BMC Biology

RESEARCH

Open Access

# The chromosome-scale genomes of two *Tinospora* species reveal differential regulation of the MEP pathway in terpenoid biosynthesis

Zhiyu Chen[1,5], Lan Xun[7], Yunyan Lu[7], Xingyu Yang[4,5], Minghui Chen[3], Tianyu Yang[1,5], Zhinan Mei[8], Yunqiang Yang[3,4,5*], Xuefei Yang[2,5,6*] and Yongping Yang[1,3,4,5*]

## Abstract

**Background** The relationship between gene family expansion and the resulting changes in plant phenotypes has shown remarkable complexity during the evolution. The gene family expansion has contributed to the diversity in plant phenotypes, specifically metabolites through neo-functionalization and sub-functionalization. However, the negative regulatory effects associated with the gene family expansion remain poorly understood.

**Results** Here, we present the chromosome-scale genomes of *Tinospora crispa* and *Tinospora sinensis*. Comparative genomic analyses demonstrated conserved chromosomal evolution within the Menispermaceae family. KEGG analysis revealed a significant enrichment of genes related to terpenoid biosynthesis in *T. sinensis*. However, *T. crispa* exhibited a higher abundance of terpenoids compared to *T. sinensis*. Detailed analysis revealed the expansion of genes encoding 1-hydroxy-2-methyl 2-(E)-butenyl 4-diphosphate synthase (HDS), a key enzyme in the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway of terpenoid biosynthesis in *T. sinensis*. TsiHDS4 retained the ancestral function of converting methylerythritol cyclic diphosphate (MEcPP) to (E)-4-hydroxy-3-methylbut-2-enyl diphosphate (HMBPP). However, the noncanonical CDS-derived small peptide TsiHDS5 was shown to interact with TsiHDS4, inhibiting its catalytic activity. This interaction reduced the levels of HMBPP and isopentenyl pyrophosphate (IPP), which represent key substrates for downstream terpenoid biosynthesis.

**Conclusions** These findings offer clues to decipher the variations in the MEP pathway of terpenoid biosynthesis between *T. crispa* and *T. sinensis* and form a basis for further detailed research on the negative regulation of expanded genes.

**Keywords** *Tinospora*, Comparative genomics, Terpenoid biosynthesis, MEP pathway, 1-hydroxy-2-methyl 2-(E)-butenyl 4-diphosphate synthase, Gene family expansion

*Correspondence:
Yunqiang Yang
yangyunqiang@xtbg.ac.cn
Xuefei Yang
xuefei@mail.kib.ac.cn
Yongping Yang
yangyp@xtbg.ac.cn
Full list of author information is available at the end of the article

## Background

Plant relatives typically share similar genomic structures but exhibit distinct phenotypic traits due to evolutionary processes such as speciation, adaptive evolution, and niche differentiation [1]. Understanding these differences is crucial for comprehending biodiversity and ecosystem complexity. Phenotypic variations, including those in metabolites, reflect adaptive evolution during species divergence. Phenotypic variations arise from genomic changes, such as copy number variations (CNVs) in protein-coding genes, single nucleotide polymorphisms (SNPs), and insertions or deletions (Indels), which affect gene expression and functions, leading to phenotypic diversity [2]. For instance, the loss of exons in the dammarenediol synthase (DDS)-encoding gene in *Araliaceae elata* after divergence from *Panax* results in the absence of dammarane-type saponin [3]. In plants, the metabolites influence phenotype and physiology and play key roles in plant-environment interactions, stress responses, and ecological adaptations [4].

Although plant relatives share similar genomic structures, they exhibit differences in metabolite content and type driven by gene family expansion, regulation changes, functional divergence, and the generation of novel genes [5, 6]. Gene family expansion, in particular, contributes significantly to phenotypic and metabolic differences. Duplicated genes can affect the phenotypes through neo-functionalization, sub-functionalization, or dosage effects [7, 8]. For example, the neo-functionalization of duplicated genes encoding enzymes like *CYP81AN15* and *SCPL* leads to the production of unique metabolites, such as ecgonine and leonurine [9, 10]. Gene family expansion, such as the *MAM* gene in *Brassica rapa* ssp. *rapa* [11] or *ADS* in *Artemisia annua* [12, 13], can increase metabolite diversity. However, studying these expanded gene families will provide valuable insights into genomic dynamics and the origins of plant diversity.

*Tinospora crispa* and *Tinospora sinensis*, both members of the Menispermaceae family, are valued in traditional medicine. *T. crispa* found in Southeast Asia, is known for its ability to lower blood sugar and treat rheumatism, while *T. sinensis*, native to southern China, India, and Sri Lanka, is used to relax tendons, activate collaterals, and relieve pain [14]. Both species are rich in terpenoids, which are their primary bioactive compounds. The diterpenoids identified in *T. crispa*, including Borapetoside A, Borapetoside C, and Borapetoside E, exhibit antidiabetic activities and show potential for the treatment of diabetes and other metabolic disorders [15–17]. Besides, 1-deacetyltinosposide A, extracted from *T. sinensis*, exhibits cytotoxicity against HeLa cells [18]. These findings established a robust chemical foundation for developing novel medicinal resources. However, the lack of genomic data for these two species limits our understanding of the molecular mechanisms underlying their metabolic diversity.

In this study, we sequenced the genomes of *T. crispa* and *T. sinensis*, revealing their phylogenetic position, metabolic differences, and molecular genetics through genome, transcriptome, metabolome, and molecular assays. Our findings provide insights into plant metabolite diversity and the unique evolutionary traits of these relatives.

## Results

### Metabolite profiles in *T. crispa* and *T. sinensis*

As two important folk medicinal plants, the relatives *T. crispa* and *T. sinensis* exhibit distinct medicinal properties, probably due to differences in their bioactive compounds. Therefore, the metabolites in the leaves and stems of these two species were analyzed to identify these differences. A total of 1178 metabolites from four samples were identified and categorized into 16 classes. The principal component analysis (PCA) showed that the samples were divided into four distinct groups, indicating differences in metabolites among the samples (Fig. 1a). *T. crispa* showed 457 and 404 differentially accumulated metabolites (DAMs) in its leaves and stems compared to *T. sinensis*, respectively (Fig. 1b, Additional file 1: Fig. S1). The DAMs, primarily terpenoids, flavonoids, and alkaloids (Additional file 1: Fig. S2, Additional file 1: Fig. S3), suggested that the metabolic diversity in the leaves and stems of the two *Tinospora* species probably contributed to the differences in their properties.

Terpenoids are key active components in both *Tinospora* species. Among the 50 terpenoids DAMs in the leaves of *T. crispa* compared with *T. sinensis*, 42 were upregulated, while 8 were downregulated, with 20 triterpenoids significantly higher in *T. crispa* leaves compared to *T. sinensis*. Besides, diterpenoids content varied significantly in two species of leaves, with isorosmanol, 7-O-methylrosmanol, and borapetoside E being 420.29, 215.39, and 56.98 times higher in *T. crispa* leaves than *T. sinensis*, respectively (Fig. 1c). Similarly, significant differences in stem metabolites were observed between the *Tinospora* species. Diterpenoids such as borapetoside E (600.79 times) and uncinatone (11.39 times), along with the triterpenoids alphitolic acid (37.17 times), were markedly higher in *T. crispa* than in *T. sinensis* (Additional file 1: Fig. S4). These findings highlight substantial variation in terpenoid content between the leaves and stems, with *T. crispa* exhibiting higher levels, suggesting genetic divergence in terpenoid biosynthesis pathways between the species.

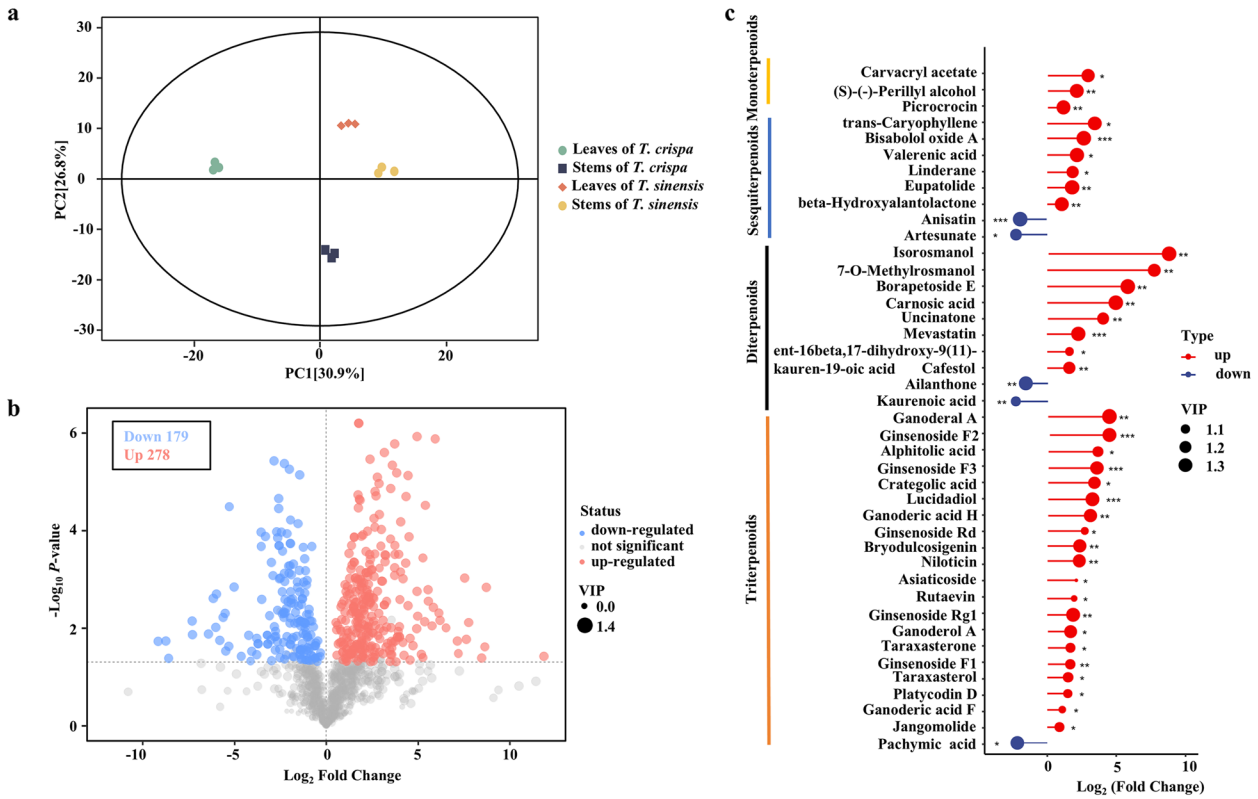


**Fig. 1** The metabolic analysis of *T. crispa* and *T. sinensis*. **a** Principal component analysis (PCA) of metabolic profiling. The axes represent relative distances, the colors and traits of the icons represent different samples. **b** Volcano plot of the number of up- and down-regulated differentially accumulated metabolites (DAMs) in the leaves of *T. crispa* compared to those of *T. sinensis*. Red color represents up-regulated DAMs and blue color represents down-regulated DAMs. The scatter size represents the VIP value. **c** Fold statistics of the content of up-regulated and down-regulated DAMs in the leaves of *T. crispa* compared to those of *T. sinensis*. Red color represents up-regulated DAMs, and blue-black color represents down-regulated DAMs (*$P < 0.05$, **$P < 0.01$)

## Genome sequencing and assembly

We sequenced and assembled the genome of two *Tinospora* species (Fig. 2a). A total of 64.94 Gb and 68.73 Gb clean Illumina data were generated for *T. crispa* and *T. sinensis*, respectively (Additional file 2: Table S1). Based on K-mer analysis, the genome size was estimated to be 999.51 Mb for *T. crispa* and 986.84 Mb for *T. sinensis*. Notably, the genomes of *T. crispa* and *T. sinensis* were assembled using approximately 73.88 Gb and 80.89 Gb Nanopore sequencing data, along with 64.94 Gb and 68.73 Gb clean Illumina data, respectively (Additional file 2: Table S1). After preliminary assembly, followed by error correction and redundancy removal, 962.85 Mb with a contig N50 of 7.42 Mb was generated for *T. crispa*, while the *T. sinensis* was assembled to 924.82 Mb with a contig N50 of 6.44 Mb (Fig. 2b, Additional file 2: Table S2). BUSCO evaluation revealed high completeness of the genome assembly with 96.34% for *T. crispa* and 96.03% for *T. sinensis* (Additional file 2: Table S3). Further assembly assessments showed that 99.45% and 99.77% of the Illumina reads were mapped to the assembled genome of *T. crispa* and *T. sinensis*, respectively (Additional file 2: Table S4). Meanwhile, 99.97% and 99.98% of the HiC-based assembly were anchored to 13 pseudochromosomes of *T. crispa* and *T. sinensis* with 94.47 Gb and 138 Gb clean Illumina reads, respectively (Additional file 1: Fig. S5, Additional file 2: Table S1). The chromosome number was consistent with that reported in the literature [19]. The length of *T. crispa* pseudochromosomes ranged from 60.79 to 85.02 Mb, while that of *T. sinensis* ranged from 55.75 to 79.03 Mb (Additional file 2: Table S5).

## Genome annotation and genome structure

We annotated the transposable elements, protein-coding genes, and non-coding RNAs in the genomes of *T. crispa* and *T. sinensis* and analyzed the genomic structural differences between the species. Approximately 759.15 Mb and 602.89 Mb of transposable elements (TE) were identified in the genome of *T. crispa* and *T. sinensis*, which accounted for 78.84% and 65.19% of the genome size, respectively (Additional file 2: Table S6). Long terminal

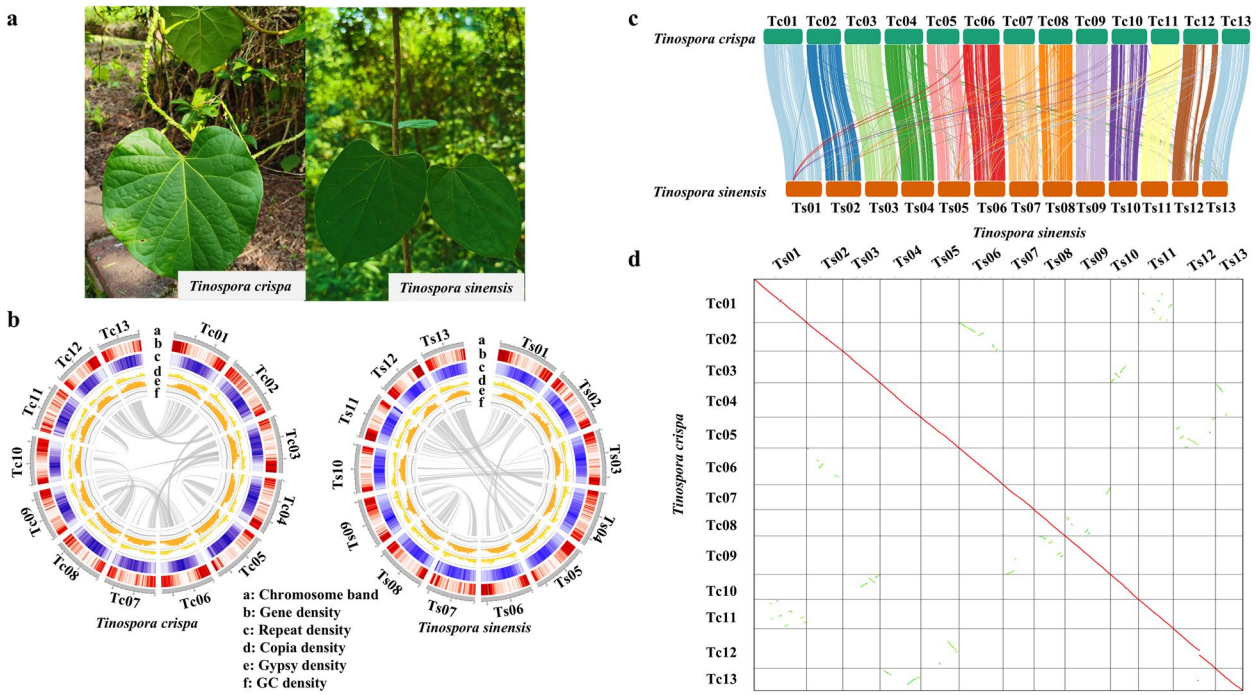


**Fig. 2** Morphology, genomic features, and synteny of *T. crispa* and *T. sinensis*. **a** Morphology of the leaf and stem in *T. crispa* and *T. sinensis*. **b** The overview of the genomic features of *T. crispa* and *T. sinensis*. The linking lines in the circle represent syntenic blocks in the genome, Tc01-Tc13 present pseudochromosomes of *T. crispa*, Ts01–Ts13 present pseudochromosomes of *T. sinensis*. **c** Synteny of *T. crispa* and *T. sinensis* based on similar sequence segments. **d** Synteny of *T. crispa* and *T. sinensis* based on syntenic blocks

repeats (LTR) constituted the largest proportion of TE, with LTR/Gypsy representing the leading share, accounting for 38.73% in *T. crispa* and 40.47% in *T. sinensis* (Additional file 2: Table S6). These results indicated that the expansion of LTR/Gypsy elements probably contributed to the increase in genome size in *T. crispa* and *T. sinensis*.

A total of 37.72 Gb and 38.04 Gb transcriptomic data were utilized for gene prediction of *T. crispa* and *T. sinensis*, respectively (Additional file 2: Table S7). Based on a combination of transcriptome-based and de novo prediction, a total of 19,526 and 19,947 genes were predicted in the genome of *T. crispa* and *T. sinensis*, respectively. BUSCO analysis showed 95.91% completeness in both genomes (Additional file 2: Table S8). Approximately 90.48% and 90.19% of the genes were functionally annotated based on Swissprot, KEGG, KOG, GO, and NR databases (Additional file 2: Table S9). In addition, noncoding RNAs, including rRNAs (644 and 676), miRNAs (1473 and 1941), and tRNAs (634 and 640), were predicted in the genome of *T. crispa* and *T. sinensis* (Additional file 2: Table S10).

Further, the synteny between the *Tinospora* species was analyzed using NGenomeSyn and WGDI software to explore the differences in their genomic structures. The results showed that similar sequence segments accounted for 39.93% and 35.96% of the genomic sequences of *T. crispa* and *T. sinensis*, respectively (Fig. 2c). Moreover, 366 syntenic blocks were identified between the two *Tinospora* species, with 79.83% and 77.56% of the genes in the genomes of *T. crispa* and *T. sinensis* were identified to be collinear, respectively (Fig. 2d). However, there were relatively large non-syntenic intervals in chromosome 12 and a version in chromosome 1 and 2 (Additional file 1: Fig. S6). These observations indicated a high genomic sequence similarity and synteny between *T. crispa* and *T. sinensis*.

## Comparative genomic analysis

A phylogenetic tree was constructed based on the sequences of 13 flowering species, including core eudicots, early-diverging eudicots, monocots, magnoliids, and *Amborella*, to elucidate the evolutionary history of *T. crispa* and *T. sinensis*. A total of 45 single-copy genes were identified by OrthoFinder and used to construct the maximum likelihood tree (Additional file 1: Fig. S7). The phylogenetic tree indicated that Menispermaceae was sister to Circaeasteraceae and Ranunculaceae and diverged from them approximately 94.02 million years ago (Mya), consistent with previous reports. Moreover, *Tinospora* species, including *T. crispa* and *T. sinensis*, had diverged from *Menispermum* (*M. dauricum*) around

53.85 Mya (Fig. 3a). The synonymous nucleotide substitutions (Ks) for homologous gene pairs in *T. crispa*, *T. sinensis*, and *C. chinensis* revealed peaks at 1.61, 1.62, and 1.53, respectively, indicating a shared ancient whole genome duplication (WGD) event within Ranunculales (Fig. 3b). The gene homology dot plot and genomic synteny further confirmed the absence of lineage-specific WGD event in the two *Tinospora* species (Additional file 1: Fig. S8, Additional file 1: Fig. S9).

Subsequently, the karyotypes of 10 eudicots were reconstructed based on the ancestral eudicot karyotype (AEK) to analyze the karyotype evolution of the two *Tinospora* species. This approach revealed that most species had experienced complex chromosome fusion and rearrangements, while *Tinospora* underwent at least one chromosome fusion between AEK chromosomes 1 and 4 (Fig. 3c). This fusion, also observed in the karyotype of *M. dauricum*, implies that it may be conserved within the Menispermaceae family. The findings highlighted the evolutionary conservation of Menispermaceae karyotype, providing valuable insights into ancestral traits. Additionally, the divergence of *T. crispa* and *T. sinensis* was dated to ~ 6.58 Mya. KEGG analysis showed that the expanded genes of *T. crispa* were significantly enriched "Isoflavonoid biosynthesis." Meanwhile, "Sesquiterpenoid and triterpenoid biosynthesis" and "Monoterpenoid biosynthesis" were enriched among the expanded genes of *T. sinensis* (Fig. 3d, Additional file 1: Fig. S10). Thus, these observations, combined with the results of metabolic profiling, suggested that the expansion of genes related to Isoflavonoid biosynthesis probably led to a higher content of flavonoids in *T. crispa*. Interestingly, while the genes involved in terpenoid biosynthesis expanded, the content of terpenoid decreased in *T. sinensis*.

#### Genes associated with terpenoid biosynthesis

Further, to elucidate the reason for the expansion of genes related to terpenoid biosynthesis and the decrease in the terpenoid content in *T. sinensis*, the genes involved in terpenoid biosynthesis were identified and their origins were analyzed across all 13 flowering plants (Fig. 4a, Additional file 1: Fig. S11). The results showed that the genes related to the MEP and Mevalonate (MVA) pathways exhibited relatively similar copy numbers in these species, except for the *HDS* in *T. sinensis* (Fig. 4a, Additional file 1: Fig. S12), encoding an enzyme that catalyzed the biosynthesis of HMBPP from MEcPP [20]. In contrast to most species with 1 or 2 *HDS* genes, *T. sinensis* contains 7 *HDS* genes (labeled as *TsiHDS1* through *TsiHDS7*), derived from dispersed duplication (Fig. 4a, Additional file 1: Fig. S13). Additionally, syntenic analysis showed good collinearity among *TsiHDS4* and *HDSs* from most species (Fig. 4b), suggesting that TsiHDS4 might have preserved the ancestral function of HDS. However, the collinear genes of other *TsiHDSs* were not observed in the genomes of *T. crispa* and other species (Additional file 1: Fig. S14). The structural analysis showed that TsiHDS4 possessed a complete structure, comprising three complete domains, including TIM barrel (A domain), insert domain (A* domain), and 4Fe4S reductase domain (B domain). However, varying degrees of loss in gene structure and disruptions in the transcript reading frame were found in other *TsiHDSs*, which had only a part of the domain (Fig. 4c, Additional file 1: Fig. S15, Fig. S16, Fig. S17). Moreover, both polypeptides encoded by *TsiHDS1* and *TsiHDS5* were less than 100 amino acids. Based on the structural characteristic [21], TsiHDS1 and TsiHDS5 were identified as noncanonical CDS small peptides, a group of short open reading frames (sORFs).

Moreover, the expression patterns of genes involved in terpenoid biosynthesis in *T. crispa* and *T. sinensis* were analyzed, showing most genes exhibited similar expression pattern, except for genes involved terpene skeletons biosynthesis, such as oxidosqualene cyclase gene (*OSCs*) for triterpenoids [22] and terpene synthase gene (*TPSs*) for monoterpenoids, sesquiterpenoids, and diterpenoids [23]. Specifically, most of the *OSCs* in *T. crispa*, such as *TciOS6*, *TciOSC4*, and *TciOSC2*, were expressed at higher levels than their orthologs in *T. sinensis*, with this difference being especially evident in the leaves (Fig. 4d). This may account for the higher content of triterpenoids in *T. crispa*. Additionally, the *TPSs* in *T. crispa* and *T. sinensis* were classified into five subfamilies based on their phylogenetic relationships. Most TPS-e/f subfamily genes responsible for diterpenoid biosynthesis in *T. crispa*, such as *Tc04G01598.1*, *Tc02G00008.1*, and *Tc02G00007.1*, showed significantly higher expression in *T. crispa*

(See figure on next page.)

**Fig. 3** The genome evolution of *T. crispa* and *T. sinensis*. **a** Phylogenetic tree constructed using 13 flowering including *T. crispa* and *T. sinensis*. The number at each node represents the divergence time with 95% credibility intervals (million years ago). The blue, yellow, and red colors of the pie chart represent the percentage of unchanged, expanded, and contracted gene families, respectively. **b** Density distribution of the synonymous substitution rate (Ks) of *T. crispa* (Tci), *T. sinensis* (Tsi), *C. chinensis* (Cch), and *A. fimbriata* (Afi). **c** Karyotype construction of 10 eudicots constructed based on 7 blocks of AEK. Purple circle for WGD events, red pentagram for WGT events. **d** KEGG enrichment analysis of the significantly expanded gene families in *T. sinensis*

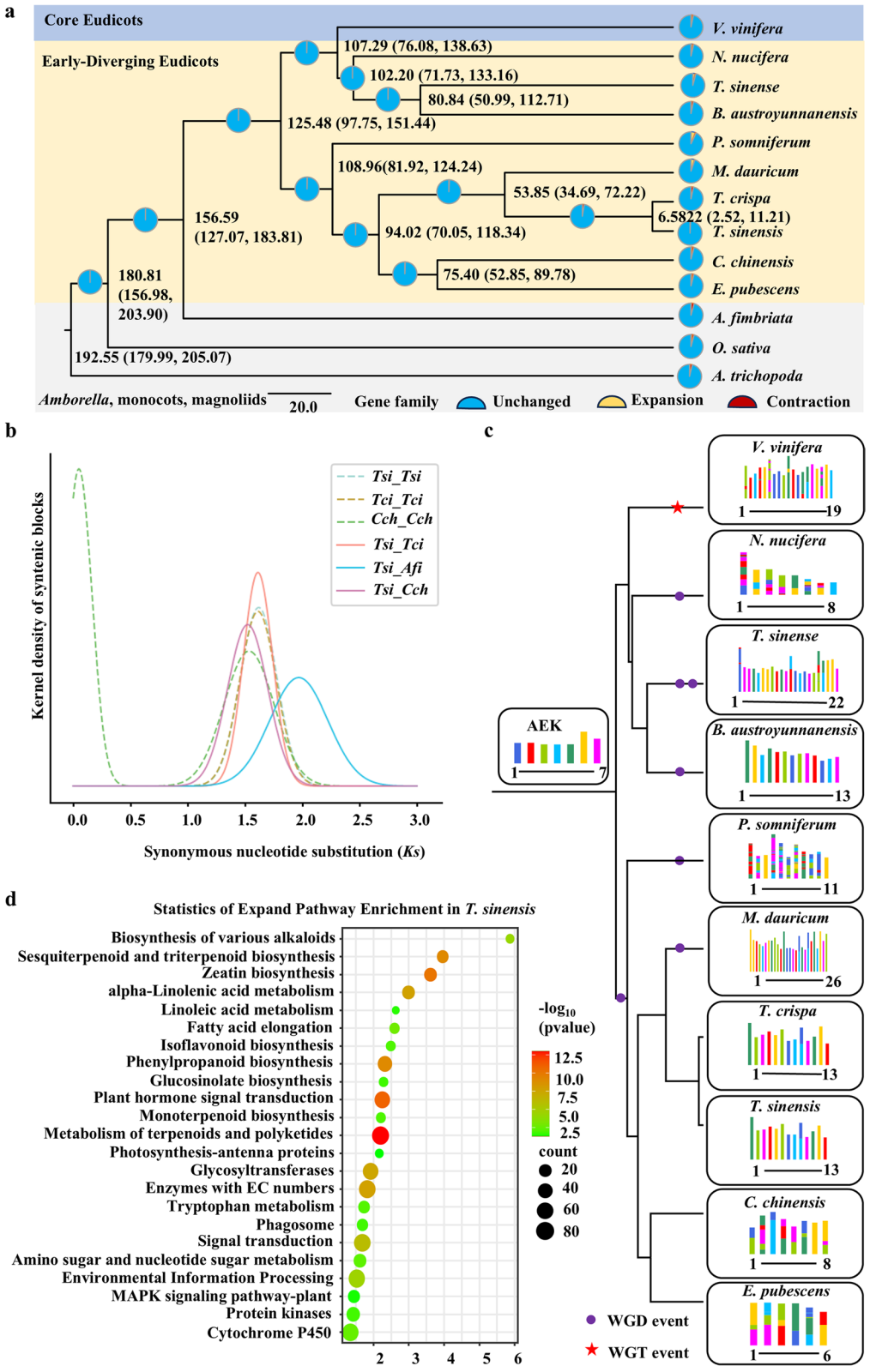


**Fig. 3** (See legend on previous page.)

compared to their orthologs in *T. sinensis* (Fig. 4e). This could be one of the important reasons for the abundant diterpenoids in *T. crispa*.

### Protein interactions of TsiHDSs in *T. sinensis*

Studies have shown that the interaction between the A* domain and A domain in HDS protein creates a "cup and ball" relationship, enhancing catalytic collaboration with the B domain [24]. Based on this finding, we investigated the interactions among HDS homologs to determine whether gene expansion contributed to terpenoid variation in *Tinospora* species. *TsiHDS4* has a complete and conserved gene structure, whereas other expanded *TsiHDSs* contain only partially conserved domains. Consequently, the interactions between TsiHDS4 and other TsiHDSs were explored by performing Luciferase Complementation Imaging (LCI) and Co-IP assays. LCI showed that TsiHDSs could not interact with themselves; however, a strong interaction signal was detected in *Nicotiana benthamiana* co-expressing TsiHDS4-cLUC and TsiHDS5-nLUC (Fig. 5a), while no fluorescence signal was detected when other TsiHDSs were co-injected with TsiHDS4. Additionally, the in vivo Co-IP assay detected TsiHDS4-GFP could be examined from co-immunoprecipitated with TsiHDS5-6FLAG after protein extracts with anti-flag M2 affinity gels (Fig. 5b; Additional file 1: Fig. S18). These results thus suggested that TsiHDS4 could physically interact with TsiHDS5.

To further identify the region of interaction between TsiHDS4 and TsiHDS5, three fragments of TsiHDS4 (A domain, A* domain, and B domain) were used to perform the LCI assays with TsiHDS5. Compared to the single construct combination, co-injected TsiHDS5-nLUC and TsiHDS4A*-cLUC resulted in strong fluorescence, whereas co-expression of TsiHDS5-nLUC with either TsiHDS4A-cLUC or TsiHDS4B-cLUC showed no interaction signal (Fig. 5c). These results proved that TsiHDS5 could interact with the insert domain (A* domain) of TsiHDS4.

### The catalytic activity of TsiHDS4 and TsiHDS5

The full-length CDS sequences of *TsiHDS4* and sORFs *TsiHDS5* were optimized and expressed in BL21 (DE3) to assess the catalytic activity of the enzymes encoded by these genes in terpenoid biosynthesis. The relevant metabolites, including the substrate MEcPP, the direct product HMBPP of HDS, and the common substrate of terpenoids IPP, were detected. Compared to the control, the overexpression of *TsiHDS4* led to the consumption of MEcPP and promoted the biosynthesis of HMBPP, and IPP. In contrast, no significant changes were detected in the content of HMBPP and IPP with the overexpression of the sORFs *TsiHDS5*. (Fig. 5d; Additional file 1: Fig. S19). These results suggested that TsiHDS4 catalyzed HMBPP biosynthesis from MEcPP, whereas TsiHDS5 did not function as a catalytic enzyme.

Then, to elucidate the impact of the interaction between TsiHDS4 and TsiHDS5 on terpenoid biosynthesis, *TsiHDS4*, *TsiHDS5*, and a mixture of *TsiHDS4* and *TsiHDS5*, were injected into *N. benthamiana* leaves for transient expression, and the levels of the metabolites MEcPP, HMBPP, and IPP were determined. Compared to the control, the expression of *TsiHDS4* alone in *N. benthamiana* could increase the biosynthesis of HMBPP and IPP, further verifying the catalytic function of TsiHDS4. On the other hand, the expression of *TsiHDS5* alone resulted in the accumulation of MEcPP and a decrease in HMBPP and IPP, suggesting an interaction of TsiHDS5 with HDS in *N. benthamiana* (NbHDS), inhibiting the catalytic function of NbHDS. Moreover, when *TsiHDS4* and sORFs *TsiHDS5* were co-injected, the content of MEcPP was significantly higher than that in the *TsiHDS4* expression alone (Fig. 5d; Additional file 1: Fig. S20), implying that TsiHDS5 could partially inhibit the function of TsiHDS4. These results proved that TsiHDS4 catalyzed the conversion of MEcPP to HMBPP, further facilitating the biosynthesis of downstream terpenoids. Additionally, sORFs *TsiHDS5* inhibits the catalytic activity of TsiHDS4, resulting in a decrease in IPP, which served as the common substrate for terpenoid biosynthesis.

(See figure on next page.)

**Fig. 4** Terpenoid biosynthetic pathway in 13 flowering species. **a** The number and duplication sources of the genes involved in terpene biosynthesis. The number in the circle represents the number of gene copies, and the color of the circle represents different sources, including Whole-genome duplication (WGD), transposed duplications (TRD), tandem duplications (TD), dispersed duplications (DSD) and proximal duplications (PD). **b** Syntenic blocks in the chromosome regions containing *HDS* genes in 13 flowering species. Syntenic blocks are connected by gray lines, in which syntenic *HDS* gene regions are highlighted in red. **c** Phylogenetic tree and structural domains of HDS in *T. crispa* and *T. sinensis*. The different colors of the rectangles represent different domains, including A domain (TIM barrel), A* domain (insert domain), and B domain (4Fe4S reductase domain). **d** Phylogenetic tree and heatmap of gene expression of *OSCs* in *T. crispa* and *T. sinensis* leaf and stem. **e** Phylogenetic tree and heatmap of gene expression of *TPSs* in *T. crispa* and *T. sinensis* leaf and stem. The different color bands represent the different subfamilies of *TPSs*, including TPS-a, TPS-b, TPS-c, TPS-g, and TPS-e/f subfamily

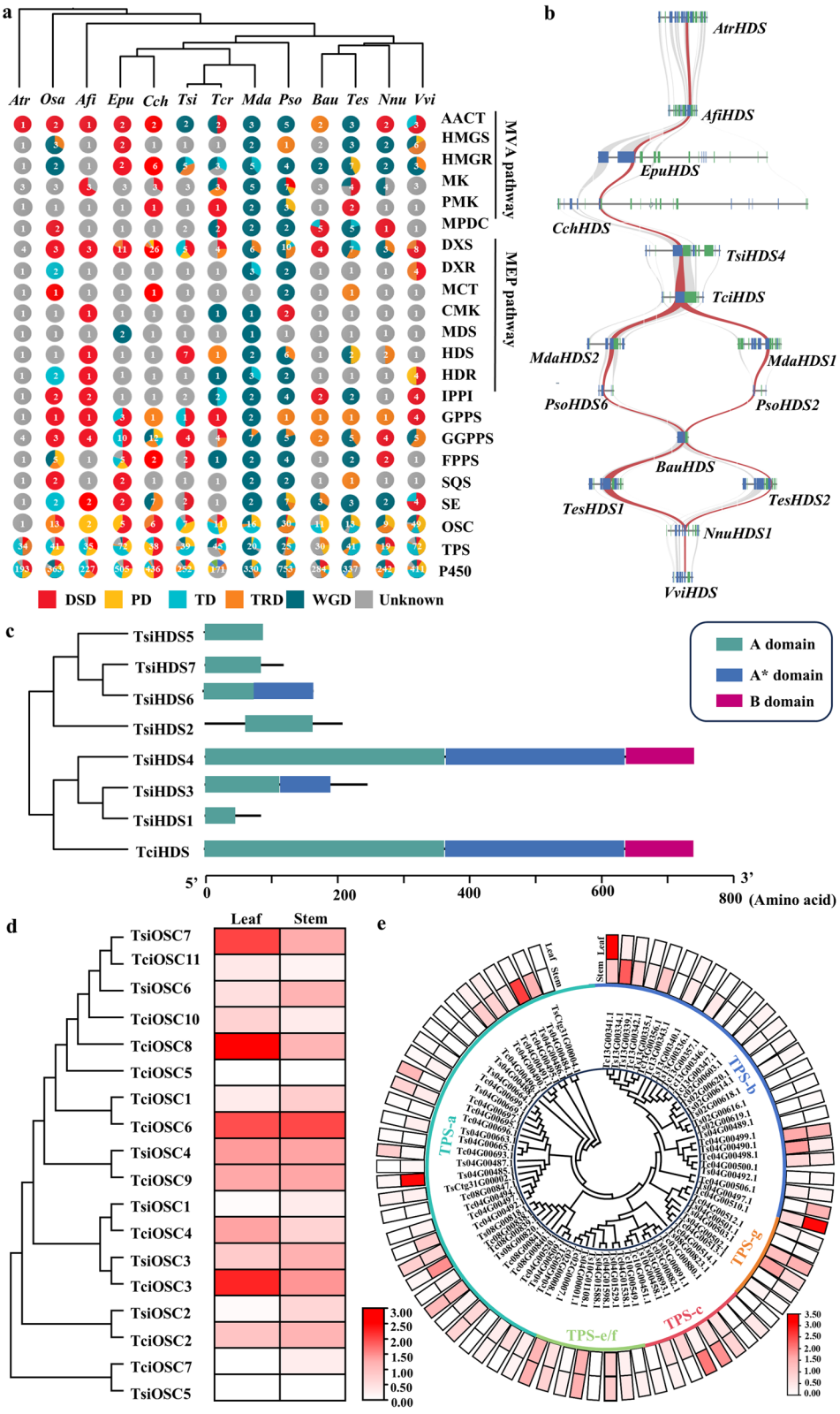


**Fig. 4** (See legend on previous page.)

## Discussion

Generally, relatives that share a common ancestor possess similar genomic structures. However, these species have diverged to exhibit distinctive phenotypic traits during evolution. The plants *T. crispa* and *T. sinensis*, mainly found in tropical regions, are important folk medicinal species. *T. crispa* is used to lower blood glucose, whereas *T. sinensis* helps relax muscles and activate blood circulation. Studies have proven that relative medicinal plants exhibit diverse effects. For instance, *Panax ginseng* is known for enhancing immunity, while *Panax quinquefolius* is commonly used for alleviating exhaustion. Metabolic profiling revealed that most terpenoid DAMs were upregulated in *T. crispa* compared to *T. sinensis*. Notably, Borapetoside E, identified as a key compound contributing to the hypoglycemic effects of *T. crispa* [15], was 56 times more abundant in the leaves of *T. crispa* than in the leaves of *T. sinensis*, and the disparity in borapetoside E content between the stems was as high as 600 times in the two species. These differences are probably responsible for the differences in the medicinal efficacy between the two *Tinospora* species. To elucidate the genetic basis underlying the variations in terpenoids, the genomes of *T. crispa* and *T. sinensis* were generated with sizes of 962.85 Mb and 924.82 Mb, contig N50 of 7.42 Mb and 6.44 Mb, respectively. Detailed analysis revealed good genomic collinearity between the two *Tinospora* species, suggesting that the differences in terpenoid biosynthesis were not due to large structural variations. Similarly, 85.3% collinearity was reported between the genome of *Wurfbainia longiligularis* and *Wurfbainia villosa*; however, the content and distribution of bornyl diphosphate-related terpenoids were significantly different in these relatives [25]. The present study confirms that despite shared similar genomic structures among relatives, their metabolite types and contents could vary dramatically.

The functional diversification of duplicated genes, resulting from the expansion of gene families, has been identified as a significant factor driving phenotypic diversification among relatives [26]. However, most researches focused on the role of expanded genes in enhancing phenotypic traits and metabolic products. Li et al. demonstrated that the increased accumulation of diosgenin saponins in *Dioscorea zingiberensis* compared to *Dioscorea sansibarensis* was primarily attributed to the gene expansion of *CYP90B*, *CYP94*, and *CYP72A* [27]. Similarly, we found that the expanded genes enriched the "Isoflavonoid biosynthesis" pathway in *T. crispa*, consistent with the higher content of flavonoids in *T. crispa* compared to *T. sinensis*. Our study also revealed a distinct phenomenon related to terpenoid biosynthesis in *T. sinensis*. The genes related to terpenoid biosynthesis had expanded in *T. sinensis*, whereas the terpenoid content had decreased in *T. sinensis*. Further analysis of the terpenoid biosynthetic genes across 13 flowering species, showing the *OSCs* and *TPSs* in the TPS-e/f subfamily in *T. crispa*, which were responsible for the biosynthesis of triterpenoids and diterpenoid skeletons, respectively, exhibited higher gene expression compared to their orthologous in *T. sinensis*. These specifically expressed genes may explain the abundant triterpenoids and diterpenoids in *T. crispa*. Interestingly, the *HDS* gene family was found to have expanded in *T. sinensis*. The *HDS* gene encodes a rate-limiting enzyme in the MEP pathway and plays a crucial role in terpenoid biosynthesis [28, 29]. *TsiHDS4* exhibited strong collinearity with *HDSs* from other species, implying that *TsiHDS4* likely retained the ancestral functions. In contrast, other *TsiHDSs* exhibited structural loss. Specifically, TsiHDS1 and TsiHDS5 were identified as noncanonical CDS small peptides. Studies have reported that small peptides could act as "competitive repressors" competing with transcription factors and potentially forming non-functional heterodimers [21, 30]. The present study found that TsiHDS4 catalyzes the conversion of MEcPP to HMBPP. TsiHDS5 lacks catalytic activity but interacts with the A* domain of TsiHDS4, inhibiting its function. This interaction decreased the content of HMBPP and downstream terpenoid biosynthetic precursors IPP, potentially inhibiting monoterpenoid and diterpenoid biosynthesis. These findings highlight how the expansion of the *HDS* gene contributed to the reduced terpenoids from the MEP pathway in *T. sinensis*, providing novel insights into the genetic mechanisms of metabolic diversity in plants.

(See figure on next page.)

**Fig. 5** Functional characterization of *TsiHDS4* and *TsiHDS5*. **a** LCI assays between TsiHDS4 and TsiHDS5 in the leaves of *N. benthamiana*. The fluorescence presents the interaction signal. **b** Co-IP assays using the leaves of *N. benthamiana* expressing TsiHDS4-GFP and TsiHDS5-6FLAG. **c** LCI assays between 3 domains of TsiHDS4 (TsiHDS4, TsiHDS4A* and TsiHDS4B) and TsiHDS5 in the leaves of *N. benthamiana*, respectively. **d** The Concentration of the intermediates on the terpenoid biosynthesis after the heterologous expression of the empty plasmids pET28a, *TsiHDS4*-pET28a, and *TsiHDS5*-pET28a in *E. coli*. **e** The concentration of the intermediates on the terpenoid biosynthesis after the injection of EHA105 with PRI101-GFP, *TsiHDS4*-GFP, *TsiHDS5*-GFP, and co-inject EHA105 with *TsiHDS4*-GFP and *TsiHDS5*-GFP in *N. benthamiana*, respectively. Bars with different letters indicate significant differences ($P < 0.05$) according to Tukey's test

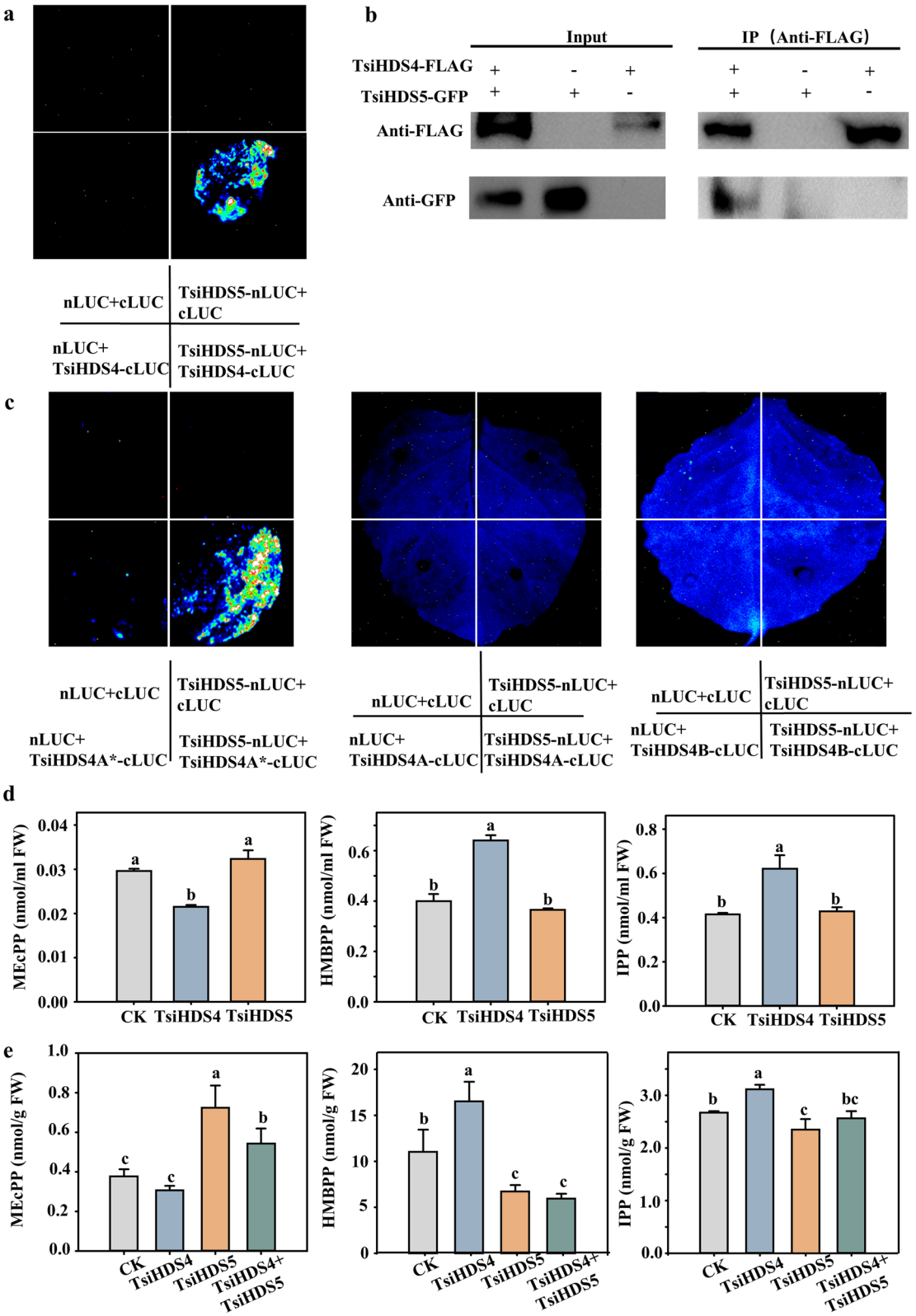


**Fig. 5** (See legend on previous page.)

## Conclusions

In summary, we generated two chromosome-level genomes of *T. crispa* and *T. sinensis* and characterized their genomic structures and phylogenetic relationship. Our study revealed that the expanded sORFs *TsiHDS5* negatively regulates the precursor of terpenoid biosynthesis in the MEP pathway by interacting with TsiHDS4. These findings offer novel insights into the relationship between gene expansion and metabolic regulation.

## Methods

### Plant materials and metabolite extraction

The leaves and stems of *T. crispa* and *T. sinensis* were collected at the Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Yunnan Province, China. For metabolomics analysis, approximately 50 mg of freeze-dried leaves and stems of *T. crispa* and *T. sinensis* were crushed and mixed with 700 µl of 75% methanol for extraction. The extract was vortexed for 30 s and sonicated for 5 min, followed by incubating overnight on a shaker at 4 ℃. The supernatant was collected, filtered, and used for liquid chromatography-tandem mass spectrometry (LC–MS/MS) analysis.

### Genome sequencing

Genomic DNA was extracted from young leaves using the QIAGEN Genomic kit following the manufacturer's instructions. The extracted DNA was used to construct an Illumina DNA library with an insertion size of 350 bp for paired-end sequencing on the Illumina Hiseq Xten platform. The Nanopore DNA library was prepared using LSK109 and sequenced on the Nanopore PromethION sequencer. The Hi-C library was constructed by fixing the fresh leaves, cross-linking the chromatin, biotinylating, and then sequenced on the Illumina NovaSeq platform.

### Genome size estimation

The raw Illumina sequencing data were filtered using Fastp v0.20.0 [31], with the parameters "-n 0 -w 6 -l 140." K-mer frequency analysis was performed on the clean Illumina reads using Jellyfish [32] with the parameter "count -m 21." Then, the genome size was estimated using Kmerfreq [33].

### Genome assembly

The raw Nanopore reads were corrected and used to construct preliminary assembly with NextDenovo v2.3.1 (https://github.com/Nextomics/NextDenovo) using default parameters. The corrected ONT reads, and the clean Illumina reads were used to perform three rounds of polishing to generate the final contig genome using Nextpolish v1.3.0 [34] with the parameters "-consensus -w window -t 4 -m 0.5 -d 30." Further, purge_dup v1.2.5 was used to remove the redundant. Finally, Benchmarking Universal Single-Copy Orthologs (BUSCO) [35] was used to assess the completeness of the contig genome. For Hi-C scaffolding, the clean Illumina reads were mapped to the contig assembly using bowtie2 [36] with the parameters "-end-to-end, -very-sensitive -L 30." And HiC-Pro was used to identify the valid interaction paired reads. Then, the contigs were further clustered, ordered, and oriented onto chromosomes by LACHESIS [37]. Finally, Juicebox [38] was used to manually adjust the orientation errors.

### Genome annotation

For repetitive elements, the tandem repeats in the genome were identified using TRF v4.09 [39] with default parameters, and the transposable elements were searched using RepeatMasker [40]. Gene prediction was conducted using a combination of transcriptome-based and de novo approaches. STAR v2.7.3a [41] was used to align the clean mRNA-seq reads to the genome, and transcript assembly was carried out using Stringtie v1.3.4d [42]. Gene structure prediction was performed using PASA v2.3.3 [43]. Augustus v3.2.1 [44] with the default parameters was used to obtain the training set and gene structure. Finally, EvidenceModeler [43] integrated the gene sets, in which low-quality genes were removed using TransposonPSI (http://transposonpsi.sourceforge.net/). Further, gene functional annotation was assigned according to BLASTP search against the NR, SwissProt, KEGG, and GO database (e value ≤ 1e-5). The non-coding RNAs (ncRNAs) in the genome were annotated using Infernal [45] by searching against the Rfam [46] database, and tRNAs and rRNAs were identified using tRNAscan-SE and RNAmmer [47], respectively.

### Transcriptome sequencing and data analysis

RNA was extracted from *T. crispa* and *T. sinensis* leaves and stems, maintaining three replicates per sample. The constructed cDNA libraries were sequenced on the Illumina HiSeq 2500 platform. The raw sequencing data were processed by removing the low quality reads, and the clean reads were generated using Fastp v0.20.0 [31]. The clean data were mapped to its reference genome using STAR v2.7.10a [41]. The transcripts per kilobase of the exon model per million mapped reads (TPM) values were calculated using RSEM v1.3.1 [48].

### Phylogenetic analyses

Phylogenetic analysis was carried out with early-diverging eudicots (*T. crispa*, *T. sinensis*, *Coptis chinensis* [49],

*Menispermum dauricum* [50], *Papaver somniferum* [51], *Epimedium pubescens* [52], *Nelumbo nucifera* subsp. *Nucifera* [53], *Tetracentron sinense* [54], and *Buxus austroyunnanensis* [55]), core eudicots (*Vitis vinifera* [56]), monocot (*Oryza sativa* [57]), magnoliids (*Aristolochia fimbriata* [58]), and using *Amborella trichopoda* [59] as the outgroup. Gene family clusters among these 13 species were identified using OrthoFinder [60] with the parameter "-M msa -S diamond." Furthermore, the amino acid sequences of the single-copy gene families were aligned using MAFFT [61] with default parameters, and the maximum likelihood tree was constructed using RAxML with 200 bootstrap replicates. The divergence time for the 13 species was estimated using the PAML package of MCMCtree [62] with the HKY85 model. The divergence time of the tree was calibrated using two fossil times from the TimeTree database (https://timetree.org/). Finally, the expansions and contractions in the gene families of these 13 species were identified using CAFE5 [63].

The syntenic blocks ("-icl") and Ks ("-ks") within and between species were determined, the dot plots of syntenic blocks ("-bk") as well as the Ks distribution peaks of paralogous pairs ("-kp") were used to infer whole genome duplication (WGD) events using WGDI pipeline [64].

#### Identification of genes involved in terpenoid biosynthesis

Local blast searches (e-value < $10^{-5}$) were performed against the genome of 13 flowering species using the genes involved in terpenoid biosynthesis in *A. thaliana* [65] to identify corresponding genes. The results were further confirmed using the website BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi). A phylogenetic tree was constructed using the MEGA7 [66] software following the maximum likelihood method. Gene structures were visualized using GSDS (Gene Structure Display Server 2.0 (gao-lab.org)).

#### Protein interaction analysis

LCI Assays were carried out to assess the interactions of the HDS protein. *TsiHDS4*, as well as three fragments of *TsiHDS4* based on the structural domains (TsiHDS4, TsiHDS4A*, and TsiHDS4B), were amplified and subcloned into the vector pCAMBIA1300-cLUC, while the CDS of *TsiHDS5* was subcloned into pCAMBIA1300-nLUC, respectively. These constructs were transferred into EHA105 and infiltrated into the *N. benthamiana* leaves as described "Transient expression in *Nicotiana benthamiana*." After four days of incubation, the luminescence intensity in the infiltrated leaves was measured using the Tanon 5200 automatic chemiluminescence image analysis system.

For the Co-IP assay, TsiHDS4-GFP and TsiHDS5-6FLAG vectors were cloned, constructed, and introduced into EHA105, followed by co-infiltrated into the leaves of *N. benthamiana*. After 4 days of infiltration, 0.18 g of the leaves were ground and mixed with 600 µl of IP-buffer (50 mM Tris–HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA 2Na, and 5% glycerin) containing protease inhibitor (1 mM cocktail, 1 mM DTT and 1mM PMSF). The extract was incubated on ice for 30 min, given ultrasonic treatment at 20% power for 14 times, and then centrifuged for 10 min. Subsequently, 30 µl of the anti-flag M2 affinity gel was added and shaken for 3 h at 4 °C to enrich the protein. Finally, the gel was washed thrice with IP buffer and analyzed using the anti-Flag and anti-GFP antibodies, respectively.

#### Heterologous expression in *Escherichia coli*

To characterize the enzyme activity, the full-length CDS sequences of *TsiHDS4* and *TsiHDS5* were constructed to pET28a vector after codon optimization, respectively. The resulting plasmids were transformed into BL21 (DE3), using the pET28a empty vector as a control to assess the enzyme activity. At an $OD_{600}$ of 0.6, the substrate MEcPP and 0.2 mM IPTG were added. The cultivation was incubated at 16 ℃ for 72 h. The metabolites were extracted as previously reported [67]. Briefly, a total of 10 ml of the cultures were centrifuged and re-suspended in 2 ml of extraction buffer (methanol: water = 4:1), sonicated for 10 min, incubated at − 20 °C for 20 min with shaking, and collected the supernatant. The extraction procedure was performed once more, the supernatant was pooled and concentrated to a final volume of 100 µl using nitrogen blowing. Finally, the extract was filtered and used for LC–MS/MS analysis.

MEcPP, HMBPP, and IPP were examined with a Waters I CLASS system coupled to a XEVO TQD mass spectrometer. The metabolites were separated with a ACQUITY UPLC BEH Amide column (2.1100 mm, 1.7 µm). Gradient elution was performed using water with 0.01% formic acid (A) and acetonitrile (B) by a gradient profile (*t*(min), %A, %B: (0, 5, 95), (1, 5, 95), (3, 30, 70), (6, 30, 70), (8, 50, 50), (9, 95, 5), (11, 95, 5), (11.5, 5, 95), (14, 5, 95)). The flow rate was kept at 0.2 ml/min and the injected volume was 5 µl. The column temperature was kept at 35 ℃. Mass spectra in negative ion mode. Samples were quantified using an external standard curve. The parameters for multiple-reaction monitoring (MRM) were as follows: the cone energy for MEcPP (276.96 > 78.9) was set at 132 V, for HMBPP (261.00 > 79.0) at 60 V, and for IPP (245.00 > 79.0) at 30 V. Correspondingly, the collision energies were set at 38 V for MEcPP, 36 V for HMBPP, and 23 V for IPP.

### Transient expression in *Nicotiana benthamiana*

The cDNA of *TsiHDS4* and *TsiHDS5* was cloned into the PRI101-GFP vector, respectively. The obtained recombinant plasmids were transformed into the *Agrobacterium tumefaciens* strain EHA105 by electroporation. The transient expression assays were conducted as previously described [68]. Briefly, the positive *Agrobacterium* transformants were incubated in LB solid medium at 28 ℃ for two days, then centrifuged, resuspended, and mixed to an $OD_{600}$ of 0.6 using infiltration buffer (10 mM MES, 10 mM $MgCl_2$, 100 μM acetosyringone, pH 5.6). This mixture was then infiltrated into the leaves of five-week-old *N. benthamiana* leaves with a needleless syringe. The leaves were harvested four days after infiltration.

Metabolite extraction was performed and optimized as previously described [69]. A 0.5 g plant sample was harvested and grounded in liquid nitrogen, followed by extraction using 3 ml of 50% acetonitrile adjusted to pH 5.5 with acetic acid twice. The supernatant was collected and dried using nitrogen blowing. The extract was then reconstituted in 200 μl of 50% acetonitrile (pH = 5.5) and filtered for LC–MS analysis. The examination of MEP metabolites was similar to that described in "Heterologous expression in *Escherichia coli*," but had a different gradient profile (t(min), %A, %B: (0, 98, 2), (1, 90, 10), (3, 50, 50), (5, 5, 95), (6, 5, 95), (7, 98, 2), (8, 98, 2)), and the flow rate was at 0.3 ml/min.

#### Abbreviations

| | |
|---|---|
| HDS | 1-Hydroxy-2-methyl 2-(E)-butenyl 4-diphosphate synthase |
| MEcPP | Methylerythritol cyclic diphosphate |
| HMBPP | (E)-4-hydroxy-3-methylbut-2-enyl diphosphate |
| IPP | Isopentenyl pyrophosphate |
| CNVs | Copy number variations |
| SNPs | Single nucleotide polymorphisms |
| Indels | Insertions or deletions |
| PCA | Principal component analysis |
| DAMs | Differentially accumulated metabolites |
| TE | Transposable elements |
| LTR | Long terminal repeats |
| Mya | Million years ago |
| Ks | Synonymous nucleotide substitutions |
| WGD | Whole genome duplication |
| AEK | Ancestral eudicot karyotype |
| MEP | 2-C-methyl-D-erythritol 4-phosphate |
| MVA | Mevalonate |
| sORFs | Short open reading frames |
| LCI | Luciferase complementation imaging |
| CYP | Cytochrome monooxygenase |

## Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12915-025-02185-z.

Additional file 1: Fig. S1. Volcano plot of the number of DAMs between stems of *T. crispa* compared to those of *T. sinensis*. Fig. S2. The number of up- and down-regulated differentially accumulated metabolites (DAMs) in the leaves of *T. crispa* compared to those of *T. sinensis*. Fig. S3. The number of up- and down-regulated differentially accumulated metabolites (DAMs) in the stems of *T. crispa* compared to those of *T. sinensis*. Fig. S4. Fold statistics of the content of up-regulated and down-regulated DAMs in the stems of *T. crispa* compared to those of *T. sinensis*. Fig. S5. The heatmap of the Hi-C assisted assembly of *T. crispa* and *T. sinensis*. Fig. S6. Structural variation detection between *T. crispa* and *T. sinensis* genome. Fig. S7. The number of gene family categories in *T. crispa*, *T. sinensis* and 11 other species. Fig. S8. Dot plots syntenic blocks between *A. fimbriata* and two *Tinospora* genome both shows 1:2 chromosomal relationship. Fig. S9. Dot plots syntenic blocks between *C. chinensis* and two *Tinospora* genome both shows 1:1 chromosomal relationship. Fig. S10. KEGG pathway enrichment analysis of expanded gene families in *T. crispa*. Fig. S11. Categories and numbers of duplication genes in 13 flowering species. Fig. S12. Heatmap of gene expression in terpenoid biosynthesis pathway in *T. crispa* and *T. sinensis*. Fig. S13. Phylogenetic tree of HDSs from 13 flowering species. Fig. S14. Syntenic blocks in the chromosome regions containing *HDS* genes of *T. crispa* and *T. sinensis*. Fig. S15. Phylogenetic tree and gene structure of *HDS* in *T. crispa* and *T. sinensis*, including exon and intron. Fig. S16. Phylogenetic tree and pfam structure of HDS in *T. crispa* and *T. sinensis*, including GcpE ng GcpE superfamily. Fig. S17. Phylogenetic tree and motif of *HDS* in *T. crispa* and *T. sinensis*. Fig. S18. Analysis of the interaction between TsiHDS4 and TsiHDS5 by Co-IP assays. Fig. S19. LC-MS spectroscopy illustrate representative compound peaks for the target products of TsiHDS4 and TsiHDS5 in *Escherichia coli*. Fig. S20. LC-MS spectroscopy illustrate representative compound peaks for the target products of TsiHDS4 and TsiHDS5 in *Nicotiana benthamiana*.

Additional file 2: Table S1. The genome sequencing data of *T. crispa* and *T. sinensis*. Table S2. Summary of genome assembly of *T. crispa* and *T. sinensis*. Table S3. Genome BUSCO result for *T. crispa* and *T. sinensis*. Table S4. Whole genome mapping rate of NGS data in *T. crispa* and *T. sinensis*. Table S5. Statistics of the pseudochromosomes length in *T. crispa* and *T. sinensis* genome. Table S6. Statistics of transposable elements in *T. crispa* and *T. sinensis* genome. Table S7. RNA-seq statistics of *T. crispa* and *T. sinensis*. Table S8. Protein BUSCO result for *T. crispa* and *T. sinensis*. Table S9. Gene functional annotations of *T. crispa* and *T. sinensis*. Table S10. Number of non-coding RNA in *T. crispa* and *T. sinensis* genome.

#### Acknowledgements

We are sincerely grateful to Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences for generously providing research plant material.

#### Authors' contribution

Y.P.Y., Y.Q.Y. and X.F.Y designed studies, Z.Y.C., L.X., Y.Y.L, and M.H.C., conducted the experiments; Z.Y.C., X.Y.Y., and Y.T.Y. carried out the bioinformatics analysis; Z.Y.C., Y.Q.Y., and Z.N.M. wrote and revised the manuscript. All authors commented on the article before submission. All authors read and approved the final manuscript.

#### Funding

This study was supported by Yunling Scholar Project to Yang Yongping, Yunnan Revitalization Talents Support Plan to Yang Yunqiang, the Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences (Grant Nos. 2015CASEABRIRG001), the Yunnan Province Science and Technology Department (Grant No. 202203AP140007), trans-boundary cooperation on biodiversity research and conservation in Gaoligong Mountains (Grant No. E1ZK251).

#### Data availability

The sequencing and assembly data during the current study are available in National Genomics Data Center with accession number PRJCA027609 [70] and PRJCA034495 [71]. The genome assembly and annotation also could be obtained from Figshare [72] (https://doi.org/10.6084/m9.figshare.26877709.v1).

## Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

**Competing interests**
The authors declare no competing interests.

**Author details**
[1]Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China. [2]Key Laboratory of Economic Plants and Biotechnology, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China. [3]Yunnan International Joint Laboratory for the Conservation and Utilization of Tropical Timber Tree Species, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan 666303, China. [4]CAS Key Laboratory of Tropical Plant Resources and Sustainable Use, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan 666303, China. [5]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China. [6]Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences, Yezin, Nay Pyi Taw 05282, Myanmar. [7]College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming 650201, Yunnan, China. [8]College of Plant Science and Technology, Huazhong Agricultural University, Wuhan 430070, China.

Received: 24 October 2024 Accepted: 4 March 2025
Published online: 20 March 2025

**References**

1. Mitchell-Olds T, Willis JH, Goldstein DB. Which evolutionary processes influence natural genetic variation for phenotypic traits? Nat Rev Genet. 2007;8(11):845–56.
2. Lin RC, Ferreira BT, Yuan YW. The molecular basis of phenotypic evolution: beyond the usual suspects. Trends Genet. 2024;40(8):668–80.
3. Wang Y, Zhang H, Ri HC, An Z, Wang X, Zhou JN, et al. Deletion and tandem duplications of biosynthetic genes drive the diversity of triterpenoids in *Aralia elata*. Nat Commun. 2022;13(1):2224.
4. Anzano A, Bonanomi G, Mazzoleni S, Lanzotti V. Plant metabolomics in biotic and abiotic stress: a critical overview. Phytochem Rev. 2021;21(2):503–24.
5. Erwin DH. Evolutionary dynamics of gene regulation. Curr Top Dev Biol. 2020;139:407–31.
6. Nowick K, Stubbs L. Lineage-specific transcription factors and the evolution of gene regulatory networks. Brief Funct Genomics. 2010;9(1):65–78.
7. Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet. 2010;11(2):97–108.
8. Kuzmin E, Taylor JS, Boone C. Retention of duplicated genes in evolution. Trends Genet. 2022;38(1):59–72.
9. Wang YJ, Tain T, Yu JY, Li J, Xu B, Chen J, et al. Genomic and structural basis for evolution of tropane alkaloid biosynthesis. Proc Natl Acad Sci U S A. 2023;120(17): e2302448120.
10. Li P, Yan MX, Liu P, Yang DJ, He ZK, Gao Y, et al. Multiomics analyses of two *Leonurus* species illuminate leonurine biosynthesis and its evolution. Mol Plant. 2024;17(1):158–77.
11. Yin X, Yang D, Zhao Y, Yang X, Zhou Z, Sun X, et al. Differences in pseudogene evolution contributed to the contrasting flavors of turnip and Chiifu, two *Brassica rapa* subspecies. Plant Commun. 2023;4(1): 100427.
12. Chen H, Guo M, Dong S, Wu X, Zhang G, He L, et al. A chromosome-scale genome assembly of *Artemisia argyi* reveals unbiased subgenome evolution and key contributions of gene duplication to volatile terpenoid diversity. Plant Commun. 2023;4(3): 100516.
13. Liao B, Shen X, Xiang L, Guo S, Chen S, Meng Y, et al. Allele-aware chromosome-level genome assembly of *Artemisia annua* reveals the correlation between *ADS* expansion and artemisinin yield. Mol Plant. 2022;15(8):1310–28.
14. Koay YC, Koay F. A review of the secondary metabolites and biological activities of *Tinospora crispa* (Menispermaceae). Trop J Pharm Res. 2013;12(4):641–9.
15. Xu Y, Niu Y, Gao Y, Wang F, Qin W, Lu Y, et al. Borapetoside E, a clerodane diterpenoid extracted from *Tinospora crispa*, improves hyperglycemia and hyperlipidemia in high-fat-diet-induced type 2 diabetes mice. J Nat Prod. 2017;80(8):2319–27.
16. Ruan CT, Lam SH, Chi TC, Lee SS, Su MJ. Borapetoside C from *Tinospora crispa* improves insulin sensitivity in diabetic mice. Phytomedicine. 2012;19(8–9):719–24.
17. Ruan CT, Lam SH, Lee SS, Su MJ. Hypoglycemic action of borapetoside A from the plant *Tinospora crispa* in mice. Phytomedicine. 2013;20(8–9):667–75.
18. Jiang H, Zhang GJ, Liao HB, Liang D. New terpenoid and phenylpropanoid glycosides from *Tinospora sinensis*. Fitoterapia. 2018;131:127–33.
19. Thakur RK, Rajpal VR, Rao SR, Singh A, Joshi L, Kaushal P, et al. Induction and evaluation of colchitetraploids of two species of *Tinospora* Miers, 1851. Comp Cytogenet. 2020;14(2):211–29.
20. Xiao Y, Zahariou G, Sanakis Y, Liu P. IspG enzyme activity in the deoxyxylulose phosphate pathway: roles of the iron-sulfur cluster. Biochemistry. 2009;48(44):10483–5.
21. Feng YZ, Zhu QF, Xue PY, Yang Y. Shining in the dark: the big world of small peptides in plants. aBIOTECH. 2023;4(3):238–56.
22. Wang J, Guo Y, Yin X, Wang X, Qi X, Xue Z. Diverse triterpene skeletons are derived from the expansion and divergent evolution of 2,3-oxidosqualene cyclases in plants. Crit Rev Biochem Mol Biol. 2021;57(2):113–32.
23. Gao Y, Honzatko RB, Peters RJ. Terpenoid synthase structures: a so far incomplete view of complex catalysis. Nat Prod Rep. 2012;29(10):1153–75.
24. Liu YL, Guerra F, Wang K, Wang W, Li J, Huang C, et al. Structure, function and inhibition of the two- and three-domain 4Fe-4S IspG proteins. Proc Natl Acad Sci U S A. 2012;109(22):8558–63.
25. Yang P, Ling XY, Zhou XF, Chen YX, Wang TT, Lin XJ, et al. Comparing genomes of *Fructus Amomi*-producing species reveals genetic basis of volatile terpenoid divergence. Plant Physiol. 2023;193(2):1244–62.
26. Guo YL. Gene family evolution in green plants with emphasis on the origination and evolution of *Arabidopsis thaliana* genes. Plant J. 2013;73(6):941–51.
27. Li Y, Tan C, Li Z, Guo J, Li S, Chen X, et al. The genome of *Dioscorea zingiberensis* sheds light on the biosynthesis, origin and evolution of the medicinally important diosgenin saponins. Hortic Res. 2022;9:uhac165.
28. Adedeji D, Hernandez H, Wiesner J, Köhler U, Jomaa H, Duin EC. Possible direct involvement of the active-site [4Fe–4S] cluster of the GcpE enzyme from *Thermus thermophilus* in the conversion of MEcPP. FEBS Lett. 2006;581(2):279–83.
29. Ma Y, McClure DD, Somerville MV, Proschogo NW, Dehghani F, Kavanagh JM, et al. Metabolic engineering of the MEP pathway in *Bacillus subtilis* for increased biosynthesis of menaquinone-7. ACS Synth Biol. 2019;8(7):1620–30.
30. Makarewich CA, Olson EN. Mining for micropeptides. Trends Cell Biol. 2017;27(9):685–96.
31. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884–90.
32. Marcais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27(6):764–70.
33. Liu B, Shi Y, Yuan J, Hu X, Zhang H, Li N, et al. Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. Quant Biol. 2013;35(s1–3):62–7.
34. Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics. 2020;36(7):2253–5.
35. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2.
36. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–9.
37. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat Biotechnol. 2013;31(12):1119.
38. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox provides a visualization system for Hi-C Contact maps with unlimited zoom. Cell Syst. 2016;3(1):99–101.
39. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80.
40. Bedell JIW. MaskerAid: a performance enhancement to RepeatMasker. Bioinformatics. 2000;16(11):1040.
41. Dobin A, Davis CA, Schlesinger F, Drenkow J, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2012;29(1):15.

42. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 2019;20(1):15–21.
43. Haas BJ, Salzberg SL, Zhu W, Pertea M. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. 2008;9(1): R7.
44. Stanke M, Diekhans M, Baertsch R, Haussler D, et al. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008;24(5):637–44.
45. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29(22):2933–5.
46. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Bateman A. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. 2005;33(SI):D121-4.
47. Lagesen K, Hallin P, Andreas E, et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35(9):3100–8.
48. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12(1):93–9.
49. Liu Y, Wang B, Shu S, Li Z, Song C, Liu D, et al. Analysis of the *Coptis chinensis* genome reveals the diversification of protoberberine-type alkaloids. Nat Commun. 2021;12(1):3276.
50. An Z, Gao R, Chen S, Tian Y, Li Q, Tian L, et al. Lineage-specific *CYP80* expansion and benzylisoquinoline alkaloid diversity in early-diverging eudicots. Adv Sci. 2024;11(19): e2309990.
51. Yang X, Gao S, Guo L, Wang B, Jia Y, Zhou J, et al. Three chromosome-scale *Papaver* genomes reveal punctuated patchwork evolution of the morphinan and noscapine biosynthesis pathway. Nat Commun. 2021;12(1):6030.
52. Shen G, Luo Y, Yao Y, Meng G, Zhang Y, Wang Y, et al. The discovery of a key prenyltransferase gene assisted by a chromosome-level *Epimedium pubescens* genome. Front Plant Sci. 2022;13: 1034943.
53. Zheng X, Wang T, Cheng T, Zhao L, Zheng X, Zhu F, et al. Genomic variation reveals demographic history and biological adaptation of the ancient relictual, lotus (Nelumbo Adans). Hortic Res. 2022;9:uhac029.
54. Chanderbali AS, Jin L, Xu Q, Zhang Y, Zhang J, Jian S, et al. *Buxus* and *Tetracentron* genomes help resolve eudicot genome history. Nat Commun. 2022;13(1):643.
55. Wang Z, Li Y, Sun P, Zhu M, Wang D, Lu Z, et al. A high-quality *Buxus austro-yunnanensis* (Buxales) genome provides new insights into karyotype evolution in early eudicots. BMC Biol. 2022;20(1):216.
56. Massonnet M, Cochetel N, Minio A, Vondras AM, Lin J, Muyle A, et al. The genetic basis of sex determination in grapes. Nat Commun. 2020;11(1):2902.
57. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, Childs KL. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice. 2013;6:4.
58. Qin L, Hu Y, Wang J, Wang X, Zhao R, Shan H, et al. Insights into angiosperm evolution, floral development and chemical biosynthesis from the *Aristolochia fimbriata* genome. Nat Plants. 2021;7(9):1239–53.
59. Amborella GP. The *Amborella* genome and the evolution of flowering plants. Science. 2013;342(6165):1241089.
60. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16(1):157.
61. Rozewicki J, Li S, Amada KM, Standley DM, Katoh K. MAFFT-DASH: integrated protein sequence and structural alignment. Nucleic Acids Res. 2019;47(W1):W5–10.
62. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.
63. Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics. 2020;36(22–23):5516–8.
64. Sun P, Jiao B, Yang Y, Shan L, Li T, Li X, et al. WGDI: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. Mol Plant. 2022;15(12):1841–51.
65. Vranova E, Coman D, Gruissem W. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. Annu Rev Plant Biol. 2013;64:665–700.
66. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4.
67. Li Q, Fan F, Gao X, Yang C, Bi C, Tang J, et al. Balanced activation of IspG and IspH to eliminate MEP intermediate accumulation and improve isoprenoids production in *Escherichia coli*. Metab Eng. 2017;44:13–21.
68. Yin X, Yang Y, Lv Y, Li Y, Yang D, Yue Y, et al. BrrICE1.1 is associated with putrescine synthesis through regulation of the arginine decarboxylase gene in freezing tolerance of turnip (Brassica rapa var. rapa). BMC Plant Biol. 2020;20(1):504.
69. Xiao Y, Savchenko T, Baidoo EE, Chehab WE, Hayden DM, Tolstikov V, et al. Retrograde signaling by the plastidial metabolite MEcPP regulates expression of nuclear stress-response genes. Cell. 2012;149(7):1525–35.
70. Chen Z, Xun L, Lu Y, Yang X, Chen M, Yang T, et al. The chromosome-scale genomes of two *Tinospora* species reveal differential regulation of the MEP pathway in terpenoid biosynthesis. National Genomics Data Center; 2025. https://ngdc.cncb.ac.cn/gsa/search?searchTerm=PRJCA027609 .
71. Chen Z, Xun L, Lu Y, Yang X, Chen M, Yang T, et al. The chromosome-scale genomes of two *Tinospora* species reveal differential regulation of the MEP pathway in terpenoid biosynthesis. National Genomics Data Center; 2025. https://ngdc.cncb.ac.cn/gwh/search/advanced/result?search_category=&search_term=&source=0&query_box=PRJCA034495.
72. Chen Z, Xun L, Lu Y, Yang X, Chen M, Yang T, et al. The chromosome-scale genomes of two Tinospora species reveal differential regulation of the MEP pathway in terpenoid biosynthesis. Figshare. 2024. https://doi.org/10.6084/m9.figshare.26877709.v1.


## Publisher's Note