

PATTERNS OF MOLECULAR EVOLUTION AND EXPRESSION DIVERGENCE OF THE POLYGALACTURONASE GENE FAMILY IN TUNG TREE (*VERNICIA FORDII* HEMSL.)

HAN, M. Y.^{2,3} – DONG, X.³ – CAO, J.³ – ZHU, H. G.¹ – MI, X. Q.⁴ – LI, W. Y.^{1,3*}

¹*College of Biology and Agricultural Resources, Huanggang Normal University, Huanggang, Hubei 438000, China*

²*Hainan Academy of Forestry (Hainan Academy of Mangrove), Haikou, Hainan, China*

³*Key Laboratory of Cultivation and Protection for Non-Wood Forest Trees of the Ministry of Education and Key Laboratory of Non-Wood Forest Products of the Forestry Ministry, Central South University of Forestry and Technology, Changsha, Hunan 410004, China*

⁴*Ecology Research and Experiment Station of Forestry Administration in Xiangxi Tujia and Miao Autonomous Prefecture, Yongshun County, Hunan 416700, China*

*Corresponding author
e-mail: wenyinli@hgnu.edu.cn

(Received 13th Apr 2024; accepted 30th Aug 2024)

Abstract. Polygalacturonase (PG; EC 3.2.1.15) genes are a major group of pectin-hydrolyzing enzymes responsible for pectin degeneration and play important roles in pollen development and male fertility. Studying the PG gene family could shed light on molecular mechanisms by which PGs regulate pollen development in tung tree (*Vernicia fordii*). A total of 35 tung PG genes (*VfPGs*) were obtained from tung genome in this study. *VfPGs* were clustered into 2 main classes which contained 6 clades by phylogenetic analysis. All Class I *VfPGs* comprised four typical conserved residues while residue III was absent in Class II *VfPGs*. Evolutionary analyses displayed that Class III PG genes that were present in the most recent common ancestors of land plants could be lost in Euphorbiaceae after their divergence. Whole genome triplication was one of the major forces of the expansion of tung PG family and duplicated *VfPGs* suffered negative selection during evolution. A total of 107 cis-regulatory elements related to phytohormone responsiveness were identified in the promoters of *VfPGs*, suggesting phytohormones may play important roles in plant development in tung trees by regulating *VfPG* activities. Transcriptome-scale gene expression analysis of *VfPGs* in various tissues displayed that *VfPGs* were expressed more in flowers than in other tissues. Several *VfPGs* were predicted to be highly correlated with anther development. This study provides clues for future works aimed at elucidating molecular mechanisms underlying PG-regulated pollen development in tung tree.

Keywords: gene loss, phytohormone regulation, flower development, pollen development

Introduction

Pectin is a type of water-soluble fiber that is found in the primary cell walls and intracellular cells of certain plants, such as the growing pollen tube apex, gel-like structure at the seed surface, intracellular layer of plant cells in apples, oranges, peaches, and other fruits. It plays a crucial role in cell-wall integrity, cell adhesion, and environmental sensing (Parre and Geitmann, 2005; Vogel, 2008; Wang et al., 2023; Xin et al., 2023). The dismantling and modification of the pectin network, which is systematically disassembled during the abscission of plant organs, fruit ripening, and anther dehiscence, is performed by a wide range of hydrolytic enzymes including

polygalacturonase (PG; EC 3.2.1.15), pectin methylesterase (PME; EC 3.1.1.11) and pectin lyase (EC 4.2.2.2) (Hadfield and Bennett, 1998; Bosch and Hepler, 2005; Ogawa et al., 2009; Duan et al., 2016).

PG genes belong to the largest hydrolase family and are a major group of pectin-hydrolyzing enzymes. By comparison of amino acid sequences, regions comprising the residues 178_NTD, 201_DD, 22_GHG, and 256_RIK (numbered throughout the open reading frame of *Aspergillus niger*) were found to be strictly conserved in PGs of bacteria, fungi, and plants (Markovic and Janecek, 2001). PGs are found to be critical for pollen development. For example, *QUARTET 2* (*QRT2*) and *QRT3* are endo-PGs in the *Arabidopsis thaliana*, and are essential for pollen grain separation and pectin wall degradation, respectively (Rhee et al., 2003; Ogawa et al., 2009). In cotton, *NO SPINE POLLEN* (*GhNSP*), encoding a PG protein homologous to AtQRT3, is mainly expressed in the tapetum cells and plays an important role in pollen nexine and spine development (Wu et al., 2022). *Restorer of reversible male sterile-2*, which is an allele of AtQRT3, was related to the degeneration of the tetrad pectin wall (Shi et al., 2021). *Brassica campestris Male Fertility 2* (*BcMF2*) is a new PG that was found to be expressed specifically in tapetum and pollen after the tetrad stage of anther development in Chinese cabbage is associated with intine development (Huang et al., 2009).

The tung tree (*Vernicia fordii* Hemsl.), belonging to the family Euphorbiaceae, is an economical plant used as a source of industrial oils. The tung oil extracted from tung seeds is one of the best quick-drying oils and is widely utilized in producing environmentally friendly waterproof finish, advanced ink, and insulation materials. The tung tree is a monoecious species with an average female-to-male flower ratio of 1:27 which may be one of the major causes of low fruit yields (Mao et al., 2017; Liu et al., 2019). Abnormal microspores were frequently observed from the meiosis telophase II stage to the two-celled pollen stage, showing irregular binucleate tetrad microspores and curved pollen grains (Li et al., 2020). Tung PG genes (*VfPGs*) should play important roles in pollen development and are critical for the fertility of pollen grains. However, little is known about the molecular mechanisms of PGs in regulating pollen development in tung tree. In this study, we conducted genome-wide identification of *VfPGs* using multiple bioinformatics methods. The gene structure, protein physicochemical properties, molecular evolution, cis-regulatory elements, and gene expression profiles of *VfPGs* were systematically analyzed, providing a comprehensive characterization of the composition, expansion, and expression divergence of *VfPGs*. Our study provides clues for future works aimed at elucidating molecular mechanisms underlying PG-regulated pollen development in tung tree.

Materials and methods

Genome-wide identification of PGs genes in the tung genome

To identify PGs genes in the tung tree, the chromosome-level tung genome v1.0 (Zhang et al., 2019) was searched by *Arabidopsis* PGs obtained from Kim et al. (2006) using NCBI BLASTP with parameters of E-value < 1e⁻²⁰ and max_target_seqs = 5. The protein sequences (PEP) of *Arabidopsis* PG gene family members were extracted from the new Araprot11 annotation (Cheng et al., 2017). The protein sequences that contained the conserved GH28 domains were scanned and predicted by HMMER

software (<http://hmmer.org/>) with GH28 profile hidden Markov model (HMM) (PF00295.19). Primitively, a total of 37 PG candidates were identified in the tung genome from the above methods. The coding sequences (CDS) of these candidates were revised using the potential transcripts identified in tung floral transcriptomes conducted by Liu et al. (2019). The revised CDS were translated into PEP by SeqKit (Shen et al., 2016). Then the presence of GH28 domains, which were found in all plant PG proteins (Kim et al., 2006), was confirmed in all VfPG candidates through the online database software of SMART (Letunic et al., 2021) and InterPro (Paysan-Lafosse et al., 2023). According to the previous study, the regions comprising the residues of NTD, DD, GHG, and RIK are conserved in all PGs (Markovic and Janecek, 2001). So, the VfPG candidates that contained at least 2 of the 4 conserved residues were kept, and 35 VfPGs were finally identified in this study.

Protein sequence analysis and cis-regulatory elements analysis of VfPG promoters

The physical and chemical parameters of VfPGs, including molecular weight (MW), isoelectric point (PI), and the number of amino acids, were calculated using online tool ExPasy (Duvaud et al., 2021). Protein subcellular localization of VfPG proteins were predicted by web-server BUSCA (<http://busca.biocomp.unibo.it/>). The possible tertiary structures of VfPGs were predicted by Phyre2 web portal (Kelley et al., 2015). The 2 Kb sequences upstream of the start codon ATG site of VfPGs were considered putative promoters. cis-Regulatory elements in these putative promoters were predicted by PlantCARE website (Lescot et al., 2002). The exon-intron structures and GH28 domains of VfPGs were visualized by using TBtools (Chen et al., 2020).

Phylogenetic and molecular evolution analyses

PG functional domain polypeptide sequences of self-identified (*V. fordii*, *Jatropha curcas*, *Manihot esculenta*, *Ricinus communis*, and *Hevea brasiliensis*) (Table S1, S2) and published (*A. thaliana*, *Brassica oleracea*, *Oryza sativa*, *Populus trichocarpa*, *Physcomitrella patens*, *Selaginella moellendorffii* and *Zea mays*) (Yang et al., 2013; Lyu et al., 2020; Lu et al., 2021) PGs were identified by Hmmer software with GH28 HMM. The domains of protein sequences were aligned using MAFFT (Katoh and Standley, 2013). The phylogenetic relationships of PGs in the tung tree, between the tung tree and *Arabidopsis*, among five species of Euphorbiaceae, and among 8 land plant species, were reconstructed using maximum likelihood by IQ-TREE (Nguyen et al., 2015) with optimal amino acid substitution models of WAG + I + G4, LG + R5, JTT + R6 and LG + R8, respectively. All trees were estimated using 1 000 bootstraps with the parameter of -bb 1 000. The online tool of iTOL (Letunic and Bork, 2021) was used for the display of PG trees. Duplicate gene pairs of VfPGs were characterized by MCscanX (Wang et al., 2012). The K_a , K_s and ω values of VfPG duplicate gene pairs were calculated using KaKs_Calculator (Zhang, 2022). The time of duplication events of VfPGs was calculated according to the formula $T = K_s/2r$ with the r value of 6.5×10^{-9} synonymous substitutions per site per year for eudicots. The collinearity and chromosome location of the VfPGs were visualized by using TBtools (Chen et al., 2020).

Gene expression and protein-protein interaction network analyses

FPKM values of VfPGs in various tissues, including the roots, stems, leaves, flowers, and endosperms, were obtained from the tung tree electronic fluorescent pictographic

(eFP) browser (http://bar.utoronto.ca/efp_tung_tree/cgi-bin/efpWeb.cgi). Gene expression patterns of VfPGs in tung flowers were analyzed using OmicShare tools (<https://www.omicshare.com/tools>). The protein-protein interaction networks among VfPGs were predicted and displayed by STRING database v12.0 (<https://string-db.org/>) and Cytoscape v3.10 (Otasek et al., 2019), respectively.

Validation of qRT-PCR

Total RNA of male and female tung flowers at 30, 20, 10 and 1 day before flowering (DBF) was obtained from the Lin group (Key Lab of Non-wood Forest Products of State Forestry Administration, College of Forestry, Central South University of Forestry and Technology, Changsha, China). Each sample has three biological replicates and the experiment was repeated in triplicate. Genomic DNA in the RNA solution was removed and then the first strand of cDNA was prepared using HiScript® II Q RT SuperMix for qPCR (+gDNA wiper) (Vazyme, China). The gene expression patterns of several VfPGs were validated by qRT-PCR which was conducted with SYBR Green Premix Pro Taq HS qPCR Kit (Accurate Biology, China) and performed on a CFX96™ Real-Time System (Bio-Rad, USA). The VfEF1 α was selected as the reference gene (Han et al., 2012).

Results

Genome-wide identification and physicochemical properties of PG proteins in the tung tree

A total of 35 full-length genes encoding putative PG proteins were obtained from the tung genome based on a combined analysis of the similarity of sequences (BLASTP) and the presence of conserved GH28 domain (Table S1). Four tung-specific conserved amino acid residues, including (I) SPNTDGI, (II) GDDCI, (III) CGPGHGISIGSLG, and (IV) RIKTW, were identified by multiple sequence alignment analysis of consensus sequences (Fig. 1A, C). The tung-specific amino acid residues comprised residues of NTD, DD, GHG, and RIK (Fig. 1A and B) which are conserved in all plant PGs (Markovic and Janecek, 2001). All putative VfPGs comprised the conserved residues I, II, and IV, but the conserved residues III were found in 74% (26 out of 35) VfPGs (Fig. 1B).

The phylogenetic relationship of VfPGs based on the GH28 domain sequences was reconstructed (Fig. 1D). The phylogenetic tree showed that VfPGs were clustered into 2 main classes which were termed Class I and II in this study with 100% bootstrap support. VfPGs can be further grouped into 6 clades which were named Clade A to Clade F based on the previous study (Park et al., 2010). Approximately 77% of the VfPGs (27 out of 35) were classed into Class I, and the rest PGs (8 out of 35) belonged to Class II. Intriguingly, the protein sequences of VfPGs belonged to Class I and comprised residues I to IV, while residue III was absent in all Class II/Clade E VfPG sequences (Fig. 1A, D). Pairwise comparison of full-length PG protein sequences showed that pairwise identity scores of all Class I PGs ranged from 29.53% to 95.56%, pairwise identity scores of all Class II PGs ranged from 45.39% to 98.74%, and pairwise identity scores between Class I and Class II PG genes ranged from 21.76% to 45.45% (Table S2), suggesting a relatively higher sequence similarity within Class II.

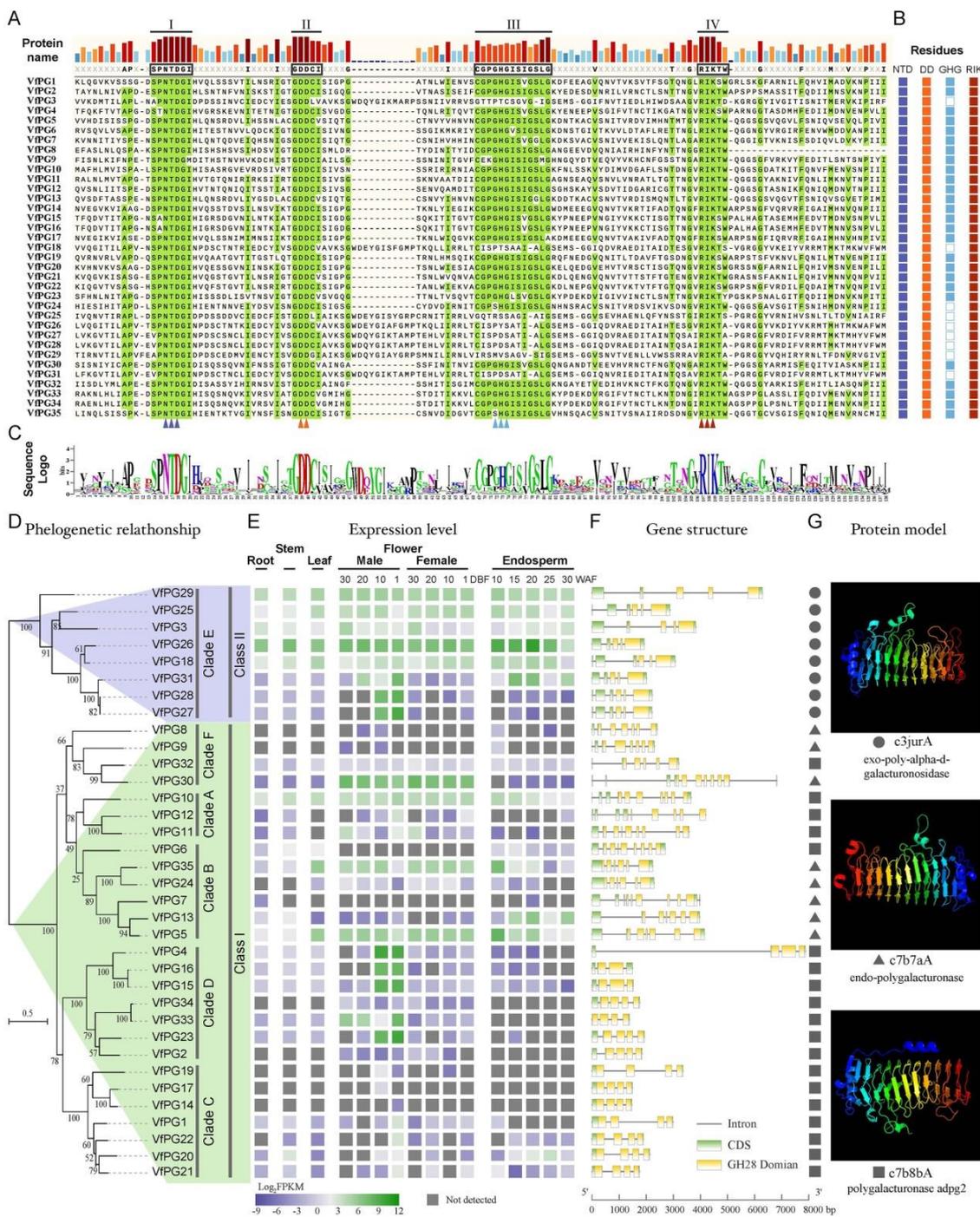


Figure 1. Sequence and structural characteristics of VfPGs. (A) Alignment of consensus protein sequences of VfPGs which contained four tung-specific conserved amino acid residues of GH28 domain (framed with black rectangles). (B) Statistics of four typical conserved amino acid residues in VfPGs. (C) Sequence logo of amino acid sequences of VfPGs. (D) Maximum likelihood-based phylogenetic relationships of VfPGs. (E) Expression patterns of VfPGs in roots, stems, leaves, flowers, and endosperms. (F) Gene structures of VfPGs. (G) Predicted tertiary structures of VfPG sequences. Triangles that were colored in purple, orange, blue, and deep red in (A) indicated the residues of NTD, DD, GHG, and RIK, respectively. Solid and hollow rectangles in (B) indicated the presence and absence of residues, respectively. DBF = Days before flowering; WAF = Weeks after flowering

The lengths of predicted proteins of the 35 VfPGs ranged from 310 to 522 amino acid residues with an average length of 426 (Table 1). The amino acid residues of GH28 domains in VfPGs were between 238 and 338, encompassing approximately 73.63% of the average PG protein sequences (Fig. 1F; Table S3). Exon numbers of VfPGs varied from 4 to 11 with an average number of 6 exons (Fig. 1F). Although exon numbers varied a lot among different clades of VfPGs, exon numbers of the same clade were usually conservative (Fig. 1F). Clade C and then Clade D were more conservative than other clades both on the sequence length and exon number. Clade C contained 7 genes and all these VfPGs had 4 exons. Exon numbers in Clade D varied between 4 and 5. Half of VfPGs in clade E had 5 exons, the other half had 6. Exon numbers in Clade A, Clade B, and Clade F varied from 6 to 11.

The molecular weights (MWs) of candidate PG proteins ranged from 33.79 to 56.70 kDa with an average weight of 46.17 kDa, and their isoelectric points (pIs) were between 4.76 and 9.29 with an average of 7.14 (Table 1). Subcellular localization prediction of 35 VfPGs showed that these proteins may be localized in 5 different cellular compartments, including chloroplast (1), endomembrane system (4), extracellular space (13), nucleus (5), and plasma membrane (12). Predicted tertiary structures of VfPG sequences were modeled with three templates based on the protein fold recognition server PHYRE2 (Fig. 1G; Table 1). Predicted tertiary structures of more than half of VfPGs were similar to c7b8bA (19 out of 35), while 8 and 8 PGs were similar to c7b7aA and c3jurA, respectively (Fig. 1G). The results revealed that Class I PGs belonged to endo-PGs, while the Class II PGs were exo-PGs. Generally, closely related genes usually had common gene structures (Fig. 1D, G).

Evolutionary analysis of PG gene family in tung tree

A rooted phylogenetic tree including 35 VfPGs and 65 *Arabidopsis* PGs was reconstructed with the maximum likelihood (ML) method using AT3G57790 as the outgroup (Fig. 2A). The tung and *Arabidopsis* PGs fell into 3 distinct classes, including 16 clades in Class I, 8 clades in Class II, and 1 clade in Class III according to the classification of AtPGs (Kim et al., 2006).

None of VfPGs was clustered into Class III. A total of 9 clades contained only AtPGs (Fig. 2A; marked with black triangles), while only 2 clades contained only VfPGs (Fig. 2A; marked with green triangles). The *Arabidopsis*-specific clades and tung-specific clades suggested several possible gene losses during the evolution of *Arabidopsis* and the tung tree, respectively. Phylogenetic relationship of PGs among 5 Euphorbiaceae plants, including the tung tree, castor bean (*R. communis*), physic nut (*J. curcas*), cassava (*M. esculenta*), and rubber tree (*H. brasiliensis*), was also analyzed in this study. Among the Euphorbiaceae species, we identified 35, 55, 48, 87, and 83 PG genes from the tung tree, castor bean, physic nut, cassava, and rubber tree genomes, respectively (Table S4). The number of PG family members in cassava, and rubber tree genome was approximately 2-fold of that in the tung tree, castor bean, and physic nut. Intriguingly, according to the ML tree reconstructed from GH28 domain sequences, the Euphorbiaceae PGs were classified into two groups, corresponding to Class I and Class II and missed Class III members (Fig. 2B). Reconstruction of phylogenetic relationships among 398 PGs from *Physcomitrella* (11), *Selaginella* (16), rice (42), maize (55), *Arabidopsis* (65), *Brassica* (99), *Populus* (75) and the tung tree (35) suggested that PGs in land plants can be classified into 3 distinct groups. All studied species, which included nonvascular plants, monocots, and eudicots, had Class III PG genes except for the tung tree.

Table 1. Summary of PG proteins identified in the tung tree

Gene name	Locus ID	Chromosome	Predicted	Molecular	Isoelectric	Exon	Subcellular	Tertiary structure prediction	
			Proteins (aa)	Weight (kDa)	Points (pI)		Localization Prediction	Template	Protein data bank molecule
VfPG1	Vf00G0359	0	415	44.17	8.16	4	Plasma membrane	c7b8bA	Polygalacturonase adpg2
VfPG2	Vf00G0716	0	409	44.26	5.99	5	Extracellular space	c7b8bA	Polygalacturonase adpg2
VfPG3	Vf01G1336	1	480	52.55	5.77	5	Nucleus	c3jurA	Exo-poly-alpha-d-galacturonosidase
VfPG4	Vf01G1744	1	391	41.56	5.17	4	Extracellular space	c7b8bA	Polygalacturonase adpg2
VfPG5	Vf01G2342	1	522	56.66	8.35	7	Extracellular space	c7b7aA	Endo-polygalacturonase
VfPG6	Vf01G2343	1	474	51.31	4.85	9	Extracellular space	c7b8bA	Polygalacturonase adpg2
VfPG7	Vf01G2365	1	456	50.05	9.12	10	Plasma membrane	c7b7aA	Endo-polygalacturonase
VfPG8	Vf02G2021	2	338	36.49	8.56	7	Extracellular space	c7b7aA	Endo-polygalacturonase
VfPG9	Vf02G2023	2	387	42.35	8.10	8	Extracellular space	c7b7aA	Endo-polygalacturonase
VfPG10	Vf02G2050	2	446	49.18	9.29	10	Nucleus	c7b8bA	Polygalacturonase adpg2
VfPG11	Vf03G0186	3	431	47.28	9.22	9	Endomembrane system	c7b8bA	Polygalacturonase adpg2
VfPG12	Vf04G0721	4	469	51.42	5.07	9	Plasma membrane	c7b8bA	Polygalacturonase adpg2
VfPG13	Vf04G1938	4	477	50.96	6.00	7	Endomembrane system	c7b7aA	Endo-polygalacturonase
VfPG14	Vf05G0737	5	389	41.71	8.50	4	Extracellular space	c7b8bA	Polygalacturonase adpg2
VfPG15	Vf05G0852	5	392	42.15	8.44	4	Extracellular space	c7b8bA	Polygalacturonase adpg2
VfPG16	Vf05G0853	5	325	35.21	8.72	4	Extracellular space	c7b8bA	Polygalacturonase adpg2
VfPG17	Vf05G1487	5	393	42.29	8.49	4	Endomembrane system	c7b8bA	Polygalacturonase adpg2
VfPG18	Vf05G1878	5	482	52.88	4.76	6	Plasma membrane	c3jurA	Exo-poly-alpha-d-galacturonosidase
VfPG19	Vf06G1194	6	389	41.64	9.10	4	Plasma membrane	c7b8bA	Polygalacturonase adpg2
VfPG20	Vf06G1199	6	390	41.24	6.06	4	Plasma membrane	c7b8bA	Polygalacturonase adpg2
VfPG21	Vf06G1200	6	371	39.19	7.49	4	Nucleus	c7b8bA	Polygalacturonase adpg2
VfPG22	Vf06G1204	6	393	41.55	8.47	4	Plasma membrane	c7b8bA	Polygalacturonase adpg2
VfPG23	Vf06G2095	6	423	44.35	6.38	5	Plasma membrane	c7b8bA	Polygalacturonase adpg2
VfPG24	Vf06G2210	6	517	56.70	5.51	6	Endomembrane system	c7b7aA	Endo-polygalacturonase
VfPG25	Vf06G2253	6	466	51.71	5.25	6	Extracellular space	c3jurA	Exo-poly-alpha-d-galacturonosidase
VfPG26	Vf06G2491	6	469	51.35	5.75	5	Extracellular space	c3jurA	Exo-poly-alpha-d-galacturonosidase
VfPG27	Vf08G0793	8	476	52.12	8.07	6	Plasma membrane	c3jurA	Exo-poly-alpha-d-galacturonosidase
VfPG28	Vf08G0796	8	476	52.04	7.09	6	Plasma membrane	c3jurA	Exo-poly-alpha-d-galacturonosidase
VfPG29	Vf08G1479	8	492	55.25	9.00	5	Chloroplast	c3jurA	Exo-poly-alpha-d-galacturonosidase
VfPG30	Vf08G1869	8	407	43.23	5.00	11	Plasma membrane	c7b7aA	Endo-polygalacturonase
VfPG31	Vf10G0347	10	442	48.35	5.57	5	Nucleus	c3jurA	Exo-poly-alpha-d-galacturonosidase
VfPG32	Vf10G2001	10	310	33.79	6.33	6	Extracellular space	c7b8bA	Polygalacturonase adpg2
VfPG33	Vf11G1871	11	316	33.91	8.69	4	Nucleus	c7b8bA	Polygalacturonase adpg2
VfPG34	Vf11G1873	11	416	44.96	8.88	5	Plasma membrane	c7b8bA	Polygalacturonase adpg2
VfPG35	Vf11G2029	11	482	51.99	4.76	6	Extracellular space	c7b7aA	Endo-polygalacturonase

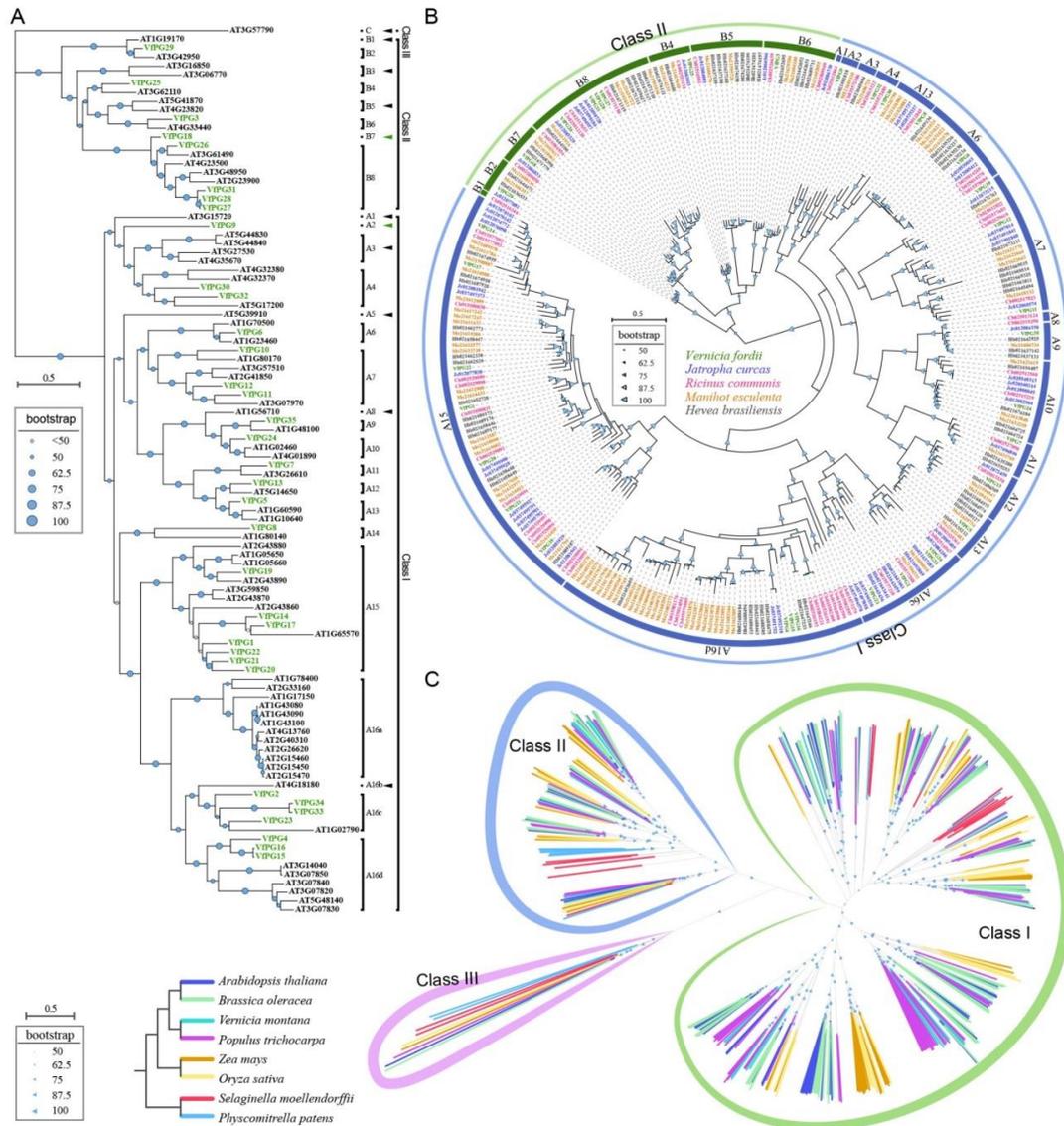


Figure 2. Phylogenetic relationship of PGs. (A) Phylogenetic tree of the tung tree and *Arabidopsis* PGs. (B) Phylogenetic tree of PGs from the tung tree and other 4 Euphorbiaceae species. (C) Phylogenetic tree of 398 PGs from 8 land plant species. Clades that only contained tung and *Arabidopsis* PG genes were marked by black and green arrows, respectively

Collinearity and selective pressure analyses

The distribution of *VfPGs* on tung chromosomes was examined. Almost all *VfPGs* (33 out of 35) were distributed unevenly on 9 tung chromosomes (Fig. 3), while the rest two genes (*VfPG1* and *VfPG2*) were localized on two unanchored scaffolds (Table 1). Gene duplications of 35 *VfPGs* were investigated by MCScanX. The *VfPGs* presented the characteristics of cluster distribution on Chromosomes 1, 2, 6, and 11. Only two pairs of *VfPGs* (*VfPG15/16* and *VfPG20/21*) were arranged in tandem repeats on chromosomes 5 and 6, respectively. A total of 12 *VfPGs* with 8 pairs associated with segmental duplications were distinguished based on conserved collinearity linkage, including *VfPG5/13*, *VfPG7/13*, *VfPG8/30*, *VfPG11/12*, *VfPG18/26*, *VfPG18/31*, *VfPG26/31* and *VfPG23/33*.

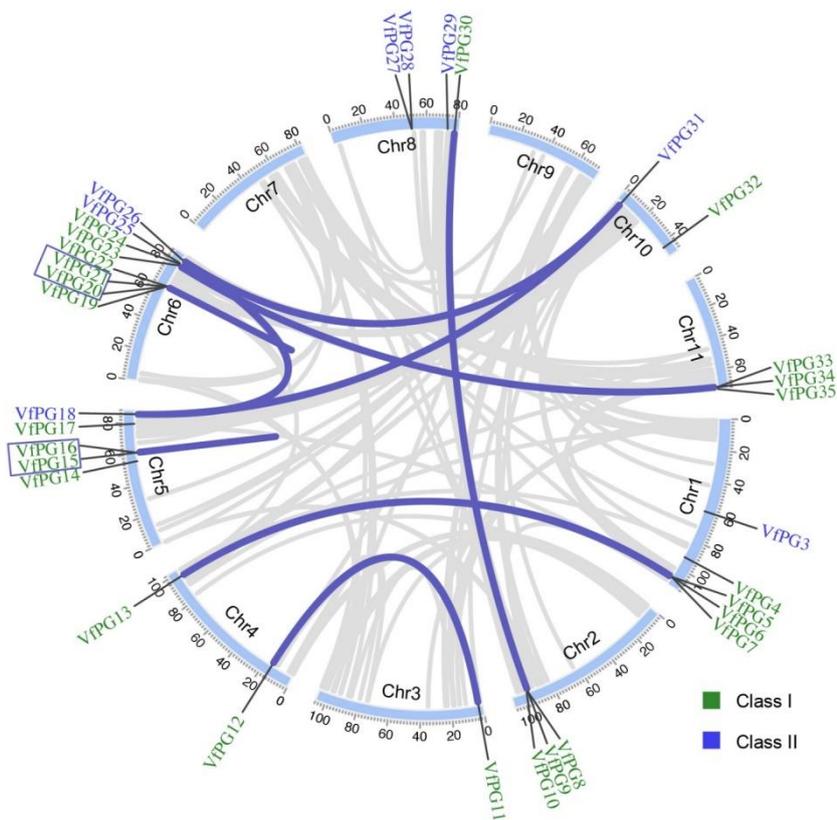


Figure 3. Genomic distribution and collinearity analysis of VfPG family members. Tandem-duplicated genes were framed with rectangles. The blue lines represented collinearity relationships of the paralog VfPGs. The grey lines represented collinearity relationships in the tung genome

The nonsynonymous substitution rate (K_a), synonymous substitution rate (K_s), and ω (K_a/K_s) of VfPG paralog pairs were calculated (Table 2). The results showed that the K_s values of tandem duplication pairs of VfPGs (ranging from 0.0614 to 0.858) were smaller than segmental duplication pairs (ranging from 0.9805 to 2.0430). The time of duplication events of VfPGs was also calculated. The latest origin time of VfPG paralog pairs (VfPG15/16) was ~4.72 million years ago (Mya), and the oldest origin time (VfPG7/13) was ~157.15 Mya. The ω values of all VfPG paralog pairs were less than 1 (ranging from 0.1576 to 0.6239), suggesting strong negative selection on duplicated VfPGs. The mean ω values of Class I and Class II VfPG pairs were 0.3162 and 0.1729, respectively, suggesting that Class II VfPGs were under stronger selection constraints than the Class I VfPGs.

cis-Regulatory element analysis of VfPG promoters

cis-Regulatory elements are DNA sequences within or flanking the gene that controls plant development and responses to the environment by regulating transcription or expression of a gene that is on the same chromosome (Marand et al., 2023). In general, a total of 280 *cis*-regulatory elements associated with plant growth and development (51), phytohormone responsiveness (107), and abiotic and biotic stresses (122) were identified from 2 kb region upstream of the start codon ATG site of VfPGs (Table S5; Fig. 4).

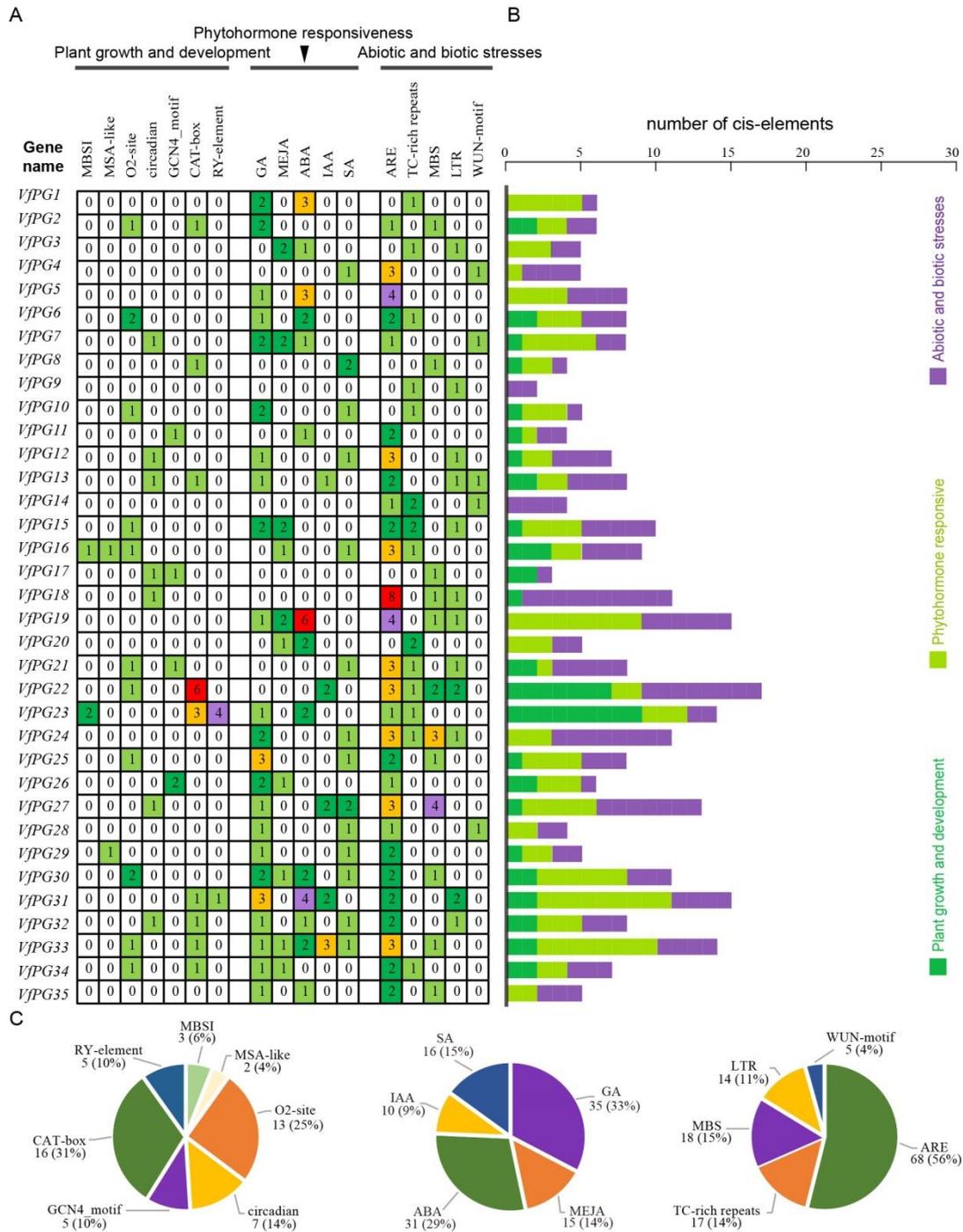


Figure 4. cis-Regulatory elements analysis of promoters of VJPGs. (A) Number of each cis-regulatory element in the promoter region of each VJPG. (B) The stacked column chart of cis-regulatory elements related to plant growth and development, phytohormone responsive, and abiotic and biotic stresses. (C) Pie charts of three groups of cis-regulatory elements

All of the VJPGs contained at least one type of defense and stress response element, including anaerobic responsive element (ARE) which is essential for the anaerobic induction, TC-rich repeats which are involved in defense and stress responsiveness, MYB binding site element (MBS) which is involved in drought inducibility, low-

temperature responsiveness (LTR), and WUN-motif which is involved in wound responsiveness. Thirty-two *VfPGs* contained phytohormone-responsive elements which were associated with responses to gibberellin (GA; TATC-box, GARE-motif, and P-box), abscisic acid (ABA; ABRE element), methyl jasmonate (MeJA; TGACG-motif/CGTCA-motif), indole-3-acetic acid (IAA; TGA-element and AuxRR-core), and salicylic acid (SA; TCA-element). Fewer *VfPGs* (24 out of 35) contained cis-elements associated with plant growth and development (Fig. 4A). MYB binding site which is involved in flavonoid biosynthetic genes regulation (MSBI), cell cycle regulation element (MSA-like) and seed-specific regulation element (RY-element) only existed in two promoters of *VfPGs*. CAT-box element (GCCACT) which is related to meristem expression existed in 9 promoters, and 11 promoters contained O2-site which is involved in zein metabolism regulation.

Table 2. The K_a , K_s and ω between *VfPG* paralog pairs of the tung tree

Paralog pairs		Duplication type	K_a	K_s	ω	Mean ω	Date (Mya)	Class
<i>VfPG5</i>	<i>VfPG13</i>	Segmental	0.2644	1.1970	0.2209	0.3162	92.08	I
<i>VfPG7</i>	<i>VfPG13</i>	Segmental	0.5192	2.0430	0.2541		157.15	I
<i>VfPG8</i>	<i>VfPG30</i>	Segmental	0.6166	1.6116	0.3826		123.97	I
<i>VfPG11</i>	<i>VfPG12</i>	Segmental	0.3169	1.2288	0.2579		94.52	I
<i>VfPG15</i>	<i>VfPG16</i>	Tandem	0.0383	0.0614	0.6239		4.72	I
<i>VfPG20</i>	<i>VfPG21</i>	Tandem	0.1940	0.8580	0.2260		66.00	I
<i>VfPG23</i>	<i>VfPG33</i>	Segmental	0.3969	1.5995	0.2481		123.04	I
<i>VfPG18</i>	<i>VfPG26</i>	Segmental	0.1787	0.9805	0.1822	0.1729	75.42	II
<i>VfPG18</i>	<i>VfPG31</i>	Segmental	0.2561	1.4328	0.1788		110.21	II
<i>VfPG26</i>	<i>VfPG31</i>	Segmental	0.2587	1.6414	0.1576		126.26	II

Gene expression and correlation analysis of *VfPGs*

In this study, FPKM values and e-fluorescence images of *VfPGs* in roots, stems, leaves, flowers, and endosperm were obtained through the tung tree electronic fluorescent pictographic (eFP) browser (Figs. 1E, and 5A). Over 85% (29 out of 35) *VfPGs* expressed in at least one tissue, except for *VfPG2*, *VfPG7-9*, *VfPG14*, and *VfPG34*. Seven *VfPGs*, including *VfPG3*, *VfPG5*, *VfPG10*, *VfPG18*, *VfPG25*, *VfPG26*, and *VfPG29*, expressed in all sequenced tissues. Compared with flowers (27), only a small number of *VfPGs* expressed in roots (9), stems (12), leaves (10), and endosperms (14), suggesting that *VfPGs* widely participated in the development of tung flowers than other tissues. The expression profiles of *VfPGs* in tung flowers were further explored. A total of 6 different expression profiles were identified, but only one profile which contained 9 *VfPGs* was significant with the p -value of $9.3E^{-3}$ (Fig. 5B; Table S6). In general, almost all genes exhibited the expression profile 4 were only expressed in late development stages (10 and 1 DBF) of male flowers. *VfPG11*, which shared a high level of homology with *AtQRT2* (*At3g07970*), only expressed at stages of 30 DBF in both male and female flowers. *VfPG12* is only expressed in male and female flowers at stage 1 DBF, indicating a possible function related to the flowering of tung flowers. *VfPG4*, *VfPG15*, *VfPG16*, *VfPG23*, *VfPG27*, and *VfPG28* are only expressed in male flowers at stages of 10 and 1 DBF with quite high FPKM values (ranging from 159.29 to 895.10)

may suggest that these *VfPGs* were highly correlated with the late development and maturity of male flowers. Only 8 PGs were identified as endo-PGs (*Fig. 1G*). Interestingly, the expression patterns of these endo-PGs in flowers were divided into two types, almost not expressed in any development stages (5 out of 8) or expressed in all development stages of male and female flowers (3 out of 8) (*Fig. 1E*). The expression profiles of *VfPG15*, *VfPG16*, *VfPG23*, *VfPG31*, and *VfPG33* which displayed relatively higher FPKM values in male flowers were further validated by qRT-PCR (quantitative real-time polymerase chain reaction) (*Fig. 5C*). The results showed that the qRT-PCR-based expression profiles of these genes were consistent with transcriptome data.

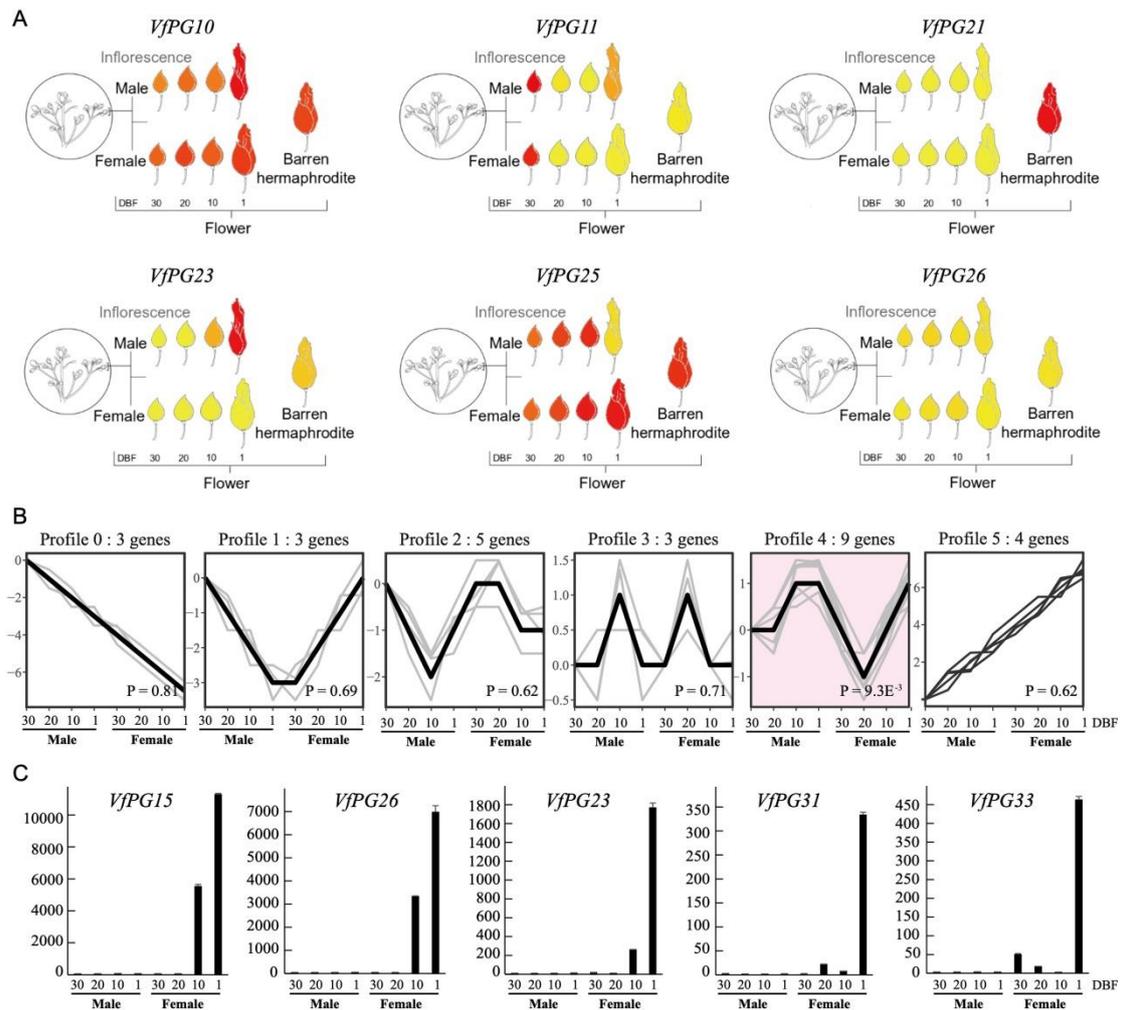


Figure 5. Gene expression Trend and qRT-PCR validation of *VfPGs* in tung flowers. (A) Electron fluorescence images of selected *VfPG* genes. (B) Expression profiles based on FPKM values. (C) Expression levels of five *VfPG* genes based on qRT-PCR. DBF = Days before flowering

The protein-protein interaction networks among *VfPGs* that are expressed in flowers were predicted by the STRING database based on the interacting protein pairs in *Arabidopsis* (*Fig. 6; Table S7*). The network was clustered into three clusters by *k*-

structures, while the rest of VfPGs lacked the conserved residues III which is suggested to be involved in the catalysis for containing the histidine residue (H) (Rao et al., 1996). Interestingly, the conserved residues III were lost in all Class II/clade E VfPGs, and the predicted tertiary structure of these PGs strongly supported the idea that these PGs were exo-PGs. VfPGs in the same clade usually have conservative gene structures, but intron lengths and numbers ranged a lot which may due to diverse autonomous transposable elements (Gozashti et al., 2022).

Phylogenetic relationships of PGs among the tung tree and other 4 representative Euphorbiaceae plants and 7 land plants were also reconstructed. It is worth noting that not any Class III PG gene was determined in the tung genome and other 4 Euphorbiaceae genomes, while several Class III members were found in *Arabidopsis* (1), *Brassica* (1), rice (1), *Populus* (1), *Physcomitrella* (2), *Selaginella* (2), and maize (1) which included nonvascular plants, monocots and eudicots, suggesting that Class III PG genes that were present in the most recent common ancestors of land plants could be lost in Euphorbiaceae after their divergence.

Gene duplication is a major force in evolution for generating new genes and creating genetic novelty that influences plant floral structures, disease resistance, adaption to stress, and so on (Magadum et al., 2013; Panchy et al., 2016). It has been reported that the majority of duplicated genes in *Arabidopsis*, rice, and 5 other model plants evolve under negative selection by comparing duplicated genes in 141 plant genomes (Qiao et al., 2019). Tandem duplications have experienced more rapid sequence divergence and stronger positive selection than segmental duplications, suggesting that tandem duplication is an essential source for the evolution of new functionalities (Qiao et al., 2019). A total of 10 paralog gene pairs, including 16 VfPGs, were identified in the tung genome. K_s values are usually used as a proxy for evolutionary time between duplication gene pairs (Maere et al., 2005). In this study, tandem gene pairs kept smaller K_s values than segmental gene pairs. The inferred origin time of segmental gene pairs ranged from 75.42 to 157.15 Mya. The whole genome triplication (WGT) peak value based on K_s values of tung paralog pairs was approximately 1.35, dating to 103.85 Mya (Zhang et al., 2019), suggesting that most of the segmental duplication pairs of VfPGs could originate from WGT. Moreover, the numbers of PG genes identified in rubber tree (83) and cassava (87) were approximately twice as much as that of the tung tree (35), physic nut (48), and castor bean (55), which was consistent with the round number of whole genome duplication (WGD) occurred in these species, suggesting that WGD was one of the major forces of the expansion of PG gene family in Euphorbiaceae. The ω ratios of all VfPG duplicated gene pairs were less than 1, suggesting strong negative selection on duplicated VfPGs.

The genomic blueprints which guarantee adequate spatiotemporal patterning of gene expression necessary for development and environmental responses are encoded by cis-regulatory elements (Marand et al., 2023). A total of 280 cis-regulatory elements related to hormone response, growth and development, and abiotic and biotic stress were identified in the promoters of VfPGs. Phytohormones are reported to participate in plant development activities by regulating PG genes. A total of 107 phytohormone responsiveness elements were identified in promoters of almost all VfPGs, including GA, ABA, SA, IAA, and MeJA. Exogenous application of ABA promotes ripening of *Vitis vinifera* with increased transcript abundance of *VvPGI* (Koyama et al., 2010). Ethylene inhibitor 1-methylcyclopropene treatment can significantly suppress gene expression of *PaPGI* and decrease its activities, thus alleviating cell wall disassembly and delaying fruit

softening of apricot (Hou et al., 2019). Therefore, phytohormone may also play important roles in plant development in tung tree by regulating *VfPG* activities.

Gene expression patterns can reflect their gene functions to some extent. Functional divergences and expression differences between paralogous genes which generated through WGD were thought to share identical expression levels due to their identical DNA sequences and chromatin environment (Lian et al., 2020). Neighboring genes, especially tandem duplicates which are crucial for the environmental adaptability of plants, are reported to tend to be co-regulated (Hanada et al., 2008; Dai et al., 2014). In general, *VfPGs* were more preferred to express in flowers than in other tissues, suggesting that PG family genes could contribute a lot to flower development in tung tree. Moreover, paralogs such as *VfPG18/VfPG26* displayed similar expression profiles, suggesting that they may have the same function.

Stamen is the male fertilizing organ of angiosperms and its fertility is determined by the early development process of meiosis, and following late developmental phases of filament elongation, pollen maturity, and anther dehiscence. Functional female tung flowers usually develop with the abortion of stamens from 30 DBF to 10 DBF, and pollen sacs of male tung flowers undergo meiosis of microsporocytes, degeneration of middle layers, tapetum, and maturity of pollen grains at the same time (Liu et al., 2019; Li et al., 2020). Previous studies have confirmed PG genes play important roles in stamen development. For example, *BcMF16* is expressed in tapetum and pollen during the late stages of pollen development (Zhang et al., 2012). *BcMF17* plays a role in pollen maturation or pollen tube growth (Zhang et al., 2011). The expression profiles of *VfPG15* and *VfPG16* (tandem duplicated gene pairs) were very similar and only highly expressed in the late development stages of male flowers, defining them were “late” genes of pollen development. *QRT2* is responsible for the release of viable pollen tetrads in *Arabidopsis* (Shi et al., 2021). In our study, *VfPG11* shared a high level of homology with *AtQRT2* and was only expressed in male flowers and female flowers with anthers in early development stages, indicating a possible role of *VfPG11* in pectin degeneration during tetrad separation in the tung tree. *PGA4* (*ATIG02790*), encoding an exo-PG, is reported to play roles in pectin degeneration, and is involved in pollen grain germination and pollen tube growth (Ye et al., 2020). *VfPG23* is highly homologous to *PGA4* and is highly expressed at important stages of anther development, suggesting *VfPG23* may be involved in the degradation of the callus around the microspores.

Conclusion

PG family genes are of great significance to the development of flowers in tung tree, especially *VfPG15*, *VfPG16*, *VfPG11* and *VfPG23*. These four genes are participated in the late development of anthers, which provides a theoretical basis for further elucidating the molecular mechanism of PG genes regulating the development of tung tree pollens.

Acknowledgment. This study was supported by Hubei Provincial Natural Science Foundation of China (Grant No. 2024AFB006), Doctoral Initiation Fund of Huanggang Normal University (Grant No. 2042024068), and Postgraduate Scientific Research Innovation Project of Hunan Province (Grant No. CX20200700). We thank Prof. Lin Zhang, Key Laboratory of Cultivation and Protection for Non-Wood Forest Trees of the Ministry of Education and Key Laboratory of Non-Wood Forest Products of the Forestry Ministry, Central South University of Forestry and Technology, for providing the reference genome of *V. fordii* in this study.

REFERENCES

- [1] Bosch, M., Hepler, P. K. (2005): Pectin methylesterases and pectin dynamics in pollen tubes. – *Plant Cell* 17(12): 3219-3226.
- [2] Cao, Y., Fan, T., Zhang, B., Li, Y. (2022): Dissection of leucine-rich repeat receptor-like protein kinases: insight into resistance to *Fusarium* wilt in tung tree. – *PeerJ* 10: e14416.
- [3] Cao, Y., Fan, T., Wang, L., Zhang, L., Li, Y. (2023): Large-scale analysis of putative Euphorbiaceae R2R3-MYB transcription factors identifies a MYB involved in seed oil biosynthesis. – *BMC Plant Biol* 23(1): 145.
- [4] Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., Xia, R. (2020): TBtools: an integrative toolkit developed for interactive analyses of big biological data. – *Mol Plant* 13(8): 1194-1202.
- [5] Cheng, C. Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., Town, C. D. (2017): Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. – *Plant J* 89(4): 789-804.
- [6] Dai, Z., Xiong, Y., Dai, X. (2014): Neighboring genes show interchromosomal colocalization after their separation. – *Mol Biol Evol* 31(5): 1166-1172.
- [7] Duan, W., Huang, Z., Song, X., Liu, T., Liu, H., Hou, X., Li, Y. (2016): Comprehensive analysis of the polygalacturonase and pectin methylesterase genes in *Brassica rapa* shed light on their different evolutionary patterns. – *Sci Rep* 6: 25107.
- [8] Duvaud, S., Gabella, C., Lisacek, F., Stockinger, H., Ioannidis, V., Durinx, C. (2021): Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users. – *Nucleic Acids Res* 49(W1): W216-W227.
- [9] Gozashti, L., Roy, S. W., Thornlow, B., Kramer, A., Ares, M., Jr., Corbett-Detig, R. (2022): Transposable elements drive intron gain in diverse eukaryotes. – *Proc Natl Acad Sci USA* 119(48): e2209766119.
- [10] Hadfield, K. A., Bennett, A. B. (1998): Polygalacturonases: many genes in search of a function. – *Plant Physiol* 117(2): 337-343.
- [11] Han, X., Lu, M., Chen, Y., Zhan, Z., Cui, Q., Wang, Y. (2012): Selection of reliable reference genes for gene expression studies using real-time PCR in tung tree during seed development. – *PLoS One* 7(8): e43084.
- [12] Hanada, K., Zou, C., Lehti-Shiu, M. D., Shinozaki, K., Shiu, S. H. (2008): Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. – *Plant Physiol* 148(2): 993-1003.
- [13] Hou, Y., Wu, F., Zhao, Y., Shi, L., Zhu, X. (2019): Cloning and expression analysis of polygalacturonase and pectin methylesterase genes during softening in apricot (*Prunus armeniaca*, L.) fruit. – *Sci Hortic* 256: 108607.
- [14] Huang, L., Cao, J., Zhang, A., Ye, Y., Zhang, Y., Liu, T. (2009): The polygalacturonase gene BcMF2 from *Brassica campestris* is associated with intine development. – *J Exp Bot* 60(1): 301-313.
- [15] Katoh, K., Standley, D. M. (2013): MAFFT multiple sequence alignment software version 7: improvements in performance and usability. – *Mol Biol Evol* 30(4): 772-780.
- [16] Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., Sternberg, M. J. (2015): The Phyre2 web portal for protein modeling, prediction and analysis. – *Nat Protoc* 10(6): 845-858.
- [17] Kim, J., Shiu, S. H., Thoma, S., Li, W. H., Patterson, S. E. (2006): Patterns of expansion and expression divergence in the plant polygalacturonase gene family. – *Genome Biol* 7(9): R87.
- [18] Koyama, K., Sadamatsu, K., Goto-Yamamoto, N. (2010): Abscisic acid stimulated ripening and gene expression in berry skins of the Cabernet Sauvignon grape. – *Funct Integr Genomics* 10(3): 367-381.
- [19] Lescot, M., Dehais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouze, P., Rombauts, S. (2002): PlantCARE, a database of plant cis-acting regulatory elements and

- a portal to tools for in silico analysis of promoter sequences. – *Nucleic Acids Res* 30(1): 325-327.
- [20] Letunic, I., Bork, P. (2021): Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. – *Nucleic Acids Res* 49(W1): W293-W296.
- [21] Letunic, I., Khedkar, S., Bork, P. (2021): SMART: recent updates, new developments and status in 2020. – *Nucleic Acids Res* 49(D1): D458-D460.
- [22] Li, W., Liu, M., Dong, X., Cao, H., Wu, Y., Shang, H., Huang, H., Zhang, L. (2020): Flower biology and ontogeny of the tung tree (*Vernicia fordii* Hemsl.). – *Trees* 34(6): 1363-1381.
- [23] Lian, S., Zhou, Y., Liu, Z., Gong, A., Cheng, L. (2020): The differential expression patterns of paralogs in response to stresses indicate expression and sequence divergences. – *BMC Plant Biol* 20(1): 277.
- [24] Liu, M., Li, W., Zhao, G., Fan, X., Long, H., Fan, Y., Shi, M., Tan, X., Zhang, L. (2019): New insights of salicylic acid into stamen abortion of female flowers in tung tree (*Vernicia fordii*). – *Front Genet* 10: 316.
- [25] Lu, L., Hou, Q., Wang, L., Zhang, T., Zhao, W., Yan, T., Zhao, L., Li, J., Wan, X. (2021): Genome-wide identification and characterization of polygalacturonase gene family in maize (*Zea mays* L.). – *Int J Mol Sci* 22(19).
- [26] Lyu, M., Iftikhar, J., Guo, R., Wu, B., Cao, J. (2020): Patterns of expansion and expression divergence of the polygalacturonase gene family in *Brassica oleracea*. – *Int J Mol Sci* 21(16).
- [27] Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., Van de Peer, Y. (2005): Modeling gene and genome duplications in eukaryotes. – *Proc Natl Acad Sci USA* 102(15): 5454-5459.
- [28] Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., Ravikesavan, R. (2013): Gene duplication as a major force in evolution. – *J genet* 92(1): 155-161.
- [29] Mao, Y., Liu, W., Chen, X., Xu, Y., Lu, W., Hou, J., Ni, J., Wang, Y., Wu, L. (2017): Flower development and sex determination between male and female flowers in *Vernicia fordii*. – *Front Plant Sci* 8: 1291.
- [30] Marand, A. P., Eveland, A. L., Kaufmann, K., Springer, N. M. (2023): cis-Regulatory elements in plant development, adaptation, and evolution. – *Annu Rev Plant Biol* 74: 111-137.
- [31] Markovic, O., Janecek, S. (2001): Pectin degrading glycoside hydrolases of family 28: sequence-structural features, specificities and evolution. – *Protein Eng* 14(9): 615-631.
- [32] Nguyen, L. T., Schmidt, H. A., von Haeseler, A., Minh, B. Q. (2015): IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. – *Mol Biol Evol* 32(1): 268-274.
- [33] Ogawa, M., Kay, P., Wilson, S., Swain, S. M. (2009): Arabidopsis dehiscence zone polygalacturonase1 (ADPG1), ADPG2, and QUARTET2 are polygalacturonases required for cell separation during reproductive development in Arabidopsis. – *Plant Cell* 21(1): 216-233.
- [34] Otasek, D., Morris, J. H., Boucas, J., Pico, A. R., Demchak, B. (2019): Cytoscape Automation: empowering workflow-based network analysis. – *Genome Biol* 20(1): 185.
- [35] Panchy, N., Lehti-Shiu, M., Shiu, S. H. (2016): Evolution of gene duplication in plants. – *Plant Physiol* 171(4): 2294-2316.
- [36] Park, K. C., Kwon, S. J., Kim, N. S. (2010): Intron loss mediated structural dynamics and functional differentiation of the polygalacturonase gene family in land plants. – *Genes & Genom* 32: 570-577.
- [37] Parre, E., Geitmann, A. (2005): Pectin and the role of the physical properties of the cell wall in pollen tube growth of *Solanum chacoense*. – *Planta* 220(4): 582-592.
- [38] Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar Gl, A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L. (2023): InterPro in 2022. – *Nucleic Acids Res* 51(D1): D418-D427.

- [39] Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., Zhang, S., Paterson, A. H. (2019): Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. – *Genome Biol* 20(1): 38.
- [40] Rao, M. N., Kembhavi, A. A., Pant, A. (1996): Implication of tryptophan and histidine in the active site of endo-polygalacturonase from *Aspergillus ustus*: elucidation of the reaction mechanism. – *Biochim Biophys Acta* 1296(2): 167-173.
- [41] Rhee, S. Y., Osborne, E., Poindexter, P. D., Somerville, C. R. (2003): Microspore separation in the quartet 3 mutants of *Arabidopsis* is impaired by a defect in a developmentally regulated polygalacturonase required for pollen mother cell wall degradation. – *Plant Physiol* 133(3): 1170-1180.
- [42] Shen, W., Le, S., Li, Y., Hu, F. (2016): SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. – *PLoS ONE* 11(10): e0163962.
- [43] Shi, Q. S., Lou, Y., Shen, S. Y., Wang, S. H., Zhou, L., Wang, J. J., Liu, X. L., Xiong, S. X., Han, Y., Zhou, H. S. (2021): A cellular mechanism underlying the restoration of thermo/photoperiod-sensitive genic male sterility. – *Mol Plant* 14(12): 2104-2114.
- [44] Vogel, J. (2008): Unique aspects of the grass cell wall. – *Curr Opin Plant Biol* 11(3): 301-307.
- [45] Wang, D., Kanyuka, K., Papp-Rupar, M. (2023): Pectin: a critical component in cell-wall-mediated immunity. – *Trends Plant Sci* 28(1): 10-13.
- [46] Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Lee, T. H., Jin, H., Marler, B., Guo, H. (2012): MCSanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. – *Nucleic Acids Res* 40(7): e49.
- [47] Wu, Y., Li, X., Li, Y., Ma, H., Chi, H., Ma, Y., Yang, J., Xie, S., Zhang, R., Liu, L. (2022): Degradation of de-esterified pectin/homogalacturonan by the polygalacturonase GhNSP is necessary for pollen exine formation and male fertility in cotton. – *Plant Biotechnol J* 20(6): 1054-1068.
- [48] Xin, Y., Liu, Z., Yang, C., Dong, C., Chen, F., Liu, K. (2023): Smart antimicrobial system based on enzyme-responsive high methoxyl pectin-whey protein isolate nanocomplex for fresh-cut apple preservation. – *Int J Biol Macromol* 253(Pt 4): 127064.
- [49] Yang, S., Wang, Y., Yin, H., Guo, H., Gao, M., Zhu, H., Chen, Y. (2015): Identification and characterization of NF-YB family genes in tung tree. – *Mol Genet Genomics* 290(6): 2187-2198.
- [50] Yang, Z. L., Liu, H. J., Wang, X. R., Zeng, Q. Y. (2013): Molecular evolution and expression divergence of the *Populus* polygalacturonase supergene family shed light on the evolution of increasingly complex organs in plants. – *New Phytol* 197(4): 1353-1365.
- [51] Ye, J., Yang, X., Yang, Z., Niu, F., Chen, Y., Zhang, L., Song, X. (2020): Comprehensive analysis of polygalacturonase gene family highlights candidate genes related to pollen development and male fertility in wheat (*Triticum aestivum*, L.). – *Planta* 252(2): 31.
- [52] Zhang, A., Chen, Q., Huang, L., Qiu, L., Cao, J. (2011): Cloning and expression of an anther-abundant polygalacturonase gene *BcMF17* from *Brassica campestris* ssp. *chinensis*. – *Plant Mol Biol Rep* 29: 943-951.
- [53] Zhang, A., Qiu, L., Huang, L., Yu, X., Lu, G., Cao, J. (2012): Isolation and characterization of an anther-specific polygalacturonase gene, *BcMF16*, in *Brassica campestris* ssp. *chinensis*. – *Plant Mol Biol Rep* 30: 330-338.
- [54] Zhang, L., Liu, M., Long, H., Dong, W., Pasha, A., Esteban, E., Li, W., Yang, X., Li, Z., Song, A. (2019): Tung tree (*Vernicia fordii*) genome provides a resource for understanding genome evolution and improved oil production. – *Genom Proteom Bioinf* 17(6): 558-575.
- [55] Zhang, Z. (2022): KaKs_Calculator 3.0: calculating selective pressure on coding and non-coding sequences. – *Genom Proteom Bioinf* 20(3): 536-540.

APPENDIX

This manuscript has an electronic appendix.