Phasing analysis of the transcriptome and epigenome in a rice hybrid reveals the inheritance and difference in DNA methylation and allelic transcription regulation

Jia-Wu Feng1,2, Yue Lu3, Lin Shao1, Jianwei Zhang1, Huan Li1,* and Ling-Ling Chen1,2,*
1National Key Laboratory of Crop Genetic Improvement, College of Informatics, Huazhong Agricultural University, Wuhan 430070, China
2College of Life Science and Technology, Guangxi University, Nanning 530004, China
3Jiangsu Key Laboratory of Crop Genetics and Physiology/Co-Innovation Center for Modern Production Technology of Grain Crops, Key Laboratory of Plant Functional Genomics of the Ministry of Education, Yangzhou University, Yangzhou 225009, China
*Correspondence: Huan Li (lihuan2729@163.com), Ling-Ling Chen (llchen@mail.hzau.edu.cn)
https://doi.org/10.1016/j.xplc.2021.100185

ABSTRACT

Hybrids are always a focus of botanical research and have a high practical value in agricultural production. To better understand allele regulation and differences in DNA methylation in hybrids, we developed a phasing pipeline for hybrid rice based on two parental genomes (PP2PG), which is applicable for Iso-Seq, RNA-Seq, and Bisulfite sequencing (BS-Seq). Using PP2PG, we analyzed differences in gene transcription, alternative splicing, and DNA methylation in an allele-specific manner between parents and progeny or different progeny alleles. The phasing of Iso-Seq data provided a great advantage in separating the whole gene structure and producing a significantly higher separation ratio than RNA-Seq. The interaction of hybrid alleles was studied by constructing an allele co-expression network that revealed the dominant allele effect in the network. The expression variation between parents and the parental alleles in progeny showed tissue- or environment-specific patterns, which implied a preference for trans-acting regulation under different conditions. In addition, by comparing allele-specific DNA methylation, we found that CG methylation was more likely to be inherited than CHG and CHH methylation, and its enrichment in genic regions was connected to gene structure. In addition to an effective phasing pipeline, we also identified differentiation in OsWAK38 gene structure that may have led to the expansion of allele functions in hybrids. In summary, we developed a phasing pipeline and provided valuable insights into alternative splicing, interaction networks, trans-acting regulation, and the inheritance of DNA methylation in hybrid rice.

Keywords: phasing, hybrid, trans-acting regulation, allele co-expression network, allele-specific expression, allele-specific DNA methylation

INTRODUCTION

Genetic variation provides resources for natural selection and breeding, and these variations can bring changes in gene structure and expression (Pastinen, 2010). Most plants and animals are nonhaploid. In a hybrid, phasing essentially distinguishes which part of its genetic information comes from the male parent and which part from the female parent (Choi et al., 2018). In addition to hybrid-related studies, phasing can also be applied to population genetics, polyploidy studies, and genome assembly. It can be used to infer recombination regions and construct recombination maps (Bosse et al., 2014), solve the assembly of highly heterozygous genomes (Giani et al., 2020; Zhang et al., 2020), settle the assembly of polyploid genomes (Chin et al., 2016; Zhang et al., 2019), and distinguish polyploid sub-genomes transcripts (Wang et al., 2018a).
Plant hybridization plays a vital role in evolution. In numerous cases, a hybrid is superior to its parents in terms of growth rate, mature biomass, and reproduction; this phenomenon is called heterosis (Goulet et al., 2017; Ottenburghs, 2018). Through hybrid phasing, researchers can identify allele-specific expressed (ASE) genes and imprinted genes from transcriptome data and appraise allele-specific DNA methylation (ASM) or allele-specific histone modification from epigenetic data. The expression of an ASE gene is biased between the two parental alleles of the hybrid. Recently, Shao and colleagues (2019) observed that some ASE genes originate from their parents and are generally under cis- rather than trans-acting regulation (Guo et al., 2008; Springer and Stupar, 2007). With advances in full-length transcriptome sequencing technology, the complete allelic gene structure can be identified by phasing based on Iso-Seq data (Deonovic et al., 2017; Wang et al., 2020). ASM refers to the difference in DNA methylation between the parental allelic regions of the hybrid, and it plays significant roles in gene imprinting and X chromosome inactivation (Tycko, 2010; Fang et al., 2012). For instance, two alleles of the F1 hybrid inherit the methylation information from its parents with nearly 100% fidelity in Nasonia (Wang et al., 2016a), and DNA methylation is stably inherited by progeny in Arabidopsis thaliana (Hofmeister et al., 2017). In addition, H3K36me3 regulates ASE genes in hybrid rice (Guo et al., 2015).

Results

Phasing pipeline based on two parental genomes

In this study, a phasing pipeline based on two parental genomes was constructed for hybrid sequencing data, including paired-end RNA-Seq, paired-end BS-Seq, and Iso-Seq data. The detailed process is as follows (Figure 1A): (1) Mapping: the sequencing data were mapped to the two parental genomes using minimap2 (Li, 2018) for Iso-Seq data (Supplemental Table 1), Bismark (Krueger and Andrews, 2011) for BS-Seq data (Supplemental Table 2), and HISAT2 for RNA-Seq data (Kim et al., 2019), and the best hit was set to one for all alignment results (Supplemental Table 3). (2) SNP Calling: allelic SNPs were acquired using MUMmer (Marçais et al., 2018) based on the reference genomes of MH63 and ZS97. (3) Scoring: reads were scored using the results of allelic SNPs and alignments. If a mismatch was generated during alignment to the MH63 genome and the mismatched base was identical to that of the ZS97 genome, then the read was scored as the ZS97 type, and vice versa. If the mismatched base did not match the other genome, it was assigned to the unknown type. Afterward, the final score for each type in each read was calculated. For BS-Seq data, given that the unmethylated C was converted to U during the bisulfite treatment, C/T and G/A SNPs were unavailable for phasing. The information on allelic SNPs that were available for phasing in MH63 and ZS97 was summarized, and available SNPs accounted for ~29.97% of the total SNPs between ZS97 and MH63 (Supplemental Table 4). (4) Classification: all reads were classified into five types: MH63-synteny reads, ZS97-synteny reads, unknown reads, MH63-only reads, and ZS97-only reads. The scoring and classification steps were performed using an in-house python script (phasing.py). Based on the alignment results, reads that mapped to only one genome were MH63-only/ZS97-only reads. Reads that did not map to either genome were identified as unknown reads. For reads that mapped to both parental genomes, if the score of the MH63 type was higher than that of the ZS97 type, the reads were assigned to the MH63-synteny reads, and vice versa. However, if the two scores were identical, then the reads were assigned to unknown reads. Based on the above steps, we developed a phasing pipeline based on the two parental genomes, PP2PG, that is freely available at GitHub (https://github.com/jia-wu-feng/PP2PG).

Evaluation of phasing results

The two parental genomes, ZS97 and MH63, were used to evaluate the error rate and separation efficiency of phasing. We calculated the error rate by MH63-ZS97-synteny reads and ZS97-MH63-synteny reads separated in parental sequencing data and counted the ratio of separated reads to total reads to evaluate the separation efficiency. For Iso-Seq data, the error rates for MH63 and ZS97 were 2.23% and 0.25%, and the separation rates were 41.55% and 41.23% (Supplemental Tables 5 and 6). For RNA-Seq data, the average error rates of MH63 and ZS97 in multiple tissues and conditions were 0.48% and 0.14%, and the average separation rates were 9.18% and 9.38% (Supplemental Figure 1A and 1B, Supplemental Tables 5 and 7). Iso-Seq data had a significantly higher separation rate than RNA-Seq data, even though their error rate was relatively high. For BS-Seq data, the error rates of the two repeated MH63 were 2.14% and 2.15%, and the separation rates were 9.27% and 9.28%. The error rates of the two repeated ZS97 were 0.57% and 0.62%, and the separation rates were 9.40% and 9.29% (Supplemental Tables 5 and 8). To further assess the local phasing results, we selected the OsMH_10G0225400 locus in Chr10 as an example for visualization (Figure 1B). The hybrid region of this locus in SY63 was clearly shown. Especially at the main isoform, we observed the obvious difference in gene structure in the first-second exon from the
Iso-Seq phasing results. The second exon was shorter in ZS97 than in MH63, and this region could be more clearly observed in Iso-Seq than in RNA-Seq. Moreover, in the BS-Seq phasing, we found that the allelic CG methylations in the first exon of OsMH_10G0225400 were different. The progeny after phasing were substantially consistent among the RNA-Seq, Iso-Seq, and BS-Seq data, suggesting that the hybrid was phased correctly. After phasing, 86.62% of the differentially expressed genes (DEGs) between the parents remained the same as before phasing (Supplemental Figure 1C), and the methylation levels were almost the same (Supplemental Figures 1D and 1E). These results suggested that our phasing pipeline can preserve the parental archetype. In addition, we compared our pipeline with SNPsplit, which directly uses SNPs for phasing based on mouse and human data (Krueger and Andrews, 2016). Separation rate was higher with SNPsplit (average 9.57%) than with our pipeline (average 9.04%), but there was a difference in the error rate.

Figure 1. Phasing pipeline and the results of multi-omics data
(A) The hybrid phasing pipeline based on two parental genomes consists of four steps: (1) mapping, (2) SNP calling, (3) scoring, and (4) classification. (B) Phasing results for the gene OsMH_10G0225400 in Iso-seq, RNA-seq, and BS-seq. In Iso-seq and RNA-seq, yellow, green, red, and blue illustrate different SNPs. In BS-seq, red indicates methylated CG, and blue indicates unmethylated CG. SY63_ZS97 indicates the ZS97-type data from the SY63 variety, and SY63_MH63 indicates the MH63-type data from the SY63 variety.
in separating the parents, especially in MH63 (Supplemental Figure 2). For SNPsplit, the average for MH63 and ZS97 was 0.16% and 3.05%, respectively, and for PP2PG, the average for MH63 and ZS97 was 0.16% and 0.52%, respectively. PP2PG performed very well with respect to error rates. Overall, the phasing pipeline developed in this study can provide high resolution and a low error rate for phasing multi-omics data in hybrid rice.

**Isolating allelic transcripts in full-length transcripts**

After phasing, TAPIS was used to correct and assemble all samples and obtain high-quality transcripts. We summarized the assembly results for transcripts and genes and showed that the overall distribution of allelic expression had no significant bias (Figure 2A and Supplemental Table 9). Through principal component analysis (PCA), we observed no clear difference in estimated expression between parents and progeny or between crosses and reciprocal crosses. Instead, we found a significant difference among different tissues (Figure 2B). These results implied that quantification after Iso-Seq phasing can accurately characterize the samples. Among the ASE genes identified in three hybrids (a total of 107 ASE genes, 29 in panicles, 60 in the crosses of flag leaf, and 66 in the reciprocal crosses of flag leaf), most had the same preference, and the bias between crosses and reciprocal crosses was essentially the same, although there were changes in the preference among tissues (Figure 2C and 2D). In addition, there was no substantial difference in the proportion of alternative splicing (AS) between parents and progeny in leaves, and differences occurred only among tissues.
Figure 3. Preference of trans-acting regulation in the hybrid

(A) DEGs between parents and progeny under different conditions. Four colors represent different regulated genes in different varieties. For instance, MH63_MH63 > SY63_MH63 refers to MH63 alleles that are downregulated in the hybrid; MH63_MH63 < SY63_MH63 refers to MH63 alleles that are upregulated in the hybrid; ZS97_ZS97 > SY63_ZS97 refers to ZS97 alleles that are downregulated in the hybrid; and ZS97_ZS97 < SY63_ZS97 refers to ZS97 alleles that are upregulated in the hybrid; ZS97_ZS97 > SY63_ZS97 refers to ZS97 alleles that are downregulated in the hybrid; and ZS97_ZS97 < SY63_ZS97 refers to ZS97 alleles that are upregulated in the hybrid; ZS97_ZS97 > SY63_ZS97 refers to ZS97 alleles that are downregulated in the hybrid; and ZS97_ZS97 < SY63_ZS97 refers to ZS97 alleles that are upregulated in the hybrid.
Allele trans-acting regulation in the transcriptome

We analyzed the transcription level of parental alleles after phasing and observed that the correlation between parents and progeny in the same tissue was generally high (Supplemental Figure 3). Interestingly, the correlation was higher in shoots under different conditions and was essentially consistent with the DEGs between parents and progeny identified in the shoots (Figure 3A). The differential expression of the same allele between parent and progeny was defined as trans-acting regulation (Springer and Stupar, 2007; Chen, 2013). We observed certain preferences of trans-regulatory DEGs in the shoot, flag leaves, and panicles under different conditions. The proportion of ZS97-type DEGs between parents and progeny was higher than that of MH63-type DEGs in panicles, but the proportion of MH63-type DEGs was higher than that of ZS97-type DEGs under different conditions in flag leaves (Figure 3A). These results suggested that the trans-acting regulation of gene transcription in the hybrid differed among tissues or environmental conditions. We identified 330 upregulated genes and 492 downregulated genes of the MH63 type and 495 upregulated genes and 518 downregulated genes of the ZS97 type in progeny compared with parents (Supplemental Data 1). GO enrichment analysis revealed that most of the downregulated genes in the progeny that overlapped in MH63 and ZS97 types were enriched in phosphate metabolism and phosphorus metabolism, phosphorylation, posttranscriptional protein modification, and protein amino acid phosphorylation, and so forth. However, few MH63-type and ZS97-type upregulated genes overlapped. The MH63-type upregulated genes were enriched in small molecular metabolism and amine biosynthesis biological processes, whereas the ZS97-type genes were enriched in polysaccharide metabolism, redox, and lipid transport biological processes (Figure 3B). We speculated that the downregulated genes may be co-acted and the upregulated genes show functional differentiation. In addition, in the 2,236 detected trans-acting regulatory events, there were 1,736 events in which the deviation of the progeny alleles was smaller than that of the parental alleles, consistent with previous reports that trans-acting regulation can narrow the difference between two alleles (Chen, 2013).

Allelic co-expression network of SY63

To further study the interaction between hybrid alleles, a co-expression network was constructed, and the samples were marked by tissues (Supplemental Figure 4). In the allele network of SY63, there were 7,278 pairs of alleles, 16,496 nodes, and 19,357,503 edges. The allele interactions were divided into three categories: (1) nodes that interacted with both parental alleles, (2) nodes that interacted only with MH63 alleles, and (3) nodes that interacted only with ZS97 alleles. Afterward, we calculated the frequencies of nodes that interacted with both alleles in the above three types (Supplemental Figure 5). Most interactions were identical, but some divergent interactions were observed. We also calculated the distribution frequency of nodes that interacted only with the MH63 allele in the divergent interactions and observed that the interactions were not evenly distributed. Based on these two distribution frequencies, we retained only those alleles that had more than 100 interaction nodes and whose divergent ratio was higher than 0.5. Based on the frequency of nodes that interacted only with MH63 alleles in divergent interactions, we divided these alleles into three types: (1) dominant interaction alleles of MH63 (0–0.25) (Figure 4A); (2) dominant interaction alleles of ZS97 (0.75–1) (Figure 4B); and (3) functional extension alleles (0.25–0.75) (Figure 4C). The numbers of the three allele types were 720, 886, and 979, respectively (Supplemental Data 2). Next, we performed GO enrichment analysis on these three allele types and found that MH63-dominant alleles were enriched in ATPase activity, kinase activity, and ATP binding, whereas ZS97-dominant alleles were enriched in methyltransferase activity and ATP binding. Functional expansion alleles were enriched in apoptosis, intracellular homeostasis, defense response, protein serine/threonine kinase activity, and so forth (Supplemental Figure 6).

Hd3a/Ahd3a (OsMH_06G0050000/OsZS_06G0051000) is an important gene that regulates rice flowering time and yield, and it was identified as an MH63-dominant allele (Huang et al., 2016) (Figure 4A). Sequence analysis showed that the MH63 genome contained the Hd3a (CC) allele, and the ZS97 genome contained the hd3a (AA) allele (Supplemental Figure 7). The reported gene dominance is consistent with our co-expression network analysis, indicating that the interaction network is able to predict the dominance-recessive relationship of alleles.

OsyWAK38 (OsMH_04G0272400/OsZS_04G0266500), a functionally extended gene, belongs to the wall-associated kinase (WAK) gene family and plays important roles in the regulation of female gametophyte development, pathogen resistance, and cold trigger responses (Zhang et al., 2005; Li et al., 2009; Wang et al., 2012a). We observed a SNP (G/A) in the MH63 genome at Chr04:17 394 272 that caused premature translation termination, resulting in a loss of 61 amino acids in the GUB_WAK_bind domain (Figure 4D, Supplemental Figures 8 and 9). By comparing the phylogenetic relationships of several rice varieties, we observed that the structure of OsMH_04G0272400 was similar to that of the gene from wild rice (Oryza nivara, Oryza glumipatula, and Oryza meridionalis), whereas that of OsZS_04G0266500 was similar to that of the gene from cultivated rice (Oryza sativa Nipponbare and Oryza sativa indica Shuhiui498) (Figure 4E). To further study the functional differentiation of these two alleles, we performed subcellular localization prediction and signal peptide prediction of their encoded proteins. OsMH_04G0272400 was predicted to localize in the plasma membrane and had a Sec/SPI signal peptide of 20 amino acids in the GUB_WAK_bind domain (Supplemental Figure 10). By contrast, OsZS_04G0266500 was predicted to localize in the chloroplast, and no signal peptide was identified. We performed functional enrichment of the allelic networks of the two alleles and observed that the MH63 allele network was enriched in protein...
Figure 4. The allele co-expression network in SY63
(A–C) represent co-expression networks of the MH63 dominant allele (Hd3a/hd3a), the ZS97 dominant allele, and the functional expansion allele (OsWAK38), respectively. Different colors indicate different types of alleles.
(D) The allelic structural difference of OsWAK38 (OsMH_04G0272400/OsZS_04G0266500).
(E) The phylogenetic relationship of OsWAK38 sequences from several varieties assessed by the maximum likelihood method.
(F) The results of subcellular localization of OsMH_04G0272400 and OsZS_04G0266500 by prediction of protein localization sites.
amino acid phosphorylation and protein serine/threonine kinase, whereas the ZS97 allele network was enriched in photosynthesis, thylakoid part, and photosynthetic membrane (Figure 4F and Supplemental Figure 11).

Inheritance and variation of allelic methylation
To accurately assess the differences in methylation levels between two parental types, we calculated the genome-wide methylation level of CG, CHG, and CHH in a 50-base pair (bp) window after removing all the error Cs used for parental assessment phasing. We compared the allelic methylation of the parents and progeny under different conditions (the same parental type, parents, and the progeny parental types), and the number of detected regions is shown in Figure 5A. After phasing, very few differentially methylated regions (DMRs) were identified in the same parental type between parents and progeny (between MH63_MH63 and

Figure 5. Inheritance and differences in allelic methylation
(A) Statistics of detected methylation regions under different conditions.
(B) Statistics of DMRs under different conditions.
(C) The inheritance mode of ASM in rice.
(D) The inheritance ratio of ASM regions in three methylation types.
(E) The relationship between ASM regions of different methylation types and gene bodies. The chi-square test was performed to detect the enrichment of ASM regions in gene bodies. ***p < 0.001 (chi-square test).
(F) The relationship between genes with different structures and ASM regions. Green, blue, and yellow indicate Gene Class 1, 2, and 3, respectively. Members of Gene Class 1 are identical in both sequence and position, with SNPs that do not cause amino acid substitutions; members of Gene Class 2 have the same lengths and syntenic positions but contain nonsynonymous substitutions with identity >80%; members of Gene Class 3 have “good collinearity” with syntenic chromosomal locations, protein sequence identity >80%, and coverage >50%. The chi-square test was performed to detect differences in differential structure genes with ASM regions. ***p < 0.001 (chi-square test). Green indicates more enrichment in the genic region, and blue indicates more enrichment in the non-genic region.
Phasing analysis for rice hybrid SY63, MH63, and between ZS97, ZS97 and SY63, ZS97, suggesting that the allelic methylation level of the progeny was mainly inherited from the parents. By contrast, significant differences were observed between the two parental types, which is referred to as ASM (Wang et al., 2016a). Although the methylation difference between parents and progeny in the same parental type is small, we still observed relatively greater variation in CHH methylation than in CG and CHG methylation (Figure 5B). We detected 2,690 DMRs of CHH in the MH63 type between parents and progeny, of which 2,068 DMRs were hyper-methylated in the progeny. We detected 1,897 DMRs of CHH in the ZS97 type, and 1,220 of them were hyper-methylated in the progeny. This result showed that most changes in CHH methylation in the progeny involved upregulation. We speculated that these changes were probably regulated by the RNA-directed DNA methylation (RdDM) pathway.

Subsequently, we analyzed the inheritance of ASM in parents and progeny (Figure 5C) and found that the inheritance ratios of CG, CHG, and CHH methylations were 81.86%, 63.65%, and 24.43%, respectively (Figure 5D and Supplemental Figure 12). That is, the ASM preference between MH63, MH63 and ZS97, ZS97 was the same as the ASM preference between SY63, MH63 and SY63, ZS97, and the methylated regions that were consistent between parents and progeny were named conservative methylated regions (Figure 5C). We analyzed the distribution of ASM regions in the SY63 genome and found that 37.52% of CG methylation in the detected regions occurred in genic regions, and the ratio increased to 46.41% for ASM regions, indicating that the ASM region of CG methylation was significantly enriched within genes. Simultaneously, we found that 49.03% of the ASM regions of CG methylation consistent with the parents were found within genes, demonstrating that these conserved CG methylated ASM regions were also enriched in genes in the parents. In addition, 41.91% of the detected CHH methylated regions were located within genes, whereas only 36.02% of the ASM regions were located within genes. For CHH methylation, 38.67% of the detected regions were located within genes; however, only 19.77% of the ASM regions were located within genes (Figure 5E). In summary, ASM regions of CG methylation were significantly enriched in genes compared with those of CHG and CHH methylations. We also found that total transposons were enriched in ASM regions of CHG and CHH methylations. DNA transposons were enriched in ASM regions of CHG and CHH methylations, whereas LTR retrotransposons were only enriched in CHG methylation (Supplemental Figure 13).

By analyzing ASM regions, we observed that 60.54% of the CG methylated regions located in genes were related to exons, whereas only 51.55% of the ASM regions were related to exons, indicating that ASM regions of CG methylation tended to be enriched in intron regions (Chi-square test, \( p \)-value <2.2e\( -16 \)). Furthermore, we evaluated the enrichment of ASM regions for CG methylation in three classes of non-TE alleles: Gene Classes 1, 2, and 3 (a detailed description of Classes 1, 2, and 3 is provided in the Materials and Methods section). Classes 1, 2, and 3 contained 15,236 alleles, 3,978 alleles, and 11,130 alleles, respectively. By comparing the proportions of genes related to ASM regions for CG methylation and genes related to CG methylated regions, we observed that the proportion of genes related to ASM regions for CG methylation increased with the increase in gene structure difference from Class 1 to Class 3. The same trend was observed for the conserved ASM regions of CG methylation. The proportion of conserved ASM regions of CG methylation in divergent allele-gene Class 3 was much higher than that of similar allele-gene Class 1 (Figure 5F). This demonstrated that there was a strong correlation between the difference in allelic gene structure and ASM regions of CG methylation, further implying that ASM regions of CG methylation helped to maintain differences in allelic structure.

**DISCUSSION**

Allele phasing is an important step in analyzing sequencing data from hybrids. In this study, we developed a phasing method based on two parental genomes, PP2PG, which not only distinguished alleles but also assigned them to different parental types. Importantly, it reduced the error rate to 0.09% to 2.23%. Our pipeline also supports phasing for multi-omics datasets, which will be helpful for understanding multiple genetic or epigenetic traits of hybrids. Despite its wide application and high accuracy, the disadvantage of this phasing pipeline is its reliance on the genomes. However, the pipeline can be extended to any read allocation in RNA-Seq, Iso-Seq, and BS-Seq with two homologous sequences. It can also be extended to the read allocation of other hybrids and read allocation between polyploid sub-genomes. Compared with SNPsplit (Krueger and Andrews, 2016), PP2PG has a lower error rate when phasing. Users must provide SNP files for SNPsplit, but PP2PG integrates this part into its pipeline and provides the phasing of long reads. Furthermore, PP2PG also provides a variety of downstream analyses after phasing. Therefore, PP2PG has the advantage when two reference sequences are available, especially in the absence of SNP information, as in the case of phasing of polyploid plants. If a complete genome is provided, users can further compare the difference in PAV regions. In addition, our study showed a higher error rate for the MH63 type than for the ZS97 type. This may be due to a preference for some specific regions in our pipeline, or it may be caused by the incompleteness of the genome assembly.

Iso-Seq is an effective way to identify isoforms and gene structures of alleles, and it is becoming a powerful tool in plants (Abdel-Ghany et al., 2016; Wang et al., 2016b, 2019; Cheng et al., 2017). The phasing results showed better separation in Iso-Seq than RNA-Seq, even though the sequencing depth of Iso-Seq was lower. Using Iso-Seq, the number of assembled transcripts in parents was nearly twice that in progeny, consistent with the original data after phasing, implying that the data was unsaturated for all transcripts. Moreover, the composition of AS showed no significant difference between the two parental types. For either parents or progeny, differences were observed only among different tissues. For instance, the panicle tissue had a higher proportion of alternative 5'/3' splice sites, as well as a lower proportion of retained introns compared with leaf and shoot tissue. This tissue-specific variance is similar to that observed in maize and sorghum (Wang et al., 2018b). Although the transcripts in Iso-Seq were insufficient, the genotype clustering revealed that differences between parents and tissues could be distinguished at an overall level. In addition, the crosses and reciprocal
Plant Communications
crosses of young leaves were almost identical according to the clustering results and ASE genes.

The RNA-Seq data used in this study have been analyzed previously for ASE genes, and ASE genes showed significant variation in different tissues and conditions (Shao et al., 2019). In this study, we focused on variation in the same allele between parents and progeny. We found that the correlation of the same allele between parents and progeny was higher in seedlings than in panicles and flag leaves, indicating that allelic expression may undergo reprogramming during the transformation from the vegetative stage to the reproductive stage. Most studies on ASE genes have discussed the dominant effects between cis- and trans-acting regulation (Springer and Stupar, 2007; Guo et al., 2008). Here, we found that trans-acting regulation based on different parents could help hybrids to change the utilization patterns of genes in different parental types. In particular, GO analysis showed that the upregulated genes in the progeny were enriched in different biological functions, but the downregulated genes showed similar functional enrichment, indicating that allelic-specific expression may confer certain functional advantages to the hybrids. Chen (2013) has suggested that most trans-acting regulation compensates for one of the two parents to achieve a balance of allelic expression, which is consistent with our results. This balance between alleles may enable plants to choose different parental biases under different conditions, thereby contributing to plant adaptation. According to the analysis of allelic interaction networks in SY63, we observed that functional expansion alleles were enriched in defense response, intracellular homeostasis, and apoptosis, implying that these alleles can improve the adaptability of hybrids by expanding their functions. However, functional expansion alleles, dominant interaction alleles of MH63, and dominant interaction alleles of ZS97 may contribute to heterosis in unique ways. Functional expansion alleles may expand the function of a single gene at the level of interaction, whereas dominant interaction alleles can supplement the dominant genes of different varieties. For instance, differential expression of OsWAK38 was reported to be associated with heterosis in different tissues (Shaleb et al., 2020). In the evolutionary analysis of OsWAK38, we found that it had the highest sequence similarity with its homolog in O. nivara, related to its cross-breeding history (Xie and Zhang, 2018). In this study, we used information on allelic co-expression networks and trans-acting regulation to demonstrate the possible model of heterosis. In fact, our allelic network analysis showed that heterosis may be caused by the differentiation of allelic functions. The trans-acting regulation and allelic network results indicated how hybrids utilize both parental alleles to achieve a growth advantage.

DNA methylation levels are thought to be stable across generations in A. thaliana (Hofmeister et al., 2017), which is basically consistent with our results in rice. In this study, there was essentially no apparent methylation variation between parents and progeny in the same parental allele. Through DMR analysis, we found that CG methylation showed a higher inheritance rate than CHG and CHH methylation between parents and progeny. We also observed significant enrichment of CG methylated regions in genes. It has been suggested that DNA methylation is strongly related to gene structure. Therefore, we speculate that these allelic methylations may have a specific function in the structure of different alleles. In addition, the variation in DNA methylation that we observed between parents and progeny was mainly in a CHH context, and most DMRs were hyper-methylated in the progeny. We speculate that this hypermethylation may be mediated by the RdDM pathway (Law and Jacobsen, 2010), that is, the small interfering RNAs (siRNAs) generated by one allele may direct CHH methylation in the other allele through RdDM. A recent study in pigeon pea reported significantly increased DNA methylation in hybrids over the genome, which was accompanied by strongly elevated 24-nt siRNA levels (Sinha et al., 2020). The CHH hypermethylation in hybrids may be caused by a global change in siRNA level, although this is currently unclear in our case. The regulation mechanism of CHH methylation in hybrid rice requires further study.

Through the phasing pipeline, we analyzed variation in transcription and DNA methylation patterns of hybrid alleles. From the transcriptome phasing analysis, it can be inferred that trans-acting regulation and dominant genes enable plants to choose different parental preference patterns under different conditions, which may improve plant adaptability. Functional expansion alleles can help hybrids to expand different functions in a limited genome. From the DNA methylation phasing analysis, we found that stable CG methylation may help to maintain the stability of hybrid genes. Taken together, these variations may help hybrid plants maintain a more stable genome while rapidly expanding gene functions.

In summary, our study describes an effective and accurate phasing pipeline for multiple types of sequencing data and provides valuable insights into AS, interaction networks, trans-acting regulation, and the inheritance of DNA methylation in hybrid rice.

METHODS
Plant materials
Oryza sativa ssp. xian/indica MH63, ZS97, SY63 (F1), and RC (reciprocal cross) were grown in a greenhouse at Huazhong Agricultural University, Wuhan (China) at 25°C under normal conditions. Iso-Seq data were obtained from the young leaves of MH63, ZS97, SY63, and RC and the panicle of SY63 using the PacBio RS II sequencing platform (Wang et al., 2016b). RNA-Seq data were obtained from 72 samples from the PRJNA550148 project. This dataset included three varieties (MH63, ZS97, and SY63) and three different tissues (seedling shoots at the four-leaf stage, flag leaves, and panicles) under four conditions: high temperature and long days (HTLD), high temperature and short days (HTSD), low temperature and short days (LTSD). Each condition was replicated twice (Shao et al., 2019). BS-Seq data were obtained from MH63, ZS97, and SY63 at the young leaf stage (two replicates each) by paired-end sequencing on the Illumina HiSeq 3000 platform. The sequences and annotation files of the two parental genomes used in the study were downloaded from RIGW (http://rice.hzau.edu.cn/rice/).

Data preprocessing and phasing evaluation
Different types of sequencing data were processed using different methods. SMRT Link (https://www.pacb.com/) was used to process the Iso-Seq raw data. The steps were as follows: (1) bax2bam was used to convert the original h5 file into a bam file; (2) ccs was used for preliminary alignment using the parameters (–nePolish –numThreads 10 –minLength = 50 –maxLength = 15,000 –minReadScore = 0.65).
Phasing analysis for rice hybrid

Phasing analysis for rice hybrid

Plant Communications

Identification of differentially expressed genes in parents and progeny

For the RNA-seq data, HTSeq (Anders et al., 2015) was used to count reads for each sample after phasing, and Pearson’s correlation coefficient was used to calculate the correlation between samples. The results of parental types were then used to identify and filter out genes with an error rate greater than 0.05. Next, DESeq2 was used to identify differentially expressed genes between parents and progeny (reads ≥ 5, q-value ≤ 0.05 and log2FoldChange > 1).

Construction of the allele co-expression network of SY63

An allele co-expression network was constructed based on weighted gene co-expression network analysis (WGCNA) (Langfelder and Horvath, 2008). The detailed steps were as follows: (1) 24 RNA-Seq samples of SY63 (3 tissues × 4 conditions × 2 replicates) were phased. (2) StringTie (Pertea et al., 2015) was used to calculate the expression of two parental alleles in the same genome. (3) Two files of allele labels were modified and merged into one, and genes whose error rates were higher than 0.05 were filtered out. (4) All 24 samples were combined into one table, and genes whose error rates were higher than 0.05 were filtered out. (5) All 24 samples were combined into one, and genes whose error rates were higher than 0.05 were filtered out. (6) The resulting network was constructed using WolF PSORT (https://wolfsort.hgc.jp/) and SignalP-5.0 (http://www.cbs.dtu.dk/services/SignalP-5.0/) were used for subcellular localization prediction and signal peptide prediction, respectively (Almagro Armenteros et al., 2019; Horton et al., 2007).

Identification of differentially methylated regions

The steps to identify the differentially methylated regions were as follows: (1) samtools was used to merge and sort the two replicates after phasing. (2) StringTie (Pertea et al., 2015) was used to calculate the expression of two parental alleles in the same genome. (3) Two files of allele labels were modified and merged into one, and genes whose error rates were higher than 0.05 were filtered out. (4) All 24 samples were combined into one table, and genes whose error rates were higher than 0.05 were filtered out. (5) The resulting network was constructed using WolF PSORT (https://wolfsort.hgc.jp/) and SignalP-5.0 (http://www.cbs.dtu.dk/services/SignalP-5.0/) were used for subcellular localization prediction and signal peptide prediction, respectively (Almagro Armenteros et al., 2019; Horton et al., 2007).
Plant Communications

(2) The bismark_methylation_extractor was used to re-call methylations based on the MH63 genome with the following parameters (p --grip --bedGraph --comprehensive --cytosine_report --CX) (Krueger and Andrews, 2011). (3) Based on CG, CHG, and CHH methylations, we separated the results of the whole genome, retained sites with more than five reads, and filtered out sites whose parental phasing was incorrectly evaluated. (4) The coverage of methylated cytosine and unmethylated cytosine was calculated across the whole genome in non-overlapping sliding windows with a window size of 50 bp. (5) Regions were retained only if they were simultaneously covered by two samples. (6) The significance of differences in the methylated regions was calculated using the Fisher exact test, and the Benjamini-Hochberg method was used to adjust the p-values. (7) The regions were screened using a false discovery rate < 0.05, CG methylation difference ≥ 0.6, CHG methylation difference ≥ 0.4, and CHH methylation difference ≥ 0.2. Finally, bedtools (Quinlan, 2014) was used to calculate the overlaps between specific regions (genes, exons, or transposons) and differentially methylated regions, ensuring that >50% of the differentially methylated regions were in specific regions. The chi-square test was performed to detect the differences of differential structure genes with ASM regions.

Identification of differentially structured alleles
We evaluated the enrichment of non-TE alleles with different structures in the ASM regions of CG methylation. Non-TE alleles with different structures were divided into three categories (Zhang et al., 2016): (1) Gene Class 1 referred to alleles in ZS97 and MH63 that were identical in both sequence and position and had SNPs that did not cause amino acid substitutions; (2) Gene Class 2 referred to alleles that had the same lengths and syntenic positions but contained nonsynonymous substitutions with identity ≥ 80%; (3) Gene Class 3 referred to alleles that had “good collinearity” with syntenic chromosomal locations, protein sequence identity ≥ 80%, and coverage >50%. MCScanX (Wang et al., 2012b) was used to calculate collinearity, and blastp was used for protein sequence alignment. The proportion of ASM genes was calculated based on the number of ASM genes among the genes that were covered by detection regions.

SNPsplit phasing process
GATK (McKenna et al., 2010) was used to call allele SNPs in all RNA-Seq data from ZS97 leaf samples and from the MH63 genome. Then bcftools was used for quality filtering with the parameters "QUAL=10 || FMT/DP < 5." Next, SNPsplit_genome_preparation was performed to create an N-masked genome, and HISAT2 (Kim et al., 2019) was used to map all the RNA-Seq data from ZS97 leaf samples and the MH63 genome. Finally, SNPsplit (Krueger and Andrews, 2016) was used for phasing. All reads were divided into six types: unaligned and hence skipped reads, unassignable reads, reads that were specific for genome one, reads that were specific for genome two, reads that did not contain one of the expected bases at known SNP positions, and reads that contained conflicting allele-specific SNPs. The reads that were specific for genome one and genome two were used to calculate error rate and separation rate.

ACCESSION NUMBERS
The Iso-Seq data for leaves (MH63, ZS97, SY63 and RC) and panicles (SY63) analyzed in the current study have been submitted to the NCBI Sequence Archive (PRJNA720049). The RNA-Seq data from three tissues and four different conditions have been submitted to the NCBI Sequence Read Archive (PRJNA550148). PP2PG is available from GitHub (https://github.com/jia-wu-feng/PP2PG/

SUPPLEMENTAL INFORMATION
Supplemental Information is available at Plant Communications Online.

FUNDING
This work was supported by the National Natural Science Foundation of China (31871269) and the Hubei Provincial Natural Science Foundation of China (2019CFA014).

AUTHOR CONTRIBUTIONS
J.-W.F., H.L., and L.-L.C. conceived and designed the project. J.-W.F., Y.L., J.Z., and L.S. performed the experiments and analysis. J.-W.F., Y.L., H.L., and L.-L.C. wrote and revised the manuscript.

ACKNOWLEDGMENTS
We thank all the reviewers for their valuable comments. The authors declare no competing interests.

Received: December 25, 2020
Revised: March 14, 2021
Accepted: April 13, 2021
Published: April 15, 2021

REFERENCES
Abdel-Ghany, S.E., Hamilton, M., Jacobi, J.L., Ngam, P., Devitt, N., Schilkey, F., Ben-Hur, A., and Reddy, A.S. (2016). A survey of the sorghum transcriptome using single-molecule long reads. Nat. Commun. 7:11706.
Almagro Armenteros, J.J., Tsirigos, K.D., Sonderby, C.K., Petersen, T.N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat. Biotechnol. 37:420–423.
Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169.
Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 30:2114–2120.
Bosse, M., Megens, H.J., Frantz, L.A., Madsen, O., Larson, G., Paudel, Y., Duijvesteijn, N., Harlizius, B., Hagemeijer, Y., Crooijmans, R.P., et al. (2014). Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. Nat. Commun. 5:4392.
Chen, Z.J. (2013). Genomic and epigenetic insights into the molecular bases of heterosis. Nat. Rev. Genet. 14:471–482.
Cheng, B., Furtado, A., and Henry, R.J. (2017). Long-read sequencing of the coffee bean transcriptome reveals the diversity of full-length transcripts. GigaScience 6:1–13.
Chin, C.S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O’Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods 13:1050–1054.
Choi, Y., Chan, A.P., Kirkness, E., Talenti, A., and Schork, N.J. (2018). Comparison of phasing strategies for whole human genomes. PLoS Genet. 14:e1007308.
Deonovic, B., Wang, Y., Weirather, J., Wang, X.J., and Au, K.F. (2017). IDP-ASE: haplotyping and quantifying allele-specific expression at the gene and gene isoform level by hybrid sequencing. Nucleic Acids Res. 45:e32.
Fang, F., Hodges, E., Molaro, A., Dean, M., Hannon, G.J., and Smith, A.D. (2012). Genomic landscape of human allele-specific DNA methylation. Proc. Natl. Acad. Sci. U S A 109:7332–7337.
Ghaleb, M.A.A., Li, C., Shahid, M.Q., Yu, H., Liang, J., Chen, R., Wu, J., and Liu, X. (2020). Heterosis analysis and underlying molecular regulatory mechanism in a wide-compatible neo-tetraploid rice line with long panicles. BMC Plant Biol. 20:83.
Phasing analysis for rice hybrid

Giani, A.M., Gallo, G.R., Gianfranceschi, L., and Formenti, G. (2020). Long walk to genomics: history and current approaches to genome sequencing and assembly. Comput. Struct. Biotechnol. J. 18:9–19.

Goff, S.A., Ricke, D., Lan, T.H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Deller, P., Varma, H., et al. (2002). A draft sequence of the rice genome (Oryza sativa L. ssp. japonica). Science 296:92–100.

Goulet, B.E., Roda, F., and Hopkins, R. (2017). Hybridization in plants: old ideas, new techniques. Plant Physiol. 173:65–78.

Guo, M., Yang, S., Rupe, M., Hu, B., Bickel, D.R., Arthur, L., and Smith, O. (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.

Guo, Z., Song, G., Liu, Z., Qu, X., Chen, R., Jiang, D., Sun, Y., Liu, C., Zhu, Y., and Yang, D. (2015). Global epigenomic analysis indicates that epialleles contribute to allele-specific expression via allele-specific histone modifications in hybrid rice. BMC Genomics 16:232.

Hofmeister, B.T., Lee, K., Rohr, N.A., Hall, D.W., and Schmitz, R.J. (2017). Stable inheritance of DNA methylation allows creation of epigenotype maps and the study of epiallele inheritance patterns in the absence of genetic variation. Genome Biol. 18:155.

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K. (2007). WoLF PSORT: protein localization predictor. Nucleic Acids Res. 35:W585–W587.

Huang, X., Yang, S., Gong, J., Zhao, Q., Feng, O., Zhan, Q., Zhao, Y., Li, W., Cheng, B., Xia, J., et al. (2016). Genomic architecture of heterosis for yield traits in rice. Nature 537:629–633.

Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genotype alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37:907–915.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-seq applications. Bioinformatics 27:1571–1572.

Krueger, F., and Andrews, S.R. (2016). SNPsplit: allele-specific splitting of alignments between genomes with known SNP genotypes. P1000Res 5:1479.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9:559.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. 11:204–220.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100.

Li, H., Zhou, S.Y., Zhao, W.S., Su, S.C., and Peng, Y.L. (2009). A novel wall-associated receptor-like protein kinase gene, OsWAK1, plays important roles in rice blast disease resistance. Plant Mol. Biol. 69:337–346.

Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., and Zimin, A. (2018). MUMmer4: a fast and versatile genome alignment system. PLoS Comput. Biol. 14:e1005944.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303.

Ottenburghs, J. (2018). Exploring the hybrid speciation continuum in birds. Ecol. Evol. 8:13027–13034.

Pastinen, T. (2010). Genome-wide allele-specific analysis: insights into regulatory variation. Nat. Rev. Genet. 11:533–538.
Xie, F., and Zhang, J. (2018). Shanyou 63: an elite mega rice hybrid in China. Rice 11:17.

Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., et al. (2002). A draft sequence of the rice genome (Oryza sativa L. ssp. indica). Science 296:79–92.

Zhang, J., Chen, L.L., Xing, F., Kudrna, D.A., Yao, W., Copetti, D., Mu, T., Li, W., Song, J.M., Xie, W., 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. U S A 113:E5163–E5171.

Zhang, S., Chen, C., Li, L., Meng, L., Singh, J., Jiang, N., Deng, X.W., He, Z.H., and Lemaux, P.G. (2005). Evolutionary expansion, gene structure, and expression of the rice wall-associated kinase gene family. Plant Physiol. 139:1107–1124.

Zhang, X., Wu, R., Wang, Y., Yu, J., and Tang, H. (2020). Unzipping haplotypes in diploid and polyploid genomes. Comput. Struct. Biotechnol. J. 18:66–72.

Zhang, X., Zhang, S., Zhao, Q., Ming, R., and Tang, H. (2019). Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. Nat. Plants 5:833–845.
Supplemental information

Phasing analysis of the transcriptome and epigenome in a rice hybrid reveals the inheritance and difference in DNA methylation and allelic transcription regulation

Jia-Wu Feng, Yue Lu, Lin Shao, Jianwei Zhang, Huan Li, and Ling-Ling Chen
Phasing Analysis of Transcriptome and Epigenome in Rice Hybrid Reveals the Inheritance and Difference of DNA Methylation and Allelic Transcription Regulation

Jia-Wu Feng¹,², Yue Lu³, Lin Shao¹, Jianwei Zhang¹, Huan Li¹, *, Ling-Ling Chen¹, ², *

¹ National Key Laboratory of Crop Genetic Improvement, College of Informatics, Huazhong Agricultural University, Wuhan 430070, China
² College of Life Science and Technology, Guangxi University, Nanning 530004, China
³ Jiangsu Key Laboratory of Crop Genetics and Physiology/Co-Innovation Center for Modern Production Technology of Grain Crops, Key Laboratory of Plant Functional Genomics of the Ministry of Education, Yangzhou University, Yangzhou, 225009, China
Consistent Ratio (100%)

Varieties

MH63  ZS97

88.14%  86.12%

86.75%  87.92%

83.35%  86.03%

86.39%  85.87%

85.99%  85.99%

RNA-Seq

MH63  SY63  ZS97

A

MH63

ZS97

Varieties

B

0.00%  20.00%  40.00%  60.00%  80.00%  100.00%

Error rate (100%)

C

Consistent Ratio

Varieties

Flag/TLD  panicle/LTLD  flag/TLD  panicle/LTLD  flag/TLD  panicle/LTLD  flag/TLD  panicle/LTLD  flag/TLD  panicle/LTLD

88.14%  86.12%  88.51%  86.75%  87.92%  83.35%  86.03%  86.36%  87.99%  86.39%  85.87%  85.99%

D

R-squared: 0.9937

R-squared: 0.984

R-squared: 0.9244

CG methylation level before phasing

CHG methylation level before phasing

CHH methylation level before phasing

E

R-squared: 0.9943

R-squared: 0.9881

R-squared: 0.9501

CG methylation level before phasing

CHG methylation level before phasing

CHH methylation level before phasing

CG methylation level after phasing

CHG methylation level after phasing

CHH methylation level after phasing
Supplemental Figure 1. Evaluation of the Phasing Results.

(A) Error rate of phasing for different varieties in RNA-Seq.

(B) Separation rate of phasing for different varieties in RNA-Seq.

(C) The ratio of DEGs after phasing is consistent with DEGs before phasing under different conditions.

(D) The correlation between CG\CHG\CHH the methylation levels in each 50 bp region before and after phasing in MH63 samples.

(E) The correlation between CG\CHG\CHH the methylation levels in each 50 bp region before and after phasing in ZS97 samples.
Supplemental Figure 2. Comparison of SNPsplit and PP2PG.
(A) Error rate of phasing for SNPsplit and PP2PG.
(B) Separation rate of phasing for SNPsplit and PP2PG.
Supplemental Figure 3. The Correlation Coefficient Between Parents and Progeny after Phasing.

(A) The correlation coefficient of flag leaves in MH63 type.
(B) The correlation coefficient of flag leaves in ZS97 type.
(C) The correlation coefficient of panicles in MH63 type.
(D) The correlation coefficient of panicles in ZS97 type.
(E) The correlation coefficient of shoots in MH63 type.
(F) The correlation coefficient of shoots in ZS97 type. From the above figures, the value of correlation coefficients between parents and progeny is higher in shoots than that in flag leaves and panicles.
Supplemental Figure 4. The Construction of Allele Co-expression Network in SY63. (A) The dendrogram and heatmap of SY63 after phasing. (B) and (C) The soft threshold of allele co-expression network in SY63 under two conditions. The red line illustrates the threshold for filtering.
Supplemental Figure 5. The Frequency Distribution of Nodes in Co-expression Network.

(A) The frequency distribution of nodes that interacted with ZS97 and MH63 alleles. The x-axis represents the frequency of nodes that interacted with MH63 and ZS97 alleles in all nodes. The y-axis represents density. The higher the frequency, the more is similar between the two interaction alleles.

(B) The frequency distribution of nodes only interacted with MH63 alleles in divergent interaction. The x-axis represents the frequency of nodes that only interact with MH63 alleles in divergent interactions. The higher the frequency, the more interactions are with the MH63 alleles than with the ZS97 alleles.
Supplemental Figure 6. The GO Analysis for Different Types of Alleles.
Supplemental Figure 7. The Alignment of cDNA Sequences for *Hd3a/hd3a* (OsMH_06G0050000/ OsZS_06G0051000). The orange color represents the difference between these two alleles.
Supplemental Figure 8. The Alignment of Protein Sequences for OsWAK38 (OsMH_04G0272400 / OsZS_04G0266500). The orange color represents the differences between these two alleles.
Supplemental Figure 9. The Mutation at Chr04:17,959,304 of ZS97 Genome Between the Two Alleles of OsWAK38. The red color indicates the mutated amino acid, and this mutation results in the deletion of 87 amino acids in its protein.
Supplemental Figure 10. The Results of Signal Peptide Prediction in OsWAK38. 
(A) and (B) are the SignalP-5.0 prediction of OsMH_04G0272400 and OsZS_04G0266500, respectively.
Supplemental Figure 11. The GO Analysis of OsWAK38 Allele Interaction Network.
Supplemental Figure 12. The Overlapping of ASM Regions Between Parents and Progeny in Three Methylations, CG (A, B), CHG (C, D) and CHH (E, F). SY63_ZS97<SY63_MH63 illustrates that the methylation of MH63 type is higher than that of ZS97 type in the ASM region of the hybrid; ZS97_ZS97<MH63_MH63 demonstrates that the methylation of MH63 type is higher than that of ZS97 type in the DMR region of the parents; SY63_ZS97>SY63_MH63 illustrates that the methylation of MH63 type is lower than that of ZS97 type in the ASM region of the hybrid; ZS97_ZS97>MH63_MH63 demonstrates that the methylation of MH63 type is lower than that of ZS97 type in the DMR region of the parents.
Supplemental Figure 13. The Relationship Between ASM Region of Different Methylation Types and Transposons.

(A) The distribution of the total transposons in different methylation types.
(B) The distribution of the DNA transposons in different methylation types.
(C) The distribution of the LTR retrotransposons in different methylation types.
### Supplemental Table 1. Summary of the Iso-Seq Data

| Tissue | Varieties | Cell Number | Library of Reads Length (Kbp) | Reads after CSS |
|--------|-----------|-------------|-------------------------------|-----------------|
| Leaf   | MH63      | 5           | 1-2                           | 251,835         |
| Leaf   | MH63      | 5           | 2-4                           | 481,165         |
| Leaf   | MH63      | 2           | 4+                            | 218,082         |
| Leaf   | ZS97      | 7           | 1-2                           | 336,171         |
| Leaf   | ZS97      | 4           | 2-4                           | 439,891         |
| Leaf   | ZS97      | 2           | 4+                            | 225,387         |
| Leaf   | SY63      | 6           | 1-2                           | 232,891         |
| Leaf   | SY63      | 4           | 2-4                           | 442,228         |
| Leaf   | SY63      | 2           | 4+                            | 214,702         |
| Leaf   | RC        | 7           | 1-2                           | 321,573         |
| Leaf   | RC        | 4           | 2-4                           | 446,093         |
| Leaf   | RC        | 2           | 4+                            | 224,200         |
| Panicle| SY63      | 1           | 1-2                           | 77,623          |
| Panicle| SY63      | 4           | 2-4                           | 222,233         |
| Panicle| SY63      | 3           | 4+                            | 102,416         |
| Sample  | Raw Reads   | Raw Bases (GB) | Clean Reads  | Clean Bases (GB) |
|---------|-------------|----------------|--------------|------------------|
| MH63-1  | 129,164,940 | 19.37          | 127,365,562  | 19.1             |
| MH63-2  | 102,181,882 | 15.33          | 100,743,922  | 15.12            |
| SY63-1  | 109,899,050 | 16.48          | 108,131,250  | 16.22            |
| SY63-2  | 131,256,968 | 19.69          | 129,045,642  | 19.36            |
| ZS97-1  | 118,535,102 | 17.78          | 116,582,542  | 17.48            |
| ZS97-2  | 107,840,302 | 16.18          | 105,766,856  | 15.86            |
### Supplemental Table 3. The Mapping Rates Based on Two Parental Genomes

| Type of dataset | Sample          | Mapping rates based on MH63 genome | Mapping rates based on ZS97 genome |
|-----------------|-----------------|------------------------------------|-----------------------------------|
| BS-Seq          | SY63-1          | 73.00%                             | 71.90%                            |
| BS-Seq          | SY63-2          | 72.10%                             | 71.00%                            |
| BS-Seq          | MH63-1          | 79.40%                             | 68.40%                            |
| BS-Seq          | MH63-2          | 79.30%                             | 68.40%                            |
| BS-Seq          | ZS97-1          | 68.50%                             | 76.90%                            |
| BS-Seq          | ZS97-2          | 69.00%                             | 77.50%                            |
| Iso-Seq         | Leaf MH63       | 99.22%                             | 98.37%                            |
| Iso-Seq         | Leaf ZS97       | 98.71%                             | 99.05%                            |
| Iso-Seq         | Leaf SY63       | 99.17%                             | 99.09%                            |
| Iso-Seq         | Leaf RC         | 98.86%                             | 98.69%                            |
| RNA-Seq         | Flagleaf SY63   | 98.59%                             | 98.47%                            |
| RNA-Seq         | Flagleaf ZS97   | 98.18%                             | 98.60%                            |
| RNA-Seq         | Flagleaf SY63   | 97.83%                             | 98.56%                            |
| RNA-Seq         | Flagleaf ZS97   | 98.26%                             | 98.58%                            |
| RNA-Seq         | Flagleaf SY63   | 96.86%                             | 97.14%                            |
| RNA-Seq         | Flagleaf SY63   | 98.32%                             | 98.61%                            |
| RNA-Seq         | Flagleaf SY63   | 98.24%                             | 98.50%                            |
| RNA-Seq         | Flagleaf SY63   | 98.47%                             | 98.71%                            |
| RNA-Seq         | Flagleaf SY63   | 98.45%                             | 98.74%                            |
| RNA-Seq         | Panicle SY63    | 98.50%                             | 98.78%                            |
| RNA-Seq         | Panicle SY63    | 98.36%                             | 98.79%                            |
| RNA-Seq         | Panicle SY63    | 98.45%                             | 98.78%                            |
| RNA-Seq         | Panicle SY63    | 98.43%                             | 98.69%                            |
| RNA-Seq         | Panicle SY63    | 98.54%                             | 98.78%                            |
| RNA-Seq         | Panicle SY63    | 98.48%                             | 98.78%                            |
| RNA-Seq         | Panicle SY63    | 98.47%                             | 98.73%                            |
| RNA-Seq         | Panicle SY63    | 98.53%                             | 98.77%                            |
| RNA-Seq         | Shoot SY63      | 98.65%                             | 98.91%                            |
| RNA-Seq         | Shoot SY63      | 98.70%                             | 98.92%                            |
| RNA-Seq         | Shoot SY63      | 98.68%                             | 98.98%                            |
| RNA-Seq         | Shoot SY63      | 98.82%                             | 98.31%                            |
| RNA-Seq         | Shoot SY63      | 98.36%                             | 98.97%                            |
| RNA-Seq         | Shoot SY63      | 98.77%                             | 98.97%                            |
| RNA-Seq         | Shoot SY63      | 98.85%                             | 98.97%                            |
| RNA-Seq         | Shoot SY63      | 98.79%                             | 99.02%                            |
| RNA-Seq         | Flagleaf ZS97   | 97.21%                             | 99.67%                            |
| RNA-Seq         | Flagleaf ZS97   | 97.34%                             | 99.69%                            |
| RNA-Seq         | Flagleaf ZS97   | 97.33%                             | 99.67%                            |
| RNA-Seq         | Flagleaf ZS97   | 97.24%                             | 99.67%                            |
| RNA-Seq         | Flagleaf ZS97   | 97.25%                             | 99.66%                            |
| RNA-Seq         | Flagleaf ZS97   | 97.41%                             | 99.66%                            |
| RNA-Seq         | Flagleaf ZS97   | 97.41%                             | 99.71%                            |
| RNA-Seq         | Flagleaf ZS97   | 97.33%                             | 99.72%                            |
| RNA-Seq         | Panicle ZS97    | 97.34%                             | 99.55%                            |
| RNA-Seq         | Panicle ZS97    | 97.53%                             | 99.65%                            |
| RNA-Seq         | Panicle ZS97    | 97.53%                             | 99.70%                            |
| RNA-Seq         | Panicle ZS97    | 97.55%                             | 99.71%                            |
| RNA-Seq         | Panicle ZS97    | 97.85%                             | 99.68%                            |
| RNA-Seq         | Panicle ZS97    | 97.29%                             | 99.60%                            |
| RNA-Seq         | Panicle ZS97    | 97.71%                             | 99.72%                            |
| RNA-Seq         | Panicle ZS97    | 97.06%                             | 99.66%                            |
| RNA-Seq         | Shoot ZS97      | 97.56%                             | 99.60%                            |
| RNA-Seq         | Shoot ZS97      | 96.81%                             | 99.53%                            |
| Type of dataset | Sample               | Mapping rates based on MH63 genome | Mapping rates based on ZS97 genome |
|----------------|----------------------|------------------------------------|-----------------------------------|
| RNA-Seq        | shoot_ZS97_HTSD_rep1 | 97.90%                             | 99.78%                            |
| RNA-Seq        | shoot_ZS97_HTSD_rep2 | 96.30%                             | 99.57%                            |
| RNA-Seq        | shoot_ZS97_LTLD_rep1 | 97.97%                             | 99.76%                            |
| RNA-Seq        | shoot_ZS97_LTLD_rep2 | 97.93%                             | 99.74%                            |
| RNA-Seq        | shoot_ZS97_LTSD_rep1 | 97.90%                             | 99.76%                            |
| RNA-Seq        | shoot_ZS97_LTSD_rep2 | 97.98%                             | 99.75%                            |
| RNA-Seq        | flagleaf_MH63_HTLD_rep1 | 99.37%                         | 97.47%                            |
| RNA-Seq        | flagleaf_MH63_HTLD_rep2 | 99.45%                         | 97.48%                            |
| RNA-Seq        | flagleaf_MH63_HTSD_rep1 | 99.40%                         | 97.63%                            |
| RNA-Seq        | flagleaf_MH63_HTSD_rep2 | 99.43%                         | 97.65%                            |
| RNA-Seq        | flagleaf_MH63_LTLD_rep1 | 99.30%                         | 97.39%                            |
| RNA-Seq        | flagleaf_MH63_LTLD_rep2 | 99.11%                         | 97.27%                            |
| RNA-Seq        | flagleaf_MH63_LTSD_rep1 | 99.06%                         | 97.53%                            |
| RNA-Seq        | flagleaf_MH63_LTSD_rep2 | 98.70%                         | 97.48%                            |
| RNA-Seq        | panicle_MH63_HTLD_rep1 | 99.30%                         | 97.86%                            |
| RNA-Seq        | panicle_MH63_HTLD_rep2 | 99.32%                         | 97.87%                            |
| RNA-Seq        | panicle_MH63_HTSD_rep1 | 99.26%                         | 97.86%                            |
| RNA-Seq        | panicle_MH63_HTSD_rep2 | 99.35%                         | 97.85%                            |
| RNA-Seq        | panicle_MH63_LTLD_rep1 | 99.06%                         | 97.68%                            |
| RNA-Seq        | panicle_MH63_LTLD_rep2 | 99.08%                         | 97.68%                            |
| RNA-Seq        | panicle_MH63_LTSD_rep1 | 99.34%                         | 97.92%                            |
| RNA-Seq        | panicle_MH63_LTSD_rep2 | 99.37%                         | 97.94%                            |
| RNA-Seq        | shoot_MH63_HTLD_rep1 | 99.47%                         | 98.08%                            |
| RNA-Seq        | shoot_MH63_HTLD_rep2 | 98.90%                         | 98.09%                            |
| RNA-Seq        | shoot_MH63_HTSD_rep1 | 99.51%                         | 98.12%                            |
| RNA-Seq        | shoot_MH63_HTSD_rep2 | 99.33%                         | 98.17%                            |
| RNA-Seq        | shoot_MH63_LTLD_rep1 | 99.43%                         | 98.25%                            |
| RNA-Seq        | shoot_MH63_LTLD_rep2 | 99.40%                         | 98.24%                            |
| RNA-Seq        | shoot_MH63_LTSD_rep1 | 99.40%                         | 98.20%                            |
| RNA-Seq        | shoot_MH63_LTSD_rep2 | 99.39%                         | 98.26%                            |
### Supplemental Table 4. Statistics of SNPs in BS-Seq Data

| Unavailable for phasing | SNP Type     | SNP Number |
|-------------------------|--------------|------------|
| Unavailable             | A G (A)      | 301,014    |
| Unavailable             | C (T) T      | 304,760    |
| Unavailable             | G (A) A      | 303,896    |
| Unavailable             | T C (T)      | 300,042    |
| Unavailable             | Total        | 1,209,712  |

| Available for phasing   | SNP Type     | SNP Number |
|-------------------------|--------------|------------|
| Available               | A C (T)      | 65,285     |
| Available               | A T          | 77,614     |
| Available               | C (T) A      | 65,965     |
| Available               | C (T) G (A)  | 49,663     |
| Available               | G (A) C (T)  | 49,576     |
| Available               | G (A) T      | 66,078     |
| Available               | T A          | 78,166     |
| Available               | T G (A)      | 65,288     |
| Available               | Total        | 517,635    |
### Supplemental Table 5. Summary of Phasing Related Genes

| Type of dataset | Sample | Related gene number |
|-----------------|--------|---------------------|
| RNA-Seq         | flagleaf_SY63_HTLD_rep1 | 7,911               |
| RNA-Seq         | flagleaf_SY63_HTLD_rep2 | 7,915               |
| RNA-Seq         | flagleaf_SY63_HTSD_rep1 | 7,872               |
| RNA-Seq         | flagleaf_SY63_HTSD_rep2 | 7,942               |
| RNA-Seq         | flagleaf_SY63_LTLD_rep1 | 7,796               |
| RNA-Seq         | flagleaf_SY63_LTLD_rep2 | 7,905               |
| RNA-Seq         | flagleaf_SY63_LTSD_rep1 | 7,617               |
| RNA-Seq         | flagleaf_SY63_LTSD_rep2 | 7,288               |
| RNA-Seq         | panicle_SY63_HTLD_rep1  | 9,702               |
| RNA-Seq         | panicle_SY63_HTLD_rep2  | 9,569               |
| RNA-Seq         | panicle_SY63_HTSD_rep1  | 9,548               |
| RNA-Seq         | panicle_SY63_HTSD_rep2  | 9,672               |
| RNA-Seq         | panicle_SY63_LTLTLD_rep1| 9,473               |
| RNA-Seq         | panicle_SY63_LTLTLD_rep2| 9,429               |
| RNA-Seq         | panicle_SY63_LTSD_rep1  | 9,640               |
| RNA-Seq         | panicle_SY63_LTSD_rep2  | 9,293               |
| RNA-Seq         | shoot_SY63_HTLD_rep1    | 9,082               |
| RNA-Seq         | shoot_SY63_HTLD_rep2    | 9,106               |
| RNA-Seq         | shoot_SY63_HTSD_rep1    | 9,187               |
| RNA-Seq         | shoot_SY63_HTSD_rep2    | 8,883               |
| RNA-Seq         | shoot_SY63_LTLTLD_rep1  | 8,727               |
| RNA-Seq         | shoot_SY63_LTLTLD_rep2  | 8,915               |
| RNA-Seq         | shoot_SY63_LTSD_rep1    | 8,805               |
| RNA-Seq         | shoot_SY63_LTSD_rep2    | 8,514               |
| Iso-Seq         | Leaf SY63              | 3,707               |
| Iso-Seq         | Pannicle SY63          | 2,351               |
| Iso-Seq         | Leaf RC                | 3,706               |
| BS-Seq          | Detection regions of SY63 CG | 18,763         |
| BS-Seq          | Detection regions of SY63 CHG | 18,852          |
| BS-Seq          | Detection regions of SY63 CHH | 19,085          |
**Supplemental Table 6. Phasing Results of the Iso-Seq Data**

| Varieties | Tissue   | MH63-synteny reads | ZS97-synteny reads | Unknown reads | MH63-only reads | ZS97-only reads | Error Rate | Separation Rate | Error Rate for -only reads | Separation Rate for -only reads |
|-----------|----------|--------------------|--------------------|---------------|----------------|----------------|------------|-----------------|-----------------------------|---------------------------------|
| MH63      | leaf     | 174,449            | 3,984              | 246,402       | 4,186          | 409            | 2.23%      | 41.55%          | 8.90%                        | 1.07%                            |
| ZS97      | leaf     | 453                | 182,286            | 258,799       | 51             | 1,656          | 0.25%      | 41.23%          | 2.99%                        | 0.39%                            |
| SY63      | leaf     | 81,268             | 84,606             | 236,647       | 1,097          | 810            |            | 41.01%          | 0.47%                        |                                 |
| RC        | leaf     | 85,489             | 89,482             | 261,764       | 1,553          | 812            |            | 39.85%          | 0.54%                        |                                 |
| SY63      | Panicle  | 50,004             | 50,250             | 134,085       | 628            | 376            |            | 42.60%          | 0.43%                        |                                 |
Supplemental Table 7. Phasing Results of the RNA-Seq Data

| Sample                | Tissue    | MH63-synteny reads | ZS97-synteny reads | Unknown reads | MH63-only reads | ZS97-only reads | Error Rate | Separation Rate | Error Rate for -only reads | Separation Rate for -only reads |
|-----------------------|-----------|--------------------|--------------------|---------------|----------------|----------------|------------|-----------------|-------------------------------|---------------------------------|
| flagleaf_ZS97_HTLD_rep1 | Flag leaf | 8,288              | 5,637,442          | 55,055,800    | 15,576         | 1,128,976      | 0.15%      | 9.13%           | 1.36%                         | 1.85%                            |
| flagleaf_ZS97_HTLD_rep2 | Flag leaf | 9,002              | 5,821,502          | 57,324,860    | 16,316         | 1,125,284      | 0.15%      | 9.07%           | 1.43%                         | 1.78%                            |
| flagleaf_ZS97_HTSD_rep1 | Flag leaf | 7,926              | 5,273,410          | 51,047,564    | 14,026         | 1,003,454      | 0.15%      | 9.21%           | 1.38%                         | 1.77%                            |
| flagleaf_ZS97_HTSD_rep2 | Flag leaf | 6,208              | 5,297,300          | 51,392,014    | 13,656         | 1,058,414      | 0.12%      | 9.18%           | 1.27%                         | 1.86%                            |
| flagleaf_ZS97_LTLD_rep1 | Flag leaf | 5,410              | 4,681,366          | 45,387,520    | 11,668         | 916,690        | 0.12%      | 9.19%           | 1.26%                         | 1.82%                            |
| flagleaf_ZS97_LTLD_rep2 | Flag leaf | 13,752             | 4,115,024          | 49,170,984    | 11,916         | 742,368        | 0.33%      | 9.13%           | 1.50%                         | 1.67%                            |
| flagleaf_ZS97_LTSD_rep1 | Flag leaf | 4,668              | 5,197,572          | 50,272,820    | 11,722         | 1,006,656      | 0.09%      | 9.21%           | 1.15%                         | 1.80%                            |
| flagleaf_ZS97_LTSD_rep2 | Flag leaf | 6,296              | 5,753,808          | 54,517,836    | 11,608         | 950,958        | 0.11%      | 9.57%           | 1.30%                         | 1.48%                            |
| flagleaf_ZS97_LTSD_rep2 | Panicle   | 6,792              | 5,923,082          | 55,864,426    | 11,952         | 860,958        | 0.12%      | 9.46%           | 1.37%                         | 1.39%                            |
| flagleaf_ZS97_LTSD_rep1 | Panicle   | 8,222              | 6,047,014          | 57,810,238    | 13,936         | 874,200        | 0.12%      | 9.35%           | 1.57%                         | 1.37%                            |
| shoot_ZS97_HTLD_rep1   | shoot     | 7,112              | 6,047,014          | 57,810,238    | 13,936         | 874,200        | 0.12%      | 9.46%           | 1.37%                         | 1.39%                            |
| shoot_ZS97_HTLD_rep2   | shoot     | 14,156             | 6,918,004          | 66,890,772    | 15,672         | 1,409,888      | 0.20%      | 9.21%           | 1.10%                         | 1.89%                            |
| shoot_ZS97_HTSD_rep1   | shoot     | 12,284             | 5,508,906          | 50,624,424    | 9,680          | 668,560        | 0.22%      | 9.87%           | 1.43%                         | 1.19%                            |
| shoot_ZS97_HTSD_rep2   | shoot     | 8,302              | 5,708,524          | 54,338,222    | 10,100         | 1,412,260      | 0.15%      | 9.30%           | 0.71%                         | 2.31%                            |
| shoot_ZS97_LTLD_rep1   | shoot     | 8,018              | 5,469,058          | 49,422,978    | 10,328         | 613,306        | 0.15%      | 9.86%           | 1.66%                         | 1.12%                            |
| shoot_ZS97_LTSD_rep1   | shoot     | 4,678              | 5,266,160          | 47,141,942    | 10,022         | 586,956        | 0.09%      | 9.94%           | 1.68%                         | 1.13%                            |
| shoot_ZS97_LTSD_rep2   | shoot     | 8,452              | 6,044,738          | 53,987,292    | 10,534         | 693,402        | 0.14%      | 9.97%           | 1.50%                         | 1.16%                            |
| flagleaf_MH63_HTLD_rep1 | Flag leaf | 4,742,248          | 29,378             | 47,748,730    | 865,478        | 145,322        | 0.62%      | 8.91%           | 14.38%                        | 1.89%                            |
| flagleaf_MH63_HTLD_rep2 | Flag leaf | 5,475,876          | 17,986             | 54,009,120    | 982,556        | 134,046        | 0.33%      | 9.06%           | 12.00%                        | 1.84%                            |
| flagleaf_MH63_HTSD_rep1 | Flag leaf | 5,671,976          | 22,428             | 57,798,828    | 972,764        | 155,822        | 0.39%      | 8.81%           | 13.81%                        | 1.75%                            |
| flagleaf_MH63_HTSD_rep2 | Flag leaf | 4,618,302          | 17,060             | 47,661,274    | 810,948        | 128,070        | 0.37%      | 8.71%           | 13.64%                        | 1.76%                            |
| flagleaf_MH63_LTLD_rep1 | Flag leaf | 5,252,882          | 16,256             | 51,606,680    | 914,936        | 125,710        | 0.31%      | 9.10%           | 12.08%                        | 1.80%                            |
| flagleaf_MH63_LTLD_rep2 | Flag leaf | 4,916,274          | 18,094             | 48,679,306    | 844,908        | 128,046        | 0.37%      | 9.04%           | 13.16%                        | 1.78%                            |
| flagleaf_MH63_LTSD_rep1 | Flag leaf | 4,036,494          | 39,076             | 40,535,252    | 656,870        | 191,742        | 0.96%      | 8.97%           | 22.59%                        | 1.87%                            |
| Sample              | Tissue     | MH63-synteny reads | ZS97-synteny reads | Unknown reads | MH63-only reads | ZS97-only reads | Error Rate | Separation Rate | Error Rate for -only reads | Separation Rate for -only reads |
|---------------------|------------|--------------------|--------------------|--------------|----------------|----------------|------------|----------------|-----------------------------|------------------------------|
| flagleaf_MH63_LTSD_rep2 | Flag leaf  | 3,255,390          | 27,590             | 33,402,444   | 518,530        | 214,924        | 0.84%      | 8.77%          | 29.30%                      | 1.96%                        |
| panicle_MH63_HTLD_rep1 | Panicle    | 6,330,992          | 14,976             | 60,271,372   | 850,306        | 232,372        | 0.24%      | 9.37%          | 21.46%                      | 1.60%                        |
| panicle_MH63_HTLD_rep2 | Panicle    | 5,335,610          | 14,786             | 52,910,306   | 743,120        | 198,912        | 0.27%      | 9.34%          | 21.12%                      | 1.59%                        |
| panicle_MH63_HTSD_rep1 | Panicle    | 5,442,718          | 15,002             | 53,029,754   | 740,766        | 217,392        | 0.27%      | 9.18%          | 22.69%                      | 1.61%                        |
| panicle_MH63_HTSD_rep2 | Panicle    | 5,806,426          | 11,624             | 56,872,276   | 811,468        | 199,360        | 0.20%      | 9.13%          | 19.64%                      | 1.59%                        |
| flagleaf_SY63_HTLD_rep1 | Flag leaf  | 2,672,692          | 2,795,610          | 53,511,028   | 489,128        | 697,272        | -          | 9.09%          | -                           | -                            |
| flagleaf_SY63_HTLD_rep2 | Flag leaf  | 2,554,844          | 2,633,714          | 51,108,420   | 463,008        | 797,720        | -          | 9.01%          | -                           | -                            |
| flagleaf_SY63_HTSD_rep1 | Flag leaf  | 2,428,282          | 2,341,420          | 47,317,034   | 386,874        | 521,318        | -          | 9.00%          | -                           | -                            |
| flagleaf_SY63_HTSD_rep2 | Flag leaf  | 2,529,766          | 2,607,762          | 48,652,118   | 368,864        | 488,046        | -          | 9.40%          | -                           | -                            |
| panicle_SY63_HTSD_rep1 | Panicle    | 2,762,716          | 2,871,280          | 52,312,058   | 361,908        | 546,200        | -          | 9.57%          | -                           | -                            |
| panicle_SY63_HTSD_rep2 | Panicle    | 2,675,310          | 2,744,752          | 50,941,570   | 355,940        | 497,190        | -          | 9.47%          | -                           | -                            |
| flagleaf_SY63_LTSD_rep1 | Flag leaf  | 2,601,744          | 2,670,998          | 50,988,342   | 394,008        | 507,836        | -          | 9.22%          | -                           | -                            |
| panicle_SY63_LTSD_rep2 | Panicle    | 2,266,208          | 2,334,754          | 43,405,562   | 320,290        | 429,590        | -          | 9.44%          | -                           | -                            |
| panicle_SY63_HTLD_rep1 | Panicle    | 3,152,894          | 3,252,670          | 60,754,518   | 426,974        | 571,398        | -          | 9.40%          | -                           | -                            |
| panicle_SY63_HTLD_rep2 | Panicle    | 2,762,716          | 2,871,280          | 52,312,058   | 361,908        | 546,200        | -          | 9.57%          | -                           | -                            |
| panicle_SY63_HTSD_rep1 | Panicle    | 2,675,310          | 2,744,752          | 50,941,570   | 355,940        | 497,190        | -          | 9.47%          | -                           | -                            |
| flagleaf_SY63_LTSD_rep1 | Flag leaf  | 2,610,744          | 2,670,998          | 50,988,342   | 394,008        | 507,836        | -          | 9.22%          | -                           | -                            |
| flagleaf_SY63_LTSD_rep2 | Flag leaf  | 2,266,208          | 2,334,754          | 43,405,562   | 320,290        | 429,590        | -          | 9.44%          | -                           | -                            |
| panicle_SY63_LTLD_rep1 | Panicle    | 3,152,894          | 3,252,670          | 60,754,518   | 426,974        | 571,398        | -          | 9.40%          | -                           | -                            |
| panicle_SY63_LTLD_rep2 | Panicle    | 2,443,280          | 2,477,496          | 47,995,708   | 327,886        | 426,606        | -          | 9.17%          | -                           | -                            |
| panicle_SY63_LTSD_rep2 | Panicle    | 2,401,932          | 2,425,560          | 47,306,224   | 321,452        | 445,320        | -          | 9.13%          | -                           | -                            |
| panicle_SY63_LTSD_rep1 | Panicle    | 2,806,028          | 2,834,120          | 55,507,666   | 406,538        | 527,874        | -          | 9.08%          | -                           | -                            |
| Sample                  | Tissue | MH63-synten reads | ZS97-synten reads | Unknown reads | MH63-only reads | ZS97-only reads | Error Rate | Separation Rate for -only reads | Error Rate for -only reads | Separation Rate for -only reads |
|------------------------|--------|------------------|------------------|--------------|----------------|----------------|-------------|---------------------------------|-----------------------------|---------------------------------|
| panicle_SY63_LTSD_rep2 | Panicle| 2,431,906        | 2,460,964        | 48,159,704   | 331,604        | 430,156        | -           | 9.09%                           | -                           | 1.42%                            |
| shoot_SY63_HTLD_rep1   | shoot  | 3,105,278        | 3,149,804        | 60,608,948   | 363,466        | 497,586        | -           | 9.24%                           | -                           | 1.27%                            |
| shoot_SY63_HTLD_rep2   | shoot  | 3,116,536        | 3,150,742        | 59,277,238   | 353,166        | 458,196        | -           | 9.44%                           | -                           | 1.22%                            |
| shoot_SY63_HTSN_rep1   | shoot  | 3,920,612        | 4,001,476        | 71,621,534   | 390,956        | 555,632        | -           | 9.84%                           | -                           | 1.18%                            |
| shoot_SY63_HTSN_rep2   | shoot  | 3,269,076        | 3,332,362        | 60,430,858   | 327,358        | 414,690        | -           | 9.74%                           | -                           | 1.09%                            |
| shoot_SY63_LTLD_rep1   | shoot  | 2,723,208        | 2,811,790        | 50,587,536   | 272,900        | 522,538        | -           | 9.72%                           | -                           | 1.40%                            |
| shoot_SY63_LTLD_rep2   | shoot  | 3,198,734        | 3,221,570        | 58,530,222   | 325,054        | 413,552        | -           | 9.77%                           | -                           | 1.12%                            |
| shoot_SY63_LTSN_rep1   | shoot  | 3,186,432        | 3,217,742        | 57,596,530   | 317,484        | 384,734        | -           | 9.90%                           | -                           | 1.09%                            |
| shoot_SY63_LTSN_rep2   | shoot  | 2,760,228        | 2,778,738        | 49,601,134   | 260,044        | 353,558        | -           | 9.93%                           | -                           | 1.10%                            |
Supplemental Table 8. Phasing Results of the BS-Seq Data

| Varieties | Tissue  | MH63-synten reads | ZS97-synten reads | Unknown reads | MH63-only reads | ZS97-only reads | Error Rate | Separation Rate | Error Rate for-only reads | Separation Rate for-only reads |
|-----------|---------|-------------------|-------------------|---------------|----------------|----------------|-----------|----------------|-----------------------------|-----------------------------|
| MH63      | leaf    | 9,482,614         | 208,188           | 73,640,878    | 17,554,190     | 3,509,980      | 2.15%     | 9.28%          | 16.66%                      | 20.18%                      |
| MH63      | leaf    | 7,472,840         | 163,550           | 58,112,692    | 13,818,662     | 2,828,436      | 2.14%     | 9.27%          | 16.99%                      | 20.20%                      |
| SY63      | leaf    | 3,972,352         | 4,046,592         | 61,448,616    | 9,153,186      | 7,963,556      | 9.26%     | 9.26%          | 19.77%                      | 19.77%                      |
| SY63      | leaf    | 4,713,842         | 4,834,190         | 72,347,190    | 10,756,116     | 9,288,474      | 9.37%     | 9.37%          | 19.66%                      | 19.66%                      |
| ZS97      | leaf    | 49,744            | 8,728,600         | 66,326,208    | 4,286,988      | 14,030,590     | 0.57%     | 9.40%          | 23.40%                      | 19.61%                      |
| ZS97      | leaf    | 49,570            | 7,894,668         | 60,600,572    | 4,068,954      | 12,924,232     | 0.62%     | 9.29%          | 23.94%                      | 19.87%                      |
**Supplemental Table 9. Statistical Results of TAPIS Process after Iso-Seq Phasing**

| Sample               | Transcript | Associated gene |
|----------------------|------------|-----------------|
| Leaf MH63_MH63       | 32,548     | 5,815           |
| Leaf SY63_MH63       | 16,427     | 4,895           |
| Leaf RC_MH63         | 16,823     | 4,903           |
| Panicle SY63_MH63    | 14,022     | 3,545           |
| Leaf ZS97_ZS97       | 32,227     | 6,206           |
| Leaf SY63_ZS97       | 16,484     | 5,117           |
| Leaf RC_ZS97         | 17,169     | 5,130           |
| Panicle SY63_ZS97    | 14,224     | 3,665           |