



Article

Characterizations of MYB Transcription Factors in *Camellia oleifera* Reveal the Key Regulators Involved in Oil Biosynthesis

Sijia Li ^{1,2,3,†} , Hu Huang ^{1,2,3,†}, Xianjin Ma ^{1,2,3}, Zhikang Hu ^{1,2,3}, Jiyuan Li ³ and Hengfu Yin ^{1,3,*}

¹ State Key Laboratory of Tree Genetics and Breeding, Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou 311400, China

² College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China

³ Key Laboratory of Forest Genetics and Breeding, Research Institute of Subtropical Forestry, Chinese Academy of Forestry, Hangzhou 311400, China

* Correspondence: hfyin@sibs.ac.cn; Tel.: +86-571-6310-5093

† These authors contributed equally to this work.

Abstract: MYB (myeloblastosis) transcription factors plays an important role in various physiological and biochemical processes in plants. However, little is known about the regulatory roles of MYB family genes underlying seed oil biosynthesis in *Camellia oleifera*. To identify potential regulators, we performed the genome-wide characterizations of the MYB family genes and their expression profiles in *C. oleifera*. A total of 186 CoMYB genes were identified, including 128 R2R3-type MYB genes that had conserved R2 and R3 domains. Phylogenetic analysis revealed the CoR2R3-MYBs formed 25 subgroups and possessed some highly conserved motifs outside the MYB DNA-binding domain. We investigated the promoter regions of CoR2R3-MYBs and revealed a series of cis-acting elements related to development, hormone response, and environmental stress response, suggesting a diversified regulatory mechanism of gene functions. In addition, we identified four tandem clusters containing eleven CoR2R3-MYBs, which indicated that tandem duplications played an important role in the expansion of the CoR2R3-MYB subfamily. Furthermore, we analyzed the global gene expression profiles at five stages during seed development and revealed seven CoR2R3-MYB genes that potentially regulated lipid metabolism and seed maturation in *C. oleifera*. These results provide new insights into understanding the function of the MYB genes and the genetic improvement of seed oil.

Keywords: genomic analysis; gene expression; gene family; MYB transcription factor; seed oil biosynthesis



Citation: Li, S.; Huang, H.; Ma, X.; Hu, Z.; Li, J.; Yin, H. Characterizations of MYB Transcription Factors in *Camellia oleifera* Reveal the Key Regulators Involved in Oil Biosynthesis. *Horticulturae* **2022**, *8*, 742. <https://doi.org/10.3390/horticulturae8080742>

Academic Editors: Dilip R. Panthee and Sergey V. Dolgov

Received: 27 June 2022

Accepted: 12 August 2022

Published: 18 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



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

The regulation of gene expression plays an important role in most biological processes, such as growth, development, and response to environmental signals [1]. Transcriptional regulation of gene expression mainly depends on the recognition of promoter elements by transcription factors [2].

MYB (myeloblastosis) is one of the largest families of transcription factors in plants. In 1987, the first plant MYB transcription factor was discovered, and a protein-coding gene at the *COLORED1* site was identified to affect the accumulation of pigments in the aleurone scutellum tissues of the kernel in maize [3]. In the model plant *Arabidopsis thaliana*, the conserved sequences of the MYB family members were characterized initially [4]. As the *Arabidopsis* genome was fully sequenced, comprehensive characterizations of the MYB genes at the genome-wide scale were reported in *A. thaliana* [5]. In recent years, MYB proteins have been identified in a range of plant species, including pear (*Pyrus bretschneideri*), poplar (*Populus trichocarpa*), and more [3,6].

The various regulatory functions of MYB members mainly depend on their sequence structure. The N-terminus of MYB members has a highly conserved MYB DNA-binding

domain, which is usually composed of 1~4 imperfect repeats (denoted as R) [7]. The C-terminal region outside the MYB DNA-binding domain is found to play important roles in the activation or inhibition of gene expression [7]. Each R contains approximately fifty-two amino acids, forming three helices. The second and third helices of each R form a helix-turn-helix (HTH) structure that binds to the DNA major groove. These repeats have 2–3 highly conserved tryptophan (W) residues that are critical in the formation of the HTH structure. Based on the number of Rs, MYB proteins are divided into four major subfamilies: the MYB proteins with a single or a partial R are named “1R-MYB/MYB-related”, with two Rs are called “R2R3-MYB”, with three Rs are called “3R-MYB” factor, and with four Rs are called “4R-MYB” [5]. Among them, 4R-MYB is found to be the smallest group and not well-known. The 3R-MYBs usually contain five members in higher plants and have been shown to play important roles in cell cycle control [8]. The 1R-MYB contains R3 and R1/R2 type genes, which are involved in secondary metabolism control and organ morphogenesis, respectively [9–11].

R2R3-MYBs have been extensively demonstrated to be involved in primary and secondary metabolism [12–14]. For instance, in *Arabidopsis*, *AtMYB52*, *AtMYB54*, *AtMYB69* (subgroup 21) and *AtMYB103* positively regulate the cell wall thickening of fiber cells [15]. *AtMYB75*, *AtMYB90*, *AtMYB113*, and *AtMYB114* (subgroup 6) involve in anthocyanin biosynthesis [12]. In poplar, *PtMYB165* is a major inhibitor of flavonoid pathways [16]. In addition to playing an independent role in the regulation of plant secondary metabolism, MYB protein can also bind bHLH and WD40 protein to form the ternary MYB-bHLH-WD40 complexes, which regulate the biosynthesis of flavonols, anthocyanins, and procyanidins [17–20].

R2R3-MYBs also play a key role in regulating lipids metabolism. *AtMYB30* has been shown to activate the synthesis of the very-long-chain fatty acids (VLCFAs) [21]. Additionally, *AtMYB92* directly activates the *BCCP2* promoter, which encodes a component of the fatty acid (FA) biosynthetic pathway [22]. However, more R2R3-MYB genes act as suppressors in lipid metabolism. The expression level of *AtMYB76* is negatively correlated with the content of FA in mature seeds [23]. *AtMYB89* inhibits seed FA accumulation by regulating *KCS11*, *WRI1*, and *BCCP1* [13]. *AtMYB123* inhibits seed FA biosynthesis by targeting *FUS3* [24]. Additionally, *AtMYB118* negatively regulates FA biosynthesis in the endosperm by repressing maturation-related genes [25].

C. oleifera is one of the trees in the world that produces edible oil, belonging to the *Camellia* genus. It has been widely cultivated as an oil crop, in many other countries, including China, the Philippines, India, Japan, Brazil, Thailand, and South Korea [26]. Moreover, the active ingredient of camellia seed cake is saponin, which is a plant pesticide that can also be used for chemical cleaning, machine rust removal, etc. [27]. *Camellia* oil, the main product of *C. oleifera* seeds, is extensively used as cooking oil in South Asia. There are mainly seven fatty acids in *Camellia* oil, including two saturated fatty acids and five unsaturated fatty acids [28]. Additionally, the rapid accumulation of oil contents begins at the stage of seed kernel maturation [29]. The unsaturated fatty acid content of *Camellia* oil is much higher than that of peanut oil and soybean oil, and the vitamin E content is twice that of olive oil [30]. It remains unclear how R2R3-MYB proteins are involved in the regulation of seed development and oil metabolism.

To uncover the potential regulators related to seed oil biosynthesis, we conducted a comprehensive characterization of MYB family genes in *C. oleifera*. The members of the MYB family were identified using the recently published genome of *C. oleifera* [31]. Subsequently, the sequence characteristics, phylogenetic analysis, chromosomal localization, and collinearity of *C. oleifera* R2R3-MYBs were studied. We also analyzed the expression patterns of R2R3-MYBs at different seed development stages using RNA-seq data. Based on the gene co-expression network analysis, we identified seven *CoR2R3-MYB* genes and six putative downstream genes of them that may play important roles in seed development and lipid metabolism. This study will help to explore the potential function of the

R2R3-MYB transcription factor in *C. oleifera* and provide important candidate genes for *C. oleifera* breeding.

2. Materials and Methods

2.1. Sequence Analysis of MYB Transcription Factors

The amino acid sequences of *C. oleifera* are derived from the whole genome (accession number: PRJNA732216, including SRR14710457 to SRR14710508). BLASTP and HmmerSearch methods were used for the MYB transcription factors identification of *C. oleifera*. Initially, the candidate MYB transcription factors were also identified by HMMER3.2.1 using the MYB DNA-binding domain HMM (hidden Markov model) profile (Pfam Number: PF00249) as the reference [32]. The HMM profile was downloaded from the Pfam database (<http://pfam.xfam.org/>, accessed on 1 June 2021). The longest amino acid sequence was selected for each gene. In parallel, we performed a sequence similarity search with the amino acid sequences from the Arabidopsis Information Resources (TAIR) database (<https://www.arabidopsis.org/index.jsp>, accessed on 1 June 2021) through BLASTP analysis by the Bioedit software (Version7.0.5.3, e-value < 1×10^{-10} ; available at <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>, accessed on 1 June 2021), in order to screen candidate MYB transcription factors with high sequence homology and to eliminate repeated sequences. Then, the result of HmmerSearch whose e-value is less than 1×10^{-10} is selected and union with the result of BLASTP. Finally, the Online website SMART (<http://smart.embl-heidelberg.de/>, accessed on 5 June 2021) was used to further verify whether it contained MYB transcription factors characteristics, and MYB type according to the number of repeats.

2.2. Sequence Features Analysis of CoR2R3-MYBs

The online website CELLO v.2.5 (<http://cello.life.nctu.edu.tw/>, accessed on 10 June 2021) was used to carry out the subcellular localization prediction. The visualization of the coding region structure analysis was based on the annotation file (GFF3, General Feature Format) of the *C. oleifera* genome. Weblogo software online (<http://weblogo.berkeley.edu/logo.cgi>, accessed on 20 June 2021) was used to draw the sequence logo of R2 and R3. SWISS-MODEL was used to draw the three-dimensional structure of the R2R3-MYB consensus sequence of *C. oleifera*. Using the online website MEME (Version5.4.1, <https://meme-suite.org/meme/doc/meme.html>, accessed on 22 June 2021) predicted the conserved motif sequence of CoR2R3-MYB transcription factors with the following parameters: number of motifs (10), expect motif sites (anr) [33]. SMART and NCBI CDD were also used to determine the R2R3-MYB domain architecture with the default parameter. All of them were visualized by Tootools software (Version1.09861, <https://github.com/CJ-Chen/TBtools>, accessed on 1 June 2021) [34]. To investigate the domain characteristics, 128 *C. oleifera* R2 and R3 repeats were aligned by ClustalW of MEGA X software [35]. To investigate cis-acting elements in the *CoR2R3-MYB* gene promoter regions, the 2 kb upstream regions of the CDS were scanned by the Search for CARE program on the PlantCARE database website (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>, accessed on 1 August 2021) and Plant Transcription Factor Database (<http://planttfdb.gao-lab.org/index.php>, e-value < 1×10^{-5} , accessed on 12 June 2022). We selected promoters of genes with MYB binding elements in both databases for display. Intron length distribution density diagram and cis-acting element diagram were plotted with R package “ggplot2” [36].

2.3. Chromosomal Location and Synteny Analysis

We visualized the localization of *CoR2R3-MYBs* on chromosomes. The tandem genes were identified by the following criteria: (1) homologous *CoR2R3-MYB* genes were located within 150 kb, and (2) these genes were separated by 2 or fewer other genes.

Multiple Collinearity Scan toolkit (MCScanX) was used to analyze gene duplication events [37]. Two model plants, *A. thaliana* and *P. trichocarpa*, were selected for interspecies collinearity analysis. Genome and annotated files of *A. thaliana* (PRJNA10719,

GCF_000001735.4) and *P. trichocarpa* (PRJNA17973, GCF_000002775.4) are downloaded from NCBI. Then, TBtools software was used to visualize the above results [34].

2.4. Phylogenetic Analysis

Multi-sequence alignment of 128 full-length CoR2R3-MYB proteins was performed by ClustalW and the phylogenetic tree was conducted by the neighbor-joining method (NJ) with 1000 bootstrap replicates [38]. They were all completed using MEGA X software [35]. Multi-sequence alignment of 126 *A. thaliana* R2R3-MYBs, 128 *C. oleifera* R2R3-MYBs, and 5 “Landmark” MYB transcription factors of other species was conducted using MUSCLE of MEGA X software [35]. MYB transcription factors of five other species are HvGAMYB (accession number: AAG22863.1), PtGAMYB (AZQ25486.1), NtMYB1 (AAB41101.1), PhMYBAN2 (BAP28593.1), AmMYBPHAN (CAA06612.1). All of them were downloaded from NCBI (<https://www.ncbi.nlm.nih.gov/>, accessed on 22 August 2021). Ttools software was used to call trimAl (Version 1.2, <http://trimal.cgenomics.org/>, accessed on 1 June 2021) and IQ-TREE (Version 1.6.12, <http://www.iqtree.org/>, accessed on 1 June 2021), trim the results of multiple sequence alignment and find the most suitable amino acid substitution model (VT+R10), and finally build a Maximum likelihood (ML) phylogenetic tree with 5000 ultrafast bootstrap replicates [34,39,40]. iTOL online tool (Interactive Tree of Life, Version 6.3.2, <https://itol.embl.de/>, accessed on 28 May 2021) was used to beautify the phylogenetic tree and 25 subgroups of Arabidopsis R2R3-MYB proteins were introduced to classify CoR2R3-MYB proteins [41].

2.5. Analysis of CoR2R3-MYBs Expression Pattern at Five Different Stages of Seed Development

RNA-seq data of seeds at five different developmental stages (PRJNA668531) were downloaded from NCBI Short Read Archive (SRA, <https://www.ncbi.nlm.nih.gov/sra/?term=>, accessed on 22 October 2021). The five developmental stages of seeds were T1, T2, T3, T4, and T5. The oil content gradually increased from T1 to T5 and reached the highest at T5 [29]. Quality control of RNA-seq data was performed using FastQC (Version 3), and qualified data were used for mapping to the reference genome of *C. oleifera* by Hisat2 (Version 2.1.0) [42]. The Sam files obtained above were converted to BAM file format using Samtools (Version 1.9) and reads in BAM files were sorted [43]. String Tie (Version 1.3.6) performs initial sequence assembly, generates a GTF file for each BAM file, combines GTF files containing transcripts information into a single GTF, and combines these transcripts into a non-redundant set of transcripts [44]. Next, these transcripts were reassembled, and gene expression abundances were estimated (normalized by FPKM, Fragments Per Kilobase Million). Finally, CoR2R3-MYBs expression levels were extracted. All expression data were converted through log₂, and heatmaps were drawn through the R package “pheatmap” to show the expression patterns of CoR2R3-MYB genes.

2.6. Weighted Correlation Network Analysis (WGCNA) and Screening of Hub Genes

The WGCNA R Shiny was used to analyze the co-expression of genes [45]. The genes whose logFPKM was less than 1 in 90% of the samples were filtered out, and the genes whose median absolute deviation was in the top 10,000 were selected to construct a scale-free network. The scale-free network was constructed with the soft threshold corresponding to $R^2 \geq 0.8$ for the first time, then modules were obtained using the default settings. The module eigengene (ME, the first principal component of a given module) values were associated with the sample. We selected the turquoise module with the largest number of genes, which is strongly correlated with the T5 stage which has the highest oil content, to explore the involvement of the CoR2R3-MYBs in lipid-related functions. The hub genes were screened according to the eigengene connectivity (kME) value. We selected the CoR2R3-MYBs with a kME value ≥ 0.9 as the module hub genes to study the possible function of the CoR2R3-MYBs. These CoR2R3-MYBs and their co-expressed genes (weight > 0.3) were annotated based on GO and Swiss-Prot (<https://www.uniprot.org/downloads>, accessed on 12 November 2021) [46]. TBtools software was used to conduct GO enrichment according to

the GO annotation information. We selected the top five terms with the smallest p -value in the three parts of GO, respectively, for display. Next, the Cytoscape software (Version 3.8.2) was used to perform the co-expression regulatory network based on the annotation of Swiss-Prot [47]. The expression profiles of hub genes and their co-expressed genes related to lipid metabolism were plotted using the R package “pheatmap”.

3. Results

3.1. Identification of MYB Transcription Factors in *C. oleifera*

Based on the BLASTP and HmmerSearch analysis, 44 1R-MYB proteins, 137 2R-MYB proteins, and 5 3R-MYB proteins were obtained (Table S1). We found that nine 2R-MYB genes were less conserved with consensus sequences, less than 40%, and displayed extra sequence features determined by the SMART analysis (Table S2). We categorized the nine 2R-MYB genes into the “Unusual MYB proteins”. In summary, we reported a total of 186 MYB transcription factors in *C. oleifera*, containing 44 1R-MYB proteins, 128 R2R3-MYB proteins, 5 3R-MYB proteins, and 9 “Unusual” MYB proteins (Table 1). We found that the number of MYB genes is comparable to the model plants, *A. thaliana*, and *P. trichocarpa*, and they all have five 3R-MYB members (Table 1) [5,6].

Table 1. Numbers of MYB transcription factors in three plant species.

Species	R2R3	3R	1R and MYB-Related	“Unusual” MYB Genes with Two or More Repeats	Total	Reference
<i>A. thaliana</i>	126	5	64	2	197	[5]
<i>P. trichocarpa</i>	196	5	152	1	354	[6]
<i>C. oleifera</i>	128	5	44	9	186	This study.

3.2. Protein Sequences Analysis of CoR2R3-MYBs

We analyzed the protein sequences and found that the longest R2R3-MYB protein (CoMYB115) contained 1444 amino acids, and the shortest (CoMYB13) contained 107 amino acids (Table S3). The prediction of subcellular localization indicated that 124 CoR2R3-MYB proteins were located in the nucleus, 3 in the mitochondrial, and 1 in the extracellular (Table S3).

To further understand the sequence features of the CoR2R3-MYB proteins from an evolutionary perspective, we constructed a phylogenetic tree based on 128 R2R3-MYBs of *C. oleifera* (Figure S1A). We found that most proteins in the same clade have similar sequence features. The protein domain analysis showed that most of them had the conserved R2 and R3 domains in the N-terminal (Figure S1C). Additionally, there are zero to five low-complexity regions on the C-terminal (Figure S1C). Consistent with this result, we showed that most members from the same clade have one or more identical motifs outside the MYB domain (Figure S1B).

3.3. Sequence Conservation of the R2R3 Domain

To investigate the features of the CoR2R3-MYB domains, we aligned the basic region of the CoR2R3-MYB proteins, consisting of about 104 residues containing the R2 and R3 sequences (Figure 1A,B). The consensus of the R2R3 domain was highly conserved to that of *Arabidopsis*, and only eight amino acid residues were different [5]. Consistent with the results of other species, both R2 and R3 repeats contained regularly spaced and highly conserved triple tryptophan residues (W) in CoR2R3-MYBs (Figure 1B,C), which were important in maintaining the HTH structure (Figure 1C). In R3 repeats, the first W residue (position 59) was usually substituted by phenylalanine (F). In addition, we found other conserved amino acid residues in CoR2R3-MYBs in R2 repeats (Figure 1A) and R3 repeats (Figure 1B). Among them, three conserved amino acid residues were located between the first and second W residues, and six were located between the second and third W residues in R2 repeats (Figure 1A). One conserved amino acid residue was located between the first and second W residues, and five were located between the second and third W

residues in the R3 repeats (Figure 1B). We used the homology modeling method to reveal the three-dimensional structure of the CoR2R3-MYB consensus and found that both R2 and R3 formed the typical HTH structure (Figure 1C). This confirmed the conservation of HTH formed by the second helix and third helix in the DNA recognition domains of MYB proteins.

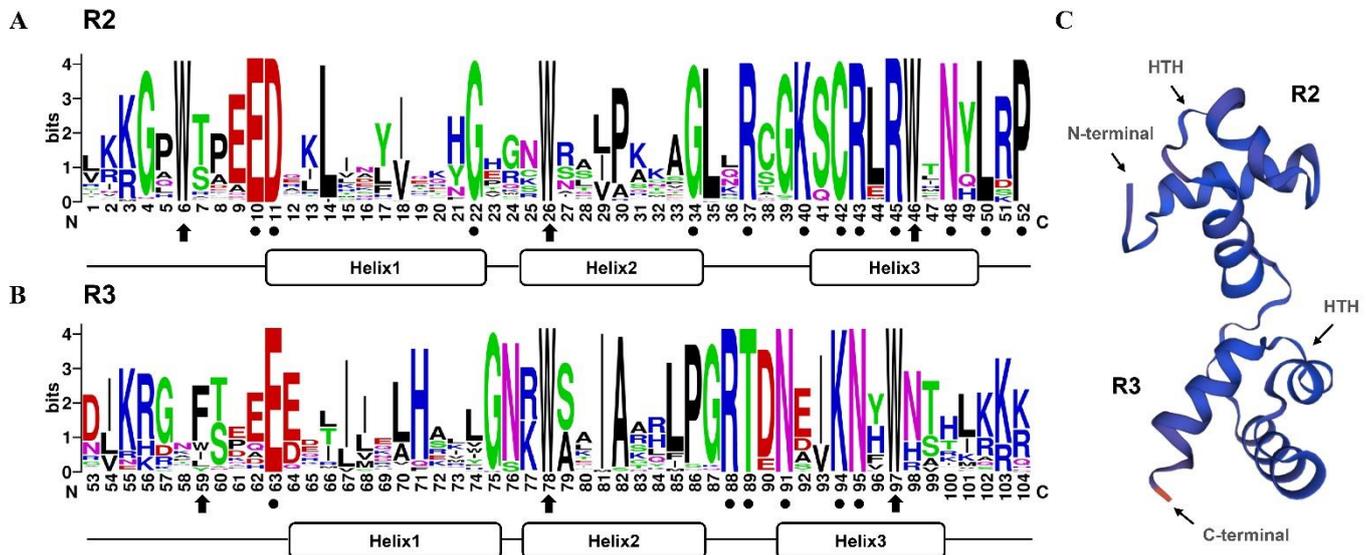


Figure 1. Sequence conservation and three-dimensional structure of 128 R2R3-MYB domains of *C. oleifera*. (A) Conservation of R2 repeats in all 128 R2R3-MYBs of *C. oleifera*. (B) Conservation of R3 repeats in all 128 R2R3-MYBs of *C. oleifera*. N and C on the horizontal axis represent the N-terminal and C-terminal; numbers represent locations. The bit score on the vertical axis represents how conservative each position in the sequence is, and the higher it is, the more conservative it is. Arrowheads indicate the typical, conserved tryptophan residues (W). The dots represent other conserved amino acid residues. (C) Three-dimensional structure of consensus sequences of all 128 R2R3-MYBs of *C. oleifera*. HTH stands for the helix-turn-helix structure.

3.4. Gene Structure Analysis of CoR2R3-MYBs

These R2R3-MYBs were divided into six categories: including one, two, three, four, five, and twenty-one exons, respectively (Figure S1, Tables S4 and S5). The majority of CoR2R3-MYB genes (88 in total) had three exons (Table S5). We found that the length of these exons is highly conserved, and most of them are between 0 and 500 bp (Figure 2A). However, the intron length was highly variable (Figures 2B and S1D). The length of most introns ranged from 0 to 4000 bp, with a small number of introns ranging from 4000 bp to 8000 bp, and only one intron exceeding 8000 bp (Figure 2B).

Although the length of these introns varied, their number, position, and phase were largely conserved. We divided the CoR2R3-MYB genes into 11 patterns (A-K) according to their number, position, and phase of introns in the DNA binding domains (Figure 2C and Table S6). We found that except for eight CoR2R3-MYBs in which the DNA binding domains had no introns (pattern A), all other CoR2R3-MYB DNA binding domains displayed different intron distribution patterns (Figure 2C). The intron phase determines which exons may or may not be targeted for alternative splicing [48]. Exons flanked by two introns of the same phase, including the pattern I and J, may undergo alternative splicing (Figure 2C). The pattern H was the most common intron distribution pattern (57.03%) (Figure 2C), which included exons flanked by two introns of different phases (Figure 2C), and alternative splicing of this pattern could result in premature stop-codons.

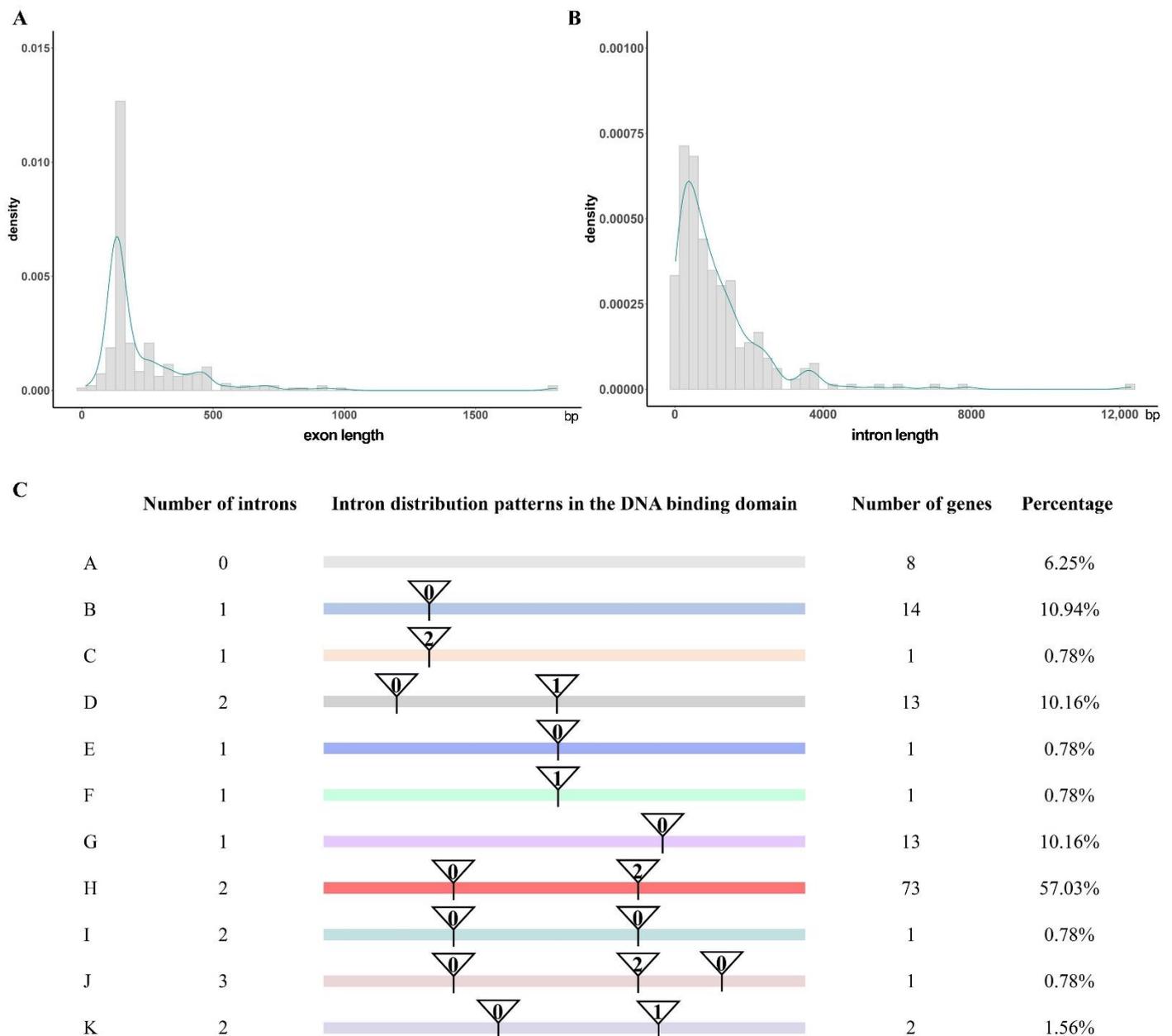


Figure 2. Distribution of exon length, distribution of intron length, and distribution patterns of introns in the DNA binding domain of *CoR2R3-MYBs*. **(A)** Distribution of exon length. **(B)** Distribution of intron length. The green line is the density of the distribution. **(C)** A total of 11 distribution patterns, named A to K. Triangles and numbers represent the positions and splicing phases of introns (0, phase 0; 1, phase 1; 2, phase 2), respectively. The horizontal bars in different colors represent the MYB DNA-binding domain in different patterns. The number of introns in each pattern is shown on the left. The number and percentage of *CoMYBs* in each pattern are shown on the right.

3.5. Promoter Cis-Acting Elements Analysis of *CoR2R3-MYBs*

We intercepted the two kb upstream region of the coding sequences (CDS) of each *CoR2R3-MYBs* as the promoter region to predict the cis-acting elements. Additionally, we classified these cis-acting elements into 27 categories (Figure 3A), which were involved in the development, phytohormone response, and environmental response (Figure S2). Among them, the number of light response elements was the largest (Figure 3A). In addition, we found many MYB binding sites in the promoters of these *R2R3-MYBs*, including the

MYB binding site involved in flavonoid biosynthetic genes regulation, light responsiveness, and drought response (Figure S2).

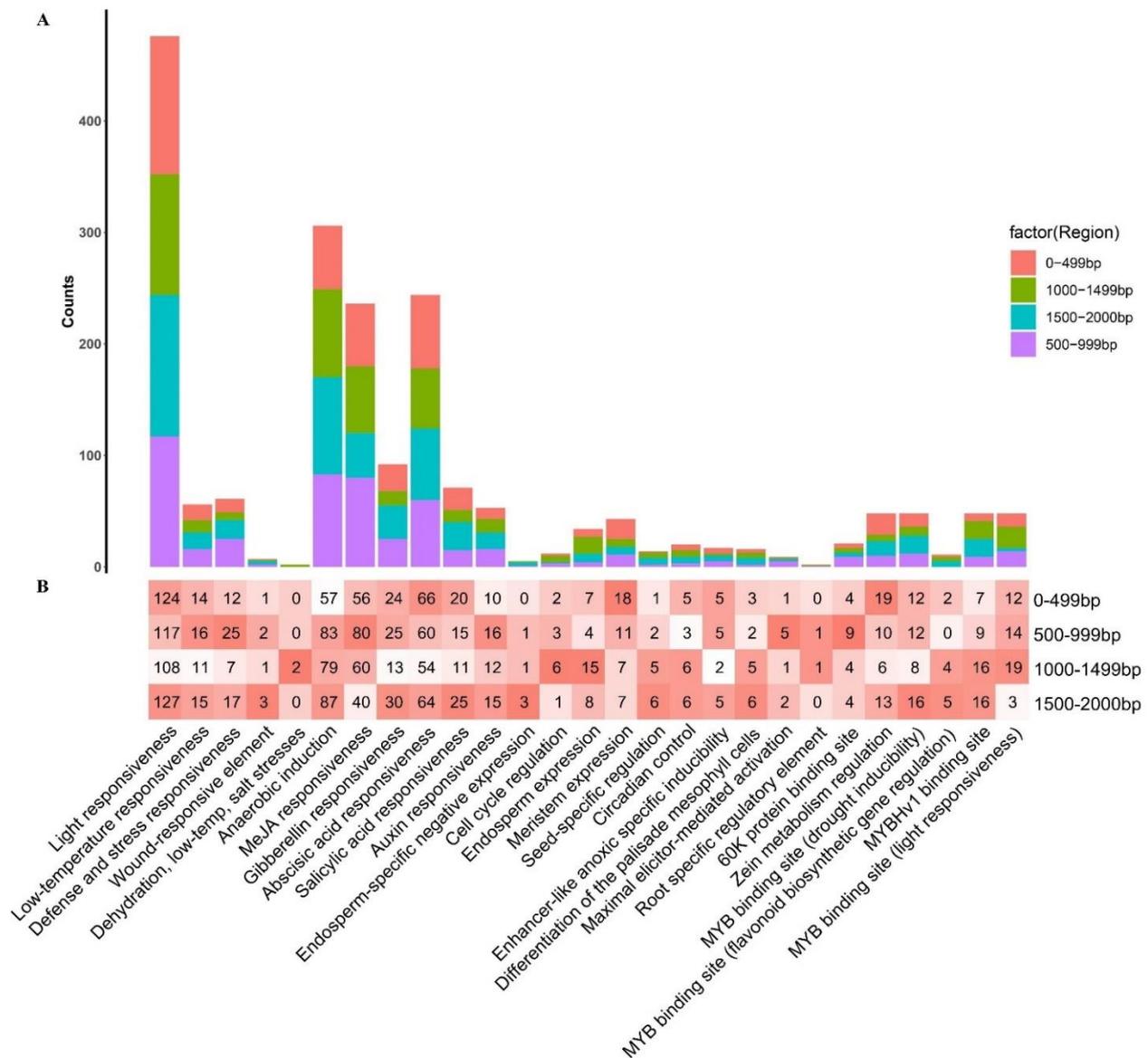


Figure 3. Prediction of the cis-acting elements of promoters of 128 *R2R3-MYB* genes in *C. oleifera*. (A) The stacked bar chart represents the sum of the cis-acting elements in each category. (B) Statistical analysis of 27 categories of cis-acting elements in 4 regions of promoters (0–499 bp, 500–999 bp, 1000–1499 bp, and 1500–2000 bp). The colors and numbers in the grid indicate the number of these cis-acting elements in the region. The darker the red color in the grid, the greater the number of cis-acting elements in that region than in other regions.

To understand the regulatory function of the *CoR2R3-MYB* genes, we divided the promoter region into four sections, starting from the five prime ends, which were 0–499 bp, 500–999 bp, 1000–1499 bp, and 1500–2000 bp, respectively, and analyzed the distribution of cis-acting elements in each region. We found that the light response elements were the largest in all regions (Figure 3B). The cis-acting elements of various hormone responses are less distributed between 1000 and 1499 bp. However, various growth and development-related cis-acting elements were enriched in the range of 1000–1499 bp (Figure 3B). Additionally, the stress response cis-acting elements were enriched at 500–999 bp and 1500–2000 bp regions. MYB binding sites are enriched at 1000–1499 bp and 1500–2000 bp (Figure 3B).

3.6. Chromosomal Location and Synteny Analysis

Chromosomal localization analysis showed that *CoR2R3-MYBs* were distributed on all 15 chromosomes of the *C. oleifera* genome (Figure 4). In addition, two *R2R3-MYB* genes were distributed in the scaffold (Table S7). The *CoR2R3-MYB* genes were renamed *CoMYB1* to *CoMYB128* based on their position on the 15 chromosomes and scaffold (Table S7). We found some *CoR2R3-MYB* genes are tightly packed in some regions of chromosomes (Figure 4). We further determined that some of these genes evolved from tandem duplication events within local chromosomes. There are four tandem duplication events in three chromosomes, containing 11 *CoR2R3-MYB* genes (Figure 4).

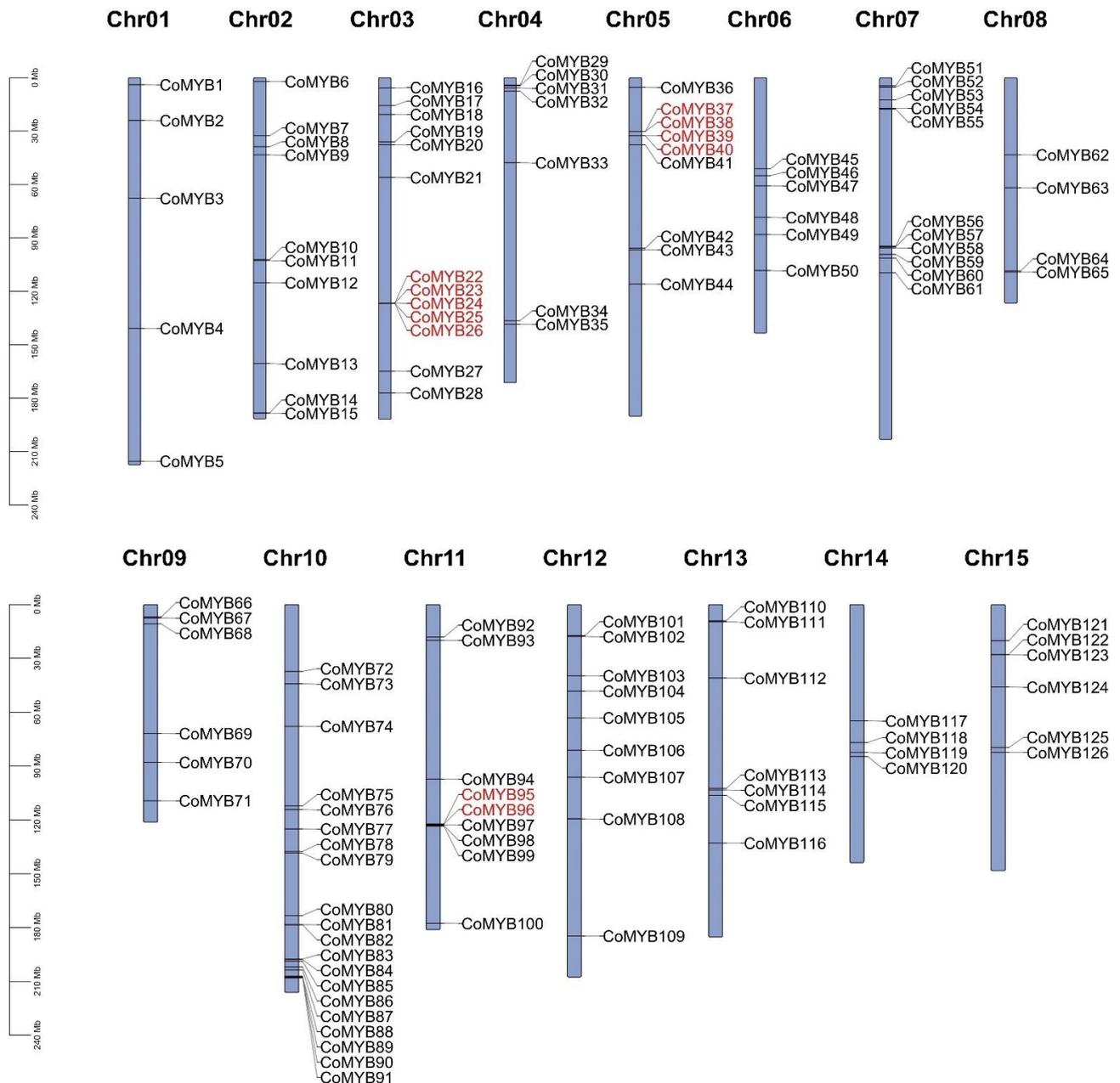


Figure 4. Chromosomal distribution of 128 *CoR2R3-MYB* genes. Each vertical bar represents one chromosome. The chromosome name is shown at the top of each chromosome. For clarity, we renamed the chromosome name using Chr1-Chr15, based on the original chromosome ID (Table S7). The gene names in red text indicate tandem duplication. The scale on the left is measured in megabase (Mb).

To further explore the potential evolutionary mechanism of the *CoR2R3-MYBs*, an intra-species collinearity analysis of *C. oleifera* was carried out. In the *C. oleifera* genome, a total of 393 segmental duplication events occurred, and 49 syngeneic *CoR2R3-MYB* gene pairs were involved (Figure S3; Table S8). To explore the evolutionary driving force of the *CoR2R3-MYB* genes, we also analyzed the non-synonymous and synonymous substitution rates (Ka and Ks) of the *CoR2R3-MYB* syngeneic gene pairs in the collinear region. The results showed that the Ka/Ks ratio was much less than 1 (Table S8), indicating that the evolution of *CoR2R3-MYB* genes was mainly driven by purification selection.

Furthermore, we performed the collinearity analysis of *C. oleifera* with two other representative species, including *Arabidopsis* and poplar (Figure S4). We found that the number of orthologous *R2R3-MYB* gene pairs between *C. oleifera* and *Arabidopsis* was 54 (Figure S4A and Table S9). Additionally, there are 92 orthologous *R2R3-MYB* gene pairs between *C. oleifera* and poplar (Figure S4B and Table S9). Some of these *R2R3-MYB* genes (52 *CoR2R3-MYBs*) were collinear with both *Arabidopsis* and poplar (Table S9), suggesting that these syntenic gene pairs may have existed before ancestral divergence. Notably, we found some syntenic gene pairs (two *CoR2R3-MYB* genes) in *C. oleifera* and *Arabidopsis* were not present in *C. oleifera* and poplar (Table S9), suggesting that these orthologous gene pairs maybe not conserved during the evolutionary process. Additionally, some syntenic gene pairs (40 *CoR2R3-MYB* genes) were found in *C. oleifera* and poplar (Table S9), which did not exist in *C. oleifera* and *Arabidopsis*. These syntenic gene pairs may have emerged after the divergence of herbaceous and woody plants in the process of evolution.

3.7. Phylogenetic Analysis

The phylogenetic tree of 126 *Arabidopsis* *R2R3-MYBs*, 128 *Camellia* *R2R3-MYBs*, and 5 “Landmark” *MYB* transcription factors in other species were constructed using the maximum likelihood method. Phylogenetic trees were grouped according to the unique conserved motifs of 25 subgroups of *Arabidopsis* *R2R3-MYBs* (Table 2) [5]. Compared with *A. thaliana*, we found that six subgroups of *C. oleifera* expanded, fourteen subgroups contracted, and five subgroups neither expanded nor contracted (Table S10). We also observed “species-specific” clades containing only *C. oleifera* or *A. thaliana* *R2R3-MYB* proteins. That implied ancestral gene duplication and loss events. One clade (between S2 and S16) contains only *R2R3-MYBs* of *C. oleifera* (Figures 5 and S5), which suggests that proteins in the clade may have specialized functions that are either lost in *Arabidopsis* or acquired after divergence from the last common ancestor. five subgroups (S3, S10, S12, S15, S19) do not contain any *CoR2R3-MYBs* (Figures 5 and S5).

Table 2. Conserved motifs of subgroups.

Subgroup	Conserved Motif
Subgroup1 (S1)	YASS
Subgroup2 (S2)	MxFW//SFW
Subgroup3 (S3)	WFKHLESELGLEExDNQQQ
Subgroup4 (S4)	LNL[E/D]L
Subgroup5 (S5)	TKAxRC
Subgroup6 (S6)	PRPRxF
Subgroup7 (S7)	Sx(14)GRT
Subgroup8 (S8)	LRKMGIDPLTHKPL
Subgroup9 (S9)	AQWESARxxAExRLxR
Subgroup10 (S10)	QxxAAAxN//KxQLxHxMxQ//DDxxSDSxWK
Subgroup11 (S11)	PRxDLLD
Subgroup12 (S12)	[L/F]LN[K/R]VA
Subgroup13 (S13)	GIDPxTHK[P/L]L[S/I]xx[E/G]
Subgroup14 (S14)	R2R3: [W]-x(20)-[W]-x(19)-[W]-x(12)-[F]-x(18)-[W]-x(18)-[W]

Table 2. Cont.

Subgroup	Conserved Motif
Subgroup15 (S15)	WVxxDxFELsXL
Subgroup16 (S16)	PxLxFxEW
Subgroup17 (S17)	QQ[F/E]QQ
Subgroup18 (S18)	GLPxyP
Subgroup19 (S19)	PxLxFSEW
Subgroup20 (S20)	WxPRL
Subgroup21 (S21)	FxDfL
Subgroup22 (S22)	QEMlxxEVRSYM
Subgroup23 (S23)	RVxRxxxF//PxxGxxGC
Subgroup24 (S24)	QxGxDPxTH
Subgroup25 (S25)	LxxYlxxxN

Assign subgroups as previously reported in Arabidopsis. However, some motifs were reinterpreted, and some of the previously defined subgroups were not obvious, because we comprehensively considered both *Arabidopsis* and *Camellia R2R3-MYB* genes. “x” represents any amino acid residue; the numbers in “()” represent the number of amino acid residues; the letters in “[]” represent amino acid residues in the same position; “/” stands for “or”; “//” stands for “and”.

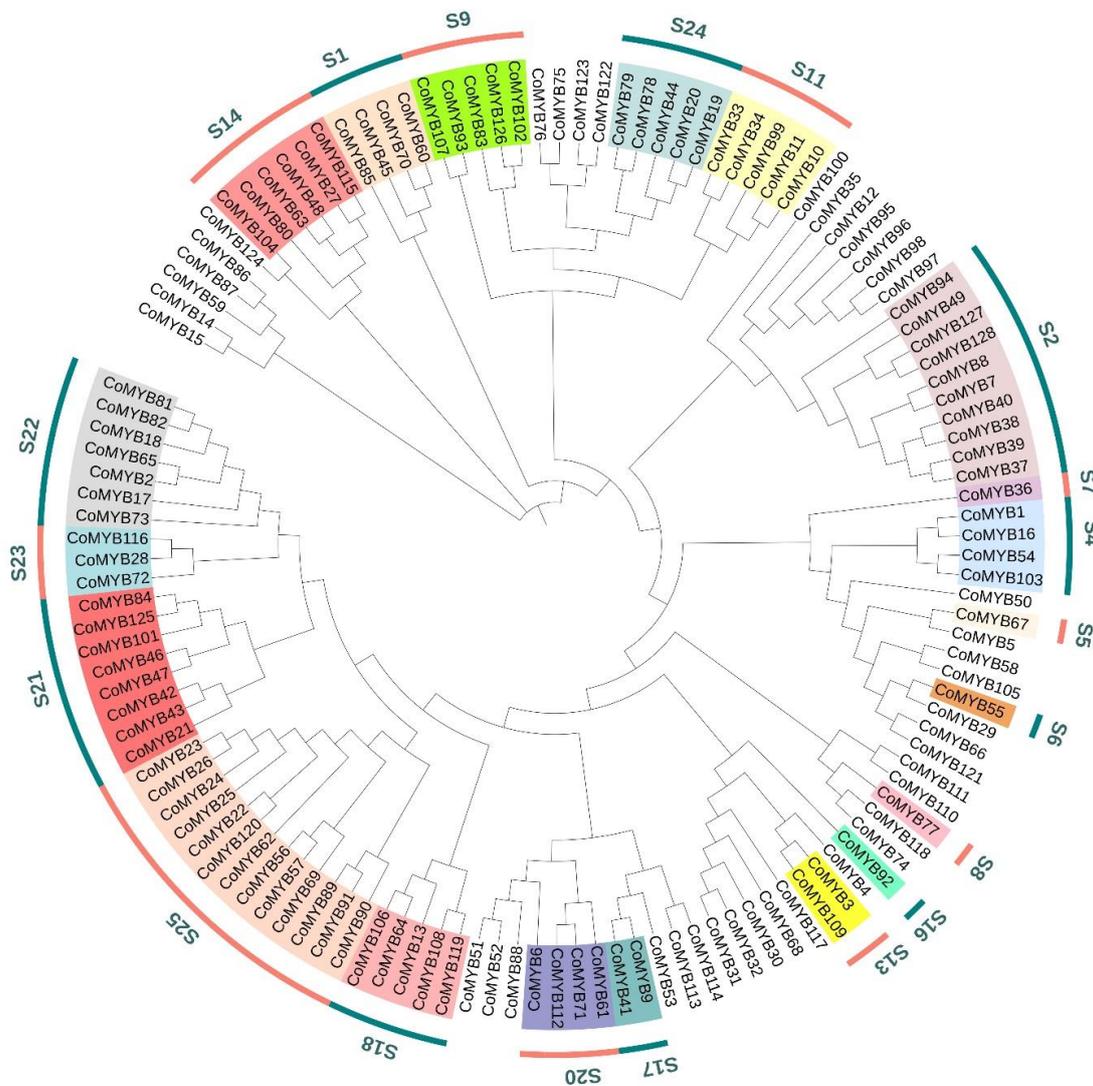


Figure 5. Phylogenetic tree of 128 R2R3-MYB transcription factors of *C. oleifera*. Different colored background colors and strips represent different subgroups. The text of the outermost wheel represents the names of the different subgroups.

3.8. Expression Analysis of CoR2R3-MYBs Genes during Seed Development

To investigate the expression pattern of the CoR2R3-MYB genes in *C. oleifera*, the transcriptional abundances of the CoR2R3-MYB genes at different seed development stages were studied using RNA-Seq. Of the 128 CoR2R3-MYB genes, 100 genes were lowly expressed (FPKM ≤ 10) or undetectable in at least 90% of the samples, and 28 genes were highly expressed (FPKM >10) in more than 10% of the samples (Table S11). Of the 28 genes, the expressions of 18 genes gradually decreased with the increase in oil content (Figure 6). On the contrary, two genes (CoMYB2 and CoMYB61) were highly expressed at stage T5, when the oil content was highest (Figure 6). In addition, one gene (CoMYB106) was expressed uniformly in five stages, five genes (CoMYB54, CoMYB60, CoMYB89, CoMYB90, CoMYB103) were highly expressed at T2 and T3 stages, one gene (CoMYB118) was highly expressed at the T4 stage, and one gene (CoMYB41) was highly expressed only at the T1 stage (Figure 6).

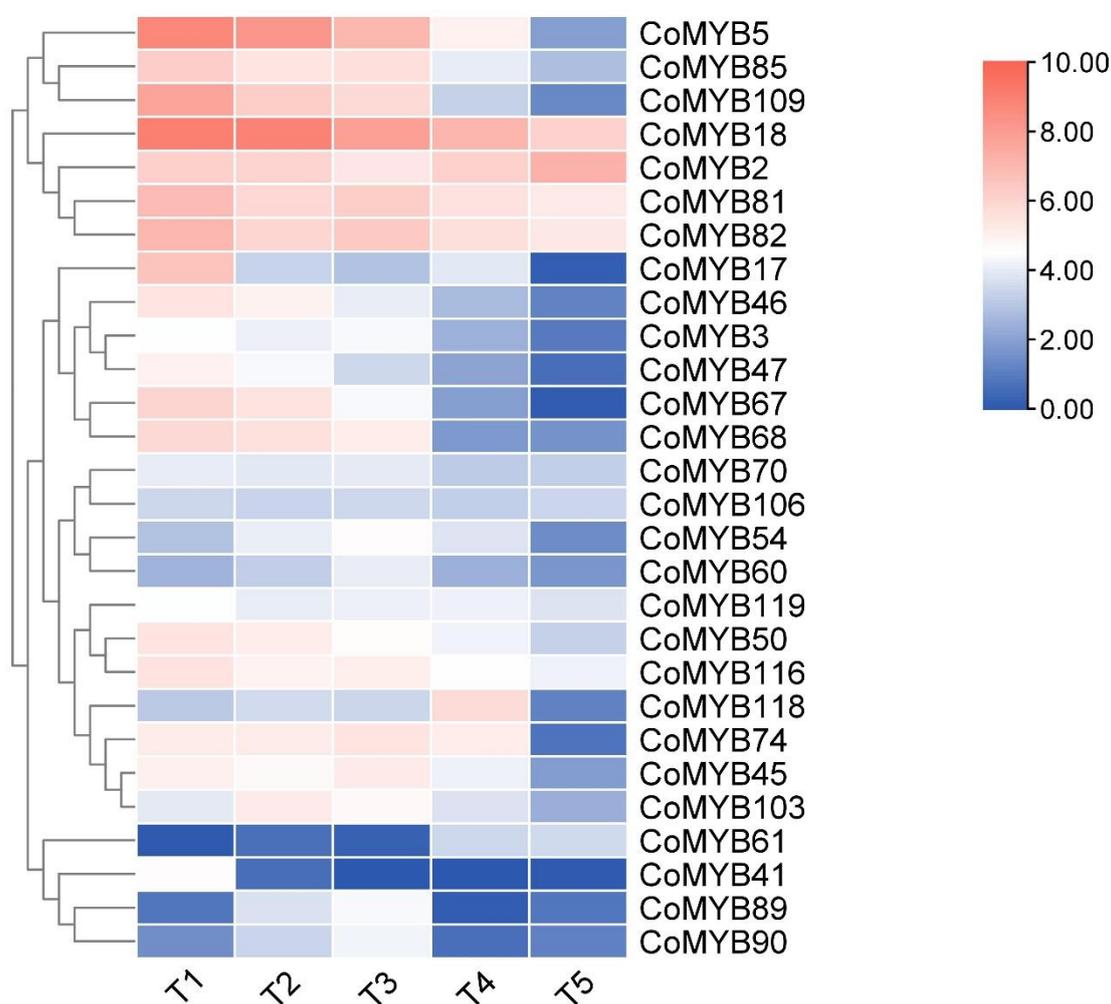


Figure 6. Hierarchical clustering of the expression patterns of 28 R2R3-MYBs in *C. oleifera* at five different stages of seed development. Red means high expression, and blue means low expression. The five developmental stages of seeds were T1 (210 DAP, days after pollination), T2 (235 DAP), T3 (258 DAP), T4 (292 DAP), and T5 (333 DAP). The oil content gradually increased from T1 to T5 and reached the highest at T5.

We evaluated the functions of *AtMYBs* that are homologous or representative in the same subgroup in the phylogenetic tree to understand the potential functions of these genes (Table 3). We classified the functions of these genes into three categories: lipid metabolism-related processes, developmental processes, and stress-related processes

(Table 3). *AtMYB89*, in the S21 subgroup, significantly inhibited seed FA accumulation by regulating *WR11* and *BCCP1*, suggesting that *CoMYB46* and *CoMYB47* in the same subgroup might also perform the same function [13]. Together with *CoMYB45*, *CoMYB60*, *CoMYB70*, and *CoMYB85*, *AtMYB30*, and *AtMYB96* in subgroup S1 regulate the biosynthesis of the VLCFAs [21,49]. *AtMYB123* in the same subgroup as *CoMYB67* inhibits seed FA biosynthesis by targeting *FUS3* [24]. *AtMYB118*, belonging to the same subgroup as *CoMYB89* and *CoMYB90*, negatively regulates FA biosynthesis in the endosperm by repressing maturation-related genes [25].

Table 3. The biological processes or potential functions that 28 CoR2R3-MYBs are involved in.

CoR2R3-MYBs	Subgroup	Representative within a Subgroup or Most Homologous R2R3-MYB Genes of Arabidopsis Thaliana	Function or Biological Process of AtMYBs	Reference
CoMYB46; CoMYB47	S21	AtMYB89; AtMYB110	Inhibit seed FA accumulation by regulating WR11, BCCP1	[13]
CoMYB67	S5	AtMYB123	Inhibit seed FA biosynthesis	[24]
CoMYB85; CoMYB45; CoMYB60; CoMYB70	S1	AtMYB30; AtMYB96	Regulate VLCFAs Biosynthesis	[21,49]
CoMYB89; CoMYB90	S25	AtMYB118; AtMYB119	Negatively regulate FA biosynthesis in the endosperm	[25]
CoMYB68	None	AtMYB26	Male sterility	[50]
CoMYB5; CoMYB50	None	AtMYB5	Control outer seed coat differentiation	[51,52]
CoMYB109; CoMYB3	S13	AtMYB61	Mucilage deposition; lignin biosynthesis	[53,54]
CoMYB116	S23	AtMYB1	Upregulate Tapetum-specific promoter A9 activity	[55]
CoMYB74; CoMYB118	None	AtMYB99	Unknown	-
CoMYB106; CoMYB119	S18	AtMYB33; AtMYB65	Stamen development; leaf development	[56]
CoMYB41	S17	AtMYB71; AtMYB79	Unknown	-
CoMYB103; CoMYB54	S4	AtMYB4	Antioxidant defense	[57]
CoMYB61	S20	AtMYB116; AtMYB62	Phosphate starvation responses	[58]
CoMYB18; CoMYB2; CoMYB81; CoMYB82; CoMYB17	S22	AtMYB44; AtMYB73	Abiotic stress response	[59,60]

“None” means not assigned to a subgroup. “-” means no relevant literature has been found. We classified the functions of these genes into three categories: lipid metabolism-related processes were indicated with an orange background, developmental processes were indicated with a blue background, and stress-related processes were indicated with a gray background.

3.9. Gene Co-Expression Network Underlying the Seed Development

We performed WGCNA to determine the genes associated with lipid metabolism or seed maturation [43]. A total of 10 gene modules were identified and labeled with different colors (Figure S6A). The number of genes in these modules ranged from 74 to 6863 (Table S12). We correlated these modules with five seed development stages with different oil content (Figure S6B). Among them, red and turquoise modules had a high correlation with oil content (Figure S6B), suggesting that they played an important role in the oil content of *C. oleifera* seeds. Next, we analyzed turquoise modules with a large number of genes. In this module, there were 17 CoR2R3-MYB genes (Table S13). Additionally, the kME value greater than 0.9 was considered as the hub gene. As a result, eight CoR2R3-MYB genes were identified as hub genes. Then, their co-expressed genes (weight > 0.3) were detected. We found no co-expressed genes with a weight value greater than 0.3 in *CoMYB45*, and *CoMYB109* had the most co-expressed genes (487) (Table S14 and Figure 7A). Thus, we constructed a gene co-expression network using the seven CoR2R3-MYB genes and their co-expression genes (Figure 7A).

To understand the biological function of genes, we performed a GO enrichment analysis on these genes. Among the top five terms with the smallest *p*-values of the three parts of GO, we found that genes related to lipid binding were significantly enriched, as well as items related to the biological processes necessary for life, chromatin assembly, and nucleosome assembly (Figure S6C). Swiss-Prot annotations were also performed to understand the function of these co-expressed genes in more detail. Among these co-expressed genes, at least 13 genes were found to be related to lipid metabolism or seed maturation (Figure 7A), including *CAC3*, encoding the α subunit of the acetyl-CoA carboxylase (ACC) which is the first enzyme in the fatty acid synthesis pathway [61]. Among them, nine genes had the same expression pattern as CoR2R3-MYB hub genes,

and four had the opposite expression pattern (Figure 7B). We detected cis-acting elements in the promoter regions of these genes and found that six genes had MYB binding sites (Figure 7C).

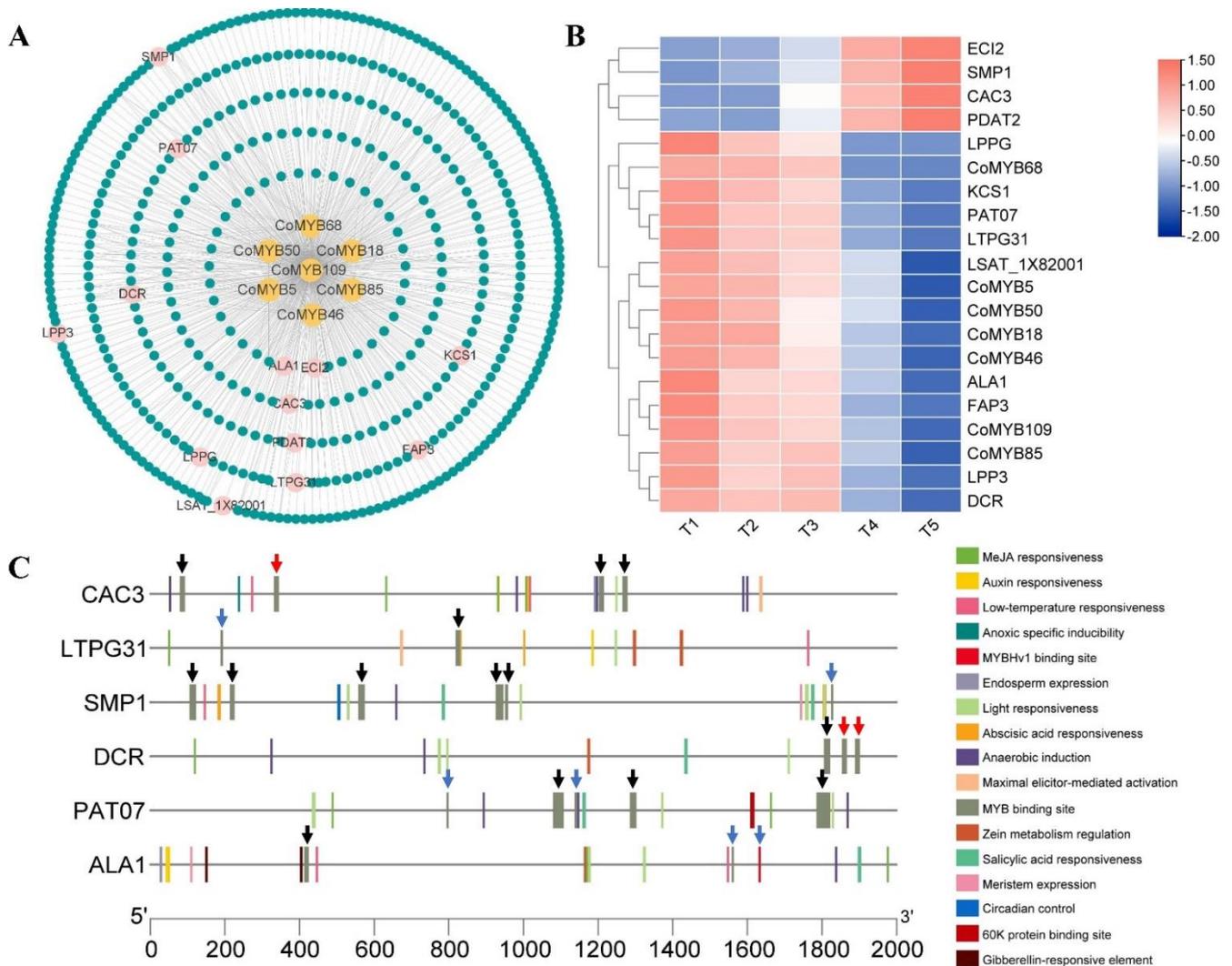


Figure 7. The co-expression network of 7 *CoR2R3-MYB* genes and expression patterns of 7 *CoR2R3-MYB* genes and their co-expressed genes related to lipid metabolism or seed maturation. (A) The co-expression network of 7 *CoR2R3-MYB* genes and their co-expression genes. Each dot represents a gene; the orange dots in the innermost circle represent the *CoR2R3-MYB* genes; the pink dots represent genes associated with seed maturation or lipid metabolism. (B) Expression patterns of 7 *CoR2R3-MYB* genes and their co-expressed genes related to lipid metabolism or seed maturation (homogenize by row). T1 (210 DAP), T2 (235 DAP), T3 (258 DAP), T4 (292 DAP), and T5 (333 DAP) represent five different stages of seed development, respectively. The oil content gradually increased from T1 to T5 and reached the highest at T5. (C) Prediction of cis-acting elements in the promoters of lipid metabolism or seed maturation-related genes containing MYB-binding elements. We selected promoters of genes with MYB binding elements in both databases (PlantCARE and Plant Transcription Factor database) for display. The MYB binding elements are indicated by arrows. The black arrow indicates that the element is from the PlantCARE database, the blue arrow indicates that the element is from the Plant Transcription Factor database, and the red arrow indicates that the element exists in both databases.

4. Discussion

The MYB is one of the most important transcription factor families in plants. *Camellia* oil is a high-quality seed oil beneficial to human health, with the reputation of “Oriental olive oil” [28]. In our study, a total of 186 MYB genes (44 1R-MYBs, 128 R2R3-MYBs, 5 3R-MYBs, and 9 “Unusual” MYBs) were identified from the genome of *C. oleifera* (Table 1). Consistent with the results of *Arabidopsis* and poplar, the CoR2R3-MYB subfamily was the largest subfamily of the MYB gene family (Table 1) [5,6]. Sequence and gene structure analysis of CoR2R3-MYB DNA-binding domains showed that the domain of R2R3-MYB was highly conserved (Figures 1 and 2). However, the C-terminal amino acid sequence was varied (Figure S1). These results suggested the conservation of the N-terminal R2R3-MYB domain which was a DNA-binding domain and the diversity of C-terminal activation or inhibition domains which endowed MYB members with a variety of regulatory functions. Although the C-terminal amino acid sequence is varied, these regions still have conserved motifs that may confer additional functions on the protein. One statement is that their position and conserved motifs in the C-terminal may play an important role in determining their binding properties and performing specific functions [62]. Previous studies in *Arabidopsis* have supported this statement: *GL1* was required to initiate trichome differentiation in *Arabidopsis* [63]. The protein encoded by the *gl1-2* alleles did not have the corresponding function, because *gl1-2* lacked that conserved motif compared with *GL1* [63].

Therefore, we grouped the phylogenetic trees according to the unique conserved motifs at the C-terminal (Figure 5), and the genes in the same subgroup may play the same function. We found that the S25 subgroup of *C. oleifera* was expanded (Figure 5 and Table S10). It was interesting that tandem replication events occurred in five genes of this subgroup (Figure 4). Additionally, *AtMYB118* of this subgroup negatively regulates the biosynthesis of endosperm FA by inhibiting mature-related genes [25]. Thus, these genes might be evolved due to the adaptation to the properties of *C. oleifera* and are associated with *Camellia* oil biosynthesis. We also observed “species-specific” clades containing only *C. oleifera* or *A. thaliana* R2R3-MYB proteins. One clade (between S2 and S16) contains only R2R3-MYBs of *C. oleifera* (Figure 5). It is important to note that we use the term “species-specific” in the context of a group of species whose genomes are currently sequenced. More genome sequences are needed to determine the presence and absence of the clade at the genus level [64]. If the clade was obtained during the divergence of the most recent common ancestor, the clade can be described as a lineage-specific expansion of *C. oleifera*, reflecting a species-specific adaptation.

Based on the RNA-seq data, we analyzed the expression patterns of 128 CoR2R3-MYBs at five different stages of seed development (Figure 6). Of the 128 CoR2R3-MYB genes, 28 genes were determined as highly expressed during seed development. We found that their expression patterns can be roughly divided into four categories: one was a high expression in the early stage of seed development and a low expression in the late stage (Figure 6); one was a high expression in the late stage of seed development and low expression in the early stage (Figure 6); one was a high expression in the middle stage of seed development, low expression in early and late stage (Figure 6); one was expressed uniformly in five stages (Figure 6). Among them, the number of the first type is the largest (including 18 genes), and with the development of the seed, the expression level gradually decreased (Figure 6). Interestingly, in these genes, together with *CoMYB46* and *CoMYB47*, *AtMYB89* in the S21 subgroup significantly inhibited seed FA accumulation by regulating *WRI1* and *BCCP1* [13]. Together with *CoMYB45*, *CoMYB70*, and *CoMYB85*, *AtMYB30* and *AtMYB96* in subgroup S1 regulate the biosynthesis of the VLCFAs [21,49]. Together with *CoMYB67*, *AtMYB123* (*TT2*) inhibits seed FA biosynthesis by targeting *FUS3* [24]. This may suggest that these CoR2R3-MYBs may play a negative regulatory role during seed maturation.

Using WGCNA, we constructed a gene co-expression network using the seven CoR2R3-MYB genes and their co-expression genes. Interestingly, we found that the number of co-

expressed genes of *CoMYB109* was significantly higher than that of several other *CoR2R3-MYB* genes (Table S13). These results suggested that *CoMYB109* might play an important role in seed maturation or lipid metabolism. We annotated the co-expressed genes and found that at least 13 of them were associated with lipid metabolism or seed maturation (Figure 7A). Among them, nine genes had the same expression pattern as the *CoR2R3-MYB* hub genes, and three genes had the opposite expression pattern as the *CoR2R3-MYB* hub genes (Figure 7B). Cis-acting elements were detected in the promoter region of these genes and six genes were found to have MYB binding sites (Figure 7C). Therefore, we proposed that *MYB* may regulate these genes through binding to the cis-acting elements of the promoter regions of these genes to affect lipid metabolism or seed maturation.

5. Conclusions

This study was the first report of the comprehensive analysis of the *CoR2R3-MYB* subfamily based on the *C. oleifera* genome. A total of 128 *CoR2R3-MYB* genes were obtained from the *C. oleifera* genome. These *CoR2R3-MYB* genes were distributed on all 15 chromosomes and have conserved R2 and R3 repeats. Tandem duplication and synteny analysis showed that tandem duplication and segmental duplication played an important role in gene family expansion. Promoter detection revealed cis-acting elements involved in various reactions. The possible functional role of the *CoR2R3-MYBs* was predicted based on its phylogenetic tree. In addition, the expression of *CoR2R3-MYBs* was diverse during seed development. Finally, we identified seven *CoR2R3-MYB* genes and six putative downstream genes of them that may play important roles in seed development and lipid metabolism. These results have an important reference value for the functional identification of *R2R3-MYBs* and genetic improvement of seed oil in *C. oleifera*.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/horticulturae8080742/s1>, Figure S1: Phylogenetic relationships, motif pattern, domain pattern, and gene structure of *R2R3-MYB* transcription factors in *C. oleifera*; Figure S2: Predicted cis-acting elements of promoters of 128 *R2R3-MYB* genes in *C. oleifera*; Figure S3: Schematic diagram of the colinear relationship within *C. oleifera*; Figure S4: Synteny analyses of the *C. oleifera* with Poplar and *Arabidopsis*; Figure S5: Phylogenetic tree of *R2R3-MYB* transcription factors from *C. oleifera*, *A. thaliana*, and five other species; Figure S6: Weighted gene co-expression network analysis of (WGCNA) of transcriptome and GO enrichment of 7 *CoMYB* genes and their co-expression genes; Table S1: The *CoMYBs*; Table S2: The nine unusual *CoMYBs*; Table S3: 128 *CoR2R3-MYBs* information; Table S4: Motifs of *CoR2R3-MYB* transcription factors; Table S5: Exon numbers within *CoR2R3-MYBs*; Table S6: Intron distribution patterns within *CoR2R3-MYB* DNA-binding domains; Table S7: Chromosomal location; Table S8: syngeneic *R2R3-MYB* gene pairs and KaKs values in *C. oleifera*; Table S9: syngeneic *CoR2R3-MYB* genes in *C. oleifera* with *Arabidopsis* and poplar; Table S10: The number of the *R2R3-MYB* members in subgroups; Table S11: The expression patterns of *CoR2R3-MYBs* (FPKM) at five different stages during seed development; Table S12: Module information; Table S13: Module turquoise_MYB hub gene_by_GS > 0.5&kME > 0.5; Table S14: Co-expression genes of 7 *CoR2R3-MYBs*.

Author Contributions: Conceptualization, H.Y.; methodology, S.L.; validation, H.H., X.M. and Z.H.; resources, J.L.; writing—original draft preparation, S.L.; writing—review and editing, H.Y.; supervision, H.Y.; funding acquisition, H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Zhejiang Science and Technology Major Program on Agricultural New Variety Breeding (2021C02070-1) and Nonprofit Research Projects (CAFYBB2021QD001) of the Chinese Academy of Forestry.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

Abbreviations

CDD	Conserved domain database
GO	Gene ontology
MeJA	Methyl jasmonate
MEME	Multiple em for motif elicitation
Pfam	Database of protein families
PlantCARE	Plant cis-acting regulatory elements database
SMART	Simple modular architecture research tool

References

- Hong, J.C. General Aspects of Plant Transcription Factor Families. In *Plant Transcription Factors*; Academic Press: Cambridge, MA, USA, 2016; pp. 35–56.
- Rombauts, S.; Florquin, K.; Lescot, M.; Marchal, K.; Rouze, P.; van de Peer, Y. Computational approaches to identify promoters and cis-regulatory elements in plant genomes. *Plant Physiol.* **2003**, *132*, 1162–1176. [[CrossRef](#)] [[PubMed](#)]
- Li, X.; Xue, C.; Li, J.; Qiao, X.; Li, L.; Yu, L.; Huang, Y.; Wu, J. Genome-Wide Identification, Evolution and Functional Divergence of MYB Transcription Factors in Chinese White Pear (*Pyrus bretschneideri*). *Plant Cell Physiol.* **2016**, *57*, 824–847. [[CrossRef](#)] [[PubMed](#)]
- Kranz, H.D.; Denekamp, M.; Greco, R.; Jin, H.; Leyva, A.; Meissner, R.C.; Petroni, K.; Urzainqui, A.; Bevan, M.; Martin, C.; et al. Towards functional characterisation of the members of the R2R3-MYB gene family from *Arabidopsis thaliana*. *Plant J.* **1998**, *16*, 263–276. [[CrossRef](#)] [[PubMed](#)]
- Stracke, R.; Werber, M.; Weisshaar, B. The R2R3-MYB gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* **2001**, *4*, 447–456. [[CrossRef](#)]
- Yang, X.; Li, J.; Guo, T.; Guo, B.; Chen, Z.; An, X. Comprehensive analysis of the R2R3-MYB transcription factor gene family in *Populus trichocarpa*. *Ind. Crops Prod.* **2021**, *168*, 113614. [[CrossRef](#)]
- Dubos, C.; Stracke, R.; Grotewold, E.; Weisshaar, B.; Martin, C.; Lepiniec, L. MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* **2010**, *15*, 573–581. [[CrossRef](#)] [[PubMed](#)]
- Haga, N.; Kato, K.; Murase, M.; Araki, S.; Kubo, M.; Demura, T.; Suzuki, K.; Muller, I.; Voss, U.; Jurgens, G.; et al. R1R2R3-Myb proteins positively regulate cytokinesis through activation of *KNOLLE* transcription in *Arabidopsis thaliana*. *Development* **2007**, *134*, 1101–1110. [[CrossRef](#)]
- Dubos, C.; Le Gourrierc, J.; Baudry, A.; Huep, G.; Lanet, E.; Debeaujon, I.; Routaboul, J.M.; Alboresi, A.; Weisshaar, B.; Lepiniec, L. MYBL2 is a new regulator of flavonoid biosynthesis in *Arabidopsis thaliana*. *Plant J.* **2008**, *55*, 940–953. [[CrossRef](#)]
- Hosoda, K.; Imamura, A.; Katoh, E.; Hatta, T.; Tachiki, M.; Yamada, H.; Mizuno, T.; Yamazaki, T. Molecular structure of the GARP family of plant Myb-related DNA binding motifs of the *Arabidopsis* response regulators. *Plant Cell* **2002**, *14*, 2015–2029. [[CrossRef](#)]
- Matsui, K.; Umemura, Y.; Ohme-Takagi, M. AtMYBL2, a protein with a single MYB domain, acts as a negative regulator of anthocyanin biosynthesis in *Arabidopsis*. *Plant J.* **2008**, *55*, 954–967. [[CrossRef](#)]
- Gonzalez, A.; Zhao, M.; Leavitt, J.M.; Lloyd, A.M. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/Myb transcriptional complex in *Arabidopsis* seedlings. *Plant J.* **2008**, *53*, 814–827. [[CrossRef](#)] [[PubMed](#)]
- Li, D.; Jin, C.; Duan, S.; Zhu, Y.; Qi, S.; Liu, K.; Gao, C.; Ma, H.; Zhang, M.; Liao, Y.; et al. MYB89 Transcription Factor Represses Seed Oil Accumulation. *Plant Physiol.* **2017**, *173*, 1211–1225. [[CrossRef](#)] [[PubMed](#)]
- Stracke, R.; Ishihara, H.; Huep, G.; Barsch, A.; Mehrtens, F.; Niehaus, K.; Weisshaar, B. Differential regulation of closely related R2R3-MYB transcription factors controls flavonol accumulation in different parts of the *Arabidopsis thaliana* seedling. *Plant J.* **2007**, *50*, 660–677. [[CrossRef](#)] [[PubMed](#)]
- Zhong, R.; Lee, C.; Zhou, J.; McCarthy, R.L.; Ye, Z.H. A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. *Plant Cell* **2008**, *20*, 2763–2782. [[CrossRef](#)] [[PubMed](#)]
- Ma, D.; Reichelt, M.; Yoshida, K.; Gershenzon, J.; Constabel, C.P. Two R2R3-MYB proteins are broad repressors of flavonoid and phenylpropanoid metabolism in poplar. *Plant J.* **2018**, *96*, 949–965. [[CrossRef](#)]
- Ben-Simhon, Z.; Judeinstein, S.; Nadler-Hassar, T.; Trainin, T.; Bar-Ya'akov, I.; Borochoy-Neori, H.; Holland, D. A pomegranate (*Punica granatum* L.) WD40-repeat gene is a functional homologue of *Arabidopsis* TTG1 and is involved in the regulation of anthocyanin biosynthesis during pomegranate fruit development. *Planta* **2011**, *234*, 865–881. [[CrossRef](#)]
- Gu, Z.; Zhu, J.; Hao, Q.; Yuan, Y.W.; Duan, Y.W.; Men, S.; Wang, Q.; Hou, Q.; Liu, Z.A.; Shu, Q.; et al. A Novel R2R3-MYB Transcription Factor Contributes to Petal Blotch Formation by Regulating Organ-Specific Expression of *PsCHS* in Tree Peony (*Paeonia suffruticosa*). *Plant Cell Physiol.* **2019**, *60*, 599–611. [[CrossRef](#)]
- Schaart, J.G.; Dubos, C.; Romero De La Fuente, I.; van Houwelingen, A.; de Vos, R.C.H.; Jonker, H.H.; Xu, W.; Routaboul, J.M.; Lepiniec, L.; Bovy, A.G. Identification and characterization of MYB-bHLH-WD40 regulatory complexes controlling proanthocyanidin biosynthesis in strawberry (*Fragaria x ananassa*) fruits. *New Phytol.* **2013**, *197*, 454–467. [[CrossRef](#)]
- Xu, W.; Dubos, C.; Lepiniec, L. Transcriptional control of flavonoid biosynthesis by MYB-bHLH-WDR complexes. *Trends Plant Sci.* **2015**, *20*, 176–185. [[CrossRef](#)]

21. Raffaele, S.; Vaillau, F.; Leger, A.; Joubes, J.; Miersch, O.; Huard, C.; Blee, E.; Mongrand, S.; Domergue, F.; Roby, D. A MYB transcription factor regulates very-long-chain fatty acid biosynthesis for activation of the hypersensitive cell death response in *Arabidopsis*. *Plant Cell* **2008**, *20*, 752–767. [[CrossRef](#)]
22. To, A.; Joubes, J.; Thueux, J.; Kazaz, S.; Lepiniec, L.; Baud, S. *AtMYB92* enhances fatty acid synthesis and suberin deposition in leaves of *Nicotiana benthamiana*. *Plant J.* **2020**, *103*, 660–676. [[CrossRef](#)] [[PubMed](#)]
23. Duan, S.; Jin, C.; Li, D.; Gao, C.; Qi, S.; Liu, K.; Hai, J.; Ma, H.; Chen, M. *MYB76* Inhibits Seed Fatty Acid Accumulation in *Arabidopsis*. *Front. Plant Sci.* **2017**, *8*, 226. [[CrossRef](#)] [[PubMed](#)]
24. Chen, M.; Wang, Z.; Zhu, Y.; Li, Z.; Hussain, N.; Xuan, L.; Guo, W.; Zhang, G.; Jiang, L. The effect of transparent *TESTA2* on seed fatty acid biosynthesis and tolerance to environmental stresses during young seedling establishment in *Arabidopsis*. *Plant Physiol.* **2012**, *160*, 1023–1036. [[CrossRef](#)] [[PubMed](#)]
25. Barthole, G.; To, A.; Marchive, C.; Brunaud, V.; Soubigou-Taconnat, L.; Berger, N.; Dubreucq, B.; Lepiniec, L.; Baud, S. *MYB118* represses endosperm maturation in seeds of *Arabidopsis*. *Plant Cell* **2014**, *26*, 3519–3537. [[CrossRef](#)] [[PubMed](#)]
26. Luan, F.; Zeng, J.; Yang, Y.; He, X.; Wang, B.; Gao, Y.; Zeng, N. Recent advances in *Camellia oleifera* Abel: A review of nutritional constituents, biofunctional properties, and potential industrial applications. *J. Funct. Foods* **2020**, *75*, 104242. [[CrossRef](#)]
27. Zhao, Y.; Su, R.; Zhang, W.; Yao, G.-L.; Chen, J. Antibacterial activity of tea saponin from *Camellia oleifera* shell by novel extraction method. *Ind. Crops Prod.* **2020**, *153*, 112604. [[CrossRef](#)]
28. Lin, P.; Wang, K.; Zhou, X.; Xie, Y.; Yao, X.; Yin, H. Seed Transcriptomics Analysis in *Camellia oleifera* Uncovers Genes Associated with Oil Content and Fatty Acid Composition. *Int. J. Mol. Sci.* **2018**, *19*, 118. [[CrossRef](#)]
29. Gong, W.; Song, Q.; Ji, K.; Gong, S.; Wang, L.; Chen, L.; Zhang, J.; Yuan, D. Full-Length Transcriptome from *Camellia oleifera* Seed Provides Insight into the Transcript Variants Involved in Oil Biosynthesis. *J. Agric. Food Chem.* **2020**, *68*, 14670–14683. [[CrossRef](#)]
30. Su, M.H.; Shih, M.C.; Lin, K.H. Chemical composition of seed oils in native Taiwanese *Camellia* species. *Food Chem.* **2014**, *156*, 369–373. [[CrossRef](#)]
31. Lin, P.; Wang, K.; Wang, Y.; Hu, Z.; Yan, C.; Huang, H.; Ma, X.; Cao, Y.; Long, W.; Liu, W.; et al. The genome of oil-Camellia and population genomics analysis provide insights into seed oil domestication. *Genome Biol.* **2022**, *23*, 14. [[CrossRef](#)]
32. Mistry, J.; Finn, R.D.; Eddy, S.R.; Bateman, A.; Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **2013**, *41*, e121. [[CrossRef](#)] [[PubMed](#)]
33. Bailey, T.L.; Williams, N.; Misleh, C.; Li, W.W. MEME: Discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* **2006**, *34*, W369–W373. [[CrossRef](#)] [[PubMed](#)]
34. 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](#)] [[PubMed](#)]
35. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [[CrossRef](#)]
36. Gómez-Rubio, V. ggplot2—Elegant Graphics for Data Analysis (2nd Edition). *J. Stat. Softw.* **2017**, *77*, 1–3. [[CrossRef](#)]
37. Wang, Y.; Tang, H.; Debarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [[CrossRef](#)]
38. Li, K.B. ClustalW-MPI: ClustalW analysis using distributed and parallel computing. *Bioinformatics* **2003**, *19*, 1585–1586. [[CrossRef](#)]
39. Capella-Gutierrez, S.; Silla-Martinez, J.M.; Gabaldon, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [[CrossRef](#)]
40. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)]
41. Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [[CrossRef](#)]
42. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*, 907–915. [[CrossRef](#)] [[PubMed](#)]
43. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)] [[PubMed](#)]
44. Pertea, M.; Pertea, G.M.; Antonescu, C.M.; Chang, T.C.; Mendell, J.T.; Salzberg, S.L. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **2015**, *33*, 290–295. [[CrossRef](#)] [[PubMed](#)]
45. Langfelder, P.; Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **2008**, *9*, 559. [[CrossRef](#)]
46. Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M.C.; Estreicher, A.; Gasteiger, E.; Martin, M.J.; Michoud, K.; O'Donovan, C.; Phan, I.; et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **2003**, *31*, 365–370. [[CrossRef](#)]
47. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)]
48. Litwack, G. *Nucleic Acids and Molecular Genetics*. In *Human Biochemistry*; Academic Press: Cambridge, MA, USA, 2018; pp. 257–317.

49. Lee, H.G.; Park, B.-Y.; Kim, H.U.; Seo, P.J. MYB96 stimulates C18 fatty acid elongation in Arabidopsis seeds. *Plant Biotechnol. Rep.* **2015**, *9*, 161–166. [[CrossRef](#)]
50. Steiner-Lange, S.; Unte, U.S.; Eckstein, L.; Yang, C.; Wilson, Z.A.; Schmelzer, E.; Dekker, K.; Saedler, H. Disruption of *Arabidopsis thaliana* MYB26 results in male sterility due to non-dehiscent anthers. *Plant J.* **2003**, *34*, 519–528. [[CrossRef](#)]
51. Gonzalez, A.; Mendenhall, J.; Huo, Y.; Lloyd, A. TTG1 complex MYBs, MYB5 and TT2, control outer seed coat differentiation. *Dev. Biol.* **2009**, *325*, 412–421. [[CrossRef](#)]
52. Li, S.F.; Milliken, O.N.; Pham, H.; Seyit, R.; Napoli, R.; Preston, J.; Koltunow, A.M.; Parish, R.W. The *Arabidopsis* MYB5 transcription factor regulates mucilage synthesis, seed coat development, and trichome morphogenesis. *Plant Cell* **2009**, *21*, 72–89. [[CrossRef](#)]
53. Newman, L.J.; Perazza, D.E.; Juda, L.; Campbell, M.M. Involvement of the R2R3-MYB, AtMYB61, in the ectopic lignification and dark-photomorphogenic components of the det3 mutant phenotype. *Plant J.* **2004**, *37*, 239–250. [[CrossRef](#)]
54. Penfield, S.; Meissner, R.C.; Shoue, D.A.; Carpita, N.C.; Bevan, M.W. MYB61 is required for mucilage deposition and extrusion in the Arabidopsis seed coat. *Plant Cell* **2001**, *13*, 2777–2791. [[CrossRef](#)] [[PubMed](#)]
55. Verma, N.; Burma, P.K. Regulation of tapetum-specific A9 promoter by transcription factors AtMYB80, AtMYB1 and AtMYB4 in *Arabidopsis thaliana* and *Nicotiana tabacum*. *Plant J.* **2017**, *92*, 481–494. [[CrossRef](#)] [[PubMed](#)]
56. Millar, A.A.; Gubler, F. The Arabidopsis GAMYB-like genes, MYB33 and MYB65, are microRNA-regulated genes that redundantly facilitate anther development. *Plant Cell* **2005**, *17*, 705–721. [[CrossRef](#)]
57. Agarwal, P.; Mitra, M.; Banerjee, S.; Roy, S. MYB4 transcription factor, a member of R2R3-subfamily of MYB domain protein, regulates cadmium tolerance via enhanced protection against oxidative damage and increases expression of PCS1 and MT1C in *Arabidopsis*. *Plant Sci.* **2020**, *297*, 110501. [[CrossRef](#)] [[PubMed](#)]
58. Devaiah, B.N.; Madhuvanthi, R.; Karthikeyan, A.S.; Raghothama, K.G. Phosphate starvation responses and gibberellic acid biosynthesis are regulated by the MYB62 transcription factor in *Arabidopsis*. *Mol. Plant* **2009**, *2*, 43–58. [[CrossRef](#)]
59. Jung, C.; Seo, J.S.; Han, S.W.; Koo, Y.J.; Kim, C.H.; Song, S.I.; Nahm, B.H.; Choi, Y.D.; Cheong, J.J. Overexpression of AtMYB44 enhances stomatal closure to confer abiotic stress tolerance in transgenic *Arabidopsis*. *Plant Physiol.* **2008**, *146*, 623–635. [[CrossRef](#)] [[PubMed](#)]
60. Kim, J.H.; Nguyen, N.H.; Jeong, C.Y.; Nguyen, N.T.; Hong, S.W.; Lee, H. Loss of the R2R3 MYB, AtMyb73, causes hyper-induction of the SOS1 and SOS3 genes in response to high salinity in *Arabidopsis*. *J. Plant Physiol.* **2013**, *170*, 1461–1465. [[CrossRef](#)]
61. Keereetaweep, J.; Liu, H.; Zhai, Z.; Shanklin, J. Biotin Attachment Domain-Containing Proteins Irreversibly Inhibit Acetyl CoA Carboxylase. *Plant Physiol.* **2018**, *177*, 208–215. [[CrossRef](#)]
62. Toll-Riera, M.; Rado-Trilla, N.; Martys, F.; Alba, M.M. Role of low-complexity sequences in the formation of novel protein coding sequences. *Mol. Biol. Evol.* **2012**, *29*, 883–886. [[CrossRef](#)]
63. Larkin, J.C.; Oppenheimer, D.G.; Pollock, S.; Marks, M.D. *Arabidopsis* GLABROUS7 Gene Requires Downstream Sequences for Function. *Plant Cell* **1993**, *5*, 1739–1748. [[CrossRef](#)] [[PubMed](#)]
64. Stracke, R.; Holtgräwe, D.; Schneider, J.; Pucker, B.; Sörensen, T.R.; Weisshaar, B. Genome-wide identification and characterisation of R2R3-MYB genes in sugar beet (*Beta vulgaris*). *BMC Plant Biol.* **2014**, *14*, 249. [[CrossRef](#)] [[PubMed](#)]