

Article

A Comprehensive Genome-Wide Investigation of the Cytochrome 71 (*OsCYP71*) Gene Family: Revealing the Impact of Promoter and Gene Variants (Ser33Leu) of *OsCYP71P6* on Yield-Related Traits in Indica Rice (*Oryza sativa* L.)

Bijayalaxmi Sahoo ^{1,2}, Itishree Nayak ^{1,3}, C. Parameswaran ^{1,*}, Mahipal Singh Kesawat ^{4,*}, Khirod Kumar Sahoo ², H. N. Subudhi ¹, Cayalvizhi Balasubramaniasai ¹, S. R. Prabhukarthikeyan ⁵, Jawahar Lal Katara ¹, Sushanta Kumar Dash ¹, Sang-Min Chung ⁶, Manzer H. Siddiqui ⁷, Saud Alamri ⁷ and Sanghamitra Samantaray ¹

- ¹ Crop Improvement Division, ICAR-National Rice Research Institute, Cuttack 753006, India; bijayalaxmisahoo7@gmail.com (B.S.); itishreenayak2616@gmail.com (I.N.); dr_hatanath_subudhi@yahoo.co.in (H.N.S.); cayalshiv@gmail.com (C.B.); jawaharbt@gmail.com (J.L.K.); skdash139@gmail.com (S.K.D.); smitraray@gmail.com (S.S.)
- ² Department of Botany, Ravenshaw University, Cuttack 753006, India; khirod555@gmail.com
- ³ Department of Botany, Utkal University, Bhubaneswar 751004, India
- ⁴ Department of Genetics and Plant Breeding, Faculty of Agriculture, Sri University, Cuttack 754006, India
- ⁵ Crop Protection Division, ICAR-National Rice Research Institute, Cuttack 753006, India; prabhukarthipat@gmail.com
- ⁶ Department of Life Science, Dongguk University-Seoul, Ilsandong-gu, Goyang-si 10326, Gyeonggi-do, Republic of Korea; smchung@dongguk.edu
- ⁷ Department of Botany and Microbiology, College of Science, King Saud University, Riyadh 11451, Saudi Arabia; mhsiddiqui@ksu.edu.sa (M.H.S.); saualamir@ksu.edu.sa (S.A.)
- * Correspondence: agriparames07@gmail.com (C.P.); mahibiotech@snu.ac.kr (M.S.K.)



Citation: Sahoo, B.; Nayak, I.; Parameswaran, C.; Kesawat, M.S.; Sahoo, K.K.; Subudhi, H.N.; Balasubramaniasai, C.; Prabhukarthikeyan, S.R.; Katara, J.L.; Dash, S.K.; et al. A Comprehensive Genome-Wide Investigation of the Cytochrome 71 (*OsCYP71*) Gene Family: Revealing the Impact of Promoter and Gene Variants (Ser33Leu) of *OsCYP71P6* on Yield-Related Traits in Indica Rice (*Oryza sativa* L.). *Plants* **2023**, *12*, 3035. <https://doi.org/10.3390/plants12173035>

Academic Editors: Wenqiang Li and Haifeng Li

Received: 13 July 2023

Revised: 17 August 2023

Accepted: 21 August 2023

Published: 23 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/>).

Abstract: The cytochrome P450 (CYP450) gene family plays a critical role in plant growth and developmental processes, nutrition, and detoxification of xenobiotics in plants. In the present research, a comprehensive set of 105 *OsCYP71* family genes was pinpointed within the genome of indica rice. These genes were categorized into twelve distinct subfamilies, where members within the same subgroup exhibited comparable gene structures and conserved motifs. In addition, 105 *OsCYP71* genes were distributed across 11 chromosomes, and 36 pairs of *OsCYP71* involved in gene duplication events. Within the promoter region of *OsCYP71*, there exists an extensive array of cis-elements that are associated with light responsiveness, hormonal regulation, and stress-related signaling. Further, transcriptome profiling revealed that a majority of the genes exhibited responsiveness to hormones and were activated across diverse tissues and developmental stages in rice. The *OsCYP71P6* gene is involved in insect resistance, senescence, and yield-related traits in rice. Hence, understanding the association between *OsCYP71P6* genetic variants and yield-related traits in rice varieties could provide novel insights for rice improvement. Through the utilization of linear regression models, a total of eight promoters were identified, and a specific gene variant (Ser33Leu) within *OsCYP71P6* was found to be linked to spikelet fertility. Additionally, different alleles of the *OsCYP71P6* gene identified through in/dels polymorphism in 131 rice varieties were validated for their allelic effects on yield-related traits. Furthermore, the single-plant yield, spikelet number, panicle length, panicle weight, and unfilled grain per panicle for the *OsCYP71P6*-1 promoter insertion variant were found to contribute 20.19%, 13.65%, 5.637%, 8.79%, and 36.86% more than the deletion variant, respectively. These findings establish a robust groundwork for delving deeper into the functions of *OsCYP71*-family genes across a range of biological processes. Moreover, these findings provide evidence that allelic variation in the promoter and amino acid substitution of Ser33Leu in the *OsCYP71P6* gene could potentially impact traits related to rice yield. Therefore, the identified promoter variants in the *OsCYP71P6* gene could be harnessed to amplify rice yields.

Keywords: allelic variants; cytochrome P450s; expression patterns; hormones; in/del markers; grain number; grain yield and plant development

1. Introduction

The cytochrome P450 (CYP450) gene family plays a critical role in plant growth and developmental processes, nutrition, and detoxification of xenobiotics in plants [1–7]. In addition, CYP450 has been implicated in diverse metabolic reactions and targets a wide range of biological molecules. These biosynthetic reactions lead to diverse plant hormones, fatty acid conjugates, lignin, secondary metabolites, and numerous defensive compounds [8,9]. For instance, CYP706B1 and CYP82D113 participate in the biosynthesis of gossypol, while CYP71AV1 is implicated in the biosynthesis of artemisinin in plants [1,10]. The CYP450 genes have been identified and characterized in various organisms such as animals, bacteria, humans, and plants [1–5,11–15]. The plant CYP450 gene family is categorized into two primary groups: the A-type and the non-A type [16,17]. Further, several research groups reclassified them into eleven classes based on the available numerous genome sequences. The A-type is grouped into the CYP71 class, while the non-A type is subdivided into ten classes (CYP51, CYP72, CYP74, CYP85, CYP86, CYP97, CYP710, CYP711, CYP727, and CYP746) following a unified nomenclature system [17–20]. The members of CYP73, CYP84, and CYP98 gene families play a role in the phenylpropanoid biosynthetic pathway, contributing to the synthesis of various phenolic compounds. These compounds include substances such as flavonoids, suberin, polyphenols, and lignin [1–5]. Further, plant CYP450 enzymes have been linked to a wide array of metabolic pathways, giving rise to both primary and secondary metabolites [20]. For example, the majority of genes associated with the biosynthesis pathways of flavonoids and isoflavonoids are members of the CYP450 gene family. Notable examples include CYP73, CYP75A, and CYP93C [21]. Remarkably, the diversity in the P450 gene family had a significant impact on the emergence of novel metabolic pathways during the evolution of land plants. A classic example is cytochrome P450 93C, which is identified in leguminous crops and enzymes belonging to the CYP93C subfamily and participates in the legume-specific isoflavonoid biosynthesis pathway [22–26].

In rice, the gene *OsCYP71P6* (BGIOSGA018523) is a homolog of *CYP71A1* (LOC_Os12g16720). This gene encodes a CYP450 monooxygenase in rice that exhibits tryptamine 5-hydroxylase activity. Its function involves catalyzing the conversion of tryptamine into serotonin [27,28]. Phyto serotonin (5-hydroxytryptamine) is found in different tissues of plants, is involved in growth and development, and regulates tolerance to different biotic and abiotic stresses [27]. Further, the synthesis of serotonin through the amino acid tryptophan is highly conserved and distinct within plants. Consequently, alterations in the serotonin biosynthesis pathway have a direct impact on various plant developmental processes. A loss-of-function mutation of *OsCYP71P6*, involved in the activity of the tryptamine-5-hydroxylase gene and serotonin synthesis, resulted in brown plant hopper (BPH) and stem borer resistance in rice by altering pest feeding behavior. However, yield traits were also affected by the loss-of-function mutation of *OsCYP71P6* in rice [28]. Rice holds a prominent position as a staple food in countries across South Asia [29,30]. The popularity of rice varieties among farmers is greatly determined by their panicle architecture, including primary and secondary branches, the count of spikelets per panicle, stable yield, and milling properties [31]. In rice, several genes have been identified and characterized, such as *Gn1*, *DEP1*, *GS3*, *SPL18*, and *IPA1*, which regulate panicle architecture, grain number, and yield [32]. Additionally, many of the genes associated with yield also play a role in regulating other crucial traits in rice. For example, elevated expression of the *IPA1* gene, which encodes *OsSPL14* in young panicles, correlates with an increased number of primary branches [33] and blast disease resistance [34]. Similarly, the *Dense and Erect Panicle 1* (*DEP1*) gene regulates nitrogen use efficiency and yield [35]. Although the *OsCYP71P6* gene provides tolerance to insect pests, reduced serotonin differentially regulates senescence and negatively affects rice yield [28]. Gene sequence polymorphisms are common occurrences in both coding and promoter regions, leading to variations in gene transcription.

These variations play a role in phenotypic adaptation among diverse rice varieties [36,37]. The availability of the rice 3K panel significantly aids in detecting single nucleotide polymorphisms and insertion–deletion (in/dels) within chromosomal regions or genes [38]. Previously, an insertion present in the 3'UTR of the TPP7 (*trehalose phosphate phosphatase 7*) gene was found to be associated with enhanced anaerobic response in rice [39].

Following recent advancements in DNA-sequencing technology, there has been a surge in the number of sequenced plant genomes. This has facilitated genome-wide analysis and characterization of the CYP450 gene family in various crop species, including *Arabidopsis thaliana* [17,40], *Glycin max* [7], *Morus notabilis* [10], *Nicotiana tabacum* [41], and *Linum usitatissimum* [42]. The CYP450 gene family is extensive, and only a limited number of its members have undergone comprehensive structural and functional characterization. Additionally, there remains a scarcity of information regarding the CYP71 gene family in rice. The evolutionary lineage, gene and motif arrangement, cis-regulatory element involvement, and transcription kinetics of the CYP71 gene family in rice are currently unexplored. A comprehensive genome-wide investigation and analysis spanning various plants from different evolutionary branches is imperative. Such information could potentially enrich our comprehension of the evolutionary lineage and biological roles of the CYP71 gene family in the plant kingdom. The complete genome sequence of indica rice (*Oryza sativa* Indica) is publicly accessible, enabling us to conduct a comprehensive genome-wide analysis of the CYP71 gene family in rice [43]. Rice holds significance as both a vital cereal crop and a model plant in scientific research [29,30,44]. Hence, in this work, we conducted a comprehensive genome-wide analysis of the CYP71 gene family in rice, employing a range of computational tools. Further, we investigated the physical and biochemical characteristics, gene structure, motif composition, chromosomal distribution, gene duplication events, 3D structural aspects, and expression patterns of members of the *OsCYP71* gene family across different tissues and developmental stages. Our hypothesis centers on the notion that the insertion and deletion (in/dels) variations within the *OsCYP71P6* gene in rice could provide insights into the gene's impact on yield-related traits. Our objective is to pinpoint specific in/dels variations unique to *OsCYP71P6* and validate these sequence polymorphisms to ascertain their effects on yield-related traits in rice. Furthermore, through linear regression models, we identified eight promoters and a gene variant (Ser33Leu) of *OsCYP71P6* that exhibited associations with spikelet fertility. Additionally, we validated various alleles of *OsCYP71P6*, identified via in/dels polymorphism in 131 rice varieties, to determine their effects on yield-related traits. The identified promoter and gene variants within *OsCYP71P6* hold potential for enhancing rice yield. Collectively, this study offers a valuable resource for further comprehending the specific functions of *OsCYP71* gene family members in rice.

2. Results

2.1. Genome-Wide Identification and Phylogenetic Tree Analysis of *OsCYP* Genes in Rice

In this study, a total of 105 CYP71 genes were successfully identified in rice (*Oryza sativa* Indica). These 105 *OsCYP71* genes were found to encode proteins ranging in length from 321 to 1165 amino acids long (Tables 1 and S1).

Table 1. The identified putative cytochrome P450 (*OsCYP71*) genes in rice and their biophysical characteristics.

Proposed Gene Name	Gene ID	Chromosome	Genomic Location	Orientation	CDS Length (bp)	Protein Length (aa)	Molecular Weight (KDa)	Isoelectric Point (pI)	GRAVY	Predicted Subcellular Localization
<i>OsCYP71K1</i>	BGIOGA001610	1	1:17410528-17412154	Reverse	1557	518	57.41	7.73	0.0144	endomembrane system
<i>OsCYP71C1</i>	BGIOGA003715	1	1:22500639-22503588	Forward	1413	470	53	5.85	−0.172	plasma membrane
<i>OsCYP71U1</i>	BGIOGA004274	1	1:32296506-32298186	Forward	1584	527	58.27	8.09	−0.092	endomembrane system
<i>OsCYP71AA1</i>	BGIOGA005209	1	1:46261115-46263199	Forward	1572	523	58.12	8.31	0.007	endomembrane system
<i>OsCYP71AA2</i>	BGIOGA005210	1	1:46266756-46268515	Forward	1605	534	59.45	8.35	−0.029	endomembrane system
<i>OsCYP71T1</i>	BGIOGA003049	1	1:7516842-7519131	Forward	1689	562	60.51	5.64	0.051	endomembrane system
<i>OsCYP71T2</i>	BGIOGA003050	1	1:7524389-7527482	Forward	1692	563	60.53	6.62	0.019	endomembrane system
<i>OsCYP71T3</i>	BGIOGA003054	1	1:7558245-7560317	Forward	1614	537	59.18	7.42	−0.018	endomembrane system
<i>OsCYP71U4</i>	BGIOGA007962	2	2:11083073-11087397	Forward	1503	500	54.15	6.47	−0.027	endomembrane system
<i>OsCYP71AB6</i>	BGIOGA008255	2	2:18910371-18913246	Forward	1518	505	56.15	8.88	−0.098	endomembrane system
<i>OsCYP71W1</i>	BGIOGA008261	2	2:19021940-19025270	Forward	1614	537	60.1	8.70	−0.063	endomembrane system
<i>OsCYP71U5</i>	BGIOGA008265	2	2:19083433-19088060	Forward	1590	529	59.11	8.17	−0.146	endomembrane system
<i>OsCYP71U6</i>	BGIOGA008266	2	2:19089067-19092362	Forward	1578	525	58.86	8.43	−0.168	endomembrane system
<i>OsCYP71Z4</i>	BGIOGA006339	2	2:20885281-20889418	Reverse	1542	513	56.9	7.98	−0.02	plasma membrane

Table 1. Cont.

Proposed Gene Name	Gene ID	Chromosome	Genomic Location	Orientation	CDS Length (bp)	Protein Length (aa)	Molecular Weight (KDa)	Isoelectric Point (pI)	GRAVY	Predicted Subcellular Localization
<i>OsCYP71T4</i>	BGIOGA008463	2	2:23403404-23404909	Forward	1506	501	55.31	8.00	0.032	endomembrane system
<i>OsCYP71Z5</i>	BGIOGA008468	2	2:23502957-23505833	Forward	1566	521	57.34	8.22	−0.004	endomembrane system
<i>OsCYP71Z6</i>	BGIOGA006215	2	2:23525585-23528377	Reverse	1557	518	57.46	7.51	−0.028	endomembrane system
<i>OsCYP71T5</i>	BGIOGA006210	2	2:23642185-23643714	Reverse	1530	509	56.07	7.03	−0.009	endomembrane system
<i>OsCYP71X1</i>	BGIOGA007002	2	2:5495789-5497484	Reverse	1560	519	57.57	7.63	−0.07	endomembrane system
<i>OsCYP71K2</i>	BGIOGA007683	2	2:5517065-5518619	Forward	966	321	36.35	5.45	−0.089	endomembrane system
<i>OsCYP71X2</i>	BGIOGA007686	2	2:5560619-5562432	Forward	1566	521	57.65	6.82	−0.031	endomembrane system
<i>OsCYP71X3</i>	BGIOGA007688	2	2:5569750-5571671	Forward	1548	515	58.05	7.74	−0.131	endomembrane system
<i>OsCYP71X4</i>	BGIOGA007691	2	2:5604868-5606669	Forward	1566	521	57.49	6.78	−0.042	endomembrane system
<i>OsCYP71K3</i>	BGIOGA007694	2	2:5642839-5644448	Forward	1530	509	56.32	8.60	0.057	endomembrane system
<i>OsCYP71AC1</i>	BGIOGA007695	2	2:5644996-5646510	Forward	1422	473	52.17	8.18	−0.055	endomembrane system
<i>OsCYP71K4</i>	BGIOGA007696	2	2:5647245-5648907	Forward	1572	523	57.78	7.54	−0.052	endomembrane system
<i>OsCYP71V1</i>	BGIOGA007791	2	2:7497274-7499270	Forward	1536	511	56.6	8.06	0.006	plasma membrane
<i>OsCYP71V2</i>	BGIOGA007792	2	2:7500786-7503251	Forward	1527	508	56.49	7.97	0.057	endomembrane system

Table 1. Cont.

Proposed Gene Name	Gene ID	Chromosome	Genomic Location	Orientation	CDS Length (bp)	Protein Length (aa)	Molecular Weight (KDa)	Isoelectric Point (pI)	GRAVY	Predicted Subcellular Localization
<i>OsCYP71P4</i>	BGIOGA011508	3	3:2163933-2165882	Reverse	1644	547	59.2	6.95	0.047	endomembrane system
<i>OsCYP71T6</i>	BGIOGA010447	3	3:21911994-21913514	Reverse	1521	506	55.75	6.45	0.009	endomembrane system
<i>OsCYP71E2</i>	BGIOGA012992	3	3:23090126-23097523	Forward	1554	517	57.42	7.58	−0.409	plasma membrane
<i>OsCYP71W2</i>	BGIOGA013057	3	3:25210900-25212580	Forward	1536	511	57.5	7.69	−0.096	endomembrane system
<i>OsCYP71W3</i>	BGIOGA013059	3	3:25254727-25257057	Forward	1614	537	60.76	7.42	−0.094	endomembrane system
<i>OsCYP71W4</i>	BGIOGA013063	3	3:25338238-25340280	Forward	1542	513	57.48	7.67	−0.063	endomembrane system
<i>OsCYP71AB7</i>	BGIOGA010140	3	3:29073720-29075333	Reverse	1614	537	57.39	7.98	0.06	endomembrane system
<i>OsCYP71V3</i>	BGIOGA013536	3	3:34645395-34647054	Forward	1506	501	55.22	9.30	0.034	endomembrane system
<i>OsCYP71U7</i>	BGIOGA013602	3	3:35634216-35637043	Forward	1425	474	52.31	7.62	−0.047	endomembrane system
<i>OsCYP71U8</i>	BGIOGA013604	3	3:35641266-35643533	Forward	1542	513	56.61	7.74	−0.012	endomembrane system
<i>OsCYP71U9</i>	BGIOGA013605	3	3:35649465-35651890	Forward	1584	527	58.39	7.46	−0.07	endomembrane system
<i>OsCYP71U10</i>	BGIOGA013606	3	3:35654380-35656113	Forward	1557	518	57.89	8.64	−0.033	endomembrane system
<i>OsCYP71E3</i>	BGIOGA013948	3	3:40194199-40200533	Forward	1584	527	58.42	9.44	−0.028	endomembrane system
<i>OsCYP71AB8</i>	BGIOGA011112	3	3:8385202-8386836	Reverse	1506	501	55.39	8.28	−0.033	endomembrane system

Table 1. Cont.

Proposed Gene Name	Gene ID	Chromosome	Genomic Location	Orientation	CDS Length (bp)	Protein Length (aa)	Molecular Weight (KDa)	Isoelectric Point (pI)	GRAVY	Predicted Subcellular Localization
<i>OsCYP71AB9</i>	BGIOGA011111	3	3:8394147-8396376	Reverse	1503	500	56.22	8.12	−0.15	endomembrane system
<i>OsCYP71AB10</i>	BGIOGA011586	3	3:954097-955621	Reverse	1446	481	52.38	7.52	0.105	endomembrane system
<i>OsCYP71Z7</i>	BGIOGA016146	4	4:13300800-13303842	Forward	1536	511	56.74	8.24	0.0001	endomembrane system
<i>OsCYP71S2</i>	BGIOGA014867	4	4:22441383-22445458	Reverse	2301	766	84.89	8.42	−0.324	endomembrane system
<i>OsCYP71S3</i>	BGIOGA014866	4	4:22450395-22452270	Reverse	1536	511	55.88	8.52	−0.074	endomembrane system
<i>OsCYP71Z8</i>	BGIOGA015504	4	4:6445565-6447178	Reverse	1524	507	56.65	6.89	−0.014	endomembrane system
<i>OsCYP71Z9</i>	BGIOGA015981	4	4:6553171-6555010	Forward	1506	501	55.76	6.04	0.011	endomembrane system
<i>OsCYP71S4</i>	BGIOGA015743	4	4:89164-90761	Reverse	1500	499	55.32	7.46	−0.05	endomembrane system
<i>OsCYP71AF2</i>	BGIOGA019698	5	5:18144104-18145713	Forward	1536	511	55.4	9.54	−0.096	endomembrane system
<i>OsCYP71AD</i>	BGIOGA018026	5	5:22038254-22039816	Reverse	1563	520	57.1	6.35	−0.087	endomembrane system
<i>OsCYP71P5</i>	BGIOGA020098	5	5:25741450-25745152	Forward	1539	512	58.36	8.73	−0.237	endomembrane system
<i>OsCYP71R1</i>	BGIOGA020185	5	5:27004831-27006483	Forward	1569	522	56.66	6.64	−0.083	endomembrane system
<i>OsCYP71P6</i>	BGIOGA018523	5	5:8791390-8793045	Reverse	1560	519	57.32	7.62	−0.067	mitochondrial membrane
<i>OsCYP71C2</i>	BGIOGA022809	6	6:13456419-13457960	Forward	1416	471	53.44	8.78	0.002	endomembrane system

Table 1. Cont.

Proposed Gene Name	Gene ID	Chromosome	Genomic Location	Orientation	CDS Length (bp)	Protein Length (aa)	Molecular Weight (KDa)	Isoelectric Point (pI)	GRAVY	Predicted Subcellular Localization
<i>OsCYP71T7</i>	BGIOGA021189	6	6:18739993-18741510	Reverse	1518	505	55.67	7.52	0.049	endomembrane system
<i>OsCYP71S1</i>	BGIOGA022122	6	6:190381-192067	Forward	1557	518	56.41	8.18	0.007	endomembrane system
<i>OsCYP71AB11</i>	BGIOGA020890	6	6:25633293-25634951	Reverse	1584	527	57.95	8.12	−0.085	endomembrane system
<i>OsCYP71K5</i>	BGIOGA023340	6	6:27270045-27274299	Forward	1560	519	57.07	8.65	−0.042	endomembrane system
<i>OsCYP71Y1</i>	BGIOGA023341	6	6:27281079-27282759	Forward	1539	512	55.46	7.93	0.089	endomembrane system
<i>OsCYP71Y2</i>	BGIOGA023345	6	6:27304346-27305982	Forward	1569	522	57.41	7.70	−0.039	endomembrane system
<i>OsCYP71Y3</i>	BGIOGA023346	6	6:27308296-27309910	Forward	1545	514	56.39	7.86	−0.008	endomembrane system
<i>OsCYP71K6</i>	BGIOGA023350	6	6:27362563-27367339	Forward	1503	500	56.9	9.67	−0.053	endomembrane system
<i>OsCYP71AC2</i>	BGIOGA023351	6	6:27374980-27377638	Forward	1632	543	60.63	7.55	−0.133	endomembrane system
<i>OsCYP71AF1</i>	BGIOGA020805	6	6:27378645-27380259	Reverse	1515	504	55.24	6.38	−0.001	endomembrane system
<i>OsCYP71X5</i>	BGIOGA020704	6	6:29562437-29565945	Reverse	1641	546	60.56	8.19	−0.125	endomembrane system
<i>OsCYP71Q1</i>	BGIOGA024456	7	7:11074316-11076970	Reverse	1113	370	42.15	5.61	−0.053	endomembrane system
<i>OsCYP71AB12</i>	BGIOGA027933	8	8:1799581-1805006	Forward	3498	1165	130.4	7.42	−0.086	mitochondrial membrane
<i>OsCYP71W5</i>	BGIOGA026880	8	8:23757559-23759121	Reverse	1563	520	58.35	7.73	−0.038	endomembrane system

Table 1. Cont.

Proposed Gene Name	Gene ID	Chromosome	Genomic Location	Orientation	CDS Length (bp)	Protein Length (aa)	Molecular Weight (KDa)	Isoelectric Point (pI)	GRAVY	Predicted Subcellular Localization
<i>OsCYP71T8</i>	BGIOGA028809	8	8:24254355-24255857	Forward	1503	500	55.99	8.22	0.037	endomembrane system
<i>OsCYP71T9</i>	BGIOGA026711	8	8:26774318-26775838	Reverse	1521	506	56.35	7.19	−0.012	endomembrane system
<i>OsCYP71U11</i>	BGIOGA026530	8	8:29302265-29303901	Reverse	1554	517	56.33	6.26	−0.008	endomembrane system
<i>OsCYP71C3</i>	BGIOGA027793	8	8:327857-329461	Reverse	1605	534	59.94	6.63	−0.003	plasma membrane
<i>OsCYP71C4</i>	BGIOGA027791	8	8:335792-337318	Reverse	1527	508	57.55	6.70	−0.012	endomembrane system
<i>OsCYP71C5</i>	BGIOGA027790	8	8:341090-342922	Reverse	1650	549	62.01	6.04	−0.192	endomembrane system
<i>OsCYP71C6</i>	BGIOGA027840	8	8:347325-348858	Forward	1389	462	52.15	7.86	−0.084	endomembrane system
<i>OsCYP71C7</i>	BGIOGA027841	8	8:358473-360517	Forward	1575	524	59.22	8.46	−0.0001	endomembrane system
<i>OsCYP71AB13</i>	BGIOGA027437	8	8:8239548-8245451	Reverse	1575	524	58.18	9.90	−0.082	endomembrane system
<i>OsCYP71W6</i>	BGIOGA030841	9	9:14907658-14909929	Forward	1545	514	58.11	8.71	−0.125	endomembrane system
<i>OsCYP71W7</i>	BGIOGA030842	9	9:14912617-14916216	Forward	1557	518	58.96	7.81	−0.107	endomembrane system
<i>OsCYP71W8</i>	BGIOGA029682	9	9:14918713-14922927	Reverse	1569	522	59.65	8.50	−0.208	endomembrane system
<i>OsCYP71E4</i>	BGIOGA029664	9	9:15258631-15261118	Reverse	1533	510	55.49	7.01	0.003	endomembrane system
<i>OsCYP71T10</i>	BGIOGA029663	9	9:15266334-15268285	Reverse	1518	505	54.62	7.39	0.031	endomembrane system

Table 1. Cont.

Proposed Gene Name	Gene ID	Chromosome	Genomic Location	Orientation	CDS Length (bp)	Protein Length (aa)	Molecular Weight (KDa)	Isoelectric Point (pI)	GRAVY	Predicted Subcellular Localization
<i>OsCYP71T11</i>	BGIOGA031134	9	9:19550851-19553047	Forward	1446	481	52.45	6.58	−0.002	endomembrane system
<i>OsCYP71AK</i>	BGIOGA031135	9	9:19556511-19558129	Forward	1521	506	54.63	8.97	−0.012	endomembrane system
<i>OsCYP71C8</i>	BGIOGA030097	9	9:5147576-5149433	Reverse	1530	509	56.53	8.78	−0.032	endomembrane system
<i>OsCYP71Z1</i>	BGIOGA031844	10	10:14240993-14243733	Reverse	1572	523	58.4	7.21	−0.019	endomembrane system
<i>OsCYP71Z2</i>	BGIOGA031843	10	10:14247493-14251099	Reverse	1575	524	58.4	7.80	−0.098	endomembrane system
<i>OsCYP71Z3</i>	BGIOGA031842	10	10:14257972-14262017	Reverse	1560	519	57.9	8.56	−0.119	endomembrane system
<i>OsCYP71AB1</i>	BGIOGA032580	10	10:3806982-3812472	Forward	1506	501	56.15	8.03	−0.045	endomembrane system
<i>OsCYP71AB2</i>	BGIOGA032583	10	10:3912905-3915694	Forward	1485	494	55.26	8.18	−0.047	endomembrane system
<i>OsCYP71AB3</i>	BGIOGA032584	10	10:3947165-3951250	Forward	1497	498	55.47	9.16	−0.038	endomembrane system
<i>OsCYP71AB4</i>	BGIOGA032599	10	10:4268768-4270367	Forward	1512	503	57.9	6.60	−0.168	endomembrane system
<i>OsCYP71AB5</i>	BGIOGA032634	10	10:5163945-5165932	Forward	1509	502	55.78	9.24	0.024	endomembrane system
<i>OsCYP71P1</i>	BGIOGA032653	10	10:5794451-5798343	Forward	1542	513	57.9	6.92	−0.155	endomembrane system
<i>OsCYP71P2</i>	BGIOGA032680	10	10:6875415-6884974	Forward	1608	535	58.49	6.56	0.02	endomembrane system
<i>OsCYP71P3</i>	BGIOGA032683	10	10:7021038-7022711	Forward	1581	526	57.52	7.67	−0.072	endomembrane system

Table 1. Cont.

Proposed Gene Name	Gene ID	Chromosome	Genomic Location	Orientation	CDS Length (bp)	Protein Length (aa)	Molecular Weight (KDa)	Isoelectric Point (pI)	GRAVY	Predicted Subcellular Localization
<i>OsCYP71E1</i>	BGIOSGA036106	12	12:16043912-16046285	Reverse	1569	522	58.42	8.83	−0.166	endomembrane system
<i>OsCYP71U2</i>	BGIOSGA035935	12	12:19695567-19697582	Reverse	1557	518	55.44	6.75	0.149	endomembrane system
<i>OsCYP71U3</i>	BGIOSGA035934	12	12:19701585-19703297	Reverse	1713	570	60.72	7.15	0.197	endomembrane system
<i>OsCYP71V4</i>	BGIOSGA037944	Scaffold	AAAA020356 82.1:5479-7227	Forward	1542	513	57.3	7.36	−0.012	endomembrane system
<i>OsCYP71W9</i>	BGIOSGA038041	Scaffold	AAAA020360 20.1:14706-17551	Forward	1530	509	57.07	8.21	−0.305	endomembrane system
<i>OsCYP71U12</i>	BGIOSGA038175	Scaffold	AAAA020367 41.1:10396-12075	Reverse	1566	521	57.9	8.24	−0.071	endomembrane system
<i>OsCYP71AB14</i>	BGIOSGA038318	Scaffold	AAAA0203760 2.1:3877-5590	Forward	1509	502	56.24	9.26	−0.25	endomembrane system

ID: identity; bp: base pair; aa: amino acids; pI: isoelectric point; MW: molecular weight; GRAVY: grand average of hydropathy; KDa: Kilo dalton.

The molecular weight of the encoded protein is between 36.35 kda (OsCYP71K2) and 130.4 kda (OsCYP71AB12). The isoelectric point sizes ranged from 5.45 (OsCYP71K2) to 9.90 (OsCYP71AB13); among them, 82 proteins encoding OsCYP71 were identified by $PI > 7$, suggesting that most OsCYP proteins have a high proportion of basic amino acid. We observed GRAVY values for 105 OsCYP71 proteins between -0.409 (OsCYP71E2) and 0.197 (OsCYP71U3). Further, subcellular localization projections indicated that 100 OsCYP71 proteins were localized in the endomembrane system. The rest are mapped to plasma membrane (OsCYP71C1, OsCYP71Z4, OsCYP71V1, OsCYP71E2, and OsCYP71C3) and mitochondrial membrane (OsCYP71P6). In addition, we also examined the molecular weight of OsCYP71 to the pI to determine the distribution of OsCYP71 proteins (Figure S1). The outcomes of this analysis unveiled a pattern where the majority of OsCYP71 proteins share a comparable molecular weight and isoelectric point. To evaluate the evolutionary connections among *OsCYP71* genes, we employed CYP71 genes from *Arabidopsis* and tomato. Using MEGA 11 software, we constructed a phylogenetic tree that encompassed 105 CYP71 proteins from rice along with 42 CYP71 proteins from *Arabidopsis* and 43 CYP71 proteins from tomato (Figure 1, Table S2).

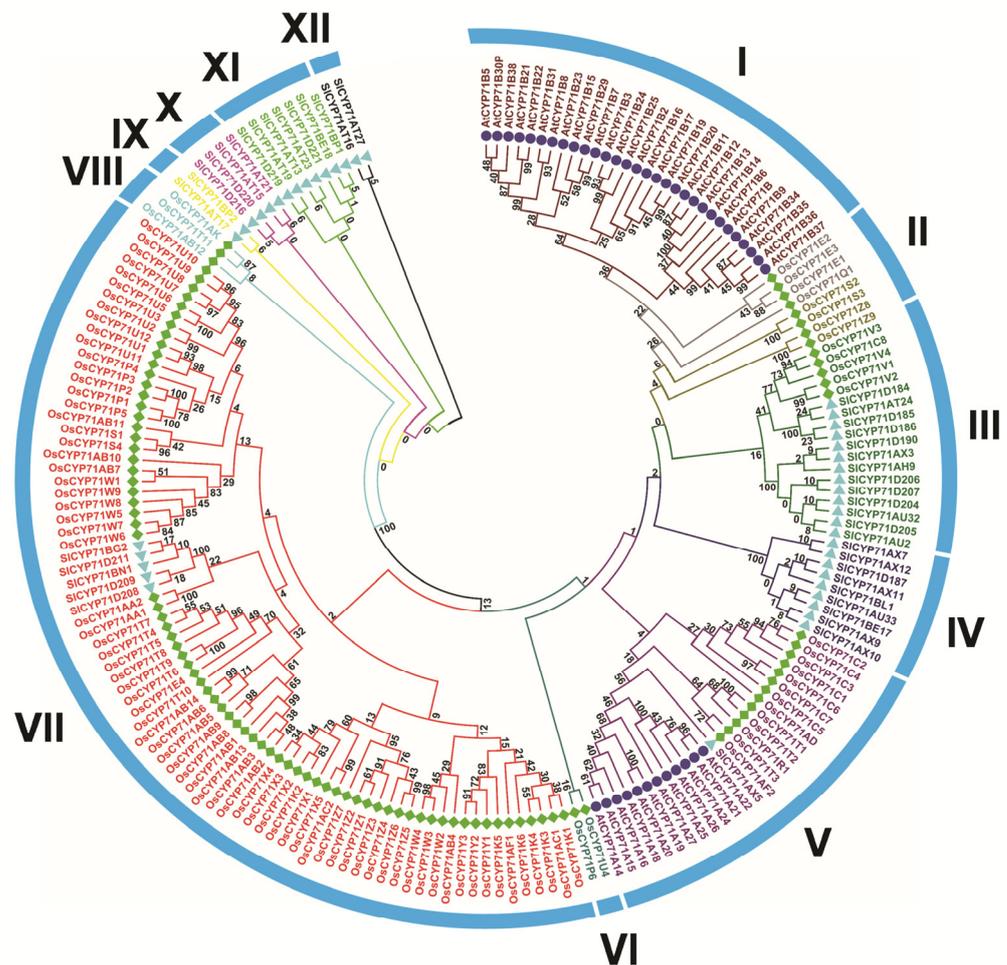


Figure 1. The phylogenetic tree of CYP proteins was established for *A. thaliana*, *O. sativa*, and *S. lycopersicum* using the neighbor-joining method in MEGA 11. To gauge the tree's reliability, 1000 bootstrap replicates were employed. *OsCYP71* genes could be divided into 12 subfamilies (I–XII).

The results showed that 105 *OsCYP71* genes could be divided into 12 subfamilies. Group VII had 74 members, whereas Groups I, II, III, IV, V, VI, VIII, IX, X, XI, and XII had 0, 8, 5, 0, 13, 2, 3, 0, 0, 0, and 0 members, respectively (Figure S2).

2.2. Chromosome Distribution and Gene Duplication Analysis of OsCYP71 Genes

To further investigate gene duplication events in rice, we performed a chromosomal mapping analysis of identified *OsCYP71* family genes. A total of 105 *OsCYP71* genes were mapped onto 11 chromosomes (Figure 2A). Chromosome 2 had the most *OsCYP71* genes (20 genes) and chromosome 7 had the least *OsCYP71* genes (1 gene). The 105 *OsCYP71* genes were distributed unevenly on 12 chromosomes of rice (Figure 2B).

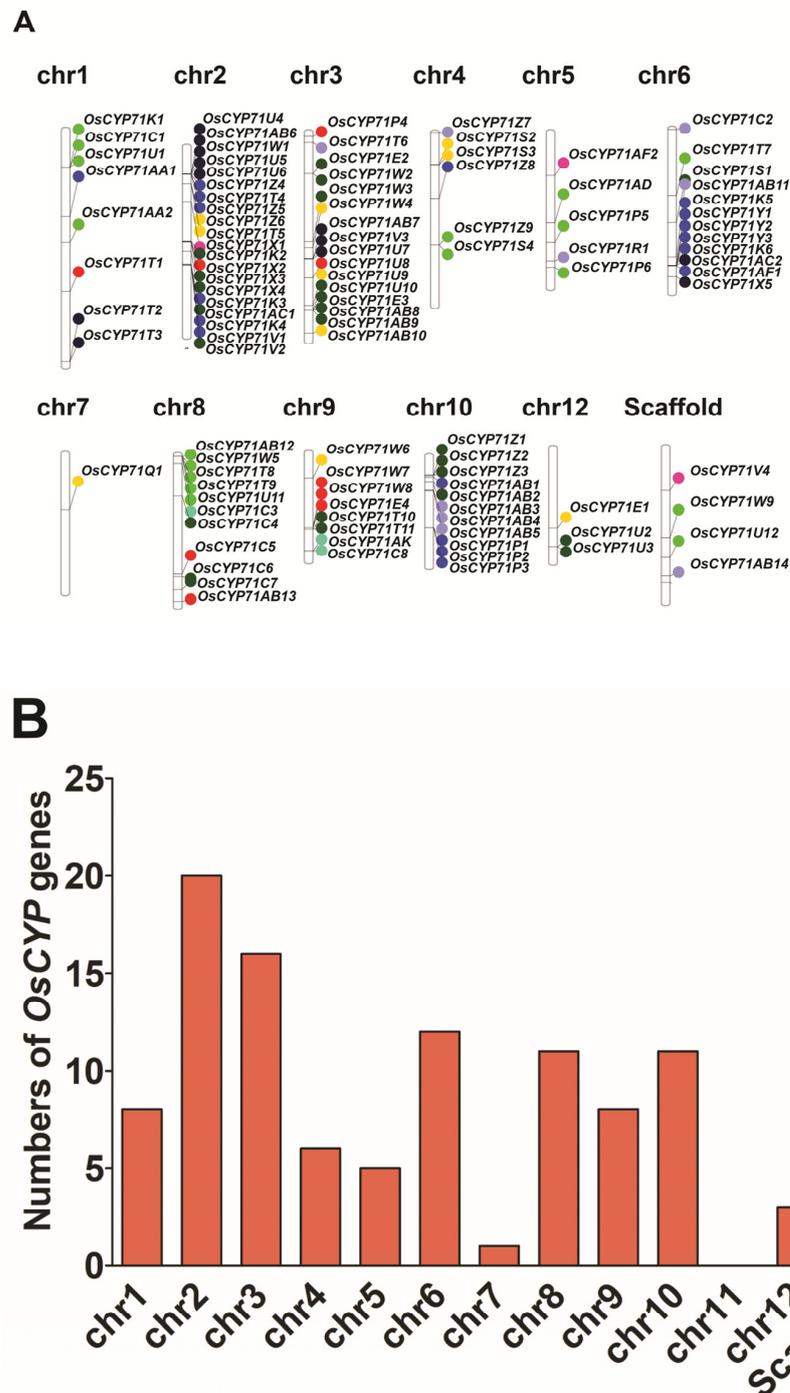


Figure 2. Chromosome distribution of identified *OsCYP71* genes. (A) Schematic illustrations of the chromosomal distribution of CYP genes on the twelve chromosomes of rice with the gene positions on each chromosome represented by a line on the right side. (B) Dispersal of CYP71 genes across eleven chromosomes of rice.

The four *OsCYP71* genes were distributed across a genomic scaffold (Table 1). Gene duplication plays a crucial role in the evolution of gene families. In order to study the evolutionary relationship of *OsCYP71* genes, the gene duplication events of *OsCYP71* genes were analyzed in this study (Figures 3 and S3).

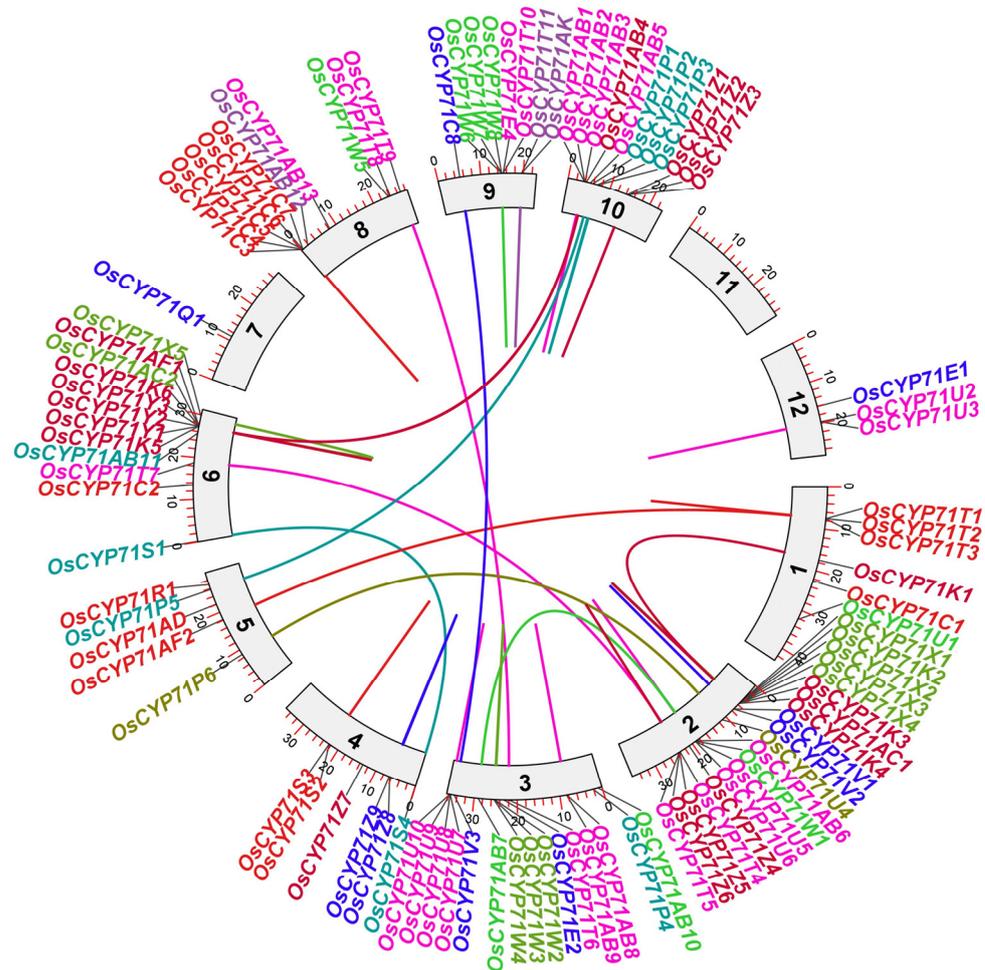


Figure 3. Chromosomal distribution and duplicated CYP gene pairs in rice. Duplicated CYP gene pairs are connected with distinct colors of lines. The figure was constructed using TB tools with numbers 1–12 denoting the rice chromosomes.

A total of 36 duplicated gene pairs were detected (Table S3). Further, the K_a/K_s score for these genes was less than one, indicating that there had been a strong purifying selection with slight changes in subsequent gene duplication. These findings highlight the fact that the *OsCYP71* family genes have undergone a conserved evolutionary process.

2.3. Gene Structure and Basic Motif Analysis of *OsCYP71* Gene

Analyzing gene structure provides valuable insights into the conserved features and evolutionary distinctions of CYP proteins in rice. Through this analysis, we observed that the number of introns varied across different subfamilies (Figure 4).

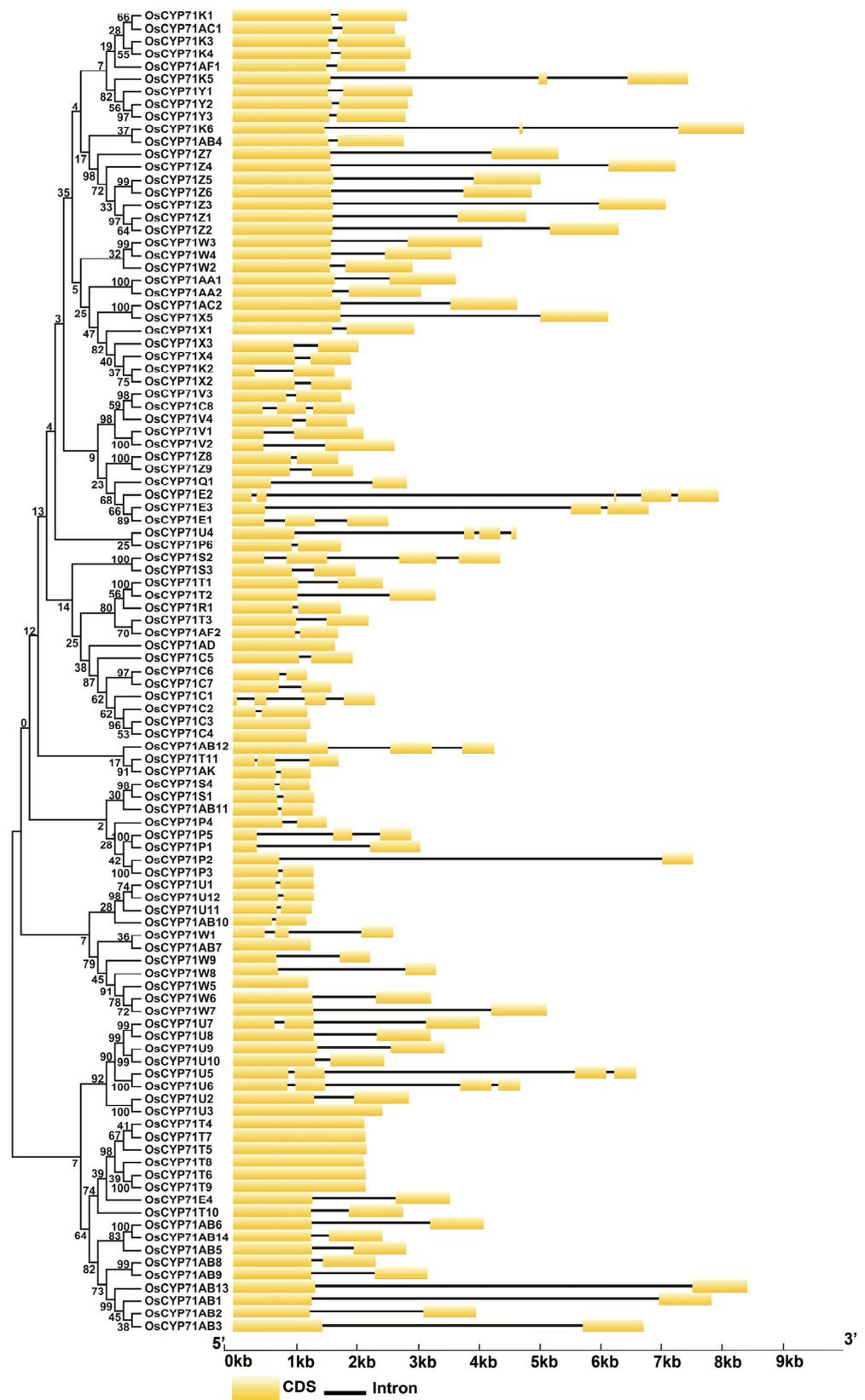


Figure 4. The intron-exon structure of the *OsCYP71* genes. The yellow boxes signify exons while black lines denote introns. The lengths of the boxes and lines are scaled based on gene length.

The number of introns of all CYP subfamily genes ranged from zero to four, whereas most of the genes have between one and two introns. The maximum four introns were

detected in *OsCYP71E2*, while *OsCYP71AD*, *OsCYP71C3*, *OsCYP71C4*, *OsCYP71AB7*, *OsCYP71W5*, *OsCYP71U3*, *OsCYP71T4*, *OsCYP71T7*, *OsCYP71T5*, *OsCYP71T8*, *OsCYP71T6*, and *OsCYP71T9* were intronless (Figure 4). Further, to investigate the characteristics of *OsCYP71* proteins, we identified ten conserved motifs in *OsCYP71* proteins (Figure 5A,B).

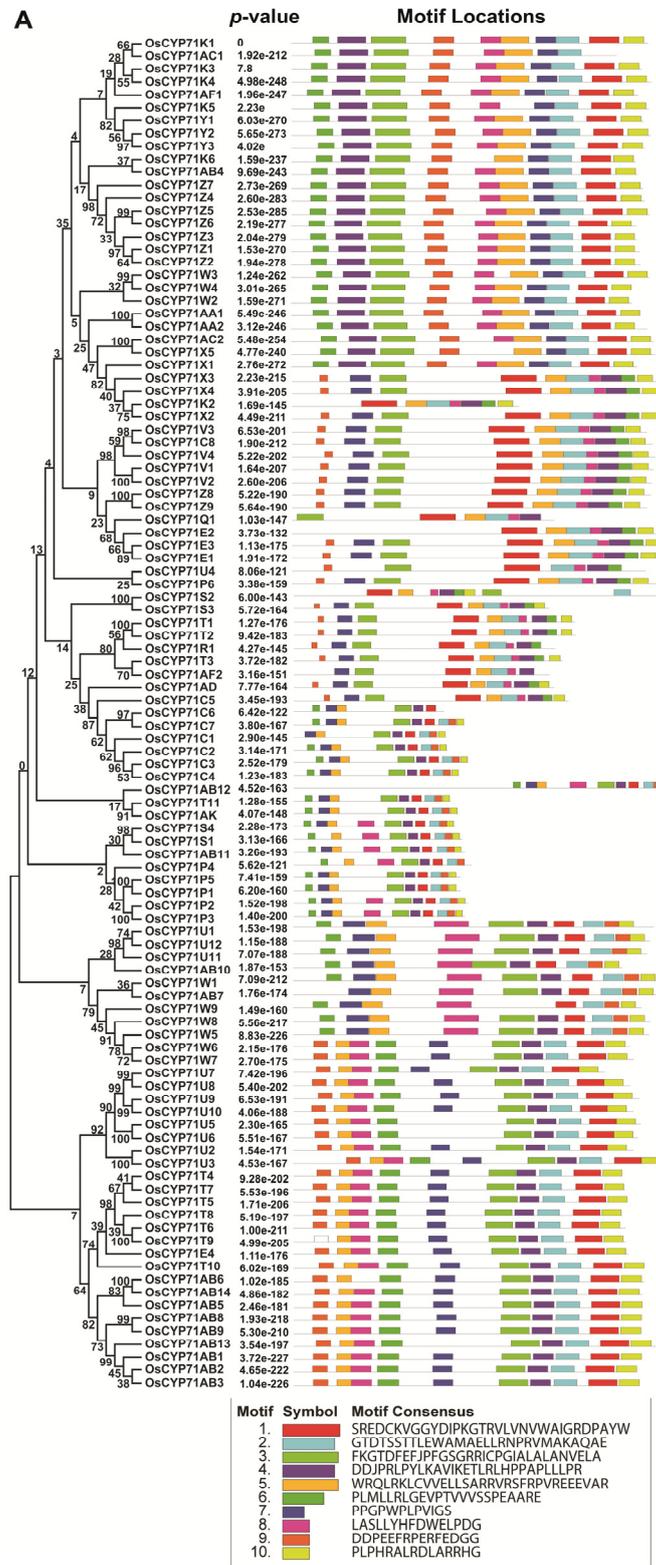


Figure 5. Cont.

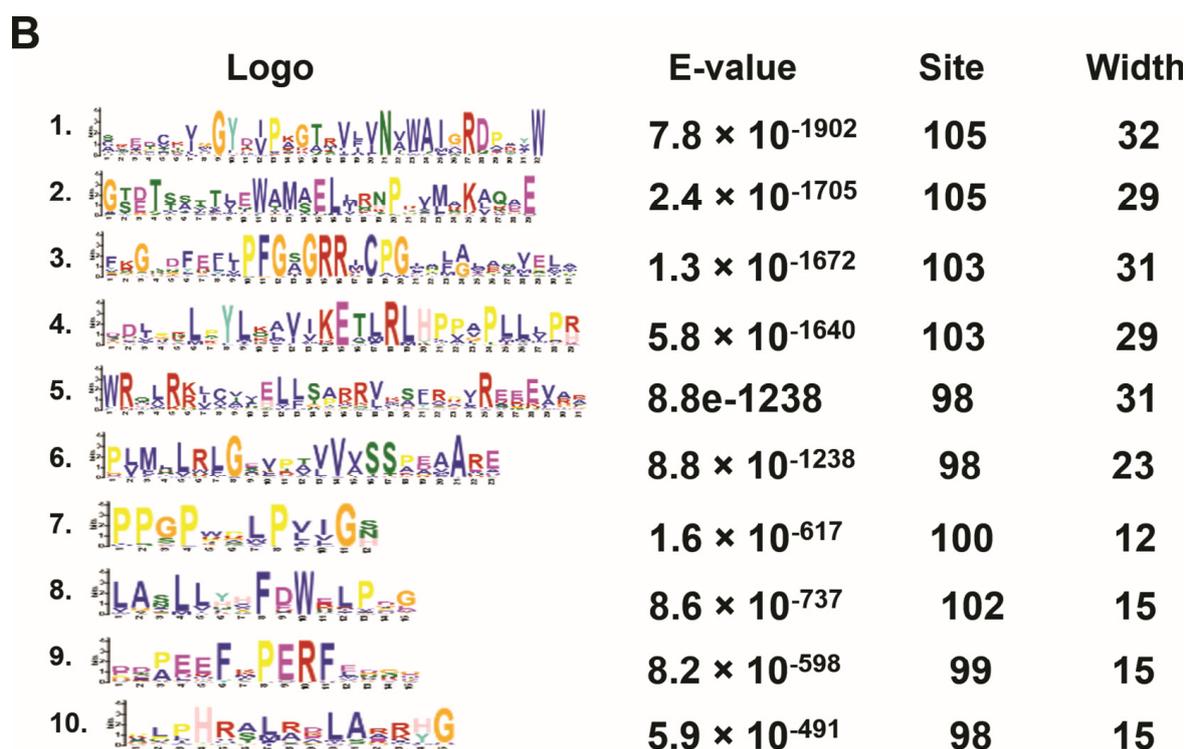


Figure 5. The conserved motifs in *OsCYP71* genes. The conserved motifs were elucidated by the MEME database. (A) The distinct colored boxes represent different conserved motifs with variable sizes and sequences. (B) Sequence logo of the conserved motif of *OsCYP71*.

In this study, it was observed that motifs 1, 2, 3, 6, 7, and 10 exhibited a high degree of conservation across the majority of *OsCYP71* proteins. Additionally, through amino acid sequence alignment, it was noted that *OsCYP71* proteins contain conserved SRS and heme-binding motifs (Figure S4A). We also conducted predictions for the 3D structure of *OsCYP71* proteins (Figure S4B). Numerous conserved motifs identified within these *OsCYP71* family proteins are likely involved in the regulation of diverse metabolic reactions.

2.4. Promoter Element and GO Enrichment Analysis

To gain insights into the potential functions of the *OsCYP71* family genes, we employed the Plant-CARE online web server to analyze the 1500 bp upstream sequences of these genes. This analysis revealed the presence of numerous cis-regulatory elements, encompassing phytohormone responsiveness, light response, cell cycle regulation, circadian rhythm, seed-specific regulation, and responses to defense and stress (Figure 6A,B).

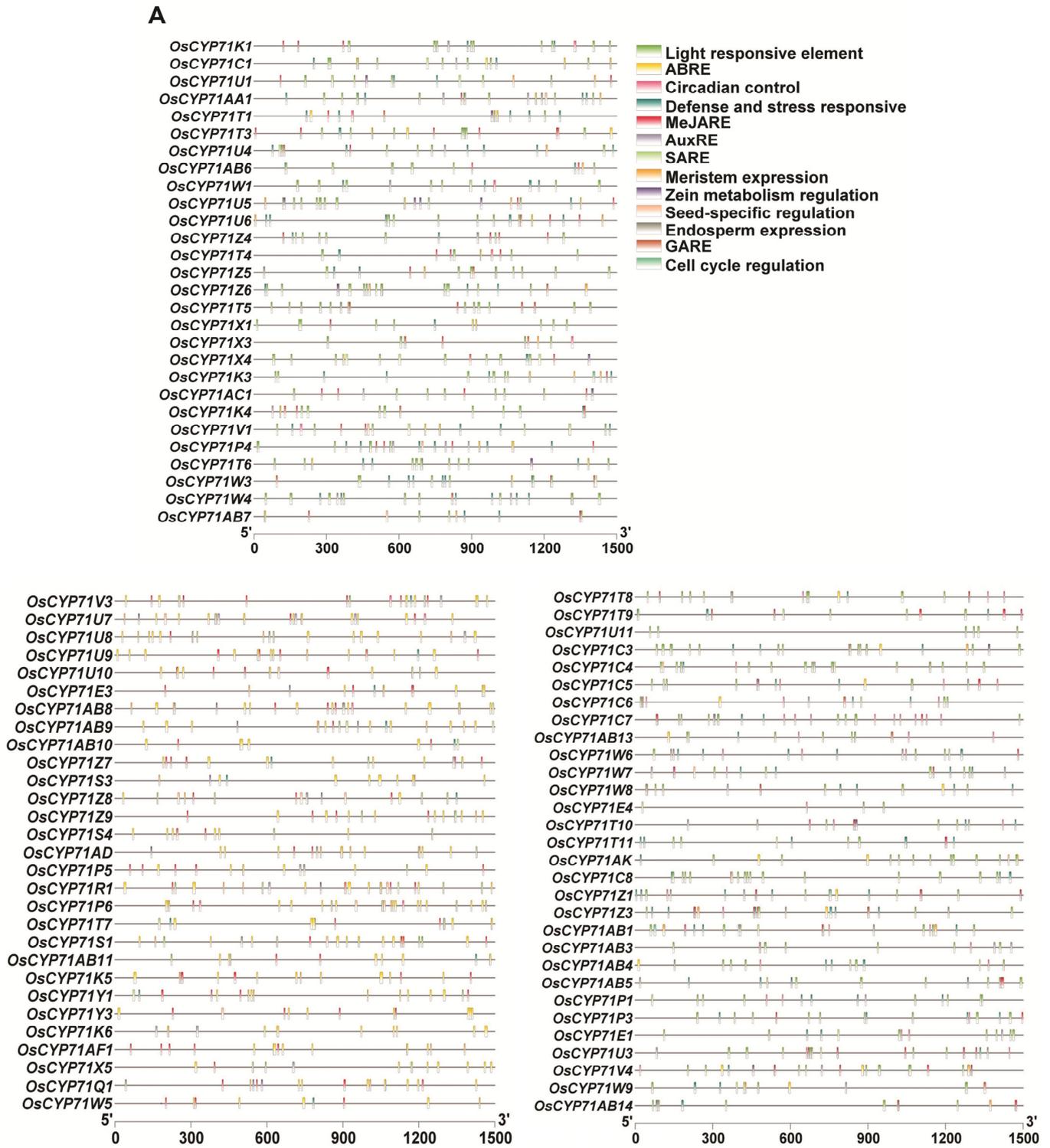


Figure 6. Cont.

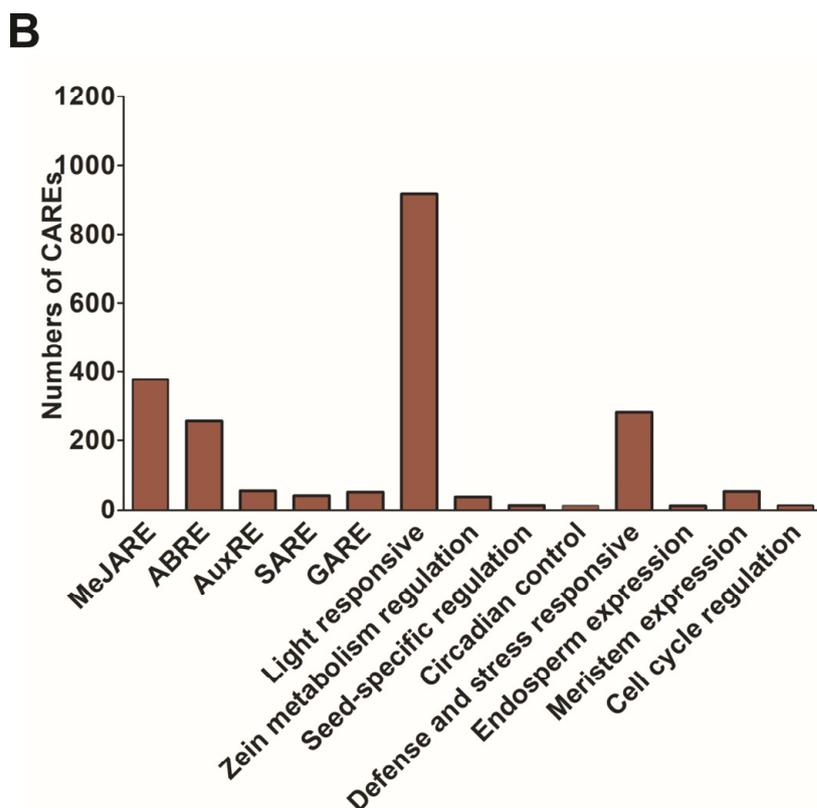


Figure 6. Identification of cis-regulatory elements in the 1500 bp promoter region of the *OsCYP71* gene family. (A) Phytohormone-responsive elements, light-responsive elements, growth- and development-related elements, stress-responsive elements and other elements with unknown functions are shown by distinct colors. (B) The numbers of different CAREs found in *OsCYP71* gene family members.

In terms of hormonal response, there are abscisic acid-responsive elements (ABRE), auxin-responsive elements (TGA-element, AuxRE, TGA-box), gibberellin-responsive elements (Tatc-box, GARE-motif, P-box, TATC-box), salicylic acid-responsive elements (TCA-element, SARE), methyl jasmonate-responsive elements (CGTCA-motif, TGACG-motif), and so on (Table S4). MeJARE, ABRE, and AuxRE were found in the promoter regions of most genes, with MeJARE and ABRE being found in almost all *OsCYP71* genes. SARE- and GARE-motif elements were found in a few promoter regions of *OsCYP71* genes. Stress response includes elements related to anaerobic induction (ARE), low temperature response (LTR), drought induction (MBS), defense and stress responses (TC-rich repeats), and low temperature and salt stress (DRE). Among them, the ARE element was found in the promoter regions of 40 genes, the LTR element was found in the promoter regions of 38 genes with 1–2 elements, and TC-rich repeat elements were found in 21 gene promoter regions with 1–2 elements. In addition, MBS elements, considered drought induction-related elements, were also present in the promoter regions of 43 genes (Table S4). These results suggest that the *OsCYP71* family genes may play an important role in plant development and stress responses through the regulation of multiple cis-regulatory elements in rice.

In order to further understand the function of *OsCYP71* family genes, we conducted a gene ontology (GO) enrichment analysis. Utilizing AgriGO, we effectively annotated and attributed GO terms to all *OsCYP71* family genes. This annotation was subsequently validated using eggNOG-Mapper (Figure S5; Tables S5 and S6), which produced almost the same results as AgriGO. *OsCYP* genes showed enrichment in metabolic processes (GO:0008152) and oxidation-reduction processes (GO:0055114) in the biological process category (Figure S5A), while in the molecular category, *OsCYP*s showed enrichment in catalytic activity (GO:0003824), binding (GO:0005488), electron carrier activity (GO:0009055), oxidoreductase activity (GO:0016491), monooxygenase activity (GO:0004497), tetrapyrrole

binding (GO:0046906), and heme binding (GO:0020037) (Figure S5B). Interestingly, no GO enrichments were identified within the cellular category. As a result, these outcomes strongly suggest that the OsCYP71 family genes are pivotal in various metabolic processes in rice.

2.5. Transcriptome Profiling of OsCYP71 Family Genes in Different Tissues and Phytohormone Treatments in Rice

To gain insights into the functions of OsCYP71 genes, we examined their expression patterns across various tissues, developmental stages, and under hormone treatments. We sourced OsCYP71 gene family expression data from the Rice Expression Profile Database (RiceXPro), which we then employed to construct heatmaps. Remarkably, among the 105 OsCYP71 family genes, we identified differential expression across different tissues, developmental stages, and in response to hormone treatments (Figure 7A–C).

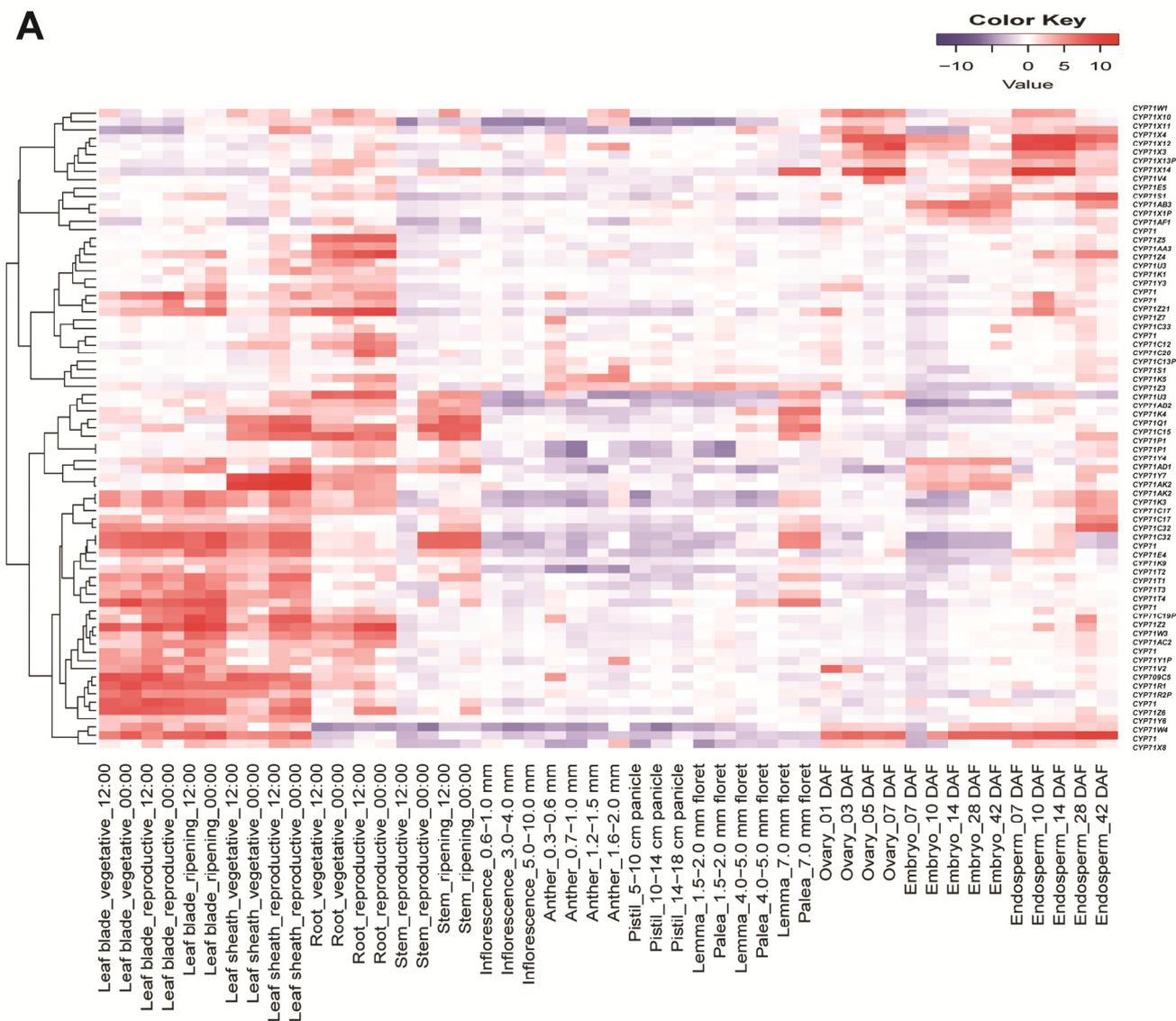


Figure 7. Cont.

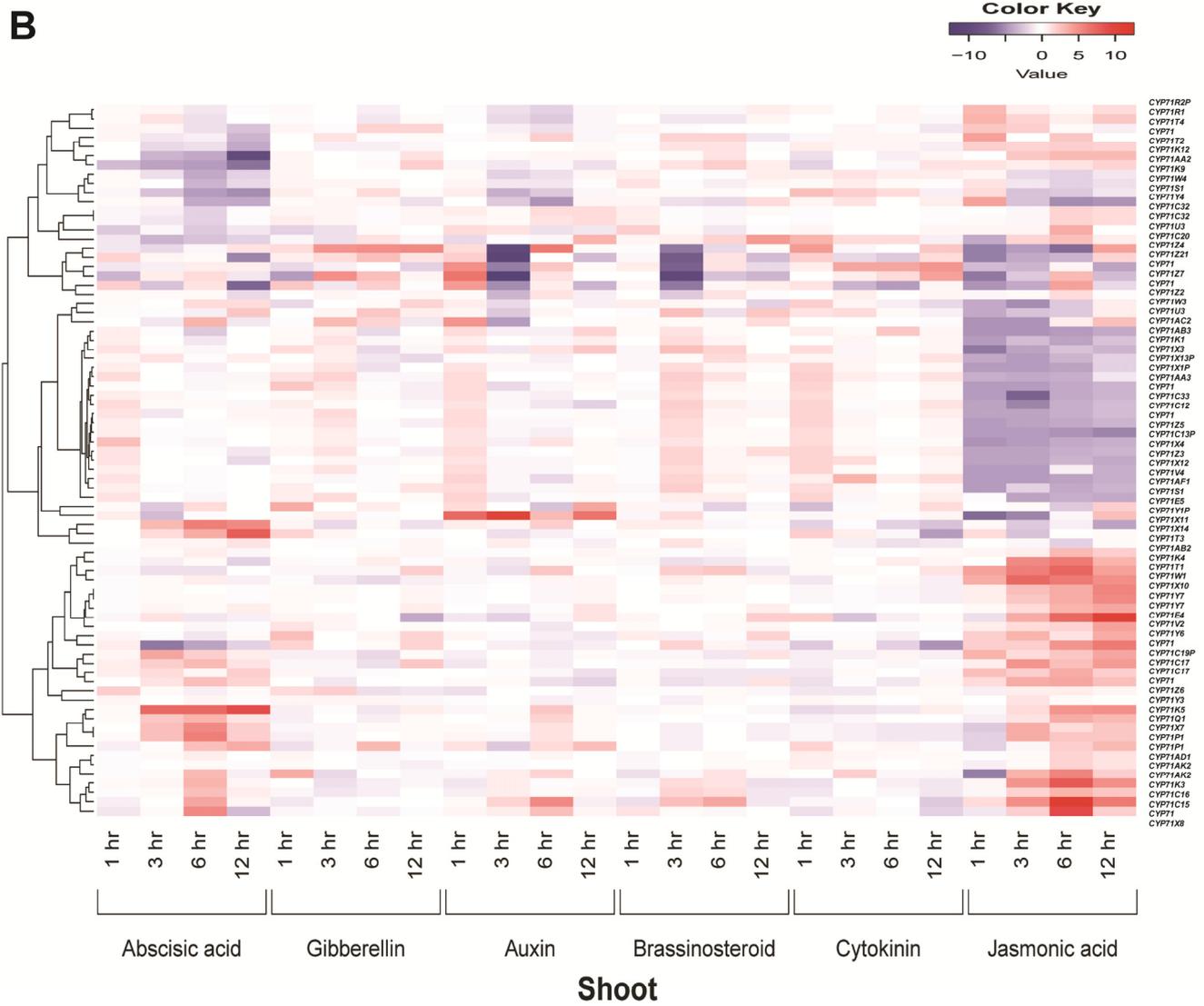


Figure 7. Cont.

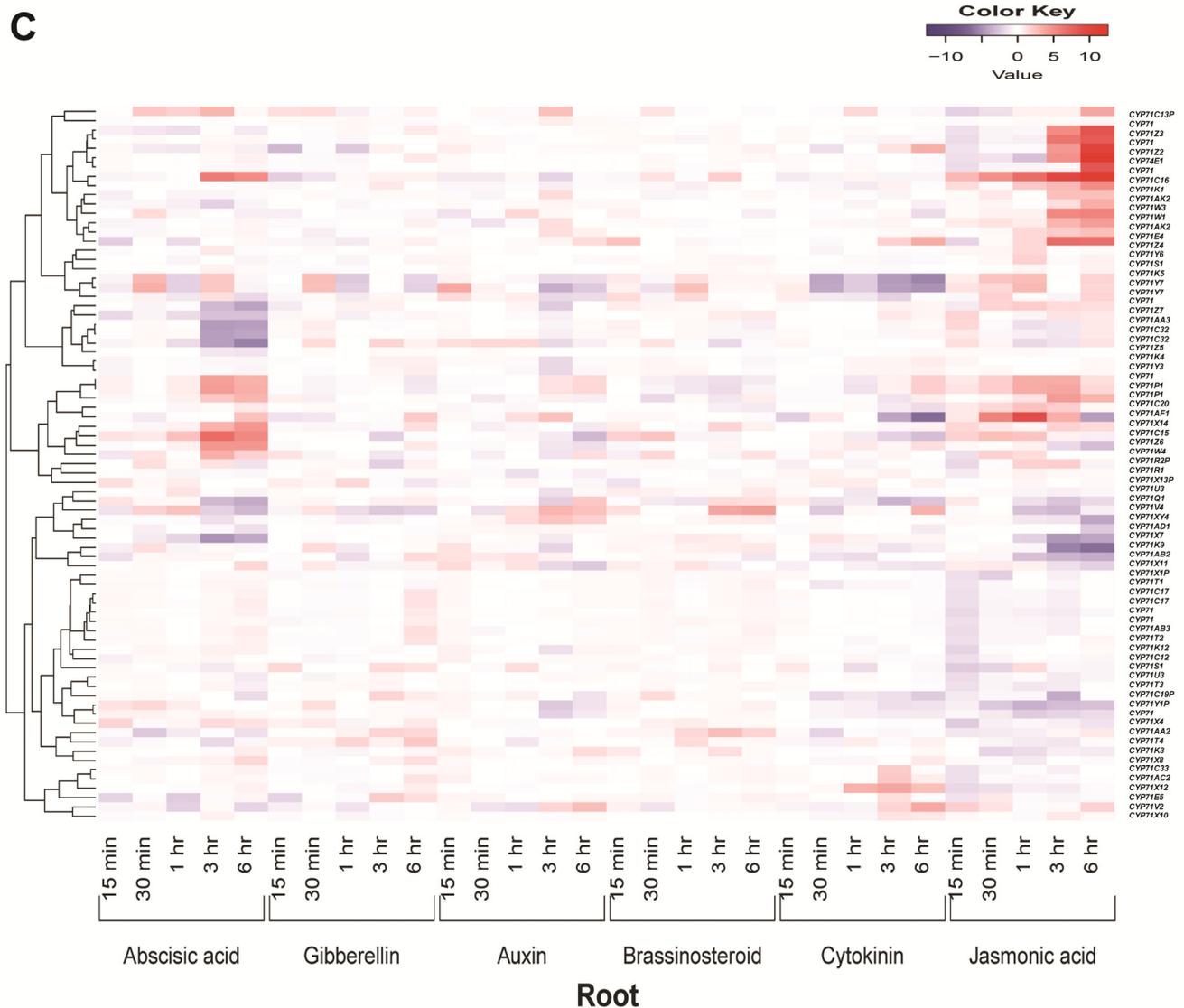


Figure 7. Expression profiles of *OsCYP71* genes in different tissues and hormone treatments. (A) The expression profile in various tissues and developmental stages. (B) The expression profile in shoots under different hormone treatments. (C) The expression profile in roots under different hormone treatments.

For instance, *CYP71X4* and *CYP71X12* expression was elevated in ovary, embryo, and endosperm tissue, while the expression of *CYP71X13P* was induced in lemma, palea, and endosperm tissue. Further, the expression of *CYP71X7* was highly induced in vegetative, reproductive, ripening leaf blade, vegetative, reproductive leaf sheath, ovary, embryo, and endosperm tissue, whereas the expression of *CYP71Z6*, *CYP71R2P*, *CYP71R1*, and *CYP71K12* was highly up-regulated in vegetative, reproductive, ripening leaf, vegetative, and reproductive leaf sheath tissue. *CYP71C19P* and *CYP71Z2* expression were increased in vegetative, reproductive, ripening leaf, vegetative, reproductive leaf sheath, root vegetative, and root reproduction tissue. Additionally, we also examined the expression pattern of *OsCYP71* in the shoots and roots under different plant hormone treatments such as auxin, gibberellin (GA), abscisic acid (ABA), cytokinin, jasmonic acid (JA), and brassinosteroid (BRS) (Figure 7B,C). *CYP71Q1* expression was elevated in the shoot in the ABA and JA treatments after 3 h, 6 h, and 12 h, while the expression level of *CYP71Z4* was induced in the shoot after 3 h, 6 h, and 12 h of GA treatment. The transcript levels of *CYP71C16*, *CYP71Z6*, *CYP71C19P*, *CYP71V2*, *CYP71X10*, *CYP71W1*, and *CYP71T1* were increased in the shoot

after 3 h, 6 h and 12 h of JA treatment. Further, expression of *CYP71X10* increased in the shoot after 3 h, 6 h, and 12 h of BRS treatment, while *CYP71X11* was elevated in the shoot after 1 h, 3 h, 6 h, and 12 h of auxin treatment (Figure 7B). The expression of *CYP71X12* increased in the root after 1 h, 3 h, and 6 h of cytokinin treatment, whereas *CYP71X14*, *CYP71P1*, and *CYP71C16* were significantly induced in the root after JA treatment for 30 min, 1 h, and 3 h. Furthermore, *CYP71Z4*, *CYP71W1*, *CYP71K5*, *CYP71Z2*, *CYP71C20*, and *CYP71Z3* were highly up-regulated in the root after JA treatment for 3 h and 6 h (Figure 7C). Collectively, these results have shown that *OsCYP71* gene family members may participate in different developmental processes and responses to phytohormones in rice.

2.6. Gene Diversity Analysis of *OsCYP71P6* Alleles

To examine gene variation, four in/del primers were validated in the one hundred and thirty-one different rice varieties. Out of four primers of the *OsCYP71P6* gene, two primers were polymorphic (*OsCYP71P6-1*; *OsCYP71P6-4*) and the other two were found to be monomorphic in rice varieties. The primer *OsCYP71P6-1* has two different allelic variants, having amplicon lengths of 320 bp and 350 bp (Figure S6). Among the 131 rice varieties examined, 113 displayed an amplicon length of 320 bp, while 18 varieties exhibited an amplicon length of 350 bp (Table S7). Similarly, the *OsCYP71P6-4* gene also showed two different alleles, with 107 rice varieties displaying amplicon lengths of 380 bp and 24 rice varieties displaying lengths of 400 bp. Further, the in/del *OsCYP71P6-1* major allelic frequency was 0.8625, which was relatively higher than that of in/del *OsCYP71P6-4* (0.8167); overall, the mean major allelic frequency for both the markers was 0.8396 (Table 2).

Table 2. Analysis of gene diversity for the two *OsCYP71P6* in/del markers across various rice varieties.

Marker	Major Allele Frequency	No. of Varieties	Allele No	Gene Diversity	Heterozygosity	PIC
CYP71P6-1	0.8626	131	2.0000	0.2370	0	0.2090
CYP71P6-4	0.8168	131	2.0000	0.2993	0	0.2545
Mean	0.8397	131	2.0000	0.2682	0	0.2317

Further, mean gene diversity for both the in/dels was found to be 0.2681, while the gene diversity of *OsCYP71P6-4* was 0.2992 and that of *OsCYP71P6-1* was 0.2370. Additionally, the PIC values of the *OsCYP71P6-1* and *OsCYP71P6-4* markers were 0.2090 and 0.2545, respectively, with a mean PIC value of 0.2317 (Table 2).

2.7. Grouping of Rice Varieties Based on *OsCYP71P6* Alleles

The unweighted pair group method with arithmetic averaging (UPGMA) was employed for distance-based diversity analysis. This analysis revealed the presence of two primary clusters, designated as Cluster A and Cluster B (Figure 8).

Furthermore, Cluster A is subdivided into two branches, namely, A1 and A2, which are further fragmented into several minor clusters. Similarly, Cluster B was also subdivided into two sub clusters (Figure 8). Cluster A encompasses a total of 24 distinct rice varieties, while Cluster B is the largest cluster, comprising 107 different rice varieties. The model-based structure analysis revealed genetic relationships between different rice varieties based on the *OsCYP71P6* alleles. The optimal K value was found to be 2 (K = 2), indicating that the 131 rice varieties were grouped into two subpopulations (Figure 9A–C).

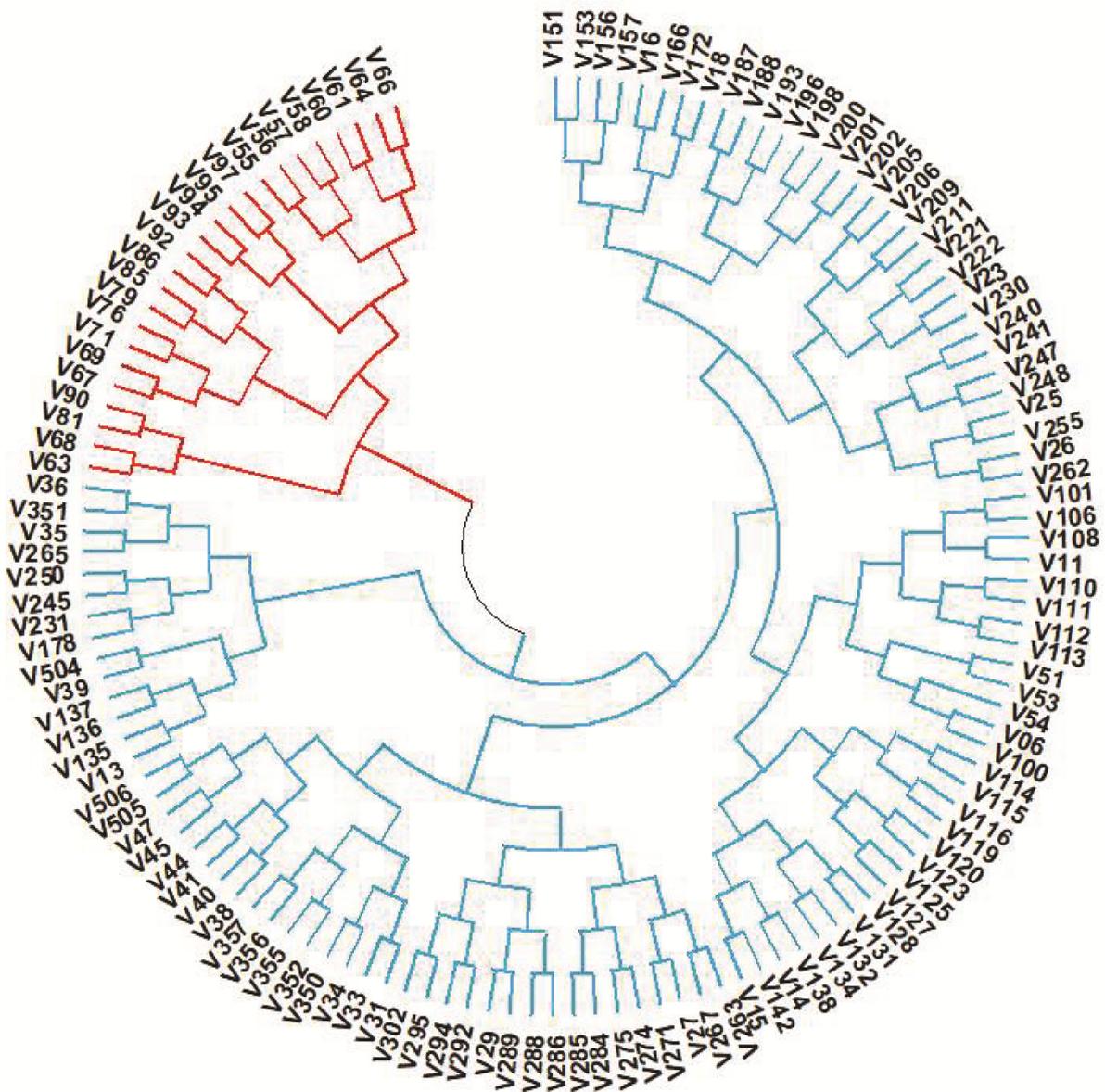


Figure 8. Phylogenetic analysis of rice varieties through the unweighted pair group method with arithmetic averaging (UPGMA) model using in/dels allelic variants of *O5CYP71P6*. Cluster A and Cluster B indicate major clusters formed using distance-based diversity analysis. V—Varieties. The names of the varieties are given in Table S7.

Additionally, within subpopulation I, a total of 29 rice varieties were identified, while subpopulation II encompassed 60 rice varieties. Moreover, 42 rice varieties were classified under the admixture category in the analysis. Notably, the analysis employed a threshold value of 60% (Table S8). AMOVA (analysis of molecular variance), based on F-statistics, revealed that 81% of the genetic variation existed between populations, while the remaining 19% of variation was observed within the population (Table 3).

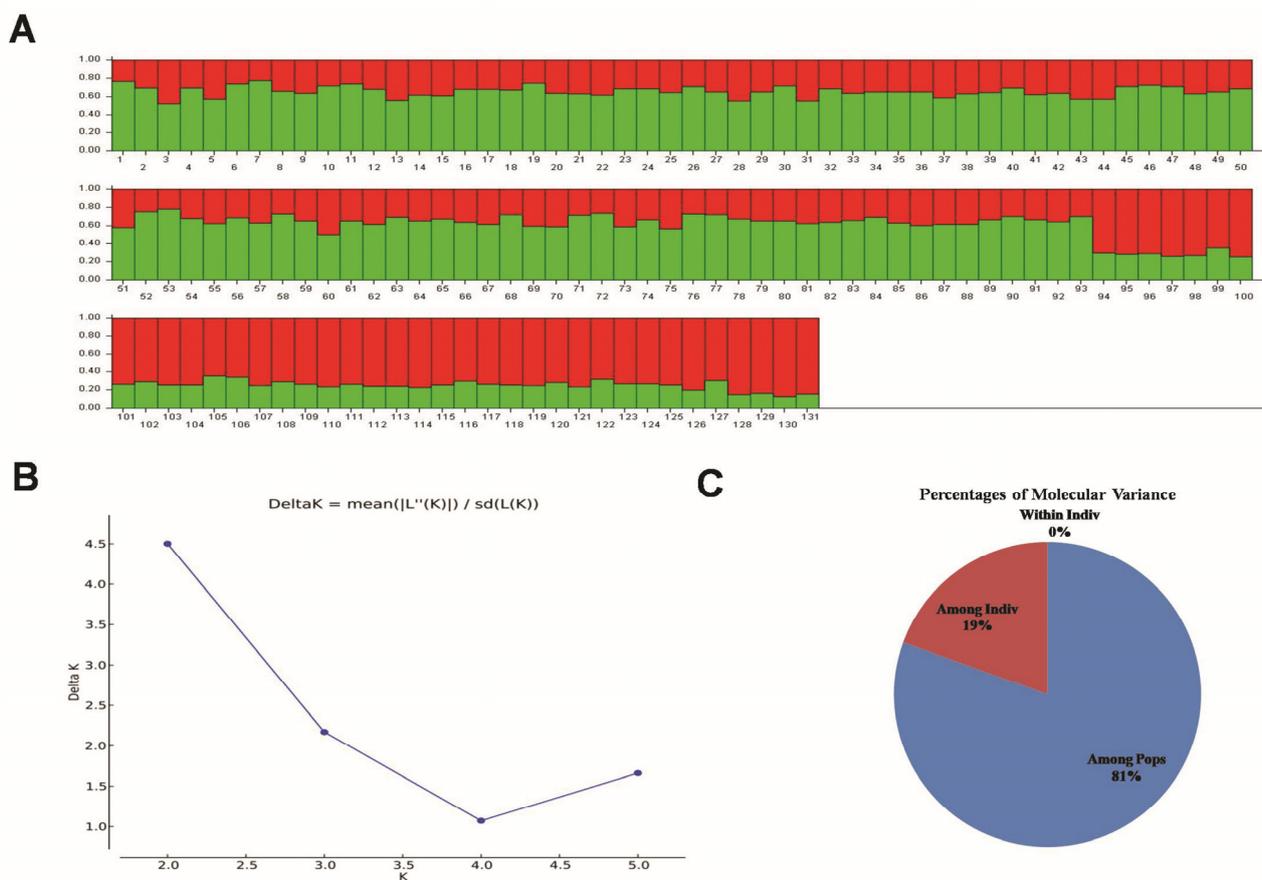


Figure 9. Model-based diversity analysis and analysis of molecular variance (AMOVA) for sub-populations. (A) Sub-structure analysis of rice varieties using the allelic variants of *OsCYP71P6*. (B) Delta K population structure plot for the rice varieties produced through the STRUCTURE harvester program. (C) Pie chart of the percent distribution of the analysis of molecular variance (AMOVA) for the *OsCYP71P6* alleles in rice varieties.

Table 3. Analysis of molecular variance (AMOVA) for two-subpopulations and admixtures of the *OsCYP71P6* in/dels variants in rice varieties.

Source of Variation	df	SS	MS	Est. Var.	Percent Variation (%)
Among Populations	1	39.256	39.256	0.498	81%
Among Individual varieties	129	31.003	0.240	0.120	19%
Within Individual varieties	131	0.000	0.000	0.000	0%
Total	261	70.260		0.618	100%

2.8. Descriptive Statistics of Yield-Related Traits for Different Alleles of *OsCYP71P6*

The descriptive statistics for the seven different traits studied in the one hundred and thirty-one rice varieties are shown in Table 4.

The varieties were grouped based on the alleles of the two polymorphic in/dels of *OsCYP71P6*. The observed trait values spanned ranges for spikelet number (163.99 to 189.93), single-plant yield (30.71 to 38.48 g), panicle weight (3.29 to 3.61 g), number of productive tillers (10.29 to 11.98), panicle length (25.89 to 27.36 cm), unfilled grain (29.2 to 47.65), filled grain (133.91 to 142.61), and 100 seed weight (2.27 to 2.4 g). The kurtosis values for these traits ranged from -0.83 to 7.58 . Notably, among the studied traits, unfilled grain exhibited the highest kurtosis value (7.58). Similarly, the skewness of various traits ranged from -0.78 to 1.85 .

Table 4. Descriptive statistics analysis of in/dels variants of the *OsCYP71P6* gene for yield-related traits in rice varieties.

Trait	Mean				Median				Mode				Kurtosis				Skewness				
	Primer		OsCYP71P6-1		OsCYP71 P6-4		OsCYP71 P6-1		OsCYP71 P6-4		OsCYP71 P6-1		OsCYP71 P6-4		OsCYP71 P6-1		OsCYP71 P6-4		OsCYP71 P6-1		OsCYP71 P6-4
Amplicon length (bp)	320	350	380	400	320	350	380	400	320	350	380	400	320	350	380	400	320	350	380	400	
No. of tillers (Nos.)	10.29	11.98 ^a	10.32	11.41	10	10.33	10	10.66	12	10.33	12	16.66	−0.08	2.19	1.62	−0.68	0.43	1.2	0.82	0.25	
Panicle length (cm)	26.15	26.24	25.89	27.36 ^b	26.2	25.96	25.9	27.21	26.46	—	26.46	—	0.78	−0.91	0.65	−0.08	−0.32	0.33	−0.27	0.41	
Single-plant yield (g)	31.21	37.9 ^a	30.71	38.48 ^b	30.06	34.7	29.8	36.26	33.65	—	33.65	—	0.81	1.11	2.64	−0.83	0.67	1.12	1.03	0.28	
No. of spikelets (Nos.)	168.26	171.81	163.99	189.93 ^b	165.66	176.83	163.33	190.16	170.33	—	170.3	165.66	0.29	−0.02	0.48	0.45	0.27	−0.18	0.27	0.12	
Unfilled grain (Nos.)	33.95	29.2	30.08	47.65 ^b	31	26.83	27.66	44	27.66	—	27.66	44	7.58	1.27	0.01	5.43	1.85	1.08	0.68	1.81	
Filled grain (Nos.)	134.3	142.61	133.91	142.27	134.66	150	134.66	150.5	133	—	133	156	0.5	−0.24	1.02	−0.07	0.39	−0.37	0.48	−0.34	
Panicle weight (g)	3.32	3.52	3.29	3.61	3.23	3.55	3.23	3.67	3.06	—	3.18	2.98	6.04	−0.03	7.09	0.27	1.37	0.29	1.52	0.1	
100 seed weight (g)	2.3	2.27	2.27	2.4	2.35	2.26	2.32	2.39	2.38	2.2	2.38	—	1.28	1.22	1.33	1.41	−0.81	0.15	−0.73	−0.78	

— Indicates not determined, bold values indicate statistically significant mean difference between the different alleles of *CYP71P6*. ^a—Significance at 5% level of significance, ^b—significance at 1% level of significance. The Z test was used for the mean difference analysis.

2.9. *OsCYP71P6* Allelic Difference in Phenotypic Traits

Further, we examined the allelic variation in the *OsCYP71P6* gene and associations with phenotypic traits. Among the two *OsCYP71P6* primers, the alternate alleles of *OsCYP71P6-4* showed statistically significant difference for traits, namely, panicle length, single-plant yield, panicle weight, number of spikelets, and unfilled grain with significant *p* values of 0.002, 0.005, 0.05, 0.004, and 0.001, respectively (Table 5).

Table 5. The population and haplotype mean difference analysis for yield-related traits in *OsCYP71P6* in/dels rice variants.

Primer Name	Traits ^a	Mean ± SD, AL320 ^b	Mean ± SD, AL350 ^b	<i>p</i> Value ^c		
<i>OsCYP71P6-1</i> (In varieties)	NT	10.29 ± 3.09	11.98 ± 4.09	0.04, *		
	SPY	31.21 ± 11.42	37.90 ± 15.38	0.03, *		
<i>OsCYP71P6-4</i> (In varieties)	Traits	Mean ± SD, AL380 ^b	Mean ± SD, AL400 ^b	<i>p</i> value		
	SPY	30.71 ± 11.38	38.48 ± 13.87	0.005, **		
	NS	163.99 ± 54.16	189.93 ± 43.86	0.004, **		
	PW	3.29 ± 0.95	3.61 ± 0.84	0.05, *		
	PL	25.89 ± 2.81	27.36 ± 2.26	0.002, **		
	UG	30.08 ± 15.86	47.65 ± 28.66	0.001, **		
<i>OsCYP71P6-1</i> and <i>OsCYP71P6-4</i> (In two subpopulations)	Traits	Mean ± SD, Sub-Pop1	Mean ± SD, Sub-Pop2	<i>p</i> value		
	SPY	37.06 ± 15.33	31.42 ± 10.48	0.03, *		
	FG	150.93 ± 28.71	135.32 ± 43.63	0.02, *		
<i>OsCYP71P6-1</i> and <i>OsCYP71P6-4</i> (In two subpopulations and admixtures)	Trait	Mean ± SD, Sub-Pop1	Mean ± SD, Sub-Pop1	Mean ± SD, Admix	<i>p</i> value	
	SPY	37.06 ± 15.33	31.42 ± 10.48	29.74 ± 11.37	0.03, *	
	NS	180.12 ± 35.8	172.97 ± 49.09	154.85 ± 42.23	0.04, *	
	FG	150.93 ± 28.71	135.32 ± 43.63	124.92 ± 35.77	0.02, *	
<i>OsCYP71P6-1</i> and <i>OsCYP71P6-4</i> (In four haplotypes)	Traits	Mean ± SD, Hap1	Mean ± SD, Hap2	Mean ± SD, Hap3	Mean ± SD, Hap4	<i>p</i> value
	SPY	29.81 ± 10.24	37.71 ± 14.41	36.64 ± 16.43	42.32 ± 11.73	0.005, **

^a—NT—No. of tillers, SPY—single-plant yield, NS—no. of spikelets, PW—panicle weight, PL—panicle length, UG—no. of unfilled grains, FG—no. of filled grains, ^b—AL—amplicon length of the in/del primer, ^c—Z test *p* value for mean difference. Double asterisk indicates 1% level of significance and single asterisk indicates significance at 5% level.

In terms of percent difference in single-plant yield, number of spikelets, panicle length, panicle weight, and unfilled grain for *OsCYP71P6-4*, the 400 bp allele in/del primer was found to contribute 20.19%, 13.65%, 5.37%, 8.79%, and 36.86% more, respectively, than the 380 bp allele in the studied rice varieties. The average single-plant yield was found to be 30.71 ± 11.38 g and 38.48 ± 13.87 g for alleles of 380 bp and 400 bp lengths, respectively, with a *p* value of 0.005. Similarly, the mean number of spikelets was found to be 163.99 ± 54.16 and 189.93 ± 43.86 for 380 bp and 400 bp alleles, respectively, and difference was found to be significant (*p* value: 0.004). The panicle weight trait also showed statistically significant differences between the alleles (*p* value: 0.05), with the mean values of 3.29 ± 0.95 and 3.61 ± 0.84 for 380 bp and 400 bp alleles, respectively. Likewise, mean panicle length between the alleles was 25.89 ± 2.81 and 27.36 ± 2.26 for 380 bp and 400 bp alleles, respectively, with a *p* value of 0.002. The 380 bp and 400 bp alleles showed a mean unfilled grain value of 30.08 ± 15.86 and 47.65 ± 28.66, respectively, and a significant *p* value (0.001). In contrast, the 350 bp allele of the *OsCYP71P6-1* primer contributes 17.64% and 14.07% in terms of

single-plant yield and number of productive tillers, respectively, relative to the 320 bp allele. Furthermore, the average yield per plant was determined to be 31.21 ± 11.42 g for one allele and 37.90 ± 15.38 g for the other allele of *OsCYP71P6-1*. This difference in yield was statistically significant, indicated by a p value of 0.03.

The mean phenotypic difference between the two subpopulations identified in the STRUCTURE analysis exhibited significant distinctions for several traits including single-plant yield, number of spikelets, filled grains, and number of productive tillers. Notably, the single-plant yield trait displayed average values of 37.06 ± 15.33 g for subpopulation I and 31.42 ± 10.48 g for subpopulation II, with a significant p value of 0.03. Similarly, the mean number of spikelets was found to be 150.93 ± 28.71 and 185.32 ± 43.63 between the populations, respectively, with a p value of 0.02. In addition, the mean difference for the subpopulations along with admixtures also showed significant difference for single-plant yield, number of spikelets, and filled grains per panicle. Moreover, all the varieties were categorized into four distinct haplotypes based on the amplicon lengths of both primers. These groups were designated as follows: Group-1 (320 and 380 bp), Group-2 (350 and 380 bp), Group-3 (320 and 400 bp), and Group-4 (350 and 400 bp). Among these haplotypes, the single-plant yields observed were 29.81 ± 10.24 g, 37.71 ± 14.41 g, 36.64 ± 16.43 g, and 42.32 ± 11.73 g for the 1st, 2nd, 3rd, and 4th haplotypes, respectively, and the p value for the mean difference between the haplotypes was found to be highly significant ($p = 0.005$) (Figure 10, Table 5).

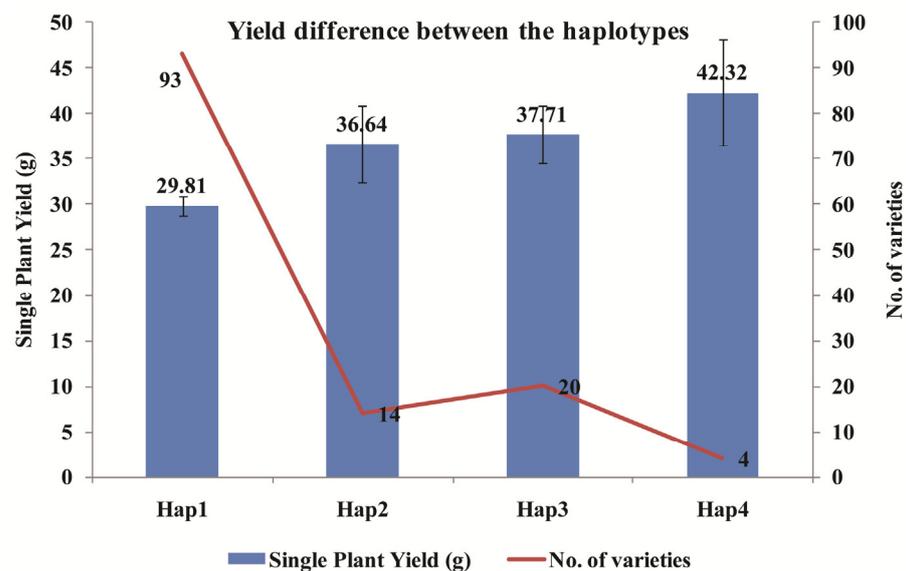


Figure 10. Yield differences between the four haplotypes of *OsCYP71P6* allelic variants. Hap 1–4 indicates the four different haplotypes of *OsCYP71P6-1* and *OsCYP71P6-4* alleles. No. of varieties present in each haplotype is represented by a line graph. Standard error was used to plot the error bar in the bar chart. The number indicates the mean yield and no. of varieties for the haplotypes. Statistical difference in the mean of the single-plant yield between the haplotypes was analyzed using the Z test for mean difference. $p = 0.005$ represents the statistically significant mean difference value calculated using the Z Test at a 1% level of significance. Hap1: *OsCYP71P6-1*_{320bp}, *OsCYP71P6-4*_{380bp}, Hap2: *OsCYP71P6-1*_{350bp}, *OsCYP71P6-4*_{380bp}, Hap3: *OsCYP71P6-1*_{320bp}, *OsCYP71P6-4*_{400bp}, Hap4: *OsCYP71P6-1*_{350bp}, *OsCYP71P6-4*_{400bp}, bp-base pair (amplicon size).

2.10. Association of *OsCYP71P6* SNPs with the Spikelet Fertility Using Linear Regression Model

We studied the 572 SNPs including in/dels of the *OsCYP71P6* gene along with the promoter (Chr12:9577747.9584014) identified among the 1858 genotypes in 3K rice database for their association with spikelet fertility scores using linear regression models. The analysis showed eight SNPs (Table 6) had significant associations, with p values less than 0.05 for spikelet fertility (Chr12: 9581604; Chr12:9582455; Chr12:9582489; Chr12:9582557; Chr12:9582591; Chr12: 9582869; Chr12:9583776; Chr12:9583083; and Chr12:9582921).

Table 6. Linear regression analysis for the association of genetic variants of *OsCYP71P6* in the 3 K database with spikelet fertility score in rice.

Sl.No	SNP Position ^a	Gene Position	Amino Acid Substitution	p Value ^d
1	Chr12:9581604	First Exon (98C>T) ^b	Ser33Leu ^c	0.005767 **
2	Chr12: 9582455	Promoter	-	0.003087 **
3	Chr12: 9582489	Promoter	-	0.019201 *
4	Chr12: 9582557	Promoter	-	0.047372 *
5	Chr12: 9582591	Promoter	-	0.004385 **
6	Chr12: 9582869	Promoter	-	0.001087 **
7	Chr12: 9583776	Promoter	-	0.020460 *
8	Chr12: 9583083	Promoter	-	0.027545 *
9	Chr12: 9582921	Promoter	-	0.003404 **

^a—SNP position indicates the nucleotide position significantly associated with the spikelet fertility score, ^b—98C>T indicates the 98th nucleotide substitution in the coding sequence from the start codon of CYP71A1, ^c—serine is substituted by leucine amino acid in the 33rd position of the protein, ^d—double asterisk indicates significance at 1% level of significance and single asterisk indicates significance at 5% level.

Within the set of eight SNPs, a solitary SNP was detected in the coding region, leading to the Ser33Leu amino acid substitution. This change occurred in the span between the signal peptide and the p450 functional domain of the *OsCYP71P6* protein. In contrast, all other SNPs were situated in the promoter region of *OsCYP71P6*. Furthermore, among these eight SNPs, six distinct haplotypes emerged. Notably, Hap2 exhibited a lower percentage (4.6%) of genotypes with a spikelet fertility score of 5, contrasting with the other haplotypes (11.9%). The specific sequence data for the *OsCYP71P6* haplotypes can be found in Table S10. Moreover, Hap4 also displayed a reduced count of genotypes (0.8%), with a spikelet fertility score of 7, relative to the other haplotypes (1.7%). The variance analysis unveiled significant distinctions in the number of genotypes across different spikelet fertility scores among the haplotypes (*p* value: 0.00006) (refer to Tables S10 and S11).

3. Discussion

Serotonin is synthesized in plants from the amino acid tryptophan via decarboxylation to tryptamine, which is then hydroxylated to form serotonin via the shikimate pathway [27]. Tryptophan decarboxylase (TDC) and tryptamine-5-hydroxylase (T-5-H) catalyze these reactions [27]. *CYP71A1* is a major gene functioning in the serotonin pathway and is reported to be involved in biotic and abiotic stress tolerance in rice [27,45]. In this work, a total of 105 *OsCYP71* family genes were identified in the indica rice genome. These genes were further categorized into twelve distinct subfamilies. Notably, members sharing the same clade exhibited resemblances in terms of gene structures and conserved motifs. In addition, 105 *OsCYP71* genes were distributed on 11 chromosomes, and 36 pairs of *OsCYP71* involved in gene duplication events were found. The promoter region of *OsCYP71* contains a multitude of cis-elements associated with responses to light, hormones, and various stress conditions. Further, transcriptome profiling revealed that a majority of the genes within the *OsCYP71* family were responsive to hormonal stimuli and exhibited induction across various tissues and developmental stages in rice. It was discovered that alleles of *OsCYP71P6* located in the promoter region play a role in regulating both spikelet fertility and yield-related traits in rice. Furthermore, a non-synonymous substitution adjacent to the signal peptide region of *OsCYP71P6* and promoter in/dels were identified as having a notable association with spikelet fertility scores in rice. The implications and the significance of these findings are discussed in detail below.

3.1. Identification and Evolutionary Analysis of the *OsCYP71* Gene Family in Rice

The cytochrome P450 (CYP450) gene family plays a critical role in plant growth and developmental processes, nutrition, and detoxification of xenobiotics [1–3,5]. In this study, a total of 105 *OsCYP71* family genes were identified in the rice genome, compared with *Medicago truncatula* (59), *Arabidopsis thaliana* (54), *Vitis vinifera* (125), *Zea mays* (55), *Sorghum*

bicolor (87), and *Glycine max* (53) [5,17,18,40,46,47]. This variation could potentially be attributed to gene duplication events. Phylogenetic relationships revealed that *OsCYP71* gene family members were divided into 12 subfamilies (Figure 1). Group VII encompassed a total of 74 members. Conversely, Groups I, II, III, IV, V, VI, VIII, IX, X, XI, and XII contained 0, 8, 5, 0, 13, 2, 3, 0, 0, 0, and 0 members, respectively (Figure S2). Based on the phylogenetic analysis, we assume that the *OsCYP71* family has a significant degree of evolutionary expansion in rice. Further, subcellular localization predicts that *OsCYP71* gene family members are mainly situated in the endomembrane system. The wide range of Mw and pI of *OsCYP71* may decide their functional diversity in different metabolic pathways (Table 1).

Chromosome mapping revealed that the distribution of *OsCYP71* genes is uneven across a total of 11 chromosomes (Figure 2A). There are hot spots or gene clusters on Chr2, Chr3, Chr6, Chr8, Chr9, and Chr10 (Figure 2B). The *OsCYP71* gene was randomly and unevenly distributed on 12 chromosomes, and WGD or segmental duplication events were detected in each chromosome. Our results suggest that gene duplication affects *OsCYP71* chromosome location and gene family expansion depends on sequence duplication, either in WGD or segmental events. Apart from the segmental type, genome-wide duplication events also contribute to the alteration in the count of family members. In general, gene duplication plays an important role in the expansion and evolution of gene families, in which tandem repeats produce gene clusters or hotspots and fragment repeats produce homologous genes [48]. In this study, we identified a total of 36 instances of gene duplication within the *OsCYP71* gene family (Figures 3 and S3). Further, different lengths of *OsCYP71* may play an important role in diversifying gene functions. The Ka/Ks score for these genes was observed to be less than one, suggesting a robust purifying selection with minor alterations following gene duplication. These findings provide evidence that the *OsCYP71* family genes have undergone a conserved evolutionary process. Furthermore, upon conducting an analysis of the exon-intron structure of *OsCYP71* genes, it became evident that the quantity of introns exhibited variation across distinct subfamilies (Figure 4). Across all CYP subfamily genes, the count of introns varies from zero to four. However, the majority of these genes typically possess between one and two introns. The maximum four introns were detected in *OsCYP71E2*, while *OsCYP71AD*, *OsCYP71C3*, *OsCYP71C4*, *OsCYP71AB7*, *OsCYP71W5*, *OsCYP71U3*, *OsCYP71T4*, *OsCYP71T7*, *OsCYP71T5*, *OsCYP71T8*, *OsCYP71T6*, and *OsCYP71T9* were intronless (Figure 4). Several studies have shown that a number of introns of CYP450 superfamily genes may have been lost during evolutionary processes in plants [13,49–51]. Additionally, a conserved motif analysis can offer deeper insights into the functional distinctions among family members. According to the findings from MEME, we identified the presence of motif 10 in *OsCYP71* proteins. Motifs 1, 2, 3, 6, 7, and 10 were observed to exhibit strong conservation across the majority of *OsCYP71* proteins (Figure 5A,B). The amino acid sequence alignment showed that *OsCYP71* proteins consist of the conserved SRS and the heme-binding motifs (Figure S4A). We also predicted the 3D structure of *OsCYP71* proteins (Figure S4B). Thus, based on these results, although certain motifs exhibit strong conservation within the *OsCYP71* family, diverse subfamilies display unique motifs that potentially participate in specialized functions.

3.2. The *OsCYP71* Gene Family Contains Multiple Cis-Regulatory Involved in Plant Developmental Processes

Cis-elements in the promoter region play a key role in the regulation of gene expression [44,52–54]. In this study, numerous cis-regulatory elements were identified within the promoter regions of *OsCYP71*. These encompassed elements responsive to phytohormones, light, cell cycle regulation, circadian rhythms, seed-specific regulation, as well as defense and stress responses (Figure 6A,B). In terms of hormonal response, ABRE, AuxRE, GARE, SARE, and MeJARE were the primary elements (Table S4), and similar results have been observed in other species [12,13,50,51]. Many researchers demonstrated that gene duplication is a key factor in gene family expansion [1,55]. Gene duplication expanded the number of gene family members under evolutionary pressures; further, mutations in

these genes may modulate the expression patterns of gene family members [56–59]. The CYP450 gene family holds a pivotal function in various aspects of plant growth and development, as well as in processes related to nutrition and the detoxification of xenobiotics in plants [1,3–5]. In addition, CYP450 has been implicated in diverse metabolic reactions and targets a wide range of biological molecules. These biosynthetic reactions lead to diverse plant hormones, fatty acid conjugates, lignin, secondary metabolites, and numerous defensive compounds [8,9]. It is intriguing to hypothesize that gene expression blueprints are evidence of their biological importance. The expression patterns of *OsCYP71* genes were thoroughly investigated across diverse tissues and developmental stages and under various hormone treatments. Our analysis revealed that the 105 *OsCYP71*-family genes studied exhibited differential expression across various tissues, developmental stages, and in response to hormone treatments (Figure 7A–C). For instance, *CYP71X4* and *CYP71X12* expression was elevated in ovary, embryo, and endosperm tissue, while the expression of *CYP71X13P* was induced in lemma, palea, and endosperm tissue. Further, the expression of *CYP71X7* was highly induced in vegetative, reproductive, ripening leaf blade, vegetative, reproductive leaf sheath, ovary, embryo and endosperm tissue, whereas the expression of *CYP71Z6*, *CYP71R2P*, *CYP71R1*, and *CYP71K12* was highly up-regulated in vegetative, reproductive, ripening leaf blade, and reproductive leaf sheath tissue. In addition, *CYP71Q1* expression was elevated in the shoot after 3 h, 6 h, and 12 h of the ABA and JA treatments, while the expression level of *CYP71Z4* was induced in the shoot after 3 h, 6 h, and 12 h of GA treatment. The transcript levels of *CYP71C16*, *CYP71Z6*, *CYP71C19P*, *CYP71V2*, *CYP71X10*, *CYP71W1*, and *CYP71T1* were induced in the shoot after 3 h, 6 h, and 12 h of JA treatment. Further, expression of *CYP71X10* increased in the shoot after 3 h, 6 h, and 12 h of BRS treatment, while *CYP71X11* was elevated in the shoot after auxin treatment for 1 h, 3 h, 6 h, and 12 h (Figure 7B). Following cytokinin treatment, the expression of *CYP71X12* increased in the root at the 1 h, 3 h, and 6 h time points. Conversely, under JA treatment, *CYP71X14*, *CYP71P1*, and *CYP71C16* were notably induced in the root after 30 min, 1 h, and 3 h. Furthermore, *CYP71Z4*, *CYP71W1*, *CYP71K5*, *CYP71Z2*, *CYP71C20*, and *CYP71Z3* were highly up-regulated in the root after JA treatment for 3 h and 6 h (Figure 7C). Similarly, the member of *BnCYP86* gene family exhibited expression in different tissues and responses to diverse environmental stresses, as did ABA in *Brassica Juncea* [13]. The levels of *SICYP71AX* and *SICYP77A20* gene expression were elevated in both the green and mature green stages of tomato fruit [50]. Collectively, these results have shown that *OsCYP71* gene family members may participate in different developmental processes and responses to phytohormones in rice. Thus, the outcome of this work may provide important insight into deciphering the biological functions of *OsCYP71* gene family members in future.

3.3. The Effect of Promoter In/Dels of *OsCYP71P6* on Yield-Related Traits

The promoter in/del alleles of *OsCYP71P6* exhibited differences of 20.19%, 13.65%, 5.37%, 8.79%, and 36.86% in relation to single-plant yield, number of spikelets per panicle, panicle length, panicle weight, and unfilled grains per panicle, respectively, across various rice varieties. In support of these findings, the previous report by Lu et al. [28] showed the association of the loss-of-function allele of *OsCYP71P6* with the yield and number of tillers in rice. This suggests that the specific promoter allele of *OsCYP71P6* could be correlated with the gene's expression level, thereby playing a role in regulating yield-related traits in rice varieties developed for cultivation in India. Previously, it was discovered that promoter insertion/deletions (in/dels) within the malate transporter 9 (Al-MT9) gene in tomato were linked to variations in malate content [60]; the Arabidopsis Fumarase gene (FUM) promoter in/dels regulates carbon assimilation and nitrogen use [61]; the wheat promoter in/del in the elongation factor (TEF-7A) was found to be associated with the grain number per spike [62]. Promoter insertion/deletion (In/Dels) polymorphisms have been linked to heterotic gene expression in hybrid varieties [63]. The promoter in/dels in these examples resulted in altered gene expression and associated phenotypes. Thus, *OsCYP71P6* gene expression variation could be related to the identified promoter in/dels,

which needs to be further confirmed. Furthermore, the *OsCYP71P6* gene participates in the serotonin biosynthesis pathway. Increasing the expression of tryptophan decarboxylase led to elevated serotonin levels. However, this led to a notable decrease in yield due to changes in rice panicle branching and the quantity of spikelets [64]. Additionally, overexpression of rice *OsSNAT1* (Serotonin N-Acetyl Transferase-1), which converts serotonin into melatonin, enhanced the number of panicles per plant and yield [65]. This supports the hypothesis that high levels of serotonin could have a negative effect on the yield. Therefore, it is feasible that elevated serotonin levels exert a negative regulatory effect on yield-related traits in rice varieties and targeting *OsCYP71P6* promoter insertions/deletions could be a potential strategy to finely adjust both gene expression and serotonin levels in rice, ultimately leading to increased yields.

3.4. The Effect of 3'-UTR In/Dels of *OsCYP71P6* on Single-Plant Yield

The 3'-untranslated region (3'-UTR) allele of *OsCYP71P6* also showed significant difference for the single-plant yield in rice varieties. The single-plant yield was 17.65% higher in the varieties having the insertion allele (350 bp) of the *OsCYP71P6-1* primer. Previously, it was reported in different genes and crops that sequence variation in the 3'-UTR region affects gene expression and was associated with major traits [66–68]. Therefore, variation in the 3'-UTR of *OsCYP71P6* might influence the expression or stability of the gene's transcripts, subsequently impacting its related yield traits in rice.

3.5. A Non-Synonymous Substitution Near the Signal Peptide Region of the *OsCYP71P6* Gene Regulates Spikelet Fertility

Haplotype 4 of *OsCYP71P6* exhibited a notably lower proportion of individuals (0.8%) with elevated spikelet fertility scores (SF7) compared to the other five haplotypes (1.61%). The SNPs leading to the amino acid substitution of Ser33Leu are situated close to the carboxy end of the trans-membrane (TM) signal peptide region within the *OsCYP71P6* protein. This indicates that the substitution might be regulating the *OsCYP71P6* cellular localization. The eukaryotic CYPs are localized in the membrane, especially the endoplasmic reticulum, and are affected by the membrane properties [69]. Recently, it was found that mutations in the TM region affected the enzymatic function of CYPs [70]. Furthermore, a comparable region within closely related homologs of *OsCYP71P6* in rice also displayed sequence variations, including deletions or substitutions affecting serine amino acids (LOC_Os09g26940). This suggests that CYP71 homologs in rice exhibit substantial diversity in their signal peptide regions, and potentially in cellular localizations as well. Consequently, it becomes crucial to determine the functional importance of the amino acid substitution of Ser33Leu in relation to spikelet fertility.

3.6. Gene Diversity of *OsCYP71P6* in Rice Varieties

In this study, distance- and model-based analysis of in/dels in the *OsCYP71P6* gene grouped the rice varieties into two major groups. In similar findings, two sub-populations were also identified for the *Dense and Erect Panicle 1* (DEP1) gene in high-yielding japonica rice varieties [71]. Further, significant differences were observed in single-plant yield and the number of spikelets per panicle (17.6% and 14.1%, respectively) between the varieties categorized within the two distinct sub-populations. This underscores the correlation between genetic variations in the *OsCYP71P6* gene and yield-related traits in rice. Of particular significance is the identification of four in/del-based haplotypes within Indian rice varieties, which exhibit substantial differences in single-plant yield, emphasizing that the rice breeding initiative for varietal enhancement possesses ample diversity in the *OsCYP71P6* gene that can be strategically harnessed for genetic improvement. In support of this observation, two to three alleles were identified in the major yield-related genes (GS3, qSW5, GS5, Gn1a, and DEP1) in high-yielding varieties grown in Northern China [72]. However, the favorable allele proportion of the *OsCYP71P6* gene identified in this study was only ~3%. Therefore, targeted breeding for introgression of the favorable allele of

OsCYP71P6 could be attempted to increase the genetic gain for yield in rice. In this regard, the favorable haplotype (Hap4) of *OsCYP71P6* identified in four rice varieties (PR106, PR113, HPR2143, Himalaya-1) could be used as a donor for marker-assisted selection for yield improvement, whereas the most common haplotypes present in the rice varieties showed the least mean single-plant yields. Previously, utilizing sequence-based methodologies on major yield-related genes in rice varieties led to the discovery of novel alleles associated with traits related to yield [73]. In comparison, the gene-specific in/dels marker approach taken in this study could also be utilized for the marker-assisted breeding program in rice. Additionally, evaluation of the *OsCYP71P6* haplotypes in various rice varieties to determine their influence on traits related to both biotic and abiotic stress tolerance holds significant importance for the effectiveness of breeding efforts and genome editing programs [28]. Nowadays, the rice 3K panel, comprised of sequence information of ~3000 rice genotypes, is used for the identification of the favorable haplotypes of key genes in rice [74]. In this study, information from the 3k database was leveraged to create in/del primers. These primers were developed with the aim of comprehending the diversity of the *OsCYP71P6* gene among rice varieties that have been cultivated in India. A similar approach was used in our previous report to identify the UTR specific in/dels in the *TPP7* gene for tolerance to germination under submerged conditions [39]. Therefore, this strategy would be immensely valuable for uncovering gene diversity within key genes in the rice genome. Similarly, major gene-specific in/dels have been identified for the yield-related gene *Gn1a* in rice [68,75]. The superior haplotypes that have been identified could be incorporated into widely grown rice varieties using marker-assisted backcross breeding techniques.

4. Materials and Methods

4.1. Identification and Characterization of *OsCYP71* Members in Rice Genome

A hidden Markov Model (HMM) profile of the *OsCYP71* conserved domain (PF00067) was obtained from the Pfam database [76] by screening in the rice genome database. *OsCYP71* family members were further identified through the NCBI-CDD [77] and SMART databases [78]. An isoelectric point calculator was utilized to analyze the theoretical isoelectric point (PI) and molecular weight (MW) of the *OsCYP71* protein [79]. PSORT and BUSCA were used to predict the subcellular localization of *OsCYP71* encoded proteins [80,81].

4.2. Phylogenetic Tree, Gene Structure, and Conserved Motif Analysis of the *OsCYP71* Family in Rice

Arabidopsis, tomato, and *OsCYP71* protein sequences were acquired from Ensembl Plants (<https://plants.ensembl.org/index.html>, (accessed on 20 June 2023)) and, subsequently, phylogenetic trees were constructed using MEGA 11 [81]. ClustalW was used for sequence alignment, the maximum likelihood method was used for phylogenetic tree construction, and the reliability of the tree obtained was determined using the bootstrap method with 1000 replicates. The exon-intron structure of *OsCYP71* genes was visualized and mapped using the Gene Structure Display Server (GSDS) [82]. The conserved motifs of *OsCYP71* protein sequences were plotted by the MEME webserver [83]. The Phyre2 web server was used to create the 3D structure of *OsCYP71* [84].

4.3. Chromosome Localization, Gene Replication, Cis-Regulatory Elements, GO Enrichment, and Expression Analysis

The acquisition of chromosome localization information for *OsCYP71* genes was carried out through Ensembl Plants (<http://plants.ensembl.org/biomart/martview>, (accessed on 20 June 2023)), aiming to facilitate their placement on the respective chromosomes. The *OsCYP71* gene family members were mapped using PhenoGram [85]. Gene replication and K_a/K_s values were analyzed using MCScan tools [86] and TBtools [87]. The 1500 bp sequence upstream of the *OsCYP71* genes was used for promoter element analysis by the PlantCARE webserver [88]. GO enrichment of *OsCYP71* proteins was performed using the singular enrichment analysis (SEA) function of a network-based agriGO program [89]. Tissue-specific expression values for the 105 *OsCYP71* genes were extracted from the RiceX-

Pro database (<http://ricexpro.dna.affrc.go.jp>, (accessed on 28 June 2023)) and, subsequently, heatmaps were generated using ClustVis [90].

4.4. Phenotyping for Different Traits

The study was carried out at ICAR National Rice Research Institute, Cuttack in January 2021. A total of 131 released rice varieties were considered for the phenotyping data analysis (Table S7). Approximately a hundred seeds of the varieties were line sown in the nursery and transplanted in the puddled field conditions, with two lines per replication and each line consisting of fifteen seedlings. Additionally, two replications and alpha lattice designs [91] to reduce the variability in field experiments were maintained for the evaluation of the yield-related traits. The standard management practices of application of fertilizer and irrigation management were followed and at the physiological maturity stage, the number of productive tillers (nos.) was directly measured in three plants per line in the field. Further, at the crop's maturity stage, various parameters including panicle length (in centimeters), panicle weight (in grams), spikelet count, both filled and unfilled grain quantities, seed weight (in grams), and single-plant yield (in grams) were assessed. This analysis involved the harvesting of three plants per line. The collected panicles were sun-dried for a duration of one week before being utilized for the assessment of panicle-related characteristics. The methodologies for measuring these traits were consistent with previously established protocols [92,93]. Briefly, panicle length was measured manually using a measuring scale from the top three panicles per plant. Further, the weight of these panicles was measured using a weighing machine. Additionally, the top three panicles per plant were also used for manually counting the number of spikelets per panicle along with filled and unfilled grains. Then, a hundred grains per line were manually counted and weighed for 100 g seed weight. Furthermore, the panicles of individual plants were subjected to threshing, and the filled spikelets were weighed to determine the yield per single plant. In parallel, young leaves were gathered for DNA extraction, rapidly frozen using liquid nitrogen, and subsequently preserved at a temperature of -80°C for future investigations.

4.5. Development of Gene Specific Polymorphic Variants-Insertion/Deletions (GPV-In/Dels)

Insertion/deletions present in the *OsCYP71P6* (CYP71A1, LOC_Os12g16720) gene, including the promoter, were identified from the SNP Seek database of rice (<https://snp-seek.irri.org/>, (accessed on 29 June 2023)). The in/del variants, present in approximately 3000 genotypes within the SNP Seek database, were detected by aligning against the Nipponbare reference genome. These variants were specific to certain genes. As a result of this development, the marker was designated as "Gene-Specific Polymorphic Variants-In/Dels" (GPV-In/Dels). Further, the CYP71A1 gene sequence of Nipponbare was retrieved from the RGAP database (<http://rice.plantbiology.msu.edu/>, (accessed 29 June 2023)) and utilized for designing the primers flanking the selected in/dels of size > 10 bp. The details of the primer sequence are given in Tables S2 and S8. In addition, the SNPs and in/dels, totaling around 572, that were discovered within the *OsCYP71P6* gene were subjected to analysis along with the spikelet fertility quality scores (ranging from 1 to 7) available in the 3 K rice database. This analysis involved investigating the associations between variants and traits through a linear regression model. The 'lm' function within the R software, specifically R Version 3.6.0, was employed for this purpose [94].

4.6. Experimental Validation of *OsCYP71P6* In/Dels

Samples of leaves from various rice varieties were gathered to validate the newly created in/del markers. The CTAB (cetyl trimethyl ammonium bromide) technique was employed to extract genomic DNA from these leaf samples [92]. Using agarose gel electrophoresis, the purity of the DNA was determined (0.8 percent). For PCR amplification, genomic DNA was diluted with nuclease-free water to a working concentration ($50\text{ ng}/\mu\text{L}$). Initially, all four gene-specific in/del markers were employed to genotype rice varieties,

facilitating an analysis of genetic diversity. The PCR was carried out in a 10 μ L reaction volume containing 1 μ L of template DNA, 1 μ L of 10 \times buffer (1.5 mM Mg in 1 \times), 1.0 μ L of dNTP (2.5 mM), 1 μ L of each in/del primer (0.2 μ M), and 0.2 U of *Taq* polymerase (Kapa Biosystem, Cape Town, South Africa), while the volume was made using double-distilled H₂O. The PCR program includes an initial denaturation for 3 min at 95 $^{\circ}$ C, which was followed by 35 cycles of denaturation for 30 s at 95 $^{\circ}$ C, annealing for 45 s at 55 $^{\circ}$ C, and an extension for 1 min at 72 $^{\circ}$ C, with a final extension for 10 min at 72 $^{\circ}$ C using a thermal master cycler (Eppendorf, Hamburg, Germany). The PCR products were resolved using a 4.0 percent agarose gel and recorded using a UV gel documentation system (Bio-Rad, Hercules, CA, USA) (Figure S6). The alleles of the in/del marker were assessed manually, taking into account their amplicon length across various rice cultivars.

4.7. Diversity Analysis

The alleles identified in 131 rice varieties for *OsCYP71P6* gene in/del markers were used for gene diversity, cluster analysis, sub-population structures, and analysis of molecular variance (AMOVA). Gene diversity and cluster analysis were performed using power marker software [95]. Further, AMOVA was performed using the GenALEX V 6.5 tool using amplicon length allelic data as variables [96]. The STRUCTURE software was utilized to analyze population sub-structure, while Structure Harvester was employed to ascertain the optimal number for interpreting the results [97].

5. Conclusions

In this study, we identified 105 *OsCYP71* genes within the indica rice genome. These genes were subsequently classified into 12 distinct subfamilies based on shared characteristics. Notably, genes within the same subfamily exhibited similar gene structures and conserved motifs. Additionally, the distribution of these 105 *OsCYP71* genes spanned across 11 chromosomes, with 36 instances of *OsCYP71* gene duplication events. The promoter regions of *OsCYP71* genes were found to harbor a substantial number of cis-elements associated with light responsiveness, hormone signaling (including Auxin, cytokinin, GA, ABA, MeJA, JA, BRS, and SA), and various stresses such as drought and low temperature. Transcriptome profiling further revealed that a majority of genes within this family demonstrated responsiveness to hormones and were induced across diverse tissues and developmental stages in rice. Employing linear regression models, we identified eight promoters along with a gene variant (Ser33Leu) within *OsCYP71P6* that exhibited a significant association with spikelet fertility. Furthermore, the allelic effects of different *OsCYP71P6* alleles, identified through in/dels polymorphism in 131 rice varieties, were validated for their impact on yield-related traits. Our investigations also revealed that the *OsCYP71P6* gene plays a role in the regulation of spikelet count, filled grains, single-plant yield, panicle length, and panicle weight in various rice varieties. These findings serve as a robust foundation for deeper exploration of the functions of *OsCYP71*-family genes in a wide array of biological processes. Additionally, the outcomes of our study underscore the potential influence of promoter allelic variation and the Ser33Leu amino acid substitution within the *OsCYP71P6* gene on yield-related traits in rice. As a result, the promoter variants of *OsCYP71P6* that we have identified hold promise for utilization in efforts aimed at enhancing rice yield.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/plants12173035/s1>: Figure S1: The molecular mass (kDa) and isoelectric point of plots of *OsCYP71* genes; Figure S2: Dispersal of *OsCYP71*s in a different group of the phylogenetic tree; Figure S3: Evolutionary analysis of *OsCYP71* genes. A phylogenetic tree was produced using MEGA7 with the NJ method and 1000 bootstrap replications. A black asterisk denotes the duplicated gene pairs; Figure S4: Amino acid sequence alignment and three-dimensional structure of the *OsCYP71*. A. The SRS and heme-binding domain is underlined in red color. B. Predicted 3D structures of *OsCYP71* proteins; Figure S5: Gene ontology term distribution in the *OsCYP71* gene family predicted using AgriGO A. Biological and B. Molecular category; Figure S6: Gel image analysis of the amplified

products of *OsCYP71P6-1* and *OsCYP71P6-4* alleles in agarose gel (4%). a) Alleles of the *OsCYP71P6-1* in rice varieties b) alleles of *OsCYP71P6-4* in rice varieties. V—Varieties. The name of the varieties is given in Table S8. Table S1: *OsCYP* genomic, CDS protein and promoter sequence; Table S2: CYP proteins from *Arabidopsis*, rice and tomato were used to generate a phylogenetic tree; Table S3: The ratio of K_a/K_s and allocation of replicated rice CYP genes; Table S4: Cis-regulatory elements are identified in the *OsCYP* gene promoter region; Table S5: Significant Go term predicted in *OsCYP* gene family by AgriGo analysis; Table S6 *OsCYP* gene annotation using eggNOGmapper; Table S7: Mean values of yield-related traits and amplicon length alleles/polymorphism of *OsCYP71P6* in/del primers in rice varieties; Table S8: In/Del Primer sequences used in the study; Table S9: Structure analysis and inferred clusters of the two sub-populations in the rice varieties for *OsCYP71P6* In/Del variants; Table S10: Proportion of genotypes in six different haplotypes from nine associated SNPs in *OsCYP71P6* gene retrieved from rice 3K database; Table S11: Analysis of variance (ANOVA) for the proportion of genotypes in different haplotypes of *OsCYP71P6* in rice.

Author Contributions: C.P. and M.S.K. designed and wrote the manuscript; C.P. acquired funding; C.P. and M.S.K. supervised the study; B.S. and I.N.: genotyping, field data collection, and primary analysis; K.K.S.: statistical analysis, manuscript editing; H.N.S.: data finalization, assisting in field experiments; C.B.: bioinformatics analysis and preparation of primary draft of manuscript; S.R.P.: data interpretation and manuscript editing; J.L.K.: analysis of data, finalization of manuscript; S.K.D.: manuscript; S.-M.C., M.H.S., S.A. and S.S. provided valuable feedback to this study. All authors have read and agreed to the published version of the manuscript.

Funding: We sincerely thank the Indian Council of Agricultural Research, New Delhi (ICAR) and National Agricultural Science Fund (NASF), ICAR, New Delhi for providing the funding support to carry out the study. Also, the study was supported by the Researchers Supporting Project number (RSP2023R194), King Saud University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is available in the manuscript and in the Supplementary Materials.

Acknowledgments: We also thank Amaresh Kumar Nayak, Director, ICAR-NRRI, Cuttack for providing the facilities required to carry out the research work. The first authors acknowledge and thank the Department of Science and Technology (DST), New Delhi for providing the INSPIRE fellowship to pursue the Ph.D. doctoral research work. The authors would like to extend their sincere appreciation to the Researchers Supporting Project number (RSP2023R194), King Saud University, Riyadh, Saudi Arabia.

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

References

- Mizutani, M.; Ohta, D. Diversification of P450 genes during land plant evolution. *Annu. Rev. Plant Biol.* **2010**, *61*, 291–315. [[CrossRef](#)]
- Nelson, D.R. The cytochrome p450 homepage. *Hum. Genom.* **2009**, *4*, 59–65. [[CrossRef](#)] [[PubMed](#)]
- Yu, J.; Tehrim, S.; Wang, L.; Dossa, K.; Zhang, X.; Ke, T.; Liao, B. Evolutionary history and functional divergence of the cytochrome P450 gene superfamily between *Arabidopsis thaliana* and Brassica species uncover effects of whole genome and tandem duplications. *BMC Genom.* **2017**, *18*, 733. [[CrossRef](#)] [[PubMed](#)]
- Xiong, R.; He, T.; Wang, Y.; Liu, S.; Gao, Y.; Yan, H.; Xiang, Y. Genome and transcriptome analysis to understand the role diversification of cytochrome P450 gene under excess nitrogen treatment. *BMC Plant Biol.* **2021**, *21*, 447. [[CrossRef](#)] [[PubMed](#)]
- Jiu, S.; Xu, Y.; Wang, J.; Wang, L.; Liu, X.; Sun, W.; Sabir, I.a.; Ma, C.; Xu, W.; Wang, S. The cytochrome P450 monooxygenase inventory of grapevine (*Vitis vinifera* L.): Genome-wide identification, evolutionary characterization and expression analysis. *Front. Genet.* **2020**, *11*, 44. [[CrossRef](#)] [[PubMed](#)]
- Ohmura, E.; Nakamura, T.; Tian, R.H.; Yahara, S.; Yoshimitsu, H.; Nohara, T. 26-Aminocholestanol derivative, a novel key intermediate of steroidal alkaloids, from *Solanum abutiloides*. *Tetrahedron Lett.* **1995**, *36*, 8443–8444. [[CrossRef](#)]
- Guttikonda, S.K.; Trupti, J.; Bisht, N.C.; Chen, H.; An, Y.-Q.C.; Pandey, S.; Xu, D.; Yu, O. Whole genome co-expression analysis of soybean cytochrome P450 genes identifies nodulation-specific P450 monooxygenases. *BMC Plant Biol.* **2010**, *10*, 243. [[CrossRef](#)]
- Schuler, M.A.; Werck-Reichhart, D. Functional genomics of P450s. *Annu. Rev. Plant Biol.* **2003**, *54*, 629–667. [[CrossRef](#)]
- Tang, B.; Cheng, Y.; Li, Y.; Li, W.; Ma, Y.; Zhou, Q.; Lu, K. Adipokinetic hormone regulates cytochrome P450-mediated imidacloprid resistance in the brown planthopper, *Nilaparvatalugens*. *Chemosphere* **2020**, *259*, 127490. [[CrossRef](#)]

10. Ma, B.; Luo, Y.; Jia, L.; Qi, X.; Zeng, Q.; Xiang, Z.; He, N. Genome-wide identification and expression analyses of cytochrome P450 genes in mulberry (*Morus notabilis*). *J. Integr. Plant Biol.* **2014**, *56*, 887–901. [[CrossRef](#)]
11. Ren, J.; Yang, L.; Li, Q.; Zhang, Q.; Sun, C.; Liu, X.; Yang, N. Global investigation of cytochrome P450 genes in the chicken genome. *Genes* **2019**, *10*, 617. [[CrossRef](#)]
12. Sun, K.; Fang, H.; Chen, Y.; Zhuang, Z.; Chen, Q.; Shan, T.; Khan, M.K.R.; Zhang, J.; Wang, B. Genome-wide analysis of the cytochrome P450 gene family involved in salt tolerance in *Gossypium hirsutum*. *Front. Plant Sci.* **2021**, *12*, 685054. [[CrossRef](#)] [[PubMed](#)]
13. Wang, Z.; Zhang, Y.; Song, M.; Tang, X.; Huang, S.; Linhu, B.; Jin, P.; Guo, W.; Li, F.; Xing, L. Genome-Wide Identification of the Cytochrome P450 Superfamily Genes and Targeted Editing of BnCYP704B1 Confers Male Sterility in Rapeseed. *Plants* **2023**, *12*, 365. [[CrossRef](#)] [[PubMed](#)]
14. Xia, Y.; Yang, J.; Ma, L.; Yan, S.; Pang, Y. Genome-wide identification and analyses of drought/salt-responsive cytochrome P450 genes in *Medicago truncatula*. *Int. J. Mol. Sci.* **2021**, *22*, 9957. [[CrossRef](#)]
15. Qin, P.; Zheng, H.; Tao, Y.; Zhang, Y.; Chu, D. Genome-Wide Identification and Expression Analysis of the Cytochrome P450 Gene Family in Bemisia tabaci MED and Their Roles in the Insecticide Resistance. *Int. J. Mol. Sci.* **2023**, *24*, 5899. [[CrossRef](#)]
16. Durst, F.; Nelson, D.R. Diversity and evolution of plant P450 and P450-reductases. *Drug Metab. Durg Interat.* **1995**, *12*, 189–206. [[CrossRef](#)] [[PubMed](#)]
17. Paquette, S.M.; Bak, S.; Feyereisen, R. Intron–exon organization and phylogeny in a large superfamily, the paralogous cytochrome P450 genes of *Arabidopsis thaliana*. *DNA Cell Biol.* **2000**, *19*, 307–317. [[CrossRef](#)] [[PubMed](#)]
18. Nelson, D.R.; Schuler, M.A.; Paquette, S.M.; Werck-Reichhart, D.; Bak, S. Comparative genomics of rice and Arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot. *Plant Physiol.* **2004**, *135*, 756–772. [[CrossRef](#)] [[PubMed](#)]
19. Nelson, D.R. Cytochrome P450 Nomenclature, 2004. *Methods Mol Biol.* **2006**, *320*, 1–10.
20. Bak, S.; Beisson, F.; Bishop, G.; Hamberger, B.; Höfer, R.; Paquette, S.; Werck-Reichhart, D. Cytochromes P450. *Arab. Book/Am. Soc. Plant Biol.* **2011**, *9*, e0144. [[CrossRef](#)]
21. Du, H.; Huang, Y.; Tang, Y. Genetic and metabolic engineering of isoflavonoid biosynthesis. *Appl. Microbiol. Biotechnol.* **2010**, *86*, 1293–1312. [[CrossRef](#)]
22. Du, Y.; Chu, H.; Chu, I.K.; Lo, C. CYP93G2 is a flavanone 2-hydroxylase required for C-glycosylflavone biosynthesis in rice. *Plant Physiol.* **2010**, *154*, 324–333. [[CrossRef](#)]
23. Sawada, Y.; Kinoshita, K.; Akashi, T.; Aoki, T.; Ayabe, S.i. Key amino acid residues required for aryl migration catalysed by the cytochrome P450 2-hydroxyisoflavanone synthase. *Plant J.* **2002**, *31*, 555–564. [[CrossRef](#)] [[PubMed](#)]
24. Cooper, L.; Doss, R.; Price, R.; Peterson, K.; Oliver, J. Application of Bruchin B to pea pods results in the up-regulation of CYP93C18, a putative isoflavone synthase gene, and an increase in the level of pisatin, an isoflavone phytoalexin. *J. Exp. Bot.* **2005**, *56*, 1229–1237. [[CrossRef](#)] [[PubMed](#)]
25. Waki, T.; Yoo, D.; Fujino, N.; Mameda, R.; Denessiouk, K.; Yamashita, S.; Motohashi, R.; Akashi, T.; Aoki, T.; Ayabe, S.-i. Identification of protein–protein interactions of isoflavonoid biosynthetic enzymes with 2-hydroxyisoflavanone synthase in soybean (*Glycine max* (L.) Merr.). *Biochem. Biophys. Res. Commun.* **2016**, *469*, 546–551. [[CrossRef](#)]
26. Chang, Z.; Wang, X.; Wei, R.; Liu, Z.; Shan, H.; Fan, G.; Hu, H. Functional expression and purification of CYP93C20, a plant membrane-associated cytochrome P450 from *Medicago truncatula*. *Protein Expr. Purif.* **2018**, *150*, 44–52. [[CrossRef](#)] [[PubMed](#)]
27. Erland, L.A.; Turi, C.E.; Saxena, P.K. Serotonin in plants: Origin, functions, and implications. *Serotonin* **2019**, 23–46. [[CrossRef](#)]
28. Lu, H.-P.; Luo, T.; Fu, H.-W.; Wang, L.; Tan, Y.-Y.; Huang, J.-Z.; Wang, Q.; Ye, G.-Y.; Gatehouse, A.M.; Lou, Y.-G. Resistance of rice to insect pests mediated by suppression of serotonin biosynthesis. *Nat. Plants* **2018**, *4*, 338–344. [[CrossRef](#)]
29. Khush, G.S. What it will take to feed 5.0 billion rice consumers in 2030. *Plant Mol. Biol.* **2005**, *59*, 1–6. [[CrossRef](#)]
30. Kesawat, M.S.; Das, B.K.; Bhaganagare, G.R.; Manorama. Genome-wide identification, evolutionary and expression analyses of putative Fe–S biogenesis genes in rice (*Oryza sativa*). *Genome* **2012**, *55*, 571–583. [[CrossRef](#)]
31. Sakamoto, T.; Matsuoka, M. Identifying and exploiting grain yield genes in rice. *Curr. Opin. Plant Biol.* **2008**, *11*, 209–214. [[CrossRef](#)] [[PubMed](#)]
32. Li, X.; Chen, Z.; Zhang, G.; Lu, H.; Qin, P.; Qi, M.; Yu, Y.; Jiao, B.; Zhao, X.; Gao, Q. Analysis of genetic architecture and favorable allele usage of agronomic traits in a large collection of Chinese rice accessions. *Sci. China Life Sci.* **2020**, *63*, 1688–1702. [[CrossRef](#)] [[PubMed](#)]
33. Jiao, Y.; Wang, Y.; Xue, D.; Wang, J.; Yan, M.; Liu, G.; Dong, G.; Zeng, D.; Lu, Z.; Zhu, X. Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nat. Genet.* **2010**, *42*, 541–544. [[CrossRef](#)] [[PubMed](#)]
34. Li, W.; Chern, M.; Yin, J.; Wang, J.; Chen, X. Recent advances in broad-spectrum resistance to the rice blast disease. *Curr. Opin. Plant Biol.* **2019**, *50*, 114–120. [[CrossRef](#)] [[PubMed](#)]
35. Huang, L.-Y.; Li, X.-X.; Zhang, Y.-B.; Fahad, S.; Fei, W. dep1 improves rice grain yield and nitrogen use efficiency simultaneously by enhancing nitrogen and dry matter translocation. *J. Integr. Agric.* **2022**, *21*, 3185–3198. [[CrossRef](#)]
36. Han, B.; Xue, Y. Genome-wide intraspecific DNA-sequence variations in rice. *Curr. Opin. Plant Biol.* **2003**, *6*, 134–138. [[CrossRef](#)]
37. Miyashita, N.T.; Yoshida, K.; Ishii, T. DNA variation in the metallothionein genes in wild rice *Oryza rufipogon*: Relationship between DNA sequence polymorphism, codon bias and gene expression. *Genes Genet. Syst.* **2005**, *80*, 173–183. [[CrossRef](#)]

38. Alexandrov, N.; Tai, S.; Wang, W.; Mansueto, L.; Palis, K.; Fuentes, R.R.; Ulat, V.J.; Chebotarov, D.; Zhang, G.; Li, Z. SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Res.* **2015**, *43*, D1023–D1027. [[CrossRef](#)]
39. Chidambaranathan, P.; Sabarinathan, S.; Sanghamitra, P.; Dash, G.K.; Lenka, D.; Balasubramania, C.; KJ, P.; Sahoo, R.K.; Samantaray, S.; BN, D. Haplogenic Quantitative Effects Regulate Flooded Germination, Subsequent Water Deficit Stress, and Recovery in Direct Seeded Rice. *Authorea* **2022**. [[CrossRef](#)]
40. Xu, W.; Bak, S.; Decker, A.; Paquette, S.M.; Feyereisen, R.; Galbraith, D.W. Microarray-based analysis of gene expression in very large gene families: The cytochrome P450 gene superfamily of *Arabidopsis thaliana*. *Gene* **2001**, *272*, 61–74. [[CrossRef](#)]
41. Xie, M.-M.; Gong, D.-P.; Li, F.-X.; Liu, G.-S.; Sun, Y.-H. Genome-wide analysis of cytochrome P450 monooxygenase genes in the tobacco. *Yi Chuan Hered.* **2013**, *35*, 379–387. [[CrossRef](#)] [[PubMed](#)]
42. Babu, P.R.; Rao, K.V.; Reddy, V.D. Structural organization and classification of cytochrome P450 genes in flax (*Linum usitatissimum* L.). *Gene* **2013**, *513*, 156–162. [[CrossRef](#)] [[PubMed](#)]
43. Yu, J.; Hu, S.; Wang, J.; Wong, G.K.-S.; Li, S.; Liu, B.; Deng, Y.; Dai, L.; Zhou, Y.; Zhang, X. A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* **2002**, *296*, 79–92. [[CrossRef](#)] [[PubMed](#)]
44. Kesawat, M.S.; Kherawat, B.S.; Katara, J.L.; Parameswaran, C.; Misra, N.; Kumar, M.; Chung, S.-M.; Alamri, S.; Siddiqui, M.H. Genome-wide analysis of proline-rich extensin-like receptor kinases (PERKs) gene family reveals their roles in plant development and stress conditions in *Oryza sativa* L. *Plant Sci.* **2023**, *334*, 111749. [[CrossRef](#)] [[PubMed](#)]
45. Fujiwara, T.; Maisonneuve, S.; Isshiki, M.; Mizutani, M.; Chen, L.; Wong, H.L.; Kawasaki, T.; Shimamoto, K. Sekiguchi lesion gene encodes a cytochrome P450 monooxygenase that catalyzes conversion of tryptamine to serotonin in rice. *J. Biol. Chem.* **2010**, *285*, 11308–11313. [[CrossRef](#)]
46. Fang, Y.; Jiang, J.; Du, Q.; Luo, L.; Li, X.; Xie, X. Cytochrome P450 superfamily: Evolutionary and functional divergence in sorghum (*Sorghum bicolor*) stress resistance. *J. Agric. Food Chem.* **2021**, *69*, 10952–10961. [[CrossRef](#)]
47. Khatri, P.; Wally, O.; Rajcan, I.; Dhaubhadel, S. Comprehensive analysis of cytochrome P450 monooxygenases reveals insight into their role in partial resistance against *Phytophthora sojae* in soybean. *Front. Plant Sci.* **2022**, *13*, 862314. [[CrossRef](#)]
48. Hughes, A.L. The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. B Biol. Sci.* **1994**, *256*, 119–124.
49. Lin, H.; Zhu, W.; Silva, J.C.; Gu, X.; Buell, C.R. Intron gain and loss in segmentally duplicated genes in rice. *Genome Biol.* **2006**, *7*, R41. [[CrossRef](#)]
50. Vasav, A.; Barvkar, V. Phylogenomic analysis of cytochrome P450 multigene family and their differential expression analysis in *Solanum lycopersicum* L. suggested tissue specific promoters. *BMC Genom.* **2019**, *20*, 116. [[CrossRef](#)]
51. Du, H.; Ran, F.; Dong, H.-L.; Wen, J.; Li, J.-N.; Liang, Z. Genome-wide analysis, classification, evolution, and expression analysis of the cytochrome P450 93 family in land plants. *PLoS ONE* **2016**, *11*, e0165020. [[CrossRef](#)] [[PubMed](#)]
52. Kesawat, M.S.; Kherawat, B.S.; Ram, C.; Singh, A.; Dey, P.; Gora, J.S.; Misra, N.; Chung, S.-M.; Kumar, M. Genome-Wide Identification and Expression Profiling of Aconitase Gene Family Members Reveals Their Roles in Plant Development and Adaptation to Diverse Stress in *Triticum aestivum* L. *Plants* **2022**, *11*, 3475. [[CrossRef](#)] [[PubMed](#)]
53. Kesawat, M.S.; Kherawat, B.S.; Singh, A.; Dey, P.; Routray, S.; Mohapatra, C.; Saha, D.; Ram, C.; Siddique, K.H.; Kumar, A. Genome-wide analysis and characterization of the proline-rich extensin-like receptor kinases (PERKs) gene family reveals their role in different developmental stages and stress conditions in wheat (*Triticum aestivum* L.). *Plants* **2022**, *11*, 496. [[CrossRef](#)] [[PubMed](#)]
54. Hapgood, J.P.; Riedemann, J.; Scherer, S.D. Regulation of gene expression by GC-rich DNA Cis-elements. *Cell Biol. Int.* **2001**, *25*, 17–31. [[CrossRef](#)] [[PubMed](#)]
55. Freeling, M. Bias in plant gene content following different sorts of duplication: Tandem, whole-genome, segmental, or by transposition. *Annu. Rev. Plant Biol.* **2009**, *60*, 433–453. [[CrossRef](#)] [[PubMed](#)]
56. Heidari, P.; Abdullah; Faraji, S.; Poczai, P. Magnesium transporter gene family: Genome-wide identification and characterization in *Theobroma cacao*, *Corchorus capsularis*, and *Gossypium hirsutum* of family Malvaceae. *Agronomy* **2021**, *11*, 1651. [[CrossRef](#)]
57. Heidari, P.; Puresmaeli, F.; Mora-Poblete, F. Genome-wide identification and molecular evolution of the magnesium transporter (MGT) gene family in *Citrullus lanatus* and *Cucumis sativus*. *Agronomy* **2022**, *12*, 2253. [[CrossRef](#)]
58. Kesawat, M.S.; Kherawat, B.S.; Singh, A.; Dey, P.; Kabi, M.; Debnath, D.; Saha, D.; Khandual, A.; Rout, S.; Manorama; et al. Genome-wide identification and characterization of the brassinazole-resistant (BZR) gene family and its expression in the various developmental stage and stress conditions in wheat (*Triticum aestivum* L.). *Int. J. Mol. Sci.* **2021**, *22*, 8743. [[CrossRef](#)]
59. Kumar, M.; Kherawat, B.S.; Dey, P.; Saha, D.; Singh, A.; Bhatia, S.K.; Ghodake, G.S.; Kadam, A.A.; Kim, H.-U.; Manorama; et al. Genome-wide identification and characterization of PIN-FORMED (PIN) gene family reveals role in developmental and various stress conditions in *Triticum aestivum* L. *Int. J. Mol. Sci.* **2021**, *22*, 7396. [[CrossRef](#)]
60. Ye, J.; Wang, X.; Hu, T.; Zhang, F.; Wang, B.; Li, C.; Yang, T.; Li, H.; Lu, Y.; Giovannoni, J.J. An InDel in the promoter of AL-ACTIVATED MALATE TRANSPORTER9 selected during tomato domestication determines fruit malate contents and aluminum tolerance. *Plant Cell* **2017**, *29*, 2249–2268. [[CrossRef](#)]
61. Pracharoenwattana, I.; Zhou, W.; Keech, O.; Francisco, P.B.; Udomchalothorn, T.; Tschoep, H.; Stitt, M.; Gibon, Y.; Smith, S.M. *Arabidopsis* has a cytosolic fumarase required for the massive allocation of photosynthate into fumaric acid and for rapid plant growth on high nitrogen. *Plant J.* **2010**, *62*, 785–795. [[CrossRef](#)] [[PubMed](#)]
62. Zheng, J.; Liu, H.; Wang, Y.; Wang, L.; Chang, X.; Jing, R.; Hao, C.; Zhang, X. TEF-7A, a transcript elongation factor gene, influences yield-related traits in bread wheat (*Triticum aestivum* L.). *J. Exp. Bot.* **2014**, *65*, 5351–5365. [[CrossRef](#)]

63. Zhang, H.-Y.; He, H.; Chen, L.-B.; Li, L.; Liang, M.-Z.; Wang, X.-F.; Liu, X.-G.; He, G.-M.; Chen, R.-S.; Ma, L.-G. A genome-wide transcription analysis reveals a close correlation of promoter INDEL polymorphism and heterotic gene expression in rice hybrids. *Mol. Plant* **2008**, *1*, 720–731. [[CrossRef](#)] [[PubMed](#)]
64. Kanjanaphachaoat, P.; Wei, B.-Y.; Lo, S.-F.; Wang, I.-W.; Wang, C.-S.; Yu, S.-M.; Yen, M.-L.; Chiu, S.-H.; Lai, C.-C.; Chen, L.-J. Serotonin accumulation in transgenic rice by over-expressing tryptophan decarboxylase results in a dark brown phenotype and stunted growth. *Plant Mol. Biol.* **2012**, *78*, 525–543. [[CrossRef](#)]
65. Lee, K.; Back, K. Overexpression of rice serotonin N-acetyltransferase 1 in transgenic rice plants confers resistance to cadmium and senescence and increases grain yield. *J. Pineal Res.* **2017**, *62*, e12392. [[CrossRef](#)] [[PubMed](#)]
66. Eveland, A.L.; McCarty, D.R.; Koch, K.E. Transcript profiling by 3'-untranslated region sequencing resolves expression of gene families. *Plant Physiol.* **2008**, *146*, 32–44. [[CrossRef](#)]
67. Vignesh, M.; Nepolean, T.; Hossain, F.; Singh, A.; Gupta, H. Sequence variation in 3' UTR region of crtRB1 gene and its effect on β -carotene accumulation in maize kernel. *J. Plant Biochem. Biotechnol.* **2013**, *22*, 401–408. [[CrossRef](#)]
68. Kim, S.-R.; Ramos, J.; Ashikari, M.; Virk, P.S.; Torres, E.A.; Nissila, E.; Hechanova, S.L.; Mauleon, R.; Jena, K.K. Development and validation of allele-specific SNP/indel markers for eight yield-enhancing genes using whole-genome sequencing strategy to increase yield potential of rice, *Oryza sativa* L. *Rice* **2016**, *9*, 12. [[CrossRef](#)]
69. Brignac-Huber, L.M.; Park, J.W.; Reed, J.R.; Backes, W.L. Cytochrome P450 organization and function are modulated by endoplasmic reticulum phospholipid heterogeneity. *Drug Metab. Dispos.* **2016**, *44*, 1859–1866. [[CrossRef](#)]
70. Mustafa, G.; Nandekar, P.P.; Camp, T.J.; Bruce, N.J.; Gregory, M.C.; Sligar, S.G.; Wade, R.C. Influence of transmembrane helix mutations on cytochrome P450-membrane interactions and function. *Biophys. J.* **2019**, *116*, 419–432. [[CrossRef](#)]
71. Zhao, M.; Sun, J.; Xiao, Z.; Cheng, F.; Xu, H.; Tang, L.; Chen, W.; Xu, Z.; Xu, Q. Variations in DENSE AND ERECT PANICLE 1 (DEP1) contribute to the diversity of the panicle trait in high-yielding japonica rice varieties in northern China. *Breed. Sci.* **2016**, *66*, 599–605. [[CrossRef](#)]
72. Dan, L.; WANG, J.-y.; WANG, X.-x.; YANG, X.-l.; Jian, S.; CHEN, W.-f. Genetic diversity and elite gene introgression reveal the japonica rice breeding in northern China. *J. Integr. Agric.* **2015**, *14*, 811–822.
73. Vemireddy, L.R.; Kadambari, G.; Reddy, G.E.; Kola, V.S.R.; Ramireddy, E.; Puram, V.R.R.; Badri, J.; Slavath, S.N.; Bollineni, S.N.; Naik, B.J. Uncovering of natural allelic variants of key yield contributing genes by targeted resequencing in rice (*Oryza sativa* L.). *Sci. Rep.* **2019**, *9*, 8192. [[CrossRef](#)]
74. Abbai, R.; Singh, V.K.; Nachimuthu, V.V.; Sinha, P.; Selvaraj, R.; Vipparla, A.K.; Singh, A.K.; Singh, U.M.; Varshney, R.K.; Kumar, A. Haplotype analysis of key genes governing grain yield and quality traits across 3K RG panel reveals scope for the development of tailor-made rice with enhanced genetic gains. *Plant Biotechnol. J.* **2019**, *17*, 1612–1622. [[CrossRef](#)] [[PubMed](#)]
75. Yan, C.-J.; Yan, S.; Yang, Y.-C.; Zeng, X.-H.; Fang, Y.-W.; Zeng, S.-Y.; Tian, C.-Y.; Sun, Y.-W.; Tang, S.-Z.; Gu, M.-H. Development of gene-tagged markers for quantitative trait loci underlying rice yield components. *Euphytica* **2009**, *169*, 215–226. [[CrossRef](#)]
76. Finn, R.; Griffiths-Jones, S.; Bateman, A. Identifying protein domains with the Pfam database. *CurrProtoc Bioinform.* **2003**, *1*, 2.5.1–2.5.19. [[CrossRef](#)]
77. Marchler-Bauer, A.; Derbyshire, M.K.; Gonzales, N.R.; Lu, S.; Chitsaz, F.; Geer, L.Y.; Geer, R.C.; He, J.; Gwadz, M.; Hurwitz, D.I. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **2015**, *43*, D222–D226. [[CrossRef](#)]
78. Letunic, I.; Doerks, T.; Bork, P. SMART 6: Recent updates and new developments. *Nucleic Acids Res.* **2009**, *37*, D229–D232. [[CrossRef](#)]
79. Kozłowski, L.P. IPC—isoelectric point calculator. *Biol. Direct* **2016**, *11*, 55. [[CrossRef](#)]
80. Nakai, K.; Horton, P. PSORT: A program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* **1999**, *24*, 34–60. [[CrossRef](#)] [[PubMed](#)]
81. Savojardo, C.; Martelli, P.L.; Fariselli, P.; Profiti, G.; Casadio, R. BUSCA: An integrative web server to predict subcellular localization of proteins. *Nucleic Acids Res.* **2018**, *46*, W459–W466. [[CrossRef](#)]
82. Hu, B.; Jin, J.; Guo, A.Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* **2015**, *31*, 1296–1297. [[CrossRef](#)]
83. Bailey, T.L.; Johnson, J.; Grant, C.E.; Noble, W.S. The MEME suite. *Nucleic Acids Res.* **2015**, *43*, W39–W49. [[CrossRef](#)]
84. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845–858. [[CrossRef](#)]
85. Wolfe, D.; Dudek, S.; Ritchie, M.D.; Pendergrass, S.A. Visualizing genomic information across chromosomes with PhenoGram. *BioData Min.* **2013**, *6*, 18. [[CrossRef](#)]
86. Wang, Y.; Tang, H.; DeBarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.-h.; Jin, H.; Marler, B.; Guo, H. MCSanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [[CrossRef](#)]
87. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.; Xia, R. TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* **2020**, *13*, 1194–1202. [[CrossRef](#)]
88. Lescot, M.; Déhais, P.; Thijs, G.; Marchal, K.; Moreau, Y.; Van de Peer, Y.; Rouzé, P.; Rombauts, S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **2002**, *30*, 325–327. [[CrossRef](#)]
89. Du, Z.; Zhou, X.; Ling, Y.; Zhang, Z.; Su, Z. agriGO: A GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* **2010**, *38*, W64–W70. [[CrossRef](#)]

90. Metsalu, T.; Vilo, J. ClustVis: A web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res.* **2015**, *43*, W566–W570. [[CrossRef](#)]
91. Yau, S. Efficiency of alpha-lattice designs in international variety yield trials of barley and wheat. *J. Agric. Sci.* **1997**, *128*, 5–9. [[CrossRef](#)]
92. Murray, M.; Thompson, W. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **1980**, *8*, 4321–4326. [[CrossRef](#)] [[PubMed](#)]
93. Chidambaranathan, P.; Balasubramaniasai, C.; Behura, N.; Purty, M.; Samantaray, S.; Subudhi, H.; Ngangkham, U.; Devanna, B.; Katara, J.L.; Kumar, A. Effects of high temperature on spikelet sterility in rice (*Oryza sativa* L.): Association between molecular markers and allelic phenotypic effect in field condition. *Genet. Resour. Crop Evol.* **2021**, *68*, 1923–1935. [[CrossRef](#)]
94. Chambers, J.M. Linear models. In *Statistical Models in S*; Routledge: Oxfordshire, UK, 2017; pp. 95–144. Available online: <https://www.routledge.com/Statistical-Models-in-S/Chambers-Hastie/p/book/9780412830402> (accessed on 13 July 2023).
95. Babu, B.K.; Meena, V.; Agarwal, V.; Agrawal, P. Population structure and genetic diversity analysis of Indian and exotic rice (*Oryza sativa* L.) accessions using SSR markers. *Mol. Biol. Rep.* **2014**, *41*, 4329–4339. [[CrossRef](#)]
96. Singh, N.; Choudhury, D.R.; Singh, A.K.; Kumar, S.; Srinivasan, K.; Tyagi, R.; Singh, N.; Singh, R. Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PLoS ONE* **2013**, *8*, e84136. [[CrossRef](#)]
97. Nachimuthu, V.V.; Muthurajan, R.; Duraialaguraja, S.; Sivakami, R.; Pandian, B.A.; Ponniah, G.; Gunasekaran, K.; Swaminathan, M.; KK, S.; Sabariappan, R. Analysis of population structure and genetic diversity in rice germplasm using SSR markers: An initiative towards association mapping of agronomic traits in *Oryza sativa*. *Rice* **2015**, *8*, 30. [[CrossRef](#)]

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.