{-20}$)	M<F≈P	1.92	HPD
Os06g13000	Ankyrin repeats ($7e^{-28}$)	P<F<M	2.63	ND
Os06g14710	Myb-like family ($2e^{-6}$)	M<F≈P	1.52	ND
Os06g24070	Golden 2-like ($2e^{-115}$)	M<F<P	2.03	Additivity

Os06g28630	Myb-like family ($3e^{-12}$)	M<F<P	2.63	Additivity
Os06g50870	DNL zinc finger ($2e^{-21}$)	M<F≈P	2.07	HPD
Os06g51260	myb-like family ($1e^{-9}$)	M<F<P	2.60	Additivity
Os07g02800	myb-like family ($2e^{-33}$)	M≈F<P	1.59	LPD
Os07g07080	Regulator of chromosome condensation ($4e^{-21}$)	M<F≈P	1.74	HPD
Os07g38750	AP2 domain ($7e^{-10}$)	P≈F<M	2.17	ND
Os07g46500	Ankyrin repeats ($3e^{-22}$)	P<F<M	1.85	Additivity
Os07g48570	Dof domain, zinc finger ($8e^{-28}$)	M≈F<P	1.62	ND
Os07g49460	Response regulator ($3e^{-53}$)	M<F≈P	1.80	HPD
Os08g06280	LSD1 zinc finger ($3e^{-4}$)	M≈F<P	1.82	LPD
Os08g08120	B-box zinc finger ($2e^{-8}$)	M<F≈P	2.90	HPD
Os08g09690	CCAAT-binding ($3e^{-23}$)	M<F≈P	1.54	HPD
Os08g38990	WRKY family ($1e^{-27}$)	M<F≈P	1.70	ND
Os08g39450	AN1-like Zinc finger ($6e^{-10}$)	M≈P<F	1.11	ODO
Os08g43340	HSF-type DNA-binding family ($4e^{-34}$)	P<F≈M	1.92	ND
Os09g01960	tuber-specific and sucrose-responsive ($3e^{-57}$)	M<P<F	1.43	ODO
Os09g03680	Ankyrin repeats ($2e^{-20}$)	M<F<P	2.84	Additivity
Os09g06460	B-box zincfinger ($3e^{-5}$)	M<F≈P	3.23	HPD
Os09g09630	WRKY transcription factor 70 ($2e^{-11}$)	M<F<P	3.49	Additivity
Os09g19570	Uncharacterized	P≈F<M	1.58	ND
Os09g25060	WRKY family ($3e^{-22}$)	F<P≈M	1.21	ODO
Os09g31390	bZIP family ($3e^{-7}$)	M<F≈P	1.51	HPD
Os09g33810	Ankyrin repeats HBP-1 ($1e^{-20}$)	P≈F<M	1.75	ND
Os09g35880	B-box zinc finger ($3e^{-20}$)	M<F<P	2.14	Additivity
Os10g26620	Dof domain, zinc finger ($3e^{-29}$)	M≈F<P	1.45	ND
Os10g30840	AP2 domain ($4e^{-8}$)	P<F<M	2.38	Additivity

Os10g34180	BTF3-like ($4e^{-55}$)	P<F≈M	1.66	HPD
Os11g11100	bZIP family (0.006)	M<F<P	1.75	Additivity
Os11g28270	Zinc finger CCCH domain ($2e^{-5}$)	M≈F<P	1.42	LPD
Os11g32110	Auxin response factor ($8e^{-32}$)	M<F<P	1.89	Additivity
Os12g06520	bZIP family ($6e^{-59}$)	M<F≈P	1.43	HPD
Os12g41880	CAAT-binding ($4e^{-21}$)	M≈F<P	1.69	ND

[a] Conserved protein domain corresponding to individual gene was searched in NCBI protein databases by using BLASTP.

[b] $P < 0.05$, M, female parent; P, male parent; F, F_1 hybrid.

[c] Fold change between parent with the highest level of gene expression and another parent is shown as a ratio.

[d] Gene expression in F_1 hybrid is classified into multiple patterns. HPD indicates high-parent dominance; LPD, low-parent dominance; ODO, overdominance; UDO, underdominance. ND, undistinguishable model from additivity and non-additivity.

Legends of Supplemental Figures

Supplemental Figures 1

Verification of microarray data by semi-quantitative RT-PCR for *Nipponbare* x *93-11* cross (**A**) and *PA64s* x *93-11* cross (**B**). Expression levels of selected genes were shown under the gene ID, which order was consistent with the ones of electrophoresis. The rice actin gene was used as loading control. * indicates that transcription level of these genes were inconsistent with the microarray data.

Supplemental Figure 2

Functional categories of differentially expressed genes. Overall differentially expressed genes (**A**) and non-additive expressed genes (**B**) from *PA64s* x *93-11* cross. Overall differentially expressed genes (**C**) and non-additive expressed genes (**D**) from *Nipponbare* x *93-11* cross.

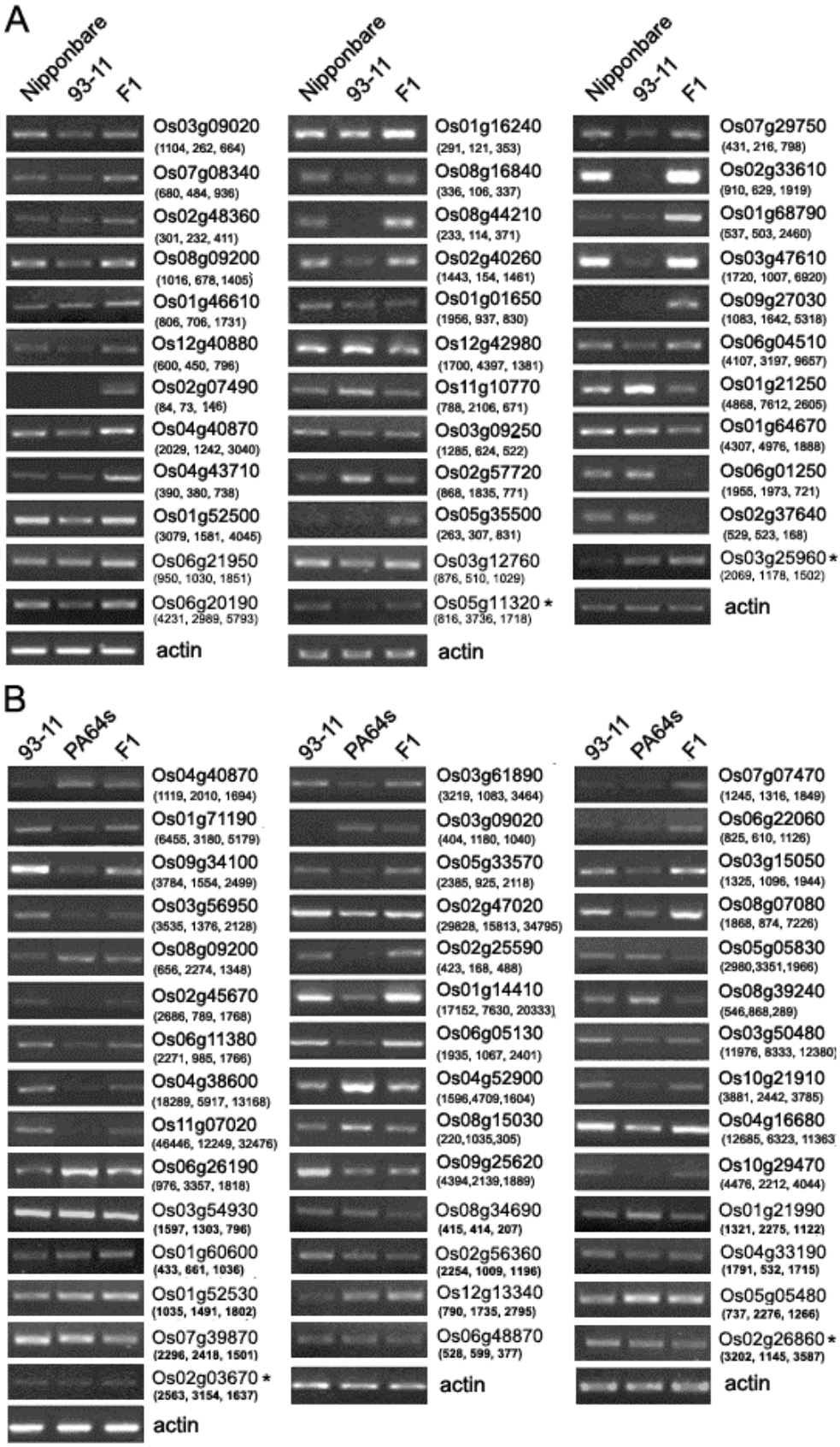
Supplemental Figure 3

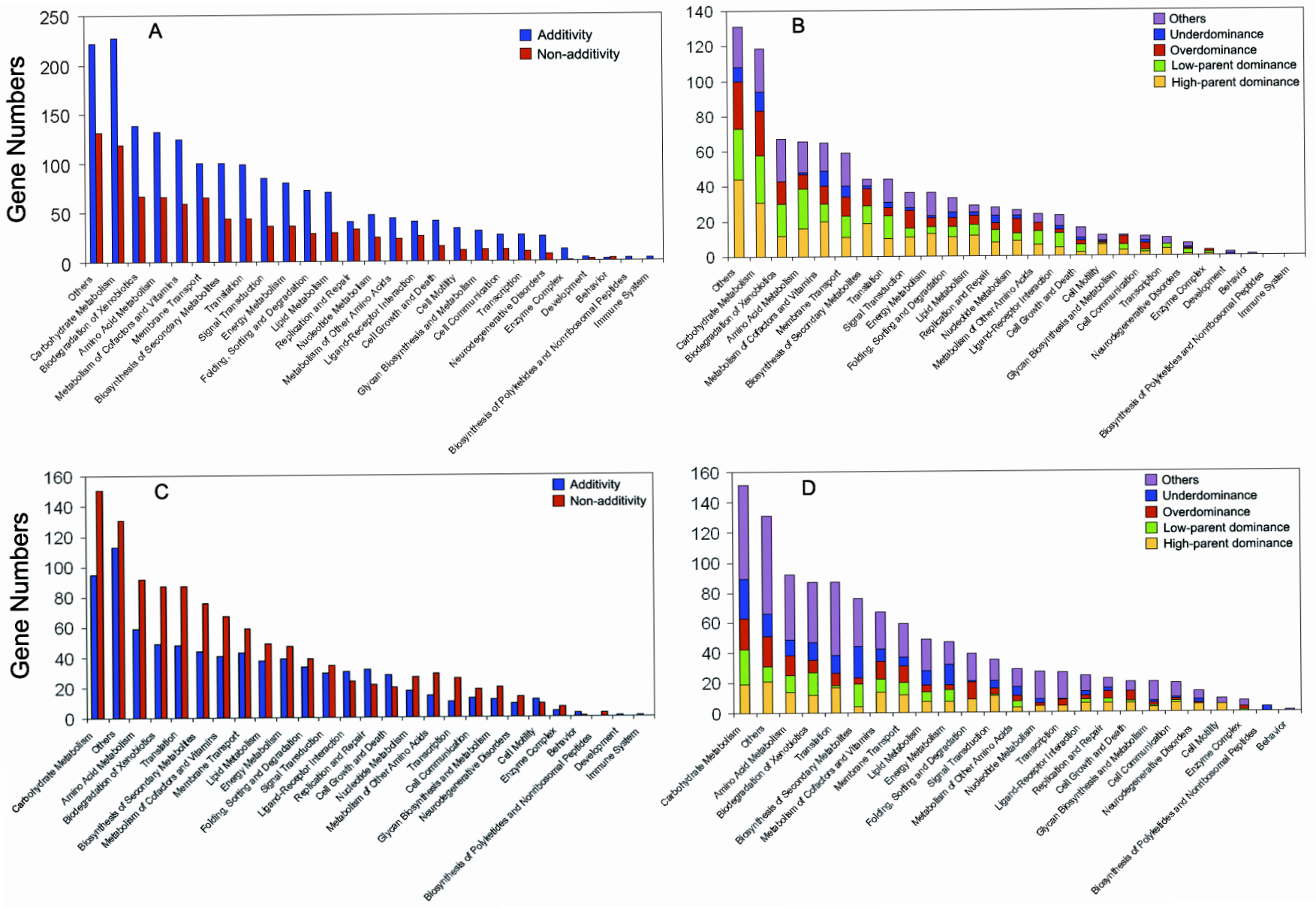
Comparison of representative biosynthesis pathways by the ratios of expression levels among inbred parents and their F1 hybrid, F1 hybrid versus female parent (Left), F1 hybrid versus male parent (Middle) and male parent versus female parent (Right). Each pathway is shown as glyphs consisting of nodes, which represent the metabolites, and lines, which represent the reaction. Expression level change of each reaction is shown in a color representing the ratio of two genotypes. Missing gene expression data, which may come from lack of annotated enzyme, lack of microarray probe, or lack of expression, are represented by gray lines. Red represents up-expression; green, down-expression; and blue, no detectable-expression change.

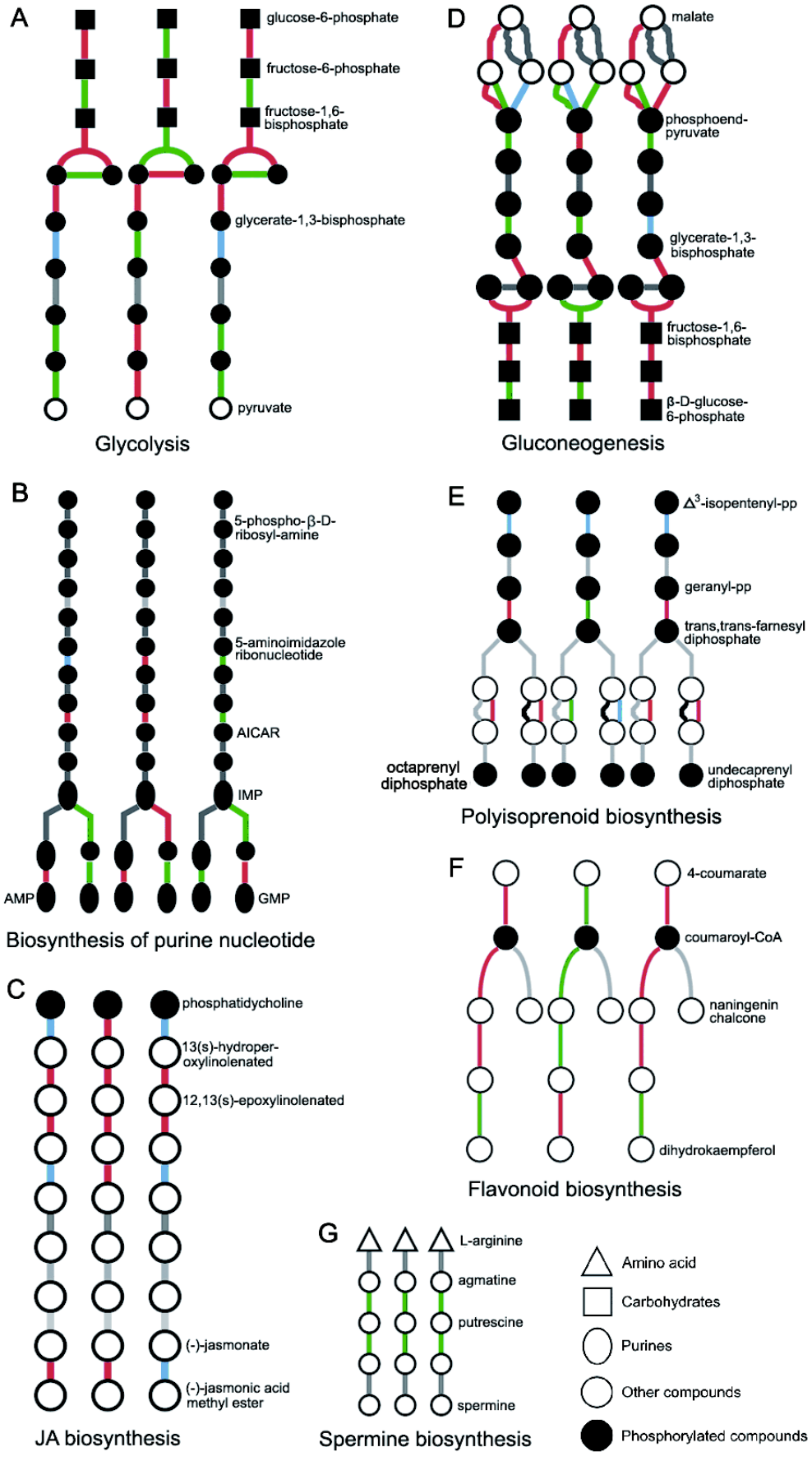
Supplemental Figure 4

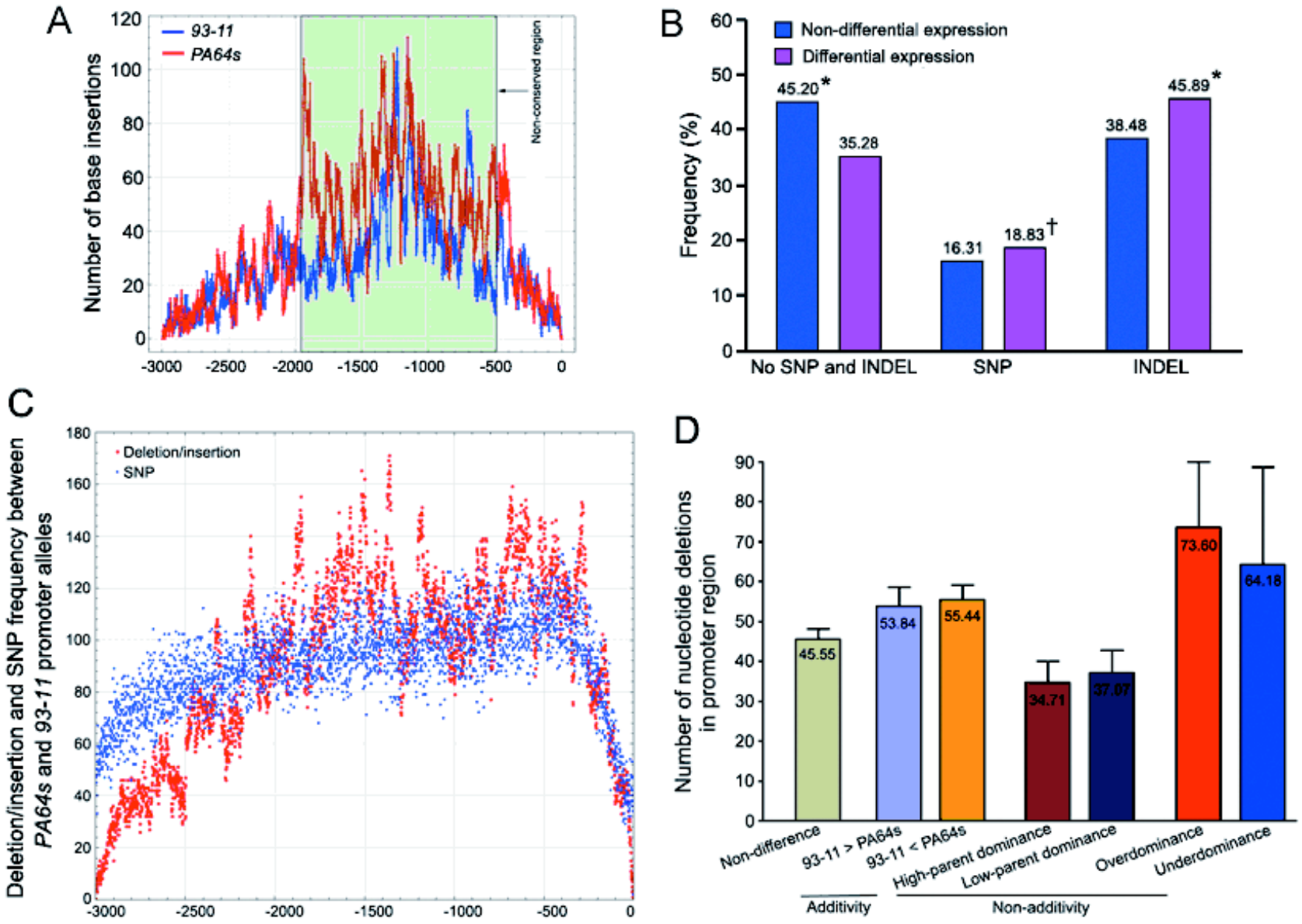
Comparative analysis of promoter variation between inbred parental *PA64s* and *93-11* promoter alleles. **(A)** Frequency of base insertions. 3 kb upstream regulatory regions of homologous gene

pairs between inbred parental genomes were extracted and aligned to detect the sequence polymorphic frequency. Aqua boxes point to the position of promoter regions with higher sequence variation (-500 to -2000 region from the start code ATG). **(B)** Frequency of promoter sequence variation with inbred parents between differentially and non-differentially expressed genes. A λ^2 test was used to determine the statistical significance between the two group genes. $^{*}P < 0.01$, † no significant difference. **(C)** Frequency comparison of deleted/inserted sequence and SNP (single nucleotide polymorphism) between inbred parental promoter alleles. Deletion/insertion and SNP are in red and blue, respectively. **(D)** Promoter deletion frequency between non-differentially and differentially expressed genes (including additive and non-additive expression pattern). Number in each column represents average length of deleted nucleotides.


A
Nipponbare
93-11
F1
Os03g09020 (1104, 262, 664)
Os07g08340 (680, 484, 936)
Os02g48360 (301, 232, 411)
Os08g09200 (1016, 678, 1405)
Os01g46610 (806, 706, 1731)
Os12g40880 (600, 450, 796)
Os02g07490 (84, 73, 146)
Os04g40870 (2029, 1242, 3040)
Os04g43710 (390, 380, 738)
Os01g52500 (3079, 1581, 4045)
Os06g21950 (950, 1030, 1851)
Os06g20190 (4231, 2989, 5793)
actin
Os01g16240 (291, 121, 353)
Os08g16840 (336, 106, 337)
Os08g44210 (233, 114, 371)
Os02g40260 (1443, 154, 1461)
Os01g01650 (1956, 937, 830)
Os12g42980 (1700, 4397, 1381)
Os11g10770 (788, 2106, 671)
Os03g09250 (1285, 624, 522)
Os02g57720 (868, 1835, 771)
Os05g35500 (263, 307, 831)
Os03g12760 (876, 510, 1029)
Os05g11320 * (816, 3736, 1718)
actin
Os07g29750 (431, 216, 798)
Os02g33610 (910, 629, 1919)
Os01g68790 (537, 503, 2460)
Os03g47610 (1720, 1007, 6920)
Os09g27030 (1083, 1642, 5318)
Os06g04510 (4107, 3197, 9657)
Os01g21250 (4868, 7612, 2605)
Os01g64670 (4307, 4976, 1888)
Os06g01250 (1955, 1973, 721)
Os02g37640 (529, 523, 168)
Os03g25960 * (2069, 1178, 1502)
actin
B
93-11
PA64s
F1
Os04g40870 (1119, 2010, 1694)
Os01g71190 (6455, 3180, 5179)
Os09g34100 (3784, 1554, 2499)
Os03g56950 (3535, 1376, 2128)
Os08g09200 (656, 2274, 1348)
Os02g45670 (2686, 789, 1768)
Os06g11380 (2271, 985, 1766)
Os04g38600 (18289, 5917, 13168)
Os11g07020 (46446, 12249, 32476)
Os06g26190 (976, 3357, 1818)
Os03g54930 (1597, 1303, 796)
Os01g60600 (433, 661, 1036)
Os01g52530 (1035, 1491, 1802)
Os07g39870 (2296, 2418, 1501)
Os02g03670 * (2563, 3154, 1637)
actin
Os03g61890 (3219, 1083, 3464)
Os03g09020 (404, 1180, 1040)
Os05g33570 (2385, 925, 2118)
Os02g47020 (29828, 15813, 34795)
Os02g25590 (423, 168, 488)
Os01g14410 (17152, 7630, 20333)
Os06g05130 (1935, 1067, 2401)
Os04g52900 (1596,4709,1604)
Os08g15030 (220,1035,305)
Os09g25620 (4394,2139,1889)
Os08g34690 (415, 414, 207)
Os02g56360 (2254, 1009, 1196)
Os12g13340 (790, 1735, 2795)
Os06g48870 (528, 599, 377)
actin
Os07g07470 (1245, 1316, 1849)
Os06g22060 (825, 610, 1126)
Os03g15050 (1325, 1096, 1944)
Os08g07080 (1868, 874, 7226)
Os05g05830 (2980,3351,1966)
Os08g39240 (546,868,289)
Os03g50480 (11976, 8333, 12380)
Os10g21910 (3881, 2442, 3785)
Os04g16680 (12685, 6323, 11363)
Os10g29470 (4476, 2212, 4044)
Os01g21990 (1321, 2275, 1122)
Os04g33190 (1791, 532, 1715)
Os05g05480 (737, 2276, 1266)
Os02g26860 * (3202, 1145, 3587)
actin


A
Gene Numbers
Additivity
Non-additivity
Others
Carbohydrate Metabolism
Biodegradation of Xenobiotics
Amino Acid Metabolism
Metabolism of Cofactors and Vitamins
Membrane Transport
Biosynthesis of Secondary Metabolites
Translation
Signal Transduction
Energy Metabolism
Folding, Sorting and Degradation
Lipid Metabolism
Replication and Repair
Nucleotide Metabolism
Metabolism of Other Amino Acids
Ligand-Receptor Interaction
Cell Growth and Death
Cell Motility
Glycan Biosynthesis and Metabolism
Cell Communication
Transcription
Neurodegenerative Disorders
Enzyme Complex
Development
Behavior
Biosynthesis of Polyketides and Nonribosomal Peptides
Immune System
B
Others
Underdominance
Overdominance
Low-parent dominance
High-parent dominance
C
Gene Numbers
Additivity
Non-additivity
D
Others
Underdominance
Overdominance
Low-parent dominance
High-parent dominance


A
glucose-6-phosphate
fructose-6-phosphate
fructose-1,6-bisphosphate
glycerate-1,3-bisphosphate
pyruvate
Glycolysis
D
malate
phosphoend-pyruvate
glycerate-1,3-bisphosphate
fructose-1,6-bisphosphate
β-D-glucose-6-phosphate
Gluconeogenesis
B
5-phospho-β-D-ribosyl-amine
5-aminoimidazole ribonucleotide
AICAR
IMP
AMP
GMP
Biosynthesis of purine nucleotide
E
Δ3-isopentenyl-pp
geranyl-pp
trans,trans-farnesyl diphosphate
octaprenyl diphosphate
undecaprenyl diphosphate
Polyisoprenoid biosynthesis
F
4-coumarate
coumaroyl-CoA
naningenin chalcone
dihydrokaempferol
Flavonoid biosynthesis
C
phosphatidycholine
13(s)-hydroper-oxylinolenated
12,13(s)-epoxylinolenated
(-)-jasmonate
(-)-jasmonic acid methyl ester
JA biosynthesis
G
L-arginine
agmatine
putrescine
spermine
Spermine biosynthesis
Amino acid
Carbohydrates
Purines
Other compounds
Phosphorylated compounds


A
93-11
PA64s
Non-conserved region
Number of base insertions
120
100
80
60
40
20
0
-3000
-2500
-2000
-1500
-1000
-500
0
B
Non-differential expression
Differential expression
Frequency (%)
60
50
40
30
20
10
0
45.20*
35.28
16.31
18.83†
38.48
45.89*
No SNP and INDEL
SNP
INDEL
C
Deletion/insertion
SNP
Deletion/insertion and SNP frequency between PA64s and 93-11 promoter alleles
180
160
140
120
100
80
60
40
20
0
-3000
-2500
-2000
-1500
-1000
-500
0
D
Number of nucleotide deletions in promoter region
90
80
70
60
50
40
30
20
10
0
45.55
53.84
55.44
34.71
37.07
73.60
64.18
Non-difference
93-11 > PA64s
93-11 < PA64s
High-parent dominance
Low-parent dominance
Overdominance
Underdominance
Additivity
Non-additivity