BMC Genomics

RESEARCH Open Access

Genome-wide identification and expression patterns of the aspartic protease gene family in *Epimedium pubescens*

Huifang Zheng[1,2], Liumeng Zheng[1], Huiying Song[3] and Xiaobo Yu[1,2*]

Abstract

Background Aspartic proteases (APs), proteolytic enzymes involved in protein maturation, degradation, and signaling, are found in various organisms. The plant *Epimedium pubescens* is known for its pharmacologically active flavonoids and its use in traditional Chinese medicine. Despite this, to date, the AP gene family in this species has not been functionally analyzed. This study aimed to uncover the roles of AP genes in *E. pubescens* (*EpAP*s), focusing in particular on their involvement in light stress responses.

Results Genome-wide analysis identified and characterized a total of 103 *EpAP*s, which were categorized into four phylogenetic groups and revealed conserved motifs crucial for their catalytic function. Structural analysis highlighted the diversity of intron-exon arrangements and the predominant role of tandem duplication in gene expansion. Promoter analysis showed an enrichment of light-responsive elements, indicating potential involvement in light stress responses. Tissue-specific expression patterns revealed specialized roles in various organs, whereas several *EpAP*s exhibited stage-specific expression during the formation of abscission zones. The analysis of protein–protein interactions identified links to reproductive development, programmed cell death, and stress responses. Under light stress, selected AP genes exhibited dynamic changes in expression, with some showing transient upregulation or recovery phases, which suggests their involvement in short-term adaptation or sustained light stress responses.

Conclusions This study provides the first comprehensive analysis of AP genes in *E. pubescens*, highlighting their potential roles in development and stress adaptation. The presence of light-responsive elements and changes in expression under light stress suggest that AP genes may serve as key regulators of environmental responses in this species. Further validation studies could inform strategies to improve light stress resistance in shade-adapted plants.

Keywords Aspartic proteases, *Epimedium pubescens*, Genome-wide identification, Abscission zone formation, Light stress response

*Correspondence:
Xiaobo Yu
yuxiaobosky@163.com
[1]Southwest Research Center for Cross Breeding of Special Economic Plants, College of Life Science, Leshan Normal University, Leshan, China
[2]Bamboo Diseases and Pest Control and Resources Development Key Laboratory of Sichuan Province, College of Life Science, Leshan Normal University, Leshan, China
[3]College of Agronomy, Sichuan Agricultural University, Chengdu, China

Introduction

Aspartic proteases (APs; EC 3.4.23) are proteolytic enzymes found in various organisms, including bacteria, fungi, animals, and plants. Their catalytic activity relies on highly conserved Asp–Thr/Ser–Gly (DTG/DSG) motifs, which enable their functions in protein maturation, degradation, and signaling [1, 2, 3, 4]. Within the MEROPS protease database, APs are classified into 17 families based on their structural characteristics and evolutionary origins and are further grouped into five following clans: AA, AB, AC, AD, and AF [5]. Most plant APs belong to the A1 family of Clan AA [4]. The APs of this family, commonly referred to as pepsin-like APs, exhibit catalytic activity under acidic conditions and are inhibited by gastric protease inhibitor A [2, 6]. However, this is not the case for all plant APs; for example, the *Arabidopsis* CDR1 protein displays peak activity at higher pH values ranging from 6.0 to 6.5 [7].

In 2004, 59 A1 members were identified in *Arabidopsis thaliana* and classified into groups A1-1 through A1-5 [8]. In 2005, another study identified 51 A1 members and divided them into three categories, namely typical APs, nucellin-like APs, and atypical APs [9]. Subsequently, in 2008, 69 AP genes were identified in *Arabidopsis*, 63 of which belonged to the AP family and 6 were classified as non-peptidases [10]. Phylogenetic and structural analyses by Duan et al.. further classified 70 APs of the A1 family in *A. thaliana* into four distinct groups (Groups I–IV), with Group IV representing a conserved group in Spermatophyta that lacks the classical DTGS/DSGT motif [11]. Moreover, studies have identified 104 A1 APs in rice (*Oryza sativa*) [12], 50 in grape (*Vitis vinifera*) [13], 67 in black cottonwood (*Populus trichocarpa*) [14], and 129 in Moso bamboo (*Phyllostachys edulis*) [15], highlighting the expansion and functional diversity of AP genes in plants.

Plant APs play multifaceted roles in plant growth, development, and stress responses. In vegetative tissues, they are crucial for protein processing and turnover. For instance, phytepsin, a vacuolar AP in barley, participates in the autolysis of tracheary elements and sieve cells [16, 17]. During leaf senescence in tobacco, CND41, an AP found in the chloroplast, facilitates the turnover of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) [18]. During seed germination, ASPARTIC PROTEASE IN GUARD CELL 1 (ASPG1) in *A. thaliana* breaks down storage proteins, assisting in seedling development [19]. APs also play critical roles in plant reproduction by regulating programmed cell death (PCD) in tapetal cells and participating in pollen development. In rice, OsAP25 and OsAP37, which are regulated by the transcription factor ETERNAL TAPETUM1 (EAT1), control PCD in tapetal cells, a function that is essential for pollen viability [20], whereas OsAP65 contributes to the biosynthesis of key compounds for pollen germination and tube growth [21]. In *A. thaliana*, the AP known as PROMOTION OF CELL SURVIVAL 1 (PCS1) regulates cell fate during embryonic development and reproduction [22], while A36 and A39 may be involved in gametogenesis and pollen guidance [23]. Studies have suggested that plant APs also play important roles in defense against biotic and abiotic stresses. APCB1 is a nucellin-like AP that induces Bcl-2-ASSOCIATED ATHANOGENE 6 (BAG6) to initiate autophagy and defense mechanisms [24]. An extracellular AP in *A. thaliana* and rice known as Constitutive disease resistance 1 (CDR1) enhances immunity through peptide signaling pathways [25, 26]. Furthermore, overexpression of *VqAP13* in grape has been shown to improve resistance to powdery mildew by modulating the salicylic acid and methyl jasmonate signaling pathways [27]. In *A. thaliana*, ASPG1, which is induced by abscisic acid (ABA), plays a role in enhancing drought resistance through ABA-dependent signaling pathways [28]. Additionally, the overexpression of *APA1* in *A. thaliana* confers greater tolerance to drought stress than that exhibited by wild-type plants [29].

Despite significant progress being made in understanding APs in model species, their roles in non-model plants remain underexplored. *Epimedium pubescens*, a medicinal herb belonging to the Berberidaceae family, is rich in pharmacologically active flavonoids such as icariin and epimedin A, B, and C [30, 31]. This species is widely used in traditional Chinese medicine to treat kidney-yang deficiency and rheumatism [32]. Although it is a shade-loving plant, recent studies have demonstrated that controlled light enhancement can significantly increase its flavonoid content. A study on *E. sagittatum* has shown that total flavonoid content peaked at 40–100 $\mu mol \cdot m^{-2} \cdot s^{-1}$, was lowest at 20–35 $\mu mol \cdot m^{-2} \cdot s^{-1}$, and declined above 160 $\mu mol \cdot m^{-2} \cdot s^{-1}$ due to inhibited synthesis [33]. In *E. pseudowushanense*, flavonoid content has also been shown to vary with light levels, with the highest epimedin C yield being recorded at 54.6 ± 2.5 $\mu mol \cdot m^{-2} \cdot s^{-1}$ [34]. In contrast to these two species, *E. pubescens* exhibits higher tolerance to strong light. A specific study has revealed that as light intensity increased from L1 (36.4 ± 5.0 $\mu mol \cdot m^{-2} \cdot s^{-1}$) to L4 ($200.2 \pm 5.0$ $\mu mol \cdot m^{-2} \cdot s^{-1}$), both branch and leaf numbers increased significantly. Moreover, the content of *Epimedium*-derived flavonoid glycosides also increased with light intensity, peaking under L3 (145.6 ± 5.0 $\mu mol \cdot m^{-2} \cdot s^{-1}$), suggesting that this level is optimal for maximizing the accumulation of medicinal compounds [35]. Some studies have found that certain APs, such as NANA [36] and APP-A1 [37], play important roles in photosynthesis. Therefore, the present study focused on the AP genes in *E. pubescens* (*EpAP*s) to explore their roles in light stress responses. Specifically, *EpAP*s were identified and their phylogeny, structure, conserved

motifs, duplication, and expression patterns were analyzed. Notably, promoter analysis revealed a substantial number of light-responsive elements, prompting us to examine their expression under light stress via real-time quantitative PCR (qPCR) to better understand their potential functions. The findings obtained contribute to the identification of key genes that could be used to enhance light stress resistance in *E. pubescens*.

Results

Genome-wide identification and characterization of AP genes in *E. pubescens*

A total of 109 and 143 candidate AP proteins in *E. pubescens* were initially identified using HMMER v3.4 and BLASTP searches against AP proteins in *A. thaliana*, respectively. After Batch-CDD domain verification, 103 *EpAP*s were confirmed (Table S1) and further validated using the SMART and InterProScan tools (Figure S1, Table S2). These genes were subsequently renamed *EpAP1–EpAP103* based on their chromosomal scaffold positions.

The predicted EpAP proteins exhibited diverse characteristics. Their amino acid lengths ranged from 211 to 795, with molecular weights ranging from 22,810.02 to 86,282.81 Da. The isoelectric points (pIs) varied between 4.46 and 9.63. The instability index varied from 24.82 to 53.48, with 63 proteins being classified as stable (instability index values below 40), while the rest exhibited slightly higher values for this index. Overall, 63 proteins exhibited hydrophilicity, as indicated by negative GRAVY scores, whereas the remaining 40 were classified as hydrophobic.

Subcellular localization predictions using WoLF PSORT indicated that the EpAP proteins were likely localized primarily in the chloroplast (40), followed by extracellular space (25), and vacuoles (17), while DeepLoc predictions revealed a predominant extracellular localization (73), followed by localization to cell membrane (9) and lysosome/vacuole (11) (Table S3). These differences are likely due to variations in the training datasets and algorithms of the two prediction tools, affecting in particular proteins with dual localization potential, such as those predicted to localize both in the chloroplast and extracellular space.

Phylogenetic analysis

A total of 173 AP proteins, 70 from *A. thaliana* and 103 from *E. pubescens*, were used to construct a phylogenetic tree, where they clustered into four distinct groups (Groups I–IV) (Fig. 1). Group III, with 77 members, represented the largest group, potentially reflecting gene expansion or functional diversification. In contrast, Groups I, II, and IV contained 2, 12, and 12 members, respectively.

Analysis of gene structure, conserved domains, and motifs

The gene structure and motif composition of the examined *EpAP*s varied significantly among groups (Fig. 2). Group I contained motifs 10, 1, 7, 6, 3, and 2, while Group II exhibited motifs 10, 1, 6, 4, 3, and 2, with motifs 4, 7, or 9 being present in most members. Group III was highly diverse and the only one exhibiting motif 8. Group IV predominantly contained motifs 6, 4, 3, and 2. Motifs 1 and 6, each 15 amino acids in length, were the shortest and most conserved (Fig. 3). Motif 1 was shown to encode the AP active site DTGS, a signature feature of APs that is essential for enzymatic activity, and was identified in 93 *EpAP*s. Motif 3, comprising 29 amino acids, was one of the most conserved, identified in 101 *EpAP*s. Another highly conserved motif, also present in 101 genes, was motif 2, which likely contributes to protein stability and enzymatic integrity. In contrast, motif 5 was the least conserved, being found in only 32 *EpAP*s and predominantly within Group III, highlighting a potential role in subgroup-specific functions.

All EpAP protein sequences contained the AP domain (PLN03146), although some were not displayed due to structural overlap (Fig. 2C). According to the results obtained via InterProScan, all 103 EpAP proteins possessed a Peptidase Family A1 domain. Additionally, Group I members EpAP103 and EpAP90 contained a Saposin-like domain. SMART domain analysis revealed that certain EpAP proteins (e.g., EpAP12, EpAP28, EpAP35, and EpAP48) lacked an N-terminal signal peptide. Furthermore, other members, including EpAP16, EpAP24, EpAP35, and EpAP40, contained transmembrane domains at either the N- or C-terminus.

The analysis of gene structure revealed intron numbers ranging from 1 to 15. A total of 48 *EpAPs* (46.6%) contained a single exon and no introns, whereas 32 had two exons and one intron, and 23 contained more than two introns (Table S4). Untranslated regions (UTRs) were absent in 78.6% of the genes, whereas 4.9% had one UTR, and 16.5% had more than one UTRs. The absence of UTRs in most *EpAP*s suggests that post-transcriptional regulation may play a significant role in controlling their expression, which is consistent with the stress-responsive roles of these genes.

Chromosomal locations, genomic duplications, and Ka/ks ratios

The chromosomal distribution of *EpAPs* revealed a non-uniform pattern across six chromosomes (Fig. 4A), with chromosome 4 (Chr4) harboring the highest number of genes (30 in total; *EpAP50–EpAP79*), predominantly clustered in the upper and middle regions. In contrast, Chr5 exhibited the lowest number, with only six evenly distributed genes (*EpAP80–EpAP85*). Chr1 and Chr2 each contained 21 genes, wherein they were clustered in

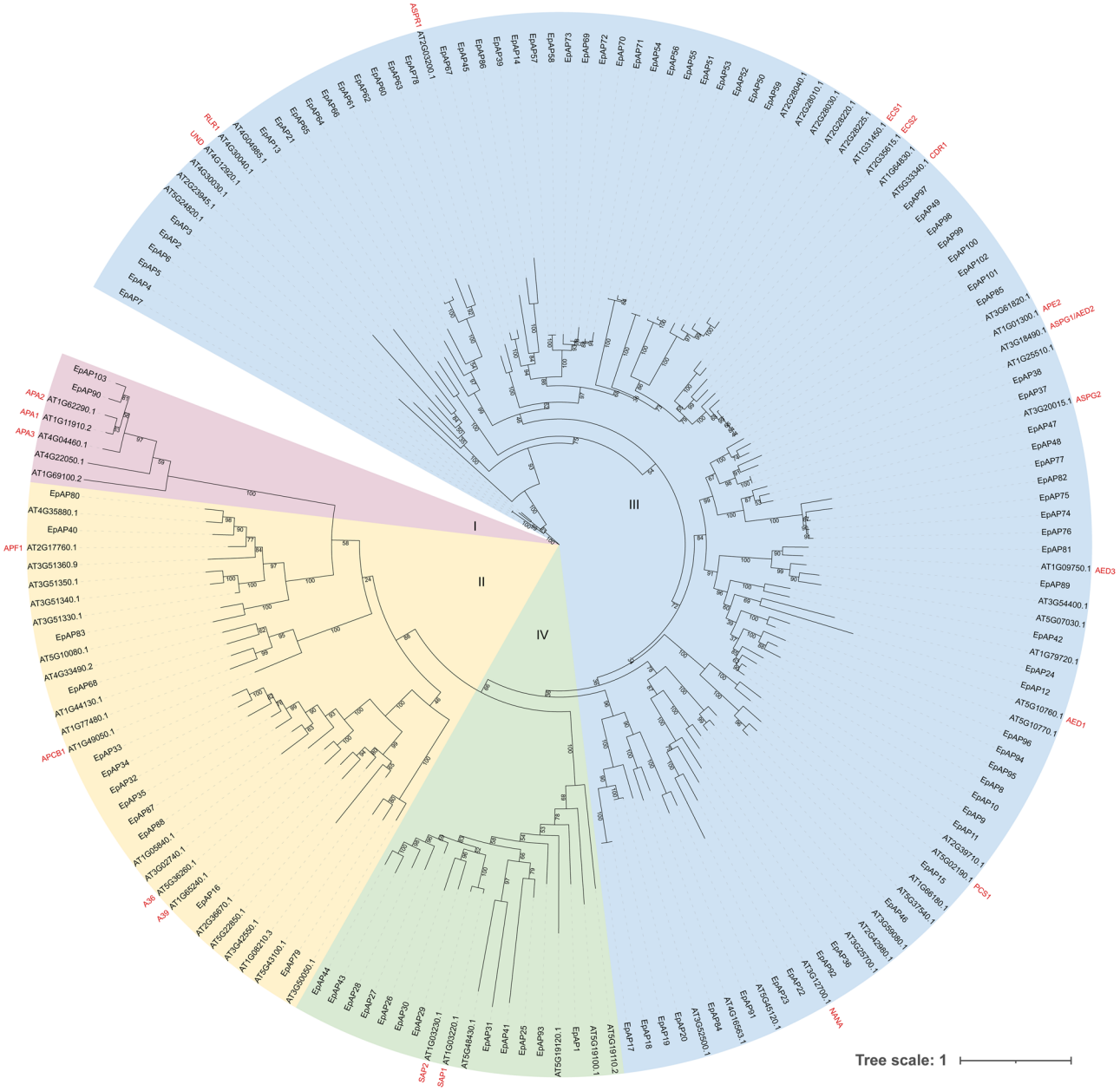

Fig. 1 Phylogenetic tree based on AP protein sequences in *E. pubescens* and *A.thaliana*

the upper regions of Chr1 (*EpAP1–EpAP21*) and central region of Chr2 (*EpAP22–EpAP42*). Chr3 contained seven genes (*EpAP43–EpAP49*) scattered on the chromosomes, whereas Chr6 contained 18 genes (*EpAP86–EpAP103*), mainly clustered at its terminal region. The abundance of tandem repeats within the examined *EpAP*s contributed to their diversity.

Our analysis identified one pair of collinear *EpAP*s (*EpAP1* and *EpAP26*; *Ka/Ks* = 0.179) (Fig. 4A, denoted by the green line) and 17 pairs of tandemly repeated *EpAP*s with *Ka/Ks* ratios ranging from 0.10 to 1.12 (Table S5). The majority of gene pairs (16/17) exhibited *Ka/Ks* ratios < 1, indicating that these genes have primarily undergone purifying selection during evolution. Notably, the gene pair *EpAP70*/*EpAP71* (*Ebr02G018630.1*/*Ebr02G018640.1*) had a *Ka/Ks* ratio slightly greater than 1 (1.125), suggesting that this duplication event may have experienced a brief period of positive selection or functional divergence. The low *Ks* values (0.045–1.710) reflected recent tandem repeat events that are potentially related to recent genomic expansion events in *E. pubescens*. In the comparison of collinear genes between *E. pubescens* and *A. thaliana*, all gene pairs showed *Ka/Ks* ratios significantly lower than 0.5 (0.104–0.419),

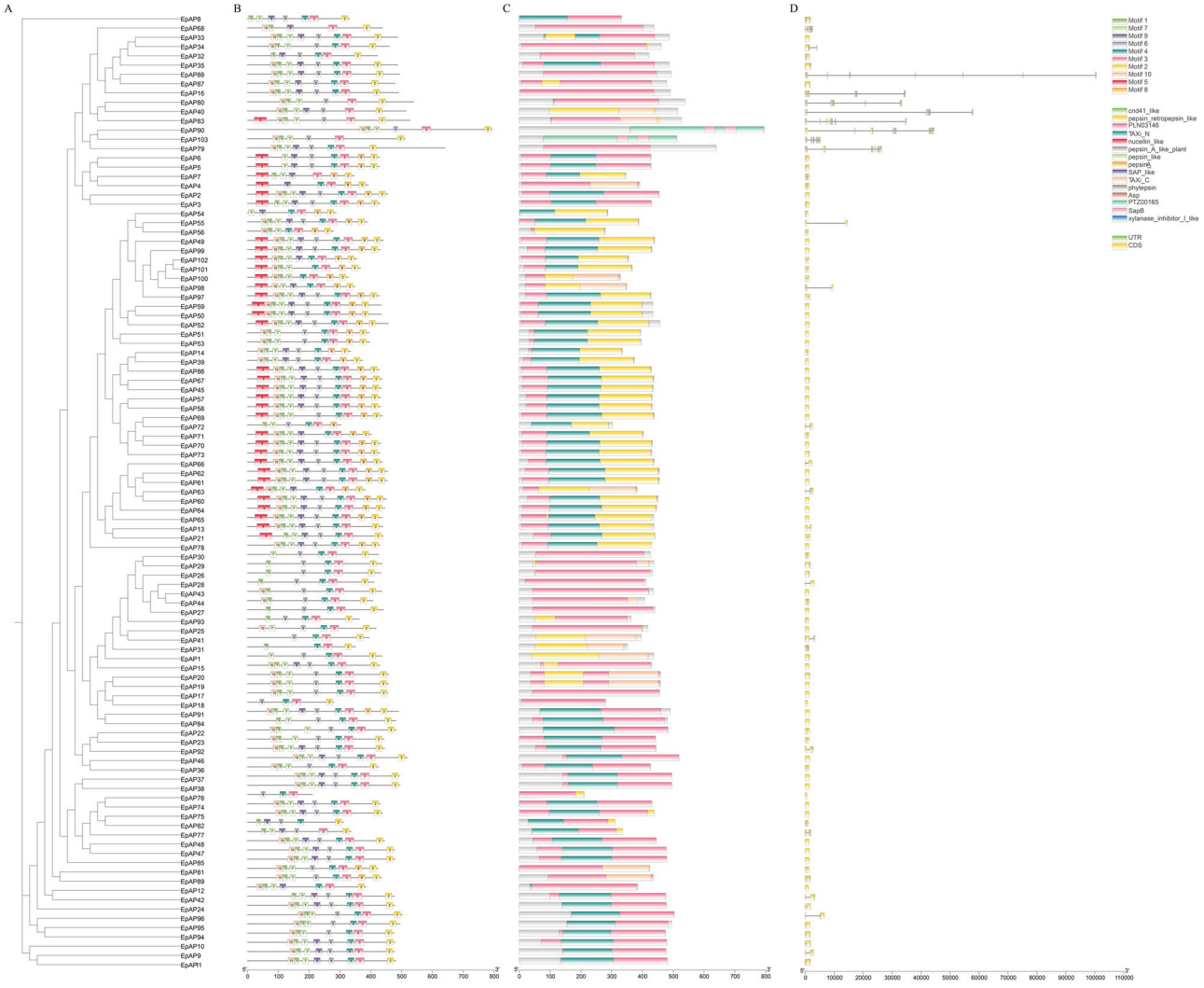

Fig. 2 Phylogenetic tree, gene structure, protein conserved motifs, and conserved domain analysis of the *EpAP* gene family. (**A**) Phylogenetic tree constructed by IQtree in TBtools with 1000 bootstrap number. (**B**) Analysis of motifs in the proteins encoded by *EpAP* genes. (**C**) Conserved domains of *EpAPs* predicted by Batch CD-Search in NCBI. (**D**) Gene structure of *EpAP* genes

indicating that AP genes exhibit a high degree of functional conservation in core eudicots (Fig. 4B). In particular, the *Ka/Ks* ratio between *EpAP103* and *AT1G11910.2* (*APA1*) was only 0.105. A total of 20 pairs of collinear AP genes were identified between *E. pubescens* and *Coptis chinensis* (Table S5). The *Ka/Ks* ratios of all gene pairs were significantly lower than 1 (range: 0.076–0.381), indicating that these genes were subjected to strong purifying selection during evolution. Among them, *EpAP37* (*Ka/Ks* = 0.077) and *EpAP38* (*Ka/Ks* = 0.076) showed extremely low ratios, suggesting that they may have assumed highly conserved physiological functions after the two species diverged.

Identification of *Cis*-elements

Promoter analysis of the *EpAP*s revealed diverse *cis*-elements associated with environmental and hormonal responsiveness within the 2000-bp upstream regions of their transcription start sites (Fig. 5; Table S6). Notably, all 103 *EpAP*s contained light-responsive elements in their promoter sequences. Other commonly identified elements included the anaerobic response element (present in 97 genes), the low-temperature response element (68 genes), and TC-rich repeats associated with defense and stress reactions (41 genes). In addition, elements related to hormonal regulation were also prevalent, with 79 genes containing methyl jasmonate-responsive motifs such as CGTCA and TGACG. These findings suggest that *EpAP* genes may play roles in the responses to environmental stressors and regulation of hormonal signaling pathways.

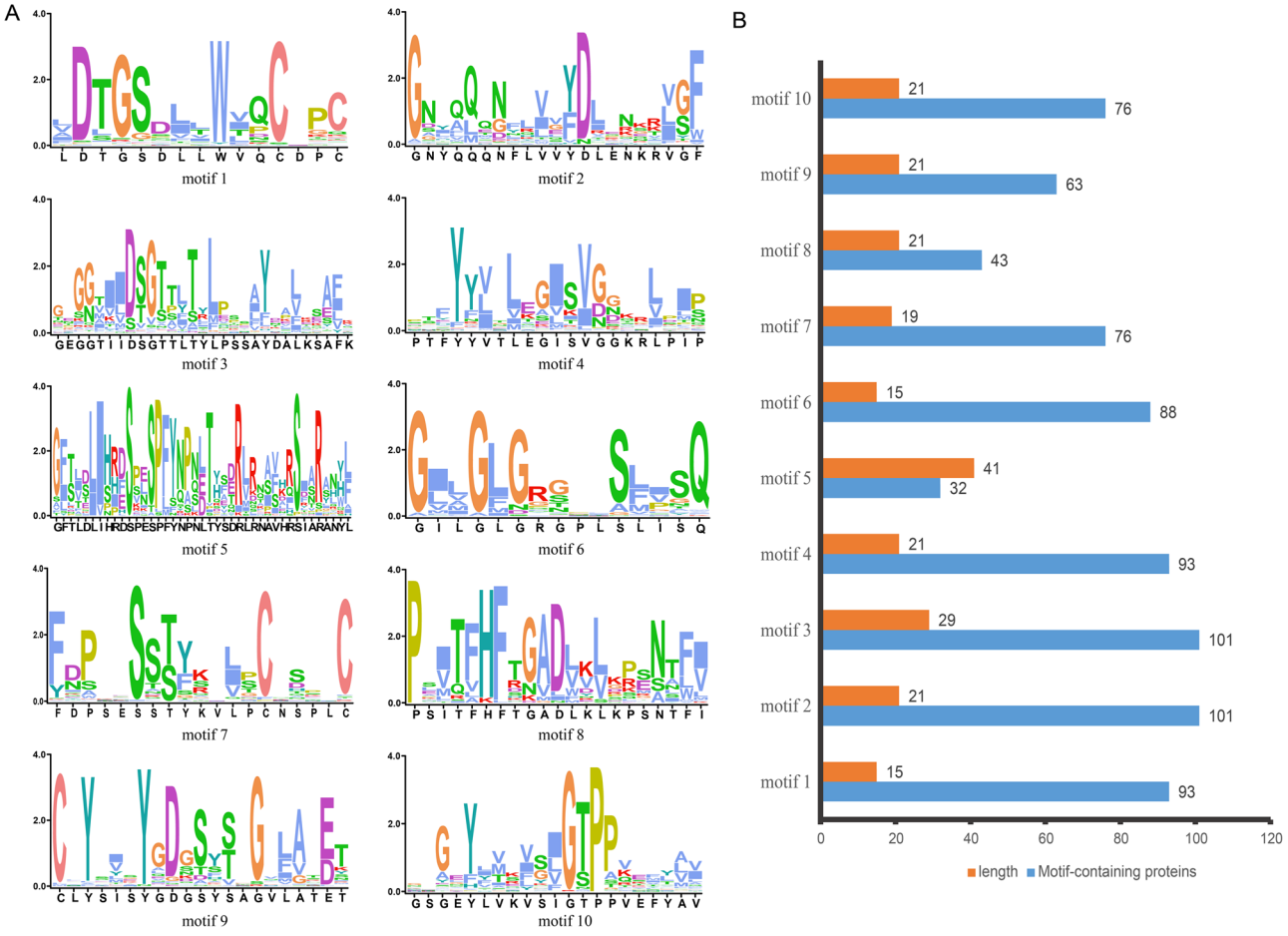

Fig. 3 Motif expression pattern in EpAP proteins. (**A**) Conserved motif logos by TBtools. (**B**) Length and distribution of motif

Expression patterns of *EpAP* genes

The expression profiles of *EpAP*s across various tissues and during the formation of abscission zones indicated their involvement in distinct biological functions. Transcriptome data revealed that 12 *EpAP*s showed no detectable expression in any tissue, while other genes exhibited tissue-specific expression patterns or high expression levels (Fig. 6). For instance, *EpAP54*, *EpAP87*, and *EpAP95* were predominantly expressed in the flowers, suggesting roles in reproductive development. *EpAP40* and *EpAP90* showed high expression levels in fruit, suggesting functions in fruit maturation or stress adaptation. Leaf-specific expression was observed for *EpAP2*, *EpAP5*, and *EpAP27*, while *EpAP30*, *EpAP47*, and *EpAP43* were highly expressed in the roots, which suggests a potential role in nutrient uptake or environmental responses. Additionally, distinct expression patterns were observed in the three stages of abscission zone formation during fruit development (Fig. 7). Genes such as *EpAP17*, *EpAP45*, *EpAP49*, and *EpAP59* were highly expressed prior to abscission, which is indicative of their involvement in the early development of abscission zones. In contrast, *EpAP15*, *EpAP46*, *EpAP68*, and *EpAP73* were predominantly expressed during abscission, indicating a role in the active separation process. Following fruit detachment, genes like *EpAP16*, *EpAP70*, *EpAP78*, and *EpAP93* showed elevated expression, which highlights their roles in cellular reorganization or recovery post-abscission.

Expression analysis of *EpAP*s at different periods after light stress

Given the abundance of light-responsive *cis*-acting elements in the promoter regions of *EpAP*s, it was hypothesized that these proteases might be involved in the plant's response to light stress. To explore gene expression patterns under light stress, real-time quantitative PCR (qPCR) was performed for 12 specific *EpAP*s (*EpAP9*, *EpAP11*, *EpAP19*, *EpAP20*, *EpAP37*, *EpAP38*, *EpAP42*, *EpAP68*, *EpAP78*, *EpAP85*, *EpAP88*, and *EpAP90*). These genes were selected based on the results of subcellular localization predictions, which suggested their localization to chloroplast, making them relevant candidates for studying responses to light stress. The qPCR results

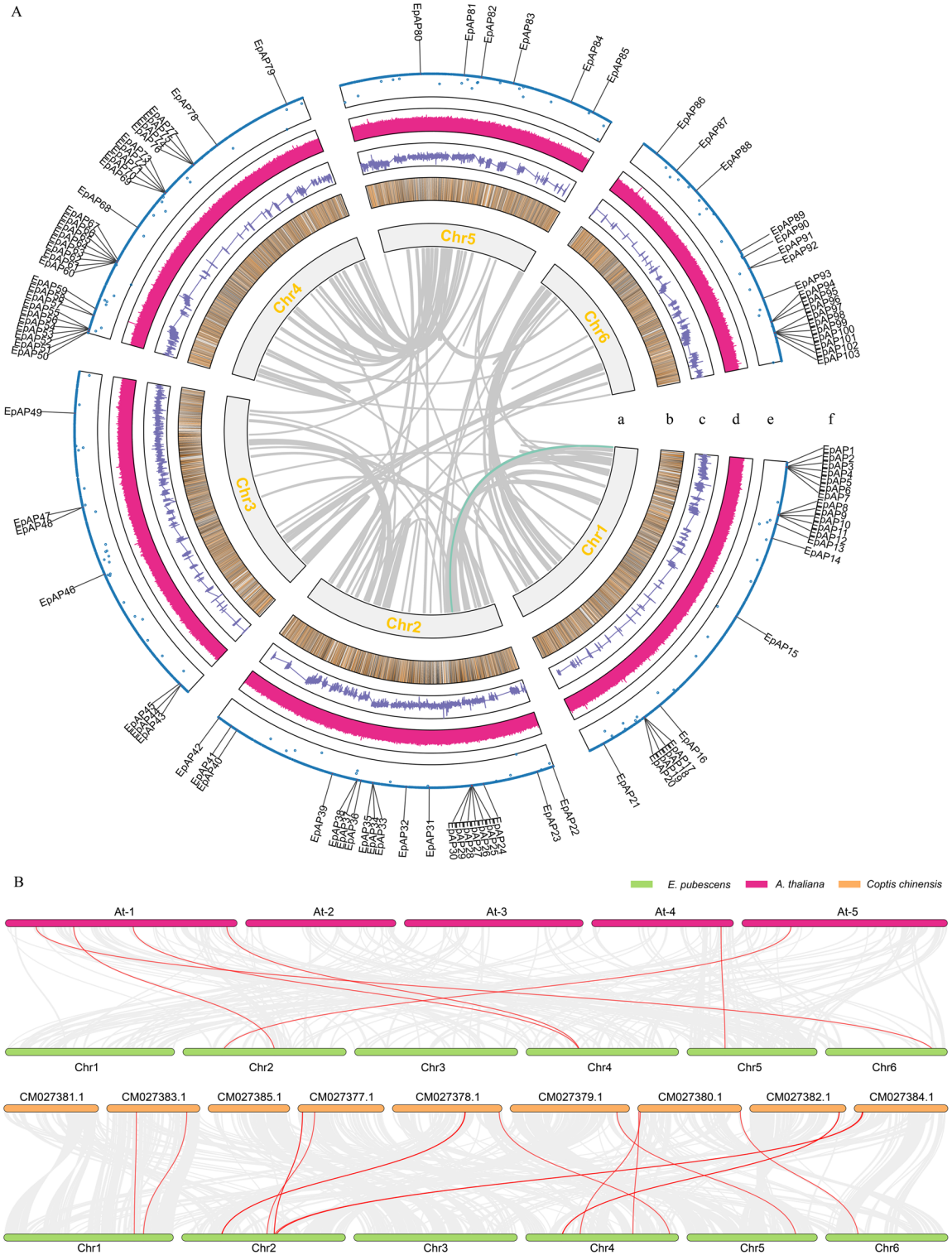

Fig. 4 Distribution of 103 EpAP genes on chromosomes and collinearity analysis. (**A**) EpAP genes on chromosome scaffolds visualized by TBtools v2.110. The central gray lines represent the collinearity of the *E. pubescens* genes, whereas the cyan line indicate the collinearity of the EpAP gene. a, chromosome scaffolds; b, gene density; c, GC ratio; d, GC skew; e, N ratio; f, localization of EpAPs. (**B**) Synteny analysis of *APs* between *E. pubescens* and *A. thaliana* and *Coptis chinensis*

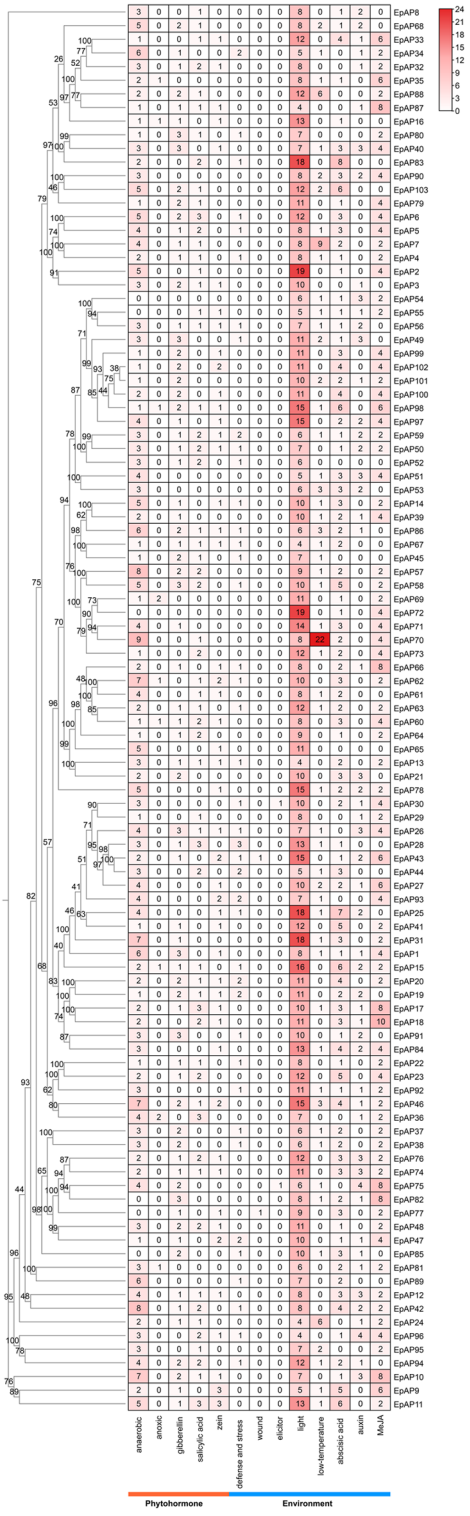

Fig. 5 *Cis*-element analysis of *EpAP* genes. This analysis was conducted on the promoter regions of 2000 bp upstream of 103 EpAP genes to predict the types and numbers of *cis*-elements

are shown in Fig. 8. *EpAP42* showed no response to light stress, with its expression remaining stable throughout the entire period. *EpAP9* exhibited an increased expression at L3 and L6, followed by a decrease back to the initial level. The other genes showed a significant reduction in expression at L3. Among them, *EpAP11*, *EpAP19*, *EpAP20*, *EpAP68*, and *EpAP78* displayed increased expression levels after 9 h of light stress (L9), but these decreased thereafter. However, only *EpAP11* and *EpAP19* exhibited expression levels at L9 that surpassed their pre-stress levels at L0. The expression levels of the other genes, while higher than those at L3, remained below or at the same level as before the treatment. Both *EpAP85* and *EpAP90* showed sustained suppression of expression under light stress.

Prediction of Protein–Protein interactions (PPIs)

The observed PPIs highlighted the potential roles of EpAP proteins in critical biological processes (Fig. 9). For instance, EpAP64 and EpAP102 were predicted to interact with Egg cell-secreted protein 1.5 (EC1.5), a protein that mediates gamete interactions during double fertilization, suggesting their potential involvement in reproductive success. Several EpAPs, including EpAP16, EpAP46, and EpAP92, were predicted to interact with RDUF1 and RDUF2 (E3 ubiquitin-protein ligases), which are implicated in protein degradation pathways essential for stress responses and development. Additionally, EpAP20, EpAP24, and other EpAP proteins were predicted to interact with cathepsin B-like protease 3 (CATHB3), a key regulator of PCD in plants. These findings collectively indicate that EpAP proteins have diverse functional roles in development, stress adaptation, and cell death regulation.

Discussion

Plant APs are critical enzymes involved in numerous biological processes, such as development, pathogen defense, and stress responses [38, 39]. Despite extensive research being conducted on APs in model plants, genome-wide identification and functional analysis of the AP gene family in *E. pubescens* have remained unexplored. The present study comprehensively analyzed the AP genes in *E. pubescens*, providing insights into their structural diversity, evolutionary conservation, and potential functional roles.

Tandem duplications are the main driver of *EpAP* gene expansion

In this study, a total of 103 *EpAP*s were identified and classified into four phylogenetic clades. Notably, proteins EpAP90 and EpAP103 in Group I were shown to contain Saposin-like domains, which play multifaceted roles in plant APs, primarily in intracellular sorting, host

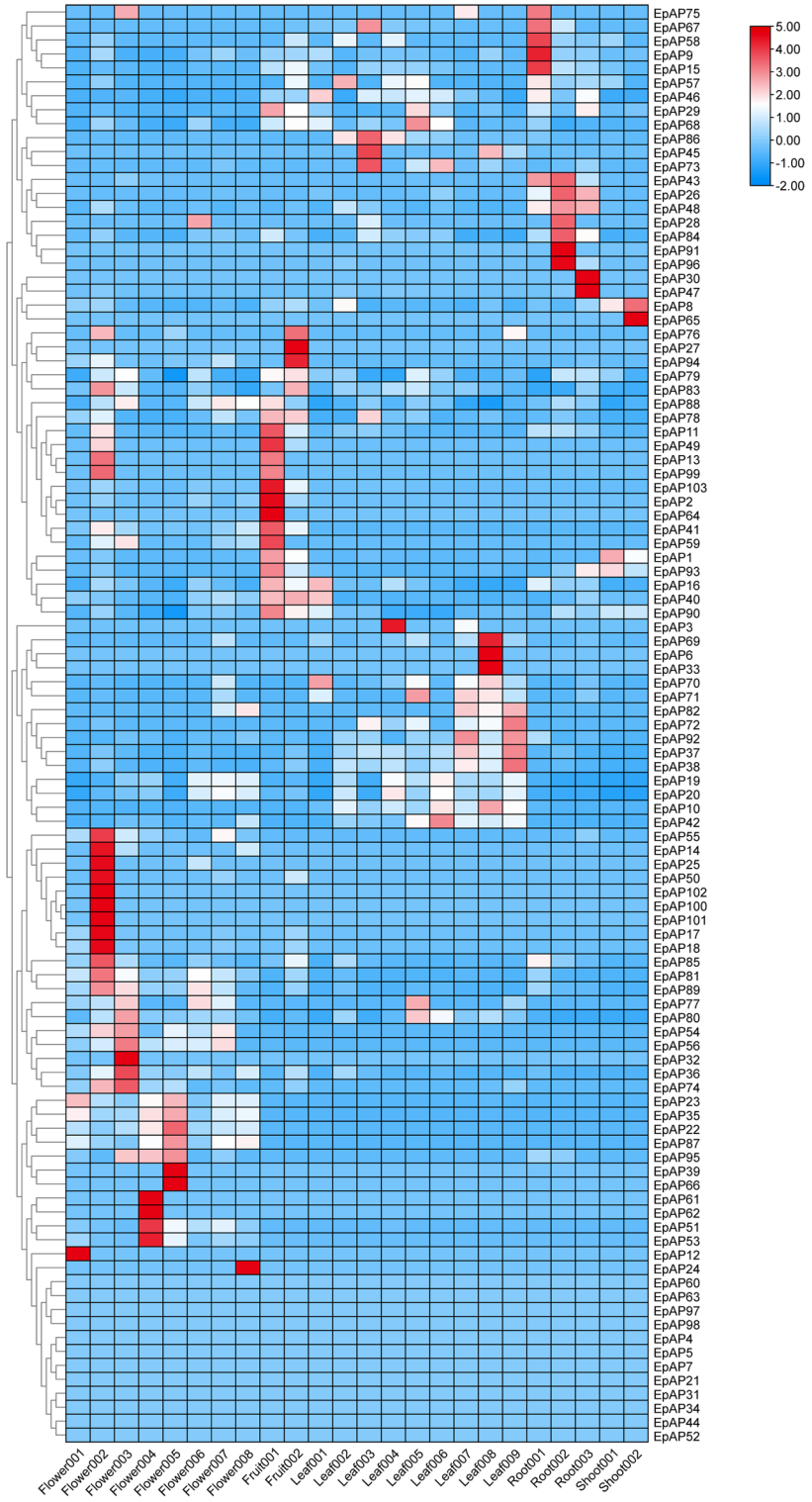

Fig. 6 Expression patterns of *EpAPs* in different tissues of *E. pubescens*. The data were obtained from the transcriptome of fully developed leaves, roots, shoots, leaves, flowers, and fruits [56]. The color scale illustrates log2 expression values, ranging from blue for low expression levels to red for high expression levels

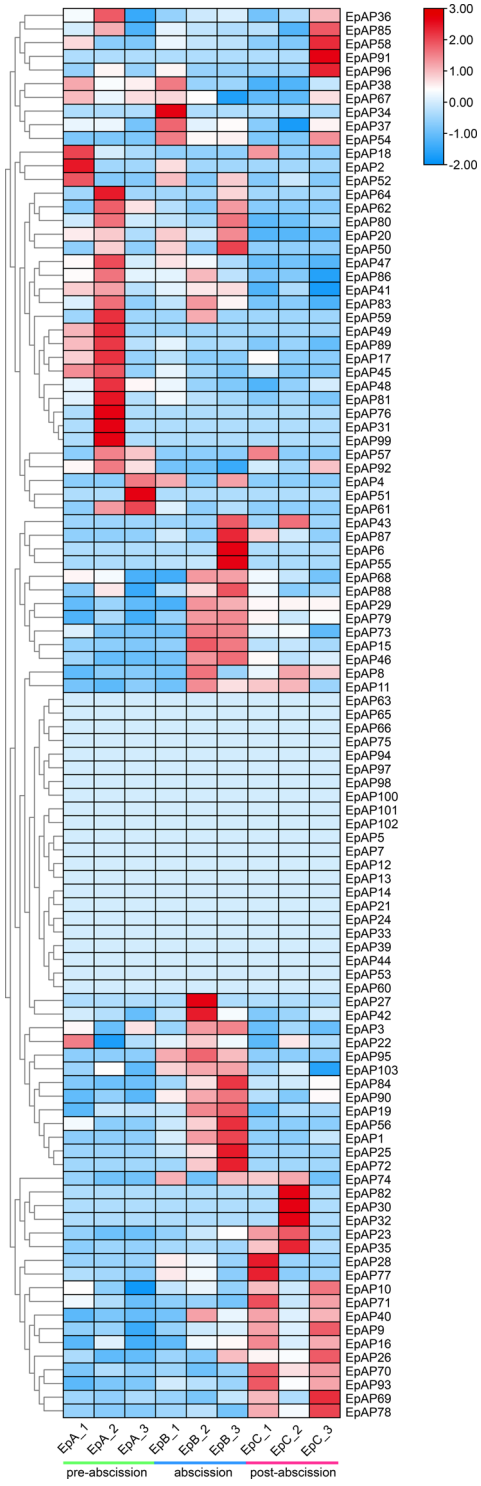

Fig. 7 Expression patterns of *EpAP*s in abscission zone of *E. pubescens*. EpA, pre-abscission stage (before abscission zone formation); EpB, abscission stage (abscission zone formation); and EpC, post-abscission stage (after abscission zone formation and pod detachment). The color scale illustrates log2 expression values, ranging from blue for low expression levels to red for high expression levels

defense, and membrane interactions [40, 41]. Additionally, consistent with previous findings in *A. thaliana* [9], Moso bamboo [15], and potato (*Solanum tuberosum*) [42], some EpAP proteins possessed signal peptides and transmembrane domains. Interestingly, except for EpAP24 in Group III, all EpAP proteins with transmembrane domains (EpAP16, EpAP35, EpAP40, EpAP79, and EpAP80) belonged to Group II. Furthermore, most EpAP proteins contained xylanase inhibitor domains (TAXI_C and TAXI_N), which overlap with the Asp domain, consistent with previous reports [15, 42].

In *E. pubescens*, the expansion of the AP gene family is primarily driven by tandem duplication, which is particularly prominent in the members of Group III. Notably, the expansion patterns of AP gene families vary significantly among plant species. For example, in *P. trichocarpa* [14] and *P. edulis* [15], segmental duplication is the primary driver of AP gene family expansion, whereas in *S. tuberosum*, tandem duplication is more prevalent [42]. This difference may reflect distinct evolutionary strategies for environmental adaptation. As a perennial herbaceous plant adapted to shaded environments, *E. pubescens* may favor rapid, localized gene duplication through tandem repeats to enhance its adaptability. Tandem duplications often facilitate the rapid expansion of gene families within a short evolutionary timespan, forming a flexible and efficient regulatory network in response to external stimuli such as light stress and pathogen invasion [43, 44].

Additionally, the numbers of introns in the *EpAP*s examined exhibited a distinct pattern across different phylogenetic groups, with an average of 13 in Group I, 5.75 in Group II, 0.65 in Group III, and 1 in Group IV. This distribution pattern is generally consistent with APs in other plant species [11]. Notably, it has previously been shown that 46.6% (48/103) of the AP genes in *E. pubescens* are intronless, a proportion comparable to that observed in *S. tuberosum* (43.5%, 27/62) [42]. This simplified gene structure may confer an adaptive advantage to *E. pubescens*, allowing it to respond swiftly to environmental changes and maintain strong competitiveness under variable ecological conditions.

*EpAP*s are enriched in light-responsive elements

It has been shown that the promoters of all *EpAP*s are enriched with light-responsive elements, including the G-box, GT1-motif, and GATA-motif [45, 46, 47]. As a shade-loving plant, *E. pubescens* likely employs *EpAP*s to mediate responses to light stress. Studies on other species have demonstrated the critical roles of APs in chloroplast function and leaf senescence [18, 48, 49, 50]. For example, the chloroplast-localized protein CND41 in tobacco is linked to the degradation of Rubisco and exhibits a negative correlation with transcriptional activity within

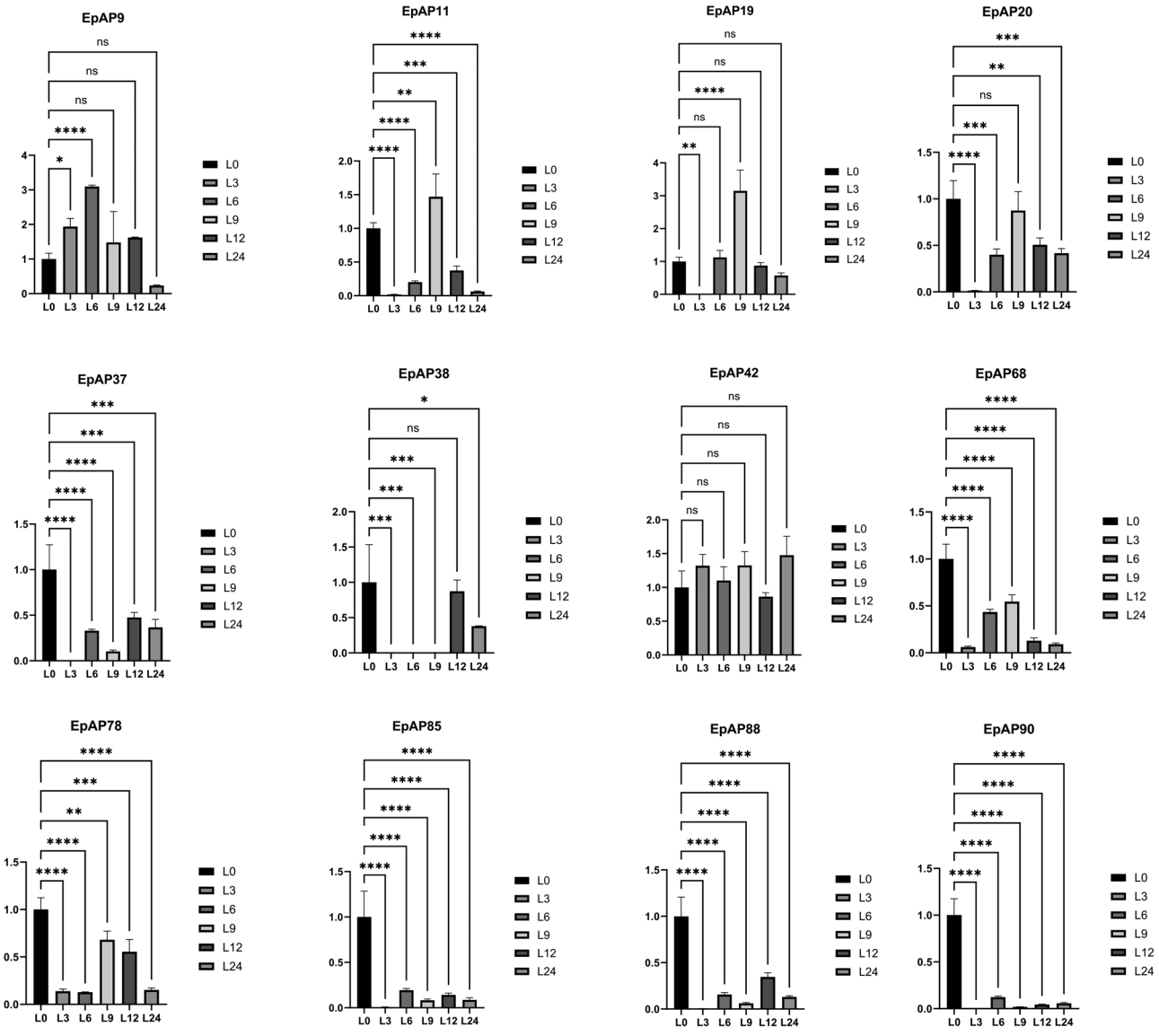

Fig. 8 qPCR result of 12 selected *EpAP* genes. Error bars indicate Standard Error (SE) from three technical replicates. Asterisks or n.s. indicate significant differences between different time point, *$p < 0.05$, **$p < 0{,}01$, **** $P < 0.0001$, ns, no significant difference, by One-way ANOVAs

chloroplasts [18, 51]. In *S. tuberosum*, the canonical AP SPAP1 protein participates in ethylene-mediated leaf senescence, functioning in coordination with multiple members of the ethylene signaling pathway [52]. In addition, the atypical AP NANA, which is localized in the chloroplast in *Arabidopsis*, plays a crucial role in maintaining chloroplast homeostasis and starch metabolism. Its precise substrate and mechanisms of action remain unclear, but it has been shown that its loss reduces chlorophyll content, impairs electron transport, and alters sugar metabolism [36]. The presence of light-responsive elements in *EpAP* promoters suggests potential involvement in these pathways, but further functional validation studies are needed to elucidate specific regulatory networks.

Tissue-specific expression and abscission zone formation

The *EpAP*s examined in this study exhibited dynamic expression patterns across various plant tissues and participated in multiple biological processes. Tissue-specific analysis showed that *EpAP102* and *EpAP64* were predominantly expressed in the flowers and fruit, and their predicted interaction with EC1.5 suggests functions in reproductive development [53]. Similarly, the significant upregulation of *EpAP15* and *EpAP46* during the formation of fruit abscission zones indicates the involvement of these genes in the regulation of organ abscission. Such role has been suggested by some studies, although this remains a poorly documented subject. For instance, in a proteome analysis of ethylene-induced pedicel abscission in tomato, AP 1 (F2VJ74) showed significant changes in

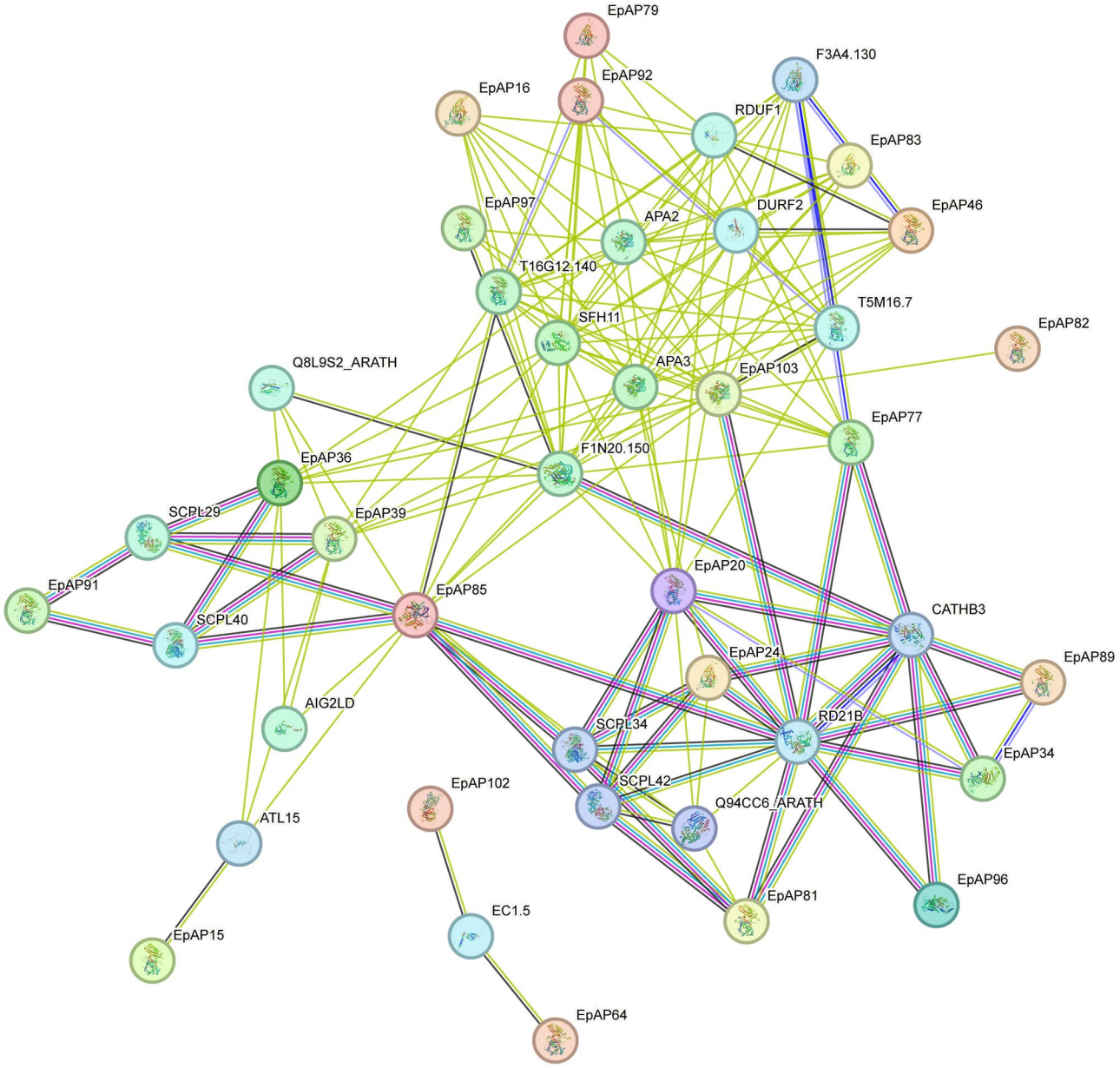

Fig. 9 Prediction of the protein–protein interaction network of EpAP proteins via STRING. The interaction score was set to medium confidence 0.400

abundance during the formation of the pedicel abscission layer [54]. Additionally, a transcriptomic analysis of the sepal abscission zone (AZ-C) in 'Hamlin' oranges identified 16 differentially expressed *APs* by comparing fruit that detached from a Huanglongbing-infected tree upon shaking with those that remained attached to the infected tree [55]. Furthermore, in the present study, PPI analysis predicted interactions between several EpAP proteins and key regulatory proteins. For example, EpAP20/24/34 were predicted to interact with the PCD-related protease CATHB3 [53], potentially contributing to cell separation in the abscission zone via substrate hydrolysis. Additionally, the observed interactions with the E3 ubiquitin-protein ligases RDUF1 and DURF2 suggest a role in stress responses, including those to drought and salinity. Collectively, these interactions highlight the multifaceted roles of *EpAP*s in maintaining cellular homeostasis and facilitating stress adaptation. However, it is important to note that PPIs can vary across species, and interactions predicted based on model plants may not always directly apply to *E. pubescens*. The PPI results obtained in this study require experimental validation, for example via yeast two-hybrid assays, co-immunoprecipitation, or bimolecular fluorescence complementation, to confirm the validity of predictions.

*EpAP*s participate in light stress responses

Consistent with the presence of light-responsive *cis*-elements in their promoters, several *EpAP*s showed dynamic changes in expression in response to different periods of exposure to light stress. *EpAP42* exhibited a stable expression pattern throughout the entire treatment period, indicating that this gene is not responsive to light stress. *EpAP9* showed a transient increase in expression at the early stages of light stress (L3 and L6), followed by a return to the initial levels. This transient upregulation may reflect a short-term adaptation or regulatory response during initial acclimation to varying light conditions. Most of the tested genes, including *EpAP11*, *EpAP19*, *EpAP20*, *EpAP68*, and *EpAP78*, displayed a significant decrease in expression at L3, followed by a rebound at L9. Notably, only *EpAP11* and *EpAP19* surpassed their baseline expression levels at L0 during this recovery phase, which indicates their potential roles in sustained light stress adaptation or recovery processes. The partial recovery of expression levels observed for the other genes suggests that they may function during specific stages of the light stress response while remaining inactive for prolonged periods.

To further elucidate the roles of *EpAP*s, it is necessary to conduct functional validation studies using reverse genetics approaches, such as RNA interference or CRISPR/Cas9. Investigations into protein targets and interaction networks will also enhance our understanding of how these proteases contribute to light stress adaptation. Additionally, examining their expression under other abiotic stresses, such as drought or temperature extremes, could reveal shared or stress-specific regulatory pathways. Considering that *E. pubescens* is a shade-adapted plant, understanding how *EpAP*s respond to different light intensities could offer valuable insights for the optimization of its growth conditions. Targeted regulation of these genes to enhance light tolerance may assist in balancing biomass accumulation and the biosynthesis of secondary metabolites, such as flavonoids. Future studies could explore genetic or agronomic approaches, such as the overexpression of specific *EpAP*s or the adjustment of light exposure strategies, to improve both stress resilience and the yield of medicinal compounds.

Conclusion

This study represents the first comprehensive genome-wide analysis of the AP gene family in *E. pubescens*. It identified 103 *EpAP* genes and provided insights into their structural diversity, evolutionary conservation, and potential functional roles. These genes were shown to exhibit significant phylogenetic diversity as well as diverse structural domains and duplication patterns, which reflects their evolutionary adaptations to various physiological and environmental challenges. Notably, the presence of light-responsive *cis*-elements in *EpAP* promoters highlighted their potential involvement in light stress responses. Furthermore, tissue-specific expression patterns and predicted interactions with key regulatory proteins suggested roles in reproductive development, PCD, and abscission zone formation. Our findings establish a foundation for future functional studies aiming to explore the precise roles of *EpAP*s, potentially contributing to improving stress tolerance and reproductive efficiency in crops.

Methods

Identification and physicochemical characterization of epap proteins

The AP genes in *E. pubescens* were identified using a comprehensive approach combining HMMER v3.4 and BLASTP in order to ensure both analytical sensitivity and specificity. The hidden Markov model profile for the AP domain (PF00026) from the Pfam database was used as a query to search the *E. pubescens* genome (downloaded from the GigaDB database [56]) in HMMER v3.4, which effectively detects conserved domains even in divergent sequences. Protein sequences with an expected value (E) < 1E − 5 were retained. In addition, 70 AP protein sequences from *A. thaliana* [11] were used as queries in BLASTP (E < 1E − 5) to identify potential *EpAP*s based on sequence similarity. To improve accuracy, the sequences retrieved using both methods were merged, and duplicates were removed. NCBI CD-Search [57], SMART [58], and InterProScan [59] were employed to further validate the identified AP genes and minimize false positives, which allowed to confirm the presence of conserved domains and structural integrity. This multi-step approach ensured a robust and reliable identification of the *EpAP* gene family. The physicochemical properties of EpAP proteins were predicted in ProtParam [60]. Subcellular localization was predicted using the WoLF PSORT [61] and DeepLoc 2.0 [62] web tools. Additionally, TargetP 2.0 [63] and DeepLoc 2.0 were used to predict the presence and location of signal peptides in EpAP protein sequences.

Phylogenetic analysis

The 103 *EpAPs* identified in this study were aligned with 70 *AtAPs* in MUSCLE v5.0, and the data were subsequently refined using TrimAl. A maximum-likelihood phylogenetic tree was constructed using IQ-tree with 1,000 bootstrap replicates in TBtools v2.110 [64] and visualized via iTOL (https://itol.embl.de/) [65].

Analysis of gene structure, motifs, and domains

Exon-intron positions were extracted from the GFF annotation file obtained for the *E. pubescens* genome. The structural domains of *EpAPs* were analyzed using the

NCBI Conserved Domain Database (https://www.ncbi.nlm.nih.gov/cdd/). Gene structures and protein domains were visualized using TBtools.

The conserved motifs in the *EpAP*s were predicted using the MEME online tool (http://meme-suite.org/tools/meme), with the predicted number of motifs set to 10. Gene and motif structures were mapped and visualized using TBtools.

Cis-elements in the promoter regions of *EpAPs*

To identify conserved *cis*-elements in the promoter regions of *EpAPs*, a screening was conducted of the sequences located 2000 bp upstream of the initiation codons, which represent the proximal promoter regions. Specifically, these sequences were analyzed using the PlantCARE database to predict promoter regions (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) [66]. Subsequently, the functional elements were categorized, and the results were visualized using TBtools.

Chromosomal locations, genomic duplications, and Ka/ks ratios

The chromosomal locations of *EpAPs* were determined using CDS coordinates from the *E. pubescens* genome. The genes were mapped onto their respective chromosomes based on their physical location and visualized using TBtools. Gene duplications within the *EpAP* gene family were illustrated using the Advanced Circos tool. One Step MCScanX was applied to identify homologous regions between *E. pubescens* and *A. thaliana* and between *E. pubescens* and *C. chinensis*. The *Ka/Ks* ratios were calculated to assess homologous relationships and estimating the divergence times of *EpAP*s. Homologous gene pairs were identified in BLAST, and their *Ka/Ks* ratios were computed using KaKs_Calculator. Based on the absolute molecular evolutionary rate (r) for Ranunculales plants (i.e., 6.98×10^{-9}), the origin time of each duplication event was calculated using the formula $T = Ks / (2r)$. The above analyses were all conducted using TBtools.

Expression patterns of *EpAP*s

To investigate the expression patterns of *EpAP*s, RNA-seq data (leaf: SRR15561380, SRR15562432, SRR15562431, SRR15562430, SRR15562429, SRR15562521, SRR15562520, SRR15562519, SRR15562518; Flower: SRR15569255, SRR15569254, SRR15569253, SRR15569252, SRR15569236, SRR15569235, SRR15569234, SRR15569232; fruit: SRR15569250, SRR15569249; root: SRR15561383, SRR15561382, SRR15561381; shoot: SRR15569248, SRR15569247) were obtained from the NCBI Sequence Read Archive database (Table S7) [56]. The transcriptome data related to fruit abscission in *E. pubescens* were obtained from the CNCB database (GSA number: CRA022394).

Predictive analysis of PPI networks

The EpAP protein sequences were analyzed on the STRING web platform (https://string-db.org/) using *A. thaliana* as a reference. The minimum required interaction score was set to 0.400, and the maximum number of interactors in the first shell was set to 20. Disconnected nodes were hidden, and the other parameters were maintained as default.

Plant materials and light stress treatments

In June 2024, 1-month-old seedlings of *E. pubescens* were obtained from the Xianling Medicinal Valley Yinyanghuo Cultivation Base (29°20 × 41.75′′ N, 103°33 × 39.60′′ E) in Shawan District, Leshan, Sichuan Province, and identified based on their morphological characteristics using the iPlant database (https://www.iplant.cn/info/Epimedium%20pubescens). The seedlings were then transferred to a greenhouse at the Southwest Characteristic Economic Plant Hybridization and Breeding Research Center of Leshan Normal University. Uniformly growing seedlings were selected and planted in pots measuring 11 cm in height, 11.5 cm in top diameter, and 8 cm in bottom diameter. The growth medium consisted of a 1:3 mixture of vermiculite and nutrient soil. Pre-treatment growth conditions were maintained at 22 °C under a light intensity of 25 $\mu mol \cdot m^{-2} \cdot s^{-1}$. After an acclimation period of 3 weeks, the seedlings were exposed to light stress (330 $\mu mol \cdot m^{-2} \cdot s^{-1}$) for 12 h. Leaf samples were collected at 0 h (pre-treatment), 3 h, 6 h, 9 h, and 12 h post-treatment. These samples were subsequently flash-frozen in liquid nitrogen and stored at – 80 °C.

RNA extraction and qPCR

Total RNA was isolated using the FastPure® Plant Total RNA Isolation Kit (Cat#RC401-01, Vazyme, Nanjing, China). cDNA was synthesized using the ABScript III RT Master Mix for qPCR with gDNA Remover (Cat#RK20428, ABclonal, Wuhan, China). qPCR amplification was conducted using the ChamQ Universal SYBR qPCR Master Mix kit (Cat#Q711-02, Vazyme, Nanjing, China), with *Actin* as the internal control. The qPCR primer sequences are listed in Table S8. The amplification protocol consisted of an initial denaturation step at 95 °C for 30 s, followed by 40 cycles at 95 °C for 10 s and at 60 °C for 30 s, with the melting curve being automatically generated. Relative gene expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method [67], with three technical replicates per sample. Data analysis and visualization were performed using GraphPad 10. Statistical significance was assessed using one-way ANOVA followed by Dunnett's multiple comparison test to determine differences

between time points at specific light stress conditions. qPCR data were normalized based on the relative expression of *EpAP*s at time point L0.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11617-7.

Supplementary Material 1

Supplementary Material 2

Acknowledgements
We would like to thank the Xianling Medicinal Valley Yinyanghuo Cultivation Base, Shawan District, Leshan City for providing the experimental materials of E. pubescens.

Author contributions
HZ: conceptualization, writing– original draft, project administration. LZ: data curation, methodology. HS: formal analysis, visualization. XY: writing– review & editing, funding acquisition. All authors reviewed the manuscript and approved the final manuscript for publication.

Funding
The Leshan Normal University Scientific Research Start-up Project for Introducing High-level Talents (No. RC2023019), Key scientific research project of Leshan Science and Technology Bureau (No. 24SZD023).

Data availability
Genome files of ***A. thaliana*** were obtained from Phytozome (https://phytozome-next.jgi.doe.gov/), genome files of ***E. pubescens*** were obtained from GigaDB (https://ngdc.cncb.ac.cn/gwh/Assembly/21845/show). Tissue-specific expression RNA-seq data were obtained from the NCBI SRA database (Accession number: SRR15561380, SRR15561381, SRR15561382, SRR15561383, SRR15562429, SRR15562430, SRR15562431, SRR15562432, SRR15562518, SRR15562519, SRR15562520, SRR15562521, SRR15569232, SRR15569252, SRR15569250, SRR15569249, SRR15569248, SRR15569247, SRR15569236, SRR15569235, SRR15569234, SRR15569253, SRR15569254, SRR15569255). The transcriptome data of ***E. pubescens*** pod abscission have been deposited in the China National Center for Bioinformation (CNCB) under GSA number CRA022394, the expression data of ***EpAPs*** in the ***E. pubescens*** abscission layer is provided in Table S9.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 22 January 2025 / Accepted: 21 April 2025
Published online: 02 May 2025

References

1. Davies DR. The structure and function of the aspartic proteinases. Annual Rev Biophys Biophys Chem. 1990;19:189–215.
2. Rawlings ND, Barrett AJ. Families of aspartic peptidases, and those of unknown catalytic mechanism. Methods Enzymol. 1995;248:105–20.
3. Mutlu A, Susannah G. Plant aspartic proteinases: enzymes on the way to a function. Physiol Plant. 1999;105(3):569–76.
4. Simões I, Faro C. Structure and function of plant aspartic proteinases. Eur J Biochem. 2004;271(11):2067–75.
5. Rawlings ND, Morton FR, Kok CY, Kong J, Barrett AJ. MEROPS: the peptidase database. Nucleic Acids Res. 2010;38(SUPPL 1):D227–233.
6. Dunn BM. Structure and mechanism of the Pepsin-like family of aspartic peptidases. Chem Rev. 2002;102(12):4431–58.
7. Simoes I, Faro R, Bur D, Faro C. Characterization of Recombinant CDR1, an *Arabidopsis* aspartic proteinase involved in disease resistance. J Biol Chem. 2007;282(43):31358–65.
8. Beers EP, Jones AM, Dickerman AW. The S8 Serine, C1A cysteine and A1 aspartic protease families in *Arabidopsis*. Phytochemistry. 2004;65(1):43–58.
9. Faro C, Gal S. Aspartic proteinase content of the *Arabidopsis* genome. Curr Protein Pept Sci. 2005;6(6):493–500.
10. Takahashi K, Niwa H, Yokota N, Kubota K, Inoue H. Widespread tissue expression of nepenthesin-like aspartic protease genes in *Arabidopsis thaliana*. Plant Physiol Biochem. 2008;46(7):724–9.
11. Duan Y, Tang H, Yu X. Phylogenetic and alphafold predicted structure analyses provide insights for A1 aspartic protease family classification in *Arabidopsis*. Front Plant Sci. 2023;14:1072168.
12. Gull S, Uddin S, Altaf A, Chen K, Jiang Y, Tang X, Yang C, Chen L, Wang S, Liu J, et al. Characterization of rice aspartic protease genes and induced expression by phytohormones and *Xanthomonas oryzae* Pv. oryzae. Volume 130. Physiological and Molecular Plant Pathology; 2024. p. 102220.
13. Guo R, Xu X, Carole B, Li X, Gao M, Zheng Y, Wang X. Genome-wide identification, evolutionary and expression analysis of the aspartic protease gene superfamily in grape. BMC Genomics. 2013;14:554.
14. Cao S, Guo M, Wang C, Xu W, Shi T, Tong G, Zhen C, Cheng H, Yang C, Elsheery NI, et al. Genome-wide characterization of aspartic protease (AP) gene family in *Populus trichocarpa* and identification of the potential PtAPs involved in wood formation. BMC Plant Biol. 2019;19(1):276.
15. Wang X, Yan X, Li S, Jing Y, Gu L, Zou S, Zhang J, Liu B. Genome-wide identification, evolution and expression analysis of the aspartic protease gene family during rapid growth of Moso bamboo (*Phyllostachys edulis*) shoots. BMC Genomics. 2021;22(1):45.
16. Runeberg-Roos P, Saarma M. Phytepsin, a barley vacuolar aspartic proteinase, is highly expressed during autolysis of developing tracheary elements and sieve cells. Plant J. 1998;15(1):139–45.
17. Kervinen J, Tobin GJ, Costa J, Waugh DS, Wlodawer A, Zdanov A. Crystal structure of plant aspartic proteinase prophytepsin: inactivation and vacuolar targeting. EMBO J. 1999;18(14):3947–55.
18. Kato Y, Murakami S, Yamamoto Y, Chatani H, Kondo Y, Nakano T, Yokota A, Sato F. The DNA-binding protease, CND41, and the degradation of ribulose-1,5-bisphosphate carboxylase/oxygenase in senescent leaves of tobacco. Planta. 2004;220(1):97–104.
19. Shen W, Yao X, Ye T, Ma S, Liu X, Yin X, Wu Y. *Arabidopsis* aspartic protease ASPG1 affects seed dormancy, seed longevity and seed germination. Plant Cell Physiol. 2018;59(7):1415–31.
20. Niu N, Liang W, Yang X, Jin W, Wilson ZA, Hu J, Zhang D. EAT1 promotes tapetal cell death by regulating aspartic proteases during male reproductive development in rice. Nat Commun. 2013;4:1445.
21. Huang J, Zhao X, Cheng K, Jiang Y, Ouyang Y, Xu C, Li X, Xiao J, Zhang Q. OsAP65, a rice aspartic protease, is essential for male fertility and plays a role in pollen germination and pollen tube growth. J Exp Bot. 2013;64(11):3351–60.
22. Ge X, Dietrich C, Matsuno M, Li G, Berg H, Xia Y. An *Arabidopsis* aspartic protease functions as an anti-cell-death component in reproduction and embryogenesis. EMBO Rep. 2005;6(3):282–8.
23. Gao H, Zhang Y, Wang W, Zhao K, Liu C, Bai L, Li R, Guo Y. Two membrane-anchored aspartic proteases contribute to pollen and ovule development. Plant Physiol. 2017;173(1):219–39.
24. Li Y, Kabbage M, Liu W, Dickman MB. Aspartyl protease-mediated cleavage of BAG6 is necessary for autophagy and fungal resistance in plants. Plant Cell. 2016;28(1):233–47.
25. Xia Y, Suzuki H, Borevitz J, Blount J, Guo Z, Patel K, Dixon RA, Lamb C. An extracellular aspartic protease functions in *Arabidopsis* disease resistance signaling. EMBO J. 2004;23(4):980–8.
26. Prasad BD, Creissen G, Lamb C, Chattoo BB. Overexpression of rice (*Oryza sativa* L.) *OsCDR1* leads to constitutive activation of defense responses in rice and *Arabidopsis*. Mol Plant Microbe Interact. 2009;22(12):1635–44.
27. Guo R, Tu M, Wang X, Zhao J, Wan R, Li Z, Wang Y, Wang X. Ectopic expression of a grape aspartic protease gene, AP13, in Arabidopsis thaliana improves resistance to powdery mildew but increases susceptibility to Botrytis cinerea. Plant Sci. 2016;248:17–27.

28. Yao X, Xiong W, Ye T, Wu Y. Overexpression of the aspartic protease *ASPG1* gene confers drought avoidance in *Arabidopsis*. J Exp Bot. 2012;63(7):2579–93.
29. Sebastián DII, Fernando FD, Raúl DG, Gabriela GM. Overexpression of *Arabidopsis* aspartic protease *APA1* gene confers drought tolerance. Plant Sci. 2020;292:110406.
30. de Smet Y, Goetghebeur P, Wanke S, Asselman P, Samain MS. Additional evidence for recent divergence of Chinese *Epimedium* (Berberidaceae) derived from AFLP, Chloroplast and nuclear data supplemented with characterisation of leaflet pubescence. Plant Ecol Evol. 2012;145(1):73–87.
31. Chase MW, Christenhusz MJM, Fay MF, Byng JW, Judd WS, Soltis DE, Mabberley DJ, Sennikov AN, Soltis PS, Stevens PF, et al. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc. 2016;181(1):1–20.
32. Zhuang W, Sun N, Gu C, Liu S, Zheng Y, Wang H, Tong X, Song J. A literature review on *Epimedium*, a medicinal plant with promising slow aging properties. Heliyon. 2023;9(11):e21226.
33. Liang Q, Wei G, Chen J, Wang Y, Huang H. Variation of medicinal components in a unique geographical accession of horny goat weed Epimedium sagittatum Maxim. (Berberidaceae). Molecules. 2012;17(11):13345–56.
34. Pan J, Guo B. Effects of light intensity on the growth, photosynthetic characteristics, and flavonoid content of *Epimedium Pseudowushanense*. B L Guo Molecules. 2016;21(11):1475.
35. MA, W-q. XU C-q, GUO B-l, ZHANG L-j. Effects of different light intensities on growth characteristics and content of icarrin flavonol glycosides of *Epimedium pubescens*. Mod Chin Med. 2023;25(4):847–53.
36. Paparelli E, Gonzali S, Parlanti S, Novi G, Giorgi FM, Licausi F, Kosmacz M, Feil R, Lunn JE, Brust H, et al. Misexpression of a Chloroplast aspartyl protease leads to severe growth defects and alters carbohydrate metabolism in *Arabidopsis*. Plant Physiol. 2012;160(3):1237–50.
37. Niu KX, Chang CY, Zhang MQ, Guo YT, Yan Y, Sun HJ, Zhang GL, Li XM, Gong YL, Ding CH, et al. Suppressing *ASPARTIC PROTEASE 1* prolongs photosynthesis and increases wheat grain weight. Nat Plants. 2023;9(6):965–77.
38. Soares A, Ribeiro Carlton SM, Simoes I. Atypical and nucellin-like aspartic proteases: emerging players in plant developmental processes and stress responses. J Exp Bot. 2019;70(7):2059–76.
39. Figueiredo L, Santos RB, Figueiredo A. Defense and offense strategies: the role of aspartic proteases in Plant-Pathogen interactions. Biology (Basel). 2021;10(2):75.
40. Bryksa BC, Yada RY. Protein structure insights into the bilayer interactions of the Saposin-Like domain of *Solanum tuberosum* aspartic protease. Sci Rep. 2017;7(1):16911.
41. De Moura DC, Bryksa BC, Yada RY. *In Silico* insights into protein-protein interactions and folding dynamics of the saposin-like domain of *Solanum tuberosum* aspartic protease. PLoS ONE. 2014;9(9):e104315.
42. Norero NS, Rey Burusco MF, D'Ippolito S, Decima Oneto CA, Massa GA, Castellote MA, Feingold SE, Guevara MG. Genome-Wide analyses of aspartic proteases on potato genome (*Solanum tuberosum*): generating new tools to improve the resistance of plants to abiotic stress. Plants (Basel). 2022;11(4):544.
43. Picart-Picolo A, Grob S, Picault N, Franek M, Llauro C, Halter T, Maier TR, Jobet E, Descombin J, Zhang P, et al. Large tandem duplications affect gene expression, 3D organization, and plant-pathogen response. Genome Res. 2020;30(11):1583–92.
44. Hu Y, Wu X, Jin G, Peng J, Leng R, Li L, Gui D, Fan C, Zhang C. Rapid genome evolution and adaptation of *Thlaspi arvense* mediated by recurrent RNA-Based and tandem gene duplications. Front Plant Sci. 2021;12:772655.
45. Martinez-Garcia JF, Huq E, Quail PH. Direct targeting of light signals to a promoter element-bound transcription factor. Science. 2000;288(5467):859–63.
46. Jing Y, Lin R. Transcriptional regulatory network of the light signaling pathways. New Phytol. 2020;227(3):683–97.
47. Manfield IW, Devlin PF, Jen CH, Westhead DR, Gilmartin PM. Conservation, convergence, and divergence of light-responsive, circadian-regulated, and tissue-specific expression patterns during evolution of the *Arabidopsis* GATA gene family. Plant Physiol. 2007;143(2):941–58.
48. Bhalerao R, Keskitalo J, Sterky F, Erlandsson R, Bjorkbacka H, Birve SJ, Karlsson J, Gardestrom P, Gustafsson P, Lundeberg J, et al. Gene expression in autumn leaves. Plant Physiol. 2003;131(2):430–42.
49. Cruz de Carvalho MH, Pham-Thi AT, Gareil M, d'Arcy-Lameta A, Fodil YZ. Isolation and characterization of an aspartic proteinase gene from Cowpea (*Vigna unguiculata* L. Walp). J Plant Physiol. 2004;161(8):971–6.
50. Diaz C, Lemaitre T, Christ A, Azzopardi M, Kato Y, Sato F, Morot-Gaudry JF, Le Dily F, Masclaux-Daubresse C. Nitrogen recycling and remobilization are differentially controlled by leaf senescence and development stage in *Arabidopsis* under low nitrogen nutrition. Plant Physiol. 2008;147(3):1437–49.
51. Kato Y, Yamamoto Y, Murakami S, Sato F. Post-translational regulation of CND41 protease activity in senescent tobacco leaves. Planta. 2005;222(4):643–51.
52. Chen HJ, Huang YH, Huang GJ, Huang SS, Chow TJ, Lin YH. Sweet potato *SPAP1* is a typical aspartic protease and participates in ethephon-mediated leaf senescence. J Plant Physiol. 2015;180:1–17.
53. Rademacher S, Sprunck S. Downregulation of egg cell-secreted EC1 is accompanied with delayed gamete fusion and polytubey. Plant Signal Behav. 2013;8(12):e27377.
54. Zhang XL, Qi MF, Xu T, Lu XJ, Li TL. Proteomics profiling of ethylene-induced tomato flower pedicel abscission. J Proteom. 2015;121:67–87.
55. Zhao W, Baldwin EA, Bai J, Plotto A, Irey M. Comparative analysis of the transcriptomes of the calyx abscission zone of sweet orange insights into the huanglongbing-associated fruit abscission. Hortic Res. 2019;6:71.
56. Shen G, Luo Y, Yao Y, Meng G, Zhang Y, Wang Y, Xu C, Liu X, Zhang C, Ding G, et al. The discovery of a key prenyltransferase gene assisted by a chromosome-level *Epimedium pubescens* genome. Front Plant Sci. 2022;13:1034943.
57. Wang J, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu S, Marchler GH, Song JS, Thanki N, Yamashita RA, et al. The conserved domain database in 2023. Nucleic Acids Res. 2023;51(D1):D384–8.
58. Letunic I, Copley RR, Schmidt S, Ciccarelli FD, Doerks T, Schultz J, Ponting CP, Bork P. SMART 4.0: towards genomic data integration. Nucleic Acids Res. 2004;32(Database issue):D142–144.
59. Blum M, Andreeva A, Florentino LC, Chuguransky SR, Grego T, Hobbs E, Pinto BL, Orr A, Paysan-Lafosse T, Ponamareva I, et al. InterPro: the protein sequence classification resource in 2025. Nucleic Acids Res. 2024;53(D1):D444–56.
60. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A. Protein Identification and Analysis Tools on the ExPASy Serve. In: Walker, J.M, editors The Proteomics Protocols Handbook. Springer Protocols Handbooks. Humana Press. 2005:571–607.
61. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, Nakai K. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 2007;35(SUPPL2):W585–587.
62. Thumuluri V, Almagro Armenteros JJ, Johansen AR, Nielsen H, Winther O. DeepLoc 2.0: multi-label subcellular localization prediction using protein Language models. Nucleic Acids Res. 2022;50(W1):W228–34.
63. Armenteros JJA, Salvatore M, Emanuelsson O, Winther O, Von Heijne G, Elofsson A, Nielsen H. Detecting sequence signals in targeting peptides using deep learning. Life Sci Alliance. 2019;2(5):e201900429.
64. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol Plant. 2020;13(8):1194–202.
65. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;52(W1):W78–82.
66. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van De Peer Y, Rouzé P, Rombauts S. PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for *in Silico* analysis of promoter sequences. Nucleic Acids Res. 2002;30(1):325–7.
67. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. Methods. 2001;25(4):402–8.

Publisher's note