Genome-Wide Single-Nucleotide Polymorphisms in CMS and Restorer Lines Discovered by Genotyping Using Sequencing and Association with Marker-Combining Ability for 12 Yield-Related Traits in *Oryza sativa* L. subsp. Japonica

Imdad U. Zaid[1], Weijie Tang[1], Erbao Liu[1], Sana U. Khan[2], Hui Wang[1], Edzesi W. Mawuli[1] and Delin Hong[1]*

[1] State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China,
[2] School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD, Australia

Heterosis or hybrid vigor is closely related with general combing ability (GCA) of parents and special combining ability (SCA) of combinations. The evaluation of GCA and SCA facilitate selection of parents and combinations in heterosis breeding. In order to improve combining ability (CA) by molecular marker assist selection, it is necessary to identify marker loci associated with the CA. To identify the single nucleotide polymorphisms (SNP) loci associated with CA in the parental genomes of japonica rice, genome-wide discovered SNP loci were tested for association with the CA of 18 parents for 12 yield-related traits. In this study, 81 hybrids were created and evaluated to calculate the CA of 18 parents. The parents were sequenced by genotyping by sequencing (GBS) method for identification of genome-wide SNPs. The analysis of GBS indicated that the successful mapping of 9.86×10^6 short reads in the Nipponbare reference genome consists of 39,001 SNPs in parental genomes at 11,085 chromosomal positions. The discovered SNPs were non-randomly distributed within and among the 12 chromosomes of rice. Overall, 20.4% (8026) of the discovered SNPs were coding types, and 8.6% (3344) and 9.9% (3951) of the SNPs revealed synonymous and non-synonymous changes, which provide valuable knowledge about the underlying performance of the parents. Furthermore, the associations between SNPs and CA indicated that 362 SNP loci were significantly related to the CA of 12 parental traits. The identified SNP loci of CA in our study were distributed genome wide and caused a positive or negative effect on the CA of traits. For the yield-related traits, such as grain thickness, days to heading, panicle length, grain length and 1000-grain weight, a maximum number of positive SNP loci of CA were found in CMS A171 and in the restorers LC64 and LR27. On an individual basis, some

of associated loci that resided on chromosomes 2, 5, 7, 9, and 11 recorded maximum positive values for the CA of traits. From our results, we suggest that heterosis in japonica rice would be improved by pyramiding the favorable SNP loci of CA and eliminating the unfavorable loci from parental genomes.

INTRODUCTION

Hybrid rice breeding has been successfully adopted in 27 rice-growing countries. China is the largest producer and consumer of hybrid rice, and considers itself a pioneer in hybrid rice breeding. With the cultivation of hybrid rice, China obtained 20% more grain yield compared to their best conventional cultivars (Yuan and Virmani, 1988; Cheng et al., 2004). In China, a total of 33 million hectares (ha) of paddy fields are available for rice cultivation, in which 25 million ha are occupied by the indica rice and 8 million ha are occupied by the japonica rice. In the total land of indica rice, 70% are used for indica hybrids rice cultivation, whereas only 5% of the total planted land is used for japonica hybrid cultivations. (Tang et al., 2008). In the past, great achievements were made to increase the growing area and yield productivity of japonica hybrids by improving culturing practices with the use of productive seeds. However, until recently, the yield potential of japonica hybrids has been low and unstable compared to the indica hybrids. Previous breeding practices have confirmed that japonica hybrids encounter several challenges due to their low standard heterosis, which may be caused by their narrow genetic background, their large panicle size with insufficient filling (poor grain plumpness), or the unavailability of elite parental combinations for developing superior hybrids (Xie and Hardy, 2009).

Commercially developed hybrid cultivars exhibit desirable characteristics of yield and quality, that are inherited from their parental lines (Cao and Zhan, 2014). Such parental characteristics are heritable and could be appear in the F_1 generation in the form of heterosis. The successful selection of proper parents, based on the combining ability of yield-related traits, contributes to yield outcomes in hybrid breeding. The combining ability is a powerful breeding test that estimates the breeding values of parents and crosses in terms of general and specific combinations (Sprague and Tatum, 1942). Conventionally, the combining ability of crossing parents is calculated by evaluating all their developed crosses, which is laborious, tedious and time-consuming (Smith et al., 2008). Moreover, when the number of parents involved in combining ability manipulation become large, their hybrids affect the experimental feasibility (Bertan et al., 2007).

With recent rapid developments in molecular marker technology, it is now feasible to genotype or identify marker loci for CA. Previous studies made this possible and discovered a genetic basis of CA with SSR markers (simple sequence repeats) (Liang et al., 2010; Huang et al., 2013; Liu E. B. et al., 2013; Qi et al., 2013; Liu Y. et al., 2015; Xie et al., 2016). These studies have revealed several SSR marker loci associated with the CA of yield and quality traits. However, such studies have been restricted to

SSR markers. To date, no SNP base analysis has been reported for the discovery of SNP locus/loci associated with the CA of parental traits.

Recent advances in next-generation sequencing (NGS) have facilitated genome-wide SNP identification and SNP characterization. The advent of NGS has reduced the cost of genome sequencing to a level, where GBS is now considered a powerful tool for inquiring into a large number of genomic variations (SNPs). The GBS platform is simple, fast and accurate and has been widely practiced for large scale mining of SNPs in many crop species, including wheat (Poland J. et al., 2012), rice (Spindel et al., 2013), soybean (Sonah et al., 2013), chickpea (Kujur et al., 2015), sorghum (Morris et al., 2012) alfalfa (Yu L. X. et al., 2016), and olive (İpek et al., 2016).

SNPs are the most common and abundant sequence variants present in both plant and animal genomes (Kwok, 2001). They are present across the genome, within coding, non-coding and intergenic regions of the genome (Jiang, 2013). They gained tremendous importance as a third-generation molecular marker for a wide range of biological applications of marker-assisted and genomic selection, associations and QTL mapping, haplotype and pedigree analysis and cultivar identification (Lee et al., 2008; McCouch et al., 2010; Subbaiyan et al., 2012).

To this end, our study will appear as a first and new approach for the discovery of selective SNP loci of CA. Prior studies have revealed that CA is a complex quantitative trait (Liu C. et al., 2015). Therefore, the identification of marker loci within the parental genomes of japonica hybrid rice can ultimately be utilized to understand the biological and genetic factors of CA. Here, we aim to develop a reliable strategy to select and determine the superior parents of hybrid rice based on the presence of CA loci, which can be achieved by using SNP data with combinations of SNP and phenotype data. The main objectives of this study were as follows: (1) to dissect japonica parents for genome-wide SNPs; (2) to evaluate japonica parents for GCA effect; and (3) to associate the identified SNPs of CMS and restorer lines with the CA of parents to determine genome-wide favorable or unfavorable SNP loci of CA related to yield-related traits.

MATERIALS AND METHODS

Library Construction and Sequencing

The total DNA of 18 japonica rice was rice extracted from fresh leaves (10–20 mg) selected at the seedling stage using the Qiagen DNeasy Plant Mini Kit. The experimental procedure for DNA isolation was the same as that previously described (Stein et al., 2001). Following the recommended protocol of GBS, optimized by Poland et al., a restriction digest buffer NEB Buffer 4 and a core

set of restriction enzymes, *Pst*I (CTGCAG) and *Msp*I (CCGG), were added to each sample to obtain a uniform distribution of cut sites across the rice genome (Poland J. A. et al., 2012). The DNA library fragment size and library concentration were analyzed using a bioanalyzer (Agilent) machine (Parson et al., 2013). To inactivate the enzymes, the samples were incubated twice at 37°C for 2 h and at 65°C for 20 min. A set of already developed adapters with 500 ng of high-quality DNA was attached to the sample ends preceding the ligation reaction. The ligated samples were pooled and selected on a 1.5% agarose gel (200–300 bp) followed by PCR-amplification in a single tube. The resulting fragments in the DNA library were sequenced using the Ion Torrent (PGM) sequencing machine with Ion Torrent kits (PGM 200 Kit v2) following the manufacturer specifications (Life Technologies, Carlsbad, CA, and U.S.A.). The sequencing data obtained in our experiment have been deposited into the NCBI short read archive (SRA) under the study accession number:SRR2758809.

Short Read Mapping, SNP Discovery and Annotation

The FASTQ format sequence reads obtained from the sequencing machine were treated for quality control. For this, the adopter sequences and low quality reads were analyzed and trimmed. The cleaned sequences of all 18 samples were aligned onto the Nipponbare genome (IRGSP-1.0) (http://rapdb.dna.affrc.go.jp/) with bowtie 2 software (Langmead, 2010). The –M4 –very-sensitive local parameter for mapping was adopted (-D 20 -R 3 -N 0 -L 20 -i S,1,0.50), where if the alignment is greater than 4, it will output only 1 alignment at a random level. The reads that mapped successfully once onto a reference genome were considered to be uniquely mapped reads, whereas the reads mapped onto multiple locations were considered to be multiple-mapped reads. The short reads that could not be mapped onto any part of the reference were called unmapped reads. The GBS data were further used for SNP discovery using the Tassel-GBS application with default parameters (Glaubitz et al., 2014). The SNPs that evolved repeatedly were deleted from the final results.

The physical locations of the discovered SNPs were annotated by SNPEff software (Cingolani et al., 2012). The SNPs that were situated within exon, intron and other corresponding regions were classified in detail. We also grouped the identified SNPs into different classes of polymorphisms, such as transition/transversions mutations and gene-base synonymous and non-synonymous nucleotide base mutations.

Field Experiment and GCA Calculation

In our experiment, all the plant materials belonged to Japonica rice, comprising nine CMS and nine restorer lines. These parental lines are widely used in China for the commercial breeding of three-line hybrid rice. For hybrid development, all the CMS and restorer lines were planted and manually crossed in a set of 9×9 combinations following the NCII mating design (North Carolina mating design II) (Shukla and Pandey, 2008). At the end of rice-cropping season, a population of 81 hybrids was obtained. The resultant hybrid seeds were grown and evaluated the next year at the Jiangpu research station, Nanjing Agricultural University, Nanjing China, with randomized complete block designs in four

replications. The seeds of hybrid and their parents (restorer lines, and maintainer lines instead of CMS lines) were sown in the rice nursery in May 10, 2015. After 30 days, the seedlings of each hybrid combination were transplanted to paddy fields. Each plot in the paddy contained 4 rows with 8 plants per row. Twenty centimeter row-to-row and plant-to-plant spacing for all hybrids and their parents was maintained. Recommended field management and cultural practices were applied accordingly to the local environment. At maturity, six plants from each plot were randomly selected to calculate the phenotypic data of 12 yield-related traits. The phenotypic data of the traits were tested for significance using the statistical method of analysis of variance, along with the coefficient of variation (CV %) as described by Gomez and Gomez (1984). The GCA effect values of each CMS and restorer line for the traits were calculated by using the phenotypic data of the developed hybrids, as per the Griffing model (1956), following method-2 (Mo, 1982). The significance of the parental GCA value was tested with an LSD test at a $P < 0.01$.

Association Analysis

To identify the SNP locus/loci associated with the CA of the yield-related traits, an association analysis between the identified SNPs and CA of 12 parental traits was performed using a computational software called CA screen 1.0 operated in the MATLAB language and developed by our laboratory* (Liang et al., 2010). The script of our association method follows the principle of single-marker analysis (SMA). This method of association led to the statistically significant identification of SNPs and their effect on CA in homozygous and heterozygous associations. Moreover, we explain the principle of this association model corresponding to the research article provided, where we developed 81 F_1 combinations by crossing nine CMS with nine restorer lines. Now, at a given locus of SNP, if the parental lines of 41 F_1 hybrids possess the heterozygous SNP genotype (for example A-G, A-T), and the parental lines of the remaining 40 F_1 hybrids possess the homozygous SNP genotype (A-A or G-G). Now, after association, if the average trait value of the 41 heterozygous crosses is significantly greater or less than the average trait value of the 40 homozygous crosses, then the SNP marker associated with the CA of the trait is significantly positive or negative. If the difference in the trait value in the heterozygous association is positive and we observed a positive effect on the CA of the trait, we consider this SNP locus to be a favorable associated marker genotype of the elite CA for the trait, and *vice versa*.

In our study, to avoid spurious associations, the significance of the associations was assessed by using a *t*-test at $P < 0.01$ (Pradeep et al., 2007). Furthermore, coefficient of determination (R^2) was calculated to determine the percentage of phenotypic variation explained by each SNP marker.

Selection of Functional SNPs

To determine the substantial number of functional SNPs, we selected associated SNPs based on their genomic positions. Only the SNP loci situated within the coding region of the genome (synonymous, non- synonymous and splice variants) were considered. Furthermore, in the NCBI, the database GENE

was used to search the corresponding genes with associated SNPs for their nomenclature, chromosomal location, gene product, protein domain, and expected functions.

RESULTS

Short Read Mapping, SNP Discovery, and Distribution

Using a GBS assay, approximately 16 million (16.3×10^6) short reads were constituted, which comprised 4.7 GB of data. After removal of low quality reads, the sequenced data were mapped onto the Nipponbare reference genome. The mapping result indicated that 60.4% (9.86×10^6) of the short reads were uniquely mapped onto the 12 chromosomes of the Nipponbare genome. Approximately 17% (2.78×10^6) of the short reads were mapped more than once (**Figure 1**). The remaining 22.4% (3.67×10^6) of the short reads could not be mapped onto any part of the reference genome.

Within the mapped sequenced reads of the nine CMS and nine restorer lines, a total of 39,001 SNPs were discovered at 11,085 sites, including 13,186 and 25,815 in the CMS and restorer lines, respectively. Despite having the same coverage in the genomes, the SNPs identified in the restorer lines were double those in the CMS lines. Of the nine CMS lines, 13,186 SNPs were identified, where the number of SNPs within the CMS lines ranged from 365 in Aizhixiang A to 3412 in Chunjiang 18A (Supplementary Table 1). The discovered SNPs in the CMS lines were distributed non-randomly on the 12 chromosomes, where chromosome 11 occupied a maximum (2740) and chromosome 5 occupied a minimum (381) of SNP numbers (**Figure 2**). The density distributions of the discovered SNPs in the CMS and restorer lines were explored by calculating the SNP frequency within a 200 kb genomic region. The average SNP densities between the CMS lines per 200 kb window were 7.5. Likewise, the density distribution varied across the chromosomes, where chromosome 9 possessed the highest SNP densities of 18.3, while chromosome 5 possessed the lowest SNP densities of 3.2. This phenomenon was also found to be uneven within the individual chromosome. For example, chromosome 11 in the CMS lines was found to be dense (750 SNPs) from the 25 to 28.8 MB genomic region, but scant SNPs (48 SNPs) were observed from the 11 to 16 MB genomic region. Similarly, on chromosome 6, the 10 MB region from 1.8 to 11.8 contained 542 SNPs, but on the same chromosome, the region from 16.6 to 22.4 MB had only 53 SNPs. On chromosome 9, we observed 34 high-density SNP sites with >30 SNPs per 200 kb region (**Figure 3**).

The discovered SNPs in nine restorer lines were counted as 25,815. These SNPs ranged from 1508 in R4179 to 5397 in LR5 (Supplementary Table 2). Due to the different chromosomal lengths, the numbers of SNPs across individual chromosomes varied. For example, for all the restorer lines, chromosome 11 was found to be rich in SNP number with 4845 SNPs, whereas a low SNP number (464) was found on chromosome 4 (**Figure 4**). The average densities of the SNPs among the nine restorer lines were 15 SNPs per 200 kb genomic region. Such densities across chromosomes were non-randomly distributed.
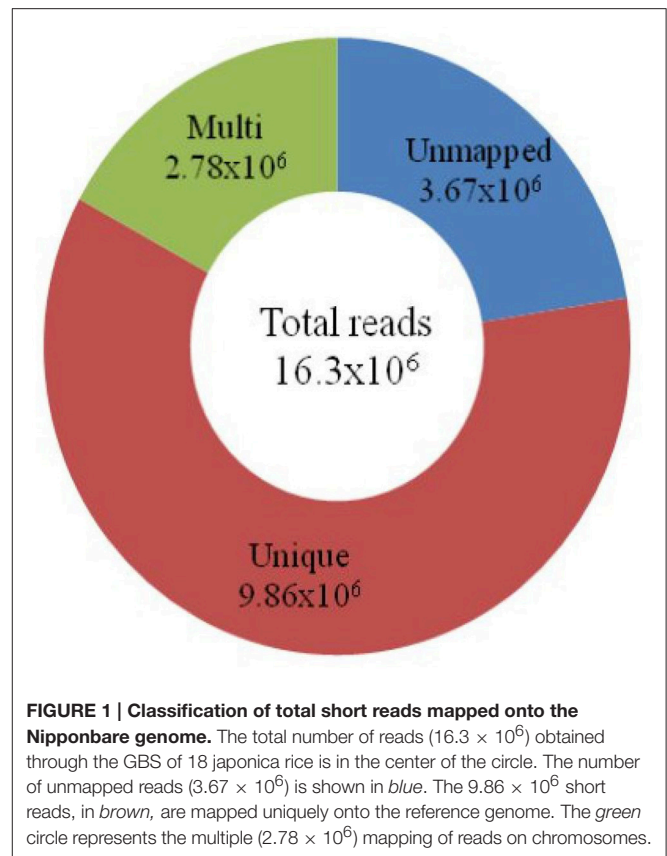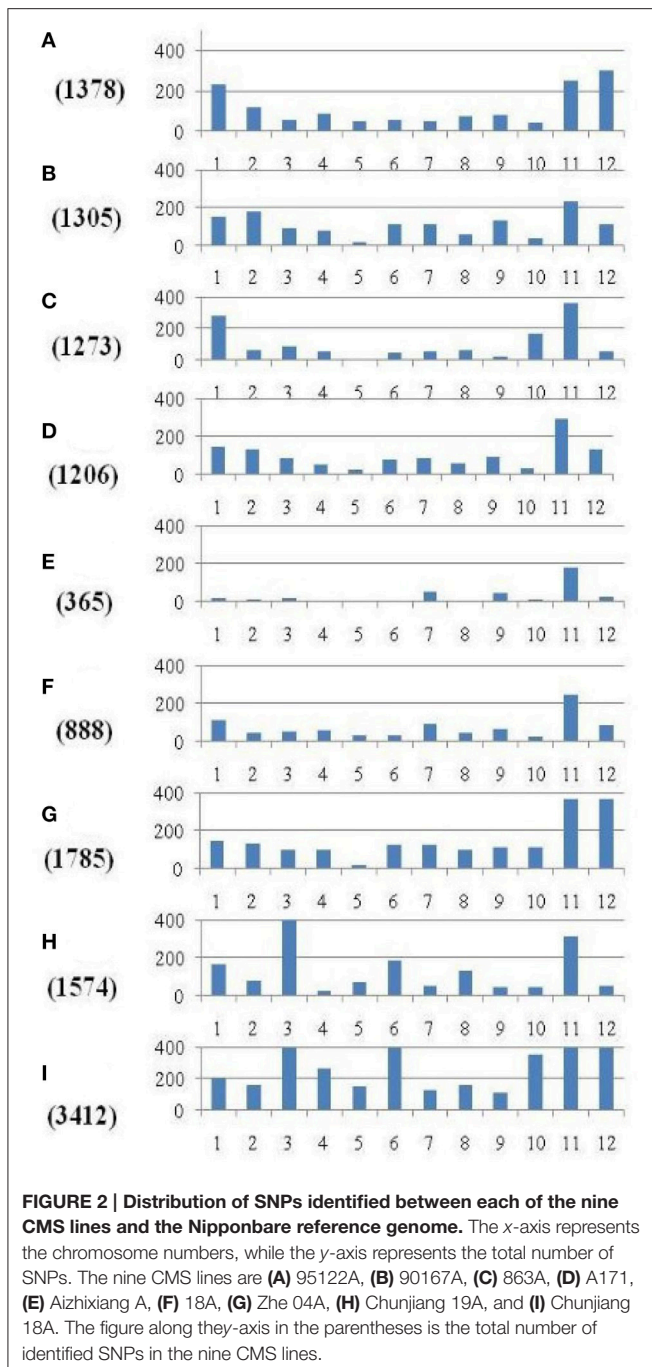

FIGURE 1 | Classification of total short reads mapped onto the Nipponbare genome. The total number of reads (16.3×10^6) obtained through the GBS of 18 japonica rice is in the center of the circle. The number of unmapped reads (3.67×10^6) is shown in *blue*. The 9.86×10^6 short reads, in *brown,* are mapped uniquely onto the reference genome. The *green* circle represents the multiple (2.78×10^6) mapping of reads on chromosomes.

The highest (35.6) SNP density was observed on chromosome 11, whereas the lowest (2.7) was observed on chromosome 4. Such densities were also found to be uneven within the individual chromosomes. In chromosome 2, the region between 5 and 22.2 MB had an unusually high (2796 SNPs) SNP density, whereas the region between 26 and 32.4 MB had a very low (24 SNPs) SNP density. The same results were observed on chromosome 11, where a high (3508 SNPs) and long SNP interval was detected between the 16 and 28.4 MB genomic region (**Figure 5**). A small (120 SNPs) and short interval was also found from the 0.4 to 4 MB region of the same chromosome. On chromosome 11 of the restorer lines, we identified a maximum of 57 high-density SNP sites, with >30 SNPs per 200 kb region.

Annotation of SNPs

The annotation of the Nipponbare rice was used to locate the distribution pattern of the identified SNPs within various genomic regions. Overall, a similar SNP distribution was observed in the CMS and restorer lines. Of the 13,186 SNPs discovered in the nine CMS lines, 7582 (57.5%) were found within intergenic spaces. The other 3212 (24.3%) and 2649 (20%) were in the intron and exon regions, respectively. Among the total SNPs in the coding regions, synonymous were counted as 1155 (8.7%), while non-synonymous were further divided into missense, 1214 (9.2%), and nonsense, 27 (0.2%) (**Table 1**, Supplementary Table 3). Moreover, the detected SNPs of the nine CMS lines were also categorized as transition (C/T and

FIGURE 2 | Distribution of SNPs identified between each of the nine CMS lines and the Nipponbare reference genome. The x-axis represents the chromosome numbers, while the y-axis represents the total number of SNPs. The nine CMS lines are (A) 95122A, (B) 90167A, (C) 863A, (D) A171, (E) Aizhixiang A, (F) 18A, (G) Zhe 04A, (H) Chunjiang 19A, and (I) Chunjiang 18A. The figure along the y-axis in the parentheses is the total number of identified SNPs in the nine CMS lines.

G/A) and transversion (G/T, T/A, A/C, C/G) nucleotide bases (Supplementary Table 5). In our results, the average ratio between the transitions and transversions was 1.4. The number of transition substitutions was significantly higher than that of transversion substitutions in the identified SNPs of all nine CMS lines (**Figure 6**).

Subsequently, the identified SNPs of the restorer lines were also annotated, where 14,970 (58%) were situated in the intergenic region of genome. However, 6226 (24%) and 5377 (20%) were detected within the intron and exon regions (**Table 2**,

Supplementary Table 4). Of the SNPs in the coding regions, 2189 (8.4%) led to synonymous while 2711 (10.5%) led to non-synonymous amino acid changes with a protein-altering effect on genes. The average ratio of the transitions and transversions in the identified SNPs was also 1.4 (Supplementary Table 6). The similar result of higher transition substitutions was also observed in the identified SNPs of all nine restorer lines (**Figure 7**).

Performance of the GCA of Parents for 12 Yield-Related Traits

The GCA performance of the 18 parents of hybrid rice for 12 yield-related traits is summarized in **Table 3**. Among the nine CMS lines, genotype Zhe04 exhibited the highest GCA effect value for plant height, whereas A171 had the highest GCA effect for days to heading and seed width. In terms of the panicle length, the number of panicles per plants and the number of grains per panicle, CMS line 863A was a good general combiner due to the highly significant and positive GCA effect value. CMS 18A recorded the maximum GCA effect value for seed thickness and grain yield per plot, whereas 90167A revealed the highest GCA effect value in terms of seed length and 1000-seed weight. Parent AizhixiangA and Chunjiang18A had the maximum GCA effect in terms of number of panicles per plant and seed setting rate, respectively (**Table 3**).

The GCA assessment of the restorer lines indicated that LR27 recorded the maximum GCA effect value for plant height, whereas Shenhui254 recorded the maximum GCA effect value for days to heading. The restorer C4115 possessed higher GCA effect values for panicle length, seed thickness and seed-setting rate. In terms of seed length, seed width and 1000-seed weight, restorer Yanhui R8 recorded the highest GCA effect value. The parent LC64 exhibited a greater value for the number of spikelets per panicle, whereas LR5 exhibited a greater GCA effect for the number of filled grains per panicle and grain yield per plot. The restorer R4179 revealed the maximum GCA for the number of panicles per plant (**Table 3**).

Genome Wide Association Analysis for the Identification SNP Loci Associated with CA

The main objective of our study was to identify SNP locus/loci of CA *via* association analysis; the discovered 39,001 SNPs at 11,085 genomic positions were integrated with the parental CA of 12 yield-related traits at $P < 0.01$. We revealed a total of 362 SNP locus/loci with the CA of parental traits that caused a positive or negative effect on F_1 trait performances. The overview and detailed information of the identified SNP loci of CA for the following traits are presented in **Figure 8** and Supplementary Table 7.

Plant Height

Fifty-three SNP loci situated on 9 different chromosomes (Chr1 Chr2, Chr3, Chr4, Chr6, Chr7, Chr8, Chr11, and Chr12) displayed significant associations with the CA of plant height (**Table 4**, Supplementary Table 7). Of these associated loci, 18 exhibited a positive and 35 exhibited a negative effect on the CA of plant height. The positively associated loci increased the CA by 6.6%, whereas the negatively associated loci reduced
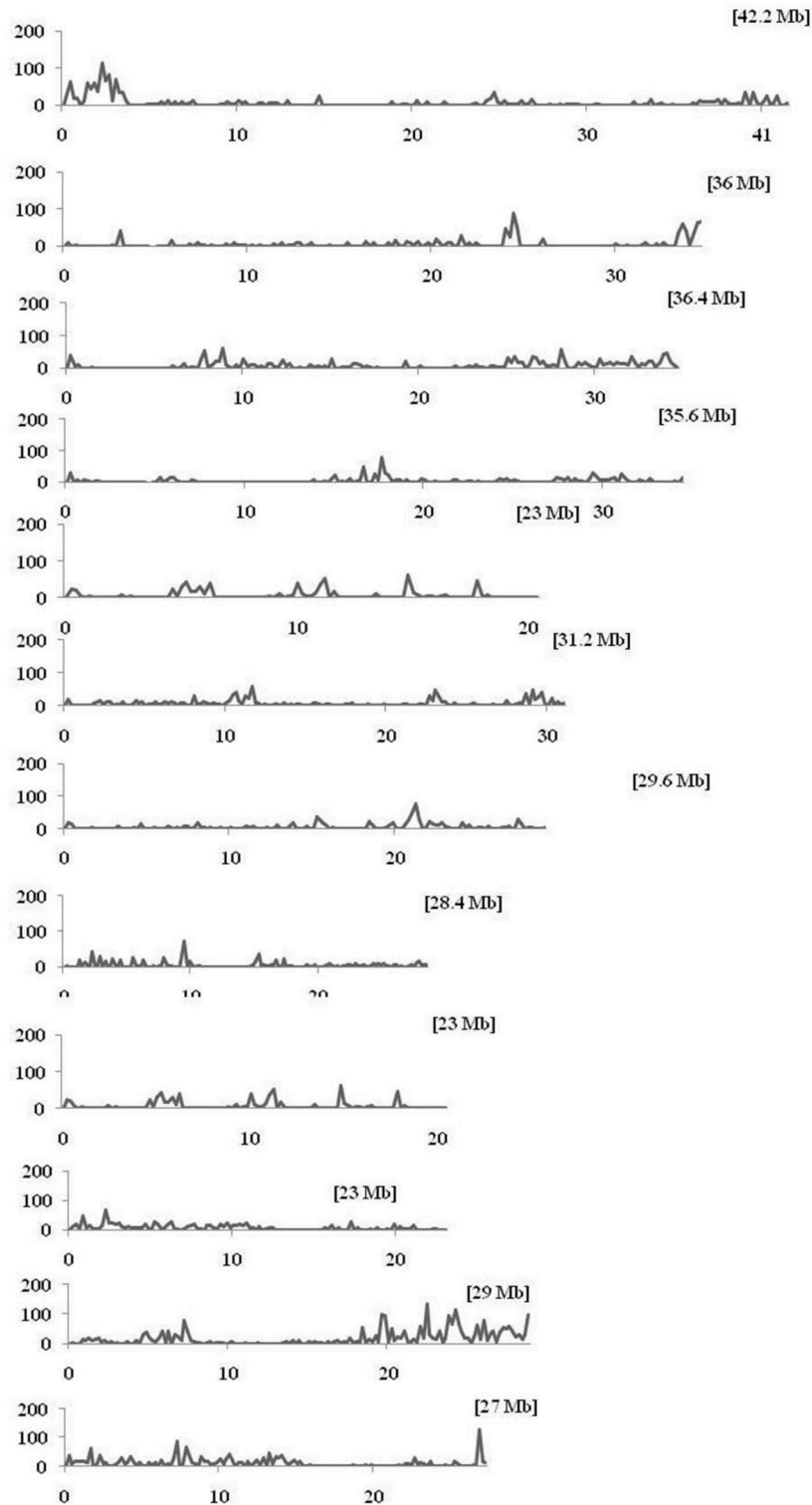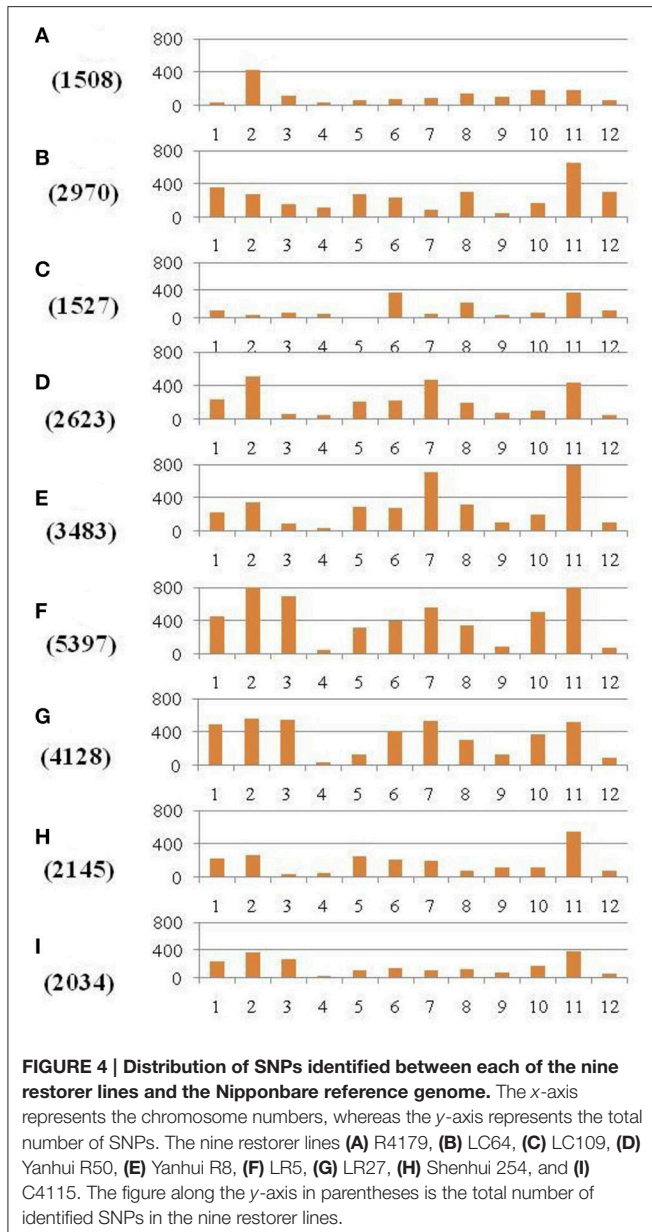
FIGURE 3 | Density distributions of SNPs detected among the nine CMS lines and Nipponbare in the 12 rice chromosomes. The x-axis indicates the physical distance of each chromosome, split into 200-kb windows. The total size of the individual chromosome is shown in the bracket. The y-axis indicates the total number of SNPs.

FIGURE 4 | Distribution of SNPs identified between each of the nine restorer lines and the Nipponbare reference genome. The x-axis represents the chromosome numbers, whereas the y-axis represents the total number of SNPs. The nine restorer lines **(A)** R4179, **(B)** LC64, **(C)** LC109, **(D)** Yanhui R50, **(E)** Yanhui R8, **(F)** LR5, **(G)** LR27, **(H)** Shenhui 254, and **(I)** C4115. The figure along the y-axis in parentheses is the total number of identified SNPs in the nine restorer lines.

the CA by 11%. Among all the parents, the restorer LC64 possessed a maximum of 35 negative CA loci of plant height in its sequence genome (**Table 5**). On an individual basis, SNP loci on chromosomes 11 and 7 with A/G and T/A alleles caused the most prominent positive effect on the CA of the trait, 8.8 and 7.9%, respectively.

Grain Width
Forty-four SNP loci were distributed over 11 different chromosomes and showed significant associations with the CA of grain width (**Table 4**, Supplementary Table 7). Of them, 40 showed an increment (5.8%), whereas 4 exhibited a reduction (3.5%). The CMS Chunjiang 18A had a maximum of 34 favorable CA loci in the genome for the CA of grain width (**Table 5**). On

an individual basis, SNP loci on chromosomes 6 and 5 with A/T and T/C alleles caused the maximum positive effect on the CA of the trait, 11 and 7.6%, respectively.

Grain Length
Twenty-seven SNP loci distributed across 10 different chromosomes revealed significant associations with the parental CA of grain length (**Table 4**, Supplementary Table 7). Among the associated loci, 17 conferred a positive effect of 5%, whereas 10 conferred a negative effect of 8%. For the number of associated SNP loci within the parents, the CMS A171 had a maximum of 13 favorable CA loci for grain length (**Table 5**). On an individual basis, SNP loci on chromosomes 7 and 11 with C/T and G/A alleles caused the maximum positive effect on the CA of the trait, 6 and 5.9%, respectively.

Grain Thickness
Ten SNP loci on 8 different chromosomes (Chr2, Chr3, Chr4, Chr5, Chr6, Chr7, Chr10, and Chr11) revealed significant relationships with the CA of grain thickness (**Table 4**, Supplementary Table 7). The effect of these loci on the CA explained that, 9 loci caused a 7.2% increment, whereas one locus caused a 4.9% reduction in the CA of grain thickness. Among the sequenced genome of the parents, the restorer LC64 exhibited a maximum of 5 positive CA loci (**Table 5**). On an individual basis, SNP loci on chromosomes 7 and 12 with C/G and A/G alleles caused the maximum positive effect on the CA, increasing its value by 16.9 and 7.4%, respectively.

1000-Grain Weight
Nineteen SNP loci situated over 9 different chromosomes (Chr1, Chr2, Chr3, Chr5, Chr6, Chr7, Chr9, Chr11, and Chr12) recorded significant associations with the CA of 1000-grain weight. Of them, 14 explained positive (5.8%), whereas 5 explained negative (5.4%), associations with the CA of the trait (**Table 4**, Supplementary Table 7). The parental genome of CMS A171 contained a maximum of 9 positive CA loci of 1000-grain weight (**Table 5**). On an individual basis, SNP loci on chromosomes 5 and 2 caused the maximum positive effect, increasing the CA trait by 7.8 and 5.5%, respectively.

Grain Yield
Five SNP loci detected on 4 different chromosomes (Chr2, Chr3, Chr9, and Chr11) were found to be significantly associated with the CA of grain yield (**Table 4**, Supplementary Table 7). Among them, one locus had a positive effect of 26.3%, whereas 4 loci had a negative effect of 27.5%. The restorer line LC64 contained a higher number of negative CA loci for grain yield (**Table 5**). Among all the associated SNPs, an SNP locus on chromosome 2 with a G/A allele had a favorable (26.3%) effect on the CA value.

Days to Heading
Twenty-seven SNP loci situated on 9 different chromosomes (Chr1, Chr2, Chr3, Chr5, Chr6, Chr7, Chr8, Chr11, and Chr12) revealed significant associations with the CA of days to heading (**Table 4**, Supplementary Table 7). Among them, 22 had a positive effect of 7.8% on the CA, whereas 5 caused a negative effect of 7.6% on the CA of the trait. We found that the restorer LC64
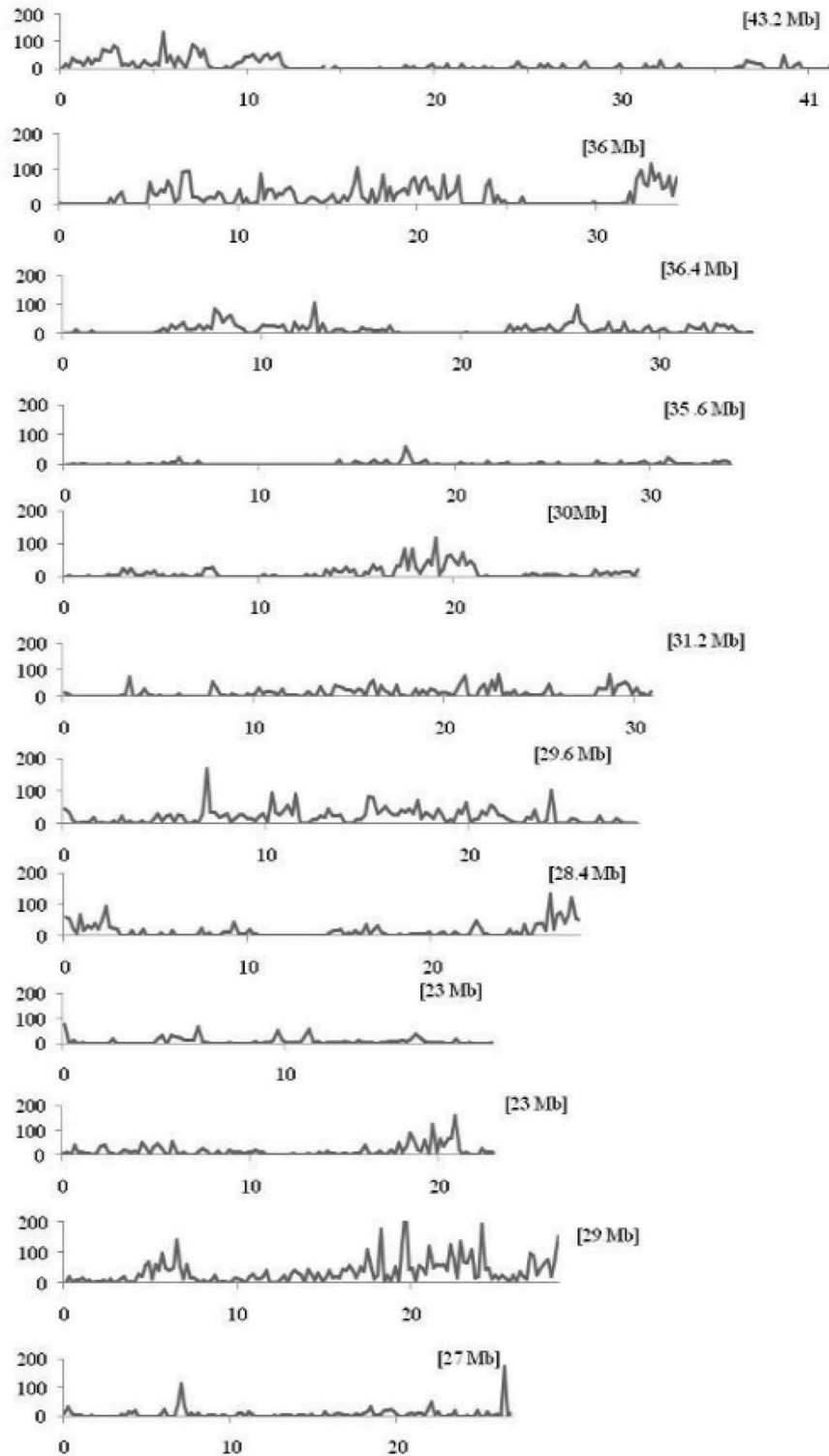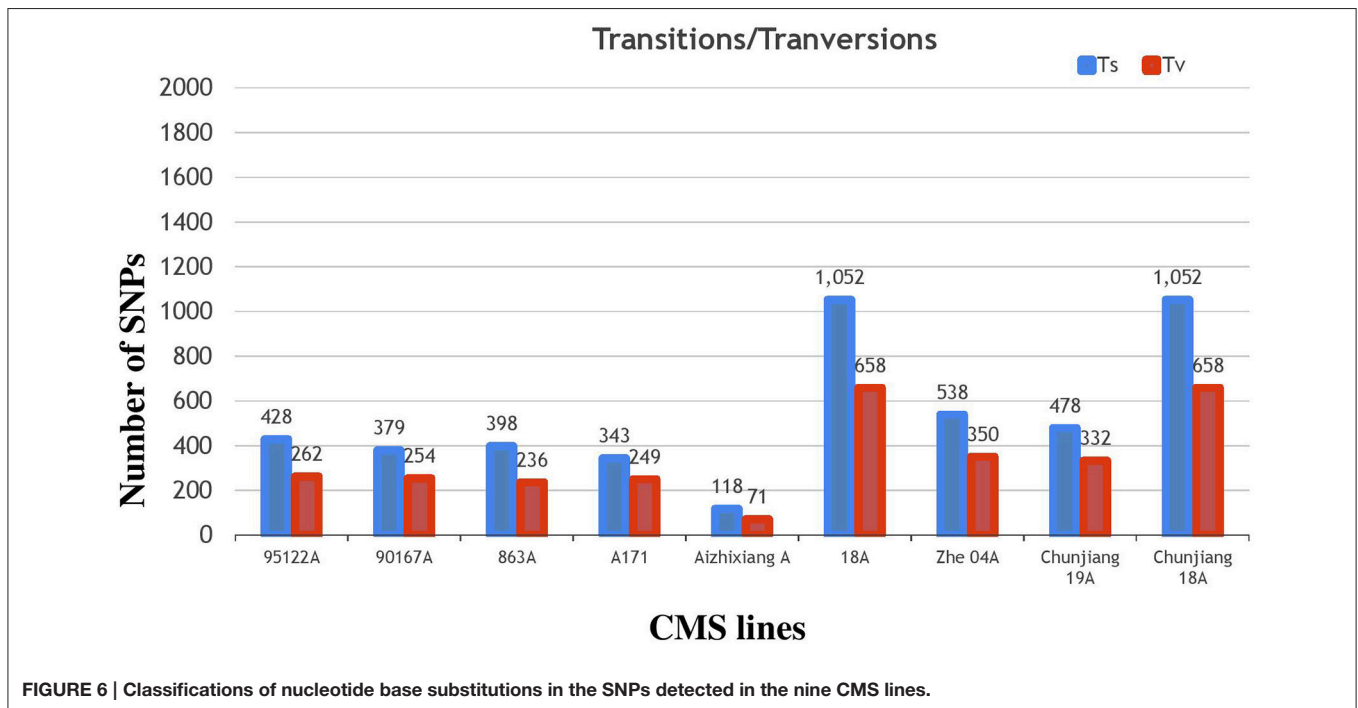
FIGURE 5 | Density distributions of SNPs detected among the nine restorer lines and Nipponbare in the 12 rice chromosomes. The x-axis indicates the physical distance of each chromosome, split into 200-kb windows. The total size of individual chromosomes is shown in the bracket. The y-axis indicates the total number of SNPs.

TABLE 1 | Annotation of single nucleotide polymorphisms (SNPs) between nine CMS lines and the Nipponbare reference genome.

CMS lines	Total	Intergenic	Splice site	Intron	Exon	Synonymous	Non-synonymous	
							Missense	Nonsense
95122A	1,378	793	13	357	276	119	139	–
90167A	1,305	770	19	328	265	121	110	2
863A	1,273	700	18	271	308	130	132	3
A171	1,206	702	12	277	259	120	116	3
Aizhixiang A	365	219	1	84	67	30	29	1
18A	888	516	27	213	169	80	73	2
Zhe 04A	1,785	1,026	20	399	362	167	158	6
Chunjiang 19A	1,574	884	21	430	288	118	149	1
Chunjiang 18A	3,412	1972	49	853	655	270	308	8

FIGURE 6 | Classifications of nucleotide base substitutions in the SNPs detected in the nine CMS lines.

had a maximum number of positive CA loci in days to heading (**Table 5**). On an individual basis, SNP loci on chromosomes 5 and 1 with A/G and G/C alleles contributed the maximum effect to the CA in terms of days to heading, 10.2 and 9.7%, respectively.

Panicle Length

One hundred and ten SNP loci distributed over all 12 chromosomes revealed significant associations with the CA of panicle length. Among them, 77 exhibited a positive (16.2%) contribution, whereas 33 exhibited a negative reduction in the CA (**Table 4**, Supplementary Table 7). For all the sequenced parents, the restorer LR27 had a maximum number of 71 favorable and 7 unfavorable CA loci in the panicle length (**Table 5**). On an individual basis, SNP loci on chromosomes 9 and 3 with A/C and T/C alleles showed the maximum positive effect of 34 and 25%, respectively.
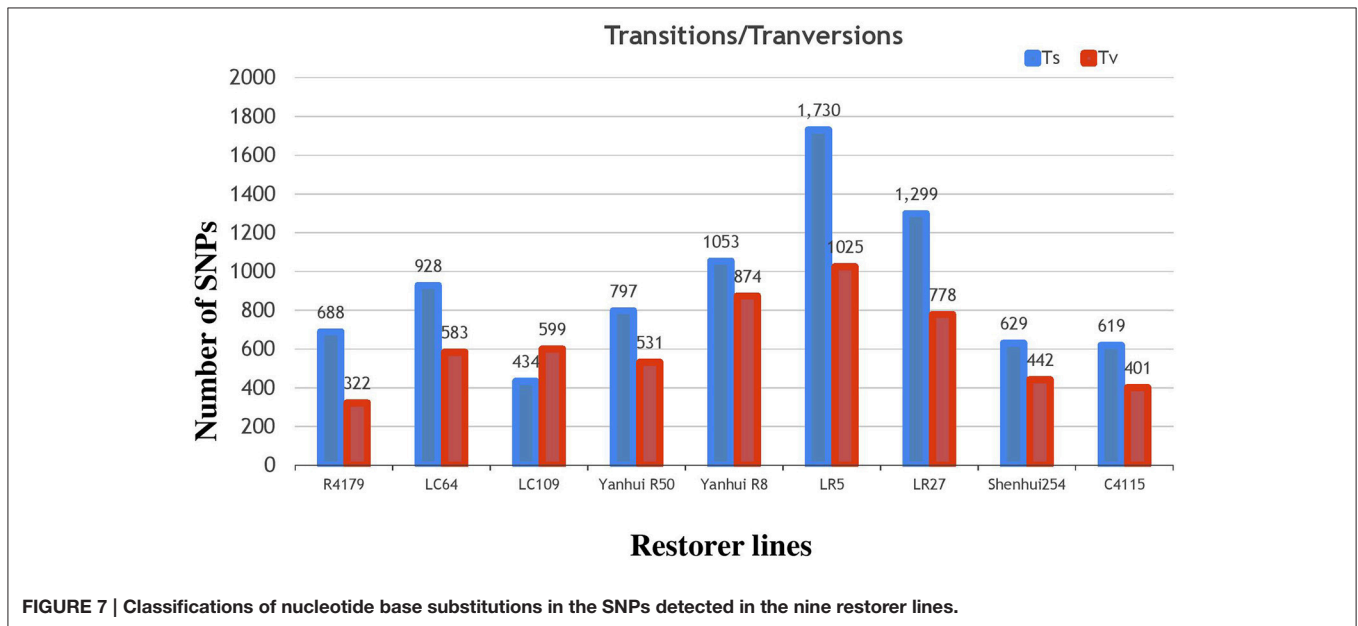
Panicle Number Per Plant

Five SNP loci located on 4 different chromosomes (Chr2, Chr5, Chr9, and Chr11) showed significant associations with the CA of panicle number per plant (**Table 4**, Supplementary Table 7). Of them, 4 had a positive (29.3%), while one locus had a negative (21.9%), effect on the CA of the trait. CMS 18A contained a maximum of 4 negative loci, causing a reduction in the CA value of the panicle number per plant (**Table 5**). Among all the SNP loci, the variants on chromosomes 5 and 9 having G/A and A/T alleles caused the maximum improvement in the CA of 39 and 31%, respectively.

Number of Spikelet's Per Panicle

Fifty-three SNP loci distributed over 10 different chromosomes were found to be significantly associated with the CA of number of spikelet's per panicle (**Table 4**, Supplementary Table 7). Of these, 2 loci had a positive (3.7%) and 51 had a negative

TABLE 2 | Annotation of single nucleotide polymorphisms (SNPs) between nine restorer lines and the Nipponbare reference genome.

Restorer lines	Total	Intergenic	Splice site	Intron	Exon	Synonymous	Non-synonymous	
							Missense	Nonsense
R4179	1,508	850	16	342	355	142	174	1
LC64	2,970	1,774	40	639	622	276	303	3
LC109	1,527	904	14	338	300	130	137	1
Yanhui R50	2,623	1,516	27	604	538	205	369	3
Yanhui R8	3,483	2,057	37	856	696	291	334	4
LR5	5,397	3,080	63	1,372	1,118	441	529	4
LR27	4,128	2,364	44	1,058	828	318	420	1
Shenhui 254	2,145	1,288	21	479	454	195	205	4
C4115	2,034	1,137	20	538	466	191	218	1

FIGURE 7 | Classifications of nucleotide base substitutions in the SNPs detected in the nine restorer lines.

(4.2%) effect on the CA of the trait. The restorer LC64 had a maximum of 47 negative CA loci (**Table 5**). Two associated loci on chromosomes 11 and 2 with A/G and C/T alleles exhibited the maximum positive effect on the CA of 19 and 15%, respectively.

Number of Filled Grains Per Panicle
Seven SNP loci were found to be associated with the CA of number of filled grains per panicle. These loci were distributed on 4 different chromosomes (Chr1, Chr2, Chr4, and Chr9) (**Table 4**, Supplementary Table 7). Of these, one locus contributed positively (20%), whereas the remaining 6 loci negatively affected (26%) the CA of the trait. It is noteworthy that, CMS 18A had a higher (5) number of negative CA loci in its genome of CA (**Table 5**). An SNP locus on chromosome 2 with a C/T allele caused a favorable increment in the CA value of the trait.

Seed Setting Rate
On chromosome 10, only one SNP locus was found to be associated with the CA of seed-setting rate (**Table 4**, Supplementary Table 7). This SNP locus reduced the CA value by 19%. Three restorers, R4179, LC109 and Yanhui R8, contained this one locus in their genomes (**Table 5**).

DISCUSSION

GBS Assay for Genome Wide SNP Discovery
The exploitation of heterosis in rice has increased the global rice-grain yield. This has led to an increase in productivity (growth, size, development, fertility and yield) of the progeny over that of both homozygous parents and is a fascinating phenomenon (Hochholdinger and Hoecker, 2007). The application of heterosis has been applied not only to rice but also to other crops such as tomatoes (Williams and Gilbert, 1960), maize (Shull, 1908), wheat (Singh et al., 2004), and soybeans (Pandini et al., 2002). In plants, heterosis is considered to be a complex trait that involves the presence of potential genetic distances, combining abilities and heterotic patterns of the parents traits (Beck et al., 1990). Rice breeders emphasized that heterosis in F_1s is caused by the presence of combining ability alleles, that are present at different

TABLE 3 | GCA effect values of 18 parental lines for 12 yield-related traits.

Parental lines	PH	GW	GL	GT	1000-GW	GYPP	DH	PL	NSPP	NPPP	NFGPP	SSR
CMS LINES												
95122A	3.94a	−0.12b	0.13a	−0.07c	−0.79cd	−0.23b	3.64b	−0.56cd	−1.77bc	−0.59b	5.85a	−0.24bc
90167A	−1.49bc	0.12a	0.18a	0.05a	1.29a	0.21ab	−4.07d	0.57c	16.65a	−0.58b	14.07a	−0.26c
863A	3.95a	0.01a	−0.01a	0.00b	−0.25c	−0.57b	3.12bc	4.14a	23.93a	−0.86b	16.07a	−0.26c
A171	−11.98d	0.19a	−0.03a	0.05ab	0.84ab	−0.48b	7.84a	−3.99e	−43.27c	0.38ab	−23.45b	−0.19b
Aizhixiang A	−0.23b	−0.13b	−0.16b	−0.01b	−0.68c	0.46a	−6.10e	2.50b	−40.83c	1.84a	−22.46b	−0.22b
18A	0.81b	0.05a	−0.04ab	0.08a	0.15bc	0.80a	1.37c	−1.70de	6.68a	0.89a	7.09a	−0.25c
Zhe 04A	4.77a	0.02a	−0.21b	−0.03bc	0.15b	−0.10b	7.23ab	−1.01d	19.57a	−1.15b	3.95a	−0.30c
Chunjiang 19A	2.61ab	−0.05ab	0.13a	−0.05c	−1.30d	−0.03b	−7.76e	0.15c	18.06a	−0.04b	4.31a	−0.30c
Chunjiang 18A	−2.386cd	−0.10b	0.02a	−0.01b	0.58b	−0.04b	−5.269de	−0.10c	0.95ab	0.10b	−5.45ab	2.04a
RESTORER LINES												
R4179	−3.76c	−0.20b	−0.05b	−0.02c	−0.69de	0.11b	−1.63d	−2.57c	−12.92c	1.52a	−16.05b	0.01b
LC64	−4.46cd	−0.06ab	−0.47c	−0.06c	−1.71e	0.74ab	−6.93e	0.84b	56.66a	0.41a	19.24a	−0.14d
LC109	−8.21d	0.034a	−0.38c	0.06ab	0.30bc	−0.78c	−6.88e	−1.43c	−69.99d	−0.18b	−67.17c	−0.10d
Yanhui R50	−0.28b	0.11a	0.34a	0.02b	0.03c	−0.73c	4.56b	0.66b	11.30b	−1.75b	14.71ab	0.01bc
Yanhui R8	−0.93bc	0.19a	0.51a	0.06a	2.08a	−0.00b	4.00bc	0.28bc	−10.18c	−0.79b	−1.34b	0.05a
LR5	5.22a	−0.03a	−0.04b	−0.04c	−0.00c	1.03a	0.53cd	0.55b	21.54b	1.35a	33.09a	0.05ab
LR27	6.06a	−0.10b	0.09b	−0.06c	−0.32c	0.00b	1.23c	1.44ab	20.83b	−0.05ab	20.64a	0.02b
Shenhui254	4.98ab	0.06a	−0.27bc	0.08a	0.72b	−0.11b	8.39a	−2.12c	−18.31cd	1.20a	−24.52bc	−0.03cd
C4115	1.39b	−0.00a	0.27ab	−0.04c	−0.40cd	−0.23bc	−3.29de	2.34a	1.08bc	−1.71b	21.41a	0.11a

The CMS and restorer line values followed by different letters are significantly different at P < 0.01. PH, plant height; GW, grain width; GL, grain length; GT, grain thickness; 1000-GW, 1000-grain weight; DH, days to heading; PL, panicle length; NPPP, number of panicles per plant; NSPP, number of spikelets per panicle; NFGPP, number of filled grains per panicle; SSR, seed setting rate; GYPP, grain yield per plot.

loci of crossing parents (Liu et al., 2002). They described how the combination of different alleles at a specific locus of crossing parents creates heterozygosity that results in the heterosis of offspring.

One part of our present study describes genome-wide SNP discovery in japonica rice. A large number of rice cultivars have been sequenced for SNP discovery to perform genetic and genomic experiments. However, only a few studies have been conducted on japonica rice (Nagasaki et al., 2010). Rice breeders have concluded that SNP discovery in japonica rice is challenging compared to that in indica rice, due to its close genetic relatedness and lower diversity (Glaszmann, 1987; Yang et al., 1994; Gao et al., 2005; Negrao et al., 2008). In our recent efforts, 18 parents of japonica rice were sequenced for SNP discovery, following optimized genotyping with a sequencing method. The use of two restriction enzymes based on a sequencing method of GBS is the most convenient, cost-effective and easy approach to the SNP discovery of large genome crop plants, including rice (Furuta et al., 2016). Nevertheless, this method also has the drawback of generating a large amount of missing data (Davey et al., 2011; Kim et al., 2016).

Here, the successful mapping of 60.4% of short reads onto the reference genome revealed 39,001 SNPs at 11,085 positions. The unique mapping of short reads plays a crucial role in the determination of SNPs (Nielsen et al., 2011). However, un-mapping reduced the chances of SNP discovery, which eventually occurs due to genomic deletions during the process of sequencing (Arai-Kichise et al., 2011). In our study, the number of SNPs among the parents varied. We observed higher polymorphisms

in the nine restorer lines compared to the CMS lines. Of the 18 sequenced parents, the CMS Aizhixiang A reported a minimum number of SNPs. An in-depth analysis of the sequence genome of Aizhixiang A indicated that 70% of its sequence genome was similar to that of the Nipponbare reference. Furthermore, 25% of the missing data were also found, which further affected the chances for SNP discovery. We also found that the SNPs of CMS and the restorer lines were non-randomly distributed across 12 chromosomes and revealed several SNP-rich and SNP-poor intervals. Such extended SNP regions were similar to those reported in rice, wheat, mei (*Prunus mume* sieb. et zucc) and peach cultivars (Ravel et al., 2006; Subbaiyan et al., 2012; Fresnedo-Ramírez et al., 2013; Sun et al., 2013). In our study, there were also non-uniform patterns of SNPs within individual chromosomes, particularly at the terminal region of chromosome. We observed this situation at the terminal region of chromosome 11 for both the CMS and restorer lines. This may possibly arise due to the low variant rate at the centromere region of the chromosome, where recombination are typically more rare than at the chromosome terminal (Guo W. et al., 2007).

Annotation of SNPs
In the sequenced genome of animals and plants, SNPs do not present equally throughout the genome. Some parts of the genome revealed more SNPs than others. The discovered SNPs in our study were annotated according to their positions in the genome. The annotation of identified SNPs indicates that more than half (57%) of the SNPs are within the intergenic region, whereas 24.4% of the SNPs have been found within the intron
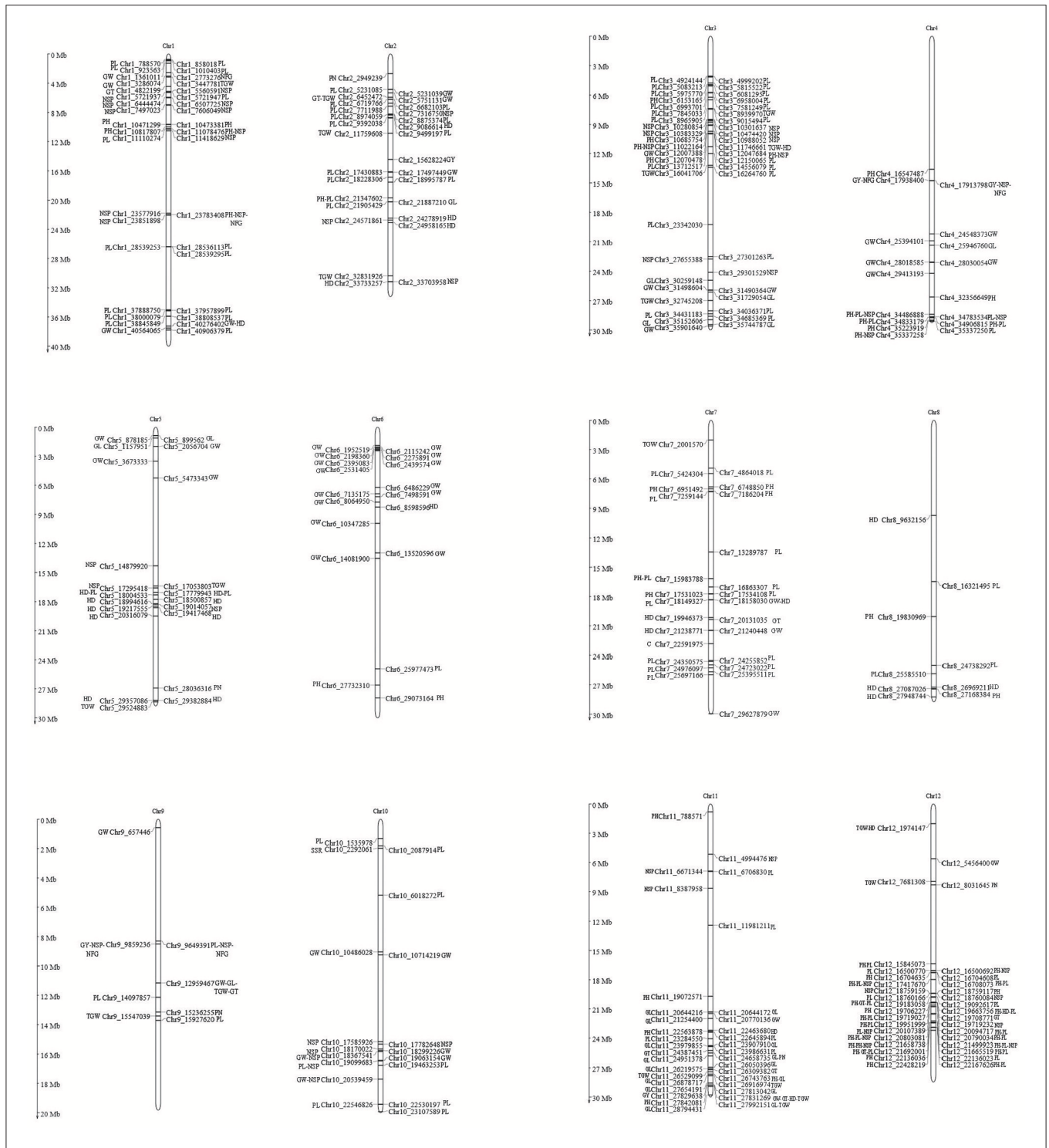
FIGURE 8 | Graphical representations of all the associated 362 SNP loci of CA and their corresponding chromosomal positions. SNP positions associated with the CA of traits are shown by alphabet, PH, plant height; GW, grain width; GL, grain length; GT, grain thickness; TGW, 1000 grain weight; GY, grain yield; HD, heading days; PL, panicle length; PN, panicle number per plant; NSP, number of spikelets per panicle; NFG, number of filled grains per panicle; SSR, seed setting rate.

region of the genome. Only 20.4% of the SNPs occurred inside the exon regions, with 8.6% synonymous and 9.9% non-synonymous in nature. The number of SNPs in the exon region was 37% less than that in the intergenic regions. This is possibly, because of the higher frequency of polymorphisms evolving in intergenic spaces rather than coding sequences (Guo X. et al., 2007). However,

TABLE 4 | Number of SNP loci significantly associated with CA and percentage of increase/decrease in the trait values.

Traits	Number of associated SNP loci	Heterozygous group		Homozygous group		Average percent heterosis at the heterozygous locus group[a]
		Number of total F_1 crosses	Average trait value of F_1 crosses	Number of total F_1 crosses	Average trait value of F_1 crosses	
PH	53	871	107.4	3,422	110	−4.9
GW	44	593	2.2	2,971	2.1	4.9
GL	27	544	8.3	1,643	8.2	−0.2
GT	10	189	3.7	621	3.5	6.1
1000 GW	19	413	26.2	1121	25.5	2.8
GYPP	6	84	25.5	321	31	−16.7
DH	27	808	110.6	1,379	105.4	10.2
PL	110	1,849	23.6	7,061	22.4	5.9
NPPP	5	68	11.4	337	9.6	19.2
NSPP	53	914	229	3,379	281.9	−17.7
NFGPP	7	114	177.4	453	222	−25.2
SSR	1	10	0.65	71	0.8	−19.1

[a]Heterosis at heterozygous locus group (%) = (average of trait value in F1 heterozygous locus group—average of trait value in F1 homozygous group)/average of trait value in F1 homozygous group ×100. PH, plant height; GW, grain width; GL, grain length; GT, grain thickness; 1000-GW, 1000-grain weight; GYPP, grain yield per plot; DH, days to heading; PL, panicle length; NPPP, number of panicles per plant; NSPP, number of spikelets per panicle; NFGPP, number of filled grains per panicle; SSR, seed setting rate.

TABLE 5 | Number of associated SNPs of CA in the parental genomes for 12 yield-related traits.

CMS lines	PH	GW	GL	GT	1000-GW	GYPP	DH	PL	NSPP	NPPP	NFGPP	SSR
95122A	3	2	1	1	4	2	2	1	2	2	2	–
90167A	–	2	3	–	2	2	1	2	4	1	4	–
863A	1	4	11	2	3	3	5	5	2	1	3	–
A171	–	3	13	3	9	3	4	2	1	1	2	–
Aizhixiang A	2	–	2	–	–	1	–	3	6	–	3	–
18A	2	1	1	–	3	3	2	4	5	2	5	–
Zhe 04A	2	1	7	1	3	2	4	3	2	–	1	1
Chunjiang 19A	3	–	10	–	3	–	4	7	2	–	1	–
Chunjiang 18A	4	34	12	1	3	3	5	24	2	–	2	1
Restorer lines												
R4179	1	8	8	2	4	4	3	13	4	–	2	–
LC64	35	5	4	5	3	4	15	35	47	–	4	–
LC109	6	2	7	–	5	3	1	6	2	–	2	1
Yanhui R50	15	7	5	2	4	4	14	11	7	–	3	–
Yanhui R8	13	9	3	1	3	3	14	16	7	–	3	–
LR5	17	8	4	2	7	4	14	62	33	2	3	–
LR27	13	6	6	4	5	2	14	78	14	–	3	–
Shenhui254	4	7	5	–	4	–	4	6	3	1	3	–
C4115	8	1	–	1	5	–	7	10	2	–	1	–

Symbol (–) indicates that none of the associated SNP loci of CA are present in the genome of parents. PH, plant height; GW, grain width; GL, grain length; GT, grain thickness; 1000-GW, 1000-grain weight; GYPP, grain yield per plot; DH, days to heading; PL, panicle length; NPPP, number of panicles per plant; NSPP, number of spikelets per panicle; NFGPP, number of filled grains per panicle; SSR, seed setting rate.

the SNPs situated inside intergenic regions are also prominent, where there may have been the possibility of the presence of some functional alleles or genes (Salvi et al., 2007). In our study, the ratio of transitions to transversions was 1.4. This ratio was in line with former sequencing experiments in flax (Kumar et al., 2012).

Evaluation of Parents for GCA

The concept of combining ability (GCA and SCA) was developed by Sprague and Tatum in 1942. They characterized the breeding values of inbred lines by testing and evaluating them by combining ability evaluation trails (Sprague and Tatum, 1942). In every hybrid-breeding program, a breeder's goal is to select

parents with good general combining ability and crosses that have high specific combining ability in their traits. From a genetic point of view, GCA evolved under additive and additive × additive gene action, whereas SCA evolved under non-additive gene action. The GCA of inbred parents is considered more important than the SCA, in which the parents were involved in random mating. The GCA evaluation of 18 parents of japonica hybrid rice indicated that most of the CMS lines exhibited maximum and positive GCA effect values for their studied traits. Previously, the same results of good general combining ability in the CMS lines were also reported (Singh et al., 1996). In contrast, most of the restorer lines also revealed greater GCA effect for their traits. The evidence of a higher GCA of restorer lines was also found in earlier studies (Rogbell et al., 1998).

SNP-CA Associations

With the discovery of genome-wide SNPs, several SNP-trait associations were reported in plant species, including maize, wheat, cotton, soybean, barley and *Arabidopsis thaliana* (Atwell et al., 2010; Yang et al., 2010; Pasam et al., 2012; Hu X. et al., 2015; Zhang et al., 2015; Su et al., 2016). In the past, most of the association studies on rice have focused on the improvement of yield-related traits (Huang et al., 2010; Zhao et al., 2011). Very few SNP associations with CA of traits have been performed. Here, we studied the heterozygous combinations of nucleotide bases (A-T/G-C) at a locus and observed its associations with the CA of traits. Moreover, we worked directly with the heterozygous association group and revealed 362 CA loci of 12 yield-related traits of japonica rice. For those results, each of the CMS and restorer lines contained a number of favorable or unfavorable SNP locus/loci of CA in the genome that caused a positive or negative effect on the traits. Few of the associated SNP loci were found to be co-associated with the CA of more than one yield-related trait. Among the CMS lines, 18A had a maximum number of negative SNP loci of CA for panicle and grain-related traits; whereas CMS A171 had a maximum number of positive SNP loci of CA for grain thickness and day to heading. For the restorer lines, LC64 and LR27 contained maximum numbers of positive SNP loci of CA for most of the studied traits. Previously, in maize, several favorable and unfavorable marker loci of CA were detected for five yield-contributing traits (Qi et al., 2013). Of these, the positive improved, while the negative decreased, the CA of the traits. Similarly, Qu et al. identified the CA loci of agronomical traits in rice by using QTL mapping of BCRILs (Qu et al., 2012). Moreover, all of the identified and associated SNP loci in our study were spread genome wide, with some on chromosomes 2, 5, 7, 9, and 11 exhibiting a maximum positive effect for the CA of traits.

The correct identification of SNPs based on annotation results made them informative. Previous studies have explained that the SNPs within the coding region can alter the promoter activities of gene expression, transcription and translation capability (LeVan et al., 2001; Alonso-Blanco et al., 2005; Ng and Henikoff, 2006; Shastry, 2009; Batley, 2015). Such identification and characterization of gene-based SNPs could be directly utilized in any breeding program regarding crop improvement (Wang et al., 2008; Huang et al., 2010). Among

the 362 SNP loci detected in this study, 32 SNP loci for CA of seven traits were situated within the reported genes (Supplementary Table 8). For CA of plant height, we found 7 SNP loci within previously identified genes. These genes encode proteins for photosynthesis, hormonal activity, lipid binding and transporter activity. Of the seven reported genes for plant height, *Os03g0309200* encodes a proteins causing significant effect on photosynthesis activity, flowering time, and also have strong effects on stem and inter-node elongation in rice. Among four SNP loci for CA of grain width detected within the characterized genes, *Os06g0225300* is directly involved in improving rice architecture for high yield. Overall, the biological function of most of the reported genes were involved in photosynthesis, flower and embryo development, post-embryonic development, multi-cellular organismal development, cellular processes and metabolic processes. Similarly, the molecular function of these genes revealed DNA binding, lipid binding, nucleotide binding, protein binding, DNA binding transcription factor activity, catalytic activity and kinase activity. Such SNP loci situated inside the genes could be used as a candidate marker locus for tagging genomic regions responsible for improving CA of yield traits.

In conclusion, we present the identification of 215 positive and 147 negative coding-based SNP loci of CA for heterosis breeding in japonica rice. Our study implies that the identified SNP loci of CA will increase the selection efficiency of rice breeders in the proper selection of inbred parents. These insights into rice will be productive for future work to improve the CA of parents for yield-related traits and develop superior japonica hybrids by utilizing breeding designs. We expect that, the parental lines with more positive SNP loci of CA and less negative SNP loci of CA will be useful in hybrid breeding, because they have much higher potential for CA improvement for the traits.

AUTHOR CONTRIBUTIONS

DH and IZ conceived the idea and designed the experiment; IZ, WT, EL, SK, HW, and EM contributed to the data collection; IZ analyzed the data and wrote the paper.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2017.00143/full#supplementary-material

REFERENCES

Alonso-Blanco, C., Mendez-Vigo, B., and Koornneef, M. (2005). From phenotypic to molecular polymorphisms involved in naturally occurring variation of plant development. *Int. J. Dev. Biol.* 49, 717–732. doi: 10.1387/ijdb.051994ca

Arai-Kichise, Y., Shiwa, Y., Nagasaki, H., Ebana, K., Yoshikawa, H., Yano, M., et al. (2011). Discovery of genome-wide DNA polymorphisms in a landrace cultivar of japonica rice by whole-genome sequencing. *Plant Cell Physiol.* 52, 274–282. doi: 10.1093/pcp/pcr003

Asano, K., Miyao, A., Hirochika, H., Kitano, H., Matsuoka, M., and Ashikari, M. (2009). SSD1, which encodes a plant-specific novel protein, controls plant elongation by regulating cell division in rice. *Proc. Jpn. Acad. B Phys. Biol. Sci.* 86, 265–273. doi: 10.2183/pjab.86.265

Asano, K., Miyao, A., Kitano, H., Matsuoka, M., Ashikari, M., and Hirochika, H. (2005). Mapping of the SWORD SHAPE DWARF1 Gene, SSD1 in *Rice*. Available online at: http://shigen.nig.ac.jp

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., et al. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465, 627–631. doi: 10.1038/nature08800

Bañuelos, M. A., Garciadeblas, B., Cubero, B., and Rodríguez-Navarro, A. (2002). Inventory and functional characterization of the HAK potassium transporters of rice. *Plant Physiol.* 130, 784–795. doi: 10.1104/pp.007781

Batley, J. (2015). *Plant Genotyping: Methods and Protocols*. New York, NY: Humana Press.

Beck, D., Vasal, S., and Crossa, J. (1990). Heterosis and combining ability of CIMMYT's tropical early and intermediate maturity maize (*Zea mays* L.) germplasm. *Maydica* 35, 279–285.

Bertan, I., Carvalho, F., and Oliveira, A. D. (2007). Parental selection strategies in plant breeding programs. *J. Crop Sci. Biotechnol.* 10, 211–222.

Cao, L., and Zhan, X. (2014). *Chinese Experiences in Breeding Three-Line, Two-Line and Super Hybrid Rice*. Rijeka: INTECH. doi: 10.5772/56821

Chen, W. R., Feng, Y., and Chao, Y. E. (2008). Genomic analysis and expression pattern of OsZIP1, OsZIP3, and OsZIP4 in two rice (*Oryza sativa* L.) genotypes with different zinc efficiency. *Russ. J. Plant Physiol.* 55, 400–409. doi: 10.1134/S1021443708030175

Cheng, S., Cao, L., Yang, S., and Zhai, H. (2004). Forty years' development of hybrid rice: China's experience. *Rice Sci.* 11, 225–230.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., and Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12, 499–510. doi: 10.1038/nrg3012

Fresnedo-Ramírez, J., Martínez-García, P. J., Parfitt, D. E., Crisosto, C. H., and Gradziel, T. M. (2013). Heterogeneity in the entire genome for three genotypes of peach [*Prunus persica* (L.) Batsch] as distinguished from sequence analysis of genomic variants. *BMC Genomics* 14:750. doi: 10.1186/1471-2164-14-750

Fu, F. F., and Xue, H. W. (2010). Coexpression analysis identifies Rice Starch Regulator1, a rice AP2/EREBP family transcription factor, as a novel rice starch biosynthesis regulator. *Plant Physiol.* 154, 927–938. doi: 10.1104/pp.110.159517

Furuta, T., Ashikari, M., Jena, K. K., Doi, K., and Reuscher, S. (2016). Adapting genotyping-by-sequencing for rice F2 populations. *bioRxiv* 055798. doi: 10.1101/055798

Gao, H., Jin, M., Zheng, X.-M., Chen, J., Yuan, D., Xin, Y., et al. (2014). Days to heading 7, a major quantitative locus determining photoperiod sensitivity and regional adaptation in rice. *Proc. Natl. Acad. Sci. U.S.A.* 111, 16337–16342. doi: 10.1073/pnas.1418204111

Gao, L.-Z., Zhang, C.-H., Chang, L.-P., Jia, J.-Z., Qiu, Z.-E., and Dong, Y.-S. (2005). Microsatellite diversity within *Oryza sativa* with emphasis on indica–japonica divergence. *Genet. Res.* 85, 1–14. doi: 10.1017/S0016672304007293

Garg, R., Jhanwar, S., Tyagi, A. K., and Jain, M. (2010). Genome-wide survey and expression analysis suggest diverse roles of glutaredoxin gene family members during development and response to various stimuli in rice. *DNA Res.* 17, 353–367. doi: 10.1093/dnares/dsq023

Glaszmann, J. (1987). Isozymes and classification of Asian rice varieties. *Theor. Appl. Genet.* 74, 21–30. doi: 10.1007/BF00290078

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q., et al. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* 9:e90346. doi: 10.1371/journal.pone.0090346

Gomez, K. A., and Gomez, A. A. (1984). *Statistical Procedures for Agricultural Research*. New York, NY: John Wiley & Sons.

Guo, W., Cai, C., Wang, C., Han, Z., Song, X., Wang, K., et al. (2007). A microsatellite-based, gene-rich linkage map reveals genome structure, function and evolution in Gossypium. *Genetics* 176, 527–541. doi: 10.1534/genetics.107.070375

Guo, X., Wang, Y., Keightley, P. D., and Fan, L. (2007). Patterns of selective constraints in noncoding DNA of rice. *BMC Evol. Biol.* 7:208. doi: 10.1186/1471-2148-7-208

Hochholdinger, F., and Hoecker, N. (2007). Towards the molecular basis of heterosis. *Trends Plant Sci.* 12, 427–432. doi: 10.1016/j.tplants.2007.08.005

Hu, X., Ren, J., Ren, X., Huang, S., Sabiel, S. A., Luo, M., et al. (2015). Association of agronomic traits with SNP markers in durum wheat (*Triticum turgidum* L. durum (Desf.)). *PLoS ONE* 10:e0130854. doi: 10.1371/journal.pone.0130854

Hu, Y., Liang, W., Yin, C., Yang, X., Ping, B., Li, A., et al. (2015). Interactions of OsMADS1 with floral homeotic genes in rice flower development. *Mol. Plant* 8, 1366–1384. doi: 10.1016/j.molp.2015.04.009

Huang, C. F., Yamaji, N., Mitani, N., Yano, M., Nagamura, Y., and Ma, J. F. (2009). A bacterial-type ABC transporter is involved in aluminum tolerance in rice. *Plant Cell* 21, 655–667. doi: 10.1105/tpc.108.064543

Huang, C. F., Yamaji, N., Ono, K., and Ma, J. F. (2012). A leucine-rich repeat receptor-like kinase gene is involved in the specification of outer cell layers in rice roots. *Plant J.* 69, 565–576. doi: 10.1111/j.1365-313X.2011.04824.x

Huang, C. L., Hung, C. Y., Chiang, Y. C., Hwang, C. C., Hsu, T. W., Huang, C. C., et al. (2012). Footprints of natural and artificial selection for photoperiod pathway genes in Oryza. *Plant J.* 70, 769–782. doi: 10.1111/j.1365-313X.2012.04915.x

Huang, J., Qi, H., Feng, X., Huang, Y., Zhu, L., and Yue, B. (2013). General combining ability of most yield-related traits had a genetic basis different from their corresponding traits per se in a set of maize introgression lines. *Genetica* 141, 453–461. doi: 10.1007/s10709-013-9744-3

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42, 961–967. doi: 10.1038/ng.695

Huang, Y. C., Huang, W. L., Hong, C. Y., Lur, H. S., and Chang, M. C. (2012). Comprehensive analysis of differentially expressed rice actin depolymerizing factor gene family and heterologous overexpression of OsADF3 confers *Arabidopsis thaliana* drought tolerance. *Rice* 5, 1–14. doi: 10.1186/1939-8433-5-33

İpek, A., Yılmaz, K., Sıkıcı, P., Tangu, N. A., Öz, A. T., Bayraktar, M., et al. (2016). SNP Discovery by GBS in Olive and the construction of a high-density genetic linkage map. *Biochem. Genet.* 54, 313–325. doi: 10.1007/s10528-016-9721-5

Ishida, S., Morita, K. I., Kishine, M., Takabayashi, A., Murakami, R., Takeda, S., et al. (2011). Allocation of Absorbed Light Energy in PSII to Thermal Dissipations in the Presence or Absence of PsbS Subunits of Rice. *Plant Cell Physiol.* 52, 1822–1831. doi: 10.1093/pcp/pcr119

Ishikawa, R., Aoki, M., Kurotani, K., Yokoi, S., Shinomura, T., Takano, M., et al. (2011). Phytochrome B regulates Heading date 1 (Hd1)-mediated expression of rice florigen Hd3a and critical day length in rice. *Mol. Genet. Genomics* 285, 461–470. doi: 10.1007/s00438-011-0621-4

Itoh, J., Hibara, K., Kojima, M., Sakakibara, H., and Nagato, Y. (2012). Rice DECUSSATE controls phyllotaxy by affecting the cytokinin signaling pathway. *Plant J.* 72, 869–881. doi: 10.1111/j.1365-313x.2012.05123.x

Iwamoto, M., Kiyota, S., Hanada, A., Yamaguchi, S., and Takano, M. (2011). The multiple contributions of phytochromes to the control of internode elongation in rice. *Plant Physiol.* 157, 1187–1195. doi: 10.1104/pp.111.184861

Jiang, G.-L. (2013). "Molecular markers and marker-assisted breeding in plants," in *Plant Breeding from Laboratories to Fields*, ed S. B. Andersen (Rijeka: INTECH), 45–83. doi: 10.5772/52583

Kim, C., Guo, H., Kong, W., Chandnani, R., Shuang, L.-S., and Paterson, A. H. (2016). Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant Sci.* 242, 14–22. doi: 10.1016/j.plantsci.2015.04.016

Kim, Y. S., Kim, I. S., Bae, M. J., Choe, Y. H., Kim, Y. H., Park, H. M., et al. (2013). Homologous expression of cytosolic dehydroascorbate reductase increases

grain yield and biomass under paddy field conditions in transgenic rice (*Oryza sativa* L. japonica). *Planta* 237, 1613–1625. doi: 10.1007/s00425-013-1862-8

Kujur, A., Bajaj, D., Upadhyaya, H. D., Das, S., Ranjan, R., Shree, T., et al. (2015). Employing genome-wide SNP discovery and genotyping strategy to extrapolate the natural allelic diversity and domestication patterns in chickpea. *Front. Plant Sci.* 6:162. doi: 10.3389/fpls.2015.00162

Kumar, S., You, F. M., and Cloutier, S. (2012). Genome wide SNP discovery in flax through next generation sequencing of reduced representation libraries. *BMC Genomics* 13:684. doi: 10.1186/1471-2164-13-684

Kwok, P. Y. (2001). Methods for genotyping single nucleotide polymorphisms. *Annu. Rev. Genomics Hum. Genet.* 2, 235–258. doi: 10.1146/annurev.genom.2.1.235

Langmead, B. (2010). Aligning short sequencing reads with Bowtie. *Curr. Protoc. Bioinformatics.* Chapter 11, Unit 11.7. doi: 10.1002/0471250953.bi1107s32

Lee, S. H., van der Werf, J. H., Hayes, B. J., Goddard, M. E., and Visscher, P. M. (2008). Predicting unobserved phenotypes for complex traits from whole-genome SNP data. *PLoS Genet.* 4:e1000231.doi: 10.1371/journal.pgen.1000231

Lee, Y. S., Jeong, D.-H. Lee, D. Y., Yi, J., Ryu, C. H., Song, L. K., et al. (2010). OsCOL4 is a constitutive flowering repressor upstream of Ehd1 and downstream of OsphyB. *Plant J. Cell Mol. Biol.* 63, 18–30. doi: 10.1111/j.1365-313x.2010.04226.x

Léran, S., Varala, K., Boyer, J. C., Chiurazzi, M., Crawford, N., Danielvedele, F., et al. (2013). A unified nomenclature of NITRATE TRANSPORTER 1/PEPTIDE TRANSPORTER family members in plants. *Trends Plant Sci.* 19, 5–9. doi: 10.1016/j.tplants.2013.08.008

LeVan, T. D., Bloom, J. W., Bailey, T. J., Karp, C. L., Halonen, M., Martinez, F. D., et al. (2001). A common single nucleotide polymorphism in the CD14 promoter decreases the affinity of Sp protein binding and enhances transcriptional activity. *J. Immunol.* 167, 5838–5844. doi: 10.4049/jimmunol.167.10.5838

Li, D., Wang, L., Wang, M., Xu, Y. Y., Luo, W., Liu, Y. J., et al. (2009). Engineering OsBAK1 gene as a molecular tool to improve rice architecture for high yield. *Plant Biotechnol. J.* 7, 791–806. doi: 10.1111/j.1467-7652.2009.00444.x

Liang, K., Huang, D. C., Kai, Ming, Z., Nguyen, P. T., Hui, X., Wenxia, M., et al. (2010). Marker genotypes for parents of japonica hybrid rice with high combining ability of yield traits. *Acta Agronom. Sin.* 36, 1270–1279. doi: 10.1016/s1875-2780(09)60064-x

Liu, C., Song, G., Zhou, Y., Qu, X., Guo, Z., Liu, Z., et al. (2015). *OsPRR37* and *Ghd7* are the major genes for general combining ability of DTH, PH and SPP in rice. *Sci. Rep.* 5:12803. doi: 10.1038/srep12803

Liu, E. B., Liu, Y., Liu, X. L., Liu, Q. M., Zhao, K. M., Edzesi, W. M., et al. (2013). Detecting marker genotypes with elite combining ability for yield traits in parents of hybrid japonica rice. *Chinese J. Rice Sci.* 27, 473–481. doi: 10.3969/j.issn.1001-7216.2013.05.004

Liu, J., Zhang, F., Zhou, J., Chen, F., Wang, B., and Xie, X. (2012). Phytochrome B control of total leaf area and stomatal density affects drought tolerance in rice. *Plant Mol. Biol.* 78, 289–300. doi: 10.1007/s11103-011-9860-3

Liu, T., Liu, H., Zhang, H., and Xing, Y. (2013). Validation and characterization of Ghd7.1, a major quantitative trait locus with pleiotropic effects on spikelets per panicle, plant height, and heading date in rice (*Oryza sativa* L.). *J. Integr. Plant Biol.* 55, 917–927. doi: 10.1111/jipb.12070

Liu, X., Ishiki, K., and Wang, W. (2002). Identification of AFLP markers favorable to heterosis in hybrid rice. *Breed. Sci.* 52, 201–206. doi: 10.1270/jsbbs.52.201

Liu, Y., Liu, E. B., Zeng, S. Y., Pu, W., Liu, Q. M., Liang, L. J., et al. (2015). Identification of molecular marker fragments associated with combining ability for quality traits in parents of hybrid japonica rice (*Oryza sativa* L.). *Chinese J.Rice Sci.* 29, 373–381. doi: 10.3969/j.issn.1001-7216.2015.04.006

McCouch, S. R., Zhao, K., Wright, M., Tung, C.-W., Ebana, K., Thomson, M., et al. (2010). Development of genome-wide SNP assays for rice. *Breed. Sci.* 60, 524–535. doi: 10.1270/jsbbs.60.524

Kapoor, M., Arora, R., Lama, T., Nijhawan, A., Khurana, J. P., Tyagi, A. K., et al. (2008). Genome-wide identification, organization and phylogenetic analysis of Dicer-like, Argonaute and RNA-dependent RNA Polymerase gene families and their expression analysis during reproductive development and stress in rice. *BMC Genomics* 9:451. doi: 10.1186/1471-2164-9-451

Mo, H. (1982). The analysis of combining ability in p×q mating pattern. *J. Jiangsu Agric. College* 3, 51–57.

Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., et al. (2012). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci. U.S.A.* 110, 453–458. doi: 10.1073/pnas.1215985110

Murakami, M., Matsushika, A., Ashikari, M., Yamashino, T., and Mizuno, T. (2005). Circadian-associated rice pseudo response regulators (PRRs): insight into the control of flowering time. *Biosci. Biotechnol. Biochem.* 69, 410–414. doi: 10.1271/bbb.69.410

Nagasaki, H., Ebana, K., Shibaya, T., Yonemaru, J.-I., and Yano, M. (2010). Core single-nucleotide polymorphisms-a tool for genetic analysis of the Japanese rice population. *Breed. Sci.* 60, 648–655. doi: 10.1270/jsbbs.60.648

Nakamura, Y., Kato, T., Yamashino, T., Murakami, M., and Mizuno, T. (2007). Characterization of a set of phytochrome-interacting factor-like bHLH proteins in *Oryza sativa*. *Biosci. Biotechnol. Biochem.* 71, 1183–1191. doi: 10.1271/bbb.60643

Negrao, S., Oliveira, M., Jena, K., and Mackill, D. (2008). Integration of genomic tools to assist breeding in the japonica subspecies of rice. *Mol. Breed.* 22, 159–168. doi: 10.1007/s11032-008-9177-3

Ng, P. C., and Henikoff, S. (2006). Predicting the effects of amino acid substitutions on protein function. *Annu. Rev. Genomics Hum. Genet.* 7, 61–80. doi: 10.1146/annurev.genom.7.080505.115630

Nielsen, R., Paul, J. S., Albrechtsen, A., and Song, Y. S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nat. Rev. Genet.* 12, 443–451. doi: 10.1038/nrg2986

Nijhawan, A., Jain, M., Tyagi, A. K., and Khurana, J. P. (2008). Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. *Plant Physiol.* 146, 333–350. doi: 10.1104/pp.107.112821

Ono, A., Yamaguchi, K., Fukada-Tanaka, S., Terada, R., Mitsui, T., and Iida, S. (2012). A null mutation of ROS1a for DNA demethylation in rice is not transmittable to progeny. *Plant J. Cell Mol. Biol.* 71, 564–574. doi: 10.1111/j.1365-313X.2012.05009.x

Pandini, F., Vello, N. A., and Lopes, . C. D. A. (2002). Heterosis in soybeans for seed yield components and associated traits. *Braz. Arch. Biol. Technol.* 45, 401–412. doi: 10.1590/S1516-89132002000600001

Parson, W., Strobl, C., Huber, G., Zimmermann, B., Gomes, S. M., Souto, L., et al. (2013). Evaluation of next generation mtGenome sequencing using the Ion Torrent Personal Genome Machine (PGM). *Foren. Sci. Int. Genet.* 7, 543–549. doi: 10.1016/j.fsigen.2013.06.003

Pasam, R. K., Sharma, R., Malosetti, M., van Eeuwijk, F. A., Haseneyer, G., Kilian, B., et al. (2012). Genome-wide association studies for agronomical traits in a world wide spring barley collection. *BMC Plant Biol.* 12:16. doi: 10.1186/1471-2229-12-16

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., et al. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5, 103–113. doi: 10.3835/plantgenome2012.06.0006

Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J.-L. (2012). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7:e32253. doi: 10.1371/journal.pone.0032253

Pradeep, A. R., Jingade, A. H., and Urs, R. S. (2007). Molecular markers for biomass traits: association, interaction and genetic divergence in silkworm *Bombyx Mori*. *Biomarker Insights* 2, 197–217.

Qi, H., Huang, J., Zheng, Q., Huang, Y., Shao, R., Zhu, L., et al. (2013). Identification of combining ability loci for five yield-related traits in maize using a set of testcrosses with introgression lines. *Theor. Appl. Genet.* 126, 369–377. doi: 10.1007/s00122-012-1985-5

Qu, Z., Li, L., Luo, J., Wang, P., Yu, S., Mou, T., et al. (2012). QTL mapping of combining ability and heterosis of agronomic traits in rice backcross recombinant inbred lines and hybrid crosses. *PLoS ONE* 7:e28463. doi: 10.1371/journal.pone.0028463

Ravel, C., Praud, S., Murigneux, A., Canaguier, A., Sapet, F., Samson, D., et al. (2006). Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.). *Genome* 49, 1131–1139. doi: 10.1139/g06-067

Rogbell, J., Subbaraman, N., and Karthikeyan, C. (1998). Heterosis in rice under saline conditions. *Crop Res. Hisar* 15, 68–72.

Ross, C. A., Liu, Y., and Shen, Q. J. (2007). The WRKY gene family in rice (*Oryza sativa*). *J. Integr. Plant Biol.* 49, 827–842. doi: 10.1111/j.1744-7909.2007.00504.x

Salvi, S., Sponza, G., Morgante, M., Tomes, D., Niu, X., Fengler, K. A., et al. (2007). Conserved noncoding genomic sequences associated with a flowering-time

quantitative trait locus in maize. *Proc. Natl. Acad. Sci. U.S.A.* 104, 11376–11381. doi: 10.1073/pnas.0704145104

Sato, Y., Nishimura, A. M., Ashikari, M., Hirano, H. Y., and Matsuoka, M. (2002). Auxin response factor family in rice. *Genes Genet. Syst.* 76, 373–380. doi: 10.1266/ggs.76.373

Shastry, B. S. (2009). SNPs: impact on gene function and phenotype. *Methods Mol. Biol.* 578, 3–22. doi: 10.1007/978-1-60327-411-1_1

Shukla, S., and Pandey, M. (2008). Combining ability and heterosis over environments for yield and yield components in two-line hybrids involving thermosensitive genic male sterile lines in rice (*Oryza sativa* L.). *Plant Breed.* 127, 28–32. doi: 10.1111/j.1439-0523.2007.01432.x

Shull, G. H. (1908). The composition of a field of maize. *J. Heredity* 4, 296–301. doi: 10.1093/jhered/os-4.1.296

Singh, H., Sharma, S., and Sain, R. (2004). Heterosis studies for yield and its components in bread wheat over environments. *Hereditas* 141, 106–114. doi: 10.1111/j.1601-5223.2004.01728.x

Singh, P., Thakur, R., Chaudhary, V., and Singh, N. (1996). Combining ability for grain yield and its components in relation to rice breeding. *Crop Res. Hisar* 11, 62–66.

Smith, J. S. C., Hussain, T., Jones, E. S., Graham, G., Podlich, D., Wall, S., et al. (2008). Use of doubled haploids in maize breeding: implications for intellectual property protection and genetic diversity in hybrid crops. *Mol. Breed.* 22, 51–59. doi: 10.1007/s11032-007-9155-1

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B., et al. (2013). An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS ONE* 8:e54603. doi: 10.1371/journal.pone.0054603

Spindel, J., Wright, M., Chen, C., Cobb, J., Gage, J., Harrington, S., et al. (2013). Bridging the genotyping gap: using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. *Theor. Appl. Genet.* 126, 2699–2716. doi: 10.1007/s00122-013-2166-x

Sprague, G. F., and Tatum, L. A. (1942). General vs. specific combining ability in single crosses of corn. *Agron. J.* 34, 923–932. doi: 10.2134/agronj1942.00021962003400100008x

Stein, N., Herren, G., and Keller, B. (2001). A new DNA extraction method for high-throughput marker analysis in a large-genome species such as *Triticum aestivum. Plant Breed.* 120, 354–356. doi: 10.1046/j.1439-0523.2001.00615.x

Su, J., Pang, C., Wei, H., Li, L., Liang, B., Wang, C., et al. (2016). Identification of favorable SNP alleles and candidate genes for traits related to early maturity via GWAS in upland cotton. *BMC Genomics* 17, 687. doi: 10.1186/s12864-016-2875-z

Subbaiyan, G. K., Waters, D. L., Katiyar, S. K., Sadananda, A. R., Vaddadi, S., and Henry, R. J. (2012). Genome-wide DNA polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. *Plant Biotechnol. J.* 10, 623–634. doi: 10.1111/j.1467-7652.2011.00676.x

Sun, L., Zhang, Q., Xu, Z., Yang, W., Guo, Y., Lu, J., et al. (2013). Genome-wide DNA polymorphisms in two cultivars of mei (*Prunus mume* sieb. et zucc.). *BMC Genet.* 14:98. doi: 10.1186/1471-2156-14-98

Sutoh, K., Washio, K., Imai, R., Wada, M., Nakai, T., and Yamauchi, D. (2015). An N-terminal region of a Myb-like protein is involved in its intracellular localization and activation of a gibberellin-inducible proteinase gene in germinated rice seeds. *Biosci. Biotechnol. Biochem.* 79, 747–759. doi: 10.1080/09168451.2014.998620

Takano, M., Inagaki, N., Xie, X., Kiyota, S., Baba-Kasai, A., Tanabata, T., et al. (2009). Phytochromes are the sole photoreceptors for perceiving red/far-red light in rice. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14705–14710. doi: 10.1073/pnas.0907378106

Takano, M., and Shinomura, T. (2005). Distinct and cooperative functions of phytochromes A, B, and C in the control of deetiolation and flowering in rice. *Plant Cell* 17, 3311–3325. doi: 10.1105/tpc.105.035899

Tang, S., Zhang, H., Liang, G., Yan, C., Liu, Q., and Gu, M. (2008). Reasons and countermeasures of slow development on three-line japonica hybrid rice. *Hybrid Rice* 23, 1–5.

Wang, D., and Al., E. (2007). Genome-wide analysis of the auxin response factors (ARF) gene family in rice (*Oryza sativa*). *Gene* 394, 13–24. doi: 10.1016/j.gene.2007.01.006

Wang, E., Wang, J., Zhu, X., Hao, W., Wang, L., Li, Q., et al. (2008). Control of rice grain-filling and yield by a gene with a potential signature of domestication. *Nat. Genet.* 40, 1370–1374. doi: 10.1038/ng.220

Williams, W., and Gilbert, N. (1960). Heterosis and the inheritance of yield in the tomato. *Heredity* 14, 133–149. doi: 10.1038/hdy.1960.11

Xie, F., and Hardy, B. (2009). Accelerating *Hybrid Rice Development.* Los Banos: International Rice Research Institute.

Xie, H., Dang, X. J., Liu, E. B., Zeng, S. Y., and Hong, D. L. (2016). Identifying SSR marker locus genotypes with elite combining ability for yield traits in backbone parents of japonica hybrid rice (*Oryza sativa* L.) in Jianghuai area. *Acta Agron. Sin.* 42, 330–343. doi: 10.3724/SP.J.1006.2016.00330

Xie, Z., and Shen, Q. J. (2005). Annotations and functional analyses of the rice WRKY gene superfamily reveal positive and negative regulators of abscisic acid signaling in aleurone cells. *Plant Physiol.* 137, 176–189. doi: 10.1104/pp.104.054312

Yan, W., Liu, H., Zhou, X., Li, Q., Zhang, J., Lu, L., et al. (2013). Natural variation in Ghd7.1 plays an important role in grain yield and adaptation in rice. *Cell Res.* 23, 969–971. doi: 10.1038/cr.2013.43

Yang, G. P., Maroof, M. A. S., Xu, C. G., Zhang, Q., and Biyashev, R. M. (1994). Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice. *Mol. Gen. Genet.* 245, 187–194. doi: 10.1007/BF002 83266

Yang, X., Yan, J., Shah, T., Warburton, M. L., Li, Q., Li, L., et al. (2010). Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. *Theor. Appl. Genet.* 121, 417–431. doi: 10.1007/s00122-010-1320-y

Yu, J., Meng, Z., Liang, W., Kudla, J., Tucker, M. R., Luo, Z., et al. (2016). A rice Ca^{2+} binding protein is required for tapetum function and pollen formation. *Plant Physiol.* 172, 1772–1786. doi: 10.1104/pp.16.01261

Yu, L. X., Liu, X., William, B., and Liu, X. P. (2016). Genome-Wide Association study identifies loci for salt tolerance during germination in autotetraploid Alfalfa (*Medicago sativa* L.) using Genotyping-by-Sequencing. *Front. Plant Sci.* 7:956. doi: 10.3389/fpls.2016.00956

Yuan, L., and Virmani, S. (1988). *Status of Hybrid Rice Research and Development.* Manila: Hybrid Rice. International Rice Research Institute.

Zhang, J., Song, Q., Cregan, P. B., Nelson, R. L., Wang, X., Wu, J., et al. (2015). Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine* max) germplasm. *BMC Genomics* 16:217. doi: 10.1186/s12864-015-1441-4

Zhao, K., Tung, C.-W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., et al. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa. Nat. Commun.* 2:467. doi: 10.1038/ncomms1467