

Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis

Lin Shao^{a,1}, Feng Xing^{a,1}, Conghao Xu^a, Qinghua Zhang^a, Jian Che^a, Xianmeng Wang^a, Jiaming Song^a, Xianghua Li^a, Jinghua Xiao^a, Ling-Ling Chen^a, Yidan Ouyang^a, and Qifa Zhang (张启发)^{a,2}

^aNational Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, 430070 Wuhan, China

Contributed by Qifa Zhang, January 24, 2019 (sent for review December 4, 2018; reviewed by Z. Jeffrey Chen, Elizabeth S. Dennis, and Mei Guo)

Utilization of heterosis has greatly increased the productivity of many crops worldwide. Although tremendous progress has been made in characterizing the genetic basis of heterosis using genomic technologies, molecular mechanisms underlying the genetic components are much less understood. Allele-specific expression (ASE), or imbalance between the expression levels of two parental alleles in the hybrid, has been suggested as a mechanism of heterosis. Here, we performed a genome-wide analysis of ASE by comparing the read ratios of the parental alleles in RNA-sequencing data of an elite rice hybrid and its parents using three tissues from plants grown under four conditions. The analysis identified a total of 3,270 genes showing ASE (ASEGs) in various ways, which can be classified into two patterns: consistent ASEs such that the ASE was biased toward one parental allele in all tissues/conditions, and inconsistent ASEs such that ASE was found in some but not all tissues/conditions, including direction-shifting ASEs in which the ASE was biased toward one parental allele in some tissues/conditions while toward the other parental allele in other tissues/conditions. The results suggested that these patterns may have distinct implications in the genetic basis of heterosis: The consistent ASEs may cause partial to full dominance effects on the traits that they regulate, and direction-shifting ASEs may cause overdominance. We also showed that ASEs were significantly enriched in genomic regions that were differentially selected during rice breeding. These ASEs provide an index of the genes for future pursuit of the genetic and molecular mechanism of heterosis.

heterosis | Shanyou 63 | RNA sequencing | allele-specific expression

Heterosis refers to the superior performance of hybrids relative to their parents. Utilization of heterosis has greatly increased productivity of many crops worldwide in the last century. Tremendous progress has been made in characterizing the genetic basis of heterosis in the last two decades with the advent of genomic technologies based on the framework of the classical dominance and overdominance hypotheses (1–4). For example, dissection of genetic components using an immortalized F_2 (IMF₂) population developed from a cross of an elite rice hybrid resolved and quantified the contributions of dominance, overdominance, and epistasis in heterosis of this hybrid (5–7). Genome-wide association studies were also applied to identify genetic controls of a large number of rice hybrids representing different decades of hybrid rice breeding, which established the link between many agronomic traits and candidate genes (3, 8). However, molecular mechanisms underlying these genetic components are much less understood.

Gene expression is a complex process that is regulated by genetic and epigenetic variations in response to developmental and environmental cues (9). Results from expression quantitative trait locus (eQTL) analysis showed that both *cis*- and *trans*-elements regulate the expression of the genes (10). Allele-specific expression (ASE) refers to the characteristic of

preferentially expressing a parental allele in the hybrid due to variations in regulatory sequences from the parental genomes (11). The expression difference caused by ASE may lead to phenotypic variation depending on the function of the genes. In mammals, ASE is often associated with epigenetic inactivation in X-chromosome and genomic imprinting as well as nonimprinted autosomal genes (12). ASE has also been demonstrated in plants, including maize (13–22), *Arabidopsis* (23–27), rice (28, 29), and barley (30). Several studies have suggested that ASE plays a role in heterosis because genetic variations frequently cause gene expression difference, which may lead to phenotypic variations (1, 14–20). The development of RNA-sequencing (RNA-seq) technologies has allowed unbiased and highly reproducible deep sequencing of whole transcriptomes. This enables the detection of single-nucleotide polymorphisms (SNPs), which can be used to distinguish parental alleles and identify genes showing ASE in heterozygotes (16). Such RNA-seq technology-based ASE identification may be greatly facilitated if high-quality genome sequences of both parents are available.

In this study, we performed a genome-wide analysis of ASE by comparing the read ratios of SNPs of the parental alleles in an elite rice hybrid (Shanyou 63) and its parents [Zhenshan 97 (ZS97) and Minghui 63 (MH63)] using RNA-seq data of seedling shoot, flag leaf, and young panicle of plants grown under

Significance

Utilization of heterosis has greatly increased productivity of many crops globally. Allele-specific expression (ASE) has been suggested as a mechanism for causing heterosis. We performed a genome-wide analysis of ASE in three tissues of an elite rice hybrid grown under four conditions. The analysis identified 3,270 genes showing various patterns of ASE in response to developmental and environmental cues, which provides a glimpse of the ASE landscape in the hybrid genome. We showed that the ASE patterns may have distinct implications in the genetic basis of heterosis, especially in light of the classical dominance and overdominance hypotheses. The genes showing ASE provide the candidates for future studies of the genetic and molecular mechanism of heterosis.

Author contributions: Qifa Zhang designed research; L.S., C.X., Qinghua Zhang, and J.C. performed research; X.L., J.X., and L.-L.C. contributed new reagents/analytic tools; L.S., F.X., X.W., J.S., and Y.O. analyzed data; and L.S. and Qifa Zhang wrote the paper.

Reviewers: Z.J.C., The University of Texas at Austin; E.S.D., Commonwealth Scientific and Industrial Research Organization; and M.G., Beidahuang KenFeng Seed Co., Ltd.

The authors declare no conflict of interest.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

¹L.S. and F.X. contributed equally to this work.

²To whom correspondence should be addressed. Email: qifazh@mail.hzau.edu.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1820513116/-DCSupplemental.

Published online March 4, 2019.

four environmental conditions. This study was made possible because of the recent availability of the high-quality reference genome sequences of the two parental lines (31). The analysis identified a total of 3,270 genes showing ASE (ASEGs) in various ways, which can be classified into two major patterns. Our analysis suggests that these patterns of ASEs may have distinct implications in the genetic and molecular basis of heterosis.

Results

Identification of ASEGs. Rice plants of ZS97, MH63, and their hybrid were planted in growth chambers set for four different conditions: high temperature (32 °C/28 °C)/long day (14 h light and 10 h darkness) (HTLD); low temperature (25 °C/22 °C)/long day (LTLD); high temperature/short day (10 h light and 14 h darkness) (HTSD); and low temperature/short day (LTSD). Seedling shoot at four-leaf stage, flag leaf, and panicle at the day of heading were collected for RNA extraction. Total RNA extracted from 72 samples (three genotypes × three tissues × four treatments × two biological replicates) was sequenced (RNA-seq) using Illumina HiSeq2000.

We used the 1,300,802 SNPs between ZS97RS1 and MH63RS1 identified previously (31) as the reference for ASE calling. A schematic overview of procedure for identifying ASEGs is depicted in *SI Appendix, Fig. S1*. The numbers of trimmed high-quality reads (see *Materials and Methods*) are assembled and listed in *Dataset S1* for two parents and in *Dataset S2* for F₁. To illustrate how ASE and ASEGs were identified, SNP read counts and probabilities for equal frequencies of the two parental sequences in the hybrid are presented for a sample of 10 genes in *Dataset S3*. Three of the 10 genes, harboring 8, 11, and 5 SNPs, were regarded as non-ASEGs because the read counts of the two parental sequences for none of the SNPs deviated significantly from the 1:1 ratio. Four genes were classified as ASEGs, in which ratios of parental read counts for all of the SNPs of each gene were significantly different and biased in the same directions. Three genes were considered to be variable ASEGs, in which the read counts of the two parental sequences were significantly biased toward one parent at some of the SNPs of the gene but not significantly biased at the rest of SNPs of the same gene. Such variations of ASE at different SNPs of the genes may be the result of alternative transcription start sites, alternative polyadenylation site usage, or allele-specific alternative splicing (32).

There were 3 to 4% of genes among the ASEGs in each tissue/condition combination (TCC) whose read counts of the two parental sequences were significantly biased toward one parent at some of the SNPs of the gene but biased toward the other parent at other SNPs of the same gene. The reason is beyond the scope of this study; thus, these genes were not included in subsequent analysis.

A total of 3,270 genes showed ASE in at least one of the TCCs (Fig. 1A). The information for loci showing ASE is listed in *Datasets S4* and *S7* for shoot, *Datasets S8* and *S11* for flag leaf, and *Datasets S12* and *S15* for panicle under four growth conditions, including their references in the ZS97RS1, MH63RS1, and Nipponbare (MSU 7) genomes, the read count ratio of MH63 allele to the total counts (MH63 allele + ZS97 allele) from all significant SNPs within an ASEG, and the *Q* values for quantitative measurements.

As illustrated in Fig. 1B–D, the number of ASEGs was the largest in shoot (1,047 to 1,339), followed by flag leaf (1,007 to 1,286) and panicle (990 to 1,105). Among the four growth conditions, the number of ASEGs was the largest under HTLD across all three tissues (1,105 to 1,339), but the number ranking of the other three conditions varied among the tissues. For ease of description, we refer to ASE showing higher expression levels of the MH63 alleles as paternal ASE, and refer to ASE showing higher expression levels of the ZS97 alleles as maternal ASE.

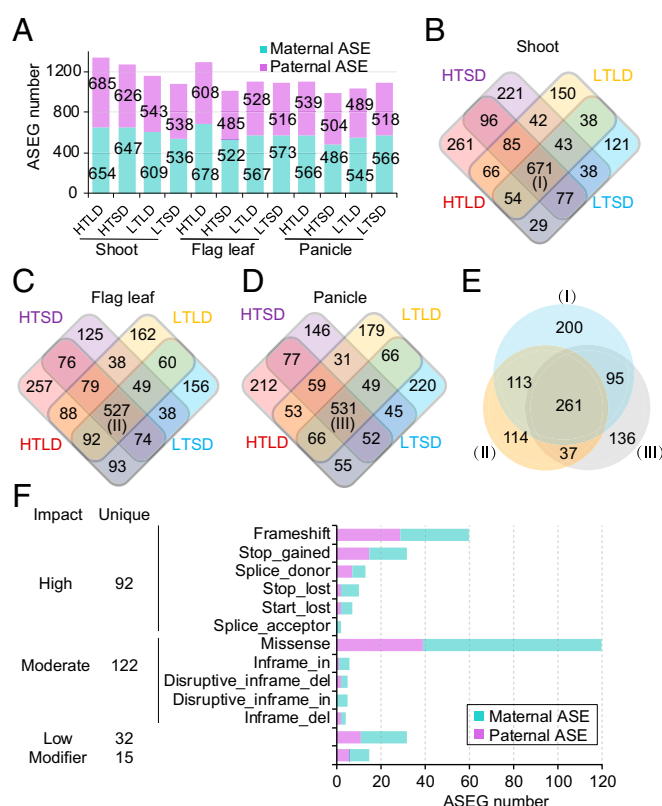


Fig. 1. Summary and features of ASEGs. (A) Numbers of ASEGs in shoot, flag leaf, and panicle under four conditions. (B–D) Four-way Venn diagrams displaying the numbers of ASEGs in shoot (B), flag leaf (C), and panicle (D) under HTLD, LTLD, HTSD, and LTSD. The numbers of ASEGs that were detected in all four conditions are indicated as shoot (I), flag leaf (II), and panicle (III). (E) Three-way Venn diagram showing the numbers of ASEGs that overlap in the three tissues based on the ASEGs of groups I, II, and III in B–D. The 261 genes that showed consistent ASE in terms of the direction of expression bias in all three tissues at all four growth conditions are regarded as consistent ASEGs. (F) Features of consistent ASEGs with high, moderate, low, and modifier impact variations that are caused by SNPs and indels between ZS97RS1 and MH63RS1. The unique numbers of the impact categories are indicated, with certain overlapping among the subclassifications within each impact category. del, deletion; in, insertion.

The numbers of maternal ASEGs were slightly larger than paternal ones in 10 of the 12 TCCs (Fig. 1A).

Genes Showing Consistent ASE Across Tissues and Conditions. A comparison of the 3,270 ASEGs revealed 261 genes that showed consistent ASE in terms of the directions of expression bias in three tissues and four growth conditions, which included 160 maternal and 101 paternal ASEGs (Fig. 1E and *Dataset S16*). The bias levels of consistent ASEGs in F₁ were almost perfectly correlated with the relative expression levels of the parental genes ($r^2 = 0.98$), indicating that the ASE was strongly affected by the expression levels of the parental genes (*SI Appendix, Fig. S2*). However, there were 25 genes whose ASE levels were not in accord to their expression levels in the parents in at least 1 of 12 TCCs (*SI Appendix, Figs. S2 and S3*). ASE levels for 5 of the 25 genes were not in accord to the relative expression levels of the parental alleles in seven or more TCCs, including MH02g0028800/ZS02g0029600 and MH09g0380200/ZS06g0335200 in 10 TCCs, MH06g0615800/ZS06g0568400 in 9 TCCs, and MH03g0174000/ZS03g0173000 and MH05g0510300/ZS05g0561000 in 7 TCCs. ASE levels of 13 genes did not accord with the expression levels of the parental alleles in only one TCC.

To gain insight into the possible impacts caused by the variations of the ASEs, we compared coding sequences between the two parental alleles of these 261 genes that showed ASE in F₁, with MH63RS1 as the reference and using SnpEff. The results showed that 92 of the 261 ASEs, including 47 maternal and 45 paternal ASEs, harbored SNPs and insertions/deletions (indels) with high impacts, which may cause protein truncation, loss of function, or triggering of nonsense-mediated decay as defined by SnpEff (Fig. 1F). The largest class ($n = 60$) was frameshift_variants due to indels of nucleotides with numbers not in multiples of three, which would cause disruptions of the translational reading frame of the genes. The percentage (35.2%) of genes with high-impact variations among the ASEs was much larger than the percentage (28.5%; 8,088/28,405) of genes with high-impact variations in the whole genome, also identified based on the MH63RS1 (31). Moreover, 122 (46.7%) of the 261 ASEs, including 83 maternal and 39 paternal ASEs, harbored SNPs and indels with moderate impacts, which were nondisruptive variants that might change protein effectiveness as defined by SnpEff (Fig. 1F). The largest class ($n = 120$) was missense_variants (nonsynonymous substitution). This percentage was again much larger than the percentage (40.4%; 11,465/28,405) of genes with moderate-impact variations in the whole genome. The remaining 47 genes belonged to two classes, including 32 (12.3%) as low-impact variations (unlikely to change protein sequence) and 15 (5.7%) as modifier (undetermined impact) (Fig. 1F).

A functional enrichment analysis of the consistent ASEs using InterPro classification showed that only NB-ARC (IPR002182) family proteins were significantly enriched ($P = 0.0192$). NB-ARC is a core nucleotide binding pocket of NBS-LRR proteins that binds specifically to and hydrolyzes ATP. The typical structure of NBS-LRR proteins consists of an N-terminal Toll/interleukin-1 receptor domain or coiled-coil (CC) domain, the NB-ARC domain, and the leucine-rich repeat (LRR) domain (33). The rice genome contains around 480 NBS-LRR genes, while the majority of the cloned *Magnaporthe oryzae* resistance (R) genes and two bacterial disease R genes encode NBS-LRR proteins (34). R proteins initiate effector-triggered immunity by recognizing highly variable avirulence effectors (35). Eight consistent ASEs were predicted to encode NBS-LRR proteins. The CC domain could be predicted in four of the proteins (CC-NBS-LRR, CNL), but not in the other four (SI Appendix, Fig. S44). The expression profile was examined in the parents and hybrid under the four conditions in each tissue; the biased levels of the parental alleles in the hybrid were in accord with the expression levels in the parents (SI Appendix, Fig. S44). Four of the eight genes showed large differences in the predicted proteins resulting from SNPs and indels (SI Appendix, Fig. S4B). For example, in the comparison of MH10g0068600 vs. ZS10g0086900 (a paternal ASE in F₁), a premature termination codon was generated by an SNP from AAG to TAG, causing a loss of Harbinger transposase-derived nuclease domain (IPR027806) in ZS10g0086900. Indeed, expression of the MH63 allele was much higher (SI Appendix, Fig. S44). In the comparison of MH11g0513800 vs. ZS11g0533900 (a maternal ASE in F₁), the MH63 protein was truncated due to premature termination that lost a PK-like domain (IPR000719); a higher transcript level of the ZS97 allele was detected (SI Appendix, Fig. S44). In the comparison of MH04g0027600 vs. ZS04g0021800 (a paternal ASE in F₁), a deletion of CGGT (position -244 to -241 from start codon) led to a sequence that was 216 aa shorter in the predicted protein of ZS04g0021800, which also affected the length of the NB-ARC domain; the transcript level of the MH63 allele was higher than that of the ZS97 allele (SI Appendix, Fig. S44). Finally, in the comparison of MH12g0303000 vs. ZS12g0325600 (a paternal ASE in F₁), five deletions were found in the ZS97 allele relative to the MH63 allele, four of

which were located in the LRR domain (SI Appendix, Fig. S4B). In addition, an insertion of CTCG (position 2788 to 2791 from start codon) resulted in a very complex structure with additional domains, including two transmembrane domains, an RX-CC-like domain, an NB-ARC domain, and an LRR domain in ZS12g0325600. The transcript level of the ZS97 allele was much lower than that of the MH63 allele, both in the parents and hybrid (SI Appendix, Fig. S44). These cases seem to provide examples for the notion that the hybrid is able to make better use of the favorable copies of the parental genes by specifically expressing them while keeping the level of the unfavored copies low, conforming to the genetic definition of dominance. This phenomenon might be related to the mechanism of nonsense-mediated mRNA decay, whereby the cell is able to terminate erroneous gene expression (36).

Genes Showing Direction-Shifting Patterns of ASE. The remaining 3,009 of the 3,270 ASEs showed inconsistent patterns in terms of directions of ASE among tissues and growth conditions, which can be divided into two major subgroups: 125 genes that showed ASE in opposite directions among the 12 TCCs (direction-shifting); and 2,884 genes that showed ASE not consistent among the 12 TCCs, but not with shifting directions.

We speculated that ASEs may show direction-shifting patterns in the hybrid relative to the parents among different TCCs such that the expression may be biased toward the paternal allele in one TCC and biased toward the maternal allele in another TCC (Dataset S17). The underlying assumption is that one of the alleles may function better in specific TCCs, and the hybrid is able to use the better allele in growth, development, and environmental adaptation, eventually leading to higher performance of the heterozygote than either of the parental homozygotes, which is referred to as genetic overdominance. We are thus specifically interested in the ASEs that displayed direction-shifting ASE, either from maternal to paternal or from paternal to maternal.

Such direction-shifting patterns were observed in a total of 67 ASEs among the four conditions in the three tissues, including 24 in shoot, 22 in flag leaf, and 27 in panicle (Fig. 2 and SI Appendix, Fig. S5); and in 105 ASEs among the three tissues under the four conditions, 47 of which overlap with the 67 ASEs in the previous category, including 32 under HTLD, 36 under HTSD, 24 under LTLD, and 33 under LTSD (Fig. 3 and

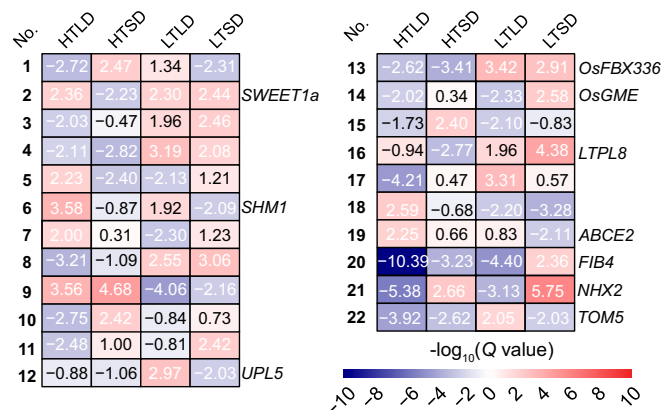


Fig. 2. Expression bias of genes showing direction-shifting patterns of ASE among four growth conditions in flag leaf. The heatmaps were generated with $-\log_{10}(Q \text{ value})$ of the significant SNP in color scale. Paternal ASE, expression bias >2 (numbers in white); maternal ASE, expression bias <-2 (numbers in white); no bias, expression bias between 2 and -2 (numbers in black). Genes are labeled with boldface ID numbers (see Dataset S17 for corresponding gene loci).

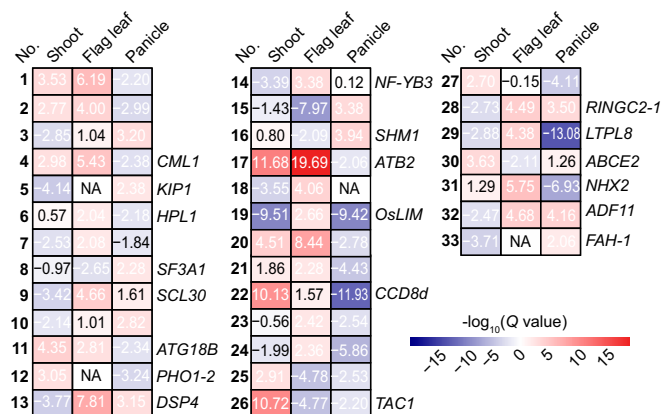


Fig. 3. Expression bias of genes showing direction-shifting patterns of ASE among three tissues under LTSD. The heatmaps were generated as depicted in Fig. 2. Paternal ASE, expression bias >2 (numbers in white); maternal ASE, expression bias <-2 (numbers in white); no bias, expression bias between 2 and -2 (numbers in black). Genes are labeled with boldface ID numbers (see Dataset S17 for corresponding gene loci).

SI Appendix, Fig. S6). Some genes showed direction-shifting patterns in two or more TCCs.

Of the ASEs showing direction-shifting patterns in response to growth conditions, we took *OsGME* for illustration. *GME* encodes a GDP-D-mannose-3,5-epimerase that catalyzes the conversion of GDP-D-mannose to GDP-L-galactose, which is a rate-limiting step of L-ascorbic acid biosynthetic pathway in plants (37). L-ascorbic acid (vitamin C) is one of the most abundant metabolites in green leaves and plays important roles as an antioxidant and an enzymatic cofactor involved in multiple processes (38, 39). In rice, the reaction products from GDP-D-mannose were GDP-L-galactose and GDP-L-glucose (40), both of which are immediate precursors in L-ascorbic acid biosynthesis. Four transcriptional isoforms of ZS97 and MH63 were identified; three of them, except transcript 1, shared identical coding sequences between ZS97 and MH63 based on the RNA-seq data (Fig. 4A). In transcript 1, a deletion of 5 bp (AAAAA) at position -60 and an insertion of 1 bp (A) at position 20 from ATG of MH10g0285600, compared with ZS10g0324200, resulted in a product 33 aa shorter in MH63 (420 aa) than that in ZS97 (453 aa). Transcript 2 was the shortest because of an alternative splicing event. The lengths of *OsGME* (378 aa) encoded by transcripts 3 and 4 were equal to that in MSU 7 (40). Despite different transcript lengths, two motifs—a nicotinamide-adenine dinucleotide (NAD^+) binding motif (GxxGxxG) and a catalytic domain (Ser and YxxxK)—were conserved among all transcription products (*SI Appendix, Fig. S7A*). In F_1 , expression of *OsGME* in flag leaf was biased to maternal allele under HTLD and LTLD but to paternal allele under LTSD (Figs. 2 and 4A). The direction-shifting pattern assessed by the second SNP under both long-day conditions and by the third SNP under LTSD indicated a differential response to photoperiod. A total of five SNPs were detected in the exons between the ZS97 and MH63 alleles of *OsGME*, but none of them affected the functional motifs. However, complex variations between ZS97 and MH63 sequences occurred in the promoter region. Several light-responsive *cis*-elements were predicted in the 5' region of two paternal sequences by the PLACE database (41), including a SORLREP3AT box, a REALPALGLHCB21 box, an AGCTT box, an IBOXCORE, a GATA box, and three GT1CONSENSUS boxes in ZS10g0324200; and a GT1CORE, two ARR1AT boxes, and a MYCCONSUSAT box in MH10g0285600, due to the SNPs and indels (*SI Appendix, Fig. S7B*). The GATA box was essential for phytochrome responsiveness and involved in Pfr-regulated gene expression (42, 43). The GT1 site is either regulated by light or acts as constitutive activating elements, depending on its context (43,

44). A heat stress-responsive *cis*-element, PRECONSCRHSP70A, was also predicted in ZS10g0324200, but not in MH10g0285600. Such differences in these *cis*-elements almost certainly cause differential responses of the genes to the environmental conditions.

An interesting example of the ASEs showing a direction-shifting pattern among the tissues was demonstrated by *Tiller Angle Control 1 (TAC1)*, a major QTL reported as regulating plant architecture by controlling tiller angle in rice such that

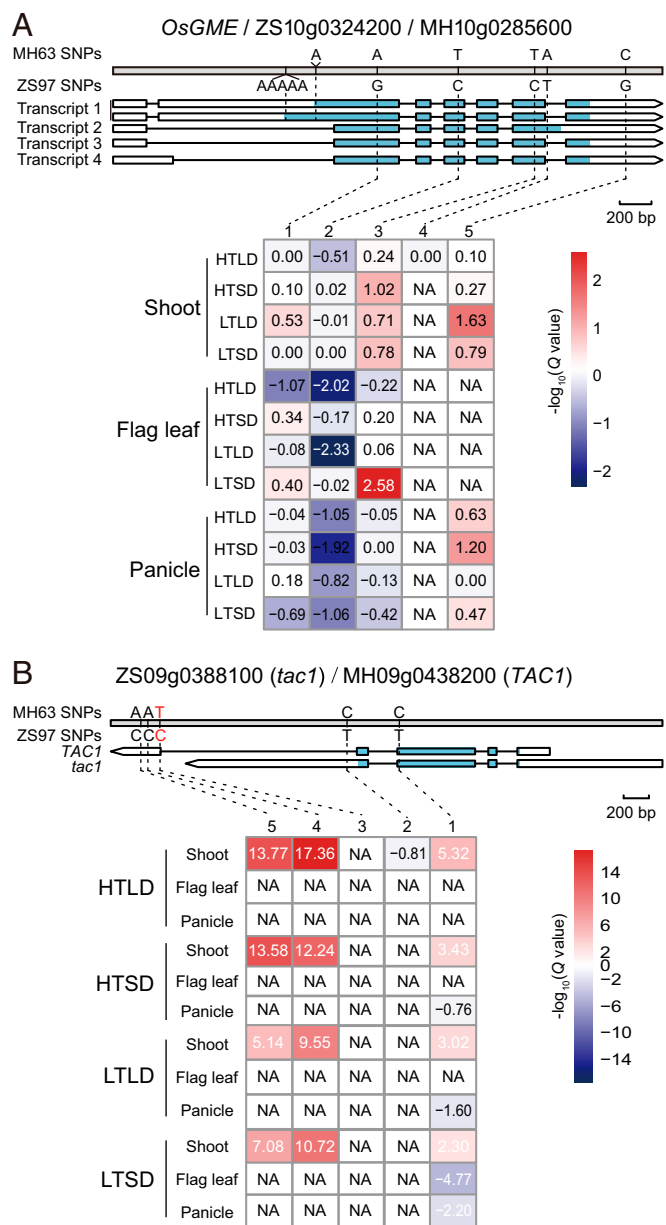


Fig. 4. Analysis of *OsGME* and *TAC1* transcripts. (A) Transcript structure of *OsGME* (ZS10g0324200/MH10g0285600) alleles based on the SNPs of genomic sequences from two parents (ZS97R51 and MH63R51). (B) Transcript structure of *TAC1* (MH09g0438200) and *tac1* (ZS09g0388100) alleles based on the SNPs of genomic sequences from two parents (ZS97R51 and MH63R51). The third SNP in red, adenine in *TAC1* and guanine in *tac1*, may affect splicing process, generating a long exon in *tac1*. The heatmaps show the expression bias of every SNP located in *OsGME* (A) and *TAC1/tac1* (B) by calculating $-\log_{10}(Q \text{ value})$ in scale, from three tissues under four conditions. The bar of sequence length is shown in each panel. NA, SNPs that were unable to meet the requirements of ASE identification process.

TAC1 corresponds to wide tillering angle, while *tac1* results in more compact plant (45). Many of the three-line hybrids are heterozygous at this locus and usually produce intermediate tiller angles compared with the parents, which may have contributed significantly to the increased productivity of hybrid rice (3). Recent studies also showed that *TAC1* homologs regulate branch angles in a range of plant species (46). Five SNPs were found in the *TAC1* gene sequence between MH63RS1 and ZS97RS1, in which the third SNP, mutated from AGGA in MH63 (*TAC1*/MH09g0438200) to GGGA in ZS97 (*tac1*/ZS09g0388100) in the 3' splicing site, affects the splicing of the fourth intron, generating a long exon in *tac1* (Fig. 4B). In F₁, the allelic bias of *TAC1* expression was detected in shoot under all four conditions, verified by the first, fourth, and fifth SNPs. However, the expression was biased to the *tac1* allele in flag leaf under LTSD and in panicle under HTLD and LTSD, as clearly indicated by the first SNP, despite lower expression levels than that in shoot (Figs. 3 and 4B).

We investigated the possible genetic effects of such dynamic differential expression on the performance of different genotypes using the immortalized F₂ population created by paired crosses of recombinant inbred lines derived from the cross between ZS97 and MH63 (5, 6). We calculated additive (A) and dominant (D) effects of Bin1244 and Bin1315 on yield, number of grains per panicle, number of spikelets per panicle, number of tillers per plant, and grain weight, which are the main determinants of yield, using data collected from the fields in 2 y (SI Appendix, Table S1). Using MH63RS1 as the reference, Bin1244 containing *TAC1* showed a significant overdominant effect ($D/|A| > 1$) on yield per plant in data from both years, mostly through changes in number of spikelets per panicle, suggesting that *TAC1* may be a locus contributing overdominance to heterosis. Thus, *TAC1* may not be simply a gene for tiller angle. It may be conjectured that the high expression of the MH63 allele for wide tillering angle in the seedling shoot allowed canopy development to rapidly cover the field at the vegetative stage, while the higher expression of the ZS97 allele in panicle and flag leaf may have a role in promoting reproductive growth and development. Bin1315 containing *OsgME* showed significant overdominant effects on grain weight in both years.

However, the precise details of the genetic effects of all of the ASEGs remain to be investigated in future studies for two reasons: A bin, on average, contains two- to three-dozen genes, and the background noise in the segregating population may often be higher than the effects of many genes. Thus, it is not possible to resolve the genetic effects to individual genes using this dataset.

ASEGs Located in Selected Regions. A previous study based on the low-coverage sequencing data of 1,479 rice accessions identified two major *indica/xian* subpopulations, *ind I* and *ind II*, and found 200 regions that were differentially selected between *ind I* and *ind II* (47). These regions spanned 7.8% of the rice genome and contained signatures of domestication or artificial selection, harboring around 4,000 nontransposable element genes, including many with functions that are associated with important agronomic traits (47). ZS97 is a member of *ind I*, and MH63 belongs to *ind II*. An enrichment analysis showed that the ASEGs identified in F₁ were highly enriched in the selected regions (408/3,270; $P = 2.07 \times 10^{-12}$, hypergeometric test), suggesting that many of the ASEGs might be targets of selection during the processes of rice breeding and production. In particular, genes containing two domains were significantly enriched in the selected regions by InterPro classification: PK, ATP binding site (IPR017441; $P = 0.0261$, $n = 18$), and P-loop containing nucleoside triphosphate hydrolase (IPR027417; $P = 0.0397$, $n = 5$).

Discussion

The genome-wide analysis of ASE of the RNA-seq data of three tissues and four growth conditions from an elite rice hybrid identified a large number of ASEGs that can be classified into two major patterns: inconsistent ASEGs (including direction-shifting ASEGs) and consistent ASEGs. From the perspective of heterosis study, these patterns may have direct and distinct implications on the classical genetic hypotheses of heterosis. In the consistent ASEGs, the expression of the gene in the hybrid is biased toward one of the parents in all of the TCCs examined in this study, allowing the possibility that some of the consistent ASEGs may become inconsistent if more tissues and growth conditions are investigated. As we showed on the basis of a limited number of genes with identifiable functional variations, such strong and consistently biased expression is likely caused by the fact that one of the parental alleles is favorable while the other allele is unfavored. This implies that the hybrid can make use of the favorable copies of the genes and express them at high levels. Such consistent biased expression of the genes would result in partially to fully dominant effects on the traits that are regulated by the genes. In a more general sense, the presence-vs.-absence type of variations also belong to this category, as exemplified by *Ghd7* (a major QTL for grain number, plant height, and heading date), which is present in MH63 but absent in ZS97 and exerts a large pleiotropic dominance effect on all of the traits (48). Complementarity as well as additivity of such dominance genetic effects between loci may explain a major portion of the genetic basis of heterosis (7).

It is also reasonable to assume that the two alleles in a hybrid, when both are functional, may perform differently in the varying developmental stages and/or environmental conditions, such that one allele may function better in some circumstances, while the other allele may be more superior in other circumstances. Accumulation of such differential advantageous effects of the two alleles in the hybrid may provide an important cause for heterosis, referred to as genetic overdominance, if the hybrid can make more use of the right allele at the right conditions by differential regulation of the two alleles. The detection of direction-shifting ASE of the genes among the tissues and growth conditions indicates that this hypothesis may possibly be correct at least for some of the genes, suggesting that the hybrid may be able to express the right allele higher in response to the environmental and developmental cues, which may result in overdominance.

However, because of the small phenotypic effects conferred by the majority of the genes, either in the category of consistent ASEGs that may cause dominant effects or in the category of direction-shifting ASEGs that may cause overdominant effects, our bin analysis of the IMF₂ data may not provide the sensitivity for detecting the effects because sizes of the bins are large, with each bin containing dozens of genes whose effects may cancel each other. In addition, this analysis may also be complicated by the noise from the segregation of the genomic background in the population.

Two *indica/xian* rice groups were identified in the Asian cultivated rice (47), *ind I* representing the germplasm from central and south China, and *ind II* from Southeast Asia, especially rice varieties from the International Rice Research Institute. The two groups were the result of differential selections in the breeding programs over decades, as clearly indicated by the existence of breeding signatures found in ~200 genomic regions. Consequently, hybrids between these two groups usually show strong heterosis, which has been widely used in hybrid rice breeding in China as well as in other countries. Enrichment of ASEGs in the selected regions indicates that ASE has also been part of the targets for selection in breeding. This also implies that such ASE

has contributed to heterosis between the two parental groups in hybrid rice breeding.

The 3,270 ASEGs made up ~6% of the gene models in both the ZS97 and MH63 genomes. However, this number should be taken as viewed from only one angle, as there were several technical limitations involved in generating the dataset besides the limited number of tissues and growth conditions. First, there is a portion of the genes that show present/absent variation between the two genomes (31). While these genes may be highly important for heterosis, they would not be detected as showing ASE because the other copies of the genes were absent. Second, genes with indel-type polymorphisms may similarly be missed in the alignment and thus not included in ASE identification. Third, in a small portion of the genes, the ASE is biased toward one parent in one (or some) of the SNPs but toward the other parent in other SNPs of the same genes in the data from the same TCC. The reason might be quite complex, including, but not limited to, short reads of RNA-seq, alternative splicing, large indels and repetitive elements, computation error, and so forth, and these genes were thus excluded from the analysis.

Nonetheless, the ASEGs have provided an index of the genes for future studies, especially with respect to the genetic and molecular mechanism of overdominance. Candidate regions may be identified by mapping the ASEGs to the whole-genome profile of dominant and overdominant effects on the traits of the IMF₂ population, especially in combination with information from other studies.

Materials and Methods

Rice varieties ZS97 and MH63 and their hybrid were the genetic materials used in the study. Rice plants were grown in growth chambers set for four different conditions: HTLD, LTLD, HTSD, and LTSD. RNA samples from seedling shoot at four-leaf stage, flag leaf, and panicle at the day of heading were collected and sequenced to identify ASE. Details of experimental methods are given in *SI Appendix, Supplementary Materials and Methods*.

ACKNOWLEDGMENTS. This work was supported by grants from the National Natural Science Foundation (31330039 and 31821005), the National Key Research and Development Program (2016YFD0100802), the Fundamental Research Funds for the Central Universities (2662017PY043) of China, and the Earmarked Fund for the China Agriculture Research System (CAARS-01-05).

- Goff SA, Zhang Q (2013) Heterosis in elite hybrid rice: Speculation on the genetic and biochemical mechanisms. *Curr Opin Plant Biol* 16:221–227.
- Seymour DK, et al. (2016) Genetic architecture of nonadditive inheritance in *Arabidopsis thaliana* hybrids. *Proc Natl Acad Sci USA* 113:E7317–E7326.
- Huang X, et al. (2016) Genomic architecture of heterosis for yield traits in rice. *Nature* 537:629–633.
- Yang M, et al. (2017) Genomic architecture of biomass heterosis in *Arabidopsis*. *Proc Natl Acad Sci USA* 114:8101–8106.
- Hua JP, et al. (2002) Genetic dissection of an elite rice hybrid revealed that heterozygotes are not always advantageous for performance. *Genetics* 162:1885–1895.
- Hua J, et al. (2003) Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid. *Proc Natl Acad Sci USA* 100:2574–2579.
- Zhou G, et al. (2012) Genetic composition of yield heterosis in an elite rice hybrid. *Proc Natl Acad Sci USA* 109:15847–15852.
- Huang X, et al. (2015) Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat Commun* 6:6258.
- Pastinen T (2010) Genome-wide allele-specific analysis: Insights into regulatory variation. *Nat Rev Genet* 11:533–538.
- Wang X, et al. (2014) Global genomic diversity of *Oryza sativa* varieties revealed by comparative physical mapping. *Genetics* 196:937–949.
- Gaur U, Li K, Mei S, Liu G (2013) Research progress in allele-specific expression and its regulatory mechanisms. *J Appl Genet* 54:271–283.
- Knight JC (2004) Allele-specific gene expression uncovered. *Trends Genet* 20:113–116.
- Guo M, Rupe MA, Danilevskaya ON, Yang X, 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, et al. (2004) Allelic variation of gene expression in maize hybrids. *Plant Cell* 16:1707–1716.
- Guo M, et al. (2006) Genome-wide transcript analysis of maize hybrids: Allelic additive gene expression and yield heterosis. *Theor Appl Genet* 113:831–845.
- Guo M, et al. (2008) Genome-wide allele-specific expression analysis using Massively Parallel Signature Sequencing (MPSS) reveals cis- and trans-effects on gene expression in maize hybrid meristem tissue. *Plant Mol Biol* 66:551–563.
- Springer NM, Stupar RM (2007) Allele-specific expression patterns reveal biases and embryo-specific parent-of-origin effects in hybrid maize. *Plant Cell* 19:2391–2402.
- Springer NM, Stupar RM (2007) Allelic variation and heterosis in maize: How do two halves make more than a whole? *Genome Res* 17:264–275.
- Paschold A, et al. (2012) Complementation contributes to transcriptome complexity in maize (*Zea mays* L.) hybrids relative to their inbred parents. *Genome Res* 22:2445–2454.
- Paschold A, et al. (2014) Nonsynthetic genes drive highly dynamic complementation of gene expression in maize hybrids. *Plant Cell* 26:3939–3948.
- Waters AJ, et al. (2013) Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. *Proc Natl Acad Sci USA* 110:19639–19644.
- Waters AJ, et al. (2017) Natural variation for gene expression responses to abiotic stress in maize. *Plant J* 89:706–717.
- Zhang X, Borevitz JO (2009) Global analysis of allele-specific expression in *Arabidopsis thaliana*. *Genetics* 182:943–954.
- Klosinska M, Picard CL, Gehring M (2016) Conserved imprinting associated with unique epigenetic signatures in the *Arabidopsis* genus. *Nat Plants* 2:16145.
- Shi X, et al. (2012) Cis- and trans-regulatory divergence between progenitor species determines gene-expression novelty in *Arabidopsis* allopolyploids. *Nat Commun* 3:950.
- Todesco M, et al. (2010) Natural allelic variation underlying a major fitness trade-off in *Arabidopsis thaliana*. *Nature* 465:632–636.
- Ng DW, et al. (2014) A role for CHH methylation in the parent-of-origin effect on altered circadian rhythms and biomass heterosis in *Arabidopsis* intraspecific hybrids. *Plant Cell* 26:2430–2440.
- He G, et al. (2010) Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. *Plant Cell* 22:17–33.
- Chodavarapu RK, et al. (2012) Transcriptome and methylome interactions in rice hybrids. *Proc Natl Acad Sci USA* 109:12040–12045.
- von Korff M, et al. (2009) Asymmetric allele-specific expression in relation to developmental variation and drought stress in barley hybrids. *Plant J* 59:14–26.
- Zhang J, et al. (2016) Extensive sequence divergence between the reference genomes of two elite indica rice varieties Zhenshan 97 and Minghui 63. *Proc Natl Acad Sci USA* 113:E5163–E5171.
- Skelly DA, Johansson M, Madeoy J, Wakefield J, Akey JM (2011) A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Res* 21:1728–1737.
- Eitas TK, Dangl JL (2010) NB-LRR proteins: Pairs, pieces, perception, partners, and pathways. *Curr Opin Plant Biol* 13:472–477.
- Liu W, Wang G-L (2016) Plant innate immunity in rice: A defense against pathogen infection. *Natl Sci Rev* 3:295–308.
- Liu W, Liu J, Triplett L, Leach JE, Wang GL (2014) Novel insights into rice innate immunity against bacterial and fungal pathogens. *Annu Rev Phytopathol* 52:213–241.
- Baker KE, Parker R (2004) Nonsense-mediated mRNA decay: Terminating erroneous gene expression. *Curr Opin Cell Biol* 16:293–299.
- Wolucka BA, et al. (2001) Partial purification and identification of GDP-mannose 3',5'-epimerase of *Arabidopsis thaliana*, a key enzyme of the plant vitamin C pathway. *Proc Natl Acad Sci USA* 98:14843–14848.
- Smirnoff N, Wheeler GL (2000) Ascorbic acid in plants: Biosynthesis and function. *Crit Rev Biochem Mol Biol* 35:291–314.
- Pastori GM, et al. (2003) Leaf vitamin C contents modulate plant defense transcripts and regulate genes that control development through hormone signaling. *Plant Cell* 15:939–951.
- Watanabe K, Suzuki K, Kitamura S (2006) Characterization of a GDP-D-mannose 3',5'-epimerase from rice. *Phytochemistry* 67:338–346.
- Higo K, Ugawa Y, Iwamoto M, Korenaga T (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999. *Nucleic Acids Res* 27:297–300.
- Kehoe DM, Degenhardt J, Winicov I, Tobin EM (1994) Two 10-bp regions are critical for phytochrome regulation of a *Lemna gibba* Lhcb gene promoter. *Plant Cell* 6:1123–1134.
- Terzaghi WB, Cashmore AR (1995) Photomorphogenesis. Seeing the light in plant development. *Curr Biol* 5:466–468.
- Zhou DX (1999) Regulatory mechanism of plant gene transcription by GT-elements and TF-factors. *Trends Plant Sci* 4:210–214.
- Yu B, et al. (2007) *TAC1*, a major quantitative trait locus controlling tiller angle in rice. *Plant J* 52:891–898.
- Hollender CA, et al. (2018) Alteration of *TAC1* expression in *Prunus* species leads to pleiotropic shoot phenotypes. *Hortic Res* 5:26.
- Xie W, et al. (2015) Breeding signatures of rice improvement revealed by a genomic variation map from a large germplasm collection. *Proc Natl Acad Sci USA* 112:E5411–E5419.
- Xue W, et al. (2008) Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nat Genet* 40:761–767.