

Article

Not peer-reviewed version

Molecular Cloning, Characterization and Expression Analysis of Cinnamate 4-Hydroxylase (C4H) from *Ferula pseudalliacea*

Pegah Shahidi , [Bahman Bahramnejad](#) * , [Yavar Vafae](#) * , [Dara Dastan](#) , [Parviz Heidari](#) *

Posted Date: 14 December 2023

doi: 10.20944/preprints202312.1026.v1

Keywords: Phenylpropanoid pathway; Cinnamate-4-hydroxylase; Apiaceae; Secondary metabolism



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Molecular Cloning, Characterization and Expression Analysis of Cinnamate 4-Hydroxylase (C4H) from *Ferula pseudalliacea*

Pegah Shahidi ¹, Bahman Bahramnejad ^{1,*}, Yavar Vafae ^{2,*}, Dara Dastan ³ and Parviz Heidari ^{4,*}

¹ Department of Plant Production and Genetics, Faculty of Agriculture, University of Kurdistan, Sanandaj, Iran; shahidi.pegah@gmail.com

² Department of Horticultural Sciences, Faculty of Agriculture, University of Kurdistan, Sanandaj, Iran

³ Department of Farmacognosy, Faculty of School of Pharmacy, Hamedan Medical University, Hamedan; Iran, d.dastan@umsha.ac.ir

⁴ Faculty of Agriculture, Shahrood University of Technology, Shahrood 3619995161, Iran

* Correspondence: b.bahramnejad@uok.ac.ir (B.B.); y.vafae@uok.ac.ir (Y.V.); heidarip@shahroodut.ac.ir (P.P.)

Abstract: Cinnamate 4-hydroxylase (C4H) is a cytochrome P450-dependent monooxygenase enzyme that catalyzes a key step in phenylpropanoid metabolism in plants. However, there is very limited information on the function of C4H in medicinal plants. In the present study, the C4H gene from the *Ferula pseudalliacea* genome (named *FpC4H*) was cloned and characterized. In addition, the expression levels of *FpC4H* in the roots, stems, leaves, flowers and seeds were investigated via real-time PCR. *FpC4H* showed the most similarity with C4H proteins from *Daucus carota*. The results predicted that *FpC4H*, as well as its orthologs, is an unstable and hydrophilic protein. A high genetic distance was observed between *FpC4H* and C4H from monocots based on evolutionary analysis. Moreover, a variable region related to the variable function of C4H was detected in conserved domain region of C4H proteins. Protein structure analysis revealed that isoleucine and leucine residues are frequently present in the binding region of *FpC4H*, and two ion binding regions, cysteine258 and asparagine217, were identified in the 3D structure of *FpC4H*. Expression analysis revealed that the *FpC4H* gene is expressed in all organs of *F. pseudalliacea*. A low expression level of *FpC4H* was recorded in leaves, and a high expression level was observed in flower tissues. Based on the expression profile, *FpC4H* seems to be linked to a high content of phenolic compounds in flowers.

Keywords: phenylpropanoid pathway; cinnamate-4-hydroxylase; Apiaceae; secondary metabolism

1. Introduction

Medicinal plants have been used for thousands of years as traditional remedies worldwide. The bioactive compounds of these plants have proven beneficial for treating human diseases and promoting health. Endemic medicinal plants found only in certain geographic regions often have unique properties that have evolved in their isolated habitats. Iran, located at the intersection of Central Asia, West Asia, and the Middle East, harbors more than 8,000 plant species, approximately 1,900 of which are endemic [1]. Many of these endemic species have a long history of use in traditional Iranian medicine for treating conditions such as infections, inflammation, pain, and digestive issues [2]. As rich sources of specialized metabolites and bioactive compounds, Iran's endemic medicinal flora represent an immense repository of plant chemical diversity waiting to be explored for new drug leads and treatment possibilities. Understanding these endemic plants through genomic, phytochemical and pharmacological studies can provide insights for developing improved therapies and advancing human health [3].

Apiaceae, also known as Umbelliferae, is an economically important family of flowering plants consisting of more than 3,700 species, including crops, herbs, spices, and medicinal plants. Major Apiaceae crops include carrots, celery, parsley, parsnip, and fennel, which are cultivated worldwide for their edible taproots, stems, leaves, or seeds that add flavor, nutrients and medicinal value to

human diets [4]. Many Apiaceae species contain bioactive compounds and essential oils that give them antimicrobial, insecticidal, and pharmaceutical properties. Popular medicinal Apiaceae plants include anise, caraway, cumin, coriander, dill, and asafetida and are used extensively in traditional medicine systems and herbal remedies globally. The Apiaceae family has a worldwide distribution but is most abundant in temperate climates. Several genera, such as *Ferula*, *Prangos* and *Kelussia*, predominantly occur around the Mediterranean and across West and Central Asia [5,6]. Understanding the phytochemical diversity of Apiaceae species through genomic and metabolomic studies can aid in selective breeding efforts and expand the utilization of medicinal members within this vital plant family for improved food security and enhanced medication use [6,7].

The genus *Ferula* belongs to the economically significant Apiaceae family and comprises more than 170 species of flowering plants. Most members of this genus are perennial herbaceous plants that grow in arid grasslands and mountains ranging from Central Asia to northern Africa [8]. The characteristic features of *Ferula* species include thick roots, sturdy hollow stems, and feathery leaves, and they produce resins and gums that contain bioactive compounds such as terpenoids, coumarins, and alkaloids [9]. These specialized metabolites confer antispasmodic, antimicrobial, and anti-inflammatory effects to *Ferula* plants, underlying their extensive use in traditional medicine systems in Iran and neighboring countries [9,10]. *Ferula pseudalliacea* is an indigenous medicinal plant species endemic to the Zagros mountainous regions of Iran's Kurdistan Province and western border areas. This species has been traditionally utilized in folk medicine for various purposes, including as an antiparasitic and antiulcer remedy [11]. Several new bioactive sesquiterpene coumarins, including kamanolol acetate, farnesiferol B, sanandajin, ethyl galbanate, methyl galbanate, and fekrynol acetate, have been isolated and identified from *F. pseudalliacea* [12]. These compounds have exhibited potent anticancer, antiplasmodial, antibacterial and strong antioxidant effects, revealing the medicinal potential of these endemic species [13,14]. However, despite harboring such valuable specialized metabolites, *F. pseudalliacea* has a narrow geographical range with scattered fragmented populations, leading to its classification as a threatened plant [15]. With rich phytochemical diversity and proven pharmacological properties, further conservation efforts and molecular research on *F. pseudalliacea* can facilitate sustainable utilization of this medicinal species as well as the discovery of novel bioactive lead compounds.

Cinnamate 4-hydroxylase (C4H) is a cytochrome P450-dependent monooxygenase that catalyzes a key step in phenylpropanoid metabolism across plant species. It controls the conversion of trans-cinnamic acid to p-coumaric acid, funneling metabolites toward myriad phenolic end products such as lignins, coumarins, flavonoids, anthocyanins, and stilbenoids [16,17]. These phenolic compounds act as structural barriers or chemical defenses in plants while also exerting beneficial antioxidant, anti-inflammatory, and antimicrobial effects. C4H belongs to the CYP73A enzyme family and is encoded by a multigene family whose members are induced differentially by biotic and abiotic stresses [18]. Complete genomic sequences and individual *C4H* gene functions have already been well characterized in model crop species using molecular genetics tools [19]. Structural modeling revealed conserved heme-binding and substrate recognition sites across C4H orthologs from diverse taxa [20]. However, *C4H* has been cloned and sequenced in several medicinal plant species, such as *Salvia miltiorrhiza* [21], *Scrophularia striata* [22], *Euphorbia maculata* [23], and *Carthamus tinctorius* [24]; its sequence has not been explored among key medicinal species, such as *Ferula*, which are rich reservoirs of phenylpropanoid-derived chemicals. Elucidating *C4H* genes in such plants can reveal their evolutionary relationships and regulatory patterns governing species-specific phenolic secondary metabolism. Overall, advances in C4H research have allowed improved metabolic engineering strategies for enhancing the biosynthesis of high-value phenolics in medicinal and crop plants. Currently, little information exists regarding the sequence and expression patterns of *C4H* in industrially and conservationally important *Ferula* species, particularly *F. pseudalliacea*. Hence, the current study aimed to i) isolate and sequence the *C4H* gene from *F. pseudalliacea* using 3'-RACE methods; ii) conduct bioinformatic analysis of the obtained C4H sequences, including conserved domains, evolutionary relationships, and predicted structural features; and iii) quantify the

expression profiles of individual *C4H* genes across roots, stems, leaves, flowers and seeds of *F. pseudalliacea* plants via real-time PCR.

2. Materials and Methods

2.1. Plant material

Multiple tissues, including roots, stems, leaves, flowers and seeds, were harvested from healthy, uniform *Ferula pseudalliacea* individuals (Figure 1) growing in their native habitat site located in Gazne village (10 km west of Sanandaj, coordinates 35°08'38.1"N 46°57'45.4"E), Iran. The collected samples were promptly frozen in liquid nitrogen upon site extraction and transported to the biotechnology laboratory at the University of Kurdistan (UOK) under cold chain maintenance at -80°C for subsequent experiments. Botanical authentication of the *F. pseudalliacea* specimens utilized in the current investigation was performed by Eng. Hossein Maroofi (Agriculture and Natural Resources Research Center). Key morphological details have been deposited in the herbarium collection housed in the Faculty of Agriculture, University of Kurdistan, Iran, for future verification.

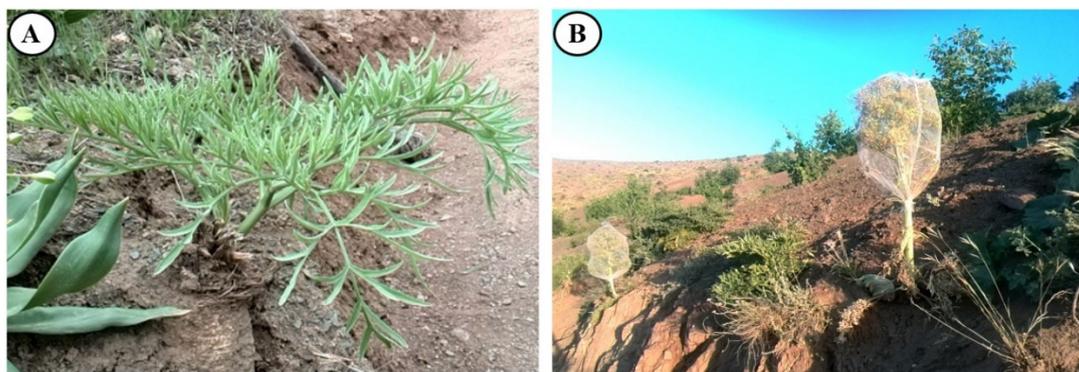


Figure 1. Morphology of *Ferula pseudalliacea*. A. Young leaf development; B. Fully developed inflorescence covered for self-pollination and protection.

2.2. DNA and RNA extraction and cDNA synthesis

Genomic DNA was extracted from the leaf samples using the cetyltrimethylammonium bromide (CTAB) protocol [25], with certain modifications. Purified DNA was subjected to digestion with Ribonuclease A (34390, SERVA, Heidelberg, Germany) at 37°C for 1 hour to eliminate contaminating RNA. Total RNA was isolated from the assorted frozen *Ferula pseudalliacea* tissues, including roots, stems, leaves, flowers and seeds, utilizing the methodology delineated by Mazzara and James [26]. The RNA samples were further treated with DNase I (Pishgam Biotech, Tehran, Iran) to eliminate DNA contamination. Quantification and integrity analyses of the extracted DNA and RNA samples were carried out with a NanoDrop OneC device (Thermo Scientific TM) coupled with agarose gel electrophoresis visualization. Subsequently, first-strand complementary DNA (cDNA) was synthesized from 1 µg of pure RNA template from each sample using a commercially available reverse transcriptase kit (Yekta Tajhiz Azma Co., Tehran, Iran) per the manufacturer's protocols.

2.3. Primer design and gene cloning and sequencing

Due to the lack of genomic resources for *F. pseudalliacea*, primer design targeted highly conserved regions within cinnamate 4-hydroxylase (C4H) coding sequences and actin reference gene sequences from other Apiaceae taxa, including *Angelica gigas*, *Angelica dahurica*, *Daucus carota*, and *Petroselinum crispum*, with available C4H protein data. Multiple sequence alignments were constructed using Clustal Omega [27] to identify consensus protein sequences, which were further utilized to develop degenerate primers via Primer3 [28]. *In silico* analyses enabled screening for self-dimers, secondary structures and mispriming sites using the Oligo Analyzer v3.1 online platform (<http://eu.idtdna.com>).

Initial PCR was used to amplify a core FpC4H fragment, which was subsequently amplified via 3'-RACE to obtain full-length transcripts via the use of an oligo-dT anchor primer paired with a specific forward primer (Table 1). Amplicons representing putative FpC4H genes were gel-purified using the N0041101 Gene MARK Kit (BioElegen Tech, Taichung, Taiwan) and TA-cloned and inserted into the pTG19-T vector. The constructs were subsequently transformed into *E. coli* DH5a cells via heat shock and subsequently plated for blue-white screening and colony PCR validation using M13 universal and gene-specific primers [29]. Bidirectional Sanger sequencing of 10 positive clone plasmids per transformation was performed (Applied Biosystems, Foster City, CA, USA), yielding common FpC4H sequences that were edited with Geneious software [30] and deposited into the NCBI under accession number MH987776.1. A BLASTx homology search was used to confirm sequence identity against nonredundant databases.

Table 1. List of primers used in the present study to clone the FpC4H gene and to study the expression patterns of the genes in *F. pseudalliacea* organs. F: forward primer, R: reverse primer, RT: real-time PCR.

Primer name	Sequence (5'-3')	tm °C	Product size
FpC4H-F	TCATGTTTGATAGAAGGTTTGAGAG	61	890
FpC4H-R	GAAGTGTCTCTCTCTCTGCTG		
Oligo(dT) primer	GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTTTTT		
PCR anchor primer	GACCACGCGTATCGATGTCGAC		
F-FpC4H.RT	ACAGCTTGAAGTGTGCCATG	56	185
R-FpC4H.RT	GCACAGTGTCCATCTCATGC		
F-FpACTIN.RT	GCCATCTATGATTGGGAATGG	56	190
R-FpACTIN.RT	GCCACCACCTTGATCTTCATG		

2.4. Phylogeny analysis

To investigate the evolutionary relationships of the FpC4H proteins, multiple sequence alignments of the deduced FpC4H proteins and known orthologs from *Arabidopsis thaliana*, *Oryza sativa*, *Brachypodium distachyon*, *Daucus carota*, *Brassica napus*, *Helianthus annuus* and *Camelina sativa* were constructed using Clustal Omega. The alignment was analyzed using the IQ-TREE web server [31] for maximum likelihood phylogenetic reconstruction with 1000 bootstrap replicates. The resulting phylogenetic tree was visualized using iTOL online software [32]. Conserved protein domain prediction of FpC4H sequences was also carried out using the InterProScan tool [33].

2.5. Sequence structure analysis

The prediction of putative phosphorylation sites in FpC4H relied on the NetPhos 3.1 server [34], for which a threshold score ≥ 0.90 was used. A protein interaction network analysis was performed for FpC4H based on the STRING database [35] and carrot ortholog information. Significant biological processes and molecular function terms linked to the network nodes were identified using an FDR ≤ 0.05 . Additionally, various physicochemical properties of the FpC4H protein and their orthologous sequences were computed using the ProtParam online tool [36]. In addition, the 3D structure and binding sites of the FpC4H protein were investigated using the online Phyre2 tool [37].

2.6. Real-time PCR analysis

Quantitative real-time PCR (qPCR) analysis was performed to examine and compare the expression profiles of FpC4H across different *F. pseudalliacea* organs, including seeds, flowers, stems, leaves, and roots. Gene-specific primers targeting 3'-UTR regions were utilized for amplification from synthesized cDNA templates. The *FpACTIN* housekeeping gene was coamplified in parallel reactions as an internal normalization control. qPCR was performed in three biological replicates with two technical replicates per sample in 15 μ L volumes by applying SYBR Green chemistry (Yekta Tajhiz

Azma, Tehran, Iran) on an ABI StepOne Plus thermal cycler (Life Technologies, CA, USA). The cycling conditions were set to initial denaturation at 95°C for 3 min, followed by 40 cycles of 95°C for 20 sec and 60°C for 30 min. Melt curve analysis validated the single peak and specific amplification. Relative quantification relied on the $\Delta\Delta C_t$ method [38] using *FpACTIN* for normalization, with the lowest expressing tissue set used as a calibrator. Statistical analyses included one-way ANOVA accompanied by Tukey's post hoc test performed with SPSS software (ver. 27.0.1.0).

3. Results

3.1. Screening and confirmation of clones

In this study, one isoform of the *C4H* gene in the *F. pseudaliacea* genome (named *FpC4H*) was identified and registered in the NCBI database with accession number MH987776. The sequence of *FpC4H* was most similar to that of *C4H* proteins from *Daucus carota*. In the present study, the physicochemical properties of *FpC4H* and its orthologs were compared (Table 2). The results revealed that all the studied *C4H* proteins are unstable and have a negative GRAVY, indicating that the *C4H*s are more hydrophilic proteins. According to the isoelectric point (pI) value, the *FpC4H* protein was predicted to be active in near-neutral environments, while other orthologs were more active under alkaline conditions.

Table 2. Physicochemical properties of *FpC4H* and its orthologs.

Gene ID	Plant	Length (aa)	MW (kDa)	Instability index	GRAVY	pI
<i>FpC4H</i> (MH987776)	<i>F. pseudaliacea</i>	299	34.54	Unstable	-0.342	5.88
DCAR_018641	<i>Daucus carota</i>	505	58.10	Unstable	-0.215	9.00
DCAR_027611	<i>Daucus carota</i>	498	57.00	Unstable	-0.216	9.21
BVRB_7g175810	<i>Beta vulgaris</i>	504	58.09	Unstable	-0.298	9.14
HanXRQr2_Chr08g0346771	<i>Helianthus annuus</i>	505	57.93	Unstable	-0.216	9.16
HanXRQr2_Chr07g0288081	<i>Helianthus annuus</i>	502	57.55	Unstable	-0.217	8.92
HanXRQr2_Chr06g0265981	<i>Helianthus annuus</i>	505	57.93	Unstable	-0.204	9.07
AT2G30490	<i>A. thaliana</i>	505	57.79	Unstable	-0.185	8.89
Os02g0467600	<i>O. sativa</i>	533	59.48	Unstable	0.007	8.77
Os05g0320700	<i>O. sativa</i>	500	56.88	Unstable	-0.102	9.02

3.2. Phylogenetic analysis

Evolutionary analysis of *C4H* genes revealed that this gene has undergone a different evolutionary process in monocotyledonous and dicotyledonous plants (Figure 2); thus, *FpC4H* orthologs in monocotyledonous plants showed greater genetic distance than orthologs in dicotyledonous plants. Based on the genetic distance, the *C4H* genes were divided into four groups. *FpC4H* and its orthologs from carrot and sugar beet were placed in group 3, which also showed less genetic distance from the sunflower *C4H* genes in group 2. *C4H*s from monocotyledonous plants were placed in group 1, and *Arabidopsis* *C4H*, along with its orthologs in rapeseed and camelina, were placed in group four. A separate multiple sequence alignment analysis was performed for the members of groups 2 and 3 along with *Arabidopsis* *C4H* to better understand the sequence structure of *C4H*s. The results revealed that there is a conserved region associated with the corresponding domain (cytochrome P450) in these proteins (Figure 3). Interestingly, a region of 11 amino acids was identified in the sequences, which was presented as a variable region (VR) that might be related to the variable function of *C4H*s in different plants, although further research is needed.

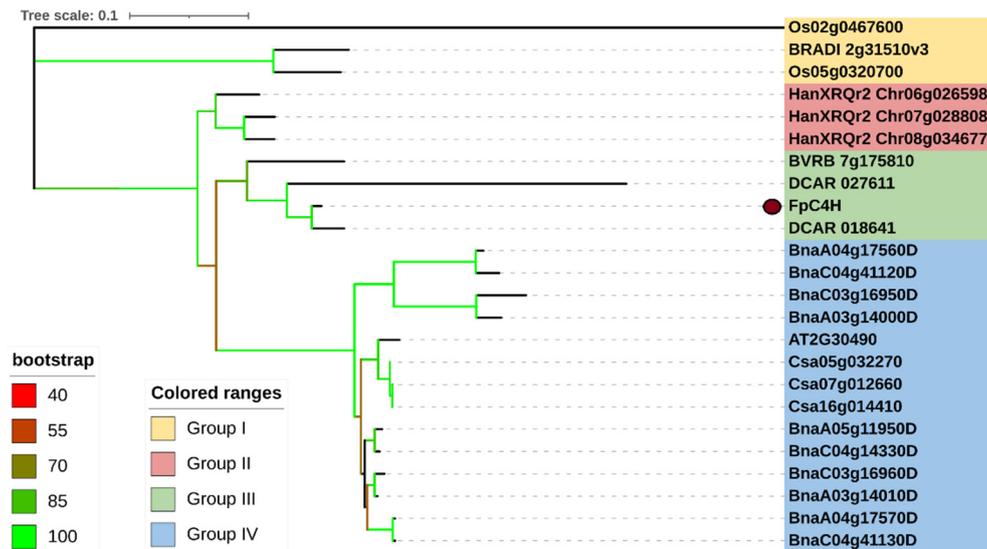


Figure 2. Phylogenetic tree of C4H in different plants. C4H genes from *O. sativa* (starting with Os), *Brachypodium distachyon* (starting with BRADI), *Helianthus annuus* (starting with HanXRQr), *Beta vulgaris* (starting with BVRB), *Daucus carota* (starting with DCAR), *B. napus* (starting with Bna), *Arabidopsis* (starting with AT), and *C. sativa* (starting with Csa) were used.

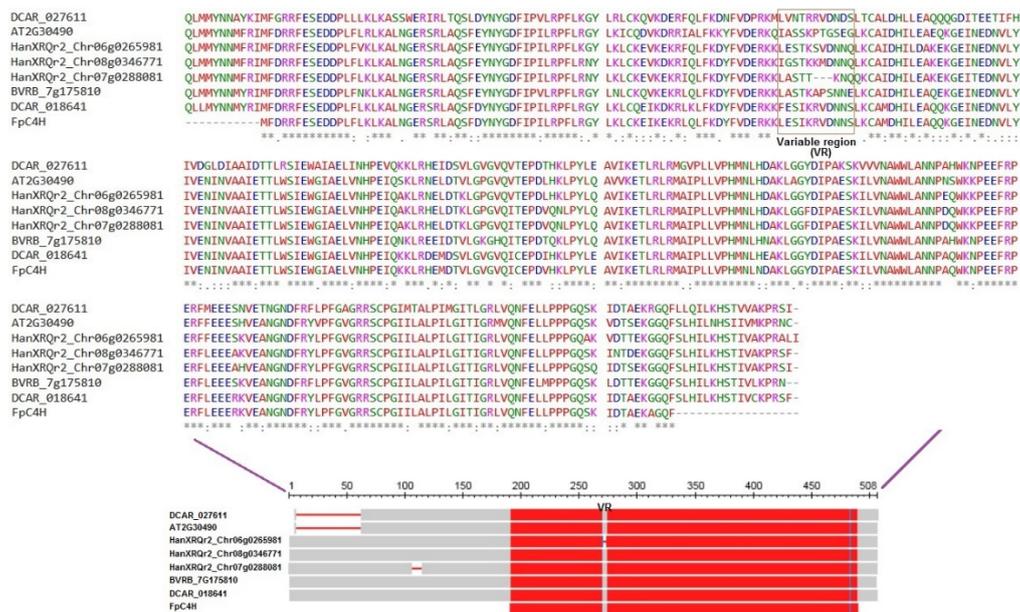


Figure 3. Multiple sequence alignment analysis of FpC4H and its close orthologs.

3.3. Three-dimensional structure and binding sites of FpC4H

The three-dimensional structure of FpC4H was predicted based on a homolog model with high identity and 100% confidence in the model (Figure 4). In addition, the binding regions that affect the activity and possible interactions were identified (Figure 4a). The results revealed that isoleucine (I) and leucine (L) residues were more abundant in the binding region of the FpC4H protein. In addition, two regions, C258 (cysteine 258) and N217 (asparagine 217), were predicted to be ion binding regions (Figure 4b). In addition, the mutation sensitivity of the FpC4H structure was analyzed, the results of which showed that the FpC4H structure has low sensitivity to mutations (Figure 4b).

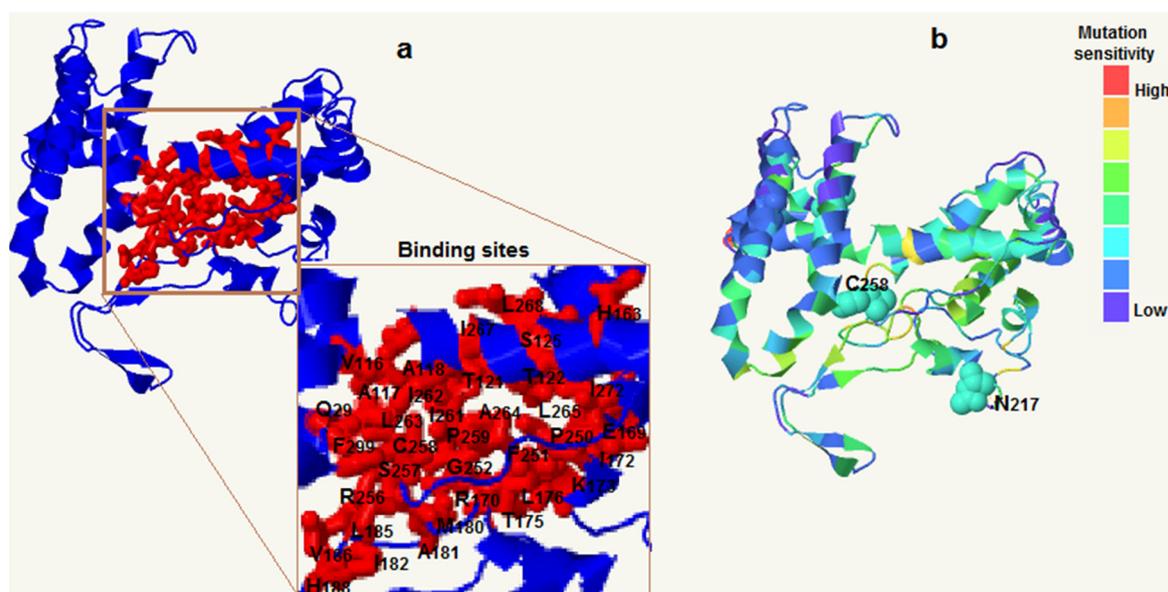


Figure 4. 3D structure of the FpC4H protein based on the homology method. Binding site of the FpC4H protein based on pocket analysis (a) and rate of mutation sensitivity of the structure of the FpC4H protein (b). Abbreviations: Alanine (A); Cysteine (C); Aspartic acid (D); Glutamic acid (E); Phenylalanine (F); Glycine (G); Histidine (H); Isoleucine (I); Lysine (K); Leucine (L); Methionine (M); Asparagine (N); Proline (P); Glutamine (Q); Arginine (R); Serine (S); Threonine (T); Valine (V); Tryptophan (W); Tyrosine (Y).

3.4. FpC4H protein phosphorylation events

Posttranslational modifications affect the activity and stability of target proteins. In the present study, the possible phosphorylation sites of FpC4H within its close orthologs were identified (Figure 5). Four serine residues were identified as phosphorylation sites in FpC4H as well as in its close ortholog in carrot (DCAR_018641). The phosphorylation sites ranged between two sites (HanXRQr2_Chr06g0265981 and Os05g0320700) and nine sites in Arabidopsis C4H (Figure 5a). The distribution pattern of the phosphorylation regions of the C4H proteins was somewhat variable, although some regions were almost common to most proteins, such as serine 266, 197, and 477 (Figure 5b). These results suggest that some posttranslational events in these proteins have remained conserved during the evolutionary process.

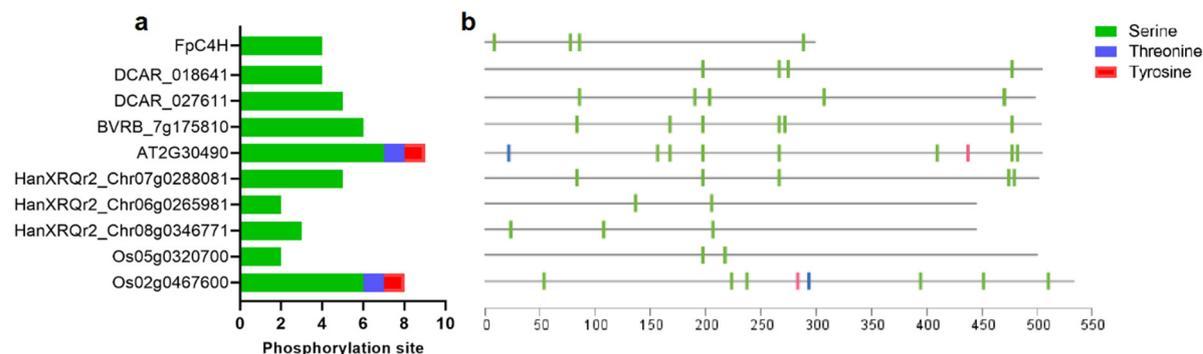


Figure 5. Frequency (a) and distribution (b) of the FpC4H phosphorylation site. Phosphorylation regions were estimated with a confidence level above 90%.

3.5. Interaction network of FpC4H

An interaction network for FpC4H was constructed based on available data from carrot plants (Figure 6). The FpC4H homolog, a key point in the network, strongly interacts with proteins involved in diverse pathways. C4H-1 interacted with other proteins from the phenylpropanoid pathway, including C4H-2, PAL1, PAL3, and PAL4 (Figure 6a). In addition, several interactions were observed between CHS (chalcone synthase) and 4CL3 (4-coumarate:coenzyme A ligase). In addition, Gene Ontology (GO) enrichment analysis was performed to determine the molecular functions and biological processes associated with the C4H network elements (Figure 6b and 6c). The molecular function terms chalcone synthase activity, oxidoreductase activity, cinnamoyl-CoA reductase activity, acyltransferase activity, 4-coumarate-CoA ligase activity and phenylalanine ammonia-lyase activity were strongly linked with the C4H network (Figure 6b). In addition, several biological process terms, such as secondary metabolic process, phenylpropanoid metabolic/biosynthetic process, polyketide biosynthetic process, flavonoid biosynthetic process, aromatic amino acid family catabolic process, and cinnamic acid biosynthetic process, were significantly enriched (Figure 6c). These findings revealed that C4H is involved in several key metabolite pathways.

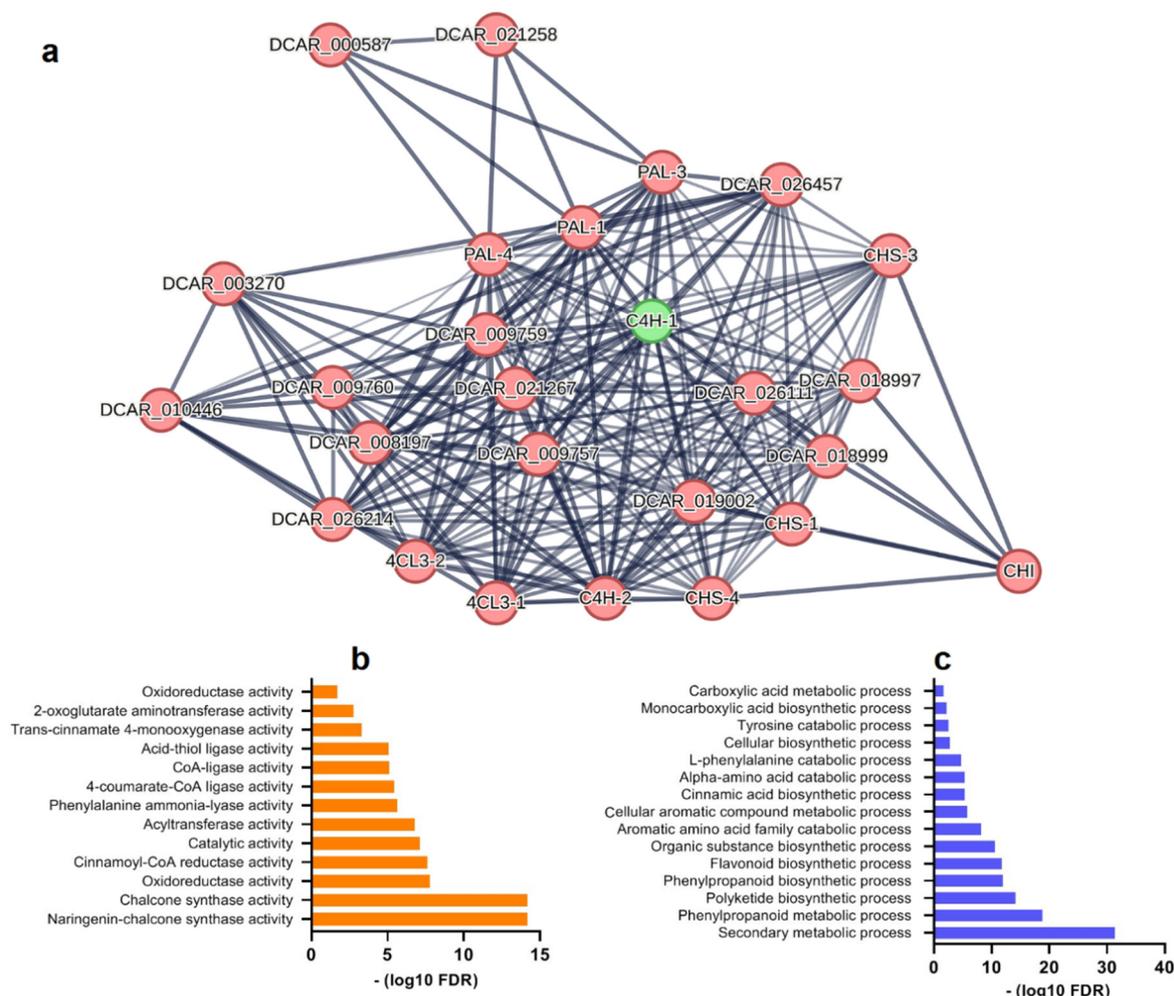


Figure 6. Interaction network of the FpC4H protein based on available data from *Daucus carota* (a). GO enrichment analysis of network nodes in terms of molecular function (b) and biological process terms (c).

3.6. Analysis of *FpC4H* gene expression

In the present study, the expression levels of the *FpC4H* gene were evaluated in different organs of *F. pseudallicea* (Figure 7). Compared with those in other tissues, *FpC4H* in leaves was expressed at lower levels. The *FpC4H* gene was more highly expressed in flower tissues, and the *FpC4H* transcript levels were equal in stem and root tissues. In addition, the *FpC4H* gene was expressed at lower levels in seed tissues than in flowers, stems, and roots. The variable expression of the *FpC4H* gene indicated that this gene is more strongly induced in reproductive organs and is involved in the increase in linked metabolites in flower tissues.

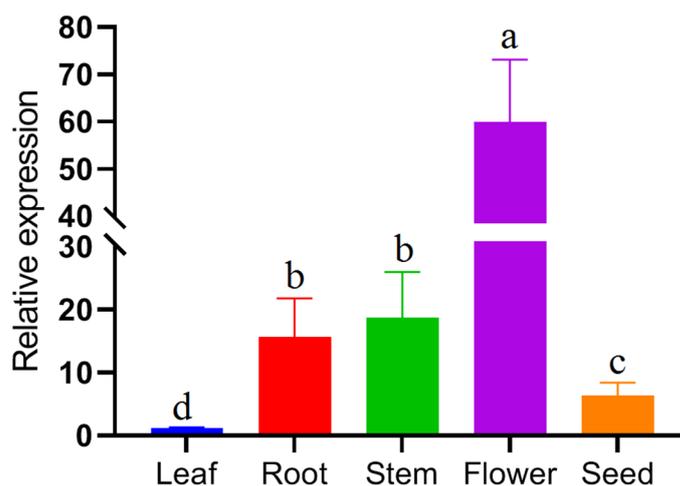


Figure 7. Expression levels of the *FpC4H* gene in different organs of *F. pseudallicea*. Columns with common letters have no significant difference at 5% probability level.

4. Discussion

C4H is involved in the biosynthesis of phenylpropanoids, which are important active plant compounds that participate in the production of valuable secondary metabolites. There is not enough information on the *C4H* gene as a key gene in the biosynthesis of several secondary metabolites in the *Ferula* genus. In this study, the isolation and identification of this gene and its expression level in different organs of *Ferula pseudallicea* plants were carried out. The results revealed that *F. pseudallicea* has one isoform of the *C4H* gene (*FpC4H*). Only one member of this gene has been identified in *Arabidopsis* [39], parsley [39], cowpea [40] and black raspberry [41]. However, in sycamore [42], subabel tree [43] and rapeseed [44], *C4H* is encoded by a small gene family. For instance, four *C4H* genes were found in the soybean genome [20] and *Populus kitakamiensis* genome [45]. The number of copies of a gene can be related to polyploidy events and segmental and tandem duplications [46,47]. Increasing the number of isoforms of a gene can be effective at increasing functional diversity and related interactions [48]. Phylogenetic analysis of *C4H* proteins revealed that there is a large genetic distance between the *C4H* genes of monocot and dicot plants, suggesting that they have undergone a different evolutionary process. Most likely, after the derivation of dicotyledons and monocotyledons, structural diversity occurred in the *C4H* genes. Despite this genetic difference, *C4H*s have similar physicochemical characteristics, revealing that the conditions and environment of activity are the same. A region (consisting of eleven amino acids) was identified in the conserved region of *C4H* proteins, which showed high variation between orthologs. This region could be involved in the functional diversity of orthologs, although further molecular studies on this region are needed.

An investigation of the *FpC4H* protein structure revealed that there is an extensive binding region in this protein. The presence of many binding regions in a protein can affect its extent of activity and interaction. Additionally, the amino acids isoleucine and leucine likely play more prominent roles in the binding and interaction of *FpC4H*. In addition, two ion binding regions, cysteine 258 and asparagine 217, were recognized in the *FpC4H* structure. The common feature of

CYP450 family members is the presence of an iron core in their protein structure, which is connected to a cysteine amino acid [49]. Additionally, four serine residues were identified as potential phosphorylation sites in the FpC4H protein. Phosphorylation is a posttranslational modification that affects the activity and life of a protein [50]. For proteins such as C4H, which is an upstream element in metabolite synthesis pathways, phosphorylation plays a critical role in controlling downstream components by influencing these proteins. The results of the interaction network confirmed that FpC4H interacts with proteins involved in the biosynthesis of secondary metabolites. For example, biological process terms, including phenylpropanoid metabolic/biosynthetic process, polyketide biosynthetic process, flavonoid biosynthetic process, aromatic amino acid family catabolic process, and cinnamic acid biosynthetic process, were significantly enriched in FpC4H (Figure 6c). Phenylpropanoids and flavonoids play key roles in the response to adverse conditions, including abiotic and biotic stresses [51], as well as in plant growth processes [52]. Previous studies have shown that the expression of the *C4H* gene in the leaves of *Lepidium apetalum* after methyl jasmonate treatment is correlated with the flavonoid biosynthesis pathway [53]. Additionally, the flavonoid content at different stages in the tea plant *Camellia sinensis* was correlated with the expression pattern of *C4Hb* [54]. In addition, downregulation of the *C4H* gene in *Nicotiana tabacum* via the use of CRISPRi could increase flavonoid biosynthesis [55]. According to these results, it is hypothesized that the *C4H* gene is an important element in the biosynthesis of secondary metabolites and that this gene affects the improvement of plant resistance to adverse environmental conditions.

The *FpC4H* gene exhibited diverse expression levels in different organs, including the roots, stems, leaves, flowers, and seeds of *F. pseudallicea*, and the expression of this gene was detected in all organs (Figure 7). The results revealed that *FpC4H* is expressed in a tissue-specific manner. For example, the *FpC4H* gene was more strongly induced in floral tissues than in other tissues. In contrast, the *C4H-2* gene was more highly expressed in rapeseed seeds than in roots and flowers [44]. Moreover, diverse expression levels were observed between four *C4H* genes in the soybean genome [20]. Metabolomics analysis of different organs of *F. pseudallicea* revealed that phenolic compounds accumulated more in flowers than in other organs [11]. The high expression level of the *FpC4H* gene in reproductive organs such as flowers suggested that this gene is positively correlated with the production of phenolic compounds, which can be considered in metabolite engineering studies.

5. Conclusion

In the present study, cinnamate 4-hydroxylase (*FpC4H*) was cloned and characterized in *Ferula pseudallicea*. The physicochemical properties revealed that FpC4H is an unstable and hydrophilic protein. Phylogenetic analysis revealed that FpC4H and C4H proteins from *Daucus carota* underwent almost the same evolutionary process. In addition, the isoleucine and leucine residues were identified as key elements affecting the interaction and function of FpC4H. Expression analysis revealed that the *FpC4H* gene can be induced in all organs of *F. pseudallicea*, and a high expression level was observed in the flower tissues. Overall, these findings suggested that *FpC4H* is associated with the concentration of phenolic compounds in flowers.

Author Contributions: Conceptualization, P.S., B.B., Y.V. and D.D.; methodology, P.S., B.B. and Y.V.; validation, B.B., Y.V. and D.D.; formal analysis, P.S., B.B., Y.V. and P.H.; investigation, P.S., B.B. and Y.V.; writing—original draft preparation, Y.V. and P.H.; writing—review and editing, Y.V. and P.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data generated or analyzed in this study are included in this article. Other materials that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments: We thank the University of Kurdistan for financially supporting and providing the needed facilities.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Buso, P.; Manfredini, S.; Reza Ahmadi-Ashtiani, H.; Sciabica, S.; Buzzi, R.; Vertuani, S.; Baldisserotto, A. Iranian medicinal plants: from ethnomedicine to actual studies. *Medicina (B. Aires)*. 2020, *56*, 97.
2. Amiri, M.S.; Yazdi, M.E.T.; Rahnama, M. Medicinal plants and phytotherapy in Iran: Glorious history, current status and future prospects. *Plant Sci. Today*, 2021, *8*, 95–111.
3. Hassanpouraghdam, M.B.; Ghorbani, H.; Esmaeilpour, M.; Alford, M.H.; Strzemeski, M.; Dresler, S. Diversity and distribution patterns of endemic medicinal and aromatic plants of Iran: implications for conservation and habitat management. *Int. J. Environ. Res. Public Health* 2022, *19*, 1552.
4. Amiri, M.S.; Joharchi, M.R. Ethnobotanical knowledge of Apiaceae family in Iran: A review. *Avicenna J. phytomedicine* 2016, *6*, 621.
5. Sayed-Ahmad, B.; Talou, T.; Saad, Z.; Hijazi, A.; Merah, O. The Apiaceae: Ethnomedicinal family as source for industrial uses. *Ind. Crops Prod.* 2017, *109*, 661–671.
6. Wang, X.-J.; Luo, Q.; Li, T.; Meng, P.-H.; Pu, Y.-T.; Liu, J.-X.; Zhang, J.; Liu, H.; Tan, G.-F.; Xiong, A.-S. Origin, evolution, breeding, and omics of Apiaceae: a family of vegetables and medicinal plants. *Hortic. Res.* 2022, *9*, uhac076.
7. Palumbo, F.; Vannozzi, A.; Barcaccia, G. Impact of genomic and transcriptomic resources on apiaceae crop breeding strategies. *Int. J. Mol. Sci.* 2021, *22*, 9713.
8. Beck, J.J.; Chou, S.-C. The structural diversity of phthalides from the Apiaceae. *J. Nat. Prod.* 2007, *70*, 891–900.
9. Salehi, M.; Naghavi, M.R.; Bahmankar, M. A review of *Ferula* species: Biochemical characteristics, pharmaceutical and industrial applications, and suggestions for biotechnologists. *Ind. Crops Prod.* 2019, *139*, 111511.
10. Mohammadhosseini, M.; Venditti, A.; Sarker, S.D.; Nahar, L.; Akbarzadeh, A. The genus *Ferula*: Ethnobotany, phytochemistry and bioactivities—A review. *Ind. Crops Prod.* 2019, *129*, 350–394.
11. Dastan, D.; Hamah-Ameen, B.A.; Salehi, P.; Ghaderi, H.; Miran, M. Chemical composition and bioactivities of essential oils from different plant parts of *Ferula pseudalliacea* Rech. f. as an endemic plant from Iran. *Nat. Prod. Res.* 2022, *36*, 1311–1316.
12. Dastan, D.; Salehi, P.; Gohari, A.R.; Zimmermann, S.; Kaiser, M.; Hamburger, M.; Khavasi, H.R.; Ebrahimi, S.N. Disesquiterpene and sesquiterpene coumarins from *Ferula pseudalliacea*, and determination of their absolute configurations. *Phytochemistry* 2012, *78*, 170–178.
13. Dastan, D.; Salehi, P.; Gohari, A.R.; Ebrahimi, S.N.; Aliahmadi, A.; Hamburger, M. Bioactive sesquiterpene coumarins from *Ferula pseudalliacea*. *Planta Med.* 2014, *80*, 1118–1123.
14. Mahaki, H.; Tanzadehpanah, H.; Abou-Zied, O.K.; Moghadam, N.H.; Bahmani, A.; Salehzadeh, S.; Dastan, D.; Saidijam, M. Cytotoxicity and antioxidant activity of Kamolonol acetate from *Ferula pseudalliacea*, and studying its interactions with calf thymus DNA (ct-DNA) and human serum albumin (HSA) by spectroscopic and molecular docking techniques. *Process Biochem.* 2019, *79*, 203–213.
15. Jafari, S.H.; Sepehry, A.; Soltanloo, C. H.; Karimian, D. A. A. Effect of topography and soil properties on distribution of *Ferula pseudalliacea* (Bitter Asafetida) in Yazd Province, Iran. *J. Rangel. Sci.* 2019, *9*, 184.
16. Ehrling, J.; Hamberger, B.; Million-Rousseau, R.; Werck-Reichhart, D. Cytochromes P450 in phenolic metabolism. *Phytochem. Rev.* 2006, *5*, 239–270.
17. Shuab, R.; Lone, R.; Koul, K.K. Cinnamate and cinnamate derivatives in plants. *Acta Physiol. Plant.* 2016, *38*, 1–9.
18. Kumar, K.; Debnath, P.; Singh, S.; Kumar, N. An overview of plant phenolics and their involvement in abiotic stress tolerance. *Stresses* 2023, *3*, 570–585.
19. Hansen, C.C.; Nelson, D.R.; Møller, B.L.; Werck-Reichhart, D. Plant cytochrome P450 plasticity and evolution. *Mol. Plant* 2021, *14*, 1244–1265.
20. Khatri, P.; Chen, L.; Rajcan, I.; Dhaubhadel, S. Functional characterization of Cinnamate 4-hydroxylase gene family in soybean (*Glycine max*). *PLoS One* 2023, *18*, e0285698.
21. Huang, B.; Duan, Y.; Yi, B.; Sun, L.; Lu, B.; Yu, X.; Sun, H.; Zhang, H.; Chen, W. Characterization and expression profiling of cinnamate 4-hydroxylase gene from *Salvia miltiorrhiza* in rosmarinic acid biosynthesis pathway. *Russ. J. Plant Physiol.* 2008, *55*, 390–399.
22. Rostami, Z.; Fazeli, A.; Hojati, Z. The isolation and expression analysis of cinnamate 4-hydroxylase and chalcone synthase genes of *Scrophularia striata* under different abiotic elicitors. *Sci. Rep.* 2022, *12*, 8128.
23. Guo, S.B.; Huang, S.H.; Zou, J.X.; Liu, H.S.; Quan, W.J.; Yang, H. Cloning and analysis of cinnamate acid 4-hydroxylase gene and its promoter from *Euphorbia maculata* L. *Mol Plant Breed* 2023, *21*, 819–825.

24. Hou, Y.; Wang, Y.; Liu, X.; Ahmad, N.; Wang, N.; Jin, L.; Yao, N.; Liu, X. A Cinnamate 4-HYDROXYLASE1 from Safflower Promotes Flavonoids Accumulation and Stimulates Antioxidant Defense System in Arabidopsis. *Int. J. Mol. Sci.* 2023, *24*, 5393.
25. Rogers, S.O.; Bendich, A.J. Extraction of DNA from plant tissues. *Plant Mol. Biol. Man.* 1989, 73–83.
26. Mazzara, M.; James, D.J. The influence of photoperiodic growth condition on isolation of RNA from strawberry (*Fragaria× ananassa* Duch.) tissue. *Mol. Biotechnol.* 2000, *15*, 237–241.
27. Sievers, F.; Higgins, D.G. Clustal omega. *Curr. Protoc. Bioinforma.* 2014, *48*, 3–13.
28. Untergasser, A.; Cutcutache, I.; Koressaar, T.; Ye, J.; Faircloth, B.C.; Remm, M.; Rozen, S.G. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 2012, *40*, e115–e115.
29. Green, M.R.; Sambrook, J. Molecular cloning. *A Lab. Man.* 4th 2012.
30. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012, *28*, 1647–1649.
31. Nguyen, L.-T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating Maximum-likelihood phylogenies. *Mol. Biol. Evol.* 2015, *32*, 268–274.
32. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 2021, *49*, W293–W296.
33. Quevillon, E.; Silventoinen, V.; Pillai, S.; Harte, N.; Mulder, N.; Apweiler, R.; Lopez, R. InterProScan: protein domains identifier. *Nucleic Acids Res.* 2005, *33*, W116–W120.
34. Blom, N.; Gammeltoft, S.; Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* 1999, *294*, 1351–1362.
35. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* 2019, *47*, D607–D613.
36. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein identification and analysis tools on the ExpASY server. In *The Proteomics Protocols Handbook*; Humana Press: Totowa, NJ, 2005; pp. 571–607.
37. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 2015, *10*, 845–858.
38. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻ΔΔCT method. *methods* 2001, *25*, 402–408.
39. Mizutani, M.; Ohta, D.; Sato, R. Isolation of a cDNA and a genomic clone encoding cinnamate 4-hydroxylase from Arabidopsis and its expression manner in planta. *Plant Physiol.* 1997, *113*, 755–763.
40. Xu, H.; Park, N. II; Li, X.; Kim, Y.K.; Lee, S.Y.; Park, S.U. Molecular cloning and characterization of phenylalanine ammonia-lyase, cinnamate 4-hydroxylase and genes involved in flavone biosynthesis in *Scutellaria baicalensis*. *Bioresour. Technol.* 2010, *101*, 9715–9722.
41. Baek, M.-H.; Chung, B.-Y.; Kim, J.-H.; Kim, J.-S.; Lee, S.-S.; An, B.-C.; Lee, I.-J.; Kim, T.-H. cDNA cloning and expression pattern of cinnamate-4-hydroxylase in the Korean black raspberry. *BMB Rep.* 2008, *41*, 529–536.
42. Li, W.; Yang, L.; Jiang, L.; Zhang, G.; Luo, Y. Molecular cloning and functional characterization of a cinnamate 4-hydroxylase-encoding gene from *Camptotheca acuminata*. *Acta Physiol. Plant.* 2016, *38*, 1–9.
43. 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.* 2013, *40*, 1265–1274.
44. 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.* 2007, *40*, 247–260.
45. Lu, S.; Zhou, Y.; Li, L.; Chiang, V.L. Distinct roles of cinnamate 4-hydroxylase genes in *Populus*. *Plant Cell Physiol.* 2006, *47*, 905–914.
46. Yaghobi, M.; Heidari, P. Genome-Wide Analysis of Aquaporin Gene Family in *Triticum turgidum* and Its Expression Profile in Response to Salt Stress. *Genes (Basel)*. 2023, *14*, 202.
47. Puresmaeli, F.; Heidari, P.; Lawson, S. Insights into the Sulfate Transporter Gene Family and Its Expression Patterns in Durum Wheat Seedlings under Salinity. *Genes (Basel)*. 2023, *14*, 333.
48. Heidari, P.; Puresmaeli, F.; Vafaee, Y.; Ahmadizadeh, M.; Ensani, M.; Ahmadinia, H. Comparative Analysis of Phospholipase D (PLD) Gene Family in *Camelina sativa* and *Brassica napus* and Its Responses in *Camelina* Seedlings under Salt Stress. *Agronomy* 2023, *13*, 2616.
49. Poulos, T.L. Heme enzyme structure and function. *Chem. Rev.* 2014, *114*, 3919–3962.
50. Rezaee, S.; Ahmadizadeh, M.; Heidari, P. Genome-wide characterization, expression profiling, and post-transcriptional study of GASA gene family. *Gene Reports* 2020, *20*, 100795.
51. Yadav, V.; Wang, Z.; Wei, C.; Amo, A.; Ahmed, B.; Yang, X.; Zhang, X. Phenylpropanoid pathway engineering: An emerging approach towards plant defense. *Pathogens* 2020, *9*, 312.

52. Levy, H.L.; Sarkissian, C.N.; Scriver, C.R. Phenylalanine ammonia lyase (PAL): From discovery to enzyme substitution therapy for phenylketonuria. *Mol. Genet. Metab.* 2018, *124*, 223–229.
53. Zhao, L.; Mali, G.; Yang, Z.A.; Feng, W.S.; Zheng, X.K. expression analysis of C4H gene from *Lepidium apetalum* [J]. *Acta Pharmacol. Sin* 2017, *5*, 821–831.
54. Xia, J.; Liu, Y.; Yao, S.; Li, M.; Zhu, M.; Huang, K.; Gao, L.; Xia, T. Characterization and expression profiling of *Camellia sinensis* cinnamate 4-hydroxylase genes in phenylpropanoid pathways. *Genes (Basel)*. 2017, *8*, 193.
55. Karlson, C.K.S.; Mohd Noor, S.N.; Khalid, N.; Tan, B.C. CRISPRi-mediated down-regulation of the cinnamate-4-hydroxylase (C4H) gene enhances the flavonoid biosynthesis in *Nicotiana tabacum*. *Biology (Basel)*. 2022, *11*, 1127.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.