Article

Whole-Genome Sequencing and RNA-Seq Reveal Differences in Genetic Mechanism for Flowering Response between Weedy Rice and Cultivated Rice

Richard S. Garcia [ID], Sapphire Coronejo [ID], Jonathan Concepcion † and Prasanta K. Subudhi *[ID]

School of Plant, Environmental, and Soil Sciences, Louisiana State University Agricultural Center, Baton Rouge, LA 70803, USA; rgarcia@agcenter.lsu.edu (R.S.G.); scoronejo@agcenter.lsu.edu (S.C.); concep10@msu.edu (J.C.)
* Correspondence: psubudhi@agcenter.lsu.edu; Tel.: +1-225-578-1303
† Current address: Department of Crop and Soil Sciences, Michigan State University, East Lansing, MI 48824, USA.

Abstract: Flowering is a key agronomic trait that influences adaptation and productivity. Previous studies have indicated the genetic complexity associated with the flowering response in a photoinsensitive weedy rice accession PSRR-1 despite the presence of a photosensitive allele of a key flowering gene *Hd1*. In this study, we used whole-genome and RNA sequencing data from both cultivated and weedy rice to add further insights. The *de novo* assembly of unaligned sequences predicted 225 genes, in which 45 were specific to PSRR-1, including two genes associated with flowering. Comparison of the variants in PSRR-1 with the 3K rice genome (RG) dataset identified unique variants within the heading date QTLs. Analyses of the RNA-Seq result under both short-day (SD) and long-day (LD) conditions revealed that many differentially expressed genes (DEGs) colocalized with the flowering QTLs, and some DEGs such as *Hd1*, *OsMADS56*, *Hd3a*, and *RFT1* had unique variants in PSRR-1. *Ehd1*, *Hd1*, *OsMADS15*, and *OsMADS56* showed different alternate splicing (AS) events between genotypes and day length conditions. *OsMADS56* was expressed in PSRR-1 but not in Cypress under both LD and SD conditions. Based on variations in both sequence and expression, the unique flowering response in PSRR-1 may be due to the high-impact variants of flowering genes, and *OsMADS56* is proposed as a key regulator for its day-neutral flowering response.

Keywords: days to heading; genetic interaction; *Oryza sativa*; photosensitivity; red rice; RNA-Sequencing

1. Introduction

Weedy rice (*Oryza sativa f. spontanea* Rosh), a conspecific relative of cultivated rice, is a major irritant for rice farmers in many parts of the world [1]. The most plausible hypotheses regarding its evolution from *japonica* or *indica* and/or *japonica* × *indica* hybrids have been supported by molecular studies [2–5]. The de-domestication hypothesis of weedy rice evolution is recently gaining much traction based on the whole-genome sequence (WGS) analyses [6–8]. There are two types of weedy rice ecotypes in the USA based on the morphology: straw-hulled (SH) weedy rice with a light, brown-colored hull; and black-hull awned (BHA) weedy rice with a black-colored hull and long awns. SH ecotypes are morphologically similar to cultivated rice, while BHA resembles wild rice *Oryza rufipogon* [9]. However, they differ in several agronomically important traits such as heading date [10]. SH weedy rice exhibited early heading compared to BHA [11].

The economic impact of weedy rice infestation in the rice growing states in southern USA is enormous due to reduced grain yield, quality, and marketability resulting from contamination with red kernels of weedy rice. As weedy rice has been noted for their persistence, researchers have focused their attention on the genetic dissection of traits such as seed shattering and seed dormancy [12,13]. Many of these adaptive features can be

exploited to improve cultivated rice. Early heading is an essential weedy trait that helps them to survive and compete with cultivated rice by contributing to the escape from being harvested [14].

Flowering is a complex trait with wide variation among cereal crops. Investigation of the flowering mechanisms of rice as a model grass species has expanded our knowledge through its relevance to other cereal crops [15–17]. However, the growth and development strategies of rice as a short-day (SD) plant compromise its relevance as a model for temperate cereals such as wheat and barley, which are all long-day (LD) plants such as *Arabidopsis*. Studies on natural variation in rice have shown orthologous photoperiod pathway genes similar to *Arabidopsis* [18]. *Heading date 1* (*Hd1*), a major gene determining day length sensitivity in rice, was orthologous to *AtCO* (*CONSTANS 1*) in *Arabidopsis* [15]. Further, *Hd3a* is also an ortholog of *Arabidopsis* floral pathway integrator *Flowering Locus T* (*FT*). *Hd1* promotes *Hd3a* expression under SD conditions and repression during LD conditions, whereas *AtCO* promotes *FT* expression under LD conditions in *Arabidopsis* [19]. Another gene, *RFT1*, a tandemly duplicated paralog of *Hd3a*, is a LD-specific florigen in rice [20]. In *Arabidopsis*, *GIGANTEA* (*GI*) plays an important role in regulating the *CO-FT* pathway [21]. However, a rice homolog *OsGI* mediates the regulation of the *Hd1-Hd3a* pathway [22]. These findings reveal that this pathway controlling photoperiodic flowering is conserved between rice and *Arabidopsis* despite their contrasting photoperiodic response. Another pathway for photoperiodic flowering in rice involves *Ehd1*, a flowering gene encoding B-type response regulator, which shows no obvious ortholog in *Arabidopsis*. *Ehd1* functions upstream of *Hd3a* and *RFT1* [23], whereas its expression is higher under LD conditions in weedy rice compared with cultivated rice [24]. In addition, *Ghd7* influences the expression levels of *Ehd1* and *Hd3a* but not *Hd1* and delays flowering by repressing their expression under LD conditions [25]. The identification of orthologous genes has been helpful for understanding flowering response pathways of both rice and *Arabidopsis*.

Analysis of the natural genetic variation in cultivated and wild germplasm at the genomic level led to an improved understanding of the evolutionary origin of weedy rice [6–8]. The understanding of the molecular differences of weedy and agronomic traits in cultivated and weedy rice can be useful for rice improvement [26]. The application of next-generation sequencing (NGS) facilitates the understanding of the variation in gene expression patterns through the identification of high-impact DNA polymorphisms such as single-nucleotide polymorphisms (SNPs), insertions/deletions (InDels), and structural variations (SVs) within coding regions. Furthermore, analyses of unmapped sequences provide an opportunity to examine novel genes unique to weedy rice [27].

The present study used a high-yielding long-grain rice cultivar Cypress and a SH weedy accession PSRR-1. PSRR-1 was used in a number of investigations related to seed dormancy, seed shattering, and days to heading [12,13,24]. Based on results from staggered planting experiments, both genotypes showed no response to photoperiod under both LD and SD conditions, but the hybrids between PSRR-1 and Cypress and the near-isogenic line (NIL) with introgressed PSRR-1 *Hd1* were highly photosensitive [24]. Despite no difference in the response to photoperiod between parents, six heading date QTLs were identified in recombinant inbred line populations under LD conditions [24]. Several of the known genes overlapping with QTLs were validated using introgression lines of PSRR-1 [28]. The comparison of genotyping and gene expression data of key flowering genes in the NIL and parents led us to suggest that the genetic interaction of *Hd1* with a novel genetic factor may be responsible for the photo-insensitive nature of PSRR-1 under LD conditions [24].

Based on genome-wide DNA polymorphisms, PSRR-1 was genetically closer to *indica* rice than the *japonica* rice and, therefore, its evolution from *indica* rice was more plausible [27]. Another WGS study revealed clear-cut differences in the genomic makeup between SH and BHA weedy rice, and the SH weedy rice diverged later than the BHA weedy rice [11]. In these studies, several high-impact SNPs and InDels were identified with an attempt to elucidate loci for adaptive traits such as seed dormancy and seed shattering [11,26].

In this study, we performed comparative analyses of SNPs and InDels using both whole-genome resequencing and 3K rice genome data to identify all the unique variants in SH weedy rice accession PSRR-1, in addition to analyses of SVs and unmapped sequences. We also compared the genetic variations in heading date loci and global transcriptomic profiles of PSRR-1 and cultivated rice Cypress to unravel shared and unique features for each genotype.

2. Results

### 2.1. Whole-Genome Sequence Analyses of PSRR-1

A total of 95,183,701 reads were generated from Cypress, while 137,685,675 reads were from PSRR-1 with assembly rates of 98% and 97%, respectively (Table S1). In total, there were 1,269,049 and 3,029,642 variants, which included SNPs, InDels, and SVs in Cypress and PSRR-1, respectively.

2.2. Unique SNPs and InDels in PSRR-1 through Comparison with 3K-Rice Genome Data Set

A comparison of 2,190,324 homozygous SNPs from PSRR-1 with 4,817,964 SNPs from the 3K RG-filtered SNP dataset revealed 1,485,635 common-to-low-frequency SNP alleles, 200,582 rare SNP alleles, and 504,107 unique alleles. Common-to-low-frequency and rare alleles were PSRR-1 SNPs that were called in the 3K RG SNP dataset with minor allele frequencies (MAFs) of ≥ 0.01 and <0.01, respectively. The SNPs that were not called in the 3K RG SNP dataset were considered unique to PSRR-1. On the other hand, there were 413,374 unique InDels in PSRR-1. The distribution of the SNPs and InDels was spread across all chromosomes (Figure 1). These unique SNP and InDels were categorized according to the predicted variant effects. Those with high-impact were considered due to their high probability of generating gain or loss-of-function and/or changes in the expression levels of the desired genes (Figure S1). In addition, structural variants (SVs) in PSRR-1 showed an unbiased distribution across all chromosomes (Figure S2). The SVs and unique variants showed several high-impact mutations within the agronomically important QTLs (Figure 2), which encompassed morphological, physiological, and stress tolerance traits. The panicle/flowering traits had the greatest number of variants. The days to heading (DTH) QTLs alone harbored several unique mutations (Figure 1, Table S2).

2.3. Unmapped Sequence of PSRR-1

Unmapped reads of PSRR-1 were assembled using rice as a reference, and there were 225 predicted genes of which only 180 genes were annotated with significant matches (Table S3). The remaining 45 predicted sequences had a very low to non-significant hits and were most likely unique to PSRR-1. The E-value was used as the main parameter for significance followed by bit score and percent identity. Twenty six of the 180 genes with E-value 0.0 were considered redundant and were therefore more likely identical, while 154 genes were closely related and/or almost identical with E-value $< 1.0 \times 10^{-10}$. From this set of significant sequence matches, 13, 31, and 136 were annotated as uncharacterized, characterized, and hypothetical proteins, respectively. Both uncharacterized and hypothetical proteins had not been experimentally validated. The genes were homologous to cultivated rice, wild rice, weedy rice, and other grass sequences. Moreover, a closer look at the protein families and domains of the predicted genes revealed that the majority of protein families belong to DISEASE RESISTANCE PROTEIN RP (PTHR23155), P-loop containing nucleoside triphosphate hydrolase (IPR027417), NB-ARC (IPR002182), Rx, N-terminal (IPR041118), and Leucine-rich repeat (IPR001611) (Table S4).
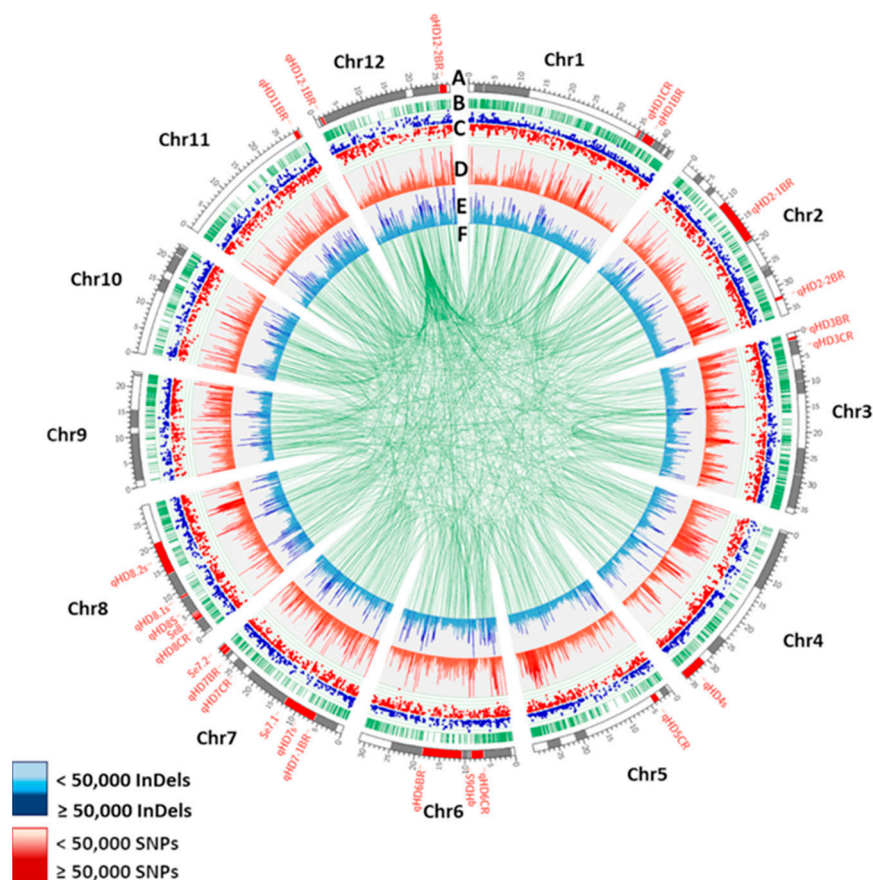
Figure 1. Genome-wide distributions of unique SNPs and InDels, and structural variations overlapping with DTH QTLs along with differentially expressed genes in PSRR-1. (**A**) DTH QTL segments identified from Gramene database (https://archive.gramene.org/qtl/) (accessed on 5 May 2020) [29]: grey bands, and DTH QTLs of weedy rice from literature [24,30–32]: red bands. (**B**) Gene distribution across chromosome (green bars). (**C**) Scatter plot of differentially expressed genes (DEGs). Blue dots and red dots represent upregulated and downregulated DEGs (| log2 fold change | > 1.5 and p-adj < 0.05) in all combined pair-wise conditions, respectively. (**D**) Unique SNP density of PSRR-1 versus 3K Rice Genome (RG) dataset at 10-kb window (red bars). (**E**) Unique InDel density of PSRR-1 versus 3K RG dataset at 10-kb window. (**F**) Structural variations in PSRR-1 (green lines).

2.4. Transcriptome Analysis of PSRR-1 and Cypress under Short and Long-Day Conditions

RNA sequencing of leaf tissues from PSRR-1 and Cypress under long-day and short-day conditions resulted in 316 and 330 million reads, respectively (Table S5). The mean assembly for both genotypes was 96%. Differentially expressed genes (DEGs) were identified for each pairwise comparison between Cypress and PSRR under different conditions. Each pairwise comparison included the genotype-specific DEGs for LD and SD conditions (LD PSRR vs. LD Cypress and SD PSRR vs. SD Cypress, respectively) and day length-specific DEGs (LD PSRR vs. SD PSRR). The total number of DEGs was 3982 (1683 upregulated and 2299 downregulated), 3936 (1476 upregulated and 2460 downregulated), and 1774 (1142 upregulated and 632 downregulated) in LD PSRR vs. LD Cypress, SD PSRR vs. SD Cypress, and LD PSRR vs. SD PSRR, respectively (Figure 3, Table S6).
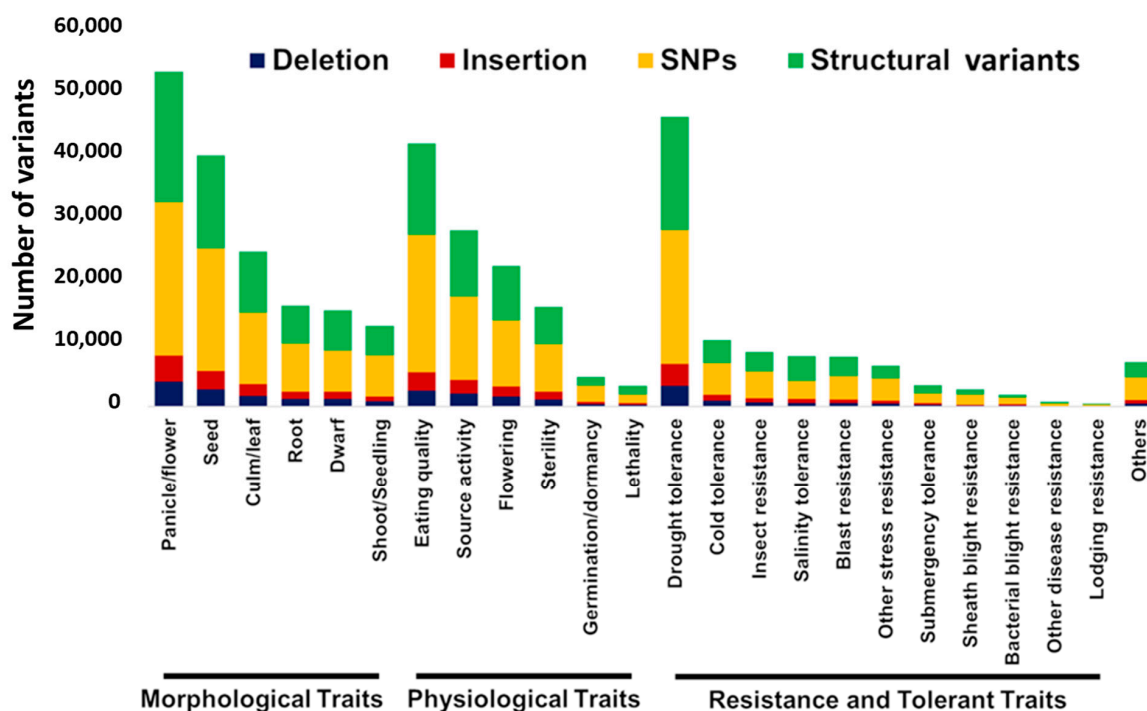
Figure 2. Unique high-impact SNPs, InDels, and structural variants (SVs) associated with QTLs for agronomically important traits mined from Q-TARO database (http://qtaro.abr.affrc.go.jp/) (accessed on 28 October 2020) [33]. Blue, red, yellow, and green bars correspond to deletion, insertion, SNPs, and SVs, respectively.
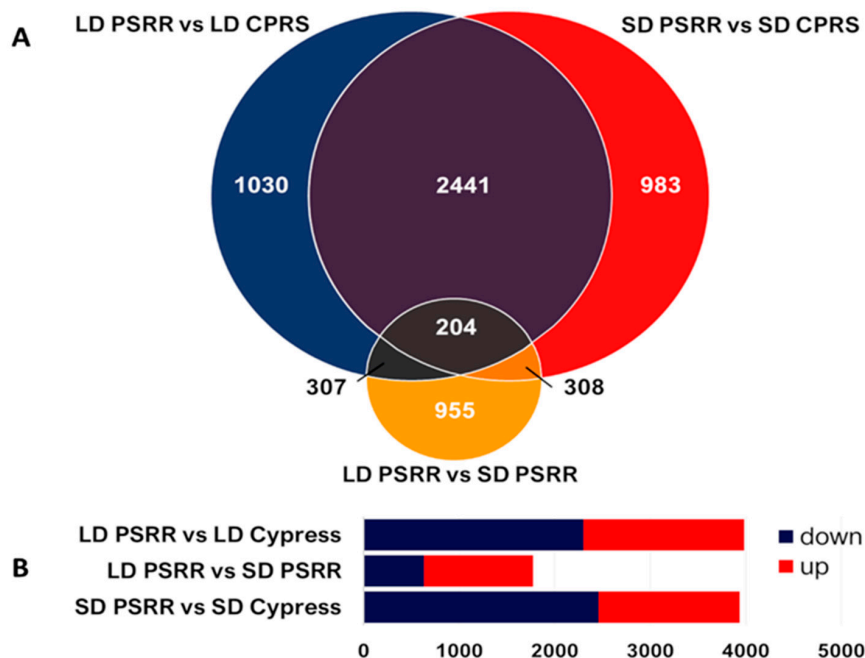
Figure 3. Differentially expressed gene (DEG) profiles of PSRR-1 and Cypress under different day length conditions. (**A**) Venn diagram showing overlapping of DEGs between PSRR-1 and Cypress (CPRS) under long-day (LD) and short-day (SD) conditions with pair-wise comparisons: LD PSRR vs. LD CPRS, LD PSRR vs. SD CPRS, and SD PSRR vs. SD CPRS (|log2 fold change| ≥ 1.5 and *padj* < 0.05). (**B**) Bar graphs showing number of DEGs from different pair-wise comparisons along with the regulation pattern.

2.5. Gene Ontology (GO) and Pathway Analysis of DEGs under SD and LD Conditions

In agriGO analysis, 78, 59, and 80 significant GO terms were identified in LD PSRR vs. LD Cypress, SD PSRR vs. SD Cypress, and LD PSRR vs. SD PSRR, respectively (Figure S3). Plant Reactome pathway analyses of DEGs showed major plant pathways involved in cellular, growth, developmental processes, metabolism and regulation, and responses to abiotic and biotic stimuli and stresses (Table 1). Both GO and pathway results indicated involvement in significant biological functions encompassing a diverse range of plant processes including reproduction. The pathways involved in flowering networking and control could be reflected from the notable number of DEGs in the 'reproductive structure development' and 'hormone signaling, transport, and metabolism' pathways. The DEGs within these categories were involved in flowering. For example, LD and SD-regulated florigens were commonly found under all conditions under 'reproductive structure development' such as *Hd3a* and *RFT1*. On the other hand, the 'hormone signaling, transport, and metabolism' pathway had DEGs involved in gibberellin and auxin signaling, which are important in flowering responses. Additionally, several of these DEGs belong to the transcription factor (TF) families that are largely involved in flowering such as CO-like and MIKC-MADS TFs (Figure S4).

Table 1. Distribution (%) of differentially expressed genes on plant pathway nodes in PSRR-1 and Cypress under short- and long-day conditions.

	Plant Pathway	LD PSRR vs. LD CPRS	SD PSRR vs. SD CPRS	LD PSRR vs. SD PSRR
Cellular processes	Protein metabolism: translation	5.0	3.7	1.1
	Cell cycle	3.5	3.3	1.9
	Cellular processes	2.3	1.8	0.6
	DNA replication: activation of the pre-replicative complex	0.6	0.4	0.0
Circadian rhythm	Circadian rhythm	0.0	0.0	0.6
Growth and developmental processes	Reproductive structure development	11.7	7.2	11.5
	Growth and developmental processes	4.2	2.7	4.7
	Vegetative structure development	1.5	0.6	2.1
	Amine and polyamine biosynthesis	1.3	0.0	0.8
Metabolism and regulation	Metabolism and regulation	17.9	20.8	19.2
	Secondary metabolism	11.9	13.0	14.9
	Hormone signaling, transport, and metabolism	10.0	11.8	13.4
	Amino acid metabolism	12.5	10.8	8.2
	Carbohydrate metabolism	5.0	8.0	6.6
	Cofactor biosyntheses	5.0	3.7	4.0
	Inorganic nutrients metabolism	2.1	4.5	2.1
	Fatty acid and lipid metabolism	2.1	2.7	2.1
	Detoxification	0.0	1.2	1.3
	Amine and polyamine biosynthesis	0.0	0.6	0.4
	Nucleotide metabolism	0.0	0.4	0.0
	Photorespiration	0.2	0.2	0.0
Responses to stimuli: abiotic stimuli and stresses	Responses to stimuli: abiotic stimuli and stresses	1.0	1.0	1.5
	Response to cold temperature	0.4	0.4	0.8
	Response to drought	0.8	0.4	0.4
	Response to heavy metals	0.4	0.4	0.8
	Response to submergence	0.0	0.4	0.0
	Response to salinity	0.2	0.0	0.2
Responses to stimuli: biotic stimuli and stresses	Responses to stimuli: biotic stimuli and stresses	0.2	0.0	0.4
	Recognition of fungal and bacterial pathogens and immunity response	0.2	0.0	0.4

SD, Short-day; LD, Long-day.

## 2.6. DEGs and Variants in Flowering Genes and Flowering-Related Pathway

Among the DEGs, there were 35 genes related to flowering under all day length conditions (Figure 4). Some of these were major rice flowering genes such as *Hd1*, *Hd3a*, *RFT1*, and *Ehd1*, while some indirectly involved in the flowering response were NF-YA and OsMADS-type flowering genes. These DEGs were constantly regulated under both LD and SD conditions, and the expression level varied due to genotype under some instances. For example, *Hd1* and *OsMADS56* of PSRR-1 showed upregulation regardless of day length conditions. Similarly, *LUX* and *RCN1* showed downregulation under all day length conditions. There were DEGs that were only regulated on specific day lengths (day length effect) such as *Ehd1, RFT1*, and *Hd3a*, which were downregulated under LD conditions. Furthermore, the differentially expressed flowering-related genes revealed unique SNPs, InDels, and SVs (Figure 4, Tables S7 and S8). For example, *Hd1*, *OsMADS56*, *Ehd1*, *Hd3a*, and *RFT1* had unique SNPs and InDels in PSRR-1 compared with Cypress.
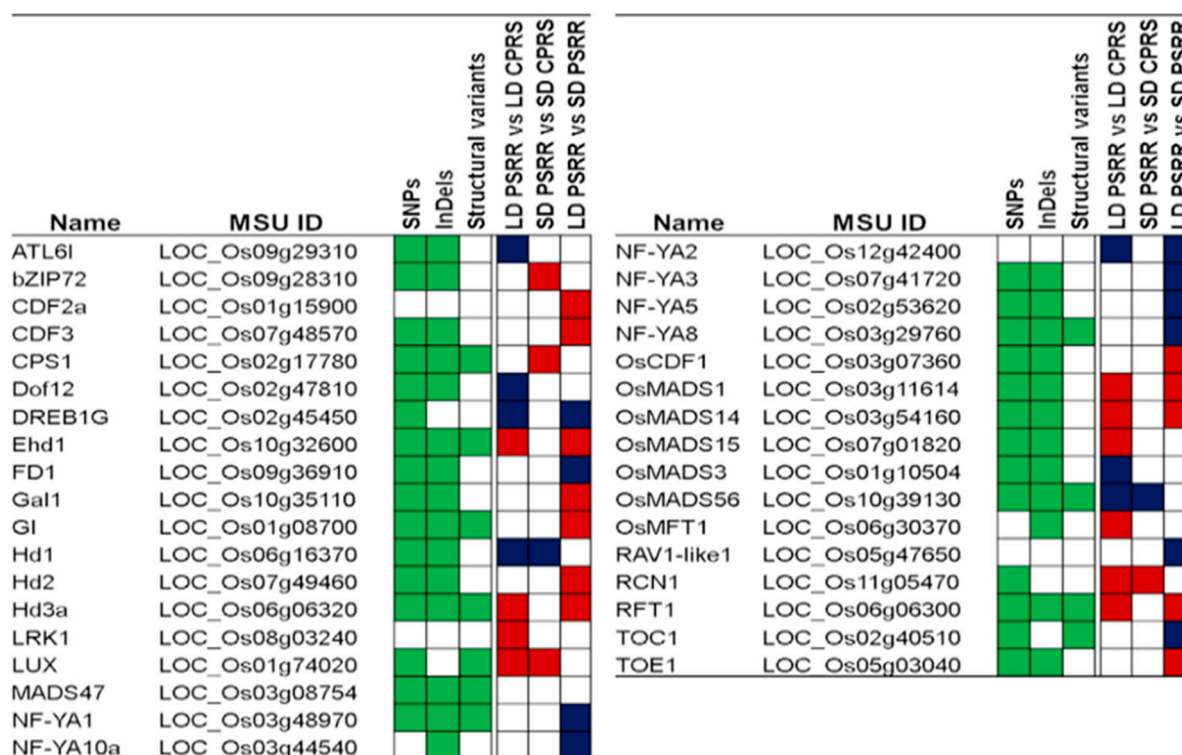
Figure 4. Flowering genes with unique SNP, InDels, and SVs, and the regulation pattern under different day length conditions. Green-filled boxes indicate variants within the gene. Blue and red-filled boxes indicate the regulation pattern (up- and downregulated, respectively) ($|\log2$ fold change$| \geq 1.5$ and *padj* < 0.05) in pairwise comparisons of day length conditions between long day (LD) and short day (SD) of PSRR-1 (PSRR) and Cypress (CPRS).

2.7. Differentially Expressed Flowering-Related Genes Overlapping DTH QTLs

Twenty of the 35 differentially expressed flowering-related genes identified from the RNA-Seq analysis (Figure 4) overlapped with DTH QTLs from previous studies involving weedy rice (Figure 5). Some of the DEGs belonged to the major flowering genes such as *RFT1* and *Hd3a*, which overlapped with *qDTH-6*. *Hd1* overlapped with *qDTH-6* and *qHD6^{BR}*, while some flowering genes such as *GI*, *OsMADS14*, and *OsMADS56* overlapped with *dth1.1*, *qDTH-3*, and *qQTL10a*, respectively.
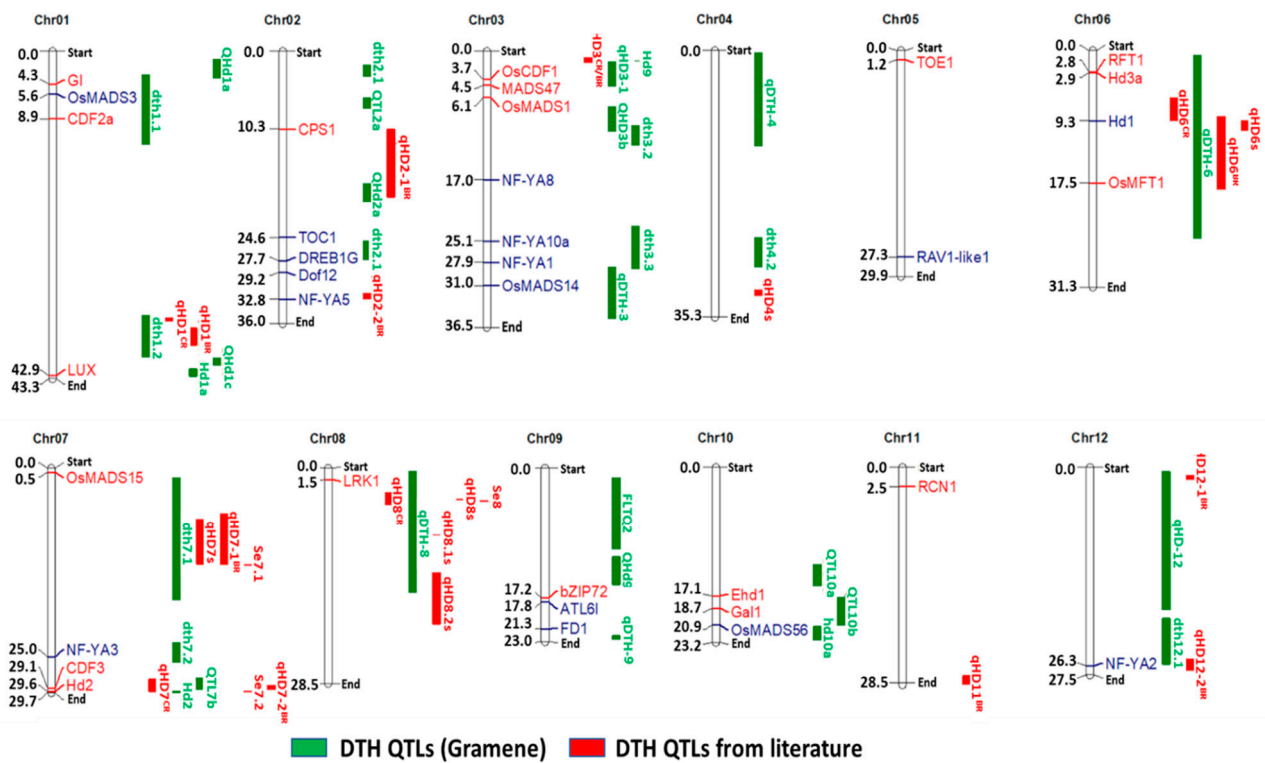
Figure 5. Overlapping of flowering-related genes with days to heading (DTH) QTLs. Upregulated and downregulated genes are shown in blue and red fonts, respectively. Vertical green bars represent published DTH QTLs from Gramene (https://archive.gramene.org/qtl/) (accessed on 5 May 2020) [29] and vertical red bars represent QTLs from earlier studies involving weedy rice [24,30–32].

2.8. Alternative Splicing Events of Flowering Genes under Short- and Long-Day Conditions

A total of 49,707 alternative splicing (AS) events were identified in PSRR-1 (25,179) and Cypress (24,528) (Table S7). The AS profiles of selected flowering genes revealed no differences between genotypes and day length conditions in *Ghd7*, *Hd3a*, *RFT1*, and *OsMADS14* (Table 2). However, *Hd1* and *OsMADS56* had AS profile differences between PSRR-1 and Cypress but not within LD and SD conditions. Further, *Ehd1* showed no AS event differences between different day lengths in Cypress, while PSRR-1 exhibited AS events only under SD conditions. *Ehd1*, *Hd1*, and *OsMADS56* showed different AS events between genotype or day length conditions, which indicated altered transcript profiles despite similar flowering patterns of PSRR-1 and Cypress.

## 2.9. qRT-PCR Validation of Flowering Genes

The qRT-PCR analysis of selected flowering genes showed similar expression profiles with RNA-Seq of Cypress and PSRR-1 under LD conditions; however, the trend was different under SD conditions (Figure 6). Under LD conditions, *RFT1*, *Hd3a*, *Ehd1*, *OsMADS14*, and *OsMADS15* were downregulated and *Hd1* showed upregulation in PSRR-1. On the other hand, *RFT1*, *Hd3a*, *Hd1*, and *OsMADS14* were upregulated, and *Ehd1* and *OsMADS15* were downregulated under SD conditions in PSRR-1. Both *Ehd1* and *OsMADS15* did not reflect the expression pattern in qRT-PCR as with RNA-seq (Figure 4), which could be due to their elimination in RNA-seq analysis because of the threshold setting ($|\log 2$ fold change$| \geq 1.5$ and *padj* < 0.05).

Table 2. Alternative splicing events of differentially expressed major flowering genes.

| Gene Symbol	SD PSRR	LD PSRR	SD Cypress	LD Cypress
Ehd1	A5SS	-	A5SS	A5SS
Ghd7	-	-	-	-
Hd1	-	-	A3SS	A3SS
Hd3a	-	-	-	-
OsMADS14	A3SS	A3SS	A3SS	A3SS
OsMADS15	-	-	-	A3SS, ES, IR, IR1 + IR2
OsMADS56	Complex IR	Complex IR	-	-
RFT1	-	-	-	-

IR, Intron retention; ES, exon skipping; A3SS, alternative 3′ splice sites; A5SS, alternative 5′ splice sites; MXE, mutually exclusive exon; Complex, A5SS + A3SS, A5SS + ES + A3SS, and the like.
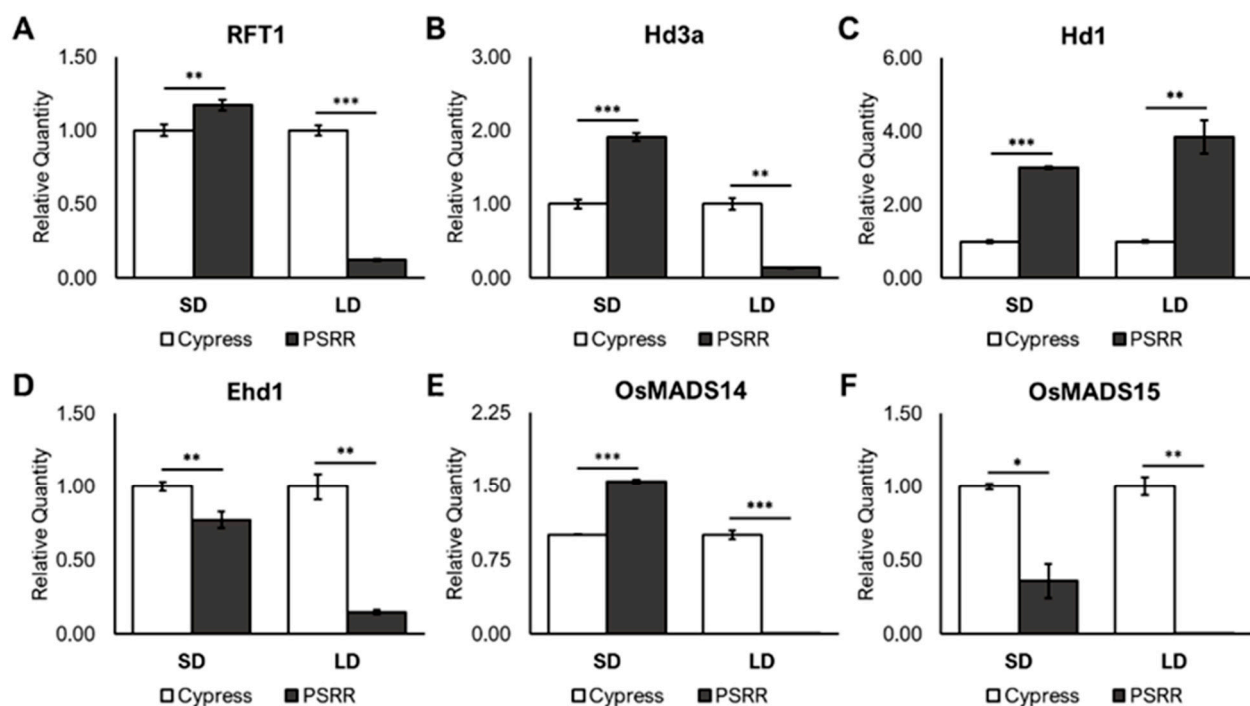
Figure 6. Relative expression levels of flowering-related genes in Cypress and PSRR-1 under short-day (SD) and long-day (LD) conditions. (**A–F**) Real-time quantitative reverse transcription PCR results of flowering genes using Ubq5 as the internal control and Cypress as the reference sample. Bar graph depicts mean relative quantity ± standard deviation. Asterisk (*) indicates significance by two-tailed Student's t-test assuming unequal variances; * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

In addition, *OsMADS56* (*LOC_Os10g39130*) showed amplification in PSRR-1 but not in Cypress (Figure 7). The lack of expression under LD or SD conditions in Cypress validated the RNA-Seq results in which BAM files showed no *OsMADS56* reads. Analysis of WGS data showed a large deletion (1008 bp at position 20,863,490–20,864,497) in Cypress, resulting in a loss of the start codon in *OsMADS56* (Figures S5 and S6).
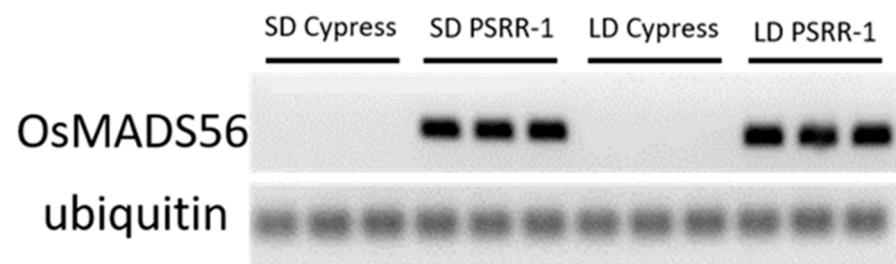
Figure 7. Semiquantitative RT-PCR amplification of *OsMADS56* in Cypress and PSRR-1 under short-day (SD) and long-day (LD) conditions.

3. Discussions

Flowering time is a key agronomic attribute that determines local adaptation and productivity in crop plants. Understanding the genetic basis of flowering was helpful because the development of varieties with uniform flowering assists in harvesting and helping to remove or reduce the response to photoperiod for expanding the adaptation and cultivation of crops to larger geographic regions. Flowering time is equally important for weedy species because it enhances their competitiveness and persistence. While early flowering allows the shedding of seeds and development of a seedbank before crop harvest, the overlapping of flowering time between weedy rice and crops ensures harvesting along with the crop and increases the chance of gene flow between weeds and crop species [17]. Genetic studies on flowering in US weedy rice have been limited to QTL mapping [24,31,32,34]. Demonstrations of the tremendous amount of hidden variation for flowering time in weedy rice have indicated the usefulness of weedy rice as a model to investigate the genes and their interaction controlling this important agronomic trait [24]. Therefore, the present investigation was carried out to provide some insights into the molecular genetic basis of flowering response in a straw hull weedy rice accession through comparison with cultivated rice using whole-genome resequencing and global transcriptomics approaches.

The de novo assembly of unmapped reads of PSRR-1 revealed novel genes as with earlier studies in rice [27]. The majority of these genes were homologous to the *indica* rice group (Table S3), confirming earlier studies regarding the origin of straw-hulled weedy rice from *indica* rice [6,26,31,32]. Among the predicted genes, a few had protein structure domains similar to heading date/flowering genes. For example, *g146.t1* contig had an RNA-binding domain (IPR035979) of *OsRRMh* (*LOC_Os09g34070*) and RNA-recognition motif (RRM) of *OsFCA* (*LOC_Os09g03610*). These two genes, *OsRRMh* and *OsFCA*, were associated with the transition from the vegetative to reproductive stage [35] and the autonomous flowering pathway in rice, respectively [36]. *OsRRMh* was also associated with increased yield [37], confirming the finding of Liu and Cai [35] who reported longer panicles in RNAi lines and reduced fertility and spikelet number in overexpressing lines compared with wild type. Studies on the ab initio gene prediction such as this help in the discovery of 'novel' genes. Although in vitro validation such as the PCR experiment was not performed, the likelihood of identifying new genes is not impossible, as illustrated in ab initio gene predictions in rats, which was confirmed through PCR validation [38]. The above genes with protein families similar to flowering-related genes suggest their involvement in flowering time variation in weedy rice.

Despite being the closest relative of cultivated rice, weedy rice exhibits wide genetic diversity and complex genetic background [39]. Although de-domestication events involving a few genes were recently hypothesized for the origin of US weedy rice [6], further investigation is needed to understand the reversion at these loci to the *indica* type, which are not cultivated in the US [40]. The hidden variation of flowering time genes in the weedy rice accession with day-neutral response could be a useful model to trace the origin of weedy rice through analysis of this domesticated trait [24].

PSRR-1 had numerous unique InDels and SNPs when compared against the 3K RG dataset, comprising several subspecies of cultivated rice and wild rice. Based on the comparison of SNPs and InDels, more variants have been reported between PSRR-1 and a *japonica* rice genotype compared to an *indica* genotype [26]. These unique variants were co-localized in the heading date QTLs reported earlier [31,32]. Similarly, heading date QTLs from our earlier study [24] involving PSRR-1 overlapped with the unique variants in $qHD1^{CR}$, $qHD3^{CR}$, $qHD5^{CR}$, $qHD6^{CR}$, $qHD7^{CR}$, and $qHD8^{CR}$. The presence of these unique variants in PSRR-1 could have arisen from spontaneous mutation during the de-domestication process. Pre-existing variations, new mutations, and variations contributed greatly to the adaptation of weedy rice [7]. Weedy rice may not have simply reverted to domesticated forms but may have employed different sets of genes for improving its adaptation depending on the selection pressure in a particular environment. For example, in addition to other adaptive genes such as *sh4*, *sh-h*, and *Rc*, flowering genes, *OsMADS51* and *Ehd4* were associated with the de-domestication process [7]. The presence of these unique variants, specifically the high-impact mutations and SVs, should be useful for functional genomic studies not only related to flowering but also with other agronomically important traits.

Among the DEGs identified from the comparative transcriptomics approach involving PSRR-1 and Cypress under SD and LD conditions, many overlapped with heading date QTLs and were related to flowering based on the literature (Figures 4 and 5). Apart from unique variants and SVs identified in the flowering genes such as *Hd3a*, *RFT1*, *OsMADS14*, *OsMADS15*, and *OsMADS56* (Tables S8 and S9), there were changes in the splicing events (Tables 2 and S7), which could have contributed to the differential expression pattern between PSRR-1 and Cypress. Differential AS events in rice were reported as coping mechanisms under adverse environments [41–43], and specifically, the flowering time in coconut was influenced by AS of the FT gene [44].

In this study, PSRR-1 did not show any delay in flowering under LD conditions despite the downregulation of *RFT1*, *Hd3a*, *Ehd1*, and the downstream genes such as *OsMADS14* and *OsMADS15*. This observation suggests a different mechanism involving genes other than *Hd1* or *Ehd1*. Based on the results from sequence variation, gene expression, and AS events, *OsMADS56* (LOC_Os10g39130) was considered as a potential candidate for regulating flowering time in PSRR-1 under LD conditions. Both *OsMADS56* and its closely related *OsMADS50* (LOC_Os03g03070) are orthologous to *SOC1* (*SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1*) of *Arabidopsis* (*AtSOC1*) [45] and regulate flowering-related genes downstream of *CO* [46], independent of the *Hd1* pathway. However, only *OsMADS56* showed a genotype-specific AS event and was upregulated in PSRR-1 compared with Cypress under both LD and SD conditions. Both encode MIKC-type MADS-box proteins [47]. While *OsMADS50* is a LD-specific flowering activator [48,49], *OsMADS56* functions as a negative regulator under LD conditions [50,51]. In *Arabidopsis*, *AtSOC1* acts downstream of *CO* and *FT* that eventually initiates flowering via the activation of *AtLFY* and *AtAP1* [18]. In contrast, *OsMADS56* operates upstream of flowering gene activators, *Ehd1*, *RFT1*, and *Hd3a*, which were downregulated in *OsMADS56*-overexpressing rice plants, leading to delayed flowering under LD conditions [52], unlike PSRR-1 in this study. Therefore, the early flowering or day-neutral response in PSRR-1 under LD conditions could be due to the increased expression of *OsMADS56* despite the downregulation of *Ehd1*, *RFT1*, *Hd3a*, *OsMADS14*, and *OsMADS15* and upregulation of *Hd1*. However, under SD conditions, *OsMADS56*, which is still upregulated, did not suppress the expression of *RFT1*, *Hd3a*, and *OsMADS14*. Considering the above results, we hypothesize that *OsMADS56* promoted flowering by acting directly in a similar fashion as *AtSOC1* under LD conditions (Figure 8). The function of *OsMADS56* was reversed in a similar manner as *Hd1*, which promotes flowering in rice under SD conditions [15] in contrast to CO promoting flowering under LD in *Arabidopsis* [53]. However, under SD conditions, *OsMADS56* had no effect in PSRR-1-like *AtCO* in *Arabidopsis*, in which all major flowering genes reverted to their normal expression mode. As all major flowering genes were active under SD conditions,

PSRR-1 likely switched back to its normal flowering mechanism. However, the role of *OsMADS56* under SD conditions could not be assessed, because it was also upregulated under those conditions.
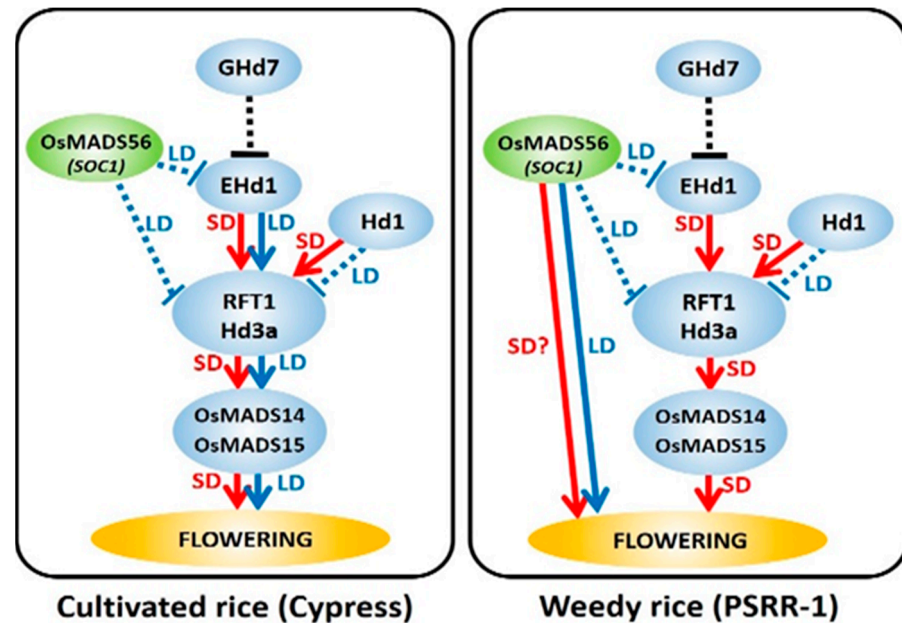
Figure 8. A simplified model for flowering in cultivated rice (Cypress) and weedy rice (PSRR-1) under long-day (LD: blue arrow lines) and short-day (SD: red arrow lines) conditions. In Cypress, *RFT1/Hd3a* signals downstream genes (*OsMADS14* and *OsMADS15*) to induce flowering. Under LD and SD, *Hd1* and *Ehd1* regulate *RFT1/Hd3a*, while *OsMADS56* acts as an upstream regulator of *Ehd1*, *RFT1*, and *Hd3a*. In weedy rice, *OsMADS56* is the major flowering activator under LD, while *RFT1/Hd3a* switches back to control flowering during SD. The role of *OsMADS56* during SD is still unclear, indicated by '?'.

This study demonstrated an approach involving both transcriptomics and WGS to analyze unique variants, expression patterns, and AS events in flowering genes, leading to the discovery of a novel flowering mechanism involving *OsMADS56* unique to weedy rice accession PSRR-1. However, further investigation is needed to gain insights into the role of *OsMADS56* and its interaction in flowering time regulation.

4. Materials and Methods

4.1. Plant Materials and Cultivation

Plant materials used in this study included SH weedy rice accession 'PSRR-1' and a rice cultivar 'Cypress'. 'Cypress' is a high-yielding long-grain rice cultivar developed from the 'L-202'/'Lemont' cross at the Louisiana Agricultural Experiment Station [54], and 'PSRR-1' is a straw-hulled US weedy rice [12]. Both PSRR-1 and Cypress are photo-insensitive or day-neutral [24]. Planting was performed under greenhouse conditions at the LSU Agricultural Center in Baton Rouge, LA (30°24′41.7″ N, 91°10′21.8″ W) around mid-April and mid-July for exposure to long- and short-day conditions, respectively.

4.2. Whole-Genome Sequencing and Analyses

The Illumina FASTQ files of PSRR-1(PRJNA413818) and Cypress (PRJNA598851) were retrieved from the sequence read archive at the National Center for Biotechnology Information (NCBI). Burrows-Wheeler Aligner (BWA; MEM-algorithm; version 0.7.8) [55,56] was used for the mapping of high-quality filtered reads against the rice reference genome (Os-Nipponbare-Reference-IRGSP-1.0), downloaded from the Rice Genome Annotation Project [57]. Genomic variant identification was performed using the Genome Analysis

Toolkit (GATK) version 4.2.3 [58], while structural variants (SV) were called using Lumpy version 0.3.1 [59]. Genome-wide coverage was estimated using SAMtools version 1.9 [54]. SnpEff version 4.3 [60] was used to predict the variant effects of SNPs, InDels, and SVs. The genome-wide distribution of DNA polymorphisms was analyzed by calculating their frequency at 10-kb intervals on each rice chromosome. R version 3.6.1 was used to visualize the distribution of SNPs and InDels on rice chromosomes. Nucleotide sequences were extracted for alignment with the consensus sequence from BAM files using IGV version 2.8.4 [61] and FastaAlternateReferenceMaker under GATK tools. Nucleotide sequences were then translated into proteins using ExPASy [62]. Alignment was performed with Jalview version 2.11.1.4 [63] using MUSCLE [64].

4.3. De Novo Assembly and Analyses of Unmapped Sequences

The unmapped reads of the PSRR-1 genome were assembled using the assembler, MaSuRCA (v3.2.1), with default options [65]. RepeatMasker (v4.0.6) [66] with default options was used to mask the assembled genome using rice as a model followed by gene prediction via Augustus v3.2.1 [67]. The annotations of predicted genes were performed using the NCBI-BLAST+ (2.11.0) [68] and UniProt [69] against plant databases. Additionally, the protein domain and family were identified based on the functional annotation by InterProScan (5.47–82.0) [70].

4.4. Comparison SNPs and InDels with 3K Rice Genome

The dataset containing 4.8 million filtered SNPs from the 3K RG was downloaded from the Rice SNP-Seek database (https://snp-seek.irri.org/) (accessed on 5 October 2020) [71]. This dataset comprised all biallelic and homozygous SNPs, which were filtered using parameters of alternative/minor allele frequency (MAF) of at least 0.01 (which covers common to low-frequency variation) and missing call proportions of at most 0.2. The SNPs were compared with PSRR-1 SNPs to enumerate the common SNPs and identify the SNPs unique to PSRR-1. The InDel dataset was downloaded from the Rice SNP-Seek database and was compared with PSRR-1 InDels. The InDels that were not called based on InDel position from the 3K RG were designated as unique InDels. From this unique dataset, the nonsynonymous and high-impact mutations that specifically caused rare amino acid variants, splice acceptor/donor variants, stop lost/gained, and start lost, were identified [60].

4.5. Agronomic Trait-Related Variants

Unique SNPs, InDels, and SVs from QTLs associated with morphological and physiological attributes were identified by co-localizing the position using the Q-TARO database (http://qtaro.abr.affrc.go.jp/) (accessed on 28 October 2020) [33]. The nonsynonymous and high-impact variants were used in the QTL co-localization to illustrate the impacts of unique alleles and SVs in PSRR-1 as a potential genomic resource.

4.6. Flowering Genes and Its Gene Signaling Network

The unique SNPs and InDels lying within the flowering genes and/or gene signaling network related to flowering between PSRR-1 and Cypress (Table S8) and structural variations with reference to Nipponbare (Table S9) were identified. The set of genes was initially obtained from the rice flowering pathway database at WikiPathways (Table S10) [72]. Subsequently, the information on differentially expressed flowering genes was used to infer the regulation pattern observed between PSRR-1 and Cypress.

4.7. RNA-Sequencing

Two sets of both Cypress and PSRR-1 were cultivated for exposure to short-day (mid-April planting) and long-day conditions (mid-July planting). These were grown under greenhouse conditions at Louisiana State University Campus Greenhouse in Baton Rouge, Louisiana (30°24′41.7″ N, 91°10′21.8″ W) with the day/night temperature set at

24 °C/13 °C, respectively. Leaf tissues from the penultimate fully expanded leaves were collected at the booting stage (~53 days after sowing). Samples were taken from three biological replicates and were stored in a −80 °C freezer until RNA extraction.

Total RNA was extracted using TRIzol (Invitrogen, CA, USA) based on the manufacturer's protocol. The quality and quantity of the extracted RNA were assessed using the ND-1000 Spectrophotometer (Thermofisher Scientific, Waltham, MA, USA) and Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). DNAse treatment was performed using PerfeCTa DNase I (QuantaBio, Beverly, MA, USA) following the manufacturer's instructions. Purified RNA samples were submitted to Novogene Corporation Inc. (Sacramento, CA, USA) for cDNA library construction and 150-bp paired-end sequencing using the Illumina Novaseq platform.

The quality of the raw reads was checked with FastQC (Babraham Bioinformatics, https://www.bioinformatics.babraham.ac.uk/) (accessed on 15 September 2021), wherein reads with a Phred quality score ≥30 were used for downstream analysis. The processed paired-end reads were mapped to the rice reference genome (Os-Nipponbare-Reference-IRGSP-1.0 downloaded from the Rice Genome Annotation Project [56,57] using HISAT2 version 2.0.1 [73]. The resulting mapped reads were processed using featureCounts [74], and differentially expressed genes (DEGs) were determined from the raw count table using the DESeq2 R package [75]. To identify the DEGs, a Benjamini–Hochberg adjusted *p*-value of <0.05 and log2 fold change $\geq$ 1.5 were set as criteria. Gene ontology was performed on the DEGs to determine the biological significance with respect to the biological processes (BPs), molecular functions (MFs), and cellular localization (CC) of their proteins using singular enrichment analysis with AgriGO v2 [76]. Plant pathway analyses were performed using Plant Reactome to identify the participation of DEGs in plant pathways affected under different conditions [77], and transcription factors were identified and classified using the PlantTFDB v5.0 database [78].

4.8. Co-localization of Flowering-Related DEGs in Heading/Flowering Date QTLs

The flowering-related DEGs identified from PSRR-1 and Cypress under both day length conditions were co-localized with published days to heading (DTH) QTLs from the Gramene QTL database (https://archive.gramene.org/qtl/) (accessed on 5 May 2020) [29] and from previous publications in weedy rice [24,31,32]. In the case of weedy-related DTH, only those QTLs linked in straw hull weedy rice mapping populations were used for co-localization. Co-localized DEGs for each QTL were illustrated using MapChart 2.3 [79].

4.9. Alternative Splicing of Flowering-Related DEGs

Transcripts for each sample mapped by HISAT2 were assembled via StringTie v.1.3.3 using the default settings [80]. The transcript assemblies were combined for each genotype per treatment through StringTie using the merge option. Subsequently, the resulting GTF files were processed using the AStalavista web server (accessed on 4 January 2022) [81] for alternative splicing (AS) event identification. The evaluation of these AS events was performed following the method of Sammeth et al. [82], which identifies events such as intron retention (IR), exon skipping (ES), alternative 3′ splice sites (A3SS), alternative 5′ splice sites (A5SS), mutually exclusive exon (MXE), and complex events (i.e., A5SS + A3SS, A5SS + ES + A3SS, and the like).

4.10. Validation of Gene Expression via qRT-PCR

To validate the gene expression patterns of selected DEGs from the RNA-seq analysis, 1.0 µg of DNase-treated RNA was used to synthesize cDNA with the iScript cDNA Synthesis Kit (Bio-Rad Laboratories, Inc., Hercules, CA, USA). The consensus sequence from the RNA-seq and published transcript sequences (Phytozome) were used for designing primers. PrimerQuest Tool (Integrated DNA Technologies, Inc., Coralville, IA, USA) was used for the primer design (Table S11) and primers were synthesized from Integrated DNA Technologies, Inc., Coralville, IA, USA.

Expression analysis of the flowering genes was performed in 96-well plates on an Applied Biosystems QuantStudio 3 Real-Time PCR system using iTaqTM Universal SYBR Green supermix (Bio-Rad Laboratories, Inc., Hercules, CA, USA). The total reaction volume was 10 μL containing the following components: 5.0 μL of 2× iTaqTM universal SYBR® Green supermix, 0.3 μL of 5 μM of each primer, 1.0 μL of 1:10 cDNA, and 3.4 μL of sterilized distilled water. PCR cycle conditions were as follows: 5 min of initial denaturation at 95 °C, 40 cycles of real-time PCR with 2-step amplification consisting of 95 °C for 15 s and 60 °C for 1 min, and a 3-step melting curve comprising 95 °C for 15 s, 60 °C for 1 min, and 95 °C for 15 s with a rate of 1.6 °C/s for each step. Reactions with no cDNA were used as a negative control. The expression level of the target genes and a reference gene (*UBQ5*) was measured in three biological replicates per tissue using gene-specific primers. Expression levels for genes were determined using the $2^{-\Delta\Delta CT}$ method [83].

References

1. Delouche, J.C.; Labrada, R.; Burgos, N.R.; Gealy, D.R. Weedy rices: Origin, biology, ecology and control. In *FAO Plant Production and Protection Paper 188*; FAO: Rome, Italy, 2007; 144p.
2. Vaughan, D.A.; Lu, B.R.; Tomooka, N. Was Asian rice (*Oryza sativa*) domesticated more than once? *Rice* **2008**, *1*, 16–24. [CrossRef]
3. Qiu, J.; Zhu, J.; Fu, F.; Ye, C.; Wang, W.; Mao, L.; Lin, X.; Chen, L.; Zhang, H.; Guo, L.; et al. Genome re-sequencing suggested a weedy rice origin from domesticated *indica-japonica* hybridization: A case study from southern China. *Planta* **2014**, *240*, 1353–1363. [CrossRef] [PubMed]
4. He, Q.; Kim, K.W.; Park, Y.J. Population genomics identifies the origin and signatures of selection of Korean weedy rice. *Plant Biotechnol. J.* **2017**, *15*, 357–366. [CrossRef] [PubMed]
5. Tong, W.; He, Q.; Park, Y.J. Genetic variation architecture of mitochondrial genome reveals the differentiation in Korean landrace and weedy rice. *Sci. Rep.* **2017**, *7*, 43327. [CrossRef] [PubMed]
6. Li, F.L.; Li, Y.L.; Jia, Y.; Caicedo, A.L.; Olsen, K.M. Signatures of adaption in weedy rice genome. *Nat. Genet.* **2017**, *49*, 5811–5816. [CrossRef]
7. Qiu, J.; Zhou, Y.; Mao, L.; Ye, C.; Wang, W.; Zhang, J.; Yu, Y.; Fu, F.; Wang, Y.; Qian, F.; et al. Genomic variations associated with local adaptation of weedy rice during de-domestication. *Nat. Commun.* **2017**, *8*, 15323. [CrossRef]
8. Sun, J.; Ma, D.; Tang, L.; Zhao, M.; Zhang, G.; Wang, W.; Song, J.; Li, X.; Liu, Z.; Zhang, W.; et al. Population genomic analysis and *de novo* assembly reveal the origin of weedy rice as an evolutionary game. *Mol. Plant* **2019**, *12*, 632–647. [CrossRef]
9. Londo, J.P.; Schaal, B.A. Origins and population genetics of weedy red rice in the USA. *Mol. Ecol.* **2007**, *16*, 4523–4535. [CrossRef]
10. Shivrain, V.K.; Burgos, N.R.; Scott, R.C.; Gbur, E.E.; Estorninos, L.E.; McClelland, M.R. Diversity of weedy red rice (*Oryza sativa* L.) in Arkansas, USA in relation to weed management. *Crop Prot.* **2010**, *29*, 721–730. [CrossRef]
11. Islam, S.; Coronejo, S.; Subudhi, P.K. Whole-genome sequencing reveals uniqueness of black-hulled and straw-hulled weedy rice genomes. *Theor. Appl. Genet.* **2020**, *133*, 2461–2475. [CrossRef]
12. Subudhi, P.K.; Parco, A.; Singh, P.K.; De Leon, T.; Karan, R.; Biradar, H.; Cohn, M.A.; Brar, D.S.; Sasaki, T. Genetic architecture of seed dormancy in US weedy rice in different genetic backgrounds. *Crop Sci.* **2012**, *52*, 2564–2575. [CrossRef]

13. Subudhi, P.K.; Singh, P.K.; De Leon, T.; Parco, A.; Karan, R.; Biradar, H.; Cohn, M.A.; Sasaki, T. Mapping of seed shattering loci provides insights into origin of weedy rice and rice domestication. *J. Hered.* **2014**, *105*, 276–287. [CrossRef] [PubMed]

14. Zhao, C.; Xu, W.; Song, X.; Dai, W.; Dai, L.; Zhang, Z.; Qiang, S. Early flowering and rapid grain filling determine early maturity and escape from harvesting in weedy rice. *Pest Manag. Sci.* **2017**, *74*, 465–476. [CrossRef] [PubMed]

15. Yano, M.; Katayose, Y.; Ashikari, M.; Yamanouchi, U.; Monna, L.; Fuse, T.; Baba, T.; Yamamoto, K.; Umehara, Y.; Nagamura, Y.; et al. *Hd1*, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the *Arabidopsis* flowering time gene CONSTANS. *Plant Cell* **2000**, *12*, 2473–2483. [CrossRef]

16. Komiya, R.; Ikegami, A.; Tamaki, S.; Yokoi, S.; Shimamoto, K. Hd3a and RFT1 are essential for flowering in rice. *Development* **2008**, *135*, 767–774. [CrossRef]

17. Zhou, S.; Zhu, S.; Cui, S.; Hou, H.; Wu, H.; Hao, B.; Cai, L.; Xu, Z.; Liu, L.; Jiang, L.; et al. Transcriptional and post-transcriptional regulation of heading date in rice. *New Phytol.* **2020**, *230*, 943–956. [CrossRef]

18. Cockram, J.; Jones, H.; Leigh, F.J.; O'Sullivan, D.; Powell, W.; Laurie, D.A.; Greenland, A.J. Control of flowering time in temperate cereals: Genes, domestication, and sustainable productivity. *J. Exp. Bot.* **2007**, *58*, 1231–1244. [CrossRef]

19. Hayama, R.; Coupland, G. The molecular basis of diversity in the photoperiodic flowering responses of *Arabidopsis* and rice. *Plant Physiol.* **2004**, *135*, 677–684. [CrossRef]

20. Matsubara, K.; Hori, K.; Tanaka-Ogiso, E.; Yano, M. Cloning of quantitative trait genes from rice reveal conservation and divergence of photoperiod flowering pathways in *Arabidopsis* and rice. *Front. Plant Sci.* **2014**, *5*, 193. [CrossRef]

21. Fowler, S.; Lee, K.; Onouchi, H.; Samach, A.; Richardson, K.; Morris, B.; Coupland, G.; Putterill, J. *GIGANTEA*: A circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. *EMBO J.* **1999**, *18*, 4679–4688. [CrossRef]

22. Hayama, R.; Yokoi, S.; Tamaki, S.; Yano, M.; Shimamoto, K. Adaptation of photoperiodic control pathways produces short-day flowering in rice. *Nature* **2003**, *422*, 719–722. [CrossRef] [PubMed]

23. Doi, K.; Izawa, T.; Fuse, T.; Yamanouchi, U.; Kubo, T.; Shimatani, Z.; Yano, M.; Yoshimura, A. *Ehd1*, a B-type response regulator in rice, confers short-day promotion of flowering and controls *FT-like* gene expression independently of *Hd1*. *Genes Dev.* **2004**, *18*, 926–936. [CrossRef] [PubMed]

24. Subudhi, P.K.; De Leon, T.B.; Tapia, R.; Chai, C.; Karan, R.; Ontoy, J.; Singh, P.K. Genetic interaction involving photoperiod-responsive *Hd1* promotes early flowering under long day condition in rice. *Sci. Rep.* **2018**, *8*, 2081. [CrossRef] [PubMed]

25. Xue, W.; Xing, Y.; Weng, X.; Zhao, Y.; Tang, W.; Wang, L.; Zhou, H.; Yu, S.; Xu, C.; Li, X.; et al. Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nat. Genet.* **2008**, *40*, 761–767. [CrossRef] [PubMed]

26. Chai, C.; Shankar, R.; Jain, M.; Subudhi, P.K. Genome-wide discovery of DNA polymorphisms by whole genome sequencing differentiates weedy and cultivated rice. *Sci. Rep.* **2018**, *8*, 14218. [CrossRef] [PubMed]

27. Lachagari, V.B.R.; Gupta, R.; Lekkala, S.P.; Mahadevan, L.; Kuriakose, B.; Chakravartty, N.; Mohan Katta, A.V.S.K.; Santhosh, S.; Reddy, A.R.; Thomas, G. Whole genome sequencing and comparative genomic analysis reveal allelic variations unique to a purple colored rice landrace (*Oryza sativa* ssp. *indica* cv. Purpleputtu). *Front. Plant Sci.* **2019**, *10*, 513. [CrossRef]

28. Subudhi, P.K.; De Leon, T.; Singh, P.K.; Parco, A.; Cohn, M.; Sasaki, T. A chromosome segment substitution library of weedy rice for genetic dissection of complex agronomic and domestication traits. *PLoS ONE* **2015**, *10*, e0130650. [CrossRef]

29. Ni, J.; Pujar, A.; Youens-Clark, K.; Yap, I.; Jaiswal, P.; Tecle, I.; Tung, C.W.; Ren, L.; Spooner, W.; Wei, X.; et al. Gramene QTL database: Development, content and applications. *Database (Oxford)* **2009**, *2009*, bap005. [CrossRef]

30. Gu, X.Y.; Foley, M.E. Epistatic interactions of three loci regulate flowering time under short and long daylengths in a backcross population of rice. *Theor. Appl. Genet.* **2007**, *114*, 745–754. [CrossRef]

31. Thurber, C.S.; Jia, M.H.; Jia, Y.; Caicedo, A.L. Similar traits, different genes? Examining convergent evolution in related weedy rice populations. *Mol. Ecol.* **2013**, *22*, 685–698. [CrossRef]

32. Qi, X.; Vigueira, C.; Liu, Y.; Caicedo, A.L. More than one way to evolve a weed: Parallel evolution of U.S. weedy rice through independent genetic mechanisms. *Mol. Ecol.* **2015**, *24*, 3329–3344. [CrossRef] [PubMed]

33. Yonemaru, J.I.; Yamamoto, T.; Fukuoka, S.; Uga, Y.; Hori, K.; Yano, M. Q-TARO: QTL annotation rice online database. *Rice* **2010**, *3*, 194–203. [CrossRef]

34. Thurber, C.S.; Reagon, M.; Olsen, K.M.; Jia, Y.; Caicedo, A.L. The evolution of flowering strategies in US weedy rice. *Am. J. Bot.* **2014**, *101*, 1737–1747. [CrossRef] [PubMed]

35. Liu, D.; Cai, X. *OsRRMh*, a Spen-like gene, plays an important role during the vegetative to reproductive transition in rice. *J. Integr. Plant Biol.* **2013**, *55*, 876–887. [CrossRef] [PubMed]

36. Lu, Q.; Xu, Z.K.; Song, R.T. *OsFY*, a homolog of *AtFY*, encodes a protein that can interact with OsFCA-gamma in rice (*Oryza sativa* L.). *Acta Biochem. Biophys. Sin.* **2006**, *38*, 492–499. [CrossRef] [PubMed]

37. Ma, X.; Feng, F.; Wei, H.; Mei, H.; Xu, K.; Chen, S.; Li, T.; Liang, X.; Liu, H.; Luo, L. Genome-wide association study for plant height and grain yield in rice under contrasting moisture regimes. *Front. Plant Sci.* **2016**, *7*, 1801. [CrossRef] [PubMed]

38. Van der Weide, R.H.; Simonis, M.; Hermsen, R.; Toonen, P.; Cuppen, E.; de Ligt, J. The genomic scrapheap challenge; extracting relevant data from unmapped whole genome sequencing reads, including strain specific genomic segments, in rats. *PLoS ONE* **2016**, *11*, e0160036. [CrossRef] [PubMed]

39. Nadir, S.; Xiong, H.B.; Zhu, Q.; Zhang, X.L.; Xu, H.Y.; Li, J.; Dongchen, W.; Henry, D.; Guo, X.Q.; Khan, S.; et al. Weedy rice in sustainable rice production: A review. *Agron. Sustain. Dev.* **2017**, *37*, 46. [CrossRef]

40. Reagon, M.; Thurber, C.S.; Gross, B.L.; Olsen, K.M.; Jia, Y.; Caicedo, A.L. Genomic patterns of nucleotide diversity in divergent population of U.S. weedy rice. *BMC Evol. Biol.* **2010**, *10*, 180. [CrossRef]

41. Dong, C.; He, F.; Berkowitz, O.; Liu, J.; Cao, P.; Tang, M.; Shi, H.; Wang, W.; Shen, Z.; Whelan, J.; et al. Alternative splicing plays a critical role in maintaining mineral nutrient homeostasis in rice (*Oryza sativa*). *Plant Cell* **2018**, *30*, 2267–2285. [CrossRef]

42. Chen, M.X.; Zhu, F.Y.; Wang, F.Z.; Ye, N.H.; Gao, B.; Chen, X.; Zhao, S.S.; Fan, T.; Cao, Y.Y.; Liu, T.Y.; et al. Alternative splicing and translation play important roles in hypoxic germination in rice. *J. Exp. Bot.* **2019**, *70*, 817–833. [CrossRef] [PubMed]

43. Yang, H.; Li, P.; Jin, G.; Gui, D.; Liu, L.; Zhang, C. Temporal regulation of alternative splicing events in rice memory under drought stress. *Plant Divers.* **2020**. [CrossRef]

44. Xia, W.; Liu, J.; Zhang, J.; Mason, A.S.; Li, Z.; Gong, S.; Zhong, Y.; Dou, Y.; Sun, X.; Fan, H.; et al. Alternative splicing of flowering time gene *FT* is associated with halving of time to flowering in coconut. *Sci. Rep.* **2020**, *10*, 11640. [CrossRef] [PubMed]

45. Nam, J.; Kaufmann, K.; Theissen, G.; Nei, M. A simple method for predicting the functional differentiation of duplicate genes and its application to MIKC-type MADS-box genes. *Nucleic Acids Res.* **2005**, *33*, e12. [CrossRef]

46. Onouchi, H.; Igeno, M.I.; Perilleux, C.; Graves, K.; Coupland, G. Mutagenesis of plants overexpressing CONSTANS demonstrates novel interactions among *Arabidopsis* flowering-time genes. *Plant Cell* **2000**, *12*, 885–900. [CrossRef]

47. Kaufmann, K.; Melzer, R.; Theissen, G. MIKC-type MADS domain proteins: Structural modularity, protein interactions and network evolution in land plants. *Gene* **2005**, *347*, 183–198. [CrossRef]

48. Tadege, M.; Sheldon, C.C.; Helliwell, C.A.; Upadhyaya, N.M.; Dennis, E.S.; Peacock, J. Reciprocal control of flowering time by *OsSOC1* in transgenic *Arabidopsis* and by *FLC* in transgenic rice. *Plant Biotechnol. J.* **2003**, *1*, 361–369. [CrossRef]

49. Lee, S.; Kim, J.; Han, J.J.; Han, M.J.; An, G. Functional analyses of the flowering time gene *OsMADS50*, the putative *SUPPRESSOR OF OVEREXPRESSION OF CO 1/AGAMOUS-LIKE 20 (SOC1/AGL20)* ortholog in rice. *Plant J.* **2004**, *38*, 754–764. [CrossRef]

50. Riechmann, J.L.; Meyerowitz, E.M. MADS domain proteins in plant development. *Biol. Chem.* **1997**, *378*, 1079–1101.

51. Alvarez-Buylla, E.R.; Pelaz, S.; Liljegren, S.J.; Gold, S.E.; Burgeff, C.; Ditta, G.S.; de Pouplana, L.R.; Martinez-Castilla, L.; Yanofsky, M.F. An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5328–5333. [CrossRef]

52. Ryu, C.H.; Lee, S.; Cho, L.H.; Kim, S.L.; Lee, Y.S.; Choi, S.C.; Jeong, H.J.; Yi, J.; Park, S.J.; Han, C.D.; et al. *OsMADS50* and *OsMADS56* function antagonistically in regulating long day (LD)-dependent flowering in rice. *Plant Cell Environ.* **2009**, *32*, 1412–1427. [CrossRef] [PubMed]

53. Putterill, J.; Robson, F.; Lee, K.; Simon, R.; Coupland, G. The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* **1995**, *80*, 847–857. [CrossRef]

54. Linscombe, S.D.; Jodari, F.; Mckenzie, K.S.; Bollich, P.K.; White, L.M.; Growth, D.E.; Dunand, R.T. Registration of Cypress rice. *Crop Sci.* **1993**, *33*, 355. [CrossRef]

55. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]

56. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **2009**, *25*, 1754–1760. [CrossRef] [PubMed]

57. Kawahara, Y.; de la Bastide, M.; Hamilton, J.P.; Kanamori, H.; McCombie, W.R.; Ouyang, S.; Schwartz, D.C.; Tanaka, T.; Wu, J.; Zhou, S.; et al. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **2013**, *6*, 4. [CrossRef] [PubMed]

58. Van Der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, 11.10.1–11.10.33.

59. Layer, R.M.; Chiang, C.; Quinlan, A.R.; Hall, I.M. LUMPY: A probabilistic framework for structural variant discovery. *Genome Biol.* **2014**, *15*, R84. [CrossRef] [PubMed]

60. Cingolani, P.; Platts, A.; Wang, L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.; Lu, X.; Ruden, D. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **2012**, *6*, 80–92. [CrossRef] [PubMed]

61. Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [CrossRef]

62. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.D.; Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31*, 3784–3788. [CrossRef] [PubMed]

63. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.A.; Clamp, M.; Barton, G.J. Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [CrossRef] [PubMed]

64. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef] [PubMed]

65. Zimin, A.V.; Marçais, G.; Puiu, D.; Roberts, M.; Salzberg, S.L.; Yorke, J.A. The MaSuRCA genome assembler. *Bioinformatics* **2013**, *29*, 2669–2677. [CrossRef]

66. Smit, A.F.A.; Hubley, R.; Green, P. RepeatMasker Open-4.0. 2015. Available online: http://www.repeatmasker.org/ (accessed on 15 September 2020).

67. Stanke, M.; Diekhans, M.; Baertsch, R.; Haussler, D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* **2008**, *24*, 637–644. [CrossRef]

68. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2008**, *10*, 421. [CrossRef]
69. The UniProt Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, 506–515. [CrossRef]
70. Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; John Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240. [CrossRef]
71. Mansueto, L.; Fuentes, R.R.; Borja, F.N.; Detras, J.; Abriol-Santos, J.M.; Chebotarov, D.; Sanciangco, M.; Palis, K.; Copetti, D.; Poliakov, A.; et al. Rice SNP-seek database update: New SNPs, indels, and queries. *Nucleic Acids Res.* **2017**, *45*, D1075–D1081. [CrossRef]
72. Slenter, D.N.; Kutmon, M.; Hanspers, K.; Riutta, A.; Windsor, J.; Nunes, N.; Mélius, J.; Cirillo, E.; Coort, S.L.; Digles, D.; et al. WikiPathways: A multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.* **2018**, *46*, D661–D667. [CrossRef]
73. Kim, D.; Landmead, B.; Salzberg, S.L. HISAT: A fast-spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [CrossRef] [PubMed]
74. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. [CrossRef] [PubMed]
75. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [CrossRef] [PubMed]
76. Tian, T.; Liu, Y.; Yan, H.; You, Q.; Yi, X.; Du, Z.; Xu, W.; Su, Z. agriGO v2.0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* **2017**, *45*, W122–W129. [CrossRef]
77. Naithani, S.; Gupta, P.; Preece, J.; D'Eustachio, P.; Elser, J.; Garg, P.; Dikeman, D.A.; Kiff, J.; Cook, J.; Olson, A.; et al. Plant Reactome: A knowledgebase and resource for comparative pathway analysis. *Nucleic Acids Res.* **2019**, *48*, D1093–D1103. [CrossRef]
78. Jin, J.P.; Tian, F.; Yang, D.C.; Meng, Y.Q.; Kong, L.; Luo, J.C.; Gao, G. PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* **2017**, *45*, D1040–D1045. [CrossRef]
79. Voorrips, R.E. MapChart: Software for the graphical presentation of linkage maps and QTLs. *Heredity* **2002**, *93*, 77–78. [CrossRef]
80. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **2015**, *33*, 290–295. [CrossRef]
81. Foissac, S.; Sammeth, M. ASTALAVISTA: Dynamic and flexible analysis of alternative splicing events in custom gene datasets. *Nucleic Acids Res.* **2007**, *35*, W297–W299. [CrossRef]
82. Sammeth, M.; Foissac, S.; Guigo, R. A general definition and nomenclature for alternative splicing events. *PLoS Comput. Biol.* **2008**, *4*, e1000147. [CrossRef]
83. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* **2001**, *25*, 402–408. [CrossRef] [PubMed]