

Article

Genome-Wide Association Study Reveals the Genetic Basis of Total Flavonoid Content in Brown Rice

Haijian Xia ^{1,†} , Xiaoying Pu ^{2,†}, Xiaoyang Zhu ¹, Xiaomeng Yang ², Haifeng Guo ¹, Henan Diao ³, Quan Zhang ¹, Yulong Wang ¹, Xingming Sun ¹, Hongliang Zhang ¹, Zhanying Zhang ¹ , Yawen Zeng ^{2,*}  and Zichao Li ^{1,4,*} 

¹ Beijing Key Laboratory of Crop Genetic Improvement, College of Agronomy and Biotechnology, China Agricultural University, Beijing 100193, China; hxiajnyx@126.com (H.X.)

² Biotechnology and Germplasm Resources Institute, Yunnan Academy of Agricultural Sciences/Agricultural Biotechnology Key Laboratory of Yunnan Province, Kunming 650205, China; puxiaoying@163.com (X.P.)

³ Heihe Branch of Heilongjiang Academy of Agricultural Sciences, Heihe 164300, China

⁴ Sanya Institute, China Agricultural University, Sanya 572025, China

* Correspondence: zengyw1967@126.com (Y.Z.); lizichao@cau.edu.cn (Z.L.)

† These authors contributed equally to this work.

Abstract: Flavonoids have anti-inflammatory, antioxidative, and anticarcinogenic effects. Breeding rice varieties rich in flavonoids can prevent chronic diseases such as cancer and cardio-cerebrovascular diseases. However, most of the genes reported are known to regulate flavonoid content in leaves or seedlings. To further elucidate the genetic basis of flavonoid content in rice grains and identify germplasm rich in flavonoids in grains, a set of rice core collections containing 633 accessions from 32 countries was used to determine total flavonoid content (TFC) in brown rice. We identified ten excellent germplasms with TFC exceeding 300 mg/100 g. Using a compressed mixed linear model, a total of 53 quantitative trait loci (QTLs) were detected through a genome-wide association study (GWAS). By combining linkage disequilibrium (LD) analysis, location of significant single nucleotide polymorphisms (SNPs), gene expression, and haplotype analysis, eight candidate genes were identified from two important QTLs (*qTFC1-6* and *qTFC9-7*), among which *LOC_Os01g59440* and *LOC_Os09g24260* are the most likely candidate genes. We also analyzed the geographic distribution and breeding utilization of favorable haplotypes of the two genes. Our findings provide insights into the genetic basis of TFC in brown rice and could facilitate the breeding of flavonoid-rich varieties, which may be a prevention and adjuvant treatment for cancer and cardio-cerebrovascular diseases.

Keywords: flavonoid; functional rice; gene; germplasm resource; QTL



Citation: Xia, H.; Pu, X.; Zhu, X.; Yang, X.; Guo, H.; Diao, H.; Zhang, Q.; Wang, Y.; Sun, X.; Zhang, H.; et al. Genome-Wide Association Study Reveals the Genetic Basis of Total Flavonoid Content in Brown Rice. *Genes* **2023**, *14*, 1684. <https://doi.org/10.3390/genes14091684>

Academic Editor: Depeng Wang

Received: 22 July 2023

Revised: 21 August 2023

Accepted: 22 August 2023

Published: 25 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

An increasing number of people are in a sub-health state and suffer from various chronic diseases, such as cancer and cardio-cerebrovascular disease [1–3]. Ten million people in the world will have died of cancer in 2020 [4]. Hypertension is the main risk factor for cardio-cerebrovascular disease [5]. Globally, more than a quarter of the population is hypertensive [6]. Therefore, reducing the harm that these diseases pose to human health is essential.

Flavonoids are important secondary metabolites widely found in plants [7] and can be classified into six major subgroups: chalcones, flavones, flavonols, flavandiol, anthocyanins, and proanthocyanidins [8]. Flavonoids have anti-inflammatory, antioxidative, and anticarcinogenic properties useful for the prevention and adjuvant treatment of cancer and cardio-cerebrovascular disease [9]. Rice is the main staple and energy source for more than half of the world's population [10]. Breeding functional rice varieties rich in flavonoids could increase daily flavonoid intake through diet without the need for drugs, which may be an easy and convenient way to prevent cancer and cardio-cerebrovascular diseases. To achieve this goal, it is essential to identify rice

germplasm with high flavonoid content and understand the genetic basis of flavonoid content in rice grains.

Previous studies identifying rice germplasm rich in flavonoids are relatively rare. Li et al. (2022) determined the total flavonoid content (TFC) of 164 accessions from the United States Department of Agriculture's (USDA) rice mini-core collection. The TFC of 905 accessions from the primary core collection in Yunnan province, China, was also identified [11]. In addition, the TFC of less than 20 Indian and South Korean rice varieties was identified [12–15]. Therefore, to identify germplasm rich in flavonoids, it is necessary to select more varieties with a wider geographical distribution and greater genetic variation.

Several studies have been conducted to elucidate the genetic basis of the flavonoid content of rice. The genes affecting flavonoid content in rice can be divided into four categories: structural, regulatory, transporter, and modifying enzyme genes [16]. Structural genes encode proteins that catalyze the conversion of phenylalanine into chalcones, flavones, flavonols, flavandiols, anthocyanins, and proanthocyanidins, such as *OsPAL06* [17], *OsF3H* [18], *OsFLS* [19], and *OsANS* [20]. Regulatory genes encode proteins that regulate the expression of structural genes, thereby affecting flavonoid content, such as *Rc* [21], *OsB1* [22], and *OsC1-MYB* [20]. The modifying enzymes catalyze the glycosylation or methylation of flavonoids to increase their stability, such as *OsUGT706C2* [23] and *OsNOMT* [24]. The transporters can transfer anthocyanins from the cytoplasm to vacuoles, such as *OsMATE34* [25].

Although several genes that affect flavonoid content in rice have been cloned, most of the reported genes are known to influence flavonoid content in leaves or seedlings [22,26–29]. To our knowledge, only 11 genes affecting flavonoid content in grains have been reported, including three structural genes (*DFR* [20], *OsF3'H* [18], and *CYP75B4* [30]) and eight regulatory genes. Dihydroflavonol 4-reductase (*DFR*) catalyzes the conversion of dihydroflavonol to flavonol in rice pericarps. *OsF3'H* catalyzes the formation of eriodictyol in the seed coat. *CYP75B4* hydroxylates leucoanthocyanins to produce anthocyanins and proanthocyanidins. Among the eight regulatory genes, five encode transcription factors (*OsMYB3* [31], *OsB2* [22], *Rc* [21], *OsTTG1* [32], and *OsVP1* [33]), and three encode non-transcription factor proteins (*OsDET1*, *OsDDB1*, and *OsCOP10* [34]). *Rc*, *OsMYB3*, *OsB2*, *OsTTG1*, and *OsVP1* regulate the expression of some structural genes to affect flavonoid content in rice pericarps. *OsDET1*, *OsDDB1*, and *OsCOP10* combine to form a complex device driver (CDD), which affects the flavonoid biosynthesis pathway in rice seeds.

To further elucidate the genetic basis of TFC in rice grains and identify flavonoid-rich varieties, we used a rice core collection population of 633 varieties to determine TFC in brown rice. We identified ten excellent germplasms with high TFC. A total of 53 (quantitative trait loci) QTLs were detected through a genome-wide association study (GWAS). Through linkage disequilibrium (LD) analysis, gene expression, and other methods, we identified eight candidate genes from two preferred QTLs. Our study sheds further light on the genetic basis of TFC in brown rice and could facilitate the breeding of flavonoid-rich varieties.

2. Materials and Methods

2.1. Plant Materials

This study used 633 cultivated rice accessions, comprising 237 accessions from the rice mini-core collection [35] and 396 accessions from the International Rice Molecular Breeding Network [36]. These accessions were collected from 32 countries across five continents (583 from Asia, 18 from Africa, 15 from America, 9 from Europe, and 8 from Oceania) (Figure 1a), representing a wide range of genetic diversity. Detailed information on the accessions is provided in Table S1.

2.2. Field Trials and Measurement of Total Flavonoid Content

Seeds from the 633 rice accessions were planted in Yuxi, Yunnan, China, in 2015, with two replications. A randomized complete block design was used. After harvesting and drying, 20 g of dehulled brown rice from each accession were ground to determine the TFC.

The TFC was determined using the $\text{NaNO}_2\text{-Al}(\text{NO}_3)_3$ method [37]. In a 16 mL centrifuge tube, a powdered rice sample (0.5 g) was collected and mixed with 5 mL of 50% ethanol. The mixture was oscillated in a vibrator (manufacturer: Jiangsu Jieruier Electric Appliance Co., Ltd., Changzhou, China; model: SHA-B) at room temperature and 200 rpm for a duration of 3 h, then centrifuged at 7000 rpm for 3 min. After centrifugation, 1 mL of supernatant was placed in a new tube, mixed with 0.6 mL of 5% NaNO_2 , and allowed to stand for 5 min. The same procedure was performed with 0.6 mL of 10% $\text{Al}(\text{NO}_3)_3$ and 4 mL of 5% NaOH , and the tubes were allowed to stand for 6 min and 10 min, respectively. After the color became red and stabilized, the absorption was measured at a wavelength of 500 nm. A standard curve was developed using different concentrations (0.05, 0.1, 0.15, 0.2, and 0.25 mg/mL) of rutin standard obtained from Guizhou Dida Science and Technology Limited Company (Guiyang, China). The TFC was estimated using the equations $y = 1.8764x - 0.0054$ ($R^2 = 0.9997$) and $y = 1.8503x - 0.0129$ ($R^2 = 0.9995$) for the two replications, respectively. The regression equations and coefficients of determination indicated a high level of precision for the measurement of TFC. To reduce error, all analyses were repeated three times. The TFC was expressed as milligrams of rutin equivalent per 100 g of dry rice flour.

2.3. Genotype

The raw SNP genotype data for the 633 accessions were obtained from the 3000 Rice Genome Project (3KRG) [38,39] with an average sequencing depth of $15\times$. SNPs with more than two alleles, a missing rate over 30%, and a minor allele frequency (MAF) less than 5% were removed, resulting in 2762746 SNPs that were used in population structure analysis and GWAS.

2.4. Population Structure Analysis

Using PLINK version 1.9 (window 50 bp, step size 5 bp, $r^2 < 0.3$) [40], 276464 SNPs in linkage equilibrium were screened and used to perform principal component analysis using Genome-wide complex trait analysis (GCTA) version 1.93.2 [41]. Kinship analysis were carried out in the R programming language version 3.5.2 using the GAPIT package [42].

2.5. GWAS

GWAS was conducted for the full, *indica*, and *japonica* populations. A total of 2762746 high-quality SNPs (MAF $\geq 5\%$, missing rate $< 30\%$) were used to perform GWAS using a compressed mixed linear model (CMLM) in the GAPIT package operated in an R environment [42]. The first three principal components and kinship were integrated simultaneously to control false positives in CMLM. The genome-wide significance threshold was determined using permutation tests with 1000 replications [43]. The significance threshold for the association analysis was set at the top 5% probability. A genome interval is defined as a QTL if it contains at least three consecutive significant SNPs and the distance between any two adjacent significant SNPs is less than 170 kb, given LD decay values of approximately 120 kb and 170 kb in *indica* and *japonica* populations, respectively [44,45]. The SNP with the minimum p -value within a QTL was considered the lead SNP of the QTL. If two QTLs detected in two populations had overlapping intervals, they were considered co-located QTLs.

2.6. Screening Candidate Genes

Screening candidate genes is a process of progressively reducing the candidate gene pool through various methods. For a selected target QTL, the LD heatmap was created using all significant SNPs in an interval of about one Mb on the flank of the lead SNP [46,47]. The region with consecutive SNPs closely linked to the lead SNP ($r^2 \geq 0.6$) was regarded as the local LD interval [48]. According to the gene annotation of the MSU Rice Genome Annotation Project Database and Resource (<http://rice.plantbiology.msu.edu/>, (accessed on 10 January 2023) [49]), all predicted genes in the LD interval were identified. Among all the predicted genes, transposons and retrotransposons were first excluded, and then genes containing significant SNPs in the entire gene region, including the promoter, exon, intron, and/or 3'UTR, were retained. For genes containing significant SNPs, their expression in grain-related tissues (ovary, embryo, and endosperm) at different time points after pollination can be examined through two rice gene expression databases, RiceXPro (<https://ricexpro.dna.affrc.go.jp/> (accessed on 15 January 2023), [50]) and MSU (<http://rice.uga.edu/expression.shtml> (accessed on 16 January 2023), [49]). Genes with no expression in grain-related tissues are unlikely to be candidate genes, while genes with some expression in these tissues may be considered candidate genes. For candidate genes that meet the expression characteristics, haplotype analysis can be performed using significant SNPs. If the TFC of different haplotypes of a gene shows significant differences in the *japonica* or *indica* subgroups, the gene may be considered a candidate gene. If there is no significant difference in the TFC of different haplotypes of a gene in both the *japonica* and *indica* subpopulations, then this gene can be excluded.

2.7. Statistical Analysis and Graphing

SPSS 19 software and a two-tailed *t*-test were used to analyze the significance difference. Venn diagrams, Manhattan plots, Q-Q plots, QTL summary plots, and geographic distribution plots were drawn using R 4.2.1. Principal component analysis plots, histograms, violin diagrams, and boxplots were drawn using Origin 2022.

3. Results

3.1. Identification of Functional Rice Germplasm Rich in Flavonoids

This study utilized 633 cultivated rice accessions from 32 countries across five continents to identify rice varieties with high flavonoid content and to conduct GWAS (Figure 1a and Table S1). The TFC in brown rice from all accessions was measured using a spectrophotometer. The full population exhibited large variations in TFC (Figure 1b), ranging from 90.6 (Tieganwu, *japonica*) to 435.9 (Yanshuichi, *indica*) mg/100 g, with a mean value of 136.0 mg/100 g and a coefficient of variation (CV) of 28.2%. The TFC of Yanshuichi is 4.8 times that of Tieganwu. The distribution of TFC was positively skewed (Figure 1b), with most accessions having low TFC (ranging from 100 to 150 mg/100 g) and only a few varieties having high TFC more than 350 mg/100 g. The 10 accessions with the highest TFC were identified and listed in Table 1: Yanshuichi (435.9 mg/100 g), Feidongtangdao (426.0 mg/100 g), Lengshuigu 2 (407.3 mg/100 g), Qiuqianbai (355.8 mg/100 g), IRAT 669 (347.8 mg/100 g), Haobayong 1 (344.5 mg/100 g), Lamujia (331.5 mg/100 g), Longhuamaohulu (330.6 mg/100 g), Tianhandao (316.4 mg/100 g), and Banjiemang (313.7 mg/100 g). These elite varieties, which comprise three *indica* and seven *japonica* varieties, mostly from southern China, and 90% of which are landraces, can serve as valuable germplasm resources for breeding flavonoid-rich functional rice.

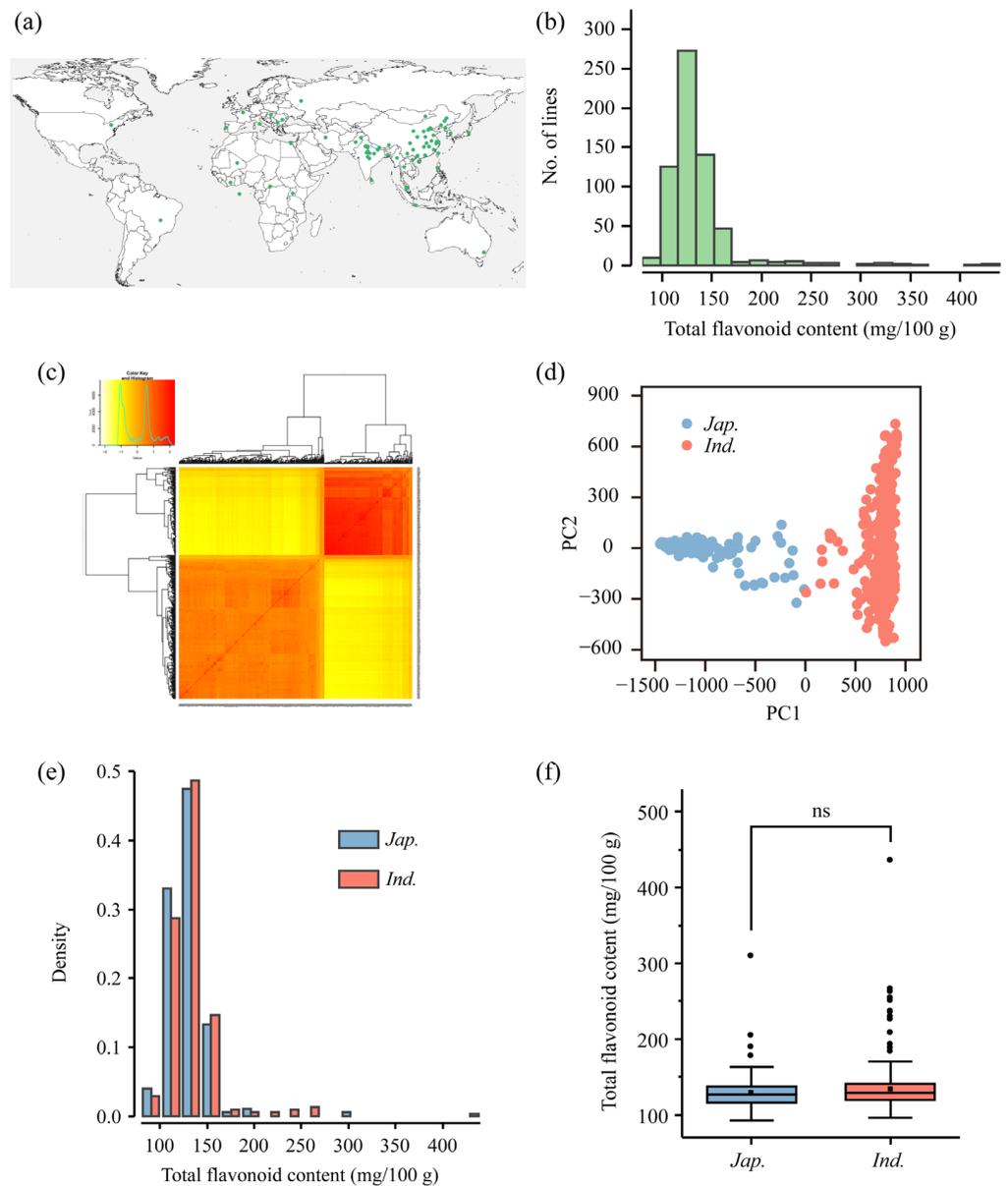


Figure 1. Phenotypic diversity of total flavonoid content (TFC) and population structure of 633 cultivated rice. **(a)** Geographic distribution of all accessions. Each green dot represents a accession. **(b)** A histogram showing the distribution of TFC in the full population. **(c)** Kinship plot of all accessions. **(d)** Principal component analysis plot. **(e)** Histogram of TFC in the *indica* and *japonica* subpopulations. The Y-axis shows the proportion of accessions falling within a certain TFC range relative to the total number of accessions within each subpopulation. **(f)** Boxplots showing the phenotypic variation of TFC in the *indica* and *japonica* subpopulations.

Based on the known variety information, the 633 accessions can be categorized into two subgroups: 383 *indica* and 250 *japonica* varieties. Kinship analysis and principal component (PC) analysis using the genotypes confirmed the division of all varieties into these two subgroups (Figure 1c,d). Both the *japonica* and *indica* subpopulations exhibited a positively skewed distribution of TFC (Figure 1e). Moreover, there was no significant difference in TFC between the *japonica* and *indica* subpopulations (Figure 1f).

3.2. GWAS for Total Flavonoid Content

To elucidate the genetic basis of the flavonoid content in rice grains, GWAS was performed on 633 accessions. To eliminate false positives from the population structure, we performed GWAS using a compressed mixed linear model (CMLM) with PC and kinship, which accounts for population structure and identifies the optimal group kinship matrix [51]. The first three PCs were used to construct the PC matrix. The CMLM model was applied to GWAS of the full, *indica*, and *japonica* populations (Figure 2a–c). The quantile-quantile (Q-Q) plots indicated that CMLM eliminated false positives and false negatives well in the three populations (Figure 2d–f). The significant thresholds of GWAS in the three populations (full, $-\log_{10}(P) = 4.97$; *indica*, $-\log_{10}(P) = 4.42$; *japonica*, $-\log_{10}(P) = 4.65$) were determined using permutation tests (Figure 2g).

Ref. [51] the full population exhibited a higher number of QTLs compared to the *indica* and *japonica* subpopulations (full, 27; *indica*, 23; *japonica*, 17) (Figure 3a,b; Table S2). We identified five co-located QTLs on chromosomes 4, 6, 7, and 8 in both the full and *indica* populations, as well as seven co-located QTLs on chromosomes 1, 5, 9, 10, and 11 in both the full and *japonica* populations (Figure 3a,b; Table S2). Additionally, a co-located QTL on chromosome three was identified in both the *indica* and *japonica* subpopulations (Figure 3a,b; Table S2). However, no QTL was detected in all three populations simultaneously. Co-located QTLs are likely to contain genes that control flavonoid content in rice grains and can serve as preferred QTLs for screening candidate genes. Table 2 provides information on all co-localized QTLs. *qTFC7-2*, *qTFC1-6*, and *qTFC9-7* exhibited strong association signals with TFC in their respective populations, with more than or equal to 30 commonly significant single nucleotide polymorphisms (SNPs) and overlapping intervals greater than 25 kb. Therefore, they are the preferred QTLs for exploring candidate genes.

Table 1. Ten rice varieties with the highest total flavonoid content.

Variety Name	Subspecies	Origin	TFC (mg/100 g)	Landrace or Improved Variety
Yanshuichi	<i>Ind.</i>	Fujian, China	435.9	landrace
Feidongtangdao	<i>Jap.</i>	Anhui, China	426.0	landrace
Lengshuigu 2	<i>Jap.</i>	Yunnan, China	407.3	landrace
Qiuqianbai	<i>Ind.</i>	Anhui, China	355.8	landrace
IRAT 669	<i>Jap.</i>	Ivory Coast	347.8	improved variety
Haobayong 1	<i>Jap.</i>	Yunnan, China	344.5	landrace
Lamujia	<i>Jap.</i>	Yunnan, China	331.5	landrace
Longhuamaohulu	<i>Jap.</i>	Hebei, China	330.6	landrace
Tianhandao	<i>Ind.</i>	Vietnam	316.4	landrace
Banjiemang	<i>Jap.</i>	Yunnan, China	313.7	landrace

To verify the reliability of our results, we investigated whether the reported genes were located within candidate QTLs. Our analysis revealed that *Rc* was detected in both *qTFC-Full-7-1* within the full population and *qTFC-Ind-7-2* within the *indica* subpopulation (Figures 2a,b and 3a; Table S2). *Rc* encodes a transcription factor that contains a basic helix-loop-helix (bHLH) domain, which affects proanthocyanidin synthesis by regulating the expression of some structural genes involved in flavonoid biosynthesis and is a major gene controlling flavonoid synthesis in rice pericarps [21,33,52,53]. The lead SNP of *qTFC-Full-7-1* and *qTFC-Ind-7-2* was the same SNP (Chr7_6069266) (Figure 3c), located in the 3' untranslated region (UTR) of *Rc*. Additionally, both *qTFC-Full-7-1* and *qTFC-Ind-7-2* exhibited the strongest association signals in their respective populations (Figures 2a,b and 3a; Table S2).

The GWAS results for both the full and *indica* populations identified 20 and 19 significant SNPs, respectively, located in *Rc* (Figure 3c and Table S3). All 19 significant SNPs of *Rc* in the *indica* subpopulation were also significant in the full population, including one non-synonymous SNP, two synonymous SNPs, 14 intron SNPs, and two 3'UTR SNPs

(Table S3). Using these significant SNPs, all accessions were divided into two haplotypes of *Rc* (Rc^{Hap1} and Rc^{Hap2}) (Figure 3d). The *indica* subpopulation contained Rc^{Hap1} and Rc^{Hap2} , while all accessions in the *japonica* subpopulation belonged to Rc^{Hap2} . The average TFC in Rc^{Hap1} was significantly higher than that of Rc^{Hap2} in the *indica* subpopulation (Figure 3d), indicating that Rc^{Hap1} was the favorable haplotype in the *indica* subpopulation. These findings support the notion that *Rc* is the functional gene for *qTFC-Full-7-1* and *qTFC-Ind-7-2* and further confirm that *Rc* is the major gene regulating flavonoid content. Overall, these results demonstrated that our GWAS results were reliable. Moreover, most QTLs with strong signals in both the *indica* and *japonica* subpopulations, such as *qTFC-Ind-7-2*, *qTFC-Ind-7-3*, *qTFC-Ind-8-1*, *qTFC-Jap-1-2*, *qTFC-Jap-1-3*, and *qTFC-Jap-9-2*, were also detected in the full population (Figure 3a; Table S2), providing further evidence of the reliability of our GWAS results.

3.3. Identification of Candidate Genes for *qTFC1-6*

qTFC1-6 was a co-located QTL of *qTFC-Full-1-5* in the full population and *qTFC-Jap-1-4* in the *japonica* subpopulation (Table 2). Both *qTFC-Full-1-5* and *qTFC-Jap-1-4* contained clustered significant SNPs and exhibited the highest association signals in their respective populations (Figure 2a,c; Table S2). The lead SNPs for *qTFC-Full-1-5* and *qTFC-Jap-1-4* had $-\log_{10}(P)$ values of 8.58 and 8.81, respectively (Table 2). There were 36 commonly significant SNPs between *qTFC-Full-1-5* and *qTFC-Jap-1-4*. However, due to the small interval size (only 26 kb) of *qTFC1-6*, the interval contained only six genes. From the Manhattan plot of the full population and the *japonica* subpopulation on chromosome 1, it was observed that there were still clustered SNPs belonging to these two QTLs below the threshold (Figure 4a). In addition, considering the average LD of rice is approximately 120~165 kb [44]. Therefore, to ensure the identification of reliable candidate genes, we adjusted the threshold for *qTFC-Full-1-5* and *qTFC-Jap-1-4* in their respective populations to $-\log_{10}(P) = 2.50$, based on the GWAS results of SNPs near these QTLs on chromosome 1 (Figure 4a). It can be seen that when the threshold was lowered to 2.50, there were not many significant SNPs near *qTFC-Full-1-5* and *qTFC-Jap-1-4* that did not belong to these two QTLs (Figure 4a), indicating that it was feasible to use the low threshold of $-\log_{10}(P) = 2.50$ when screening candidate genes for *qTFC1-6*. Using this new threshold, the intervals of *qTFC-Full-1-5* and *qTFC-Jap-1-4* were expanded to 186 kb (Chr1_34313030–Chr1_34499004) and 144 kb (Chr1_34355283–Chr1_34499004), respectively. Consequently, we selected the maximum interval involved in *qTFC-Full-1-5* and *qTFC-Jap-1-4* as the interval for *qTFC1-6*, which spanned from Chr1_34313030 to Chr1_34499004 and has a length of 186 kb. *qTFC1-6* contained 49 commonly significant SNPs in both the full and *japonica* populations.

Local LD analysis using all significant SNPs revealed that the LD block for *qTFC1-6* spanned from Chr1_34363101 to Chr1_34499004 (Figure 4b), covering a distance of 136 kb, and contains 22 genes, of which *LOC_Os01g59580* is a transposon that can be excluded. Among the remaining 21 genes, 13 have significant SNPs in their promoter or genebody regions. The gene responsible for regulating TFC in brown rice is likely to be expressed in grain-related tissues such as the ovary, embryo, and endosperm. To identify the promising candidate genes, we examined the expression levels of the 13 genes in these tissues at different stages after pollination using two rice gene expression databases, RiceXPro [50] and MSU [49]. It was found that *LOC_Os01g59440* and *LOC_Os01g59600* had the highest expression levels in the ovary, embryo, and endosperm at different stages after pollination, while *LOC_Os01g59610* had moderate expression, and *LOC_Os01g59630*, *LOC_Os01g59620*, *LOC_Os01g59530*, and *LOC_Os01g59420* exhibited extremely low expression levels (Figure 4c,d). *LOC_Os01g59430*, *LOC_Os01g59450*, *LOC_Os01g59460*, *LOC_Os01g59470*, *LOC_Os01g59480*, and *LOC_Os01g59640* were not expressed in these tissues at different stages after pollination (Figure 4c,d). Therefore, *LOC_Os01g59440*, *LOC_Os01g59600*, and *LOC_Os01g59610* are the most likely candidates for *qTFC1-6*. Haplotype analysis of these three genes using significant SNPs revealed that they were divided into two haplotypes (Hap1 and Hap2) in all germplasms. In the *japonica* subpopulation, the average TFC of Hap1 was significantly

higher than that of Hap2, while in the *indica* subpopulation, there was no significant difference in the mean TFC between Hap1 and Hap2 (Figure 4e). This result further supports the candidacy of these three genes. Therefore, *LOC_Os01g59440*, *LOC_Os01g59600*, and *LOC_Os01g59610* were the three candidate genes for *qTFC1-6* (Table S4). *LOC_Os01g59440* encodes a receptor-like kinase. *OsRLCK160* also encodes a receptor-like kinase that can interact with *OsZIP48* and phosphorylate it, thereby regulating the accumulation of flavonoids in rice leaves. Therefore, *LOC_Os01g59440* is likely to have a similar function to *OsRLCK160* and thus is more likely to be the functional gene for *qTFC1-6*; however, there is not enough evidence to exclude *LOC_Os01g59600* and *LOC_Os01g59610* from the candidates.

Table 2. Information on all co-located QTLs.

QTL Name Merged	QTL Name of the Respective Populations	Chr.	Lead SNP	$-\log_{10}(P)$	Left SNP	Right SNP	Length of Overlapping Interval	No. of Commonly sig. SNP	Cloned Gene
<i>qTFC7-2</i>	<i>qTFC-Full-7-1</i>	7	6141196	7.91	6062744	6468059	371872	104	<i>Rc</i>
	<i>qTFC-Ind-7-2</i>	7	6069266	15.76	5921983	6434616			<i>Rc</i>
<i>qTFC1-6</i>	<i>qTFC-Full-1-5</i>	1	34468941	8.58	34464710	34491418	26708	36	
	<i>qTFC-Jap-1-4</i>	1	34467581	8.81	34464710	34491418			
<i>qTFC9-7</i>	<i>qTFC-Full-9-5</i>	9	14419332	8.29	14014590	14471720	348111	30	
	<i>qTFC-Jap-9-2</i>	9	14210477	7.66	14111737	14459848			
<i>qTFC1-2</i>	<i>qTFC-Full-1-1</i>	1	6313712	5.75	6251537	6340212	88675	21	
	<i>qTFC-Jap-1-2</i>	1	6368734	8.65	6226226	6461894			
<i>qTFC4-3</i>	<i>qTFC-Full-4-1</i>	4	25197717	6.79	25029953	25399507	369554	11	
	<i>qTFC-Ind-4-2</i>	4	25197717	6.50	24981902	25399507			
<i>qTFC7-3</i>	<i>qTFC-Full-7-2</i>	7	6749334	6.18	6749334	6953995	129030	7	
	<i>qTFC-Ind-7-3</i>	7	6782095	7.96	6723681	6878364			
<i>qTFC5-4</i>	<i>qTFC-Full-5-1</i>	5	22309638	8.23	22280681	22468262	47736	2	
	<i>qTFC-Jap-5-1</i>	5	22317452	5.61	22309638	22357374			
<i>qTFC6-1</i>	<i>qTFC-Full-6-1</i>	6	2046161	6.35	2046147	2046165	18	4	
	<i>qTFC-Ind-6-1</i>	6	2046161	6.25	2046147	2046688			
<i>qTFC1-4</i>	<i>qTFC-Full-1-3</i>	1	25460333	6.54	25454033	25496676	16256	0	
	<i>qTFC-Jap-1-3</i>	1	25479651	5.11	25478991	25495247			
<i>qTFC3-3</i>	<i>qTFC-Jap-3-1</i>	3	10736942	5.67	10736930	11139080	155123	0	
	<i>qTFC-Ind-3-2</i>	3	11092283	5.66	10983957	11271479			
<i>qTFC8-3</i>	<i>qTFC-Full-8-1</i>	8	20869072	5.63	20844005	20939712	95707	0	
	<i>qTFC-Ind-8-1</i>	8	20653491	9.11	20648615	20942578			
<i>qTFC10-2</i>	<i>qTFC-Full-10-1</i>	10	11225165	7.48	11220698	11225165	4467	0	
	<i>qTFC-Jap-10-1</i>	10	11228250	6.96	11137616	11704622			
<i>qTFC11-4</i>	<i>qTFC-Full-11-1</i>	11	23285861	5.44	23111961	23406804	114658	0	
	<i>qTFC-Jap-11-2</i>	11	23170730	5.56	23170584	23285242			

3.4. Identification of Candidate Genes for *qTFC9-7*

qTFC9-7 was a co-located QTL of *qTFC-Full-9-5* in the full population and *qTFC-Jap-9-2* in the *japonica* subpopulation, with an overlapping interval of 78 kb (Chr9_14393345–Chr9_14459848) (Table 2). Both *qTFC-Full-9-5* and *qTFC-Jap-9-2* contained clusters of significant SNPs. *qTFC-Full-9-5* had the second highest association signal in the full population, and *qTFC-Jap-9-2* had the third highest association signal in the *japonica* subpopulation (Figure 2a,c; Table S2). The lead SNPs for *qTFC-Full-9-5* and *qTFC-Jap-9-2* had $-\log_{10}(P)$ values of 8.29 and 7.66, respectively (Table 2). There were 30 commonly significant SNPs between *qTFC-Full-9-5* and *qTFC-Jap-9-2*. To ensure the identification of reliable candidate genes, we lowered the threshold to $-\log_{10}(P) = 4.00$ for *qTFC-Full-9-6* and $-\log_{10}(P) = 3.70$ for *qTFC-Jap-9-3* based on the GWAS results of SNPs near these two QTLs on chromosome 9 (Figure 5a). It can be seen that after the threshold was lowered, there were not many significant SNPs near *qTFC-Full-9-7* and *qTFC-Jap-9-2* that did not belong to these two QTLs (Figure 5a), indicating that it was feasible to lower the threshold when screening candidate genes for *qTFC9-7*. Based on the new threshold, the interval for *qTFC9-7* spanned from Chr9_13979678 to Chr9_14475412, covering a distance of 496 kb, and contained 79 commonly significant SNPs in both the full and *japonica* populations.

Local LD analysis using all significant SNPs showed that the LD block for *qTFC9-7* spanned from Chr9_14210097 to Chr9_14475412, covering a distance of 265 kb and containing 52 genes. After excluding transposons, retrotransposons, and hypothetical proteins, 33 genes remained, of which 13 have significant SNPs in their promoter or genebody regions. The gene responsible for regulating TFC in brown rice is likely expressed in grain-related tissues such as the ovary, embryo, and endosperm. By examining the expression levels of these 13 genes in these tissues at different stages after pollination using RiceXPro [50] and MSU [49], it was found that *LOC_Os09g24260*, *LOC_Os09g24200*, and *LOC_Os09g24250* had high expression levels in the ovary, embryo, and endosperm at different stages after pollination (Figure 5c,d). *LOC_Os09g24190*, *LOC_Os09g24210*, *LOC_Os09g24290*, *LOC_Os09g24320*, and *LOC_Os09g24330* had moderate expression, while other genes had very low or no expression (Figure 5c,d). Therefore, the eight genes with high and moderate expression in grain-related tissues were possible candidate genes. Among these eight genes, five have commonly significant SNPs, which are *LOC_Os09g24200*, *LOC_Os09g24250*, *LOC_Os09g24260*, *LOC_Os09g24290*, and *LOC_Os09g24320* (Table S5). They were stably detected in GWAS of both the full and *japonica* populations. Haplotype analysis of these five genes using commonly significant SNPs revealed that they were divided into two haplotypes (Hap1 and Hap2) in all germplasms. In the *japonica* subpopulation, the average TFC of Hap1 was significantly higher than that of Hap2, while in the *indica* subpopulation, all accessions belonged to Hap2 (Figures 5e and S1). This result further supports the candidacy of these five genes for *qTFC9-7*. Therefore, *LOC_Os09g24200*, *LOC_Os09g24250*, *LOC_Os09g24260*, *LOC_Os09g24290*, and *LOC_Os09g24320* were candidate genes for *qTFC9-7*. Among these five genes, *LOC_Os09g24260* encodes a protein containing a WD domain and G-beta repeat domain, which is closest to the lead SNP in the full population and has more commonly significant SNPs (Table S5). *OsWD40/OsTTG1* also encodes a WD40 protein, and its mutations significantly reduced the accumulation of flavonoids in various organs of rice [32]. The WD40 protein is a component of the MYB-bHLH-WD40 (MBW) transcription factor complex that regulates the expression of some structural genes involved in flavonoid synthesis [54,55]. Therefore, *LOC_Os09g24260* is the most likely candidate gene in *qTFC9-7*; however, there is not enough evidence to exclude the other four genes from the list of candidates.

3.5. Geographic Distribution and Breeding Utilization Analysis of Favorable Haplotypes of Preferred Candidate Genes of *qTFC1-6* and *qTFC9-7*

LOC_Os01g59440 and *LOC_Os09g24260* are the most promising candidate genes for *qTFC1-6* and *qTFC9-7*, respectively, and they are divided into two haplotypes (Hap1 and Hap2) in all germplasms (Figures 4e and 5e). *LOC_Os01g59440*^{Hap1} is the favorable haplotype of *LOC_Os01g59440* in the *japonica* subpopulation, and *LOC_Os09g24260*^{Hap1} is the favorable haplotype of *LOC_Os09g24260* in the *japonica* subpopulation. However, almost all *indica* germplasm belonged to *LOC_Os01g59440*^{Hap2} and *LOC_Os09g24260*^{Hap2}.

To investigate the geographic distribution and breeding utilization of *LOC_Os01g59440*^{Hap1} and *LOC_Os09g24260*^{Hap1} in *japonica*, we conducted geographical distribution analysis and breeding utilization analysis on them. The results showed that germplasms belonging to *LOC_Os01g59440*^{Hap1} in the *japonica* subpopulation was widely distributed across multiple regions worldwide, including areas with varying longitudes and latitudes (Figure 6a). Similarly, germplasm belonging to *LOC_Os09g24260*^{Hap1} in the *japonica* subpopulation was also widely distributed across multiple regions worldwide, including areas with varying longitudes and latitudes (Figure 6b). The proportion of *LOC_Os01g59440*^{Hap1} in landraces of the *japonica* subpopulation was 32%, while it was 9% in improved varieties of the *japonica* subpopulation (Figure 6c). This indicates that the utilization rate of *LOC_Os01g59440*^{Hap1} in improved varieties is still relatively low, and there is great potential for future applications of this favorable haplotype. Similarly, the proportion of *LOC_Os09g24260*^{Hap1} in landraces of the *japonica* subpopulation was 33%, while it was 20% in improved varieties of the *japonica* subpopulation (Figure 6d). This indicates that the utilization rate of *LOC_Os09g24260*^{Hap1}

in improved varieties is also relatively low, and there is significant potential for future applications of this favorable haplotype.

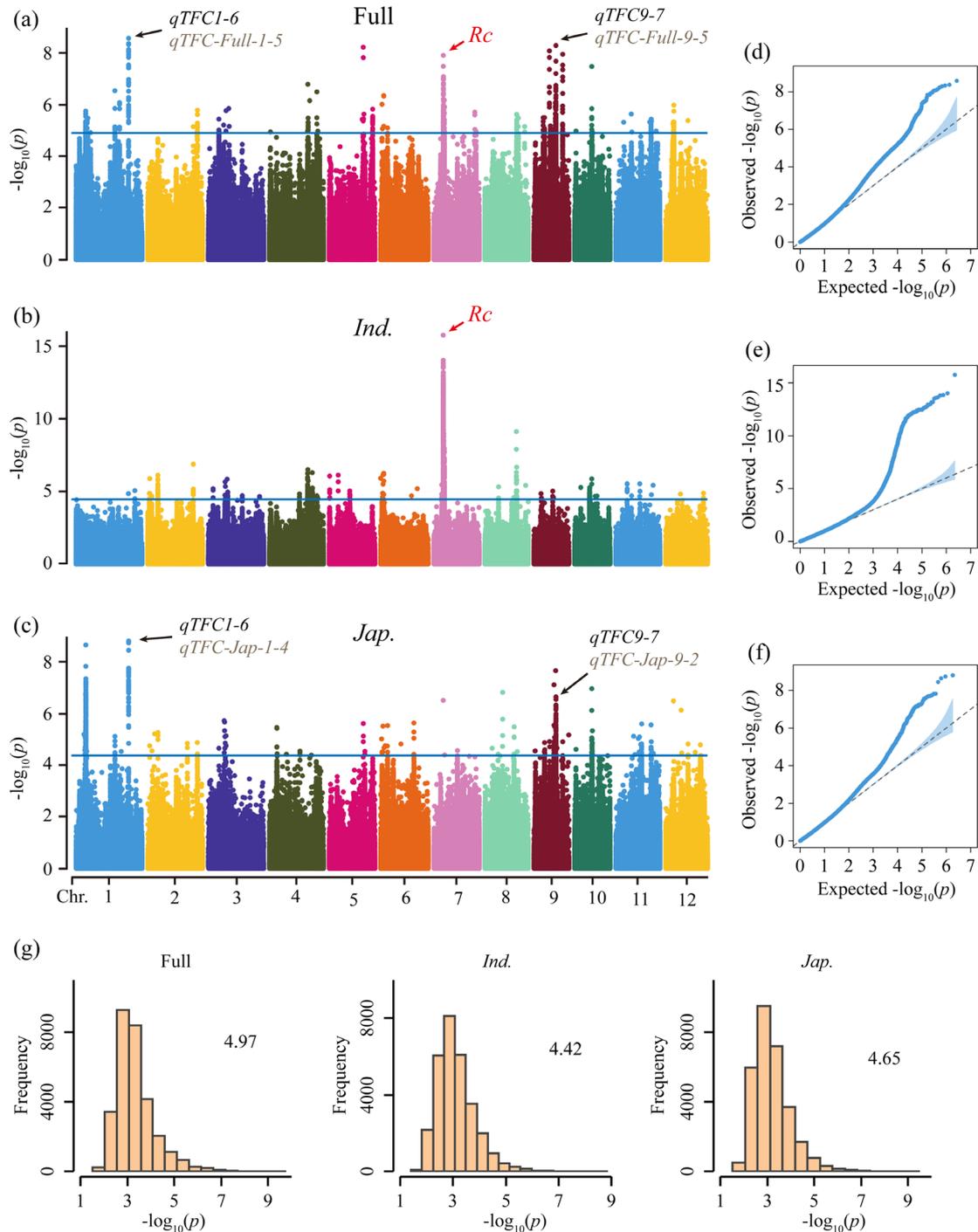


Figure 2. Genome-wide association study (GWAS) of TFC in brown rice in the full, *indica*, and *japonica* populations. (a–c) Manhattan plots showing the GWAS results for the full, *indica*, and *japonica* populations, respectively. (d–f) Quantile-quantile (Q-Q) plots illustrating the distribution of observed p -values compared to expected p -values in the full, *indica*, and *japonica* populations, respectively. (g) Histogram displaying the distribution of the maximum $-\log_{10}(P)$ values from 1000 permutations. The significance thresholds for GWAS are provided on the right side of each plot.

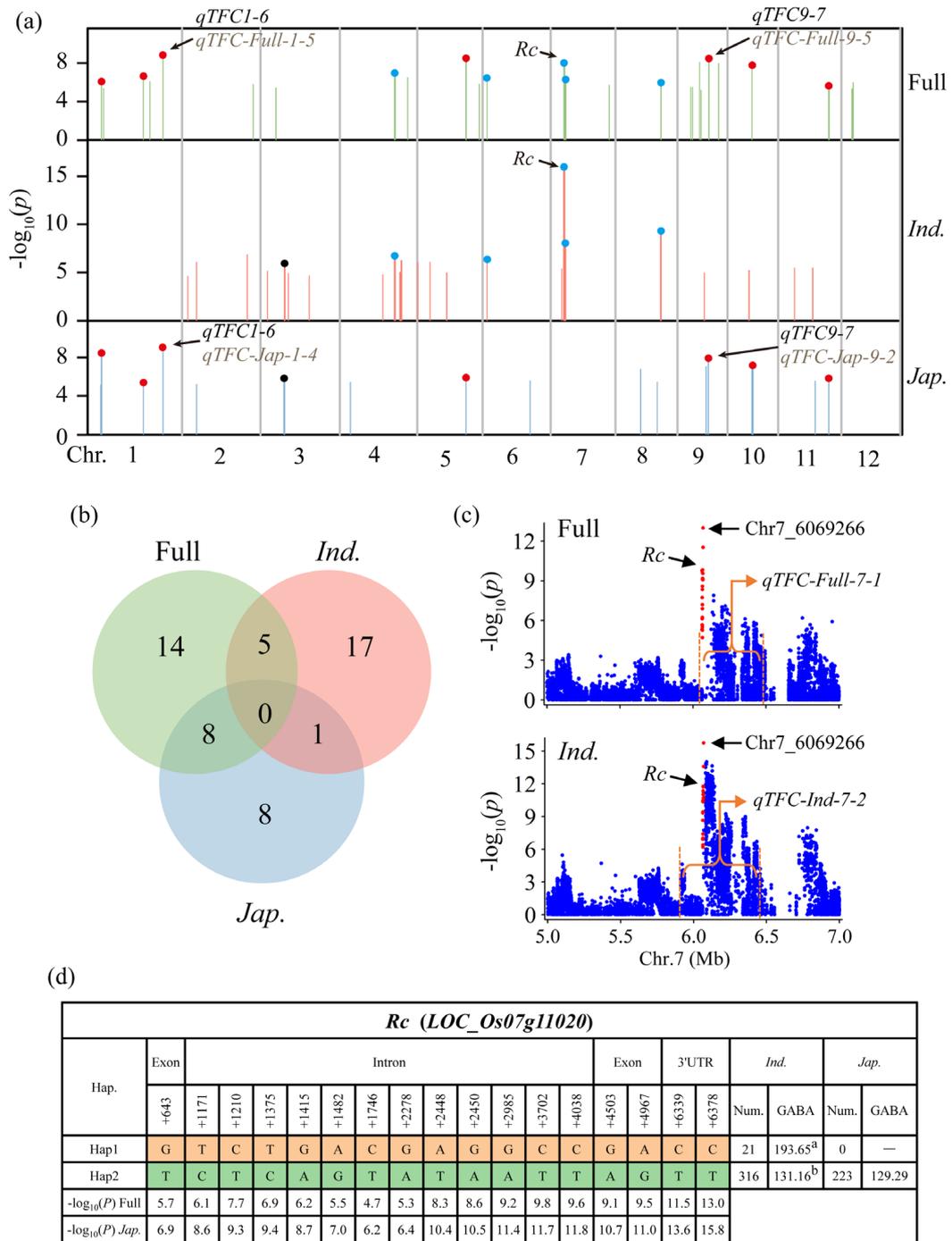


Figure 3. Summary of quantitative trait loci (QTLs). (a) Summary of QTLs in the full, *indica*, and *japonica* populations. Each bar represents a QTL. The blue dots indicate the QTLs co-located in both the full and *indica* populations. The red dots indicate the QTLs co-located in both the full and *japonica* populations. The black dots indicate the QTLs co-located in both the *indica* and *japonica* populations. (b) A venn diagram showing the number of QTLs identified in each population and the overlap between them. (c) Local Manhattan plot of chromosome 7 from 5 Mb to 7 Mb in the full population (top) and *indica* subpopulation (bottom). The interval inside the orange dotted line indicates the QTL. The red dots indicate significant SNPs located on *Rc*. Chr7_6069266 is the same lead SNP as *qTFC-Full-7-1* and *qTFC-Ind-7-2*. (d) Haplotype analysis of *Rc*. Different letters indicate significant differences at $p < 0.05$ according to double-tailed Student's *t*-tests.

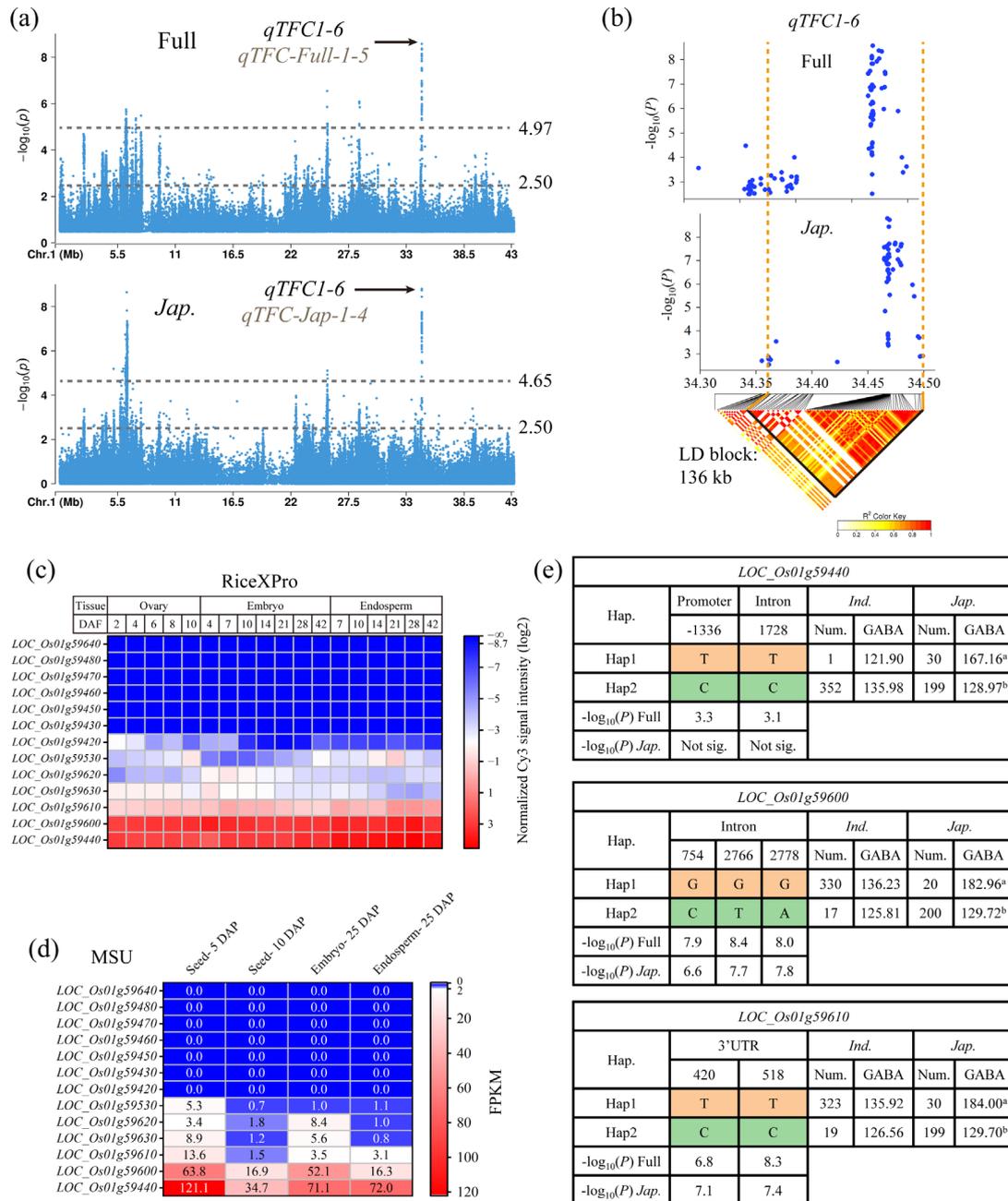


Figure 4. Exploration of candidate genes in *qTFC1-6*. (a) Manhattan plots of the full and *japonica* populations on chromosome 1. (b) Local Manhattan plots and linkage disequilibrium plots of *qTFC1-6* based on significant single nucleotide polymorphisms (SNPs) in the full and *japonica* populations. (c) Relative expression levels of the 13 primary candidate genes of *qTFC1-6* in grain-related tissues in the RiceXPro gene expression database. DAF represents the number of days after flowering. The relative expression levels are represented by the normalized Cy3 fluorescence signal intensity. Since the normalization is based on a logarithmic scale of two, genes with no expression are represented as negative infinity ($-\infty$) in blue. (d) Relative expression levels of the 13 primary candidate genes of *qTFC1-6* in grain-related tissues in the MSU gene expression database. The relative expression levels are represented by fragments per kilobase of the exon model per million mapped fragments (FPKM). (e) Haplotype analysis of three candidate genes. Different letters indicate significant differences at $p < 0.05$ according to double-tailed Student's *t*-tests.

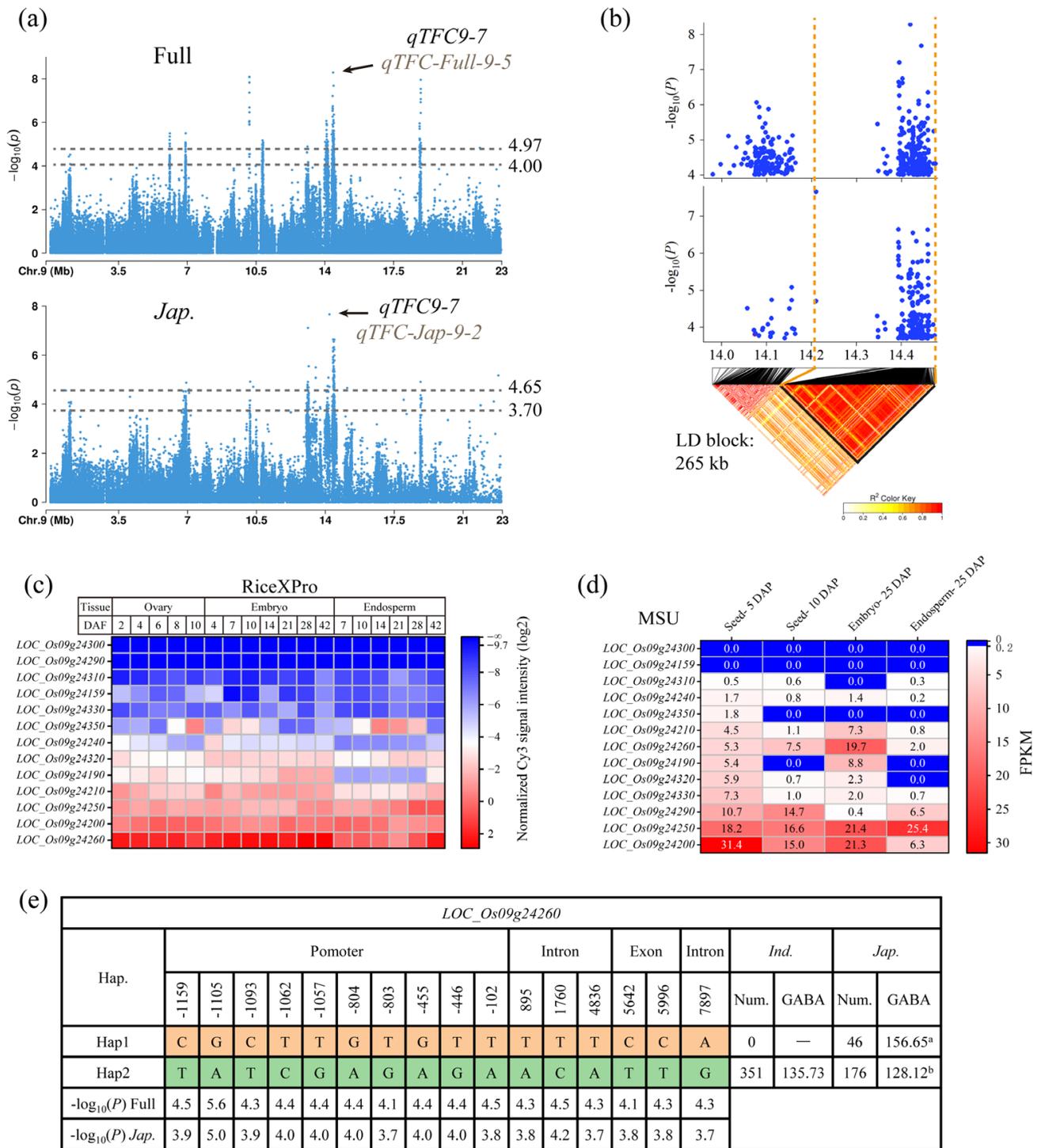


Figure 5. Exploration of candidate genes in *qTFC9-7*. (a) Manhattan plots of the full and *japonica* populations on chromosome 9. (b) Local Manhattan plots and linkage disequilibrium plots of *qTFC9-7* based on significant SNPs in the full and *japonica* populations. (c) Relative expression levels of the 13 primary candidate genes of *qTFC9-7* in grain-related tissues in the RiceXPro gene expression database. (d) Relative expression levels of the five primary candidate genes of *qTFC9-7* in grain-related tissues in the MSU gene expression database. (e) Haplotype analysis of *LOC_Os09g24260*. Different letters indicate significant differences at $p < 0.05$ according to double-tailed Student's *t*-tests.

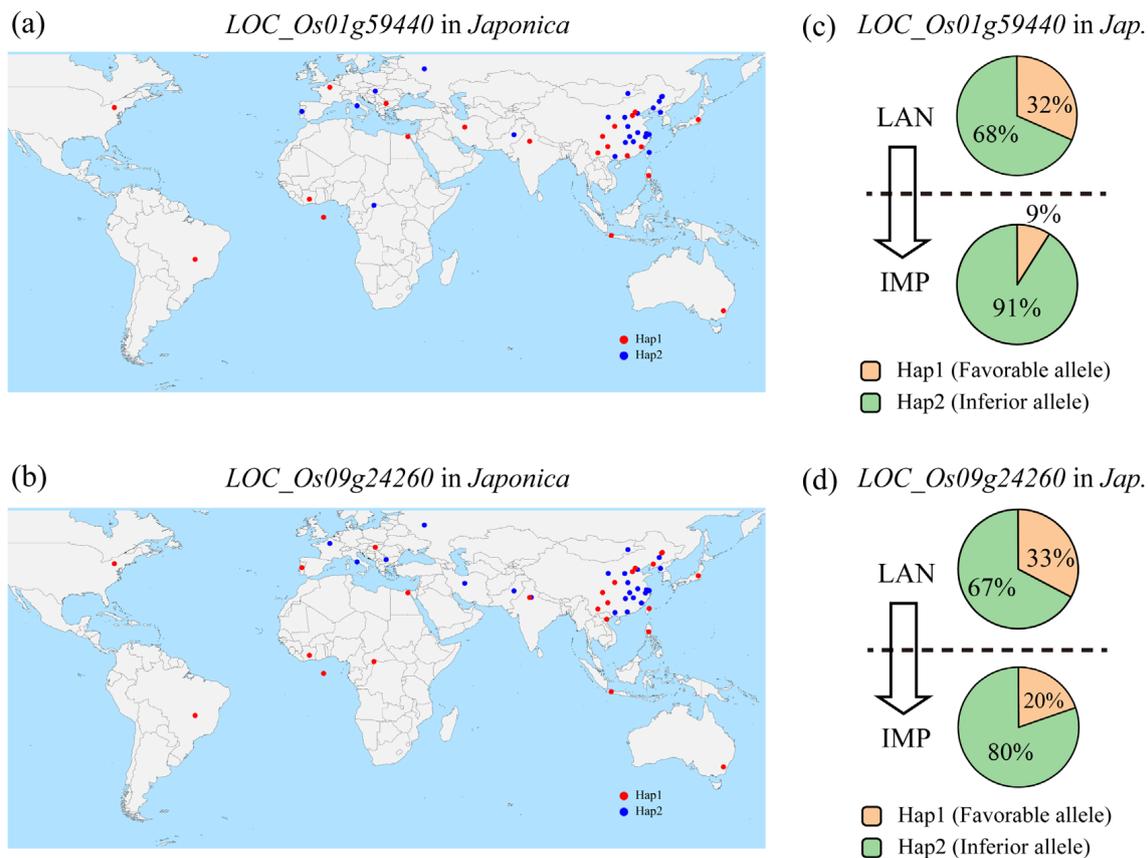


Figure 6. The geographic distribution and breeding utilization analysis of *LOC_Os01g59440*^{Hap1} and *LOC_Os09g24260*^{Hap1}. (a) The geographic distribution of *LOC_Os01g59440*^{Hap1} in the *japonica* subpopulation. (b) The geographic distribution of *LOC_Os09g24260*^{Hap1} in the *japonica* subpopulation. (c) The breeding utilization of *LOC_Os01g59440*^{Hap1} in the *japonica* subpopulation. LAN represents a landrace. IMP represents improved variety. (d) The breeding utilization of *LOC_Os09g24260*^{Hap1} in the *japonica* subpopulation.

4. Discussion

Flavonoids are known to have anti-inflammatory, antioxidative, and anticarcinogenic effects. Breeding functional rice varieties rich in flavonoids may prevent chronic diseases such as cancer and cardio-cerebrovascular disease.

In a previous study, the TFC of 164 accessions from the USDA rice mini-core collection ranged from 43.4 to 411.9 mg/100 g [53]. In our study, the TFC ranged from 90.6 to 435.9 mg/100 g. Although the range of TFC in our study did not exceed that of the previous study, we identified elite accessions with higher TFC, such as Yanshuichi (TFC = 435.91 mg/100 g), which could serve as valuable germplasm resources for breeding. The TFC in the full population showed a positively skewed distribution because the TFC of 11 accessions exceeded 300 mg/100 g, and 93% of the accessions ranged from 90 to 170 mg/100 g (Figure 1b). This suggests that a few accessions may contain favorable alleles that control flavonoid content.

We found no significant difference in TFC between the *indica* and *japonica* subpopulations (Figure 1f), and the 10 varieties with the highest TFC included three *indica* and seven *japonica* varieties, indicating no apparent relationship between TFC and geographic origin. Our findings are consistent with those of a previous study that found no significant difference in TFC between the *indica* and *japonica* subpopulations in the primary core collection from the Yunnan province of China [11]. However, Dong et al. (2014) reported significant differences between the two subpopulations for specific flavonoids [56]. For instance, two flavone C-pentosides in flag leaves, apigenin C-pentoside and luteolin

C-pentoside, showed over-accumulation by 200 times in the *indica* subpopulation compared to the *japonica* subpopulation, and coumaroyl derivatives of flavone C-hexosyl-O-hexosides in flag leaves were found to have lower levels in the *indica* subpopulation compared to the *japonica* subpopulation. These results suggest that although we found no significant difference in TFC between the *indica* and *japonica* subpopulations, there may still be significant differences in specific flavonoids.

Our analysis found that among the 10 accessions with high flavonoid content, nine were landrace varieties, indicating that landraces are a promising source for identifying excellent varieties with high flavonoid content. Therefore, it is crucial to carefully collect and phenotypically identify landrace germplasm.

When looking for genes that have been reported, although some genes affecting flavonoid content were reported to be expressed in grains, no direct evidence that these genes affect flavonoid content in grains was found. Thus, they could not be regarded as genes affecting the flavonoid content of rice grains. For example, although *OsCHS24* and *OsbZIP48* are expressed in grains and leaves and affect the content of flavonoids in leaves [28,57], no evidence that *OsCHS24* and *OsbZIP48* affect the flavonoid content in grains was found. Therefore, they do not belong to the genes that affect the flavonoid content of rice grains.

When screening candidate genes, previous studies considered that if significant SNPs were located in the promoter of certain genes or belonged to non-synonymous SNPs in exons, such genes might be candidate genes [58,59], which would ignore genes with significant SNPs located in introns and 3'UTRs. However, variations in the intron or 3'UTR can also lead to changes in gene expression, thus affecting gene function [60–62]. To avoid disregarding potential candidate genes, we considered that genes with significant SNPs in the entire gene region, including the promoter, exon, intron, and/or 3'UTR, might be candidates. Therefore, in our study, we included *LOC_Os01g59600* and *LOC_Os01g59610* as candidate genes, even though their significant SNPs were located in the intron and 3'UTR (Figure 4e; Table S4).

The genes involved in the synthesis and regulation of flavonoids in rice have been widely studied, including structural, regulatory, transporter, and modifying enzyme genes. However, among the genes that have been reported to affect the flavonoid content in rice, there are still a few that do not belong to the known pathways of flavonoid synthesis or regulation, such as *OsRLCK160*, *POT*, and *OsCRY1b* [57,63,64], which encode a receptor-like kinase, proton-dependent oligopeptide transporter, and cryptochrome, respectively. The relationship between these proteins and flavonoid synthesis or regulatory pathways is still unknown, indicating a lack of knowledge about the mechanisms of flavonoid content in rice. Therefore, even though *LOC_Os01g59440*, *LOC_Os01g59600*, *LOC_Os01g59610*, *LOC_Os09g24200*, *LOC_Os09g24250*, *LOC_Os09g24290*, and *LOC_Os09g24320*, the seven candidate genes identified in our study for *qTFC1-6* and *qTFC9-7*, encode a brassinosteroid insensitive 1-associated receptor kinase 1 precursor, T1 family peptidase, KAZ1-Kazal-type serine protease inhibitor precursor, RAD23 DNA repair protein, agenet domain containing protein, and 2-oxo acid dehydrogenase acyltransferase domain containing protein, respectively, which appear to be unrelated to known flavonoid synthesis and regulatory pathways, they could also affect flavonoid content through other unknown pathways. For example, *LOC_Os01g59440*, encoding a brassinosteroid insensitive 1-associated receptor kinase 1 precursor, is similar to *OsRLCK160*, which encodes a receptor-like kinase that interacts with *OsbZIP48* and phosphorylates it, thereby regulating the accumulation of flavonoids in rice leaves [57]. Therefore, *LOC_Os01g59440* could regulate TFC in brown rice through similar unknown mechanisms.

Another candidate gene identified in our study, *LOC_Os09g24260* of *qTFC9-7*, encodes a WD domain and G-beta repeat domain-containing protein, which is related to the known flavonoid content pathway. The WD40 protein is a component of the MYB-bHLH-WD40 (MBW) transcription factor complex that regulates the expression of some structural genes

involved in flavonoid synthesis [54,55]. Therefore, *LOC_Os09g24260* is the most likely candidate gene for *qTFC9-7*.

5. Conclusions

In this study, we identified ten excellent rice varieties with high TFC and detected 53 QTLs associated with TFC in brown rice. From two preferred co-located QTLs, *qTFC1-6* and *qTFC9-7*, we identified eight candidate genes: *LOC_Os01g59440*, *LOC_Os01g59600*, *LOC_Os01g59610*, *LOC_Os09g24200*, *LOC_Os09g24250*, *LOC_Os09g24260*, *LOC_Os09g24290*, and *LOC_Os09g24320*. Among these genes, *LOC_Os01g59440* and *LOC_Os09g24260*, which encode a receptor-like kinase and a WD domain and G-beta repeat domain-containing protein, respectively, are the most likely candidate genes. We further analyzed the geographic distribution and breeding utilization of favorable haplotypes of these two genes. This study elucidates the genetic basis of total flavonoid content in brown rice, which could help in developing rice cultivars with high flavonoid levels that offer benefits in the prevention and adjuvant treatment of cardio-cerebrovascular diseases and cancer.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/genes14091684/s1>. xxx. Figure S1. Haplotype analysis of the other four candidate genes of *qTFC9-7*. Table S1. The information of 633 cultivated rice accessions. Table S2. Information of the QTLs associated with total flavonoid content in brown rice identified by GWAS in the full, *indica*, and *japonica* populations. Table S3. Significant SNPs of *Rc* in the full and *indica* populations. Table S4. All significant SNPs from the full and *indica* populations within the LD interval of *qTFC1-6*. Table S5. All significant SNPs from the full and *japonica* populations within the LD interval of *qTFC9-7*.

Author Contributions: Conceptualization, Y.Z. and Z.L.; Data curation, H.X. and X.Z.; Formal analysis, H.X. and X.P.; Funding acquisition, H.Z., Z.Z. and Y.Z.; Investigation, X.P., X.Y. and H.D.; Methodology, X.Z. and H.G.; Project administration, H.Z.; Resources, Z.Z.; Software, Y.W.; Supervision, Y.Z. and Z.L.; Visualization, X.P., H.G. and Y.W.; Writing—original draft, H.X. and X.P.; Writing—review and editing, Q.Z. and X.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by grants from the Ministry of Science and Technology of China (2021YFD1200502) and the National Natural Science Foundation of China (31760376), as well as the AgroST Project (NK2022050103).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data in the present study are available in the public database, as mentioned in Section 4.

Acknowledgments: The authors are grateful to all the laboratory members for their continuous technical advice and helpful discussion.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Meetoo, D. Chronic diseases: The silent global epidemic. *Br. J. Nurs.* **2008**, *17*, 1320–1325. [[CrossRef](#)] [[PubMed](#)]
2. Nugent, R. Chronic diseases in developing countries: Health and economic burdens. *Ann. N. Y. Acad. Sci.* **2008**, *1136*, 70–79. [[CrossRef](#)]
3. Yach, D.; Leeder, S.R.; Bell, J.; Kistnasamy, B. Global chronic diseases. *Science* **2005**, *307*, 317. [[CrossRef](#)]
4. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)]
5. Ferdinand, K.C.; Reddy, T.K.; Vo, T.N. Global interventions in hypertension: New and emerging concepts. *Curr. Opin. Cardiol.* **2021**, *36*, 436–443. [[CrossRef](#)]
6. Hengel, F.E.; Sommer, C.; Wenzel, U. Arterielle Hypertonie—Eine Übersicht für den ärztlichen Alltag. *DMW Dtsch. Med. Wochenschr.* **2022**, *147*, 414–428. [[CrossRef](#)]

7. Liu, W.; Feng, Y.; Yu, S.; Fan, Z.; Li, X.; Li, J.; Yin, H. The Flavonoid Biosynthesis Network in Plants. *Int. J. Mol. Sci.* **2021**, *22*, 12824. [[CrossRef](#)]
8. Ferreyra, M.L.F.; Casas, M.I.; Questa, J.I.; Herrera, A.L.; DeBlasio, S.; Wang, J.; Jackson, D.; Grotewold, E.; Casati, P. Evolution and expression of tandem duplicated maize flavonol synthase genes. *Front. Plant Sci.* **2012**, *3*, 101. [[CrossRef](#)]
9. Ramesh, P.; Jagadeesan, R.; Sekaran, S.; Dhanasekaran, A.; Vimalraj, S. Flavonoids: Classification, Function, and Molecular Mechanisms Involved in Bone Remodelling. *Front. Endocrinol.* **2021**, *12*, 779638. [[CrossRef](#)]
10. Verma, V.; Vishal, B.; Kohli, A.; Kumar, P.P. Systems-based rice improvement approaches for sustainable food and nutritional security. *Plant Cell Rep.* **2021**, *40*, 2021–2036. [[CrossRef](#)] [[PubMed](#)]
11. Zeng, Y.-W.; Du, J.; Yang, S.-M.; Pu, X.-Y.; Wang, Y.-C.; Yang, T.; Sun, Z.-H.; Xin, P.-Y. The Zonal Characteristics and Cultivated Types Difference of Functional Components in Brown Rice for Core Collection of Yunnan Rice. *Spectrosc. Spectr. Anal.* **2010**, *30*, 3388–3394. [[CrossRef](#)]
12. Ghosh, P.; Roychoudhury, A. Nutrition and antioxidant profiling in the unpolished and polished grains of eleven indigenous aromatic rice cultivars. *3 Biotech* **2020**, *10*, 548. [[CrossRef](#)]
13. Kim, J.K.; Lee, S.Y.; Chu, S.M.; Lim, S.H.; Suh, S.-C.; Lee, Y.-T.; Cho, H.S.; Ha, S.-H. Variation and correlation analysis of flavonoids and carotenoids in korean pigmented rice (*Oryza sativa* L.) cultivars. *J. Agric. Food Chem.* **2010**, *58*, 12804–12809. [[CrossRef](#)] [[PubMed](#)]
14. Nayeem, S.; Venkidasamy, B.; Sundararajan, S.; Kuppuraj, S.P.; Ramalingam, S. Differential expression of flavonoid biosynthesis genes and biochemical composition in different tissues of pigmented and non-pigmented rice. *J. Food Sci. Technol.* **2020**, *58*, 884–893. [[CrossRef](#)] [[PubMed](#)]
15. Singh, S.P.; Vanlalsanga; Mehta, S.; Singh, Y.T. New insight into the pigmented rice of northeast India revealed high antioxidant and mineral compositions for better human health. *Heliyon* **2022**, *8*, e10464. [[CrossRef](#)] [[PubMed](#)]
16. Mackon, E.; Mackon, G.C.J.D.E.; Ma, Y.; Kashif, M.H.; Ali, N.; Usman, B.; Liu, P. Recent Insights into Anthocyanin Pigmentation, Synthesis, Trafficking, and Regulatory Mechanisms in Rice (*Oryza sativa* L.) Caryopsis. *Biomolecules* **2021**, *11*, 394. [[CrossRef](#)]
17. Duan, L.; Liu, H.; Li, X.; Xiao, J.; Wang, S. Multiple phytohormones and phytoalexins are involved in disease resistance to *Magnaporthe oryzae* invaded from roots in rice. *Physiol. Plant.* **2014**, *152*, 486–500. [[CrossRef](#)]
18. Park, S.; Choi, M.J.; Lee, J.Y.; Kim, J.K.; Ha, S.-H.; Lim, S.-H. Molecular and Biochemical Analysis of Two Rice Flavonoid 3'-Hydroxylase to Evaluate Their Roles in Flavonoid Biosynthesis in Rice Grain. *Int. J. Mol. Sci.* **2016**, *17*, 1549. [[CrossRef](#)]
19. Park, S.; Kim, D.-H.; Park, B.-R.; Lee, J.-Y.; Lim, S.-H. Molecular and Functional Characterization of *Oryza sativa* Flavonol Synthase (OsFLS), a Bifunctional Dioxygenase. *J. Agric. Food Chem.* **2019**, *67*, 7399–7409. [[CrossRef](#)]
20. Ithal, N.; Reddy, A.R. Rice flavonoid pathway genes, OsDfr and OsAns, are induced by dehydration, high salt and ABA, and contain stress responsive promoter elements that interact with the transcription activator, OsC1-MYB. *Plant Sci.* **2004**, *166*, 1505–1513. [[CrossRef](#)]
21. Sweeney, M.T.; Thomson, M.J.; Pfeil, B.E.; McCouch, S. Caught Red-Handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* **2006**, *18*, 283–294. [[CrossRef](#)] [[PubMed](#)]
22. Sakamoto, W.; Ohmori, T.; Kageyama, K.; Miyazaki, C.; Saito, A.; Murata, M.; Noda, K.; Maekawa, M. The Purple leaf (Pl) Locus of rice: The plw allele has a complex organization and includes two genes encoding basic helix-loop-helix proteins involved in anthocyanin biosynthesis. *Plant Cell Physiol.* **2001**, *42*, 982–991. [[CrossRef](#)] [[PubMed](#)]
23. Zhang, F.; Guo, H.; Huang, J.; Yang, C.; Li, Y.; Wang, X.; Qu, L.; Liu, X.; Luo, J. A UV-B-responsive glycosyltransferase, OsUGT706C2, modulates flavonoid metabolism in rice. *Sci. China Life Sci.* **2020**, *63*, 1037–1052. [[CrossRef](#)]
24. Shimizu, T.; Lin, F.; Hasegawa, M.; Okada, K.; Nojiri, H.; Yamane, H. Purification and identification of naringenin 7-O-Methyltransferase, a key enzyme in biosynthesis of flavonoid phytoalexin sakuranetin in rice. *J. Biol. Chem.* **2012**, *287*, 19315–19325. [[CrossRef](#)]
25. Mackon, E.; Ma, Y.; Mackon, G.C.J.D.E.; Usman, B.; Zhao, Y.; Li, Q.; Liu, P. Computational and Transcriptomic Analysis Unraveled OsMATE34 as a Putative Anthocyanin Transporter in Black Rice (*Oryza sativa* L.) Caryopsis. *Genes* **2021**, *12*, 583. [[CrossRef](#)] [[PubMed](#)]
26. Akhter, D.; Qin, R.; Nath, U.K.; Eshag, J.; Jin, X.; Shi, C. A rice gene, OsPL, encoding a MYB family transcription factor confers anthocyanin synthesis, heat stress response and hormonal signaling. *Gene* **2019**, *699*, 62–72. [[CrossRef](#)] [[PubMed](#)]
27. Ko, J.H.; Kim, B.G.; Kim, J.H.; Kim, H.; Lim, C.E.; Lim, J.; Lee, C.; Lim, Y.; Ahn, J.-H. Four glucosyltransferases from rice: cDNA cloning, expression, and characterization. *J. Plant Physiol.* **2008**, *165*, 435–444. [[CrossRef](#)]
28. Park, H.L.; Yoo, Y.; Bhoo, S.H.; Lee, T.-H.; Lee, S.-W.; Cho, M.-H. Two Chalcone Synthase Isozymes Participate Redundantly in UV-Induced Sakuranetin Synthesis in Rice. *Int. J. Mol. Sci.* **2020**, *21*, 3777. [[CrossRef](#)]
29. Yang, Z.; Li, N.; Kitano, T.; Li, P.; Spindel, J.E.; Wang, L.; Bai, G.; Xiao, Y.; McCouch, S.R.; Ishihara, A.; et al. Genetic mapping identifies a rice naringenin O-glucosyltransferase that influences insect resistance. *Plant J.* **2021**, *106*, 1401–1413. [[CrossRef](#)]
30. Lam, P.Y.; Liu, H.; Lo, C. Completion of Tricin Biosynthesis Pathway in Rice: Cytochrome P450 75B4 Is a Unique Chrysoeriol 5'-Hydroxylase. *Plant Physiol.* **2015**, *168*, 1527–1536. [[CrossRef](#)]
31. Maeda, H.; Yamaguchi, T.; Omoteno, M.; Takarada, T.; Fujita, K.; Murata, K.; Iyama, Y.; Kojima, Y.; Morikawa, M.; Ozaki, H.; et al. Genetic dissection of black grain rice by the development of a near isogenic line. *Breed. Sci.* **2014**, *64*, 134–141. [[CrossRef](#)] [[PubMed](#)]

32. Yang, X.; Wang, J.; Xia, X.; Zhang, Z.; He, J.; Nong, B.; Luo, T.; Feng, R.; Wu, Y.; Pan, Y.; et al. *OsTTG1*, a WD40 repeat gene, regulates anthocyanin biosynthesis in rice. *Plant J.* **2021**, *107*, 198–214. [[CrossRef](#)] [[PubMed](#)]
33. Wang, J.; Deng, Q.; Li, Y.; Yu, Y.; Liu, X.; Han, Y.; Luo, X.; Wu, X.; Ju, L.; Sun, J.; et al. Transcription Factors *Rc* and *OsVP1* Coordinately Regulate Preharvest Sprouting Tolerance in Red Pericarp Rice. *J. Agric. Food Chem.* **2020**, *68*, 14748–14757. [[CrossRef](#)] [[PubMed](#)]
34. Kim, B.; Lee, Y.; Nam, J.-Y.; Lee, G.; Seo, J.; Lee, D.; Cho, Y.-H.; Kwon, S.-W.; Koh, H.-J. Mutations in *OsDET1*, *OsCOP10*, and *OsDDB1* confer embryonic lethality and alter flavonoid accumulation in Rice (*Oryza sativa* L.) seed. *Front. Plant Sci.* **2022**, *13*, 952856. [[CrossRef](#)]
35. Zhang, H.; Zhang, D.; Wang, M.; Sun, J.; Qi, Y.; Li, J.; Wei, X.; Han, L.; Qiu, Z.; Tang, S.; et al. A core collection and mini core collection of *Oryza sativa* L. in China. *Theor. Appl. Genet.* **2010**, *122*, 49–61. [[CrossRef](#)]
36. Yu, S.B.; Xu, W.J.; Vijayakumar, C.H.M.; Ali, J.; Fu, B.Y.; Xu, J.L.; Jiang, Y.Z.; Marghirang, R.; Domingo, J.; Aquino, C.; et al. Molecular diversity and multilocus organization of the parental lines used in the International Rice Molecular Breeding Program. *Theor. Appl. Genet.* **2003**, *108*, 131–140. [[CrossRef](#)]
37. He, S.M.; Liu, J.L. Study on the determination method of flavone content in tea. *Chin. J. Anal. Chem.* **2007**, *35*, 1365–1368.
38. Alexandrov, N.; Tai, S.; Wang, W.; Mansueto, L.; Palis, K.; Fuentes, R.R.; Ulat, V.J.; Chebotarov, D.; Zhang, G.; Li, Z.; et al. SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Res.* **2014**, *43*, D1023–D1027. [[CrossRef](#)]
39. Wang, W.; Mauleon, R.; Hu, Z.; Chebotarov, D.; Tai, S.; Wu, Z.; Li, M.; Zheng, T.; Fuentes, R.R.; Zhang, F.; et al. Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* **2018**, *557*, 43–49. [[CrossRef](#)]
40. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [[CrossRef](#)]
41. Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **2011**, *88*, 76–82. [[CrossRef](#)] [[PubMed](#)]
42. Tang, Y.; Liu, X.; Wang, J.; Li, M.; Wang, Q.; Tian, F.; Su, Z.; Pan, Y.; Liu, D.; Lipka, A.E.; et al. GAPIT Version 2: An Enhanced Integrated Tool for Genomic Association and Prediction. *Plant Genome* **2016**, *9*, plantgenome2015-11. [[CrossRef](#)] [[PubMed](#)]
43. Zhao, Y.; Zhang, H.; Xu, J.; Jiang, C.; Yin, Z.; Xiong, H.; Xie, J.; Wang, X.; Zhu, X.; 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](#)]
44. Huang, X.; Wei, X.; Sang, T.; Zhao, Q.; Feng, Q.; Zhao, Y.; Li, C.; Zhu, C.; Lu, T.; Zhang, Z.; et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **2010**, *42*, 961–967. [[CrossRef](#)]
45. Mather, A.K.; Caicedo, A.L.; Polato, N.R.; Olsen, K.M.; McCouch, S.; Purugganan, M.D. The Extent of Linkage Disequilibrium in Rice (*Oryza sativa* L.). *Genetics* **2007**, *177*, 2223–2232. [[CrossRef](#)]
46. Guo, H.; Zeng, Y.; Li, J.; Ma, X.; Zhang, Z.; Lou, Q.; Li, J.; Gu, Y.; Zhang, H.; Li, J.; et al. Differentiation, evolution and utilization of natural alleles for cold adaptability at the reproductive stage in rice. *Plant Biotechnol. J.* **2020**, *18*, 2491–2503. [[CrossRef](#)] [[PubMed](#)]
47. Shin, J.H.; Blay, S.; McNeney, B.; Graham, J. Ldheatmap: An R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J. Stat. Softw.* **2006**, *16*, 1–9. [[CrossRef](#)]
48. Yano, K.; Yamamoto, E.; Aya, K.; Takeuchi, H.; Lo, P.-C.; Hu, L.; Yamasaki, M.; Yoshida, S.; Kitano, H.; Hirano, K.; et al. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* **2016**, *48*, 927–934. [[CrossRef](#)]
49. 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](#)]
50. Sato, Y.; Antonio, B.A.; Namiki, N.; Takehisa, H.; Minami, H.; Kamatsuki, K.; Sugimoto, K.; Shimizu, Y.; Hirochika, H.; Nagamura, Y. RiceXPro: A platform for monitoring gene expression in japonica rice grown under natural field conditions. *Nucleic Acids Res.* **2010**, *39*, D1141–D1148. [[CrossRef](#)]
51. Zhang, Z.; Ersoz, E.; Lai, C.-Q.; Todhunter, R.J.; Tiwari, H.K.; Gore, M.A.; Bradbury, P.J.; Yu, J.; Arnett, D.K.; Ordo vas, J.M.; et al. Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* **2010**, *42*, 355–360. [[CrossRef](#)] [[PubMed](#)]
52. Furukawa, T.; Maekawa, M.; Oki, T.; Suda, I.; Iida, S.; Shimada, H.; Takamura, I.; Kadowaki, K.-I. The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp. *Plant J.* **2007**, *49*, 91–102. [[CrossRef](#)] [[PubMed](#)]
53. Li, K.; Li, Q.; Wang, L.Y.; Ren, H.; Ge, Y. Genetic variation and association mapping of phenolic, flavonoid content and antioxidant capacity in USDA rice mini-core collection. *Genet. Resour. Crop. Evol.* **2022**, *69*, 1685–1694. [[CrossRef](#)]
54. Xu, W.; Dubos, C.; Lepiniec, L. Transcriptional control of flavonoid biosynthesis by MYB–bHLH–WDR complexes. *Trends Plant Sci.* **2015**, *20*, 176–185. [[CrossRef](#)] [[PubMed](#)]
55. Petroni, K.; Tonelli, C. Recent advances on the regulation of anthocyanin synthesis in reproductive organs. *Plant Sci.* **2011**, *181*, 219–229. [[CrossRef](#)]
56. Dong, X.; Chen, W.; Wang, W.; Zhang, H.; Liu, X.; Luo, J. Comprehensive profiling and natural variation of flavonoids in rice. *J. Integr. Plant Biol.* **2014**, *56*, 876–886. [[CrossRef](#)]
57. Zhang, F.; Huang, J.; Guo, H.; Yang, C.; Li, Y.; Shen, S.; Zhan, C.; Qu, L.; Liu, X.; Wang, S.; et al. *OsRLCK160* contributes to flavonoid accumulation and UV-B tolerance by regulating *OsbZIP48* in rice. *Sci. China Life Sci.* **2022**, *65*, 1380–1394. [[CrossRef](#)]

58. Ma, X.; Li, F.; Zhang, Q.; Wang, X.; Guo, H.; Xie, J.; Zhu, X.; Khan, N.U.; Zhang, Z.; Li, J.; et al. Genetic architecture to cause dynamic change in tiller and panicle numbers revealed by genome-wide association study and transcriptome profile in rice. *Plant J.* **2020**, *104*, 1603–1616. [[CrossRef](#)] [[PubMed](#)]
59. Zhao, Y.; Zhao, W.; Jiang, C.; Wang, X.; Xiong, H.; Todorovska, E.G.; Yin, Z.; Chen, Y.; Wang, X.; Xie, J.; et al. Genetic Architecture and Candidate Genes for Deep-Sowing Tolerance in Rice Revealed by Non-syn GWAS. *Front. Plant Sci.* **2018**, *9*, 332. [[CrossRef](#)]
60. Chorev, M.; Carmel, L. The Function of Introns. *Front. Genet.* **2012**, *3*, 55. [[CrossRef](#)]
61. Jo, B.-S.; Choi, S.S. Introns: The Functional Benefits of Introns in Genomes. *Genom. Inform.* **2015**, *13*, 112–118. [[CrossRef](#)] [[PubMed](#)]
62. Mayr, C. What Are 3' UTRs Doing? *Cold Spring Harb. Perspect. Biol.* **2018**, *11*, a034728. [[CrossRef](#)] [[PubMed](#)]
63. Zhang, Y.-C.; Gong, S.-F.; Li, Q.-H.; Sang, Y.; Yang, H.-Q. Functional and signaling mechanism analysis of rice CRYPTOCHROME 1. *Plant J.* **2006**, *46*, 971–983. [[CrossRef](#)] [[PubMed](#)]
64. Zhang, Y.-C.; He, R.-R.; Lian, J.-P.; Zhou, Y.-F.; Zhang, F.; Li, Q.-F.; Yu, Y.; Feng, Y.-Z.; Yang, Y.-W.; Lei, M.-Q.; et al. OsmiR528 regulates rice-pollen intine formation by targeting an uclacyanin to influence flavonoid metabolism. *Proc. Natl. Acad. Sci. USA* **2019**, *117*, 727–732. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.