the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2023) **115,** 139–154 doi: 10.1111/tpj.16217

Transcriptomic and methylomic analyses provide insights into the molecular mechanism and prediction of heterosis in rice

Chongyun Fu[1,2,3,4,*,†], Ce Ma[5,†], Manshan Zhu[1,2,3,4], Wuge Liu[1,2,3,4], Xiaozhi Ma[1,2,3,4], Jinhua Li[1,2,3,4], Yilong Liao[1,2,3,4], Dilin Liu[1,2,3,4], Xiaofeng Gu[6,*], Haiyang Wang[7,*] and Feng Wang[1,2,3,4,*]

[1]*Rice Research Institute, Guangdong Academy of Agricultural Sciences, Guangzhou, China,*
[2]*Guangdong Key Laboratory of New Technology in Rice Breeding, Guangzhou, China,*
[3]*Guangdong Rice Engineering Laboratory, Guangzhou, China,*
[4]*Key Laboratory of Genetics and Breeding of High Quality Rice in Southern China, Ministry of Agriculture and Rural Affairs Beijing, China,*
[5]*Novogene Biotechnology Inc, Beijing, China,*
[6]*Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing, China, and*
[7]*State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, South China Agricultural University, Guangzhou 510642, China*

Received 1 September 2022; revised 14 March 2023; accepted 24 March 2023; published online 30 March 2023.
*For correspondence (e-mail fcysd2901@126.com; guxiaofeng@caas.cn; whyang@scau.edu.cn; fwang1631@163.com).
[†]These authors contributed equally to this work

SUMMARY

Heterosis has been widely used in multiple crops. However, the molecular mechanism and prediction of heterosis remains elusive. We generated five F_1 hybrids [four showing better-parent heterosis (BPH) and one showing mid-parent heterosis], and performed the transcriptomic and methylomic analyses to identify the candidate genes for BPH and explore the molecular mechanism of heterosis and the potential predictors for heterosis. Transcriptomic results showed that most of the differentially expressed genes shared in the four better-parent hybrids were significantly enriched into the terms of molecular function, and the additive and dominant effects played crucial roles for BPH. DNA methylation level, especially in CG context, significantly and positively correlated with grain yield per plant. The ratios of differentially methylated regions in CG context in exons to transcription start sites between the parents exhibited significantly negative correlation with the heterosis levels of their hybrids, as was further confirmed in 24 pairwise comparisons of other rice lines, implying that this ratio could be a feasible predictor for heterosis level, and this ratio of less than 5 between parents in early growth stages might be a critical index for judging that their F_1 hybrids would show BPH. Additionally, we identified some important genes showing differential expression and methylation, such as *OsDCL2*, *Pi5*, *DTH2*, *DTH8*, *Hd1* and *GLW7* in the four better-parent hybrids as the candidate genes for BPH. Our findings helped shed more light on the molecular mechanism and heterosis prediction.

Keywords: ***Indica*** **rice, heterosis, prediction, transcriptome, methylome.**

INTRODUCTION

Heterosis or hybrid vigor refers to the phenomenon that the F_1 hybrid has better performance than its parents in terms of biomass, yield, rate of growth and stress tolerance. Since its initial discovery by Darwin in the early 19th century, hybrid breeding has been widely used in agricultural production in multiple crops, such as maize, rice, sorghum, oilseed rape and tomato, etc. (Gai et al., 2016). Since the 1970s, first three-line hybrid rice, then two-line hybrid rice, have been developed and widely used in China first and then spread to other Southeast Asian countries and the USA, which dramatically boosted rice production (Cheng et al., 2007).

Despite the fact that heterosis has been successfully applied in crop breeding and agricultural production for more than 100 years, the molecular mechanisms of heterosis still remain a mystery, despite three major hypotheses: dominance (Davenport, 1908), overdominance (East, 1908) and epistasis (Powers, 1944) being put forward to explain heterosis. However, these genetic models are largely

conceptual and cannot sufficiently explain the molecular basis of heterosis (Birchler et al., 2010; He et al., 2013; Miller et al., 2015). Interestingly, several earlier studies using the same elite rice hybrid population produced by crossing Zhenshan97B and Minghui63 documented evidence for different genetic models of heterosis including partial dominance and overdominance (Li et al., 2008), epistasis (Yu et al., 1997), and overdominance and pseudo-overdominance (Zhou et al., 2012), suggesting that different genetic models may all contribute to heterosis. Recently, genome-wide association studies and other omics approaches have been utilized to explore the molecular basis of heterosis (Huang et al., 2015; Yang et al., 2017; Zhen et al., 2017; Zhu et al., 2016). For example, Huang et al. (2015) found that the overall heterozygosity did not play a significant role in heterosis, but instead heterozygosity at some specific loci was important to heterosis in rice. Recent studies also suggested roles of small RNAs and epigenetic regulation in heterosis (Chen, 2013; Kawanabe et al., 2016; Lauss et al., 2018; Li et al., 2014; Shen et al., 2012; Zhang et al., 2014; Zhang et al., 2015). Despite the progress, overall elucidating the molecular mechanisms of heterosis still remains a daunting task.

Screening of commercial valued F_1 hybrids still requires a lot of manual labor to carry out the test-crossing trials. To reduce the workload for crossing and evaluating F_1 hybrids, breeders and geneticists have been trying to find ways to predict the magnitude of heterosis between different parental lines. Laborda et al. (2005) proposed that the parental lines were clustered into different heterotic groups based on genetic diversity evaluated by molecular markers to guide the breeding of productive hybrids. Currently, the genetic distance between parents has been proven to be not completely correlated with the heterosis level of their corresponding F_1 hybrids in maize, rice, rapeseed, melon and *Arabidopsis* (Dafna et al., 2021; Fernandes et al., 2015; Silva et al., 2019; Yang et al., 2017; Younas et al., 2012; Wang et al., 2015). Recently, several studies on prediction of heterosis have been reported in different aspects, including natural variation of stress-responsive gene expression (Miller et al., 2015), non-linear phenotypic variation (Vasseur et al., 2019) and metabolomic pathway (Dan et al., 2021). Additionally, methylation information showed a predictive relationship with complex quantitative traits, such as plant height (PH; Hu et al., 2015) and human longevity (Horvath & Raj, 2018). However, most of these relationships of prediction were based on quantitative genetics and regression analysis, and a convenient and effective method to predict the heterosis level has still not been developed.

In the previous studies, generally, only one or two F_1 hybrids and their parental lines were used to survey the genetic mechanism, and mid-parent heterosis (MPH) and better-parent heterosis (BPH) could not be effectively distinguished. Actually, BPH is more valuable than MPH for breeding commercially valued hybrids in crops and vegetables. In this study, we selected two elite rice restorer lines: Guanghui 998 (R998) and Guanghui 308 (R308), and two maintainer lines: Wufeng B (WFB) and Rongfeng B (RFB), representing the excellent hybrid rice parental lines in Southern China (their hybrid combinations have a cumulative planting area of more than 22 million hectares to the end of 2020) to make five hybrids. Four of the hybrids showed BPH, whereas one hybrid derived from the two maintainer lines showed MPH. So, we could investigate the change of transcriptome and methylome in these hybrids showing different heterosis levels relative to their parents, and reveal the molecular mechanism and explore the potential predictor of BPH.

RESULTS

Evaluation of heterosis for several yield-related traits

In order to facilitate the investigation of yield-related traits, four rice inbred paternal lines including two restorer lines R998 and R308 and two maintainer lines WFB and RFB were used to cross with each other to generate the following five F_1 hybrids: WY998 (WFB/R998); WY308 (WFB/R308); RY998 (RFB/R998); RY308 (RFB/R308); and WR (WFB/RFB). We evaluated the heterosis levels of five F_1 hybrids for panicle number (PN), PH, grain number per panicle (GNPP), kilo-grain weight (KGW) and grain yield per plant (GYPP), respectively (Table S1). We observed that the four hybrids WY998, WY308, RY998 and RY308 showed higher PH (0.6–9% for BPH), greater GNPP (1.1–24.7% for BPH), larger KGW (2.3–7.1% for BPH) and greater GYPP (9.6–28.6% for BPH), while PN among the five F_1 hybrids and their parents ranged only from 7.03 to 8.53, thus is not suitable for evaluating the heterosis level. The performance of the hybrid WR was similar to the mid-parent values for the above four traits. Overall, for PH, GNPP, KGW and GYPP, the four commercially valued hybrids WY998, WY308, RY998 and RY308 showed BPH, whereas the hybrid WR derived from two maintainer lines WFB and RFB showed only MPH. Hence, understanding the differences of transcriptomic and methylomic profiles between the better-parent hybrids and the mid-parent hybrid could help to identify the candidate genes for BPH.

Identification of candidate genes for BPH based on transcriptomic analysis

We performed global transcriptomic sequencing of young leaves (20 days after transplanting), young panicles (10 days before heading) and filling panicles (15 days after anthesis) with three replicates for five hybrids and their four parents. Approximately 81.3–122 million mRNA-seq reads were obtained from the libraries of each genotype, and over 85% reads were mapped to the reference genome

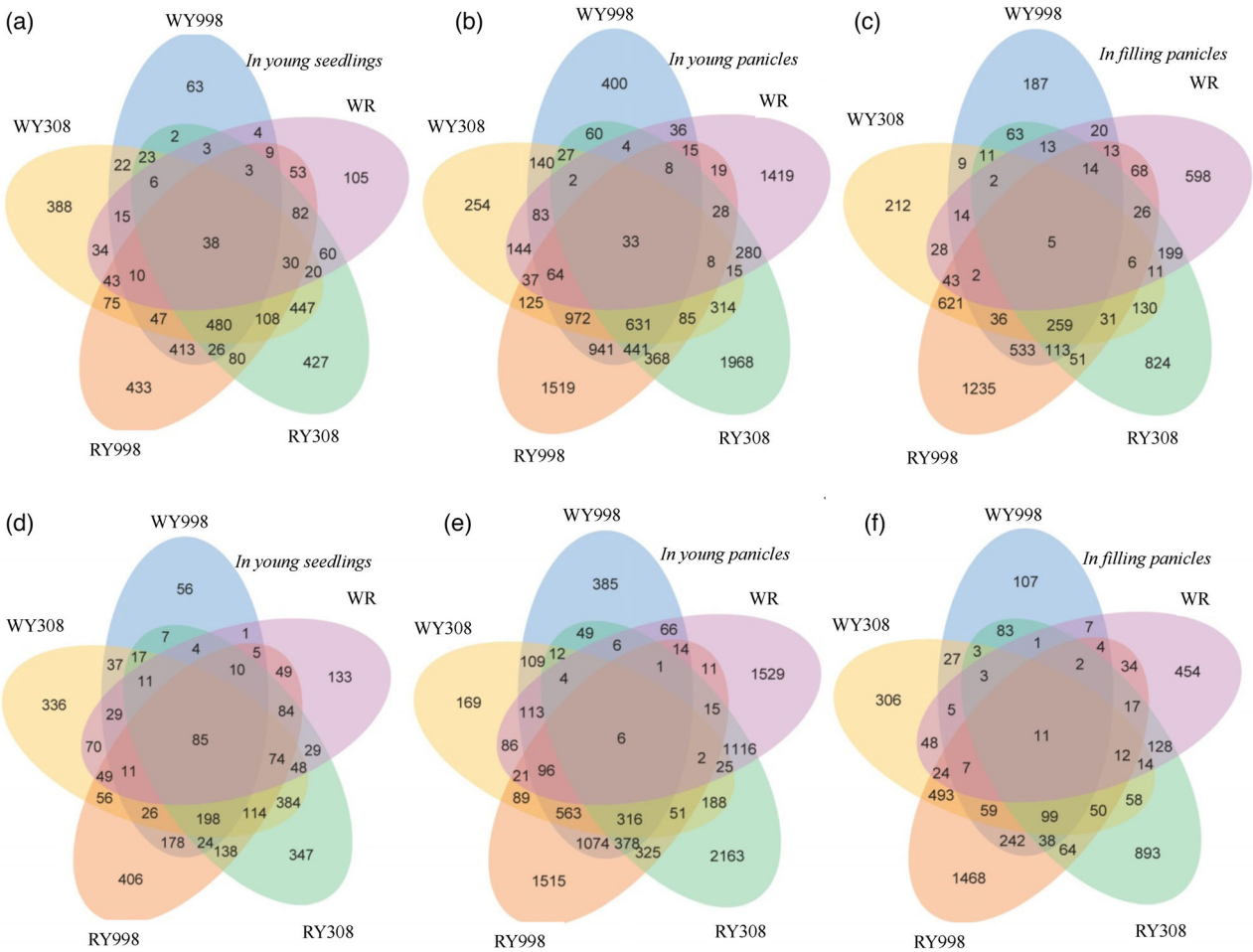

Figure 1. Venn diagrams showing overlapped differentially expressed genes (DEGs) in five hybrids.
(a–c) Venn diagrams showing upregulated DEGs that are overlapped among the young seedlings, young panicles and filling panicles of five hybrids.
(d–f) Venn diagrams showing downregulated DEGs that are overlapped among young seedlings, young panicles and filling panicles of five hybrids.

(http://mbkbase.org/R498/; Table S2). The mRNA-seq data among three replicates showed high correlations ($R^2 >$ 0.869; Figure S1).

We investigated the differentially expressed genes (DEGs) in young seedlings, young panicles and filling panicles of five F_1 hybrids relative to their parents, and detected a large number of DEGs [false discovery rate (FDR) < 0.05] in three types of tested tissues (Table S3). To screen for the potential DEGs associated with BPH, we focused on the DEGs overlapped only among four better-parent hybrids as potential candidate genes for BPH, and identified 480, 631 and 259 upregulated DEGs (Figure 1a–c) and 198, 316 and 99 downregulated DEGs in young seedlings, young panicles and filling panicles, respectively (Figure 1d–f; Table S4). We noticed that about 40% overlapped upregulated genes were not expressed in one of both parents of F_1 hybrids in young seedlings, indicating that the dominance effect might play important roles in heterosis in early growth stage. Surprisingly, gene ontology (GO) enrichment analysis indicated that most of the overlapped DEGs were grouped into only the terms of molecular function (FDR < 0.05) in three types of tested tissues (Figure 2a–c), and only a few upregulated DEGs were significantly clustered into the terms of biological processes related to metabolic process, phosphorylation and amino acid biosynthesis in young seedlings and panicles (Figure 2a,b). However, several previously reported blast resistance loci, including *OsDCL2* (Zhang et al., 2015), *Pi33* (Berruyer et al., 2003), *Pid2* (Chen et al., 2006) and *Pi5* (Lee et al., 2009), were detected to be enriched in terms of macromolecule metabolic process and carbohydrate derivative binding. Expression pattern analysis indicated that in the four better-parent hybrids, *OsDCL2*, *Pi33* and *Pi5* exhibited the additive expression patterns, and *Pid2* was close to that of the male parents R998 and R308 and showed dominant expression patterns (Figure 2d). These results implied that the resistance loci might be important heterotic genes.

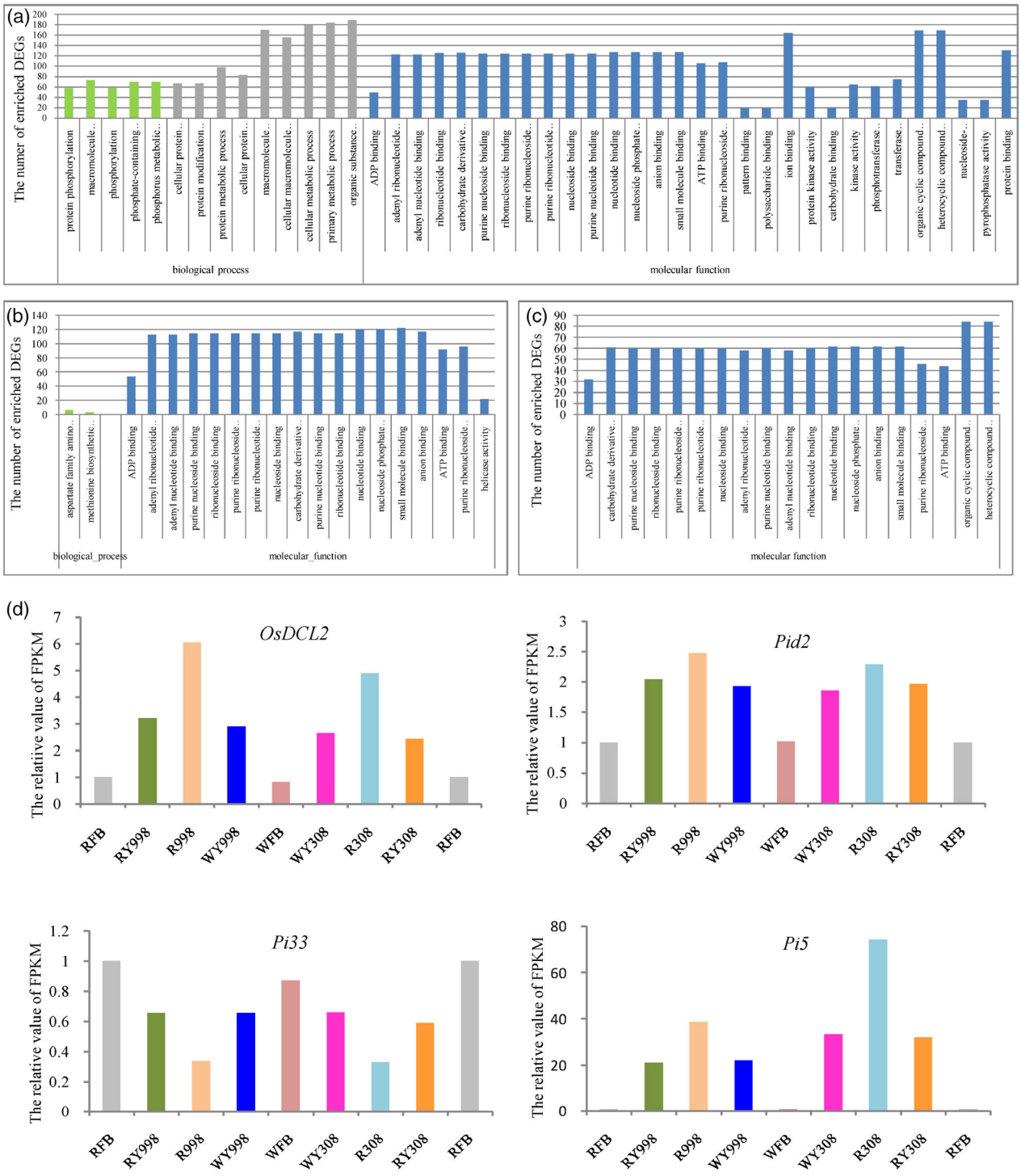

Figure 2. Function enrichment and expression patterns of the differentially expressed genes (DEGs) only overlapped among the four better-parent hybrids. (a–c) Gene ontology (GO) enrichment of overlapped DEGs among young seedlings, young panicles and filling panicles of the four better-parent hybrids. (d) The expression patterns of four overlapped DEGs among young seedlings of the four better-parent hybrids.

Analysis of the genetic models and candidate genes for heterosis based on allele-specific expression (ASE)

The ASE between the expression levels of two parental alleles in the hybrid was considered as a mechanism of heterosis (Shao et al., 2019). To understand the potential roles of ASE for heterosis, we performed the analysis of ASE on the differential expression between parents. The ASE genes (ASEGs) in F_1 hybrids were classified into three categories based on their expression patterns: (i) additive ASEGs: expression level close to the mid-parent value; (ii) dominant ASEGs: expression level significantly different from one parent and similar to another one. When expression level is similar to the male parent, it is called male-dominant gene, and if similar to female parent called female-dominant gene; and (iii) overdominant ASEGs: expression level above the better parent or below the lower parent.

A large number of single-nucleotide polymorphisms (SNPs) were detected between the four parents using the genome sequence of R498 (http://mbkbase.org/R498/) as the reference sequence for variant detection with the software GATK3 (Figure S2). We randomly selected 49 SNPs in R308 relative to the reference genome sequences and designed the primers to verify these variants (Table S5). Among the 49 pairs of primers designed, 45 pairs showed successful specific amplification. Their polymerase chain reaction (PCR) products were sequenced and the detected variants were fully consistent with the information of RNA-seq, suggesting that the detected SNPs between the parents can be used to analyze ASE in the F_1 hybrids.

A number of ASEGs were identified in F_1 hybrids (Table S6). We investigated the percentage of every type of ASEGs and found that, for four better-parent hybrids, about 46.8%, 43.4% and 31.3% were additive ASEGs in young seedlings, young panicles and filling panicles, respectively, about 49.4%, 52.9%, and 62.9% were dominant ASEGs, and only 3.8%, 3.8% and 5.9% were overdominant-effect genes in young seedlings, young panicles and filling panicles, respectively (Figure 3a–c), while for the mid-parent hybrid WR, most of the ASEGs (58.1%, 77.7% and 78.0%) showed a dominant effect in young seedlings, young panicles and filling panicles, respectively. Also, about 26.4%, 1.3% and 14.4% were overdominant ASEGs in young seedlings, young panicles and filling panicles, respectively (Figure 3d–f). The proportions of additive ASEGs (46.8%, 43.4%, 31.3%) in the better-parent hybrids were far more than that (15.5%, 21.0%, 7.6%) in the mid-parent hybrid WR in young seedlings, young panicles and filling panicles, respectively, while the total proportions of dominant and overdominant ASEGs in the better-parent hybrids (53.2%, 56.6%, 68.7%) were less than that in the mid-parent hybrid WR (84.5%, 79.0%, 92.4%) in three tested tissues (Table S6), indicating that the additive and dominant ASEGs play important roles for heterosis, which is consistent with the previous study (Guo et al., 2006).

Among the dominant ASEGs shared in the four better-parent hybrids, 94, 38 and 93 overlapped ASEGs showing male-dominance (Figure 3g) and 70, 36 and 3 shared ASEGs showing female-dominance (Figure 3h) were identified in three tested tissues, respectively (Table S7). Also, the male-dominant ASEGs shared only in the four better-parent hybrids were more than the female-dominant ASEGs, especially in young seedlings and filling panicles, indicating that the male parents (the restorer lines) intended to contribute more than female parents (the sterile lines) to heterosis. And several important genes such as *OsMADS51*, *Pid2*, *Hd3a*, *Ehd1* and *OsGME1* were detected in the overlapped male-dominant genes of young seedlings and should be very important for the formation of BPH (Table S7). This finding is consistent with previous studies (Huang et al., 2015; Wei et al., 2021).

DNA methylation of F_1 hybrids and their parents

To explore the role of DNA methylation in heterosis, especially BPH, we investigated the methylation profiles in young seedlings, young panicles and filling panicles of five hybrids and their parents (Table S8). It was observed that the four better-parent hybrids showed increased total DNA methylation and CG and CHG methylation in young seedlings compared with their parents, while the methylation level of the mid-parent F_1 hybrid WR was similar to the mean value (MV) of its two parents in young seedlings (Figure 4a,b), suggesting that the methylation level in hybrids could be associated with their heterosis level. In addition, we also observed that the methylation level of all hybrids was increased in the panicles compared with their parents (Figure S3). Hence, we investigated the correlation between the methylation level and their GYPP in nine genotypes (five hybrids and four parents). The positive correlation was observed in all cytosine contexts, especially in CG context in young seedlings, and the correlation coefficients (R^2) were 0.645 for mC, 0.919 for mCpG, 0.739 for mCHG and 0.181 for mCHH, respectively (Figure 4c–f). However, the correlation between the methylation level in panicles and GYPP appeared to be not significant or weak (Figure S4), suggesting that the methylation status in early growth stage might play more important roles for the formation of heterosis.

We further investigated the methylation status in different genomic functional regions in all cytosine contexts. We found different patterns of methylation exhibited in the various genomic functional regions among the hybrids and their parents (Figure 4g). In the regions of gene body, especially exons, CG methylation levels differed significantly among these lines, but the CHG and CHH methylation levels showed no significant difference. In the

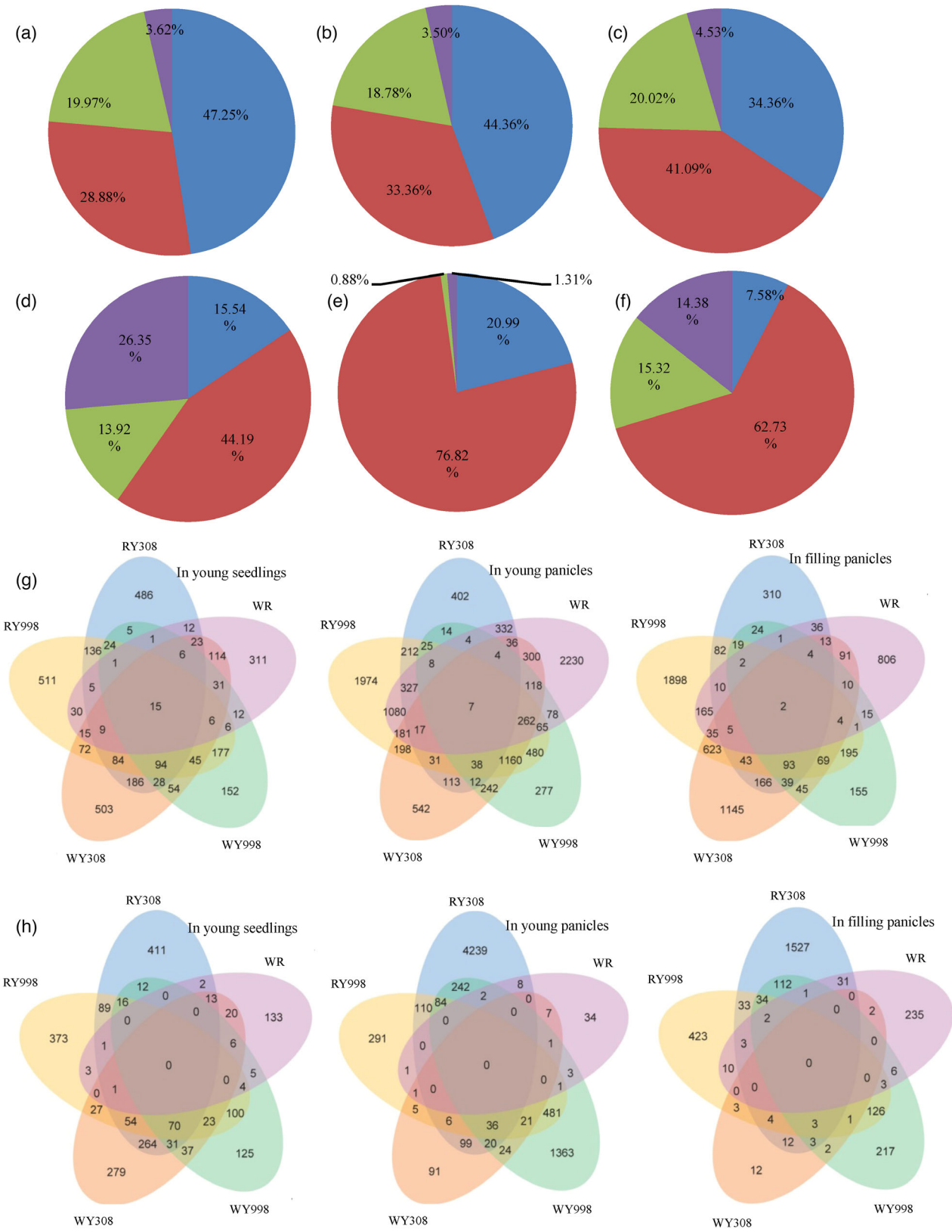

Figure 3. The analysis of allele-specific expression genes (ASEGs) in five hybrids.
(a–c) Average percentage of the additive ASEGs (blue), male-dominant ASEGs (red), female-dominant ASEGs (green) and overdominant ASEGs (purple) in young seedlings, young panicles and filling panicles of the four better-parent hybrids.
(d–f) Percentage of the additive ASEGs (blue), male-dominant ASEGs (red), female-dominant ASEGs (green) and overdominant ASEGs (purple) in young seedlings, young panicles and filling panicles of the mid-parent heterosis (MPH) hybrid WR.
(g) Venn diagrams showing male-dominant ASEGs among young seedlings, young panicles and filling panicles of the five hybrids.
(h) Venn diagrams showing female-dominant ASEGs among young seedlings, young panicles and filling panicles of the five hybrids.

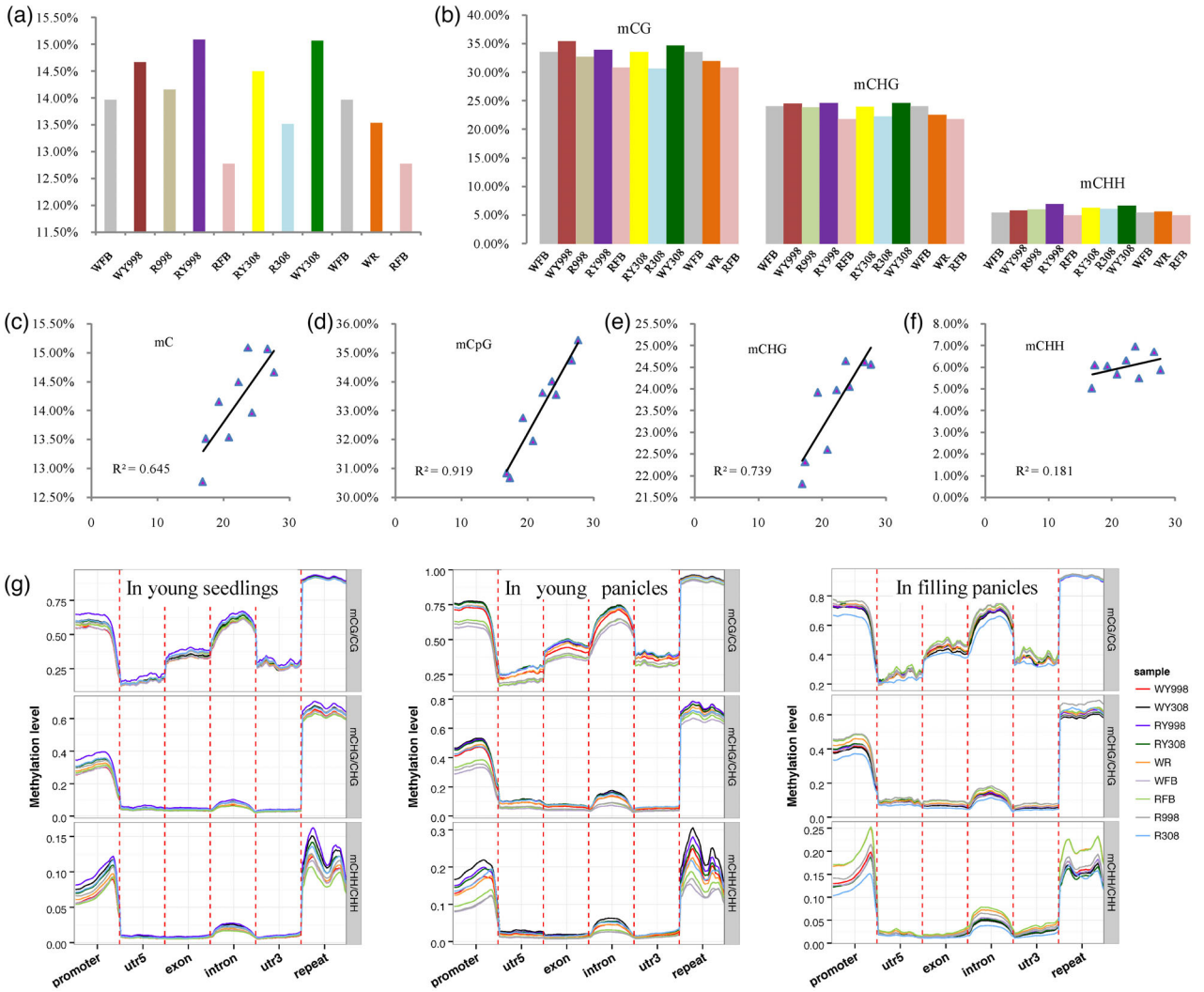

Figure 4. The characteristics of methylation, and the correlation between methylation level and grain yield per plant (GYPP).
(a) The total level of methylation in young seedlings of the five hybrids and their parents.
(b) The level of methylation in three cytosine contexts in young seedlings of the five hybrids and their parents.
(c–f) Analysis of the correlation between GYPP of the five hybrids and their parents, and the level of methylation including total mC, mCG, mCHG and mCHH in their young seedlings.
(g) The level of methylation in different genomic regions of the five hybrids and their parents.

repeated regions including transposable elements (TEs), non-CG methylation level displayed significant differences among these lines, while the level of CG methylation in TEs was almost identical. In the promoter and intron regions, these lines had similar patterns of overall methylation levels among the three cytosine contexts. Thus, the patterns of CG methylation in gene bodies, especially in exons and non-CG methylation in TEs could play crucial roles in heterosis.

Association analysis between transcriptomic and methylomic profiles

Because the levels of CG and CHG methylation were significantly correlated with GYPP, we further analyzed the

Table 1 The association analysis between differential gene expression and methylation

Tissues	Combinations	CG		CHG		CHH	
		Hyper-methylation	Hypo-methylation	Hyper-methylation	Hypo-methylation	Hyper-methylation	Hypo-methylation
Young seedlings	WY998_vs_WFB	−0.06	−0.24	−0.32	−0.36	−0.42	−0.31
	WY998_vs_R998	N	−0.18	−0.29	−0.51	−0.37	−0.3
	WY308_vs_WFB	N	−0.23	−0.38	−0.37	−0.42	−0.35
	WY308_vs_R308	N	−0.23	−0.3	−0.51	−0.4	−0.35
	RY998_vs_RFB	−0.02	−0.27	−0.41	−0.29	−0.26	N
	RY998_vs_R998	N	−0.2	−0.31	−0.46	−0.28	−0.35
	RY308_vs_RFB	N	−0.18	−0.38	−0.3	−0.34	−0.27
	RY308_vs_R308	N	−0.21	−0.29	−0.47	−0.34	−0.3
	WR_vs_RFB	−0.26	N	−0.41	−0.53	−0.37	−0.24
	WR_vs_WFB	−0.15	N	−0.39	−0.49	−0.42	−0.32
Young panicles	WY998_vs_WFB	−0.12	−0.34	−0.37	−0.34	−0.22	N
	WY998_vs_R998	−0.1	−0.28	−0.3	−0.52	−0.37	N
	WY308_vs_WFB	−0.15	−0.33	−0.37	−0.38	−0.18	N
	WY308_vs_R308	−0.12	−0.39	−0.29	−0.41	−0.05	N
	RY998_vs_RFB	−0.11	−0.35	−0.37	−0.32	−0.28	−0.34
	RY998_vs_R998	−0.17	−0.3	−0.31	−0.5	−0.27	N
	RY308_vs_RFB	−0.09	−0.33	−0.34	−0.34	−0.35	−0.19
	RY308_vs_R308	−0.14	−0.38	−0.24	−0.47	−0.11	−0.42
	WR_vs_RFB	−0.08	−0.13	−0.38	−0.5	−0.4	−0.35
	WR_vs_WFB	−0.17	−0.1	−0.37	−0.27	−0.42	−0.21
Filling panicles	WY998_vs_WFB	−0.07	−0.27	−0.36	−0.33	−0.2	−0.37
	WY998_vs_R998	−0.11	−0.26	−0.31	−0.42	−0.23	−0.32
	WY308_vs_WFB	−0.12	−0.3	−0.35	−0.34	−0.18	−0.32
	WY308_vs_R308	−0.15	−0.35	−0.28	−0.42	−0.3	−0.18
	RY998_vs_RFB	−0.11	−0.31	−0.3	−0.38	−0.29	−0.15
	RY998_vs_R998	−0.12	−0.34	−0.35	−0.5	−0.32	−0.28
	RY308_vs_RFB	−0.11	−0.3	−0.27	−0.37	−0.27	−0.19
	RY308_vs_R308	−0.08	−0.39	−0.18	−0.4	−0.27	−0.28
	WR_vs_RFB	−0.18	−0.36	−0.31	−0.41	−0.19	−0.35
	WR_vs_WFB	−0.12	−0.31	−0.42	−0.46	−0.09	−0.55

correlation between differential expression and CG and CHG methylation. The results showed that the differential expression was significantly and negatively correlated with CHG methylation and its correlation coefficient was larger than that with CG methylation, but the coefficients of CHG methylation were nearly identical among the five hybrids (Table 1), which suggested that CHG methylation might have larger effects than CG methylation on the differential expression of genes in the F_1 hybrids, but may not be associated with BPH. CG methylation, especially CG hypo-methylation, was significantly associated with differential expression of genes in better-parent F_1 hybrids, while the corresponding correlations were not significant or very weak (0.13 and 0.1) in the mid-parent hybrid WR. Additionally, the correlation coefficients of hyper-methylation were much smaller than those of hypo-methylation in four better-parent hybrids, which may, at least partially, explain why the upregulated genes were more than the downregulated genes in better-parent hybrids (Table S3).

Thereafter, we surveyed the shared genes between DEGs and differentially methylated genes (DMGs) in corresponding pairwise combinations as the candidate genes for heterosis, and observed that about one-third DEGs were directly associated with the differential methylation between the four better-parent hybrids and their parents (Table S9), which is consistent with previous studies (Eichten et al., 2013; Li et al., 2015; Schmitz et al., 2013). We thus focused on the genes shared between DMGs and DEGs in young seedlings of the four better-parent hybrids relative to their parents, and identified some yield-related or heading date genes, such as *DTH2* (Wu et al., 2013), *hwh1* (Jiang et al., 2008), *GLW7/OsSPL13* (Si et al., 2016; Wei et al., 2021), *Hd5* (Fujino et al., 2013) and *Hd1* (Zhang et al., 2012; Figure 5), which could be key candidate genes for the formation of heterosis. Notably, *DHT2* was detected in all the pairwise combinations between the four better-parent hybrids and their maternal line WFB or RFB. Noteworthy, its gene body and promoter were differently methylated in five hybrids and their parents, and the deletion of genomic DNA sequence was detected in the promoter region of WFB and RFB (Figures 6 and S5; Table S10), suggesting that the difference of methylation might result from structural variation in DNA sequences.

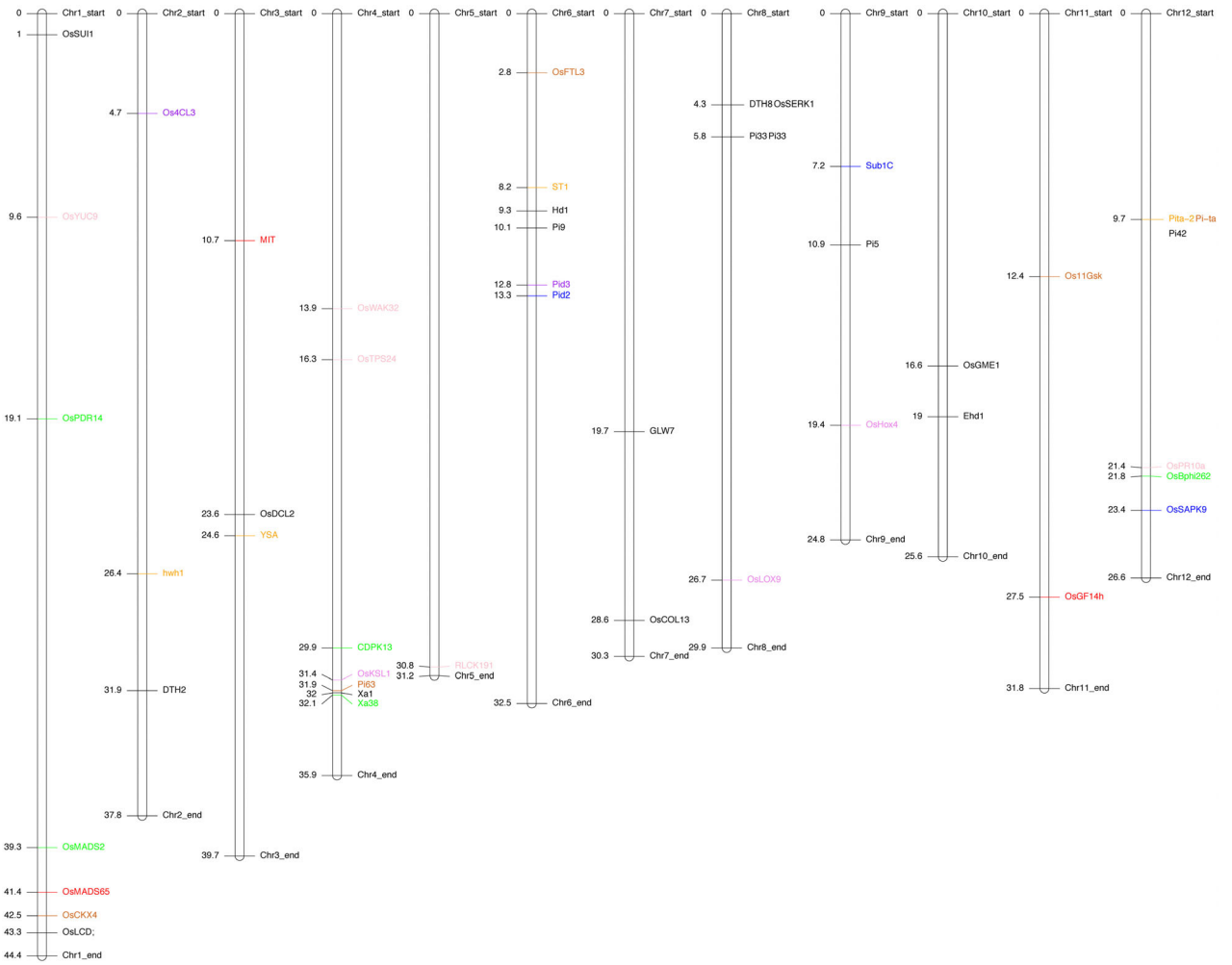

Figure 5. The genomic locations of overlapped genes between differentially expressed genes (DEGs) and differentially methylated genes (DMGs) in pairwise combinations between the four better-parent hybrids and their parents in young seedlings are illustrated along the 12 rice chromosomes. WY998_vs_R998 in green; WY998_vs_WFB in blue; WY308_vs_R308 in pink; WY308_vs_WFB in red; RY998_vs_R998 in violet; RY998_vs_RFB in purple; RY308_vs_R308 in orange; RY308_vs_RFB in chocolate. The genes shared among the above combinations are marked in black.

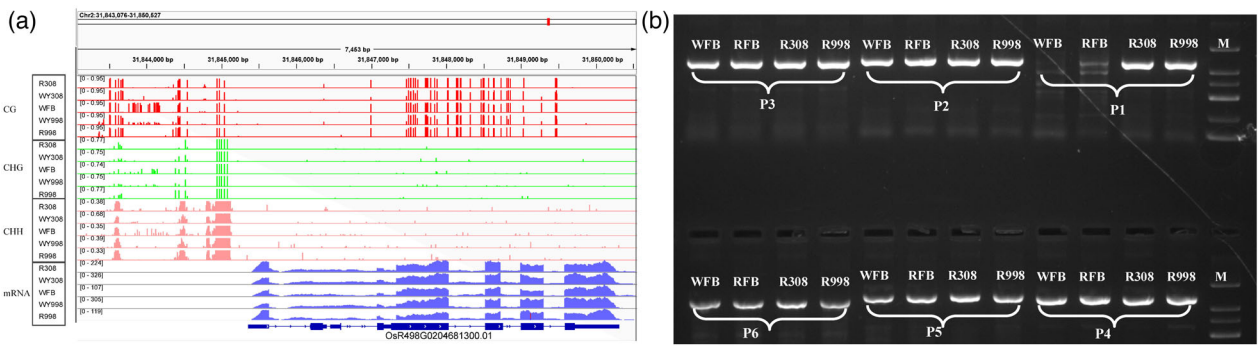

Figure 6. The genomic characteristics of *DTH2*.
(a) DNA methylation profiles at three cytosine contexts and correlated mRNA profiles in *DTH2* including the promoter and gene body regions in young seedlings of the WY308 and WY998 hybrids and their parents.
(b) The amplification of *DTH2* using six pairs of the specific primers covering the gene body and its 2-kb upstream region in R998, R308, WFB and RFB.

Analysis of differentially methylated regions (DMRs) and heterosis predictor

We called the DMRs and annotated the related genes (non_TE genes) and TEs in the identified DMRs between the four parents and between the parents with their five F_1 hybrids using a genome-wide sliding window approach (window size of 200 bp, step size of 50 bp). A large number of DMRs were identified in different pairwise comparisons, and DMRs between the parents were much more than those between F_1 hybrids and their parents (Table S11), suggesting that not all methylation difference is related to heterosis. Then, we focused on the characteristics of DMRs between the parental lines in different cytosine contexts to explore potential clues for predicting the heterosis level of their F_1 hybrids. For the four better-parent hybrids, CG DMRs were more than CHH DMRs between their corresponding parents (R998_vs_WFB, R308_vs_WFB, R998_vs_RFB and R308_vs_RFB) in young seedlings and young panicles, while for the mid-parent F_1 hybrid WR, CG DMRs between its parents were less than those in CHH context in all three tested tissues. Thus, the characteristics of DMRs in different cytosine contexts between the parental lines in the early growth stages could help purposefully screen for parents to develop novel hybrids with BPH in crop breeding.

Because the overall levels of CG gene body methylation (gbM) of the five F_1 hybrids and their four parental lines were highly consistent with the performance of yield-related traits, we further surveyed the distribution of DMRs in pairwise comparisons, especially between the parental lines in the different functional regions including exons, introns and transcription start sites (TSSs; Table S12). We observed that the CG DMRs in exons, introns and TSSs between the four better-parent hybrids and their corresponding parents were more than those between the mid-parent hybrid WR and its parents, especially in young seedlings. Interestingly, we noticed that the DMRs in exons between the parents of the four better-parent hybrids were significantly less than those between the parents of the mid-parent hybrid WR, while the DMRs in TSSs between the parents of the four better-parent hybrids were far more than those between the parents of the mid-parent hybrid WR in all three tested tissues. No significant difference was observed in the number of DMRs in introns among the parents of the five hybrids. Thus, the distribution characteristics of CG DMRs in exons and TSSs between different rice parental lines seem to be correlated with the heterosis levels of their hybrids. The further correlation analysis indicated that the number of CG DMRs in exons between parents was significantly and negatively ($R^2 = 0.960$) correlated with the heterosis level of their hybrids, while the number of CG DMRs in TSSs between parents was positively ($R^2 = 0.623$) correlated with the heterosis level of their hybrids (Figure 7a,b).

Because CG DMRs in exons and TSSs between the parents exhibited different correlation with the heterosis level of their F_1 hybrids, we calculated the ratio of the number of CG DMRs in exons to that in TSSs in every pairwise combination to evaluate the correlation (Table S12). We found that the ratios of the number of CG DMRs in exons to that in TSSs between the parents of the four better-parent hybrids were less than 5 (ranging from 3.18 to 4.73 among all tested tissues), and those between the parents WFB and RFB of the mid-parent hybrid WR were more than 8 (ranging from 8.12 to 16.34 among three tested tissues). Especially in young seedlings, the ratio (16.34) between WFB and RFB was far larger than those (from 4.03 to 4.73, less than 5) between the parents of the four better-parent hybrids. The correlation analysis showed that there was significantly negative correlation between the ratios of the number of CG DMRs in exons to that in TSSs between parents and the magnitude of heterosis of their hybrids, and the correlation coefficients were up to 0.982, 0.962 and 0.991 in young seedlings, young panicles and filling panicles, respectively (Figure 7c–e), which implied that the ratio of the number of CG DMRs in exons to that in TSSs between the parental lines might be used as a feasible predictor of heterosis level of their F_1 hybrids.

To validate the above speculation, we selected 19 rice lines including 17 parental lines of hybrid rice widely used in commercial production in China (Table S13) and performed methylomic sequencing for the 20 days after transplanting (DATs) young seedlings. In 25 pairwise combinations including *Japonica_vs_Japonica*, *Japonica_vs_Indica* and *Indica_vs_Indica*, we noticed that the ratios of the number of CG DMRs in exons to that in TSSs between parents were less than 3, less than 5, more than 5 and up to 19.34 in *Japonica_vs_Indica, Indica* maintainer lines_vs_restorer lines (their corresponding F_1 hybrids have been widely used in rice production), *Indica* maintainer lines_vs_maintainer lines and *Japonica_vs_Japonica*, respectively (Figure 7f; Table S14). The trend of the ratios among different types of rice lines was quite opposite to the magnitude of heterosis reported as *indica-japonica* F_1 hybrids > *indica-indica* F_1 hybrids > *japonica-japonica* F_1 hybrids (Shen et al., 2015). Thus, it can be seen that the negative correlation between the ratios of the number of CG DMRs in exons to that in TSSs and the heterosis level has been further confirmed. Based on our findings, we speculate that this ratio could be viewed as an intuitive indicator to predict the heterosis level in hybrid rice breeding, and the ratio of less than 5 could be the critical index for judging F_1 hybrids of BPH in the early stages of growth and development.

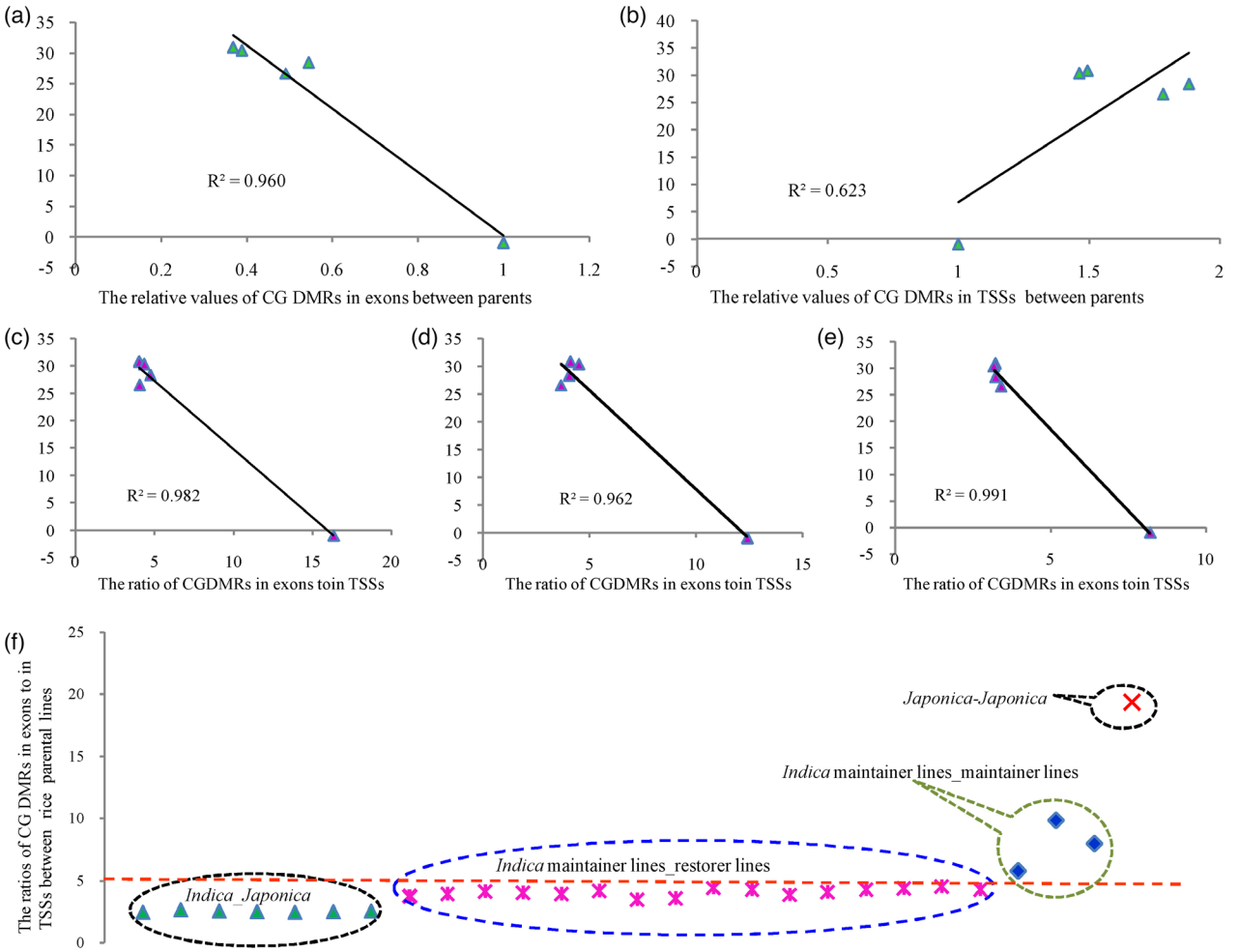

Figure 7. The characteristics of CG differentially methylated regions (DMRs) in exons and transcription start sites (TSSs) between the rice parental lines.
(a) Analysis of the correlation between the number of CG DMRs in exons between four rice parents R998, R308, WFB and RFB and the heterosis level of their F_1 hybrids.
(b) Analysis of the correlation between the number of CG DMRs in TSSs between four rice parents R998, R308, WFB and RFB and the heterosis level of their F_1 hybrids.
(c–e) Analysis of the correlation between the ratios of CG DMRs in exons to TSSs between four rice parents R998, R308, WFB and RFB and the heterosis level of their F_1 hybrids in young seedling, young panicles and filling panicles.
(f) Scatter plot of the ratios of CG DMRs in exons to TSSs in 25 pairwise combinations between 19 different types of rice parental lines.

DISCUSSION

Candidate genes for BPH

In this study, transcriptomic analyses of four better-parent hybrids and one mid-parent hybrid were used to screen for DEGs only shared among the four better-parent hybrids as the candidate genes for BPH. Hundreds of shared genes were identified, but most of them were enriched into the terms of molecular function, and only a small subset of upregulated DEGs was enriched into the metabolic and phosphorylation processes. These results further verified the inference that most of the loci contributing to heterosis have minor effects on important agronomic traits (Schnable & Springer, 2013).

The levels of methylation, especially in CG and CHG contexts, were significantly positively correlated with GYPP. The association analysis between transcriptomic and methylomic profiles showed the significant correlation. About one-third of DEGs in four better-parent hybrids also displayed the differential methylation, implying that the differential expression of these genes should be directly caused by methylation. Hence, we focused on the genes overlapped between DEGs and DMGs in four better-parent hybrids as the key candidate genes for BPH. In these shared genes, some important genes regulating disease resistance, heading date and yield-related traits, such as *OsDCL2*, *Pi33*, *Pi5*, *DTH2*, *Hd1*, *Ehd1*, *DTH8*, *OsGME1*and *GLW7*, were identified (Figure 5 marked in black), and

these genes were previously reported to affect yield-related traits or be related to heterosis (Huang et al., 2015; Si et al., 2016; Urayama et al., 2010; Wei et al., 2010; Wei et al., 2021). The susceptibility loci of resistance genes may have important biological function, and their mutations generally cause undesired pleiotropic effects on plant growth, development and crop yields (Li et al., 2022). The expression of stress-responsive genes was induced against pathogen invasions under stress conditions and to promote biomass heterosis under non-stress conditions by epigenetic mechanisms and circadian rhythm (Miller et al., 2015; Yang et al., 2021). Additionally, *DTH2* was also detected in the mid-parents hybrid and its sequence variants were also verified in four parents. *DTH2* was reported to correlate with early heading and increase reproductive fitness under long-day conditions (Wu et al., 2013). It was suggested that *DTH2* could be a superior heterotic gene in South-China double cropping rice regions.

Genomic DNA methylation and the molecular mechanism of heterosis

Although heterosis has been successfully used in crop breeding and agricultural production for several decades, the molecular mechanism remains elusive. Growing evidence indicates that genomic DNA methylation has a great impact on heterosis (Groszmann et al., 2013; He et al., 2013; Lauss et al., 2018; Shen et al., 2012). Our work revealed that the overall methylation level, especially in CG and CHG contexts, in five hybrids and their four parents significantly and positively correlated with their GYPP.

Compared with their parents, the F_1 hybrids exhibited differential CG methylation in gene body and non-CG methylation in TEs, suggesting that the levels of DNA methylation, especially gene body CG methylation and TE non-CG methylation, could be critical for heterosis. It has been shown that DNA methylation on genes and TEs in different cytosine contexts has different regulatory effects in plants (Zhang et al., 2018). gbM is largely conserved across plant species and predominantly enriched into constitutively expressed house-keeping genes (To et al., 2015; Bewick & Schmitz, 2017), indicating that body-methylated genes are functionally important. Our association analysis between methylome and transcriptome showed that CG gbM, especially CG hypo-gbM and CHG gbM, significantly and negatively correlated with the differential expression of genes in F_1 hybrids. The coefficients of CHG gbM were bigger than those of CG gbM, which seems to indicate that CHG methylation plays important roles for heterosis (Ma et al., 2021). However, in our study, the correlation coefficients of CG gbM in the five F_1 hybrids are consistent with the heterosis levels of five F_1 hybrids, while the correlation coefficients of CHG gbM are nearly identical in the five F_1 hybrids. These results suggest that the CHG gbM might play more roles for the differential expression than CG gbM in hybrids, but CG gbM was more important for the formation of heterosis, especially BPH, than CHG gbM.

DNA methylation signatures and the prediction of heterosis

It is known that screening for F_1 hybrids with strong heterosis is still a laborious task so far. Geneticists and breeders have been trying to find relevant clues to predict heterosis at different levels, including genetic distance (genomic level; Laborda et al., 2005), gene expression (transcriptomic level; Miller et al., 2015), non-linear phenotypic variation (phenomic level; Vasseur et al., 2019) and metabolomic pathway (metabolomic level; Dan et al., 2021). More and more studies showed that genetic distance was not completely correlated with the degree of heterosis in multiple species (Fernandes et al., 2015; Silva et al., 2019; Wang et al., 2015; Yang et al., 2017). Several other omics studies cannot easily and intuitively predict the level of heterosis. Methylation information has been used to predict the complex quantitative traits such as PH and human longevity (Horvath & Raj, 2018; Hu et al., 2015). In our study, we observed the significant correlation between methylation level and GYPP and the different distribution characteristics of CG DMRs in the different functional regions of gene body between parents of better-parent and mid-parents F_1 hybrids. It was reported that gbM preferentially occurs at exons and introns, and is rare at the TSSs and transcription stop sites (Takuno & Gaut, 2013). Correlation analysis also showed that CG DMRs in exons and TSSs between parental lines were also negatively and positively correlated with the magnitudes of heterosis of their F_1 hybrids, respectively. It is not easy and accurate to evaluate the level of heterosis using the absolute numbers of CG DMRs in exons or TSSs between any two parental lines. Their ratios also significantly correlated with the heterosis level of the corresponding F_1 hybrids, and the correlation coefficients were also larger than those of the absolute number of CG DMRs in exons and TSSs. This correlation was also confirmed in 24 pairwise combinations from an additional 19 different types of rice lines. Hence, we thought that the ratio of CG DMRs in exons to TSSs between parental lines was a potential predictor for heterosis, and the ratio of less than 5 may be the critical indicator to judge a hybrid with BPH based on a total of 23 parental lines used in our study.

EXPERIMENTAL PROCEDURES

Plant materials

Four rice inbred paternal lines including two restorer lines Guanghui 998 (R998) and Guanghui 308 (R308) and two maintainer lines Wufeng B (WFB) and Rongfeng B (RFB) instead of the corresponding sterile lines Wufeng A (WFA) and Rongfeng A (RFA) were used to generate five hybrids: WY998 (WFB/R998), WY308 (WFB/R308), RY998 (RFB/R998), RY308 (RFB/R308) and WR (WFB/RFB). Up to date, more than 140 F_1 hybrids from them (95 from

WFA, 27 from RFA, 15 from R998 and seven from R308) were approved above the provincial level and cumulatively planted more than 30 million hectares in Southern China. Of them, WY308 was recognized as a super rice variety by Ministry of Agriculture and Rural Affairs of China and was used as a control variety for the national variety regional experiment. These lines were grown into three independent replicates (100 plants per line per replicate) in the paddy field of the Dafeng experimental base in Guangzhou, China in 2017. Thirty plants from each line per replicate were randomly selected to investigate yield-related traits. The sample pools from 30 plants from each line were used to perform whole-genome bisulfite sequencing (WGBS; two independent biological replicates) and transcriptome sequencing (three independent biological replicates) by Novogene Bioinfromatics Technology.

Evaluation of heterosis for six yield-related traits

The five hybrids and their four parents (100 plants per line) were grown for assessing the phenotypic performance in the paddy field of the Dafeng experimental base in Guangzhou, China in 2017. We monitored 270 plants (30 replicate plants per hybrid or parental line) for a range of yield-related traits, including tiller number (TN), PN, PH, GNPP, KGW and GYPP. TN per line was scored every 5 days until heading from 15 DAT to understand the tillering dynamics. PH was measured after heading, and PN, GNPP, KGW and GYPP were scored after mixed harvesting of 30 plants.

Calculation of heterosis is as follows:

$$\text{BPH } H_{bp} = (F1 - HP)/HP \times 100\%.$$

Note: HP means the high-value parent.

$$\text{MPH } H_{mp} = (F1 - M)/M \times 100\%.$$

Note: M means the average value of both parents.

Sample preparing and transcriptome sequencing

We collected the young leaves 20 DAT, young panicles 10 days before heading and filling panicles 15 days after anthesis, respectively, and three biological replicate samples, each containing 30 corresponding tissues from 30 plants, were prepared to isolate total RNAs. Total RNA samples were extracted using TRIzol reagent (Invitrogen) and treated with RNase-free DNase I (Takara, Dalian, China) to remove genomic DNA. mRNA libraries were constructed according to the standard protocols provided by Illumina. The quality of mRNA including purity, quantity and integrity was tested using Nanodrop, Qubit and Agilent 2100. mRNA extraction was performed using Dynabeads oligo (dT; Dynal; Invitrogen) and fragmented using fragmentation buffer. Double-stranded cDNAs were synthesized using reverse transcriptase (Superscript II; Invitrogen) and random hexamer primers, and further purified using AMPure XP beads. Finally, the purified double-stranded cDNA samples were further enriched by PCR to construct the final cDNA libraries that were sequenced using Hiseq4000 (150-bp paired ends) by Novogene (Beijing, China).

Differential expression analysis and gene expression models

Adaptor sequences and low-quality sequences were removed from the raw reads ($Q < 20$). Paired-end clean reads were aligned to the *indica* rice R498 genome (http://mbkbase.org/R498/) using Hisat2 (v2.0.4). HTSeq (v0.6.1) was used to count the reads numbers mapped to each gene. Then, FPKM (fragments per kilobase of transcript per million mapped reads) of each gene was calculated based on the length of the gene and reads count mapped to this gene. Differential expression analysis of two lines from five hybrids and their four parents (three biological replicates per line) was performed using the DESeq R package (1.10.1). DESeq provide statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting *P*-values were adjusted using the Benjamini and Hochberg's approach for controlling the FDR. Genes with an adjusted $P < 0.05$ found by DESeq were assigned as differentially expressed. Gene expression models, including additivity, expression-level dominance and overdominance, were classified according to Rapp et al. (2009) and Li et al. (2014).

SNP calling and ASE identification

Picard-tools (v1.96) and SAMtools (v0.1.18) were used to sort and mark duplicated reads and reorder the bam alignment results of each sample. ASE reads separation and identification were carried out as previously described (Shao et al., 2019). R498 genome sequence was used as the reference sequence, and GATK3 software was used to perform SNP calling using default parameters. The identification of ASE for each gene was based on the SNPs between two parental genomes, and unreliable SNPs were filtered out according to the following criteria: (i) all reads uniquely match both parents' genomes; (ii) all reads from one parent produce a consensus base at the SNP position but different from another parent; (iii) the SNP is represented by at least 10 reads. A gene showing ASE of more than one SNP was referred to as an ASEG. Allelic bias in hybrids was identified by determining for each SNP whether there was significant deviation from the binomial distribution of parental alleles (i.e. the allele ratio in the hybrids deviated from 1.0) according to Hu et al. (2016).

WGBS and differentially methylated analysis

Genomic DNA was fragmented by sonication to 200–300 bp with Covaris S220, followed by end repair and adenylation. Cytosine-methylated barcodes were ligated to sonicated DNA as per manufacturer's instructions. Then these DNA fragments were treated twice with bisulfite using EZ DNA Methylation-GoldTMKit (Zymo Research, Orange County, CA, USA), before the resulting single-strand DNA fragments were PCR amplificated using KAPA HiFi-HotStart Uracil + ReadyMix (2 ×). Library concentration was quantified by Qubit® 2.0 Flurometer (Life Technologies, Carlsbad, CA, USA) and quantitative PCR, and the insert size was assayed on Agilent Bioanalyzer 2100 system.

The library preparations were sequenced on an Illumina Hiseq4000 and 150-bp and paired-end reads were generated by Novogene (China). Raw data were filtered, and the low-quality data were removed through Trimmomatic (Trimmomatic-0.36) software use the parameter (SLIDINGWINDOW:4:15; LEADING:3, TRAILING:3; ILLUMINACLIP:adapter.fa: 2:30: 10;MINLEN:36). FastQC were used to perform basic statistics on the quality of the clean data reads. The reference genomes were firstly transformed into bisulfite-converted version (C-to-T and G-to-A converted). Bismark software (version 0.16.3) was used to perform alignments of clean data reads to indica rice R498 genome (http://mbkbase.org/R498/). The sequences were divided into multiple bins with 10-kb size to calculate the methylation level. DMRs were identified using the DSS software using the parameter (smoothing.span = 200, delta = 0, p.threshold = 1e-05, minlen = 50, minCG = 3, dis.merge = 100, pct.sig = 0.5).

TEs characterization and analysis

The TEs were screened and annotated using the RepeatMasker program (http://www.repeatmasker.org/). The parameters were set

as follows: match:2; mismatch: 7; delta: 7; PM:80; PI:10; minscore: 50; maximum period: 500. Based on the RepeatMasker annotation, all genes were divided into two groups: genes with TEs insertion or not.

AUTHOR CONTRIBUTIONS

CF, writing, analyzing the data of sequencing and designing the experiment; CM, analyzing the data of sequencing; FW, designing the experiment and revising the manuscript; XG and HW, revising the manuscript; MZ, preparing the experimental materials; WL, XM, JL, YL and DL scoring the yield-related traits.

ACKNOWLEDGEMENTS

This work was supported in part by grants from State Key R & D Program (2017YFD0100102), Laboratory of Lingnan modern agriculture project (NT2021001), the earmarked fund for Modern Agro-industry Technology Research System (CARS-01-18), the Joint Funds of the National Natural Science Foundation of China (U20A2035), Agricultural competitive industry discipline team building project of Guangdong academy of agricultural sciences and Guangdong Scientific Program Project (2020B1212060047).

DATA AVAILABILITY STATEMENT

The raw data generated by our study are available in SequenceRead Archive (SRA) database of the National Center for Biotechnology Information (https://trace.ncbi.nlm.nih.gov/Traces/sra/) under the accession number PRJNA774890.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Pearson correlation between samples in RNA-seq.

Figure S2. The Venn diagrams displaying the number of detected SNPs in every rice parental lines based on the reference genome of R498.

Figure S3. The methylation level in the panicles of five hybrids and their parents.

(a) The total level of methylation in the panicles of nine genotypes. (b) The level of methylation in three cytosine contexts in young panicles of nine genotypes. (c) The level of methylation in three cytosine contexts in filling panicles of nine genotypes.

Figure S4. The correlation between methylation level and grain yield per plant (GYPP) in panicles.

(a–d) the correlation between methylation levels in mC, mCpG, mCHG, mCHH and GYPP in young panicles, respectively; (e–h) the correlation between methylation levels in mC, mCpG, mCHG, mCHH and GYPP in young panicles, respectively. The *x*-axis represents methylation levels, and the *y*-axis represents GYPP.

Figure S5. The distribution of methylation sites in the regions of gene body and 2 kb upstream of OsR498G0204681300.01 (*DTH2*) and the information of its expression.

Table S1 The performance of several agronomic traits in four hybrid rice parental lines and their five hybrids

Table S2 The produced reads of transcriptomic sequencing and the reads mapped on reference genome

Table S3 DEGs detected in pairwise combinations including between parents, and between the hybrids and their parents

Table S5 The information of tested 49 SNPs

Table S6 The expression patterns of ASEGs in five F_1 hybrids

Table S7 The dominant ASEGs shared in better-parent hybrids

Table S8 Coverage information of cytosine base in different cytosine contexts of genome

Table S9 The analysis of the shared genes between DMRs and DEGs in pairwise comparisons

Table S10 The primers for amplifying *DTH2* and its promoter regions

Table S11 The detected DMRs among F_1 hybrids and their parents in three cytosine contexts

Table S12 DMRs in different function regions of gene body of pairwise combinations between parents, and between five hybrids and their parents, and the ratios of DMRs in exons to TSSs

Table S13 The information of 19 rice lines used for methylome sequencing

Table S14 CG DMRs detected in exons and TSSs in the pair combinations of different types of parental lines and the ratios of CG DMRs in TSSs

Table S4 The shared DEGs only among four better-parent hybrids

REFERENCES

Berruyer, R., **Adreit, H.**, **Milazzo, J.**, **Gaillard, S.**, **Berger, A.**, **Dioh, W.** ***et al.*** (2003) Identification and fine mapping of *Pi33*, the rice resistance gene corresponding to the *Magnaporthe grisea* avirulence gene *ACE1*. *Theoretical and Applied Genetics*, **107**, 1139–1147. Available from: https://doi.org/10.1007/s00122-003-1349-2

Bewick, A.J. & **Schmitz, R.J.** (2017) Gene body DNA methylation in plants. *Current Opinion in Plant Biology*, **36**, 103–110. Available from: https://doi.org/10.1016/j.pbi.2016.12.007

Birchler, J.A., **Yao, H.**, **Chudalayandi, S.**, **Vaiman, D.** & **Veitia, R.A.** (2010) Heterosis. *Plant Cell*, **22**, 2105–2112. Available from: https://doi.org/10.1105/tpc.110.076133

Chen, X., **Shang, J.**, **Chen, D.**, **Lei, C.**, **Zou, Y.**, **Zhai, W.** ***et al.*** (2006) A B-lectin receptor kinase gene conferring rice blast resistance. *The Plant Journal*, **46**(5), 794–804. Available from: https://doi.org/10.1111/j.1365-313X.2006.02739.x

Chen, Z.J. (2013) Genomic and epigenetic insights into the molecular bases of heterosis. *Nature Reviews. Genetics*, **14**(7), 471–482. Available from: https://doi.org/10.1038/nrg3503

Cheng, S.H., **Zhuang, J.Y.**, **Fan, Y.Y.**, **Du, J.H.** & **Cao, L.Y.** (2007) Progress in research and development on hybrid rice: a super-domesticate in China. *Annals of Botany*, **100**, 959–966. Available from: https://doi.org/10.1093/aob/mcm121

Dafna, A., **Halperin, I.**, **Oren, E.**, **Isaacson, T.**, **Tzuri, G.**, **Meir, A.** ***et al.*** (2021) Underground heterosis for yield improvement in melon. *Journal of Experimental Botany*, **72**(18), 6205–6218. Available from: https://doi.org/10.1093/jxb/erab219

Dan, Z., **Chen, Y.**, **Li, H.**, **Zeng, Y.**, **Xu, W.**, **Zhao, W.** ***et al.*** (2021) The metabolomic landscape of rice heterosis highlights pathway biomarkers for predicting complex phenotypes. *Plant Physiology*, **187**(2), 1011–1025. Available from: https://doi.org/10.1093/plphys/kiab273

Davenport, C.B. (1908) Degeneration, albinism and inbreeding. *Science*, **28**, 454–455. Available from: https://doi.org/10.1126/science.28.718.454-b

East, E.M. (1908) Inbreeding in corn. In: *Report of the connecticut agricultural experiment station for years 1907–1908*. New Haven, CT, pp. 419–428.

Eichten, S.R., **Briskine, R.**, **Song, J.**, **Li, Q.**, **Swanson-Wagner, R.**, **Hermanson, P.J.** ***et al.*** (2013) Epigenetic and genetic influences on DNA methylation variation in maize populations. *Plant Cell*, **25**(8), 2783–2797. Available from: https://doi.org/10.1105/tpc.113.114793

Fernandes, E.H., **Schuster, I.**, **Scapim, C.A.**, **Vieira, E.S.N.** & **Coan, M.M.D.** (2015) Genetic diversity in elite inbred lines of maizeand its association

with heterosis. *Genetics and Molecular Research*, **14**(2), 6509–6517. Available from: https://doi.org/10.4238/2015.June.12.3

Fujino, K., **Yamanouchi, U.** & **Yano, M.** (2013) Roles of the *Hd5* gene controlling heading date for adaptation to the northern limits of rice cultivation. *Theoretical and Applied Genetics*, **126**(3), 611–618. Available from: https://doi.org/10.1007/s00122-012-2005-5

Gai, J.Y., **Dai, J.R.**, **Li, J.S.** & **Sun, Q.X.** (2016) *Utilization of heterosis in crops*. Beijing, China: High Education Press. (in Chinese).

Groszmann, M., **Greaves, I.K.**, **Fujimoto, R.**, **Peacock, W.J.** & **Dennis, E.S.** (2013) The role of epigenetics in hybrid vigour. *Trends in Genetics*, **12**, 684–690. Available from: https://doi.org/10.1016/j.tig.2013.07.004

Guo, M., **Rupe, M.A.**, **Yang, X.**, **Crasta, O.**, **Zinselmeier, C.**, **Smith, O.S.** ***et al.*** (2006) Genome-wide transcript analysis of maize hybrids: allelic additive gene expression and yield heterosis. *Theoretical and Applied Genetics*, **113**(5), 831–845. Available from: https://doi.org/10.1007/s00122-006-0335-x

He, G.M., **He, H.** & **Deng, X.W.** (2013) Epigenetic variations in plant hybrids and their potential roles in heterosis. *Journal of Genetics and Genomics*, **40**, 205–210. Available from: https://doi.org/10.1016/j.jgg.2013.03.011

Horvath, S. & **Raj, K.** (2018) DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews. Genetics*, **19**, 371–384. Available from: https://doi.org/10.1038/s41576-018-0004-3

Hu, X., **Wang, H.**, **Diao, X.**, **Liu, Z.**, **Li, K.**, **Wu, Y.** ***et al.*** (2016) Transcriptome profiling and comparison of maize ear heterosis during the spikelet and floret differentiation stages. *BMC Genomics*, **17**(1), 959. Available from: https://doi.org/10.1186/s12864-016-3296-8

Hu, Y., **Morota, G.**, **Rosa, G.J.** & **Gianola, D.** (2015) Prediction of plant height in Arabidopsis thaliana using DNA methylation data. *Genetics*, **201**(2), 779–793. Available from: https://doi.org/10.1534/genetics.115.177204

Huang, X., **Yang, S.**, **Gong, J.**, **Zhao, Y.**, **Feng, Q.**, **Gong, H.** ***et al.*** (2015) Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nature Communications*, **6**, 6258. Available from: https://doi.org/10.1038/ncomms7258

Jiang, W., **Chu, S.H.**, **Piao, R.**, **Chin, J.H.**, **Jin, Y.M.**, **Lee, J.** ***et al.*** (2008) Fine mapping and candidate gene analysis of *hwh1* and *hwh2*, a set of complementary genes controlling hybrid breakdown in rice. *Theoretical and Applied Genetics*, **116**(8), 1117–1127. Available from: https://doi.org/10.1007/s00122-008-0740-4

Kawanabe, T., **Ishikura, S.**, **Miyaji, N.**, **Sasaki, T.**, **Wu, L.M.**, **Itabashi, E.** ***et al.*** (2016) Role of DNA methylation in hybrid vigor in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, **113**(43), E6704–E6711. Available from: https://doi.org/10.1073/pnas.1613372113

Laborda, P.R., **Oliveira, K.M.**, **Garcia, A.A.F.**, **Paterniani, M.E.A.G.Z.** & **de Souza, A.P.** (2005) Tropical maize germoplasm: what can we say about its genetic diversity in the light of molecular markers? *Theoretical and Applied Genetics*, **111**, 1288–1299. Available from: https://doi.org/10.1007/s00122-005-0055-7

Lauss, K., **Wardenaar, R.**, **Oka, R.**, **van Hulten, M.H.A.**, **Guryev, V.**, **Keurentjes, J.J.B.** ***et al.*** (2018) Parental DNA methylation states are associated with heterosis in epigenetic hybrids. *Plant Physiology*, **176**, 1627–1645. Available from: https://doi.org/10.1104/pp.17.01054

Lee, S.K., **Song, M.Y.**, **Seo, Y.S.**, **Kim, H.K.**, **Ko, S.**, **Cao, P.J.** ***et al.*** (2009) Rice *Pi5*-mediated resistance to *Magnaporthe oryzae* requires the presence of two coiled-coil-nucleotide-binding-leucine-rich repeat genes. *Genetics*, **181**(4), 1627–1638.

Li, A., **Liu, D.**, **Wu, J.**, **Zhao, X.**, **Hao, M.**, **Geng, S.** ***et al.*** (2014) mRNA and small RNA transcriptomes reveal insights into dynamic homoeolog regulation of allopolyploid heterosis in Nascent Hexaploid Wheat. *Plant Cell*, **26**(5), 1878–1900.

Li, L., **Lu, K.**, **Chen, Z.**, **Mu, T.**, **Hu, Z.** & **Li, X.** (2008) Dominance, overdominance and epistasis condition the heterosis in two heterotic rice hybrids. *Genetics*, **180**, 1725–1742. Available from: https://doi.org/10.1534/genetics.108.091942

Li, Q., **Song, J.**, **West, P.T.**, **Zynda, G.**, **Eichten, S.R.**, **Vaughn, M.W.** ***et al.*** (2015) Examining the causes and consequences of context-specific differential DNA methylation in maize. *Plant Physiology*, **168**(4), 1262–1274. Available from: https://doi.org/10.1104/pp.15.00052

Li, S., **Lin, D.**, **Zhang, Y.**, **Deng, M.**, **Chen, Y.**, **Lv, B.** ***et al.*** (2022) Genome-edited powdery mildew resistance in wheat without growth penalties. *Nature*, **602**(7897), 455–460. Available from: https://doi.org/10.1038/s41586-022-04395-9

Ma, X., **Xing, F.**, **Jia, Q.**, **Zhang, Q.**, **Hu, T.**, **Wu, B.** ***et al.*** (2021) Parental variation in CHG methylation is associated with allelic-specific expression in elite hybrid rice. *Plant Physiology*, **186**(2), 1025–1041. Available from: https://doi.org/10.1093/plphys/kiab088

Miller, M., **Song, Q.X.**, **Shi, X.L.**, **Juenger, T.E.** & **Chen, Z.J.** (2015) Natural variation in timing of stress-responsive gene expression predicts heterosis in intraspecific hybrids of Arabidopsis. *Nature Communications*, **6**, 7453. Available from: https://doi.org/10.1038/ncomms8453

Powers, L. (1944) An expansion of Jones' theory for the explanation of heterosis. *The American Naturalist*, **78**, 275–280.

Rapp, R.A., **Udall, J.A.** & **Wendel, J.F.** (2009) Genomic expression dominance in allopolyploids. *BMC Biology*, **7**, 18. Available from: https://doi.org/10.1186/1741-7007-7-18

Schmitz, R.J., **Schultz, M.D.**, **Urich, M.A.**, **Nery, J.R.**, **Pelizzola, M.**, **Libiger, O.** ***et al.*** (2013) Patterns of population epigenomic diversity. *Nature*, **495** (7440), 193–198. Available from: https://doi.org/10.1038/nature11968

Schnable, P.S. & **Springer, N.M.** (2013) Progress toward understanding heterosis in crop plants. *Annual Review of Plant Biology*, **64**, 71–88. Available from: https://doi.org/10.1146/annurev-arplant-042110-103827

Shao, L., **Xing, F.**, **Xu, C.**, **Zhang, Q.**, **Che, J.**, **Wang, X.** ***et al.*** (2019) Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *Proceedings of the National Academy of Sciences of the United States of America*, **16**(2), 5653–5658. Available from: https://doi.org/10.1073/pnas.1820513116

Shen, G.J., **Hu, W.**, **Zhang, B.** & **Xing, Y.Z.** (2015) The regulatory network mediated by circadian clock genes is related to heterosis in rice. *Journal of Integrative Plant Biology*, **57**(3), 300–312. Available from: https://doi.org/10.1111/jipb.12240

Shen, H., **He, H.**, **Li, J.**, **Chen, W.**, **Wang, X.**, **Guo, L.** ***et al.*** (2012) Genome-wide analysis of DNA methylation and gene expression changes in two Arabidopsis ecotypes and their reciprocal hybrids. *Plant Cell*, **24**, 875–892. Available from: https://doi.org/10.1105/tpc.111.094870

Si, L., **Chen, J.**, **Huang, X.**, **Gong, H.**, **Luo, J.**, **Hou, Q.** ***et al.*** (2016) *OsSPL13* controls grain size in cultivated rice. *Nature Genetics*, **48**(4), 447–456. Available from: https://doi.org/10.1038/ng.3518

Silva, K.J., **Guimarães, C.T.**, **Guilhen, J.H.S.**, **de Oliveira Guimarães, P.E.**, **Parentoni, S.N.**, **dos Santos Trindade, R.** ***et al.*** (2019) High-density SNP-based genetic diversity and heterotic patterns of tropical maize breeding lines. *Crop Science*, **60**, 779–787. Available from: https://doi.org/10.1002/csc2.20018

Takuno, S. & **Gaut, B.S.** (2013) Gene body methylation is conserved between plant orthologs and is of evolutionary consequence. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 1797–1802. Available from: https://doi.org/10.1073/pnas.1215380110

To, T.K., **Saze, H.** & **Kakutani, T.** (2015) DNA methylation within transcribed regions. *Plant Physiology*, **168**, 1219–1225. Available from: https://doi.org/10.1104/pp.15.00543

Urayama, S., **Moriyama, H.**, **Aoki, N.**, **Nakazawa, Y.**, **Okada, R.**, **Kiyota, E.** ***et al.*** (2010) Knock-down of *OsDCL2* in rice negatively affects maintenance of the endogenous dsRNA virus, *Oryza sativa* endornavirus. *Plant & Cell Physiology*, **51**(1), 58–67. Available from: https://doi.org/10.1093/pcp/pcp167

Vasseur, F., **Fouqueau, L.**, **de Vienne, D.**, **Nidelet, T.**, **Violle, C.** & **Weigel, D.** (2019) Nonlinear phenotypic variation uncovers the emergence of heterosis in Arabidopsis thaliana. *PLoS Biology*, **17**(4), e3000214. Available from: https://doi.org/10.1371/journal.pbio.3000214

Wang, K., **Qiu, F.**, **Larazo, W.**, **Dela Paz, M.A.** & **Xie, F.** (2015) Heterotic groups of tropical indica rice germplasm. *Theoretical and Applied Genetics*, **128**, 421–430. Available from: https://doi.org/10.1007/s00122-014-2441-5

Wei, X., **Qiu, J.**, **Yong, K.**, **Fan, J.**, **Zhang, Q.**, **Hua, H.** ***et al.*** (2021) A quantitative genomics map of rice provides genetic insights and guides breeding. *Nature Genetics*, **53**(2), 243–253. Available from: https://doi.org/10.1038/s41588-020-00769-9

Wei, X., **Xu, J.**, **Guo, H.**, **Jiang, L.**, **Chen, S.**, **Yu, C.** ***et al.*** (2010) DTH8 suppresses flowering in rice, influencing plant height and yield potential simultaneously. *Plant Physiology*, **153**(4), 1747–1758. Available from: https://doi.org/10.1104/pp.110.156943

Wu, W., **Zheng, X.M.**, **Lu, G.**, **Zhong, Z.**, **Gao, H.**, **Chen, L.** ***et al.*** (2013) Association of functional nucleotide polymorphisms at *DTH2* with the northward expansion of rice cultivation in Asia. *Proceedings of the National Academy of Sciences of the United States of America*, **110**(8), 2775–2780. Available from: https://doi.org/10.1073/pnas.1213962110

Yang, L., **Liu, P.**, **Wang, X.**, **Jia, A.**, **Ren, D.**, **Tang, Y.** ***et al.*** (2021) A central circadian oscillator confers defense heterosis in hybrids without growth vigor costs. *Nature Communications*, **12**(1), 2317. Available from: https://doi.org/10.1038/s41467-021-22268-z

Yang, M., **Wang, X.**, **Ren, D.**, **Huang, H.**, **Xu, M.**, **He, G.** ***et al.*** (2017) Genomic architecture of biomass heterosis in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, 8101–8106. Available from: https://doi.org/10.1073/pnas.1705423114

Younas, M., **Xiao, Y.**, **Cai, D.**, **Yang, W.**, **Ye, W.**, **Wu, J.** ***et al.*** (2012) Molecular characterization of oilseed rape accessions collected from multi continents for exploitation of potential heterotic group through SSR markers. *Molecular Biology Reports*, **39**(5), 5105–5113. Available from: https://doi.org/10.1007/s11033-011-1306-0

Yu, S.B., **Li, J.X.**, **Xu, C.G.**, **Tan, Y.F.**, **Gao, Y.J.**, **Li, X.H.** ***et al.*** (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 9226–9231. Available from: https://doi.org/10.1073/pnas.94.17.9226

Zhang, D., **Liu, M.**, **Tang, M.**, **Dong, B.**, **Wu, D.**, **Zhang, Z.** ***et al.*** (2015) Repression of microRNA biogenesis by silencing of *OsDCL1* activates the basal resistance to *Magnaporthe oryzae* in rice. *Plant Science*, **237**, 24–32. Available from: https://doi.org/10.1016/j.plantsci.2015.05.002

Zhang, H., **Lang, Z.** & **Zhu, J.K.** (2018) Dynamics and function of DNA methylation in plants. *Nature Reviews. Molecular Cell Biology*, **19**, 489–506. Available from: https://doi.org/10.1038/s41580-018-0016-z

Zhang, L., **Peng, Y.**, **Wei, X.**, **Dai, Y.**, **Yuan, D.**, **Lu, Y.** ***et al.*** (2014) Small RNAs as important regulators for the hybrid vigour of super-hybrid rice. *Journal of Experimental Botany*, **65**, 5989–6002. Available from: https://doi.org/10.1093/jxb/eru337

Zhang, Z.H., **Wang, K.**, **Guo, L.**, **Zhu, Y.J.**, **Fan, Y.Y.**, **Cheng, S.H.** ***et al.*** (2012) Pleiotropism of the photoperiod-insensitive allele of *Hd1* on heading date, plant height and yield traits in rice. *PLoS One*, **7**(12), e52538. Available from: https://doi.org/10.1371/journal.pone.0052538

Zhen, G., **Qin, P.**, **Liu, K.Y.**, **Nie, D.Y.**, **Yang, Y.Z.**, **Deng, X.W.** ***et al.*** (2017) Genome-wide dissection of heterosis for yield traits in two-line hybrid rice populations. *Scientific Reports*, **7**, 7635. Available from: https://doi.org/10.1038/s41598-017-06742-7

Zhou, G., **Chen, Y.**, **Yao, W.**, **Zhang, C.**, **Xie, W.**, **Hua, J.** ***et al.*** (2012) Genetic composition of yield heterosis in an elite rice hybrid. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 15847–15852. Available from: https://doi.org/10.1073/pnas.1214141109

Zhu, Y.J., **Huang, D.R.**, **Fan, Y.Y.**, **Zhang, Z.H.**, **Ying, J.Z.** & **Zhuang, J.Y.** (2016) Detection of QTLs for yield heterosis in rice using a RIL population and its testcross population. *International Journal of Genomics*, **2016**, 2587823. Available from: https://doi.org/10.1155/2016/2587823