



OPEN Genome-wide characterization of *PAL*, *C4H*, and *4CL* genes regulating the phenylpropanoid pathway in *Vanilla planifolia*

Arshpreet Kaur¹, Karan Sharma¹, Sandip V. Pawar² & Jaspreet K. Sembi¹✉

Phenylalanine ammonia lyase (PAL), *cinnamate 4-hydroxylase (C4H)* and *4-coumarate: CoA ligase (4CL)* genes encode the enzymes catalyzing the steps of the phenylpropanoid pathway which is responsible for the biosynthesis of a diverse range of therapeutically important phenylpropanoids. In the present study, identification and characterization of the *PAL*, *C4H* and *4CL* gene family in an economically and medicinally important orchid species, *Vanilla planifolia* was done. Six *PAL*, two *C4H* and five *4CL* proteins have been identified in *Vanilla planifolia*. All the amino acid residues related to the enzymatic activity were found to be conserved in all the identified proteins. Subcellular localization of Vp*PAL*, Vp*C4H* and Vp*4CL* proteins predicted their location in the cytoplasm, endoplasmic reticulum and peroxisome, respectively. Alpha helices and random coils predominated the secondary structure of these proteins. Gene structure analysis showed the presence of two introns in *C4H* genes while *PAL* and *4CL* genes had one and four introns present, respectively in the majority of members. The analysis of promoter sequences predicted *cis*-regulatory elements regulated by light, plant growth and development, phytohormones and abiotic and biotic stress conditions. Expression profiling of genes revealed variable relative expression for all the identified genes in various vegetative and reproductive tissues, suggesting their overall role in growth and development.

Keywords PAL, C4H, 4CL, Phenylpropanoids, Secondary metabolism, Orchids

Abbreviations

4CL	4-coumarate: CoA ligase
C4H	cinnamate 4-hydroxylase
MIO	4-methylidene-imidazolone-5-one
PAL	phenylalanine ammonia lyase

Phenylpropanoids comprise a diverse group of compounds involved mainly in plant defense, structural support, survival and adaptation to environmental perturbations^{1,2}. They also protect the plant against UV radiations, herbivores, and pathogens and mediate the plant-pollinator interactions by producing different floral pigments and scented products^{2,3}. Their biosynthesis occurs through the phenylpropanoid pathway. In the core phenylpropanoid pathway, phenylalanine is converted into activated hydroxycinnamic acid derivatives via the sequential action of *PAL* (phenylalanine ammonia lyase), *C4H* (cinnamate 4-hydroxylase) and *4CL* (4-coumarate: CoA ligase) enzymes (Fig. 1). The end product of this pathway acts as a precursor molecule for the biosynthesis of various secondary metabolites such as lignins, coumarins, benzoic acids, stilbenes, and flavonoids etc^{1,3}. Thus, this pathway originates from phenylalanine and ends up with the synthesis of a large class of phytochemicals.

In the first step of the pathway, phenylalanine is converted into *trans*-cinnamic acid in the presence of *PAL* enzyme via non-oxidative deamination of phenylalanine^{2,4,5}. This step basically channels the flow of carbon from primary metabolism into secondary metabolism, thereby interconnecting these two physiological processes in plants⁶. The enzyme *PAL* is present in all plants, some fungi and bacteria but is not present in animals⁷. The first *PAL* was identified in *Hordeum vulgare*⁴. Since then, researchers had gained interest in studying the *PAL* gene regulating this enzyme in numerous species of the plant kingdom like *Citrullus lanatus*⁸, *Eucalyptus grandis*⁹, *Malus domestica*¹⁰, and *Camellia sinensis*¹¹ etc. Numerous studies have consistently demonstrated that the *PAL*

¹Department of Botany, Panjab University, Chandigarh 160014, India. ²University Institute of Pharmaceutical Sciences, Panjab University, Chandigarh 160014, India. ✉email: jaspreet.sembi@pu.ac.in

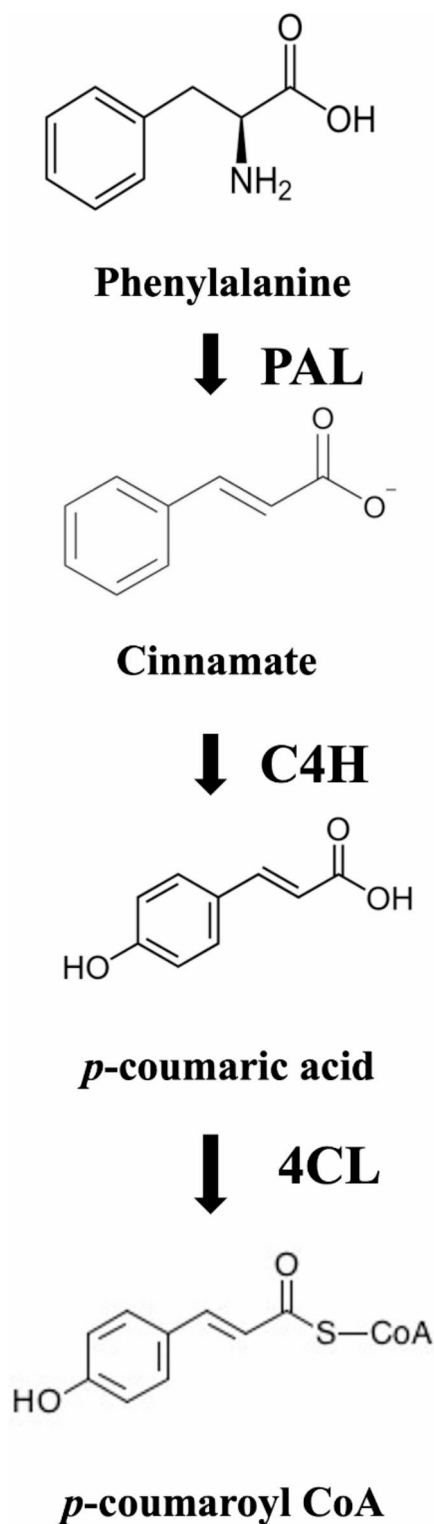


Fig. 1. Flowchart depicting the phenylpropanoid pathway with positions of PAL: phenylalanine ammonia-lyase; C4H: cinnamate 4-hydroxylase and 4CL: 4-coumarate: CoA ligase enzymes in the pathway.

gene exhibits a stress-responsive behavior. It is known to be activated by a range of environmental factors, including UV radiation¹², pathogen infections^{12,13}, tissue injury¹³, extreme temperatures¹⁴, nutrient depletion¹⁵, long term phosphate starvation¹⁶, salinity and water stress¹⁷, and other similar stimuli.

In the second step of the phenylpropanoid pathway, C4H catalyzes the hydroxylation of cinnamic acid or cinnamate, thus, yielding *p*-coumaric acid or 4-coumarate¹⁸. C4H, a cytochrome P-450 dependent monooxygenase was initially discovered in 1967 in pea seedlings¹⁹. Later on, studies related to C4H have been

conducted in various model plants such as *Oryza sativa*²⁰ and *Arabidopsis thaliana*^{5,21}. C4H proteins have been divided into two groups, C4H class I and C4H class II, wherein the class I members play a major role in lignin biosynthesis while class II members have been associated with stress responses in plants^{18,21,22}. Additionally, the expression profiling of C4H genes in various tissues during different growth stages has been evaluated in *Populus tremuloides*²³, *Populus trichocarpa*²⁰, *Leucaena leucocephala*²⁴, *Dryopteris fragrans*²⁵, and *Eucalyptus grandis*⁹. Further, a change in the expression of the C4H gene has also been observed in *Morus notabilis* in response to heavy metal stress²⁶ and in *Camellia sinensis* in response to wounding and abiotic stress conditions^{27,28}, thus highlighting the stress-responsive nature of C4H genes.

The last step of the pathway is marked with the formation of *p*-coumaroyl CoA from *p*-coumaric acid which is catalyzed by 4-coumarate: CoA ligase (4CL). Similar to C4H, the 4CL enzymes have also been divided into three different clusters namely, class I, class II and class III on the basis of evolutionary analysis^{29,30}. 4CL directs the flow of carbon from the core phenylpropanoid pathway into the biosynthesis of numerous phenylpropanoid-derived compounds²⁹. The first 4CL gene was cloned from *Petroselinum crispum*³¹. The characterization of 4CL genes has been done in a variety of plant species like *Glycine max*³², *Panicum virgatum*³³, *Populus tomentosa*³⁴, *Boehmeria nivea*³⁵, and *Citrus sinensis*³⁶ etc. The expression patterns of 4CL genes have been investigated in various tissues of several plants, during different stages of development^{29,35,37–41}, and in response to various triggers such as elicitors/phytohormones⁴², abiotic stress^{43,44}, and UV exposure³⁹ etc.

The number of members in the PAL, C4H and 4CL gene family varies considerably amongst different plants (Table 1). In the PAL gene family, 17 genes were found in *Brassica napus* which was the maximum number in this gene family⁴⁵. In the case of the C4H gene family, a single C4H member was identified in *Arabidopsis thaliana* while *Brassica napus* showed the presence of 10 C4H members^{5,21,46}. In a similar manner, the highest number of 4CL genes have been reported in *Malus domestica* which showed the presence of 69 genes⁴⁷. Thus, variations in the numerical strength of these gene families suggested that different members may play a role in production of several kinds of phenylpropanoids in plants. Moreover, different expression patterns in a variety of vegetative and reproductive tissues of PAL, C4H and 4CL genes depicted their role in plant growth and development.

Orchids are one of the largest families of flowering plants and harbour a wide variety of bioactive compounds known for their therapeutic importance^{90–92}. One such orchid capturing a huge market size and forming a multibillion-dollar market is *Vanilla planifolia*. Its highly valued phytochemical, vanillin, has a wide range of applications like usage as flavors and fragrance ingredients in ice-creams, confectionaries, milk products, perfumes etc. Vanillin is also exploited for its multiple therapeutic properties namely, anticancer and neuroprotective activity^{93–95}. A byproduct (a C₆-C₃ phenylpropanoid) of the phenylpropanoid pathway serves as a precursor in vanillin biosynthesis⁹⁶. Additionally, reports have demonstrated a positive correlation between the upregulated expression of the PAL and C4H genes and the accumulation of vanillin in *Vanilla planifolia*^{97,98}. Hence, the genome of *Vanilla planifolia* was sourced from NCBI⁹⁹ and PAL, C4H and 4CL genes were identified in the phenylpropanoid pathway and subjected to in silico characterization studies. The detailed study of these pivotal genes of the phenylpropanoid pathway will lay the groundwork for investigating its functional aspect in the production of vanillin in *V. planifolia*. In addition, it could also serve as a template for studying these genes in other plants and lead to the identification and characterization of novel, bioactive compounds through varied biotechnological approaches.

Materials and methods

Identification of PAL, C4H and 4CL proteins

For the current analysis, genome data of *Vanilla planifolia* [Accession: PRJNA753216; <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA753216/>] was taken from National Center for Biotechnology Information (NCBI) (<https://www.ncbi.nlm.nih.gov/>). BLASTp searches were conducted using protein sequences of PALs^{16,49}, C4Hs^{20,21} and 4CLs^{50,73} of *Arabidopsis thaliana* and *Oryza sativa* as query against the protein sequences of *V. planifolia* in NCBI, keeping all the parameters at their default values.

Analysis of conserved domains, motifs and multiple sequence alignment

The identification of protein sequences in *Vanilla planifolia* was confirmed by checking the presence of conserved domain, Lyase_aromatic (PF00221) for PAL, p450 (PF00067) for C4H and AMP-binding (PF00501) and AMP-binding_C (PF13193) domains for 4CL using SMART server (<http://smart.embl-heidelberg.de/>)¹⁰⁰. Further, the protein architecture of those proteins that had the conserved domain present was built using the My Domains – Image Creator tool in the ExPASy PROSITE server (<https://prosite.expasy.org/>)¹⁰¹. Conserved motifs and their location within the identified PAL, C4H and 4CL *Vanilla planifolia* proteins along with their counterparts in *Arabidopsis thaliana* and *Oryza sativa* were predicted by using Multiple Expectation Maximization for Motif Elicitation (MEME) Suite (version 5.5.5) (<https://meme-suite.org/meme/>)¹⁰². Any number of repetitions for a motif was allowed in the analysis and the maximum number of motifs to be detected was selected as 10. The minimum and the maximum motif width were set as 6 and 50, respectively, keeping all other parameters at the program's default values. In addition, multiple sequence alignment was also performed by aligning the protein sequences of *Vanilla planifolia*, *Arabidopsis thaliana* and *Oryza sativa* using MultAlin (<http://multalin.toulouse.inra.fr/multalin/>)¹⁰³ for further confirmation of the identified proteins by checking the presence of conserved regions.

In silico prediction of physicochemical properties

The sequence length (total amino acids), molecular weight, isoelectric point, instability index, aliphatic index and GRAVY (grand average of hydropathicity index) value of *Vanilla planifolia* proteins were predicted using PROTPARAM (<https://web.expasy.org/protparam/>)¹⁰⁴ and the sub-cellular localization was predicted by using Plant-mPLoc server (<http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/>)¹⁰⁵.

Plant Name	Family	Number of genes			Reference
		PAL	C4H	4CL	
<i>Apostasia shenzhenica</i>	Orchidaceae	5	-	-	48
<i>Arabidopsis thaliana</i>	Brassicaceae	4	1	4	5,21,49,50
<i>Arachis hypogaea</i>	Fabaceae	10	-	-	51
<i>Bletilla striata</i>	Orchidaceae	-	-	21	52
<i>Boehmeria nivea</i>	Urticaceae	-	-	3	35
<i>Brassica napus</i>	Brassicaceae	17	10	-	45,46
<i>Brassica oleracea</i>	Brassicaceae	8	-	-	53
<i>Brassica rapa</i>	Brassicaceae	7	-	-	53
<i>Camellia sinensis</i>	Theaceae	7	3	-	11,28
<i>Camellia sinensis</i> 'Huangdan'	Theaceae	-	-	8	54
<i>Camellia sinensis</i> 'Tie-guanyin'	Theaceae	-	-	9	54
<i>Carya illinoensis</i>	Juglandaceae	7	-	-	55
<i>Cephalotaxus hainanensis</i>	Taxaceae	4	-	-	56
<i>Citrullus lanatus</i>	Cucurbitaceae	12	-	-	8
<i>Citrus sinensis</i>	Rutaceae	4	1	3	36,57
<i>Coffea arabica</i>	Rubiaceae	6	-	-	58
<i>Coffea canephora</i>	Rubiaceae	4	-	-	58
<i>Coffea eugenioides</i>	Rubiaceae	4	-	-	58
<i>Coffea liberica</i>	Rubiaceae	3	-	-	58
<i>Corchorus olitorius</i>	Malvaceae	4	-	-	58
<i>Cucumis melo</i>	Cucurbitaceae	13	-	-	60
<i>Cucumis sativus</i>	Cucurbitaceae	10	-	-	61
<i>Cypripedium formosanum</i>	Orchidaceae	2	-	-	48
<i>Dendrobium catenatum</i>	Orchidaceae	2	-	-	48
<i>Eucalyptus grandis</i>	Myrtaceae	9	2	-	9
<i>Euryale ferox</i>	Nymphaeaceae	12	-	-	62
<i>Fagopyrum tataricum</i>	Polygonaceae	-	8	-	63
<i>Fallopia multiflora</i>	Polygonaceae	3	-	-	64
<i>Fragaria vesca</i>	Rosaceae	2	-	-	10
<i>Glycine max</i>	Fabaceae	-	4	20	65,66
<i>Hippophae rhamnoides</i>	Elaeagnaceae	-	-	7	67
<i>Juglans regia</i>	Juglandaceae	12	-	36	68,69
<i>Juglans mandshurica</i>	Juglandaceae	-	-	31	69
<i>Leucaena leucocephala</i>	Fabaceae	-	2	-	24
<i>Malus domestica</i>	Rosaceae	6	-	69	10,47
<i>Marchantia paleacea</i>	Marchantiaceae	-	-	4	70
<i>Medicago sativa</i>	Fabaceae	7	-	-	17
<i>Morus alba</i>	Moraceae	-	3	-	26
<i>Musa acuminata</i>	Musaceae	-	-	18	71
<i>Nelumbo nucifera</i>	Nelumbonaceae	3	-	-	72
<i>Oryza sativa</i>	Poaceae	9	4	5	16,20,73
<i>Panicum virgatum</i>	Poaceae	-	2	2	33,74
<i>Petroselinum crispum</i>	Apiaceae	-	1	-	75
<i>Phalaenopsis aphrodite</i>	Orchidaceae	2	-	-	48
<i>Phalaenopsis bellina</i>	Orchidaceae	2	-	-	48
<i>Phalaenopsis equestris</i>	Orchidaceae	2	-	-	48
<i>Phalaenopsis lueddemanniana</i>	Orchidaceae	2	-	-	48
<i>Phalaenopsis modesta</i>	Orchidaceae	2	-	-	48
<i>Phalaenopsis schilleriana</i>	Orchidaceae	2	-	-	48
<i>Physcomitrella patens</i>	Funariaceae	-	-	4	76
<i>Piper nigrum</i>	Piperaceae	-	-	3	41
<i>Populus kitakamiensis</i>	Salicaceae	-	3	-	77
<i>Populus tomentosa</i>	Salicaceae	-	-	5	34
<i>Populus trichocarpa</i>	Salicaceae	-	3	-	20
<i>Prunus mume</i>	Rosaceae	2	-	-	10
Continued					

Plant Name	Family	Number of genes			Reference
		PAL	C4H	4CL	
<i>Prunus persica</i>	Rosaceae	3	-	-	¹⁰
<i>Punica granatum</i>	Lythraceae	-	-	12	⁷⁸
<i>Pyrus bretschneideri</i>	Rosaceae	2	3	18	⁷⁹
<i>Raphanus sativus</i>	Brassicaceae	5	-	-	⁸⁰
<i>Rubus idaeus</i>	Rosaceae			3	³⁸
<i>Salix matsudana</i>	Salicaceae	-	9	-	⁸¹
<i>Salvia miltiorrhiza</i>	Lamiaceae	3	2	2	^{40,82,83}
<i>Scutellaria baicalensis</i>	Lamiaceae	-	1	2	⁸⁴
<i>Solanum lycopersicum</i>	Solanaceae	14	-	14	⁸⁵
<i>Solanum pennellii</i>	Solanaceae	11	-	11	⁸⁵
<i>Solanum tuberosum</i>	Solanaceae	-	-	10	⁸⁶
<i>Sorghum bicolor</i>	Poaceae	8	-	-	⁸⁷
<i>Triticum aestivum</i>	Poaceae	10	-	-	⁸⁸
<i>Vitis vinifera</i>	Vitaceae	15	-	-	⁸⁹
<i>Zea mays</i>	Poaceae	13	-	-	¹⁴

Table 1. Quantitative strength of PAL, C4H and 4CL gene family across various plant species.

Secondary structure prediction

The secondary structure of the *Vanilla planifolia* proteins depicting the percentages of alpha helices, extended strands, beta turns and random coils in the proteins was predicted by using the SOPMA online server (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html)¹⁰⁶.

Phylogenetic analysis

The full-length amino acid sequences of PAL proteins of *Vanilla planifolia* were aligned with the PAL proteins of *Cephalotaxus hainanensis*⁵⁶, *Arabidopsis thaliana*²¹, *Apostasia shenzhenica*⁴⁸, *Dendrobium catenatum*⁴⁸, *Phalaenopsis aphrodite*⁴⁸, *Phalaenopsis bellina*⁴⁸, *Phalaenopsis equestris*⁴⁸, *Phalaenopsis lueddemanniana*⁴⁸, *Phalaenopsis modesta*⁴⁸, *Phalaenopsis schilleriana*⁴⁸, *Sorghum bicolor*⁸⁷ and *Oryza sativa*¹⁶ using Muscle software with gaps. Then, the phylogenetic tree was constructed using the neighbor-joining method with pairwise deletion and 1000 bootstrap replicates by using Molecular Evolutionary Genetics Analysis (MEGA-XI)¹⁰⁷. In a similar manner, using the same parameters, a phylogenetic tree was constructed for C4H proteins of *Vanilla planifolia* along with the C4H proteins of *Arabidopsis thaliana*²¹, *Camellia sinensis*²⁸, *Fagopyrum tataricum*⁶³ and *Oryza sativa*²⁰. Further, to explore the evolutionary relationships amongst the various identified Vpl4CL proteins along with the 4CLs of *A. thaliana*⁵⁰, *Panicum virgatum*³³, *Scutellaria baicalensis*⁸⁴, *Solanum tuberosum*⁸⁶ and *O. sativa*⁷³, a phylogenetic analysis was also carried out.

Gene structure analysis

The genomic and CDS (coding) sequences of *Vanilla planifolia* genes were retrieved from the NCBI database. The gene architecture was constructed to analyze the distribution of exons and introns for each gene by comparing the genomic and coding sequences using Gene Structure Display Server (GSDS v.2.0) (<http://gsds.gao-lab.org/>)¹⁰⁸.

Promoter analysis

Promoter region sequences, 1.5 kb upstream from the start codon were retrieved from the NCBI database. Later on, the obtained promoter sequences of the genes were subjected to the PlantCARE database (<https://bioinformatics.psb.ugent.be/webtools/plantcare/html/>)¹⁰⁹ for predicting the type and position of *cis*-regulatory elements.

Gene expression analysis The CDS (coding sequences) of *V. planifolia* genes were used in order to perform a NCBI Sequence Read Archive (SRA) BLASTn search against the RNA sequence data derived from the tissues of *Vanilla planifolia* [stem (SRX11714032), leaf (SRX11714030), soil root (SRX11714033), aerial root (SRX11714034), flower bud (SRX11714036), flower (SRX11714031), ovary (SRX11714037), and fruit (SRX11714029)]. This sequence data is a part of the NCBI Bioproject accession: PRJNA753216⁹⁹. The number of hits were counted and subsequently, the RPKM values were estimated as $(C \times 10^9)/(N \times L)$, 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 gene¹¹⁰. Heat-maps were generated individually for VplPAL, VplC4H and Vpl4CL genes using the ClustVis visualization tool (<https://biit.cs.ut.ee/clustvis/>)¹¹¹ in order to study the relative expression of these genes in different tissues.

Results

Identification of PAL, C4H and 4CL proteins and conserved domain & motif analysis

On performing a BLASTp search using *Arabidopsis thaliana* and *Oryza sativa* PAL, C4H and 4CL protein sequences, a total of 16, 8 and 9 sequences of PAL, C4H and 4CL gene family, respectively were predicted in *Vanilla planifolia* based on query coverage, percentage identity, alignment score and e-value. Domain analysis

depicted that Lyase_aromatic (PF00221), the conserved domain of PAL proteins, was found to be absent in eight sequences and two members had a partial sequence and hence, these sequences were not taken for further analysis. Thus, the remaining six PAL sequences were considered for further characterization. Similarly, two non-redundant and complete C4H proteins were identified in *V. planifolia* by checking the presence of the conserved domain, p450 (PF00067). On these lines, only five 4CL sequences were identified that showed the presence of both AMP-binding (PF00501) and AMP-binding_C (PF13193) domains. Further, domain architecture showed the presence and position of the conserved domains in all the proteins of *V. planifolia* (Supplementary Table S1, Figs. 2a, 3a and 4a). A total of 10 conserved motifs (marked as 1–10) were predicted in VplPAL, VplC4H and Vpl4CL proteins. The lyase_aromatic domain of PAL proteins was depicted by motifs 1, 2, 3, 4 and 8 which were found to be present in all the PAL members of *V. planifolia*, *A. thaliana* and *O. sativa*. All these above motifs also consisted of catalytically active essential residues of PAL proteins (Fig. 2b). For C4H proteins, motif 4 which represented the p450 domain and consisted of the hinge motif sequence (PPGP) was present in all the C4H proteins (Fig. 3b). In case of 4CL proteins, motif 1, 2, 4, 6 and 8 occupied in the AMP_binding domain and were present in all the proteins except Vpl4CL5 where motif 6 was absent. Further, motifs 1 and 2 had Box II sequence present in them and substrate binding residues resided in motifs 2, 4, and 8. Motif 5 which was present in all the proteins represented the AMP_binding_C domain and had the residue essential for the enzymatic function of 4CL present in it (Fig. 4b).

Multiple sequence alignment

Multiple sequence alignment of VplPAL1-6 proteins along with their counterparts in *Arabidopsis thaliana* and *Oryza sativa* showed the presence of five conserved domains [N-terminal domain, MIO (4-methylidene-imidazolone-5-one) domain, core domain, shielding domain and C-terminal domain], that were characterized in all the identified PALs through MultAlin. A high degree of sequence conservation was depicted in all these domains except for some sequence divergence in the N-terminal domain. All the PAL proteins depicted that the three amino acid residues 'Ala-Ser-Gly' (ASG), playing a significant role in substrate binding and catalysis of the MIO-domain, were completely conserved. In addition, the 'FL' residue known for imparting substrate specificity to the PAL enzymes was also seen to be conserved. Alongside these conserved residues, other catalytically active sites such as GLALVNG, NDN, and HNQD were predominantly conserved in the majority of PAL proteins (Fig. 5).

The alignment of the protein sequences of VplC4H1 and VplC4H2 along with C4H proteins of *Arabidopsis thaliana* and *Oryza sativa* showed that all the residues related to C4H activity such as substrate recognition sites, ERR triad, heme-iron binding domain, hinge motif and enzymatic active sites were conserved. A total of five substrate binding sites were found in all proteins along with ERR triad and enzymatic active sites. The hinge motif denoted by the sequence (PPGP) and heme-iron binding domain (PFGVGRRSCPG) were also found in all the proteins (Fig. 6).

In 4CL proteins, the two signature motifs, Box I (SSGTTGLPKGV) and Box II (GEICIRG) were found to be conserved in Vpl4CL proteins upon aligning with the 4CL proteins of *A. thaliana* and *O. sativa*. Additionally, the amino acid residues involved in substrate binding and enzymatic function were conserved in the majority of the proteins (Fig. 7).

In silico prediction of physicochemical properties

The physicochemical properties of VplPAL, VplC4H and Vpl4CL were evaluated using various in silico tools (Table 2). The average length was found to be 717, 497 and 524 for VplPAL, VplC4H and Vpl4CL proteins, respectively. Molecular weight showed an average of 77.61 kDa for VplPAL proteins, 57.06 kDa for VplC4H proteins and 56.69 kDa for Vpl4CL proteins. The isoelectric point ranged from 5.57 in Vpl4CL4 to 9.18 in VplC4H2. The GRAVY value of both VplPAL and VplC4H proteins was negative. However, in the case of Vpl4CL proteins, all the proteins had positive GRAVY value except for Vpl4CL4. The instability index was below 40 for a majority of proteins except for VplPAL6, VplC4H1 and VplC4H2. Transmembrane helices were absent in all the proteins. Subcellular localization prediction revealed the localization of VplPAL, VplC4H and Vpl4CL proteins in the cytoplasm, endoplasmic reticulum and peroxisome, respectively.

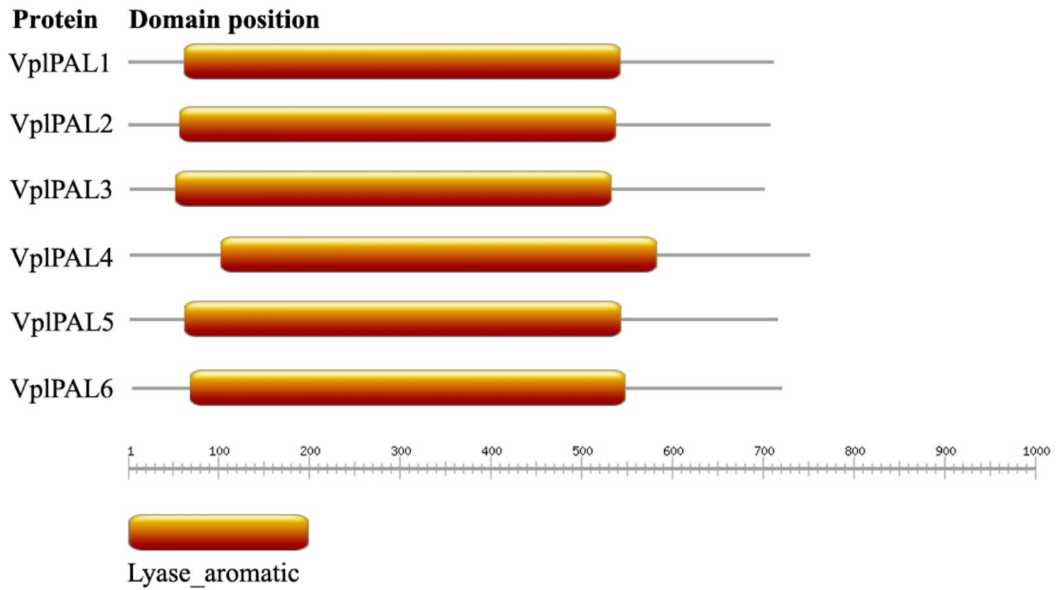
Secondary structure prediction

Analysis of secondary structures of VplPAL, VplC4H and Vpl4CL proteins showed that the alpha helices and random coils predominate the secondary structure of VplPAL and VplC4H proteins followed by low percentages of extended strands and beta turns. However, in Vpl4CL proteins, the percentage of alpha helix and random coil were almost similar. Additionally, the distribution of extended strands was also found to be higher in Vpl4CL proteins compared to the VplPAL and VplC4H proteins (Table 3; Fig. 8).

Phylogenetic analysis

The evolutionary analysis for PAL proteins showed that the PAL proteins belonging to gymnosperms, dicots and monocots formed three separate clades and orchids clustered together in the monocot clade (Fig. 9). In the case of C4H proteins, the VplC4H proteins clustered together with the class I members of C4H proteins of *A. thaliana* (AtC4H) and *O. sativa* (OsC4H1 and OsC4H4) (Fig. 10). Similarly, phylogenetic analysis in 4CL proteins depicted close clustering of Vpl4CL1 with class II members of 4CL proteins of *A. thaliana* (At4CL3) and *O. sativa* (Os4CL2) and Vpl4CL2-5 clustered with class III 4CL proteins of *O. sativa* (Os4CL1, 3, 4 and 5) and *Panicum virgatum* (Pv4CL1) (Fig. 11).

a



b

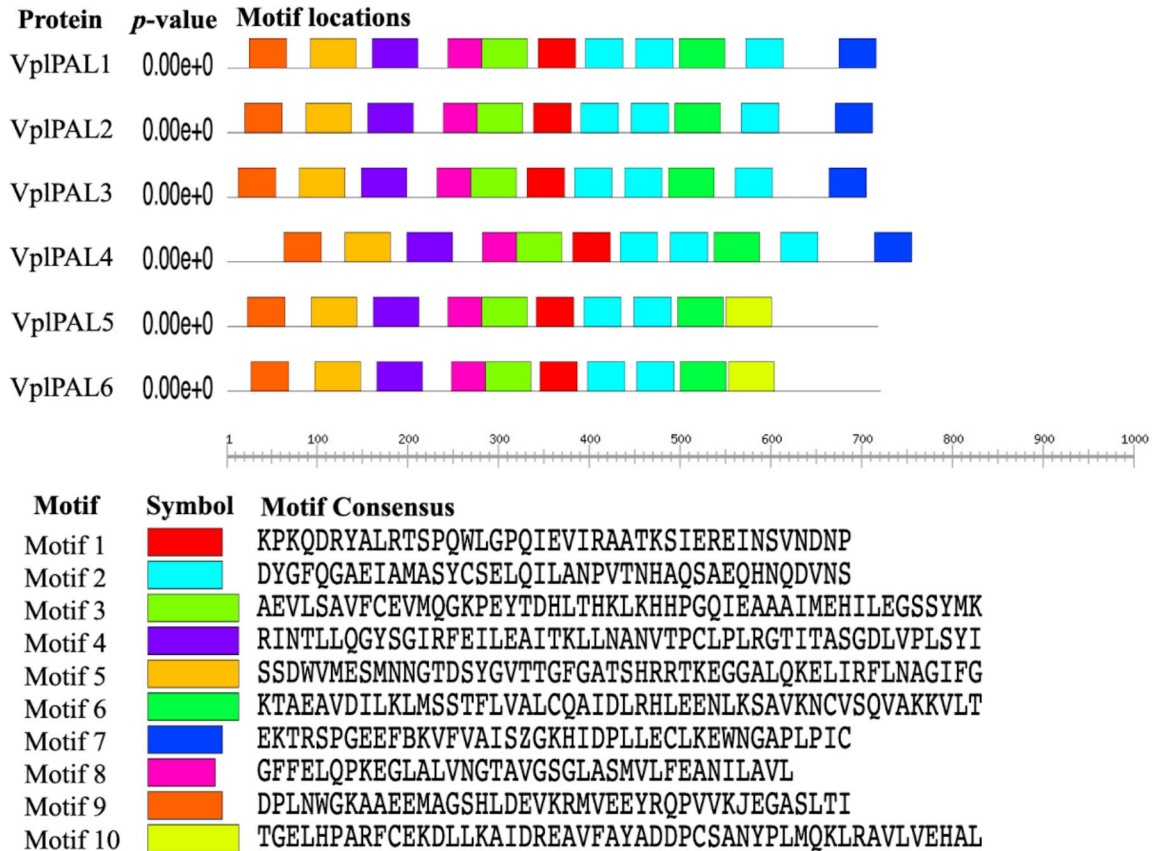
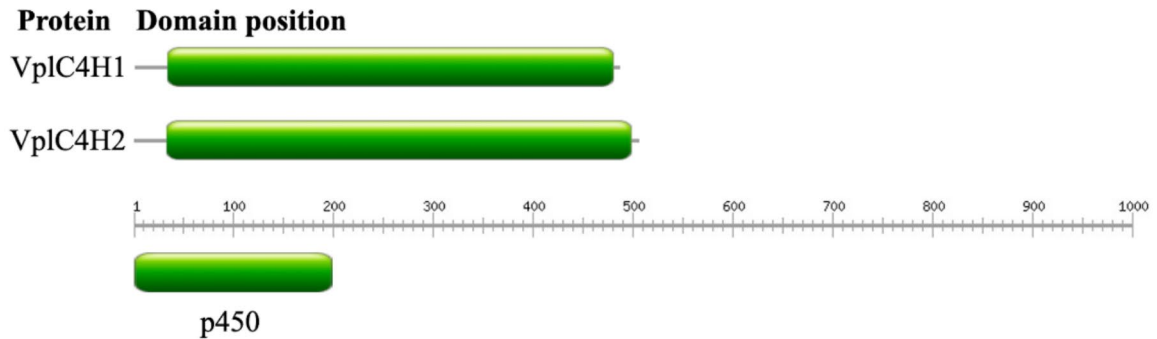


Fig. 2. VplPAL proteins showing the position of conserved domains and motifs.

Gene structure analysis

The exon-intron organization was similar in all the *VplPAL* genes with the presence of one intron in the biphasic phase except for the *VplPAL6* gene which had two introns present. An identical arrangement of exons and introns was observed for *VplCAH* genes as both the genes had two introns present; one in the monophasic intronic phase and the other in the biphasic intronic phase. Interestingly, multiple introns were present in *Vpl4CL* genes. *Vpl4CL1* showed the presence of six introns and seven exons while *Vpl4CL2-5* genes had four introns (Supplementary Table S2, Fig. 12).

a



b

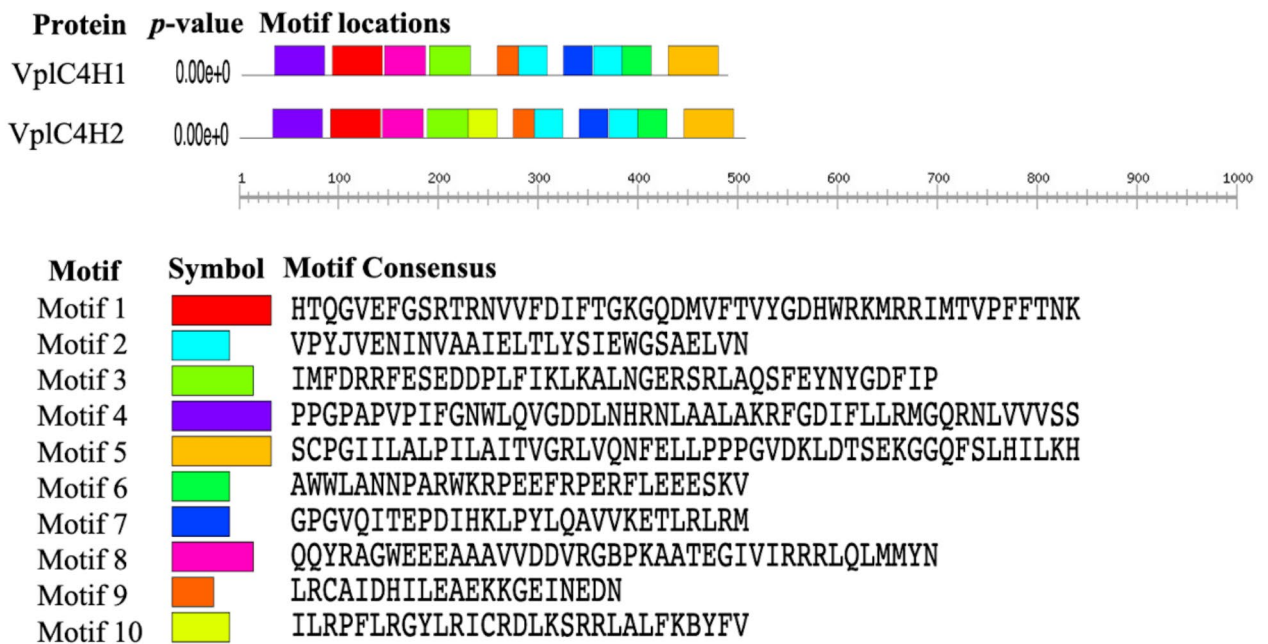


Fig. 3. VplC4H proteins showing the position of conserved domains and motifs.

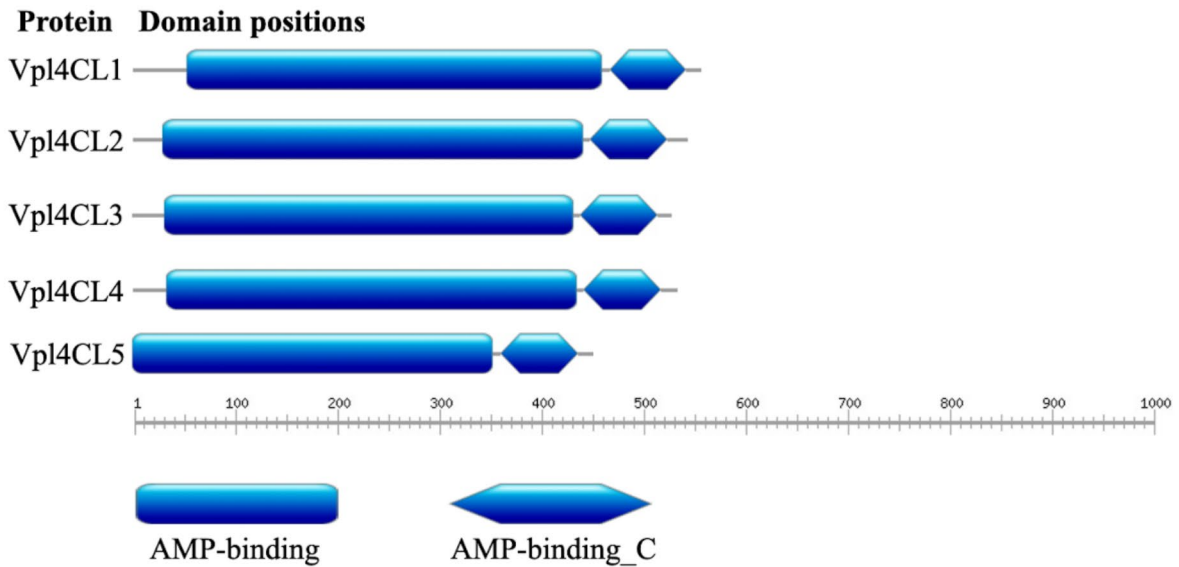
Promoter analysis

On evaluating the promoter sequences of all the *VplPAL*, *VplC4H* and *Vpl4CL* genes, apart from the *cis*-acting elements such as CAAT and TATA boxes which are found commonly in all the genes, various other elements were also identified. These elements regulate four basic responses in plants; plant growth and development, phytohormone response, abiotic and biotic stress response and light response (Supplementary Table S3). Stress-responsive and light-responsive elements were found in more abundance in comparison to elements regulating phytohormone responses and plant growth and development. The *cis*-acting elements like ACI, ACII and O2-site identified in the present study played a critical role in plant growth and development. Elements that showed phytohormone responsiveness are ABRE, CGTCA-motif, ERE, P-box, TGACG-motif and TCA element. Some stress-responsive elements that were detected included ARE, DRE core, MYB, MYC, MYB-like sequence, WRE and WUN-motif. Along with them, some light responsive *cis*-acting elements such as AE Box, Box 4, G-Box, GA motif, GATA motif, GT1-motif, MRE and chs-CMA1a were also identified (Fig. 13).

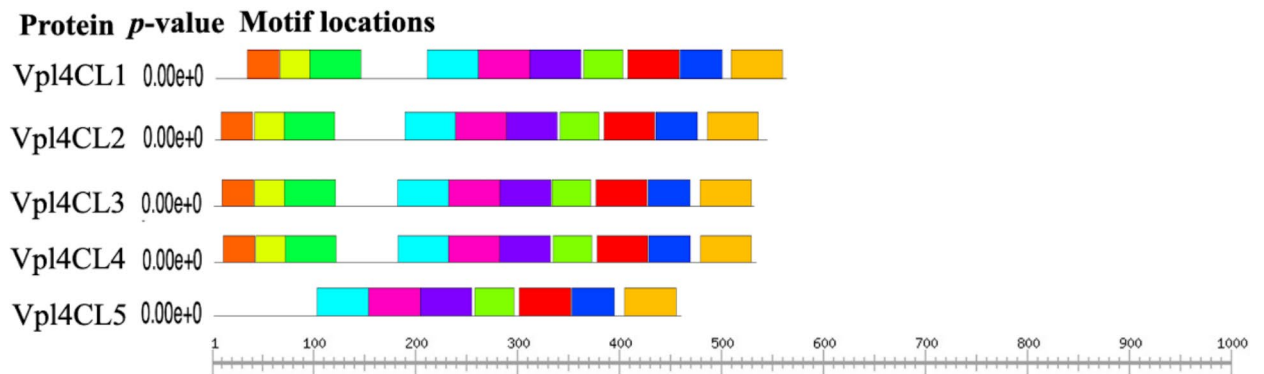
Gene expression analysis

The relative expression patterns of *VplPAL*, *VplC4H* and *Vpl4CL* genes were investigated in a variety of vegetative and reproductive tissues of *Vanilla planifolia* (Supplementary Table S4). *VplPAL5* and *VplPAL6* genes were expressed at elevated levels in the ovary while *VplPAL2-4* and *Vpl4CL2* showed high expression in fruit relative to the other tissues. Both *VplC4H1* and *VplC4H2* shared a similar expression profile by depicting high expression in the flower and relatively lower expression in the rest of the tissues under study. Lower expression of *VplPAL*, *VplC4H* and *Vpl4CL* genes was observed in the vegetative tissues compared to the reproductive tissues (Fig. 14).

a



b



Motif	Symbol	Motif Consensus
Motif 1	[Red]	GEICIRGEQIMKGYLNDPEATARTIDKEGWLHTGDIGYVDDDDDEIFIVDR
Motif 2	[Cyan]	SGTTGLPKGVMLTHRGLVTSVAQQVDGENPNLYFNEDDVILCVLPMFHIY
Motif 3	[Light Green]	AMCLAFAKEPFPVKSGSCGTVVRNAELKIVDPDTGASL
Motif 4	[Purple]	KSPAVEAYDLSSVRMVMGAAPLGKELZDALMAKLPNAVLGQYGMTEAG
Motif 5	[Orange]	GSEITEDEIKQYVAKQVVFYKRJHKVFFVDSIPKSPSGKILRKDLRARLA
Motif 6	[Green]	HRLGVRQGDVVMLLLNSPEFAFSFLAASRLGAVTTTANPFYTPAEIHKQ
Motif 7	[Blue]	KELIKYKGFQVAPAELEALLIAHPSIADAADVPMKDEVAGE
Motif 8	[Pink]	SLNSVLLCGLRAGAAIVIMRKFELGAMLELIZRHKVTVAPFVPPIVEJA
Motif 9	[Light Orange]	EVIFRSKLPDIEIPNHLPLHDYCFERLPEVA
Motif 10	[Yellow]	RPCLIDGATGEIYTYADVHVLRRVAAGL

Fig. 4. Vpl4CL proteins showing the position of conserved domains and motifs.

Discussion

PALs, C4Hs and 4CLs are the key enzymes of the core phenylpropanoid pathway which contribute towards the synthesis of phenylpropanoids that act as precursor molecules for the production of a myriad of compounds that play a role in plant growth and development, defense against pathogens and response to environmental cues^{2,3}. Although PALs, C4Hs, and 4CLs play crucial roles in plant metabolism, they are not characterized by a large number of members in most plant species. In silico characterization of genes has emerged as a crucial research technique in molecular biology to comprehend the different metabolic pathways and there are no earlier reports on the characterization of PAL, C4H and 4CL gene family in *Vanilla planifolia*. In the present

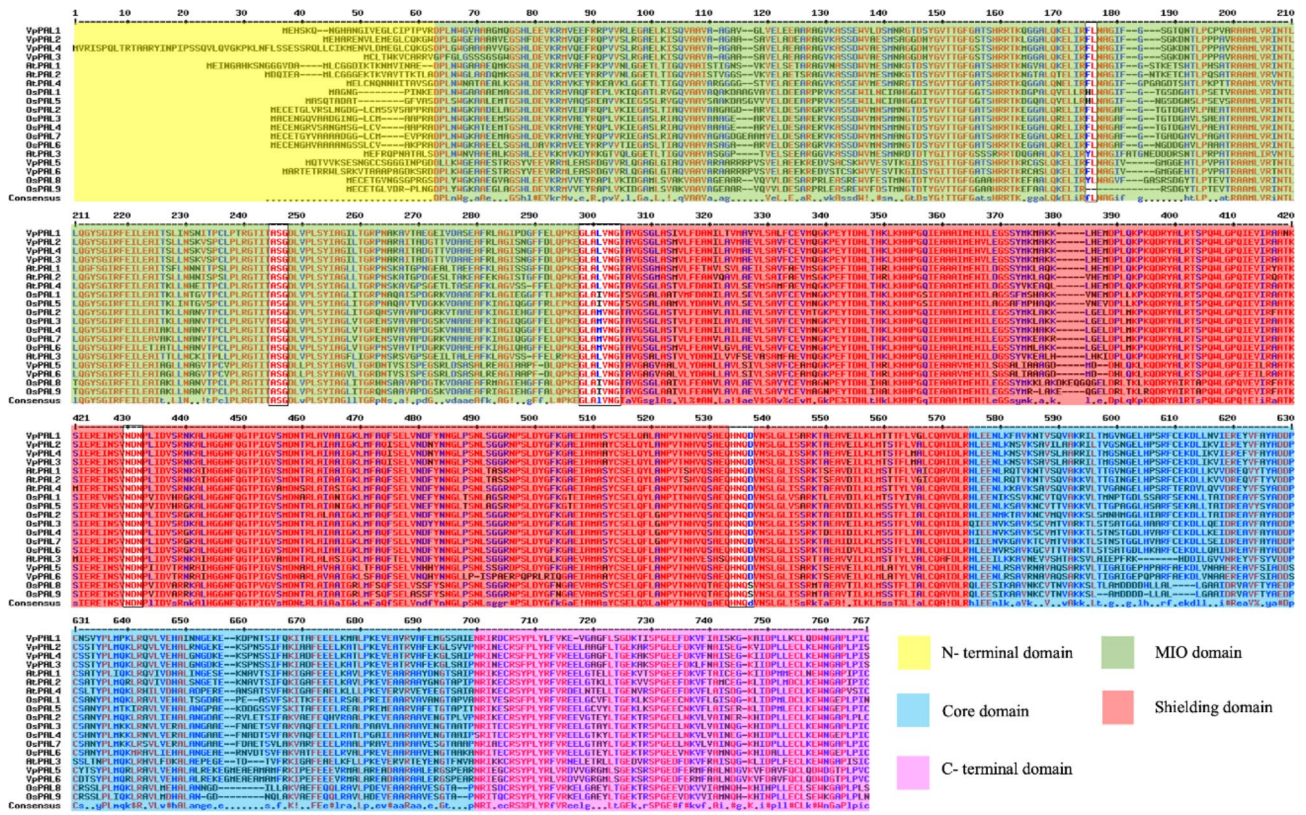


Fig. 5. Multiple sequence alignment showing the conserved regions along with the catalytically essential active residues of PAL proteins.

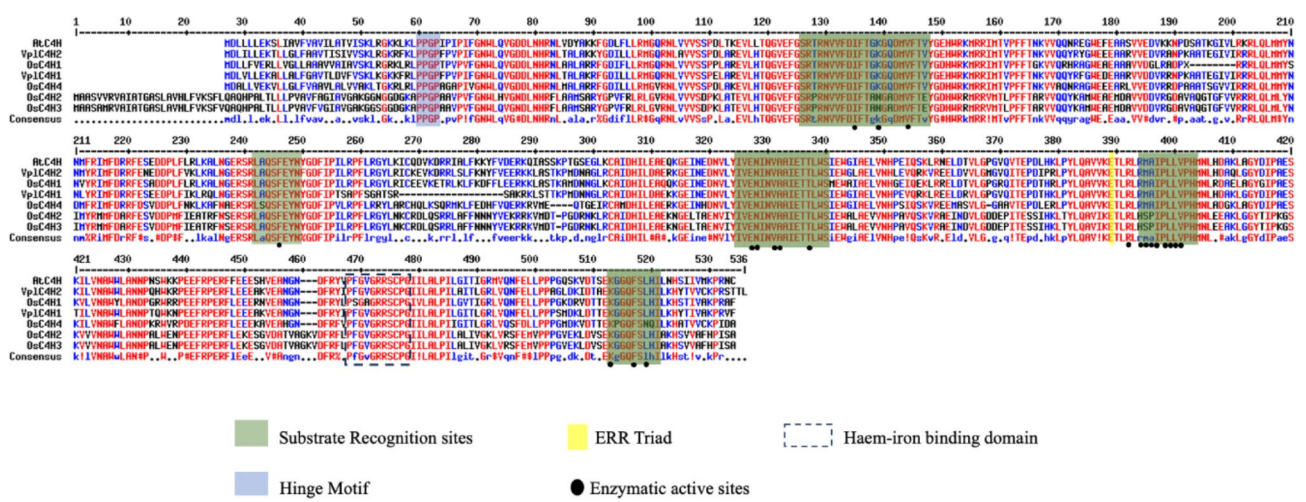


Fig. 6. Multiple sequence alignment showing the presence of conserved residues required for the enzymatic activity of C4H proteins.

research, six PAL genes have been identified in *Vanilla planifolia* similar to the identification of six PALs in *Malus domestica*¹⁰ *Eucalyptus grandis*⁹, *Leucaena leucocephala*²⁴ and *Salvia miltiorrhiza*³³ consist of two C4H genes which is consistent with the number of C4H genes characterized in *Vanilla planifolia*. Further, five 4CL genes have been predicted in *Vanilla planifolia* similar to the model plant *Oryza sativa*⁷³ and another plant *Populus tomentosa*³⁴ which also possessed five 4CL genes. Thus, a varied number of PALs, C4Hs and 4CLs have been reported amongst various plant species.

All the VpPAL and VpL4C proteins showed the presence of lyase_aromatic and p450 domain, respectively while in VpL4C proteins two domains, AMP-binding and AMP-binding_C were predicted. The presence

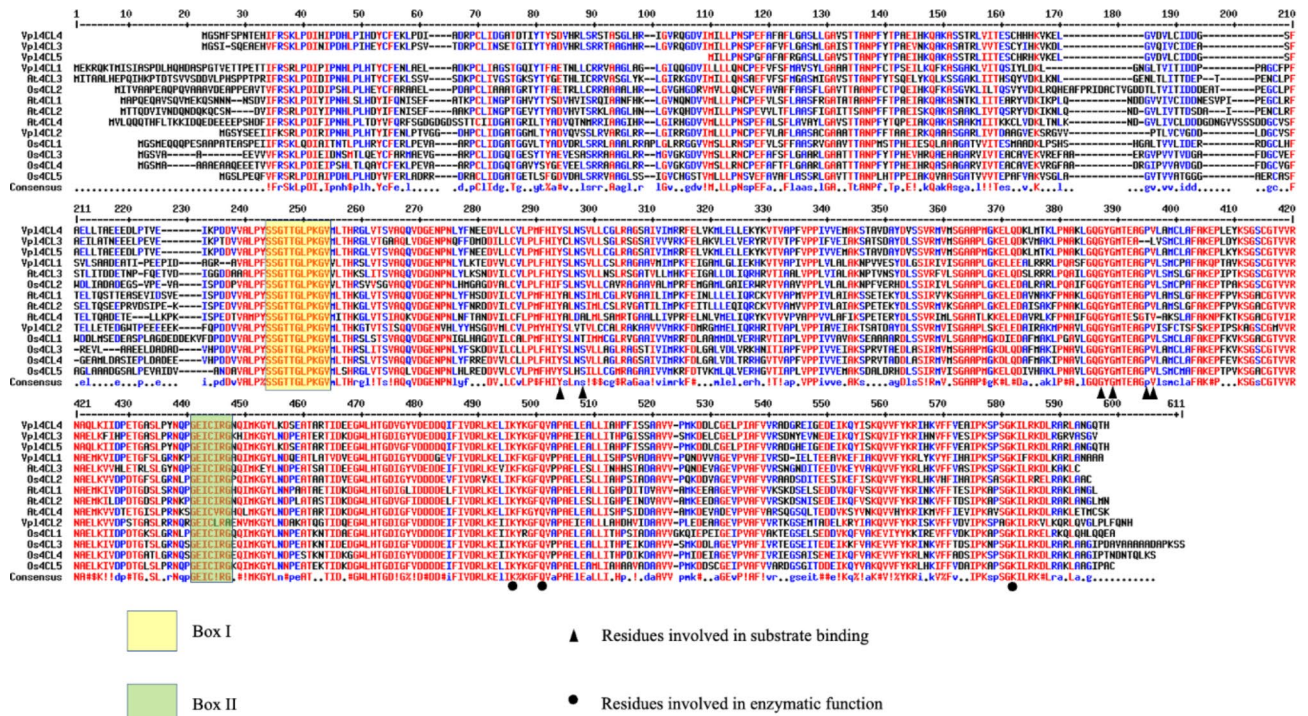


Fig. 7. Sequence alignment of Vpl4CL proteins with At4CL and Os4CL proteins.

Gene	NCBI ID	Protein length (aa)	MW (kDa)	pI	Instability index	Aliphatic index	GRAVY value	TM	Subcellular localization
VplPAL1	KAG04447363.1	712	77.31	6.04	35.31	94.55	-0.073	0	Cytoplasm
VplPAL2	KAG0488634.1	708	76.78	6.08	36.51	92.92	-0.118	0	Cytoplasm
VplPAL3	KAG0490360.1	701	75.85	6.78	37.06	91.36	-0.112	0	Cytoplasm
VplPAL4	KAG0488630.1	751	81.23	6.78	36.5	93.06	-0.103	0	Cytoplasm
VplPAL5	KAG0486323.1	714	76.88	6.04	39.92	94.22	-0.081	0	Cytoplasm
VplPAL6	KAG0484528.1	717	77.65	6.44	43.95	94.8	-0.116	0	Cytoplasm
VplC4H1	KAG0487879.1	487	55.75	9.07	47.42	95.85	-0.244	0	Endoplasmic reticulum
VplC4H2	KAG0484021.1	507	58.37	9.18	46.23	101.3	-0.18	0	Endoplasmic reticulum
Vpl4CL1	KAG0474598.1	558	59.93	5.69	38.32	103.66	0.205	0	Peroxisome
Vpl4CL2	KAG0475152.1	544	59.08	5.82	38.51	98.57	0.089	0	Peroxisome
Vpl4CL3	KAG0479627.1	529	58.23	5.71	39.61	99.49	0.024	0	Peroxisome
Vpl4CL4	KAG0470711.1	534	58.71	5.57	37.85	97.68	-0.001	0	Peroxisome
Vpl4CL5	KAG0472291.1	453	49.52	5.77	35.99	99.01	0.053	0	Peroxisome

Table 2. Physicochemical parameters of VplPAL, VplC4H and Vpl4CL proteins.

of the conserved domains implied a structural similarity between all the members of a gene family. Further, motif analysis showed that the predicted motifs consisted of the catalytically essential and substrate binding amino acid residues and the conserved distribution and arrangement of PAL, C4H and 4CL specific motifs also pinpointed the highly conserved nature of PAL, C4H and 4CL gene families making them an important part of the phenylpropanoid pathway.

Multiple sequence alignment of VplPAL proteins along with PALs of *Arabidopsis thaliana* and *Oryza sativa* showed a high degree of sequence similarity among them and the existence of five conserved domains [N-terminal, MIO (4-methylidene-imidazole-5-one) domain, core domain, shielding domain and C-terminal domain]. The Ala-Ser-Gly triad of the MIO domain, which is crucial for the enzymatic activity of PALs, was also present in all of the proteins^{112,113}. As the protein domains are distinct protein sequences forming discrete tertiary structures linked to specific functions like catalysis or binding, identifying the conserved domains in proteins is indicative of its molecular or cellular roles. All PALs shared the Phe-Hle residue that gives the PAL enzyme its substrate specificity, indicating that they all accept phenylalanine as a substrate¹¹⁴. The conservation of other catalytically active residues (GLALVNG, YNDN, and HNQD) point towards the catalytic activity and conserved nature of all the discovered VplPALs and is in line with the research on PALs in *Dendrobium*

Sequence ID	Alpha helix	Extended strand	Beta turn	Random coil
VplPAL1	397 (55.76%)	53 (7.44%)	49 (6.88%)	213 (29.92%)
VplPAL2	420 (59.32%)	41 (5.79%)	42 (5.93%)	505 (28.95%)
VplPAL3	399 (56.92%)	50 (7.13%)	44 (6.28%)	208 (29.67%)
VplPAL4	414 (55.13%)	63 (8.39%)	43 (5.73%)	231 (30.76%)
VplPAL5	396 (55.46%)	60 (8.40%)	46 (6.44%)	212 (29.69%)
VplPAL6	412 (57.46%)	56 (7.81%)	45 (6.28%)	204 (28.45%)
VplC4H1	230 (47.23%)	67 (13.76%)	21 (4.31%)	169 (34.70%)
VplC4H2	242 (47.73%)	65 (12.82%)	23 (4.54%)	177 (34.91%)
Vpl4CL1	178 (31.90%)	118 (21.15%)	39 (6.99%)	223 (39.96%)
Vpl4CL2	180 (33.09%)	108 (19.85%)	46 (8.46%)	210 (38.6%)
Vpl4CL3	173 (32.7%)	109 (20.60%)	37 (6.99%)	210 (39.70%)
Vpl4CL4	171 (32.02%)	110 (20.60%)	40 (7.49%)	213 (39.89%)
Vpl4CL5	145 (32.01%)	103 (22.74%)	35 (7.73%)	170 (37.53%)

Table 3. Distribution of secondary structures in VplPAL, VplC4H and Vpl4CL proteins.

*candidum*¹¹⁵ and *Vanda coerulea*¹¹⁶. Similarly, VplC4H proteins showed the presence of all residues related to C4H activity such as substrate binding sites, enzymatic active site, ERR triad, hinge motif and heme-iron binding domain similar to *Camellia sinensis* (28) and *Saccharum spontaneum*¹¹⁷. Further, sequence alignment of Vpl4CL proteins depicted the conservation of the two signature motifs Box I (SSGTTGLPKGV) and Box II (GEICIRG) of the 4CL proteins which are present in 4CL proteins in all plants^{43,78,118–122}. Box I is the AMP (Adenosine monophosphate) nucleotide binding motif and is conserved in all the proteins belonging to the adenylate forming enzyme family^{123,124}. In addition, the cysteine in the GEICIRG motif has a role in the stability and catalytic activity of 4CLs¹²⁵. Thus, the conserved nature of amino acid residues in all these proteins highlighted the consistency in the functionality of these proteins.

A bioinformatics approach was employed to characterize the physicochemical properties of the deduced proteins. The average molecular weight and isoelectric point range of PAL proteins were similar to findings from research on other plants like *Salvia miltiorrhiza*⁸², *Citrullus lanatus*⁸, *Salix viminalis*¹²⁶, *Cucumis sativus* and *Cucumis melo*⁶⁰. The amino acid length of VplC4H2 protein (507aa) was found to be in equivalence to SmC4H1 (504aa) of *Salvia miltiorrhiza*^{83,127}, SsC4H4.1–1 A (505aa) of *Saccharum spontaneum*¹¹⁷ and CsC4Ha and CsC4Hb (505aa) of *Camellia sinensis*²⁸, and C4H8 of *Fagopyrum tataricum*⁶³. The molecular weight and isoelectric point of SmC4H1 of *Salvia miltiorrhiza*^{83,127} was also comparable to the average molecular weight and isoelectric point of VplC4H proteins, respectively. Similarly, the protein length and molecular weight of Vpl4CL1 was similar to 4CL of *Neosinocalamus affinis*¹²⁸. The cytoplasmic localization of PAL proteins in the current study was consistent with earlier reports^{55,59,88,129,130}. Similarly, VplC4H proteins were located in the endoplasmic reticulum which is in line with GmC4Hs of *Glycine max*⁶⁵ and the localization of Vpl4CL proteins in the peroxisome is conforming to reports on many 4CL members of *Gossypium hirsutum*⁴³ and Pg4CL10 of *Punica granatum*⁷⁸. Proteins identified in *Vanilla planifolia* in the present study had no transmembrane regions and likewise in other plants such as *Salvia miltiorrhiza*⁸³, *Boehmeria nivea* (35) and *Fagopyrum tataricum*⁶³, the trans-membrane helices are absent in PALs, C4Hs and 4CL members. The majority of the discovered PAL, C4H and 4CL proteins had instability indices less than 40, indicating towards their stable nature. Further, negative GRAVY value in specific proteins pointed towards their polar and hydrophilic character. PAL proteins of hydrophilic nature have been detected in *Ornithogalum saundersiae*¹³¹ and *Cephalotaxus hainanensis*⁵⁶. However, positive GRAVY value in all the Vpl4CL proteins except Vpl4CL4 is in conformity with a previous report of hydrophobicity of Bn4CL3 of *Boehmeria nivea*³⁵ and Pg4CL1-3, Pg4CL6 and Pg4CL8-11 of *Punica granatum*⁷⁸. Thus, similarities in the physical parameters of the deduced PALs, C4Hs and 4CLs in the present study to the already identified members in other plant species corroborated the conserved nature of these proteins.

Secondary structure prediction showed that alpha helices and random coils constituted a major proportion in VplPAL and VplC4H proteins, hinting towards the importance of these secondary structure elements for structural stability and catalytic function. These results were also comparable with PAL and C4H proteins identified in other plant species^{14,27,56,72,130,132}.

According to the phylogenetic analysis, the VplC4H proteins were predicted as putative class I members as they clustered along with the class I members of *A. thaliana* and *O. sativa*, thus suggesting their role in lignification in plants²¹. In the case of dicots, the 4CL proteins are classified into class I and class II wherein the class I proteins are linked to lignin accumulation and class II proteins play a role in the metabolism of other phenolic compounds²⁹. However, in the case of monocots like *Oryza sativa*, a new phylogenetic category, class III is also present. It is speculated that this divergent evolution may be due to the variation in the phenolic compound composition in monocots and dicots and different substrate specificity of 4CL enzymes among these two groups of plants⁷³. In the present study, the majority of the 4CL proteins of *Vanilla planifolia* (a monocotyledonous species) clustered in the class III clade. In addition, close evolutionary ties between the proteins were depicted by their close clustering in the phylogenetic tree.

The gene structure analysis showed the presence of one intron in almost all VplPAL genes which is in line with the previous studies on PAL genes in other plants like *Oryza sativa* and *Carya illinoensis*^{55,61}. In a similar

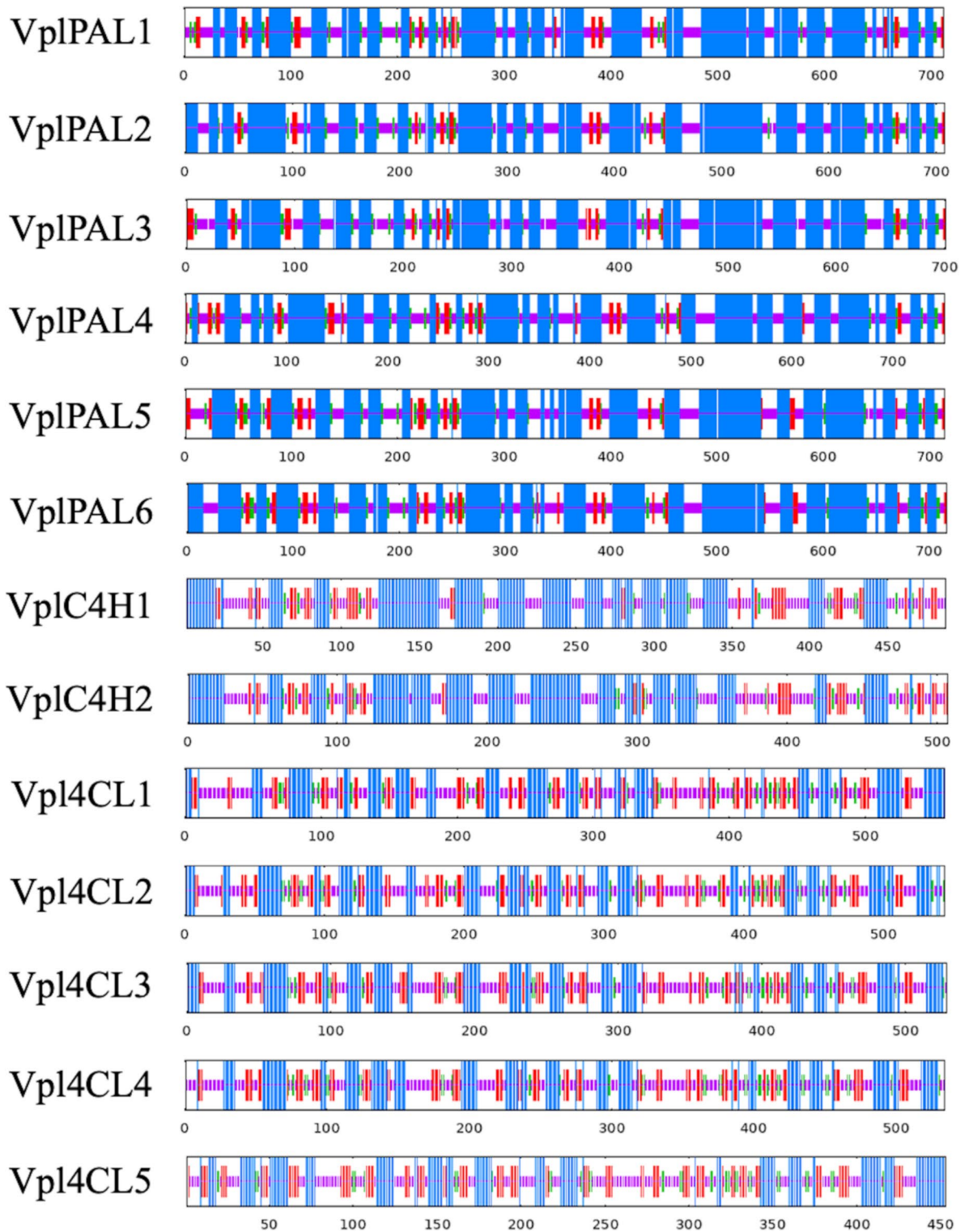


Fig. 8. Analysis of secondary structures showing the presence of alpha helices, extended strands, beta turns and random coils by blue, pink, red and green colour, respectively.

manner, the results of gene structure analysis for *VplCAH* genes and *Vpl4CL* genes were in conformity with *CAH* genes of *Brassica napus*¹³² and *Salvia miltiorrhiza*⁸³ and *4CL* genes of *Physcomitrella patens*⁷⁶ respectively. Additionally, for *Vpl4CL* genes, all the members shared a similar structural pattern except *Vpl4CL1* which belonged to class I of 4CLs. Thus, the degree of similarity was more in genes belonging to the same class in the phylogenetic tree. Moreover, the similar organization of gene structure within a gene family indicates towards its plausible conservation during the plant evolution.

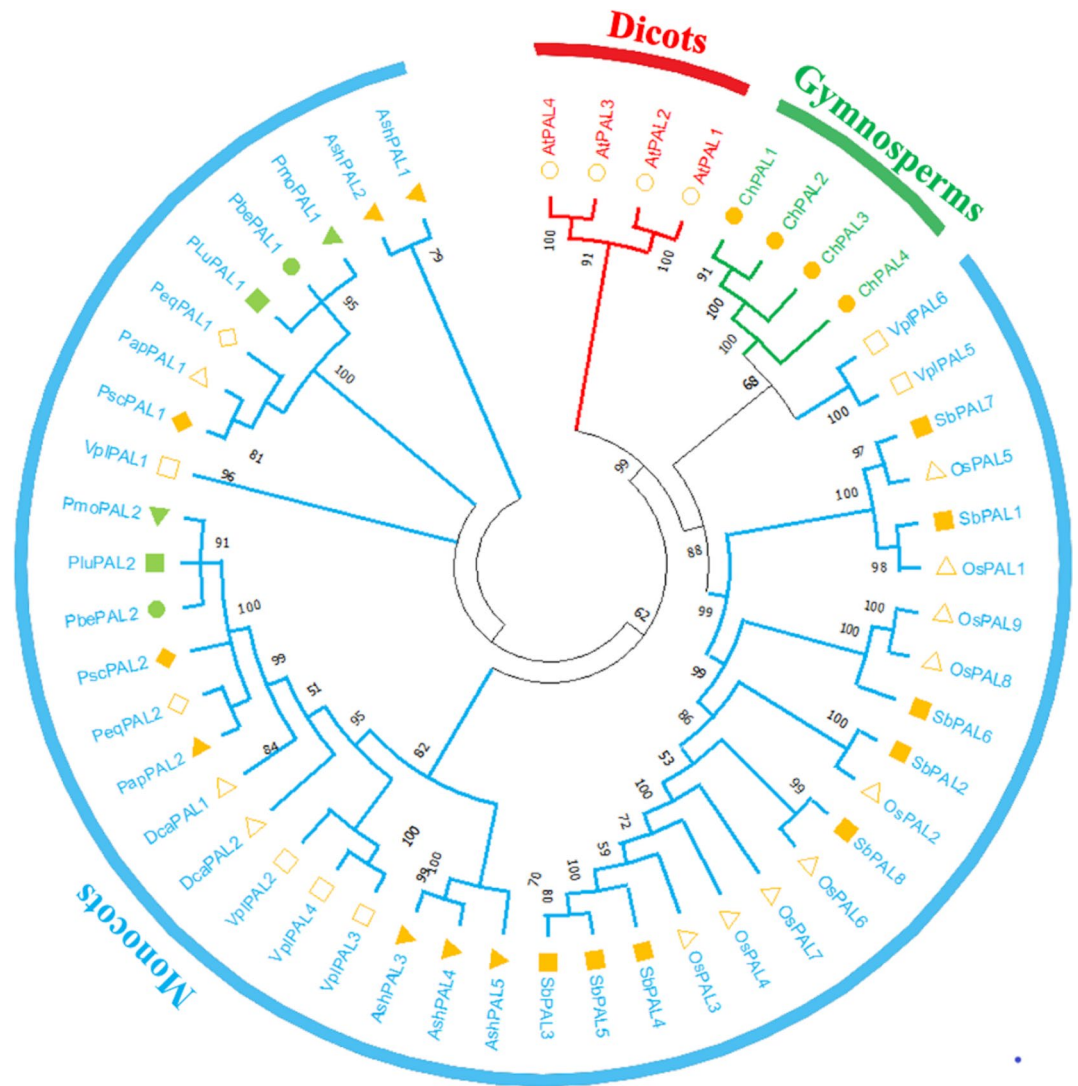


Fig. 9. Phylogenetic tree of PAL proteins showing the distinction between gymnosperms, monocots and dicots. PAL proteins belonging to different plants have been marked with different symbol annotations.

The presence of *cis*-regulatory elements typical of the *VplPAL*, *VplC4H* and *Vpl4CL* genes involved in the phenylpropanoid pathway was elucidated through promoter analysis. The analysis of the promoter regions revealed that these genes consisted of numerous phytohormone-responsive, abiotic and biotic stress-responsive, plant growth and development inducers and light-responsive *cis*-elements. In *Punica granatum* *4CL* genes⁷⁸ and *Glycine max* *C4H* genes¹³³, similar *cis*-acting elements belonging to the aforementioned four categories were predicted. The presence of AC-II element in *VplPAL1* and *Vpl4CL3* is similar to the promoter of *Na4CL* of *Neosinocalamus affinis*¹²⁸. The *cis*-acting elements found in the promoter sequences of *VplPAL*, *VplC4H* and *Vpl4CL* genes were also similar to those found in *Salvia miltiorrhiza* *4CL* genes⁴⁰. The presence of *cis*-acting elements regulating stress and phytohormone responses was also reported in *4CL* genes of *Gossypium hirsutum*⁴³ and *Eucommia ulmoides*⁴⁴ and *PAL* genes of *Salvia miltiorrhiza*⁸² and *Triticum aestivum*⁸⁸. In addition, specifically MeJA responsive elements were found in *C4H* members of *Salvia miltiorrhiza*⁸³ and *Saccharum spontaneum*¹¹⁷ and *PAL* members of *Carya illinoensis*⁵⁵ similar to many *PAL*, *C4H* and *4CL* members in our study. Apart from MeJA, *cis*-elements induced by gibberellins, abscisic acid and salicylic acid were also predicted during the present analysis which is in line with *C4H* genes in *Saccharum spontaneum*¹¹⁷. Thus, it is foreseeable that phytohormones may regulate the expression of *PAL*, *C4H* and *4CL* genes.

The spatial variation in the expression of *PAL*, *C4H* and *4CL* genes influences the spatial diversity of phenylpropanoids in plants. Hence, the expression profile for *VplPAL*, *VplC4H* and *Vpl4CL* genes in different plants tissues of *Vanilla planifolia* was analysed. The expression profile of *VplPAL* genes was found to be divergent amongst the various vegetative and reproductive tissues. Reduced expression in leaf tissue of *ChPAL* genes of *Cephalotaxus hainanensis*⁵⁶, *DcPAL* of *Dendrobium candidum*¹¹⁵ and *DcPAL1* of *Dendrobium catenatum*¹³⁴ was in accordance with the expression of *VplPAL*s. In addition, *VplPAL2-4* were strongly expressed in fruit which could be corroborated with the high expression of *PbPAL2* gene of *Pyrus bretschneideri*¹³⁵ in the same tissue.

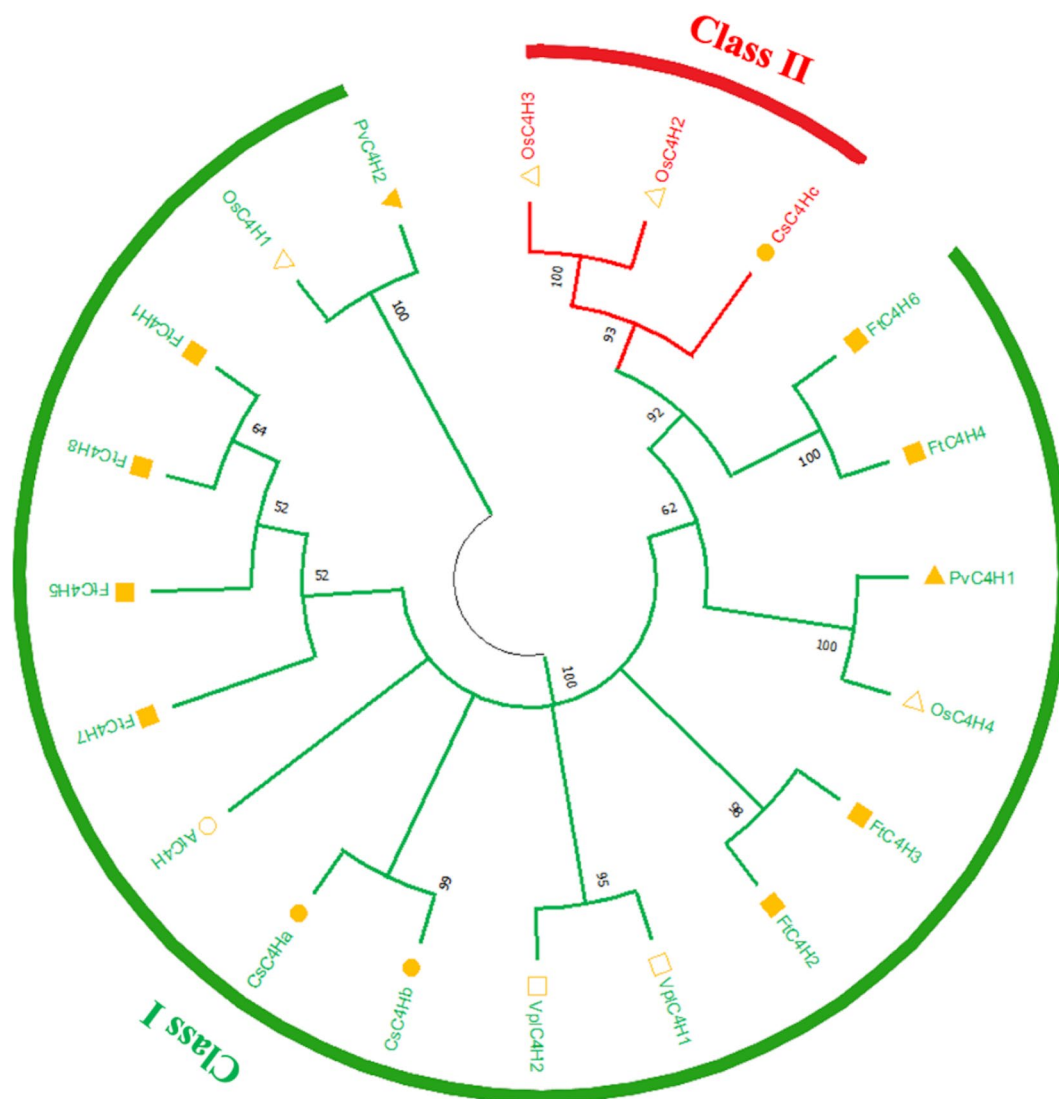


Fig. 10. Phylogenetic tree of C4H proteins showing the distinction between class I and class II. C4H proteins belonging to different plants have been marked with different symbol annotations.

This suggests that these genes may play a function in the mechanism of fruit development. The expression analysis of *C4H* genes of *Vanilla planifolia* showed that *VpC4H1* and *VpC4H2*, both genes were expressed in elevated levels in the flower tissue. Likewise, *SmC4H1* gene of *Salvia miltiorrhiza*⁸³ and *FtC4H6* of *Fagopyrum tataricum*⁶³ were highly expressed in flowers, which might indicate the role of these genes in the mechanism of flowering. However, both genes showed very less expression in the soil root which is in line with the expression of *BoC4H.2* and *BoC4H.4* of *Brassica oleracea*¹³⁶. Similarly, the overlapping and divergent expression pattern of *Vp4CL* genes in different tissues was also analyzed. Additionally, high levels of expression were observed for *Vp4CL2* in fruit and likewise, *Ri4CL2-3* of *Rubus idaeus*³⁸ also showed enhanced expression in fruit. Thus, the results point towards the functional dispersion in *PAL*, *C4H* and *4CL* gene family in *Vanilla planifolia*.

Thus, insights into plant metabolic pathways and their adaptation to various environmental situations can be gained by further study of the regulation of the enzymes involved in the pathways. Furthermore, understanding the mechanisms underlying the function of *PALs*, *C4Hs*, and *4CLs* may also enable targeted genetic manipulations to enhance the overaccumulation of desired phenylpropanoid compounds in orchid plants.

Conclusions

An in-silico genome-wide characterization of *PAL*, *C4H* and *4CL* gene family regulating the phenylpropanoid pathway in *Vanilla planifolia* was carried out and a total of six *PAL*, two *C4H* and five *4CL* genes were identified. Domain analysis, multiple sequence alignment, conserved motif prediction, phylogenetic analysis, secondary structure prediction, gene structural analysis substantiated the highly conserved nature of these three gene families. *Cis*-regulatory elements prediction and expression profiling highlighted the role of these genes in plant growth and development. This study lays a basic frame for functional characterization of *PAL*, *C4H* and

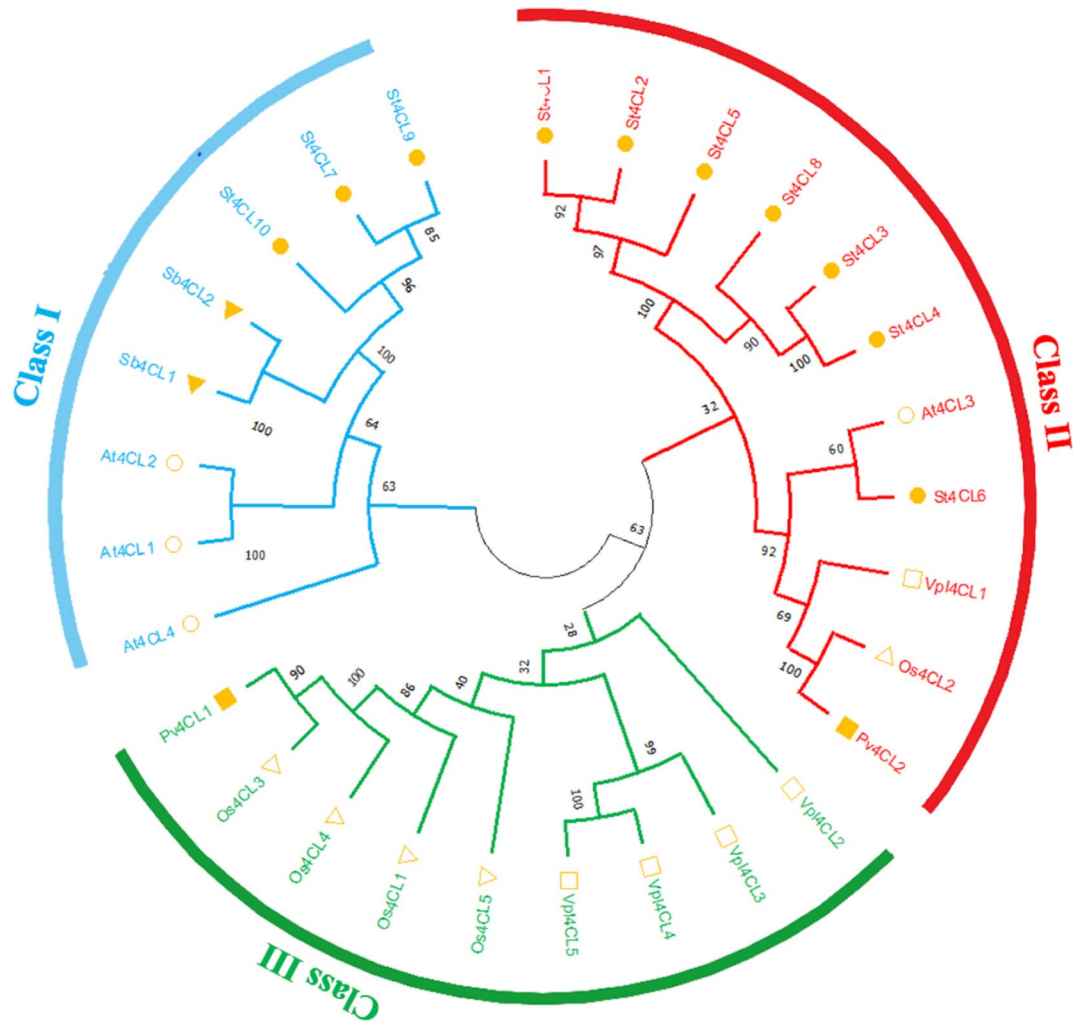


Fig. 11. Phylogenetic tree of 4CL proteins showing the distinction between class I, class II and class III. 4CL proteins belonging to different plants have been marked with different symbol annotations.

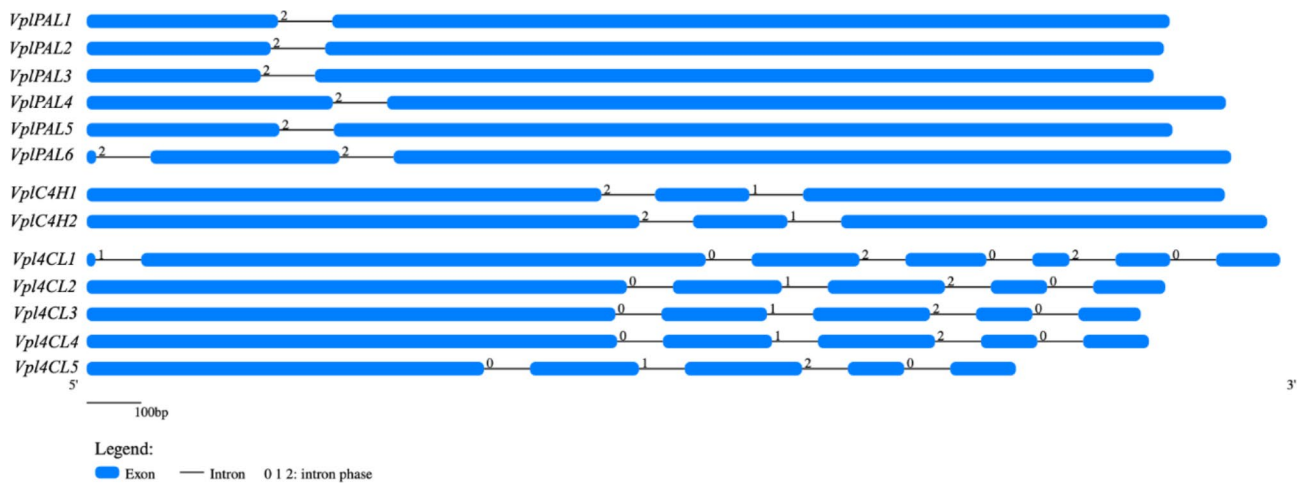


Fig. 12. Exon-intron arrangement in *VplPAL*, *VplCAH* and *Vpl4CL* genes.

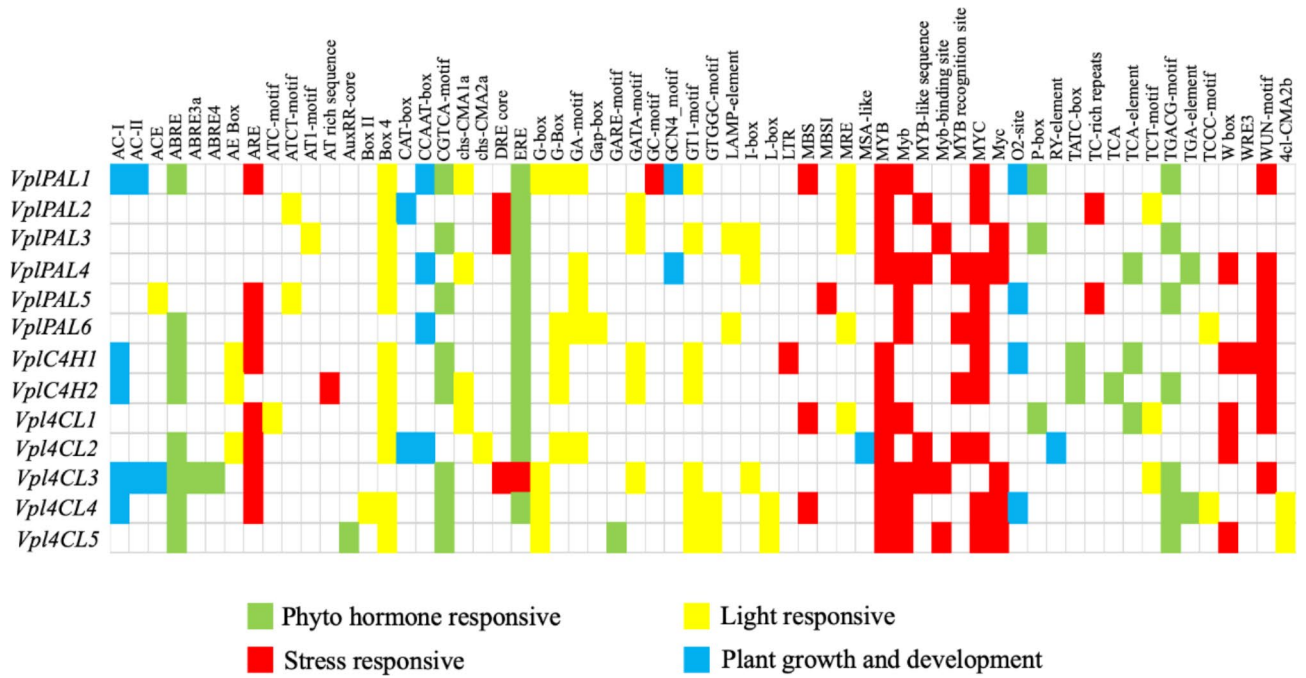


Fig. 13. Distribution of *cis*-regulatory elements into different categories in *VplPAL*, *VplC4H* and *Vpl4CL* genes.

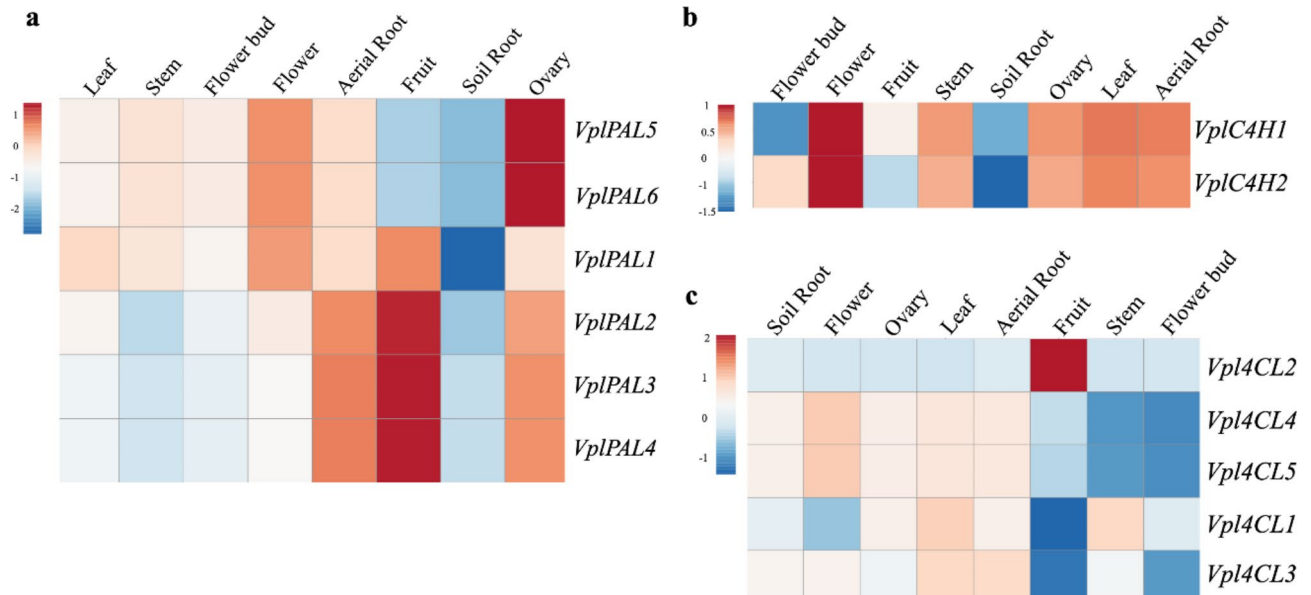


Fig. 14. Expression analysis of *VplPAL*, *VplC4H* and *Vpl4CL* genes in various tissues of *Vanilla planifolia*.

4CL genes in orchids which would further help us in understanding the correlation of individual specific gene expression with the biosynthesis of phenylpropanoids in orchids.

Data availability

All data generated or analysed during this study are included in this published article [and its supplementary information files].

Received: 19 July 2024; Accepted: 2 December 2024

Published online: 28 March 2025

References

- Dixon, R. A. et al. The phenylpropanoid pathway and plant defence—a genomics perspective. *Mol. Plant. Pathol.* **3**, 371–390. <https://doi.org/10.1046/j.1364-3703.2002.00131.x> (2002).
- Vogt, T. Phenylpropanoid biosynthesis. *Mol. Plant.* **3**, 2–20. <https://doi.org/10.1093/mp/ssp106> (2010).
- Ferrer, J. L., Austin, M. B., Stewart Jr, C. & Noel, J. P. Structure and function of enzymes involved in the biosynthesis of phenylpropanoids. *Plant. Physiol. Biochem.* **46**, 356–370. <https://doi.org/10.1016/j.plaphy.2007.12.009> (2008).
- Koukol, J. & Conn, E. E. The metabolism of aromatic compounds in higher plants: IV. Purification and properties of the phenylalanine deaminase of *Hordeum vulgare*. *J. Biol. Chem.* **236**, 2692–2698. [https://doi.org/10.1016/S0021-9258\(19\)61721-7](https://doi.org/10.1016/S0021-9258(19)61721-7) (1961).
- Fraser, C. M. & Chapple, C. The phenylpropanoid pathway in *Arabidopsis*. *Arabidopsis Book*. **9**, e0152. <https://doi.org/10.11999/2Ftab.0152> (2011).
- Hyun, M. W., Yun, Y. H., Kim, J. Y. & Kim, S. H. Fungal and plant phenylalanine ammonia-lyase. *Mycobiology* **39**, 257–265. <https://doi.org/10.5941/MYCO.2011.39.4.257> (2011).
- MacDonald, M. J. & D’Cunha, G. B. A modern view of phenylalanine ammonia lyase. *Biochem. Cell. Biol.* **85**, 273–282. <https://doi.org/10.1139/O07-018> (2007).
- Dong, C. J. & Shang, Q. M. Genome-wide characterization of phenylalanine ammonia-lyase gene family in watermelon (*Citrullus lanatus*). *Planta* **238**, 35–49. <https://doi.org/10.1007/s00425-013-1869-1> (2013).
- Carocha, V. et al. Genome-wide analysis of the lignin toolbox of *Eucalyptus grandis*. *New Phytol.* **206**, 1297–1313. <https://doi.org/10.1111/nph.13313> (2015).
- Li, G. et al. Comparative genomic analysis of the PAL genes in five Rosaceae species and functional identification of Chinese white pear. *PeerJ* **7**, e8064. <https://doi.org/10.7717/peerj.8064> (2019).
- Chen, X. et al. Identification of PAL genes related to anthocyanin synthesis in tea plants and its correlation with anthocyanin content. *Hortic. Plant. J.* **8**, 381–394. <https://doi.org/10.1016/j.hpj.2021.12.005> (2022).
- Huang, J. et al. Functional analysis of the *Arabidopsis* PAL gene family in plant growth, development, and response to environmental stress. *Plant. Physiol.* **153**, 1526–1538. <https://doi.org/10.1104/pp.110.157370> (2010).
- Liang, X. W., Dron, M., Cramer, C. L., Dixon, R. A. & Lamb, C. J. Differential regulation of phenylalanine ammonia-lyase genes during plant development and by environmental cues. *J. Biol. Chem.* **264**, 14486–14492. [https://doi.org/10.1016/S0021-9258\(18\)71704-3](https://doi.org/10.1016/S0021-9258(18)71704-3) (1989).
- Wu, D. G. et al. Genome-wide identification and analysis of maize pal gene family and its expression profile in response to high-temperature stress. *Pak J. Bot.* **52**, 1577–1587. <https://doi.org/10.30848/PJB2020-528> (2020).
- Olsen, K. M., Lea, U. S., Slimestad, R., Verheul, M. & Lillo, C. Differential expression of four *Arabidopsis* PAL genes; PAL1 and PAL2 have functional specialization in abiotic environmental-triggered flavonoid synthesis. *J. Plant. Physiol.* **165**, 1491–1499. <https://doi.org/10.1016/j.jplph.2007.11.005> (2008).
- Gho, Y. S., Kim, S. J. & Jung, K. H. Phenylalanine ammonia-lyase family is closely associated with response to phosphate deficiency in rice. *Genes Genomics*. **42**, 67–76. <https://doi.org/10.1007/s13258-019-00879-7> (2020).
- Feng, Y. et al. Molecular characterisation of PAL gene family reveals their role in abiotic stress response in lucerne (*Medicago sativa*). *Crop Pasture Sci.* **73**, 300–311. <https://doi.org/10.1071/CP21558> (2022).
- Ehrling, J., Hamberger, B., Million-Rousseau, R. & Werck-Reichhart, D. Cytochromes P450 in phenolic metabolism. *Phytochem Rev.* **5**, 239–270. <https://doi.org/10.1007/s11101-006-9025-1> (2006).
- Russell, D. W. & Conn, E. E. The cinnamic acid 4-hydroxylase of pea seedlings. *Arch. Biochem. Biophys.* **122**, 256–258. [https://doi.org/10.1016/0003-9861\(67\)90150-6](https://doi.org/10.1016/0003-9861(67)90150-6) (1967).
- Hamberger, B. et al. Genome-wide analyses of phenylpropanoid-related genes in *Populus trichocarpa*, *Arabidopsis thaliana*, and *Oryza sativa*: the *Populus* lignin toolbox and conservation and diversification of angiosperm gene families. *Botany* **85**, 1182–1201. <https://doi.org/10.1139/B07-098> (2007).
- Raes, J., Rohde, A., Christensen, J. H., Van de Peer, Y. & Boerjan, W. Genome-wide characterization of the lignification toolbox in *Arabidopsis*. *Plant. Physiol.* **133**, 1051–1071. <https://doi.org/10.1104/pp.103.026484> (2003).
- Nedelkina, S. et al. Novel characteristics and regulation of a divergent cinnamate 4-hydroxylase (CYP73A15) from French bean: engineering expression in yeast. *Plant. Mol. Biol.* **39**, 1079. <https://doi.org/10.1023/A:1006156216654> (1999).
- Lu, S., Zhou, Y., Li, L. & Chiang, V. L. Distinct roles of cinnamate 4-hydroxylase genes in *Populus*. *Plant. Cell. Physiol.* **47**, 905–914. <https://doi.org/10.1093/pcp/pcj063> (2006).
- Kumar, S., Omer, S., Patel, K. & Khan, B. M. Cinnamate 4-Hydroxylase (C4H) genes from *Leucaena leucocephala*: a pulp yielding leguminous tree. *Mol. Biol. Rep.* **40**, 1265–1274. <https://doi.org/10.1007/s11033-012-2169-8> (2013).
- Li, Y. et al. Cloning and expression analysis of phenylalanine ammonia-lyase (PAL) gene family and cinnamate 4-hydroxylase (C4H) from *Dryopteris fragrans*. *Biologia* **70**, 606–614. <https://doi.org/10.1515/biolog-2015-0083> (2015).
- Chao, N., Yu, T., Hou, C., Liu, L. & Zhang, L. Genome-wide analysis of the lignin toolbox for morus and the roles of lignin related genes in response to zinc stress. *PeerJ* **9**, e11964. <https://doi.org/10.7717/peerj.11964> (2021).
- Singh, K., Kumar, S., Rani, A., Gulati, A. & Ahuja, P. S. Phenylalanine ammonia-lyase (PAL) and cinnamate 4-hydroxylase (C4H) and catechins (flavan-3-ols) accumulation in tea. *Funct. Integr. Genomic.* **9**, 125–134. <https://doi.org/10.1007/s10142-008-0092-9> (2009).
- Xia, J. et al. Characterization and expression profiling of *Camellia sinensis* cinnamate 4-hydroxylase genes in phenylpropanoid pathways. *Genes* **8**, 193. <https://doi.org/10.3390/genes8080193> (2017).
- Ehrling, J. et al. Three 4-coumarate: coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant. J.* **19**, 9–20. <https://doi.org/10.1046/j.1365-313X.1999.00491.x> (1999).
- Costa, M. A. et al. 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. [https://doi.org/10.1016/S0031-9422\(03\)00517-X](https://doi.org/10.1016/S0031-9422(03)00517-X) (2003).
- Douglas, C., Hoffmann, H., Schulz, W. & Hahlbrock, K. Structure and elicitor or uv-light-stimulated expression of two 4-coumarate: CoA ligase genes in parsley. *EMBO J.* **6**, 1189–1195. <https://doi.org/10.1002/j.1460-2075.1987.tb02353.x> (1987).
- Lindermayr, C. et al. 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. <https://doi.org/10.1046/j.1432-1033.2002.02775.x> (2002).
- Xu, B. et al. 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. <https://doi.org/10.1111/j.1469-8137.2011.03830.x> (2011).
- Rao, G. et al. Divergent and overlapping function of five 4-coumarate/coenzyme A ligases from *Populus tomentosa*. *Plant. Mol. Biol. Rep.* **33**, 841–854. <https://doi.org/10.1007/s11105-014-0803-4> (2015).
- Tang, Y. H. et al. Cloning and characterization of the key 4-coumarate CoA ligase genes in *Boehmeria nivea*. *S Afr. J. Bot.* **116**, 123–130. <https://doi.org/10.1016/j.sajb.2018.02.398> (2018).
- Zhou, G. et al. 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. <https://doi.org/10.1016/j.scienta.2021.110763> (2022).
- Hu, W. J. et al. Compartmentalized expression of two structurally and functionally distinct 4-coumarate: CoA ligase genes in aspen (*Populus tremuloides*). *Proc. Natl. Acad. Sci.* **95**, 5407–5412. <https://doi.org/10.1073/pnas.95.9.5407> (1998).

38. Kumar, A. & Ellis, B. E. 4-Coumarate: CoA ligase gene family in *Rubus idaeus*: cDNA structures, evolution, and expression. *Plant Mol. Biol.* **51**, 327–340. <https://doi.org/10.1023/A:1022004923982> (2003).
39. Di, P. et al. Characterization and the expression profile of 4-coumarate: CoA ligase (Ii4CL) from hairy roots of *Isatis Indigotica*. *Afr. J. Pharma Pharmacol.* **6**, 2166–2175. <https://doi.org/10.5897/AJPP12.852> (2012).
40. Jin, X. Q., Chen, Z. W., Tan, R. H., Zhao, S. J. & Hu, Z. B. Isolation and functional analysis of 4-coumarate: coenzyme a ligase gene promoters from *Salvia miltiorrhiza*. *Biol. Plant.* **56**, 261–268. <https://doi.org/10.1007/s10535-012-0085-3> (2012).
41. Jin, Z. et al. 4-Coumarate: coenzyme a ligase isoform 3 from *Piper nigrum* (Pn4CL3) catalyzes the CoA thioester formation of 3, 4-methylenedioxy-cinnamic and piperic acids. *Biochem. J.* **477**, 61–74. <https://doi.org/10.1042/BCJ20190527> (2020).
42. Awasthi, P. et al. Characterization of the gene encoding 4-coumarate: CoA ligase in *Coleus Forskohlii*. *J. Plant. Biochem. Biotechnol.* **28**, 203–210. <https://doi.org/10.1007/s13562-018-0468-4> (2019).
43. Sun, S. C. et al. Characterization of the Gh4CL gene family reveals a role of Gh4CL7 in drought tolerance. *BMC Plant. Biol.* **20**, 1–15. <https://doi.org/10.1186/s12870-020-2329-2> (2020).
44. Zhong, J. et al. Genome-wide identification and expression analyses of the 4-Coumarate: CoA ligase (4CL) Gene Family in *Eucommia ulmoides*. *Forests* **13**, 1253. <https://doi.org/10.3390/f13081253> (2022).
45. Zhang, H. et al. Genome-wide identification and expression analysis of phenylalanine ammonia-lyase (PAL) family in rapeseed (*Brassica napus* L.). *BMC Plant. Biol.* **23**, 481. <https://doi.org/10.1186/s12870-023-04472-9> (2023).
46. Qu, C. et al. Genome-wide survey of flavonoid biosynthesis genes and gene expression analysis between black-and yellow-seeded *Brassica napus*. *Front. Plant. Sci.* **7**, 1755. <https://doi.org/10.3389/fpls.2016.01755> (2016).
47. Ma, Z. H., Nan, X. T., Li, W. F., Mao, J. & Chen, B. H. Comprehensive genomic identification and expression analysis 4CL gene family in apple. *Gene* **858**, 147197. <https://doi.org/10.1016/j.gene.2023.147197> (2023).
48. Kaur, A., Yadav, V. G., Pawar, S. V. & Sembi, J. K. Insights to Phenylalanine Ammonia Lyase (PAL) and secondary metabolism in Orchids: an in silico Approach. *Biochem. Genet.* **62**, 413–435. <https://doi.org/10.1007/s10528-023-10428-3> (2024).
49. Cochrane, F. C., Davin, L. B. & Lewis, N. G. The *Arabidopsis* phenylalanine ammonia lyase gene family: kinetic characterization of the four PAL isoforms. *Phytochemistry* **65**, 1557–1564. <https://doi.org/10.1016/j.phytochem.2004.05.006> (2004).
50. Li, Y., Kim, J. I., Pysh, L. & Chapple, C. Four isoforms of *Arabidopsis* 4-coumarate: CoA ligase have overlapping yet distinct roles in phenylpropanoid metabolism. *Plant. Physiol.* **169**, 2409–2421. <https://doi.org/10.1104/pp.15.00838> (2015).
51. Chai, P. et al. Genome-wide characterization of the phenylalanine Ammonia-lyase Gene Family and their potential roles in response to *Aspergillus Flavus* L. infection in cultivated peanut (*Arachis hypogaea* L.). *Genes* **15**, 265. <https://doi.org/10.3390/gene15030265> (2024).
52. Huang, C. et al. Identification, characterization and expression analysis of the 4-coumarate-coa ligase gene family in *Bletilla striata*. *Gene Rep.* **32**, 101785. <https://doi.org/10.1016/j.gene.2023.101785> (2023).
53. Tian, J. et al. Identification of PAL Gene in Purple Cabbage and functional analysis related to anthocyanin synthesis. *Horticulturae* **9**, 469. <https://doi.org/10.3390/horticulturae9040469> (2023).
54. Ren, W. W. et al. Genome-wide identification and expression analysis of 4CL Gene Family in *Camellia sinensis*. *Acta Bot. Boreali-Occident Sin.* **43**, 1459–1469 (2023).
55. Zhang, C., Yao, X., Ren, H., Wang, K. & Chang, J. Genome-wide identification and characterization of the phenylalanine ammonia-lyase gene family in pecan (*Carya illinoensis*). *Sci. Hortic.* **295**, 110800. <https://doi.org/10.1016/j.scienta.2021.110800> (2022).
56. He, Y. et al. Characterisation, expression and functional analysis of PAL gene family in *Cephalotaxus hainanensis*. *Plant. Physiol. Biochem.* **156**, 461–470. <https://doi.org/10.1016/j.plaphy.2020.09.030> (2020).
57. Wei, L. et al. Genome-wide identification of the CsPAL gene family and functional analysis for strengthening green mold resistance in citrus fruit. *Postharvest Biol. Technol.* **196**, 112178. <https://doi.org/10.1016/j.postharvbio.2022.112178> (2023).
58. Huang, X. et al. De novo transcriptome assembly of *Coffea liberica* reveals phylogeny and expression atlas of phenylalanine ammonia-lyase genes in *Coffea* species. *Ind. Crops Prod.* **192**, 116029. <https://doi.org/10.1016/j.indcrop.2022.116029> (2023).
59. Hossain, M. S. et al. Phenylalanine ammonia-lyase gene family (PAL): genome wide characterization and transcriptional expression in jute (*Corchorus olitorius*). *J. Biosci. Agric. Res.* **26**, 2185–2191. <https://doi.org/10.18801/jbar.260220.267> (2020).
60. Dong, C. J., Ning, C. A. O., Zhang, Z. G. & Shang, Q. M. Phenylalanine ammonia-lyase gene families in cucurbit species: structure, evolution, and expression. *J. Integr. Agric.* **15**, 1239–1255. [https://doi.org/10.1016/S2095-3119\(16\)61329-1](https://doi.org/10.1016/S2095-3119(16)61329-1) (2016).
61. Rezaee, S., Gharanjik, S. & Mojerlou, S. Structure and upstream region analysis of phenylalanine ammonia-lyase gene in rice (*Oryza sativa* L. ssp. *Japonica*) and cucumber (*Cucumis sativus* L. Cv. Chinese long). *Arch. Phytopathol. Plant. Prot.* **53**, 355–378. <https://doi.org/10.1080/03235408.2020.1740504> (2020).
62. Liu, A. et al. Molecular identification of phenylalanine ammonia lyase-encoding genes EfPALs and EfPAL2-interacting transcription factors in *Euryale ferox*. *Front. Plant. Sci.* **14**, 1114345. <https://doi.org/10.3389/fpls.2023.1114345> (2023).
63. Yao, Y. et al. Genome-wide investigation of major enzyme-encoding genes in the flavonoid metabolic pathway in Tartary buckwheat (*Fagopyrum tataricum*). *J. Mol. Evol.* **89**, 269–286. <https://doi.org/10.1007/s00239-021-10004-6> (2021).
64. Yang, Z. et al. Identification and functional characterization of three phenylalanine Ammonia-lyase genes from *Fallopia multiflora* (Thunb.) Harald. *Russ J. Bioorg. Chem.* **49**, 655–663. <https://doi.org/10.1134/S1068162023030263> (2023).
65. Khatiri, P., Chen, L., Rajcan, I. & Dhaubhadel, S. Functional characterization of cinnamate 4-hydroxylase gene family in soybean (*Glycine max*). *Plos One.* **18**, e0285698. <https://doi.org/10.1371/journal.pone.0285698> (2023).
66. Yang, Z. et al. Genome-wide identification and expression profiling of 4-coumarate: coenzyme a ligase genes influencing soybean isoflavones at the seedling stage. *Crop Pasture Sci.* **75**, CP23147. <https://doi.org/10.1071/CP23147> (2023).
67. Song, Y., Zhang, G., Chen, N., Zhang, J. & He, C. Metabolomic and transcriptomic analyses provide insights into the flavonoid biosynthesis in sea buckthorn (*Hippophae rhamnoides* L.). *LWT* **187**, 115276. <https://doi.org/10.1016/j.lwt.2023.115276> (2023).
68. Yan, F., Li, H. & Zhao, P. Genome-wide identification and transcriptional expression of the PAL Gene family in common Walnut (*Juglans regia* L.). *Genes* **10**, 46. <https://doi.org/10.3390/gene10010046> (2019).
69. Ma, J. et al. 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. <https://doi.org/10.1186/s12870-024-04899-8> (2024).
70. Gao, S. et al. X. 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. <https://doi.org/10.1371/journal.pone.0296079> (2024).
71. Wang, Z. et al. 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. <https://doi.org/10.3389/fpls.2023.1072086> (2023).
72. Wu, Z., Gui, S., Wang, S. & Ding, Y. Molecular evolution and functional characterisation of an ancient phenylalanine ammonia-lyase gene (NnPAL1) from *Nelumbo nucifera*: novel insight into the evolution of the PAL family in angiosperms. *BMC Evol. Biol.* **14**, 1–14. <https://doi.org/10.1186/1471-2148-14-100> (2014).
73. Gui, J., Shen, J. & Li, L. Functional characterization of evolutionarily divergent 4-coumarate: coenzyme a ligases in rice. *Plant. Physiol.* **157**, 574–586. <https://doi.org/10.1104/pp.111.178301> (2011).
74. Shen, H. et al. A genomics approach to deciphering lignin biosynthesis in switchgrass. *Plant. Cell.* **25**, 4342–4361. <https://doi.org/10.1105/tpc.113.118828> (2013).
75. Koopmann, E., Logemann, E. & Hahlbrock, K. Regulation and functional expression of cinnamate 4-hydroxylase from parsley. *Plant. Physiol.* **119**, 49–56. <https://doi.org/10.1104/pp.119.1.49> (1999).

76. Silber, M. V., Meimberg, H. & Ebel, J. Identification of a 4-coumarate: CoA ligase gene family in the moss, *Physcomitrella patens*. *Phytochemistry* **69**, 2449–2456. <https://doi.org/10.1016/j.phytochem.2008.06.014> (2008).
77. Kawai, S. et al. Isolation and analysis of cinnamic acid 4-hydroxylase homologous genes from a hybrid aspen, *Populus kitakamiensis*. *Biosci. Biotech. Biochem.* **60**, 1586–1597. <https://doi.org/10.1271/bbb.60.1586> (1996).
78. Wang, Y., Guo, L., Zhao, Y., Zhao, X. & Yuan, Z. 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. <https://doi.org/10.3390/ijms23073509> (2022).
79. Zhang, L. et al. Genome-wide identification and Expression Analysis of Fifteen Gene Families Involved in anthocyanin synthesis in Pear. *Horticulturae* **10**, 335. <https://doi.org/10.3390/horticulturae10040335> (2024).
80. Lai, B. et al. Genome-wide identification and expression analyses of PAL genes in different Color Radish. *Comput. Mol. Biol.* **12**, 1. <https://doi.org/10.5376/cmb.2022.12.0001> (2022).
81. Liu, G. et al. Genome-wide identification and analysis of monolignol biosynthesis genes in *Salix matsudana* Koidz and their relationship to accelerated growth. *Res.* **1**, 1–11. <https://doi.org/10.48130/FR-2021-0008> (2021).
82. Hou, X., Shao, F., Ma, Y. & Lu, S. The phenylalanine ammonia-lyase gene family in *Salvia miltiorrhiza*: genome-wide characterization, molecular cloning and expression analysis. *Mol. Biol. Rep.* **40**, 4301–4310. <https://doi.org/10.1007/s11033-013-2517-3> (2013).
83. Wang, B. et al. Genome-wide identification of phenolic acid biosynthetic genes in *Salvia miltiorrhiza*. *Planta* **241**, 711–725. <https://doi.org/10.1007/s00425-014-2212-1> (2015).
84. Xu, H. et al. U. Molecular cloning and characterization of phenylalanine ammonia-lyase, cinnamate 4-hydroxylase and genes involved in flavone biosynthesis in *Scutellaria baicalensis*. *Bioresour Technol.* **101**, 9715–9722. <https://doi.org/10.1016/j.biortech.2010.07.083> (2010).
85. Zhang, F. et al. Genome-wide identification and expression analyses of phenylalanine ammonia-lyase gene family members from tomato (*Solanum lycopersicum*) reveal their role in root-knot nematode infection. *Front. Plant. Sci.* **14**, 1204990. <https://doi.org/10.3389/fpls.2023.1204990> (2023).
86. Nie, T. et al. Genome-wide identification and expression analysis of the 4-Coumarate: CoA ligase Gene Family in *Solanum tuberosum*. *Inter J. Mol. Sci.* **24**, 1642. <https://doi.org/10.3390/ijms24021642> (2023).
87. Pant, S. & Huang, Y. Genome-wide studies of PAL genes in sorghum and their responses to aphid infestation. *Sci. Rep.* **12**, 22537. <https://doi.org/10.1038/s41598-022-25214-1> (2022).
88. Rasool, F. et al. Phenylalanine ammonia-lyase (PAL) genes family in wheat (*Triticum aestivum* L.): genome-wide characterization and expression profiling. *Agronomy* **11**, 2511. <https://doi.org/10.3390/agronomy11122511> (2021).
89. Zhao, T. et al. Genome-wide identification and characterisation of phenylalanine ammonia-lyase gene family in grapevine. *J. Hort. Sci. Biotechnol.* **96**, 456–468. <https://doi.org/10.1080/14620316.2021.1879685> (2021).
90. Hossain, M. M. Therapeutic orchids: traditional uses and recent advances-an overview. *Fitoterapia* **82**, 102–140. <https://doi.org/10.1016/j.fitote.2010.09.007> (2011).
91. Sut, S., Maggi, F. & Dall'Acqua, S. Bioactive secondary metabolites from orchids (Orchidaceae). *Chem. Biodivers.* **14**, e1700172. <https://doi.org/10.1002/cbdv.201700172> (2017).
92. Kaur, A., Verma, J., Yadav, V. G., Pawar, S. V. & Sembi, J. K. Exploring the potential of in vitro cultures as an aid to the production of secondary metabolites in Medicinal orchids. In *Tiwari P, Chen JT (eds) Advances in Orchid Biology, Biotechnology and Omics*, Springer, Singapore, 163–185 https://doi.org/10.1007/978-981-99-1079-3_5 (2023).
93. Zhang, S. & Mueller, C. Comparative analysis of volatiles in traditionally cured bourbon and Ugandan vanilla bean (*Vanilla planifolia*) extracts. *J. Agric. Food Chem.* **60**, 10433–10444. <https://doi.org/10.1021/jf302615s> (2012).
94. Gallage, N. J., Möller, B. L. & Vanilla The most popular flavour. In: *Schwab W, Lange B, Wüst M (eds) Biotechnology of Natural Products*, Springer, Cham, 3–24 https://doi.org/10.1007/978-3-319-67903-7_1 (2018).
95. Arya, S. S., Rookes, J. E., Cahill, D. M. & Lenka, S. K. Vanillin: a review on the therapeutic prospects of a popular flavouring molecule. *Adv. Tradit. Med.* **21**, 1–17. <https://doi.org/10.1007/s13596-020-00531-w> (2021).
96. Havkin-Frenkel, D. & Belanger, F. C. Application of metabolic engineering to Vanillin biosynthetic pathways in *Vanilla planifolia*. In: *Verpoorte R, Alfermann A, Johnson T (eds) Applications of Plant Metabolic Engineering*, Springer, Dordrecht, 175–196 https://doi.org/10.1007/978-1-4020-6031-1_7 (2007).
97. Fock-Bastide, I. et al. Expression profiles of key phenylpropanoid genes during *Vanilla planifolia* pod development reveal a positive correlation between PAL gene expression and vanillin biosynthesis. *Plant. Physiol. Biochem.* **74**, 304–314. <https://doi.org/10.1016/j.plaphy.2013.11.026> (2014).
98. Yang, H. et al. A re-evaluation of the final step of vanillin biosynthesis in the orchid *Vanilla planifolia*. *Phytochemistry* **139**, 33–46. <https://doi.org/10.1016/j.phytochem.2017.04.003> (2017).
99. Piet, Q. et al. A chromosome-level, haplotype-phased *Vanilla planifolia* genome highlights the challenge of partial endoreplication for accurate whole-genome assembly. *Plant. Commun.* **3**, 100330. <https://doi.org/10.1016/j.xplc.2022.100330> (2022).
100. Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P. & Bork, P. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* **28**, 231–234. <https://doi.org/10.1093/nar/28.1.231> (2000).
101. Sigrist, C. J. et al. New and continuing developments at PROSITE. *Nucleic Acids Res.* **41**, D344–D347. <https://doi.org/10.1093/nar/gks1067> (2012).
102. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208. <https://doi.org/10.1093/nar/gkp335> (2009).
103. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* **16**, 10881–10890. <https://doi.org/10.1093/nar/16.22.10881> (1988).
104. Gasteiger, E. et al. Protein identification and analysis tools on the ExPASy server. In: *Walker JM (eds) The Proteomics Protocols Handbook*. Springer Protocols Handbooks, Humana Press, 571–607 <https://doi.org/10.1385/1-59259-890-0:571> (2005).
105. Chou, K. C. & Shen, H. B. Plant-mPLOC: a top-down strategy to augment the power for predicting plant protein subcellular localization. *PLoS One.* **5**, e11335. <https://doi.org/10.1371/journal.pone.0011335> (2010).
106. Sapay, N., Guerneur, Y. & Deléage, G. Prediction of amphipathic in-plane membrane anchors in monotopic proteins using a SVM classifier. *BMC Bioinform.* **7**, 1–11. <https://doi.org/10.1186/1471-2105-7-255> (2006).
107. Tamura, K., Stecher, G. & Kumar, S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3302. <https://doi.org/10.1093/molbev/msab120> (2021).
108. Hu, B. et al. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* **31**, 1296–1297. <https://doi.org/10.1093/bioinformatics/btu817> (2015).
109. Lescot, M. et al. 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. <https://doi.org/10.1093/nar/30.1.325> (2002).
110. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods.* **5**, 621–628. <https://doi.org/10.1038/nmeth.1226> (2008).
111. Metsalu, T. & Vilo, J. ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Res.* **43**, W566–W570. <https://doi.org/10.1093/nar/gkv468> (2015).
112. Calabrese, J. C., Jordan, D. B., Boodhoo, A., Sariaslani, S. & Vannelli, T. Crystal structure of phenylalanine ammonia lyase: multiple helix dipoles implicated in catalysis. *Biochem* **43**, 11403–11416. <https://doi.org/10.1021/bi049053+> (2004).
113. Ritter, H. & Schulz, G. E. Structural basis for the entrance into the phenylpropanoid metabolism catalyzed by phenylalanine ammonia-lyase. *Plant. Cell.* **16**, 3426–3436. <https://doi.org/10.1105/tpc.104.025288> (2004).

114. Watts, K. T., Mijts, B. N., Lee, P. C., Manning, A. J. & Schmidt-Dannert, C. Discovery of a substrate selectivity switch in tyrosine ammonia-lyase, a member of the aromatic amino acid lyase family. *Chem. Biol.* **13**, 1317–1326. <https://doi.org/10.1016/j.chembiol.2006.10.008> (2006).
115. Jin, Q., Yao, Y., Cai, Y. & Lin, Y. Molecular cloning and sequence analysis of a phenylalanine ammonia-lyase gene from *Dendrobium*. *PLoS One* **8**, e62352. <https://doi.org/10.1371/journal.pone.0062352> (2013).
116. Nag, S. & Kumaria, S. In silico characterization and transcriptional modulation of phenylalanine ammonia lyase (PAL) by abiotic stresses in the medicinal orchid *Vanda coerulea* Griff. Ex Lindl. *Phytochemistry* **156**, 176–183. <https://doi.org/10.1016/j.phytochem.2018.09.012> (2018).
117. Jardim-Messeder, D. et al. Genome-wide analysis of general phenylpropanoid and monolignol-specific metabolism genes in sugarcane. *Funct. Integr. Genomics* **21**, 73–99. <https://doi.org/10.1007/s10142-020-00762-9> (2021).
118. Uhlmann, A. & Ebel, J. 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. <https://doi.org/10.1104/pp.102.4.1147> (1993).
119. Lee, D., Ellard, M., Wanner, L. A., Davis, K. R. & Douglas, C. J. 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. <https://doi.org/10.1007/BF00042072> (1995).
120. Wei, H. Y., Rao, G. D., Wang, Y. K., Zhang, L. & Lu, H. Cloning and analysis of a new 4CL-like gene in *Populus tomentosa*. *Sci. Pract.* **15**, 98–104. <https://doi.org/10.1007/s11632-013-0204-z> (2013).
121. Gao, S., Yu, H. N., Xu, R. X., Cheng, A. X. & Lou, H. X. Cloning and functional characterization of a 4-coumarate CoA ligase from liverwort *Plagiochasma appendiculatum*. *Phytochemistry* **111**, 48–58. <https://doi.org/10.1016/j.phytochem.2014.12.017> (2015).
122. Zhang, C., Zang, Y., Liu, P., Zheng, Z. & Ouyang, J. Characterization, functional analysis and application of 4-Coumarate: CoA ligase genes from *Populus trichocarpa*. *J. Biotechnol.* **302**, 92–100. <https://doi.org/10.1016/j.jbiotec.2019.06.300> (2019).
123. Allina, S. M., Pri-Hadash, A., Theilmann, D. A., Ellis, B. E. & Douglas, C. J. 4-Coumarate: coenzyme a ligase in hybrid poplar: properties of native enzymes, cDNA cloning, and analysis of recombinant enzymes. *Plant. Physiol.* **116**, 743–754. <https://doi.org/10.1104/pp.116.2.743> (1998).
124. Lavhale, S. G., Kalunke, R. M. & Giri, A. P. Structural, functional and evolutionary diversity of 4-coumarate-CoA ligase in plants. *Planta* **248**, 1063–1078. <https://doi.org/10.1007/s00425-018-2965-z> (2018).
125. Stuble, H. P., Büttner, D., Ehlting, J., Hahlbrock, K. & Kombrink, E. 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. [https://doi.org/10.1016/S0014-5793\(00\)01133-9](https://doi.org/10.1016/S0014-5793(00)01133-9) (2000).
126. de Jong, F., Hanley, S. J., Beale, M. H. & Karp, A. Characterisation of the willow phenylalanine ammonia-lyase (PAL) gene family reveals expression differences compared with poplar. *Phytochemistry* **117**, 90–97. <https://doi.org/10.1016/j.phytochem.2015.06.005> (2015).
127. Huang, B. et al. Characterization and expression profiling of cinnamate 4-hydroxylase gene from *Salvia miltiorrhiza* in rosmarinic acid biosynthesis pathway. *Russ J. Plant. Physiol.* **55**, 390–399. <https://doi.org/10.1134/S1021443708030163> (2008).
128. Cao, Y., Hu, S. L., Huang, S. X., Ren, P. & Lu, X. Q. Molecular cloning, expression pattern, and putative cis-acting elements of a 4-coumarate: CoA ligase gene in bamboo (*Neosinocalamus Affinis*). *Electron. J. Biotechnol.* **15**, 9–9. <https://doi.org/10.2225/vol15-issue5-fulltext-10> (2012).
129. Achnine, L., Blancaflor, E. B., Rasmussen, S. & Dixon, R. A. Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. *Plant. Cell.* **16**, 3098–3109. <https://doi.org/10.1105/tpc.104.024406> (2004).
130. Rui-Fang, M. A. et al. E. N. The phenylalanine ammonia-lyase gene family in *Isatis Indigotica* Fort.: molecular cloning, characterization, and expression analysis. *Chin. J. Nat. Med.* **14**, 801–812. [https://doi.org/10.1016/S1875-5364\(16\)30097-8](https://doi.org/10.1016/S1875-5364(16)30097-8) (2016).
131. Wang, Z. B., Chen, X., Wang, W., Cheng, K. D. & Kong, J. Q. Transcriptome-wide identification and characterization of *Ornithogalum saundersiae* phenylalanine ammonia lyase gene family. *RSC Adv.* **4**, 27159–27175. <https://doi.org/10.1039/C4RA03385J> (2014).
132. Chen, A. H., Chai, Y. R., Li, J. N. & Chen, L. Molecular cloning of two genes encoding cinnamate 4-hydroxylase (C4H) from oilseed rape (*Brassica napus*). *J. Biochem. Mol. Biol.* **40**, 247–260. <https://doi.org/10.5483/bmbrep.2007.40.2.247> (2007).
133. Reinprecht, Y., Perry, G. E. & Peter Pauls, K. A comparison of phenylpropanoid pathway gene families in common bean. Focus on P450 and C4H genes. In: Pérez de la Vega M, Santalla M, Marsolais F (eds) *The Common Bean Genome. Compendium of Plant Genomes*. Springer, Cham, pp 219–261 (2017). https://doi.org/10.1007/978-3-319-63526-2_11
134. Vishwakarma, S. K., Singh, N. & Kumaria, S. Genome-wide identification and analysis of the PAL genes from the orchids *Apostasia Shenzhenica*, *Dendrobium catenatum* and *Phalaenopsis Equestris*. *J. Biomol. Struct. Dyn.* **41**, 1295–1308. <https://doi.org/10.1080/07391102.2021.2019120> (2021).
135. Cao, Y., Li, X. & Jiang, L. Integrative analysis of the core fruit lignification toolbox in pear reveals targets for fruit quality bioengineering. *Biomol* **9**, 504. <https://doi.org/10.3390/biom9090504> (2019).
136. Han, F. et al. Genome-wide characterization and analysis of the anthocyanin biosynthetic genes in *Brassica oleracea*. *Planta* **254**, 1–14. <https://doi.org/10.1007/s00425-021-03746-6> (2021).

Author contributions

JKS and AK conceptualized the work. AK and KS performed the analysis. AK prepared the original draft. SVP and JKS reviewed and edited the draft. All the authors have read and approved the final version.

Funding

INSPIRE Fellowship from Department of Science and Technology (DST) for AK (File No. DST/IN-SPIRE/03/2021/002638). Partial financial support from Department of Science and Technology, Government of India under Promotion of University Research and Scientific Excellence (PURSE) grant scheme.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-81968-w>.

Correspondence and requests for materials should be addressed to J.K.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024