forests

MDPI

Article

Genome-Wide Identification and Expression Analyses of the 4-Coumarate: CoA Ligase (*4CL*) Gene Family in *Eucommia ulmoides*

Jian Zhong [1,2,†], Jun Qing [1,2,†], Qi Wang [3], Chenlu Liu [3], Hongyan Du [1,2], Panfeng Liu [1,2], Qingxin Du [1,2], Lanying Du [1,2] and Lu Wang [1,2,*]

[1] Research Institute of Non-Timber Forestry, Chinese Academy of Forestry (CAF), Zhengzhou 450003, China
[2] Key Laboratory of Non-Timber Forest Germplasm Enhancement & Utilization of State Forestry and Grassland Administration, Zhengzhou 450003, China
[3] Non-Timber Forest R&D Center, Central South University of Forestry and Technology, Changsha 410004, China
* Correspondence: wanglu181716@163.com
† These authors contributed equally to this work.

Abstract: 4-Coumarate: coenzyme A ligase (*4CL*) is an important rate-limiting enzyme in phenylpropanoid metabolism that plays an important role in the biosynthesis of phenylpropanoid, flavonoid, lignin and other secondary metabolites in plants. However, the family members and functions have not been identified in *Eucommia ulmoides* (*E. ulmoides*). In this study, 35 Eu4CLs were identified in the *E. ulmoides* genome, and the gene structure, conserved domain, evolutionary relationship and expression pattern were comprehensively analyzed. The results show that 35 Eu4CLs were assembled into three subgroups according to the classification in Arabidopsis, where Eu4CLs in the same subgroup had similar gene structures and conserved protein motifs. Putative cis-element analysis of Eu4CL promoter regions uncovered numerous elements related to the response of stress and plant hormones. Expression patterns showed that Eu4CL4/5/13/34 expression levels were positively related to chlorogenic acid content in different periods, which indicate that the synthesis of chlorogenic acid in *E. ulmoides* was regulated by multiple genes, and the genes regulating the synthesis of chlorogenic acid in different tissues were different. In addition, nine selected Eu4CL genes showed different expression patterns under cold, WeJA (methyl jasmonate), and ethylene by quantitative reverse transcription-PCR (qRT-PCR) assay, suggesting that Eu4CL genes not only play an important role in the synthesis of chlorogenic acid, but also plays an important role in the process of biotic and abiotic stress. Taken together, these findings provide theoretical reference for further exploring the molecular characteristics and biological functions of Eu4CL genes.

Keywords: *Eucommia ulmoides*; *4CL* gene family; chlorogenic acid; bioinformatic analysis; secondary metabolites

Citation: Zhong, J.; Qing, J.; Wang, Q.; Liu, C.; Du, H.; Liu, P.; Du, Q.; Du, L.; Wang, L. Genome-Wide Identification and Expression Analyses of the 4-Coumarate: CoA Ligase (*4CL*) Gene Family in *Eucommia ulmoides*. *Forests* **2022**, *13*, 1253. https://doi.org/10.3390/f13081253

Academic Editor: Richard Dodd

Received: 22 June 2022
Accepted: 4 August 2022
Published: 8 August 2022

1. Introduction

Eucommia ulmoides, a plant belonging to the Eucommiaceae family, has been used for more than 2000 years. It is a unique economic tree species in China and has high medicinal value and ecological maintenance function [1,2]. The main medicinal components of *E. ulmoides* include phenylpropanoids, flavonoids, lignans and iridoids, among which chlorogenic acid is a secondary metabolite of phenylpropanoids and an important dietary polyphenol with biological activity [3]. It has the functions of antioxidation, anti-inflammation, nerve protection, anti-obesity, anti-virus and anti-hypertension [4,5]. In addition, chlorogenic acid is a natural antioxidant, which plays a key role in plant response to abiotic and biotic stresses such as drought, low temperature, pests and pathogen infection, and enhances plant adaptability to severe environments due to its strong ability

to scavenge free radicals [6,7]. In recent years, the synthesis pathway and related gene function of chlorogenic acid in plants have been gradually clarified. It is found that the expression of key enzymes in the synthesis pathway of chlorogenic acid can be regulated by abiotic and biotic treatment, thereby increasing the content of chlorogenic acid, which has become one of the effective ways to obtain chlorogenic acid [8,9].

The key enzymes of the chlorogenic acid synthesis pathway mainly contain phenylalanine ammonia lyase (*PAL*), cinnamic acid 4-hydroxylase (*C4H*), 4-coumaric coenzyme A ligase (*4CL*) and quinic acid hydroxycinnamoyl transferase (*HQT*) [10]. Among them, *4CLs* are the key enzymes in the third step of the phenylpropanoid metabolic pathway that catalyze cinnamic acid to generate the corresponding coenzyme A ester [11]. They participate in phenylpropanoid synthesis pathways and synthesize chlorogenic acid and other compounds [12,13]. Since the first *4CL* gene was cloned from Petroselinum hortense in 1987, it has been studied in several species [14]. *A. thaliana* contains four *4CL* genes encoding catalytically active proteins, in which *At4CL3* is highly expressed in mature leaves and flowers, mainly regulating the synthesis of flavonoids [15]. There are two completely different *4CL* genes in *P. tomentosa*. *Pt4CL1* is specifically expressed in lignification tissues and participates in lignin biosynthesis. *Pt4CL2* is specifically expressed in the epidermis of stems and leaves and is involved in the biosynthesis of flavonoids and other phenolic compounds [16].

In summary, the current studies on the function of *4CL* genes mainly focus on regulating the biosynthesis of lignin and flavonoids, while there are few studies on the synthesis pathway of chlorogenic acid. Therefore, in this study, *E. ulmoides* with high chlorogenic acid content was used as the experimental material. Based on the whole genome of *E. ulmoides*, the members of the *4CL* gene family were identified, and their gene structure, conserved motif, evolutionary relationship and expression pattern were analyzed. This study provides theoretical reference for further exploring the molecular characteristics and biological functions of the *E. ulmoides 4CL* gene.

2. Materials and Methods

2.1. Plant Materials

The leaves and bark of *E. ulmoides* were obtained from the Mengzhou Test Base, Institute of Economic Forestry, Chinese Academy of Forestry. 'Huazhong 8' *E. ulmoides* were selected as typical sample plants and sampled every month after 15 April. Sampling occurred 6 times on 15 April, 15 May, 15 June, 15 July, 15 August and 15 September. The leaves without diseases and pests were randomly collected. The bark was selected as strips of length of 50 cm and width of 5 cm from the main cadres approximately 130 cm above the ground, and then the leaves and bark were brought back to the laboratory and stored in a ventilated dry room for natural air drying.

2.2. Determination of Chlorogenic Acid

After natural drying, the plant materials were crushed and screened, and the content of chlorogenic acid in the bark and leaves was determined by high-performance liquid chromatography (HPLC) [17].

2.2.1. Materials Treatment

Amounts of 0.1 g of the bark and leaf powder were weighed and added into 50% methanol solution to fix the volume to 5 mL, then sonicated for 40 min, and finally filtered through a 0.22 μm microporous membrane to obtain the test solution.

2.2.2. Reaction Conditions

Thermo Hypersil Gold (250 mm × 4.6 mm, 5 μm), methanol (A)—0.5% phosphoric acid aqueous solution (B) as mobile phase, gradient elution, 0–30 min, 5%–15% A; 30–70 min, 15%–30% A; 70–100 min, 30%–54% A; detection wavelength 206 nm (0–15 min), 236 nm

(15–55 min), 255 nm (55–100 min); volume flow rate 1.0 mL/min; injection volume 8 µL; column temperature 30 °C.

2.3. Identification of the 4CL Gene Family in E. ulmoides

Eucommia genome data were downloaded from Genome Warehouse (https://ngdc.cncb.ac.cn/gwh/Assembly/13/show, (accessed on 21 June 2022)) (PRJCA000677). *Arabidopsis 4CL* gene sequences were downloaded from TATR (https://www.arabidopsis.org/, (accessed on 21 June 2022)). Two *4CL* genes (*EUC10101* and *EUC15175*) annotated by KEGG were used for the blast search, and the *4CL* conserved domain (PF00501) HMM file was downloaded from the Pfam database (http://pfam.sanger.ac.uk/, (accessed on 21 June 2022)). The HMM file was used to query the 4CLs in *E. ulmoides* using the hmmsearch program of HMMER 3.0 software, with an E-value threshold of 1.0E-5. Three incomplete sequences were removed by online alignment with NCBI CDD (http://www.ncbi.nlm.nih.gov/cdd/, (accessed on 21 June 2022)) and Pfam (https://pfam.xfam.org/, (accessed on 21 June 2022)), resulting in 35 sequences with the 4CL conserved domain.

2.4. Bioinformatic Analysis of EU4CL Genes

Based on Eu4CL gene full-length sequences and CDS sequences of *E. ulmoides* genome data, exon–intron structure analysis was performed using the online Gene Structure Display Server (GSDS) (http://gsds.cbi.pku.edu.cn/, (accessed on 21 June 2022)). MEME (http://meme.nbcr.net/meme/tools/meme, (accessed on 21 June 2022)) was used to analyze the conserved domain of Eu4CL proteins with the following parameters: number of repeats, any; maximum number of motifs, 10; and optimal length of each motif, 6–30. ExPASy (https://web.expasy.org/protparam/, (accessed on 21 June 2022)) was used to predict protein physicochemical features. Subcellular localization of EU4CL genes was predicted by online software CELLO v. 2.5 (http://cello.life.nctu.edu.tw/, (accessed on 21 June 2022)) [18]. Thirty-five Eu4CL protein sequences were compared with the 4CL sequences of A. thaliana using clustalW, and the phylogenetic tree of the 4CL gene family was constructed using MEGA 7.0 Maximum Likelihood (ML) and 500 bootstrap tests (Bootstrap = 500). The 1500 bp upstream region of the 4CL sequences was intercepted and predicted by the plantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/, (accessed on 21 June 2022)) website.

2.5. RNA Extraction and Quantitative Reverse Transcription-PCR (qRT-PCR) Assay

EZgene Plant Easy Spin RNA Miniprep Kit (R6611) was used to extract total RNA from *E. ulmoides* leaves and bark at different growth stages. Reverse transcription of RNA into cDNA was conducted according to Novoprotein E047-01B reverse transcription kit instructions. The primers were designed by Primer 5, and the reference gene was actin. Primers are shown in Table 1. Expression analysis was performed using ChamQ Universal SYBR qRT-PCR Master Mix fluorescent dye with the DLAB Precision 96 fluorescence quantitative PCR instrument. The reaction program was as follows: 95 °C 5 min, 95 °C 10 s, 60 °C 10 s, 72 °C 10 s, 75 °C 5 s, a total of 40 cycles. Repeat 3 times. SPSS software was used to analyze the significant difference (* means significant difference ($p < 0.05$), * means extremely significant difference ($p < 0.01$).

Table 1. The primers used in qRT-PCR analysis.

Name	Primer Sequences
Eu4CL4-F	TCGTTGGCGACTCTTCCATC
Eu4CL4-R	GCGGGATCCTCAATTGACCT
Eu4CL5-F	GTTCGCCTCGTCTTGTCGTA
Eu4CL5-R	CGTCGTGCCGGAGGTATAAT
Eu4CL13-F	CGCTATTACGGTTCGCTGGA

Table 1. *Cont.*

Name	Primer Sequences
Eu4CL13-R	TATTCTGTGCACGCCGACTT
Eu4CL15-F	TCTTGCCTTTACTGTGGCTG
Eu4CL15-R	GCCGCAGTGTCTAGCTGAT
Eu4CL22-F	ATTCTTGCCGGTGGCTTGTT
Eu4CL22-R	GAACGGAAGGCCTCCACAAT
Eu4CL26-F	CTTATCTCGTCCGCTTCGCA
Eu4CL26-R	GTGCTAGAGGGACTGCAACA
Eu4CL32-F	GGCATGCAATGCTCATGGTC
Eu4CL32-R	AAGAAACTCCGCCGAAGCTC
Eu4CL34-F	AGGGGATACCGAAGGGTTCA
Eu4CL34-R	ACTTGACAGCATCTGGACCG
Eu4CL35-F	ATGGGCAGGTGTTTGATCGT
Eu4CL35-R	CACTCCGCTCGAGTTTCAGA
Actin-F	TTGTTAGCAACTGGGATGATATGG
Actin-R	CAGGGTGTTCTTCAGGAGCAA

2.6. Analysis of 4CL Gene Expression Patterns in E. ulmoides

Raw RNA-seq data of salt stress in different tissues (fruit, leaves and bark) of *E. ulmoides* were downloaded from NCBI (https://www.ncbi.nlm.nih.gov, (accessed on 21 June 2022)) with the landing numbers PRJNA329457, SRR2170964 and PRJNA321358, respectively. The transcriptome data of male and female flower bud differentiation were obtained from the E ulmoides team of the Economic Forest Research Institute, Chinese Academy of Forestry Sciences [19]. The TBtools program was used to display the heatmap of gene expression [20]. The STRING protein interaction database (http://string-db.org/, (accessed on 21 June 2022)) was used to analyze the interactions of 35 Eu4CL proteins, and the plant model was Arabidopsis. Pearson correlation analysis was performed between Eu4CL genes with expression levels greater than 1 and chlorogenic acid accumulation using Spass 20.

3. Results

3.1. Identification and Protein Physicochemical Feature Analysis of 4CL in E. ulmoides

Based on the hidden Markov model and the two existing genes, BLAST and HMMER3.0 were used to search for the protein sequences of *E. ulmoides*, and ClustalW was used to construct the specific 4CL model. HMMER3.0 was then used to search for the whole-genome protein sequences of *E. ulmoides*. The E-value was set to be less than 0.1, and 35 sequences were finally obtained. The 35 genes were renamed *Eu4CL1–Eu4CL35*, and the gene structure and distribution of these genes were further analyzed.

The physicochemical features of *Eu4CL* proteins were analyzed using the ExPasy online website (Table 2). The results showed that the average length of *Eu4CL* proteins was 518 aa, the longest was *Eu4CL4* (802 aa), and the shortest was *Eu4CL1* (119 aa). The molecular weight of *Eu4CL* proteins was 12.61–89.16 kDa; the theoretical isoelectric point ranged from 5.31 to 9.52, of which 23 proteins' pI values were less than 7, indicating that most were alkaline proteins. The instability index of *Eu4CL* proteins was 25.02–48.55, and the instability index of 10 proteins was greater than 40, indicating unstable proteins. The instability index of 25 proteins was less than 40, indicating stable proteins. The aliphatic index was between 78.27 and 104.04. The average hydrophobic index was −0.344 to 0.246. Subcellular localization of proteins can provide important information for protein function research [18]. Subcellular localization prediction showed that *Eu4CL* proteins were mainly distributed in plasma membrane, cytoplasm and chloroplast, and *Eu4CL17* and *Eu4CL29* were located in mitochondria and nuclei, respectively. It is speculated that the different distribution positions of *Eu4CL* proteins in plants may be related to functional diversity.

Table 2. Detailed physicochemical characteristics of 4CL proteins of *Eucommia ulmoides*.

Gene Name	Number of Amino Acids	Molecular Weight (kDa)	Theoretical pI	Instability Index	Aliphatic Index	Grand Average of Hydropathicity	Subcellular Location
EU4CL1	119	12.61	6.56	25.02	91.85	0.246	plasma membrane
EU4CL2	378	41.06	9.13	39.44	95.16	0.044	cytoplasm
EU4CL3	254	27.78	6.33	42.75	78.27	−0.315	plasma membrane
EU4CL4	802	89.16	6.41	37.2	81.2	−0.23	cytoplasm
EU4CL5	555	61.04	6.93	35.41	88.22	−0.033	chloroplast
EU4CL6	254	28.01	7.02	43.38	78.27	−0.344	cytoplasm
EU4CL7	258	28.52	8.25	38	81.94	−0.29	cytoplasm
EU4CL8	695	76.85	6.89	29.23	88.78	−0.074	plasma membrane
EU4CL9	515	55.54	5.79	39.23	92.83	−0.062	plasma membrane
EU4CL10	521	57.76	6.69	36.28	89.81	−0.139	cytoplasm
EU4CL11	317	35.04	7.14	35.34	84.79	−0.205	cytoplasm
EU4CL12	423	45.84	6.32	47.09	99.15	0.115	cytoplasm
EU4CL13	771	85.09	9.52	46.67	84.16	−0.129	plasma membrane
EU4CL14	552	60.61	6.15	35.05	96.92	0.084	chloroplast
EU4CL15	564	61.94	6.17	36.24	101.44	0.109	chloroplast
EU4CL16	703	77.19	6.39	34.65	92.36	−0.08	chloroplast
EU4CL17	520	58.48	5.35	34.28	96.15	−0.054	mitochondrion
EU4CL18	623	70.14	8.21	38.97	92.3	−0.21	plasma membrane
EU4CL19	518	56.67	6.38	34.99	103.46	0.078	plasma membrane
EU4CL20	475	51.62	5.54	40.09	96.25	0.029	cytoplasm
EU4CL21	192	21.54	8.77	42.66	87.34	−0.264	chloroplast
EU4CL22	668	74.74	8.29	48.55	96.32	−0.183	cytoplasm
EU4CL23	658	73.50	6.04	31.5	91.09	−0.071	chloroplast
EU4CL24	346	37.73	5.85	40.53	94.36	0.046	cytoplasm
EU4CL25	375	40.42	8.52	39.11	91.55	0.023	plasma membrane
EU4CL26	627	68.64	6.66	36.45	89.76	−0.137	plasma membrane
EU4CL27	674	75.82	5.79	38.76	83.04	−0.199	chloroplast
EU4CL28	553	60.72	8.2	40.53	88.12	−0.06	chloroplast
EU4CL29	551	60.13	6.4	36.81	86.46	−0.055	nucleus
EC4CL30	612	66.96	9.01	32.81	100.18	0.174	chloroplast
EU4CL31	554	60.35	6.44	44.79	95.13	0.022	chloroplast
EU4CL32	665	74.00	5.9	35.46	89.17	−0.126	cytoplasm
EU4CL33	443	48.09	5.31	27.4	104.04	0.147	chloroplast
EU4CL34	693	76.44	6.45	33.66	91.96	−0.038	chloroplast
EU4CL35	731	79.93	8.26	30.33	91	−0.068	cytoplasm

3.2. Conserved Motifs and Gene Structure Analysis

To further study the structural characteristics and distribution of *Eu4CL* proteins, the amino acid sequences of *Eu4CL* proteins were analyzed using MEME online software. Finally, 20 motifs (Figure 1a,b) were identified, of which four were conserved motifs. The motifs 1, 2, 3 and 4 (Figure 1b) correspond to the conserved domains Box IV [GWLHTGD], Box I [SSGTTGLPKGV], Box II [GEICIRG] and Box III [QGYGMTE], respectively. Box I and Box II were AMP-binding regions and enzyme catalytic sites [21]. The *Eu4CL1* sequence contained only one conserved motif and three motifs in total. It was also the protein with the least number of motifs. The other *Eu4CL* sequences contained two or more conserved motifs. Multiple sequence alignment analysis showed that the *4CL* genes of *E. ulmoides* could be divided into three groups. Among them, 12 *Eu4CL* sequences in Group I mainly contained motifs 2/5/8/13/4/3, 10 *Eu4CL* sequences in Group II mainly contained motifs 4/10/3/1/6/16/7/9, and 13 EU4CL sequences in Group III mainly contained motifs 17/11/2/13/4/12/3/1/6. There were differences in the number of motifs in different sequences, which might be related to the functional diversity of the *4CL* gene family in *E. ulmoides*.

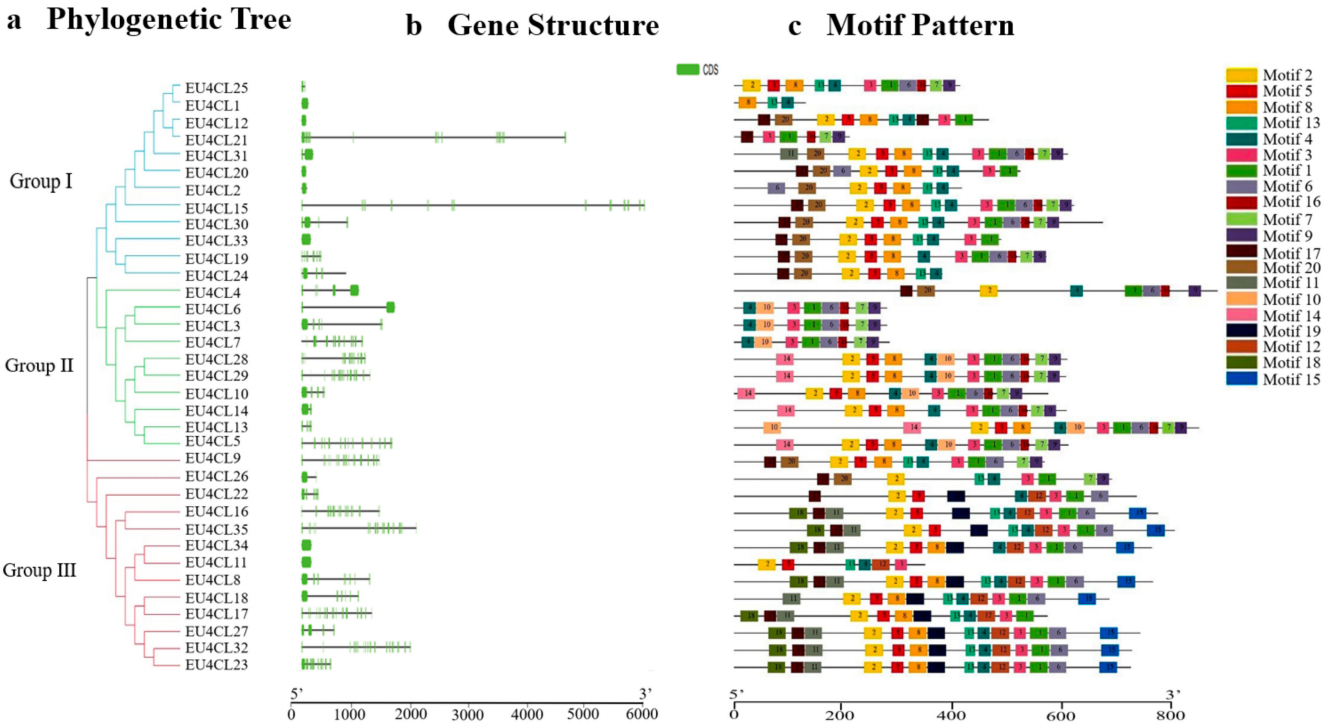

Figure 1. Phylogenetic relationships, gene structure and architecture of conserved protein motifs in *4CL* genes from *E. ulmoides*. (**a**) The phylogenetic tree was constructed based on the full-length sequences of *E. ulmoides 4CL* proteins using MEGA 7 software. Details of clusters are shown in different colors. (**b**) Exon–intron structure of *E. ulmoides 4CL* genes, green boxes indicate exons; gray lines indicate introns. The number indicates the phases of corresponding introns. (**c**) The motif composition of *E. ulmoides 4CL* proteins. The motifs, numbers 1–20, are displayed in different-colored boxes. The length of protein can be estimated using the scale at the bottom.

Gene structure analysis showed that the average length of the *Eu4CL* genes was 1556 bp. The *Eu4CL4* gene had the longest length of 2406 bp, containing 18 exons. The *EU4CL1* gene had the shortest length of 357 bp, containing two exons. There were significant differences in the number of introns and exons in the *Eu4CL* genes, and the number of exons was between 1 and 23. Among them, the most abundant exons were found in *Eu4CL8* and *Eu4CL34*, and there were seven genes with only one exon (Figure 1c). Therefore, the number of exons and introns had no significant correlation with gene length.

3.3. Phylogenetic Analysis of the 4CL Gene Family in E. ulmoides

To study the phylogenetic relation of *Eu4CL*s, 35 *Eu4CL*s and 45 *At4CL* amino acid sequences were analyzed using MEGA7.0. The *4CL* gene family of *E. ulmoides* was divided into six categories (Figure 2). There were nine *Eu4CL*s in Class 1, three *Eu4CL*s in Class 2, two *Eu4CL*s in Class 3, ten *Eu4CL*s in Class 4, only one *Eu4CL* in Class 5 and ten *Eu4CL*s in Class 6. The *4CL* genes evolved to different degrees in the evolutionary process. At the same time, the *4CL* genes of *A. thaliana* and *E. ulmoides* existed in each group, indicating that they had high homology.

The different-colored arcs indicate different groups (or subgroups) of *4CL* domains. The green solid circles and red circles represent the *4CL* domain from *E. ulmoides* and *Arabidopsis*, respectively. *4CL* proteins from *Arabidopsis* with the prefix 'At' indicate '*At4CL*'.

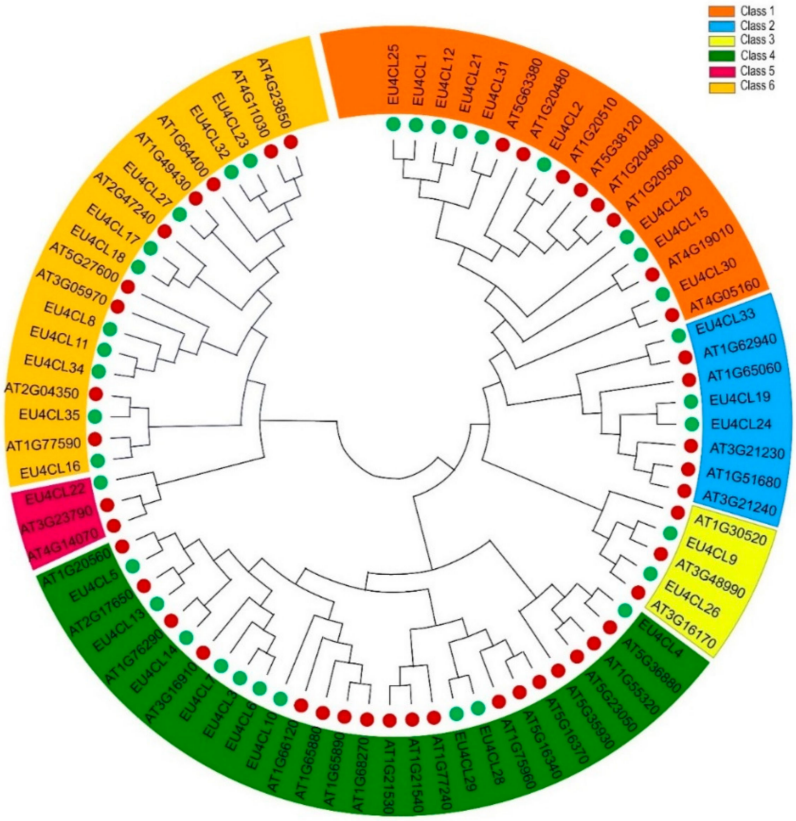

Figure 2. Unrooted phylogenetic tree representing relationships among *4CL* domains of *E. ulmoides* and *Arabidopsis*.

3.4. Interaction Network and Promoter Cis-Acting Element Analysis of Eu4CLs

The STRING protein interaction database was used to predict the functional relationship and co-expression of 35 *Eu4CL* proteins, with *Arabidopsis* as the model plant (Figure 3). Except for AAE5, AAE11, AAE1 and AAE3 proteins, the proteins had correlations with each other, and each protein interacted with at least two or more proteins. According to the degree values, the top six *4CL* proteins were selected as *4CL8/11/16/23/32/35*, which indicated that they played significant roles in the whole protein interaction network. Therefore, it is speculated that the *4CL* gene family completes a metabolic process through the cooperation of multiple proteins during plant development [22].

Promoters play an important role in the regulation of gene expression, and cis-acting elements regulate the initiation and transcription efficiency of gene transcription by combining with transcription factors [23,24]. PlantCARE was used to predict the cis-acting elements of *Eu4CLs* by extracting the promoter region sequence of 1500 bp upstream of *Eu4CLs*. The promoter region of *Eu4CL* contained 13 cis-elements, including drought-inducible elements, stress-responsive elements, light-responsive elements, low-temperature-responsive elements, methyl jasmonate-responsive elements, circadian rhythm regulatory elements and salicylic acid-responsive elements (Figure 4). Among these, the light-responsive cis-acting elements were the most abundant. The promoter regions of genes contained light-responsive elements, except for *Eu4CL1/23/27*. The *Eu4CL2* gene contained the least cis-acting elements (only two), while *Eu4CL20* contained the most cis-acting elements (21). Different 4CL gene family members have different types and numbers of cis-acting elements, which may explain the different functions and expressions of *Eu4CL* gene family members.

Blocks of different colors represent light-responsive elements, low-temperature-responsive elements, salicylic acid-responsive elements, abscisic acid-responsive elements, drought-responsive elements, MeJA-responsive elements, auxin-responsive elements, endosperm-responsive elements, gibberellin-responsive elements, and defense and stress-responsive elements.

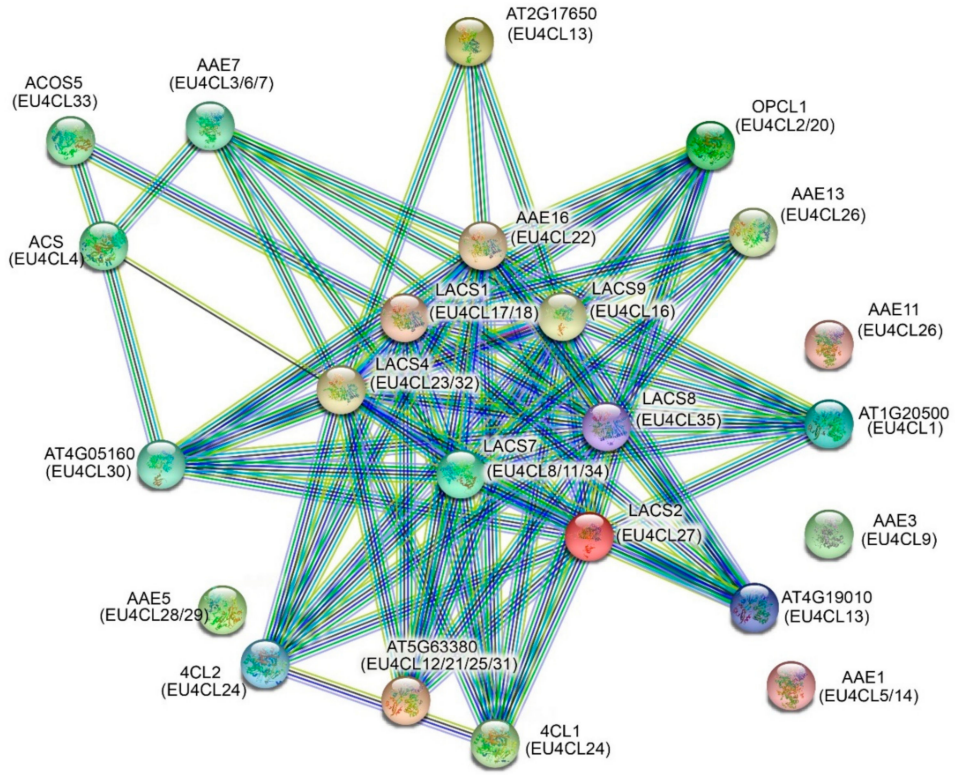

Figure 3. The predicted protein interaction network of the Eu4CL protein based on orthologs in Arabidopsis using STRING database.

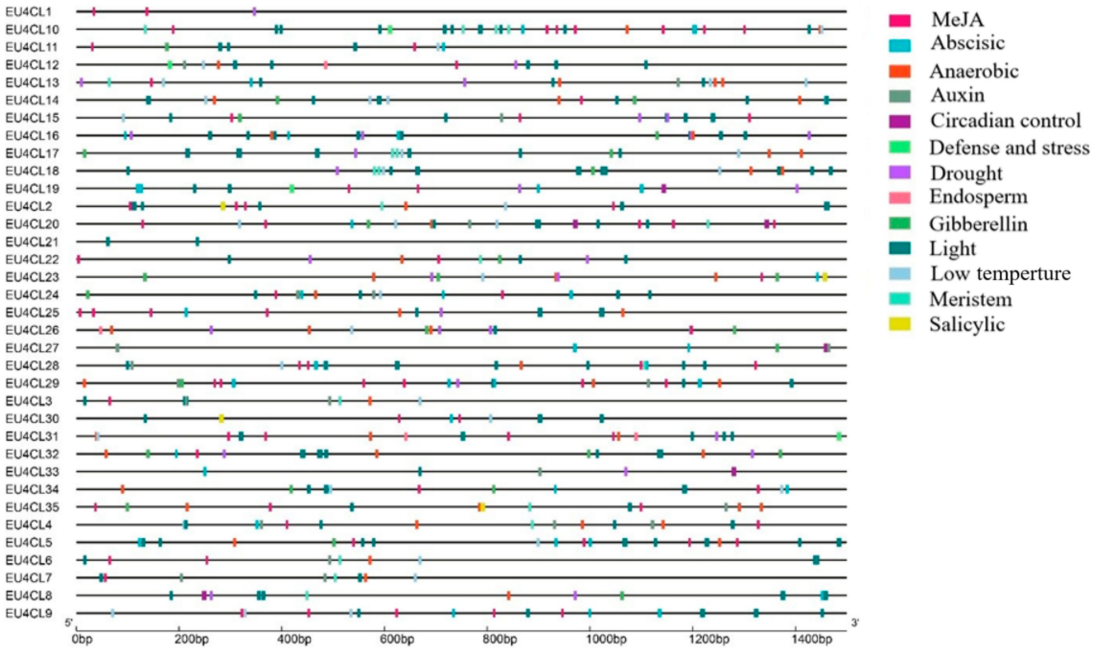

Figure 4. The cis-acting elements Analysis of Eu4CL genes promoters.

3.5. Analysis of 4CL Expression Patterns in E. ulmoides

The expression levels of *Eu4CL* genes were analyzed based on transcriptome data, and the results showed that *Eu4CL* genes were differentially expressed in bark, leaves, flower buds and fruit of *E. ulmoides* at different developmental stages (Figure 5). *Eu4CL32* was highly expressed in all tissues. The expression trend of *Eu4CL15/1618* was significantly downregulated during the development of bark, fruit and leaves, while the expression

trend of *Eu4CL28/9* was significantly upregulated. The expression trend of *Eu4CL23/22* was significantly downregulated with the development of floral buds, and the expression trend of *Eu4CL8/16/29* was upregulated, indicating that these genes may be related to floral organ development. However, *Eu4CL17/15/9* was specifically expressed on 14 June (the critical period of floral bud differentiation), suggesting that they may be associated with key hormones involved in floral organ differentiation. The expression of *EU4CLs* in the roots was significantly different from that in other tissues. The expression of *Eu4CL6/11/10/7/21/28/2* was specifically expressed in the roots, while the expression of *Eu4CL30/18/15/16/8/28/27* was significantly lower than that in other tissues, indicating that the *Eu4CL* gene family has tissue-specific expression and functional diversity.

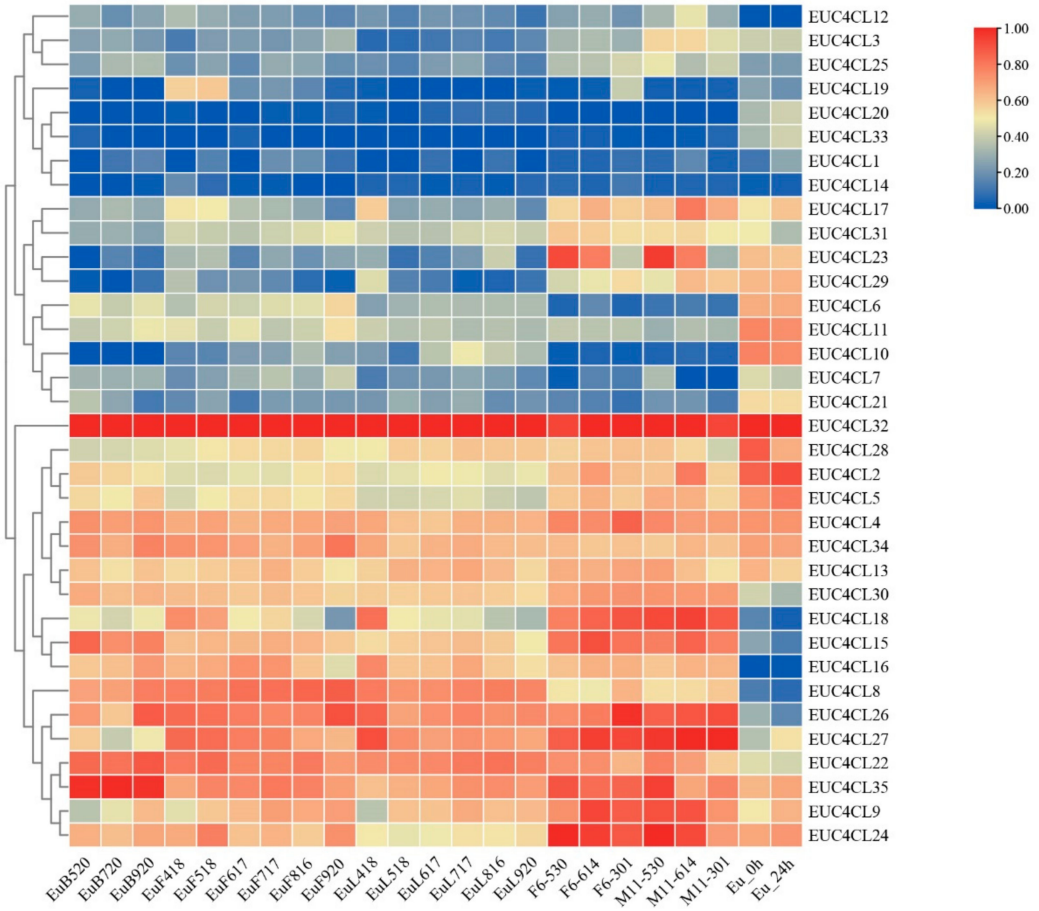

Figure 5. Expression profiles of Eu4CL genes during the development of flowers and roots under salt treatment.

Note: EuL, EuF and EuB indicate leaves, fruit and bark, respectively. The following number indicates the date; Eu0h_1 and Eu24h_1 indicate CK and salt treatment 24 h later. F6 indicates female flowers, and M11 indicates male flower buds. The following number indicates the date. The clustering tree was constructed by hierarchical clustering using the average linkage method.

To further confirm whether the expression of *Eu4CL* genes was influenced by different abiotic stresses and hormonal treatments, nine *Eu4CL* members, whose expression levels were relatively high across different tissues, were carefully selected from 35 *E. ulmoides Eu4CL* genes. qRT-PCR experiments were further performed to analyze their expression patterns in response to different treatments (Figures 6–8). Overall, some Eu4CL genes were significantly induced/repressed by multiple treatments. For instance, Eu4CL13 and *Eu4CL26* significantly responded to cold, WeJA (methyl jasmonate), and ethylene treatments. *Eu4CL15* and *Eu4CL32* were induced by all tested treatments except WeJA

stress. In contrast, multiple *Eu4CL* genes were simultaneously induced by one treatment. For example, *Eu4CL4*, *Eu4CL22* and *Eu4CL35* were induced by ethylene, WeJA and cold treatment, respectively. Interestingly, the transcript levels of many *Eu4CL* genes, such as *Eu4CL13*, *Eu4CL15* and *Eu4CL26*, were upregulated by ethylene stress treatment. Several genes showed opposing expression patterns under different treatments. For instance, *Eu4CL13* was significantly induced by ethylene, but was repressed by WeJA treatment.

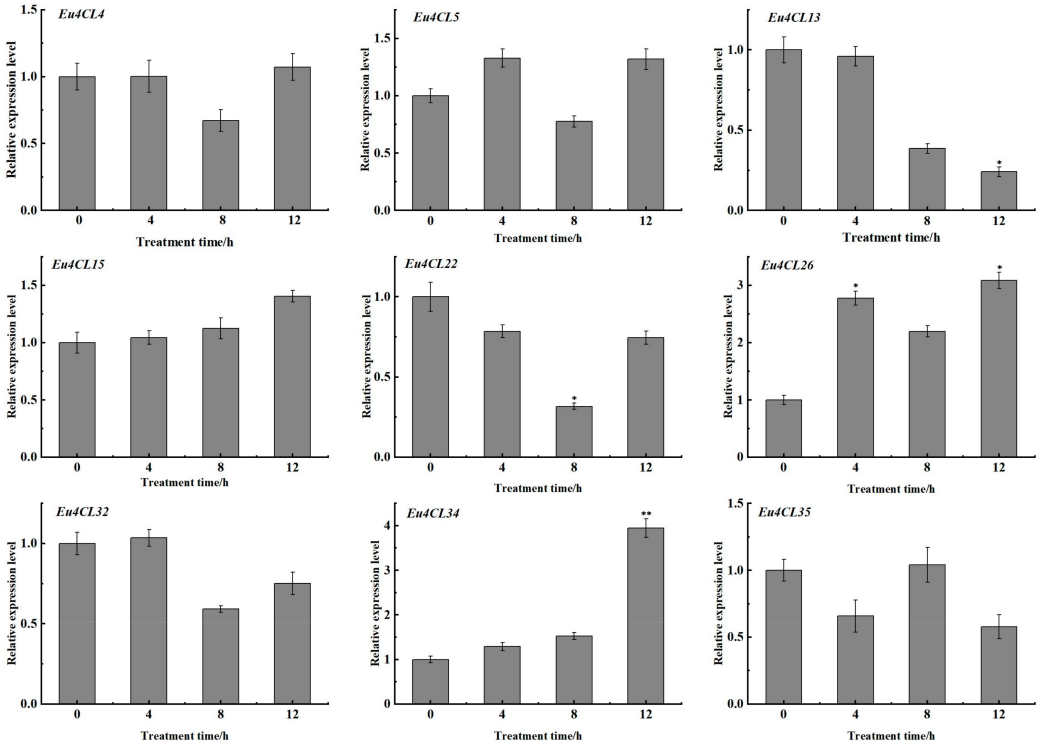

Figure 6. Expression profiles of selected Eu4CL genes under WeJA stress conditions. Error bars show the standard deviation of three biological replicates. * means significant correlation at a level of 0.05. ** means significant correlation at a level of 0.01.

To explore the correlation between *Eu4CL* genes' expression pattern and chlorogenic acid accumulation in *E. ulmoides*, the content of chlorogenic acid and the expression pattern of *Eu4CL* genes in bark and leaves of *E. ulmoides* at different stages were detected. The results showed that there were differences in the content of chlorogenic acid in bark and leaves of Eucommia ulmoides at different developmental stages (Figure 9). The content of chlorogenic acid in leaves was high, and increased rapidly at first and then decreased gradually, and finally reached the maximum in mid-June. However, the content of chlorogenic acid in bark was low and increased slowly at first and then decreased gradually, reaching its maximum in mid-August. Selected genes were expressed to varying degrees in bark and leaves of *E.ulmoides* at different developmental stages (Figure 10). Correlation analysis showed that (Figure 11) *Eu4CL4* gene expression was significantly negatively correlated with chlorogenic acid accumulation in the leaves, and the absolute value of the correlation index was greater than 0.9. The expression levels of *Eu4CL5* and *Eu4CL13* were significantly positively correlated with chlorogenic acid accumulation in the leaves, and the correlation index was greater than 0.8. *Eu4CL34* was positively correlated with the accumulation of chlorogenic acid in *E. ulmoides* bark, and the correlation index was 0.999. It was speculated that *Eu4CL4/5/13/34* were significant genes in regulating chlorogenic acid biosynthesis in *E. ulmoides*.

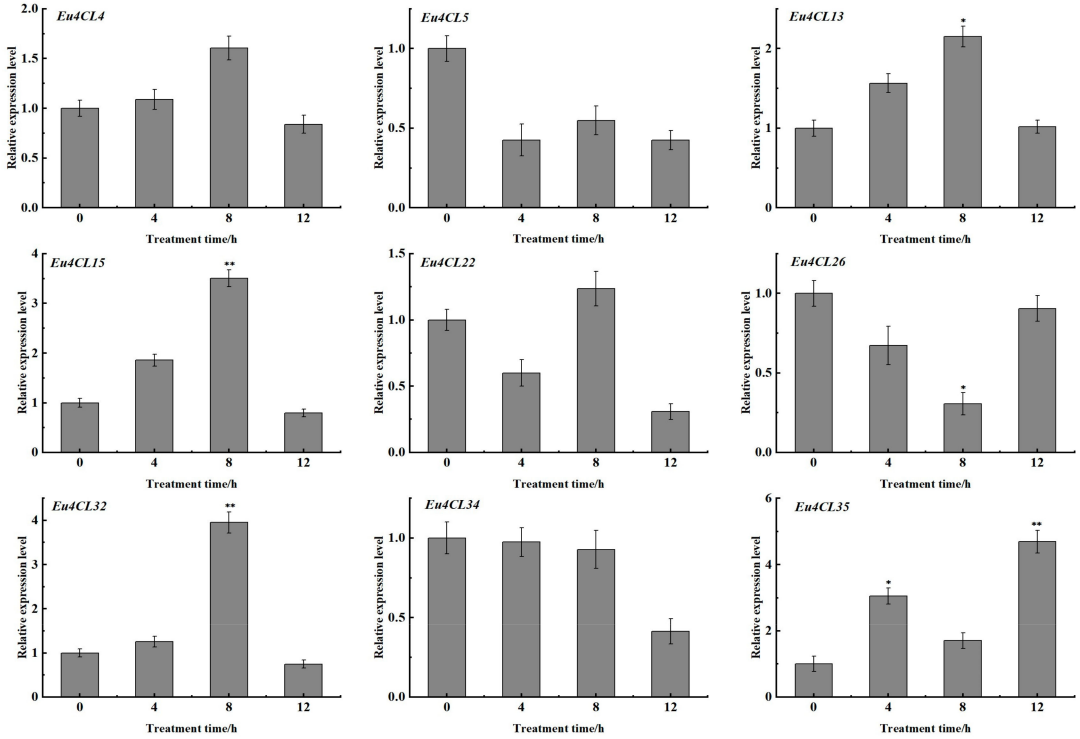

Figure 7. Expression profiles of selected Eu4CL genes under cold stress conditions. Error bars show the standard deviation of three biological replicates. * means significant correlation at a level of 0.05. ** means significant correlation at a level of 0.01.

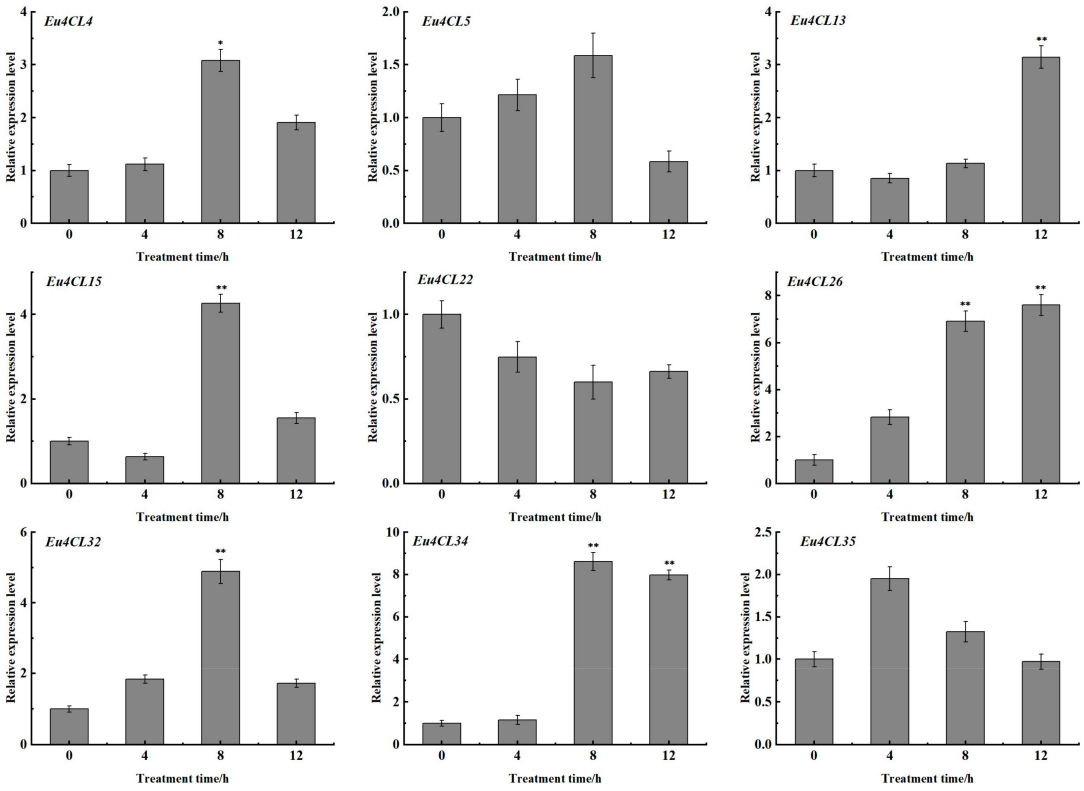

Figure 8. Expression profiles of selected *Eu4CL* genes under ethylene stress conditions. Error bars show the standard deviation of three biological replicates. * means significant correlation at a level of 0.05. ** means significant correlation at a level of 0.01.

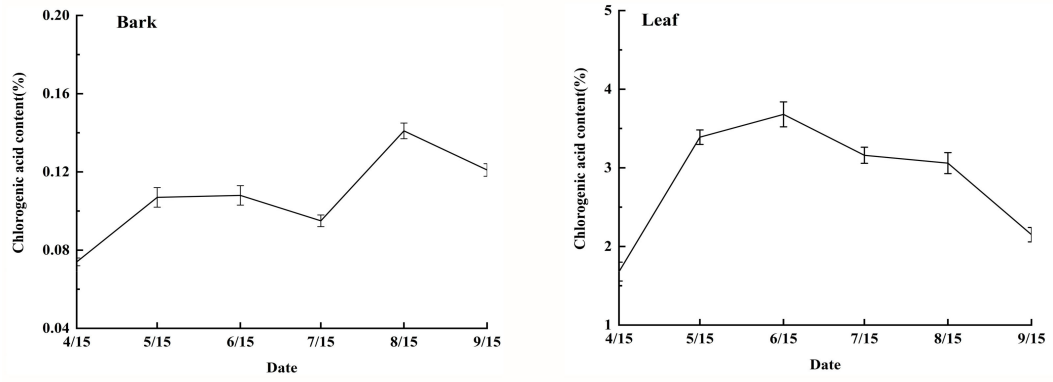

Figure 9. Changes of chlorogenic acid content in different developmental stages.

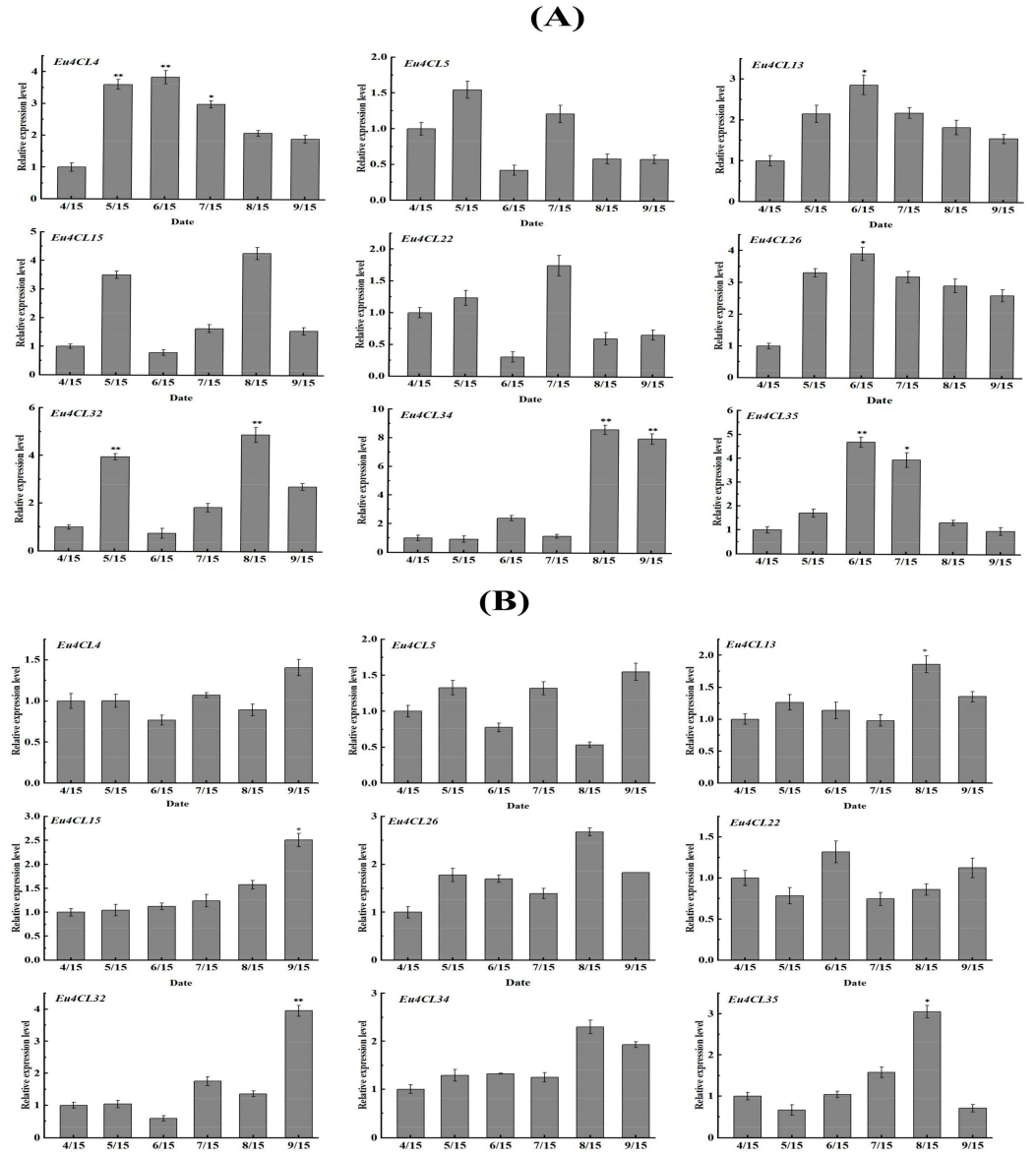

Figure 10. Expression profiles of selected *Eu4CL* genes in different developmental stages. (**A**) represents the expression of 4CL gene in leaves at different developmental stages. (**B**) represents the expression of 4CL genes in bark at different developmental stages. * means significant correlation at a level of 0.05. ** means significant correlation at a level of 0.01.

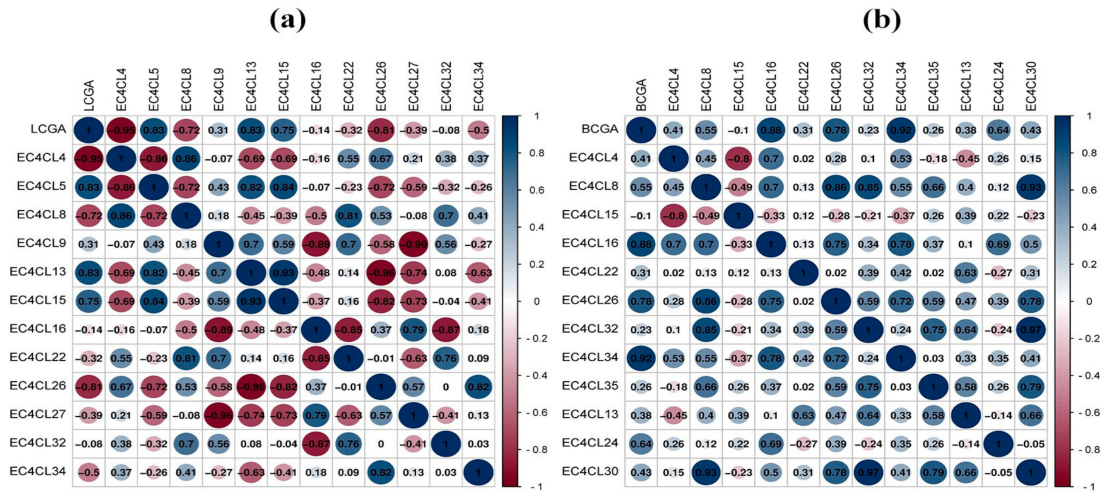

Figure 11. Correlation analysis of chlorogenic acid accumulation and *Eu4CL* expression in *Eucommia ulmoides* leaves (**a**) and bark (**b**).

4. Discussion

4-coumarate coenzyme A ligase (*4CL*) is an important rate-limiting enzyme in phenylpropanoid metabolic pathways and plays an important regulatory role in the synthesis of secondary metabolites such as chlorogenic acid, flavonoids and lignin [25,26]. At present, *4CL* has been studied in honeysuckle, tobacco, *Arabidopsis thaliana* and *Camellia oleifera* [27,28]. However, as one of the species with high chlorogenic acid content, *E. ulmoides* has not been reported regarding its relation to *4CL* genes. Therefore, this study identified 35 *4CL* genes based on the whole-genome data of *E. ulmoides*, and analyzed their gene structure, conserved motif, expression pattern and correlation with chlorogenic acid content, which provides theoretical reference for further exploring the biological function and molecular characteristics of *E. ulmoides 4CL* genes.

Members of the *Eu4CL* gene family in *E. ulmoides* were grouped into six subfamilies according to their relatedness with Arabidopsis, in line with earlier studies in other species, such as *Glycine max*, *Citrus clementina*, *Lonicera japonica* and *Nicotiana tabacum* [29,30]. The gene structure and conserved motif analysis showed that *Eu4CL* genes had structural similarity in the same group, and different subfamilies contained various differences, suggesting a variety of functions of *Eu4CL* genes [31]. Previous studies have shown that 4CL not only plays a key role in the flow of carbon sources into specific branches in phenylpropanoid metabolic pathways, but also plays an important role in plant responses to biotic and abiotic stresses. For example, *At4CL1*, *At4CL2*, *Pt4CL1* and *Os4CL2* are induced by cold, drought, WeJA and ethylene [32]. In this study, we investigated the cis-elements of *Eu4CL* genes on their promoters and the expression levels under different stress. Most of the *Eu4CL* genes exhibited various cis-elements on their promoters, such as light-responsive elements, low-temperature-responsive elements, drought-responsive elements, methyl jasmonate-responsive elements, circadian rhythm-regulating elements and salicylic acid-responsive elements, and they were induced by cold, WeJA and ethylene, which indicates their functional roles in stress resistance in *E. ulmoides* [33]. Among these *Eu4CL* genes, expressions of *Eu4CL26*, *Eu4CL32* and *Eu4CL34* were obviously induced by almost all stresses, especially the expression levels of *Eu4CL34*, which were upregulated more than 6-fold and 10-fold after the WeJA and ethylene treatment. These results indicated that *Eu4CL* genes played crucial roles in the regulation of stress resistance in *E. ulmoides* [34].

Chlorogenic acid is the main phenylpropanoid metabolite in *E. ulmoides*, and its accumulation has developmental stages and tissue specificity. [35,36]. This study found that the content of chlorogenic acid in leaves at different developmental stages was significantly different, and higher than that in bark [37]. To explore the *4CL* genes related to chlorogenic acid synthesis in leaves and bark of *E. ulmoides*, this study analyzed the *4CL* genes in

leaves and bark of *E. ulmoides* at different developmental stages based on transcriptome data and qRT-PCR [38]. The results showed that the expression of *Eu4CL* had tissue and developmental specificity. The expression levels of *Eu4CL13*, *Eu4CL26* and *Eu4CL34* were positively correlated with the content of chlorogenic acid in bark, and the expression levels of *Eu4CL4*, *Eu4CL13* and *Eu4CL26* were positively correlated with the content of chlorogenic acid in leaves. It was indicated that *Eu4CL4*, *Eu4CL13*, *Eu4CL26* and *Eu4CL34* might be involved in the synthesis of chlorogenic acid from *E. ulmoides* [39]. Previous studies have found that *Lm4CL1* and *Lm4CL2* genes are highly expressed in leaves and flowers of *Lonicera macranthoides*, respectively, and are positively correlated with chlorogenic acid content in leaves and flowers, which further indicates that *Eu4CL* expression is tissue-specific and chlorogenic acid synthesis is regulated by multiple genes [40–42]. In summary, *Eu4CL* genes not only participate in a variety of biotic and abiotic stresses during the growth and development of *E. ulmoides*, but also play an important regulatory role in chlorogenic acid synthesis [43]. This study provides theoretical reference for further exploring the molecular characteristics and biological functions of *Eu4CL* genes.

5. Conclusions

A comprehensive analysis of *4CL* gene families in *E. ulmoides* was carried out in the present study. A total of 35 *Eu4CL* genes were characterized and further classified into six main groups, with high similar exon–intron structures and motif compositions within the same groups and subgroups. *Eu4CL* genes played important roles in *E. ulmoides* growth and development as indicated by their expression patterns in different tissues and in response to various treatments. The synthesis of chlorogenic acid in *E. ulmoides* is regulated by multiple genes, and the genes regulating the synthesis of chlorogenic acid in different tissues are different. *Eu4CL13*, *Eu4CL26* and *Eu4CL34* are mainly involved in the synthesis of chlorogenic acid in bark, and *Eu4CL4*, *Eu4CL13* and *Eu4CL26* are involved in the synthesis of chlorogenic acid in leaves. These results provide theoretical reference for further exploring the molecular characteristics and biological functions of *Eu4CL* genes.

Author Contributions: Conceptualization, J.Z. and J.Q.; methodology, Q.W.; software, J.Z. and C.L.; validation, P.L., Q.D. and J.Z.; formal analysis, J.Z. and J.Q.; investigation, L.D. and H.D.; resources, L.W.; data curation, J.Q.; writing—original draft preparation, J.Z.; writing—review and editing, J.Z.; visualization, J.Q.; supervision, L.W.; project administration, H.D.; funding acquisition, H.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China (Grant No. 2017YFD0600702).

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

References

1. Du, Q.X.; Wei, Y.X.; Liu, P.F.; Du, H.Y. Diversity of the Content of Main Active Components in *Eucommia ulmoides* Male Flowers. *Sci. Silvae Sin.* **2017**, *53*, 35–43.
2. Qing, J.; Wei, Y.X.; Du, H.Y.; Du, L.Y. Study on Genetic Variation of Main Active Components in Leaves of *Eucommia ulmoides*. *Acta. Bot. Boreali-Occident. Sin.* **2018**, *38*, 316–323.
3. Dong, J.; Ma, X.; Fu, Z.; Guo, Y. Effects of microwave drying on the contents of functional constituents of *Eucommia ulmoides* flower tea. *Ind. Crops Prod.* **2011**, *34*, 1102–1110. [CrossRef]
4. Naveed, M.; Hejazi, V.; Abbas, K.; Ali, K. Chlorogenic acid (CGA): A pharmacological review and call for further research. *Biomed. Pharmacother.* **2018**, *34*, 267–283. [CrossRef]
5. Li, F.M.; Chen, L.X.; Li, J.J. Identification and Bioinformatics Analysis of Gene for Chlorogenic Acid in *Lonicera japonica* Thunb. *Mol. Plant Breed.* **2021**, *5*, e3781.
6. Kundu, A.; Vadassery, J. Chlorogenic acid-mediated chemical defence of plants against insect herbivores. *Plant Biol.* **2019**, *21*, 185–189. [CrossRef]
7. Santana-Gálvez, J.; Cisneros-Zevallos, L.; Jacobo-Velázquez, D. Chlorogenic Acid: Recent Advances on Its Dual Role as a Food Additive and a Nutraceutical against Metabolic Syndrome. *Molecules* **2017**, *22*, 358. [CrossRef]
8. Xu, P.; Xu, X.B.; Guo, H.; Zhong, C.H. Research Progress on Accumulation and Biosynthesis of Chlorogenic Acid in *Lonicerae japonicae*. *Food Serv.* **2021**, *42*, 252–256.

9. Luo, Y.E.; Tian, J.X.; Ye, J.M.; Li, J.C. Analysis and Regulation of Biosynthetic Pathways of Chlorogenic Acid in Fungi. *Sci. Technol. Eng.* **2021**, *21*, 14029–14039.
10. Endler, A.; Martens, S.; Wellmann, F.; Matern, U. Unusually divergent 4-coumarate: CoA-ligases from *Ruta graveolens* L. *Plant Mol. Biol.* **2008**, *67*, 335–346. [CrossRef]
11. Zhang, J.; Yang, Y.; Zheng, K.; Xie, M.; Feng, K.; Jawdy, S.S. Genome-wide association studies and expression-based quantitative trait loci analyses reveal roles of HCT2 in caffeoylquinic acid biosynthesis and its regulation by defense-responsive transcription factors in *Populus*. *New Phytol.* **2018**, *220*, 502–516. [CrossRef]
12. Kiewning, H. Structure and evolution of 4-coumarate: Coenzyme A ligase (*4CL*) gene families. *Biol. Chem.* **2001**, *382*, 645–654.
13. He, L.; Yan, X.; Yuan, Y.; Zhong, J.J. Genome-wide Identification and Bioinformatics Analysis of the *4CL* Gene Family in Pepper (*Capsicum annuum*). *Mol. Plant Breed.* **2022**, *16*, 1–15.
14. Ragg, H.; Kunhn, D.N.; Hahlbrock, K. Coordinated regulation of 4-coumarate-coA ligase and phenylalanine ammonia-lyase messenger-RNAs in cultured plant-cells. *J. Biol. Chem.* **1981**, *256*, 10061–10065. [CrossRef]
15. Ehlting, J.; Buttner, D.; Wang, Q.; Douglas, C.J. Three 4-coumarate: Coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant J.* **1999**, *19*, 9–20. [CrossRef]
16. Hu, W.J.; Kawaoka, A.; Tsai, C.J.; Lung, J.; Osakabe, K.; Ebinuma, H.; Chiang, V.L. Compartmentalized expression of two structurally and functionally distinct 4-coumarate: CoA ligase genes in aspen (*Populus tremuloides*). *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 5407–5412. [CrossRef]
17. Liu, C.Y. Research on the formation and accumulation of main active components and nutritional regulation of *Eucommia ulmoides*. *Chin. Acad. For. Sci.* **2011**, *12*, 43–91.
18. Yu, C.S.; Lin, C.J.; Hwang, J.K. Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci.* **2004**, *13*, 1402–1406. [CrossRef]
19. Qing, J.; Zhong, J.; Du, H.Y.; Wang, L. Genome-wide analysis of the long noncoding RNAs and mRNAs involved in flower bud development of *Eucommia ulmoides* Oliver. *Sci. Hortic. Amst.* **2022**, *300*, 111038. [CrossRef]
20. Chen, C.; Chen, H.; Zhang, Y.; Thomas, H.R.; Xia, R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol. Plant* **2020**, *13*, 1194–1202. [CrossRef]
21. Hamberger, B.; Hahlbrock, K. The 4-coumarate: CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 2209–2214. [CrossRef]
22. Lavhale, S.G.; Kalunke, R.M.; Giri, A.P. Structural, functional and evolutionary diversity of 4-coumarate-CoA ligase in plants. *Planta* **2018**, *248*, 1063–1078. [CrossRef]
23. Fang, H.; Qi, X.; Li, Y.; Yu, X.; Liu, X. De novo transcriptomic analysis of light-induced flavonoid pathway, transcription factors in the flower buds of *Lonicera japonica*. *Trees* **2020**, *34*, 267–283. [CrossRef]
24. Guo, J.Y.; Zheng, X.Y.; Zou, C.X. Research Progress of cis-elements of Abiotic Stress Inducible Promoters and Associated Transcription Factors. *Biotechnol. Bull.* **2011**, *4*, 16–20.
25. Xu, J.; Lin, Y.H.; Wang, X.N.; Han, Y.S.; Tang, L.Q.; Wang, X.H.; Zhu, H.L. Bioinformatic Identification and Expression Analysis of *4CL* Genes in *Ipomoea batatas*. *Acta Bot. Boreali-Occident. Sin.* **2020**, *40*, 581–587.
26. Harding, S.A.; Leshkevich, J.; Chiang, V.L.; Tsai, C.J. Differential substrate inhibition couples kinetically distinct 4-coumarate: Coenzyme A ligases with spatially distinct metabolic roles in quaking aspen. *Plant Physiol.* **2002**, *128*, 428–438. [CrossRef] [PubMed]
27. Li, Y.; Kim, J.I.; Pysh, L.; Chapple, C. Four isoforms of *Arabidopsis thaliana* 4-coumarate: CoA ligase (*4CL*) have overlapping yet distinct roles in phenylpropanoid metabolism. *Plant Physiol.* **2015**, *35*, 838–2015. [CrossRef] [PubMed]
28. Chen, X.H.; Peng, X.X.; Liu, Z.; Xu, X.L. Bioinformatics Analysis on Tobacco *Nt4CL* Gene Family. *Guizhou Agric. Sci.* **2018**, *46*, 11–15.
29. Park, J.J.; Yoo, C.G.; Flanagan, Y.Q.; Debnath, Y.X. Defined tetra-allelic gene disruption of the 4-coumarate: Coenzyme A ligase 1 (*Pv4CL1*) gene by CRISPR/Cas9 in switchgrass results in lignin reduction and improved sugar release. *Biotechnol. Biofuels* **2017**, *10*, 284. [CrossRef]
30. Sun, S.; Xiong, X.; Zhang, X.; Feng, H.; Li, Y. Characterization of the *Gh4CL* gene family reveals a role of *Gh4CL7* in drought tolerance. *BMC Plant Biol.* **2020**, *45*, 337–346. [CrossRef]
31. Zhang, Q.; Chen, T.; Wang, X.; Wang, J.H.; Gu, K.D. Genome-wide identification and expression analyses of homeodomain-leucine zipper family genes reveal their involvement in stress response in apple (*Malus* × *domestica*). *Hortic. Plant J.* **2022**, *8*, 261–278. [CrossRef]
32. Zhang, T.; Li, X.M.; Zhao, Q.; Shi, Y.; Hao, Y.J. Genome-wide identification and functional characterization of the MdCLE peptide family in apple (*Malus* × *domestica*). *Hortic. Plant J.* **2022**, *8*, 279–288. [CrossRef]
33. Gui, J.; Shen, J.; Li, L. Functional Characterization of Evolutionarily Divergent 4-Coumarate: Coenzyme A Ligases in Rice. *Plant Physiol.* **2011**, *157*, 574–586. [CrossRef]
34. Soltani, B.M.; Ehlting, J.; Hamberger, B.; Douglas, C.J. Multiple cis-regulatory elements regulate distinct and complex patterns of developmental and wound-induced expression of *Arabidopsis thaliana 4CL* gene family members. *Planta* **2006**, *224*, 1239–1240. [CrossRef]

35. Chen, H.; Song, J.N.; Wang, J.; Lin, Y.C.; Liu, J.; Li, Q.Z. Systems Biology of Lignin Biosynthesis in *Populus trichocarpa*: Heteromeric 4-Coumaric Acid: Coenzyme A Ligase Protein Complex Formation, Regulation, and Numerical Modeling. *Plant Cell* **2014**, *26*, 876–893. [CrossRef]
36. Egea, C.; Ahmed, A.S.; Candela, M.; Candela, M.E. Elicitation of peroxidase activity and lignin biosynthesis in pepper suspension cells by *Phytophthora capsici*. *Plant Physiol.* **2001**, *158*, 151–158. [CrossRef]
37. Burbulis, I.E.; Winkel-Shirley, B. Interactions among enzymes of the *Arabidopsis* flavonoid biosynthetic pathway. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 12929–12934. [CrossRef]
38. Guan, Y.; Hu, W.; Jiang, A.; Xu, Y.; Yang, X. Effect of Methyl Jasmonate on Phenolic Accumulation in Wounded Broccoli. *Molecules* **2019**, *24*, 3537. [CrossRef]
39. Cao, Y.; Han, Y.; Li, D.; Lin, Y.; Cai, Y. Systematic Analysis of the 4-Coumarate: Coenzyme A Ligase (*4CL*) Related Genes and Expression Profiling during Fruit Development in the Chinese Pear. *Genes* **2016**, *7*, 89. [CrossRef]
40. Jung, J.H.; Kannan, B.; Dermawan, H.; Moxley, G.W.; Altpeter, F. Precision breeding for RNAi suppression of a major 4-coumarate: Coenzyme A ligase gene improves cell wall saccharification from field grown sugarcane. *Plant Mol. Biol.* **2016**, *92*, 505–517. [CrossRef]
41. Zhi, L.; Nair, S. Structural Basis for Specificity and Flexibility in a Plant 4-Coumarate: CoA Ligase. *Structure* **2015**, *23*, 2032–2042.
42. Zhang, C.H.; Ma, T.; Luo, W.C.; Xu, J.M.; Liu, J.Q. Identification of *4CL* Genes in Desert Poplars and Their Changes in Expression in Response to Salt Stress. *Genes* **2015**, *6*, 901–917. [CrossRef] [PubMed]
43. Tsai, C.J.; Harding, S.A.; Tschaplinski, T.J.; Lindroth, R.L.; Yuan, Y. Genome-wide analysis of the structural genes regulating defense phenylpropanoid metabolism in *Populus*. *New Phytol.* **2010**, *172*, 47–62. [CrossRef] [PubMed]