



Article

Genome-Wide Comparative Analysis of the Fasciclin-like Arabinogalactan Proteins (FLAs) in *Salicacea* and Identification of Secondary Tissue Development-Related Genes

Yingying Zhang, Fangwei Zhou, Hui Wang, Yingnan Chen, Tongming Yin and Huaitong Wu *

Key Laboratory for Tree Breeding and Germplasm Improvement, Southern Modern Forestry Collaborative Innovation Center, College of Forestry, Nanjing Forestry University, Nanjing 210037, China; 15361901286@163.com (Y.Z.); zhoufangwei@njfu.edu.cn (F.Z.); zwh045687@163.com (H.W.); chenyingnan@njfu.edu.cn (Y.C.); tmyin@njfu.com.cn (T.Y.)

* Correspondence: wuht@njfu.edu.cn; Tel.: +86-150-5052-1502

Abstract: Fasciclin-like arabinogalactan proteins (FLAs) are a subclass of arabinogalactan proteins (AGPs) containing both AGP-like glycosylated domains and fasciclin (FAS) domains, which are involved in plant growth and development and synthesis of the cell wall. However, these proteins have not been identified or analyzed in willow, *Salix*, the sister genus of *Populus*. In this study, we performed a whole genome study of the FLA gene family of *Salix suchowensis* and compared it with the FLA gene family of *Populus deltoides*. The results showed the presence of 40 and 46 FLA genes in *P. deltoides* and *S. suchowensis*, distributed on 17 and 16 chromosomes, respectively. Four pairs of tandem repeat genes were found in willow, while poplar had no tandem repeat genes. Twelve and thirteen pairs of duplicated gene fragments were identified in poplar and willow, respectively. The multispecies phylogenetic tree showed that the FLA gene family could be divided into four groups (I–IV), with Group 1 showing significant expansion in woody plants. A gene expression analysis showed that *PdeFLA19/27* in Group I of poplar was highly expressed, specifically during the secondary growth period of the stem and the rapid elongation of seed hairs. In the Group I genes of *S. suchowensis*, *SsuFLA25/26/28* was also highly expressed during the secondary growth period, whereas increased expression of *SsuFLA35* was associated with seed hair tissue. These results provide important clues about the differences in the FLA gene family during the evolution of herbs and woody plants, and suggest that the FLA gene family may play an essential role in regulating the secondary growth of woody plants. It also provides a reference for further studies on the regulation of secondary growth and seed hair development by FLA genes in poplar and willow.

Keywords: *Populus deltoides*; *Salix suchowensis*; fasciclin-like arabinogalactan proteins; gene family; secondary growth; seed hair



Citation: Zhang, Y.; Zhou, F.; Wang, H.; Chen, Y.; Yin, T.; Wu, H. Genome-Wide Comparative Analysis of the Fasciclin-like Arabinogalactan Proteins (FLAs) in *Salicacea* and Identification of Secondary Tissue Development-Related Genes. *Int. J. Mol. Sci.* **2023**, *24*, 1481. <https://doi.org/10.3390/ijms24021481>

Academic Editor: Bartolome Sabater

Received: 2 December 2022

Revised: 24 December 2022

Accepted: 10 January 2023

Published: 12 January 2023



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

1. Introduction

The hydroxyproline-rich glycoproteins (HRGPs) are a class of structural proteins usually found in the cell walls of higher plants [1–4]. Arabinogalactan proteins (AGPs) are a subfamily of HRGPs involved in many aspects of plant development, including tissue differentiation, reproduction, plant hormone interactions, responses to biotic stress, cell expansion, and secondary cell wall deposition [5,6]. AGPs are subdivided according to the structure of their protein backbone into classical, lysine-rich, AG peptide, fasciclin-like (FLAs), non-classical, and chimeric AGPs [2,7,8]. The fasciclin-like arabinogalactan proteins (FLAs) are characterized by having fasciclin (FAS) domains which were first detected in fruit flies (*Drosophila melanogaster*) and later in other organisms including algae, lichens, plants, and animals. FLAs are found in numerous species. Twenty-one FLAs have been identified in *Arabidopsis thaliana* [9], 30 in *Musa spp.* [10], 33 in *Brassica rapa* L. [11], 23 in *Cannabis sativa* L. [12], 19 in *Corchorus olitorius* [13], 19 in *Gossypium hirsutum* [14], 18 in

Eucalyptus grandis [15], 38 in *Nicotiana benthamiana* [16], 35 in *Populus trichocarpa* [17], 27 in *Oryza sativa* [3], 34 in *Triticum aestivum* [18], and 17 in *Cuscuta campestris* [19].

Studies on plant FLAs have focused on tissue-specific functions, pollination, and embryogenesis, as well as general responses to biotic and abiotic stresses [6,15,17,20]. In addition, numerous studies have demonstrated the importance of FLAs in aspects related to cell wall biosynthesis. For example, in Arabidopsis, *AtFLA4* induced aberrant cell expansion, adhesion, cell wall synthesis, and seed coat pectin mucilage [21]. *AtFLA11* and *AtFLA12* played a role in stem tensile strength, biomechanics, and elastic modulus, thus affecting the composition and structure of the cell walls [20]. Studies of *AtFLA16* mutants showed that deletion of *AtFLA16* resulted in reduced stem length and altered biomechanical properties [22]. Studies on poplar showed that the *PtFLA6* gene was associated with the poplar's xylem fiber cells, cell wall composition, and stem biomechanics [23,24]. Woody plants have more developed lignified stems than herbaceous plants, and their lignified stems are mainly composed of secondary wall-thickened cells. Therefore, the prediction and study of FLA gene functions in woody plants could provide essential clues to improve the properties of wood.

Populus and *Salix* are sister genera belonging to the Salicaceae family. They are useful species for maintaining the ecological environment and for solving the shortage of wood [25,26]. Although the FLA gene family has been identified and analyzed in poplar [17], it has not been analyzed in willow. Here, we performed a genome-wide comparative analysis of the FLAs in *Populus* and *Salix*. A multi-species phylogenetic tree of the FLA gene family to explore differences in gene expansion in herbaceous and woody plants was constructed. The expression patterns of FLA genes in different tissues of both poplar and willow were combined and those that were strongly and specifically expressed during stem lignification and seed hair development were identified, thus providing reliable candidate genes for subsequent verification of the gene functions.

2. Results and Discussion

2.1. Genome-Wide Identification and Sequence Analysis of *Populus* and *Salix* FLA Genes

Based on the sequence structural characteristics of the FLA gene family, 40 and 46 FLA gene family members were identified in the genomes of *P. deltoides* and *S. suchowensis*, respectively. They were named *PdeFLA1-PdeFLA40* and *SsuFLA1-SsuFLA46*, respectively, according to their positions on the chromosomes. All the genes contained the conserved FLA structural domain (Supplementary Tables S1 and S2). In *P. deltoides*, 29 proteins were predicted by the BIG-PI online program to have C-terminal glycosylphosphatidylinositol (GPI) anchors, and 34 N-terminal signal peptides were predicted by the SignalP online program (Supplementary Table S3). The proteins contained between 176 and 466 amino acids, with isoelectric points in the range of 4.50–10.77 and molecular weights of between 18,211.68 and 51,122.1 Da. In *S. suchowensis*, 32 proteins were predicted to have C-terminal GPI anchors by the BIG-PI prediction software, and 41 to were predicted to have N-terminal signal peptides by the SignalP online program (Supplementary Table S4). The amino acid lengths ranged from 206 to 468, the isoelectric points ranged from 4.33 to 10.81, and the molecular weights were between 21,708.68 and 51,646.93 Da. The gene names, gene IDs, chromosomal location, amino acid lengths, isoelectric point (pIs), and molecular weights (Mw) are listed in Table 1. This study comprehensively analyzed the FLA gene family in *P. deltoides* and *S. suchowensis*, identifying 40 and 46 genes, respectively. These gene numbers were 1.9-fold and 2.2-fold higher than the number of FLA genes in Arabidopsis [9]. Previous research has shown that early flowering species have a faster replacement rate than species with long generations [27,28]. This suggests that these two species have more genes than Arabidopsis, either because the poplar and willow genomes were larger or the replacement rate was much lower.

Table 1. Putative FLA genes identified in the genomes of *P. deltooides* and *S. suchowensis*.

Gene Symbol	Gene ID	Chr	Start	End	Amino Acid	pI	Mw
<i>PdeFLA1</i>	EVM0033614	chr1	2863,934	2,864,908	281	7.79	28,804.55
<i>PdeFLA2</i>	EVM0013530	chr1	31,953,859	31,954,589	213	5.58	22,696.96
<i>PdeFLA3</i>	EVM0016507	chr1	37,804,827	37,807,480	406	5.49	43,144.42
<i>PdeFLA4</i>	EVM0019309	chr1	46,418,162	46,419,355	397	5.06	42,295.82
<i>PdeFLA5</i>	EVM0012430	chr2	14,552,755	14,553,795	176	6.38	18,211.68
<i>PdeFLA6</i>	EVM0003247	chr2	14,743,911	14,744,707	254	4.50	26,606.07
<i>PdeFLA7</i>	EVM0003107	chr4	20,721,639	20,727,689	323	8.59	34,386.37
<i>PdeFLA8</i>	EVM0000914	chr5	6,121,373	6,122,701	442	6.03	47,578.60
<i>PdeFLA9</i>	EVM0035476	chr6	10,905,263	10,906,724	227	4.75	24,112.12
<i>PdeFLA10</i>	EVM0035420	chr6	17,439,680	17,441,342	426	5.23	44,980.52
<i>PdeFLA11</i>	EVM0030864	chr6	20,717,497	20,720,696	466	6.07	50,934.06
<i>PdeFLA12</i>	EVM0005454	chr6	26,560,554	26,561,810	394	8.92	41,861.21
<i>PdeFLA13</i>	EVM0010041	chr8	598,740	601,878	440	5.98	48,356.80
<i>PdeFLA14</i>	EVM0021839	chr9	2,003,076	2,004,125	263	8.44	28,699.16
<i>PdeFLA15</i>	EVM0017387	chr9	2,005,847	2,007,233	269	8.40	28,404.59
<i>PdeFLA16</i>	EVM0034822	chr10	21,572,863	21,578,206	460	6.21	50,769.91
<i>PdeFLA17</i>	EVM0012659	chr11	11,492,405	11,495,560	408	5.69	43,677.05
<i>PdeFLA18</i>	EVM0036111	chr12	1,843,829	1,848,143	229	7.77	23,800.59
<i>PdeFLA19</i>	EVM0016248	chr12	14,176,837	14,177,559	240	5.33	25,352.79
<i>PdeFLA20</i>	EVM0035889	chr13	1,002,373	1,003,172	255	7.02	27,077.81
<i>PdeFLA21</i>	EVM0037601	chr13	13,975,549	13,976,510	241	6.55	25,114.65
<i>PdeFLA22</i>	EVM0020019	chr13	15,988,602	16,005,827	269	6.64	28,265.15
<i>PdeFLA23</i>	EVM0022991	chr13	16,027,081	16,028,142	353	9.32	39,230.62
<i>PdeFLA24</i>	EVM0034823	chr14	5,118,954	5,120,795	423	5.36	43,390.45
<i>PdeFLA25</i>	EVM0004019	chr14	12,708,764	12,710,721	262	5.56	27,645.30
<i>PdeFLA26</i>	EVM0001785	chr14	13,240,881	13,242,143	376	7.65	40,562.30
<i>PdeFLA27</i>	EVM0017855	chr15	13,612,965	13,614,288	240	6.72	25,379.90
<i>PdeFLA28</i>	EVM0008767	chr16	4,738,845	4,742,370	466	5.90	51,122.10
<i>PdeFLA29</i>	EVM0000252	chr16	7,252,981	7,254,142	239	5.73	25,172.56
<i>PdeFLA30</i>	EVM0030626	chr17	12,475,605	12,476,664	339	4.50	37,001.51
<i>PdeFLA31</i>	EVM0029903	chr18	11,808,901	11,810,496	427	5.29	44,971.60
<i>PdeFLA32</i>	EVM0013339	chr19	1,874,378	1,875,226	282	10.77	30,578.54
<i>PdeFLA33</i>	EVM0014374	chr19	9,899,008	9,899,766	252	7.76	27,151.55
<i>PdeFLA34</i>	EVM0017850	chr19	15,221,093	15,221,992	245	6.72	25,421.84
<i>PdeFLA35</i>	EVM0023928	chr19	17,054,292	17,069,477	261	6.50	27,443.27
<i>PdeFLA36</i>	EVM0017906	chr19	17,076,489	17,081,736	212	9.68	22,669.99
<i>PdeFLA37</i>	EVM0003708	Contig00307	137	987	211	6.18	22,266.27
<i>PdeFLA38</i>	EVM0009936	Contig00307	14,016	15,179	263	6.50	27,647.50
<i>PdeFLA39</i>	EVM0004735	Contig00342	28,127	29,763	397	5.51	42,607.59
<i>PdeFLA40</i>	EVM0030058	Contig02018	2408	3139	243	9.04	25,611.18
<i>SsuFLA1</i>	EVM0026394	chr1	6,636,804	6,639,178	454	6.29	49,805.69
<i>SsuFLA2</i>	EVM0024166	chr1	9,407,953	9,408,990	240	5.29	25,207.58
<i>SsuFLA3</i>	EVM0006591	chr1	25,816,885	25,817,328	340	5.19	35,939.01
<i>SsuFLA4</i>	EVM0006450	chr1	32,542,611	32,543,813	400	5.47	42,028.31
<i>SsuFLA5</i>	EVM0021770	chr4	15,902,015	15,902,939	242	8.89	26,127.29
<i>SsuFLA6</i>	EVM0006635	chr4	15,905,288	15,906,103	271	8.55	28,341.48
<i>SsuFLA7</i>	EVM0003490	chr6	8,865,503	8,866,458	241	6.55	25,146.69
<i>SsuFLA8</i>	EVM0000130	chr6	10,047,978	10,048,840	257	7.85	27,187.19
<i>SsuFLA9</i>	EVM0014792	chr6	12,547,684	12,549,068	388	5.16	40,792.62
<i>SsuFLA10</i>	EVM0022300	chr6	14,817,544	14,820,039	468	7.77	51,646.93
<i>SsuFLA11</i>	EVM0039496	chr6	14,859,433	14,862,830	444	6.06	48,952.77
<i>SsuFLA12</i>	EVM0030612	chr6	14,870,417	14,872,828	465	6.25	51,410.46
<i>SsuFLA13</i>	EVM0002771	chr6	20,137,156	20,138,004	282	8.98	29,097.38
<i>SsuFLA14</i>	EVM0038485	chr7	9,014,791	9,016,143	450	5.95	48,762.99
<i>SsuFLA15</i>	EVM0004828	chr7	9,076,360	9,077,712	450	5.95	48,775.05
<i>SsuFLA16</i>	EVM0004641	chr8	467,485	470,609	428	5.78	47,346.78
<i>SsuFLA17</i>	EVM0030085	chr8	7,324,477	7,325,427	316	4.68	34,441.38

Table 1. Cont.

Gene Symbol	Gene ID	Chr	Start	End	Amino Acid	pI	Mw
<i>SsuFLA18</i>	EVM0025887	chr9	1,053,546	1,054,711	269	6.41	28,307.37
<i>SsuFLA19</i>	EVM0026596	chr10	16,349,980	16,352,801	461	5.61	50,880.71
<i>SsuFLA20</i>	EVM0032967	chr11	9,355,631	9,358,661	344	6.14	37,231.76
<i>SsuFLA21</i>	EVM0003643	chr12	927,908	931,248	275	6.41	28,692.16
<i>SsuFLA22</i>	EVM0039047	chr12	9,693,341	9,693,356	206	5.78	21,712.59
<i>SsuFLA23</i>	EVM0028769	chr13	792,822	793,782	257	7.74	27,166.11
<i>SsuFLA24</i>	EVM0029643	chr13	12,512,025	12,512,743	211	4.73	21,708.68
<i>SsuFLA25</i>	EVM0007673	chr13	14,193,258	14,194,071	245	6.51	25,665.23
<i>SsuFLA26</i>	EVM0039302	chr13	14,197,604	14,198,398	264	7.87	27,766.61
<i>SsuFLA27</i>	EVM0003847	chr13	14,200,804	14,202,010	245	8.67	25,785.40
<i>SsuFLA28</i>	EVM0021248	chr13	14,285,373	14,286,276	265	7.88	27,869.72
<i>SsuFLA29</i>	EVM0029154	chr13	14,289,871	14,290,923	259	7.06	27,479.26
<i>SsuFLA30</i>	EVM0019512	chr13	14,308,009	14,309,070	353	8.91	39,166.45
<i>SsuFLA31</i>	EVM0039706	chr14	4,440,309	4,441,573	411	5.37	42,242.28
<i>SsuFLA32</i>	EVM0024403	chr14	10,667,361	10,669,307	260	4.73	27,416.04
<i>SsuFLA33</i>	EVM0016971	chr14	11,035,380	11,037,015	391	7.09	41,605.74
<i>SsuFLA34</i>	EVM0008734	chr15	781,936	782,981	267	6.50	28,227.71
<i>SsuFLA35</i>	EVM0034733	chr15	11,756,924	11,758,526	240	6.57	25,349.91
<i>SsuFLA36</i>	EVM0008341	chr16	2,217,768	2,218,745	325	6.57	33,166.50
<i>SsuFLA37</i>	EVM0021655	chr17	10,904,748	10,905,794	325	4.33	35,075.06
<i>SsuFLA38</i>	EVM0000963	chr18	324,610	325,410	266	5.45	27,430.66
<i>SsuFLA39</i>	EVM0000740	chr18	9,619,792	9,620,553	397	4.96	42,023.18
<i>SsuFLA40</i>	EVM0022185	chr19	3,251,016	3,251,879	287	10.81	31,346.24
<i>SsuFLA41</i>	EVM0019293	chr19	9,639,274	9,639,996	240	8.41	25,832.92
<i>SsuFLA42</i>	EVM0007092	chr19	14,655,296	14,656,329	243	6.27	25,492.98
<i>SsuFLA43</i>	EVM0013439	chr19	16,018,464	16,019,599	263	6.97	27,756.63
<i>SsuFLA44</i>	EVM0030289	chr19	16,037,704	16,038,793	263	6.58	27,414.32
<i>SsuFLA45</i>	EVM0013597	chr19	16,041,720	16,042,511	263	6.97	27,756.63
<i>SsuFLA46</i>	EVM0006563	Contig04279	21,081	21,802	223	7.06	24,277.03

2.2. Chromosomal Distribution and Identification of Gene Duplication in *Populus* and *Salix* FLA Genes

To determine the distribution of FLA gene family members on the chromosomes of the *P. deltoides* and *S. suchowensis* genomes, and to evaluate the gene duplication relationships among the genes, we mapped the 40 poplar and 46 willow FLA genes onto the 19 chromosomes of their respective genomes using nucleotide sequence alignment. Four poplar genes and two willow genes failed to map to chromosomes (Figure 1), but were mapped to unassembled scaffold fragments, probably due to incomplete genome assembly in these plant tissues.

We found that the distribution of FLA genes on the chromosomes in the two studied genomes was uneven. The maximum number of FLA gene members distributed on the poplar chromosomes I, VI, XIII, and XIX was four, and no FLA genes were distributed on chromosomes III and VII (Figure 1A). There were relatively more FLA genes distributed on willow chromosomes VI, XIII, and XIX, with 7, 8, and 6 members (Figure 1B), respectively. No FLA genes were distributed on chromosomes II, III, and V. It was also found that the same number of FLA genes were distributed in the same positions on chromosomes I, X, XI, XII, XIV, and XVII of the two genomes, and it was speculated that they were orthologous genes. We analyzed the orthologous relationship between *P. deltoides* and *S. suchowensis*. Thirty-six pairs of orthologous genes were identified (Supplementary Table S5). The synteny analysis showed that 28 pairs of FLA genes were distributed in the same chromosomal position of *P. deltoides* and *S. suchowensis*, five pairs of genes were distributed in different positions, and three pairs of genes could not be determined linearly at the time because they were not yet mounted to the chromosome (Figure 2A). Our results demonstrated that the distribution of FLA genes is similar overall on both genomes, although there were large distribution differences on certain chromosomes. For example, compared with the poplar

genome, there were no genes on chromosome II in the willow genome (Figure 1B), which might be because the evolution of willow species was faster than that of poplar species, and some genes that were not adapted to the environment were gradually eliminated.

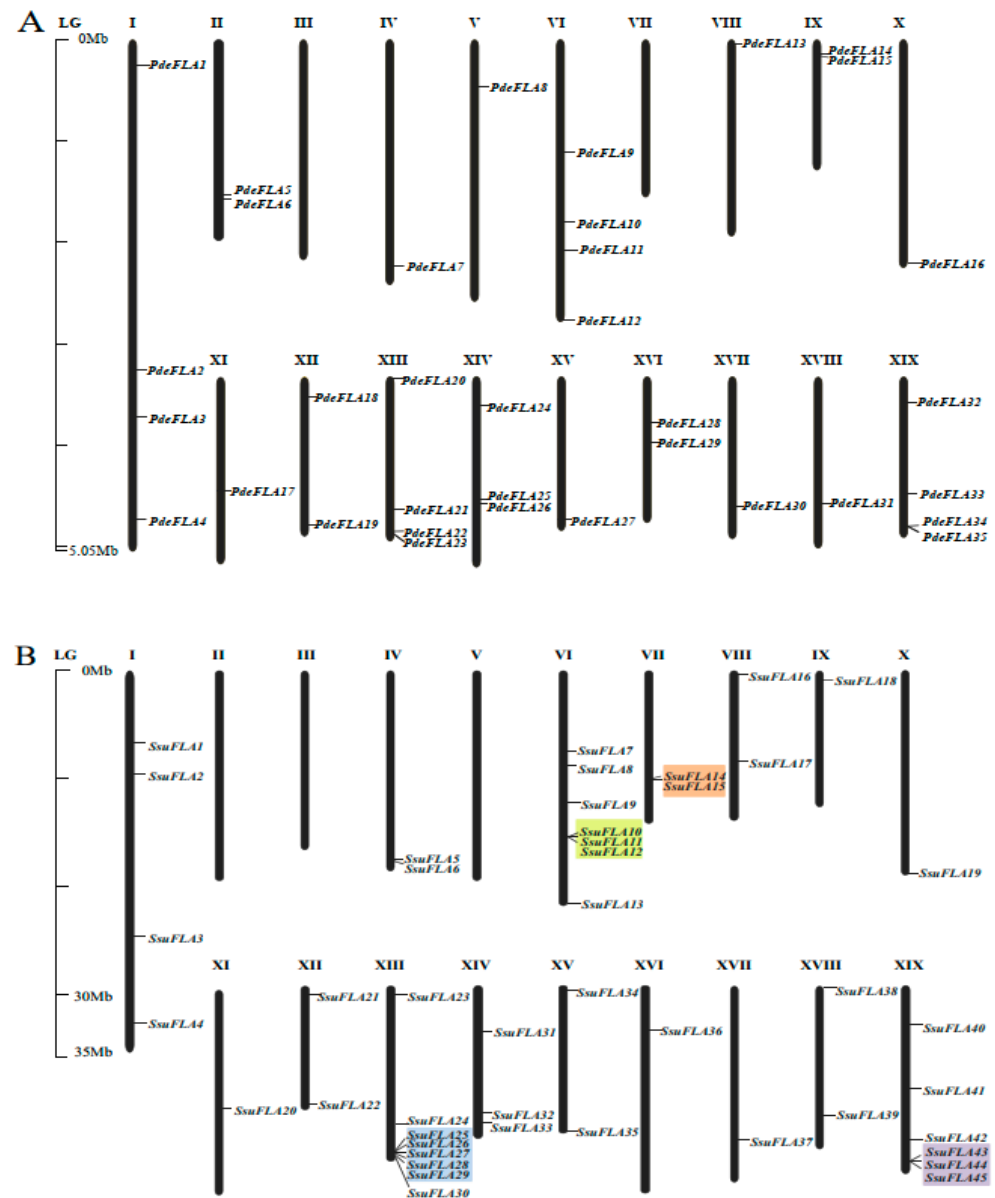


Figure 1. Location of FLA genes on chromosomes. Distribution of FLA gene family members on chromosomes in *P. deltoides* (A) and *S. suchowensis* (B). Genes are named according to their order on the chromosomes. Identified tandem repeat genes are highlighted using a colored background.

Segmental and tandem duplication are the main drivers of gene family expansion. We found four pairs of tandem duplicated genes (*SsuFLA10*, *SsuFLA11* and *SsuFLA12*; *SsuFLA14* and *SsuFLA15*; *SsuFLA25*, *SsuFLA26*, *SsuFLA27*, *SsuFLA28* and *SsuFLA29*; and *SsuFLA43*, *SsuFLA44*, and *SsuFLA45*) in *S. suchowensis*, distributed on chromosomes VI, VII, XIII, and XIX (Figure 1B). No tandem duplicated genes were found in *P. deltoides*. Twelve segmental duplication events were identified in *P. deltoides*, involving 23 genes (Figure 2B and Supplemental Table S6). Thirteen segmental duplication events were identified in *S. suchowensis*, involving 25 genes (Figure 2C and Supplemental Table S6). All gene pairs, apart from *SsuFLA43/45* in *S. suchowensis*, which showed complete sequence identity, had Ka/Ks values less than 1 (Supplemental Table S6), indicating that the FLA gene family underwent purifying selection. A comparison of the gene duplication results between

poplar and willow showed that the number of segmental duplication events was similar, but the location of segmental duplication differed greatly, with only six segmental duplications having the same chromosomal location. In addition, tandem duplication of FLA genes was more frequent in willow than poplar. The above results suggest that FLA genes may have undergone six segmental duplications within the common ancestral species of poplar and willow. After divergence of the two species, the average rate of gene replacement was significantly higher in the willow genome than in the poplar genome [27,28], and the rate of evolution was also higher [26,29]. This rapid expansion allowed the preservation of these newly generated species. This may also be why willow trees had more FLA genes than poplars.

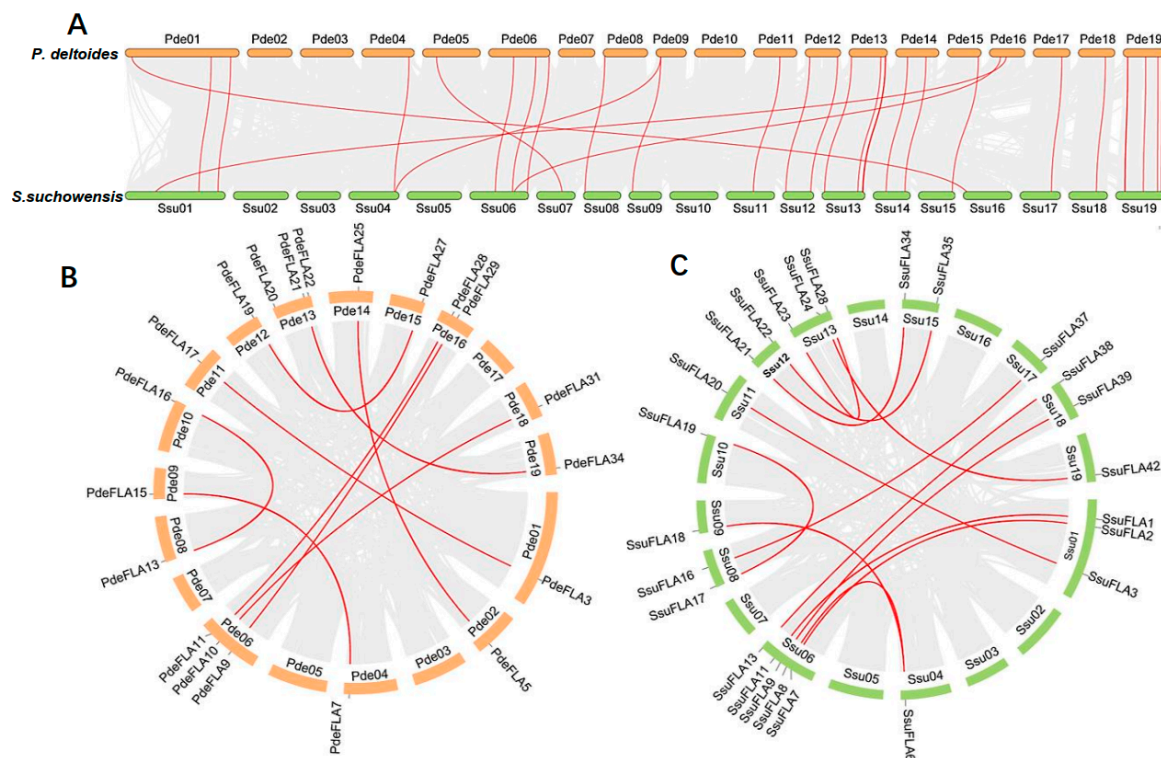


Figure 2. Synteny analysis of FLA genes. Segmental duplication of FLA gene pairs in *P. deltoides* (A) and *S. suchowensis* (B), with red lines indicating the correspondence of duplicated genes. (C) Gene duplication and synteny relationship of FLA genes between *P. deltoides* and *S. suchowensis*. The red lines indicate the syntenic FLA gene pairs, and the light gray indicates collinear blocks.

2.3. Phylogeny, Conserved Gene Structures, and Protein Motif Analysis of *Populus* and *Salix* FLA Genes

Both poplar and willow evolved from the same ancestral species, and during their respective evolution after species divergence, both retained orthologous genes inherited from their ancestors. At the same time, replication occurred in both genomes, generating new paralogous genes. To compare the sequences and structures of these homologous genes, we constructed a phylogenetic tree based on the full-length sequences of the FLA genes. We combined gene structure and conserved sequence prediction to present a comprehensive picture of the homology of FLA genes between *P. deltoides* and *S. suchowensis*. The analysis showed that there were 22 pairs of *P. deltoides* and *S. suchowensis* FLA genes at the extreme ends of the evolutionary tree branches, including *PdeFLA2* and *SsuFLA46*, *PdeFLA35* and *SsuFLA27*, and *PdeFLA18* and *SsuFLA15* (Figure 3A). In addition, multiple willow FLA genes were found to cluster with one poplar FLA gene on one evolutionary branch (Figure 3A). For example, *SsuFLA43*, *SsuFLA44*, and *SsuFLA45* clustered with *PdeFLA36* on the same branch, and this, combined with their chromosomal distributions,

confirmed that *SsuFLA43*, *SsuFLA44*, and *SsuFLA45* were duplicates of genes generated by tandem repeats during willow evolution.

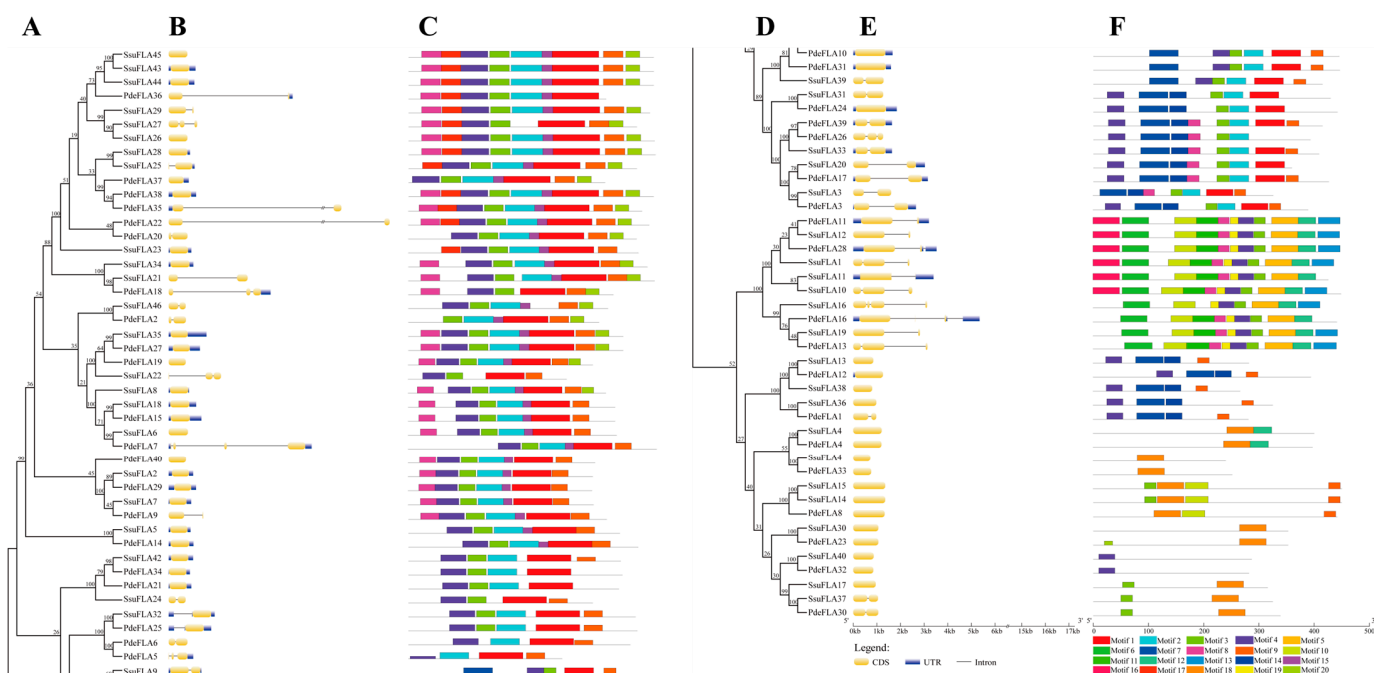


Figure 3. Phylogenetic relationships, gene structures, and motif compositions of FLA genes in *P. deltoides* and *S. suchowensis*. (A) Phylogenetics of FLA genes. (B) Genetic structure of exons and introns in each FLA gene. The yellow boxes represent exons and the single black lines represent introns. (C) The conserved motifs contained in each FLA protein. (D) Continuation of (A). (E) Continuation of (B). (F) Continuation of (C). Each motif is indicated using colored boxes and single black lines indicate non-conserved sequences. The bottom scale allows estimation of the size of each motif.

Normally, the FLA binding domain shows that introns were lost during plant evolution. Nevertheless, in our analysis, we found that 34.88% (30/86) of *P. deltoides* and *S. suchowensis* FLA genes contained one intron, 11.63% (10/86) had two introns, and 2.33% (2/86) had more than two introns, with only 51.16% (44/86) having no introns (Figure 3B). Overall, the intron structure varied widely among the FLA gene family members. However, combined with the phylogenetic tree analysis, our results indicated that genes on the same evolutionary branches of the *P. deltoides* and *S. suchowensis* FLA gene families had similar intron patterns. Furthermore, while some of the closest genes showed similar structures, a small subset showed different intron-exon arrangements. For example, *SsuFLA6* had no introns, while its nearby homologous gene, *PdeFLA7*, contained three, even though their evolutionary relationships reached 99% bootstrap values, respectively.

We further predicted the conserved motifs in the *P. deltoides* and *S. suchowensis* FLA proteins using the MEME online program to determine the specific regions of the FLA proteins. A comparison of the 20 conserved motifs found in *P. deltoides* and *S. suchowensis* (Figure 3C) revealed that the major motifs of the two species were similar, and the composition of the motifs was not significantly different. The lengths of the 20 conserved motifs ranged from 11 to 50 amino acids (Supplementary Table S7), and the number of motifs in each FLA protein ranged from 1 to 11. Most FLA proteins had motifs 1, 2, 3, 4, 8, and 9 (Figure 3C). Motifs 1, 2, 3, and 4 were included within the fasciclin domain, and motifs 8 and 9 were observed in the AGP-like glycosylation regions. However, while motifs 1, 2, 3, and 4 followed the fasciclin domain in the sequence and were particularly conserved in most clades, some motifs were selectively distributed between specific clades of the phylogenetic tree. For example, *PdeFLA32* and *SsuFLA40* contained only motif 4, and

SsuFLA41, *SsuFLA30*, and *PdeFLA33* contained only motif 18. These results suggested that these conserved motifs played key roles in specific or similar functions.

2.4. Phylogenetic Analysis and Functional Prediction of the FLA Gene Family

We have investigated the phylogenetic relationships of the FLA proteins in different plants and constructed a phylogenetic tree including five woody plants (*P. deltooides*, *S. suchowensis*, *Betula platyphylla*, *Quercus rubra* L., and *Cinnamomum kanehirae*) and five herbaceous plants (*Setaria italica*, *Zea mays* L., *Cucumis sativus* L., *Oryza sativa* L., and *Arabidopsis thaliana*). In all, 277 FLA gene members were classified using the neighbor-joining (NJ) method. The phylogenetic tree showed that the FLA gene family was divided into four evolutionary branches, which were named Group I, II, III, and IV, respectively (Figure 4).

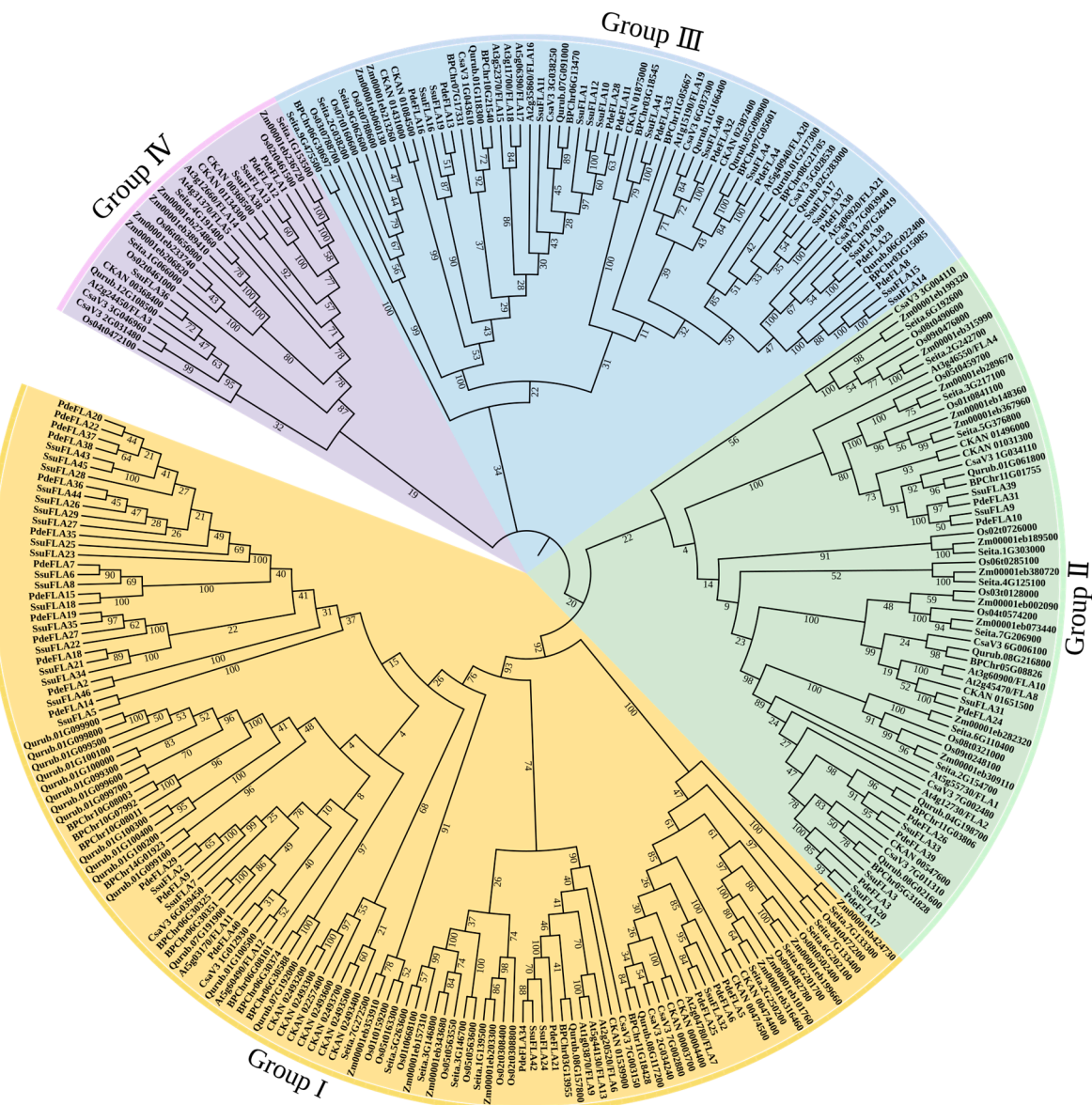


Figure 4. The phylogenetic tree of FLA proteins from *P. deltooides*, *S. suchowensis*, *Corylus avellana* L., *Quercus rubra* L., *Cinnamomum kanehirae*, *Setaria italica*, *Zea mays* L., *Cucumis sativus* L., *Oryza sativa* L., and *Arabidopsis thaliana*. A total of 277 FLA proteins were analyzed using ClustalX 2.1, and the Neighbor-Joining tree was constructed using MEGA 11 software. The bootstrap value was 1000 replicates. The FLA proteins were clustered into four distinct groups (I–IV), represented by different colors.

The statistics of the gene members in each group showed that Group I was the largest group, containing 123 genes, while Group IV was the smallest, with 26 genes. The number of genes in each group was counted for each species used in the construction of the evolutionary tree, and it was found that the woody plants had more Group I genes than herbaceous plants. Group I genes made up an average of 50.49% of the FLA genes in woody plants, while the average proportion in herbaceous plants was 33.16% (Supplementary Table S8). Additionally, Group I FLA made up an average of 0.04584% of all genes in the genomes of woody plants, but only 0.02118% in herbaceous plants (Supplementary Table S8). The differences in these proportions between woody and herbaceous plants were found to be highly significant (Supplementary Figure S1). This suggests that among the plants selected for evolutionary analysis, FLA genes in the Group I branch of woody plants underwent a major expansion. It also implies that FLA genes play important roles in secondary growth in woody plants.

Combined with the functional conservation of homologous genes, genes with similar or identical functions were classified into the same groups, providing a reliable reference basis for predicting the functions of related genes in the gene family. To predict the biological functions of the poplar and willow FLA genes, we referred to the functions of FLA genes that have been verified in *Arabidopsis*. The gene members in Group I belonged to the same subgroup as the *AtFLA11* and *AtFLA12* genes that are associated with xylem development. Related studies have found a correlation between the abundance of *AtFLA11* and *AtFLA12* transcripts containing a single FAS domain and the onset of secondary cell wall cellulose synthase expression in *Arabidopsis* stems [30,31]. In addition, the phenotypes of *AtFLA11* mutants showed the presence of a mild collapsed vessel phenotype and reduced stem cellulose content [30–32]. These analyses indicated that the FLA members in Group I were associated with secondary wall and cellulose synthesis in the stem. Group II proteins fall into the same subgroup as *AtFLA1*. Several studies have described a T-DNA insertion mutant in the *AtFLA1* gene under standard growth conditions, showing that although the FLA1-1 mutants had no distinct phenotype, the gene may play a role in the ability of the callus to induce regeneration, shown by in vitro shoot induction assays [33]. Therefore, we speculate that the Group II proteins might be involved in the development of buds. *AtFLA16* belongs to Group III, and studies on *AtFLA16* mutants showed that the deletion of the gene resulted in reduced stem lengths and altered biomechanical properties [22]. We thus speculate that Group III members are responsible for maintaining and regulating stem strength and support. Group IV includes *AtFLA3*, and overexpression of the gene led to defective elongation of stamen filaments and reduced female fertility resulting in siliques with low seed settings, suggesting that *AtFLA3* was involved in microspore development and might affect the pollen lining by participating in cellulose deposition [34]. Therefore, we speculate that the members of Group IV might have functions involved in the development of microspores.

A combination of the phylogenetic analysis and functional predictions of FLA genes strongly suggested that genes in Group I were likely to be associated with cell wall synthesis and secondary growth in plants. In particular, two sets of tandem repeat genes from willow (*SsuFLA10*, *SsuFLA11*, and *SsuFLA12* and *SsuFLA43*, *SsuFLA44*, and *SsuFLA45*) were included in Group I (Figure 4). In addition, it has been shown that a subset of FLAs containing a single FAS domain contributed to plant stem strength by affecting cellulose deposition. It has also been suggested that this influences the stem elastic modulus by affecting the integrity of the cell wall matrix [20]. We further found that FLA genes in Group I contained only one FAS domain and had undergone duplication. Thus, these analyses supported our hypothesis that Group I genes were involved in plant cellulose and cell wall formation, were significantly amplified in woody plants, and were involved in wood formation. These results suggested that these genes were good candidates for further functional verification and phylogenetic analysis.

2.5. Identification of FLA Candidate Genes Related to Stem Development in *Populus* and *Salix*

Wood production relies on the activity of vascular bundle-forming layers. A primary vascular system consists of a discrete set of vascular bundles that contain bundle-forming layers, primary phloem, and primary xylem. The primary vascular bundle originates from proto-formation layer cells on the periphery of the rib region of the shoot apical meristem (SAM). In perennial woody plants, secondary growth arises from the meristem activity of the vascular formation layer, and the bundle-like formation layer located at the center of the primary vascular bundle extends to the interbundled region, tangentially generating an interbundled formation layer to form a vascular formation layer ring [35,36]. The meristem activity of the vascular forming layer then leads to the continuous production of cylindrical secondary vascular tissue (wood). Therefore, we selected tissues with different degrees of lignification in *P. deltoides* and *S. suchowensis* stems and analyzed the expression patterns of Group I members in the FLA gene family. Since some genes had high sequence similarity, it was impossible to design specific primers to distinguish them, so universal primers were designed for these genes to analyze their expression patterns. For example, *PdeFLA19* and *PdeFLA27* were 91.42% similar; *PdeFLA35* and *PdeFLA38* were 96.09% similar; *SsuFLA43*, *SsuFLA44* and *SsuFLA45* were 97.99% alike; and *SsuFLA25*, *SsuFLA26* and *SsuFLA28* were 93.38% identical.

In *P. deltoides*, five parts of the stem, internodes 1-2, 3-4, 5-7, 8-9, and 10-11, were selected from the stem tip downwards and named P_S1, P_S2, P_S3, P_S4, and P_S5, respectively (Figure 5A). Expression patterns were analyzed for members of some FLA gene families in Group I. RT-qPCR analysis showed (Figure 5C) that *PdeFLA2/15/19/20/22/27/35/36/38* were expressed significantly in the elongated region of the stem (P_S3) but were significantly reduced in the tip of the stem (P_S1) and the secondary growth region of the stem (P_S4). On the other hand, *PdeFLA7/14/37* was expressed mainly in young stems (P_S2) and elongated regions of stems (P_S3). In particular, *PdeFLA19/27* had the highest relative expression. These results suggested that *PdeFLA19/27* was associated with secondary thickening growth and could thus serve as a major candidate gene for later studies on poplar wood growth and development.

We selected tissues from *S. suchowensis*, as we did in poplar. They were named S_S1 to S_S5 and were analyzed by RT-qPCR to determine the expression patterns of the Group I FLA family members. The results of the stem cross-sections showed that the willow had a faster lignification rate than the poplar in the same nodal tissues, especially in S4 and S5, where the secondary vascular tissue formed a thicker structure (Figure 5B). The RT-qPCR analysis showed that a total of 17 genes (Figure 5D), all members of the Group I branch in willow, were positively correlated with stem node lignification and had a very low expression in shoot leaves and stem tips (S_S1). The expression of *SsuFLA5/6/18/23/27* increased gradually from S_S1 to S_S5, while the expression levels of *SsuFLA21/22/25/26/28/29/34/35/43/44/45/46* were highest in S_S4 tissue, but tended to decrease in S_S5 tissue (Figure 5D). Therefore, it was inferred that rapid lignification of willow stems occurred at internodes 8-9 (S_S4). The Group I FLA genes showed greater functional redundancy in the regulation of stem lignification function. Of these, the three tandem repeats of *SsuFLA25/26/28* were expressed at high levels and played a dominant role in willow stem lignification. These results suggested that these could be important candidate genes for further investigation of wood development and formation in willow.

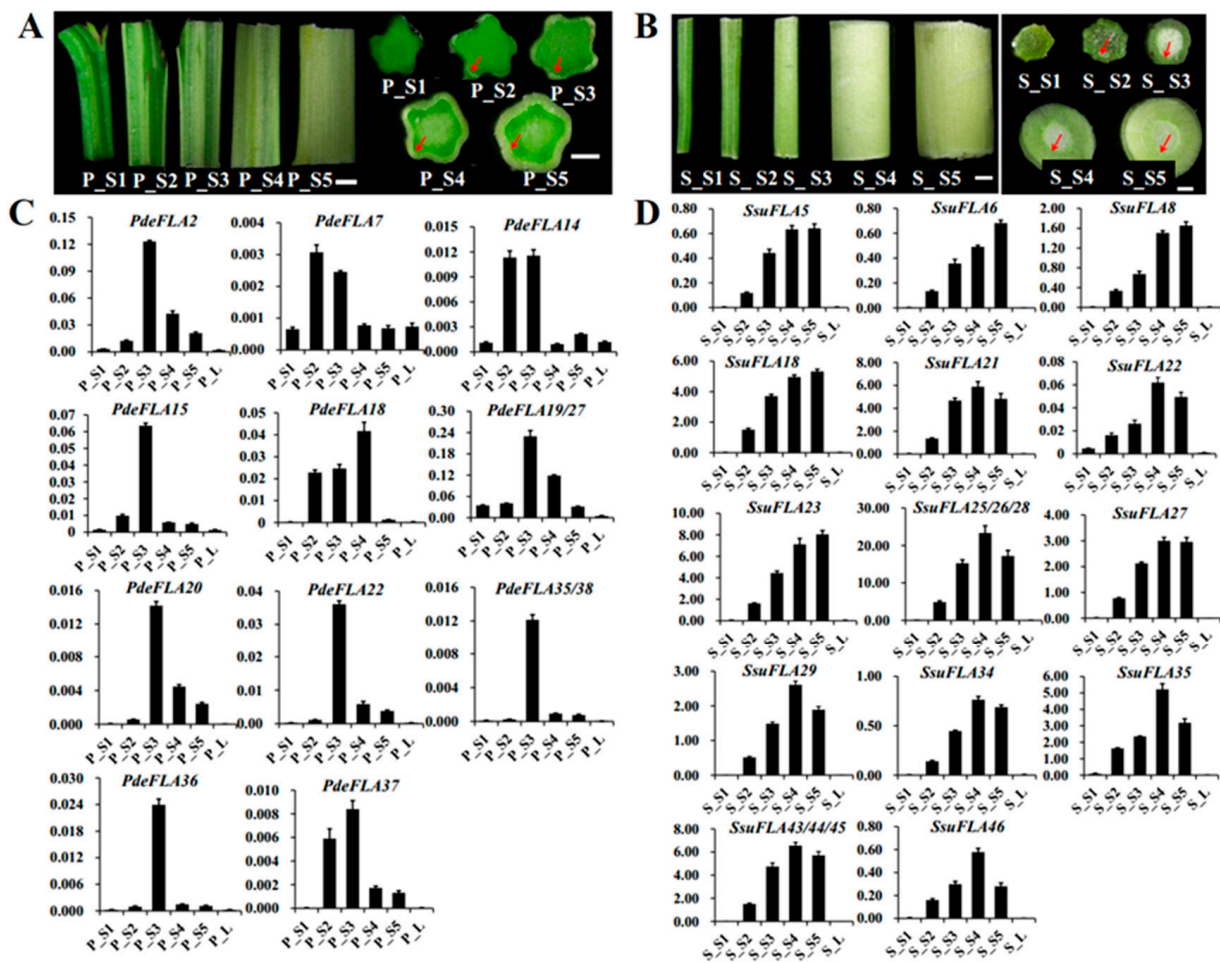


Figure 5. Expression analysis of Group I members of the FLA gene family in *P. deltoides* and *S. suchowensis* at different stages of stem lignification development. (A,B) Observations of *P. deltoides* and *S. suchowensis* at stem lignification development sites. The red arrow points to the xylem. Scale bar = 0.5 mm. (C,D) Expression pattern analysis of Group I members in *P. deltoides* and *S. suchowensis*. The X-axis represents different tissues and the Y-axis represents relative expression. Values shown are means \pm SD of three biological replicates.

2.6. Identification of FLA Candidate Genes Related to Seed Hair Development in *Populus* and *Salix*

Seed hairs are composed of flocculent fibers produced by poplar and willow trees to assist in seed dispersal after reaching sexual maturity. They are formed by the projection of placentation epidermal cells and their cell walls are largely composed of cellulose, hemicellulose, and pectin. As in the investigation of the regulatory roles of FLA genes in plant cell walls, we collected tissues from different periods of poplar seed hair development to analyze the expression of the Group I genes. The results showed that poplar seed hairs produced obvious flocculent fibers that elongated rapidly 3 days after pollination (DAP). Furthermore, the length of seed hairs was sufficient to completely wrap the ovules by 5 DAP, indicating that the 3–5 DAP period was associated with the rapid elongation of seed hairs (Figure 6A).

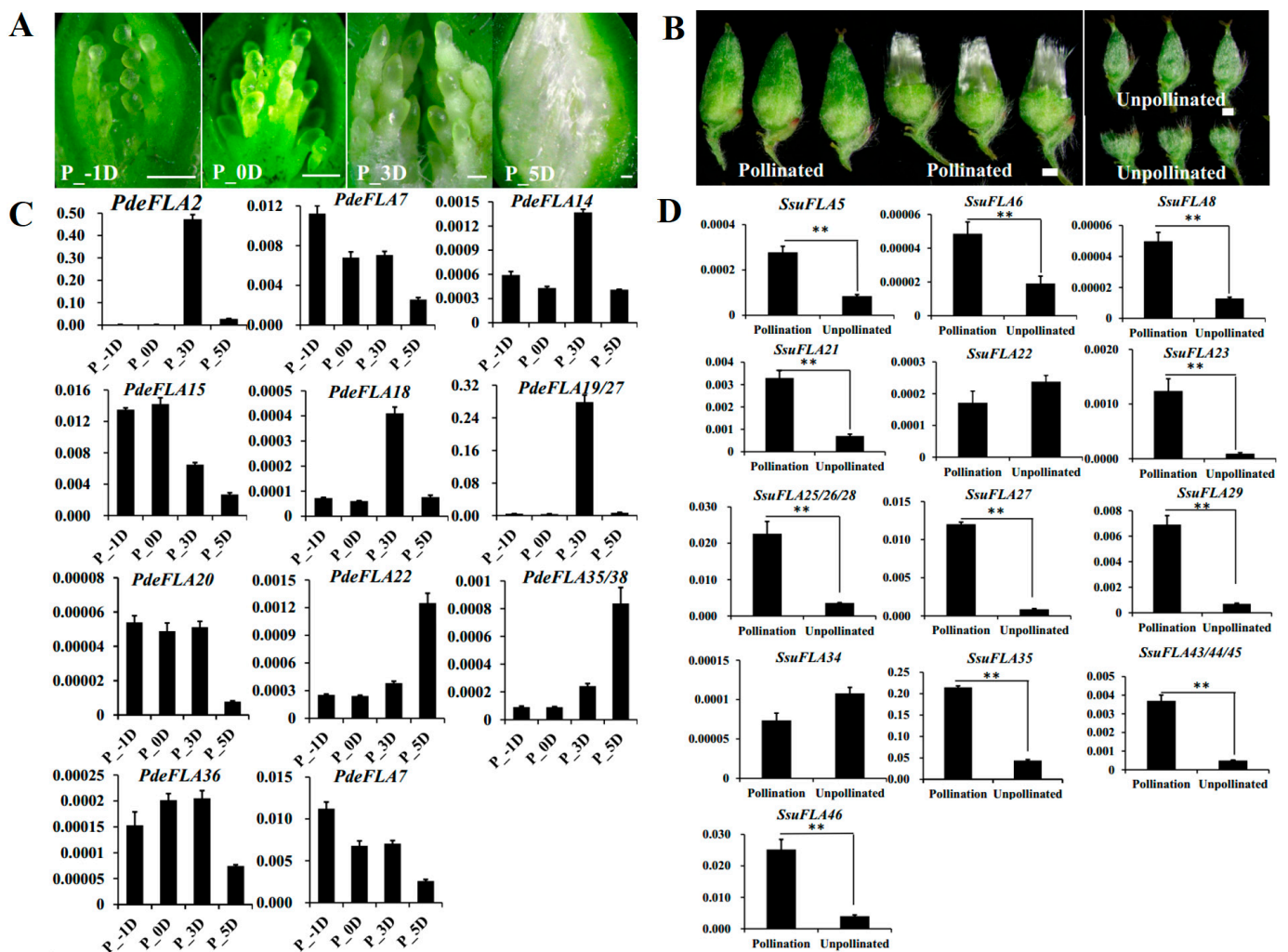


Figure 6. Analysis of the expression of *P. deltoides* and *S. suchowensis* FLA Group I genes involved in seed hair development. (A) Observation of different stages of seed hair development in *P. deltoides*. Scale bar = 0.5 mm. (B) Observation of seed hair phenotypes of *S. suchowensis* florets in pollinated versus unpollinated treatments. Scale bar = 0.5 mm. (C,D) Expression patterns of Group I FLA genes in *P. deltoides* and *S. suchowensis* (*p*-values are provided in Supplementary Table S9). The X-axis represents different tissues and the Y-axis represents relative expression. Values shown are means \pm SD of three biological replicates (** *p* < 0.01, two-sided Student's *t*-test).

Furthermore, the RT-qPCR analysis showed that *PdeFLA2/18/19/27* genes were highly expressed specifically at 3 DAP, which was consistent with the rapid elongation development of seed hairs. The highest relative expression was seen in the *PdeFLA19/27* genes (Figure 6C). We therefore speculate that these genes have important regulatory roles in the development of cell walls during the rapid elongation of poplar seed hairs.

In *S. suchowensis*, we collected pollinated and unpollinated florets in the same catkin to analyze the expression of Group I FLA genes. The results showed that pollinated florets developed numerous seed hairs on 7 DAP, while unpollinated florets did not produce seed hairs (Figure 6B). Furthermore, RT-qPCR analysis showed that 15 of the 17 gene Group I genes were expressed at significantly higher levels in pollinated florets than in unpollinated tissue, except for *SsuFLA22/34* (Figure 6D). Of these, the *SsuFLA35* gene showed the highest relative expression in pollinated florets, and it is speculated that it may be involved in regulating the development of willow seed hair tissue.

3. Materials and Methods

3.1. Genome-Wide Identification and Sequence Analysis of *Populus* and *Salix* FLA Genes

The genome sequences of *P. deltoides* (PRJNA598948) and *S. suchowensis* (ASM1755242v1) were downloaded from the NCBI website (<https://www.ncbi.nlm.nih.gov/> accessed on 8 March 2021). The fasciclin domain genome sequence (PF02469) was downloaded from the Pfam database (<http://pfam.xfam.org/> accessed on 8 March 2021). HMM3.0 software was used to screen proteins containing this domain in the genome sequences of *P. deltoides* and *S. suchowensis* [37]. In addition, a local database was constructed from the whole genome sequences of *P. deltoides* and *S. suchowensis*, and a BLASTP search was performed using an FLA protein sequence of *A. thaliana*. The candidate protein sequences identified by the above two methods were combined to remove duplicates, incomplete sequencing, and protein sequences without complete coding frames. Candidate proteins were analyzed using the Pfam (<http://pfam.xfam.org/> accessed on 8 March 2021, PF02469) [38] and SMART (<http://smart.embl.de/> accessed on 8 March 2021) [39] databases, and protein sequences with incomplete domains were removed.

Prediction of the N-terminal signal peptides was performed on the SignalP 5.0 website (<https://services.healthtech.dtu.dk/service.php?SignalP-5.0> accessed on 8 March 2021) [40]. Prediction of the attachment sites of the C-terminal GPI anchor was performed in the BIG-PI database (http://mendel.imp.ac.at/sat/gpi/gpi_server.html accessed on 8 March 2021) [41]. The fasciclin domain, N-terminal signal peptide, and C-terminal GPI anchor attachment site were removed for further screening for AGP-like glycosylation sites. Finally, the sequences containing the fasciclin domain and AGP-like glycosylation regions were considered the *P. deltoides* and *S. suchowensis* FLA gene family sequences. ExPASy (https://web.expasy.org/compute_pi accessed on 8 March 2021) was used for the prediction of the pI (isoelectric point) and molecular weight (Mw) [42].

The nucleotide and protein sequences of the identified FLA gene family members were aligned with the whole genome sequences of *P. deltoides* and *S. suchowensis* using BLAST, and the corresponding genome sequences, chromosomal positions, and exon/intron distribution patterns of each member were obtained. In addition, the genome sequences of six other species were downloaded from different websites, including *Corylus avellana* (https://hardwoodgenomics.org/bio_data/4302786?tripal_pane=group_downloads accessed on 14 May 2021), *Quercus rubra* L. (https://phytozome-next.jgi.doe.gov/info/Quercus_v2.1 accessed on 14 May 2021), *Cinnamomum kanehirae* (https://phytozome-next.jgi.doe.gov/info/Ckanehirae_v3 accessed on 14 May 2021), *Setaria italica* (https://phytozome-next.jgi.doe.gov/info/Sitalica_v2.2 accessed on 14 May 2021), *Zea mays* L. (http://plants.ensembl.org/Zea_mays/Info/Index accessed on 14 May 2021), and *Cucumis sativus* L. (http://cucurbitgenomics.org/ftp/genome/cucumber/Chinese_long/v3/ accessed on 14 May 2021), and their respective FLA gene families were identified. Members of the FLA gene family of *A. thaliana* (21 sequences) [9] and *Oryza sativa* L. (27 sequences) [3] were also downloaded (Supplementary Sequence S1).

3.2. Chromosomal Distribution of FLA Genes

Using the annotation files of the *P. deltoides* and *S. suchowensis* genomes, the chromosome location information of the FLA gene family was extracted, and the length of each chromosome was obtained. MapChart software was used to map the chromosomal locations and relative distances in the FLA genes. Tandem duplications of FLA genes were confirmed based on two criteria: (1) genes distributed in the 100 kb range on chromosomes and (2) similarity of >70% of the sequence alignment of the two genes [43,44]. Synteny analysis between species and gene duplication within species were performed using MCScanX software. Visual mapping was completed using Tbttools and KaKs_Calculator 2.0 was used to calculate synonymous (Ks) and non-synonymous (Ka) substitution of FLA gene pairs.

3.3. Phylogenetic Analysis and Functional Prediction of FLA Genes

Multiple sequence comparisons of FLA full-length protein sequences of *P. deltooides* and *S. suchowensis* were performed using ClustalX 2.1 with the default parameters [45]. Neighbor-joining (NJ) phylogenetic trees were constructed using MAGE 11 with parameters set to Poisson correction, pairwise deletion, and bootstrap analysis with 1000 replicates [46]. All FLA protein sequences from *P. deltooides*, *S. suchowensis*, *C. avellana*, *Q. rubra* L., *C. kanehirae*, *S. italica*, *Z. mays* L., *C. sativus* L., *O. sativa* L., and *A. thaliana* were clustered by MEGA 11 software for multiple sequence alignment and construction of the phylogenetic tree. The phylogenetic tree was visualized using the online program iTOL (<https://itol.embl.de/login.cgi> accessed on 11 August 2021), and a folded phylogenetic tree consisting of 10 species was constructed [47].

3.4. Analysis of Gene Structure and Protein Motif Identification

The intron and exon positions of the FLA genes were extracted, and the number and length of introns and exons were counted in generic feature format (GFF) files released for the whole genomes of *P. deltooides* and *S. suchowensis*. The exon and intron structures were displayed through the online program Gene Structure Display Server (<http://gsds.gao-lab.org/index.php> accessed on 8 March 2021) [48]. The MEME online program (<https://meme-suite.org/meme/tools/meme> accessed on 8 March 2021) was used to identify conserved motifs in the FLA proteins. The applied inclusion criteria were: (1) each FLA protein sequence appeared with at least 1 and a maximum of 20 motifs; (2) each motif had a minimum of 2 and a maximum of 250 amino acids in the sequence; and (3) other default parameters were used. The IDs of FLA members were kept the same as previously named.

3.5. Preparation of Plant Materials

Stems with different degrees of lignification were collected. Six-month-old *P. deltooides* and *S. suchowensis* were selected at the Plant Growth Laboratory of Nanjing Forestry University (China, Nanjing). Five parts of the stem, internodes 1–2, 3–4, 5–7, 8–9, and 10–11, were selected from the stem tip downwards and named S1, S2, S3, S4, and S5. Tissues at different stages of poplar seed hair development were also collected; all were from female *P. deltooides* flowering branches. Hydroponics was used to promote flowering. Pollination was performed with the stigma fully expanded, termed 0 days after pollination (0 DAP). Inflorescence ovaries were collected at –1 DAP, 0 DAP, 3 DAP, and 5 DAP in sequence. For the collection of willow seed hair tissue, male and female flowering branches of *S. suchowensis* were selected to promote flowering using hydroponics. After opening of the female inflorescences, some of the florets were pollinated. Seven days after pollination, pollinated and unpollinated florets in the same inflorescence were collected separately. Three replicates of all samples were collected and were immediately frozen in liquid nitrogen after collection and stored at –80 °C for RNA extraction.

3.6. Primer Design, RNA Extraction, and qRT-PCR

The primers used for analyzing the expression of FLA genes in *P. deltooides* and *S. suchowensis* were designed using Primer 5.0. The specificity of the primers was verified by agarose gel electrophoresis and the primer sequences are shown in Supplementary Table S10. *PtUBQ* (Potri.015G013600.1) [49] and *ACT7* (SapurV1A.0231s0320.1) [50] were used as internal reference genes for *P. deltooides* and *S. suchowensis*, respectively. Total RNA was extracted from the *P. deltooides* and *S. suchowensis* tissues using the RNA Easy Fast Plant Tissue Kit (TIANGEN, Nanjing, China). One microgram of RNA from each sample was reverse transcribed to cDNA using the One-Step gDNA Removal and cDNA Synthesis SuperMix (TIANGEN, Nanjing, China). Quantitative RT-PCR was performed on a 7500 Fast Real-Time PCR System (Applied Biosystems, USA). The amplification reaction system was as follows: 1 µL each of forward and reverse primers (10 µmol/L), 2 µL cDNA (5-fold dilution), 10 µL PowerUp™ SYBR™ Green Master Mix (Applied Biosystems), and RNase-free water was added to make the solution up to 20 µL. The amplification procedure

was as follows: 95 °C for 15 s, 60 °C for 30 s, and 72 °C for 30 s, with 40 cycles. Finally, the relative expression of each FLA gene was calculated using the CT method [51].

4. Conclusions

In this study, we performed a systematic analysis of the FLA gene family in *P. deltoides* and *S. suchowensis*, including the identification of the FLA genes and analysis of their structure and chromosomal distribution, as well as the identification of conserved domains and the evolution of the gene family in multiple species. The analysis revealed that FLA genes on the Group I branch of the phylogenetic tree were involved in the regulation of secondary growth processes, such as lignification in poplar and willow, and further showed that these genes had undergone significant expansion in woody plants, suggesting that this group of genes may have been important in the evolution of woody plants from herbaceous plants. Finally, analysis of the expression of the Group I genes in different tissues confirmed that the *PdeFLA19/27* gene was highly expressed in poplar, specifically during stem lignification and seed hair development, suggesting its important regulatory role in stem or seed hair development. Furthermore, it was verified that a set of duplicated genes, *SsuFLA25/26/28*, in willow were important candidates for stem lignification development, and that *SsuFLA35* may be involved in the regulation of germplasm hair development. In summary, this study provides a multi-faceted reference for subsequent studies on the functions of FLA genes in poplar and willow.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms24021481/s1>.

Author Contributions: Conceptualization, H.W. (Huitong Wu) and T.Y.; software, Y.Z. and F.Z.; validation, Y.Z., H.W. (Hui Wang) and Y.C.; data curation, Y.Z. and H.W. (Huitong Wu); writing—original draft preparation, Y.Z.; writing—review and editing, H.W. (Huitong Wu); visualization, Y.Z.; supervision, H.W. (Huitong Wu); project administration, H.W. (Huitong Wu); funding acquisition, T.Y., Y.C., and H.W. (Huitong Wu). All authors have read and agreed to the published version of the manuscript.

Funding: This study was financially supported by the Key Research and Development Project of Jiangsu Province, China (grant number BE2021366); the Natural Science Foundation of China (grant numbers 32071795); and the Youth Science and Technology Innovation Fund of Nanjing Forestry University (grant number CX2018008).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article or Supplementary Material.

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

References

1. Kieliszewski, M.J.; Shpak, E. Synthetic genes for the elucidation of glycosylation codes for arabinogalactan-proteins and other hydroxyproline-rich glycoproteins. *Cell. Mol. Life Sci.* **2001**, *58*, 1386–1398. [[CrossRef](#)] [[PubMed](#)]
2. Schultz, C.J.; Rumsewicz, M.P.; Johnson, K.L.; Jones, B.J.; Gaspar, Y.M.; Bacic, A. Using genomic resources to guide research directions. The arabinogalactan protein gene family as a test case. *Plant Physiol.* **2002**, *129*, 1448–1463. [[CrossRef](#)] [[PubMed](#)]
3. Ma, H.; Zhao, J. Genome-wide identification, classification, and expression analysis of the arabinogalactan protein gene family in rice (*Oryza sativa* L.). *J. Exp. Bot.* **2010**, *61*, 2647–2668. [[CrossRef](#)]
4. Showalter, A.M.; Keppler, B.; Lichtenberg, J.; Gu, D.; Welch, L.R. A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. *Plant Physiol.* **2010**, *153*, 485–513. [[CrossRef](#)]
5. Seifert, G.J.; Roberts, K. The biology of arabinogalactan proteins. *Annu. Rev. Plant Biol.* **2007**, *58*, 137–161. [[CrossRef](#)]
6. Pereira, A.M.; Pereira, L.G.; Coimbra, S. Arabinogalactan proteins: Rising attention from plant biologists. *Plant Reprod.* **2015**, *28*, 1–15. [[CrossRef](#)]
7. Gaspar, Y.; Johnson, K.L.; McKenna, J.A.; Bacic, A.; Schultz, C.J. The complex structures of arabinogalactan-proteins and the journey towards understanding function. *Plant Mol. Biol.* **2001**, *47*, 161–176. [[CrossRef](#)]
8. Tan, L.; Leykam, J.F.; Kieliszewski, M.J. Glycosylation motifs that direct arabinogalactan addition to arabinogalactan-proteins. *Plant Physiol.* **2003**, *132*, 1362–1369. [[CrossRef](#)]

9. Johnson, K.L.; Jones, B.J.; Bacic, A.; Schultz, C.J. The fasciclin-like arabinogalactan proteins of Arabidopsis. A multigene family of putative cell adhesion molecules. *Plant Physiol.* **2003**, *133*, 1911–1925. [[CrossRef](#)]
10. Meng, J.; Hu, B.; Yi, G.; Li, X.; Chen, H.; Wang, Y.; Yuan, W.; Xing, Y.; Sheng, Q.; Su, Z.; et al. Genome-wide analyses of banana fasciclin-like AGP genes and their differential expression under low-temperature stress in chilling sensitive and tolerant cultivars. *Plant Cell Rep.* **2020**, *39*, 693–708. [[CrossRef](#)]
11. Jun, L.; Xiaoming, W. Genome-wide identification, classification and expression analysis of genes encoding putative fasciclin-like arabinogalactan proteins in Chinese cabbage (*Brassica rapa* L.). *Mol. Biol. Rep.* **2012**, *39*, 10541–10555. [[CrossRef](#)] [[PubMed](#)]
12. Guerriero, G.; Mangeot-Peter, L.; Legay, S.; Behr, M.; Lutts, S.; Siddiqui, K.S.; Hausman, J.F. Identification of fasciclin-like arabinogalactan proteins in textile hemp (*Cannabis sativa* L.): In silico analyses and gene expression patterns in different tissues. *BMC Genom.* **2017**, *18*, 1–13. [[CrossRef](#)] [[PubMed](#)]
13. Hossain, M.; Ahmed, B.; Ullah, M.; Aktar, N.; Haque, M.; Islam, M. Genome-wide identification of fasciclin-like arabinogalactan proteins in jute and their expression pattern during fiber formation. *Mol. Biol. Rep.* **2020**, *47*, 7815–7829. [[CrossRef](#)] [[PubMed](#)]
14. Huang, G.Q.; Xu, W.L.; Gong, S.Y.; Li, B.; Wang, X.L.; Xu, D.; Li, X.B. Characterization of 19 novel cotton FLA genes and their expression profiling in fiber development and in response to phytohormones and salt stress. *Physiol. Plant.* **2008**, *134*, 348–359. [[CrossRef](#)] [[PubMed](#)]
15. MacMillan, C.P.; Taylor, L.; Bi, Y.; Southerton, S.G.; Evans, R.; Spokevicius, A. The fasciclin-like arabinogalactan protein family of Eucalyptus grandis contains members that impact wood biology and biomechanics. *New Phytol.* **2015**, *206*, 1314–1327. [[CrossRef](#)]
16. Wu, X.; Lai, Y.; Lv, L.; Ji, M.; Han, K.; Yan, D.; Lu, Y.; Peng, J.; Rao, S.; Yan, F.; et al. Fasciclin-like arabinogalactan gene family in Nicotiana benthamiana: Genome-wide identification, classification and expression in response to pathogens. *BMC Plant Biol.* **2015**, *20*, 1–15. [[CrossRef](#)]
17. Zang, L.; Zheng, T.; Chu, Y.; Ding, C.; Zhang, W.; Huang, Q.; Su, X. Genome-wide analysis of the fasciclin-like arabinogalactan protein gene family reveals differential expression patterns, localization, and salt stress response in Populus. *Front Plant Sci.* **2015**, *6*, 1140. [[CrossRef](#)]
18. Faik, A.; Abouzouhair, J.; Sarhan, F. Putative fasciclin-like arabinogalactan-proteins (FLA) in wheat (*Triticum aestivum*) and rice (*Oryza sativa*): Identification and bioinformatic analyses. *Mol. Genet. Genom.* **2006**, *276*, 478–494. [[CrossRef](#)]
19. Hozumi, A.; Bera, S.; Fujiwara, D.; Obayashi, T.; Yokoyama, R.; Nishitani, K.; Aoki, K. Arabinogalactan proteins accumulate in the cell walls of searching hyphae of the stem parasitic plants, *Cuscuta campestris* and *Cuscuta japonica*. *Plant Cell Physiol.* **2017**, *58*, 1868–1877. [[CrossRef](#)]
20. MacMillan, C.P.; Mansfield, S.D.; Stachurski, Z.H.; Evans, R.; Southerton, S.G. Fasciclin-like arabinogalactan proteins: Specialization for stem biomechanics and cell wall architecture in Arabidopsis and Eucalyptus. *Plant J.* **2010**, *62*, 689–703. [[CrossRef](#)]
21. Shi, H.; Kim, Y.; Guo, Y.; Stevenson, B.; Zhu, J.K. The Arabidopsis SOS5 locus encodes a putative cell surface adhesion protein and is required for normal cell expansion. *Plant Cell* **2003**, *15*, 19–32. [[CrossRef](#)]
22. Liu, E.; MacMillan, C.P.; Shafee, T.; Ma, Y.; Ratcliffe, J.; Van de Meene, A.; Bacic, A.; Humphries, J.; Johnson, K.L. Fasciclin-like arabinogalactan-protein 16 (FLA16) is required for stem development in Arabidopsis. *Front. Plant Sci.* **2020**, *11*, 615392. [[CrossRef](#)]
23. Wang, H.; Jiang, C.; Wang, C.; Yang, Y.; Yang, L.; Gao, X.; Zhang, H. Antisense expression of the fasciclin-like arabinogalactan protein FLA6 gene in Populus inhibits expression of its homologous genes and alters stem biomechanics and cell wall composition in transgenic trees. *J. Exp. Bot.* **2015**, *66*, 1291–1302. [[CrossRef](#)]
24. Wang, H.; Jin, Y.; Wang, C.; Li, B.; Jiang, C.; Sun, Z.; Zhang, Z.; Kong, F.; Zhang, H. Fasciclin-like arabinogalactan proteins, PtFLAs, play important roles in GA-mediated tension wood formation in Populus. *Sci. Rep.* **2017**, *7*, 6182. [[CrossRef](#)]
25. Zhang, L.; Xi, Z.; Wang, M.; Guo, X.; Ma, T. Plastome phylogeny and lineage diversification of Salicaceae with focus on poplars and willows. *Ecol. Evol.* **2018**, *8*, 7817–7823. [[CrossRef](#)]
26. Hou, J.; Wei, S.; Pan, H.; Zhuge, Q.; Yin, T. Uneven selection pressure accelerating divergence of Populus and Salix. *Hortic. Res.* **2019**, *6*, 37. [[CrossRef](#)]
27. Gaut, B.; Yang, L.; Takuno, S.; Eguiarte, L.E. The patterns and causes of variation in plant nucleotide substitution rates. *Annu. Rev. Ecol. Evol. Syst.* **2011**, *42*, 245–266. [[CrossRef](#)]
28. Smith, S.A.; Donoghue, M.J. Rates of molecular evolution are linked to life history in flowering plants. *Science* **2008**, *322*, 86–89. [[CrossRef](#)]
29. Dai, X.; Hu, Q.; Cai, Q.; Feng, K.; Ye, N.; Tuskan, G.A.; Milne, R.; Chen, Y.; Wan, Z.; Wang, Z.; et al. The willow genome and divergent evolution from poplar after the common genome duplication. *Cell Res.* **2014**, *24*, 1274–1277. [[CrossRef](#)]
30. Brown, D.M.; Zeef, L.A.; Ellis, J.; Goodacre, R.; Turner, S.R. Identification of novel genes in Arabidopsis involved in secondary cell wall formation using expression profiling and reverse genetics. *Plant Cell* **2005**, *17*, 2281–2295. [[CrossRef](#)]
31. Persson, S.; Wei, H.; Milne, J.; Page, G.P.; Somerville, C.R. Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 8633–8638. [[CrossRef](#)] [[PubMed](#)]
32. Ito, S.; Suzuki, Y.; Miyamoto, K.; Ueda, J.; Yamaguchi, I. AtFLA11, a fasciclin-like arabinogalactan-protein, specifically localized in sclerenchyma cells. *Biosci. Biotechnol. Biochem.* **2005**, *69*, 1963–1969. [[CrossRef](#)] [[PubMed](#)]
33. Johnson, K.L.; Kibble, N.A.; Bacic, A.; Schultz, C.J. A Fasciclin-Like Arabinogalactan-Protein (FLA) Mutant of Arabidopsis thaliana, fla1, Shows Defects in Shoot Regeneration. *PLoS ONE* **2011**, *6*, e25154. [[CrossRef](#)] [[PubMed](#)]
34. Li, J.; Yu, M.; Geng, L.L.; Zhao, J. The fasciclin-like arabinogalactan protein gene, FLA3, is involved in microspore development of Arabidopsis. *Plant J.* **2010**, *64*, 482–497. [[CrossRef](#)]

35. Nieminen, K.; Blomster, T.; Helariutta, Y.; Mähönen, A.P. Vascular cambium development. *Arab. Book* **2015**, *2015*, 13. [[CrossRef](#)]
36. Sehr, E.M.; Agusti, J.; Lehner, R.; Farmer, E.E.; Schwarz, M.; Greb, T. Analysis of secondary growth in the Arabidopsis shoot reveals a positive role of jasmonate signalling in cambium formation. *Plant J.* **2010**, *63*, 811–822. [[CrossRef](#)]
37. Eddy, S.R. Profile hidden Markov models. *Bioinformatics* **1998**, *14*, 755–763. [[CrossRef](#)]
38. Finn, R.D.; Mistry, J.; Schuster-Böckler, B.; Griffiths-Jones, S.; Hollich, V.; Lassmann, T.; Moxon, S.; Marshall, M.; Khanna, A.; Durbin, R.; et al. Pfam: Clans, web tools and services. *Nucleic Acids Res.* **2006**, *34*, D247–D251. [[CrossRef](#)]
39. Letunic, I.; Doerks, T.; Bork, P. SMART 7: Recent updates to the protein domain annotation resource. *Nucleic Acids Res.* **2012**, *40*, D302–D305. [[CrossRef](#)]
40. Almagro Armenteros, J.J.; Tsirigos, K.D.; Sønderby, C.K.; Petersen, T.N.; Winther, O.; Brunak, S.; Von Heijne, G.; Nielsen, H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.* **2019**, *37*, 420–423. [[CrossRef](#)]
41. Eisenhaber, B.; Wildpaner, M.; Schultz, C.J.; Borner, G.H.; Dupree, P.; Eisenhaber, F. Glycosylphosphatidylinositol lipid anchoring of plant proteins. Sensitive prediction from sequence-and genome-wide studies for Arabidopsis and rice. *Plant Physiol.* **2003**, *133*, 1691–1701. [[CrossRef](#)] [[PubMed](#)]
42. Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; De Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E.; et al. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **2012**, *40*, W597–W603. [[CrossRef](#)]
43. Gu, Z.; Cavalcanti, A.; Chen, F.C.; Bouman, P.; Li, W.H. Extent of gene duplication in the genomes of Drosophila, nematode, and yeast. *Mol. Biol. Evol.* **2002**, *19*, 256–262. [[CrossRef](#)]
44. Yang, S.; Zhang, X.; Yue, J.X.; Tian, D.; Chen, J.Q. Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol. Genet. Genom.* **2008**, *280*, 187–198. [[CrossRef](#)]
45. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. [[CrossRef](#)]
46. Kumar, S.; Stecher, G.; Li, M.; Nnyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547. [[CrossRef](#)]
47. Subramanian, B.; Gao, S.; Lercher, M.J.; Hu, S.; Chen, W.H. Evolvview v3: A webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.* **2019**, *47*, W270–W275. [[CrossRef](#)]
48. Hu, B.; Jin, J.; Guo, A.Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* **2015**, *31*, 1296–1297. [[CrossRef](#)]
49. Gutierrez, L.; Mauriat, M.; Guénin, S.; Pelloux, J.; Lefebvre, J.F.; Louvet, R.; Rusterucci, C.; Moritz, T.; Guérineau, F.; Bellini, C.; et al. The lack of a systematic validation of reference genes: A serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. *Plant Biotechnol. J.* **2008**, *6*, 609–618. [[CrossRef](#)]
50. Li, J.; Jia, H.; Han, X.; Zhang, J.; Sun, P.; Lu, M.; Hu, J. Selection of reliable reference genes for gene expression analysis under abiotic stresses in the desert biomass willow, *Salix psammophila*. *Front. Plant Sci.* **2016**, *7*, 1505. [[CrossRef](#)]
51. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2[−]ΔΔCT method. *Methods* **2001**, *25*, 402–408. [[CrossRef](#)] [[PubMed](#)]

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