



OPEN ACCESS

EDITED BY
Xiupeng Mei,
Southwest University, China

REVIEWED BY
Muhammad Zulfiqar Ahmad,
Jiangsu Academy of Agricultural Sciences
(JAAS), China
Jianing Xu,
Shandong University of Technology,
China

*CORRESPONDENCE
Jaspreet K. Sembi
✉ jaspreet.sembi@pu.ac.in

RECEIVED 30 May 2026
REVISED 03 July 2026
ACCEPTED 07 July 2026
PUBLISHED 23 July 2026

CITATION
Kaur A, Pawar SV and Sembi JK (2026)
Molecular characterization of
4-Coumarate: CoA Ligase gene family in
orchids: an *in silico* study.
Front. Plant Sci. 17:1895523.
doi: 10.3389/fpls.2026.1895523

COPYRIGHT
© 2026 Kaur, Pawar and Sembi. This is an
open-access article distributed under the
terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which does
not comply with these terms.

Molecular characterization of 4-Coumarate: CoA Ligase gene family in orchids: an *in silico* study

Arshpreet Kaur^{1,2}, Sandip V. Pawar³ and Jaspreet K. Sembi^{1*}

¹Department of Botany, Panjab University, Chandigarh, India, ²Zakir Husain Delhi College, University of Delhi, New Delhi, India, ³University Institute of Pharmaceutical Sciences, Panjab University, Chandigarh, India

The 4-Coumarate: CoA Ligase (4CL) is a crucial enzyme in the phenylpropanoid pathway, which leads to the biosynthesis of a diverse array of secondary metabolites. The present research identified 29 4CL genes encoding this enzyme in eight orchid species (*Apostasia shenzhenica*, *Dendrobium catenatum*, *Phalaenopsis aphrodite*, *Phalaenopsis bellina*, *Phalaenopsis equestris*, *Phalaenopsis modesta*, *Phalaenopsis lueddemanniana* and *Phalaenopsis schilleriana*). Domain and motif analyses revealed the relatively conserved nature of the identified proteins, and all were localized to the peroxisome. Phylogenetic analysis showed that they clustered into three different groups. A total of 54 types of *cis*-acting regulatory elements across four categories were found in the promoters of these genes, indicating their functional diversity. Varied expression profiles were observed for the genes in vegetative and reproductive tissues of the orchids under study, with 11 genes showing high expression in the root, underscoring their role in plant growth and development. To conclude, this research lays the groundwork for future studies involving the functional characterization of 4CL genes and their correlation with the related phenylpropanoid synthesis.

KEYWORDS

4-Coumarate: CoA Ligase (4CL), orchids, *p*-coumaric acid, phenylpropanoid pathway, secondary metabolites

Introduction

Orchids belong to the family Orchidaceae, which is one of the largest and most advanced groups of flowering plants (Govaerts et al., 2026). Orchids have an astonishing therapeutic potential due to the presence of phytochemicals that can be extracted from different parts of the plant. Several classes of phytoconstituents, such as alkaloids, phenolics, terpenes, etc., have been isolated from therapeutically important orchids (Ghai et al., 2021). Amongst these diverse phytoconstituents, phenolics are a vast group of compounds consisting of one or more aromatic rings with one or more hydroxyl groups and synthesized through the phenylpropanoid pathway. Thus, this metabolic pathway leads to the synthesis

of lignins, flavonoids, coumarins, suberins, and bibenzyls, etc., that possess therapeutic potential and also provide protection to the plants against various environmental stress (biotic and abiotic) conditions (Vogt, 2010). Recent studies have also highlighted the role of phenylpropanoid pathway genes in stress tolerance. For instance, in *Zea mays*, *flavanone 3-hydroxylase (F3H)* gene (*ZmF3H6*) enhanced salt stress tolerance by the overaccumulation of flavonols (Xiao et al., 2026).

The phenylpropanoid pathway is regulated by several key players, including phenylalanine ammonia lyase (PAL), which catalyzes the first step of deamination of phenylalanine, resulting in the production of *trans*-cinnamate (Vogt, 2010), which is further transformed into *p*-coumaroyl CoA through the action of cinnamate 4-hydroxylase (C4H) and 4-Coumarate: CoA Ligase (4CL). The 4-Coumarate: CoA Ligase (4CL) enzyme (EC 6.2.1.12) was first discovered in *Petroselinum crispum* (Douglas et al., 1987) to catalyze the transformation of *p*-coumaric acid into *p*-coumaroyl CoA (Ehlting et al., 1999). It plays a crucial role in directing carbon flux from the core phenylpropanoid pathway into the synthesis of various phenylpropanoid-derived compounds (Ehlting et al., 1999). Further, the products formed by the action of 4CL enzyme act as precursors for the synthesis of varied metabolites, playing a role in tolerance towards abiotic and biotic stresses, lignification and UV protection (Lavhale et al., 2018).

The 4CL enzymes have been classified into three clusters: class I, class II, and class III, based on evolutionary analysis (Ehlting et al., 1999; Costa et al., 2003). Sequence alignments have consistently revealed the presence of conserved peptide motifs in all available 4-coumarate: CoA ligase (4CL) amino acid sequences (Becker-André et al., 1991). Among these motifs, the Box I motif, represented by SSGTTGLPKGKV, exhibits near-absolute conservation in 4CLs, and similar motifs are prevalent in diverse enzymes, including luciferases, acetyl-CoA synthetases, long-chain fatty acyl-CoA synthetases, and peptide synthetases. Furthermore, the Box II motif, GEICIRG, is universally conserved across all 4CLs (Becker-André et al., 1991; Uhlmann and Ebel, 1993; Hu et al., 1998). The central cysteine residue within this motif has been proposed to play a direct role in catalytic processes (Stuible et al., 2000). This highlights the significance of these motifs in the functional and structural aspects of 4CLs.

The number of members included in the 4CL gene family also varies amongst different species of the plant kingdom like *Arabidopsis thaliana* (4; Li et al., 2015), *Boehmeria nivea* (3; Tang et al., 2018), *Citrus sinensis* (3; Zhou et al., 2022), *Glycine max* (3; Lindermayr et al., 2002), *Oryza sativa* (5; Gui et al., 2011), *Panicum virgatum* (2; Xu et al., 2011), *Physcomitrella patens* (4; Silber et al., 2008), *Piper nigrum* (3; Jin et al., 2020), *Populus tomentosa* (5; Rao et al., 2015), *Punica granatum* (12; Wang et al., 2022), *Rubus idaeus* (3; Kumar and Ellis, 2003), and *Salvia miltiorrhiza* (2; Jin et al., 2012). To date, the maximum number of 4CL genes has been reported in *Malus domestica* (69; Ma et al., 2023) (Table 1). These variations in the number of genes in the 4CL gene family in different plants suggest that different 4CL genes may lead to the biosynthesis of different kinds of phenylpropanoids.

Researchers have extensively investigated the expression patterns of 4CL genes in multiple plant species across various tissues and developmental stages (Hu et al., 1998; Ehlting et al., 1999;

TABLE 1 Numerical strength of the 4CL gene family across various plant species.

Name of the plant	Name of the family	Number of members	Reference
<i>Arabidopsis thaliana</i>	Brassicaceae	4	Li et al., 2015
<i>Beta vulgaris</i>	Amaranthaceae	27	Lan et al., 2026
<i>Bletilla striata</i>	Orchidaceae	21	Huang et al., 2023
<i>Boehmeria nivea</i>	Urticaceae	3	Tang et al., 2018
<i>Brassica carinata</i>	Brassicaceae	10	Zhang M. et al., 2025
<i>Brassica juncea</i>	Brassicaceae	13	Zhang M. et al., 2025
<i>Brassica napus</i>	Brassicaceae	12	Zhang M. et al., 2025
<i>Brassica nigra</i>	Brassicaceae	6	Zhang M. et al., 2025
<i>Brassica oleracea</i>	Brassicaceae	5	Zhang M. et al., 2025
<i>Brassica rapa</i>	Brassicaceae	7	Zhang M. et al., 2025
<i>Camellia sinensis</i> 'Huangdan'	Theaceae	8	Ren et al., 2023
<i>Camellia sinensis</i> 'Tie-guanyin'	Theaceae	9	Ren et al., 2023
<i>Citrus sinensis</i>	Rutaceae	3	Zhou et al., 2022
<i>Gastrodia elata</i>	Orchidaceae	14	Sha et al., 2025
<i>Glycine max</i>	Fabaceae	20	Yang et al., 2024
<i>Hippophae rhamnoides</i>	Elaeagnaceae	7	Song et al., 2023
<i>Juglans regia</i>	Juglandaceae	36	Ma et al., 2024
<i>Juglans mandshurica</i>	Juglandaceae	31	Ma et al., 2024
<i>Lithospermum erythrorhizon</i>	Boraginaceae	8	Nakanishi et al., 2024
<i>Malus domestica</i>	Rosaceae	69	Ma et al., 2023
<i>Manihot esculenta</i>	Euphorbiaceae	50	Ran et al., 2024
<i>Marchantia paleacea</i>	Marchantiaceae	4	Gao et al., 2024
<i>Miscanthus lutarioriparius</i>	Poaceae	20	Huang et al., 2026
<i>Musa acuminata</i>	Musaceae	20	Wang et al., 2024
<i>Nicotiana tabacum</i>	Solanaceae	24	Ma et al., 2026
<i>Oryza sativa</i>	Poaceae	5	Gui et al., 2011
<i>Panicum virgatum</i>	Poaceae	2	Xu et al., 2011
<i>Physcomitrella patens</i>	Funariaceae	4	Silber et al., 2008

(Continued)

TABLE 1 Continued

Name of the plant	Name of the family	Number of members	Reference
<i>Piper nigrum</i>	Piperaceae	3	Jin et al., 2020
<i>Populus tomentosa</i>	Salicaceae	5	Rao et al., 2015
<i>Punica granatum</i>	Lythraceae	12	Wang et al., 2022
<i>Pyrus bretschneideri</i>	Rosaceae	18	Zhang et al., 2024
<i>Rubus idaeus</i>	Rosaceae	3	Kumar and Ellis, 2003
<i>Salvia miltiorrhiza</i>	Lamiaceae	10	Fu et al., 2025
<i>Solanum tuberosum</i>	Solanaceae	10	Nie et al., 2023
<i>Triticum aestivum</i>	Poaceae	40	Gholizadeh et al., 2025
<i>Zea mays</i>	Poaceae	32	Zhang Z et al., 2025a

Kumar and Ellis, 2003; Di et al., 2012; Jin et al., 2012; Tang et al., 2018; Jin et al., 2020). Furthermore, the response of 4CL genes to different stimuli has been explored. Investigations have focused on their expression changes in relation to elicitors/phytohormones, as demonstrated by the work of Awasthi et al. (2019) in *Coleus forskohlii*. Additionally, the impact of abiotic stress on the expression of 4CL genes has been studied by Sun et al. (2020) in *Gossypium hirsutum* and Zhong et al. (2022) in *Eucommia ulmoides* and Ma et al. (2026) in *Nicotiana tabacum*. The role of 4CL genes in antifungal response has also been reported in *Gastrodia elata* (Sha et al., 2025). Another factor that has been investigated is the influence of UV exposure on the expression patterns of 4CL genes in *Isatis indigotica*, as observed in the study conducted by Di et al. (2012). By examining their expression patterns across diverse tissues, developmental stages, and in response to various environmental triggers, scientists have gained a better understanding of the roles and regulatory mechanisms of 4CL genes in plant growth, development, and stress responses. Additionally, overexpression of specific 4CL genes leading to increased metabolite synthesis or tolerance to stress conditions has been reported in certain plants, such as *Fraxinus mandshurica*, wherein overexpression of the *Fm4CL2* gene conferred resistance to osmotic and drought stress (Chen et al., 2020). Thus, these studies have improved our understanding of the genetic diversity, functional roles, and potential applications of 4CL genes in different plant species, highlighting their importance in plant biology.

Thus, keeping both the significance of 4CL genes and the therapeutic importance of orchids in view, the present study on the identification and characterization of 4CL genes in eight selected orchid species (*Apostasia shenzhenica*, *Dendrobium catenatum*, *Phalaenopsis aphrodite*, *Phalaenopsis bellina*, *Phalaenopsis equestris*, *Phalaenopsis modesta*, *Phalaenopsis lueddemanniana*, and *Phalaenopsis schilleriana*) was conducted. Further, this study acts as

a reference template for conducting similar studies on these genes in various plants to delve into the plausible functions of the 4CL genes.

Materials and methods

Identification of 4CL genes in selected orchid species

For the current study, genome data of *A. shenzhenica* [Accession: PRJNA310678], *D. catenatum* [Accession: PRJNA262478] and *P. equestris* [Accession: PRJNA192198, PRJNA382149] was taken from the National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/>). 4CLs (Gui et al., 2011; Li et al., 2015) of *Arabidopsis thaliana* and *Oryza sativa* were retrieved from NCBI and Phytozome using their accession numbers (Supplementary Table 1). These sequences were used as a query to execute a BLASTp search against the protein sequences of *A. shenzhenica*, *D. catenatum* and *P. equestris* using the NCBI Protein BLAST web server against the ClusteredNR database. The search was conducted using the default BLASTp settings, including the BLOSUM62 substitution matrix, conditional compositional score matrix adjustment and default gap costs. The hits showing an E-value of zero and sequence identity greater than 50% were subjected to further validation. The 4CL protein sequences of *A. thaliana* and *O. sativa* (Supplementary Table 1) were used to conduct a tBLASTn search for the identification of these specific proteins from the transcriptomes of *P. aphrodite*, *P. bellina*, *P. lueddemanniana*, *P. modesta* and *P. schilleriana* in the Orchidstra 2.0 database (Chao et al., 2017; <http://orchidstra2.abrc.sinica.edu.tw/orchidstra2/index.php>).

Domain analysis and multiple sequence alignment

Conserved domain search was employed using the SMART (Letunic et al., 2015; <http://smart.embl-heidelberg.de/>) tool to confirm the presence of the AMP-binding (PF00501) and AMP-binding_C (PF13193) domains, which are conserved for 4CL proteins. After the recognition of the domain, all the protein sequences were subjected to Expasy-PROSITE (Sigrist et al., 2012; <https://prosite.expasy.org/>) for the construction of domain architecture. The conserved region of 4CL proteins was located and marked by using the MultAlin server (Corpet, 1988; <http://multalin.toulouse.inra.fr/multalin/>) by sequence alignment of the identified 4CL proteins along with the 4CLs of *Arabidopsis thaliana* and *Oryza sativa*.

Conserved motif analysis

Conserved motifs and their positional distribution within the identified 4CL proteins from *A. shenzhenica*, *D. catenatum*, *P. aphrodite*, *P. bellina*, *P. equestris*, *P. lueddemanniana*, *P. modesta* and *P. schilleriana* along with their counterparts from *Arabidopsis*

thaliana and *Oryza sativa*, were predicted using the Multiple Expectation Maximization for Motif Elicitation (MEME) Suite (version 5.5.5) (Bailey et al., 2015; <https://meme-suite.org/meme/>). The analysis permitted any number of motif repetitions, with the maximum number of motifs set to 10. The minimum and maximum motif widths were defined as 6 and 50 amino acids, respectively, while all other parameters were maintained at their default settings.

Physicochemical characterization and sub-cellular localization

Protein physicochemical properties, including sequence length (total amino acids), molecular weight, isoelectric point (pI), instability index, aliphatic index, and GRAVY (grand average of hydropathicity) values, were calculated using ProtParam (Gasteiger et al., 2005; <https://web.expasy.org/protparam/>). Subcellular localization of the proteins was predicted using the Plant-mPLOC server (Chou and Shen, 2010; <http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/>).

Secondary and tertiary structure prediction

The secondary structural features of the proteins, including the relative proportions of α -helices, extended strands, β -turns, and random coils, were predicted using the SOPMA online server (Geourjon and Deleage, 1995; https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html). All identified proteins were subsequently submitted to the SWISS-MODEL server (Waterhouse et al., 2018; <https://swissmodel.expasy.org/>) to generate three-dimensional models, where the best-fitting and highest-similarity homology templates were selected. The templates with high GMQE and sequence identity were then employed for structural visualization and model construction using PyMOL (DeLano, 2002; <https://pymol.org/>).

Phylogenetic analysis

The full-length amino acid sequences of 4CL proteins from *A. shenzhenica*, *D. catenatum*, *P. aphrodite*, *P. bellina*, *P. equestris*, *P. lueddemanniana*, *P. modesta* and *P. schilleriana* were aligned with their counterparts in *A. thaliana* (Li et al., 2015), *O. sativa* (Gui et al., 2011) and *Vanilla planifolia* (Kaur et al., 2025) using MUSCLE (Multiple Sequence Comparison by Log-Expectation) with gaps. Phylogenetic trees were then constructed using the neighbor-joining method in MEGAX (Kumar et al., 2018; <http://www.megasoftware.net/>) with pairwise deletion and 1000 bootstrap replicates.

Gene structure and promoter analysis

The genomic and coding sequences (CDS) of 4CL genes from *A. shenzhenica*, *D. catenatum*, and *P. equestris* were retrieved from the NCBI database. Gene structures were analyzed by comparing the genomic and coding sequences to determine the distribution of

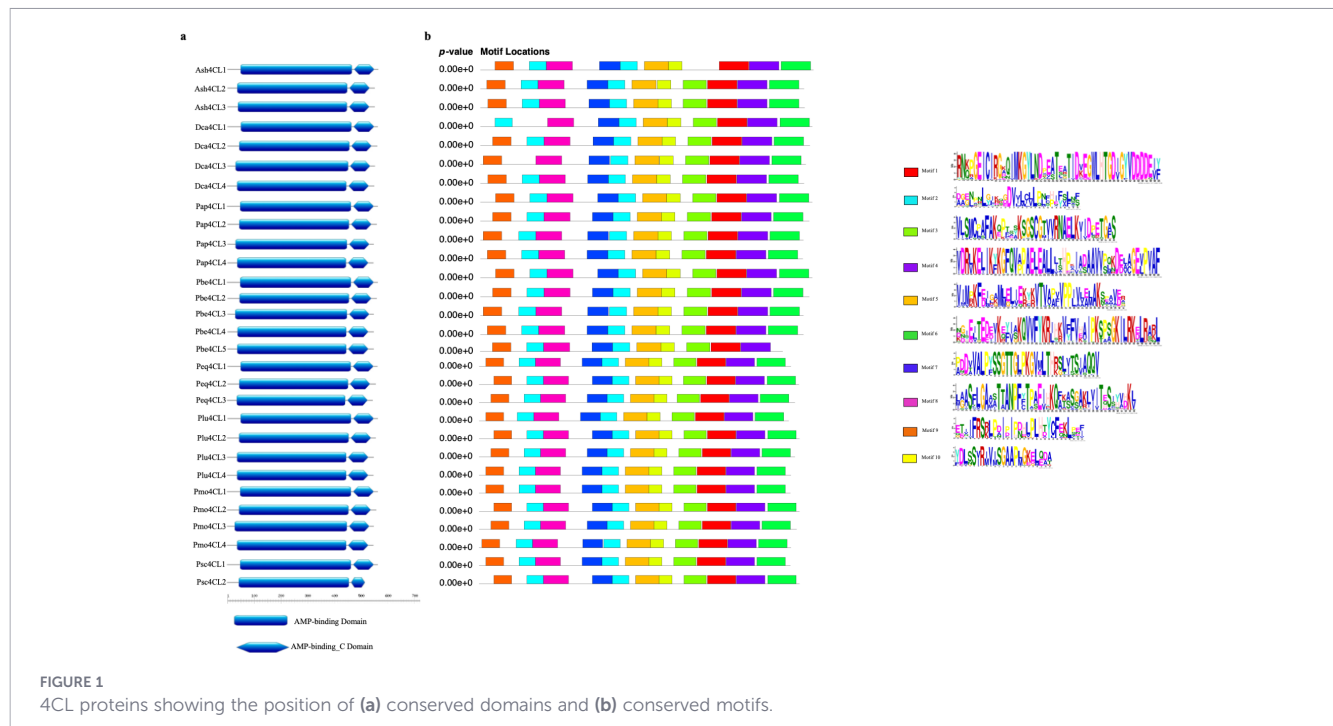
exons and introns using the Gene Structure Display Server (GSDS v2.0) (Hu et al., 2015; <http://gsds.gao-lab.org/>). Additionally, promoter sequences located 1.5 kb upstream of the start codon for these genes were obtained from NCBI. These promoter sequences were then analyzed using the PlantCARE database (Lescot et al., 2002; <https://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) to predict the type and positions of *cis*-regulatory elements. The gene structure and promoter elements studies were only done for *A. shenzhenica*, *D. catenatum* and *P. equestris*, as genomic data were available for only these three out of the eight orchids under study.

Expression profiling

The CDS (coding sequences) of genes were used in order to perform an NCBI Sequence Read Archive (SRA). BLASTn search was done against the RNA sequence data derived from the tissues of *A. shenzhenica* [pollen (SRX2938652), seed (SRX2938653), tuber (SRX2938654)], *D. catenatum* [root (SRX2938667), column (SRX2251512), sepal (SRX2251513), stem (SRX2251516), leaf (SRX2251517), lip (SRX2251518), flower bud (SRX2251519), pollinia (SRX2938662)], *P. aphrodite* [root (SRX2439763), leaf (SRX2439762), fully opened flower (SRX2439761), large flower bud (SRX2196321), small flower bud (SRX2439759)], *P. bellina* [leaf (SRX2210823), root (SRX2210822), large flower bud (SRX2210821), fully opened flower (SRX2210820)], *P. equestris* [sepal (SRX1806366), petal (SRX1806365), lip (SRX1806348), column (SRX1805894), flower bud (SRX1074880), leaf (SRX1074879), stem (SRX1074876), root (SRX1074875)], *P. lueddemanniana* [leaf (SRX2210814), root (SRX2210813), small bud (SRX2210812), large bud (SRX2210811), fully opened flower (SRX2210810)], *P. modesta* [small bud (SRX2210819), large bud (SRX2210818), fully opened flower (SRX2210817)] and *P. schilleriana* [large bud (SRX2210807), fully opened flower (SRX2210806)]. The number of hits were counted, and RPKM values were estimated as $(C \cdot 10^9)/(N \cdot L)$, where C denotes the number of hits corresponding to a particular sequence, N denotes the total number of reads in that specific RNA-seq experiment and L denotes the length of the CDS sequence for the particular candidate gene (Mortazavi et al., 2008). Heatmaps were generated using the ClustVis tool (Metsalu and Vilo, 2015; <https://biit.cs.ut.ee/clustvis/>) to analyze the relative expression of all identified genes across different tissues in the selected orchid species.

Results

Extensive BLASTp and tBLASTn searches led to the identification of 4CL proteins in selected orchid species using 4CL proteins of *Arabidopsis* and Rice as query sequences. The existence of two characteristic conserved domains, AMP-binding (PF00501) and AMP-binding_C (PF13193), was confirmed with the help of SMART server for validation of the predicted protein sequences. A total of three 4CL proteins in *A. shenzhenica* and *P. equestris* each, four in *D. catenatum*, *P. aphrodite*, *P. lueddemanniana* and *P.*



modesta each, five in *P. bellina* and two in *P. schilleriana* were identified. Domain analysis and visualization to depict the presence of the conserved domain were done (Supplementary Table 2; Figure 1).

Upon aligning all the 4CL proteins, it was observed that the two distinct motifs, Box I (SSGTTGLPKGV) and Box II (GEICIRG), which are characteristic of 4CL proteins, were found conserved in all the proteins. Moreover, the conserved amino acid residues crucial for substrate binding and enzymatic function were also observed in the majority of the proteins (Figure 2).

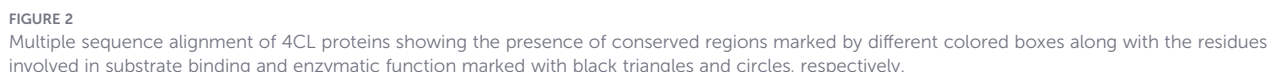
In 4CL proteins, motifs 1 and 7 represented the AMP_binding domain and were consistently identified in all the proteins. Notably, motifs 7 and 1 contained the Box I (SSGTTGLPKGV) and Box II (GEICIRG) sequences, respectively, which are characteristic conserved regions of the 4CL proteins. Motif 4, present across all proteins, represented the AMP binding_C domain and contained the essential residues required for the enzymatic function of 4CLs (Figure 1).

Physicochemical characterization of 4CL proteins was done. The average length of these proteins was found to be 553 amino acids (aa). The smallest predicted protein was Psc4CL2 with a length of 514 aa, while the longest was Ash4CL1 with a length of 565 aa. The average molecular weight was predicted to be 59.97 kDa. Notably, Ash4CL1 exhibited the highest molecular weight (61.27 kDa) among all the 4CL proteins. The isoelectric point across all 4CL proteins ranged from 5.25 (Psc4CL2) to 6.6 (Peq4CL2), with an average of 5.86. Interestingly, all the proteins exhibited a positive GRAVY value except Pbe4CL5. Seventeen proteins were found to be stable based on protein stability analysis. The aliphatic index of the identified 4CL proteins ranged from 94.98 to 106.45, with the majority of proteins exhibiting values above 95, indicating the high abundance of aliphatic amino acids in these proteins. All 4CL proteins were localized in the peroxisome (Table 2).

The comparative analysis revealed a high degree of similarity in the predicted secondary structure of 4CL proteins. Interestingly, in 4CL proteins, random coils constituted the highest percentage compared to other secondary structure elements. The prevalence of random coils ranged from 37.23% in Pbe4CL4 and Pbe4CL5 to 43.37% in Pmo4CL2 (Table 3; Supplementary Figure 1). The prediction of tertiary structures of the identified 4CL proteins showed that the proteins showed homology to various PDB structure templates. However, all the templates represented 4-coumarate: CoA ligase proteins (Supplementary Table 3). The three-dimensional visualization of 4CL proteins based on the most commonly occurring and high-similarity templates was done using PyMOL (Figure 3).

MEGA software was employed to explore the evolutionary connections among the 4CL protein sequences of *A. schenzenica*, *D. catenatum*, *P. aphrodite*, *P. bellina*, *P. equestris*, *P. lueddemanniana*, *P. modesta*, *P. schilleriana*, *V. planifolia*, *A. thaliana* and *O. sativa*. The phylogenetic analysis of 4CL proteins revealed a tight clustering of Ash4CL2, Ash4CL3, Dca4CL4, Pap4CL4, Pbe4CL4-5, Pe4CL3, Pmo4CL4, Plu4CL4, Vpl4CL3-5 with class III 4CL proteins of *O. sativa* (Os4CL1, Os4CL3, Os4CL4 and Os4CL5) while the remaining orchid proteins clustered with class II members of 4CL proteins in *Arabidopsis thaliana* (At4CL3) and *Oryza sativa* (Os4CL2). However, no orchid protein sequence clustered in class I of 4CL proteins comprising At4CL1, At4CL2 and At4CL4 (Figure 4).

The Gene Structure Display Server (GSDS) online tool was used for gene structural analysis. Notably, 4CL genes exhibited a diverse intron-exon structure. *Dca4CL3-4* and *Peq4CL3* genes featured four introns and five exons, while *Ash4CL1*, *Ash4CL3*, *Dca4CL1-2* and *Peq4CL1-2* displayed five introns and six exons and *Ash4CL2* exhibited three introns and four exons (Figure 5). Comprehensive



motif, and TCA-element. These elements can be induced by different hormones such as abscisic acid, gibberellins, salicylic acid and methyljasmonate. Stress-responsive elements detected encompassed ARE, DRE core, LTR, MYB, MYC, MYB-like sequence, W-Box, WRE3, and WUN-motif. Additionally, various light-responsive *cis*-acting elements, such as Box 4, G-Box, GA motif, GATA motif, GT1-motif, MRE and chs-CMA1a were also identified (Figure 6).

Expression analysis played a pivotal role in discerning the variations in 4CL genes expression across diverse vegetative and reproductive tissues of orchids under study. *Ash4CL1* and *Ash4CL2* were expressed highly in pollen in comparison to the seed and tuber. *Dca4CL2* of *Dendrobium catenatum* also showed high expression in pollinia. *Pbe4CL1-5* of *Phalaenopsis bellina*, *Pap4CL2-4* of *Phalaenopsis aphrodite*, *Peq4CL3* of *Phalaenopsis equestris* and *Plu4CL3-4* of *Phalaenopsis lueddemanniana* were

TABLE 2 Physicochemical parameters of 4CL proteins.

Protein	AA	MW (kDa)	pI	Instability index	Aliphatic index	GRAVY	Subcellular localization
Ash4CL1	565	61.27	5.5	39.01	99.24	0.132	Peroxisome
Ash4CL2	549	59.55	5.78	37.3	101.73	0.087	Peroxisome
Ash4CL3	550	60.01	5.46	39.51	97.53	0.05	Peroxisome
Dca4CL1	563	60.53	5.4	39.93	100.85	0.194	Peroxisome
Dca4CL2	559	60.41	5.78	45.16	97.87	0.113	Peroxisome
Dca4CL3	552	60.40	5.9	42.19	96.9	0.043	Peroxisome
Dca4CL4	549	59.76	5.43	37.13	102.66	0.09	Peroxisome
Pap4CL1	563	60.64	5.59	40.06	104.49	0.198	Peroxisome
Pap4CL2	558	59.90	6.32	41.13	103.98	0.23	Peroxisome
Pap4CL3	548	59.90	5.99	40.5	96.42	0.009	Peroxisome
Pap4CL4	547	59.63	6.04	37.81	99.27	0.023	Peroxisome
Pbe4CL1	563	60.53	5.83	39.56	102.61	0.193	Peroxisome
Pbe4CL2	558	60.10	6.26	40.81	106.45	0.27	Peroxisome
Pbe4CL3	548	60.05	6.09	43.38	94.98	0.005	Peroxisome
Pbe4CL4	548	60.29	5.87	39.73	100.31	0.013	Peroxisome
Pbe4CL5	548	60.28	5.8	40.64	99.62	-0.006	Peroxisome
Peq4CL1	562	60.57	5.76	39.3	102.78	0.183	Peroxisome
Peq4CL2	555	59.60	6.6	42.13	101.89	0.18	Peroxisome
Peq4CL3	544	59.27	6.04	38.07	98.38	0.022	Peroxisome
Plu4CL1	563	60.62	5.64	38.58	106.06	0.226	Peroxisome
Plu4CL2	555	59.83	6.47	40.77	104.4	0.239	Peroxisome
Plu4CL3	548	60.27	5.88	39.01	101.02	0.01	Peroxisome
Plu4CL4	547	59.65	6.14	38.12	99.8	0.029	Peroxisome
Pmo4CL1	563	60.59	5.74	38.37	104.16	0.199	Peroxisome
Pmo4CL2	558	60.09	6.53	42.14	105.93	0.261	Peroxisome
Pmo4CL3	548	60.07	6.29	41.51	97.66	0.029	Peroxisome
Pmo4CL4	547	59.72	6.04	37.39	99.09	0.029	Peroxisome
Psc4CL1	563	60.70	5.57	39.51	103.62	0.19	Peroxisome
Psc4CL2	514	55.02	5.25	39.69	104.53	0.256	Peroxisome

AA, Amino acid; MW, Molecular weight (in kDa); pI, Theoretical isoelectric point; GRAVY, Grand average of hydropathicity index.

highly expressed in the root compared to other tissues. *Pmo4CL1*, *Pmo4CL2* and *Pmo4CL4* showed maximum expression in the small bud, while *Psc4CL1* and *Psc4CL2* displayed elevated expression in the large bud (Supplementary Table 5; Figure 7).

Discussion

The rich diversity of secondary metabolites in orchids underlies their remarkable curative potential, highlighting the importance of investigating their biosynthetic pathways. Phenolics are a large and important class of secondary metabolites, and the pathway involved

in the biosynthesis of phenolic compounds is the phenylpropanoid pathway. The current research was aimed at identification and characterization of 4CL, one of the key enzymes operating at a terminal stage of the phenylpropanoid metabolic pathway and contributing to the synthesis of phenylpropanoids, which act as precursor molecules for the production of a wide range of compounds involved in plant growth and development, defense against pathogens, and response to environmental cues (Ferrer et al., 2008; Vogt, 2010). The number of 4CL genes identified across various plant species has been quite variable. The prediction of five 4CL genes in *P. bellina* corresponds to the reports in *O. sativa* (Gui et al., 2011) and *Populus tomentosa* (Rao et al., 2015), where both of these plants also exhibited five 4CL genes. The 4CL gene family in *D.*

TABLE 3 Prediction of Secondary structures of 4CL proteins.

Sequence ID	Alpha helix	Random coil	Extended strand	Beta turn
Ash4CL1	181 (32.04%)	233 (41.24%)	109 (19.29%)	42 (7.43%)
Ash4CL2	167 (30.42%)	223 (40.62%)	113 (20.58%)	46 (8.38%)
Ash4CL3	168 (30.55%)	226 (41.09%)	114 (20.73%)	42 (7.64%)
Dca4CL1	163 (28.95%)	235 (41.74%)	118 (20.96%)	47 (8.35%)
Dca4CL2	166 (29.70%)	240 (42.93%)	109 (19.50%)	44 (7.87%)
Dca4CL3	172 (31.16%)	222 (40.22%)	112 (20.29%)	46 (8.33%)
Dca4CL4	170 (30.97%)	232 (42.26%)	104 (18.94%)	43 (7.83%)
Pap4CL1	170 (30.20%)	233 (41.39%)	116 (20.60%)	44 (7.82%)
Pap4CL2	168 (30.11%)	236 (42.29%)	116 (20.79%)	38 (6.81%)
Pap4CL3	168 (30.66%)	221 (40.33%)	117 (21.35%)	42 (7.66%)
Pap4CL4	180 (32.91%)	217 (39.67%)	110 (20.11%)	40 (7.31%)
Pbe4CL1	176 (31.26%)	235 (41.74%)	111 (19.72%)	41 (7.28%)
Pbe4CL2	172 (30.82%)	239 (42.83%)	108 (19.35%)	39 (6.99%)
Pbe4CL3	166 (30.29%)	225 (41.06%)	112 (20.44%)	45 (8.21%)
Pbe4CL4	192 (35.04%)	204 (37.23%)	109 (19.89%)	43 (7.85%)
Pbe4CL5	193 (35.22%)	204 (37.23%)	107 (19.53%)	44 (8.03%)
Peq4CL1	172 (30.60%)	232 (41.28%)	110 (19.57%)	48 (8.54%)
Peq4CL2	161 (29.01%)	238 (42.88%)	112 (20.18%)	44 (7.93%)
Peq4CL3	181 (33.27%)	211 (38.79%)	109 (20.04%)	43 (7.90%)
Plu4CL1	177 (31.44%)	232 (41.21%)	113 (20.07%)	41 (7.28%)
Plu4CL2	176 (31.71%)	232 (41.80%)	109 (19.64%)	38 (6.85%)
Plu4CL3	190 (34.67%)	199 (36.31%)	116 (21.17%)	43 (7.85%)
Plu4CL4	183 (33.46%)	217 (39.67%)	108 (19.74%)	39 (7.13%)
Pmo4CL1	176 (31.26%)	234 (41.56%)	109 (19.36%)	44 (7.82%)
Pmo4CL2	162 (29.03%)	242 (43.37%)	114 (20.43%)	40 (7.17%)
Pmo4CL3	172 (31.39%)	226 (41.24%)	107 (19.53%)	43 (7.85%)
Pmo4CL4	181 (33.09%)	212 (38.76%)	110 (20.11%)	44 (8.04%)
Psc4CL1	177 (31.44%)	235 (41.74%)	114 (20.25%)	37 (6.57%)
Psc4CL2	147 (28.60%)	216 (42.02%)	117 (22.76%)	34 (6.61%)

catenatum, *P. aphrodite*, *P. modesta* and *P. lueddemanniana* comprises four members, which is comparable to the presence of the 4CL gene family in the model plant *A. thaliana* (Li et al., 2015). Three 4CL genes have been identified in *A. shenzhenica* and *P. equestris* in the present work. Other plants which show such reduced numbers include *Rubus idaeus* (Kumar and Ellis, 2003), *Boehmeria nivea* (Tang et al., 2018), *Piper nigrum* (Jin et al., 2020), and *Citrus sinensis* (Zhou et al., 2022). Similarity in the quantitative strength of 4CL genes in *P. schilleriana* (two 4CL genes) is observed with *Panicum virgatum* (Xu et al., 2011) and *Salvia miltiorrhiza* (Jin et al., 2012). However, there are plant species in which the number of 4CL genes is quite high, which may be due to evolutionary differences caused by gene duplication events.

Further, protein domain and motif studies shed insights into numerous conserved characteristics of the proteins. In 4CL

proteins, two domains were predicted: AMP-binding and AMP-binding_C. The two signature motifs of the 4CL proteins, Box I (SSGTTGLPKGV) and Box II (GEICIRG), are present in 4CL proteins in all plants, and sequence alignment of the orchid 4CL proteins also showed the same conservation, thus it is consistent with the studies in other plants (Uhlmann and Ebel, 1993; Lee et al., 1995; Wei et al., 2013; Gao et al., 2015; Zhang et al., 2019; Sun et al., 2020; Wang et al., 2022, 2023). This highlights the evolutionary stability of 4CL proteins across various plant species. All proteins in the family of adenylate-forming enzymes have Box I, which is the AMP (adenosine monophosphate) nucleotide binding motif (Allina et al., 1998; Lavhale et al., 2018). Furthermore, according to Stuible et al. (2000), the cysteine in the GEICIRG motif contributes to the stability and catalytic activity of 4CLs. Further, identical motif organization was detected in 4CL proteins of orchids compared

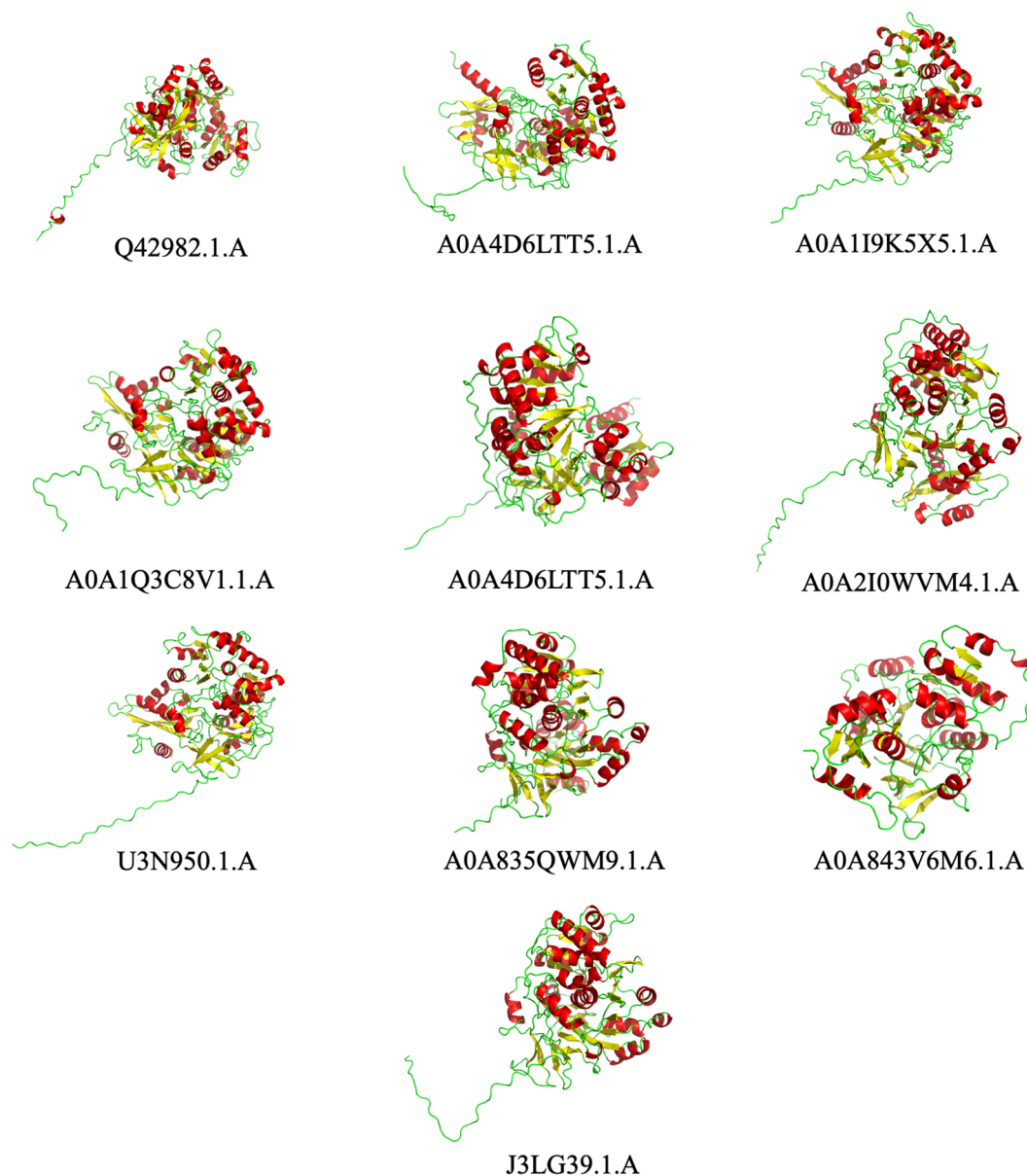


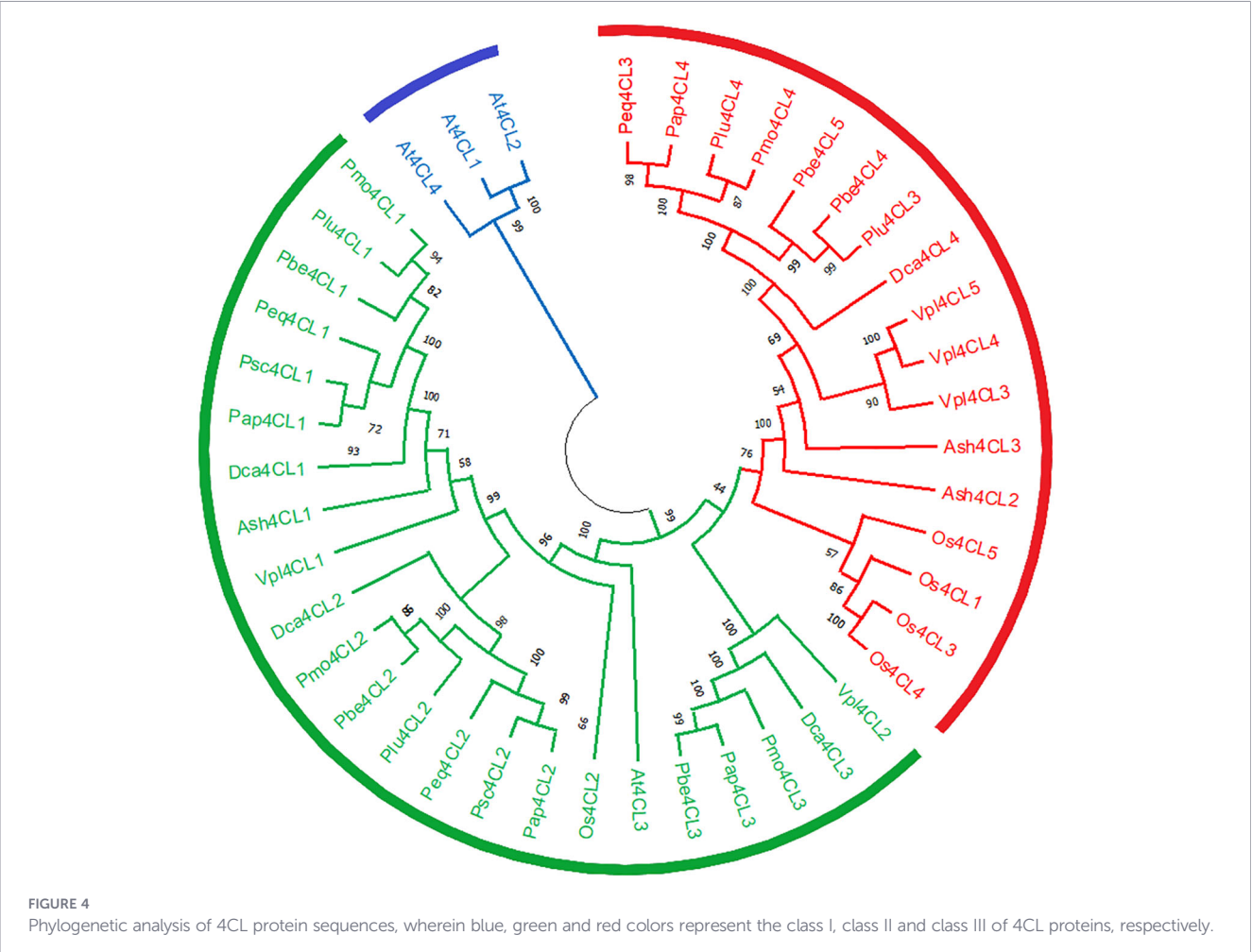
FIGURE 3

Three-dimensional model of selected templates of 4CL proteins.

to *Arabidopsis thaliana* and *Oryza sativa* 4CLs, which corroborated that the gene family is very conserved and the proteins are likely to perform a similar role in the phenylpropanoid pathway.

The characteristics of the inferred proteins were described using various bioinformatics methods. 4CL proteins were found in the peroxisome, resembling Pg4CL10 of *Punica granatum*, numerous 4CL members of *Gossypium hirsutum* (Sun et al., 2020), M14CL1–10 of *Miscanthus lutarioriparius* (Huang et al., 2026) and Le4CL3 and Le4CL4 of *Lithospermum erythrorhizon* (Nakanishi et al., 2024). All of the 4CL proteins, except Pbe4CL5 had positive GRAVY values, which is consistent with a prior study on the hydrophobicity of Bn4CL3 from *Boehmeria nivea* (Tang et al., 2018), Pg4CL1–3, Pg4CL6, and Pg4CL8–11 from *Punica granatum* (Wang et al., 2022), Ma4CLs of *Musa*

acuminata (Wang et al., 2024), Vpl4CL1–3 and Vpl4CL5 of *Vanilla planifolia* (Kaur et al., 2025) and M14CLs of *Miscanthus lutarioriparius* (Huang et al., 2026). Further, twelve 4CL proteins under study exhibited an instability index of more than 40, which indicates that these proteins may require post-translational modifications for their functioning (Lee et al., 2023). Additionally, the high aliphatic index for the majority of proteins indicated their high thermostability (Ikai, 1980). Molecular modelling techniques differentiated the various proteins under study on the basis of the amount of beta turns, extended strands, random coils, and alpha helices. In addition, 3-D structure analysis showed the resemblance of the proteins to 4-coumarate: CoA ligase templates based on homology analysis.



In the case of 4CL proteins identified in the present study in orchid species, the proteins belonging to classes I, II, and III formed different clades. In dicots, 4CL proteins are divided into 2 classes, with Class I proteins playing a role in lignin biosynthesis while the Class II proteins are involved in the biosynthesis of other phenolic compounds (Ehlting et al., 1999). But in the case of *Oryza sativa*, a monocot, the proteins are clustered into 3 clades, Class I, II and III, similar to our present results. Thus, this indicates that there are variations in substrate specificity of 4CL enzymes in monocots and dicots, eventually leading to different compositions of compounds

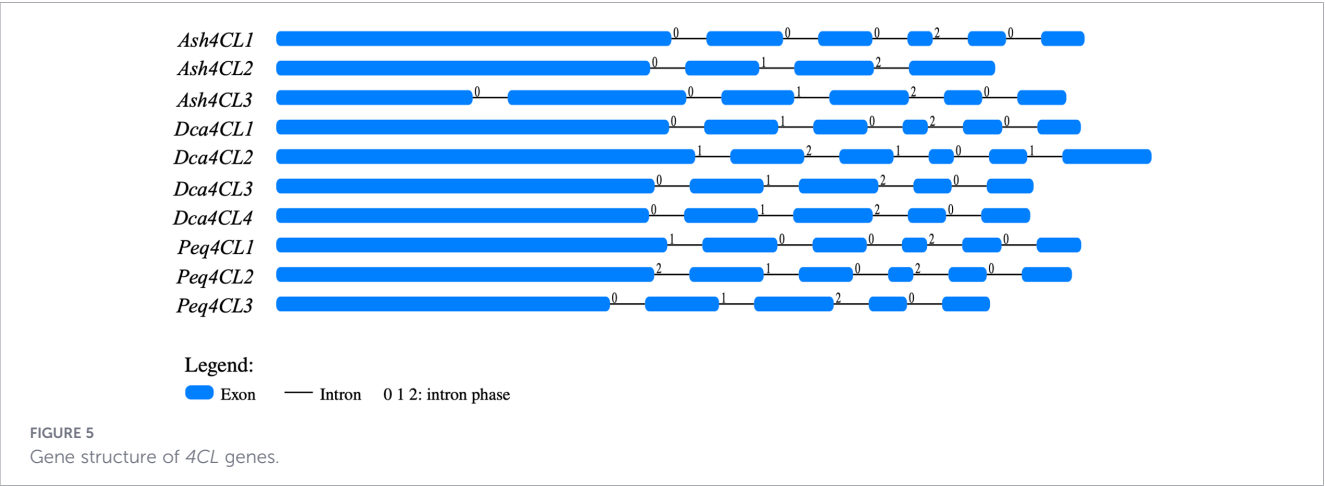


TABLE 4 Characterization of *Ash4CL*, *Dca4CL* and *Peq4CL* genes.

Gene	Location in genome	Gene stretch	Number of exons	Number of introns
<i>Ash4CL1</i>	Loc: fragScaff_scaffold_95	c1187152-1176385	6	5
<i>Ash4CL2</i>	Loc: fragScaff_scaffold_80	c433741-426298	4	3
<i>Ash4CL3</i>	Loc: fragScaff_scaffold_155	c1160315-1152316	6	5
<i>Dca4CL1</i>	Loc: fragScaff_scaffold_716	20371336-20376294	6	5
<i>Dca4CL2</i>	Loc: fragScaff_scaffold_601	c227168-221460	6	5
<i>Dca4CL3</i>	Loc: fragScaff_scaffold_1126	c695704-693488	5	4
<i>Dca4CL4</i>	Loc: fragScaff_scaffold_621	c11330335-11325764	5	4
<i>Peq4CL1</i>	Loc: Scaffold000369	c774561-766809	6	5
<i>Peq4CL2</i>	Loc: Scaffold000202	1738006-1751433	6	5
<i>Peq4CL3</i>	Loc: Scaffold000002	14068300-14073361	5	4

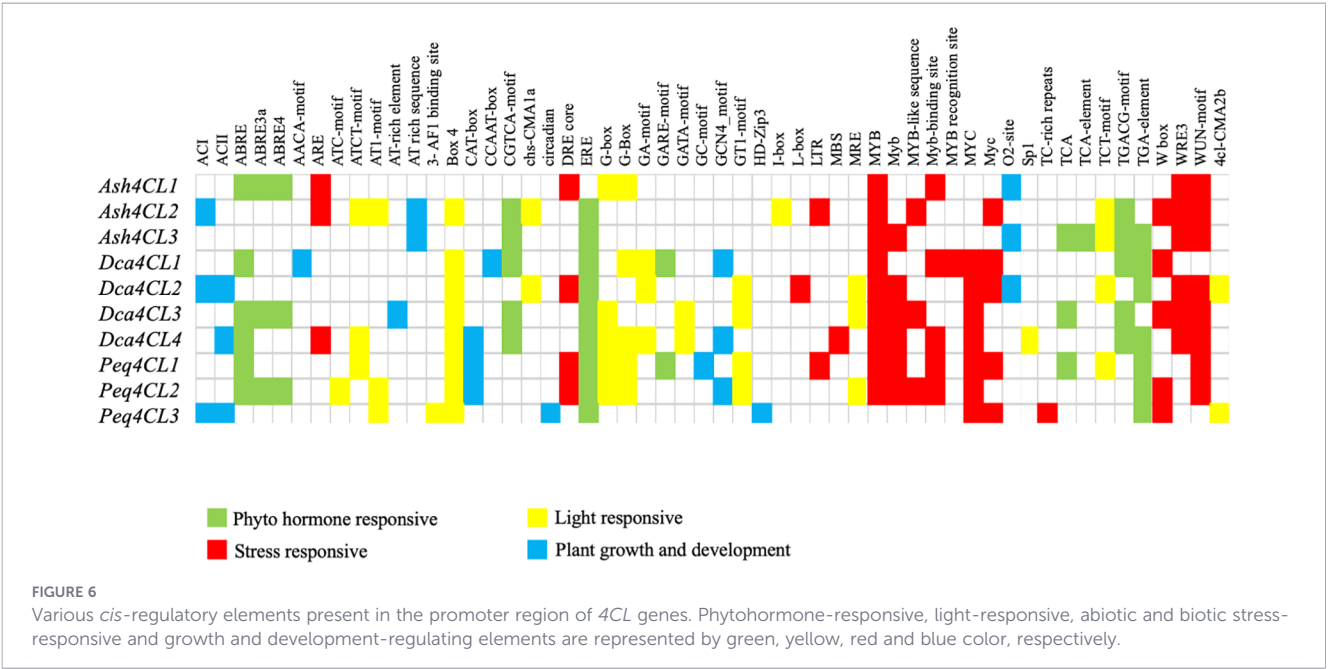
*c in the gene stretch column in the above table stands for complementary strand.

produced amongst these groups of plants. Furthermore, a close grouping of specific proteins in the phylogenetic tree illustrated their strong evolutionary links.

Gene architecture studies were performed by evaluating the number of exons/introns as well as intronic phases. 60% of the orchid *4CL* genes under study exhibited five introns and 6 exons, similar to *Ma4CL1-1b*, *Ma4CL2*, *Ma4CL3-2*, *Ma4CL5-1*, *Ma4CL5-2*, *Ma4CL8-1*, *Ma4CL11-1*, and *Ma4CL11-3* of *Musa acuminata* (Wang et al., 2024) and *MI4CL7* and *MI4CL8* of *Miscanthus lutarioriparius* (Huang et al., 2026). *4CL* genes identified in the present work were found to have a variety of phytohormone-responsive, abiotic and biotic stress-responsive, inducers of plant growth and development, and light-responsive *cis*-elements in their promoter regions. Similar *cis*-acting elements of 4 categories were predicted for the *Punica granatum* (Wang et al., 2022), *Triticum aestivum* (Gholizadeh et al., 2025), *Gastrodia elata* (Sha et al., 2025) and *Zea mays* (Zhang Z. et al., 2025a) *4CL* genes. It has also been shown that the *4CL* genes of *Gossypium hirsutum* (Sun et al., 2020),

Eucommia ulmoides (Zhong et al., 2022) and *Nicotiana tabacum* (Ma et al., 2026) contain *cis*-acting elements that regulate stress and phytohormone responses. The presence of several *cis*-regulatory elements related to phytohormones and stress responses, including MYB and MYC binding sites, suggests that *4CL* genes may be involved in plant responses to various biotic and abiotic stresses. MYB and MYC transcription factors are known to play important roles in regulating stress responses in plants, supporting the possible involvement of *4CL* genes in plant resilience under changing environmental conditions (Liu et al., 2024).

The expression profiles of *4CL* genes varied depending on the different vegetative and reproductive organs. *Ash4CL1*, *Ash4CL2* and *Dca4CL2* in the present study expressed highly in the male reproductive tissues, similar to the high expression of *Os4CL2* in anthers of *O. sativa* (Sun et al., 2013) and *Zm4CL27* in anthers of *Zea mays* (Zhang Z. et al., 2025a). This indicates their potential function in male tissue development. Roots showed increased expression of *Pbe4CL1-5* of *P. bellina*, *Pap4CL2-4* of *P. aphrodite*, *Peq4CL3* of *P.*



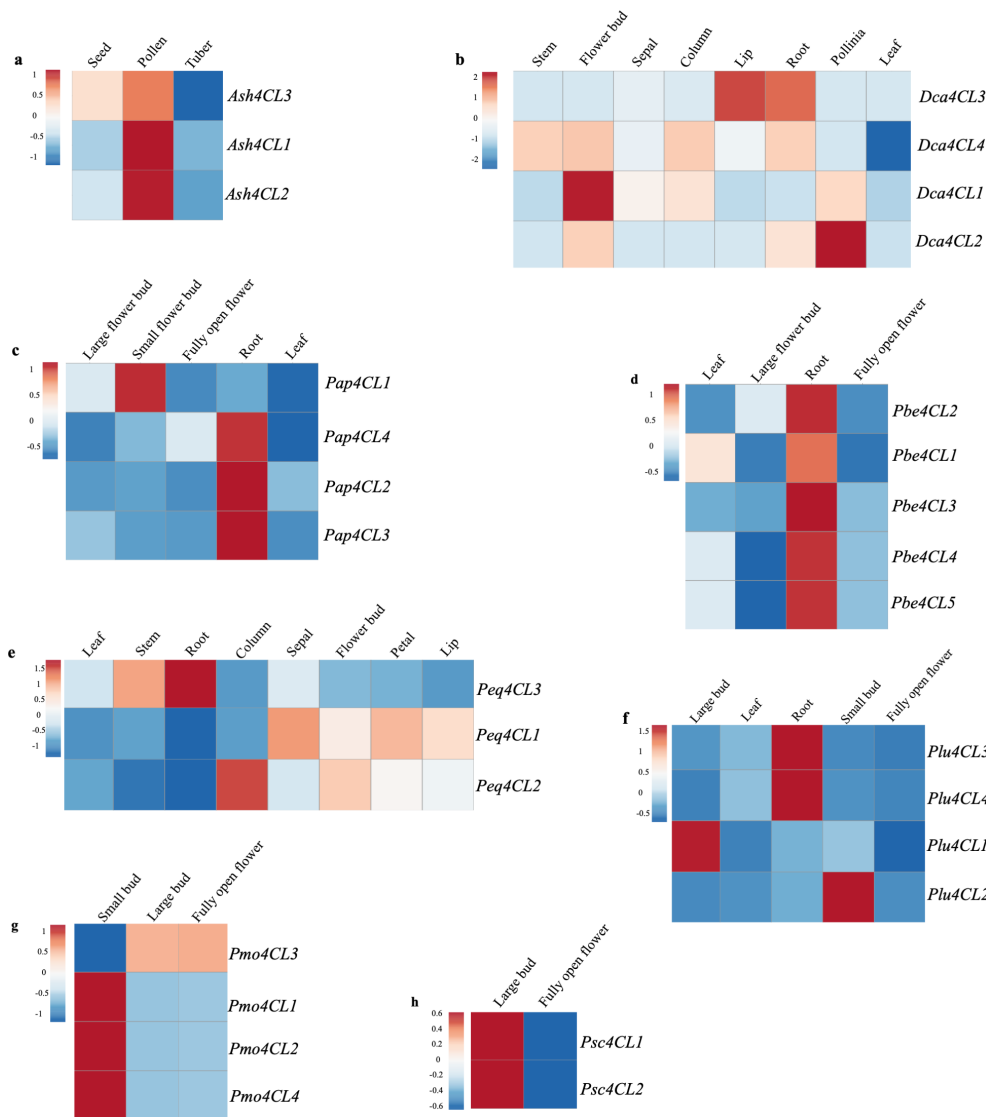


FIGURE 7

Expression profile of 4CL genes: (A) *Apostasia shenzhenica*, (B) *Dendrobium catenatum*, (C) *Phalaenopsis aphrodite*, (D) *Phalaenopsis bellina*, (E) *Phalaenopsis equestris*, (F) *Phalaenopsis lueddemanniana*, (G) *Phalaenopsis modesta*, (H) *Phalaenopsis schilleriana*.

equestris and *Plu4CL3–4* of *P. lueddemanniana*. *At4CL1* and *At4CL2* of *A. thaliana* (Lee et al., 1995), *Pto4CL1*, *Pto4CL4*, and *Pto4CL5* of *Populus tomentosa* (Rao et al., 2015). 4CL genes like *Ii4CL* of *Isatis indigotica* (Di et al., 2012), *Pg4CL5* of *Punica granatum* (Wang et al., 2022), *Ri4CL2–3* of *Rubus idaeus* (Kumar and Ellis, 2003), *Le4CL3* and *Le4CL4* of *Lithospermum erythrorhizon* (Nakanishi et al., 2024), *Nt4CL5*, *Nt4CL7*, and *Nt4CL8* of *Nicotiana tabacum* (Ma et al., 2026) and *Ta4CL90*, *Ta4CL96*, and *Ta4CL99* of *Triticum aestivum* (Zhang Z. et al., 2025b) all showed similar findings for high expression in the root which suggests the role of these genes in roots. Similarly, *Bna4CL1*, *Bna4CL3*, *Bna4CL4*, *Bna4CL8* and *Bna4CL11* of *Brassica napus* showed significantly higher expression in roots and stems (Zhang M. et al., 2025). This suggests that these genes are likely also involved in the vegetative growth of plants.

Further, more research on the regulation of the 4CL enzymes involved in the phenylpropanoid pathway can provide insights into plant metabolic processes and how they adapt to different

environmental conditions. Additionally, by comprehending the mechanisms that underlie the activity of 4CLs, specific targeted genetic modifications may also be made to improve the abundance of desired substances in plants.

Conclusions

In the present study, we identified and characterized 29 4CL genes in eight orchid species. The number and physicochemical properties of the identified proteins were found to be variable amongst the different orchid species under study. Phylogenetic analysis showed the clustering of 4CL genes into three clades. Our results provided an understanding of the potential functions of 4CL genes. However, the functional characterization and validation of the identified genes would be the part of future research.

Data availability statement

The original contributions presented in the study are included in the article/[Supplementary Material](#). Further inquiries can be directed to the corresponding author.

Author contributions

AK: Methodology, Investigation, Data curation, Writing – original draft. SP: Writing – review & editing. JS: Investigation, Conceptualization, Funding acquisition, Writing – review & editing, Supervision.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. Infrastructure support from Rashtriya Uchchatar Shiksha Abhiyan (RUSA).

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Allina, S. M., Pri-Hadash, A., Theilmann, D. A., Ellis, B. E., and Douglas, C. J. (1998). 4-coumarate: coenzyme A ligase in hybrid poplar: properties of native enzymes, cDNA cloning, and analysis of recombinant enzymes. *Plant Physiol.* 116, 743–754. doi: 10.1104/pp.116.2.743
- Awasthi, P., Mahajan, V., Jamwal, V. L., Chouhan, R., Kapoor, N., Bedi, Y. S., et al. (2019). Characterization of the gene encoding 4-coumarate: CoA ligase in *Coleus forskohlii*. *J. Plant Biochem. Biotechnol.* 28, 203–210. doi: 10.1007/s13562-018-0468-4
- Bailey, T. L., Johnson, J., Grant, C. E., and Noble, W. S. (2015). The MEME suite. *Nucleic Acids Res.* 43, W39–W49. doi: 10.1093/nar/gkv416
- Becker-André, M., Schulze-Lefert, P., and Hahlbrock, K. (1991). Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate: CoA ligase genes in potato. *J. Biol. Chem.* 266, 8551–8559. doi: 10.1016/s0021-9258(18)93010-3
- Chao, Y. T., Yen, S. H., Yeh, J. H., Chen, W. C., and Shih, M. C. (2017). Orchidstra 2.0-a transcriptomics resource for the orchid family. *Plant Cell. Physiol.* 58, e9–e9. doi: 10.1093/pcp/pcw220
- Chen, X., Su, W., Zhang, H., Zhan, Y., and Zeng, F. (2020). Fraxinus mandshurica 4-coumarate-CoA ligase 2 enhances drought and osmotic stress tolerance of tobacco by increasing coniferyl alcohol content. *Plant Physiol. Biochem.* 155, 697–708. doi: 10.1016/j.plaphy.2020.08.031
- Chou, K. C., and Shen, H. B. (2010). Plant-mPLoc: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS One* 5, e11335. doi: 10.1371/journal.pone.0011335
- Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* 16, 10881–10890. doi: 10.1093/nar/16.22.10881
- Costa, M. A., Collins, R. E., Anterola, A. M., Cochrane, F. C., Davin, L. B., and Lewis, N. G. (2003). An in silico assessment of gene function and organization of the phenylpropanoid pathway metabolic networks in *Arabidopsis thaliana* and limitations thereof. *Phytochemistry* 64, 1097–1112. doi: 10.1016/S0031-9422(03)00517-X
- DeLano, W. L. (2002). Pymol: an open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* 40, 82–92.
- Di, P., Hu, Y., Xuan, H., Xiao, Y., Chen, J., Zhang, L., et al. (2012). Characterization and the expression profile of 4-coumarate: CoA ligase (li4CL) from hairy roots of *Isatis indigotica*. *Afr. J. Pharm. Pharmacol.* 6, 2166–2175. doi: 10.5897/ajpp12.852
- Douglas, C., Hoffmann, H., Schulz, W., and Hahlbrock, K. (1987). Structure and elicitor or uv-light-stimulated expression of two 4-coumarate: CoA ligase genes in parsley. *EMBO J.* 6, 1189–1195. doi: 10.1002/j.1460-2075.1987.tb02353.x
- Ehrling, J., Büttner, D., Wang, Q., Douglas, C. J., Somssich, I. E., and Kombrink, E. (1999). Three 4-coumarate: coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant J.* 19, 9–20. doi: 10.1046/j.1365-3113.1999.00491.x
- Ferrer, J. L., Austin, M. B., Stewart, C. Jr, and Noel, J. P. (2008). Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant Physiol. Biochem.* 46, 356–370. doi: 10.1016/j.plaphy.2007.12.009
- Fu, F., Qin, H., Xin, Y., Li, Q., Kang, H., Han, L., et al. (2025). Characterization of 4-coumarate CoA ligase (4CL) gene family and functional study of Sm4CL2/3/7/9 in *Salvia miltiorrhiza*. *Funct. Integr. Genom.* 25, 35. doi: 10.1007/s10142-025-01541-0
- Gao, S., Liu, X. Y., Ni, R., Fu, J., Tan, H., Cheng, A. X., et al. (2024). Molecular cloning and functional analysis of 4-coumarate: CoA ligases from *Marchantia paleacea* and their roles in lignin and flavanone biosynthesis. *PLoS One* 19, e0296079. doi: 10.1371/journal.pone.0296079
- Gao, S., Yu, H. N., Xu, R. X., Cheng, A. X., and Lou, H. X. (2015). Cloning and functional characterization of a 4-coumarate CoA ligase from liverwort *Plagiochasma appendiculatum*. *Phytochemistry* 111, 48–58. doi: 10.1016/j.phytochem.2014.12.017
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S. E., Wilkins, M. R., Appel, R. D., et al. (2005). “Protein identification and analysis tools on the ExPASy server,” in *The Proteomics Protocols Handbook*. Ed. J. M. Walker (Totowa, New Jersey: Humana Press), 571–607.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2026.1895523/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

4CL proteins showing the distribution of secondary structures.

- Geourjon, C., and Deleage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics* 11, 681–684. doi: 10.1093/bioinformatics/11.6.681
- Ghai, D., Verma, J., Kaur, A., Thakur, K., Pawar, S. V., and Sembi, J. K. (2021). “Bioprospection of orchids and appraisal of their therapeutic indications,” in *Bioprospecting of Plant Biodiversity for Industrial Molecules*. Eds. S. K. Upadhyay and S. P. Singh (Hoboken, New Jersey: John Wiley & Sons Ltd), 401–424.
- Gholizadeh, F., Szalai, G., and Janda, T. (2025). Genome-wide analysis of the 4-coumarate-CoA ligase gene family in wheat: Expression patterns under salt stress and light-regulated cold acclimation. *Plant Stress* 16, 100879. doi: 10.1016/j.stress.2025.100879
- Govaerts, R., Bernet, P., Kratochvil, K., Gerlach, G., Carr, G., Alrich, P., et al. (2026). World checklist of orchidaceae. *Kew: Facilitated by the Royal Botanic Gardens*. Available online at: <https://wcp.science.kew.org/> (Accessed January 30, 2026).
- Gui, J., Shen, J., and Li, L. (2011). Functional characterization of evolutionarily divergent 4-coumarate: coenzyme A ligases in rice. *Plant Physiol.* 157, 574–586. doi: 10.1104/pp.111.178301
- Hu, B., Jin, J., Guo, A. Y., Zhang, H., Luo, J., and Gao, G. (2015). GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31, 1296–1297. doi: 10.1093/bioinformatics/btu817
- Hu, W. J., Kawaoka, A., Tsai, C. J., Lung, J., Osakabe, K., Ebinuma, H., et al. (1998). Compartmentalized expression of two structurally and functionally distinct 4-coumarate: CoA ligase genes in aspen (*Populus tremuloides*). *Proc. Natl. Acad. Sci.* 95, 5407–5412. doi: 10.1073/pnas.95.9.5407
- Huang, C., Wen, W., Li, Q., Wang, M., Xiao, S., Zhang, X., et al. (2023). Identification, characterization and expression analysis of the 4-coumarate-CoA ligase gene family in *Bletilla striata*. *Gene Rep.* 32, 101785. doi: 10.1016/j.genrep.2023.101785
- Huang, X., Zhou, X., Peng, Y., Fu, T., Li, M., Yi, Z., et al. (2026). Genome-wide identification and characterization of the 4-coumarate: CoA ligase (4CL) gene family in *Miscanthus lutarioriparius*: Transcriptional response to cadmium stress. *Agronomy* 16, 855. doi: 10.3390/agronomy16090855
- Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. *J. Biochem.* 88, 1895–1898. doi: 10.1002/047120918x.emb1245
- Jin, X. Q., Chen, Z. W., Tan, R. H., Zhao, S. J., and Hu, Z. B. (2012). Isolation and functional analysis of 4-coumarate: coenzyme A ligase gene promoters from *Salvia miltiorrhiza*. *Biol. Plant* 56, 261–268. doi: 10.1007/s10535-012-0085-3
- Jin, Z., Wungsintaweekul, J., Kim, S. H., Kim, J. H., Shin, Y., Ro, D. K., et al. (2020). 4-coumarate: coenzyme A ligase isoform 3 from *Piper nigrum* (Pn4CL3) catalyzes the CoA thioester formation of 3, 4-methylenedioxycinnamic and piperic acids. *Biochem. J.* 477, 61–74. doi: 10.1042/BCJ20190527
- Kaur, A., Sharma, K., Pawar, S. V., and Sembi, J. K. (2025). Genome-wide characterization of PAL, C4H, and 4CL genes regulating the phenylpropanoid pathway in *Vanilla planifolia*. *Sci. Rep.* 15, 10714. doi: 10.1038/s41598-024-81968-w
- Kumar, A., and Ellis, B. E. (2003). 4-coumarate: CoA ligase gene family in *Rubus idaeus*: cDNA structures, evolution, and expression. *Plant Mol. Biol.* 51, 327–340. doi: 10.1023/A:1022004923982
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547. doi: 10.1093/molbev/msy096
- Lan, X., Liu, N., Zhou, Q., Jin, M., Liu, Y., and Tang, L. (2026). 4CL gene family in red beet: Genome-wide identification, stress-responsive expression patterns, and their role in coordinating secondary metabolism and growth regulation. *Sugar Tech.* 28, 45–58. doi: 10.1007/s12355-025-01641-y
- Lavhale, S. G., Kalunke, R. M., and Giri, A. P. (2018). Structural, functional and evolutionary diversity of 4-coumarate-CoA ligase in plants. *Planta* 248, 1063–1078. doi: 10.1007/s00425-018-2965-z
- Lee, D., Ellard, M., Wanner, L. A., Davis, K. R., and Douglas, C. J. (1995). The *Arabidopsis thaliana* 4-coumarate: CoA ligase (4CL) gene: stress and developmentally regulated expression and nucleotide sequence of its cDNA. *Plant Mol. Biol.* 28, 871–884. doi: 10.1007/BF00042072
- Lee, J. M., Hammarén, H. M., Savitski, M. M., and Baek, S. H. (2023). Control of protein stability by post-translational modifications. *Nat. Commun.* 14, 201. doi: 10.1038/s41467-023-35795-8
- Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (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, 325–327. doi: 10.1093/nar/30.1.325
- Letunic, I., Doerks, T., and Bork, P. (2015). SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.* 43, D257–D260. doi: 10.1093/nar/gku949
- Li, Y., Kim, J. I., Pysh, L., and Chapple, C. (2015). Four isoforms of *Arabidopsis* 4-coumarate: CoA ligase have overlapping yet distinct roles in phenylpropanoid metabolism. *Plant Physiol.* 169, 2409–2421. doi: 10.1007/s13131-015-0588-9
- Lindermayr, C., Möllers, F., Fliegmann, J., Uhlmann, A., Lottspeich, F., Meimberg, H., et al. (2002). Divergent members of a soybean (*Glycine max* L.) 4-coumarate: coenzyme A ligase gene family: Primary structures, catalytic properties, and differential expression. *Eur. J. Biochem.* 269, 1304–1315. doi: 10.1046/j.1432-1033.2002.02775.x
- Liu, F., Xi, M., Liu, T., Wu, X., Ju, L., and Wang, D. (2024). The central role of transcription factors in bridging biotic and abiotic stress responses for plants’ resilience. *New crops.* 1, 100005. doi: 10.1016/j.ncrops.2023.11.003
- Ma, B., Song, W., Zhang, H., Peng, Y., Sun, S., Liu, W., et al. (2026). Characterization of the Nt4CL gene family in tobacco (*Nicotiana tabacum*) and reveals Nt4CL9 response to cold stress. *Ind. Crops Prod.* 242, 122870. doi: 10.1016/j.indcrop.2026.122870
- Ma, J., Zuo, D., Zhang, X., Li, H., Ye, H., Zhang, N., et al. (2024). Genome-wide identification analysis of the 4-coumarate: CoA ligase (4CL) gene family expression profiles in *Juglans regia* and its wild relatives *J. mandshurica* resistance and salt stress. *BMC Plant Biol.* 24, 211. doi: 10.1186/s12870-024-04899-8
- Ma, Z. H., Nan, X. T., Li, W. F., Mao, J., and Chen, B. H. (2023). Comprehensive genomic identification and expression analysis 4CL gene family in apple. *Gene* 858, 147197. doi: 10.1016/j.gene.2023.147197
- Metsalu, T., and Vilo, J. (2015). ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Res.* 43, W566–W570. doi: 10.1093/nar/gkv468
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226
- Nakanishi, K., Li, H., Ichino, T., Tatsumi, K., Osakabe, K., Watanabe, B., et al. (2024). Peroxisomal 4-coumaroyl-CoA ligases participate in shikonin production in *Lithospermum erythrorhizon*. *Plant Physiol.* 195, 2843–2859. doi: 10.1093/plphys/kiad157
- Nie, T., Sun, X., Wang, S., Wang, D., Ren, Y., and Chen, Q. (2023). Genome-wide identification and expression analysis of the 4-coumarate: CoA ligase gene family in *Solanum tuberosum*. *Int. J. Mol. Sci.* 24, 1642. doi: 10.3390/ijms24021642
- Ran, F., Xiang, C., Wang, C., Zang, Y., Liu, L., Wu, S., et al. (2024). Identification of the 4CL family in cassava (*Manihot esculenta* Crantz) and expression pattern analysis of the Me4CL32 gene. *Biochem. Biophys. Res. Commun.* 735, 150731. doi: 10.2139/ssrn.4659303
- Rao, G., Pan, X., Xu, F., Zhang, Y., Cao, S., Jiang, X., et al. (2015). Divergent and overlapping function of five 4-coumarate/coenzyme A ligases from *Populus tomentosa*. *Plant Mol. Biol. Rep.* 33, 841–854. doi: 10.1007/s11105-014-0803-4
- Ren, W. W., Gao, T., Shao, S. X., Hou, B. H., Liao, X. S., and Ye, N. X. (2023). Genome-wide identification and expression analysis of 4CL gene family in *Camellia sinensis*. *Acta Bot. Boreal.-Occid. Sin.* 43, 1459–1469. doi: 10.21203/rs.3.rs-1921086/v1
- Sha, S., Mu, K., Luo, Q., Yao, S., Tang, T., Sun, W., et al. (2025). Genome-wide identification and characterization of the 4-coumarate-CoA ligase (4CL) gene family in *Gastrodia elata* and their transcriptional response to fungal infection. *Int. J. Mol. Sci.* 26, 7610. doi: 10.3390/ijms26157610
- Sigrist, C. J., De Castro, E., Cerutti, L., Cucho, B. A., Hulo, N., Bridge, A., et al. (2012). New and continuing developments at PROSITE. *Nucleic Acids Res.* 41, D344–D347. doi: 10.1093/nar/gks1067
- Silber, M. V., Meimberg, H., and Ebel, J. (2008). Identification of a 4-coumarate: CoA ligase gene family in the moss, *Physcomitrella patens*. *Phytochemistry* 69, 2449–2456. doi: 10.1016/j.phytochem.2008.06.014
- Song, Y., Zhang, G., Chen, N., Zhang, J., and He, C. (2023). Metabolomic and transcriptomic analyses provide insights into the flavonoid biosynthesis in sea buckthorn (*Hippophae rhamnoides* L.). *LWT* 187, 115276. doi: 10.1016/j.lwt.2023.115276
- Stuible, H. P., Büttner, D., Ehlting, J., Hahlbrock, K., and Kombrink, E. (2000). Mutational analysis of 4-coumarate: CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. *FEBS Lett.* 467, 117–122. doi: 10.1016/S0014-5793(00)01133-9
- Sun, H., Li, Y., Feng, S., Zou, W., Guo, K., Fan, C., et al. (2013). Analysis of five rice 4-coumarate: coenzyme A ligase enzyme activity and stress response for potential roles in lignin and flavonoid biosynthesis in rice. *Biochem. Biophys. Res. Commun.* 430, 1151–1156. doi: 10.1016/j.bbrc.2012.12.019
- Sun, S. C., Xiong, X. P., Zhang, X. L., Feng, H. J., Zhu, Q. H., Sun, J., et al. (2020). Characterization of the Gh4CL gene family reveals a role of Gh4CL7 in drought tolerance. *BMC Plant Biol.* 20, 125. doi: 10.1186/s12870-020-2329-2
- Tang, Y. H., Liu, F., Mao, K. Q., Xing, H. C., Chen, J. R., and Guo, Q. Q. (2018). Cloning and characterization of the key 4-coumarate CoA ligase genes in *Boehmeria nivea*. *S. Afr. J. Bot.* 116, 123–130. doi: 10.1016/j.sajb.2018.02.398
- Uhlmann, A., and Ebel, J. (1993). Molecular cloning and expression of 4-coumarate: coenzyme A ligase, an enzyme involved in the resistance response of soybean (*Glycine max* L.) against pathogen attack. *Plant Physiol.* 102, 1147–1156. doi: 10.1104/pp.102.4.1147
- Vogt, T. (2010). Phenylpropanoid biosynthesis. *Mol. Plant* 3, 2–20. doi: 10.1093/mp/ssp106
- Wang, Y., Guo, L., Zhao, Y., Zhao, X., and Yuan, Z. (2022). Systematic analysis and expression profiles of the 4-coumarate: CoA ligase (4CL) gene family in pomegranate (*Punica granatum* L.). *Int. J. Mol. Sci.* 23, 3509. doi: 10.3390/ijms23073509
- Wang, M., Jia, L., Liu, Y., Zhang, C., Chen, Y., Lin, Y., et al. (2024). Genome-wide identification, evolutionary analysis of 4CL gene family and expression analysis in banana. *Food Agric. Immunol.* 35, 2287958. doi: 10.1080/09540105.2023.2287958

- Wang, Z., Yao, X. M., Jia, C. H., Xu, B. Y., Wang, J. Y., Liu, J. H., et al. (2023). Identification and analysis of lignin biosynthesis genes related to fruit ripening and stress response in banana (*Musa acuminata* L. AAA group, cv. Cavendish). *Front. Plant Sci.* 14, 1072086. doi: 10.3389/fpls.2023.1072086
- Wei, H. Y., Rao, G. D., Wang, Y. K., Zhang, L., and Lu, H. (2013). Cloning and analysis of a new 4CL-like gene in *Populus tomentosa*. *For. Sci. Pract.* 15, 98–104. doi: 10.1007/s11632-013-0204-z
- Xiao, T., Zhang, S., Gu, Y., Hu, H., Sun, L., Lu, C., et al. (2026). Systematic analysis of the F3H family in maize reveals a role for ZmF3H6 in salt stress tolerance. *New Crops*. 3, 100082. doi: 10.1016/j.ncrops.2025.100082
- Xu, B., Escamilla-Treviño, L. L., Sathitsuksanoh, N., Shen, Z., Shen, H., Percival Zhang, Y. H., et al. (2011). Silencing of 4-coumarate: coenzyme A ligase in switchgrass leads to reduced lignin content and improved fermentable sugar yields for biofuel production. *New Phytol.* 192, 611–625. doi: 10.1111/j.1469-8137.2011.03830.x
- Yang, Z., Wu, X., Ma, J., Yuan, M., Zhan, Y., Li, Y., et al. (2024). Genome-wide identification and expression profiling of 4-coumarate: coenzyme A ligase genes influencing soybean isoflavones at the seedling stage. *Crop Pasture Sci.* 75. doi: 10.1071/cp23147
- Zhang, L., Song, B., Li, B., Zhang, S., Liu, Y., Chen, G., et al. (2024). Genome-wide identification and expression analysis of fifteen gene families involved in anthocyanin synthesis in pear. *Horticulturae* 10, 335. doi: 10.3390/horticulturae10040335
- Zhang, M., Tian, M., Weng, Z., Yang, Y., Pan, N., Shen, S., et al. (2025). Genome-wide identification analysis of the 4-coumarate: CoA ligase (4CL) gene family in *Brassica U*'s triangle species and its potential role in the accumulation of flavonoids in *Brassica napus* L. *Plants* 14, 714. doi: 10.3390/plants14050714
- Zhang, Z., Wang, Y., and Li, R. (2025a). Genome-wide analysis of Zm4CL genes identifies Zm4CL8 regulating drought and salt tolerance in maize. *Agronomy* 15, 1100. doi: 10.3390/agronomy15051100
- Zhang, C., Zang, Y., Liu, P., Zheng, Z., and Ouyang, J. (2019). Characterization, functional analysis and application of 4-coumarate: CoA ligase genes from *Populus trichocarpa*. *J. Biotechnol.* 302, 92–100. doi: 10.1016/j.jbiotec.2019.06.300
- Zhang, Z., Yang, X., Ning, D., and Li, R. (2025b). Characterization of 4-coumarate-CoA ligase (4CL) genes in wheat uncovers Ta4CL91's role in drought and salt stress adaptation. *Plants* 14, 1301. doi: 10.3390/plants14091301
- Zhong, J., Qing, J., Wang, Q., Liu, C., Du, H., Liu, P., et al. (2022). Genome-wide identification and expression analyses of the 4-coumarate: CoA ligase (4CL) gene family in *Eucommia ulmoides*. *Forests* 13, 1253. doi: 10.3390/f13081253
- Zhou, G., Sun, X., Zhang, L., Zeng, X., Liu, G., and Sheng, O. (2022). Lignin metabolism plays an essential role in the formation of corky split vein caused by boron deficiency in 'Newhall' navel orange (*Citrus sinensis* Osb.). *Sci. Hortic.* 294, 110763. doi: 10.1016/j.scienta.2021.110763