



Integrative omics reveals mechanisms of biosynthesis and regulation of floral scent in *Cymbidium tracyanum*

Mengling Tu^{1,2,†}, Ningyawen Liu^{2,3,†}, Zheng-Shan He^{4,†}, Xiu-Mei Dong¹, Tian-Yang Gao^{1,2}, Andan Zhu^{4,*} , Jun-Bo Yang^{4,*} and Shi-Bao Zhang^{1,*} 

¹Key Laboratory for Economic Plants and Biotechnology, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China

²University of Chinese Academy of Sciences, Beijing, China

³National Key Laboratory of Genetic Evolution & Animal Models, Kunming Natural History Museum of Zoology, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China

⁴Germplasm Bank of Wild Species & Yunnan Key Laboratory of Crop Wild Relatives Omics, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, China

Received 1 October 2024;

revised 20 January 2025;

accepted 10 February 2025.

*Correspondence (Tel/fax +86 871

65238370; email zhuandan@mail.kib.ac.cn

(A.Z.); Tel/fax +86 871 65223139;

email jbyang@mail.kib.ac.cn (J.-B.Y.); Tel/fax

+86 871 65223002; email sbzhzhang@mail.kib.ac.cn

(S.-B.Z.))

[†]These authors contributed equally to this work.

Summary

Flower scent is a crucial determiner in pollinator attraction and a significant horticultural trait in ornamental plants. Orchids, which have long been of interest in evolutionary biology and horticulture, exhibit remarkable diversity in floral scent type and intensity. However, the mechanisms underlying floral scent biosynthesis and regulation in orchids remain largely unexplored. In this study, we focus on floral scent in *Cymbidium tracyanum*, a wild species known for its strong floral fragrance and as a primary breeding parent of commercial *Cymbidium* hybrids. We present a chromosome-level genome assembly of *C. tracyanum*, totaling 3.79 Gb in size. Comparative genomic analyses reveal significant expansion of gene families associated with terpenoid biosynthesis and related metabolic pathways in *C. tracyanum*. Integrative analysis of genomic, volatilomic and transcriptomic data identified terpenoids as the predominant volatile components in the flowers of *C. tracyanum*. We characterized the spatiotemporal patterns of these volatiles and identified *CtTPS* genes responsible for volatile terpenoid biosynthesis, validating their catalytic functions *in vitro*. Dual-luciferase reporter assays, yeast one-hybrid assays and EMSA experiments confirmed that *CtTPS2*, *CtTPS3*, and *CtTPS8* could be activated by various transcription factors (i.e., *CtAP2/ERF1*, *CtbZIP1*, *CtMYB2*, *CtMYB3* and *CtAP2/ERF4*), thereby regulating the production of corresponding monoterpenes and sesquiterpenes. Our study elucidates the biosynthetic and regulatory mechanisms of floral scent in *C. tracyanum*, which is of great significance for the breeding of fragrant *Cymbidium* varieties and understanding the ecological adaptability of orchids. This study also highlights the importance of integrating multi-omics data in deciphering key horticultural traits in orchids.

Keywords: orchid, *cymbidium*

tracyanum, multi-omics, floral scent, terpene synthase.

Introduction

Orchidaceae is one of the most diverse families of angiosperms, comprising nearly 800 genera and over 29 000 species worldwide (Christenhusz and Byng, 2016; Govaerts *et al.*, 2021), many of which are of considerable ornamental and medicinal value. The floral organs of orchids have evolved remarkably diverse morphology, colour and scent, enabling deceptive pollination strategies in nearly one-third of orchid species (Cozzolino and Widmer, 2005; Jersáková *et al.*, 2006). Floral scent is a key horticultural trait for ornamental plants and affects plant adaptation to environmental disturbances, including defence against pathogens and folivores (Holopainen and Gershenzon, 2010; Huang *et al.*, 2011), pollinator attraction (Byers *et al.*, 2014) and mediating plant-environment interactions (Heil and Silva Bueno, 2007; Kegge and Pierik, 2010). These complex interactions between environmental disturbances and diversified floral features make orchids invaluable for research in evolutionary biology and horticulture (Pérez-Escobar *et al.*, 2024; Waterman and Bidartondo, 2008).

Floral scent is composed of various volatile organic compounds (VOCs): terpenoids, phenylpropanoids/benzenoids and fatty acid derivatives, based on their biosynthetic origins (Dudareva *et al.*, 2013). The most diverse and largest proportion of VOCs is the terpenoids, including terpenes and their modified forms (Chen *et al.*, 2011; Tholl, 2015). Terpenoids are synthesized from two common 5-carbon precursors—*isopentenyl diphosphate* (IPP) and its allylic isomer, *dimethylallyl diphosphate* (DMAPP) (McGarvey and Croteau, 1995). These C5-isoprene units are generated from two independent and compartmentalized pathways. The methylerythritol phosphate (MEP) pathway, which occurs in plastids, produces both IPP and DMAPP. The mevalonic acid (MVA) pathway, which occurs in the cytosol, endoplasmic reticulum and peroxisomes, primarily produces IPP. Prenyltransferases use IPP and DMAPP as substrates to catalyse head-to-tail coupling condensation in different proportions to yield prenyl diphosphate precursors—*farnesyl diphosphate* (FPP) in the cytosol, *geranyl diphosphate* (GPP) and *geranylgeranyl diphosphate* (GGPP) in plastids.

Terpenoid precursors are converted into the basic carbon skeletons of monoterpenes, sesquiterpenes and diterpenes by

terpene synthases (TPSs) (Chen *et al.*, 2011; Jia *et al.*, 2022). Because many TPSs are multifunctional and located at key branching points in the isoprenoid pathway, they produce a variety of main products and by-products, contributing significantly to the structural diversity of terpenoids (Jia *et al.*, 2022). Recent biochemical and molecular studies of TPSs have provided significant insights into their evolutionary, structural and mechanistic properties, as well as their regulatory mechanisms. However, functional identification of *TPS* genes in orchids remains limited (Chuang *et al.*, 2018; Yu *et al.*, 2020).

Synthesis and release of terpenoids in plants is spatially (tissues and organs) and temporally specific (circadian rhythm and developmental stages) (Dudareva *et al.*, 2003; Jørgensen *et al.*, 2005). Previous studies have shown that higher expression levels of terpenoid biosynthesis-related genes often correlate with the primary sites of floral scent production (Bergougnoux *et al.*, 2007; Nagegowda *et al.*, 2008), indicating that the synthesis of floral scent components relies on key transcription factors (TFs) that determine the spatiotemporal expression of enzyme genes (Colquhoun and Clark, 2011). To date, at least six families of TFs have been implicated in the regulation of terpenoid biosynthesis: AP2/ERF, bHLH, bZIP, ARF, MYB and WRKY (Hong *et al.*, 2012; Samad *et al.*, 2017; Tan *et al.*, 2015). For instance, studies have shown that members of the bHLH family regulate sesquiterpene production by directly binding to promoters of the sesquiterpene synthase genes *AtTPS11* and *AtTPS21* (Hong *et al.*, 2012). In addition, TPS activity in terpenoid biosynthesis is differentially regulated at the post-transcriptional or translational levels (Picazo-Aragón *et al.*, 2020).

Floral VOCs have been characterized (Xu *et al.*, 2017) and *TPS* genes have been identified in various orchid species, for example, *Apostasia shenzhenica* (Huang *et al.*, 2021; Yang *et al.*, 2021; Zhang *et al.*, 2016), *Phalaenopsis equestris* (Huang *et al.*, 2021; Tsai *et al.*, 2017; Yang *et al.*, 2021; Zhang *et al.*, 2016), *Vanilla planifolia* (Huang *et al.*, 2021) and *Dendrobium officinale* (Huang *et al.*, 2021; Tsai *et al.*, 2017; Yang *et al.*, 2021; Zhang *et al.*, 2016). In *Cymbidium*, the dominant floral volatiles have been shown to vary between species. For example, the dominant floral volatiles include farnesol, methyl epi-jasmonate, (*E*)- β -farnesene and nerolidol in *C. goeringii* (Ramya *et al.*, 2019), whereas they are methyl jasmonate, acacia alcohol and linalool in *C. ensifolium* (Ai *et al.*, 2021). The remarkable diversity in both the type and intensity of fragrance across *Cymbidium* species makes this genus ideal for examining the evolution of molecular mechanisms that regulate floral scent.

Cymbidium are world-renowned ornamental orchids, with terrestrial species (commonly known as Chinese orchids) cultivated in China for over 1000 years. These orchids are highly prized for their fragrant flowers, diverse flower shapes and elegant plant forms in Asian countries (Liu *et al.*, 2006). Over the past century, hybrids derived from epiphytic *Cymbidium* species have dominated the global flower market, but artificial selection for other traits has often resulted in a loss of fragrance, even from aromatic parents (Dudareva and Pichersky, 2008; Ramya *et al.*, 2019). Importantly, *Cymbidium* includes scentless orchids (e.g., *C. lowianum*), which enable comparative analysis between orchid species to identify VOCs, synthetic enzymes and differentially expressed TFs that may play a role in scent biosynthesis. Furthermore, several *Cymbidium* orchid genomes have recently been sequenced, which facilitate the identification of related genomic changes.

In this study, we presented a chromosome-level genome assembly of *C. tracyanum*, a species characterized by strongly fragrant flowers with more than 10 flowers per scape, making it a key parent in hybrid breeding. Comparative genome analysis with other orchids revealed a significant expansion of gene families associated with terpenoid biosynthesis. Furthermore, through comparative transcriptomic and metabolomic analyses between the fragrant *C. tracyanum* and the scentless *C. lowianum*, we elucidated the biosynthetic pathways and transcriptional regulation of floral scent. We identified, isolated, cloned and functionally analysed *TPS* genes in *C. tracyanum*, uncovering the potential regulatory network of terpenoid biosynthesis. These findings shed light on the breeding of fragrant varieties and reveal the evolution and adaptation of orchids.

Results

Chromosome-level genome assembly and annotation of *C. tracyanum*

To elucidate the genetic mechanisms underlying floral scent, we first assembled the genome of *C. tracyanum* at the chromosome scale, laying a foundation for our research. The *C. tracyanum* genome assembly spans 3.79 Gb, comprising 16 288 contigs with a contig N50 of 1.66 Mb (Tables S1 and S2). Using chromatin interaction signals from Hi-C data, 20 pseudochromosomes were constructed, with the longest pseudochromosome measuring 220.09 Mb and the shortest 106.15 Mb (Figures 1a, S1 and Table S2). The assembled genome size aligned with size estimates based on flow cytometry (3.95 Gb) and *k*-mer frequency distribution (3.86 Gb) (Figure S2). Approximately 88.57% of the *C. tracyanum* genome consisted of repetitive elements, a proportion higher than that observed in currently sequenced orchids, such as *C. mannii* (82.8%) and *D. nobile* (61.07%) (Figure 1b). In *C. tracyanum*, most of the repetitive elements were transposable elements (TEs) (Table S3), with Long Terminal Repeats (LTRs) accounting for 60.27% of the genome, representing the largest proportion. Of these LTRs, *Gypsy* retrotransposons were the most prevalent, followed by *Copia* elements, accounting for 23.85% and 3.83% of the genome, respectively (Figures 1c and S3).

A total of 42 249 protein-coding genes were annotated in the *C. tracyanum* genome with an average gene length of 7871 bp, coding sequence (CDS) length of 823 bp and an average of 4.13 exons per gene (Table S4). TrEMBL analysis revealed that 32 855 (77.77%) genes were functionally annotated, 24 735 of which encoded metabolic enzymes according to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Table S5). The annotation quality was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO). A total of 1570 complete gene models out of 1614 (97.3%) were recovered, including 92.6% single-copy genes and 4.7% duplicates, indicating high annotation quality and completeness (Table S6).

Whole-genome duplication (WGD) events are prevalent among angiosperms and have significantly influenced agronomic or specialized phenotypic traits (Van de Peer *et al.*, 2009). The distribution of synonymous substitutions per synonymous site (*K*s) of all paralogs in the *C. tracyanum* genome showed two distinct peaks in *K*s values (~ 1.0 and ~ 1.9) (Figure 1d), indicating two separate WGD events. The most recent WGD, inferred from the *K*s peak of approximately 1.0 and consistent with other genome-sequenced orchids, likely occurred before the divergence

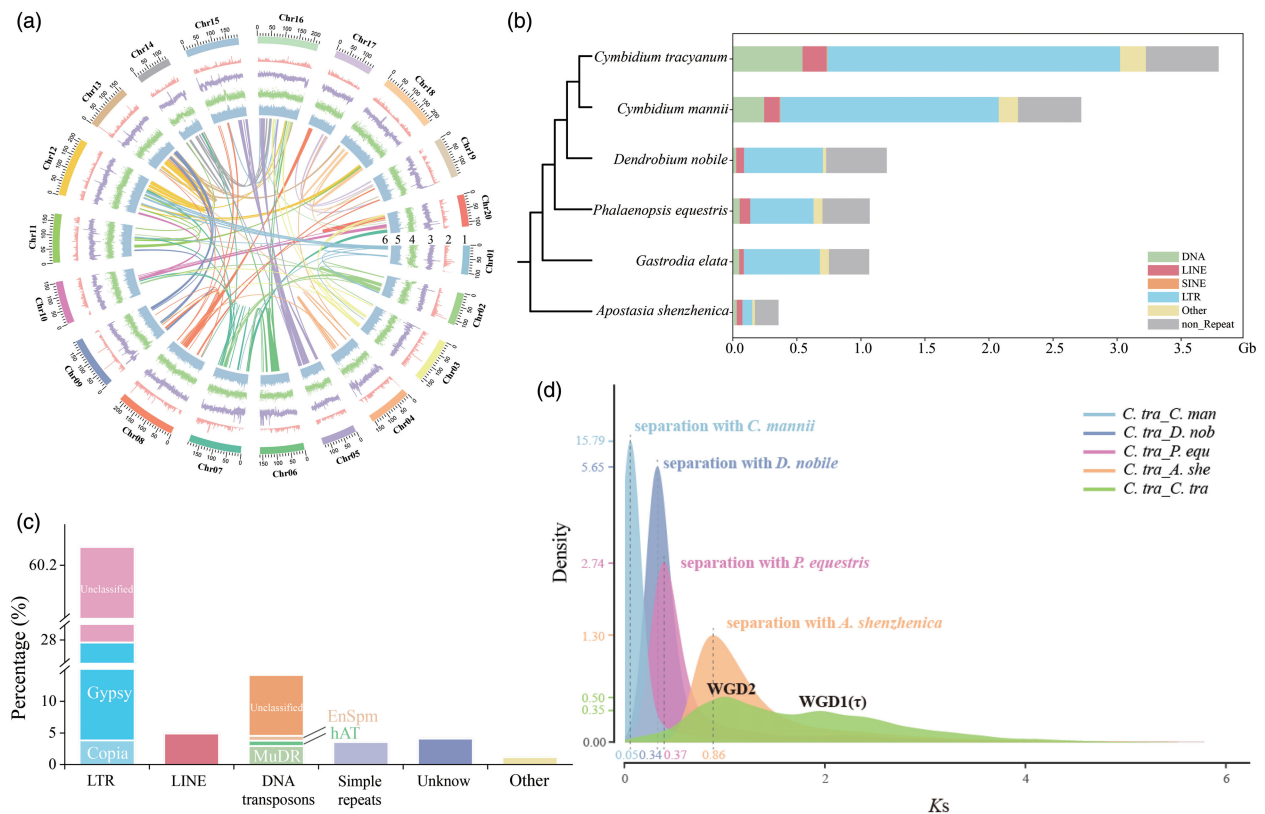


Figure 1 Genome features of *Cymbidium tracyanum* and the evolutionary history of whole-genome duplication events. (a) Genomic features of *C. tracyanum*. The outer to inner circles represent: (1) chromosome length, with tick marks at 50 Mb intervals; (2–5) densities of genes, GC content, transposable elements (TEs) and long terminal repeats (LTRs), respectively, all plotted using a 100-kb sliding window; (6) homologous regions among chromosomes, displayed with connecting lines colour-coded according to chromosome number. (b) Comparison of repetitive sequences in orchids. (c) Relative contents of different repetitive element classes, with types constituting less than 0.5% of the genome not shown. (d) K_s distribution and whole-genome duplication (WGD) events in *C. tracyanum*. The K_s distribution reveals two peaks: one at approximately 1.0 (WGD2), indicating the most recent WGD event shared by all orchids, and another at approximately 1.9 (WGD1), likely corresponding to the more ancient τ event shared by most monocots.

between *C. tracyanum* and *A. shenzhenica*, suggesting that this WGD event was shared by all extant orchids (Cai *et al.*, 2014; Fan *et al.*, 2023; Zhang *et al.*, 2017; Xu *et al.*, 2022). Intragenomic collinearity analyses revealed that several chromosomal regions in *C. tracyanum* had one other syntenic region in the genome attributable to the orchid-specific WGD, while some chromosomes, such as Chr08, showed up to three homologous syntenic regions, providing evidence for an even more ancient WGD (Figure S4), likely the more ancient τ event shared by most monocots (Jiao *et al.*, 2014). Comparison with the chromosome-level assembled genomes of *C. mannii* and *D. nobile* revealed excellent one-to-one correspondence (Figure S4), supporting the hypothesis that these species share the two WGD events observed in *C. tracyanum*.

Phylogenomic analyses of *C. tracyanum*

Orchids represent one of the most remarkable species radiations of flowering plants (Serna-Sánchez *et al.*, 2021). To elucidate the evolutionary history of *C. tracyanum*, we used 351 single-copy orthologs across 19 plant species, selected based on their phylogenetic positions and floral scent phenotypes, to construct a phylogenetic tree and calculate divergence times (Table S7). As expected, *C. tracyanum* clustered with other orchids within the

monocotyledonous clade, and the divergence times aligned with previous reports (Sun *et al.*, 2021; Yang *et al.*, 2021). As shown in Figure 2a, the divergence time between *C. tracyanum* and *C. mannii* was estimated to be 8 million years ago (Mya). The divergence time between *Cymbidium* and *Phalaenopsis* was estimated at 29 Mya, during the early Oligocene, and the origin of Orchidaceae was tracked back to approximately 110 Mya, during the mid-Cretaceous.

The expansion or contraction of gene families has a profound role in driving phenotypic diversity and adaptive evolution in flowering plants (Lai *et al.*, 2024). Based on our analysis of orchid phylogenetic relationships and divergence times, we identified gene families that significantly expanded or contracted at each ancestor node of 19 representative species (Figure 2a). Notably, a total of 1357 gene families (comprising 6861 genes) were specific to *C. tracyanum*. Additionally, 1397 gene families expanded, while 1750 families contracted. Intriguingly, Gene Ontology (GO) enrichment analysis revealed that the significantly expanded gene families were especially enriched in 'terpenoid biosynthetic process', 'diterpenoid biosynthetic process' and 'diterpenoid metabolic process' (Figure 2b and Table S8), which were apparently compatible with the strong aromatic quality specific to *C. tracyanum*. Terpenoid metabolites are not only essential for

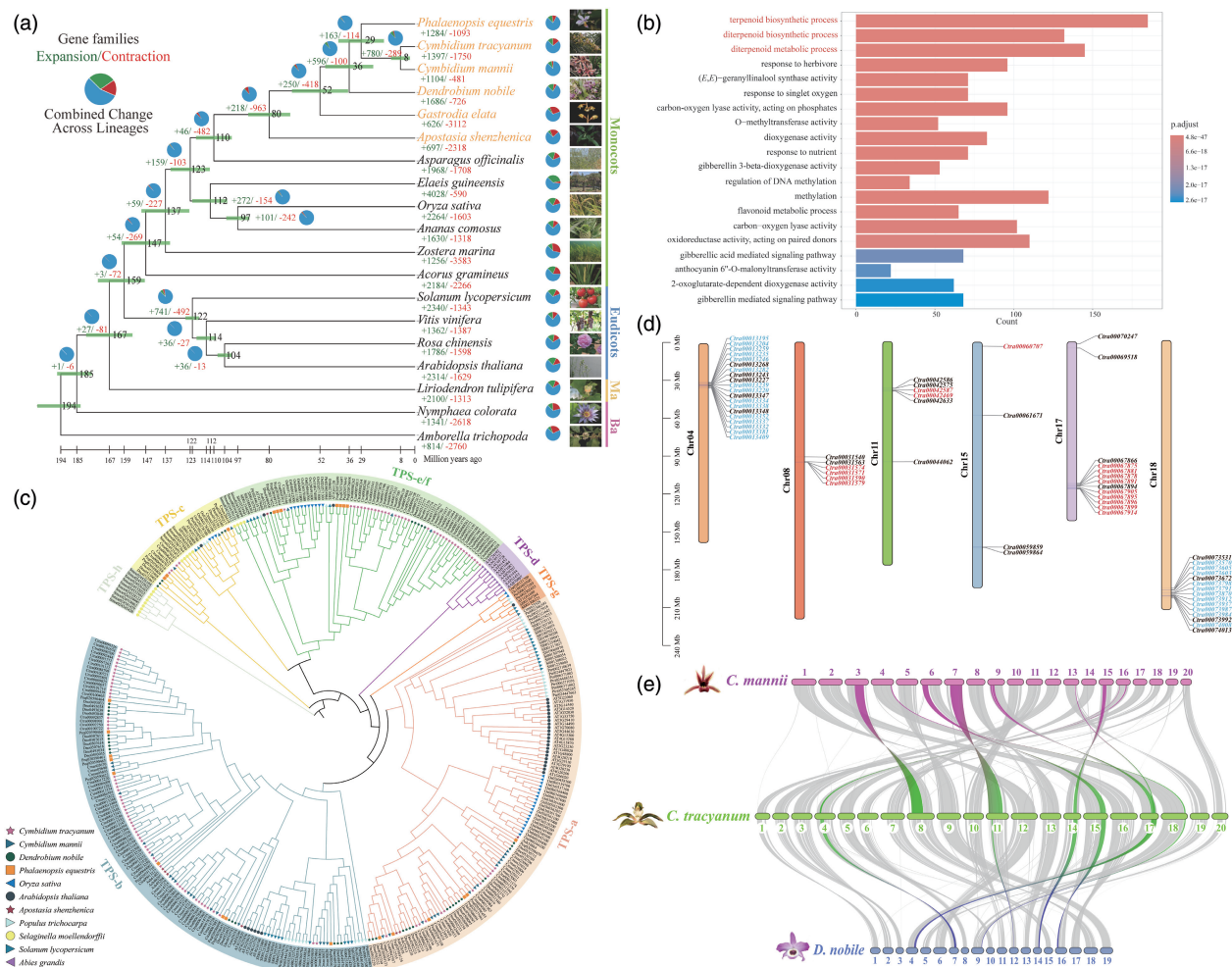


Figure 2 Phylogenomic analyses of *Cymbidium tracyanum*. (a) Phylogenetic tree of 19 plant species. Numbers represent divergence times at each node (Mya, million years ago), with green bars indicating 95% confidence intervals. Pie charts on branches show proportions of gene families undergoing expansion (green) and contraction (red). Numbers below pie charts correspond to the total number of expanded and contracted gene families, with colours matching pie charts. Ma and Ba represent Magnoliids and Basal angiosperms respectively. (b) Gene Ontology (GO) functional enrichment of expanded genes in the *C. tracyanum* genome (see Table S8 for details). (c) Phylogenetic tree of TPS genes from *Apostasia shenzhenica*, *Arabidopsis thaliana*, *Cymbidium mannii*, *C. tracyanum*, *Dendrobium nobile*, *Oryza sativa*, *Phalaenopsis equestris*, *Populus trichocarpa*, *Selaginella moellendorffii*, *Solanum lycopersicum* and *Abies grandis*. Detailed TPS proteins information is available in Table S10. (d) Schematic representation of the distribution of TPS genes in clusters on *C. tracyanum* chromosomes. Twenty TPS genes distributed across a stretch of 4.36 Mb on chromosome 4, 6 TPS genes distributed across a stretch of 4.23 Mb on chromosome 6, 5 TPS genes distributed across a stretch of 2.33 Mb on chromosome 11, 11 TPS genes distributed across a stretch of 4.43 Mb on chromosome 17, and 15 TPS genes occurred in 8.07-Mb stretch on chromosome 18. Blue and red labels indicate segmental and tandem duplication events, respectively (see Table S11). (e) Synteny patterns between genomic regions from *C. tracyanum*, *C. mannii* and *D. nobile*. Highlighted segments denote collinear blocks containing TPS genes of *C. tracyanum*.

plant growth and development (e.g., gibberellins as phytohormones), but also serve as important intermediaries in various plant-environment interactions (Tholl, 2006). This suggests that the terpene synthases (TPSs) may play a critical role in environmental adaptability of *C. tracyanum* by influencing the biosynthesis of floral volatiles.

Significant expansion of the TPS gene family via tandem and segmental duplications

Plants have evolved diverse TPS gene families and subfamilies to synthesize specific terpenoid compounds, enabling them to interact effectively with both biotic and abiotic environments (Jiang et al., 2019). To understand this evolutionary diversification in the scented *C. tracyanum*, we identified and characterized 110

TPS genes—more than in any other of the sequenced orchid genome (Table S9). The TPS gene family is generally classified into seven clades (designated TPS-a to TPS-h) based on sequence similarity, functional assessment and gene structure (Chen et al., 2011). Hence, we constructed a phylogenetic tree incorporating TPS genes in orchids and other representative plant species to explore the potential functions of TPSs in *C. tracyanum* (Figure 2c). Orchid TPS genes were grouped into TPS-a, -b, -c and -e/f subfamilies, excluding the TPS-d subfamily, which is specific to gymnosperms, and the TPS-h subfamily, which is unique to the lycophyte *Selaginella moellendorffii* (Chen et al., 2011). More than half of the TPSs in *C. tracyanum* belonged to the TPS-b subfamily, which encodes monoterpene synthases (monoTPSs) in angiosperms, indicating that the diversification of monoTPSs likely

contributed to the increase in the abundance of monoterpenes. The second most abundant *TPS* genes clustered into the *TPS-e/f* subfamily, which encodes a variety of terpene synthases, including copalyl diphosphate synthases and kaurene synthases that are involved in the biosynthesis of gibberellic acid (Chen *et al.*, 2011). Several *TPS* genes grouped into the *TPS-a* subfamily, encoding sesquiterpene synthases (sesquiTPSs) in both monocots and dicots. Four *TPSs* of *C. tracyanum* were categorized into the *TPS-c* subfamily, closely related to the *TPS-e/f* subfamily and are conserved across land plants with diterpene synthase (diTPS) activity.

In many plant genomes, *TPS* genes frequently occur in tandem arrays, functioning as gene clusters (Chen *et al.*, 2011; Martin *et al.*, 2010). We analysed the distribution of *TPS* genes across their chromosomes in *C. tracyanum* (Figure 2d) and found that there were more than five *TPS* genes distributed in clusters on five chromosomes. These gene clusters were likely the result of tandem or segmental duplications (Figure 2d and Table S10). Notably, these clusters of genes belonged to the same *TPS* subfamily, respectively. In contrast, *TPS* genes distributed further apart on the same chromosome had significantly lower sequence similarity (Table S9).

Furthermore, we conducted a collinearity analysis between *C. tracyanum*, *C. mannii* and *D. nobile*, focusing on collinear blocks that contain *TPS* genes in *C. tracyanum* (Figure 2d). All chromosomes with clustered distributions of *TPS* genes in *C. tracyanum* (chromosomes 4, 8, 11, 17 and 18) exhibited collinear blocks with the other two species, which were also evident by the synteny dot plots (Figure S4).

Profiles of floral volatiles in scented *C. tracyanum* and scentless *C. lowianum*

To identify components responsible for floral scent in *C. tracyanum*, we conducted volatolomics by GC–MS to measure volatile compounds in different parts of *C. tracyanum* flowers at the full-blooming stage. These results were compared to those in *C. lowianum*, a species within the same genus that lacks scent (Figure 3a). We found that petals were the primary site of scent release. In total, 230 volatile compounds were identified (Table S11), including aldehydes (37), alcohols (44), terpenes (21), ketones (42), hydrocarbons (20), heterocyclic compounds (14), phenols (5), esters (37) and acids (10) (Figures 3b and S5).

Comparative analysis of volatile compounds in the petals of *C. tracyanum* and *C. lowianum* at the bud and full-blooming stages identified several compounds with higher content in *C. tracyanum*. The vast majority of these compounds were terpenes, including monoterpenes such as α -thujene, α -phellandrene, sabinene, α -terpinene, terpinolene and sesquiterpenes such as valencene and (*E*)- β -caryophyllene, as well as some alcohols (e.g., linalool, (-)-terpinen-4-ol), a few esters (e.g., methyl (*E*)-cinnamate) and ketones (e.g., (-)-verbenone). These compounds were the significantly altered metabolites in the petals of full-blooming flowers in *C. tracyanum* and *C. lowianum* (Table S12).

The emission patterns of terpenoids (terpenes and their oxygenated derivatives) were further characterized throughout flower development in *C. tracyanum*. We found that the production of terpenoids varied during flowering. Terpenoid emission, which was absent in floral buds, began on the day of anthesis (D + 1), increased rapidly during flower maturation (D + 15), peaked at full-blooming stage (D + 27), and gradually decreased thereafter with flower decaying (Table S13).

(*E*)- β -Caryophyllene accounted for the highest proportion of terpenoids. The OAV (odour activity value) of most terpenoids was greater than 1, indicating that they contributed to the floral scent of *C. tracyanum*. This was consistent with their differential expression detected by volatolomics (Table S12). In *C. lowianum* floral buds and blooming flowers, no terpenoids were detected (Figure S6). Additionally, volatile terpenoids were likely released directly from the epidermal cells of sepals and petals through a diffuse liberation mode, because no obvious osmophore was observed (Maffei, 2010; Vogel, 1983) (Figure S7).

Comparative transcriptomics reveals key steps in terpenoid biosynthesis in *Cymbidium*

To explore the mechanisms that underlie differences in terpenoid biosynthesis between *C. tracyanum* and *C. lowianum*, we conducted a comparative analysis of their floral transcriptomes. To minimize tissue-specific bias, transcriptomic data were generated from petals at various developmental stages. Samples from six developmental stages (Bud, D + 1, D + 7, D + 15, D + 27 and D + 45) of *C. tracyanum*, as well as floral bud and full-blooming stages in *C. lowianum*, were examined (Table S14). KEGG enrichment analysis revealed that a total of 3862 differentially expressed genes in the petals of both orchids at full-blooming stage were significantly enriched in pathways such as 'Monoterpenoid biosynthesis', 'Sesquiterpenoid and triterpenoid biosynthesis', and 'Diterpenoid biosynthesis' (Figure 3c). This finding was consistent with the observation that terpenoids were the main differential metabolites in floral volatiles. Similar enrichment patterns were also observed in the petals of *C. tracyanum* across different developmental stages, particularly during the transition from buds to the day of anthesis (Figure S8), when terpenoid emission began to increase rapidly (Table S13).

The putative genes encoding each step of the MVA and MEP pathways were further identified in both orchids (Table S15, Figures S9 and 4a). Additionally, the expression patterns of several key enzyme-encoding genes across the floral developmental stages of the two orchids were validated through qRT-PCR analysis (Figure S10). In the MVA pathway, HMG-CoA synthase (HMGS) catalyses the production of HMG-CoA from acetoacetyl-CoA (AcAc-CoA), a crucial step in both feedback regulation and stress responses (Tholl, 2015). Two *HMGS* genes were identified in *C. tracyanum* and *C. lowianum*, and neither exhibited significant changes across developmental stages in either species, nor did they show notable differences in the petals at the full-blooming stage (Figures 4a and S10). HMG-CoA reductase (HMGR) catalyses the irreversible formation of mevalonate from HMG-CoA, which is considered a rate-limiting step (Re *et al.*, 1995). *C. tracyanum* possessed two *HMGR* genes: one showed an expression pattern consistent with flower development, peaking at the D + 7 before declining, whereas the other showed a continuous decline in expression from the bud stage to flower decay. In contrast, there was almost no difference in the expression levels of these two *HMGR* genes between the bud stage and the full-blooming stage of *C. lowianum*, both of which remained at relatively low levels (Figures 4a, S10 and S11). In the MEP pathway, 1-deoxy-D-xylulose 5-phosphate (DXP) synthase (DXS) functions as an important regulatory and rate-limiting enzyme to form DXP (Estévez *et al.*, 2001). The four *DXS* genes in *C. tracyanum* belong to three clades, with enzymes in clade 2 mainly involved in the formation of plant secondary metabolites including terpenoids (Paetzold *et al.*, 2010; Walter *et al.*, 2002). This enzyme-encoding gene (*novel.3190*) was also the only one

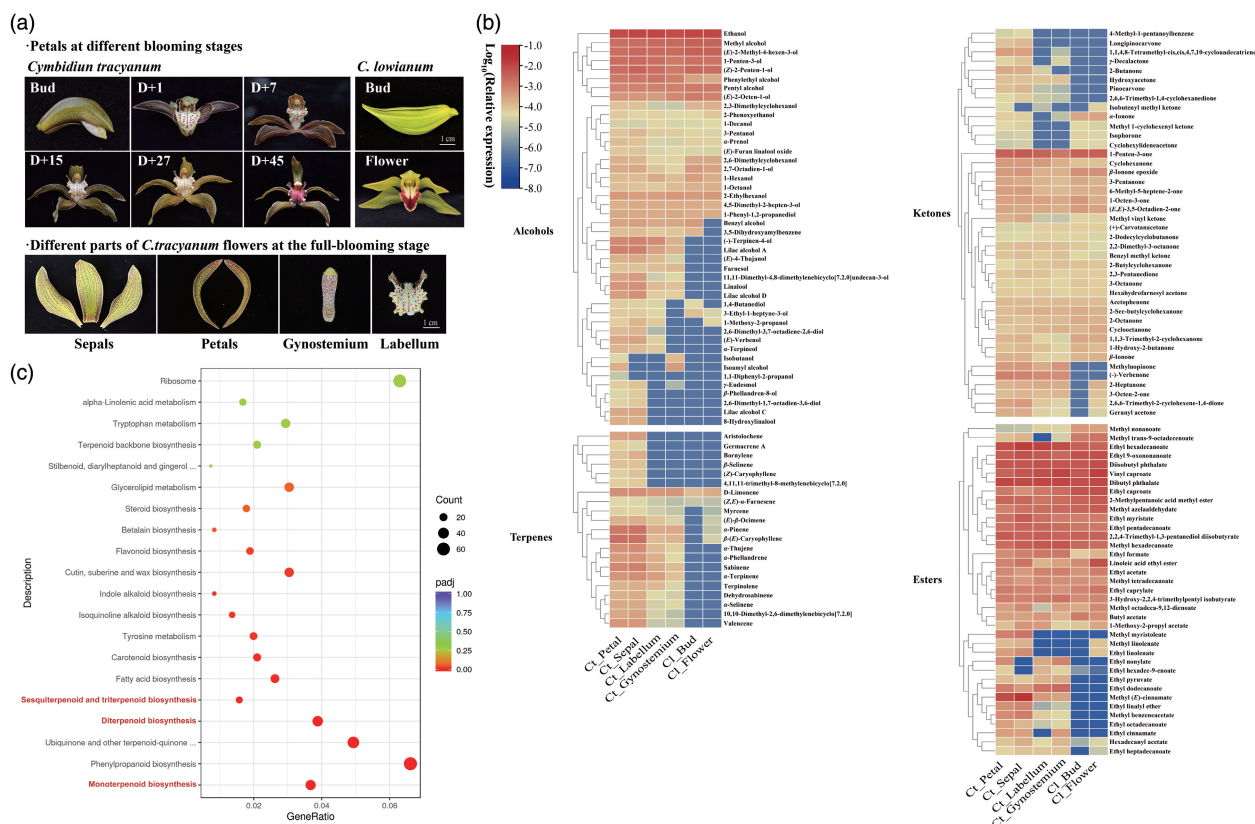


Figure 3 Profiles of floral volatiles in scented *Cymbidium tracyanum* and scentless *C. lowianum*. (a) Flowers at different developmental stages. Top: single flowers of *C. tracyanum* at six developmental stages (left), including the first day of flowering (D + 1), and 7 days (D + 7), 15 days (D + 15), 27 days (D + 27) and 45 days (D + 45) after flowering; Bud and full-blooming flower (15 days after flowering) of *C. lowianum* (right); Below: different parts of *C. tracyanum* flowers at D + 15; Scale bars represent 1 cm; (b) Composition and content of volatile compounds detected in volatolomics of different parts of *C. tracyanum* flowers at the full-blooming stage, and the petals of *C. lowianum* at the bud and full-blooming stages (see Figure S5 and Table S12). (c) KEGG functional enrichment of differentially expressed genes in petals of *C. tracyanum* and *C. lowianum* at full-blooming stage.

that exhibited differential expression across developmental stages in both orchids. However, its expression levels were comparable at the full-blooming stage in both species (Figures 4a, S10 and S12). Another key rate-limiting enzyme, DXR reductoisomerase (DXR), which converts DXP to MEP by an intramolecular rearrangement, exhibited significantly differential expression compared to the bud stage during the entire flowering process of *C. tracyanum*. Whereas at different developmental stages, this gene expression in *C. lowianum* remained consistently high, comparable to that at the full-blooming stage of *C. tracyanum* (Figures 4a and S10).

Transcriptome analysis of different floral parts of *C. tracyanum* flowers at the full-blooming stage revealed that most genes (22/32) in MVA and MEP pathways, including the aforementioned rate-limiting enzyme genes, exhibited significantly higher expression in petals and sepals compared to the labellum and gynostemium. This pattern was consistent with the profiles of terpenoids detected in the volatolomic analysis (Figure S13). As shown in Figure 4a, only one DXS enzyme-encoding gene exhibited significant differential expression in the petals of *C. lowianum* between the full-blooming and bud stages, while other genes maintained a stable expression pattern throughout the flowering process of *C. lowianum*. Additionally, across the entire pathway, only three genes (*Ctra00032652*, *Ctra00014764* and *Ctra00078867*) showed significantly higher expression in the

petals of *C. tracyanum* at the full-bloom stage compared to *C. lowianum*. This suggested that the differences in terpenoid content among different floral parts of *C. tracyanum* might be primarily due to variations in substrate synthesis supplied to TPSs. In contrast, the differences in terpenoid content between the petals of the two orchids were less influenced by this factor but were more likely driven by the substrate-specific catalytic activity of downstream TPSs, which further contributed to variations in both the types and amounts of terpenoids.

We also analysed the expression levels of 110 *TPS* genes identified in *C. tracyanum* (Figure S14), focusing on 42 genes with an average FPKM > 1 in different floral parts at the full-blooming stage. Among them, half of the genes exhibited significantly higher expression levels in petals and sepals compared to the labellum and gynostemium. Moreover, we found that the expression of several *TPS* genes in petals changed with floral developmental stages in *C. tracyanum* and *C. lowianum* (Table S16 and Figure 4b).

Co-expression network related to terpenoid biosynthesis in *C. tracyanum* identifies candidate genes involved

To further clarify the key *TPS* genes and regulatory elements responsible for terpenoid biosynthesis in *C. tracyanum*, we conducted a weighted correlation network analysis (WGCNA) by combining the floral transcriptome data from different

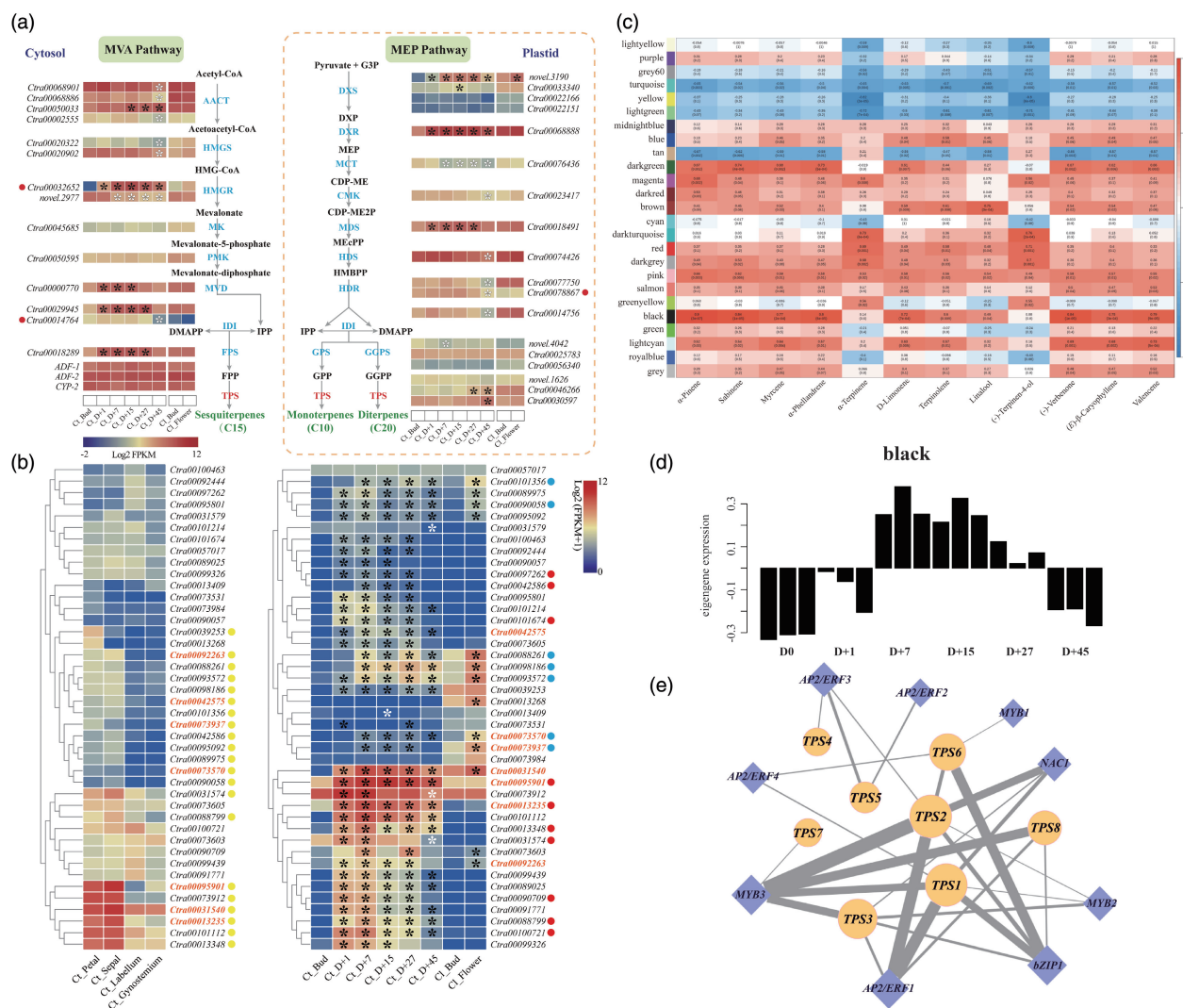


Figure 4 Gene expression profiles in terpenoid biosynthesis pathways and co-expression analysis. (a) Comparative expression profiles of putative genes encoding enzymes involved in terpenoid biosynthesis in petals of *Cymbidium tracyanum* and *C. lowianum* at different developmental stages. Abbreviations for enzymes at each catalytic step are highlighted in blue and red. We set the threshold for differential gene expression as $|\log_2(\text{fold change})| \geq 1$ and $\text{padj} \leq 0.05$. Genes meeting this threshold at different developmental stages are marked with asterisks at the corresponding stages, with solid asterisks indicating significant upregulation and hollow asterisks indicating significant downregulation. Red dots denote genes that are significantly highly expressed in the petals of *C. tracyanum* at the full-blooming stage compared to *C. lowianum* (see Table S15). (b) Heatmaps depicting the expression patterns of *CtTPS* genes with FPKM > 1 in different floral parts of *C. tracyanum* at the full-blooming stage (left) and petals of *C. tracyanum* and *C. lowianum* at different developmental stages (right). Yellow dots represent genes that are significantly highly expressed in petals and sepals compared to the labellum and gynostemium in *C. tracyanum* at the full-blooming stage. Red dots indicate genes that are significantly highly expressed in the petals of *C. tracyanum* compared to *C. lowianum*, while blue dots indicate significantly lower expression (see Table S16). The genes marked in orange represent *TPS* genes in the black module identified through the correlation analysis. (c) Matrix of module-metabolite associations, with abscissa representing terpenoids and ordinate representing modules. Correlation coefficients and *P*-values between modules and metabolites are shown at the row-column intersections. (d) The expression pattern of genes within the black module is represented by its eigengene. (e) Co-expression network of nine transcription factors (CtAP2/ERF1-4, CtMYB1-3, CtNAC1 and CtMZP1) and eight *CtTPS* genes in the black module. Diamonds represent transcription factors, and circles represent *CtTPS* genes. Node size indicates the importance of connectivity in the network, and edge thickness indicates the strength of the connectivity.

developmental stages and corresponding terpenoid content measurements (Figure S15 and Table S13). The analysis revealed correlations between 25 modules and 12 characteristic terpenoids detected throughout flower development in *C. tracyanum* (Figure 4c). Modules with larger correlation coefficients and lower *P*-values were highly correlated with specific phenotypes. The black module exhibited the strongest correlation with the majority

of terpenoids, including α -pinene, sabinene, myrcene, D-limonene, (-)-verbenone, (E)- β -caryophyllene and valencene. The expression of this module, represented by its eigengene, increased during the blooming process, peaked at the full-blooming stage and then declined (Figure 4d). Eight *TPS* genes (*Ctra00013235*, *Ctra00031540*, *Ctra00042575*, *Ctra00073570*, *Ctra00073937*, *Ctra00092263*, *Ctra00094506* and

Ctra00095901) were identified (designated as *CtTPS1*–*CtTPS8*). These genes were expressed at relatively higher levels in the petals and sepals of full-blooming flowers (Figure 4b).

The emission of terpenoids is not only affected by TPS, but also regulated by various *cis*-elements at the transcriptional level. Here, a total of 642 genes were annotated as bHLH, bZIP, EIL, MYB, NAC, WRKY and AP2/ERF transcription factors. Among these, nine transcription factors were identified as potential regulators of terpenoid synthesis, based on their connectivity with eight *CtTPS* genes in the black module. These TFs were nominated as *CtAP2/ERF1*–4, *CtMYB1*–3, *CtNAC1* and *CtbZIP1* (Table S17). The regulatory network mediated by these TFs and *CtTPS* genes was shown in Figure 4e. Notably, *CtAP2/ERF1*, *CtbZIP1*, *CtNAC1* and *CtMYB2,3*, displayed high co-expression with *CtTPS1, 2, 3, 8*, suggesting their roles as hub genes in the network.

To validate the expression patterns of genes identified through co-expression analysis, the transcript abundances of eight selected *CtTPSs* and nine *CtTFs* in the petals of *C. tracyanum* at various floral developmental stages were analysed by qRT-PCR. Notably, *CtTPS5* was expressed at lower levels at D + 27 than at the bud stage, while the expression of other structural genes involved in terpenoid biosynthesis generally increased, peaking during flowering before declining. This pattern mirrored the trend of terpenoid emissions. Although the expression levels of *CtTPS2* and *CtTPS8* fluctuated during floral development, they consistently showed significantly higher expression levels at all blooming stages compared to the bud stage (Figure 5a). The expression of the 9 TFs either peaked on D + 15 (*CtbZIP1*, *CtMYB2* and *CtNAC1*), D + 27 (*CtAP2/ERF2, 3*) or D + 7 (*CtAP2/ERF1, 4* and *CtMYB2, 3*), preceding the expression of the structural genes (Figure 5b).

We also compared the gene expression across different floral parts of *C. tracyanum* at the full-blooming stage (D + 15). All *CtTPS* genes showed significantly higher expression levels in petal and sepal compared to labellum and gynostemium, with *CtTPS1* expression approximately 50-fold higher in petal and sepal. *CtTPS5* exhibited the least variation in expression across different flower parts, lacking clear tissue specificity (Figure S16). For transcription factors, *CtbZIP1*, *CtMYB1, 2, 3* and *CtAP2/ERF4* were highly expressed in petal and sepal. *CtAP2/ERF1* showed comparable expression levels in petal, sepal and gynostemium, while *CtAP2/ERF3* exhibited the highest expression in gynostemium. *CtNAC1* had similar expression levels in the petal and gynostemium but was significantly higher than in the sepal (Figure S17). Structural genes and transcription factors involved in terpenoid biosynthesis in *C. tracyanum* are expected to be expressed at higher levels in scented *C. tracyanum* than in scentless *C. lowianum*. We found that *CtTPS1, 2, 3, 5, 6* and *8* were highly expressed in *C. tracyanum* but were rarely detectable in scentless *C. lowianum*. However, *CtTPS4* and *CtTPS7* showed higher expression in *C. lowianum* than in *C. tracyanum*. The expression of *CtAP2/ERF2, 3* did not differ significantly between the two orchids, but *CtAP2/ERF1, 4*, *CtbZIP1*, *CtMYB1, 2, 3* and *CtNAC1* were expressed at remarkably higher levels in scented *C. tracyanum* than in scentless *C. lowianum* (Figure 5c).

Based on these findings, we identified four *CtTPSs* (*CtTPS1, 2, 3, 8*) and six *CtTFs* (*CtAP2/ERF1, 4*, *CtbZIP1* and *CtMYB1, 2, 3*) as key structural genes and regulators, showing specifically high expression in the petal and sepal of *C. tracyanum*. Their expression patterns were consistent with terpenoid emissions during flower development.

Versatile and diverse functions of *CtTPS* genes

Phylogenetic analysis was conducted to infer the potential catalytic functions of the four *CtTPS* proteins by comparing them with other functionally validated plant TPSs (Figure 6b and Table S18). *CtTPS2* and *CtTPS3* clustered into the TPS-a subfamily, which represents sesquiterpenes in monocots and dicots. *CtTPS1* was classified into the TPS-b subfamily, representing monoTPSs in angiosperms. Notably, DoTPS10, closely related to *CtTPS1*, is one of the few TPSs in orchids that have been identified to catalyse linalool synthesis *in vitro* (Yu *et al.*, 2020). Additionally, *CtTPS8* belongs to the subfamily TPS-e/f that has enzyme activities of monoTPSs, sesquiterpenes and diTPSs in vascular plants (Chen *et al.*, 2011). Subcellular localization analysis was performed by fusing the full-length sequences with the eGFP protein. The results revealed that *CtTPS1* and *CtTPS3* were localized in the plastids, whereas *CtTPS2* and *CtTPS8* were distributed in the cytoplasm (Figure 6a). However, the subcellular localization of terpene synthases and their functional sites is not entirely consistent (Bao *et al.*, 2023; Conart *et al.*, 2023).

Previous studies have suggested that *TPS* genes evolve rapidly, and even with high sequence similarity, their catalytic products can vary considerably (Jia *et al.*, 2022). To directly confirm the enzymatic properties of these *CtTPS* proteins, we performed *in vitro* assays using GPP or FPP as substrates. Recombinant *CtTPS* proteins were expressed heterologously in *E. coli* for biochemical analysis (Figure S18). As expected, no products were detected in reactions where heat-inactivated recombinant proteins were added to reaction mixtures supplemented with both substrates. Thus, only crude protein extracts from the *E. coli* expression system containing empty vector were used as controls. Upon incubation with GPP as a substrate, preliminary identification based on comparison with the NIST database indicated that *CtTPS2* produced (*E*)- β -ocimene, α -terpineol and terpinolene, all of which were detected in the *C. tracyanum* flowers. *CtTPS3* catalysed the formation of neo- α -ocimene, whereas *CtTPS8* not only catalysed neo- α -ocimene but also produced comparable amounts of terpinolene. Furthermore, the production of neo- α -ocimene, terpinolene and α -terpineol was confirmed using authentic standards (Figure 6c and Table S19). The catalytic activity of recombinant *CtTPS* proteins was also performed using FPP as a substrate. Preliminary comparison results indicated that *CtTPS2* could catalyse the production of a variety of sesquiterpenes, including β -sesquiphellandrene, α -curcumen, α -farnesene, α -bisabolene and nerolidol, demonstrating the diverse catalytic functions of *CtTPS2*. *CtTPS3* could produce β -humulene and (*Z*, *E*)-farnesol. The main product of *CtTPS8* was (*E*)- β -caryophyllene, which was also the most abundant compound during the flowering process of *C. tracyanum*. The formation of α -farnesene and nerolidol was further confirmed using authentic standards (Figure 6d and Table S19). The *in vitro* catalytic assays of these recombinant *CtTPS* proteins partially explain the biosynthesis of certain floral scent compounds in *C. tracyanum*. It was also evident that *CtTPS2*, *CtTPS3* and *CtTPS8* are all bifunctional enzymes *in vitro*, while their actual product profiles in planta could be influenced by substrate availability and enzyme localization.

CtTPS genes directly regulated by selected TFs

After elucidating the functions of *CtTPS* proteins, we sought to understand the regulatory mechanisms by which transcription factors control these terpene synthases. Initial subcellular

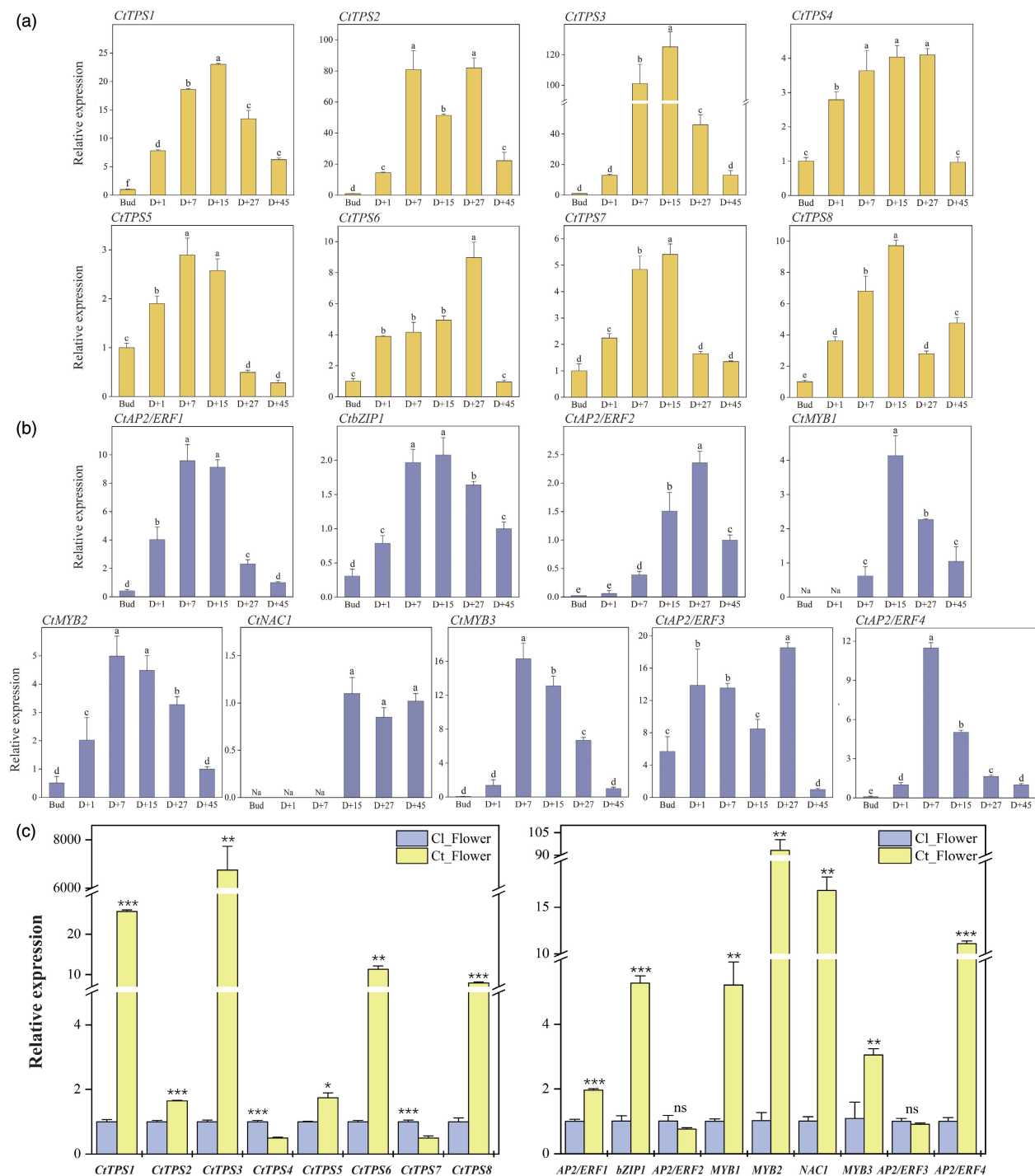


Figure 5 Gene expression levels of *CtTPSs* and *CtTFs* at different developmental stages analysed by qRT-PCR. (a, b) Expression levels of *CtTPS* genes (a) and *CtTF* genes (b) at different developmental stages in the petals of *C. tracyanum*. *CtCYP2* was used as an internal control. Different letters represent significant differences between stages, calculated through one-way ANOVA ($P < 0.05$). Error bars indicate the standard deviation (SD) of three replicates. (c) Expression levels of eight structural genes (left) and nine candidate TFs (right) in the petals of scentless *C. lowianum* and scented *C. tracyanum* (at the D + 15 stage). Statistical analysis of the expression levels of genes between the two orchids was performed by using Student's *t*-test (* $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$). Expression levels are normalized to *CtCYP2*. Data are represented as means \pm SD from three replicates. Primers are listed in Table S21.

localization analysis confirmed that all six transcription factors were localized to the cell nucleus (Figure S19). We then cloned the promoter sequences (~2 kb upstream) of the four *CtTPS* genes and identified several *cis*-elements within these regions,

indicating potential binding sites for candidate TFs (CtAP2/ERF1, 4, CtZIP1 and CtMYB1, 2, 3). These *cis*-elements mainly included the AP2/ERF transcription factor-specific binding GCC-box, AP2 and DRE/CRT, the bZIP transcription factor-specific binding

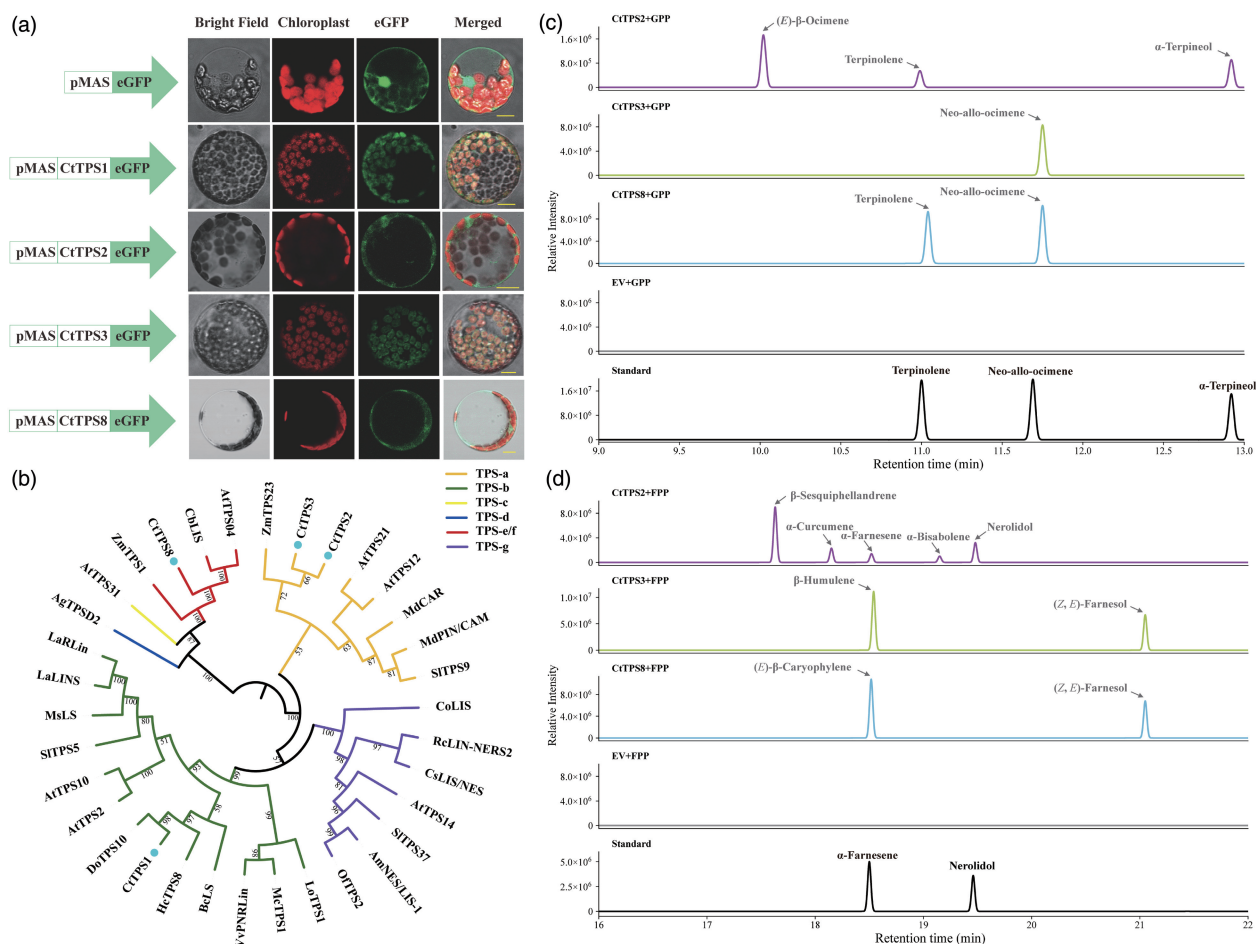


Figure 6 Functional characterization of the enzymes encoded by *CtTPS* genes. (a) Subcellular localization of free eGFP and four *CtTPS*-eGFP fusions in *Nicotiana benthamiana* leaf protoplasts. eGFP fluorescence was detected in the green channel, while chlorophyll autofluorescence was detected in the red channel; Merged, brightfield image combined with green and red channels; scale bar indicates 10 μ m. (b) Phylogenetic analysis of TPS proteins from *Cymbidium tracyanum* (*CtTPS*1, 2, 3, 8) and other plants using the maximum likelihood method. The *CtTPS* proteins in this study are highlighted by circles. The TPS-a, TPS-b, TPS-c, TPS-d, TPS-e/f and TPS-g clades are highlighted with shaded lines. Plant species are as follows: *Arabidopsis thaliana* (At); *Antirrhinum majus* (Am); *Abies grandis* (Ag); *Backhousia citriodora* (Bc); *Clarkia breweri* (Cb); *Cinnamomum osmophloeum* (Co); *Camellia sinensis* (Cs); *Dendrobium officinale* (Do); *Hedychium coronarium* (Hc); *Lavandula angustifolia* (La); *Lilium* hybrid cultivar (Lo); *Magnolia champaca* (Mc); *Malus domestica* (Md); *Mentha spicata* (Ms); *Osmanthus fragrans* (Of); *Rosa chinensis* (Rc); *Solanum lycopersicum* (Sl); *Vitis vinifera* (Vv); *Zea mays* (Zm). Detailed information on TPS proteins is provided in Table S18. (c, d) Identification of enzymatic products of recombinant *CtTPS* proteins using geranyl diphosphate (GPP)/farnesyl diphosphate (FPP) as substrates. Volatile terpenoids were analysed by GC–MS and extracts from the *E. coli* expression system containing empty vector were used as controls. Arrows indicate compounds identified based on database matching, while “Standard” refers to the detection results of standard samples under the same conditions (Table S19).

G-box, C-box and ABRE, and binding sites for MYB transcription factors (Figure 7a).

Subsequently, we constructed promoter activation dual-luciferase reporter assay vectors with the coding sequences of the TFs under the control of the CaMV 35S promoter. Promoter fragments of *CtTPS*s were integrated into the pGreenII-0800-LUC vectors to perform the promoter activation experiment in tobacco leaf protoplasts (Figure 7b). As shown in Figure 7c, based on the expression ratio of LUC/REN compared to controls (empty vector without TF), we found that CtAP2/ERF1 activated the promoters of *CtTPS*1 and *CtTPS*8, CtZIP1 only activated the promoter of *CtTPS*8, CtMYB2 activated the promoters of *CtTPS*3 and *CtTPS*8, and CtAP2/ERF4 activated the promoters of *CtTPS*2 and *CtTPS*8. Additionally, CtMYB3 exhibited activation of the promoters of

*CtTPS*2, 3 and 8. Further examination of the activation of the *CtTPS*s promoter regions by selected TFs using yeast one-hybrid assays yielded consistent binding results (Figure 7d). In addition, electrophoretic mobility shift assays (EMSA) were conducted to provide direct biochemical evidence of TF binding, reinforcing their regulatory roles in *CtTPS* gene expression (Figure S20 and Table S21). As shown in Figure 7e, CtAP2/ERF1 was able to directly bind to the AP2 domain in the *CtTPS*1 promoter region and the DRE element in the promoter of *CtTPS*8. Similarly, CtAP2/ERF4 was directly bound to the AP2 domain in the *CtTPS*2 promoter region and the DRE element in the promoter of *CtTPS*8. The activation of *CtTPS*8 by CtZIP1 was attributed to its binding to the G-box and ABA response element. CtMYB2 and CtMYB3 bound to the MYB sites in their respective gene promoter regions, with CtMYB2 specifically

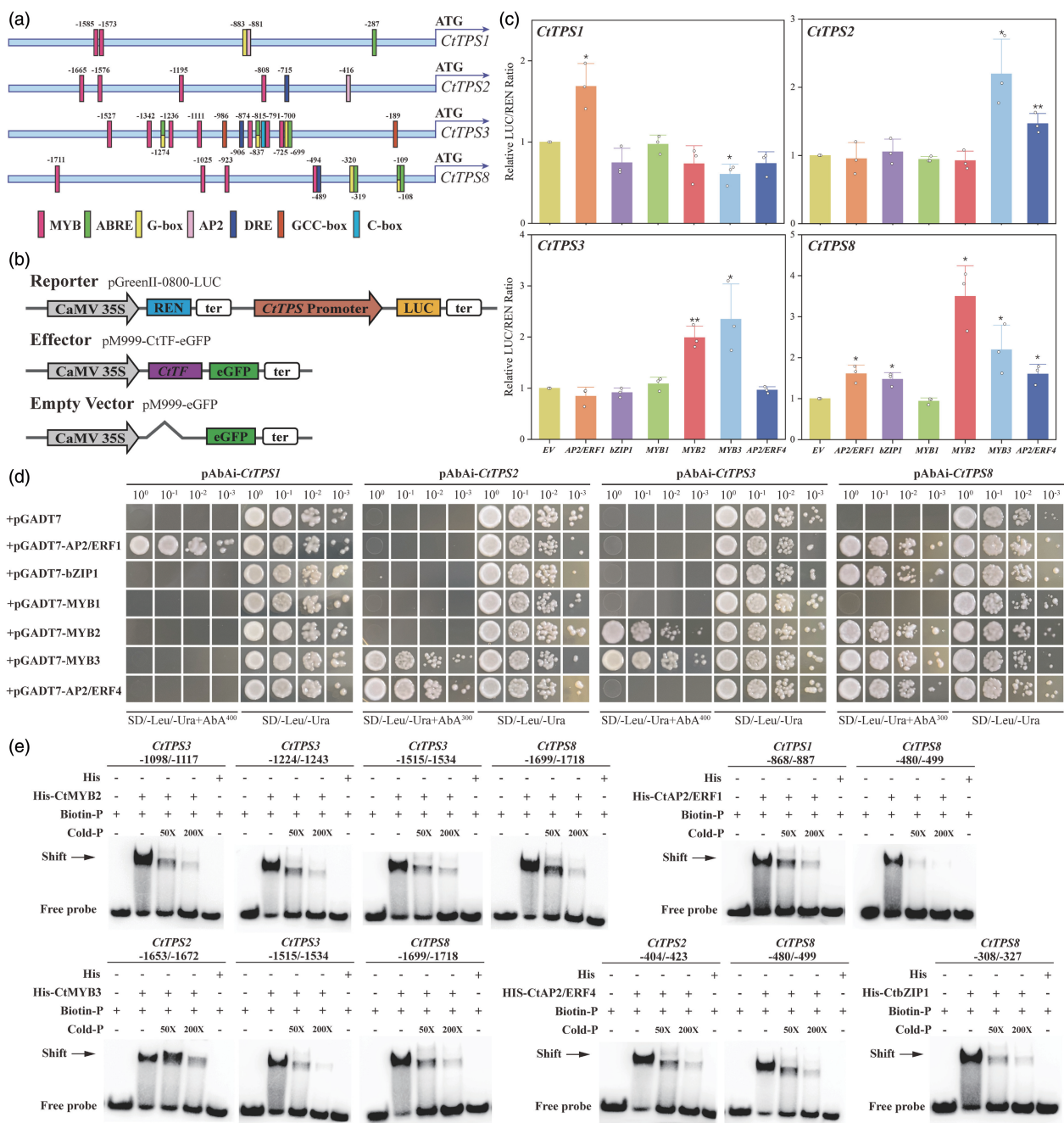


Figure 7 Confirmation of *CtTPS* genes directly regulated by TFs. (a) Schematic diagram of cis-elements in the promoter region of four *CtTPS* genes. MYB: MYB structural domain; ABRE: ABA response elements; G-box: CACGTG; AP2: AP2 domain; DRE: GCCGAC; GCC-box: GCCGCC; C-box: GACGTC. (b) Schematic diagram of vector construct. Reporter vector: pGreenII-0800-LUC-pro*CtTPS*; effector vector: pM999-CtTF-eGFP; empty vector: pM999-eGFP. (c) Activation of six CtTFs on four *CtTPS* genes was assessed by dual-luciferase assays in tobacco leaf protoplasts. The empty pM999-eGFP vector served as a negative control. Error bars represent the SD of three replicates. Asterisks denote significance levels calculated by Student's *t*-test compared with negative control (**P* < 0.05, ***P* < 0.01 and ****P* < 0.001). (d) Yeast one-hybrid assays demonstrating that CtTFs specifically bind to the promoters of different *CtTPS* genes. The empty vector pGADT7 vector served as the negative control. SD medium is the Ura and Leu double deficient medium. Superscripted numbers indicate the concentration of aureobasidin A used for screening. (e) Electrophoretic mobility shift assays (EMSAs) showing direct binding of His-CtTFs to the promoter of *CtTPS* genes. Each element, along with its upstream and downstream flanking sequences, was biotin-labelled and used as a probe. A shifted band (indicated by an arrow) represents the protein-DNA complex. Competition assays were performed using an excess of unlabeled probe (50-fold and 200-fold) to confirm binding specificity. A His-tagged control protein was used as a negative control. Experiments were repeated at least twice with consistent results.

interacting with three MYB sites in the *CtTPS3* promoter. It was worth noting that the results of transcription factor activation were partially consistent with our putative co-expression network (Figure 4e), with some TFs exhibiting regulatory roles as predicted. These findings verified that the selected TFs regulate the expression of terpenoid biosynthesis-related genes.

Discussion

Here, we employed multi-omics, molecular biology and phytochemistry to elucidate the mechanisms underlying the biosynthesis and regulation of floral scent in *C. tracyanum*. Comparative genomic analysis through high-quality chromosome-level genome assembly revealed that the significantly expanded gene families in *C. tracyanum* were primarily enriched in pathways related to terpenoid biosynthesis and metabolism. Our integration of comparative transcriptome and metabolome analyses identified several essential terpene synthases (i.e. *CtTPS1*, *CtTPS2*, *CtTPS3* and *CtTPS8*). We verified the catalytic function of *CtTPSs* and identified transcription factors that regulate *CtTPS* gene expression, including *CtAP2/ERF1*, *CtZIP1*, *CtMYB2*, *CtMYB3* and *CtAP2/ERF4*.

The contribution of significantly expanded *TPS* genes in *C. tracyanum* to volatile terpenoid biosynthesis and its implications for adaptive evolution

Our comparative genomic analyses, based on the newly released *C. tracyanum* genome, revealed that significantly expanded gene families in *C. tracyanum* are predominantly enriched in pathways related to terpenoid biosynthesis and metabolism. This finding is consistent with its strong floral scent phenotype, suggesting that terpenoids have played a crucial role in the environmental adaptability of *C. tracyanum* throughout its evolutionary history. Our finding that *TPS*-a and *TPS*-b subfamily genes have undergone extensive expansion in *C. tracyanum* is consistent with previous studies, which indicate that their expansion contributes to the diversification and accumulation of mono-terpenes and sesquiterpenes (Huang *et al.*, 2021; Shang *et al.*, 2020; Zhang *et al.*, 2021). GO enrichment analysis of the contracted gene families revealed significant enrichment in pathways related to transcription factor or transcription regulator activity, stress responses, as well as pathways associated with plant organ morphogenesis. This might also reflect the strategies of *C. tracyanum* at the genomic level in pathway regulation and adaptation to specific life forms during evolution (Figure S21 and Table S20).

Previous studies have shown that the *TPS* family has undergone several expansions due to gene duplications during plant evolution, leading to structural remodelling, differential expression, subcellular segregation and neofunctionalization (Jia *et al.*, 2022; Karunanithi and Zerbe, 2019). Our results revealed that *C. tracyanum* has experienced two major duplication events in its evolutionary history, with an ancient event shared by monocots and a more recent event shared by all species within the Orchidaceae (Figure 1d). Yet, the number of *TPS* gene family members in *C. tracyanum* surpasses that of any other sequenced orchid (Table S10), likely due to a combination of tandem and segmental duplications (Table S11). Further research is needed to confirm the existence of functional redundancy or divergence among *TPS* genes in *C. tracyanum* from different sources, particularly the potential role of those distributed in clusters on chromosomes (Figure S9).

Interestingly, some *TPS* genes were scarcely expressed in *C. tracyanum* flowers (Figure S14), suggesting potential functional divergence or even neofunctionalization. Beyond their role as floral volatiles, terpenoids also participate in plant growth and development as essential components and regulatory substances (Christianson, 2006; Pichersky *et al.*, 2006; Tholl, 2006). In

addition, volatile terpenoids may compete with other substances for metabolic flux. For instance, floral colours, which are often discussed alongside floral scents, are biochemically connected with VOCs despite their distinct biosynthetic pathways. They exhibit synergistic interactions during flower development and respond to biotic and abiotic stresses (Li *et al.*, 2025). Similarly, the differentially expressed genes in the petals of *C. tracyanum* and *C. lowianum* at the full-blooming stage were significantly enriched in the phenylpropanoid biosynthesis pathway (Figure 3c). Compared to the lack of corresponding differences in volatolomics, this phenomenon was more likely attributed to the differences in flower colour between the two species (Figure 3a). It also reflects the diverse roles of phenylpropanoid pathway products in the structural, physiological and ecological functions of the plants (Vogt, 2010; Wang *et al.*, 2014).

Mechanisms of biosynthesis and regulation of volatile terpenoids in *C. tracyanum*

The floral terpenoids in *C. tracyanum* exhibited significant tissue and spatiotemporal specificity throughout the developmental stages (Figures 3b, 55 and Table S13). The highest levels of VOCs were detected in the petal and sepal, whereas the lowest levels were detected in the labellum and gynostemium, a pattern that may avoid damaging critical pollination structures. During flower development, VOC content first increased and then decreased. This pattern has been observed in other plants and reflects an initial attraction of pollinators followed by the avoidance of further damage to pollinated flowers (Cancino and Damon, 2007). However, this speculation needs to be further analysed by examining changes in floral volatiles before and after pollination and orchestrating specific pollinator attraction experiments.

Terpenoids are one of the most diverse classes of plant metabolites, with significant physiological and ecological functions. *TPSs*, as pivotal enzymes that initiate terpenoid biosynthesis, have drawn considerable attention. Although the identification of *TPSs* across a broad range of plant lineages, including angiosperms, gymnosperms, bryophytes, lycophytes and ferns, has extended our understanding of the origins and evolution of *TPSs* (Jia *et al.*, 2022), model plants, such as *Arabidopsis thaliana* and *Oryza sativa*, have served as paradigms for in-depth studies of the spatial organization and temporal regulation of *TPSs* in relation to various biological functions (Chen *et al.*, 2014; Huang *et al.*, 2011). Few studies have examined the regulation of terpenoid biosynthesis in orchids, although one report on *Phalaenopsis* demonstrated that *PbbHLH4* regulates floral monoterpene biosynthesis (Chuang *et al.*, 2018).

Here, we examined the expression patterns of *CtTPSs* at different developmental stages and in different parts of *C. tracyanum* flowers. We also used co-expression analysis to screen for transcription factors that may regulate *CtTPS* genes (Figure 4e). Our analysis of *cis*-element binding sites in the promoter region of the *CtTPS* genes (Figure 7a) and the subsequent validation experiments through dual-luciferase reporter assays, yeast one-hybrid experiments and EMSA assays (Figure 7c–e) identified several TFs that activate the promoter regions of multiple *CtTPS* genes. Specifically, *CtAP2/ERF1* activated the promoters of *CtTPS1* and *CtTPS8*; *CtMYB2* activated the promoters of *CtTPS3* and *CtTPS8*, and *CtAP2/ERF4* activated the promoters of *CtTPS2* and *CtTPS8*. *CtMYB3* activated the promoters of *CtTPS2*, *CtTPS3* and *CtTPS8*. This suggested that *CtMYB3* is likely a key transcription factor involved in regulating the floral scent volatiles of *C. tracyanum* and warrants further

investigation. These findings showed that *CtTPS* promoters could be bound by multiple transcription factors. Meanwhile, our study also provided a solid foundation for the functional analysis of other terpene synthases in *C. tracyanum*. It remains unclear whether these transcription factors simultaneously activate and/or repress promoter regions and how their combined actions regulate the synthesis of terpenoids in *C. tracyanum*. Currently, the lack of a suitable genetic transformation system limits our ability to directly verify the regulatory effects of transcription factors on *CtTPSs* in *C. tracyanum*.

Our study provides a solid reference for in-depth exploration of floral scent and other unique adaptive traits in orchids

Since Darwin's publication of *Fertilisation of Orchids* in 1862, orchids have garnered significant attention from evolutionary biologists. The genome of *C. tracyanum* offers new avenues to delve into the genomic basis of the physiological characteristics in orchids. Although studies on orchid floral scent are common, they predominantly focus on the detection of volatile compounds or the identification of genes in related pathways. Comprehensive studies that encompass the spatiotemporal changes of volatile compounds, genome assembly and comparative analysis, expression patterns of structural genes involved in terpenoid biosynthesis, as well as the verification of their catalytic functions and transcriptional regulation, are quite rare in orchids. However, many aspects of terpenoid biosynthesis in orchids still warrant in-depth exploration.

First, attention should be given to the biosynthesis of several other floral VOCs, such as those in the benzenoid/phenylpropanoid biosynthesis pathway and the biosynthetic pathway of volatiles derived from fatty acids. Although these VOCs are not as abundant as terpenoids, they may still impact the floral scent of *C. tracyanum*. For example, in *Caladenia plicata*, a new floral volatile constituent, 2-hydroxy-6-methylacetophenone, has been shown to cooperate with the commonly reported (S)- β -citronellol in the attraction of pollinators (Xu *et al.*, 2017). Secondly, the focus should be placed on the role of floral VOCs as signalling molecules that attract pollinators and facilitate communication between plants and the environment, especially the conversion between volatile and non-volatile substances, such as the identification of subsequent modification enzymes. Modification of TPS products by oxidation, peroxidation, methylation, acylation, or cleavage changes their physical properties and may alter their biological activities (Chen *et al.*, 2011). Furthermore, in some plants, such as *Freesia* (Bao *et al.*, 2023), the functions of natural allelic variants of TPS among different varieties within the same genus have been identified. In rose, differences in the content of citronellol in different species have been shown to be due to *NUDX1-1a* copy numbers, and the evolution of biosynthetic pathways of terpene scent compounds within the genus has been elucidated (Shang *et al.*, 2024). For *Cymbidium* species with such rich floral scent resources, understanding the differences in the biosynthesis of floral scent compounds among different varieties, as well as the evolutionary mechanisms within this genus, is crucial for guiding breeding practices and understanding their adaptability to the environment.

Conclusions

In the present study, we identified terpenoids as the primary contributors to floral scent in *C. tracyanum* and uncovered a

significant expansion of gene families related to terpenoid biosynthesis and metabolism in its genome. We further validated the roles of several *CtTPS* genes in the biosynthesis of volatile terpenoids in *C. tracyanum* flowers and explored the potential regulatory networks involving these genes and associated transcription factors. The integration of multi-omics data significantly advances our understanding of floral scent biosynthesis in orchids. Moreover, the availability of a newly generated genome of *C. tracyanum* provides a valuable resource for comparative genomics studies, shedding light on the diversity of ornamental traits and the evolutionary mechanisms of adaptive traits in orchids at the genomic level. These findings are of great importance for breeding fragrant orchids and understanding the adaptation and evolution of orchids.

Materials and methods

Plant materials and growth conditions

All plants used in this study were cultivated in a greenhouse at Kunming Institute of Botany, Kunming, China. The greenhouse conditions were maintained at an air temperature of 18–24 °C, with relative humidity (RH) of 50%–70% and exposure to 30% of full sunlight. For *C. tracyanum*, flower samples were collected at six developmental stages: buds (Bud), the first day of flowering (D + 1), and 7 days (D + 7), 15 days (D + 15), 27 days (D + 27) and 45 days (D + 45) after flowering. A portion of fresh flowers from six developmental stages was used for the measurement of VOCs, and petals at each stage were collected for RNA extraction. In addition, the flowers at D + 15 were further divided into four parts: petal, sepal, labellum and gynostemium. Each part was collected separately for tissue sample subsets. For *C. lowianum*, petals were collected from buds and the full-blooming stage (15 days after flowering). All samples were used for volatolomics detection and RNA extraction. Samples were immediately frozen in liquid nitrogen and stored at –80 °C until required.

Genomic DNA preparation and sequencing

Healthy leaves from the same *C. tracyanum* individual were collected, immediately transferred to liquid nitrogen, and stored at –80 °C until DNA extraction. High-molecular-weight genomic DNA used for both short-read and long-read DNA sequencing was extracted using the cetyltrimethylammonium bromide (CTAB) method. The quality of the DNA was checked using 0.5% agarose gel electrophoresis, and the concentration of the DNA was determined by Qubit fluorimeter (Invitrogen, MA, USA). Genomic DNA sequencing was performed at Wuhan BGI Technology Co., Ltd. on both HiSeq X Ten (Illumina, CA, USA) and Sequel (Pacific Biosciences, CA, USA) platforms. Two Illumina libraries produced 462 Gb of short reads, and 43 SMRTbell libraries produced 403 Gb of long reads (Table S1).

Genome assembly and evaluation

The genome size of *C. tracyanum* was estimated by analysing *k*-mer (17-mer) frequencies in whole-genome sequencing data produced previously, and further experimentally validated by flow cytometry. Contigs were assembled using FALCON v0.5 (Chin *et al.*, 2016), followed by polishing and error correction with PacBio and Illumina sequencing reads by using *Arrow* (SMRTlink release_6.0.0.47841) and Pilon v1.22 (Walker *et al.*, 2014). The TrimDup module in Rabbit was used to remove redundant and heterozygous sequences (Chen *et al.*, 2020). The completeness of the assembled *C. tracyanum* genome was assessed using BUSCOs

v5.5.0 (Manni *et al.*, 2021) against the embryophyta_odb10 dataset.

Pseudochromosome construction with Hi-C data

To construct pseudochromosomes, *C. tracyanum* leaf cells were treated with formaldehyde to fix the cross-linked complexes. The cross-linked DNA was then digested using the restriction enzyme *Mbol*. The sticky ends were labelled with biotin and ligated to form circular chimeric molecule. Purified and fragmented DNA was selected for library construction. Four libraries were sequenced on Illumina HiSeq X Ten and generated 544 Gb of paired-end reads. Uniquely mapped read pairs were analysed using HiC-Pro (Servant *et al.*, 2015). The assembled contigs were anchored and ordered into a chromosome-level assembly using Juicer v1.5 (Durand *et al.*, 2016b) and 3D-DNA v180922 (Dudchenko *et al.*, 2017). Finally, we reviewed and refined the assembly with Juicebox v1.11.08 (Durand *et al.*, 2016a).

Genome annotation

The annotation of repetitive sequences, including tandem repeats and TEs, was performed prior to the gene annotation of *C. tracyanum*. Tandem repeats were identified by Tandem Repeats Finder v4.09 (Benson, 1999). Transposable elements (TEs) were identified by a combination of homologue and *de novo* approaches. RepeatMasker v4.0.7 and RepeatProteinMask were used to annotate the TEs in the genome (Tarailo-Graovac and Chen, 2009). The *de novo* repeat library was predicted by RepeatModeler v1.0.4 and LTR_FINDER v1.0.6 (Xu and Wang, 2007). For *de novo* gene model prediction, AUGUSTUS v3.2.1 (Stanke *et al.*, 2006) and SNAP (Johnson *et al.*, 2008) were employed. For homology-based annotation, protein sequences from five sequenced plants, *Arabidopsis thaliana* (Cheng *et al.*, 2017), *Oryza sativa* (Ouyang *et al.*, 2007), *Asparagus officinalis* (Harkess *et al.*, 2017), *Phalaenopsis aphrodite* (Chao *et al.*, 2018) and *Gastrodia elata* (Yuan *et al.*, 2018) were downloaded and mapped onto the *C. tracyanum* genome using TBLASTN (Camacho *et al.*, 2009) followed by inferring the exon-intron boundaries using Exonerate v2.2.0 (Slater and Birney, 2005). Eight healthy tissues (root, root tip, stem, leaf, bract, flower bud, opening flower and fruit) were collected for RNA extraction and sequences from eight cDNA libraries were used for genome annotation on the HiSeq 2500 platform. Mixed RNA from these eight tissues was used for IsoSeq sequencing. cDNA products were amplified using KAPA HiFi PCR kits (KAPA Biosystems, Cape Town, South Africa), followed by purification using the SMRTbell Template Prep Kit (Pacific Biosciences, CA, USA). Libraries were sequenced on the PacBio Sequel II platform (Table S1). Finally, *de novo*, transcriptome-based and homology-based approaches were combined to predict gene function by Maker v2.31.8 (Holt and Yandell, 2011). Subsequently, the Swiss-Prot, TrEMBL (Boeckmann, 2003), Kyoto Encyclopedia of Genes and Genomes (Ogata *et al.*, 1999), InterPro (Zdobnov and Apweiler, 2001) and Gene Ontology (Ashburner *et al.*, 2000) databases were used for functional annotation of predicted gene models. tRNAscan-SE v1.3.1 (Chan and Lowe, 2019) was used to annotate tRNAs, and BLASTN v2.2.31 was used to identify rRNAs. Rfam/Infernal v1.1 (Nawrocki and Eddy, 2013) was used to predict microRNAs (miRNAs) and small nuclear RNAs (snRNAs) in the genome. Additionally, the functional annotation of genes was performed using eggNOG-mapper v-2.1.12-1/eggNOG DB v5.0.2 (Cantalapiedra *et al.*, 2021; Huerta-Cepas *et al.*, 2018). We then used a custom script for gene GO and KEGG annotations.

Phylogenomic analysis and detection of gene family expansion and contraction

We used protein-coding genes to analyse the phylogenetic relationships between *C. tracyanum* and 18 species, including *Acorus gramineus*, *Amborella trichopoda*, *Ananas comosus*, *Apostasia shenzhenica*, *Arabidopsis thaliana*, *Asparagus officinalis*, *Dendrobium nobile*, *Elaeis guineensis*, *Gastrodia elata*, *Liriodendron tulipifera*, *Nymphaea colorata*, *Oryza sativa*, *Phalaenopsis equestris*, *Rosa chinensis*, *Solanum lycopersicum*, *Vitis vinifera*, *Zostera marina*, *C. mannii* and *C. tracyanum* (Table S7). OrthoFinder v2.5.5 (Emms and Kelly, 2019) was used to perform ortholog inference analysis. As a result, a total of 26 697 gene clusters were identified, including 351 one-to-one single-copy families.

We aligned and trimmed each single-copy orthogroup. Specifically, the sequences in each orthogroup were aligned by MAFFT v7.505 (Katoh and Standley, 2013) and trimmed by Gblocks v0.91b (Talavera and Castresana, 2007). Subsequently, IQ-TREE 2 v2.2.2.7 was used with all 351 aligned loci to estimate a maximum likelihood concatenated tree with 1000 bootstrap replicates (Minh *et al.*, 2020). Then, the MCMCTree program within PAML v4.10.7 was used to determine the divergence times within the generated phylogenetic tree (Yang, 2007). Calibration points, obtained from publications and the TimeTree website (<http://www.timetree.org>), were utilized as normal to constrain the age of the nodes between *A. thaliana* and *R. chinensis* (102.0–112.5 Mya), *A. comosus* and *O. sativa* (94.1–117.0 Mya), *C. mannii* and *C. tracyanum* (9.3–45.0 Mya), *C. tracyanum* and *D. nobile* (12.3–51.0 Mya), *C. tracyanum* and *A. officinalis* (92.5–118.5 Mya), *C. tracyanum* and *A. shenzhenica* (72.2–114.1 Mya), *C. tracyanum* and *A. trichopoda* (179.9–205.0 Mya), *C. tracyanum* and *N. colorata* (168.4–191.6 Mya).

The expansions and contractions of gene families were identified through a comparison of the differences in cluster size between the ancestor and each species employing CAFÉ v5.1 (Mendes *et al.*, 2020). The phylogenetic relationships, divergence times and expansions and contractions of gene families were visualized and edited with the assistance of iTOL v5 (Letunic and Bork, 2021).

Collinearity analysis and whole-genome duplication

Paralogs and orthologs were detected using the Best Reciprocal Hit (BRH) method. Firstly, the proteomes of five orchids (*A. shenzhenica*, *P. equestris*, *C. tracyanum*, *C. mannii* and *D. nobile*) were aligned by BLASTP v2.15.0 (Camacho *et al.*, 2009) with an E-value of a maximal $1e^{-10}$ to search all potential homologous gene pairs of protein sequences. The resulting BLASTP outputs and GFF annotation files of the genomes were processed using MCScanX v2 (Wang *et al.*, 2012) with default parameters to assess the duplications within a species genome and syntenic regions between two different species. All orthologs defined by the BRH method had their corresponding protein-coding DNA sequences (CDS) aligned using MAFFT v7.526 with default parameters (Katoh and Standley, 2013). The number of synonymous substitutions per synonymous site (*Ks*) for each BRH was calculated by KaKs_Calculator v3.0 using the approximate Nei-Gojobori method (Zhang, 2022). The *Ks* distribution was plotted using the ggplot package in R v4.3.2 with the density function.

Synteny plots were drawn based on the results of the previous genome synteny blocks between *C. tracyanum* and *C. mannii*,

C. tracyanum and *D. nobile*. Meanwhile, MCscan (Python version) was used to complement the assessment ([https://github.com/tanghaibao/jvarkit/wiki/MCscan-\(Python-version\)](https://github.com/tanghaibao/jvarkit/wiki/MCscan-(Python-version))) (Tang et al., 2008). Specifically, the LAST output was filtered to remove tandem duplications and weak hits. Single linkage clustering was performed on the LAST output to cluster anchors into synteny blocks (Kielbasa et al., 2011).

The BLASTP results of all protein sequences in the *C. tracyanum* genome, along with the GFF annotation file, were input into MCScanX. The duplicate_gene_classifier module was then employed to identify and classify the gene duplication modes (Wang et al., 2012).

Identification of TPS gene family

We identified TPS gene families from the genomes of 10 species (*A. shenzhenica*, *A. thaliana*, *C. mannii*, *C. tracyanum*, *D. nobile*, *O. sativa*, *P. equestris*, *Populus trichocarpa*, *S. moellendorffii* and *S. lycopersicum*). The hidden Markov models of PF01397 (N-terminal) and PF03936 (C-terminal) (Starks et al., 1997) were downloaded from the Pfam database (<http://pfam.xfam.org/>), and HMMER v3.1b2 software was used to search the TPS gene family sequences. BLAST was utilized to search for genes homologous to TPS genes in *A. thaliana*. These results were combined to identify TPS genes. For *Abies grandis*, due to the lack of genome information, we referred to the results reported in other studies to obtain TPS gene sequences (Huang et al., 2021; Yu et al., 2020). The expression heatmaps and distribution of TPS genes on chromosomes were visualized by TBtools (Chen et al., 2023).

Volatolomics analysis of floral volatile compounds

Different parts of *C. tracyanum* flowers at the full-blooming stage and the petals of *C. lowianum* at the bud and full-blooming stages were frozen, and then ground into powder. A total of 50 ± 1 mg of each sample was placed into a 20-mL headspace bottle, and 10 μ L of 2-Octanol (10 mg/L) was added as an internal standard. In the SPME cycle of the PAL rail system, samples were incubated at 60 °C for 30 min, following a 15-min pre-heat. the desorption time was 4 min. GC–MS analysis was performed using an Agilent 7890 gas chromatograph system coupled with a 5977B mass spectrometer (Agilent technologies, CA, USA). The system utilized a DB-Wax and was injected in splitless mode. Helium was used as the carrier gas. The gas flow rate through the column was 1 mL/min. The initial temperature was kept at 40 °C for 4 min, then raised to 245 °C at a rate of 5 °C/min, and kept for 5 min. Chroma TOF 4.3X software of LECO Corporation (MI, USA) and the NIST database (National Institute of Standards and Technology, MD, USA) were used for raw peak extraction, data baseline filtration and calibration, peak alignment, deconvolution analysis, peak identification, integration and spectrum matching of the peak area. Principal component analysis (PCA) of the identified metabolites was performed using the R package (www.r-project.org). Based on the variable importance in projection (VIP) scores obtained from the OPLS-DA model, metabolites with VIP ≥ 1.0 and *P*-value ≤ 0.05 were defined as significantly changed metabolites (SCMs).

Qualitative and quantitative analysis of volatile terpenoids in fresh flowers of *C. tracyanum*

Floral volatiles were analysed using HS-SPME-GC–MS in HP6890GC/5973MS system (Agilent Technologies, CA, USA). Briefly, fresh flowers of *C. tracyanum* at six developmental stages were excised, enclosed in clean glass vials, and heated in a 45 °C

water bath for 20 min. Then, a 50/30 μ m DVB/CAR/PDMS extraction fibre, fixed on an SPME holder (Sigma-Aldrich, MA, USA) was inserted into the headspace of the glass vial and extracted for 40 min before injection. The GC was equipped with an HP-5MS column (30 m \times 0.25 mm \times 0.25 μ m). Temperature was held at 40 °C for 2 min, raised to 160 °C at 2.5 °C/min and then raised to 280 °C at 15 °C/min. The injector and detector temperatures were maintained at 250 °C. The carrier gas helium flow rate was 1.0 mL/min. Identification of HS-SPME-GC–MS was performed by comparison with n-alkane standards and the NIST 14 (National Institute of Standards and Technology, MD, USA) mass spectral library. Retention indices (RI) of the compounds were determined by using the Kovat index. $RI = 100n + 100 [T_{R(x)} - T_{R(n)}] / [T_{R(n+1)} - T_{R(n)}]$, where $T_{R(x)}$, $T_{R(n)}$ and $T_{R(n+1)}$ represent the retention times of the compound and the normal alkanes with carbon numbers *n* and (*n* + 1), respectively. GC–MS data of volatile compounds values were shown as means \pm SD of triplicates. The standard was diluted with dichloromethane solution in five sequential gradients from the stock solution, and the standard curve was calculated. The characteristic ion peak area response value of the target volatile compound was substituted into the standard curve to calculate the content (Table S22). The odour activity value (OAV) was calculated by the following formula: $OAV = C/OT$, where C is the concentration of the volatile compound and OT is its odour threshold. The odour threshold values have been cited from *Odour thresholds: compilations of odour threshold values in air, water and other media* (van Gemert, 2011). Compounds with $OAV \geq 1$ were considered potential contributors to the floral scent profile. In general, higher OAV values indicate compounds that contribute more significantly to the volatile profile.

Transcriptome construction, assembly and annotation

Total RNA was extracted from petals of *C. tracyanum* and *C. lowianum* at different developmental stages, as well as from different parts of full-blooming flowers of *C. tracyanum* (three biological replicates). Raw data were obtained using the Illumina Novaseq 6000 platform. Raw data in fastq format were first processed through in-house Perl scripts, and clean data were obtained by removing reads containing adapter, reads containing N bases and low-quality reads from the raw data. Then paired-end clean reads were aligned to the reference genome using HISAT2 v2.0.5 (Kim et al., 2019). featureCounts v1.5.0-p3 (Liao et al., 2013) was used to count the read numbers mapped to each gene. FPKM values of each gene were calculated based on the length of the gene and the read count mapped to the gene. Differentially expressed genes (DEGs) were screened by $|\log_2(\text{foldchange})| \geq 1$ and $\text{padj} \leq 0.05$ by applying DESeq2 R package 1.20.0 (Love et al., 2014).

Enrichment analysis

To further understand the functions of expanded and contracted genes in the *C. tracyanum* genome and differentially expressed genes in the transcriptome, we first collated GO and KEGG annotations of these genes. Then, clusterProfiler 4.0 (Wu et al., 2021) was used to perform enrichment analysis and visualization.

Identification of structural genes and TFs related to terpenoid biosynthesis

We identified genes encoding enzymes related to terpenoid biosynthesis according to the KEGG annotation (<https://www.genome.jp/kegg/annotation/>). The related genes in the

transcriptomes were identified by using a local TBLASTN algorithm with an *E*-value cut-off of $1e^{-5}$ and confirmed using BLAST on the National Center for Biotechnology Information website (<https://www.ncbi.nlm.nih.gov>). To isolate TFs and regulators, all proteins obtained from transcriptomes were annotated and classified by using iTAK (Zheng *et al.*, 2016). Among these, 642 genes annotated as bHLH, bZIP, EIL, NAC, MYB, WRKY and AP2/ERF were isolated. Heatmaps of genes related to terpenoid biosynthesis pathways were generated by TBtools (Chen *et al.*, 2023).

Weighted gene co-expression network analysis (WGCNA)

WGCNA was performed using the WGCNA R package (Langfelder and Horvath, 2008). We first screened genes from the transcriptome of petals of *C. tracyanum* at different developmental stages, selecting the top 75% with a median absolute deviation ≥ 0.01 . After filtering, the abundance of 25 364 genes and 12 terpenoid metabolites was used to build a co-expression network by calculating correlation coefficients. The soft threshold power of the correlation network was set to 18, the minimum module size was set to 30, and the cutHeight for merging modules was set to 0.5. The eigengene value was calculated for each module and used to test the association with each metabolite. The co-expression network was visualized with Cytoscape v3.8.2 (Shannon *et al.*, 2003).

Quantitative real-time PCR analysis

All qRT-PCR primers were designed using the Primer Premier 6.0 program (Primer Biosoft Inc., QC, Canada) and are listed in Table S21. Quantitative real-time PCR was carried out using the Bio-Rad real-time PCR System (Bio-Rad, CA, USA). We configured the reaction mixtures according to the manufacturer's instructions of iTaq Universal SYBR[®] Green Supermix (Bio-Rad, CA, USA). Relative gene expression was calculated using the $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen, 2001). Three biological replicates were used for each analysis.

Subcellular localization of CtTPS proteins

The intact ORF sequences of *CtTPS* genes were integrated into a pCambia Super 1300-eGFP vector by replacing the termination codons with sequences encoding enhanced green fluorescent protein (eGFP), under the control of a mannopine synthetase (*mas*) promoter (Table S21). The resulting vector construct was transformed into *Agrobacterium tumefaciens* (strain GV3101) and used to infiltrate 4-week-old *N. benthamiana* leaves. Three days post-infiltration, protoplasts were isolated from the tobacco leaves. Protoplasts expressing eGFP fusions of *CtTPS* proteins were visualized by confocal laser scanning microscopy LSM 900 (Carl Zeiss, Jena, Germany).

In vitro characterization of recombinant CtTPS proteins

DNA fragments of *CtTPS* genes were inserted into a pET-32a vector. The recombinant plasmids were then transformed into *E. coli* strain BL21 (DE3). Recombinant proteins were induced by the application of 0.2 mM isopropyl β -D-thiogalactoside (IPTG) at 16 °C overnight. Induced cells were harvested by centrifugation, resuspended in Tris-buffered saline and disrupted by sonication on ice. After centrifugation, the supernatants were purified using Ni-NTA Agarose (Qiagen, Venlo, Netherlands). Purified proteins were collected and concentrated before enzyme assays. The purity of the isolated proteins was verified by densitometry of SDS-PAGE gels after Coomassie Brilliant Blue staining. Protein

concentrations were estimated using the Detergent Compatible Bradford Protein Assay Kit (Beyotime, Shanghai, China).

Assays for recombinant *CtTPS* protein activity were carried out in a 500 μ L assay buffer (50 mM HEPES, pH 7.2, 10% [v/v] glycerol, 10 mM $MgCl_2$, 1.25 mM $MnCl_2$, 5 mM DTT) containing 10 μ g purified recombinant *CtTPS* proteins and 20 μ g GPP/FPP. The mixture was incubated at 30 °C for 2 h and then mixed vigorously at 60 °C for 5 min to obtain enzymatic products. The catalytic products were collected using the DVB/CAR/PDMS headspace sampler and analysed by GC-MS. Extracts from *E. coli* transformed with the pET-32a empty vector served as controls and standard samples were analysed under the same conditions (Table S19). Unfortunately, *CtTPS1* failed to be induced as a soluble protein in the supernatant, and therefore, no enzyme activity assay was performed.

Dual-luciferase assays

The full-length CDS of *CtTF* genes was cloned into the pM999-eGFP vector under the control of the CaMV 35S promoter as effectors. The promoter fragments of *CtTPS* genes were ligated into the binary vector pGreenII-0800-LUC as the double-reporter vector (Hellens *et al.*, 2005). The pM999-eGFP vector without *CtTF* genes was used as a negative control. Primers used in this assay are listed in Table S21. The constructed effector and reporter vectors were co-transformed into tobacco (*N. benthamiana*) leaf protoplasts using the polyethylene glycol 4000 (PEG 4000) method, as previously described (Abel and Theologis, 1994). The transformed protoplasts were incubated at 23 °C for 16 h in darkness, and dual-luciferase assays were performed using the Dual-Luciferase[®] Reporter Assay System (Vazyme, Nanjing, China). Luciferase activity was measured using the Infinite[®] 200 PRO plate reader (TECAN Group, Switzerland). Finally, the LUC:REN ratio was calculated and normalized to the control vector as the final value. The pM999-eGFP empty vector was used as a negative control. At least three replicates were used for each dual-luciferase assay.

Yeast one-hybrid assays

Promoter fragments within 2 kb of *CtTPS* genes were amplified via PCR using corresponding primers (Table S21). The amplified products were verified by sanger sequencing and cloned into the pAbAi vector as the baits, and the full-length CDS of *CtTF* genes were subcloned from the pM999-eGFP vector into the pGADT7 vector to construct the prey. The linearized recombinant pAbAi vector was first transformed into the Y1H Gold yeast strain via the homologous recombination method and selected on SD/-Ura medium. Subsequently, the recombinant pGADT7 vector was introduced into the yeast strain, and transformants were further selected for resistance concentrations using SD/-Leu/-Ura medium with proper concentrations of aureobasidin A. According to the growth ability of the yeast colonies, the protein-DNA interaction was determined. The pGADT7 vector without the *CtTF* gene sequence was used as a negative control.

Electrophoretic mobility shift assays

Electrophoretic mobility shift assays (EMSAs) were performed using an EMSA/Gel-Shift kit (Beyotime) according to the manufacturer's instructions. The coding sequences of *CtAP2/ERF1*, *CtZIP1*, *CtMYB2*, *CtMYB3* and *CtAP2/ERF4* were amplified and cloned into the pET-32a vector for fusion with a His tag. The purification of recombinant proteins was performed as described in the section "In vitro characterization of recombinant *CtTPS* proteins". Oligonucleotide probes were 5' end-labelled with biotin. For

competition assays, unlabeled competitors were added to the reaction at 50- and 200-fold (sequences are listed in Table S21). The DNA–protein complexes were separated by gel electrophoresis and transferred to a nylon membrane (GE Healthcare, Chicago, IL, USA). After UV crosslinking, the biotin signal was detected according to the manufacturer's instructions. Experiments were repeated at least twice with consistent results.

Statistical analysis

One-way ANOVA was used to determine the significance of gene expression levels at different developmental stages and in different parts of *C. tracyanum*. The threshold for statistical significance was set at $P < 0.05$. Statistical analysis of gene expression levels between *C. tracyanum* and *C. lowianum* was performed using Student's *t*-test ($*P < 0.05$, $**P < 0.01$ and $***P < 0.001$). For the dual-luciferase assay, the significance of the LUC/REN ratio was calculated by Student's *t*-test compared with the negative control ($*P < 0.05$, $**P < 0.01$ and $***P < 0.001$). Data are presented as means (\pm SD) from three replicates. All statistical analyses were performed using R software.

Acknowledgements

We thank Prof. Wei Xu (Kunming Institute of Botany, Chinese Academy of Sciences) for providing the plasmid pGreenII-0800-LUC for this study. We also thank Dr. Fei Li (Kunming Institute of Botany, Chinese Academy of Sciences) for his assistance with the GC–MS analysis. We thank Yanyu Zhang (Yunnan University) for helping with the dual-luciferase assays. We are very grateful to Dr. Raymond Porter for his help in polishing the language of the manuscript.

Author contributions

S.-B.Z., A.Z. and J.-B.Y. conceived the project and designed the research. M.T., N.L. and Z.-S.H. performed the analyses of the genome and transcriptome sequence. M.T., X.-M.D. and T.-Y.G. collected the samples and performed the experiments. M.T., N.L. and Z.-S.H. wrote the manuscript. S.-B.Z., A.Z. and J.-B.Y. revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Key Research and Development Program of China (2024YFF1306703), the National Natural Science Foundation of China (32170393), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB31000000), the Key Research and Development Program of Yunnan Province (202403AC100032), the Yunnan Fundamental Research Project (202201AU070123 and 202301AT070306) and the High-level Talent Support Plan of Yunnan Province (YNWR-CYJS-2020-023).

Competing interests

The authors declare that they have no competing interests.

Data availability statement

All raw sequencing reads have been deposited in the NCBI Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) under project PRJNA1145103.

References

- Abel, S. and Theologis, A. (1994) Transient transformation of *Arabidopsis* leaf protoplasts: a versatile experimental system to study gene expression. *Plant J.* **5**, 421–427.
- Ai, Y., Li, Z., Sun, W.-H., Chen, J., Zhang, D., Ma, L. *et al.* (2021) The *Cymbidium* genome reveals the evolution of unique morphological traits. *Hortic. Res.* **8**, 255.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P. *et al.* (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29.
- Bao, T., Kimani, S., Li, Y., Li, H., Yang, S., Zhang, J., Wang, Q. *et al.* (2023) Allelic variation of terpene synthases drives terpene diversity in the wild species of the *Freesia* genus. *Plant Physiol.* **192**, 2419–2435.
- Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580.
- Bergougnoux, V., Caissard, J.-C., Jullien, F., Magnard, J.-L., Scalliet, G., Cock, J.M., Huguency, P. *et al.* (2007) Both the adaxial and abaxial epidermal layers of the rose petal emit volatile scent compounds. *Planta* **226**, 853–866.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.-C., Estreicher, A., Gasteiger, E. *et al.* (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365–370.
- Byers, K.J.R.P., Vela, J.P., Peng, F., Riffell, J.A. and Bradshaw, H.D. (2014) Floral volatile alleles can contribute to pollinator-mediated reproductive isolation in monkeyflowers (*Mimulus*). *Plant J.* **80**, 1031–1042.
- Cai, J., Liu, X., Vanneste, K., Proost, S., Tsai, W.-C., Liu, K.-W. *et al.* (2014) The genome sequence of the orchid *Phalaenopsis equestris*. *Nat. Genet.* **47**, 65–72.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421.
- Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P. and Huerta-Cepas, J. (2021) eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829.
- Chan, P.P. and Lowe, T.M. (2019) tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol. Biol.* **1962**, 1–14.
- Chao, Y.-T., Chen, W.-C., Chen, C.-Y., Ho, H.-Y., Yeh, C.-H., Kuo, Y.-T., Su, C.L. *et al.* (2018) Chromosome-level assembly, genetic and physical mapping of *Phalaenopsis aphrodite* genome provides new insights into species adaptation and resources for orchid breeding. *Plant Biotechnol. J.* **16**, 2027–2041.
- Chen, F., Tholl, D., Bohlmann, J. and Pichersky, E. (2011) The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* **66**, 212–229.
- Chen, H., Li, G., Köllner, T.G., Jia, Q., Gershenzon, J. and Chen, F. (2014) Positive Darwinian selection is a driving force for the diversification of terpenoid biosynthesis in the genus *Oryza*. *BMC Plant Biol.* **14**, 239.
- Chen, S.-P., Sun, W.-H., Xiong, Y.-F., Jiang, Y.-T., Liu, X.-D., Liao, X.-Y. *et al.* (2020) The *Phoebe* genome sheds light on the evolution of magnoliids. *Hortic. Res.* **7**, 146.
- Chen, C., Wu, Y., Li, J., Wang, X., Zeng, Z., Xu, J., Liu, Y. *et al.* (2023) TBtools-II: A “one for all, all for one” bioinformatics platform for biological big-data mining. *Mol. Plant* **16**, 1733–1742.
- Cheng, C.-Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S. and Town, C.D. (2017) Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J.* **89**, 789–804.
- Chin, C.-S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C. *et al.* (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **13**, 1050–1054.
- Christenhusz, M.J.M. and Byng, J.W. (2016) The number of known plants species in the world and its annual increase. *Phytotaxa* **261**, 201–217.
- Christianson, D.W. (2006) Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* **106**, 3412–3442.
- Chuang, Y.-C., Hung, Y.-C., Tsai, W.-C., Chen, W.-H. and Chen, H.-H. (2018) PbbHLH4 regulates floral monoterpene biosynthesis in *Phalaenopsis* orchids. *J. Exp. Bot.* **69**, 4363–4377.

- Colquhoun, T.A. and Clark, D.G. (2011) Unraveling the regulation of floral fragrance biosynthesis. *Plant Signal. Behav.* **6**, 378–381.
- Conart, C., Bomzan, D.P., Huang, X.-Q., Bassard, J.-E., Paramita, S.N., Saint-Marcoux, D. *et al.* (2023) A cytosolic bifunctional geranyl/farnesyl diphosphate synthase provides MVA-derived GPP for geraniol biosynthesis in rose flowers. *Proc. Natl. Acad. Sci. USA* **120**, e2221440120.
- Cozzolino, S. and Widmer, A. (2005) Orchid diversity: an evolutionary consequence of deception? *Trends Ecol. Evol.* **20**, 487–494.
- Cancino, A.D.M. and Damon, A. (2007) Fragrance analysis of euglossine bee pollinated orchids from Soconusco, south-east Mexico. *Plant Species Biol.* **22**, 129–134.
- Dudareva, N. and Pichersky, E. (2008) Metabolic engineering of plant volatiles. *Curr. Opin. Biotechnol.* **19**, 181–189.
- Dudareva, N., Martin, D., Kish, C.M., Kolosova, N., Gorenstein, N., Fäldt, J., Miller, B. *et al.* (2003) (*E*)- β -ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: function and expression of three terpene synthase genes of a new terpene synthase subfamily. *Plant Cell* **15**, 1227–1241.
- Dudareva, N., Klempien, A., Muhlemann, J.K. and Kaplan, I. (2013) Biosynthesis, function and metabolic engineering of plant volatile organic compounds. *New Phytol.* **198**, 16–32.
- Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S. *et al.* (2017) De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95.
- Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S. and Aiden, E.L. (2016a) Juicebox provides a visualization system for Hi-C contact maps with Unlimited Zoom. *Cell Syst.* **3**, 99–101.
- Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S. and Aiden, E.L. (2016b) Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98.
- Emms, D.M. and Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238.
- Estévez, J.M., Cantero, A., Reindl, A., Reichler, S. and León, P. (2001) 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. *J. Biol. Chem.* **276**, 22901–22909.
- Fan, W., He, Z.-S., Zhe, M., Feng, J.-Q., Zhang, L., Huang, Y. *et al.* (2023) High-quality *Cymbidium mannii* genome and multifaceted regulation of crassulacean acid metabolism in epiphytes. *Plant. Commun.* **4**, 100564.
- van Gemert, L.J. (2011) Odour thresholds: compilations of odour threshold values in air, water and other media. The Netherlands: Oliemans, Punter & Partners B.V.
- Govaerts, R., Nic Lughadha, E., Black, N., Turner, R. and Paton, A. (2021) The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Sci. Data* **8**, 215.
- Harkess, A., Zhou, J., Xu, C., Bowers, J.E., Van der Hulst, R., Ayyampalayam, S. *et al.* (2017) The asparagus genome sheds light on the origin and evolution of a young Y chromosome. *Nat. Commun.* **8**, 1279.
- Heil, M. and Silva Bueno, J.C. (2007) Within-plant signaling by volatiles leads to induction and priming of an indirect plant defense in nature. *Proc. Natl. Acad. Sci. USA* **104**, 5467–5472.
- Hellens, R., Allan, A., Friel, E., Bolitho, K., Grafton, K., Templeton, M. *et al.* (2005) Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. *Plant Methods* **1**, 13.
- Holopainen, J.K. and Gershenzon, J. (2010) Multiple stress factors and the emission of plant VOCs. *Trends Plant Sci.* **15**, 176–184.
- Holt, C. and Yandell, M. (2011) MAKER2: an annotation pipeline and genome database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491.
- Hong, G.-J., Xue, X.-Y., Mao, Y.-B., Wang, L.-J. and Chen, X.-Y. (2012) *Arabidopsis* MYC2 interacts with DELLA proteins in regulating sesquiterpene synthase gene expression. *Plant Cell* **24**, 2635–2648.
- Huang, M., Sanchez-Moreiras, A.M., Abel, C., Sohrabi, R., Lee, S., Gershenzon, J. and Tholl, D. (2011) The major volatile organic compound emitted from *Arabidopsis thaliana* flowers, the sesquiterpene (*E*)- β -caryophyllene, is a defense against a bacterial pathogen. *New Phytol.* **193**, 997–1008.
- Huang, L.-M., Huang, H., Chuang, Y.-C., Chen, W.-H., Wang, C.-N. and Chen, H.-H. (2021) Evolution of terpene synthases in Orchidaceae. *Int. J. Mol. Sci.* **22**, 6947.
- Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H. *et al.* (2018) eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314.
- Jersáková, J., Johnson, S.D. and Kindlmann, P. (2006) Mechanisms and evolution of deceptive pollination in orchids. *Biol. Rev. Camb. Philos. Soc.* **81**, 219–235.
- Jia, Q., Brown, R., Köllner, T.G., Fu, J., Chen, X., Wong, G.K.-S. *et al.* (2022) Origin and early evolution of the plant terpene synthase family. *Proc. Natl. Acad. Sci. USA* **119**, e2100361119.
- Jiang, S.-Y., Jin, J., Sarojam, R. and Ramachandran, S. (2019) A Comprehensive survey on the terpene synthase gene family provides new insight into its evolutionary patterns. *Genome Biol. Evol.* **11**, 2078–2098.
- Jiao, Y., Li, J., Tang, H. and Paterson, A.H. (2014) Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. *Plant Cell* **26**, 2792–2802.
- Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J. and de Bakker, P.I.W. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24**, 2938–2939.
- Jørgensen, K., Rasmussen, A.V., Morant, M., Nielsen, A.H., Bjørnholt, N., Bak, S. and Möller, B.L. (2005) Metabolite formation and metabolic channeling in the biosynthesis of plant natural products. *Curr. Opin. Plant Biol.* **8**, 280–291.
- Karunanithi, P.S. and Zerbe, P. (2019) Terpene synthases as metabolic gatekeepers in the evolution of plant terpenoid chemical diversity. *Front. Plant Sci.* **10**, 1166.
- Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780.
- Kegge, W. and Pierik, R. (2010) Biogenic volatile organic compounds and plant competition. *Trends Plant Sci.* **15**, 126–132.
- Kielbasa, S.M., Wan, R., Sato, K., Horton, P. and Frith, M.C. (2011) Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493.
- Kim, D., Paggi, J.M., Park, C., Bennett, C. and Salzberg, S.L. (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915.
- Lai, Y., Ma, J., Zhang, X., Xuan, X., Zhu, F., Ding, S., Shang, F. *et al.* (2024) High-quality chromosome-level genome assembly and multi-omics analysis of rosemary (*Salvia rosmarinus*) reveals new insights into the environmental and genome adaptation. *Plant Biotechnol. J.* **22**, 1833–1847.
- Langfelder, P. and Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559.
- Letunic, I. and Bork, P. (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296.
- Li, Y., Bao, T., Zhang, J., Li, H., Shan, X., Yan, H. *et al.* (2025) The coordinated interaction or regulation between floral pigments and volatile organic compounds. *Hortic. Plant J.* **11**, 463–485.
- Liao, Y., Smyth, G.K. and Shi, W. (2013) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930.
- Liu, Z., Chen, S. and Ru, Z. (2006) *The genus Cymbidium in China*. Beijing: Science Press.
- Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408.
- Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.
- Maffei, M.E. (2010) Sites of synthesis, biochemistry and functional role of plant volatiles. *S. Afr. J. Bot.* **76**, 612–631.
- Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A. and Zdobnov, E.M. (2021) BUSCO Update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654.
- Martin, D.M., Aubourg, S., Schouwey, M.B., Daviet, L., Schalk, M., Toub, O. *et al.* (2010) Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. *BMC Plant Biol.* **10**, 226.

- McGarvey, D.J. and Croteau, R. (1995) Terpenoid metabolism. *Plant Cell* **7**, 1015–1026.
- Mendes, F.K., Vanderpool, D., Fulton, B. and Hahn, M.W. (2020) CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**, 5516–5518.
- Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A. and Lanfear, R. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534.
- Nagegowda, D.A., Gutensohn, M., Wilkerson, C.G. and Dudareva, N. (2008) Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. *Plant J.* **55**, 224–239.
- Nawrocki, E.P. and Eddy, S.R. (2013) Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935.
- Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H. and Kanehisa, M. (1999) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27**, 29–34.
- Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F. et al. (2007) The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Res.* **35**, D883–D887.
- Paetzold, H., Garms, S., Bartram, S., Wieczorek, J., Urós-Gracia, E.-M., Rodríguez-Concepción, M., Boland, W. et al. (2010) The isogene 1-deoxy-D-xylulose 5-phosphate Synthase 2 controls isoprenoid profiles, precursor pathway allocation, and density of tomato trichomes. *Mol. Plant* **3**, 904–916.
- Pérez-Escobar, O.A., Bogarín, D., Przelomska, N.A.S., Ackerman, J.D., Balbuena, J.A., Bellot, S., Bühlmann, R.P. et al. (2024) The origin and speciation of orchids. *New Phytol.* **242**, 700–716.
- Picazo-Aragón, J., Terrab, A. and Balao, F. (2020) Plant volatile organic compounds evolution: transcriptional regulation, epigenetics and polyploidy. *Int. J. Mol. Sci.* **21**, 8956.
- Pichersky, E., Noel, J.P. and Dudareva, N. (2006) Biosynthesis of plant volatiles: nature's diversity and ingenuity. *Science* **311**, 808–811.
- Ramya, M., Park, P.H., Chuang, Y.-C., Kwon, O.K., An, H.R., Park, P.M. et al. (2019) RNA sequencing analysis of *Cymbidium goeringii* identifies floral scent biosynthesis related genes. *BMC Plant Biol.* **19**, 337.
- Re, E.B., Jones, D. and Learned, R.M. (1995) Co-expression of native and introduced genes reveals cryptic regulation of HMG CoA reductase expression in *Arabidopsis*. *Plant J.* **7**, 771–784.
- Samad, A.F.A., Sajad, M., Nazaruddin, N., Fauzi, I.A., Murad, A.M.A., Zainal, Z. and Ismail, I. (2017) MicroRNA and transcription factor: key players in plant regulatory network. *Front. Plant Sci.* **8**, 565.
- Serna-Sánchez, M.A., Pérez-Escobar, O.A., Bogarín, D., Torres-Jimenez, M.F., Alvarez-Yela, A.C., Arcila-Galvis, J.E. et al. (2021) Plastid phylogenomics resolves ambiguous relationships within the orchid family and provides a solid timeframe for biogeography and macroevolution. *Sci. Rep.* **11**, 6858.
- Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P. et al. (2015) HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259.
- Shang, J., Tian, J., Cheng, H., Yan, Q., Li, L., Jamal, A. et al. (2020) The chromosome-level wintersweet (*Chimonanthus praecox*) genome provides insights into floral scent biosynthesis and flowering in winter. *Genome Biol.* **21**, 200.
- Shang, J., Feng, D., Liu, H., Niu, L., Li, R., Li, Y. et al. (2024) Evolution of the biosynthetic pathways of terpene scent compounds in roses. *Curr. Biol.* **34**, 3550–3563.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504.
- Slater, G.S.C. and Birney, E. (2005) Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31.
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S. and Morgenstern, B. (2006) AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439.
- Starks, C.M., Back, K., Chappell, J. and Noel, J.P. (1997) Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science* **277**, 1815–1820.
- Sun, Y., Chen, G.-Z., Huang, J., Liu, D.-K., Xue, F., Chen, X.-L. et al. (2021) The *Cymbidium goeringii* genome provides insight into organ development and adaptive evolution in orchids. *Ornam. Plant Res.* **1**, 10.
- Talavera, G. and Castresana, J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577.
- Tan, H., Xiao, L., Gao, S., Li, Q., Chen, J., Xiao, Y., Ji, Q. et al. (2015) *TRICHOME AND ARTEMISININ REGULATOR 1* is required for trichome development and artemisinin biosynthesis in *Artemisia annua*. *Mol. Plant* **8**, 1396–1411.
- Tang, H., Bowers, J.E., Wang, X., Ming, R., Alam, M. and Paterson, A.H. (2008) Synteny and collinearity in plant genomes. *Science* **320**, 486–488.
- Tarailo-Graovac, M. and Chen, N. (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **4**, 4.10.1–4.10.14.
- Tholl, D. (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr. Opin. Plant Biol.* **9**, 297–304.
- Tholl, D. (2015) Biosynthesis and biological functions of terpenoids in plants. *Adv. Biochem. Eng. Biotechnol.* **148**, 63–106.
- Tsai, W.-C., Dievart, A., Hsu, C.-C., Hsiao, Y.-Y., Chiou, S.-Y., Huang, H. and Chen, H.-H. (2017) Post genomics era for orchid research. *Bot. Stud.* **58**, 61.
- Van de Peer, Y., Maere, S. and Meyer, A. (2009) The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10**, 725–732.
- Vogel, S. (1983) Ecophysiology of zoophilic pollination. In *Physiological Plant Ecology III*, pp. 559–624. Berlin, Heidelberg: Springer.
- Vogt, T. (2010) Phenylpropanoid biosynthesis. *Mol. Plant* **3**, 2–20.
- Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S. et al. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963.
- Walter, M.H., Hans, J. and Strack, D. (2002) Two distantly related genes encoding 1-deoxy-D-xylulose 5-phosphate synthases: differential regulation in shoots and apocarotenoid-accumulating mycorrhizal roots. *Plant J.* **31**, 243–254.
- Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X. et al. (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49.
- Wang, L., Albert, N.W., Zhang, H., Arathoon, S., Boase, M.R., Ngo, H., Schwinn, K.E. et al. (2014) Temporal and spatial regulation of anthocyanin biosynthesis provide diverse flower colour intensities and patterning in *Cymbidium* orchid. *Planta* **240**, 983–1002.
- Waterman, R.J. and Bidartondo, M.I. (2008) Deception above, deception below: linking pollination and mycorrhizal biology of orchids. *J. Exp. Bot.* **59**, 1085–1096.
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z. et al. (2021) clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovations* **2**, 100141.
- Xu, Z. and Wang, H. (2007) LTR-FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268.
- Xu, H., Bohman, B., Wong, D.C.J., Rodriguez-Delgado, C., Scaffidi, A., Flematti, G.R. et al. (2017) Complex sexual deception in an orchid is achieved by co-opting two independent biosynthetic pathways for pollinator attraction. *Curr. Biol.* **27**, 1867–1877.
- Xu, Q., Niu, S.-C., Li, K.-L., Zheng, P.-J., Zhang, X.-J., Jia, Y. et al. (2022) Chromosome-scale assembly of the *Dendrobium nobile* genome provides insights into the molecular mechanism of the biosynthesis of the medicinal active ingredient of *Dendrobium*. *Front. Genet.* **13**, 844622.
- Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591.
- Yang, F.-X., Gao, J., Wei, Y.-L., Ren, R., Zhang, G.-Q., Lu, C.-Q. et al. (2021) The genome of *Cymbidium sinense* revealed the evolution of orchid traits. *Plant Biotechnol. J.* **19**, 2501–2516.
- Yu, Z., Zhao, C., Zhang, G., Teixeira da Silva, J.A. and Duan, J. (2020) Genome-wide identification and expression profile of TPS gene family in *Dendrobium officinale* and the role of *DoTPS10* in linalool biosynthesis. *Int. J. Mol. Sci.* **21**, 5419.
- Yuan, Y., Jin, X., Liu, J., Zhao, X., Zhou, J., Wang, X. et al. (2018) The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. *Nat. Commun.* **9**, 1615.
- Zdobnov, E.M. and Apweiler, R. (2001) InterProScan – an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848.
- Zhang, Z. (2022) KaKs_calculator 3.0: Calculating selective pressure on coding and non-coding sequences. *Genomics Proteomics Bioinformatics* **20**, 536–540.

- Zhang, G.-Q., Xu, Q., Bian, C., Tsai, W.-C., Yeh, C.-M., Liu, K.-W. *et al.* (2016) The *Dendrobium catenatum* Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. *Sci. Rep.* **6**, 19029.
- Zhang, G.-Q., Liu, K.-W., Li, Z., Lohaus, R., Hsiao, Y.-Y., Niu, S.-C. *et al.* (2017) The *Apostasia* genome and the evolution of orchids. *Nature* **549**, 379–383.
- Zhang, Y., Li, C., Wang, S., Yuan, M., Li, B., Niu, L. and Shi, Q. (2021) Transcriptome and volatile compounds profiling analyses provide insights into the molecular mechanism underlying the floral fragrance of tree peony. *Ind. Crop. Prod.* **162**, 113286.
- Zheng, Y., Jiao, C., Sun, H., Rosli, H.G., Pombo, M.A., Zhang, P., Banf, M. *et al.* (2016) iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* **9**, 1667–1670.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

- Table S1** Summary of sequenced data from multiple platforms in this study.
- Table S2** Summary of *C. tracyanum* genome assemblies.
- Table S3** Repetitive elements identified in *C. tracyanum* genome.
- Table S4** Summary of annotated protein-coding genes in *C. tracyanum*.
- Table S5** Functional annotation in *C. tracyanum* genome.
- Table S6** BUSCO estimates of gene annotation completeness.
- Table S7** Species information used in phylogenomic analysis.
- Table S8** GO enrichment analysis of the expanded genes in *C. tracyanum* genome.
- Table S9** *TPS* genes used for phylogenetic tree construction.
- Table S10** Duplication types of *TPS* genes in the genome of *C. tracyanum*.
- Table S11** Summary of the 230 detected metabolites in volatilomics.
- Table S12** Differential metabolites in *C. tracyanum* and *C. lowianum* petals at the full-blooming stage.
- Table S13** Terpenoids detected in fresh flowers of *C. tracyanum* at six developmental stages.
- Table S14** Transcriptome reads mapping rate.
- Table S15** The genes annotated as putative enzymes involved in the terpenoid biosynthesis pathway.
- Table S16** Expression statistics of genes with an average FPKM >1 in different floral parts at the full-blooming stage of *C. tracyanum*.
- Table S17** Information about candidate *CtTPS* genes, transcription factors and reference genes.
- Table S18** *TPS* proteins from other plant species used in phylogenetic analysis.
- Table S19** *In vitro* enzymatic reaction products and authentic standard information.

Table S20 GO enrichment analysis of the contracted genes in *C. tracyanum* genome.

Table S21 Primers and probes used in this study.

Table S22 Information about standard curves.

Figure S1 Hi-C map of the *C. tracyanum* genome showing genome-wide all-by-all interactions.

Figure S2 Estimation of *C. tracyanum* genome size by flow cytometry and *k*-mer analysis.

Figure S3 The estimated time distribution of LTR-RTs insertions.

Figure S4 Gene pairs in the synteny region among *C. tracyanum*, *C. mannii* and *D. nobile*.

Figure S5 Volatile compound profiles in *C. tracyanum* and *C. lowianum*.

Figure S6 Gas chromatogram of floral volatiles from the fresh flowers of *C. tracyanum* and *C. lowianum*.

Figure S7 Scanning Electron Microscope (SEM) images of different parts of *C. tracyanum* flowers.

Figure S8 Enrichment analysis of differentially expressed genes in the petals of *C. tracyanum* at Bud and D+1 stages.

Figure S9 Schematic representation of the distribution of *TPS* genes on *C. tracyanum* chromosomes and together with the locations of key genes involved in terpenoid biosynthesis.

Figure S10 Expression levels of genes encoding key enzymes in the terpenoid backbone biosynthesis pathway in the petals of *C. tracyanum* and *C. lowianum* at different developmental stages analyzed by qRT-PCR.

Figure S11 Phylogenetic analysis of HMGR proteins from *C. tracyanum* and other plants.

Figure S12 Phylogenetic analysis of DXS proteins from *C. tracyanum* and other plants.

Figure S13 Expression profiles of putative genes encoding enzymes for terpenoid biosynthesis in different parts of *C. tracyanum* flowers at full-blooming stage.

Figure S14 Heatmap of the expression patterns of *TPS* genes from different parts of *C. tracyanum* flowers at the full-blooming stage.

Figure S15 Hierarchical cluster tree displaying 25 modules of co-expressed genes.

Figure S16 Expression levels of eight *CtTPS* genes in different parts of *C. tracyanum* flowers (at the D+15 stage) by qRT-PCR analysis.

Figure S17 Expression levels of candidate TFs in different parts of *C. tracyanum* flowers (at the D+15 stage) by qRT-PCR analysis.

Figure S18 SDS-PAGE analysis of recombinant terpene synthase proteins.

Figure S19 Subcellular localization of free eGFP and six *CtTF*-eGFP fusions in tobacco leaf protoplasts.

Figure S20 SDS-PAGE analysis of recombinant transcription factor proteins.

Figure S21 GO pathway enrichment distribution of the contracted genes in *C. tracyanum* genome.