



Identification and function analysis of *fasciclin-like arabinogalactan protein* family genes in pear (*Pyrus bretschneideri*)

Xiaoqiang Li¹ · Mengyu Cheng¹ · Chao Tang¹ · Xiaoxuan Zhu¹ · Kaijie Qi¹ · Shaoling Zhang¹ · Juyou Wu¹ · Peng Wang¹

Received: 13 January 2021 / Accepted: 5 June 2021 / Published online: 29 June 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2021

Abstract

Plant-specific fasciclin-like arabinogalactan proteins (FLAs) are a subclass of the arabinogalactan proteins (AGPs) superfamily. In addition to AGP-like glycosylated regions, FLAs have conserved fasciclin (FAS) domains. Here, we identified 220 *FLA* genes from seven Rosaceae species, including 38 *FLA* gene in pear. Based on gene structure and phylogenetic analysis, the Rosaceae *FLA* genes can be divided into four classes. The Ks and 4DTv values suggested that the *PbrFLA* gene family had undergone two whole-genome duplication events occurring at 30–45 MYA and ~140 MYA, respectively. Whole-genome duplication (WGD) and transposed duplication (TRD) events mainly drove the evolution of *PbrFLA* gene family. Most pear *FLAs* from pear had no intron in their genomic DNA sequences. Pear *FLAs* possess two highly conserved regions (H1 and H2) and the conserved [Tyr Phe] His ([Y/F]H) motif locating between these two regions. Based on gene expression analysis, most pear *FLAs* exhibited tissue-specific patterns. *PbrFLA10*, *PbrFLA20*, and *PbrFLA21* were highly expressed in pollen tubes and self-pollinated styles, indicating *FLAs* play roles in pollen tube growth and self-incompatibility response. Repression of *PbrFLA10/20/21* resulted in the acceleration of pear pollen tube growth. Taken together, our results provided information for understanding the evolution of the *PbrFLA* gene family and identified the key *PbrFLAs* genes regulating pollen tube growth.

Keywords Evolution · *FLA* · Gene expression · Pear · Pollen tube

Introduction

The pollen tube is one of the fastest growing organs in flowering plants, which responsible for delivering sperm cells through the pistil to the ovules. During its rapid growth, new components of the cell wall and plasma membrane are

continuously deposited at the tip of the growing pollen tube. The role of different structural proteins and polysaccharides such as hydroxyproline-rich glycoproteins (HRGPs) are crucial to maintain the integrity of the pollen tube cell wall (Sede et al. 2020).

HRGPs are typical cell-wall proteins that participate in plant growth, development and immunity (Castilleux et al. 2018). The HRGP superfamily can be classified into three subfamilies: arabinogalactan proteins (AGPs),

Handling editor: Christian Parisod.

Xiaoqiang Li and Mengyu Cheng have contributed equally.

✉ Peng Wang
wangpeng@njau.edu.cn

Xiaoqiang Li
2018104032@njau.edu.cn

Mengyu Cheng
2018104033@njau.edu.cn

Chao Tang
617627294@qq.com

Xiaoxuan Zhu
2018104031@njau.edu.cn

Kaijie Qi
qikaijie@njau.edu.cn

Shaoling Zhang
slzhang@njau.edu.cn

Juyou Wu
juyouwu@njau.edu.cn

¹ State Key Laboratory of Crop Genetics and Germplasm Enhancement, Centre of Pear Engineering Technology Research, Nanjing Agricultural University, Nanjing 210095, China

pro-rich proteins (PRPs) and extensins (EXTs) (Wu et al. 2020). Among the three subfamilies, AGPs have the highest degree of glycosylation. Previous studies speculated that two *HRGPs* (*Pbr036330.1* and *Pbr010506.1*) might be related to gametophytic self-incompatibility (GSI) in pear (Jiao et al. 2018).

AGPs as one of the major family of glycoproteins in plant cell wall may function in gametophytic cell differentiation, male–female interactions and pollen tube growth (Pereira et al. 2016; Su and Higashiyama 2018). Most of AGPs are predicted to be glycosylphosphatidylinositol (GPI) anchored which is a possible mode of release of AGPs to the extracellular medium where AGPs may exert their functions (Pereira et al. 2016). In *Arabidopsis*, the expression of *AGP6* and *AGP11* genes is related to the development of pollen grains and the start of the germination process (Coimbra et al. 2009). In tobacco, AGP components of the stylar extracellular matrix (ECM) directly interact with pollen tubes (Cruz-Garcia et al. 2005). It is indicated that AGPs may be involved in pollen grain formation, initiation of pollen grain germination and pollen tube guidance in the transmitting tract of pistil.

Fasciclin-like AGPs (FLAs) are one of the subfamilies of AGP. To date, genome-wide identification and comparative analysis of *FLA* gene families have been reported in many plant species, including *Arabidopsis thaliana* (Johnson et al. 2003), rice (*Oryza sativa*) (Ma and Zhao 2010), wheat (*Triticum aestivum*) (Faik et al. 2006), cotton (*Gossypium hirsutum*) (Huang et al. 2008), Chinese cabbage (*Brassica rapa*) (Jun and Xiaoming 2012), eucalypt (*Eucalyptus grandis*) (MacMillan et al. 2015), poplar (*Populus trichocarpa*) (Showalter et al. 2016), textile hemp (*Cannabis sativa*) (Guerriero et al. 2017). The members of the *FLA* proteins possess at least one fasciclin domains that consist of 110 to 150 amino acids (aa) that contain two highly conserved regions (H1 and H2) and one [Phe/Tyr]-His motif (Johnson et al. 2011). This domain has been shown to function as adhesion motifs in extracellular matrix (Kim et al. 2000). Most of *FLAs* also contain a GPI membrane anchor that allows anchoring the *FLA* molecule to the cell membrane. Evidence from several studies indicates that *FLAs* are likely to be important during plant growth and development, such as fiber development (Wang et al. 2015), cell wall regeneration (Showalter et al. 2016), cell-to-cell communication (Jun and Xiaoming 2012), glycan stabilization (Xue et al. 2017) and in response to stresses from cold (Takahashi et al. 2016) and hydrogen peroxide (Zhang et al. 2015). Otherwise, *AtFLA3* and *AtFLA5* may perform function in pollen development and affect pollen intine formation in *Arabidopsis* (Li et al. 2010). However, the function of *FLA* in plant sexual reproduction and the development of pollen tube has never been reported in Rosaceae.

Pear, one of the typical GSI plant in Rosaceae fruit trees, is widely cultivated all over the world (Wu et al. 2013). In this study, we have identified 220 *FLA* family genes from pear and other six species of Rosaceae, and performed comprehensive analysis on their evolutionary history. Moreover, the expression patterns of *PbrFLA10/20/21* in reproductive tissues were investigated to speculate their functional roles in pollen tube growth. Suppressing the expression of *PbrFLA10/20/21* resulted in the acceleration of pollen tube growth in pear. Overall, the identification and bioinformatic analysis of *FLAs* in pear will open the way for studying their biological functions in pollen tube growth and development as well as evolutionary history.

Materials and methods

Plant materials

Pear (*Pyrus bretschneideri*, ‘Dangshansuli’ variety) trees grown in Fruit Experimental Yard of Nanjing Agricultural University, China were used in this study. Root (from seedling), leaf (from bearing tree), fruit (mature fruit), different developmental stages of pollen tubes, the styles (on the day of flowering) and the styles with different pollination were collected and frozen immediately in liquid nitrogen and stored at -80°C . Anthers of the ‘Dangshansuli’ variety were collected and dried for one day and then stored in silica gel at -20°C . Pollen grains were cultured in sealed tube with liquid medium at 24°C for hydration, germination and growth. The medium components were 0.5 mM $\text{Ca}(\text{NO}_3)_2$, 1.5 mM H_3BO_3 , 450 mM sucrose, 15% (w/v) PEG4000 and 25 mM 2-(N-morpholino) ethanesulfonic acid hydrate, with pH 6.0–6.5 (pH was adjusted with KOH). Pollen tubes longer than their diameters were considered germinated. The length of pollen tubes was measured by Image-Pro software, with approximately 150 pollen tubes measured for length. A great number of pollen grains (0.3 g mature dry pollen per sample) were cultured for different time and used for RNA isolating. The different developmental stages of pollen tubes, mature pollen grains (MP) with 0 min post-cultured, hydrated pollen grains (HP) with 40 min post-cultured, growing pollen tubes (PT) with 6 h post-cultured and stopped-growth pollen tubes (SPT) with 15 h post-cultured, were immediately frozen in liquid nitrogen and kept at -80°C for RNA extractions (Zhou et al. 2016). As for the styles with different pollination, we collected the styles of the ‘Dangshansuli’ variety at 24 h after self-pollination (SP) and cross-pollination with ‘Cuiguan’ variety (CP), respectively. The styles of ‘Dangshansuli’ variety without pollination as control (UP). The pistils were isolated and stored in liquid nitrogen.

Whole-genome identification of *FLA* genes in seven species of Rosaceae

The genome sequences of seven species of Rosaceae were downloaded from each genome project (Online Resource 4). Then, the 21 protein sequences of identified AtFLAs were obtained (Schultz et al. 2002). Subsequently, the alignment file for the fasciclin domain (PF02469) obtained from the Pfam database (<https://pfam.sanger.ac.uk/>) (Finn et al. 2010) was used to build a hidden Markov model (HMM) file. Then, a HMM search was conducted against the seven species of Rosaceae protein database by using HMMER3 software (Eddy 2011). Furthermore, NCBI Batch CD-Search (<https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>) (Marchler-Bauer et al. 2015, 2017) was used to retain the candidate protein sequences based on the existence of the fasciclin domain (smart00554, cl02663, pfam02469). Subsequently, the GPI anchor was identified by big-PI Plant Predictor (https://mendel.imp.ac.at/sat/gpi/gpi_server.html) (Eisenhaber et al. 1999). The ExPASy (https://web.expasy.org/compute_pi) was used to predict the pI (isoelectric point) and MW (molecular weight).

Phylogenetic analysis of the *FLA* genes

Maximum-likelihood (ML) phylogenetic tree was constructed with all 241 full-length FLA protein sequences of seven Rosaceae species and *Arabidopsis* by IQ-TREE (Trifinopoulos et al. 2016). The best-fit substitution model VT + F + R7 was determined by ModelFinder (Kalyaanamoorthy et al. 2017), and the bootstrap values were 1000. Evolview (version 2) (Zhang et al. 2012a, b), an online visualization tool, was used to beautify phylogenetic tree.

Chromosomal localization and synteny, K_a , K_s and 4DTv analysis of *FLA*

Five duplication events, whole-genome duplication (WGD)/segmental duplication, singleton duplication (SD), tandem duplication (TD), proximal duplication (PD), and transposed duplication (TRD), of gene pairs among the seven species of Rosaceae were identified by using the DupGen_finder pipeline (Qiao et al. 2019). Kaks_calculator 2.0 (Wang et al. 2010) was used to calculate the K_a , K_s substitution rates with the Yang and Nielsen method (YN method). ParaAT 2.0 (Zhang et al. 2012a, b) was used to calculate corrected 4DTv values of *FLA* with the Hasegawa, Kishino and Yano method (HKY method) (Hasegawa et al. 1985). The density plot of the distribution of K_s and 4DTv values was plotted using ggplot2 package in R.

Gene structure, conserved motifs and multiple sequence alignment analysis of the pear *FLA* genes

Gene structural information for *PbrFLA* genes was obtained from the pear genome database and displayed using TBtools software (Chen et al. 2020). Conserved motifs of the genes were analyzed by the Multiple EM for Motif Elicitation (MEME) program (Liu et al. 2016) with the following parameters: optimum motif width was set to 30–70, and the maximum number of motifs was set to identify 15 motifs. Multiple sequence alignments corresponding to conserved motifs characteristic of the PbrFLA proteins were determined by Jalview with muscle method (Waterhouse et al. 2009).

Expression pattern analysis of *PbrFLA* genes by transcriptome sequencing

RNA-seq data of pear pollen (PRJNA299117) were acquired from Sequence Read Archive database (SRA, <https://www.ncbi.nlm.nih.gov/sra>). Clean reads were obtained by fastp software (Chen et al. 2018a, b). Hisat2 (Kim et al. 2015) was used to align clean reads to the reference genome, and featureCounts (Liao et al. 2014) was used to estimate transcript abundance levels. The transcripts per million (TPM) values were used to measure the expression levels of the *PbrFLA* genes. The RNA-seq expression profile of pear styles with different pollination (<https://figshare.com/s/49b94f0fc5ff1d926afe>) was obtained from the previous study (Shi et al. 2017). Then, the expression levels were transformed from fragments per kilobase million (FPKM) to TPM using R code. The pheatmap, a function in R, was used to plot heatmaps based the TPM values of each *PbrFLA* genes with normalization method of Z-score standardization.

Quantitative real-time PCR analysis (qRT-PCR)

The expression patterns of the *PbrFLA* genes were analyzed by qRT-PCR analysis. Total RNA was extracted using Plant Total RNA Isolation Kit Plus (FOREGENE Co. Ltd.). Then, the total RNA was adjusted to the same concentration, and based on the adjusted RNA, first-strand cDNA was synthesized by using TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen Biotech Co. Ltd.). The specific primers were designed by using NCBI online software (<https://www.ncbi.nlm.nih.gov/tools/primer-blast>), which amplified 38 *PbrFLA* gene sequences (Online Resource 13). qRT-PCR analysis was performed by using the LightCycler 480 SYBR GREEN I Master (Roche). A 20 μ l mixed reaction system was performed, each contained with 150 ng of template cDNA, 0.5 μ M of each pair primers and 10 μ l of LightCycler 480 SYBR GREEN I Master. All reactions were performed in 96-well plates and each cDNA

sample had three replicates. qRT-PCR procedure was set as follows: first 5 min at 95 °C for pre-incubation, 55 cycles at 95 °C for 3 s, 60 °C for 10 s, and 72 °C for 30 s, and then 3 min at 72 °C for extension. Finally, the step of fluorescence signal data collection was carried out at 60 °C. *Pyrus* UBC was used as the internal control genes (Chen et al. 2015). The data were analyzed by LightCycler 480 Software v1.5.0 (Roche) using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001). The expression data were analyzed and graphed by R.

Antisense oligo deoxynucleotide experiment

Both phosphorothioate antisense oligo deoxynucleotide (as-ODN) and sense control (s-ODN) were designed using RNA fold Web Server (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>). Soligo software was used to calculate efficient candidate as-ODN sequences for suitable target regions (<http://sfold.wadsworth.org/soligo.pl>). The antisense oligonucleotide experiment was performed as previously described (Moutinho et al. 2001; Chen et al. 2018a, b). The length of the pollen tubes was measured using a Leica DM6B microscope with LASX software. The length of at least 100 pollen tubes was measured on each condition. The sequences of oligonucleotide primers are listed in Online Resource 13.

Construction of gene co-expression network

To investigate the regulatory network between target genes and transcription factors (TFs), random forest regression was estimated for each gene based on the TFs as inputs using the GENIE3 package in Python3 with default parameters (Huynh-Thu et al. 2010). By the GENIE3 algorithm, we calculated the connectivity between target genes and TFs based on the RNA-seq data (different development of pollen tubes and different pollinated styles). For each co-expressed gene pair, the value of connectivity > 0.005 was extracted (Ramirez-Gonzalez et al. 2018). Networks were visualized by using Cytoscape software (Maere et al. 2005).

Statistical analysis

All experimental data were analyzed by t.test() function in R. Two groups of samples were compared using Student's *t*-test.

Results

Whole-genome characterization and classification of the *FLA* genes in seven Rosaceae species

Most *FLA* proteins have low overall similarity among plant (He et al. 2019); therefore, it is difficult to find

family members using Basic Local Alignment Search Tool (BLAST) searches. In order to comprehensively identify the candidate *FLA* genes in pear genome, we searched for proteins with the fasciclin domain across the whole-genome sequence of pear. The seed file of fasciclin (FAS) (PF02469) from Pfam (<http://pfam.janelia.org/>) was used to obtain the hidden Markov model (HMM) sequence file; then, HMM searches were performed in HMMER3.0 software against the pear protein database (<http://peargenome.njau.edu.cn/>). A total of 41 candidate PbrFLA proteins were identified. To further verify the reliability of candidate *FLA* proteins, NCBI Batch CD-Search was used to detect the completeness of fasciclin domain of candidate proteins. As a result, three candidate proteins without FAS domain were removed. Consequently, a total of 38 non-redundant and complete *FLA* proteins were identified in pear genome for further analysis. To distinguish members of the *FLA* family, we named each of *PbrFLA* genes from *PbrFLA1* to *PbrFLA38*, based on their order on the chromosomes (Online Resource 6). Among these 38 *PbrFLA* genes, 30 genes could be mapped on chromosomes, and eight genes on the unanchored scaffolds.

Using the same strategies, we identified 33, 28, 33, 37, 31 and 20 *FLA* genes from apple (*Malus domestica*), peach (*Prunus persica*), Chinese plum (*Prunus mume*), sweet cherry (*Prunus avium*), strawberry (*Fragaria vesca*) and black raspberry (*Rubus occidentalis*) genomes, respectively (Online Resource 4). The lengths of the *FLA*s protein sequences ranged from 100 to 1197 amino acids. The molecular weights and protein isoelectric points (PI) were ranged from 11.02 to 126.24 kD and 4.11 to 11.10, respectively (Online Resource 6).

To clarify the evolutionary relationship of *FLA* family in seven Rosaceae species and 21 identified protein sequences of *Arabidopsis* (Schultz et al. 2002), a ML phylogenetic tree for the putative *FLA* proteins was inferred. Based on the classification of subfamilies in *Arabidopsis*, a total of 241 *FLA* proteins from eight species were phylogenetically categorized into four groups, namely class I, class II, class III and class IV. Most *FLA* genes were contained in class IV (96 of 241), followed by class III (70 of 241), class I (60 of 241) and class II (15 of 241) (Fig. 1, Online Resource 5). Presence and number of fasciclin domains and GPI were also used to interpret the classification (Online Resource 6). As previously described, members of class I possess a single fasciclin domain and most (81.48%) have GPI anchored signal. The smallest group is class II and members contain two fasciclin domains (except Ro07_G07830 and MD07G1214900) but have no C-terminal GPI anchor site. Members of subclass III have either one to five fasciclin domains, and most (70.97%) have a C-terminal GPI

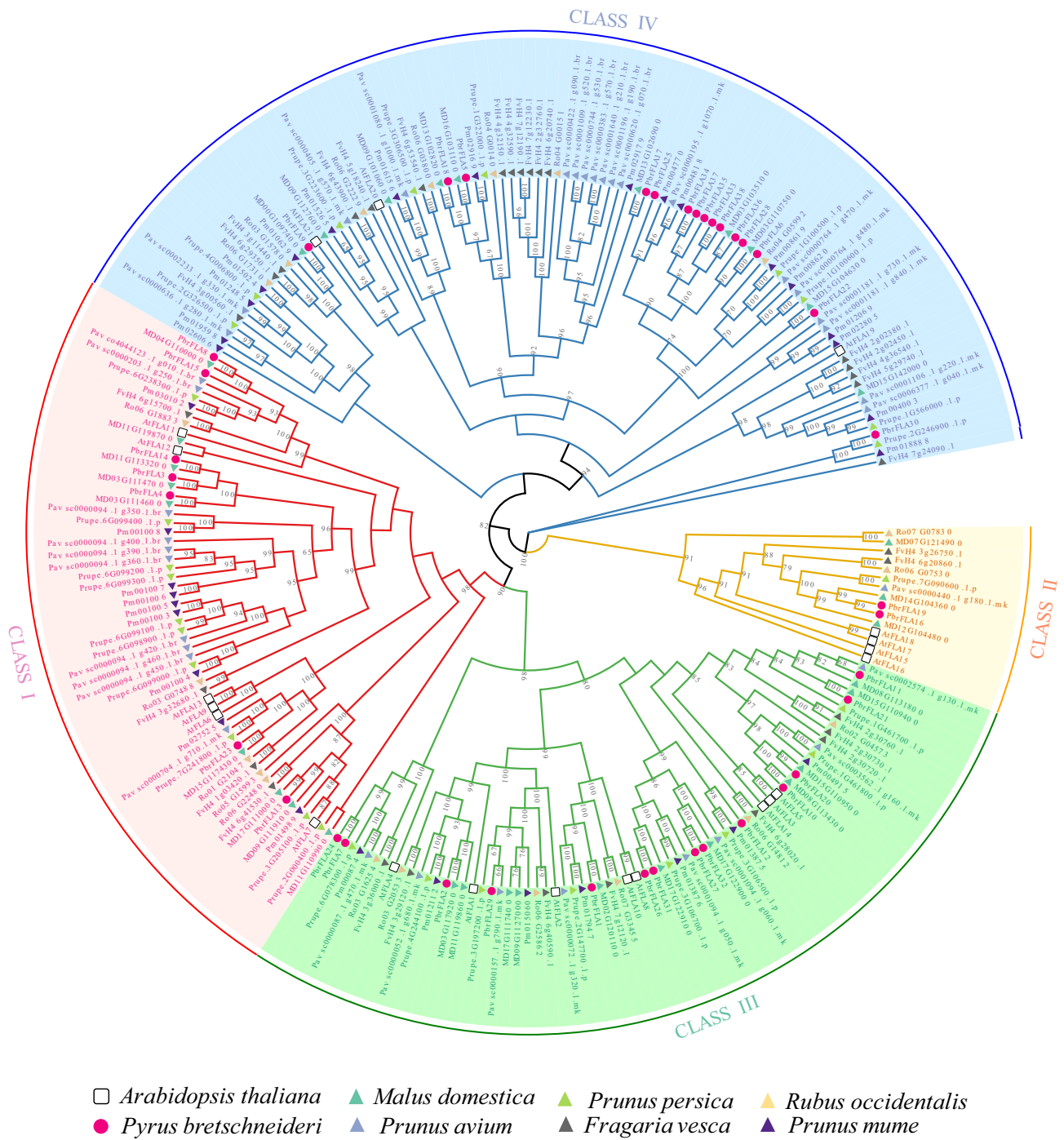


Fig. 1 Maximum-likelihood phylogenetic tree and protein features of *FLA* proteins from *Arabidopsis* and Rosaceae. The red, yellow, green and blue indicate the four subfamilies (class I, class II, class III and

class IV), respectively. The different shapes and colors indicate different species

anchor site. The remaining *FLA* proteins constitute subclass IV, which contains *FLAs* that are quite distantly related to the other *FLAs* and which have no consistent pattern in the number of fasciclin domains or the presence of a GPI signal.

Synten analysis reveals the dates and driving forces for evolution of the *FLA* gene family

Duplication events have contributed to the evolution of novel functions, such as the production of floral structures, induction of disease resistance, and adaptation to stress (Panchy

et al. 2016). To investigate the origin of *FLA* family genes, we analyzed five modes of gene duplication, whole-genome duplication (WGD), tandem duplication (TD), proximal duplication (PD), transposed duplication (TRD) and dispersed duplication (DSD), in seven Rosaceae species and identified their contributions to the expansion of the *FLA* gene family. A total of 267 duplicated gene pairs were found in seven Rosaceae species, with the maximum number of gene pairs derived from DSDs (168 gene pairs), followed by TRDs (33 gene pairs), WGDs (31 gene pairs), TDs (27 gene pairs) and PDs (8 gene pairs) (Table 1). Although a high frequency of DSD-genes was detected, we should not ignore the important roles of other types of single-gene duplication such as WGD and TRD in the expansion of the *PbrFLA* gene family. Notably, the percentage of gene pairs retained following WGD duplication in pear (20.97%) and apple (23.91%) was higher than that in peach (2.94%), strawberry (3.23%), Japanese apricot (5.56%), sweet cherry (5.26%), and black raspberry (5.00%).

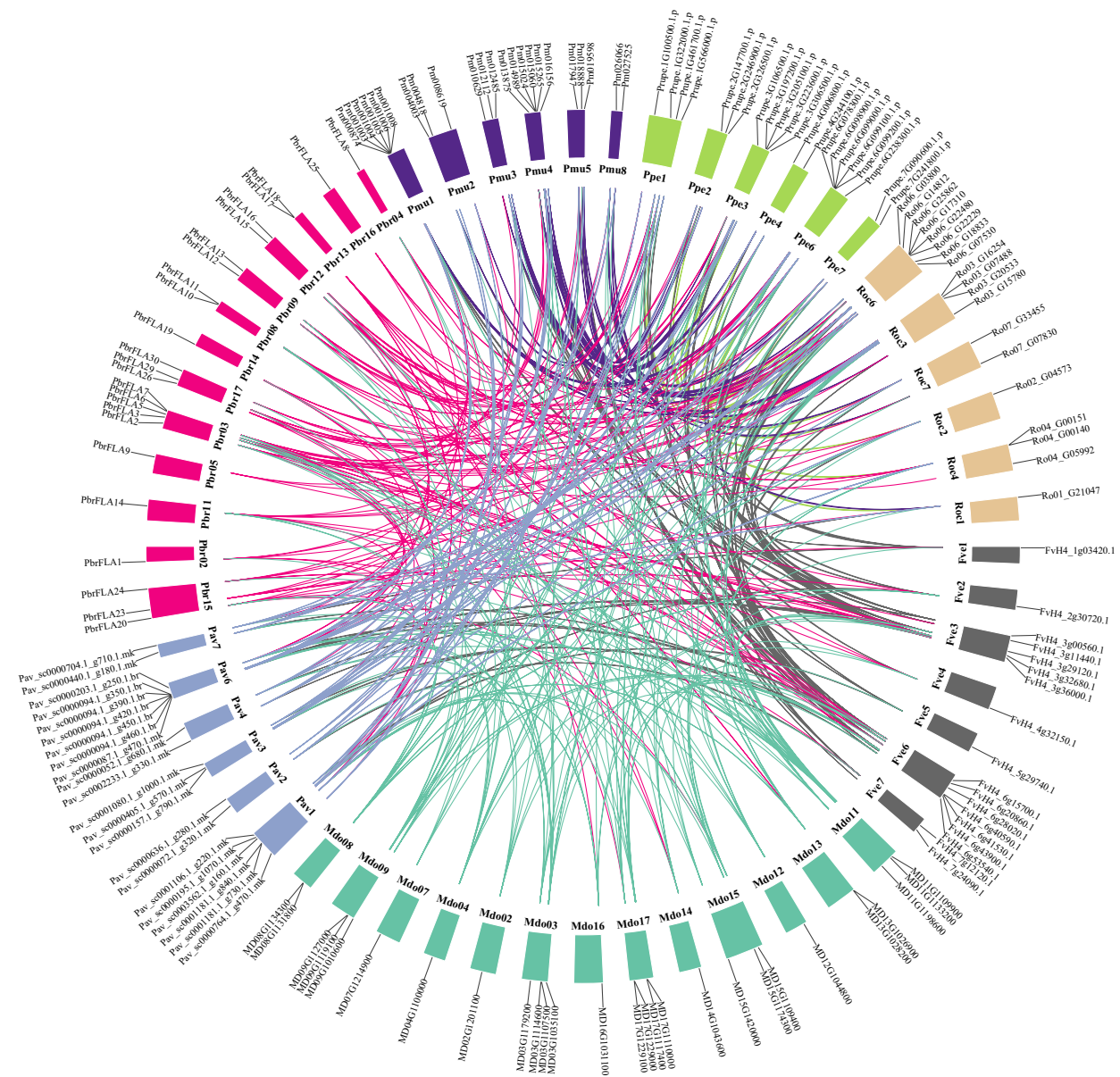
To reveal the evolutionary routes of the *FLA* gene family, we evaluated distribution of the *FLA* genes on the chromosomes and both inter- and intra-genomic synteny analyses within seven Rosaceae species. The landscape of 454 inter-species collinear orthologous *FLA* gene pairs among Rosaceae species presented in Fig. 2, suggesting the strong homology of *FLA* genes in Rosaceae species. Further, we detected six genes (*FvH4_2g30720.1*, *MD08G1134300*, *MD15G1109400*, *Pav_sc0003562.1_g160.1.mk*, *Prupe.1G461700.1.p*, *Ro02_G04573*) that have collinear relationship with *PbrFLA10* and seven genes (*FvH4_2g30720.1*, *MD08G1134300*, *MD08G1131800*, *MD15G1109400*, *Pav_sc0003562.1_g160.1.mk*, *Prupe.1G461700.1.p*, *Ro02_G04573*) that have collinear relationship with *PbrFLA20* (Online Resource 9). After that, intra-genomic collinearity of five duplication events was also investigated in each species, and 62 pairs were found in pear, 46 pairs in apple, 34 pairs in peach, 38 pairs in sweet cherry, 36 pairs in Japanese apricot, 31 pairs in strawberry, and 20 pairs in black raspberry (Fig. 3, Table 1). The chromosome localization of collinear orthologous *FLA* gene pairs was detected (Fig. 3). In pear and apple, the gene pairs were

mainly distributed on chromosomes 3 and 15. In pear and apple, the gene pairs were mainly distributed on chromosomes 3 and 15. In Japanese apricot, the gene pairs were mainly distributed on chromosomes 1. In the rest of the four species, the gene pairs were mainly distributed on chromosomes 6. Thus, the *FLA* of the seven Rosaceae species was distributed unevenly on chromosomes.

The *Ks* value is used to estimate the evolutionary history of WGD events (Blanc and Wolfe 2004). In this study, there were two peaks with *Ks* value at approximately 0.15–0.30 and 1.5–1.8 in pear and apple, respectively (Fig. 4b). In detail, nine WGD-pairs of pear and seven WGD-pairs of apple fell in the range from about 0.15 to 0.30. Simultaneously, three WGD-pairs of pear and three WGD-pairs of apple mapped to the region with values from about 1.5 to 1.8 (Online Resource 8). Furthermore, the recent WGD event with 4DTv (four-fold degenerate site transversion) of ~0.08 and an ancient WGD event with 4DTv of ~0.5 (Wu et al. 2013). From the distribution of 4DTv (Fig. 4b, Online Resource 8), there were two peaks with a 4DTv value at approximately 0.08 and 0.5 in pear and apple, respectively. This result further supported the conclusions of *Ks* analysis, which indicated the *FLA* gene family had undergone two WGD events in pear and apple. By combining *Ks* and 4DTv analysis, we concluded that six (*PbrFLA26-PbrFLA12*, *PbrFLA12-PbrFLA31*, *PbrFLA16-PbrFLA19*, *MD09G1119100-MD17G1110000*, *MD13G1028200-MD16G1031100*, *MD03G1179200-MD11G1198600*) and three WGD-pairs (*PbrFLA20-PbrFLA11*, *MD03G1114600-MD04G1100000*, *MD08G1134300-MD15G1109400*) were duplicated during the recent and ancient WGD duplication event, respectively. The *Ka/Ks* value equals to one showed neutral evolution, positive selection when the *Ka/Ks* value is greater than one, and purifying selection when the *Ka/Ks* value is lower than one (Yang 2007). Except *Pav_sc0000094.1_g360.1.br-Pav_sc0000094.1_g390.1.br*, *Pav_sc0001106.1_g220.1.mk-Pav_sc0006377.1_g040.1.mk*, *Pav_sc0001009.1_g520.1.br-Pav_sc0001640.1_g210.1.br*, most *Ka/Ks* ratios of the *FLA* gene pairs were lower than one demonstrating *FLA* primarily evolved under purifying selection (Fig. 4a, Online Resource 8).

Table 1 Numbers of *FLA* gene pairs from different duplication events in seven Rosaceae intra-genomics

	<i>Fragaria vesca</i>	<i>Malus domestica</i>	<i>Prunus avium</i>	<i>Pyrus bretschneideri</i>	<i>Prunus mume</i>	<i>Prunus persica</i>	<i>Rubus occidentalis</i>	SUM
DSD	24	23	29	31	24	22	15	168
WGD	1	11	2	13	2	1	1	31
TRD	3	8	2	10	4	3	3	33
TD	1	4	5	4	6	7	0	27
PD	2	0	0	4	0	1	1	8
SUM	31	46	38	62	36	34	20	267



▲ *Pyrus bretschneideri* ▲ *Malus domestica* ▲ *Prunus persica* ▲ *Rubus occidentalis*
 ▲ *Prunus avium* ▲ *Fragaria vesca* ▲ *Prunus mume*

Fig. 2 Inter-species collinear orthologous *FLA* gene pairs among seven Rosaceae species. Black lines along the circumference of the circle mark the gene positions. The lines in different colors inside the circle indicate collinearity relationships among the genes from seven Rosaceae species

Gene structural and conserved motif analysis of *PbrFLA*

Gene structure analysis provides valuable information to interpret the phylogenetic relationship of members within gene family (Li et al. 2009). To identify gene structures and evolutionary trajectories of *FLA* genes in pear, we investigated exon–intron compositions of the 38 *PbrFLAs*.

Analysis of the genomic DNA sequences showed that none of the *PbrFLAs* contained intron except that *PbrFLA12*, *PbrFLA19* and *PbrFLA29* had one intron (Fig. 5a). As previously reported that introns in *FLA* genes were lost during plant evolution, especially from green algae to land plants (He et al. 2019). Most members (79%) of class III had two untranslated regions (UTR), whereas no UTR regions were detected from other three classes except *PbrFLA14* and

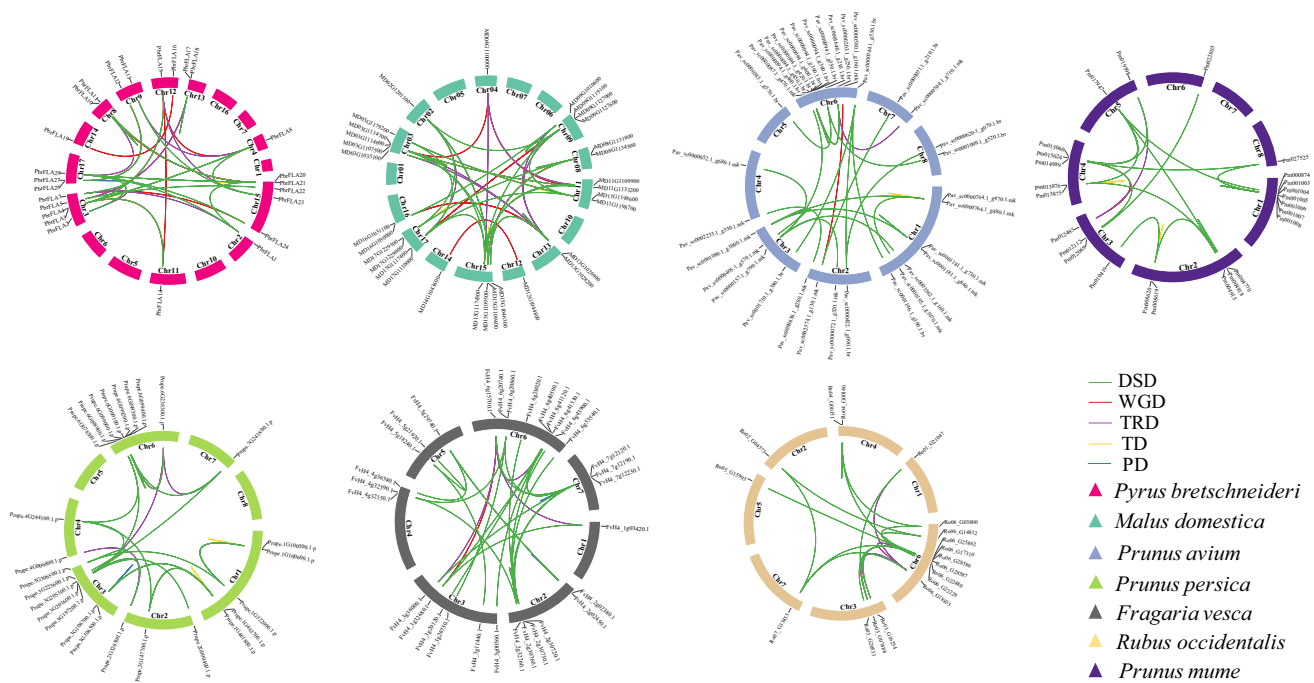


Fig. 3 Intra-genomic collinearity and localization of the *FLA* genes in seven Rosaceae. The circle in different colors indicate different species and the lines in different colors inside the circle indicate different duplication modes. Whole-genome duplication (WGD), tandem

duplication (TD), proximal duplication (PD), transposed duplication (TRD), dispersed duplication (DSD). Intra-genomic collinearity *FLA* genes pairs in seven Rosaceae species were mapped on the different chromosomes

PbrFLA30 (Fig. 5a), suggesting that *PbrFLAs* in class III might have unique and significant biological functions.

To further provide information to understand the sequence and structure of *PbrFLA* proteins in pear, MEME tool was used to predict the conserved motifs of *PbrFLA* proteins. A total of 15 conserved motifs were identified in our study, named from motif 1 to motif 15 (Online Resource 7). As shown in Fig. 5b, most of members of *PbrFLA* protein contained conserved motifs 1, 4, 6 and 8, while the others were unique to one or few classes. For example, all members of class I contained both motifs 1, 4 and 6; motifs 10 were present only in class II; except for *PbrFLA11* and *PbrFLA12*, the N-terminal and C-terminal of all proteins had motif 1 and motif 8 in class III, respectively; class IV was clearly less closely related to the other subclasses, and motifs 2, 5, 7, 9, 14 and 15 were unique to this class. It is noteworthy that all genes except for class IV had motif 4, suggesting that motif 4 was critical for class I, class II and class III members function.

Multiple sequence alignment of the putative fasciclin domains in 38 *PbrFLA* proteins were conducted using Jalview software. The predicted fasciclin domains of *PbrFLAs* possess two highly conserved regions (H1 and H2) of approximately 10 amino acids each, and the [Y/F]H motif locating between these two regions (Online Resource 1). The Thr residue in the H1 region is highly

conserved and the other conserved residues such as Val/Ile/Leu and Asn/Asp that may play a role in maintaining the structure of the fasciclin domain and/or cell adhesion (Kim et al. 2000). Small hydrophobic amino acids, such as Leu, Val and Ile, are abundant in the H2 region. In the [Y/F]H motif, His and Pro residues are also relatively conserved. In most cases, H1 and H2 also contain some other conserved residues, such as Val, Ile and Leu. There are about 105 amino acids between H1 and H2. On the contrary, the similarity of amino acid sequence between H1 and H2 is low, though some residues, such as Gly, Asn, Thr and Ser, are found to be conserved in most *PbrFLAs*.

Expression patterns of the *PbrFLA* family

The expression patterns of *PbrFLAs* were investigated in six different tissues of pear by qRT-PCR. Except *PbrFLA12/15/35* which showed no expression in any of the six different tissues, 35 *PbrFLAs* exhibited obviously tissue-specific patterns and were mainly expressed in pollen, root and stem (Fig. 6a). Of note, *PbrFLAs* in class III were preferentially expressed in the pollen compared to other classes (Fig. 6d). Meanwhile, *PbrFLA10/20/21* presented a relatively higher expression level in pollen (Online Resource 10). Subsequently, RNA-seq data were utilized to analyze the expression patterns of *PbrFLA* during different stages of pollen tube and different pollinated

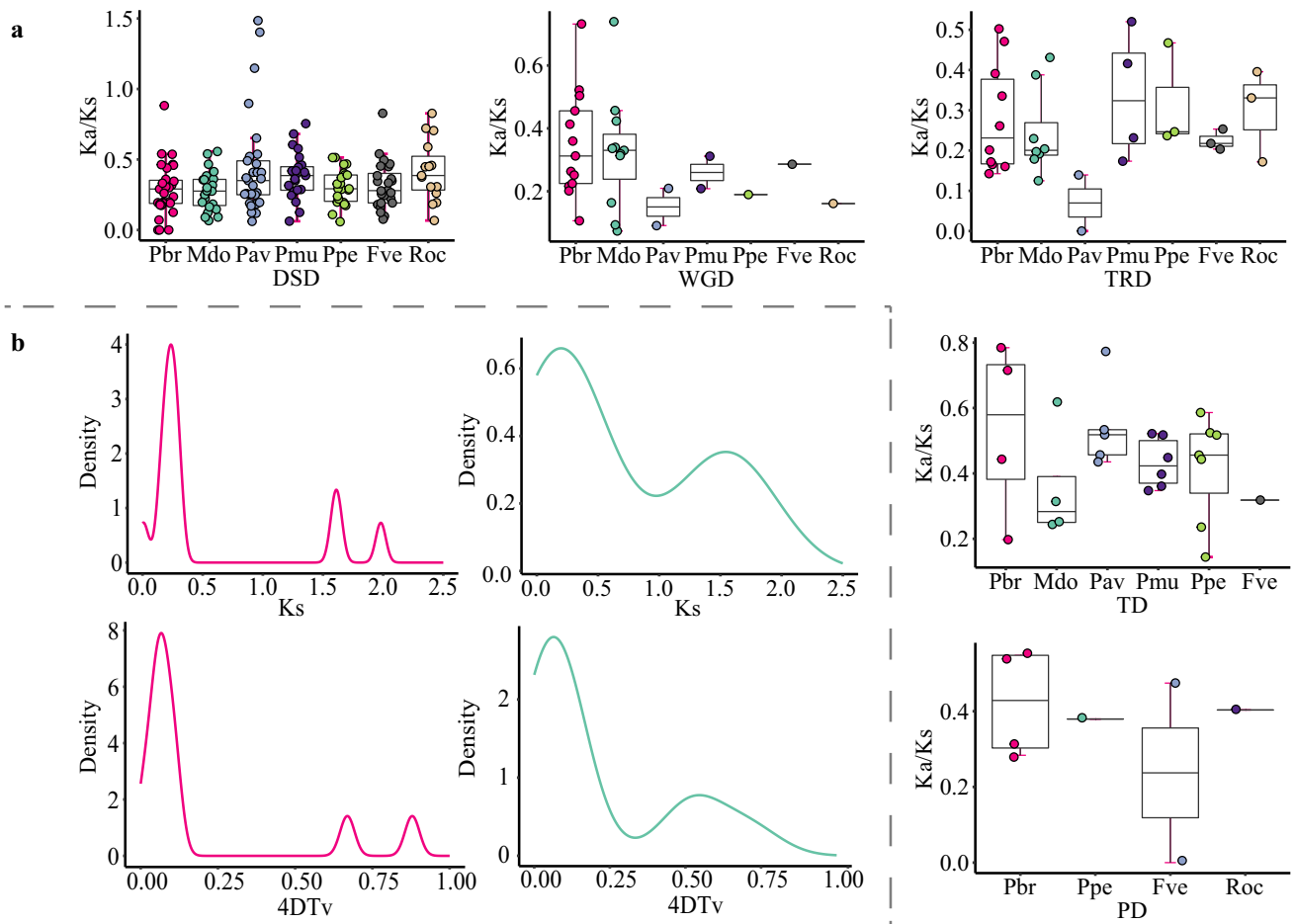


Fig. 4 The evolution analysis of *PbrFLA* gene family. **a** Comparison of Ka/Ks values for different modes of gene duplications. **b** The distribution of mean Ks and 4DTv value of *FLA* duplicated gene pairs in pear (red line) and apple (blue line)

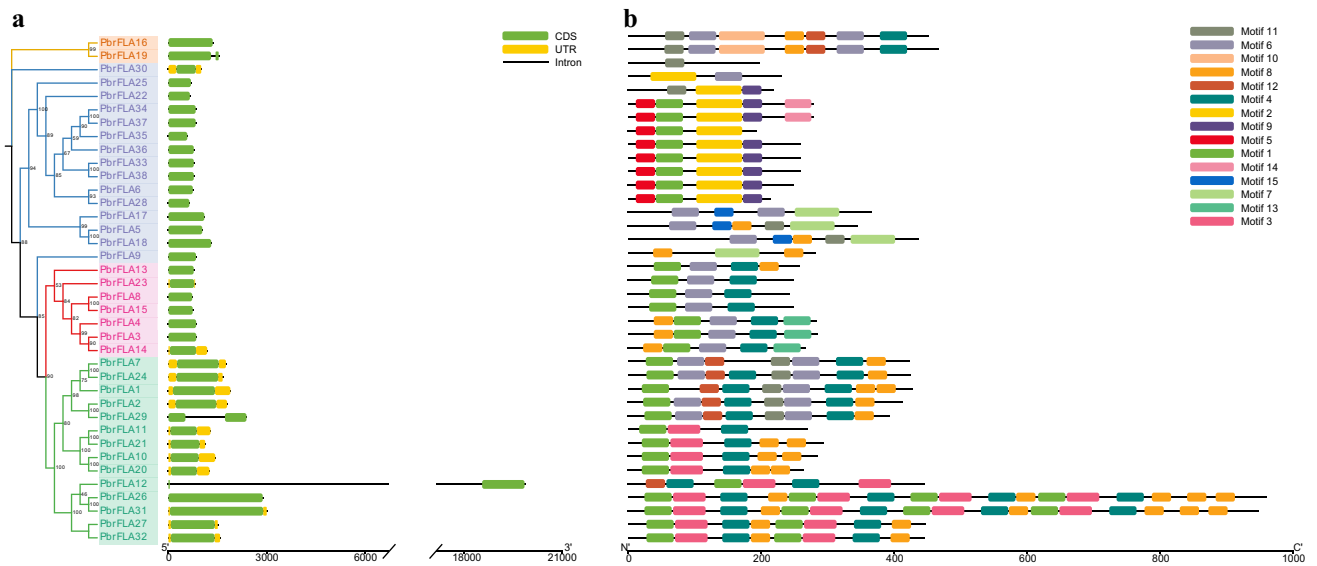


Fig. 5 The phylogenetic relationship, conserved motifs of *PbrFLAs* and gene structure of *PbrFLAs*. **a** The gene structure of *PbrFLA* genes, including intron, CDS, UTR. The black line indicates intron;

the green rectangle indicates CDS; the yellow rectangle UTR. **b** The conserved motifs in *PbrFLA* proteins of pear identified by MEME. Fifteen motifs were indicated by different colors

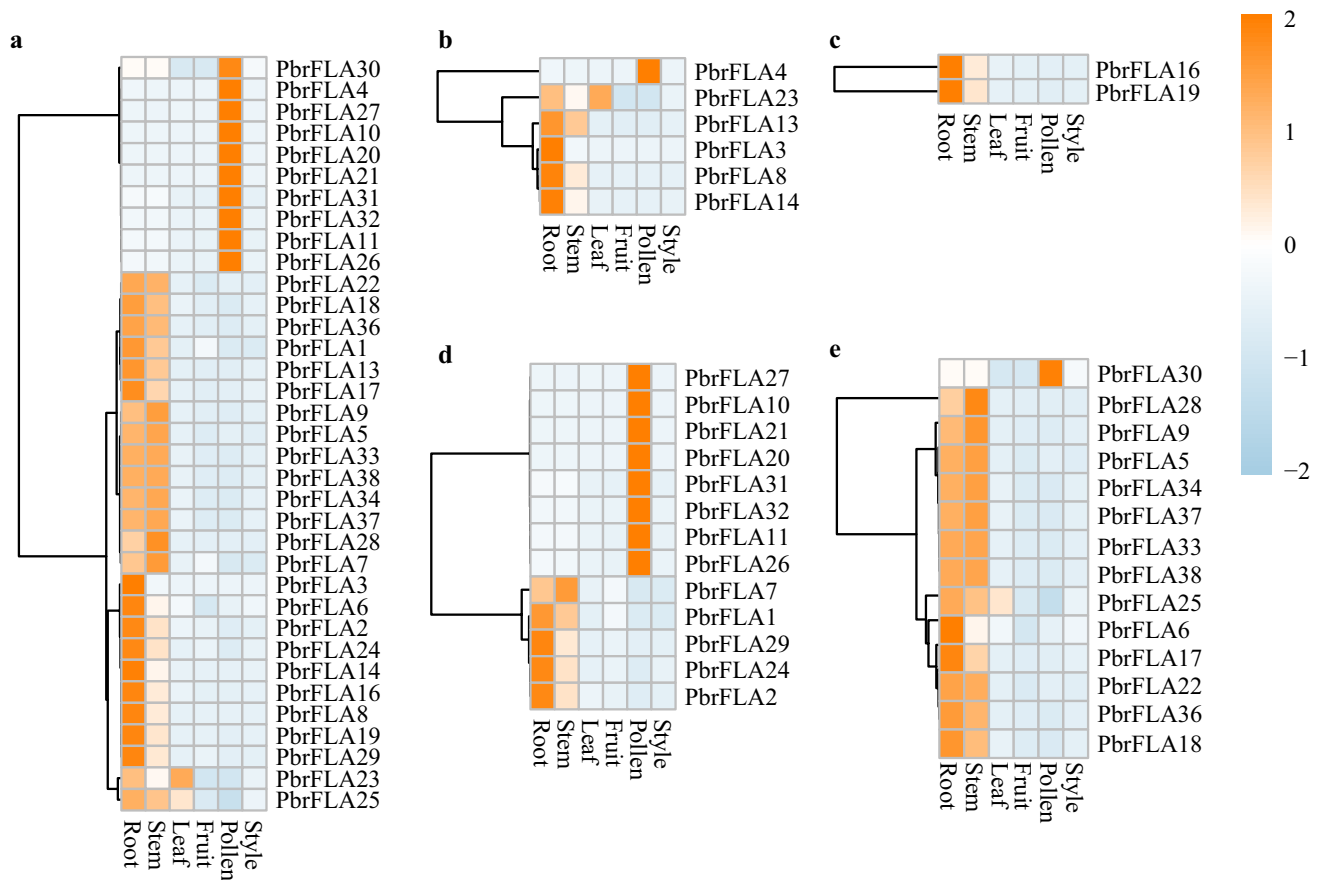


Fig. 6 Expression analysis of *PbrFLAs* by qRT-PCR in root, stem, leaf, fruit, pollen, and style. The color scale represents the calculated Z-score. Orange indicates high expression. Root (from seedling); leaf (from bearing tree); fruit (mature fruit); pollen (mature pollen grain);

style (the style on the day of flowering). **a** Expression of *PbrFLAs* in all class. **b** Expression of *PbrFLAs* in class I. **c** Expression of *PbrFLAs* in class II. **d** Expression of *PbrFLAs* in class III. **e** Expression of *PbrFLAs* in class IV

styles of pear (Zhou et al. 2016; Shi et al. 2017). As shown in Online Resource 2, the heatmap of the *PbrFLA* genes expression levels was hierarchically, and 23 (60.53%) *PbrFLAs* were expressed in pollinated pistil, and 25 (66.79%) were expressed in pollen, indicating the significant role of *PbrFLAs* in male–female interactions and growth of pollen tube. Notably, the three genes (*PbrFLA10/20/21*) have high expression in self-pollinated style (SP) and in stop growing pollen tube (STP) (Fig. 7, Online Resource 11, Online Resource 12). To verify the reliability of the RNA-seq data, nine *PbrFLA* genes were selected to conduct a qRT-PCR experiment to investigate the expression levels in growing pollen tubes and styles after pollination, respectively. The expression levels of most *PbrFLA* genes were consistent with the RNA-seq data (Fig. 7).

Functional analysis of candidate *PbrFLA* genes in pear pollen tube growth

In order to confirm the contribution of *PbrFLA10/20/21* to pear pollen tube growth, we applied antisense oligo

deoxynucleotide (ODN) technique (Mizuta and Higashiyama 2014; Chen et al. 2018a, b) in pear pollen. As shown in Fig. 8c, comparing the negative control sense ODN (s-ODN), antisense ODN (as-ODN) was able to downregulate the expression levels of *PbrFLA10/20/21*. The specific knockdown of *PbrFLA10*, *PbrFLA20* and *PbrFLA21* led to a significant acceleration in pollen tube length comparing to the control groups (control, cytofectin and s-ODN) (Fig. 8a, Fig. 8b). It is suggested that pear pollen tube growth might be regulated by *PbrFLA10*, *PbrFLA20* and *PbrFLA21*.

Gene co-expression network analysis

To further investigate the expression regulation of the *PbrFLA10*, *PbrFLA20* and *PbrFLA21* in pear, a co-expression network of the *FLA* genes was constructed. First, we obtained all the transcription factors (TFs) in pear from PlantTFDB (<http://planttfdb.cbi.pku.edu.cn/>). Then, based on the RNA-seq data, we calculated co-expression relationship between the target genes (*PbrFLA10/20/21*) and TFs by

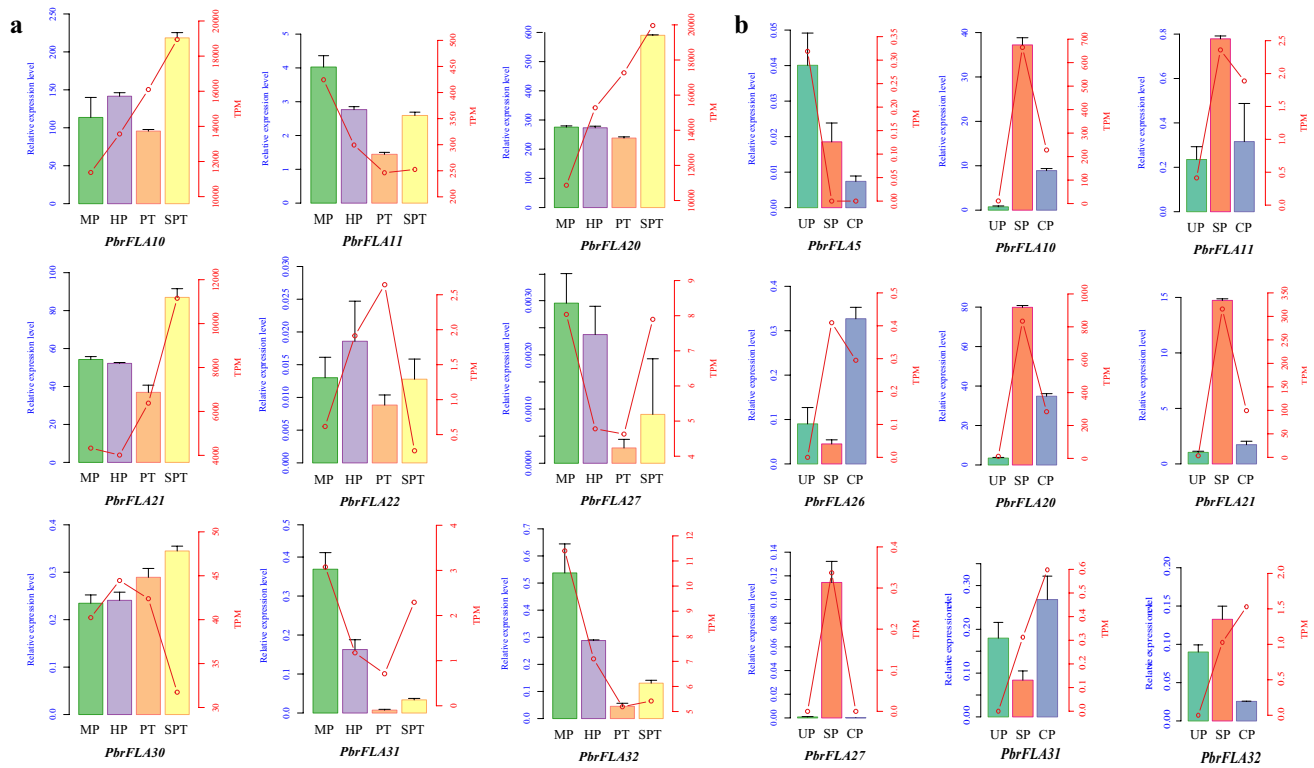


Fig. 7 qRT-PCR and RNA-seq data analysis of the *PbrFLA* genes in different developmental stages of pear pollen tubes and the styles with different pollination. The right y-axis represents the TPM values of RNA-seq; the left y-axis represents the relative expression levels. **a** The x-axis represents the different developmental stages of pollen tubes. HP, PT, and SPT correspond to three different developmental

stages: HP hydrated pollen grains, PT growing pollen tubes, and SPT stop growing pollen tubes. **b** The x-axis represents the different pollinated styles. UP, SP and CP correspond to three different pollinated style: UP un-pollinated styles, SP self-pollinated styles, CP cross-pollinated styles

GENIE3 algorithm that exploits variable importance scores derived from random forests to identify the regulators of each target gene (Huynh-Thu et al. 2010). 50 pairs between *PbrFLA10/20/21* and TFs have co-expression relationships (Online Resource 14). Gene pairs were used to construct the co-expression network (Online Resource 3). As a result, there were seven, five and four TFs that revealed a strong co-expression relationship with *PbrFLA10*, *PbrFLA20* and *PbrFLA21*, respectively. According to the connectivity values, *Pbr005306.1* (C2H2), *Pbr000646.1* (EIL) and *Pbr001365.1* (MYB_related) have the strongest co-expression correlation with *PbrFLA10*, *PbrFLA20* and *PbrFLA21*, respectively.

Discussion

As crucial role of the plant cell development, some members of the *FLA* gene family have been extensively studied in plants. However, there have been few such efforts to annotate the *FLA* gene family in Rosaceae. In this study, we have identified 220 *FLA* genes which encoded a conserved fasciclin domain in pear and other six species of Rosaceae.

Except for black raspberry, the number of *FLA* genes in the Rosaceae was much higher than those reported in *Arabidopsis thaliana*. According to the classification of *FLAs* in *Arabidopsis* (Johnson et al. 2003), phylogenetic analysis classified the Rosaceae *FLA* proteins into four subgroups (class I, class II, class III and class IV), suggesting that the evolution of different subfamilies was relatively independent. Previous studies have reported that class III genes may play crucial roles in strong growth tissues, such as male gametophyte (He et al. 2019). For example, *AtFLA3* was involved in microspore development, and its knock-down plants showed reduced female fertility (Li et al. 2010). *MTR1* regulates male reproductive development in rice (Tan et al. 2012). *AtFLA1* might play an important role in root development (Johnson et al. 2011).

WGD event, important for plant adaptability in stress, is an extreme mechanism of gene duplication that leads to a sudden increase in both genome size and the entire gene set (Panchy et al. 2016). TRD event is presumed to arise through distantly transposed duplications occurred by DNA-based or RNA-based mechanisms, and is significant for plant to obtain new functions or to partition existing functions

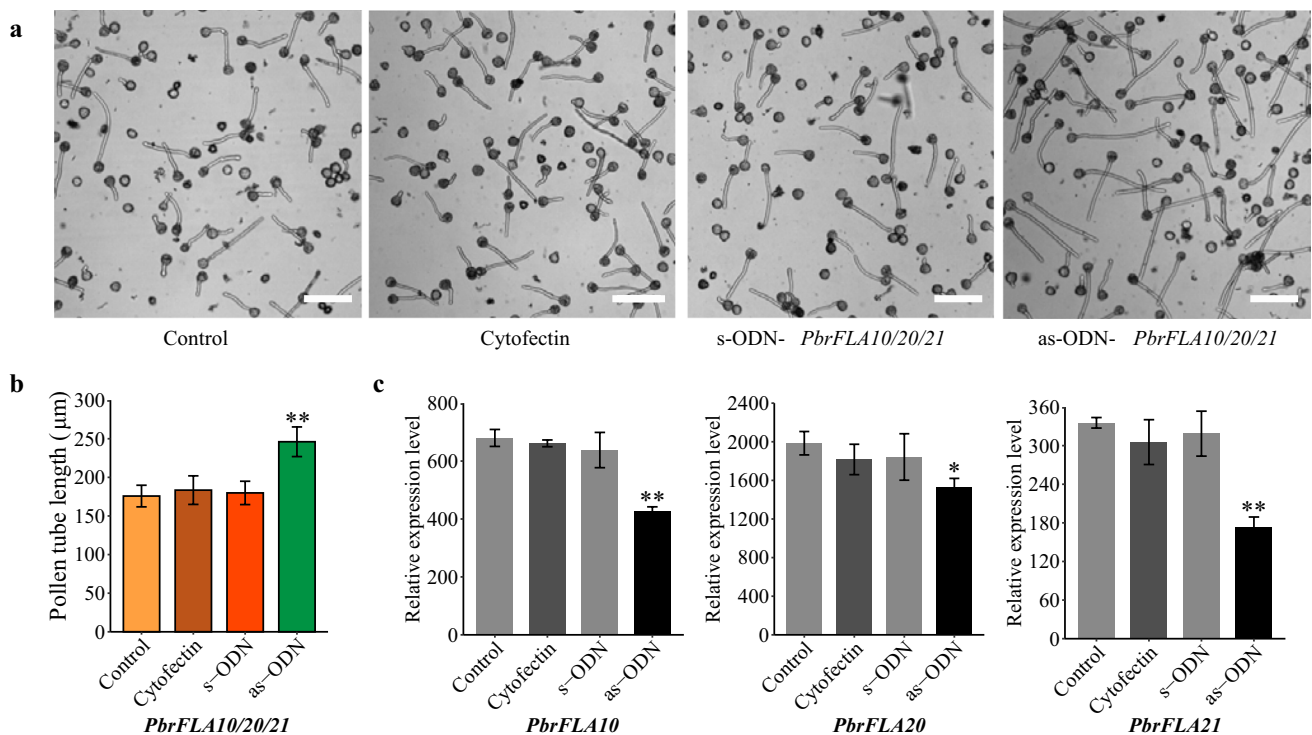


Fig. 8 Knockdown of the expression level of *PbrFLA10/20/21* caused the acceleration of pear pollen tube growth. **a** Pollen tube growth when treated with as-ODN and control groups (control, cytofectin and s-ODN). Bars equal 100 µm. **b** Treatment with *PbrFLA10/20/21* antisense oligonucleotides mixture (as-ODN) resulted in the acceleration of pollen tube growth. s-ODN was transfected

with sense oligonucleotides in the presence of cytofectin for the negative control. ** $p < 0.01$ and * $p < 0.05$ by Student's *t*-test. Data are the mean \pm s.e.m., $n = 3$ biological replicates. **c** Expression levels between control groups and as-ODN-mediated knockdown of *PbrFLA10/20/21* expression

of ancestral genes (Wang et al. 2012). DSD events happen through unpredictable and random patterns by mechanisms that remain unclear, generating two gene copies that are neither neighboring nor collinear (Ganko et al. 2007). In this study, although a high frequency of DSD-genes was detected, WGD and TRD events were also the crucial force of expansion for *FLAs* in the seven Rosaceae species. Especially, WGD duplication played an important role in the expansion of the *FLA* gene family in pear and apple. Based on previous reports in our lab, the genome of pear and apple has undergone two rounds of duplications: the recent WGD duplication occurred approximately 30–45 MYA ($Ks \sim 0.15\text{--}0.30$) in pear and apple, while the ancient WGD from γ triplication event occurred approximately ~ 140 MYA ($Ks \sim 1.5\text{--}1.8$) (Wu et al. 2013). Our results of Ks and 4DTV values of *FLA* paralogs showed that the *FLA* gene family had undergone the two WGD/segmental duplication events, suggesting that *FLA* gene family was of early origin and differentiated by adaptation to different types of environment in the evolution process in pear and apple. The Ka/Ks ratios of all duplicated *FLA* pairs were less than one (except three gene pairs in sweet cherry), which implied that the *FLA* is mainly undergoing purifying selection.

In general, the gene expression patterns can provide important clues to predict the gene function. Except for the function of *AtFLA3* and *MTR1* in reproductive organs and *AtFLA1* in root. *AtFLA8* and *AtFLA10*, two class III *FLAs*, are specifically expressed during seed development (Costa et al. 2019). Some *EgrFLAs* were also highly expressed in stems (MacMillan et al. 2015). In our study, qRT-PCR expression profiling was conducted to explore *PbrFLA* gene expression patterns in six different tissues. This result indicated that most *PbrFLA* transcripts were detected in transcriptional abundance in root and pollen, suggesting these genes might have important roles in tissues undergoing growth. Comparing other tissues, 10 *PbrFLAs* that most belonged to class III were differentially high expressed in the pollen, suggesting class III of *FLAs* might be related to the development of reproductive organs.

Previous study reported that *FLAs* containing single-FAS domain may have conserved roles in secondary cell wall biology and properties (MacMillan et al. 2015). In *Arabidopsis*, *AtFLA3*, which also contains a single-FAS domain and only one region of arabinogalactan addition sites (a class III *AtFLA*), was required for cellulose deposition in the intine layer of pollen grain (Li et al. 2010). The phylogenetic

Table 2 Features of FLA genes identified in pear

ID	ChrID	Start	End	CDS length (bp)	Amino acid (aa)	Genomic length (bp)	MW (kDa)	pI	GPI (YES/NO)	FAS domain
<i>PbrFLA1 (Pbr023359.1)</i>	Pbr02	929,709	931,596	1278	425	1887	43.08	6.61	Y	2
<i>PbrFLA2 (Pbr017840.1)</i>	Pbr03	7,949,606	7,951,395	1236	411	1789	43.93	6.37	Y	2
<i>PbrFLA3 (Pbr029013.1)</i>	Pbr03	16,459,713	16,460,561	849	282	848	29.54	8.79	N	1
<i>PbrFLA4 (Pbr029014.1)</i>	Pbr03	16,466,287	16,467,132	846	281	845	29.52	7.5	Y	1
<i>PbrFLA5 (Pbr013397.1)</i>	Pbr03	20,519,651	20,520,682	1032	343	1031	36.92	4.11	N	2
<i>PbrFLA6 (Pbr033944.1)</i>	Pbr03	24,681,076	24,681,819	744	247	743	27.41	5.54	N	1
<i>PbrFLA7 (Pbr032866.1)</i>	Pbr03	25,497,353	25,499,106	1266	421	1753	44.45	6.3	N	1
<i>PbrFLA8 (Pbr021135.1)</i>	Pbr04	6,617,645	6,618,370	726	241	725	25.07	6.51	Y	1
<i>PbrFLA9 (Pbr042015.1)</i>	Pbr05	23,313,537	23,314,376	840	279	839	29.51	6.12	N	1
<i>PbrFLA10 (Pbr021264.1)</i>	Pbr08	10,077,715	10,079,142	852	283	1427	28.65	6	Y	1
<i>PbrFLA11 (Pbr029779.1)</i>	Pbr08	11,023,604	11,024,886	804	267	1282	27.79	10.02	Y	1
<i>PbrFLA12 (Pbr005362.1)</i>	Pbr09	6,107,794	6,127,499	1335	444	19,705	48.25	5.54	N	1
<i>PbrFLA13 (Pbr005226.1)</i>	Pbr09	15,011,145	15,011,915	771	256	770	26.82	4.94	Y	1
<i>PbrFLA14 (Pbr025607.1)</i>	Pbr11	20,612,275	20,613,457	798	265	1182	27.79	6.95	Y	1
<i>PbrFLA15 (Pbr008332.1)</i>	Pbr12	8,899,219	8,899,959	741	246	740	25.33	6.07	Y	1
<i>PbrFLA16 (Pbr001318.1)</i>	Pbr12	18,175,587	18,176,936	1350	449	1349	49.21	6.68	N	2
<i>PbrFLA17 (Pbr028730.1)</i>	Pbr13	1,506,689	1,507,783	1095	364	1094	40.31	6.32	N	2
<i>PbrFLA18 (Pbr028740.1)</i>	Pbr13	1,555,411	1,556,715	1305	434	1304	47.15	5.69	N	2
<i>PbrFLA19 (Pbr010304.2)</i>	Pbr14	1,712,788	1,714,318	1395	464	1530	50.92	6.74	N	2
<i>PbrFLA20 (Pbr002661.1)</i>	Pbr15	1,129,204	1,130,455	789	262	1251	26.98	8.99	Y	1
<i>PbrFLA21 (Pbr002662.1)</i>	Pbr15	1,134,708	1,135,828	876	291	1120	30.65	9.49	Y	1
<i>PbrFLA22 (Pbr019945.1)</i>	Pbr15	5,833,381	5,834,034	654	217	653	23.82	6.62	N	1
<i>PbrFLA23 (Pbr010617.1)</i>	Pbr15	13,049,723	13,050,536	744	247	813	25.68	7.82	Y	1
<i>PbrFLA24 (Pbr020817.1)</i>	Pbr15	41,967,182	41,968,834	1269	422	1652	44.83	6.79	N	1
<i>PbrFLA25 (Pbr038002.2)</i>	Pbr16	14,229,262	14,229,948	687	228	686	26.37	9.79	N	1
<i>PbrFLA26 (Pbr022454.1)</i>	Pbr17	3,066,428	3,069,301	2874	957	2873	101.42	5.17	Y	3
<i>PbrFLA27 (Pbr022452.1)</i>	Pbr17	3,075,224	3,076,727	1338	445	1503	47.46	7.69	Y	2
<i>PbrFLA28 (Pbr008678.1)</i>	Pbr17	5,807,739	5,808,377	639	212	638	23.64	8.92	N	1
<i>PbrFLA29 (Pbr010280.1)</i>	Pbr17	8,475,925	8,478,303	1176	391	2378	41.3	6.51	Y	2
<i>PbrFLA30 (Pbr011341.1)</i>	Pbr17	21,703,603	21,704,603	588	195	1000	21.92	5.11	N	1
<i>PbrFLA31 (Pbr012855.1)</i>	scaffold197.0	483,526	486,522	2841	946	2996	100.48	5.17	Y	3
<i>PbrFLA32 (Pbr012857.1)</i>	scaffold197.0	492,770	494,344	1332	443	1574	47.34	7.73	Y	2
<i>PbrFLA33 (Pbr042430.1)</i>	scaffold982.0	1741	2514	774	257	773	28.52	5.54	N	1
<i>PbrFLA34 (Pbr042432.1)</i>	scaffold982.0	5277	6107	831	276	830	30.52	6.51	N	1

Table 2 (continued)

ID	ChrID	Start	End	CDS length (bp)	Amino acid (aa)	Genomic length (bp)	MW (kDa)	pI	GPI (YES/NO)	FAS domain
<i>PbrFLA35 (Pbr042433.1)</i>	scaffold982.0	9565	10,140	576	191	575	21.28	8.24	N	1
<i>PbrFLA36 (Pbr042446.1)</i>	scaffold982.0	128,961	129,734	774	257	773	28.57	6.3	N	1
<i>PbrFLA37 (Pbr042447.1)</i>	scaffold982.0	133,117	133,947	831	276	830	30.52	6.51	N	1
<i>PbrFLA38 (Pbr042449.1)</i>	scaffold982.0	136,710	137,483	774	257	773	28.52	5.54	N	1

ChrID chromosomal number, MW molecular weight, pI isoelectric point, GPI glycosylphosphatidylinositol

tree and conserved motifs showed that *PbrFLA10*, *PbrFLA20* and *PbrFLA21* were grouped into the same subclass with *AtFLA3* (Fig. 1), contained highly conserved domains (Fig. 5) and shared similar gene structure and protein lengths (Table 2). Thus, we explored their expression patterns in differently stages of pollen tubes and in styles with different states of pollination. *PbrFLA10*, *PbrFLA20* and *PbrFLA21* were abundantly expressed in the self-pollinated styles (SP) and stop growing pollen tubes (STP) (Fig. 7). Repression of *PbrFLA10/20/21* resulted in the acceleration of pear pollen tube growth. Based on previous researches that FLA function exists a genetic redundancy (MacMillan et al. 2010). We speculated that *PbrFLA10/20/21* might be functional redundancy and play roles in pollen tube growth by affecting cellulose deposition. The high expression of *PbrFLA10/20/21* seems to be involved in fundamental self-incompatibility response mechanisms. More detailed analyses are required to confirm what types of metabolism are affected and if full loss of function of the *PbrFLA10/20/21* locus will lead to a more severe defect on pollen tube function.

MYB, C2H2 and EIL transcription factors play important roles in reproductive tissues development. In *Arabidopsis thaliana*, *AtMYB26* (Steiner-Lange et al. 2003), *AtMYB103* (Higginson et al. 2003), *AtMYB32* (Preston et al. 2003) and *AtMYB80* (Phan et al. 2011) are reported to be involved in regulation of anther or pollen development. In apple, *MdMYB39L* is a key role in sorbitol-modulated stamen development and pollen tube growth by regulating its downstream target genes (Meng et al. 2018). In Chinese cabbage, *BcMF20*, a C2H2-type zinc-finger transcription factor, may transcriptionally regulate tapetal-specific genes involved in pollen wall development, which then regulate pollen wall material secretion and subsequent wall development and tapetal programmed cell death (Han et al. 2018). In tobacco, *TEIL (Tobacco EIN3-Like)* gene is a tobacco homologue of *Arabidopsis Ethylene Insensitive 3 (EIN3)* and participates in the regulation of flower shape development (Hibi et al. 2007). Combined with the connectivity values of the gene co-expression analysis (Online Resource 3, Online Resource 14), we predicted that *Pbr005306.1* (C2H2), *Pbr000646.1* (EIL) and *Pbr001365.1* (MYB_related) were the most likely to regulate *PbrFLA10*, *PbrFLA20* and *PbrFLA21*, respectively.

Conclusion

In conclusion, this work presented the phylogenetic relationships, evolutionary history, and expression patterns of *PbrFLA* family genes. Functional validation indicated that *PbrFLA10/20/21* could affect the growth of pear pollen tubes, suggesting their important roles in the growth of the pear pollen tube. This work was able to help us

to understand the evolution of the *FLA* gene family and provided a foundation for future molecular biology on *FLA* genes during pollen tube growth.

Information on Electronic Supplementary Material

Online Resource 1. Multiple sequence alignment of the fasciclin domains of *PbrFLAs*.

Online Resource 2. RNA-seq data analysis of *PbrFLAs* in different developmental stages of pear pollen tube and different pollinated style.

Online Resource 3. Co-expression relationships between *PbrFLA10/20/21* genes and transcription factor genes.

Online Resource 4. List of genome information of seven Rosaceae species.

Online Resource 5. List of the number of genes of each subfamily in *Arabidopsis thaliana* and seven Rosaceae species.

Online Resource 6. List of gene features of all identified *FLA* genes in seven Rosaceae.

Online Resource 7. List of motif sequences identified by MEME tools in pear *FLA*.

Online Resource 8. List of collinearity relationship and Ka/Ks analysis among *FLA* genes in the same species.

Online Resource 9. List of orthologous pairs of *FLA* genes between two Rosaceae species.

Online Resource 10. List of the relative expression levels of *PbrFLA* genes expression in six tissues.

Online Resource 11. List of the TPM value of *PbrFLA* genes expression in different development of pollen.

Online Resource 12. List of the TPM value of *PbrFLA* genes expression in different pollinated style.

Online Resource 13. List of primers for qRT-PCR of *PbrFLA* candidate genes and antisense oligo deoxynucleotide experiment in pear.

Online Resource 14. List of gene co-expression network analysis of *PbrFLA* genes in pear.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00606-021-01769-w>.

Acknowledgements This work was funded by the National Key Research and Development Program of China (Grant No. 2020YFE0202900), National Natural Science Foundation of China (31772256, 31772276), Fundamental Research Funds for the Central Universities (JCQY201903, KYLH202002), Jiangsu Province Science and Technology Support Program (BE2018389), the Priority Academic Program Development of Jiangsu Higher Education Institutions. Bioinformatic analysis was supported by the Bioinformatics Center of Nanjing Agricultural University. Usage of microscope was guided by Yuehua Ma.

Author contributions Peng Wang conceived and designed the experiments, Xiaoqiang Li and Mengyu Cheng performed the experiments, and Xiaoqiang Li wrote the manuscript. The manuscript was revised and confirmed by Chao Tang, Xiaoxuan Zu, Kaijie Qi, Shaoling Zhang and Juyou Wu. All those involved have read and confirmed the paper.

Availability of data and materials All coding and protein sequences in this study can be found in the Genome Database for Rosaceae (<https://www.rosaceae.org/>) and Pear Genome Project (<http://peargenome.njau.edu.cn/>).

Declarations

Competing interests The authors declare that they have no competing interests.

References

- Blanc G, Wolfe KH (2004) Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *PL Cell* 16:1667–1678. <https://doi.org/10.1105/tpc.021345>
- Castilleux R, Plancot B, Ropitiaux M, Carreras A, Leprince J, Boulogne I, Follet-Gueye ML, Popper ZA, Driouich A, Vicre M (2018) Cell wall extensins in root-microbe interactions and root secretions. *J Exp Bot* 69:4235–4247. <https://doi.org/10.1093/jxb/ery238>
- Chen J, Li X, Wang D, Li L, Zhou H, Liu Z, Wu J, Wang P, Jiang X, Fabrice MR, Zhang S, Wu J (2015) Identification and testing of reference genes for gene expression analysis in pollen of *Pyrus bretschneideri*. *Sci Hort* 190:43–56. <https://doi.org/10.1016/j.scienta.2015.04.010>
- Chen J, Wang P, de Graaf B, Zhang H, Jiao H, Tang C, Zhang S, Wu J (2018a) Phosphatidic acid counteracts S-RNase signaling in pollen by stabilizing the actin cytoskeleton. *PL Cell* 30:1023–1039. <https://doi.org/10.1105/tpc.18.00021>
- Chen S, Zhou Y, Chen Y, Gu J (2018b) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Pl* 13:1194–1202. <https://doi.org/10.1016/j.molp.2020.06.009>
- Coimbra S, Costa M, Jones B, Mendes MA, Pereira LG (2009) Pollen grain development is compromised in *Arabidopsis agp6 agp11* null mutants. *J Exp Bot* 60:3133–3142. <https://doi.org/10.1093/jxb/erp148>
- Costa M, Pereira AM, Pinto SC, Silva J, Pereira LG, Coimbra S (2019) In silico and expression analyses of fasciclin-like arabinogalactan proteins reveal functional conservation during embryo and seed development. *PL Reprod* 32:353–370. <https://doi.org/10.1007/s00497-019-00376-7>
- Cruz-Garcia F, Nathan Hancock C, Kim D, McClure B (2005) Styler glycoproteins bind to S-RNase in vitro. *PL J* 42:295–304. <https://doi.org/10.1111/j.1365-313X.2005.02375.x>
- Eddy SR (2011) Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>
- Eisenhaber B, Bork P, Eisenhaber F (1999) Prediction of potential GPI-modification sites in proprotein sequences. *J Molec Biol* 292:741–758. <https://doi.org/10.1006/jmbi.1999.3069>
- Faik A, Abouzouhair J, Sarhan F (2006) Putative fasciclin-like arabinogalactan-proteins (FLA) in wheat (*Triticum aestivum*) and rice (*Oryza sativa*): identification and bioinformatic analyses. *Molec Genet Genomics* 276:478–494. <https://doi.org/10.1007/s00438-006-0159-z>
- Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer EL, Eddy SR, Bateman A (2010) The Pfam protein families database. *Nucl Acids Res* 38:D211–222. <https://doi.org/10.1093/nar/gkp985>
- Guerriero G, Mangeot-Peter L, Legay S, Behr M, Lutts S, Siddiqui KS, Hausman JF (2017) Identification of fasciclin-like arabinogalactan proteins in textile hemp (*Cannabis sativa* L.): in silico analyses and gene expression patterns in different tissues. *BMC Genom* 18:741. <https://doi.org/10.1186/s12864-017-3970-5>

- Han YY, Zhou HY, Xu LA, Liu XY, Fan SX, Cao JS (2018) The zinc-finger transcription factor *BcMF20* and its orthologs in Cruciferae which are required for pollen development. *Biochem Biophys Res Commun* 503:998–1003. <https://doi.org/10.1016/j.bbrc.2018.06.108>
- Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Molec Evol* 22:160–174. <https://doi.org/10.1007/BF02101694>
- He J, Zhao H, Cheng Z, Ke Y, Liu J, Ma H (2019) Evolution analysis of the fasciclin-like arabinogalactan proteins in plants shows variable fasciclin-AGP domain constitutions. *Int J Molec Sci* 20:1945. <https://doi.org/10.3390/ijms20081945>
- Hibi T, Kosugi S, Iwai T, Kawata M, Seo S, Mitsuhashi I, Ohashi Y (2007) Involvement of *EIN3* homologues in basic PR gene expression and flower development in tobacco plants. *J Exp Bot* 58:3671–3678. <https://doi.org/10.1093/jxb/erm216>
- Higginson T, Li SF, Parish RW (2003) AtMYB103 regulates tapetum and trichome development in *Arabidopsis thaliana* . *Pl J* 35:177–192. <https://doi.org/10.1046/j.1365-313x.2003.01791.x>
- Huang GQ, Xu WL, Gong SY, Li B, Wang XL, Xu D, Li XB (2008) Characterization of 19 novel cotton *FLA* genes and their expression profiling in fiber development and in response to phytohormones and salt stress. *Physiol Pl* 134:348–359. <https://doi.org/10.1111/j.1399-3054.2008.01139.x>
- Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P (2010) Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* 5:12776. <https://doi.org/10.1371/journal.pone.0012776>
- Jiao H, Liu X, Sun S, Wang P, Qiao X, Li J, Tang C, Wu J, Zhang S, Tao S (2018) The unique evolutionary pattern of the hydroxyproline-rich glycoproteins superfamily in Chinese white pear (*Pyrus bretschneideri*). *BMC Pl Biol* 18:36. <https://doi.org/10.1186/s12870-018-1252-2>
- Johnson KL, Jones BJ, Bacic A, Schultz CJ (2003) The fasciclin-like arabinogalactan proteins of *Arabidopsis* . A multigene family of putative cell adhesion molecules. *Pl Physiol* 133:1911–1925. <https://doi.org/10.1104/pp.103.031237>
- Johnson KL, Kibble NA, Bacic A, Schultz CJ (2011) A fasciclin-like arabinogalactan-protein (FLA) mutant of *Arabidopsis thaliana* , *fla1* , shows defects in shoot regeneration. *PLoS ONE* 6:e25154. <https://doi.org/10.1371/journal.pone.0025154>
- Jun L, Xiaoming W (2012) Genome-wide identification, classification and expression analysis of genes encoding putative fasciclin-like arabinogalactan proteins in Chinese cabbage (*Brassica rapa* L.). *Molec Biol Reprod* 39:10541–10555. <https://doi.org/10.1007/s11033-012-1940-1>
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Meth* 14:587–589. <https://doi.org/10.1038/nmeth.4285>
- Kim JE, Kim SJ, Lee BH, Park RW, Kim KS, Kim IS (2000) Identification of motifs for cell adhesion within the repeated domains of transforming growth factor-beta-induced gene, betaig-h3. *J Biol Chem* 275:30907–30915. <https://doi.org/10.1074/jbc.M002752200>
- Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature Meth* 12:357–360. <https://doi.org/10.1038/nmeth.3317>
- Li W, Liu B, Yu L, Feng D, Wang H, Wang J (2009) Phylogenetic analysis, structural evolution and functional divergence of the 12-oxo-phytodienoate acid reductase gene family in plants. *BMC Evol Biol* 9:90. <https://doi.org/10.1186/1471-2148-9-90>
- Li J, Yu M, Geng LL, Zhao J (2010) The fasciclin-like arabinogalactan protein gene, *FLA3* , is involved in microspore development of *Arabidopsis* . *Pl J* 64:482–497. <https://doi.org/10.1111/j.1365-313X.2010.04344.x>
- Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930. <https://doi.org/10.1093/bioinformatics/btt656>
- Liu F, Xu Y, Jiang H, Jiang C, Du Y, Gong C, Wang W, Zhu S, Han G, Cheng B (2016) Systematic identification, evolution and expression analysis of the *Zea mays PHT1* gene family reveals several new members involved in root colonization by arbuscular mycorrhizal fungi. *Int J Molec Sci* 17:930. <https://doi.org/10.3390/ijms17060930>
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)^{-Delta Delta C} method. *Methods* 25:402–408. <https://doi.org/10.1006/meth.2001.1262>
- Ma H, Zhao J (2010) Genome-wide identification, classification, and expression analysis of the arabinogalactan protein gene family in rice (*Oryza sativa* L.). *J Exp Bot* 61:2647–2668. <https://doi.org/10.1093/jxb/erq104>
- MacMillan CP, Mansfield SD, Stachurski ZH, Evans R, Southerton SG (2010) Fasciclin-like arabinogalactan proteins: specialization for stem biomechanics and cell wall architecture in *Arabidopsis* and *Eucalyptus* . *Pl J* 62:689–703. <https://doi.org/10.1111/j.1365-313X.2010.04181.x>
- MacMillan CP, Taylor L, Bi Y, Southerton SG, Evans R, Spokevicius A (2015) The fasciclin-like arabinogalactan protein family of *Eucalyptus grandis* contains members that impact wood biology and biomechanics. *New Phytol* 206:1314–1327. <https://doi.org/10.1111/nph.13320>
- Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21:3448–3449. <https://doi.org/10.1093/bioinformatics/bti551>
- Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH (2015) CDD: NCBI's conserved domain database. *Nucl Acids Res* 43:D222–226. <https://doi.org/10.1093/nar/gku1221>
- Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Geer LY, Bryant SH (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucl Acids Res* 45:D200–D203. <https://doi.org/10.1093/nar/gkw1129>
- Meng D, He M, Bai Y, Xu H, Dandekar AM, Fei Z, Cheng L (2018) Decreased sorbitol synthesis leads to abnormal stamen development and reduced pollen tube growth via an MYB transcription factor, *MdMYB39L* , in apple (*Malus domestica*). *New Phytol* 217:641–656. <https://doi.org/10.1111/nph.14824>
- Mizuta Y, Higashiyama T (2014) Antisense gene inhibition by phosphorothioate antisense oligonucleotide in *Arabidopsis* pollen tubes. *Pl J* 78:516–526. <https://doi.org/10.1111/tpj.12461>
- Moutinho A, Camacho L, Haley A, Pais MS, Trewavas A, Malhó R (2001) Antisense perturbation of protein function in living pollen tubes. *Sexual Pl Reprod* 14:101–104. <https://doi.org/10.1007/s004970100086>
- Panchy N, Lehti-Shiu M, Shiu SH (2016) Evolution of gene duplication in plants. *Pl Physiol* 171:2294–2316. <https://doi.org/10.1104/pp.16.00523>
- Pereira AM, Lopes AL, Coimbra S (2016) Arabinogalactan proteins as interactors along the crosstalk between the pollen tube and the female tissues. *Frontiers Pl Sci* 7:1895. <https://doi.org/10.3389/fpls.2016.01895>
- Phan HA, Iacuone S, Li SF, Parish RW (2011) The MYB80 transcription factor is required for pollen development and the regulation

- of tapetal programmed cell death in *Arabidopsis thaliana*. *Pl Cell* 23:2209–2224. <https://doi.org/10.1105/tpc.110.082651>
- Preston J, Wheeler J, Heazlewood J, Li SF, Parish RW (2004) AtMYB32 is required for normal pollen development in *Arabidopsis thaliana*. *Pl J* 40:979–995. <https://doi.org/10.1111/j.1365-313X.2004.02280.x>
- Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, Zhang S, Paterson AH (2019) Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biol* 20:38. <https://doi.org/10.1186/s13059-019-1650-2>
- Ramirez Gonzalez RH, Borrill P, Lang D, Harrington SA, Brinton J, Venturini L, Davey M, Jacobs J, van Ex F, Pasha A, Khedikar Y, Robinson SJ, Cory AT, Florio T, Concia L, Juery C, Schoonbeek H, Steuernagel B, Xiang D, Ridout CJ, Chalhoub B, Mayer KFX, Benhamed M, Latrasse D, Bendahmane A, International Wheat Genome Sequencing C, Wulff BBH, Appels R, Tiwari V, Datla R, Choulet F, Pozniak CJ, Provart NJ, Sharpe AG, Paux E, Spannagl M, Brautigam A, Uauy C (2018) The transcriptional landscape of polyploid wheat. *Science* 361:6089. <https://doi.org/10.1126/science.aar6089>
- Schultz CJ, Rumsewicz MP, Johnson KL, Jones BJ, Gaspar YM, Bacic A (2002) Using genomic resources to guide research directions. The arabinogalactan protein gene family as a test case. *Pl Physiol* 129:1448–1463. <https://doi.org/10.1104/pp.003459>
- Sede AR, Wengier DL, Borassi C, Estevez JM, Muschietti JP (2020) Imaging and analysis of the content of callose, pectin, and cellulose in the cell wall of *Arabidopsis* pollen tubes grown in vitro. *Meth Molec Biol* 2160:233–242. https://doi.org/10.1007/978-1-0716-0672-8_17
- Shi D, Tang C, Wang R, Gu C, Wu X, Hu S, Jiao J, Zhang S (2017) Transcriptome and phytohormone analysis reveals a comprehensive phytohormone and pathogen defence response in pear self-/cross-pollination. *Pl Cell Rep* 36:1785–1799. <https://doi.org/10.1007/s00299-017-2194-0>
- Showalter AM, Keppler BD, Liu X, Lichtenberg J, Welch LR (2016) Bioinformatic identification and analysis of hydroxyproline-rich glycoproteins in populus trichocarpa. *BMC Pl Biol* 16:229. <https://doi.org/10.1186/s12870-016-0912-3>
- Steiner-Lange S, Unte US, Eckstein L, Yang C, Wilson ZA, Schmelzer E, Dekker K, Saedler H (2003) Disruption of *Arabidopsis thaliana* MYB26 results in male sterility due to non-dehiscent anthers. *Pl J* 34:519–528. <https://doi.org/10.1046/j.1365-313x.2003.01745.x>
- Su S, Higashiyama T (2018) Arabinogalactan proteins and their sugar chains: functions in plant reproduction, research methods, and biosynthesis. *Pl Reprod* 31:67–75. <https://doi.org/10.1007/s00497-018-0329-2>
- Takahashi D, Kawamura Y, Uemura M (2016) Cold acclimation is accompanied by complex responses of glycosylphosphatidylinositol (GPI)-anchored proteins in *Arabidopsis*. *J Exp Bot* 67:5203–5215. <https://doi.org/10.1093/jxb/erw279>
- Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucl Acids Res* 44:W232–W235. <https://doi.org/10.1093/nar/gkw256>
- Wang D, Zhang Y, Zhang Z, Zhu J, Yu J (2010) KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genom Proteom Bioinf* 8:77–80. [https://doi.org/10.1016/S1672-0229\(10\)60008-3](https://doi.org/10.1016/S1672-0229(10)60008-3)
- Wang Y, Wang X, Paterson AH (2012) Genome and gene duplications and gene expression divergence: a view from plants. *Ann New York Acad Sci* 1256:1–14. <https://doi.org/10.1111/j.1749-6632.2011.06384.x>
- Wang H, Jiang C, Wang C, Yang Y, Yang L, Gao X, Zhang H (2015) 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* 66:1291–1302. <https://doi.org/10.1093/jxb/eru479>
- Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>
- Wu J, Wang Z, Shi Z, Zhang S, Ming R, Zhu S, Khan MA, Tao S, Korban SS, Wang H, Chen NJ, Nishio T, Xu X, Cong L, Qi K, Huang X, Wang Y, Zhao X, Wu J, Deng C, Gou C, Zhou W, Yin H, Qin G, Sha Y, Tao Y, Chen H, Yang Y, Song Y, Zhan D, Wang J, Li L, Dai M, Gu C, Wang Y, Shi D, Wang X, Zhang H, Zeng L, Zheng D, Wang C, Chen M, Wang G, Xie L, Sovero V, Sha S, Huang W, Zhang S, Zhang M, Sun J, Xu L, Li Y, Liu X, Li Q, Shen J, Wang J, Paull RE, Bennetzen JL, Wang J, Zhang S (2013) The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res* 23:396–408. <https://doi.org/10.1101/gr.144311.112>
- Wu X, Lai Y, Lv L, Ji M, Han K, Yan D, Lu Y, Peng J, Rao S, Yan F, Zheng H, Chen J (2020) Fasciclin-like arabinogalactan gene family in *Nicotiana benthamiana*: genome-wide identification, classification and expression in response to pathogens. *BMC Pl Biol* 20:305. <https://doi.org/10.1186/s12870-020-02501-5>
- Xue H, Veit C, Abas L, Tryfona T, Maresch D, Ricardi MM, Estevez JM, Strasser R, Seifert GJ (2017) *Arabidopsis thaliana* *FLA4* functions as a glycan-stabilized soluble factor via its carboxy-proximal Fasciclin 1 domain. *Pl J* 91:613–630. <https://doi.org/10.1111/tpj.13591>
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Molec Biol Evol* 24:1586–1591. <https://doi.org/10.1093/molbev/msm088>
- Zhang Z, Xiao J, Wu J, Zhang H, Liu G, Wang X, Dai L (2012b) ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. *Biochem Biophys Res Commun* 419:779–781. <https://doi.org/10.1016/j.bbrc.2012.02.101>
- Zhang H, Gao S, Lercher MJ, Hu S, Chen W-H (2012a) EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucl Acids Res* 40:W569–W572. <https://doi.org/10.1093/nar/gks576>
- Zhang Z, Xin W, Wang S, Zhang X, Dai H, Sun R, Frazier T, Zhang B, Wang Q (2015) Xylem sap in cotton contains proteins that contribute to environmental stress response and cell wall development. *Funct Integr Genomics* 15:17–26. <https://doi.org/10.1007/s10142-014-0395-y>
- Zhou H, Yin H, Chen J, Liu X, Gao Y, Wu J, Zhang S (2016) Gene-expression profile of developing pollen tube of *Pyrus bretschneideri*. *Gene Expr Patterns* 20:11–21. <https://doi.org/10.1016/j.gep.2015.10.004>