Article

The Landscape of Presence/Absence Variations during the Improvement of Rice

Xia Zhou [1,†], Chenggen Qiang [2,†], Lei Chen [3], Dongjin Qing [3], Juan Huang [3], Jilong Li [2,*] and Yinghua Pan [3,*]

[1] Urban Construction School, Beijing City University, Beijing 101300, China; whzhouxia@163.com
[2] State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China; qiangcg@ibcas.ac.cn
[3] Guangxi Key Laboratory of Rice Genetics and Breeding, Rice Research Institute, Guangxi Academy of Agricultural Sciences, Nanning 530007, China; chenlei@gxzaas.net (L.C.); qingdongjin169@gxaas.net (D.Q.); huangjuan@gxaas.net (J.H.)
[*] Correspondence: lijilong@ibcas.ac.cn (J.L.); panyinghua@gxaas.net (Y.P.)
[†] These authors contributed equally to this work.

Abstract: Rice is one of the most important staple crops in the world; therefore, the improvement of rice holds great significance for enhancing agricultural production and addressing food security challenges. Although there have been numerous studies on the role of single-nucleotide polymorphisms (SNPs) in rice improvement with the development of next-generation sequencing technologies, research on the role of presence/absence variations (PAVs) in the improvement of rice is limited. In particular, there is a scarcity of studies exploring the traits and genes that may be affected by PAVs in rice. Here, we extracted PAVs utilizing resequencing data from 148 improved rice varieties distributed in Asia. We detected a total of 33,220 PAVs and found that the number of variations decreased gradually as the length of the PAVs increased. The number of PAVs was the highest on chromosome 1. Furthermore, we identified a 6 Mb hotspot region on chromosome 11 containing 1091 PAVs in which there were 29 genes related to defense responses. By conducting a genome-wide association study (GWAS) using PAV variation data and phenotypic data for five traits (flowering time, plant height, flag leaf length, flag leaf width, and panicle number) across all materials, we identified 186 significantly associated PAVs involving 20 cloned genes. A haplotype analysis and expression analysis of candidate genes revealed that important genes might be affected by PAVs, such as the flowering time gene *OsSFL1* and the flag leaf width gene *NAL1*. Our work investigated the pattern in PAVs and explored important PAV key functional genes associated with agronomic traits. Consequently, these results provide potential and exploitable genetic resources for rice breeding.

Keywords: rice improvement; presence/absence variations; genome-wide association study

1. Introduction

Asian cultivated rice (*Oryza sativa* L.) is an important crop globally which consists of two subspecies, *indica* and *japonica*. More than half of the world's population relies on Asian cultivated rice as a primary food source [1]. As one of the oldest domesticated cereals, rice has undergone two major processes, domestication and improvement, in its transformation from wild ancestors to cultivated varieties [2]. The former refers to the shift from wild rice to landraces, while the latter denotes to the transition from landraces to modern breeding [3]. Although the improvement of rice has gone been conducted for only a few decades, strong artificial selection has exerted a substantial influenced on the agronomic traits of rice [3]. Among these agronomic traits, flowering time plays a crucial role in rice improvement. Firstly, the flowering time directly affects rice growth, exhibiting a direct correlation with yield [4,5]. Secondly, an appropriate flowering time can enhance the adaptability of rice [6,7]. Flag leaf shape, another important trait, has an impact on rice yield and quality. Some studies have shown that there is a positive correlation between flag

leaf length and the panicle number, grains, and yield of rice [8–10]. As for flag leaf width, a wider flag leaf can enhance photosynthesis and nutrient absorption, thus improving the quality of rice [10,11]. Therefore, it is of importance to focus on agronomic traits such as flowering time and flag leaf shape when improving rice.

Significant achievements have been made in understanding the genetic basis of various traits in rice in recent years [12,13], especially with the sequencing of vast of rice materials. For example, the sequencing of 3010 rice genomes has provided us with abundant genomic resources [14]. Utilizing these data, extensive studies have been performed on rice domestication and improvement, including population genetic analyses and gene function exploration [15–17]. So far, research studies have mainly focused on the significant role of SNPs in rice improvement [18]. Compared to SNPs, structural variations (>50 bp) generate larger and more diverse types of variations, thus exerting a greater impact on the genome [19]. Structural variations are mainly classified as insertions, deletions, copy number variations, inversions, and translocations [20]. Among them, insertions and deletions (presence/absence variations (PAVs)) are easier to detect due to their low complexity, especially when using short-read sequencing data. Although some studies have analyzed the role of PAVs in rice improvement [21–27], reports on the functional genes affected by them are still limited [23,24,26,27]. For instance, Li et al. reported that 1116 bp deletions in the *DTH8* gene affect the rice heading date [28]. Additionally, 520 bp deletions in the promoter of *DNR1* in HJX74 decreased its expression levels and improved nitrogen uptake rates [29]. Qin et al. found that an insertion at 643 bp upstream of *OsGLP2-1* in the *japonica* background was related to seed dormancy [23]. These findings indicated that PAVs play a crucial role in the phenotypic diversity of rice. Therefore, it is particularly important to investigate the role of PAVs and the potential genes associated with agronomic traits in rice improvement.

In this study, we conducted a comprehensive PAV analysis using resequencing data from 148 modern rice varieties in Asia. We analyzed the variation patterns of PAVs in the genome, including the number, size, and distribution. Combined with five agronomic traits (days to heading, panicle height, panicle number, flag leaf length, and flag leaf width), we performed a GWAS and identified some significant candidate genes. Our results explored the role of PAVs in rice improvement and uncovered important genes related to PAVs. These findings will provide genetic resources and breeding materials for the improvement of rice.

2. Materials and Methods

2.1. Materials and Phenotype

We selected 148 improved rice cultivars distributed in Asian countries, including 109 *indica* and 39 *japonica* cultivars, from a 3000-rice-genomes project based on the Genesys website (https://www.genesys-pgr.org/, accessed on 30 October 2023) and reports in the literature [3] in which information about the improved rice cultivars was easier to distinguish. Phenotypic data for 148 rice accessions were obtained from RFGB v2.0 (Rice Functional Genomics & Breeding, https://www.rmbreeding.cn/, accessed on 30 October 2023) [30]. Considering the consistency of the phenotypes, we selected five traits (days to heading, plant height, flag leaf length, flag leaf width, and panicle number) measured in Hainan Province, China, in 2018 for further analysis. Each rice cultivar contained at least three replications.

2.2. Sequence Data

The sequence data for the 148 rice accessions were downloaded from the 3000 Rice Genomes Project [14]. Adapters and low-quality sequences of raw reads were removed using Trimmomatic (v0.36) [31]. We used BWA (v0.7.10) and SAMtools (v1.1) to build an index for the Nipponbare reference genome (IRGSP-1.0) [32,33]. Clean reads were mapped to the Nipponbare genome using BWA (v0.7.17-r1188) and the MEM algorithm. Using Genome Analysis Toolkit 4 (GATK4, v4.1, https://gatk.broadinstitute.org/hc/en-us,

accessed on 2 November 2023), we transformed the format of the alignment file from SAM to BAM, sorted the aligned reads, masked duplicate reads, and built an index for each BAM file. The number of mapping reads and sequencing depths were calculated using SAMtools (v1.1) [33] and GATK3 (v3.8), respectively.

2.3. PAV Calling

PAVs (presence/absence variations) were identified using Delly (v0.8.7) and Manta (v1.6.0) using mapping results from resequencing data in a BAM format [34,35]. When using Delly, we ran PAVs, calling on each accession, and then merged the results into one VCF file as a guiding reference. Next, we ran PAVs again, calling with the guidance of the combined VCF file. Furthermore, we filtered the PAVs with PASS tag for further analysis. Finally, PAVs from all accessions were combined using BCFtools (v1.9) [36]. For Manta, we used the recommended workflow to perform PAV calling for each individual. SURVIVOR (v1.0.7) [37] was used to merge all PAV calling files. For more comprehensive PAV data, we merged the PAV calls detected by Delly and Manta using an in-house Perl script. PAVs from all rice cultivars were filtered using VCFtools (v0.1.16) [38] based on the following criteria: (1) minor allele frequencies (MAFs ≥ 0.01); (2) missing rate < 40% [27,39].

2.4. Distribution of PAVs Relative to Gene Position

According to the PAV breakpoints and gene locations in the genome, we classified the PAVs into five categories as follows: overlapping with (1) ± 3 kb of gene regions (3 kb upstream of the start codon and 3 kb downstream of the stop codon), (2) coding and UTR regions, (3) introns, (4) and intergenic regions. The percentage of each category was calculated.

2.5. Identification of PAV Hotspot Regions

We calculated the distribution of PAV breakpoints for each 1000 kb window with a 500 kb step size along each chromosome. Then, according to the number of PAVs within the window, all 1000 kb windows were ranked in descending order. The top 10% of windows with the highest frequency of PAV breakpoints were defined as PAV hotspots. Furthermore, all continuous hotspot windows were merged into a hotspot region [23].

2.6. GWAS and Genetic Correlation Analysis of PAVs

A genome-wide association analysis (GWAS) of PAVs and five phenotypic types of data (days to heading, plant height, flag leaf length, flag leaf width, and panicle number) were performed using BLINK-R implemented in GAPIT [40,41]. The first three principal components from GAPIT were used as covariates to control the population structure. The proportion of variance explained by each component was computed by dividing the variance in each component (eigenvalue) by the total variance in the dataset. To determine the significance threshold for each trait, we applied a conditional permutation method, as described in Zhao et al. [42]. Considering that BLINK-C is about 20 times faster than BLINK-R [41], we performed 1000 conditional permutation tests with BLINK-C to improve computational efficiency. And 95th quantile of -log (p) was selected as the threshold for each trait, respectively. As for the statistical analysis of different haplotypes, comparisons were conducted using Student's t-test in R [43].

3. Results

3.1. PAV Identification

To explore the role of structural variations in the improvement of cultivated rice distributed in Asia, we downloaded the whole-genome sequencing data of 148 rice varieties from the 3K database, including 109 *indica* and 39 *japonica* varieties (Table S1). Reads from all rice varieties were aligned to the Nipponbare (IRGSP1.0) genome. Due to the short reads of the sequencing data and the complexity of the structural variations found, we selected two software, Manta (v1.6.0) and Delly (v0.8.7), to call structural variations, focusing only

on deletions and insertions (presence/absence variations (PAVs)). We assessed the coverage and sequencing depth of the resequencing data for rice varieties, revealing an average sequencing depth of 14× (5.22–32.64×) and an average coverage of 91.51% (83.2–98.7%) (Table S2). A total of 33,220 PAVs were called, including 32,237 deletions and 983 insertions. Furthermore, an analysis of the PAVs' distribution in terms of size revealed that the number of variations decreased gradually as the length of the PAVs increased (Figure 1). Short PAVs (50–1000 bp) accounted for 74.6%, indicating that short PAVs may play a predominant role in the improvement of rice cultivars.
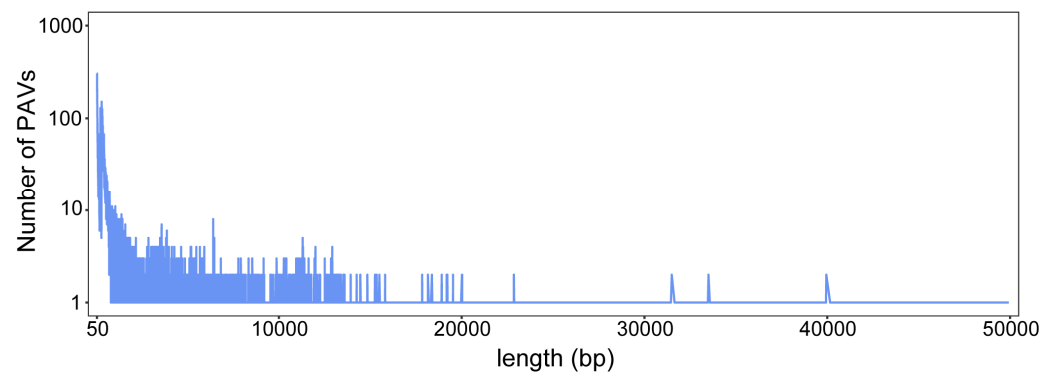
Figure 1. The number of PAVs of different sizes.

3.2. Distribution of PAVs

To study the distribution pattern of the PAVs, we first analyzed the number of PAVs on chromosomes and found an average of 2768 PAVs per chromosome (Figure 2a). The number of PAVs was the highest on chromosome 1, with 3887, while on chromosome 5, the number was the lowest, with only 1996 (Figure 2a). Furthermore, we found uneven distributions of PAVs along chromosomes and identified 27 PAV hotspots region, suggesting the occurrence of multiple independent PAVs in these regions (Figure 2b). Interestingly, we identified a 6 Mb hotspot region on chromosome 11 containing 1091 PAVs. It revealed that there were 29 genes related to defense responses within this region. Consistent with previous studies [23], these findings indicated that variations located in PAV hotspots are more likely to play a role in the environmental adaptation of cultivated rice compared to other genomic regions. Subsequently, we examined the distribution of PAVs across the genome, and the results showed that 67.4% of PAVs were in genic regions (3 kb upstream and downstream of genes, UTR regions, coding regions, and introns), while 46.1% of PAVs were distributed in intergenic regions (Figure 2c). Within the genic regions, 72.6% of PAVs were found in the 3 kb upstream and downstream regions of genes, suggesting that PAVs in regulatory regions are more likely kept and may play important roles in the improvement of rice.

3.3. The Impact of PAVs on Phenotypes

To investigate the impact of PAVs on phenotypes in rice improvement, we downloaded the data of 148 rice materials with respect to five phenotypes, the days to heading, plant height, panicle number, flag leaf length, and flag leaf width, from the RFGB database (Table S3). We analyzed the distribution of these trait among the 148 materials. The average flowering time was 96.8 days, with the shortest being 68 days and the longest being 128.5 days. The average plant height was 91.3 cm, with the shortest being 55.5 cm and the tallest being 148.3 cm. The average number of panicles was 9.2, with the lowest being 3.7 and the highest being 14.5. The average flag leaf length was 26.9 cm, with the shortest being 16.5 cm and the longest being 40.6 cm. The average width of the flag leaf was 1.4 cm, with the narrowest being 0.96 cm and the widest being 2.1 cm. Then, based on the phenotypic statistics of each individual, we found that the 148 materials exhibited a continuous and approximately normal distribution for the five traits (Figure S1), suggesting that each trait is a quantitative trait controlled by multiple genes. These results indicated

that the selected samples were representative with large variations and abundant genetic information resources. To explore the role of PAVs in these traits of rice improvement, we used 33,220 PAVs as genotypes and conducted a genome-wide association analysis for the five traits. Considering the potential impact of population structure (Figure S2), we ultimately selected the BLINK model in the GAPIT (Version 3) software package for subsequent analyses. In addition, to accurately identify outlier loci, we performed 1000 permutation tests for each trait, employing a significance threshold of α 0.05. As shown in the figures, we found a total of 29 significantly associated PAVs for flowering time, 9 for plant height, 37 for panicle number, 82 for flag leaf length, and 29 for flag leaf width (Figures 3a, 4a, S3a and S4a,c). These results indicate that PAVs play an important role in the improvement of agronomic traits in rice.
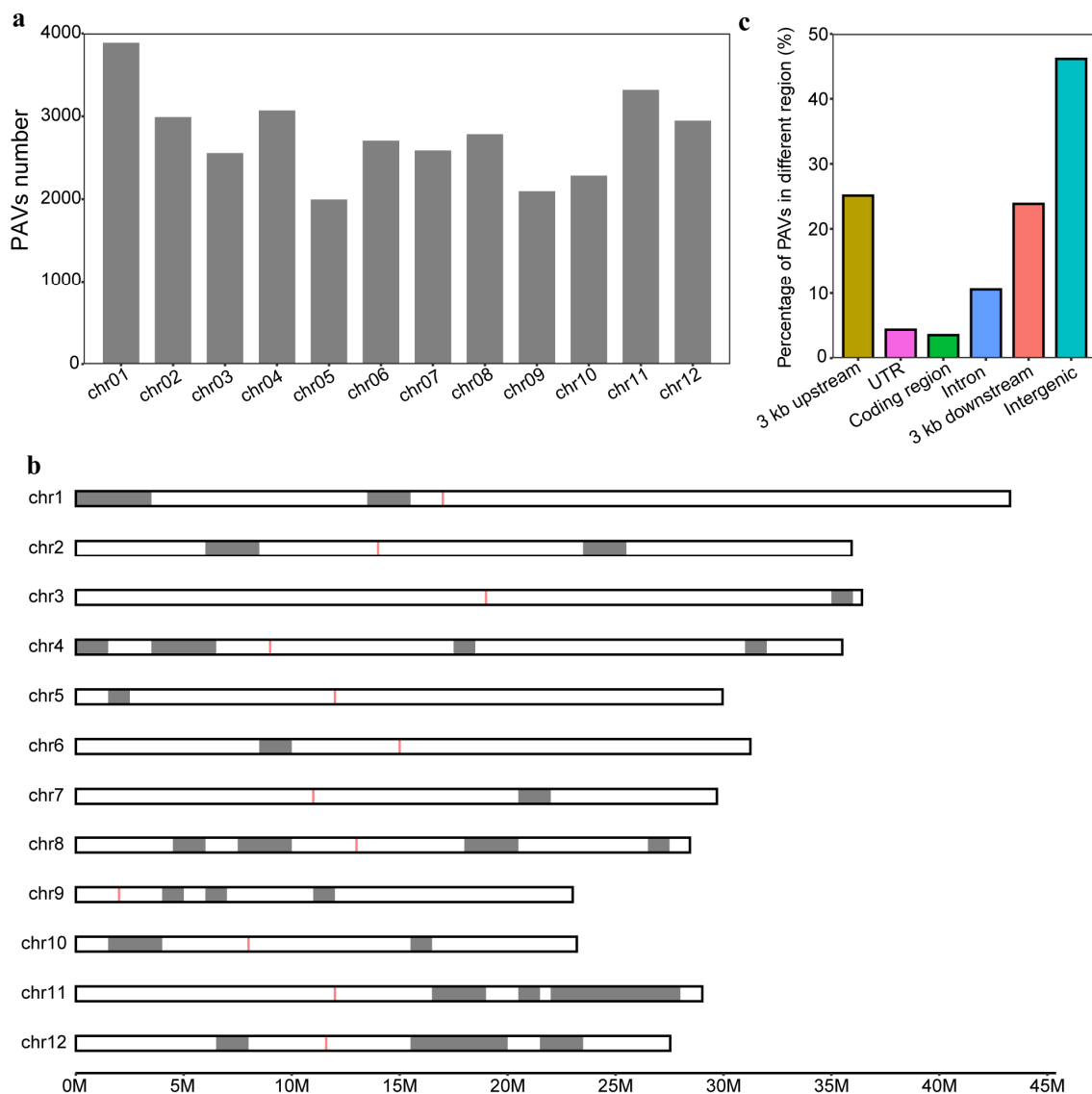
Figure 2. Distribution pattern of PAVs. (**a**) Number of PAVs on different chromosomes. (**b**) Distribution of hotspot regions for PAVs on chromosomes. Red boxes (**b**) represent centromere region. (**c**) Percentage of PAVs overlapping with different genomic regions.
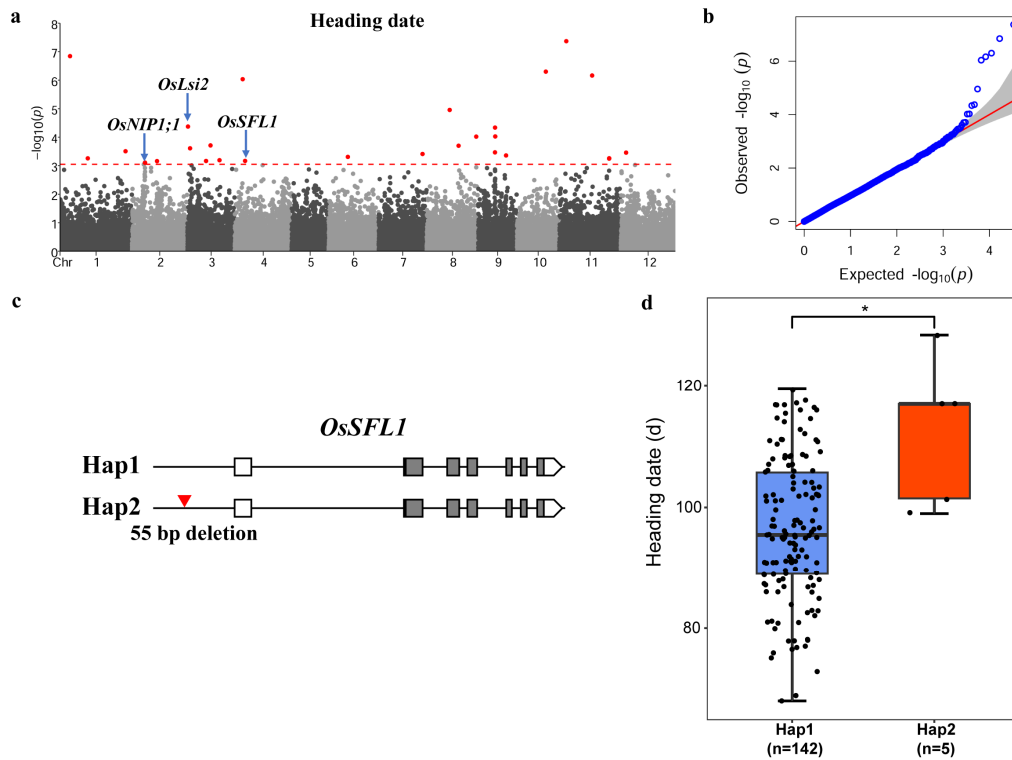
Figure 3. GWAS analysis for heading date. (**a**,**b**) Manhattan plot and quantile–quantile (QQ) plots of PAV and heading date association in 148 rice varieties. Red dots indicate a significant correlation with phenotype, and red dashed lines show threshold (**a**). (**c**) A 55 bp deletion occurred within *OsSFL1*, exhibiting two haplotypes. (**d**) Comparison of heading date in Hap1 and Hap2. Statistical significance was calculated using Student's *t*-test; * *p* < 0.05.
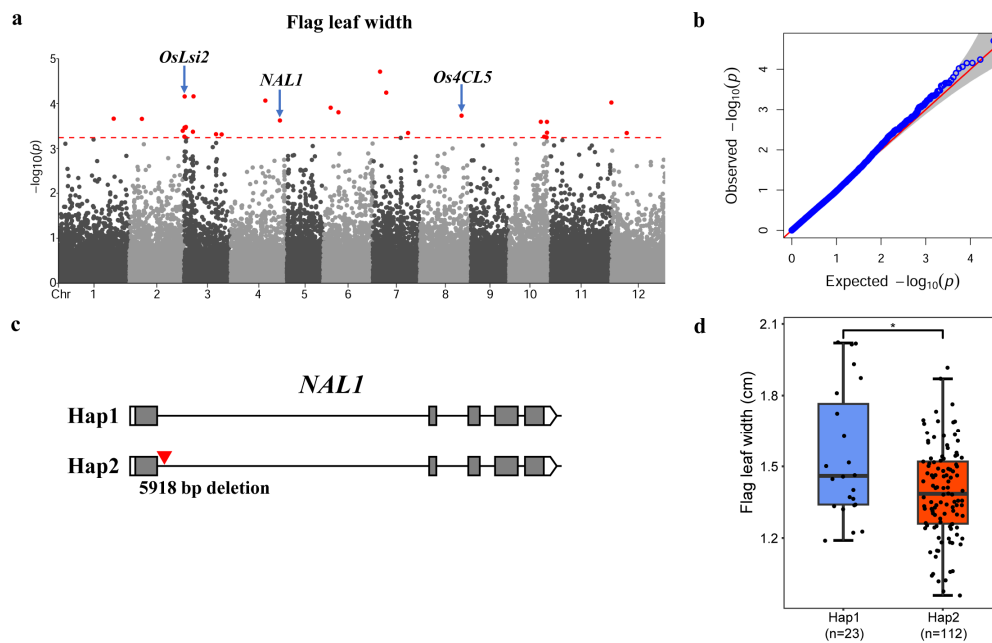

Figure 4. GWAS analysis for flag leaf width. (**a**,**b**) Manhattan plot and quantile–quantile (QQ) plots of PAV and flag leaf width association in 148 rice varieties. Red dots indicate significant correlation with phenotype, and red dashed lines show threshold (**a**). (**c**) A 5918 bp deletion occurred within the *NAL1*, exhibiting two haplotypes. (**d**) Comparison of heading date in Hap1 and Hap2. The statistical significance was calculated by Student's *t*-test. * *p* < 0.05.

3.4. Important Genes Associated with PAVs

To explore the genes affected by PAVs in rice improvement, we further analyzed the genes corresponding to significant PAVs according to the annotation information (Table S4). We selected genes based on the following criteria: (1) significant PAVs located in the gene region, including the 3 kb upstream and downstream regions, UTR region, coding region, and intron region, and (2) genes with functions reported in the literature. By integrating the positions of significant PAVs and gene functions, we focused on genes that regulate flowering time and flag leaf width. For flowering time, three cloned genes, *OsNIP1;1*, *OsLsi2*, and *OsSFL1* contained significant PAVs [44–47]. Previous studies have shown that the *OsSFL1* gene positively regulates flowering time in rice. The flowering time of the *ossfl1* mutant was delayed by nearly two weeks compared to the wild-type Nipponbare, and the *ossfl1-T* mutant was delayed by approximately 17 days compared to the wild-type DJ [45]. In this study, we found that some materials had a 55 bp deletion at 615 bp upstream of *OsSFL1* compared to the reference genome (Figure 3c). Furthermore, we divided the materials into two haplotypes based on this deletion. A statistical analysis revealed that materials with the 55 bp deletion variant exhibited significantly delayed flowering times compared to materials without the deletion (Figure 3d). Therefore, we inferred that the 55 bp deletion might delay flowering time by reducing the expression level of *OsSFL1*. For the flag leaf width, significant PAVs involved three cloned genes, *OsLsi2*, *NAL1*, and *Os4CL5* [44,47,48]. Previous research showed that *NAL1* positively regulates leaf width, and the null mutant exhibits reduced numbers of longitudinal veins and narrow leaves [48,49]. In our study, we found that some materials had a 5918 bp deletion in the first intron of *NAL1* compared to the reference genome (Figure 4c). Furthermore, we divided the materials into two haplotypes according to the deletion. A statistical analysis indicated that materials with the 5918 bp deletion variant exhibited significantly smaller flag leaf widths compared to materials without the deletion (Figure 4d). Subsequently, we obtained expression data of *NAL1* in two materials, Nip (Hap1) and MH63 (Hap2), from previous research [50]. We found that the expression level of *NAL1* in the mature leaves of MH63 was significantly lower than that in Nip, suggesting that the 5918 bp deletion in the first intron of *NAL1* might reduce the expression of *NAL1* in leaves, thereby decreasing the flag leaf width (Figure S5). The above results provide substantial evidence that PAVs contributed to phenotypic variation in agronomic traits in modern rice and played a crucial role in the breeding improvement of rice.

4. Discussion

4.1. The Distribution Pattern of PAVs on the Genome

In recent years, with the application of sequencing technology, there have been increasing reports on PAVs in plants. However, less attention has been paid to the distribution patterns of PAVs on the genome. Previous research has identified a widespread distribution of PAVs on the genome without further analyzing their distribution patterns [51–54]. Only a few studies have found that PAVs exhibited a clustered distribution on the genome. In *Mytilus galloprovincialis*, PAVs showed an uneven distribution on chromosomes [55]. Here, we also found similar results, with the presence of multiple hotspot regions. Studies on PAVs in plants have discovered that certain gene families, such as defense response-related genes [51,56], contain more PAVs suggesting that the distribution of these genes on chromosomes might relate to the formation of PAV hotspots. Furthermore, we found that most PAVs are distributed in intergenic regions or regulatory regions. Consistent with our findings, Göktay et al. reported that the number of PAVs in intergenic regions is much greater than in genic regions in *Arabidopsis* [57]. Genic regions are more conserved compared to intergenic regions, and newly generated PAVs are more likely deleterious. Therefore, PAVs in regulatory regions might affect phenotypes by changing gene expression levels in a more subtle manner, thus avoiding direct effects on gene function and making it easier to retain during the improvement of rice.

4.2. PAVs Influenced Multiple Agronomic Traits

PAVs played an important role in the breeding and improvement of crops including rice, maize, and wheat [51,53,58]. In rice, several studies have shown that PAVs affected phenotypes by altering the functions of genes at the expression or protein level. For example, Shang et al. (2022) showed that a 127 bp insertion in the upstream of *HGW* was associated with the expression level of the gene and decreased the thousand-grain weight compared to materials without the insertion [24]. Qin et al. reported a 66.6 kb or 43.3 kb deletion in rice which includes a known negatively regulated disease-resistant gene, *WAK112d*, suggesting that the large fragment deletion may enhance the disease resistance of these materials [23,59]. In this study, we revealed significant association signals for five traits using a genome-wide association analysis involving multiple cloned genes. In addition to the previously mentioned flowering time and flag leaf width, we also analyzed the flag leaf length and found that significant PAVs involved seven cloned genes: *OsMFT2*, *OsPP1a*, *OsPHT1*, *OsPP2A-1*, *OsLLB*, *OsCAB1R*, and *OsDEP1* [60–66] (Figure S3a). Previous studies showed that *OsDEP1* regulates panicle shape, plant height, yield, and nitrogen response in rice [62,67]. Our study discovered that some materials had a 637 bp deletion in the fifth exon of *OsDEP1* compared to the reference genome (Figure S3c). Furthermore, based on this deletion, the materials were divided into two haplotypes. A genetic correlation analysis showed that the flag leaf length of materials with the 637 bp deletion was significantly shorter than that of non-deletion materials, suggesting that the 637 bp deletion might reduce flag leaf length by altering the structure of the *OsDEP1* protein (Figure S3d). Interestingly, we found a significant locus that is shared between the flowering time and flag leaf width, corresponding to the cloned gene *OsLsi2* [44] (Figures 3a and 4a). Compared to the reference genome, some materials had a 303 bp deletion in the upstream of the gene. Previous studies had reported that the growth of *lsi2* mutants is inhibited under field conditions, with a slightly reduced plant height and changed glume color, and its yield was only 40% of that of the wild type [44], suggesting that *OsLsi2* might be a pleiotropic gene and the 303 bp deletion possibly affects both flowering time and flag leaf width by changing the expression level of *OsLsi2*.

4.3. PAVs Play an Important Role in Rice Improvement

With the release of large-scale sequencing data in rice in recent years, an increasing number of PAVs have been discovered, and the impact of PAVs on traits has attracted attention. For example, a researcher found that PAVs could participate in the combination of heterosis genotypes for two-line and three-line hybrid rice by influencing gene expression patterns [53]. Another study on hybrid incompatibility between rice subspecies showed that negatively interacting genes fix different PAVs in different subspecies, causing incompatibility during hybridization [68]. Many studies have discovered that PAVs are related to disease resistance in plants including in sorghum, *Brassica napus*, *Arabidopsis*, and rice [56,57,68,69]. Furthermore, it was discovered by an association analysis that many PAVs were an important source of phenotypic variation in crops [70]. Therefore, uncovering PAVs related to phenotypic variation will provide more usable genetic information for breeding. Here, we conducted GWAS using PAVs data and phenotypic data for five traits (flowering time, plant height, flag leaf length, flag leaf width, and panicle number), identifying 186 significantly associated PAVs. Considering that the varieties selected in this study were all improved varieties, it is implied that these PAVs play significant roles in rice improvement. Additionally, our study revealed new variants for some known functional genes, including *OsSFL1*, *NAL1*, and *OsLsi2*, which will provide important gene resources for rice breeding.

5. Conclusions

Our work investigated the pattern of PAVs and explored important PAVs' key functional genes associated with agronomic traits. Consequently, these results provide potentially exploitable genetic resources for rice breeding.

References

1. Ahmadi, N.; Audebert, A.; Bennett, M.J.; Bishopp, A.; de Oliveira, A.C.; Courtois, B.; Diedhiou, A.; Diévart, A.; Gantet, P.; Ghesquière, A.; et al. The roots of future rice harvests. *Rice* **2014**, *7*, 1–9. [CrossRef] [PubMed]
2. Meyer, R.S.; Purugganan, M.D. Evolution of crop species: Genetics of domestication and diversification. *Nat. Rev. Genet.* **2013**, *14*, 840–852. [CrossRef] [PubMed]
3. Wang, X.Q.; Wang, W.S.; Tai, S.S.; Li, M.; Gao, Q.; Hu, Z.Q.; Hu, W.S.; Wu, Z.C.; Zhu, X.Y.; Xie, J.Y.; et al. Selective and comparative genome architecture of Asian cultivated rice (*Oryza sativa* L.) attributed to domestication and modern breeding. *J. Adv. Res.* **2022**, *42*, 1–16. [CrossRef] [PubMed]
4. Wei, S.B.; Li, X.; Lu, Z.F.; Zhang, H.; Ye, X.Y.; Zhou, Y.J.; Li, J.; Yan, Y.Y.; Pei, H.G.; Duan, F.Y.; et al. A transcriptional regulator that boosts grain yields and shortens the growth duration of rice. *Science* **2022**, *377*, eabi8455. [CrossRef]
5. Xue, W.Y.; Xing, Y.Z.; Weng, X.Y.; Zhao, Y.; Tang, W.J.; Wang, L.; Zhou, H.J.; Yu, S.B.; Xu, C.G.; Li, X.H.; 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]
6. 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*. *Gene Dev.* **2004**, *18*, 926–936. [CrossRef]
7. 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] [PubMed]
8. Itoh, J.I.; Hasegawa, A.; Kitano, H.; Nagato, Y. A recessive heterochronic mutation, *plastochron1*, shortens the plastochron and elongates the vegetative phase in rice. *Plant Cell* **1998**, *10*, 1511–1521. [CrossRef]

9. Makino, Y.; Hirooka, Y.; Homma, K.; Kondoa, R.; Liu, T.S.; Tang, L.; Nakazakia, T.; Xu, Z.J.; Shiraiwa, T. Effect of flag leaf length of erect panicle rice on the canopy structure and biomass production after heading. *Plant Prod. Sci.* **2022**, *25*, 1–10. [CrossRef]

10. Saitoh, K.; Yonetani, K.; Murota, T.; Kuroda, T. Effects of flag leaves and panicles on light interception and canopy photosynthesis in high-yielding rice cultivars. *Plant Prod. Sci.* **2002**, *5*, 275–280. [CrossRef]

11. Rong, C.Y.; Liu, Y.X.; Chang, Z.Y.; Liu, Z.Y.; Ding, Y.F.; Ding, C.Q. Cytokinin oxidase/dehydrogenase family genes exhibit functional divergence and overlap in rice growth and development, especially in control of tillering. *J. Exp. Bot.* **2022**, *73*, 3552–3568. [CrossRef] [PubMed]

12. Chen, R.Z.; Deng, Y.W.; Ding, Y.L.; Guo, J.X.; Qiu, J.; Wang, B.; Wang, C.S.; Xie, Y.Y.; Zhang, Z.H.; Chen, J.X.; et al. Rice functional genomics: Decades' efforts and roads ahead. *Sci. China Life Sci.* **2022**, *65*, 33–92. [CrossRef] [PubMed]

13. Sasaki, T.; Ashikari, M. *Rice Genomics, Genetics and Breeding*; Springer: Berlin/Heidelberg, Germany, 2018.

14. Wang, W.S.; Mauleon, R.; Hu, Z.Q.; Chebotarov, D.; Tai, S.S.; Wu, Z.C.; Li, M.; Zheng, T.Q.; Fuentes, R.R.; Zhang, F.; et al. Genomic variation in 3010 diverse accessions of Asian cultivated rice. *Nature* **2018**, *557*, 43–49. [CrossRef] [PubMed]

15. Hour, A.L.; Hsieh, W.H.; Chang, S.H.; Wu, Y.P.; Chin, H.S.; Lin, Y.R. Genetic diversity of landraces and improved varieties of rice (*Oryza sativa* L.) in Taiwan. *Rice* **2020**, *13*, 82. [CrossRef] [PubMed]

16. Ren, M.M.; Huang, M.H.; Qiu, H.Y.; Chun, Y.; Li, L.; Kumar, A.; Fang, J.J.; Zhao, J.F.; He, H.; Li, X.Y. Genome-wide association study of the genetic basis of effective tiller number in rice. *Rice* **2021**, *14*, 56. [PubMed]

17. Wang, Y.; Zhai, L.Y.; Chen, K.; Shen, C.C.; Liang, Y.T.; Wang, C.C.; Zhao, X.Q.; Wang, S.; Xu, J.L. Natural sequence variations and combinations of *GNP1* and *NAL1* determine the grain number per panicle in rice. *Rice* **2020**, *13*, 14. [CrossRef] [PubMed]

18. Lv, Q.M.; Li, W.G.; Sun, Z.Z.; Ouyang, N.; Jing, X.; He, Q.; Wu, J.; Zheng, J.K.; Zheng, J.T.; Tang, S.Q.; et al. Resequencing of 1143 *indica* rice accessions reveals important genetic variations and different heterosis patterns. *Nat. Commun.* **2020**, *11*, 4778. [CrossRef] [PubMed]

19. Chiang, C.; Scott, A.J.; Davis, J.R.; Tsang, E.K.; Li, X.; Kim, Y.; Hadzic, T.; Damani, F.N.; Ganel, L.; Consortium, G.; et al. The impact of structural variation on human gene expression. *Nat. Genet.* **2017**, *49*, 692–699. [CrossRef]

20. Yuan, Y.X.; Bayer, P.E.; Batley, J.; Edwards, D. Current status of structural variation studies in plants. *Plant Biotechnol. J.* **2021**, *19*, 2153–2163. [CrossRef]

21. Fuentes, R.R.; Chebotarov, D.; Duitama, J.; Smith, S.; De la Hoz, J.F.; Mohiyuddin, M.; Wing, R.A.; McNally, K.L.; Tatarinova, T.; Grigoriev, A.; et al. Structural variants in 3000 rice genomes. *Genome Res.* **2019**, *29*, 870–880. [CrossRef]

22. Kou, Y.X.; Liao, Y.; Toivainen, T.; Lv, Y.D.; Tian, X.M.; Emerson, J.J.; Gaut, B.S.; Zhou, Y.F. Evolutionary genomics of structural variation in Asian rice (*Oryza sativa*) domestication. *Mol. Biol. Evol.* **2020**, *37*, 3507–3524. [CrossRef] [PubMed]

23. Qin, P.; Lu, H.W.; Du, H.L.; Wang, H.; Chen, W.L.; Chen, Z.; He, Q.; Ou, S.J.; Zhang, H.Y.; Li, X.Z.; et al. Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* **2021**, *184*, 3542–3558. [CrossRef] [PubMed]

24. Shang, L.G.; Li, X.X.; He, H.Y.; Yuan, Q.L.; Song, Y.N.; Wei, Z.R.; Lin, H.; Hu, M.; Zhao, F.L.; Zhang, C.; et al. A super pan-genomic landscape of rice. *Cell Res.* **2022**, *32*, 878–896. [CrossRef] [PubMed]

25. Wang, T.Y.; He, W.C.; Li, X.X.; Zhang, C.; He, H.Y.; Yuan, Q.L.; Zhang, B.; Zhang, H.; Leng, Y.; Wei, H.; et al. A rice variation map derived from 10,548 rice accessions reveals the importance of rare variants. *Nucleic Acids Res.* **2023**, *51*, 10924–10933. [CrossRef] [PubMed]

26. Wang, Y.; Li, F.C.; Zhang, F.; Wu, L.; Xu, N.; Sun, Q.; Chen, H.; Yu, Z.W.; Lu, J.H.; Jiang, K.; et al. Time-ordering *japonica/geng* genomes analysis indicates the importance of large structural variants in rice breeding. *Plant Biotechnol. J.* **2023**, *21*, 202–218. [CrossRef]

27. Zheng, X.M.; Zhong, L.M.; Pang, H.B.; Wen, S.Y.; Li, F.; Lou, D.J.; Ge, J.Y.; Fan, W.Y.; Wang, T.Y.; Han, Z.Y.; et al. Lost genome segments associate with trait diversity during rice domestication. *BMC Biol.* **2023**, *21*, 20. [CrossRef] [PubMed]

28. Li, D.Y.; Huang, Z.Y.; Song, S.H.; Xin, Y.Y.; Mao, D.H.; Lv, Q.M.; Zhou, M.; Tian, D.M.; Tang, M.F.; Wu, Q.; et al. Integrated analysis of phenome, genome, and transcriptome of hybrid rice uncovered multiple heterosis-related loci for yield increase. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, E6026–E6035. [PubMed]

29. Zhang, S.Y.; Zhu, L.M.; Shen, C.B.; Ji, Z.; Zhang, H.P.; Zhang, T.; Li, Y.; Yu, J.P.; Yang, N.; He, Y.B.; et al. Natural allelic variation in a modulator of auxin homeostasis improves grain yield and nitrogen use efficiency in rice. *Plant Cell* **2021**, *33*, 566–580. [CrossRef]

30. Wang, C.C.; Yu, H.; Huang, J.; Wang, W.S.; Faruquee, M.; Zhang, F.; Zhao, X.Q.; Fu, B.Y.; Chen, K.; Zhang, H.L.; et al. Towards a deeper haplotype mining of complex traits in rice with RFGB v2.0. *Plant Biotechnol. J.* **2020**, *18*, 14–16. [CrossRef]

31. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]

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

33. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genomic Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]

34. Rausch, T.; Zichner, T.; Schlattl, A.; Stütz, A.M.; Benes, V.; Jan, O.; Korbel, J.O. DELLY: Structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **2012**, *28*, i333–i339. [CrossRef] [PubMed]

35. Chen, X.Y.; Schulz-Trieglaff, O.; Shaw, R.; Barnes, B.; Schlesinger, F.; Källberg, M.; Cox, A.J.; Kruglyak, S.; Saunders, C.T. Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **2016**, *32*, 1220–1222. [CrossRef] [PubMed]

36. Danecek, P.; McCarthy, S.A. BCFtools/csq: Haplotype-aware variant consequences. *Bioinformatics* **2017**, *33*, 2037–2039. [CrossRef] [PubMed]

37. Jeffares, D.C.; Jolly, C.; Hoti, M.; Speed, D.; Shaw, L.; Rallis, C.; Balloux, F.; Dessimoz, C.; Bähler, J.; Sedlazeck, F.J. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat. Commun.* **2017**, *8*, 14061. [CrossRef] [PubMed]

38. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef] [PubMed]

39. Zhao, G.W.; Lian, Q.; Zhang, Z.H.; Fu, Q.S.; He, Y.H.; Ma, S.W.; Ruggieri, V.; Monforte, A.J.; Zhao, G.W.; Lian, Q.; et al. A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. *Nat. Genet.* **2019**, *51*, 1607–1615. [CrossRef] [PubMed]

40. Lipka, A.E.; Tian, F.; Wang, Q.S.; Peiffer, J.; Li, M.; Bradbury, P.J.; Gore, M.A.; Buckler, E.S.; Zhang, Z.W. GAPIT: Genome association and prediction integrated tool. *Bioinformatics* **2012**, *28*, 2397–2399. [CrossRef]

41. Huang, M.; Liu, X.L.; Zhou, Y.; Summers, R.M.; Zhang, Z.W. BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* **2019**, *8*, giy154. [CrossRef]

42. Zhao, Q.; Feng, Q.; Lu, H.Y.; Li, Y.; Wang, A.H.; Tian, Q.L.; Zhan, Q.L.; Lu, Y.Q.; Zhang, L.; Huang, T.; et al. Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. *Nat. Genet.* **2018**, *50*, 278–284. [CrossRef] [PubMed]

43. The R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.

44. Chen, Y.; Moore, K.L.; Miller, A.J.; McGrath, S.P.; Ma, J.F.; Zhao, F.J. The role of nodes in arsenic storage and distribution in rice. *J. Exp. Bot.* **2015**, *66*, 3717–3724. [CrossRef] [PubMed]

45. Geng, Y.K.; Zhang, P.X.; Liu, Q.; Wei, Z.W.; Riaz, A.; Wei, Z.; Riaz, A.; Chachar, S.; Gu, X.F. Rice homolog of Sin3-associated polypeptide 30, OsSFL1, mediates histone deacetylation to regulate flowering time during short days. *Plant Biotechnol. J.* **2020**, *18*, 325–327. [CrossRef] [PubMed]

46. Sakurai, J.; Ishikawa, F.; Yamaguchi, T.; Uemura, M.; Maeshima, M. Identification of 33 rice aquaporin genes and analysis of their expression and function. *Plant Cell Physiol.* **2005**, *46*, 1568–1577. [CrossRef] [PubMed]

47. Gui, J.S.; Shen, J.H.; Li, L.G. Functional characterization of evolutionarily divergent 4-coumarate: Coenzyme A ligases in rice. *Plant Physiol.* **2011**, *157*, 574–586. [CrossRef] [PubMed]

48. Qi, J.; Qian, Q.; Bu, Q.Y.; Li, S.Y.; Chen, Q.; Sun, J.Q.; Liang, W.X.; Zhou, Y.H.; Chu, C.C.; Li, X.G.; et al. Mutation of the rice *Narrow leaf1* gene, which encodes a novel protein, affects vein patterning and polar auxin transport. *Plant Physiol.* **2008**, *147*, 1947–1959. [CrossRef]

49. Cho, S.H.; Yoo, S.C.; Zhang, H.T.; Lim, J.H.; Paek, N.C. Rice *NARROW LEAF1* regulates leaf and adventitious root development. *Plant Mol. Biol. Report.* **2014**, *32*, 270–281. [CrossRef]

50. Xie, L.; Liu, M.H.; Zhao, L.; Cao, K.; Wang, P.; Xu, W.H.; Sung, W.K.; Li, X.W.; Li, G.L. RiceENCODE: A comprehensive epigenomic database as a rice Encyclopedia of DNA Elements. *Mol. Plant.* **2021**, *14*, 1604–1606. [CrossRef]

51. Springer, N.M.; Ying, K.; Fu, Y.; Ji, T.M.; Yeh, C.T.; Jia, Y.; Wu, W.; Richmond, T.; Kitzman, J.; Rosenbaum, H.; et al. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. *PLoS Genet.* **2009**, *5*, e1000734. [CrossRef]

52. Zhang, L.M.; Luo, H.; Liu, Z.Q.; Zhao, Y.; Luo, J.C.; Hao, D.Y.; Jing, H.C. Genome-wide patterns of large-size presence/absence variants in sorghum. *J. Integr. Plant Biol.* **2014**, *56*, 24–37. [CrossRef]

53. Zhang, Y.L.; Fu, J.; Wang, K.; Han, X.; Yan, T.Z.; Su, Y.M.; Li, Y.F.; Lin, Z.C.; Qin, P.; Fu, C.J.; et al. The telomere-to-telomere gap-free genome of four rice parents reveals SV and PAV patterns in hybrid rice breeding. *Plant Biotechnol. J.* **2022**, *20*, 1642–1644. [CrossRef] [PubMed]

54. Zhao, Y.; Zhang, H.L.; Xu, J.L.; Jiang, C.H.; Yin, Z.G.; Xiong, H.Y.; Xie, J.Y.; Wang, X.Q.; Zhu, X.Y.; Li, Y.; et al. Loci and natural alleles underlying robust roots and adaptive domestication of upland ecotype rice in aerobic conditions. *PLoS Genet.* **2018**, *14*, e1007521. [CrossRef] [PubMed]

55. Saco, A.; Rey-Campos, M.; Gallardo-Escárate, C.; Gerdol, M.; Novoa, B.; Figueras, A. Gene presence/absence variation in *Mytilus galloprovincialis* and its implications in gene expression and adaptation. *iScience* **2023**, *26*, 107827. [CrossRef] [PubMed]

56. Zheng, L.Y.; Guo, X.S.; He, B.; Sun, L.J.; Peng, Y.; Dong, S.S.; Liu, T.F.; Jiang, S.Y.; Ramachandran, S.; Liu, C.M.; et al. Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*). *Genome Biol.* **2011**, *12*, R114. [CrossRef] [PubMed]

57. Göktay, M.; Fulgione, A.; Hancock, A.M. A new catalog of structural variants in 1301 *A. thaliana* lines from Africa, Eurasia, and North America reveals a signature of balancing selection at defense response genes. *Mol. Biol. Evol.* **2021**, *38*, 1498–1511. [CrossRef] [PubMed]

58. Fitzgerald, T.L.; Kazan, K.; Li, Z.Y.; Morell, M.K.; Manners, J.M. A high-throughput method for the detection of homoeologous gene deletions in hexaploid wheat. *BMC Plant Biol.* **2010**, *10*, 264. [CrossRef]

59. Delteil, A.; Gobbato, E.; Cayrol, B.; Estevan, J.; Michel-Romiti, C.; Dievart, A.; Kroj, T.; Morel, J.B. Several wall-associated kinases participate positively and negatively in basal defense against rice blast fungus. *BMC Plant Biol.* **2016**, *16*, 17. [CrossRef]

60. Anandan, A.; Parameswaran, C.; Mahender, A.; Nayak, A.K.; Vellaikumar, S.; Balasubramaniasai, C.; Ali, J. Trait variations and expression profiling of *OsPHT1* gene family at the early growth-stages under phosphorus-limited conditions. *Sci. Rep.* **2021**, *11*, 13563. [CrossRef]

61. Chang, M.C.; Wang, B.Y.; Chen, X.F.; Wu, R. Molecular characterization of catalytic-subunit cDNA sequences encoding protein phosphatases 1 and 2A and study of their roles in the gibberellin-dependent Osamy-c expression in rice. *Plant Mol. Biol.* **1999**, *39*, 105–115. [CrossRef]

62. Huang, X.Z.; Qian, Q.; Liu, Z.B.; Sun, H.Y.; He, S.Y.; Luo, D.; Xia, G.M.; Chu, C.C.; Li, J.Y.; Fu, X.D. Natural variation at the *DEP1* locus enhances grain yield in rice. *Nat. Genet.* **2009**, *41*, 494–497. [CrossRef]

63. Liao, Y.D.; Lin, K.H.; Chen, C.C.; Chiang, C.M. *Oryza sativa* protein phosphatase 1a (*OsPP1a*) involved in salt stress tolerance in transgenic rice. *Mol. Breed.* **2016**, *36*, 22. [CrossRef]

64. Song, S.; Wang, G.F.; Wu, H.; Fan, X.W.; Liang, L.W.; Zhao, H.; Li, S.L.; Hu, Y.; Liu, H.Y.; Ayaad, M.; et al. OsMFT2 is involved in the regulation of ABA signaling-mediated seed germination through interacting with OsbZIP23/66/72 in rice. *Plant J.* **2020**, *103*, 532–546. [CrossRef] [PubMed]

65. Sugiyama, N.; Izawa, T.; Oikawa, T.; Shimamoto, K. Light regulation of circadian clock-controlled gene expression in rice. *Plant J.* **2001**, *26*, 607–615. [CrossRef] [PubMed]

66. Tamiru, M.; Takagi, H.; Abe, A.; Yokota, T.; Kanzaki, H.; Okamoto, H.; Saitoh, H.; Takahashi, H.; Fujisaki, K.; Oikawa, K.; et al. A chloroplast-localized protein LESION AND LAMINA BENDING affects defence and growth responses in rice. *New Phytol.* **2016**, *210*, 1282–1297. [CrossRef] [PubMed]

67. Sun, H.Y.; Qian, Q.; Wu, K.; Luo, J.J.; Wang, S.S.; Zhang, C.W.; Ma, Y.F.; Liu, Q.; Huang, X.Z.; Yuan, Q.B.; et al. Heterotrimeric G proteins regulate nitrogen-use efficiency in rice. *Nat. Genet.* **2014**, *46*, 652–656. [CrossRef] [PubMed]

68. Li, J.; Lee, C.R. The role of gene presence-absence variations on genetic incompatibility in Asian rice. *New Phytol.* **2023**, *239*, 778–791. [CrossRef] [PubMed]

69. Dolatabadian, A.; Bayer, P.E.; Tirnaz, S.; Hurgobin, B.; Edwards, D.; Batley, J. Characterization of disease resistance genes in the *Brassica napus* pangenome reveals significant structural variation. *Plant Biotechnol. J.* **2020**, *18*, 969–982. [CrossRef]

70. Zhang, F.; Xue, H.Z.; Dong, X.R.; Li, M.; Zheng, X.M.; Li, Z.K.; Xu, J.L.; Wang, W.S.; Wei, C.C. Long-read sequencing of 111 rice genomes reveals significantly larger pan-genomes. *Genome Res.* **2022**, *32*, 853–863. [CrossRef]