



OPEN ACCESS

EDITED BY

Peng Wang,
Jiangsu Province and Chinese Academy of
Sciences, China

REVIEWED BY

Yao Jun,
Beijing Forestry University, China
Lina Jiang,
Henan Institute of Science and Technology,
China

*CORRESPONDENCE

Ya Ning
✉ ningya0428@126.com

Wei-bin Wang
✉ wangwb@dongyang-lab.org

Yang Dong

✉ loyalyang@163.com

[†]These authors have contributed
equally to this work and share
first authorship

RECEIVED 24 January 2025

ACCEPTED 12 March 2025

PUBLISHED 28 March 2025

CITATION

Luo K-y, Wang S-p, Yang L, Luo S-l, Cheng J,
Dong Y, Ning Y and Wang W-b (2025)
Evolutionary landscape of plant chalcone
isomerase-fold gene families.
Front. Plant Sci. 16:1559547.
doi: 10.3389/fpls.2025.1559547

COPYRIGHT

© 2025 Luo, Wang, Yang, Luo, Cheng, Dong,
Ning and Wang. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Evolutionary landscape of plant chalcone isomerase-fold gene families

Kai-yong Luo^{1,2†}, Shi-ping Wang^{2,3†}, Ling Yang³, Sen-lin Luo^{1,2},
Jia Cheng^{1,2}, Yang Dong^{2*}, Ya Ning^{4,5*} and Wei-bin Wang^{2,5*}

¹College of Food Science and Technology, Yunnan Agricultural University, Kunming, Yunnan, China, ²Yunnan Provincial Key Laboratory of Biological Big Data, Yunnan Agricultural University, Kunming, Yunnan, China, ³Institute of Agro-Products of Processing and Design, Hainan Academy of Agricultural Sciences, Haikou, China, ⁴Department of Pain Management, The Second Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, China, ⁵College of Science, Yunnan Agricultural University, Kunming, Yunnan, China

Flavonoids are crucial for plant survival and adaptive evolution, and chalcone isomerase (CHI) genes serve as key rate-limiting gene in the flavonoid biosynthesis pathway. It is important for plant adaptive evolution to comprehensively study the evolution and diversity of the CHI gene families. However, the CHI gene families in many plant lineages remain elusive. This study systematically identified CHI genes from 259 species including algae, bryophytes, ferns, gymnosperms, and angiosperms. A total of 1,738 CHI gene family members were discovered. We analyzed the diversity, distribution trajectory, and the driving forces of gene duplication during the evolution of the plant lineages. The present study is the first to identify potential type II and type IV CHI genes in the extant liverwort model species *Marchantia polymorpha*. The distribution pattern of CHI genes across the plant kingdom reveals that the origin of type II CHI can be traced back to the last common ancestor of bryophytes and vascular plants, and type III CHI may represent the ancestral form of the CHI gene family. The identification of conserved motifs showed significant differences in motif distribution among different CHI gene types. It was found that the drivers of gene duplication varied across plant lineages: dispersed duplications (DSD) were predominant in algae and bryophytes, whole-genome duplication (WGD) was the main driver in basal angiosperms and monocots, while tandem duplications (TD) predominating in eudicots. Structural clustering analysis demonstrated the 3-layer sandwich structure in the CHI-fold proteins remained conserved in the central region, while repeated loss of N-terminal sequences contributed to structural diversity. This study provides a deeper understanding of the evolution and diversity of the CHI-fold proteins and lays a theoretical foundation for further studies of their function and the identification of new functional CHI genes.

KEYWORDS

chalcone isomerase, evolution, flavonoids, diversity, structural cluster analysis

Introduction

During the colonization and evolution of terrestrial plants, a vast array of specialized metabolites has been developed to facilitate adaptation and response to a diverse range of biotic and abiotic stresses (Winkel-Shirley, 2002; Agati et al., 2012; Petrusa et al., 2013; Dias et al., 2021). Flavonoids are one of the largest groups of (poly)phenolic compounds, which are extensively distributed in the plant kingdom, and played a pivotal role in terrestrial colonization and adaptation in plant evolution (Mouradov and Spangenberg, 2014; Yonekura-Sakakibara et al., 2019). Flavonoids in plants are critical for a variety of biological processes, including protection against ultraviolet-B (UV-B) radiation, pollinators attraction, phytoalexins, signaling molecules, and auxin transport and fertility regulation (Agati et al., 2012). It is hypothesized that in early flavonoid-producing plants, their primary functions were to provide defense against UV-B exposure and the regulation of plant hormone activity (Rausher, 2006). Over the course of plant evolution, these functions have diversified considerably. The biosynthesis of plant flavonoids primarily originates from phenylalanine metabolism and branches off from the general phenylpropanoid pathway through the catalytic actions of chalcone synthase (CHS) and chalcone isomerase (CHI) (Liu et al., 2021). As the pivotal enzyme in the flavonoid biosynthetic pathway, CHI catalyzes the intramolecular and stereospecific cyclization of chalcone through the Michael addition reaction, thereby facilitating the formation of the fundamental scaffold for subsequent flavonoid synthesis. Subsequently, a diverse array of flavonoids is produced through the modification of superfamilies such as cytochrome P450, oxidoreductase, and UDP glycosyltransferase (Yin et al., 2019).

Chalcone isomerase (CHI-fold proteins, CHI, EC: 5.5.1.6) was first identified and isolated in the study of the mechanism of synthesis of Phenylpropanoid-like natural products in *Phaseolus vulgaris* under environmental stress, and the key role of chalcone isomerase in the synthesis of flavonoids was discovered (Mehdy and Lamb, 1987). After that, CHI has been cloned and characterized in many plants, including *Arabidopsis thaliana* (Shirley et al., 1992), rice (Druka et al., 2003), *Petunia hybrida* (Van Tunen et al., 1988), soybean (Ralston et al., 2005), and alfalfa (McKhann and Hirsch, 1994). Based on sequence similarity and function, CHI-fold proteins were classified into four subfamilies: type I, type II, type III, and type IV (Zamora et al., 2013; Zu et al., 2019; Lin et al., 2021; Wang et al., 2022b). Studies on CHI-fold proteins in *Arabidopsis* further categorized the type III CHI subfamily into three subclasses, AtFAP1, AtFAP2, and AtFAP3 (Ngaki et al., 2012). Recent studies have also categorized type II CHI into two subgroups, CHIA and CHIB, based on their differentiation in the function of symbiotic nitrogen fixation in rhizomes (Liu et al., 2024).

Together, type I and type II CHIs are referred to as *bona fide* CHIs or active CHIs, catalyzes the cyclization of 4,2',4',6'-tetrahydroxychalcone (also known as chalcone) to form naringenin. Although this process is capable of spontaneous reaction, the reaction catalyzed by CHI is 10^7 -fold more efficient and produces specific 2S-stereoisomers (important precursors for

downstream hydroxylation), which lead to the synthesis of various flavonoids and isoflavonoids (Jez et al., 2000). In addition, type II CHIs have acquired a new function over type I, due to a substitution of key catalytic residues, enabling the additional cyclization of isoliquiritigenin to 5-deoxy flavanone, i.e., (2S)-liquiritigenin (Shimada et al., 2003; Dastmalchi and Dhaubhadel, 2015). Previous research has shown that the presence of multiple catalytic residues within the CHI protein, along with the hydrogen bonding interactions between the protein and its substrate, is crucial for its catalytic activity (Jez et al., 2000, 2002). In contrast, types III and IV are not catalytically active because they lack one or more key active sites (Dastmalchi and Dhaubhadel, 2015). Although type IV CHI (also known as CHI-like) does not have chalcone isomerase activity, unlike type III CHI, it still plays a crucial role in the flavonoid biosynthetic pathway (Sugimoto et al., 2024). For example, in *Humulus lupulus*, type IV CHI increased the production of 2',4,4',6'-tetrahydroxychalcone and reduced the formation of by-products by binding to CHS (Ban et al., 2018). This function has since been shown to be widespread across plants (Waki et al., 2020).

The essential functions of CHI-fold proteins and flavonoid compounds have led to their retention during plant adaptive evolution. Type I CHI is present in almost all vascular plants. Type IV CHI is predominantly identified in terrestrial plants. In contrast, type III CHI exhibits a broad distribution across the plant kingdom and is also present in homologous structures within certain fungi and bacteria (Gensheimer and Mushegian, 2004). However, the origin and distribution of type II CHI have been subjects of considerable debate. Initial investigations posited that type II CHI was confined to leguminous plants (Shimada et al., 2003); subsequent research, however, has demonstrated its occurrence in ferns and even in bryophyte lineages (Cheng et al., 2018; Ni et al., 2020). This complex distribution pattern, along with the origin and evolution of early CHI-fold proteins, remains an unresolved puzzle. Subsequent studies, including the analysis of the higher structure of type III CHI members in *Arabidopsis*, combined with phylogenetic analyses, have revealed that CHI-fold proteins likely originated from fatty acid-binding proteins (FAPs) (Ngaki et al., 2012). Since then, the evolution of CHI-fold proteins has been widely studied. Research on the type CHI I and II genes in 52 fern species suggests that the emergence of type I CHI may coincide with the divergence of ferns (Ni et al., 2020). Additionally, studies on CHI-fold proteins in 15 species also indicate that type III CHI is the common ancestor of the other CHI types, which evolved from the common ancestor FAP3 (Wang et al., 2022b). This evolutionary path, in green plants, has undergone continual structural evolution and gene duplication under natural selection pressures.

Despite substantial research on the evolution and diversity of CHI-fold proteins in certain model plants and specific plant groups, a comprehensive exploration of their evolutionary history and diversity across the entire plant kingdom remains incomplete. This study aims to elucidate the origin and evolutionary trajectory of the CHI-fold proteins in plants by conducting an extensive investigation into the distribution of CHI-fold protein members across various lineages. It also examines the mechanisms

driving gene duplication and the evolutionary relationships, utilizing a wide array of plant and algal genomic data. Furthermore, AlphaFold was utilized to predict the tertiary structures of CHI-fold protein members across diverse lineages, further enhancing our understanding of the structural variations. This research enhances our comprehension of the evolution and diversity of the CHI-fold proteins and establishes a theoretical foundation for further investigation into the functions of the CHI-fold proteins, as well as the identification of novel functional CHI genes.

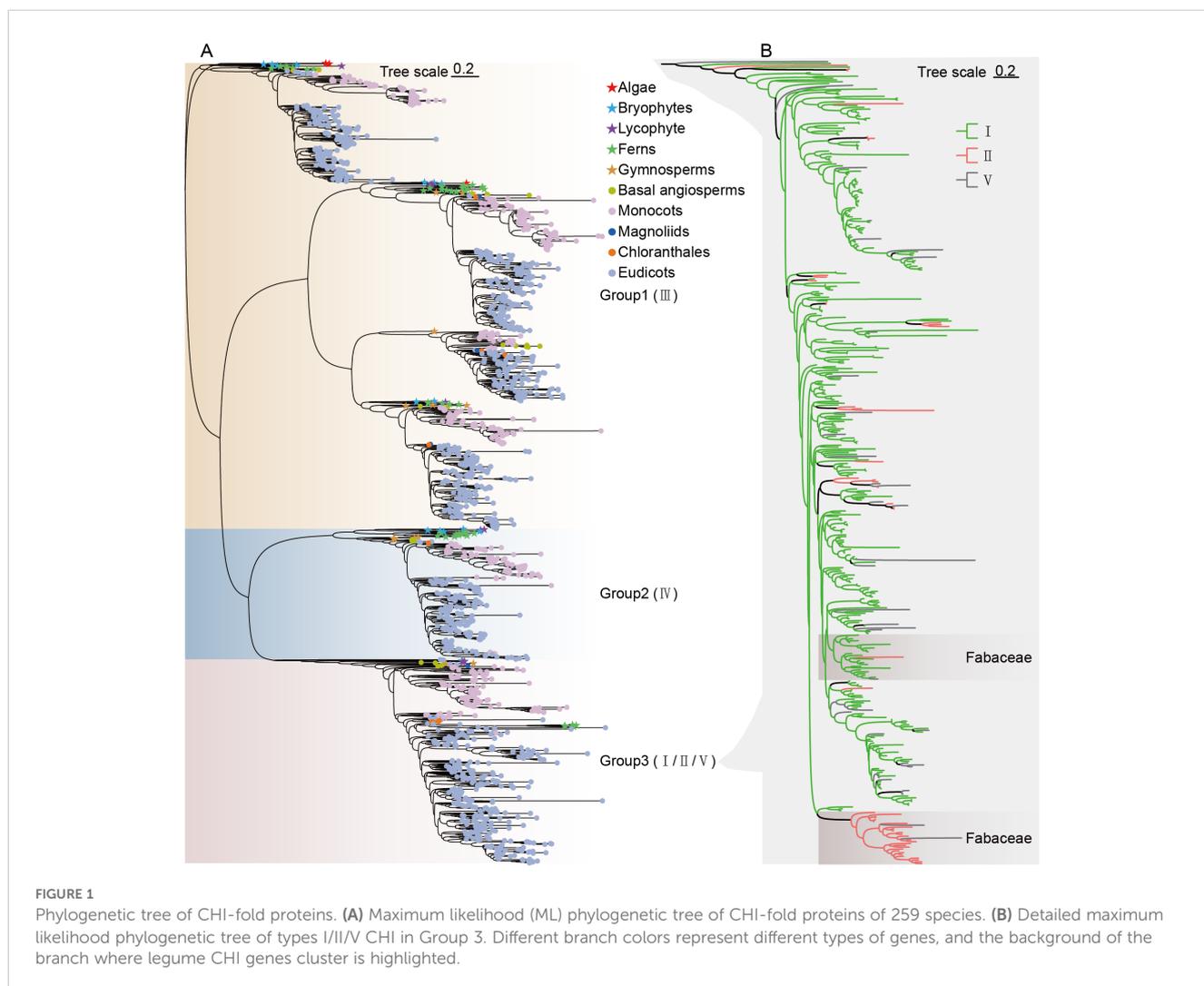
Results

Identification and phylogenetic analysis of CHI-fold protein families in plants

To systematically investigate the CHI gene families in plants, Hidden Markov Model (HMM) methodologies were employed to identify CHI protein sequences from 259 species genomes that cover different plant lineages (Supplementary Table S1). In total,

1,738 CHI sequences were identified (Supplementary Table S2). CHIs were present in all species investigated, ranging from algae to angiosperms. To investigate the evolutionary relationships, a maximum likelihood (ML) phylogenetic tree was constructed using IQ-TREE, which classified 1,738 CHI sequences into three groups (Figure 1A; Supplementary Figure S2). Group 1 consists predominantly of type III CHIs (1016), Group 2 comprises type IV (282), and Group 3 includes a heterogeneous mix of CHI types. Notably, type I and type II CHIs did not fully resolve into separate branches in Group 3. Previous studies have indicated that the types I and II CHIs can be effectively distinguished by analyzing the amino acids at the active site, particularly at position 190 (Dastmalchi and Dhaubhadel, 2015; Ni et al., 2020). Specifically, a threonine (T) residue at position 190 is characteristic of type II CHI (56), while a serine (S) residue indicates type I CHI (347). In the sequences of Group 3, we observed other amino acid residues besides threonine and serine at position 190, which will be classified as type V CHI (37) (Figure 1B; Supplementary Figures S3–S5).

Group 1, encompassing type III CHIs, is widely distributed across a wide range of plant lineages including algae, bryophytes,



lycophytes and ferns, and spermatophytes. This extensive phylogenetic distribution implies that type III CHI genes may represent the earliest evolutionary form within the CHI family, having been extensively conserved throughout subsequent plant evolutionary processes. Group 2 is the type IV CHIs, which have been primarily retained in land plants, and have undergone radiation and expansion in the angiosperm lineage, highlighting their potential role in early terrestrial adaptation of plant lineages. The distinction between types I, II, and V CHIs within Group 3 showed extensive divergence, especially for type II CHIs, which are notably enriched in legumes (Figure 1B).

Additionally, gene characteristics such as length, isoelectric point, molecular weight, and subcellular localization were analyzed for each CHI type (Supplementary Tables S2, S3). The mean value of the isoelectric point of gene members was 7.04 (neutral), of molecular weight was 30,160 Da, and of gene length was 275 amino acids (aa). Notably, type III CHIs are significantly longer, averaging 309 amino acids, while the other types exhibit minimal length variation. Subcellular localization analysis revealed that types I, II, and IV CHIs are predominantly cytoplasmic (cyto), while types III and V are mostly chloroplastic (chlo).

Distribution and evolution of CHI-fold protein families in plants

To explore the origin, distribution, and evolutionary history of the CHI-fold proteins in plants, we constructed a phylogenetic tree for 259 species (Supplementary Figure S1). Using the BUSCO dataset, we identified 196 single-copy homologous genes with more than 80% coverage. ASTRAL merging results indicated that quartet trees from 85% of the gene trees appeared in the final species tree. The phylogenetic tree reconstructed in this study is highly similar to the recently published tree of life (Zuntini et al., 2024). Subsequently, we assessed the number of CHI gene copies (Contains types I/II/III/IV/V) in each species and calculated the proportion of each CHI type (Figure 2). Algae (Charophyceae, Florideophyceae, Bangiophyceae, and Mamiellophyceae) contain 1–3 copies of CHI genes, while bryophytes, including Bryopsida and Marchantiopsida, have an average of 5 copies, suggesting CHI gene expansion may be related to terrestrial adaptation (such as increased UV exposure, desiccation, and pathogen resistance). Among lycophytes (Lycopodiopsida) and ferns (Polypodiopsida), the average number of CHI gene copies further increases to 7.71, and *Marsilea vestita* possesses up to 10 copies. Moreover, most species possess at least one copy of the type I or type II CHI gene. Considering that the synthesis of naringenin from 4,2',4',6'-tetrahydroxychalcone occurs much less efficiently when the process occurs spontaneously than under the action of types I/II CHI. Lycophytes and ferns may have benefitted from elevated CHI gene copy numbers and the evolution of active CHI to support more robust flavonoid biosynthesis, essential for vascular development and stress adaptation. Gymnosperms display relatively lower CHI copy numbers (2–4), whereas angiosperms exhibit considerable variation, ranging from 1 to 25. This substantial variability likely

reflects their ecological diversity and adaptive requirements for specialized flavonoid functions.

We classified members of the CHI-fold proteins into five types, I to V, based on phylogenetic analyses and amino acid residue analyses. These types have emerged progressively across diverse plant lineages, illustrating their functional diversification and emphasizing the crucial role of flavonoid metabolism in plant adaptation to terrestrial habitats (Figure 2; Supplementary Figure S6). In algae, only the non-catalytic type III CHI gene is present, implying that type III may represent the ancestral form of the CHI-fold proteins in plants. During the process of plant terrestrialization, other types of CHIs gradually evolved. Types II and IV CHI genes were first identified in basal bryophyte liverwort (*M. polymorpha*), where their presence likely represents metabolic adaptations necessary for coping with environmental stresses associated with terrestrial habitats, such as ultraviolet radiation and desiccation. Type IV CHI is nearly ubiquitous across land plants, indicating its essential role in plant evolution. However, type II CHI displays a more sporadic distribution, being present only in ferns, bryophytes, legumes, and a limited number of core angiosperms, while absent in gymnosperms and basal angiosperms, suggesting it has lineage-specific retention. Type I CHI, the primary enzyme catalyzing chalcone isomerization in plants, was identified in ferns, and then widely conserved across vascular plants, further underscoring the importance of flavonoid pathways in plant adaptive evolution. Type V CHIs, which are evolutionarily recent, appear sporadically in only a few core angiosperms. In summary, the CHI-fold proteins have experienced numerous expansions and diversifications throughout the evolutionary transition from aquatic to terrestrial habitats. Notably, bryophytes already exhibit a CHI-fold protein repertoire that closely resembles that of angiosperms.

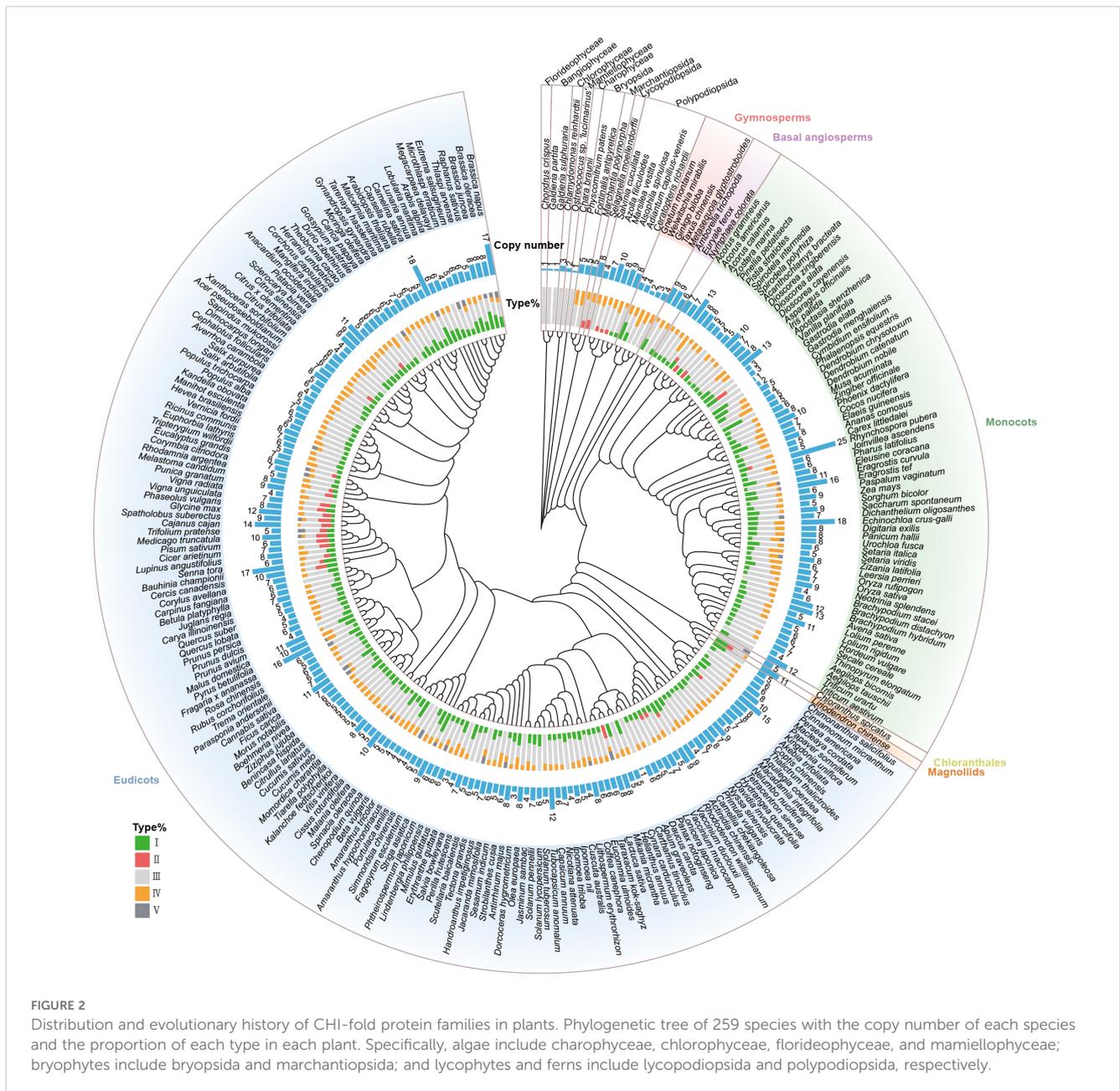
Gene structural characteristics of different CHI types in representative species

The conserved motifs of CHI-fold proteins were analyzed for 15 representative species using the MEME software, with the maximum number of conserved motifs set to 15 (Supplementary Table S4; Supplementary Figure S8). There were large variations in the motifs among the different types of CHI genes (Figure 3). Overall, there are three shared motifs (motif 3, motif 5, and motif 7) in all CHI genes, which originated in early algae and remain highly conserved across all plant lineages. The type III CHI subfamily exhibits two distinctly diverged clades, showing significant variability in conserved motifs. Group 2-1 may be an important intermediate transitional CHI type containing specific motifs (motif 4, motif 8, motif 10, motif 11), which are conserved in type IV, type II, and type I CHIs. These motifs were already present in algae. In contrast, another branch of type III (Group 1) is primarily characterized by motif 1 and motif 2. Motif 9 was exclusive to type III and absent in types I, II, and IV. The motif distribution of type I and II CHIs were identical, probably due to their high similarity in conserved domain sequences. Types I, II, and IV shared a conserved motif 13, which originated from bryophytes.

Type IV CHIs have an additional motif 12 or motif 15 upon the active CHI (types I and II). The motif 15 appeared initially in gymnosperms, whereas the loosed motifs at the corresponding position in the active CHI has not been replaced by other motifs. Motif 4, motif 8, and motif 10 appeared in some sequences of early algae and remain conserved in Group 2-1 of type III CHI, while completely absent in other branches. Interestingly, despite their distant evolutionary relationship, type II and type IV CHI in bryophytes exhibit highly similar motif structures to those found in vascular plants, including type I, type II, and type IV. This suggests that these structures may have originated from the last common ancestor of the two lineages.

Analysis of duplication types of CHI families

Gene duplication serves as the main driver of metabolic pathway diversification and gene neofunctionalization, and exploring gene duplication types could further elucidate the evolution of the CHI-fold proteins (Long et al., 2003). Duplication event types were identified at the whole-genome and gene family levels in 259 species using Dupgene_finder (Qiao et al., 2019). Five types of gene duplication events were identified, including whole-genome duplications (WGD), tandem duplications (TD), transposed duplications (TRD), proximal duplications (PD), and dispersed



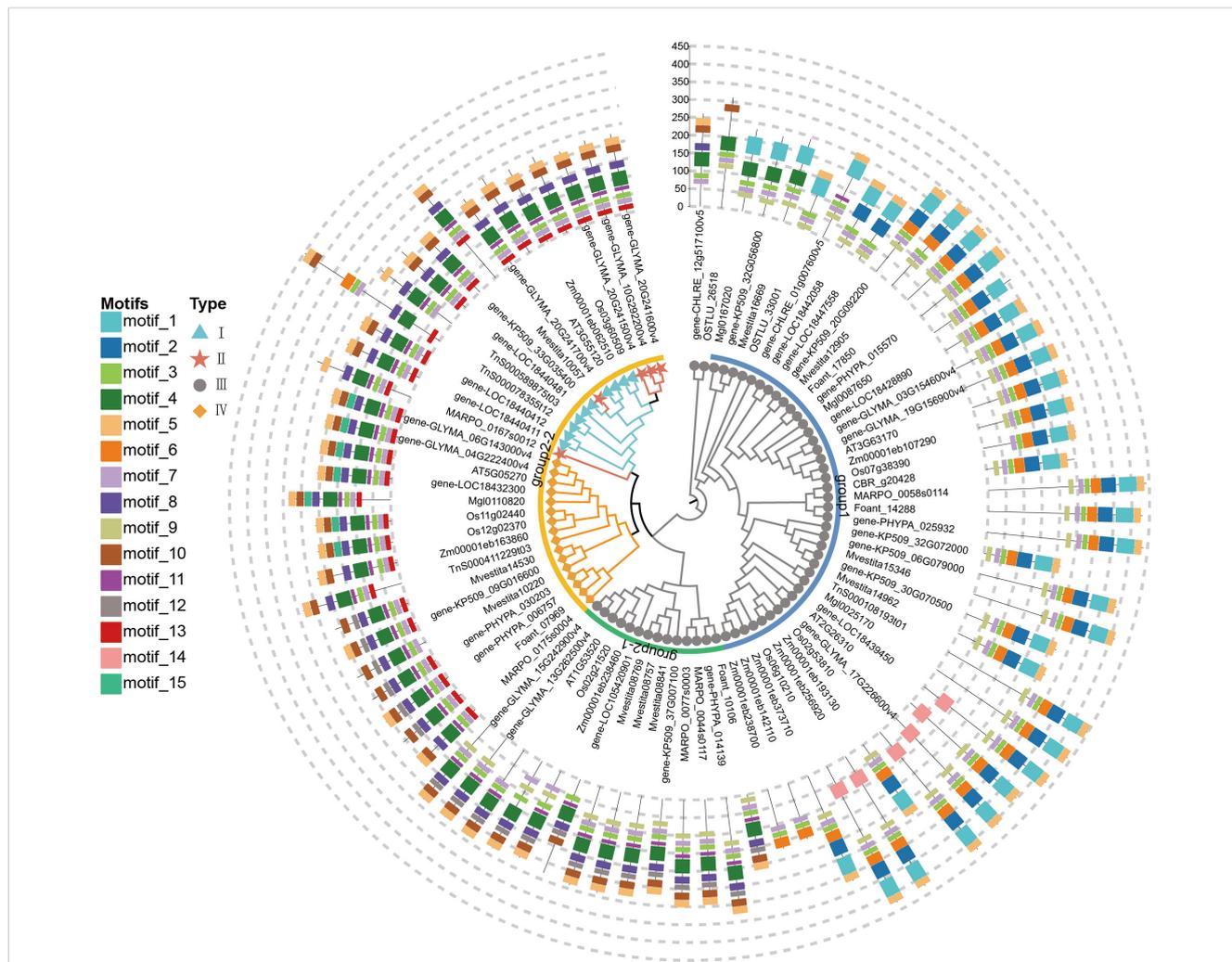


FIGURE 3 Conserved motif identification and distribution of representative species studied. The phylogenetic tree was obtained using the maximum likelihood method. Each gene's corresponding taxa are distinguished based on the prefix of its label name: two algal species (OSTLU and gene-CHILRE), three bryophyte species (MARPO, Foant, and gene-PHYPA), two fern species (Mvestita and gene-KP509), two gymnosperm species (Mgl and Tns), one basal angiosperm species (gene-LOC), two monocot species (Os and Zm), and two eudicot species (gene-GLYMA and AT).

duplications (DSD) (Supplementary Tables S5–S9; Supplementary Figure S9). The significance analysis was conducted on all identified CHI genes to explore differences in duplication types at the genome-wide level and their specificity across various plant lineages (Figure 4).

The analysis revealed that, among the five major duplication types, tandem duplication genes were significantly enriched in the CHI-fold proteins, exhibiting the highest percentage of enrichment (13.9%) across 36 species, and no significant reduction was observed (Figure 4A; Supplementary Table S6). No significant difference was detected in the remaining 223 species. The DSD was the second richest, significantly enriched in 34 species, 6 species significantly reduced, and 219 species without significant difference. The WGD, PD, and TRD were significantly enriched in 11, 21, and 24 species, respectively, significantly reduced in 9, 0, and 0 species, and not significantly different in 239, 238, and 235 species (Supplementary

Tables S5, S7, S8). Overall, species with no significant differences accounted for a high proportion of all examined types.

Statistical analyses showed that DSD was the main type in algae and bryophytes. Only whole genome duplication was detected in basal angiosperms. In ferns and gymnosperms PD and DSD dominated, with similar proportions. Monocots showed significant enrichment in all five duplication types, with WGD being the most frequent. In eudicot, the highest proportion of TD, followed by DSD, PD and TRD, while WGD had the lowest proportion (Figure 4B). In terms of significant reduction duplications were not found in algae, bryophytes, ferns, gymnosperms, and basal angiosperms. However, monocots and eudicots showed significant reductions in DSD and WGD (Figure 4C). In summary, DSD was the main duplication type of CHI gene expansion in algae and bryophytes. In ferns and gymnosperms, PD and DSD were the main types. In basal angiosperms and monocots WGD was the main force

driving the expansion of the CHI-fold proteins, whereas in eudicots TD was the main driver.

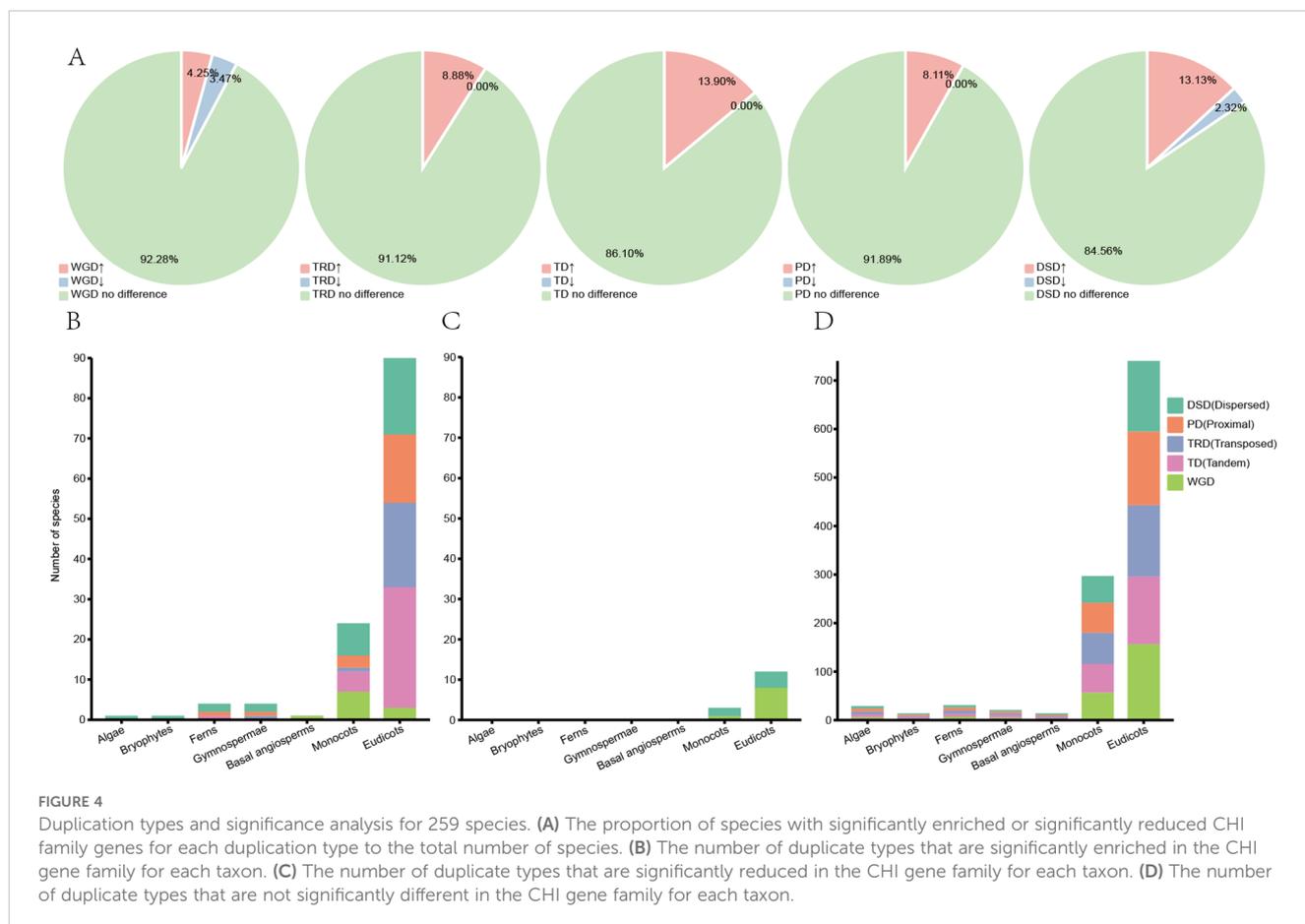
Identification of duplicated gene types and Ka/Ks calculations

The diversification of the chalcone isomerase-fold proteins has been significantly enhanced by gene duplication and subfunctionalization, particularly among eudicots, where whole-genome duplication (WGD) and tandem duplication have served as the primary driving forces. To explore the selection pressures driving CHI-fold protein expansion and diversification, we calculated the Ka and Ks values for CHI gene pairs derived from WGD and TD events using the `calculate_Ka_Ks_pipe.pl` programme of `DupGen_finder` (Supplementary Tables S10, S11). Species with no detectable gene pairs generated by WGD and TD were removed, and Ka/Ks values were calculated for the remaining 103 species duplicate gene pairs. The results showed that homozygous gene pairs with Ka/Ks values more than 1 accounted for 6.14% of the total gene pairs, while homozygous gene pairs with Ka/Ks values less than 1 accounted for 93.86% of the total gene pairs, suggesting that the CHI-fold protein was subjected to strongly purifying selection. However, in lycophyte ferns, basal angiosperm, monocots, and eudicots there were individual species, respectively, that had CHI

gene pairs with Ka/Ks values greater than 1. A group of type I gene pairs in *Camelina sativa* had a Ka/Ks value of type I gene pair of 4.3, suggesting that this gene pair may have experienced strong selection pressure in this species (Figure 5; Supplementary Table S10).

Structural diversity and evolutionary analysis of the CHI-fold proteins

To explore the structural diversity of CHI proteins, protein structure clustering was referred to the method of Caixia Gao et al (Huang et al., 2023). The three-dimensional structures of 51 protein sequences from eight representative species were predicted by AlphaFold3, with two fabp proteins containing the PF00061 domain selected as outgroups. Protein structure alignment was performed using US-align based on TM-value, and a normalized similarity matrix was generated. The UPGMA hierarchical clustering method was used to construct a dendrogram to display structural similarities. The result of clustering revealed that CHI proteins were classified into five groups, with differences in the structure of each group (Figure 6). The 3-layer sandwich structure was located in the central region of all CHI family members and showed high structural conservation. N-terminal structures of all CHI proteins displayed abundant diversity, whereas the C-terminal region was conserved.



Group 1 exhibited the most complex protein folding structure, present in all lineages and consisting entirely of type III CHI. Group 2 was found in all lineages except algae, and displayed a similar structure to Group 1. However, Group 2 was characterized by a significantly longer and irregularly coiled structure at the N-terminal end. Furthermore, the irregularly coiled region of Group 1 comprises two segments of α -helices. Group 3, Group 4, and Group 5 had significantly shorter N-terminal irregularly curled structures compared to Group 1 and Group 2, but still retained some segments. In particular, Group 2 contains a type II CHI gene (MARPO_0167s0012) from the bryophyte *M. polymorpha*, which exhibited an overall structural similarity to the other members within the group. This observation implies that early active CHI proteins may have preserved an N-terminal disordered region akin to that found in type III CHI, or alternatively, the mechanisms underlying the retention of the N-terminal structure of these genes exhibit lineage-specificity across different plant lineages. The type III CHI genes were classified into three groups, each demonstrating notable variations in their N-terminal regions. Among them two type III CHI genes from algae were identified in Group 1 and Group 3, respectively, indicating that the structural diversity of type III CHI genes had emerged during the algae.

Discussion

Chalcone isomerase is a key rate-limiting enzyme in the flavonoid biosynthesis pathway and is crucial for plant survival and adaptation (Yin et al., 2019; Wang et al., 2022b). Cross-lineage studies among gene families not only provide a more comprehensive understanding of gene family conservation and variability during evolution but also greatly reduce the possibility of biased conclusions due to the limitations of the results derived from studies on single or few species (Naake et al., 2021; Liu et al., 2024). In this study, a total of 1,738 CHI-fold proteins were identified across 259 species spanning the green plant lineage, and the basic information statistics showed that almost all of the CHI type III were longer than the other types, which was related to the irregularly curled structure at their N-terminal regions at the N-end (Figure 6). Subcellular localization analysis showed that types III and V are predominately localized in chloroplasts, consistent with their role in fatty acid synthesis (Ngaki et al., 2012). Whereas, types I, II, and IV CHIs are mainly localized in the cytoplasm, consistent with the fact that flavonoid biosynthesis takes place in the cytoplasm (Kitamura, 2006). 12 CHI genes were identified in *Glycine Max*, an important species for the study of the CHI family, consistent with the findings of Mehran DastMalchi et al (Dastmalchi and Dhaubhadel, 2015). where 9 and 5 CHIs were identified in maize and grape, respectively. The number of CHI family copies showed an increasing trend from lower to higher plants, especially during the plant terrestrialization, where the active CHI-fold protein members (types I and II) went from absent to abundantly expanded (Figure 2A; Supplementary Figure S1). Considering the vital role of stress resistance of flavonoid compounds in plants (Shomali et al., 2022). this might be closely

related to the significant biotic and abiotic stresses that plants faced during their evolution from aquatic to terrestrial habitats.

Early evolutionary studies of the CHI family member in *A. thaliana* demonstrated that non-catalytically active type III CHI are proteins involved in fatty acid binding and elucidated, in terms of three-dimensional structure, the adaptive evolutionary process by which the non-catalytically active CHI ancestor gradually evolved

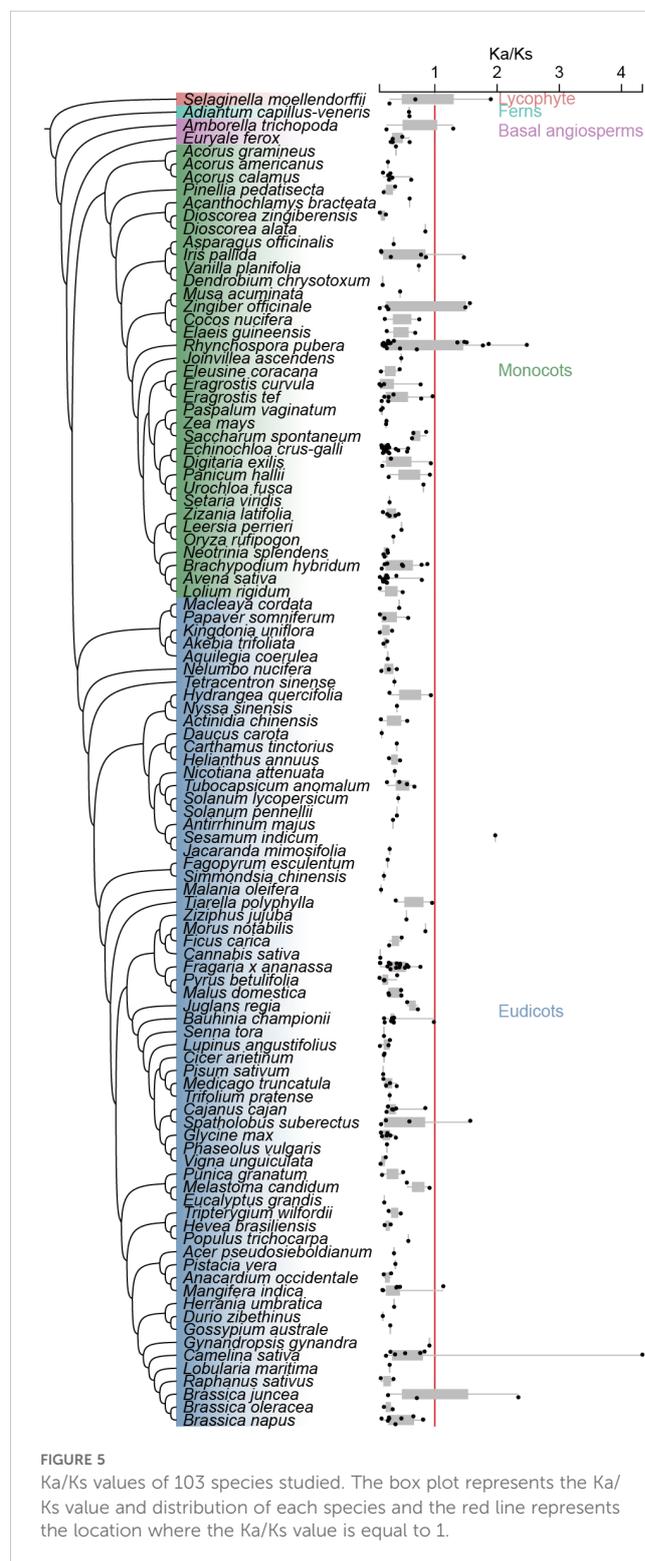
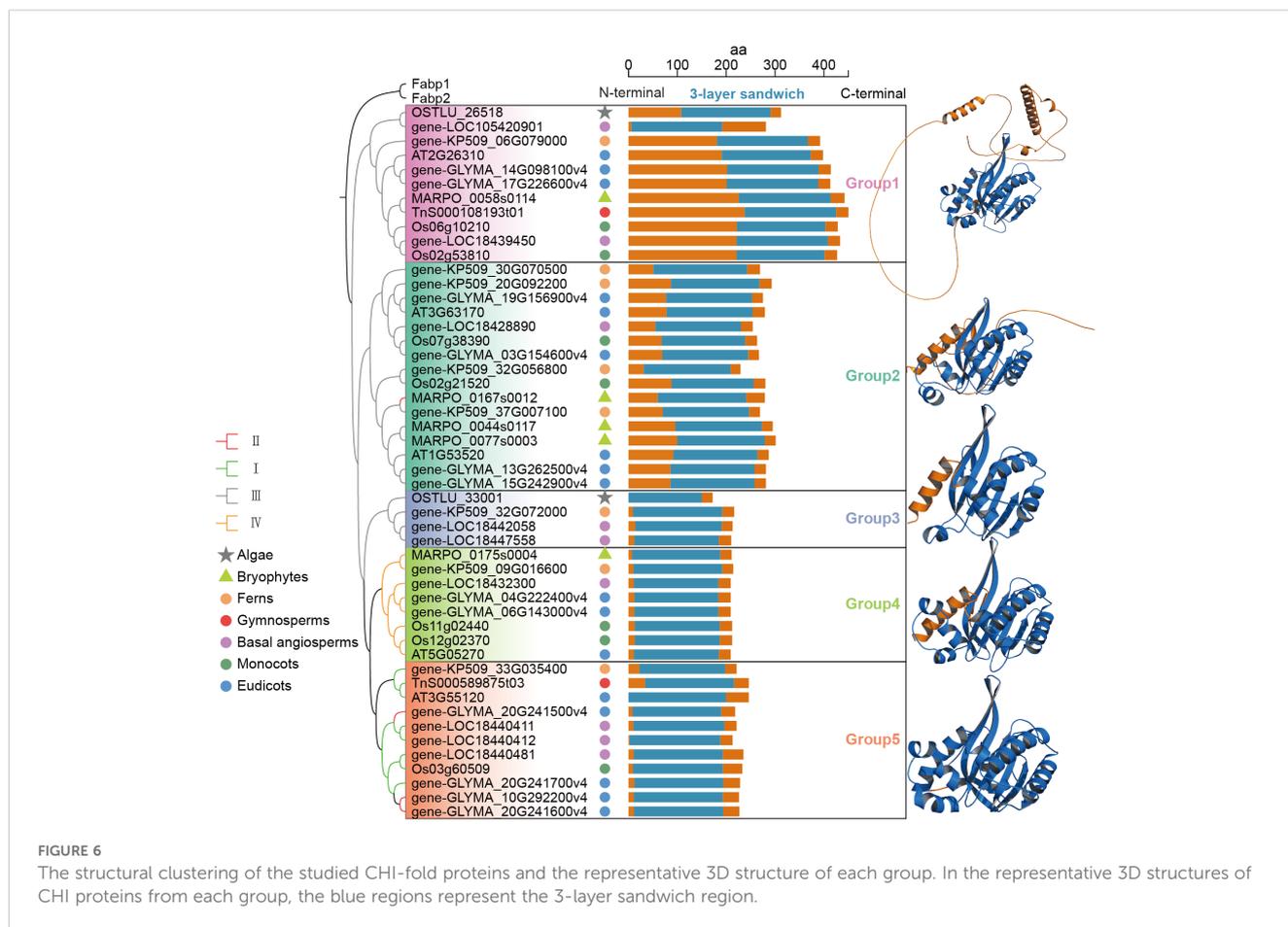


FIGURE 5
Ka/Ks values of 103 species studied. The box plot represents the Ka/Ks value and distribution of each species and the red line represents the location where the Ka/Ks value is equal to 1.



catalytically active through plant evolution (Ngaki et al., 2012). In this study, phylogenetic analyses of CHI family members from cross-lineage species (Figures 1, 2). Firstly, the large-scale identification of different lineages of CHI genes revealed that, although type II CHI was significantly enriched in legumes, it was also found in many other plants, initially in bryophytes (*M. polymorpha*), and then in vascular plants with repeated loss and reappearance. Therefore, type II CHI is not exclusive to leguminous plants (Wang et al., 2022b). This result disproves the theory that type II CHIs evolved with the emergence of leguminous plants (Shimada et al., 2003). Bryophytes are classified into liverworts, mosses, and hornworts (Wang et al., 2022c). Previous studies reported the presence of type IV CHI in the moss *Physcomitrella patens* (Wang et al., 2022b). This study further revealed that type IV CHI is also present in liverworts, particularly in the model species *M. polymorpha*. Type IV CHI, although not catalytically active, has been extensively shown to play an important role in the flavonoid biosynthesis pathway (Jiang et al., 2015; Waki et al., 2020; Sugimoto et al., 2024). Subsequently, all the genes of *M. polymorpha* were used to construct a phylogenetic tree (Supplementary Figure S7), to investigate the evolutionary relationship of CHI types II and IV. From the result, we can infer that in the bryophyte lineage, CHIs type II and type IV similarly evolved from type III CHI, consistent with those in Arabidopsis. It follows that, the evolutionary relationships between different types of CHI genes were the same as those in previous studies and that both type II and type IV CHIs evolved from type III CHIs (Kaltenbach

et al., 2018). Moreover, the types II and IV CHI genes of *M. polymorpha* have the matching conserved gene structure of the types I and IV CHI genes of Arabidopsis, each being distinct and characteristic (Figure 3; Supplementary Table S4). The extant bryophyte and angiosperm lineages are the same age, dating back to their last common ancestor (McDaniel, 2021). Thus, it is probable that these genes were present in the last common ancestor of both lineages and have been maintained since then during lineage-specific evolution (with modifications to type II or I from whichever was the progenitor gene in the last common ancestor).

Motifs are short sequences appearing in the gene family with certain functions or structures, which may be related to the identification of transcription factors and regulation of gene expression (Bailey et al., 2009). In this study, we identified conserved motifs (Figure 3; Supplementary Table S4) in 15 representative species, which illustrate the differences and similarities among different types of CHI family members. Notably, the conserved strategy of motif 4, motif 8, motif 10, and motif 11 during evolution seems to reveal that some of the essential structures of the active CHIs (types I and II) gene originated from ancient algae. However, the origins of motif 12, motif 13, and motif 15 are relatively recent. Motif 13 is present in land plants, motif 12 is found in the lineages of bryophytes and ferns, while motif 15 is only present in seed plants. More specifically, motif 12 first appeared concurrently in the type III CHI branch of Group 2-1 and the type IV CHI of Group 2-2. Based on the inference, that type IV evolved from type III, motif

12 may have originated from CHI type III of Early land plants and evolved to appear in CHI type IV in a short period (Kaltenbach et al., 2018; Liu et al., 2024). Motif 15 originated in basal angiosperms and was accompanied by the loss of motif 12 in bryophytes and fern plants at the same position of gene structure. Thus, this could be the result of diversification of the auxiliary functions among CHI type IV due to severe selective pressure in angiosperms (Wu et al., 2020). E.g. in *A. thaliana* CHI type IV acts as a unique enhancer in synergistic collaboration with TT5 (a CHI type I) to promote flavonoid biosynthesis, which can also promote chalcone synthase and membrane-bound terpene transferase (PT1L) activity in *H. lupulus*, as well as a general role in plants that bind to chalcone synthase to reduce byproduct formation thereby correcting CHS non-specificity (Jiang et al., 2015; Waki et al., 2020). Interestingly, the motif structures of active CHI (types I and II) and type IV CHI genes are similar, with the difference that at the same position, active CHI lacks motif12 or motif15, and this position is vacant (Figure 3). In conclusion, the loss of motif 12 or motif 15 and the appearance of motif 13 in land plants may have gradually provided the conditions for CHI to possess catalytic activity.

Gene family diversification is mostly driven by gene duplication, where subsequent mutations following gene duplication lead to the possibility of divergence or the emergence of neofunctionalization of genes (Lynch and Conery, 2000). In this study, we conducted a detailed analysis of the replication types that contribute to the expansion and functional diversification of the CHI-fold proteins, revealing significant variation across different taxa (Figure 5). In algae and bryophytes, CHI gene duplication is predominantly driven by dispersed repeats, while in ferns and gymnosperms, duplication is primarily mediated by both dispersed and proximal repeats. Basal angiosperms and monocots exhibit CHI gene expansion mainly driven by WGD, whereas tandem duplications play a dominant role in eudicots. The function of CHI genes plays an important role in plant adaptive evolution. Due to their relatively compact genomes and stable ecological niches, WGD are infrequent in algae and bryophytes, making dispersed repeats the primary mode of gene duplication (Lynch and Conery, 2000; Liu et al., 2024). Conversely, WGDs are prevalent throughout angiosperms, enabling rapid generation of extensive gene copies that facilitate adaptive evolution through natural selection (Lyons et al., 2008; Soltis et al., 2009; Renny-Byfield and Wendel, 2014; Soltis et al., 2014). The evolutionary path of true dicotyledons is notably complex and diverse, with tandem duplications serving as a strategic mechanism for swift adaptation to environmental changes, exemplified by the sub-functionalization of type II CHI, a pivotal factor in their evolutionary process (Liu et al., 2024). This observation elucidates the mechanism behind CHI gene expansion in these plants.

For those sequences with low similarity that diverge significantly over evolutionary time, structural clustering offers a better approach for representing gene conservation and variability. With the assistance of accurate prediction of protein structure by artificial intelligence, we can study the relationship between the variability of the 3D folding structure of genes and their functions (Huang et al., 2023). Exploring the relationship between gene family structure and function based on structural clustering is an

important tool for probing the evolution of gene families (Esposito et al., 2021; Duan et al., 2023; Liao et al., 2023). The results of this study show that the overall structure of CHI is diverse, with this diversity mainly arising from the non-conserved N-terminal region, and most notably observed in type III CHI. Moreover, structural clustering analysis showed that all active CHIs, except the MARPO_0167s0012 gene from bryophytes, retained minimal sequences at their N-terminal. Previous evolutionary studies suggest that catalytically active CHI proteins may have evolved from non-catalytic ancestors (Kaltenbach et al., 2018). Thus, it is probable that after diverging from the last common ancestor, the N-terminus of type II CHI evolved lineage specificity. The type II CHI in the bryophyte lineage retains the ancestral N-terminal structural features, whereas, in the vascular plant lineage, it gradually evolved to become shorter. The unique folding structure, 3-layer sandwich, of the CHI family is the basis for its catalytic activity (Jez et al., 2000). We found that such structure is present in all CHI genes including the ancient algae CHI, and is almost exclusively located in the center of the sequences (Figure 6). It is evident that despite non-conserved terminal regions, the critical role of CHI genes in plant adaptation underlines the strong selective pressure for conserving the 3-layer sandwich fold.

This study classified chalcone isomerases (CHIs) into five types (I–V) based on phylogenetic analysis of the gene family and screening of active sites. Type I and type II CHIs are the representative functional enzymes in the flavonoid biosynthesis pathway, both catalyzing the isomerization of naringenin chalcone, resulting in the formation of (2S)-naringenin (Jez et al., 2000; Yin et al., 2019). However, type II CHI possesses an additional function—it can also catalyze the conversion of isoliquiritigenin to (2S)-liquiritigenin, which serves as a key distinguishing feature between type I and II CHIs (Shimada et al., 2003; Cheng et al., 2018). Type III CHI does not participate in flavonoid biosynthesis but has been confirmed to be involved in fatty acid synthesis (Ngaki et al., 2012). Knockout of type III CHI in *Arabidopsis thaliana* disrupts fatty acid metabolism, leading to reproductive defects (Ngaki et al., 2012). Notably, in flavonoid biosynthesis, type IV CHI interacts with chalcone synthase (CHS) via protein-protein interactions, enhancing THC production while reducing CTAL formation (Waki et al., 2020; Wolf-Saxon et al., 2023). Researchers typically further validate CHIL based on this biological function (Xu et al., 2022; Lewis et al., 2024; Zhang et al., 2024). Due to the lack of key active sites, type V CHI is presumed to lack chalcone isomerase activity. In summary, the biological functions of different types of CHI genes exhibit significant differences, which may provide some valuable reference for the future screening of functional CHI genes.

Materials and methods

Genomic data file collection and processing

Genome files (.fa) and annotation files (gff/gff3) of 259 species were obtained from JGI (<https://phytozome-next.jgi.doe.gov/>)

(Goodstein et al., 2012), Ensemble plant (<https://plants.ensembl.org/>), NCBI (<https://www.ncbi.nlm.nih.gov/datasets/genome/>) (Bolser et al., 2016), NGDC (<https://download.cncb.ac.cn/gwh/Plants/>), and Published Plant Genomes databases (https://www.plabipd.de/plant_genomes_pa.ep), and species were classified according to the NCBI classification system, which was quickly completed with the help of the R package taxize (Federhen, 2012; Chamberlain and Szöcs, 2013). Differential splicings were excluded, and only the longest transcripts of each gene were retained for subsequent analyses.

Phylogenetic analysis of 259 species

Species phylogenetic trees were reconstructed using the coalescence method by first obtaining single-copy orthologous gene ensembles for each species from the benchmarking universal single-copy ortholog BUSCO gene sets (Simão et al., 2015), with the parameters “-m prot -l viridiplantae_odb10”. Then, single-copy orthologous genes with over 80% coverage among 259 species were obtained, and multiple sequence alignments for each of these single-copy orthologous genes were performed using MAFFT v7.429 software (Katoh and Standley, 2013), with the parameters “-maxiterate 1000 -localpair”. The results of multiple sequence alignments were adjusted using trimAl v1.5 software (Capella-Gutiérrez et al., 2009), with the parameter “-gt 0.5” to limit the number of gaps in the results. The phylogenetic trees were constructed for each adjusted multiple sequence alignments result using the IQ-TREE v1.6.11 tool of the ML (maximum likelihood) method (Nguyen et al., 2015), with the parameters “-bb 1000 -m TEST -nt AUTO”. Finally, the astral-tree-5.7.8-1 software was used to construct the species phylogenetic tree based on the multi-species coalescent model, which infers species phylogenetic trees from an extensive collection of gene phylogenies (Zhang et al., 2018), with the command “java -jar astral.5.7.8.jar -i input file -o output file”. The Timetree fossil time tree serves as a reference to ascertain the root position of the phylogenetic tree (Kumar et al., 2017). The figure’s beautification was made online using the Chiplot website (<https://www.chiplot.online/>) (Xie et al., 2023).

CHI-fold proteins identification

First, a protein sequence library of the longest protein coding sequence of each species was constructed using the default parameters of the makeblastdb module of the ncbi -blast -2.9.0+ software (Johnson et al., 2008). The amino acid sequences of the seven CHI-fold protein members that have been identified and characterized were used as queries for monocots (Park et al., 2021), while the identified CHI-fold protein members in *A. thaliana* downloaded from the TAIR database (<https://www.arabidopsis.org/>) were used for other species. The ncbi -blast -2.9.0+ software’s blastp module, with an e-value of 10^{-5} ,

was employed to search all protein libraries, and the search results were used as candidate gene family members. The seed alignment files for the CHI domain (PF021431, PF16036, PF16035), obtained from pfam database (<https://www.ebi.ac.uk/interpro/entry/pfam/>), were used to build an HMM file using the HMMER3 (v3.3) software package (Mistry et al., 2013). Subsequently, genes lacking CHI domains were excluded based on annotations from the Conserved Domain Database (<https://www.ncbi.nlm.nih.gov/cdd/>) and SMART (<https://smart.embl.de/>) (Marchler-Bauer et al., 2010; Letunic et al., 2021). Redundant genes were then eliminated using seqkit rmdup with parameter “-s” and finally obtained CHI-fold protein members (Shen et al., 2016). The information and sequences of gene length were extracted by script bash, information of isoelectric point and molecular was searched and obtained online by ExPASy (https://web.expasy.org/compute_pi/), and subcellular localization was predicted by WoLF PSORT (<https://wolfsort.hgc.jp/>) website.

Phylogenetic analysis of CHI-fold protein members

Following the integration of all identified members into a gene set, high-throughput rapid multiple sequence alignments were performed using the MAFFT v7.429 software with default parameters, and trimAl v1.5 software was used to adjust the result. Phylogenetic trees were constructed using the IQ-TREE v1.6.11 tool with the maximum likelihood (ML) method for each adjusted result. The appropriate models were automatically selected using the parameter “-TEST”, and the Bootstrap values with parameter “-bb 1000” to assess phylogenetic tree reliability (Horowitz, 2001). The phylogenetic tree was classified by the similarity of clustering relationship among genes and sequences of chalcone isomerase-fold proteins family members, whose functions had been verified, and then further classified using amino acid residues recognizing types I and II identities (Dastmalchi and Dhaubhadel, 2015; Ni et al., 2020). All the CHIs belonging to Group 3 were classified by a custom Python script for each type of gene, following a detailed sequential alignment performed by MAFFT v7.429 software with parameter “-maxiterate 1000 -localpair”.

Conserved motif identification and analysis of the representative species

Members of the CHI-fold proteins from 15 representative species were analyzed using the MEME Suite online platform (<https://meme-suite.org/meme/tools/meme>) for motif prediction, with parameters “the maximum number of motifs is set to 15”, “motif length set to 6 to 200aa”, and “the maximum width set to 50” (Bailey et al., 2009). The logo plots of the motif were obtained directly from the program running file. The phylogenetic tree is done through Chiplot (Xie et al., 2023).

Identification and analysis of the duplication types of CHI gene families

Initially, the whole-genome duplication genes of 259 species were identified using the DupGen_finder-unique.pl program, a component of the DupGen_finder tool, with the default parameter (Qiao et al., 2019). Setting a reference species for each taxon, including *Chlamydomonas reinhardtii* for with algae, *M. polymorpha* for with bryophyte, *Selaginella moellendorffii* for with ferns, *Gnetum montanum* for with Gymnosperms, and the angiosperms (*Amborella trichopoda* for with basal angiosperm, *Acorus gramineus* for with monocotyledon, *Vitis vinifera* for with dicotyledon, and two species of the orders *Chloranthales* and *Magnoliids* were referenced for with each other). Then, statistics on the type and number of replicates to which members of the CHI-fold proteins belonged were obtained through a Python custom script. The Ka and Ks values of duplication gene pairs of the CHI-fold proteins were obtained through the calculate_Ka_Ks_pipe.pl program of the DupGen_finder tool with default parameter. The chi-square independence test was calculated using the chi2_contingency function in the scipy library of Python, with a *P* value of <0.05 being considered as a significant difference, and the significance of the result was determined by the magnitude of the ratios of the number of replicated types to the number of genome-wide genes at the genome-wide level (Wang et al., 2022a). As the taxa *Chloranthales* and *Magnoliales* are represented by only one species, which is not considered representative, they are not shown in Figures 4B–D.

Analysis of structural diversity of CHI-fold proteins

The protein folding structures of the representative species in CHI-fold protein members were predicted using the default setting of AlphaFold 3 (Abramson et al., 2024). All predictions with the highest-ranking score were selected for further analyses, and proteins with low confidence were screened using a ranking score ≥ 0.7 as a threshold. Protein structure clustering was referred to the method of Caixia Gao et al, where all predicted proteins were first annotated by InterPro to ensure the presence of CHI family structural domains (IPR016087) (Jones et al., 2014; Huang et al., 2023). Structural alignment was performed using the TM-score method of the US-align tool (Zhang et al., 2022). With parameters “-a 1 -outfmt 1”. followed by the construction of a whole structural similarity matrix based on the TM values and processed using the min-max scaling. The matrix with structural similarity was clustered by the Unweighted Pair-Group Method with Arithmetic Means (UPGMA) with the Bray-Curtis dissimilarity index as the basis for clustering, and the entire clustering process was done using the vegan and phangorn packages of the R language (Ihaka and

Gentleman, 1996). Structural visualization of proteins was used PyMOL v4.6.0 (DeLano, 2002).

Conclusion

This study examined 1,738 members of chalcone isomerase (CHI) across 259 species spanning the green plant lineages, revealing an increase in the number of gene copies during the evolutionary transition from aquatic to terrestrial species. The phylogenetic distribution of CHI genes throughout the plant kingdom indicates that the origin of type II CHI can be traced back to the last common ancestor shared by bryophytes and vascular plants. We discovered a gene set in the active CHI group with 190 amino acid residues different from those of type I and II CHIs, which were categorized as type V CHIs. The study explored the primary types of gene duplication driving CHI-fold protein expansion across different groups. It was found that dispersed duplication was the main driver in early algae and bryophytes, while ferns and gymnosperms primarily exhibited dispersed and proximal duplications. Basal angiosperms and monocots showed expansions mainly driven byWGD, whereas tandem duplication was predominant in eudicots. The findings also confirmed that CHI genes in plants are under strong purifying selection. The structural diversity of CHI family proteins is mainly attributed to the non-conserved nature of the N-terminal regions, which contribute to variations in the CHI-fold structure, most notably observed in type III CHI. This study enhances the understanding of the evolutionary patterns and functional diversification of the CHI-fold proteins, providing a theoretical basis for future investigations into their biological functions and potential applications.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

KL: Data curation, Investigation, Software, Visualization, Writing – original draft, Writing – review & editing. SW: Data curation, Methodology, Writing – review & editing. LY: Conceptualization, Investigation, Writing – review & editing. SL: Data curation, Writing – review & editing. JC: Data curation, Writing – review & editing. YD: Conceptualization, Writing – original draft, Writing – review & editing. YN: Conceptualization, Writing – review & editing. WW: Conceptualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Acknowledgments

The authors would like to thank the other members of the Yunnan Provincial Key Laboratory of Biological Big Data, particularly Shengchang Duan and Baozheng Chen for their maintenance of the supercomputer.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be constructed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

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

References

- Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630 (8016), 493–500. doi: 10.1038/s41586-024-07487-w
- Agati, G., Azzarello, E., Pollastri, S., and Tattini, M. (2012). Flavonoids as antioxidants in plants: location and functional significance. *Plant Sci.* 196, 67–76. doi: 10.1016/j.plantsci.2012.07.014
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335
- Ban, Z., Qin, H., Mitchell, A. J., Liu, B., Zhang, F., Weng, J.-K., et al. (2018). Nuncatalytic chalcone isomerase-fold proteins in *Humulus lupulus* are auxiliary components in prenylated flavonoid biosynthesis. *Proc. Natl. Acad. Sci.* 115, E5223–E5232. doi: 10.1073/pnas.1802223115
- Bolser, D., Staines, D. M., Pritchard, E., and Kersey, P. (2016). Ensembl plants: integrating tools for visualizing, mining, and analyzing plant genomics data. *Plant bioinform.: Methods Protoc.* 1374, 115–140. doi: 10.1007/978-1-4939-3167-5_6
- Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348
- Chamberlain, S. A., and Szöcs, E. (2013). taxize: taxonomic search and retrieval in R. *F1000Research* 2, 191. doi: 10.12688/f1000research
- Cheng, A. X., Zhang, X., Han, X. J., Zhang, Y. Y., Gao, S., Liu, C. J., et al. (2018). Identification of chalcone isomerase in the basal land plants reveals an ancient evolution of enzymatic cyclization activity for synthesis of flavonoids. *New Phytol.* 217, 909–924. doi: 10.1111/nph.2018.217.issue-2
- Dastmalchi, M., and Dhaubhadel, S. (2015). Soybean chalcone isomerase: evolution of the fold, and the differential expression and localization of the gene family. *Planta* 241, 507–523. doi: 10.1007/s00425-014-2200-5
- DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* 40, 82–92. Available online at: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=ab82608e9a44c17b60d7f908565fa628295dc72#page=44>.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2025.1559547/full#supplementary-material>

SUPPLEMENTARY TABLE 1

The genomic data source of 259 species.

SUPPLEMENTARY TABLE 2

The basic statistics of CHI-fold protein family members.

SUPPLEMENTARY TABLE 3

Statistics of subcellular localization results of different types of CHI-fold protein family.

SUPPLEMENTARY TABLE 4

The motif analysis of different CHI types in representative species.

SUPPLEMENTARY TABLE 5

The statistics and significance analysis of the number of CHI WGD duplication (WGD) genes in 259 species with P-value<0.05.

SUPPLEMENTARY TABLE 6

The statistics and significance analysis of the number of CHI tandem duplication (TD) genes in 259 species with P-value<0.05.

SUPPLEMENTARY TABLE 7

The statistics and significance analysis of the number of CHI proximal duplication (PD) genes in 259 species with P-value<0.05.

SUPPLEMENTARY TABLE 8

The statistics and significance analysis of the number of CHI transposed duplication (TRD) genes in 259 species with P-value<0.05.

SUPPLEMENTARY TABLE 9

The statistics and significance analysis of the number of CHI dispersed duplication (DSD) genes in 259 species with P-value<0.05.

SUPPLEMENTARY TABLE 10

The WGD duplicate gene pairs and Ka/Ks values in the CHI-fold protein family.

SUPPLEMENTARY TABLE 11

The tandem duplicate gene pairs and Ka/Ks values in the CHI-fold protein family.

- Dias, M. C., Pinto, D. C., and Silva, A. M. (2021). Plant flavonoids: Chemical characteristics and biological activity. *Molecules* 26, 5377. doi: 10.3390/molecules26175377
- Druka, A., Kudrna, D., Rostoks, N., Bruggeman, R., von Wettstein, D., and Kleinohs, A. (2003). Chalcone isomerase gene from rice (*Oryza sativa*) and barley (*Hordeum vulgare*): physical, genetic and mutation mapping. *Gene* 302, 171–178. doi: 10.1016/S0378-1119(02)01105-8
- Duan, Y., Tang, H., and Yu, X. (2023). Phylogenetic and AlphaFold predicted structure analyses provide insights for A1 aspartic protease family classification in *Arabidopsis*. *Front. Plant Sci.* 14, 1072168. doi: 10.3389/fpls.2023.1072168
- Esposito, L., Balasco, N., Smaldone, G., Berisio, R., Ruggiero, A., and Vitagliano, L. (2021). AlphaFold-Predicted structures of KCTD proteins unravel previously undetected relationships among the members of the family. *Biomolecules* 11, 1862. doi: 10.3390/biom11121862
- Federhen, S. (2012). The NCBI taxonomy database. *Nucleic Acids Res.* 40, D136–D143. doi: 10.1093/nar/gkr1178
- Gensheimer, M., and Mushegian, A. (2004). Chalcone isomerase family and fold: no longer unique to plants. *Protein Sci.* 13, 540–544. doi: 10.1110/ps.03395404
- Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Horowitz, J. L. (2001). "The Bootstrap," in *Handbook of Econometrics*, eds. J.J. Heckman and E. Leamer (Amsterdam: Elsevier), 3159–3228. doi: 10.1016/S1573-4412(01)05005-X
- Huang, J., Lin, Q., Fei, H., He, Z., Xu, H., Li, Y., et al. (2023). Discovery of deaminase functions by structure-based protein clustering. *Cell* 186, 3182–3195. e3114. doi: 10.1016/j.cell.2023.05.041
- Ihaka, R., and Gentleman, R. (1996). R: a language for data analysis and graphics. *J. Comput. Graph. Stat.* 5, 299–314. doi: 10.1080/10618600.1996.10474713
- Jez, J. M., Bowman, M. E., Dixon, R. A., and Noel, J. P. (2000). Structure and mechanism of the evolutionarily unique plant enzyme chalcone isomerase. *Nat. Struct. Biol.* 7, 786–791. doi: 10.1038/79025
- Jez, J. M., Bowman, M. E., and Noel, J. P. (2002). Role of hydrogen bonds in the reaction mechanism of chalcone isomerase. *Biochemistry* 41, 5168–5176. doi: 10.1021/bi0255266
- Jiang, W., Yin, Q., Wu, R., Zheng, G., Liu, J., Dixon, R. A., et al. (2015). Role of a chalcone isomerase-like protein in flavonoid biosynthesis in *Arabidopsis thaliana*. *J. Exp. Bot.* 66, 7165–7179. doi: 10.1093/jxb/erv413
- Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res.* 36, W5–W9. doi: 10.1093/nar/gkn201
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Kaltenbach, M., Burke, J. R., Dindo, M., Pabis, A., Munsberg, F. S., Rabin, A., et al. (2018). Evolution of chalcone isomerase from a noncatalytic ancestor. *Nat. Chem. Biol.* 14, 548–555. doi: 10.1038/s41589-018-0042-3
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kitamura, S. (2006). "Transport of flavonoids: from cytosolic synthesis to vacuolar accumulation," in *The science of flavonoids*, ed. E. Grotewold. (New York, NY: Springer New York), 123–146.
- Kumar, S., Stecher, G., Suleski, M., and Hedges, S. B. (2017). TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* 34, 1812–1819. doi: 10.1093/molbev/msx116
- Letunic, I., Khedkar, S., and Bork, P. (2021). SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res.* 49, D458–D460. doi: 10.1093/nar/gkaa937
- Lewis, J. A., Jacobo, E. P., Palmer, N., Vermerris, W., Sattler, S. E., Brozik, J. A., et al. (2024). Structural and interactional analysis of the flavonoid pathway proteins: chalcone synthase, chalcone isomerase and chalcone isomerase-like protein. *Int. J. Mol. Sci.* 25, 5651. doi: 10.3390/ijms25115651
- Liao, B., Wang, C., Li, X., Man, Y., Ruan, H., and Zhao, Y. (2023). Genome-wide analysis of the *Populus trichocarpa* laccase gene family and functional identification of PtrLAC23. *Front. Plant Sci.* 13, 1063813. doi: 10.3389/fpls.2022.1063813
- Lin, L.-M., Guo, H.-Y., Song, X., Zhang, D.-D., Long, Y.-H., and Xing, Z.-B. (2021). Adaptive evolution of Chalcone Isomerase superfamily in Fagaceae. *Biochem. Genet.* 59, 491–505. doi: 10.1007/s10528-020-10012-z
- Liu, T., Liu, H., Xian, W., Liu, Z., Yuan, Y., Fan, J., et al. (2024). Duplication and sub-functionalization of flavonoid biosynthesis genes plays important role in Leguminosae root nodule symbiosis evolution. *J. Integr. Plant Biol.* 66, 2191–2207. doi: 10.1111/jipb.v66.10
- Liu, W., Feng, Y., Yu, S., Fan, Z., Li, X., Li, J., et al. (2021). The flavonoid biosynthesis network in plants. *Int. J. Mol. Sci.* 22, 12824. doi: 10.3390/ijms222312824
- Long, M., Betrán, E., Thornton, K., and Wang, W. (2003). The origin of new genes: glimpses from the young and old. *Nat. Rev. Genet.* 4, 865–875. doi: 10.1038/nrg1204
- Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *science* 290, 1151–1155. doi: 10.1126/science.290.5494.1151
- Lyons, E., Pedersen, B., Kane, J., Alam, M., Ming, R., Tang, H., et al. (2008). Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: CoGe with rosids. *Plant Physiol.* 148, 1772–1781. doi: 10.1104/pp.108.124867
- Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., et al. (2010). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* 39, D225–D229. doi: 10.1093/nar/gkq1189
- McDaniel, S. F. (2021). Bryophytes are not early diverging land plants. *New Phytol.* 230, 1300–1304. doi: 10.1111/nph.v230.4
- McKhann, H. I., and Hirsch, A. M. (1994). Isolation of chalcone synthase and chalcone isomerase cDNAs from alfalfa (*Medicago sativa* L.): highest transcript levels occur in young roots and root tips. *Plant Mol. Biol.* 24, 767–777. doi: 10.1007/BF00029858
- Mehdy, M. C., and Lamb, C. J. (1987). Chalcone isomerase cDNA cloning and mRNA induction by fungal elicitor, wounding and infection. *EMBO J.* 6, 1527–1533. doi: 10.1002/j.1460-2075.1987.tb02396.x
- Mistry, J., Finn, R. D., Eddy, S. R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 41, e121. doi: 10.1093/nar/gkt263
- Mouradov, A., and Spangenberg, G. (2014). Flavonoids: a metabolic network mediating plants adaptation to their real estate. *Front. Plant Sci.* 5, 620. doi: 10.3389/fpls.2014.00620
- Naake, T., Maeda, H. A., Proost, S., Tohge, T., and Fernie, A. R. (2021). Kingdom-wide analysis of the evolution of the plant type III polyketide synthase superfamily. *Plant Physiol.* 185, 857–875. doi: 10.1093/plphys/kiaa086
- Ngaki, M. N., Louie, G. V., Philippe, R. N., Manning, G., Pojer, F., Bowman, M. E., et al. (2012). Evolution of the chalcone-isomerase fold from fatty-acid binding to stereospecific catalysis. *Nature* 485, 530–533. doi: 10.1038/nature11009
- Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Ni, R., Zhu, T.-T., Zhang, X.-S., Wang, P.-Y., Sun, C.-J., Qiao, Y.-N., et al. (2020). Identification and evolutionary analysis of chalcone isomerase-fold proteins in ferns. *J. Exp. Bot.* 71, 290–304. doi: 10.1093/jxb/erz425
- Park, S.-I., Park, H.-L., Bhoo, S.-H., Lee, S.-W., and Cho, M.-H. (2021). Biochemical and molecular characterization of the rice chalcone isomerase family. *Plants* 10, 2064. doi: 10.3390/plants10102064
- Petrussa, E., Braidot, E., Zancani, M., Peresson, C., Bertolini, A., Patui, S., et al. (2013). Plant flavonoids—biosynthesis, transport and involvement in stress responses. *Int. J. Mol. Sci.* 14, 14950–14973. doi: 10.3390/ijms140714950
- Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., et al. (2019). Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. *Genome Biol.* 20, 1–23. doi: 10.1186/s13059-019-1650-2
- Ralston, L., Subramanian, S., Matsuno, M., and Yu, O. (2005). Partial reconstruction of flavonoid and isoflavonoid biosynthesis in yeast using soybean type I and type II chalcone isomerases. *Plant Physiol.* 137, 1375–1388. doi: 10.1104/pp.104.054502
- Rausher, M. D. (2006). "The evolution of flavonoids and their genes," in *The science of flavonoids*, ed. E. Grotewold. (New York, NY: Springer New York), 175–211.
- Renny-Byfield, S., and Wendel, J. F. (2014). Doubling down on genomes: polyploidy and crop plants. *Am. J. Bot.* 101, 1711–1725. doi: 10.3732/ajb.1400119
- Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* 11, e0163962. doi: 10.1371/journal.pone.0163962
- Shimada, N., Aoki, T., Sato, S., Nakamura, Y., Tabata, S., and Ayabe, S.-i. (2003). A cluster of genes encodes the two types of chalcone isomerase involved in the biosynthesis of general flavonoids and legume-specific 5-deoxy (iso) flavonoids in *Lotus japonicus*. *Plant Physiol.* 131, 941–951. doi: 10.1104/pp.004820
- Shirley, B. W., Hanley, S., and Goodman, H. M. (1992). Effects of ionizing radiation on a plant genome: analysis of two *Arabidopsis* transparent testa mutations. *Plant Cell* 4, 333–347. doi: 10.1105/tpc.4.3.333
- Shomali, A., Das, S., Arif, N., Sarraf, M., Zahra, N., Yadav, V., et al. (2022). Diverse physiological roles of flavonoids in plant environmental stress responses and tolerance. *Plants* 11, 3158. doi: 10.3390/plants11223158
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., et al. (2009). Polyploidy and angiosperm diversification. *Am. J. Bot.* 96, 336–348. doi: 10.3732/ajb.0800079

- Soltis, D. E., Visger, C. J., and Soltis, P. S. (2014). The polyploidy revolution then ... and now: Stebbins revisited. *Am. J. Bot.* 101, 1057–1078. doi: 10.3732/ajb.1400178
- Sugimoto, K., Irani, N. G., Grotewold, E., and Howe, G. A. (2024). Catalytically impaired chalcone isomerase retains flavonoid biosynthetic capacity. *Plant Physiol.* 195, 1143–1147. doi: 10.1093/plphys/kiac096
- Van Tunen, A., Koes, R., Spelt, C., van der Krol, A., Stuitje, A., and Mol, J. (1988). Cloning of the two chalcone flavanone isomerase genes from *Petunia hybrida*: coordinate, light-regulated and differential expression of flavonoid genes. *EMBO J.* 7, 1257–1263. doi: 10.1002/j.1460-2075.1988.tb02939.x
- Waki, T., Mameda, R., Nakano, T., Yamada, S., Terashita, M., Ito, K., et al. (2020). A conserved strategy of chalcone isomerase-like protein to rectify promiscuous chalcone synthase specificity. *Nat. Commun.* 11, 870. doi: 10.1038/s41467-020-14558-9
- Wang, J., Jiang, Y., Sun, T., Zhang, C., Liu, X., and Li, Y. (2022b). Genome-wide classification and evolutionary analysis reveal diverged patterns of chalcone isomerase in plants. *Biomolecules* 12, 961. doi: 10.3390/biom12070961
- Wang, H.-L., Manolas, C., and Xanthidis, D. (2022a). “Statistical Analysis with Python,” in *Handbook of Computer Programming with Python*, 1st Edn, eds. C. Bernhardt. (Boca Raton, FL: Chapman and Hall/CRC), 373–408.
- Wang, Q.-H., Zhang, J., Liu, Y., Jia, Y., Jiao, Y.-N., Xu, B., et al. (2022c). Diversity, phylogeny, and adaptation of bryophytes: insights from genomic and transcriptomic data. *J. Exp. Bot.* 73, 4306–4322. doi: 10.1093/jxb/erac127
- Winkel-Shirley, B. (2002). Biosynthesis of flavonoids and effects of stress. *Curr. Opin. Plant Biol.* 5, 218–223. doi: 10.1016/S1369-5266(02)00256-X
- Wolf-Saxon, E. R., Moorman, C. C., Castro, A., Ruiz-Rivera, A., Mallari, J. P., and Burke, J. R. (2023). Regulatory ligand binding in plant chalcone isomerase-like (CHIL) proteins. *J. Biol. Chem.* 299, 104804. doi: 10.1016/j.jbc.2023.104804
- Wu, S., Han, B., and Jiao, Y. (2020). Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Mol. Plant* 13, 59–71. doi: 10.1016/j.molp.2019.10.012
- Xie, J., Chen, Y., Cai, G., Cai, R., Hu, Z., and Wang, H. (2023). Tree Visualization By One Table (tvBOT): a web application for visualizing, modifying and annotating phylogenetic trees. *Nucleic Acids Res.* 51, W587–W592. doi: 10.1093/nar/gkad359
- Xu, H., Lan, Y., Xing, J., Li, Y., Liu, L., and Wang, Y. (2022). AfCHIL, a type IV chalcone isomerase, enhances the biosynthesis of naringenin in metabolic engineering. *Front. Plant Sci.* 13, 891066. doi: 10.3389/fpls.2022.891066
- Yin, Y.-c., Zhang, X.-d., Gao, Z.-q., Hu, T., and Liu, Y. (2019). The research progress of chalcone isomerase (CHI) in plants. *Mol. Biotechnol.* 61, 32–52. doi: 10.1007/s12033-018-0130-3
- Yonekura-Sakakibara, K., Higashi, Y., and Nakabayashi, R. (2019). The origin and evolution of plant flavonoid metabolism. *Front. Plant Sci.* 10, 943. doi: 10.3389/fpls.2019.00943
- Zamora, P., Pardo, A., Fierro, A., Prieto, H., and Zúñiga, G. E. (2013). Molecular characterization of the chalcone isomerase gene family in *Deschampsia Antarctica*. *Polar Biol.* 36, 1269–1280. doi: 10.1007/s00300-013-1346-0
- Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinf.* 19, 15–30. doi: 10.1186/s12859-018-2129-y
- Zhang, C., Shine, M., Pyle, A. M., and Zhang, Y. (2022). US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nat. Methods* 19, 1109–1115. doi: 10.1038/s41592-022-01585-1
- Zhang, S., Su, J., Chiu, T.-Y., Fang, J., Liang, X., He, Z., et al. (2024). CcCHIL, a type IV chalcone isomerase that can improve (2S)-naringenin production in *Saccharomyces cerevisiae*. *Food Bioprod. Process.* 148, 229–239. doi: 10.1016/j.fbp.2024.09.016
- Zu, Q.-L., Qu, Y.-Y., Ni, Z.-Y., Zheng, K., Chen, Q., and Chen, Q.-J. (2019). The chalcone isomerase family in cotton: Whole-genome bioinformatic and expression analyses of the *Gossypium barbadense* L. response to fusarium wilt infection. *Genes* 10, 1006. doi: 10.3390/genes10121006
- Zuntini, A. R., Carruthers, T., Maurin, O., Bailey, P. C., Leempoel, K., Brewer, G. E., et al. (2024). Phylogenomics and the rise of the angiosperms. *Nature* 629(8013), 843–850. doi: 10.1038/s41586-024-07324-0