Preprints.org

Article

Genome-Wide Identification and Evolution-Profiling Analysis of TPS Gene Family in *Camphora longepaniculata*

Xin Liu [*] , Minghu Zhang , Xin Zhao , Li Yang , Mang Shen , Ruizhang Feng , Qin Wei

Article

Genome-Wide Identification and Evolution-Profiling Analysis of TPS Gene Family in Camphora longepaniculata

Xin Liu [1,2,*], Minghu Zhang [1,2], Xin Zhao [1,2], Li Yang [1,2], Mang Shen [3], Ruizhang Feng [1,2] and Qin Wei [1,2]

[1] Sichuan Oil Cinnamon Engineering Technology Research Center, Yibin 644000, Sichuan, China
[2] Faculty of Agriculture, Forestry and Food Engineering, Yibin University, Yibin 644000, Sichuan, China
[3] State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Sichuan Agricultural University, Wenjiang, Chengdu, 611130, China
* Correspondence: liuxin.sau@gmail.com

Abstract: *Camphora longepaniculata* is an important economic crop renowned for its rich volatile terpene compounds. Terpene synthases (TPS) are key enzymes in the biosynthesis of these compounds, playing significant roles in plant growth, development, and secondary metabolism. In this study, a total of 86 TPS genes were identified in *Camphora longepaniculata*, which were classified into five groups based on their evolutionary relationships. Analysis of cis-regulatory elements revealed associations between TPS genes and processes related to plant growth, development, and environmental stress responses. Gene Ontology (GO) enrichment analysis indicated that these TPS genes are predominantly linked to various enzymatic activities. Furthermore, analysis of duplication events revealed that tandem duplications (TD) and whole genome duplications (WGD) are major driving forces in the evolution of the TPS gene family. Notably, 18 TPS genes were found to be upregulated in high essential oil content varieties of *Camphora longepaniculata*. These findings lay a solid foundation for further exploration of the functions of TPS genes in *Camphora longepaniculata*.

Keywords: *Camphora longepaniculata*; TPS gene family; evolution; gene duplicated events; expression profiles

Introduction

Terpenoids are the largest and most diverse class of plant secondary metabolites, with over 40,000 known structures[1]. They play crucial roles in plant growth, development, and ecological interactions[2]. Terpenoids function as phytoalexins in direct defense and as signals in indirect defense responses involving herbivores and their natural enemies[3]. Their biosynthesis involves the formation of C5 precursors, immediate diphosphate precursors, and diverse end products, with terpene synthases playing a key role[4]. Terpenoids have been traditionally used in food, pharmaceuticals, and chemical industries, and are gaining attention for sustainable pest control and abiotic stress protection [5]. Recent research has focused on elucidating their ecological functions, characterizing biosynthetic enzymes, and exploring their potential in various aspects of human life[6]. Metabolic engineering approaches are being developed to increase terpenoid production for industrial applications[5].

Terpenoids, or isoprenoids, are a highly diverse group of compounds derived from isoprene units (C5H8)n and classified based on the number of isoprene units they contain: hemiterpenes (C5), monoterpenes (C10), sesquiterpenes (C15), diterpenes (C20), triterpenes (C30), tetraterpenes (C40), and polyterpenes[7]. These compounds play vital roles in both primary and secondary metabolism, exhibiting various biological activities, including anticancer and antimalarial effects. Terpene synthases (TPSs) are key enzymes that facilitate the conversion of precursors such as geranyl diphosphate (GPP), farnesyl diphosphate (FDP), and geranylgeranyl diphosphate (GGDP) into a wide array of mono-, sesqui-, and diterpenes[8,9]. Additionally, other enzymes like cytochrome P450s

and transferases further modify terpenoids, enhancing their functional diversity. The biosynthesis of terpenoids involves complex pathways that guide carbon flow and catalyze cyclization reactions, allowing plants to produce hundreds of distinct terpenoids, with the total estimated in the thousands[10,11]. This vast diversity is enabled by the unique biosynthetic mechanisms in plants, which facilitate genetic and biochemical changes[12]. Terpenoids are essential for plant defense, communication, and the attraction of pollinators and seed dispersers, making them crucial in ecological interactions. Their importance in sustainable pest control and abiotic stress mitigation has become a significant research focus, underscoring their multifaceted roles in plant ecology.

Terpene synthase (TPS) catalyzes the production of terpenoids, which are closely linked to the abundance of TPS genes in plants. Most plant genomes contain approximately 20 to 152 TPS genes, with the exception of the moss *Physcomitrella* patens, which has only one functional TPS gene[13]. The TPS gene family, categorized into class I and II, plays crucial roles in plant metabolism, stress response, and development. Significant expansion of this gene family has occurred through segmental duplications[14]. Evolutionary analyses suggest that TPS genes likely originated in the ancestor of land plants, with potential gene transfer occurring between microbes and plants[15]. Plant TPS proteins contain both TPS and TPP domains; however, many studies have not detected TPP activity. It appears that the TPP domain in plant TPS proteins has lost its enzymatic activity during the course of evolution[14,16]. Additionally, TPS genes demonstrate tissue-specific and stress-responsive expression patterns, indicating functional divergence[17]. Phylogenetic studies have identified distinct groups within the TPS gene family, revealing differential evolution between subfamilies involved in primary and secondary metabolism[13,18]. These findings enhance our understanding of the evolutionary history and functional diversification of the TPS gene family in plants.

Camphora longepaniculata (Gamble) Y. Yang, Bing Liu, and Zhi Yang , often referred to as "China Eucalyptus Oil," is a culturally and economically significant evergreen tree species extensively cultivated in the Yibin region of southwestern China, possessing high economic value for the pharmaceutical, fragrance, and daily chemical industries[19]. The leaves and twigs of *C. longepaniculatum* are harvested primarily for essential oil extraction, which is rich in terpenoids, comprising over 85% of its composition[20]. Key components of its essential oils include 1,8-cineole, α-terpilenol, and γ-terpinen, which are valued for their antibacterial, anti-inflammatory, and antioxidant properties[19]. Research has demonstrated that these essential oils exhibit significant antibacterial activity against various pathogens, with α-terpineol being particularly effective due to its mechanism of disrupting cellular integrity[21,22]. The composition of essential oils varies depending on factors such as growth environment, phenological stages, and plant varieties [22]. Furthermore, transcriptomic and metabolomic analyses suggest that variations in essential oil content and leaf phenotypes are linked to phenylalanine metabolism, specifically in flavonoid and terpenoid pathways[23].

This study conducted a comprehensive genome-wide identification and characterization of the TPS gene family in *C. longepaniculata* based on the latest genomic information. A total of 86 CLTPSs were identified. Computational methods were used to analyze gene structure, motif composition, subcellular localization, protein 3D modeling, and expression profiles. The study examined the synteny of TPS genes among four *Lauraceae* species, as well as the role of cis-regulatory elements in the promoter regions during stress responses and developmental processes. Additionally, it elucidated the expression patterns of this gene family in *C. longepaniculata* varieties rich in terpenoid substances through transcriptome data analysis, providing valuable insights for research on the TPS gene family in this species.

Results

Characterization of TPS Gene Family

To comprehensively identify the terpene synthase (TPS) genes in four species of the *Lauraceae* family, we conducted a genome-wide search utilizing hidden Markov models (HMM) specific to the TPS domains (PF01397; PF03936). Our analysis revealed the presence of 81 TPS genes in *C. camphora*,

72 in *C. chago*, 88 in *C. kanehirae*, and 86 in *C. longepaniculatum*. The protein lengths of these TPS genes ranged from 79 to 278 amino acids, with isoelectric points varying from 4.41 to 9.81 and instability indices between 21.69 and 65.53. Except for *CchigoTPS52*, *CkTPS14*, and *CkTPS51*, the remaining TPS proteins exhibited hydrophilic characteristics (with a grand average of hydropathicity, GRAVY < 0) (Figure 1, Table S1). Subcellular localization predictions indicated that the majority of the TPS proteins (74.3%) are localized in the cytoplasm, while 24.3% are found in plastids. Only three TPS proteins were localized to the endoplasmic reticulum and extracellular regions (Table S1).
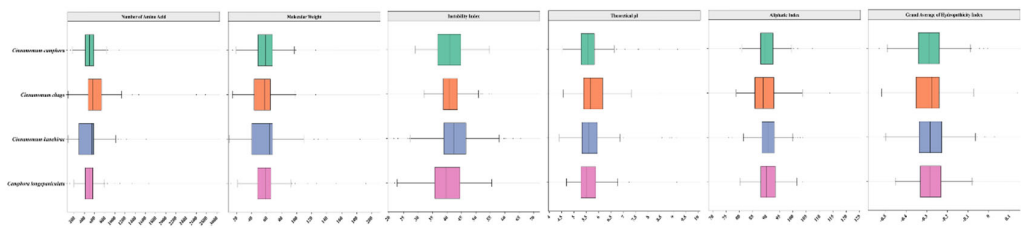

Figure 1. Physicochemical properties of TPS genes in four *Lauraceae* species.

Analysis of Phylogenetic Relationships of TPS Gene Family

A multiple sequence alignment of TPS sequences from four *Lauraceae* species and *A. thaliana* was performed using ClustalW to investigate the evolutionary relationships among TPS gene members. A phylogenetic tree was constructed based on the full-length amino acid sequences, revealing five distinct subfamilies (Figure 2A). The largest subfamily, Group 2, comprised 166 members, while Group 3 was the smallest, containing only 18 members. Notably, Group 4 included only members from *Arabidopsis*. The distribution of TPS gene members among the four *Lauraceae* species exhibited similar patterns across the different subfamilies, though some variations were evident. In Group 2, three species (*C. camphora*, *C. kanehirae*, and *C. longepaniculatum*) each harbored more than 40 members, whereas *C. chago* had fewer. Conversely, in Group 1, *C. chago* displayed the highest number of members, totaling 17 (Figure 2B). These findings highlight the evolutionary divergence of TPS genes within the *Lauraceae* family, as well as the distinct distribution patterns among species. Such insights contribute to a deeper understanding of the functional diversity and evolutionary history of TPS genes in relation to terpenoid biosynthesis.
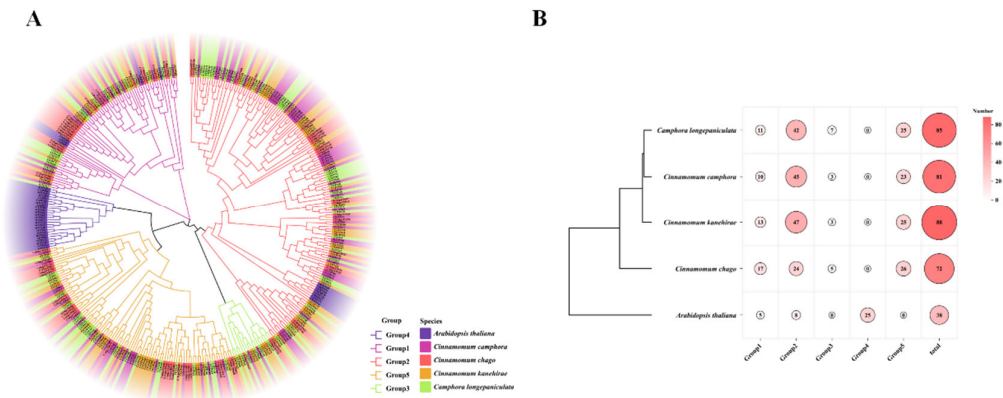

Figure 2. Phylogenetic tree of TPS genes in 5 species. (A) The evolutionary tree was constructed by the maximum likelihood method (bootstrap values: 1000 replicates) using MEGA X. (B) Statistical results of different subfamilies in different species.

Conserved Motifs Analysis of TPS Gene Family

To identify the conserved motifs of the TPS gene family in four *Lauraceae* species, the MEME suite was utilized to analyze the full-length protein sequences of TPS from each species. This analysis revealed a total of 10 distinct motifs (Figure 3). The results indicated that *C. chago* exhibited 5 similar motifs compared to *C. longepaniculatum*, while both *C. camphora* and *C. kanehirae* displayed 7 motifs that were similar to those in *C. longepaniculatum*. Notably, four motifs (motif 1, motif 3, motif 4, and motif 7) were conserved across all four species, suggesting their potential significance in the evolutionary processes within the *Lauraceae* family. Interestingly, motif 8 was exclusively identified in *Cinnamomum camphora*, indicating unique evolutionary adaptations in this species. These findings enhance our understanding of the structural and functional diversity of TPS genes among *Lauraceae* species.
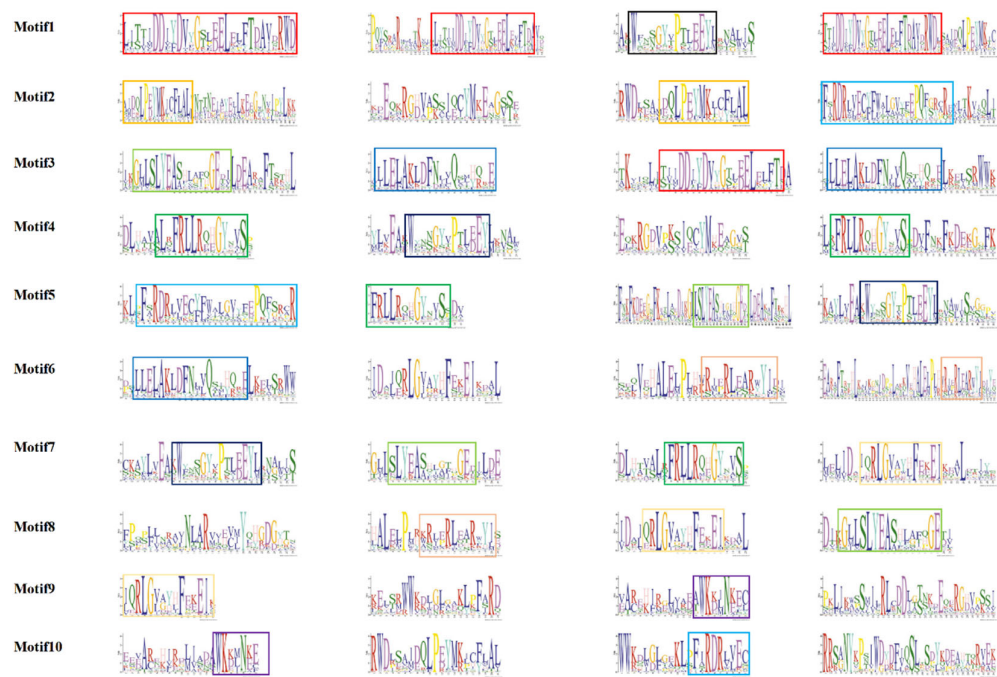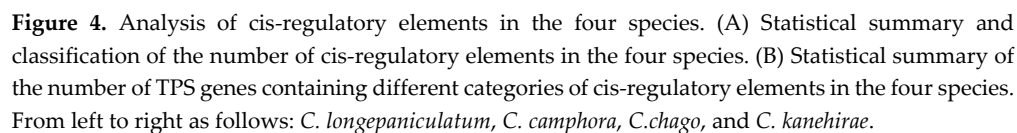

Figure 3. The conserved motifs among the four species are presented from left to right as follows: *C. longepaniculatum*, *C.chago*, *C. kanehirae*, and *C. camphora*. Boxes of the same color indicate similar motifs.

Promoters Analysis of TPS Genes

To further investigate the transcriptional regulation and potential functional roles of TPS genes, we predicted the presence of cis-regulatory elements in their promoter regions. Beyond the core cis-elements, we identified a total of 69 distinct cis-acting elements in the 2000 bp upstream region of the transcription start site of the TPS genes (Table S2). These elements play critical roles in various biological processes, including stress responsive, hormone responsive, metabolic regulation, and growth and developmental processes (Figure 4A). Among the identified elements, transcription factor binding sites were the most abundant, followed by light-responsive and hormone-responsive elements (Figure 4A). With the exception of *CcTPS36*, *CkTPS49*, and *CLTPS73*, all other TPS genes were found to contain at least one hormone-responsive element, the most prevalent of which was the abscisic acid response element (ABRE). Furthermore, the TPS promoter regions harbored several hormone-responsive elements, including auxin response elements (AuxRR core, TGA-element), MeJA response elements (CGTCA-motif, TGACG-motif), gibberellin response elements (TATC-box, P-box, GARE-motif), and salicylic acid response elements (TCA, SARE). The analysis also highlighted several types of stress response elements, including hypoxia-responsive elements (GC-motif), drought-responsive elements (MBS), wound response elements (WUN-motif), low-temperature

responsive elements (LTR), and defense/stress-responsive elements (TC-rich repeats). Among the four species analyzed, some genes were found to contain all types of cis-elements, with *C. longepaniculatum* exhibiting 47, *C. camphora* exhibiting 68, *C. chago* exhibiting 54, and *C. kanehirae* exhibiting 75 genes (Figure 4B). The findings suggest that TPS genes may play a role in the transcriptional control of plant growth, hormone regulation, and stress responses.

Figure 4. Analysis of cis-regulatory elements in the four species. (A) Statistical summary and classification of the number of cis-regulatory elements in the four species. (B) Statistical summary of the number of TPS genes containing different categories of cis-regulatory elements in the four species. From left to right as follows: *C. longepaniculatum*, *C. camphora*, *C.chago*, and *C. kanehirae*.

Duplication Events Analysis of TPS Genes

This study extracted the physical locations of TPS genes based on the reference genome annotation files of *C. longepaniculatum* and three other species within the *Lauraceae*. The results indicate that the distribution of TPS genes on chromosomes is not uniform, appearing primarily in gene clusters. *C. longepaniculatum* exhibits clustering of TPS genes mainly on chromosomes 3, 4, 7, and 9, whereas the other three species show clustering on chromosomes 2, 4, 7, and 10 (Figure 5A-D). This uneven distribution suggests a potential evolutionary adaptation of TPS genes within these species, highlighting the importance of genomic architecture in the regul.

Tandem duplications (TD) and whole-genome duplications (WGD) are significant drivers of gene family evolution and expansion. In this study, we analyzed the duplication events of TPS genes across four species. The results revealed 35, 35, 23, and 45 pairs of TD genes in species *C. longepaniculatum*, *C. camphora*, *C. chago*, and *C. kanehirae*, respectively, along with 8, 16, 7, and 12 pairs of WGD genes (Figure 5A-D). These two types of duplicated genes account for 50.0%, 62.9%, 41.6%, and 64.7% of the total TPS genes in each species. This indicates that both TD and WGD have made substantial contributions to the amplification of TPS genes throughout evolutionary processes, highlighting their importance in the diversification of this gene family.

Additionally, we employed KaKs_Calculator to compute the Ka, Ks, and Ka/Ks ratios for the duplicated gene pairs to analyze the evolutionary selection pressures on the TPS gene family. Most gene pairs displayed Ka/Ks ratios below 1 (Figure 5E), suggesting a predominant negative selection acting on these TPS genes, which may help maintain essential functions and prevent deleterious mutations. Conversely, the genes *CcTPS69*, *CcTPS70*, and *CchigoTPS4* demonstrated Ka/Ks ratios exceeding 1, implying they have undergone positive selection, potentially driving their diversification and adaptation in response to specific environmental pressures. This variation in selection pressures highlights the evolutionary dynamics within the TPS gene family.
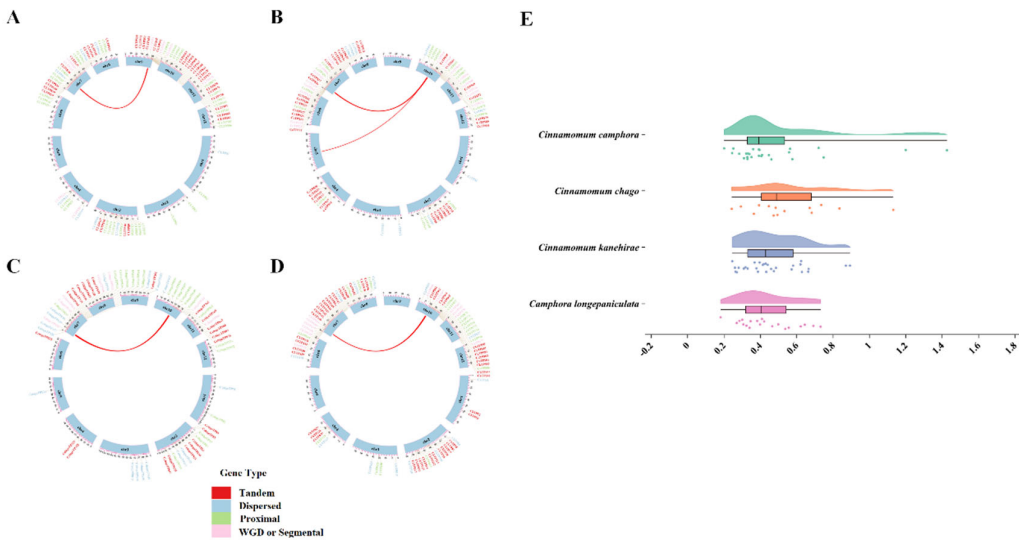
Figure 5. The duplicated gene pair and Ka/Ks values of TPS genes in 4 species. (A) *C. longepaniculatum*, (B) *C. camphora*, (C) *C. chago*,(D) *C. kanehirae*. (E) The Ka/Ks values of TPS genes in 4 species.

Collinearity Analysis of TPS Genes

To further investigate the evolutionary mechanisms of the TPS genes, we analyzed the homology relationships between *C. longepaniculatum* and the other three species. The results indicated that *C. longepaniculatum* shares 38, 36, and 42 direct homologous gene pairs with species *C. camphora*, *C. chago*, and *C. kanehirae*, respectively (Figure 6A). These homologous gene pairs are primarily concentrated on chromosomes 3 and 7 of *C. longepaniculatum* (Table S3), suggesting a potential genomic basis for the functional diversity observed within the TPS gene family across these species . Additionally, it was observed that among the collinear TPS gene pairs in *C. longepaniculatum*, 16 genes (Figure 6B) exhibited collinearity with homologous genes in all three other species. This finding highlights the conservation of these specific TPS genes across the examined species, suggesting that they may play critical roles in maintaining similar functions or regulatory pathways throughout evolution. The presence of such conserved collinear genes underscores the importance of these TPS genes in the evolutionary dynamics of the *Lauraceae* family.

We also compared the structural variations between the genomes of *C. longepaniculatum* and *C. camphora*, revealing a high degree of collinearity between the two genomes, accompanied by certain structural variations. The predominant type of variation observed was inversions, which were concentrated on chromosomes 1, 2, 6, and 7 (Figure 6C). Additionally, instances of translocations were noted, with specific segments on chromosomes 1 and 10 of *C. longepaniculatum* corresponding to translocated regions on chromosomes 2 and 9 of *C. camphora* (Figure 6C). These findings indicate that while the overall genomic structure is conserved, rearrangements have occurred, which may influence the evolutionary trajectory and functional divergence of these species.
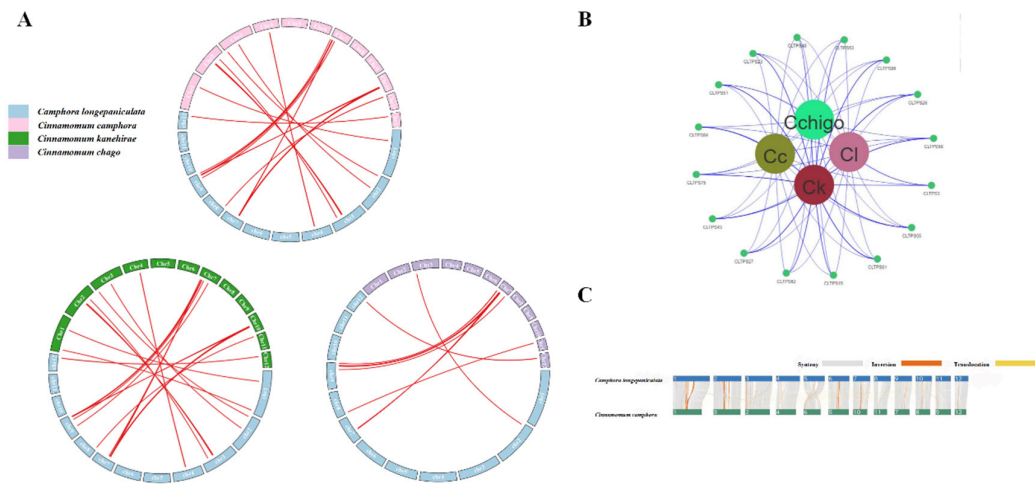
Figure 6. Syntenic relationships analysis of *TPS* genes in four *Lauraceae* species. (A) Syntenic relationships analysis of TPS genes between *C. longepaniculatum* and three other *Lauraceae* species, (B) The collinearity relationship between four genomes, (C) Analysis of structural variations in *C. longepaniculatum* and *C. camphora.*

Prediction of Protein 3D Structure

For each of the four TPS gene groups in *C. longepaniculatum* (noting that Group 4 has no members), we selected one TPS gene and conducted homology modeling of its three-dimensional structure using Swiss-Model (Figure 7). The analysis revealed that all four proteins predominantly feature α-helices as their main structural elements. The domains PF01397 and PF03936 were identified as key components of these TPS proteins, containing essential structures such as α-helices. We also predicted the protein binding pocket (Figure 7, Table S4), where the surrounding amino acid residues are crucial in determining the physicochemical properties, conformation, and functionality of the proteins. These findings provide insights into the potential biochemical activities of the TPS proteins and their roles in terpenoid biosynthesis.
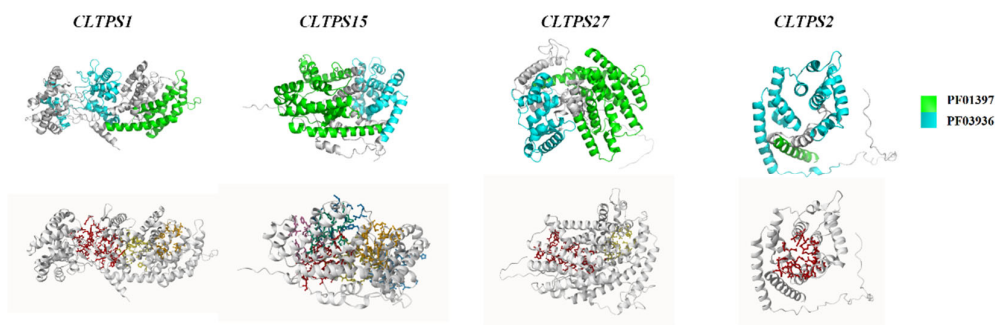

Figure 7. The 3D structure modeling and ligand binding site prediction in *CLTPSs.*

Gene Ontology Functional Annotation of CLTPSs

To further explore the potential biological functions of TPS genes, we utilized the eggNOG-mapper database to perform Gene Ontology (GO) functional annotation on the TPS protein sequences from *C. longepaniculatum*, followed by a comprehensive statistical analysis of these annotations (Figure 8, Table S5).

The results indicated that, in the category of Molecular Function (MF), the TPS proteins are primarily associated with the activity of various enzymes, including cyclase, S-linalool synthase, terpene synthase, sesquiterpene synthase, and pinene synthase. In terms of Cellular Component (CC), the proteins are related to cellular structures, organelles, and chloroplasts. In the Biological Process (BP) category, the TPS genes are involved in the biosynthesis of metabolites and responses to stimuli. Notably, the key components of *C. longepaniculatum* essential oil include various monoterpenols and terpenes, highlighting the significant role of TPS genes in the synthesis of these important phytochemicals. This functional characterization underscores the relevance of TPS genes in the biosynthesis of bioactive compounds that contribute to the plant's ecological interactions and potential applications in pharmaceuticals and cosmetics.
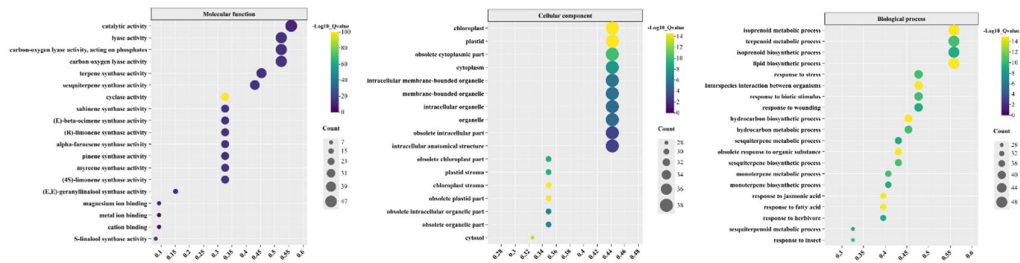

Figure 8. GO functional annotations of TPSs in *C. longepaniculatum*.

Expression Analysis of CLTPSs in C. longepaniculatum Varieties with Different Essential Oil Contents

To investigate the role of TPS genes in the production of essential oil in *C. longepaniculatum*, we analyzed transcriptome data to examine the expression patterns of TPS genes in high essential oil content (CLL) and ordinary (CLH) varieties. A total of 22 differentially expressed TPS genes were identified between the two varieties, with 18 genes being upregulated in the CLL variety (Figure 9).

This upregulation suggests that TPS genes play a significant role in the biosynthesis of essential oils, highlighting their importance in the high oil content phenotype of *C. longepaniculatum*. These findings contribute to our understanding of the molecular mechanisms underlying essential oil production and may inform breeding strategies aimed at enhancing oil yield in this species.
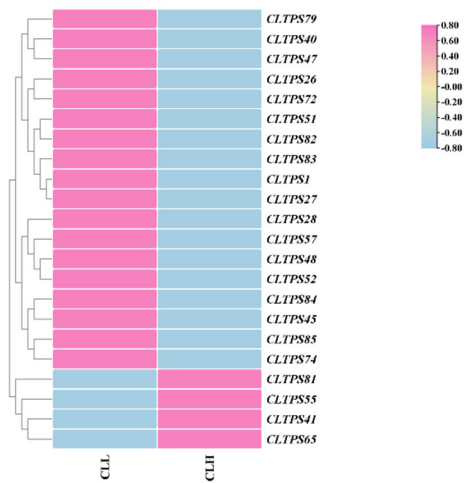

Figure 9. Gene expression analysis of TPS genes in in high essential oil content (CLL) and ordinary (CLH) varieties of *C. longepaniculatum*.

Discussion

Terpenes and terpenoid compounds are key natural substances involved in plant defense, development, and the composition of essential oils. The synthesis and accumulation of a wide variety of volatile terpenoids in these plants can significantly alter the quality and flavor of the oils, presenting substantial commercial value for oil-producing plants[24]. TPS are crucial enzymes responsible for the diversity of terpenes. Studying the TPS gene family provides valuable theoretical support for the genetic improvement of oil-producing plants. Terpenoid compounds play a vital role in enhancing plant resistance, and TPS enzymes are key factors in the structural diversity of these natural compounds. They catalyze the conversion of isoprenyl diphosphates into various forms, including volatile monoterpenes, sesquiterpenes, semi-volatile compounds, and non-volatile diterpenes[5,25]. This study comprehensively identified 86 TPS genes in *C. longepaniculatum* and investigated their phylogenetic relationships, gene duplication events, cis-regulatory elements in promoter sequences, and expression patterns of the TPS gene family in *C. longepaniculatum* with varying essential oil contents. These findings enhance our understanding of the molecular mechanisms governing terpenoid biosynthesis and their implications for plant resilience and essential oil production.

Phylogenetic analysis categorized the four *Lauraceae* species and *A. thaliana* into five distinct groups. Group 2 contained the largest number of members, comprising over 50% of the *Lauraceae* species analyzed, while Group 4 was unique to *A. thaliana*. Among the 10 identified motifs, 5 were present in all four *Lauraceae* species, with both *C. camphora*, and *C. kanehirae* sharing 7 similar motifs with *C. longepaniculatum*. This indicates a significant level of conservation and similarity in the TPS gene family across these species.

Cis-regulatory elements within plant TPS genes play a crucial role in mediating responses to environmental stress and signaling pathways during growth and development, and they have been extensively studied across various plant species[26,27]. In addition to core cis-elements, this analysis identified a total of 69 distinct cis-regulatory elements within the 2000 bp upstream region of the transcription start sites of TPS genes. The results indicate that TPS genes are not only key enzymes in terpenoid biosynthesis but also likely participate in various biological processes. These include plant growth and development, where they may influence developmental pathways; stress responses, where they contribute to the plant's ability to cope with adverse conditions such as drought, salinity, or temperature extremes; and light signaling, which is crucial for optimizing photosynthesis and energy use. The presence of multiple hormone-responsive elements further suggests that TPS genes are integrated into complex regulatory networks, allowing them to modulate metabolic pathways in response to internal and external stimuli. Additionally, GO enrichment analysis results also validate this point, highlighting the involvement of TPS genes in various biological processes, including metabolic pathways, stress responses, and developmental regulation. This multifaceted role underscores the importance of TPS genes in enhancing plant resilience and adaptability in changing environments.

WGD and small-scale duplication events are not only defining features of eukaryotic genome evolution but also significant drivers for the emergence of new functions in genomic and genetic evolutionary systems[28]. In the four analyzed species of the *Lauraceae* family, the duplicated TPS genes accounted for 50.0% (*C. longepaniculatum*), 62.9% (*C. camphora*), 41.6% (*C. chago*), and 64.7% (*C. kanehirae*) of the total TPS gene complement, respectively. This substantial representation of duplicated genes contributes significantly to the uneven distribution of TPS genes across different chromosomes. It is generally accepted that, due to selective pressures, amino acid sequences with important biological functions exhibit relatively low variability among different species[29]. Most TPS genes have undergone strong negative selection, indicating that these genes are relatively conserved. This conservation is essential for maintaining the functional integrity of terpenoid biosynthesis pathways, which are critical for plant defense, communication, and adaptation to environmental stresses.

C. longepaniculatum, native to the Yibin region of Sichuan in Southwest China, is an economically significant tree species with considerable local importance. Various parts of the plant are utilized for

essential oil extraction[19]. Key components of the leaf essential oil include 1,8-cineole, α-pinene, and γ-terpinene, which are highly regarded for their excellent antibacterial, anti-inflammatory, and antioxidant properties[22]. To investigate the role of TPS genes in the essential oil content of *C. longepaniculatum*, this study analyzed the expression profiles of TPS genes in high essential oil content lines (CLL) compared to those with ordinary content (CLH). The results revealed that 18 TPS genes were significantly upregulated in the CLL, indicating that these genes may play a role in influencing the essential oil yield of this species. This suggests the potential for TPS genes to be targeted in genetic improvement efforts aimed at enhancing the economic value of *C. longepaniculatum* through increased essential oil production.

Materials and Methods

Identification and Chromosomal Location of TPS Gene Family in C. longepaniculata

Genomic data and GFF file for C. longepaniculata[30] was downloaded from Figshare (https://figshare.com/s/ff6a0f810527f61ef63c), and for C. chago[31] from the National Genomics Data Center (accession number PRJCA022354; https://ngdc.cncb.ac.cn/gwh). The genomes of C. kanehirae[32] (accession number PRJNA477266; https://www.ncbi.nlm.nih.gov) and C. camphora[33] (https://doi.org/10.6084/m9.figshare.20647452.v1) were also obtained. The typical domains of terpene synthases (TPS) were retrieved from the Pfam database (http://pfam.xfam.org/) under Pfam IDs PF01397 and PF03936. Hidden Markov model (HMM) searches were performed using HMMER 3.0[34], with an expected value (E-value) threshold set to 1e-5. Potential TPS protein sequences were further validated for TPS domain presence using the NCBI Conserved Domain Database (CDD; https://www.ncbi.nlm.nih.gov/cdd/), the SMART database (http://smart.embl-heidelberg.de/), and the Pfam database. Biochemical parameters of the identified TPS proteins were computed using the ExPASy online tool (https://web.expasy.org/protparam/). The subcellular localization of the TPS proteins was predicted using DeepLoc 2.1 (https://services.healthtech.dtu.dk/services/DeepLoc-2.1/).

Phylogenetic Relationship and Conserved Motifs Analysis

The maximum likelihood (ML) method implemented in MEGA X[35] was utilized to construct a phylogenetic tree for TPS proteins. A bootstrap analysis with 1000 resampling iterations was performed to assess the robustness of the tree. The analysis included TPS proteins from *A. thaliana*, *C. longepaniculata*, *C. chago*, *C. kanehirae*, and *C. camphora*. The resulting phylogenetic tree file (*.nwk) was uploaded to the iTOL website (http://itol.embl.de/) for enhanced visualization. The conserved motifs of *C. longepaniculata* TPS proteins were analyzed using the MEME Suite[36] (http://meme-suite.org/meme), with the maximum number of motifs set to 10.

Analysis of Promoter Cis-Acting Elements

The 2000-bp upstream sequences of TPS genes from *C. longepaniculata*, *C. chago*, *C. kanehirae*, and *C. camphora* were employed to define the promoter regions. To identify potential cis-regulatory elements within these promoters, the Plant CARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was utilized. This database offers tools for the prediction and analysis of cis-regulatory elements, which are crucial for understanding gene regulation and expression patterns. The identified elements will be systematically classified and organized based on their functional categories, such as light responsiveness, hormone responsiveness, and stress-related elements. Following this classification, the results will be visualized using the ggplot2 package[37] in R.

Synteny Analysis for TPS Genes Analyze the gene collinearity among *C. longepaniculata*, *C. chago*, *C. kanehirae*, and *C. camphora* based on their amino acid sequences and chromosome locations using MCScanX software[38]. The synteny circos plots was created using TBtools (v2.097)[39] and the Venn diagram was generated using EVenn (http://www.ehbio.com/test/venn/)[40] to show the genes that exhibit synteny with *C. longepaniculata* TPS genes across several species. Subsequently, we employed the approximate method implemented in KaKs Calculator version 1.2[41] to estimate the synonymous (Ks) and nonsynonymous (Ka) substitution rates of the orthologous genes among the four species. Identification and visualization of structural variations in the genomes of *C. longepaniculata* and *C. camphora* using GenomeSyn[42].

Prediction of Protein Pocket Binding Sites

The three-dimensional structures of the candidate *CLTPSs* in *C. longepaniculata* were predicted through Swiss-Model server (https://swissmodel.expasy.org/), which utilizes homology modeling techniques to generate reliable structural models based on known protein structures. Docking analysis of ligand-binding regions in the predicted protein models was performed using the P2Rank tool (https://prankweb.cz/), followed by visualization in PyMOL[43].

Gene Ontology

The TPS genes of *C. longepaniculata* were annotated using the eggNOG mapper server[44], which provides functional annotation based on evolutionary relationships and gene orthology. Following the annotation, Gene Ontology (GO) enrichment analysis is conducted using TBtools. The enrichment results will be classified according to molecular function (MF), cellular component (CC), and biological process (BP), selecting the top 20 entries from each category for visualization.

Expression Patterns of Cltpss in C. longepaniculata With High Terpenoid Content

Transcriptome data for various *C. longepaniculata* varieties exhibiting different terpenoid contents were obtained from the Sequence Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra/) using the accession number PRJNA804339. The raw sequencing data were processed using Trimmomatic[45], which effectively filtered out adapter sequences and low-quality reads to ensure data integrity. Subsequently, Kallisto[46] was employed for aligning the cleaned reads to the reference genome, followed by quantification of gene expression levels. This analysis yielded comprehensive expression data that reflect the transcriptional profiles of the different varieties.

Author Contributions: Xin Liu conducted most of the analysis and wrote the manuscript; Minghu Zhang, Xin Zhao, Mang Shen, and Yue Yan performed part of the data collection and analysis; Ruizhang Feng provided guidance on manuscript writing, Qin Wei designed the experiments and revised the manuscript. All authors reviewed the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within this article or the Supplementary Materials, further inquiries can be directed to the corresponding author/s.

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

References

1. Bohlmann, J.; Keeling, C.I. Terpenoid Biomaterials. *The Plant Journal* **2008**, *54*, 656–669.
2. Loreto, F.; Dicke, M.; Schnitzler, J.-P.; Turlings, T.C. Plant Volatiles and the Environment. *Plant, cell & environment* **2014**, *37*, 1905–1908.
3. Cheng, A.-X.; Lou, Y.-G.; Mao, Y.-B.; Lu, S.; Wang, L.-J.; Chen, X.-Y. Plant Terpenoids: Biosynthesis and Ecological Functions. *Journal of integrative plant biology* **2007**, *49*, 179–186.
4. McGarvey, D.J.; Croteau, R. Terpenoid Metabolism. *The plant cell* **1995**, *7*, 1015.
5. Li, C.; Zha, W.; Li, W.; Wang, J.; You, A. Advances in the Biosynthesis of Terpenoids and Their Ecological Functions in Plant Resistance. *International Journal of Molecular Sciences* **2023**, *24*, 11561.
6. Jahangeer, M.; Fatima, R.; Ashiq, M.; Basharat, A.; Qamar, S.A.; Bilal, M.; Iqbal, H. Therapeutic and Biomedical Potentialities of Terpenoids-A Review. *Journal of Pure & Applied Microbiology* **2021**, *15*.
7. Mabou, F.D.; Yossa, I.B.N. Terpenes: Structural Classification and Biological Activities. *IOSR J Pharm Biol Sci* **2021**, *16*, 25–40.
8. O'maille, P.E.; Malone, A.; Dellas, N.; Andes Hess Jr, B.; Smentek, L.; Sheehan, I.; Greenhagen, B.T.; Chappell, J.; Manning, G.; Noel, J.P. Quantitative Exploration of the Catalytic Landscape Separating Divergent Plant Sesquiterpene Synthases. *Nature chemical biology* **2008**, *4*, 617–623.
9. Irmisch, S.; Krause, S.T.; Kunert, G.; Gershenzon, J.; Degenhardt, J.; Köllner, T.G. The Organ-Specific Expression of Terpene Synthase Genes Contributes to the Terpene Hydrocarbon Composition of Chamomile Essential Oils. *BMC Plant Biology* **2012**, *12*, 1–13.
10. Pichersky, E.; Raguso, R.A. Why Do Plants Produce so Many Terpenoid Compounds? *New Phytologist* **2018**, *220*, 692–702.
11. Liu, D.; Huang, X.; Jing, W.; An, X.; Zhang, Q.; Zhang, H.; Zhou, J.; Zhang, Y.; Guo, Y. Identification and Functional Analysis of Two P450 Enzymes of *Gossypium Hirsutum* Involved in DMNT and TMTT Biosynthesis. *Plant Biotechnology Journal* **2018**, *16*, 581–590.
12. Tholl, D. Biosynthesis and Biological Functions of Terpenoids in Plants. *Biotechnology of isoprenoids* **2015**, 63–106.
13. Degenhardt, J.; Köllner, T.G.; Gershenzon, J. Monoterpene and Sesquiterpene Synthases and the Origin of Terpene Skeletal Diversity in Plants. *Phytochemistry* **2009**, *70*, 1621–1637.
14. Yang, H.-L.; Liu, Y.-J.; Wang, C.-L.; Zeng, Q.-Y. Molecular Evolution of Trehalose-6-Phosphate Synthase (TPS) Gene Family in Populus, *Arabidopsis* and Rice. **2012**.
15. Chen, F.; Tholl, D.; Bohlmann, J.; Pichersky, E. The Family of Terpene Synthases in Plants: A Mid-Size Family of Genes for Specialized Metabolism That Is Highly Diversified throughout the Kingdom. *The Plant Journal* **2011**, *66*, 212–229.
16. Vandesteene, L.; Ramon, M.; Le Roy, K.; Van Dijck, P.; Rolland, F. A Single Active Trehalose-6-P Synthase (TPS) and a Family of Putative Regulatory TPS-like Proteins in *Arabidopsis*. *Molecular plant* **2010**, *3*, 406–419.
17. Xu, Y.; Wang, Y.; Mattson, N.; Yang, L.; Jin, Q. Genome-Wide Analysis of the *Solanum Tuberosum* (Potato) Trehalose-6-Phosphate Synthase (TPS) Gene Family: Evolution and Differential Expression during Development and Stress. *BMC genomics* **2017**, *18*, 1–15.
18. Whittington, D.A.; Wise, M.L.; Urbansky, M.; Coates, R.M.; Croteau, R.B.; Christianson, D.W. Bornyl Diphosphate Synthase: Structure and Strategy for Carbocation Manipulation by a Terpenoid Cyclase. *Proceedings of the National Academy of Sciences* **2002**, *99*, 15375–15380.
19. Wu, D.; Zhu, P.; Wu, H.; Dai, H. Industry Development Status and Prospect of *Cinnamomum Longepaniculatum*. *Open Access Library Journal* **2022**, *9*, 1–10.
20. Hu, W.; Gao, H.; Jiang, X.; Yang, H. Analysis on Constituents and Contents in Leaf Essential Oil from Three Chemical Types of *Cinnamum Camphora*. *J Cent South Univ Technol* **2012**, *32*, 186–194.
21. Li, N.; Zu, Y.; Wang, W. Antibacterial and Antioxidant of Celery Seed Essential Oil. *Chin Condiment* **2012**, *37*, 28–30.
22. Li, L.; Li, Z.-W.; Yin, Z.-Q.; Wei, Q.; Jia, R.-Y.; Zhou, L.-J.; Xu, J.; Song, X.; Zhou, Y.; Du, Y.-H.; et al. Antibacterial Activity of Leaf Essential Oil and Its Constituents from *Cinnamomum Longepaniculatum*. *International journal of clinical and experimental medicine* **2014**, *7*, 1721.
23. Zhao, X.; Yan, Y.; Zhou, W.; Feng, R.; Shuai, Y.; Yang, L.; Liu, M.; He, X.; Wei, Q. Transcriptome and Metabolome Reveal the Accumulation of Secondary Metabolites in Different Varieties of *Cinnamomum Longepaniculatum*. *BMC Plant Biology* **2022**, *22*, 243.
24. Yan, X.-M.; Zhou, S.-S.; Liu, H.; Zhao, S.-W.; Tian, X.-C.; Shi, T.-L.; Bao, Y.-T.; Li, Z.-C.; Jia, K.-H.; Nie, S.; et al. Unraveling the Evolutionary Dynamics of the TPS Gene Family in Land Plants. *Frontiers in Plant Science* **2023**, *14*, 1273648.
25. Gutiérrez-del-Río, I.; López-Ibáñez, S.; Magadán-Corpas, P.; Fernández-Calleja, L.; Pérez-Valero, Á.; Tuñón-Granda, M.; Miguélez, E.M.; Villar, C.J.; Lombó, F. Terpenoids and Polyphenols as Natural Antioxidant Agents in Food Preservation. *Antioxidants* **2021**, *10*, 1264.

26. Yadav, S.K.; Santosh Kumar, V.V.; Verma, R.K.; Yadav, P.; Saroha, A.; Wankhede, D.P.; Chaudhary, B.; Chinnusamy, V. Genome-Wide Identification and Characterization of ABA Receptor PYL Gene Family in Rice. *BMC genomics* **2020**, *21*, 1–27.

27. Zhu, L.; Yan, K.; Ren, J.; Chen, Z.; Ma, Q.; Du, Y.; Wang, Y.; Li, S.; Li, Q. Genome-Wide Investigation of the PYL Genes in *Acer Palmatum* and Their Role in Freezing Tolerance. *Industrial Crops and Products* **2024**, *210*, 118107.

28. Wang, Y.; Wang, X.; Paterson, A.H. Genome and Gene Duplications and Gene Expression Divergence: A View from Plants. *Annals of the New York Academy of Sciences* **2012**, *1256*, 1–14.

29. Delph, L.F.; Kelly, J.K. On the Importance of Balancing Selection in Plants. *New phytologist* **2014**, *201*, 45–56.

30. Yan, K.; Zhu, H.; Cao, G.; Meng, L.; Li, J.; Zhang, J.; Liu, S.; Wang, Y.; Feng, R.; Soaud, S.A.; et al. Chromosome Genome Assembly of the *Camphora Longepaniculata* (Gamble) with PacBio and Hi-C Sequencing Data. *Frontiers in Plant Science* **2024**, *15*, 1372127.

31. Tao, L.; Guo, S.; Xiong, Z.; Zhang, R.; Sun, W. Chromosome-Level Genome Assembly of the Threatened Resource Plant *Cinnamomum Chago*. *Scientific Data* **2024**, *11*, 447.

32. Chaw, S.-M.; Liu, Y.-C.; Wu, Y.-W.; Wang, H.-Y.; Lin, C.-Y.I.; Wu, C.-S.; Ke, H.-M.; Chang, L.-Y.; Hsu, C.-Y.; Yang, H.-T.; et al. Stout Camphor Tree Genome Fills Gaps in Understanding of Flowering Plant Genome Evolution. *Nature plants* **2019**, *5*, 63–73.

33. Li, D.; Lin, H.-Y.; Wang, X.; Bi, B.; Gao, Y.; Shao, L.; Zhang, R.; Liang, Y.; Xia, Y.; Zhao, Y.-P.; et al. Genome and Whole-Genome Resequencing of *Cinnamomum Camphora* Elucidate Its Dominance in Subtropical Urban Landscapes. *BMC biology* **2023**, *21*, 192.

34. Mistry, J.; Finn, R.D.; Eddy, S.R.; Bateman, A.; Punta, M. Challenges in Homology Search: HMMER3 and Convergent Evolution of Coiled-Coil Regions. *Nucleic acids research* **2013**, *41*, e121–e121.

35. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular biology and evolution* **2018**, *35*, 1547.

36. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for Motif Discovery and Searching. *Nucleic acids research* **2009**, *37*, W202–W208.

37. Villanueva, R.A.M.; Chen, Z.J. Ggplot2: Elegant Graphics for Data Analysis 2019.

38. Wang, Y.; Tang, H.; DeBarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity. *Nucleic acids research* **2012**, *40*, e49–e49.

39. Chen, C.; Wu, Y.; Li, J.; Wang, X.; Zeng, Z.; Xu, J.; Liu, Y.; Feng, J.; Chen, H.; He, Y.; et al. TBtools-II: A "One for All, All for One" Bioinformatics Platform for Biological Big-Data Mining. *Molecular Plant* **2023**, *16*, 1733–1742.

40. Yang, M.; Chen, T.; Liu, Y.-X.; Huang, L. Visualizing Set Relationships: EVenn's Comprehensive Approach to Venn Diagrams. *iMeta* **2024**, e184.

41. Zhang, Z.; Li, J.; Zhao, X.-Q.; Wang, J.; Wong, G.K.-S.; Yu, J. KaKs_Calculator: Calculating Ka and Ks through Model Selection and Model Averaging. *Genomics, proteomics and bioinformatics* **2006**, *4*, 259–263.

42. Zhou, Z.-W.; Yu, Z.-G.; Huang, X.-M.; Liu, J.-S.; Guo, Y.-X.; Chen, L.-L.; Song, J.-M. GenomeSyn: A Bioinformatics Tool for Visualizing Genome Synteny and Structural Variations. *Journal of genetics and genomics= Yi chuan xue bao* **2022**, *49*, 1174–1176.

43. DeLano, W.L.; others Pymol: An Open-Source Molecular Graphics Tool. *CCP4 Newsl. Protein Crystallogr* **2002**, *40*, 82–92.

44. Huerta-Cepas, J.; Forslund, K.; Coelho, L.P.; Szklarczyk, D.; Jensen, L.J.; Von Mering, C.; Bork, P. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular biology and evolution* **2017**, *34*, 2115–2122.

45. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120.

46. Bray, N.L.; Pimentel, H.; Melsted, P.; Pachter, L. Near-Optimal Probabilistic RNA-Seq Quantification. *Nature biotechnology* **2016**, *34*, 525–527.