



Article

Phylogenetic, Structural and Functional Evolution of the LHC Gene Family in Plant Species

Yanhong Lan ¹, Yao Song ¹, Fei Zhao ¹, Yu Cao ¹, Dening Luo ², Dairong Qiao ¹, Yi Cao ^{1,*} and Hui Xu ^{1,*}

¹ Microbiology and Metabolic Engineering Key Laboratory of Sichuan Province, Key Laboratory of Bio-Resources and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu 610065, China

² School of Automation, Chengdu University of Information Technology, Chengdu 610255, China

* Correspondence: geneium@scu.edu (Y.C.); xuhui_scu@scu.edu.cn (H.X.)

Abstract: Light-harvesting chlorophyll a/b-binding (LHC) superfamily proteins play a vital role in photosynthesis. Although the physiological and biochemical functions of LHC genes have been well-characterized, the structural evolution and functional differentiation of the products need to be further studied. In this paper, we report the genome-wide identification and phylogenetic analysis of LHC genes in photosynthetic organisms. A total of 1222 non-redundant members of the LHC family were identified from 42 species. According to the phylogenetic clustering of their homologues with *Arabidopsis thaliana*, they can be divided into four subfamilies. In the subsequent evolution of land plants, a whole-genome replication (WGD) event was the driving force for the evolution and expansion of the LHC superfamily, with its copy numbers rapidly increasing in angiosperms. The selection pressure of photosystem II sub-unit S (PsbS) and ferredoxin (FCII) families were higher than other subfamilies. In addition, the transcriptional expression profiles of LHC gene family members in different tissues and their expression patterns under exogenous abiotic stress conditions significantly differed, and the LHC genes are highly expressed in mature leaves, which is consistent with the conclusion that LHC is mainly involved in the capture and transmission of light energy in photosynthesis. According to the expression pattern and copy number of LHC genes in land plants, we propose different evolutionary trajectories in this gene family. This study provides a basis for understanding the molecular evolutionary characteristics and evolution patterns of plant LHCs.

Keywords: divergence; evolution; expansion; phylogenetic analysis; collinearity analysis; whole-genome duplication



Citation: Lan, Y.; Song, Y.; Zhao, F.; Cao, Y.; Luo, D.; Qiao, D.; Cao, Y.; Xu, H. Phylogenetic, Structural and Functional Evolution of the LHC Gene Family in Plant Species. *Int. J. Mol. Sci.* **2023**, *24*, 488. <https://doi.org/10.3390/ijms24010488>

Academic Editor: Richard R.-C. Wang

Received: 8 December 2022

Revised: 13 December 2022

Accepted: 21 December 2022

Published: 28 December 2022



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Light-harvesting chlorophyll a/b binding (LHC) superfamily proteins, as antenna proteins, play an essential role in capturing solar energy, as well as critical roles in plant growth, development, and abiotic stress responses [1]. The LHC superfamily, containing the conserved chlorophyll-binding (CB) domain, is divided into four subfamilies in photosynthetic plants: Lhc, Lil (light-harvesting-like), PsbS, and FCII [2,3]. The Lhc subfamily comprises two groups, named Lhca and Lhcb [4,5]. Similarly, the Lil subfamily is further divided into four subfamilies: SEP (stress-enhanced protein), OHP (one-helix protein), ELIP (early light-induced protein), and Psb33 (photosystem II protein 33) [6,7]. However, The PsbS and FCII subfamilies each include only a single group [8].

Most research on plant LHCs has focused on the Lhc subfamilies Lhca and Lhcb, which play important functions in PSI (photosystem I) and PSII (photosystem II) [9], respectively. The combination of LHC protein with antenna pigments chlorophyll and carotenoids enhances the light capturing ability of plants [10,11]. In *Arabidopsis*, several Lhc members have been studied in terms of capturing solar energy, seed germination, growth, and plant adaptation to environmental changes [12–14]. The overexpression of

the tea *CsLhc* gene in the Arabidopsis Lhcb mutant facilitated chlorophyll accumulation and promoted leaf regreening by increasing expression levels of chlorophyll biosynthesis-related genes [15]. The knockout mutant plants of *AtLhcb6* and *AtLhcb5*, as well as *AtLhcb4*, exhibited significantly lower chlorophyll contents in Arabidopsis [16]. When *AcLhcb3.1/3.2* of kiwifruit were over-expressed in tobacco leaves, the results indicated that *AcLhcb3.1/3.2* had a higher chlorophyll a content and total chlorophyll content than *AcLhcb1.5* [17]. When *AtLhca2* and *AtLhca3* were over-expressed in bacteria, the results indicated that *AtLhca2* had a higher chlorophyll b content than *AtLhca3* [18]. The heterologous over-expression of tomato *LeLhcb2* in tobacco significantly enhanced the ability of transgenic plants to cope with chilling stress tolerance and alleviated photo-oxidation of PSII [19]. Over-expression of kiwifruit's *AcLhcb3.1/3.2* in tobacco leaves significantly increased the content of chlorophyll a [17]. In *Gossypium hirsutum*, down-regulated expression of *GhLhcb2.3* by VIGS (virus-induced gene silencing) significantly decreased the content of chlorophyll a. Additionally, *GhLhcb2.3* and *GhLhcb1.4* exhibited complementary functions during total chlorophyll synthesis in photosynthetic light-harvesting [20]. Although a functional characterization of the LHC family has been carried out, it lacks a more comprehensive phylogenetic, structural and functional evolution of the LHC gene family in plant species.

To date, the LHC superfamily has mainly been studied from the perspective of a single species and family [20,21]. However, the evolutionary origin and expansion of the LHC superfamily across the major lineages of green plants have not been reported systematically. With the availability of increasing numbers of genomes, we currently have the ability to study the LHC gene family from a wider perspective. In this study, BLASTP and the domain architecture of Arabidopsis homologues were considered for plant LHC gene superfamily identification at the whole-genome level. The evolutionary relationships, expansion, and expression patterns of different members of the LHC superfamily were investigated. Furthermore, we performed a functional diversity analysis and positive selection analysis, in order to explore the evolution of their structure and function. Based on the research results, we provide a theoretical basis for further research on the function of LHC genes in plants.

2. Results

2.1. Identification and Distribution of LHC Genes in Plants

To identify putative LHC genes throughout photosynthetic organisms, sequences from 12 Arabidopsis LHC proteins were used as queries in a BLASTp [22] search against 42 species with relatively complete genome annotations, and, as a result, 1222 non-redundant LHC family members and the chromosomal distribution of the LHC genes were identified (Tables 1 and S1). The results indicated that LHC superfamily members were distributed in each selected species, and their numbers varied from species to species. The Lhc and Lil subfamilies had the most members, while FCII had the least (Table 1). The FCII subfamily was found in all species, and there were single copies in algae and bacteria, while the members of the Psbs subfamily were only found in green algae and land plants, the Psb33 subfamily genes were only found in land plants, and the tetraploid *Glycine max* genome contained the largest number of LHC genes; meanwhile, the lower alga, plants, and bacteria only contained one member. In the monocots, *Triticum aestivum* had the second-highest number of LHC genes. We observed that the expansion of the LHC superfamily was concentrated in the angiosperms. In the land plants' genomes, the LHC genes were randomly distributed on chromosomes or scaffolds, and gene numbers varied between 1 and 12 for all chromosomes. In the *Sorghum bicolor* genome, chromosome 2 included the maximum gene number (twelve LHC genes) (Table S1).

Table 1. Distribution of the members of the LHC gene family in 42 plant species.

Taxa	Species	Total	Lhca/b	Lil				PsbS	FCII
				OHP	SEP	ELIP	Psb33		
Eudictos	<i>Arabidopsis thaliana</i>	32	22	2	2	2	1	1	2
	<i>Brassica rapa</i>	44	29	4	3	4	0	2	2
	<i>Gossypium raimondii</i>	44	24	5	2	3	1	2	7
	<i>Theobroma cacao</i>	25	15	2	2	1	2	1	2
	<i>Eucalyptus grandis</i>	30	17	5	3	2	0	1	2
	<i>Malus domestica</i>	42	26	4	2	3	1	2	4
	<i>Glycine max</i>	62	34	7	7	3	1	3	7
	<i>Populus trichocarpa</i>	36	22	3	3	3	1	1	3
	<i>Vitis vinifera</i>	31	20	2	2	3	1	1	2
	<i>Solanum lycopersicum</i>	40	24	3	2	4	1	1	5
	<i>Chenopodium quinoa</i>	49	29	4	4	2	2	2	6
Monocots	<i>Zostera marina</i>	32	22	1	2	2	1	1	3
	<i>Phoenix dactylifera</i>	33	21	3	3	2	0	1	3
	<i>Triticum aestivum</i>	58	31	6	3	3	6	3	6
	<i>Hordeum vulgare</i>	30	22	2	1	1	1	1	2
	<i>Zea mays</i>	37	25	3	2	2	1	1	3
	<i>Sorghum bicolor</i>	29	17	3	2	1	1	1	4
	<i>Oryza sativa</i>	25	14	3	1	1	2	2	2
Basal angiosperms	<i>Amborella trichopoda</i>	27	15	2	3	3	0	1	3
Ferns	<i>Adiantum capillsveneris</i>	42	35	2	0	1	0	1	3
	<i>Ceratopteris richardii</i>	54	40	4	1	2	0	2	5
Lycophytes	<i>Selaginella moellendorffii</i>	25	16	1	0	2	0	2	4
Mosses	<i>Sphagnum magellanicum</i>	53	37	5	0	4	0	4	3
	<i>Ceratodon purpureus</i>	41	34	2	0	3	0	1	1
Liverworts	<i>Marchantia polymorpha</i>	54	45	2	0	2	0	3	2
	<i>Physcomitrium patens</i>	51	40	2	0	4	0	1	4
Streptophyta	<i>Marchantia paleacea</i>	56	47	2	0	2	1	3	1
	<i>Chara braunii</i>	16	12	1	0	0	0	2	1
Chlorophyta	<i>Chlamydomonas reinhardtii</i>	28	23	1	0	1	0	2	1
	<i>Volvox carter</i>	30	27	0	1	0	0	1	1
	<i>Ostreococcus</i> sp	11	9	0	0	1	0	0	1
	<i>Coccomyxa subellipsoidea</i>	22	20	0	0	1	0	0	1
	<i>Dunaliella salina</i>	16	13	0	0	1	0	1	1
Rhodophyta	<i>Cyanidioschyzon merolae</i>	1	0	0	0	0	0	0	1
	<i>Ectocarpus siliculosus</i>	9	8	0	0	0	0	0	1
Bacteria	<i>Bathycoccus prasinos</i>	1	0	0	0	0	0	0	1
	<i>Panttoea ananatis</i>	1	0	0	0	0	0	0	1
	<i>Erwinia gerundensis</i>	1	0	0	0	0	0	0	1
	<i>Halomonas sulfidaeris</i>	1	0	0	0	0	0	0	1
	<i>Cronobacter sakazakii</i>	1	0	0	0	0	0	0	1
	<i>Geobacter sulfurreducens</i>	1	0	0	0	0	0	0	1
	<i>Gloeobacter violaceus</i>	1	0	0	0	0	0	0	1

2.2. Phylogenetic Analysis and Conserved Motifs of the Plant LHC Gene Family Members

To keep the analysis reliable, an independent unrooted phylogenetic tree was constructed for each subfamily by using LHC full-length proteins to explore the phylogenetic relationship and evolutionary pattern of photosynthetic organisms' LHC genes. Consistent with the classification of AtLHC proteins, all LHC proteins were grouped into four distinct subfamilies, namely Lhc, Lil, PsbS, and FCII. The results revealed that Lhc, Lil, PsbS, and FCII proteins were clustered into 2, 4, 1, and 1 groups, respectively (Figures 1–7, Table 1). To infer structural variations and a possible functional divergence, coding sequences of the LHC genes were analyzed by MEME. The results showed that 24 conserved motifs were obtained, where the type and quantity of motifs were highly variable among subfamilies or groups, and the FCII subfamily contained the highest number of motifs compared to PsbS, Lhc, and Lil (Table S2, Figure S1). The protein motifs in each sub-group were highly conserved, and some differences were found between different sub-groups. Except for the

Psb33 subfamily, motif 3 was widely distributed in all the LHC proteins and located in the conserved CB domain (Figures 1–7 and S2). Some specific motifs only existed in specific subfamilies. For example, motif 15 was only present in the ELIP subfamily, motif 14 was only present in the PsbS subfamily, and the OHP subfamily only contained motif 18. The LHC subfamilies contained specific motifs, suggesting that they also had specific functions.

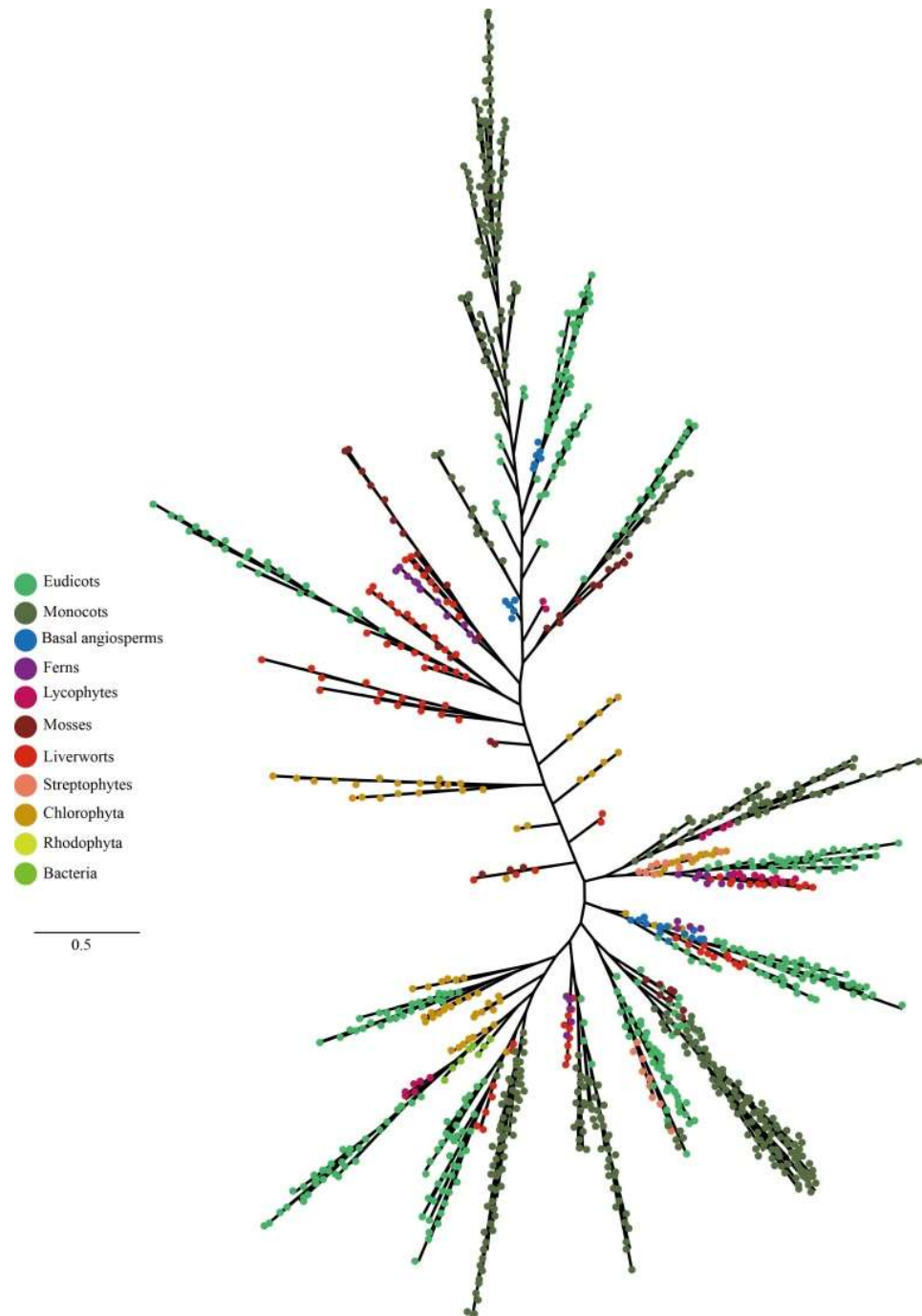


Figure 1. Phylogenetic analysis of Lhc gene subfamily from 42 species.

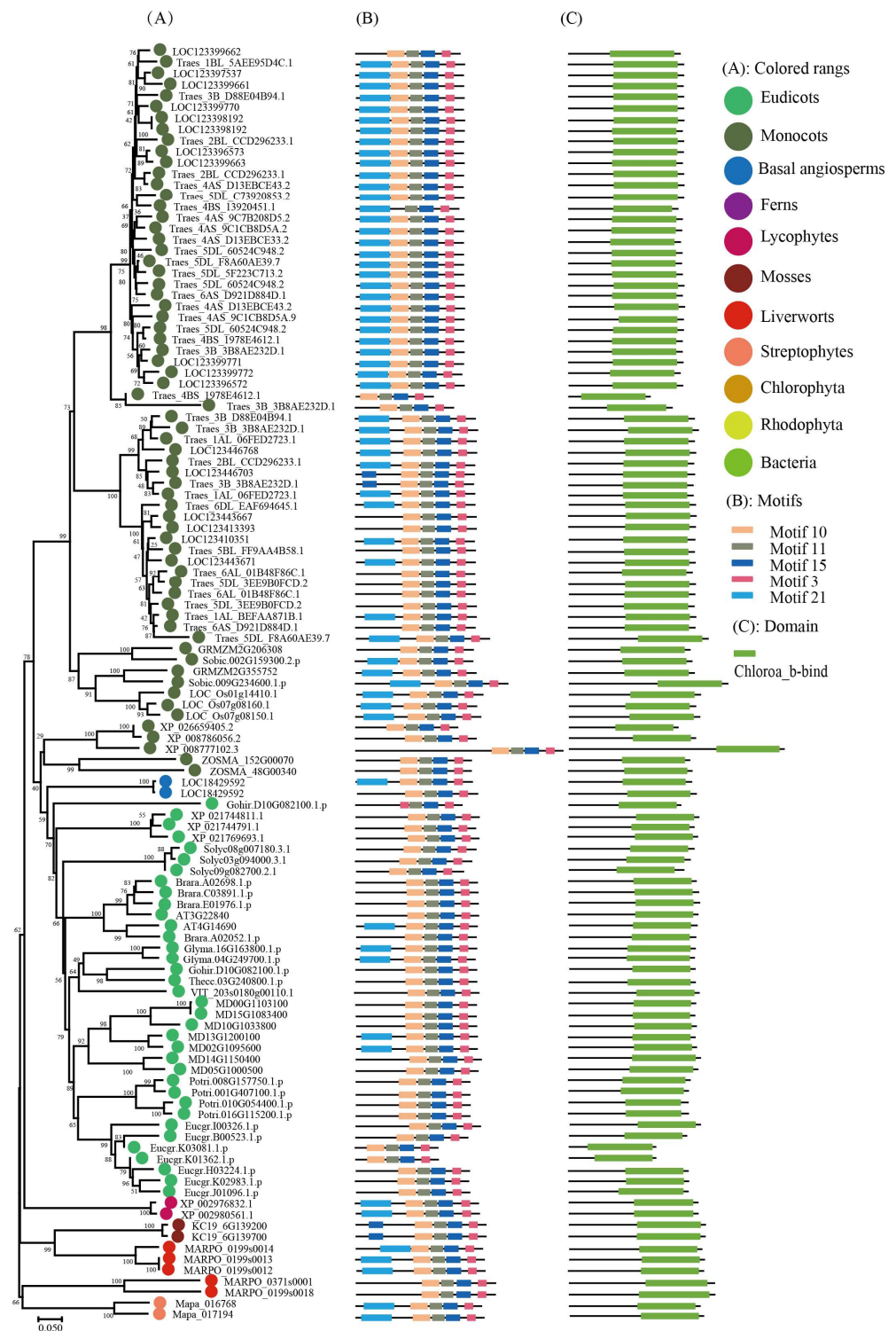


Figure 2. Phylogenetic evolutionary tree, protein motifs, and domains of ELIP subfamily members: (A) Unrooted phylogenetic tree of the ELIP subfamily; (B) motifs of the ELIP proteins. The different colors indicate five motifs (left panel); (C) protein domains of the ELIP genes.

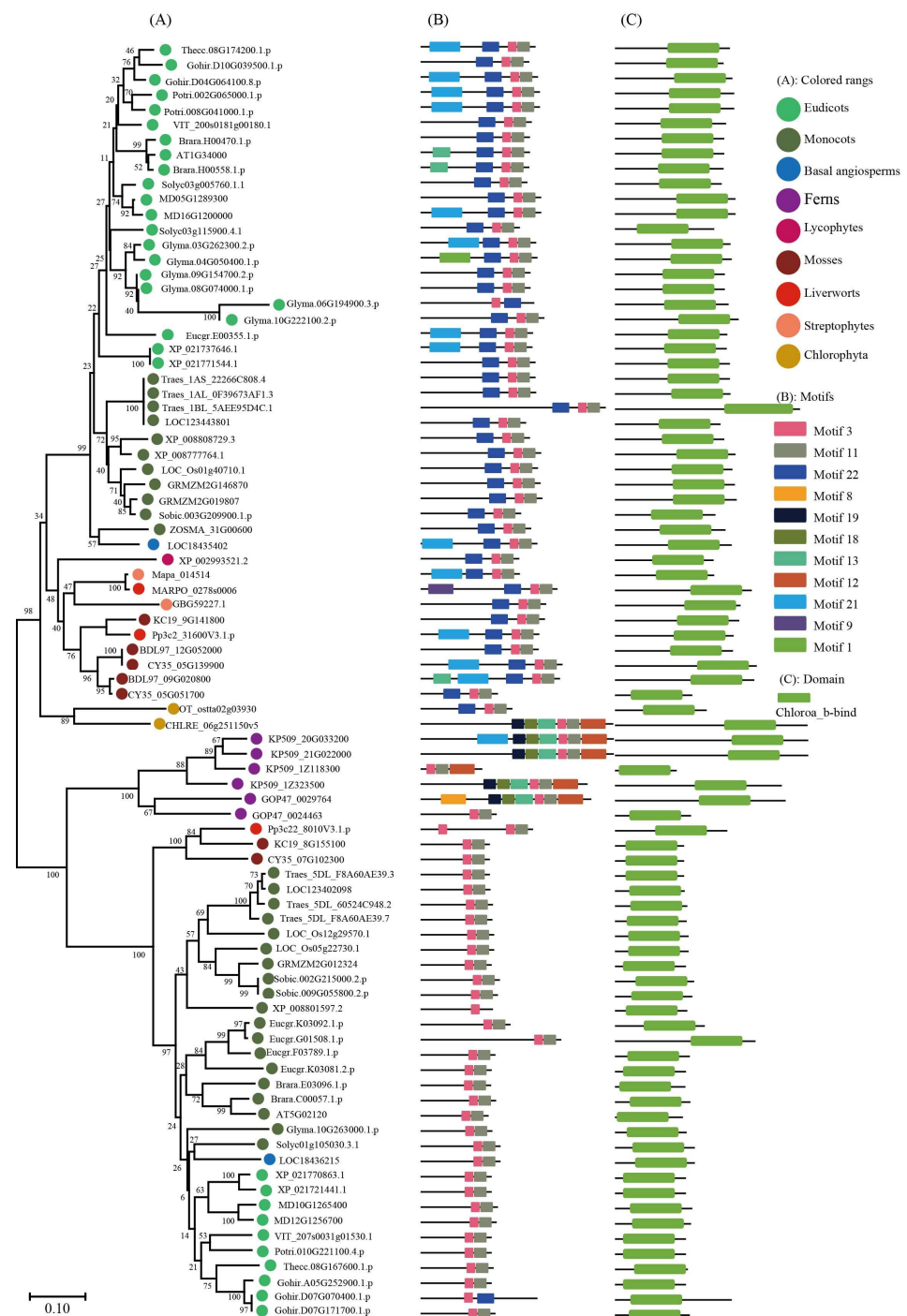


Figure 3. Phylogenetic evolutionary tree, protein motifs, and domains of the OHP subfamily members: (A) Unrooted phylogenetic tree of the OHP subfamily; (B) motifs of the OHP proteins. The different colors indicate 11 motifs (left panel); and (C) protein domains of the OHP genes.

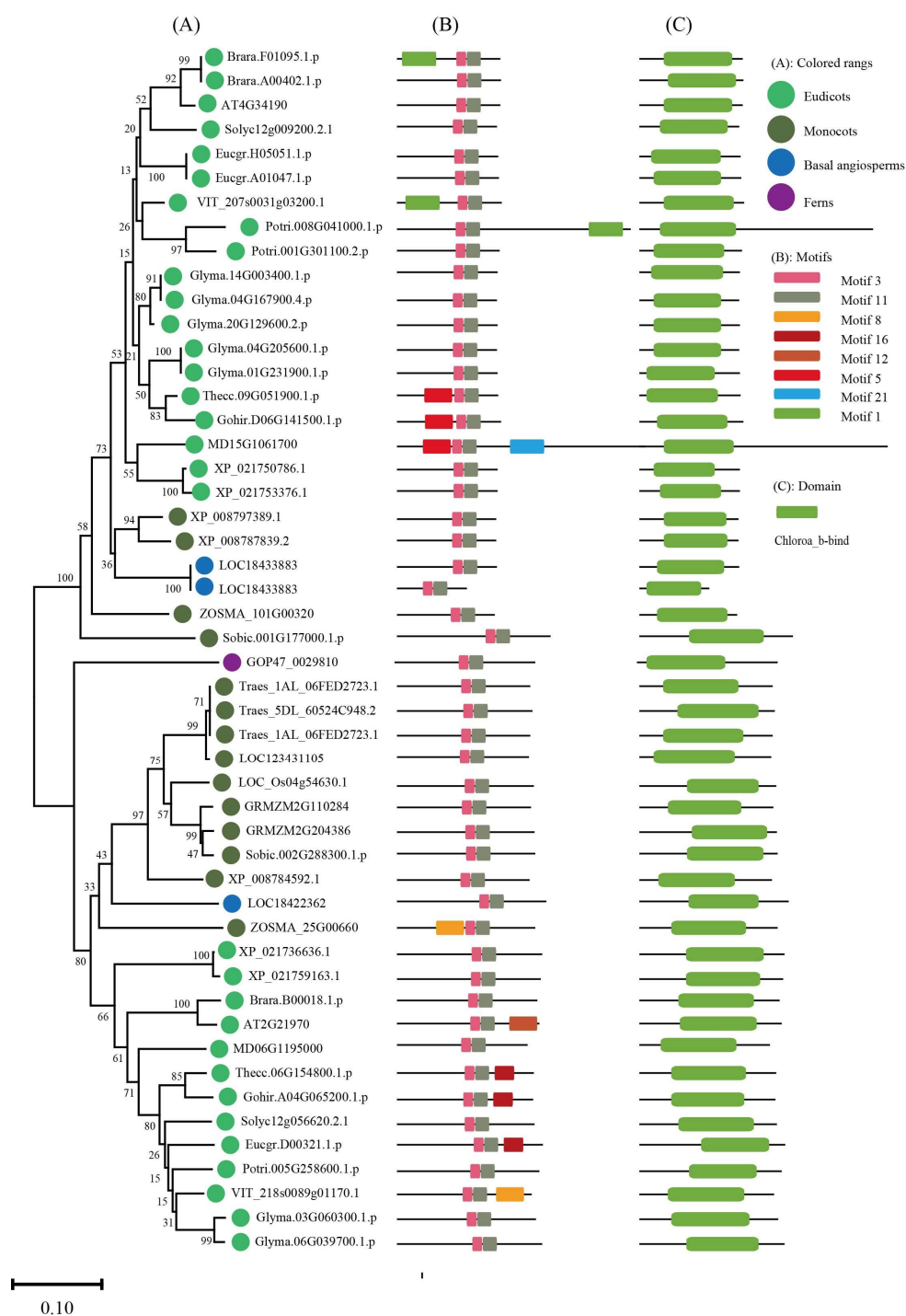


Figure 4. Phylogenetic evolutionary tree, protein motifs, and domains of the SEP subfamily members: (A) Unrooted phylogenetic tree of the SEP subfamily; (B) motifs of the SEP proteins. The different colors indicate eight motifs (left panel); (C) protein domains of the SEP genes.

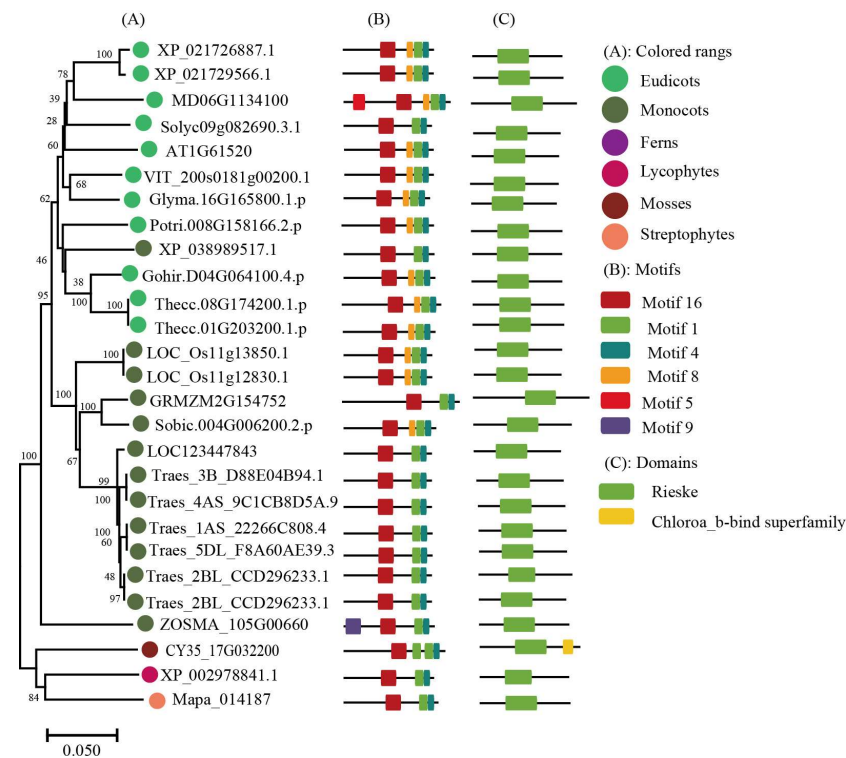


Figure 5. Phylogenetic evolutionary tree, protein motifs, and domains of the Psb33 subfamily members: (A) Unrooted phylogenetic tree of the Psb33 subfamily; (B) motifs of the Psb33 proteins. The different colors indicate six motifs (left panel); (C) protein domains of the Psb33 genes.

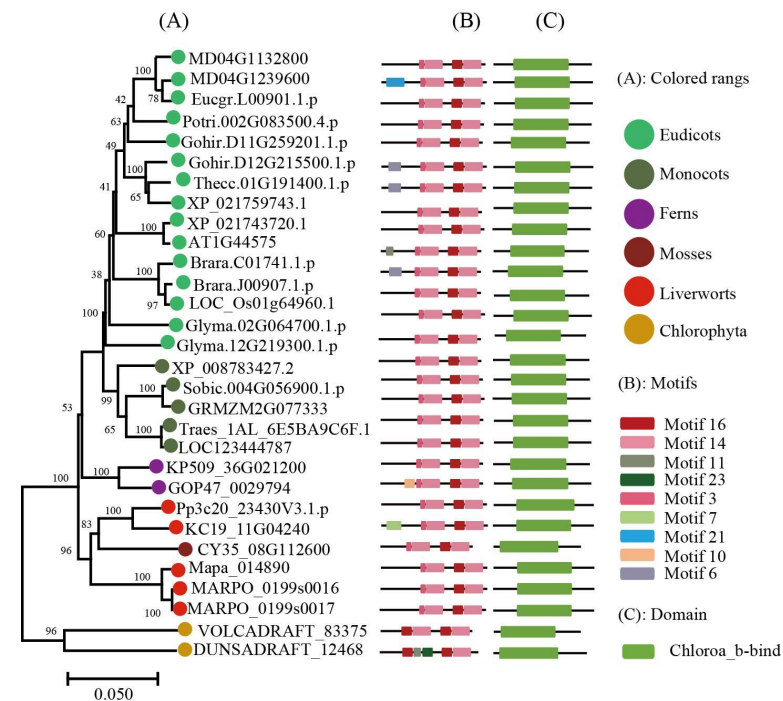


Figure 6. Phylogenetic evolutionary tree, protein motifs, and domains of the PsbS subfamily members: (A) Unrooted phylogenetic tree of the PsbS subfamily; (B) motifs of the PsbS proteins. The different colors indicate nine motifs (left panel); (C) protein domains of the PsbS genes.

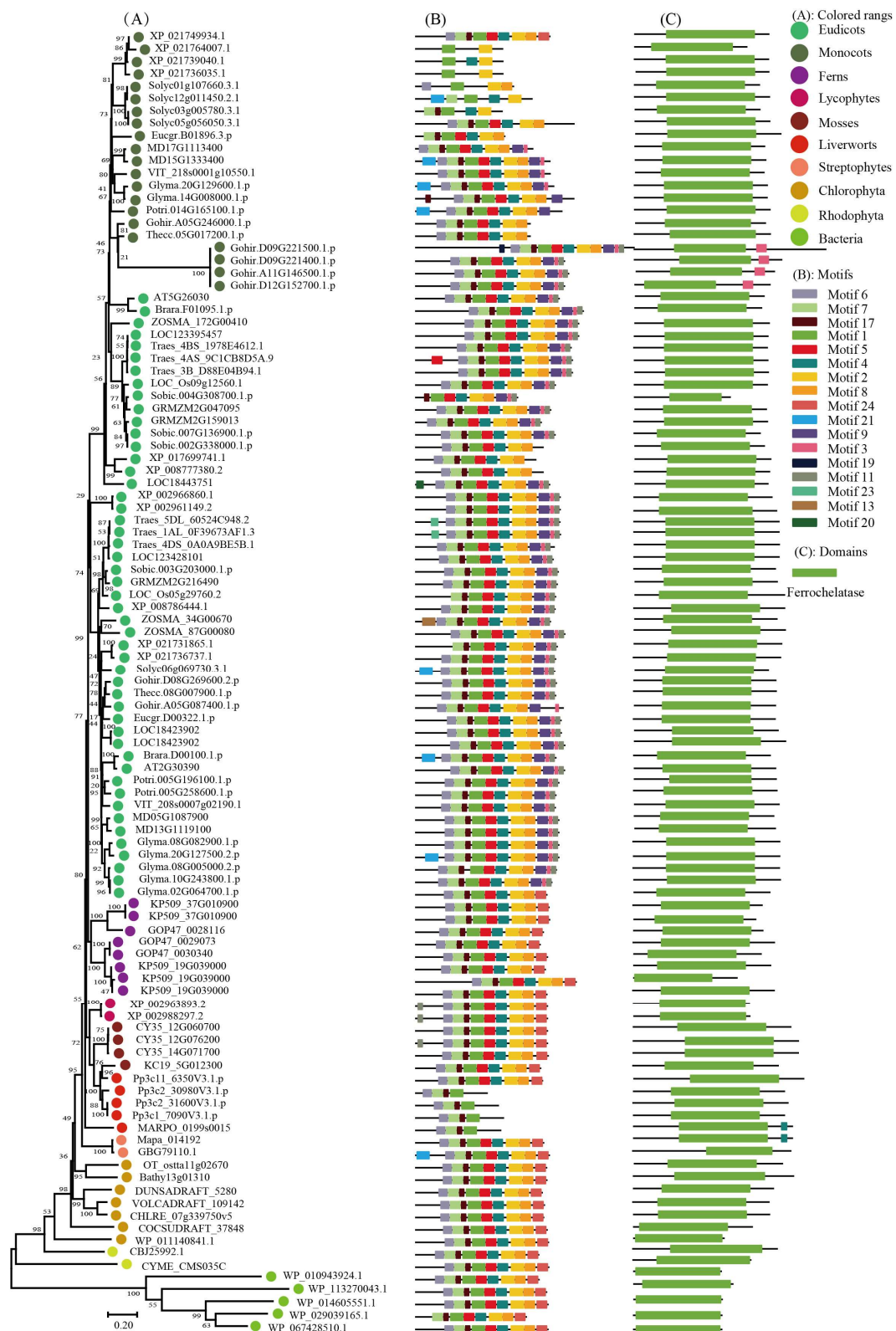


Figure 7. Phylogenetic evolutionary tree, protein motifs, and domains of the FCII subfamily members: (A) Unrooted phylogenetic tree of the FCII subfamily; (B) motifs of the FCII proteins. The different colors indicate seventeen motifs (left panel); (C) protein domains of the FCII genes.

2.3. Expansion Pattern of LHC Genes and Collinearity Analysis

To further study the duplication events in the evolutionary history of the LHC gene superfamily, *G. max* was used as a model, as it had the largest number of members from this superfamily. According to the species' evolutionary relationships, we analyzed the genome collinearity among *G. max*, *Populus trichocarpa*, *Vitis vinifera*, and *Amborella trichopoda*, and found that the number of LHC genes increased following the WGDs in these species, including the four subfamilies in the LHC superfamily (Figure 8A).

A total of 31/20 LHC genes in *G. max*/*P. trichocarpa*, 25/17 SPX genes in *P. trichocarpa*/*V. vinifera*, and 14/12 LHC genes in *V. vinifera*/*A. trichopoda* had a high degree of collinearity. The collinear regions for the PsbS and FCII subfamilies were only present in *G. max*/*P. trichocarpa*, the collinear regions for Lil and Lhc subfamilies were present in *G. max*/*P. trichocarpa* and *P. trichocarpa*/*V. vinifera*, and *V. vinifera*/*A. trichopoda* only presented an Lhc collinearity relationship (Figure 8A). However, there were 30 replication events in the *G. max*/*G. max* genome (Figure 8B).

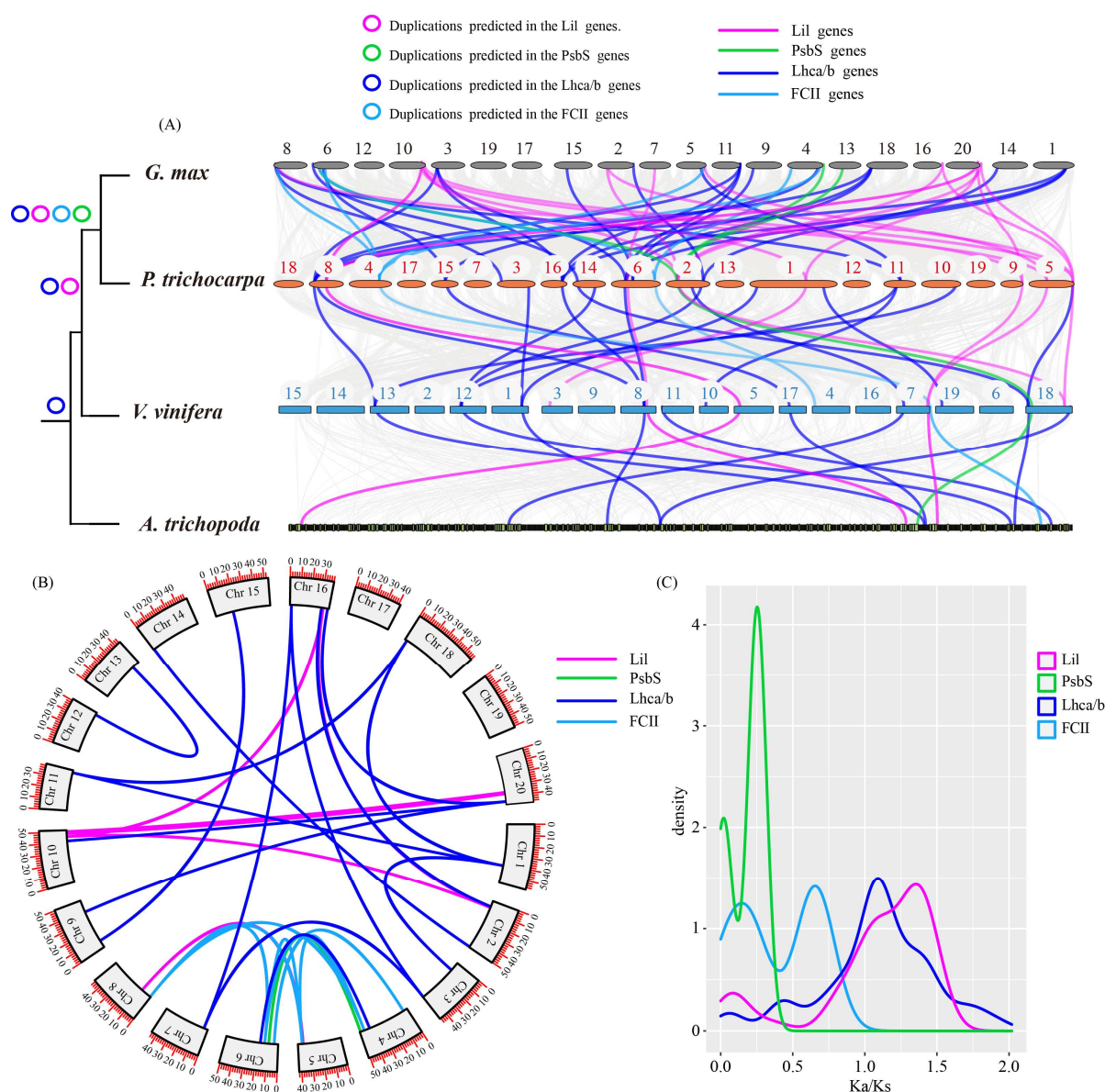


Figure 8. Duplication events of LHC genes in *G. max*. (A) Syntenic relationships of LHC genes among *Glycine max*, *Populus trichocarpa*, *Vitis vinifera*, and *Amborella trichopoda*; (B) syntenic relationships of LHC genes in *Glycine max*/*Glycine max*; (C) distribution of Ka/Ks values among LHC genes in *Glycine max*.

To further study the conservation of the structure and function of the soybean LHC gene superfamily, the selection pressure was systematically analyzed. The results indicated that there were great differences in selection pressures among different sub-families of the soybean LHCs' superfamily. In addition, the Ks value of duplicated gene pairs varied from 0.02 to 1.67 (Figure 8C), suggesting that duplicated pairs have different evolution rates. The higher Ks value of dispersed duplication when compared to whole-genome duplication (WGD) suggested that it may have happened earlier than the WGD event. Most duplicated gene pairs underwent WGD, indicating that WGD played a vital role in expanding LHCs in soybeans. The Ka/Ks of the PsbS gene family was the lowest (the average was 0.175, less than 1), followed by FCII (Ka/Ks was 0.384), suggesting that the selection pressure on these two subfamilies was high. The Ka/Ks ratios for LHC and Lil were higher (an average of 1.077 and 1.067, respectively), suggesting that the genes in these two gene families may have experienced positive selection (Figure 8C).

2.4. Expression Patterns of LHCs in Plants

To further study the potential functions of the LHC superfamily, the tissue-specific and abiotic stress expression patterns of several representative LHC genes from *Oryza sativa*, model *A. thaliana*, and *G. max* were analyzed, using microarray data of LHC genes from the BAR database across different tissue types (e.g., root, leaf, flower, and SAM) and in response to abiotic stresses (e.g., drought, salt, light, heat, and cold).

The results indicated that 18 LHC genes were expressed in leaf tissues, with a strong tissue-specific expression pattern. Lil genes were highly expressed in leaves, such as Glyma.16G163800.1.p, Glyma.04G249700.1.p, Glyma.03G262300.2.p, LOC_Os01g41710, and LOC_Os01g64960, suggesting that these genes have important functions in leaves. For the Lil gene in soybeans (Glyma.03G262300.2.p), the highest expression was detected in flowers, and the lowest expression was in the root, suggesting that this gene may play a key role in flower development (Figure 9A–C). *A. thaliana* Lil genes (AT4G14690 and AT3G22840) were highly expressed in shoot and root tissues under cold stress, and their expression levels were relatively high under salt and UV-B stress in shoot tissues (Figure 9D–E); however, the expression of these two genes was low under drought and heat stress. For LHC genes in rice, the expression was low under abiotic stress, among which LOC_Os07g08150, LOC_Os01g14410, and LOC_Os04g59440 had the lowest expression under cold stress, suggesting that these genes were functional in the plant response to these stimuli. LOC_Os01g41710 was expressed highly in tissues, with no significant change under abiotic stress.

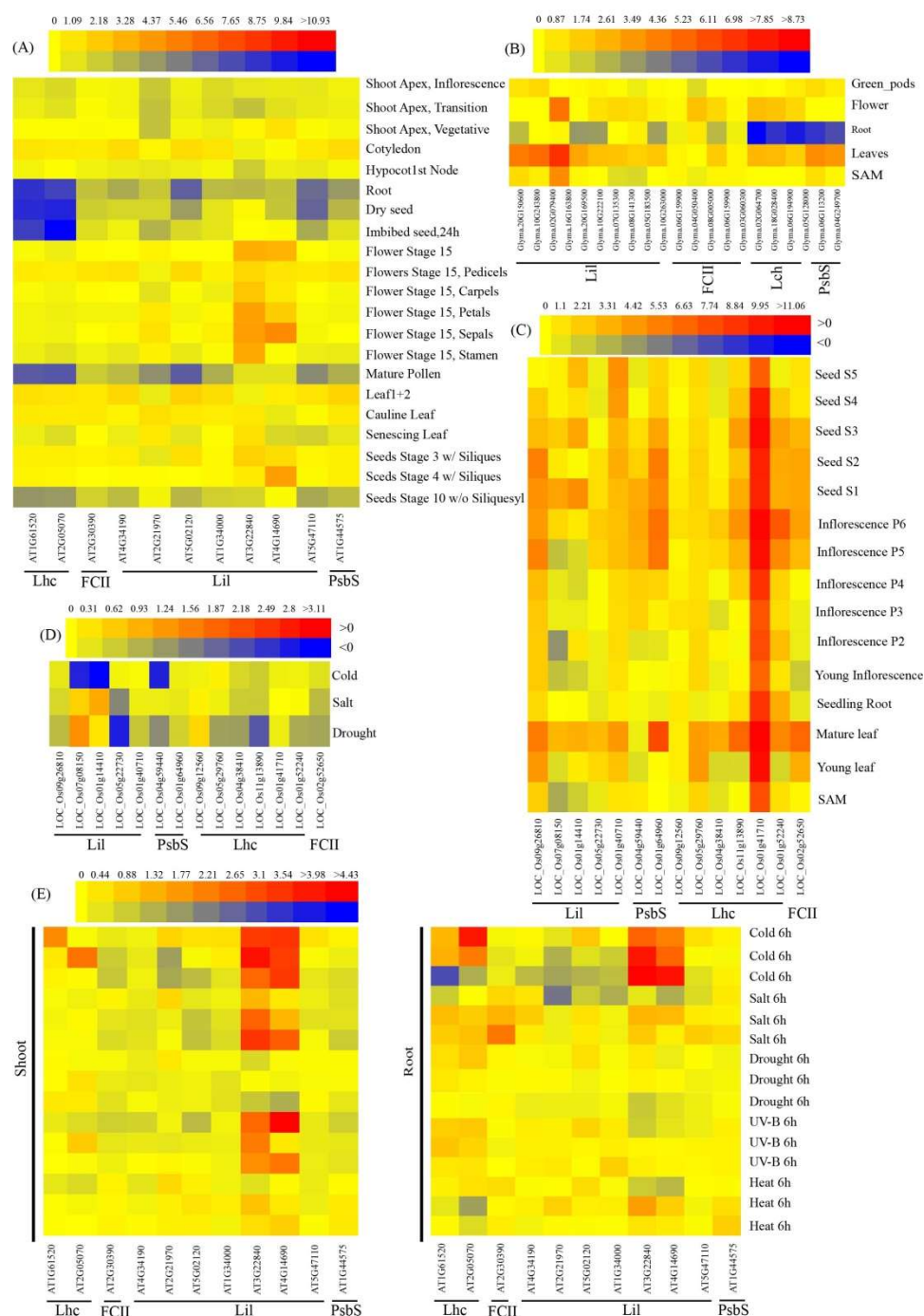


Figure 9. Expression profiles of LHC genes in representative species: (A–C) Spatial and temporal expressions of LHC genes from *Arabidopsis*, *Glycine max* (soybean), and *Oryza sativa* (rice); (D,E) Expression of LHC genes under different stress treatments.

3. Discussion

Light-harvesting chlorophyll a/b binding protein (LHC) is the most abundant protein complex on the thylakoid membrane, and it plays an important role in plant growth and development (e.g., capturing and transforming light during photosynthesis and having an association with oxidative stress) [23–25]. Members of the LHC superfamily have been identified and characterized in land plants, but there has still no systematic and comprehensive report on the photosynthetic LHC superfamily organism. In the present study, we performed a genome-wide analysis of LHC gene family members in 42 species. Phylogenetic analysis indicated that the LHC superfamily was mainly divided into four

subfamilies (Lhc, Lil, PsbS, and FCII), where the genetic origins of these subfamilies were inconsistent and each type had experienced a long-term independent evolution process. The CB domain was highly conserved in the LHC family in different plants [26]. The topology of the evolutionary tree indicated that LHC genes from all land plants are generally conserved and that they share a similar pattern in terms of protein domain and motifs. Bacteria and algal LHC were distantly associated with those of land plants, as supported by differences in the number and organization of conserved protein motifs.

Researchers have reported that the duplication, loss, pseudogene, and other changes of gene family members during the process of evolution are common phenomena [27]. Genomic replication often results in an expansion of the gene family [28–30]. In this study, we performed a synteny analysis of Lhc superfamily genes in *G. max*, *P. trichocarpa*, *V. vinifera*, and *A. trichopoda* (Figure 8). Genome-wide replication analysis showed that LHC gene family members experienced multiple WGD events between genomes. However, the difference in the number of LHC members in the superfamily and subfamilies of the plants suggested that they may have experienced different evolution patterns. In addition, an analysis of the selection pressure in *G. max* indicated that the LHC gene family had different evolutionary directions due to different selection pressures during its evolution, thus forming a diverse population. The average gene Ka/Ks ratio for the Lhc and FCII subfamilies was more than 1, suggesting that the gene family was greatly affected by positive selection, while those of the other gene families were all under 1, suggesting that the gene family has been greatly affected by negative selection [17]. After gene fragment replication, a strong purification selection was experienced. These results indicate that the genes of the PsbS and Lil subfamilies are relatively conservative in structure.

The expression analysis demonstrated that multiple LHC genes were specifically expressed in different tissues and stresses; however, the differentiation of expression patterns was mainly concentrated among different subfamily genes. With an increase in the number of gene family members in the species, the differentiation of expression patterns was more obvious [31]. The expression of soybean LHC family genes was the highest in mature leaves, followed by flowers, and it was the lowest in roots; this distribution was consistent with the conclusion that LHC—as a light-harvesting pigment-binding protein gene—is expressed in the thylakoids of chloroplasts, the light reaction site of photosynthesis. Photosynthesis mainly occurs in the leaves of plants [32]. A large amount of light energy is absorbed by the leaves, while the root hardly receives light. LOC_Os01g41710 is highly expressed in the mature leaves of rice, but it is hardly expressed in roots, suggesting that this gene may be necessary for plant photosynthesis. Analysis of the response of LHC family members to abiotic stress indicated that the genes of LHC family members in Arabidopsis were expressed at low levels under stress; therefore, stress treatment would inhibit plant capture light efficiency and affect photosynthesis. Meanwhile, rice LHC family members were highly expressed under cold stress, suggesting that this gene has a functional role in the plant response to such stimuli.

Overall, 1222 LHC genes from 42 species were comprehensively and systematically for the first time, to the best of our knowledge. The LHC superfamily is divided into four subfamilies, Lhc, Lil, PsbS, and FCII, according to the phylogeny of its homologues with *A. thaliana*. The results of the whole-genome replication event indicated that the molecular evolution of LHC has been characterized by large-scale indirect amplification and gene loss. The number of LHC gene members was the largest in soybeans, and the molecular evolution of the LHC family in the genome showed indirect amplification and gene loss. LHC genes were purified and screened, but a strong selectivity was observed in the Lhc gene subfamily. The tissue and stress-condition-specific expression patterns indicated that LHCs play important roles in multiple biological processes. This study provided novel viewpoints regarding the functional evolution and expansion of the LHC gene family in various important species.

4. Methods and Materials

4.1. Identification of LHC Gene Family Members

In order to explore the evolutionary divergence and expansion of the LHC superfamily across photosynthetic organisms, 42 representative species with relatively complete annotated genome data were used as research subjects for the assessment of the LHCs' taxonomy and phylogenetic relationships. The genomic data were obtained from the Plant JGI Database phytozome v12.1 (<https://phytozome.jgi.doe.gov/pz/portal.html> (accessed on 6 October 2022)) [33] and National Center for Biotechnology Information Genome database (<https://ncbi.nlm.nih.gov/genome/> (accessed on 22 December 2022)). Then, twelve previously identified and characterized LHC proteins from *Arabidopsis thaliana* [21] (GenBank: Lhca1, AT3G54890; Lhca2, AT3G61470; Lhcb1, AT1G29920; Lhcb3, AT5G54270; ELIP1, AT3G22840; ELIP2, AT4G14690; SEP1, AT4G34190; SEP2, AT2G21970; OHP1, AT5G02120; OHP2, AT1G34000; Psb33, AT1G71500; PsbS, AT1G44575) were employed to construct a query protein set. The candidate LHCs were identified using BLASTp (BLAST plus: architecture and applications) [22], with a cutoff score of ≥ 100 and an e-value of $\leq 1 \times 10^{-10}$. The Conserved Domain Database (CDD) (<https://ncbi.nlm.nih.gov/cdd> (accessed on 8 January 2020)) [34], Pfam [35], and SMART (<http://smart.embl-heidelberg.de> (accessed on 26 October 2020)) [36] were used to confirm the identity of LHC proteins containing a chlorophyll-binding (CB) domain. The location of all species' LHC genes was extracted from the corresponding GFF file using an in-house Perl script.

4.2. Phylogeny Analysis and Identification of Conserved Motifs

Multiple sequence alignments of the amino acid sequences of identified LHC proteins were performed using ClustalW and ClustalX with default parameters [37]. Unrooted phylogenetic trees were constructed based on the neighbor-joining (NJ) method with a JTT (Jones–Taylor–Thornton) model and 1000 bootstrap replicates, using the MEGA software (v7.0.26, download from <https://www.megasoftware.net/> (accessed on 9 August 2021)) [38]. The phylogenetic trees were visualized using FigTree v1.4.3 (<http://tree.bio.ed.ac.uk/> (accessed on 13 June 2019)). To obtain insight into the divergence and function of the LHC proteins, the twelve most conserved motifs were identified using the MEME tool (<http://meme-suite.org/tools/meme> (accessed on 17 December 2022)) [39], and the results were imported to the TBtools [40] to visualize protein structures.

4.3. Microarray Based Expression Analysis

The original RNA-Seq expression data of LHC family members from different tissues and abiotic stress were obtained from BARPB (Bio-Analytic Resource for Plant Biology <http://bar.utoronto.ca/> (accessed on 7 January 2022)). Subsequently, the BAR HeatMapper Plus tool (http://bar.utoronto.ca/ntools/cgi-bin/ntools_heatmapper_plus.cgi (accessed on 7 January 2022)) was used to analyze the expression levels of selected plant LHC genes.

4.4. Gene Duplication and Synteny Analysis

The MCScanX [41] software was used to investigate the synteny, collinearity, and gene duplication patterns of plant LHCs with default parameters. TBtools [40] was used to calculate the synonymous (Ks) and nonsynonymous (Ka) mutation rates of the duplicated LHC gene pairs [42], and the Ka/Ks ratio was calculated to measure the selection pressure (Ka/Ks < 1, purifying; or Ka/Ks > 1, positive) during evolution [43]. The syntenic relationships between identified genes were visualized using TBtools.

Supplementary Materials: The supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms24010488/s1>.

Author Contributions: Conceptualization, Y.L., H.X. and Y.C. (Yi Cao); Investigation, Y.L., Y.S. and F.Z.; Visualization, Y.L., Y.S.; Writing-original draft, Y.L. and H.X.; Writing-review editing, Y.L., D.Q., H.X. and Y.C. (Yu Cao). D.L. participated in the visualization of the article. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (32171473, 32071479, 32271535, and 31670078), Fundamental Research Funds for the Central Universities (No.2018SCUH0072); Sichuan Science and Technology Bureau (2021YJ0024, 2022NSFSC0243, and 2022NSFSC0119), Chengdu Science and Technology Bureau (2020-GH02-00037-HZ), and the Sichuan Provincial S&T Projects, China (No. 2022NSFSC0934).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We thank Carlos Flores for his critical review.

Conflicts of Interest: The authors have no financial conflict of interest to declare.

References

- Jiang, Q.; Xu, Z.S.; Wang, F.; Li, M.Y.; Ma, J.; Xiong, A.S. Effects of abiotic stresses on the expression of Lhcb1 gene and photosynthesis of *Oenanthis javanica* and *Apium graveolens*. *Biol. Plantarum*. **2014**, *58*, 256–264. [\[CrossRef\]](#)
- Klimmek, F.; Sjödin, A.; Noutsos, C.; Leister, D.; Jansson, S. Abundantly and rarely expressed Lhc protein genes exhibit distinct regulation patterns in plants. *Plant Physiol.* **2006**, *140*, 793–804. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zou, Z.; Yang, J. Genomics analysis of the light-harvesting chlorophyll a/b-binding (Lhc) superfamily in cassava (*Manihot esculenta* Crantz). *Gene* **2019**, *702*, 171–181. [\[CrossRef\]](#) [\[PubMed\]](#)
- Jansson, S. A guide to the Lhc genes and their relatives in Arabidopsis/IT>. *Trends Plant Sci.* **1999**, *4*, 236–240. [\[CrossRef\]](#)
- Green, B.R.; Pichersky, E.; Kloppstech, K. Chlorophyll a/b-binding proteins: An extended family. *Trends Biochem. Sci.* **1991**, *16*, 181–186. [\[CrossRef\]](#)
- Engelken, J.; Brinkmann, H.; Adamska, I. Taxonomic distribution and origins of the extended LHC (light-harvesting complex) antenna protein superfamily. *BMC Evol. Biol.* **2010**, *10*, 233. [\[CrossRef\]](#)
- Zou, Z. Mining gene families in the castor bean genome. In *The Castor Bean Genome*; Rabinowicz, P., Kole, C., Eds.; Springer: Cham, Switzerland, 2018; pp. 135–173.
- Zhao, Y.; Kong, H.; Guo, Y.; Zou, Z. Light-harvesting chlorophyll a/b-binding protein-coding genes in jatropha and the comparison with castor, cassava and Arabidopsis. *PeerJ*. **2020**, *8*, e8465. [\[CrossRef\]](#)
- Neilson, A.; Durnford, G. Structural and functional diversification of the light-harvesting complexes in photosynthetic eukaryotes. *Photosynth Res.* **2010**, *106*, 57–71. [\[CrossRef\]](#)
- Mogen, K.; Eide, J.; Duysen, M.; Eskins, K. Chloramphenicol Stimulates the Accumulation of Light-Harvesting Chlorophyll a/b Protein II by Affecting Posttranscriptional Events in the Chlorina CD3 Mutant Wheat. *Plant Physiol.* **1990**, *92*, 1233–1240. [\[CrossRef\]](#)
- Passarini, F.; Xu, P.; Caffarri, S.; Hille, J.; Croce, R. Towards in vivo mutation analysis: Knock-out of specific chlorophylls bound to the light-harvesting complexes of *Arabidopsis thaliana*-the case of CP24 (Lhcb6). *Biochim. Biophys. Acta* **2014**, *1837*, 1500–1506. [\[CrossRef\]](#)
- Xu, H.; Liu, R.; Yan, L.; Liu, Q.; Jiang, C.; Shen, Y.; Wang, F.; Zhang, P. Light-harvesting chlorophyll a/b-binding proteins are required for stomatal response to abscisic acid in Arabidopsis. *J. Exp. Bot.* **2012**, *63*, 1095–1106. [\[CrossRef\]](#) [\[PubMed\]](#)
- Wang, S.-S.; Song, Z.-B.; Sun, Z.; Zhang, J.; Mei, Y.; Nian, H.-J.; Li, K.-Z.; Chen, L.-M. Effects of Formaldehyde Stress on Physiological Characteristics and Gene Expression Associated with Photosynthesis in *Arabidopsis thaliana*. *Plant Mol. Biol.* **2012**, *30*, 1291–1302. [\[CrossRef\]](#)
- Peterson, R.B.; Schultes, N.P. Light-harvesting complex B7 shifts the irradiance response of photosynthetic light-harvesting regulation in leaves of *Arabidopsis thaliana*. *J. Plant Physiol.* **2014**, *171*, 311–318. [\[CrossRef\]](#) [\[PubMed\]](#)
- Chen, X.; Li, J.; Yu, Y.; Kou, X.; Periakaruppan, R.; Chen, X.; Li, X. STAY-GREEN and light-harvesting complex II chlorophyll a/b binding protein are involved in albinism of a novel albino tea germplasm ‘Huabai 1’. *Sci. Hortic.* **2022**, *293*, 110653. [\[CrossRef\]](#)
- De Bianchi, S.; Betterle, N.; Kouril, R.; Cazzaniga, S.; Boekema, E.; Bassi, R.; Dall’Osto, L. Arabidopsis mutants deleted in the light-harvesting protein Lhcb4 have a disrupted photosystem II macrostructure and are defective in photoprotection. *Plant Cell* **2011**, *23*, 2659–2679. [\[CrossRef\]](#) [\[PubMed\]](#)
- Luo, J.; Abid, M.; Tu, J.; Gao, P.; Wang, Z.; Huang, H. Genome-Wide Identification of the LHC Gene Family in Kiwifruit and Regulatory Role of AcLhcb3.1/3.2 for Chlorophyll a Content. *Int. J. Mol. Sci.* **2022**, *23*, 6528. [\[CrossRef\]](#) [\[PubMed\]](#)
- Castelletti, S.; Morosinotto, T.; Robert, B.; Caffarri, S.; Bassi, R.; Croce, R. Recombinant Lhca2 and Lhca3 subunits of the photosystem I antenna system. *Biochemistry* **2003**, *42*, 4226–4234. [\[CrossRef\]](#)
- Deng, S.; Kong, Y.; Zhou, B.; Zhang, S.; Yue, M.; Meng, W. Heterology expression of the tomato LeLhcb2 gene confers elevated tolerance to chilling stress in transgenic tobacco. *Plant Physiol. Biochem.* **2014**, *80*, 318–327. [\[CrossRef\]](#)
- Zhang, Q.; Ma, C.; Wang, X.; Ma, Q.; Fan, S.; Zhang, C. Genome-wide identification of the light-harvesting chlorophyll a/b binding (Lhc) family in *Gossypium hirsutum* reveals the influence of GhLhcb2.3 on chlorophyll a synthesis. *Plant Biol.* **2021**, *23*, 831–842. [\[CrossRef\]](#)

21. Umate, P. Genome-wide analysis of the family of light-harvesting chlorophyll a/b-binding proteins in Arabidopsis and rice. *Plant Signal. Behav.* **2010**, *5*, 1537–1542. [[CrossRef](#)]
22. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST plus: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [[CrossRef](#)] [[PubMed](#)]
23. Wang, P.; Grimm, B. Connecting Chlorophyll Metabolism with Accumulation of the Photosynthetic Apparatus. *Trends Plant Sci.* **2021**, *26*, 484–495. [[CrossRef](#)] [[PubMed](#)]
24. De Bianchi, S.; Dall’Osto, L.; Tognon, G.; Morosinotto, T.; Bassi, R. Minor antenna proteins CP24 and CP26 affect the interactions between photosystem II subunits and the electron transport rate in grana membranes of Arabidopsis. *Plant Cell.* **2008**, *20*, 1012–1028. [[CrossRef](#)] [[PubMed](#)]
25. Chen, Y.E.; Ma, J.; Wu, N.; Su, Y.Q.; Zhang, Z.W.; Yuan, M.; Zhang, H.Y.; Zeng, X.Y.; Yuan, S. The roles of Arabidopsis proteins of Lhcb4, Lhcb5 and Lhcb6 in oxidative stress under natural light conditions. *Plant Physiol. Biochem.* **2018**, *130*, 267–276. [[CrossRef](#)] [[PubMed](#)]
26. Zou, Z.; Li, M.; Jia, R.; Zhao, H.; He, P.; Zhang, Y.; Guo, A. Genes encoding light-harvesting chlorophyll a/b-binding proteins in papaya (*Carica papaya* L.) and insight into lineage-specific evolution in Brassicaceae. *Gene* **2020**, *748*, 144685. [[CrossRef](#)]
27. Pan, Z.; Chen, L.; Wang, F.; Song, W.; Cao, A.; Xie, S.; Chen, X.; Jin, X.; Li, H. Genome-wide identification and expression analysis of the ascorbate oxidase gene family in *Gossypium hirsutum* reveals the critical role of GhAO1A in delaying dark-induced leaf senescence. *Int. J. Mol. Sci.* **2019**, *20*, 6167. [[CrossRef](#)]
28. Wang, X.; Wang, H.; Wang, J.; Sun, R.; Wu, J.; Liu, S.; Bai, Y.; Mun, J.H.; Bancroft, I.; Cheng, F.; et al. The Genome of the Mesopolyploid Crop Species *Brassica Rapa*. *Nat. Genet.* **2011**, *43*, 1035–1040. [[CrossRef](#)] [[PubMed](#)]
29. Cheng, F.; Mandáková, T.; Wu, J.; Xie, Q.; Lysak, M.A.; Wang, X. Deciphering the Diploid Ancestral Genome of the Mesohexaploid *Brassica Rapa*. *Plant Cell.* **2013**, *25*, 1541–1554. [[CrossRef](#)] [[PubMed](#)]
30. Li, K.; Ma, B.; Shen, J.; Zhao, S.; Ma, X.; Wang, Z.; Fan, Y.; Tang, Q.; Wei, D. The Evolution of the Expansin Gene Family in *Brassica* Species. *Plant Physiol. Biochem.* **2021**, *167*, 630–638. [[CrossRef](#)]
31. Wang, D.; Zhang, S.; He, F.; Zhu, J.; Hu, S.; Yu, J. How do variable substitution rates influence Ka and Ks calculations? *Genom. Proteom. Bioinform.* **2009**, *7*, 116–127. [[CrossRef](#)]
32. Yu, F.; Tang, C.Q.; Xin, Y.Y.; Peng, D.C.; Xu, Y.N.; Li, L.B.; Kuang, T.Y. Progress in the structural and functional study of photosystem I. *Chin. Bull. Bot.* **2001**, *18*, 266–275.
33. Goodstein, D.M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R.D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N.; et al. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **2012**, *40*, D1178–D1186. [[CrossRef](#)] [[PubMed](#)]
34. 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.; et al. CDD: NCBI’s conserved domain database. *Nucleic Acids Res.* **2015**, *43*, D222–D226. [[CrossRef](#)] [[PubMed](#)]
35. Bateman, A.; Coin, L.; Durbin, R.; Finn, R.D.; Hollich, V.; Griffiths-Jones, S.; Khanna, A.; Marshall, M.; Moxon, S.; Sonnhammer, E.L.L.; et al. The Pfam protein families database. *Nucleic Acids Res.* **2004**, *32* (Suppl. S1), D138–D141. [[CrossRef](#)] [[PubMed](#)]
36. Letunic, I.; Doerks, T.; Bork, P. SMART: Recent updates, new developments and status in 2015. *Nucleic Acids Res.* **2015**, *43*, D257–D260. [[CrossRef](#)]
37. Thompson, J.D.; Gibson, T.J.; Higgins, D.G. Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinform.* **2002**, *32*, 1792–7. [[CrossRef](#)]
38. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)]
39. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [[CrossRef](#)]
40. 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](#)]
41. Wang, Y.; Tang, H.; Debarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [[CrossRef](#)]
42. Cui, L.; Feng, K.; Wang, M.; Wang, M.; Deng, P.; Song, W.; Nie, X. Genome-wide identification, phylogeny and expression analysis of AP2/ERF transcription factors family in *Brachypodium distachyon*. *BMC Genom.* **2016**, *17*, 636. [[CrossRef](#)] [[PubMed](#)]
43. Lynch, M.; Conery, J.S. The evolutionary fate and consequences of duplicate genes. *Science* **2000**, *290*, 1151–1155. [[CrossRef](#)] [[PubMed](#)]

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