Acta Pharmaceutica Sinica B 2026;16(8):5487−5503

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com

ORIGINAL ARTICLE

Expansion and functional diversification of terpene synthases shape volatile terpenoid landscape in *Cannabis sativa*

Yaolei Mi[a,b,c,d,†], Mofan Zhang[b,d,†], Kaiquan Zhang[a], Jun Li[b,d], Song Yan[e], Xuewen Zhu[b,d], Liwei Wang[a], Yupeng Du[a], Sifan Wang[b,d], Xue Cao[b,d], Shilin Chen[f,*], Zhichao Xu[a,*], Wei Sun[b,d,*]

[a]*Key Laboratory of National Forestry and Grassland Administration on Chinese Herbal Medicine, College of Life Science, Northeast Forestry University, Harbin 150040, China*
[b]*State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China*
[c]*Institute of Chinese Materia Medica and Artemisinin Research Center, China Academy of Chinese Medical Sciences, Beijing 100700, China*
[d]*Beijing Key Laboratory of Psychoactive Substances Discovery and Control in Chinese Herbal Medicines, Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China*
[e]*Pharmacy of College, Heilongjiang University of Chinese Medicine, Harbin 150040, China*
[f]*Institute of Herbgenomics, Innovative Institute of Chinese Medicine and Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China*

Received 22 September 2025; received in revised form 9 March 2026; accepted 10 March 2026

KEY WORDS

Terpene synthase;
Cannabis sativa;
Volatile terpenes;
Genetic expansion;
Haplotype-resolved genome

Abstract Terpenoids are key specialized metabolites in *Cannabis sativa*, shaping cultivar-specific aromas and potentially modulating cannabinoid effects. This study provided a comprehensive analysis of the terpene synthase (TPS) gene family in *C. sativa*, integrating haplotype-resolved genomic data, volatile terpene profiling, transcriptomics, and functional assays. The comprehensive volatile terpene profiling across 28 spatiotemporal samples spanning weekly developmental intervals and distinct maturity stages from six cultivars identified 227 cannabis volatile terpenes, including 88 monoterpenes and 139 sesquiterpenes, exhibiting distinct tissue, developmental stage, and cultivar-specific patterns. Comparative expression and co-expression network analysis revealed coordinated regulation between MEP/MVA pathways, TPS, and cannabinoid genes, underscoring a shared metabolic foundation. Genome

*Corresponding authors.
E-mail addresses: slchen@cdutcm.edu.cn (Shilin Chen), zcxu@nefu.edu.cn (Zhichao Xu), wsun@icmm.ac.cn (Wei Sun).
[†]These authors made equal contributions to this work.
Peer review under the responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

https://doi.org/10.1016/j.apsb.2026.03.035

annotation identified 41 full-length CsTPSDK genes, exhibiting extensive expansion and subfamily-specific clusters with structural divergence between haplotypes. Transcript profiling across developmental stages and cultivars distinguished a core set of highly expressed inflorescence-associated CsTPSs from genes exhibiting cultivar-specific regulation. Functional characterization of six previously unreported CsTPSDKs uncovered diverse mono- and sesquiterpene synthase activities, including unexpected substrate promiscuity across subfamilies. These findings deliver the most comprehensive functional annotation of the *C. sativa* TPS repertoire to date, elucidating the genetic and biochemical bases of terpene diversity and providing a foundation for targeted metabolic engineering and cultivar improvement.

1. Introduction

Terpenoids constitute the largest and most structurally diverse class of plant specialized metabolites. These compounds play pivotal ecological roles, from attracting pollinators to deterring herbivores and facilitating adaptation to environmental stresses[1−5]. Their biosynthesis is compartmentalized into two distinct pathways: the plastidial 2-*C*-methyl-D-erythritol-4-phosphate (MEP) pathway and the cytosolic mevalonate (MVA) pathway, which supply precursors for terpene synthases (TPSs) to generate the vast structural diversity of mono- and sesquiterpenes[6−10]. These skeletal structures can subsequently be modified by various enzymes, including cytochrome P450 oxygenases (CYP450), methyltransferases, acyltransferases, and glycosyltransferases (GT), resulting in a more diverse range of compounds[11−13]. Throughout plant evolution, the TPS family has undergone extensive expansion and functional diversification, allowing plants to adapt to diverse ecological niches[14]. In angiosperms, TPS enzymes are categorized into distinct subfamilies. Among them, the TPS-a and TPS-b clades ate responsible for the majority of sesquiterpene and monoterpene production, respectively.

Among terpene-rich plants, *Cannabis sativa* has garnered global attention for its significant medicinal and recreational value[15]. While cannabinoids are the primary pharmacological agents, volatile terpenes constitute the second largest class of metabolites in cannabis in the resin of glandular trichomes[7]. Notably, both cannabinoids and monoterpenes compete for the same isoprenoid precursor pool, reflecting a shared metabolic foundation[16,17]. Beyond shaping the distinctive aroma of different cultivars, cannabis terpenes modulate therapeutic outcomes through the entourage effect, acting either independently or synergistically with cannabinoids[18,19]. Therefore, dissecting the genetic and biochemical basis of terpene biosynthesis is critical for both molecular breeding and metabolic engineering.

Despite the importance of cannabis terpenes, a complete functional mapp of the CsTPS family remains elusive. Previous studies have identified approximately 30−50 *CsTPS* genes per genome and functionally characterized around 29 enzymes across different cannabis cultivars[7,20-24]. However, the highly heterozygous cannabis genome has constrained these efforts. This complexity often causes allelic variations to be collapsed or misannotated in consensus reference genomes, obscuring the true size and structure of the gene family. Additionally, sampling has been historically limited. Most metabolic profiling has focused exclusively on mature floral tissues[21,25]. Consequently, the spatiotemporal dynamics of terpene biosynthesis, particularly in vegetative organs or during early developmental transitions, remain largely unexplored[26].

In this study, we address these challenges by leveraging a high-quality, haplotype-resolved genome of the auto-flowering cultivar Dinamed Kush (DK). This genomic resource allowed us to dissect a fundamental yet overlooked dimension of cannabis specialized metabolism: how the structural and functional asymmetry between haplotypes dictates catalytic divergence and the resulting volatile terpenoid landscape in *C. sativa*. By integrating non-targeted metabolomics with transcriptomics across an unprecedented 28-sample spatiotemporal dataset from six cultivars we delineated the metabolic reprogramming between inflorescence hotspots and the volatile silence of roots and seeds. Beyond genomic annotation, we expanded the functional repertoire of the cannabis TPS family by characterizing six previously unreported enzymes, bridging the gap between haplotype architecture and chemical chemodiversity through multi-cultivar comparative analysis. Together, our findings deliver the most comprehensive functional map of the *CsTPS* family to date, providing a mechanistic foundation for understanding the genetic and biochemical bases of terpenes diveristy in *C. sativa*.

2. Materials and methods

2.1. Plant materials and chemicals

The feminized auto-flowering cannabis hybrid DK (Purple Kush × Dinamed CBD), and five commercial cultivars (Painkiller, PK; Swisse Dream, SD; Terra italia, TI; George gloira, GG; and Red pure, RP) were cultivated under controlled conditions (26 °C, 70% relative humidity) as described previously[27,28]. The characteristics of the six cannabis cultivars, includingthe genetic background, flowering type, chemotype, and aroma profiles were displayed in Supporting Information Table S1. Sampling was conducted in two distinct sets to investigate spatiotemporal dynamics and cultivar differences.

To characterize the developmental profile of DK, ten representative samples were collected throughout its life cycle. Since DK initiates flowering approximately 30 days after germination (DAG), tissues were harvested at weekly intervals during the rapid growth phases. Specifically, leaf samples collected weekly from 7 DAG until the onset of flowering were designated as the pre-flowering stage (PF-L). Similarly, flowers, bracts, and leaves collected weekly during the first three weeks of flowering (<50% pistils browned) were classified as the early stage of anthesis (EsA). In contrast, samples for the late stages of anthesis (LsA)

comprised flowers, bracts, leaves, roots, stems, and mature seeds. Additionally, EsA flowers, bracts, and leaves were collected at two discrete time points to represent advanced maturity. The LsA samples included flowers, bracts, leaves, roots, stems, and mature seeds from DK.

For comparative profiling, flowers, bracts, and leaves were collected from the five commercial cultivars as well as DK specifically at the LsA stage. In total, 28 distinct spatiotemporal samples were obtained. All samples were flash-frozen in liquid nitrogen and stored at −80 °C until volatile terpene profiling and transcriptome data analyses. Detailed sample information is listed in Supporting Information Table S2, and analytical standards are provided in Supporting Information Table S3.

2.2. Volatile terpene profiling in C. sativa by GC−MS/MS

Frozen tissue was ground to a fine powder in liquid nitrogen. Approximately 20 mg of powder was allocated to a 20 mL headspace vial and spiked with 2 μg 2-nonanone and 20 ng *n*-Tridecane as the internal standards. Volatile terpenes were analyzed on an Agilent 8890 Accurate-Mass Quadrupole Time-of-Flight GC/MS system coupled with a flame ionization detector, fitted with an HP-5MS UI column (15 mm × 0.25 mm, 0.25 μm film). Headspace solid-phase microextraction (HS-SPME)[29] used a DVB/Carbon WR/PDMS SPME arrow, underwent an equilibration process at 50 °C for 10 min (300 rpm) and then conditioned at 250 °C for 10 min, followed by adsorption for 40 min and desorption for 10 min. The inlet was 300 °C. The oven program was: 36 °C (1.5 min), 3 °C/min to 90 °C (hold 6 min), 3 °C/min to 210 °C, then 10 °C/min to 270 °C (hold 10 min). Mass spectra scanning was executed over *m/z* 35−500 at a scan rate of 781 u/s. Data were processed in Agilent MassHunter Unknows Analysis. For the identification of terpenes, in addition to comparing mass spectra with the NIST mass spectrometry library, the retention index (RI) was determined. The experimental RI values were calculated using the retention times of the target compounds and the *n*-Tridecane under the same chromatographic conditions. Compounds were considered tentatively identified only when their experimental RI values matched those reported in the NIST Chemistry WebBook and when the forward match factor was ≥60 and characteristic ion fragment match was >80%. All identifications were manually curated, and selected compounds were confirmed with authentic standards to exclude co-elution artifacts.

2.3. Quantification of volatile terpenes

Quantification was performed using two distinct methods depending on the availability of authentic standards. For terpene profiling in different developmental stages and cannabis cultivars, the peak area was calculated and normalized relative to the internal standard response. For the absolute quantification for 22 target terpenes, the accurate quantification was achieved using the external standard method. A series of mixed standard working solutions were prepared at concentrations of 0.1, 0.25, 0.5, 1, 2.5, 5, 10, 25, 50, 100, 250, 500, 1000, 2500, and 5000 ng/mL. Each concentration level was analyzed in triplicates, and the peak areas were recorded. Calibration curves were constructed for each compound by plotting the peak area against standard concentrations.

2.4. Identification of CsTPSDK and terpene-biosynthesis-related genes

Protein and nucleotide queries for upstream genes in the terpene biosynthetic pathway from multiple species were retrieved from NCBI and Uniprot. The sequences included genes related to the MVA and MEP pathways. Specifically, protein sequences of characterized enzymes from *Arabidopsis thaliana* were used as queries for BLASTP searches (*E*-value ≤ 1e-5) against the two haplotype-resolved DK genome assemblies generated in our laboratory (BioProject:PRJCA029174). Candidate hits were further verified by checking for conserved domains using the NCBI Conserved Domain Database. Additionally, TPS sequences from other species and other cannabis cultivars[7,12,20−23] were retrieved from public databases as listed in Supporting Information Table S4. Candidate TPS loci were validated with Hidden Markov Model (HMM) using two Pfam profiles, including Terpene_synth_C (PF03936) and Terpene synthase N-terminal domain (PF01397).

2.5. RNA isolation, cDNA synthesis and gene cloning

RNA isolation and cDNA synthesis followed our previous study[27,28]. The full-length coding sequences (CDS) of 20 previously uncharacterized CsTPSs and DK homologs of two reported CsTPSs genes were targeted. Eight genes, including CsTPS3DK, CsTPS6DK, CsTPS19DK, CsTPS23DK, CsTPS26DK, CsTPS27DK, CsTPS28DK, and CsTPS37DK, were successfully cloned by high-fidelity PCR on cDNA templates using gene-specific primers (Supporting Information Table S5).

2.6. Multiple sequence alignment and phylogenetic analysis

To resolve TPS phylogeny, 102 TPS amino acid sequences from nine cannabis cultivars and 40 additional plant species (Table S4) were aligned with MUSCLE (default settings). A neighbor-joining algorithm (NJ) tree was generated in MEGA X with 1000 bootstrap replicates. In parallel, a maximum-likelihood (ML) phylogenetic tree was constructed from a MAFFT alignment using IQ-TREE[30] with 1000 bootstrap replicates. The scale bar corresponded to 20% amino-acid divergence. Trees were visualized with iTOL (http://itol.embl.de/). Gene duplication modes were inferred with DupGen_Finder v1.12[31] (default parameters).

To perform comparative evolutionary analysis, the genome sequences and annotation files of *Humulus lupulus* and *Morus notabilis* were retrieved from the NCBI database. The TPS genes in these species were identified using a two-step strategy. First, BLASTN searches were performed against the respective genomes using characterized CsTPS sequences as queries (*E*-value < 1e-5). Second, the resulting hits were validated by CD-Search (NCBI Conserved Domain Database) to confirm the presence of characteristic Terpene Synthase N-terminal (PF01397) and C-terminal (PF03936) domains. Only full-length sequences possessing both domains were retained for phylogenetic analysis.

2.7. Gene expression pattern analysis

Transcriptomes from the DK developmental series and diverse tissues/cultivars (this study, lab-generated) were analyzed together with public glandular-trichome datasets from additional cultivars retrieved from NCBI[7,23]. Transcript abundance was quantified as Fragments Per Kilobase of exon model per Million mapped

fragments (FPKM) using the rsem-calculate-expression tool from the RESM software package. To evaluate allele-specific expression (ASE), we employed a combined diploid mapping strategy. Specifically, RNA-seq reads were mapped against a combined reference comprising the sequences from both HiFi and Hi-C assembled haplotypes. Because the two haplotypes were assembled independently, natural sequence variants (SNPs and InDels) were preserved. RSEM leveraged these variants to accurately distinguish allele-specific reads using its Expectation-Maximization algorithm. Furthermore, by competitively mapping reads against both haplotypes simultaneously rather than a single collapsed haploid reference, we inherently avoided reference mapping bias, eliminating the need for further mathematical bias correction. Expression matrices were normalized and visualized in TBtools[32]. Pearson correlation coefficients were computed between CsTPS genes and all other genes; highly correlated pairs ($|r| \geq 0.9$) were visualized as networks in Cytoscape (Supporting Information Table S6).

2.8. Subcellular localization

Subcellular localization was predicted with WoLF-PSORT (https://wolfpsort.hgc.jp/), TargetP-2.0 (https://services.healthtech.dtu.dk/services/TargetP-2.0/), and DeepLoc-2.1 (https://services.healthtech. dtu.dk/services/DeepLoc-2.1/). Full-length CDS of *CsGPPSs* including a large subunit and two small subunit copies, and eight candidate CsTPSs including *CsTPS3*, *CsTPS6*, *CsTPS19*, *CsTPS23*, *CsTPS26*, *CsTPS27*, *CsTPS28*, and *CsTPS37*, were cloned into pCambia1300-eGFP vector to generate C-terminal eGFP fusions. Constructs were introduced into *Nicotiana benthamiana* leaves *via Agrobacterium tumefaciens* GV3101 infiltration. After 72−96 h of incubation, fluorescence signals were imaged on a Zeiss LSM900 confocal microscope. Excitation wavelengths were set at 488 nm for GFP and 633 nm for chlorophyll autofluorescence. All transient assays were performed in at least three independent biological replicates.

2.9. GPPS functional characterization and interaction verification

To verify the enzymatic function of the candidate GPPS subunits, a combinatorial expression assay was performed using a metabolic engineering approach. A specific monoterpene synthase CsTPS6DK was used as a reporter to detect the *in vivo* supply of GPP. The coding sequences of *CsGPPSsu1* and *CsGPPSsu3* were individually cloned into the pRSFDuet-MlspA vector by replacing the original GPPS and co-transformed with the reporter TPS construct into *E. coli* strain EIP81. The transformed cells were cultured and induced based on the same instructions as TPS enzymatic assays.

To investigate the protein-protein interaction assay between CsGPPSsu1 and CsGPPSlu *in vivo*, a firefly luciferase complementation imaging assay (LCA) was performed. The full-length coding sequences of *CsGPPSsu3* and *CsGPPSlu* were amplified and individually cloned into the pCAMBIA-nLUC and pCAMBIA-cLUC vectors, to generate fusion proteins with the N-terminal (nLUC) and C-terminal (cLUC) fragments of firefly luciferase. Combinations of each construct with the corresponding empty vector served as negative controls. After 72 h of post-infiltration incubation, the tobacco leaves were sprayed with 1 mmol/L D-luciferin to detect luciferase activity. The luminescence signals were captured using a Berthold LB985.

2.10. TPSs enzymatic assays

The engineered *E. coli* strain EIP81 was designed to elevate intracellular IPP/DMAPP levels. For monoterpene production, CsTPSDK genes were cloned into the pRSFDuet-MIspA vector (harboring a GPPS coding gene) and transformed into the EIP81 strain[33]. The pMevT-MBIS plasmid was developed as a *de novo* biosynthetic construct to ensure adequate FPP supply[34]. For sesquiterpene production, CsTPSDK genes were cloned into the pET28a vector and co-transformed into *E. coli* strain BL21 (DE3) with pMEVT-MBIS. Protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 1 mmol/L $MgCl_2$ at 16 °C for 76 h. Two milliliters of induced culture were subjected to HS-SPEM-GC−MS analysis. Chromatographic conditions followed those described in Section 2.2, with specific modification for injection: monoterpene assays utilized splitless injection, while sesquiterpene assays employed either splitless injection or a split injection mode (ratio 5:1) to optimize peak separation and quantification of major products. The oven program was: 40 °C (1 min), 8 °C/min to 250 °C (hold 5 min), 30 °C/min to 300 °C.

3. Results

3.1. Tissue-, stage-, and cultivar-resolved volatile terpene profiling in cannabis

We profiled cannabis volatile terpenes (CVTs) in 28 distinct spatiotemporal samples from the six cultivars. These samples comprised a developmental series for the DK cultivar (10 samples) and a comparative collection for all six cultivars at the late stage of anthesis (18 samples). A GC−MS-based non-targeted metabolomics analysis (Fig. 1A and Supporting Information Fig. S1A−S1D) yielded 227 CVTs after filtering, comprising 88 monoterpenes and 139 sesquiterpenes (Supporting Information Tables S7−S9).

For the DK cultivar, samples were collected at weekly intervals during the pre-flowering (PF) and early anthesis (EsA) stages to capture rapid metabolic transitions, while the late stage of anthesis (LsA) was represented by two discrete time points to characterize advanced maturity. In total, we detected 142 CVTs across DK developmental stages. Analysis of terpene distribution revealed distinct metabolic traits across tissues (Fig. S1E and S1F). Flowers and bracts served as metabolic hotspots, accumulating the majority of compounds with high intensity. In contrast, root and mature seeds exhibited volatile metabolic silence, with only one and three volatile terpenes detected, respectively. This spatial restriction confirms that volatile terpene diversity is almost exclusively confined to aerial trichome-rich tissues. Tissues from LsA displayed terpene profiles significantly distinct from those at EsA, indicating developmental reprogramming of terpene metabolism (Fig. 1B).

To investigate inter-cultivar diversity, we compared the LsA floral and leaf samples of DK with five commercial cultivars (PK, SD, TI, GG, RP). These cultivars were selected to represent diverse chemotypes (high-CBD, high-THC, and balanced) and aroma profiles, as detailed in Table S1. While each cultivar produced a similar number of total CVTs, their compositional "fingerprints" varied significantly (Fig. 1C and D). Quantitatively, DK, RP, and TI exhibited balanced mono-/sesquiterpene ratios, whereas PK, GG, and SD showed a higher proportion of sesquiterpenes (Supporting Information Fig. S2A).

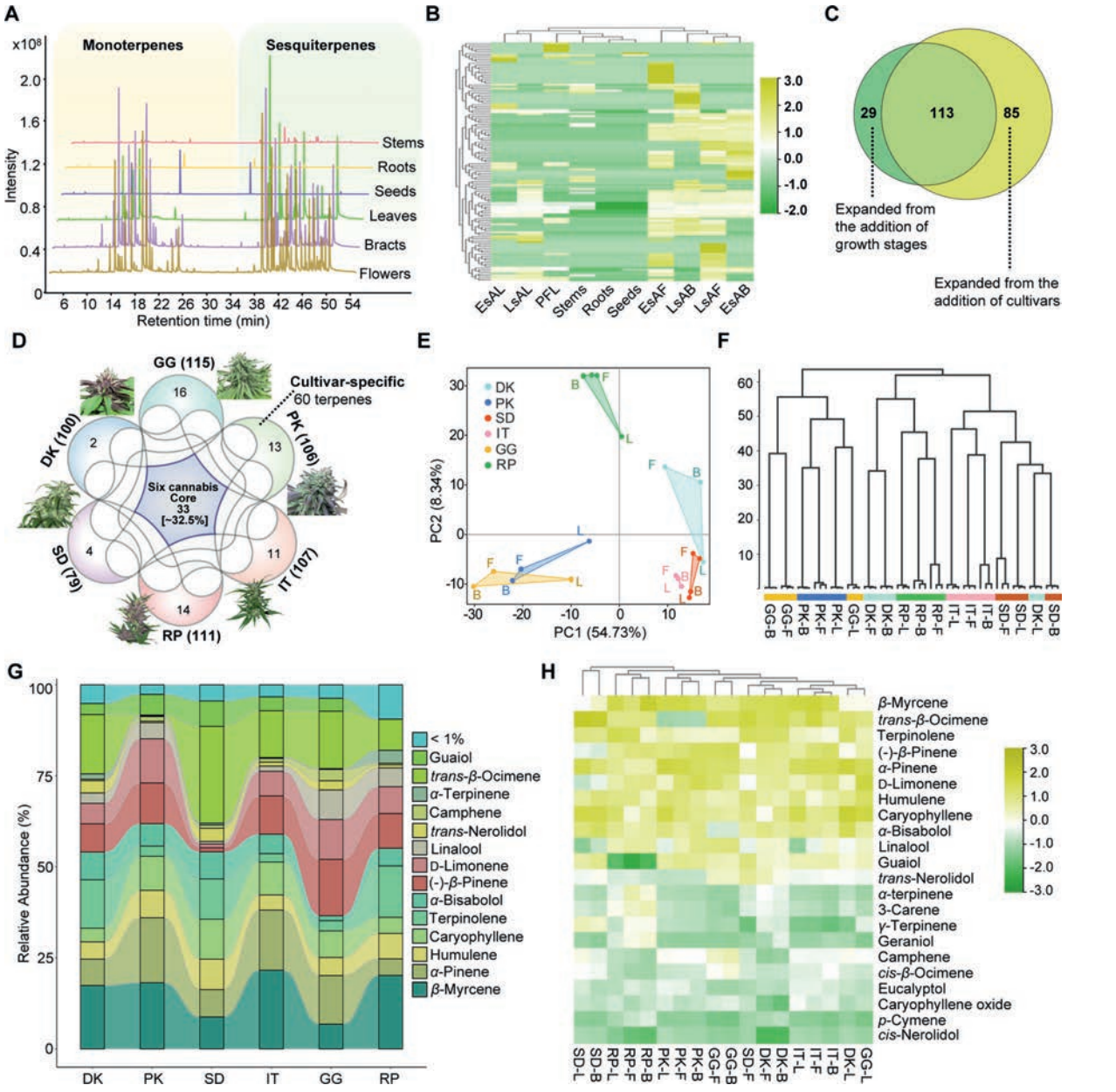

Figure 1 The distribution of CVTs in six different cannabis cultivars and the developmental stages of the DK cultivar. (A) The representative total ion chromatograms of volatile substances analyzed by GC−MS across different tissues of the DK cultivar. (B) The distribution of the detected CVTs in different tissues at different developmental stages of the DK cultivar. The stages included EsA (early stage of anthesis), LsA (late stage of anthesis), as well as tissues such as leaves (L), flowers (F) and bracts (B). (C) Differences in the number of CVTs between when samples from different developmental stages were combined and when differnent cannabis cultivars were combined. A total of 227 compounds were identified with 113 compounds being co-detected, 29 compounds being newly identified at different developmental stages, and 85 compounds identified in other cultivars. (D) Venn diagram showing the overlap of identified CVTs acorss six commercial cannabis cultivars. Among them, 33 core CVTs were co-detected in all six cultivars, while 60 terpenes were found to be specific to individual cultivars. (E) PCA of CVT profiles from flowers, bracts, and leaves of six cannabis cultivars. (F) A dendrogram analysis of CVT profiles from flowers, bracts, and leaves of six cannabis cultivars. (G) The relative abundance of terpenes across different cannabis cultivars. Terpenes with the relative abundance less than 1% were not shown. (H) The accumulation pattern of these 22 terpenes in flowers, bracts, and leaves across the six cannabis cultivars.

Principal component analysis (PCA) and hierarchical dendrogram analysis highlighted cultivar-specific metabolic signatures (Fig. 1E and F, Fig. S2B and S2C). Interestingly, the terpene-based clustering revealed that terpene-based grouping did not strictly align with cannabinoid chemotypes. The balanced cultivar PK clustered closely with the high-THC cultivar GG, suggesting shared terpene biosynthetic modules despite their divergence in cannabinoid pathways. In contrast, RP formed a distinct outlier group. OPLS-DA loading analysis further identified specific biomarkers driving this differentiation. RP was distinguished by high levels of β-myrcene and terpinene, GG by guaiol and camphene, and DK by caryphyllene and humulene. These results underscore

that cultivar-specific terpene landscapes are shaped by complex genetic regulation beyond simple chemotype classifications.

3.2. Signature terpenes and gene-metabolite concordance

To elucidate the molecular basis of the observed metabolic diversity, we quantified 22 representative terpenes and ranked them based on DK content (Fig. 1G and Supporting Information Table S10). Across cultivar, the five most prevalent terpenes were β-myrcene, *trans*-β-ocimene, terpinolene, (−)-β-pinene, and α-pinene. While these components generally indicated the overall terpene profile (Fig. 1H), the dynamic range of individual terpenes varied markedly. A striking example of this variation was *trans*-β-ocimene. While it emerged as a dominant terpene in DK, SD, and GG cultivars, it exhibited significantly low levels in the PK cultivar. To explain this discrepancy, we explored the expression levels of key enzymes responsible for the biosynthesis of *trans*-β-ocimene. The enzyme CsTPS38FN from the Finola cultivar was previously identified as producing *trans*-β-ocimene as the principal product[23]. A genome-wide search confimed that Cs_C01H2G021590 (*CsTPSDK50-2*) is the sole homolog in the DK genome sharing high identity (>98%) with CsTPS38FN, whereas all other TPS genes shared less than 70% identity, ruling out functional redundancy. Transcriptomic analysis revealed that CsTPSDK50-2 expression was approximately 20-fold lower in PK than in other five cultivars, supporting gene-metabolite coupling (Fig. S2D). These data identified signature terpenes that effectively report cultivar specificities and implicate specific TPS homologs in their control.

3.3. MEP/MVA pathway architecture and tissue-partitioned routing underpin terpene precursor supply

Because MEP/MVA pathway outputs supply GPP/FPP precursors for terpene biosynthesis, we inventoried relevant genes and integrated transcriptome-metabolome data. These genes exhibited distinct copy numbers; for example, three 1-deoxy-D-xylulose 5-phosphate synthase (*DXS*) genes, two 1-deoxy-D-xylulose 5-phosphate reductoisomerase (*DXR*) genes, and one 3-hydroxy-3-methylglutaryl-CoA synthase (*HMGS*) gene (Fig. 2A and Supporting Information Table S11). Co-expression mapping ($|r| \geq 0.9$) linked 3496 genes to at least one terpene, including 18 TPSs, seven terpene-upstream genes, 159 transcription factors, and multiple cannabinoid-related pathway genes (CRPGs) (Fig. 2B and Table S6). This phenomenon can be explained by the fact that cannabinoids and monoterpenes were co-accumulated in secretory resin of trichomes and shared GPP from the MEP pathway[23,35]. Gene ontology terms were enriched for "monoterpene biosynthetic/metabolic process" and "secondary metabolic process" (Fig. 2C). Kyoto Encyclopedia of Genes and Genomes analysis pointed to terpenoid synthesis pathways (Supporting Information Fig. S3). An analysis of the origins of these multiple copies using *H. lupulus* as an outgroup indicated that they mainly originated from dispersed duplication (DSD), transposed duplication (TRD), or proximal duplication (PD). We found no evidence of recent whole-genome duplication (WGD), consistent with the lack of a recent WGD in cannabis[36].

Expression profiling across DK tissues and secretory glandular trichomes (SGT) showed distinct patterns between MEP and MVA pathway (Fig. 2A). Single copy genes in MEP pathway displayed high expression levels in flowers, bracts, leaves, and SGT. For multi-copy genes, at least one paralog was highly expressed in these tissues. For instance, *CsDXS3* was the most highly expressed DXS gene, likely dominating DXP production in aerial tissues along with *CsDXS1*, whereas *CsDXS2* was root-biased. By contrast, genes of the cytosolic MVA pathway were broadly expressed across all tissues, with several root-enriched members. This pattern is consistent with their role in supplying precursors for non-volatile sesquiterpene-derived metabolites and triterpenoids required for general growth[20,22,23,37].

Downstream of the MEP pathway genes, the condensation of IPP and DMAPP into GPP represents the critical branch point for partitioning between monoterpene and cannabinoid biosynthesis. To elucide how this flux is regulated, we characterized the GPPS machinery in *C. sativa*. We identifical one CsGPPS large subunit (*CsGPPSlu*) and three small-subunit (*CsGPPSsu*) genes. Among them, while *CsGPPSsu2* showed negligible expression levels in all tested tissues, *CsGPPSsu3* and *CsGPPSls* were highly expressed in aerial tissues except mature seeds. *CsGPPSsu1* exhibited specific high expression in flowers, bracts, and SGTs (Fig. 3A), a high-intensity co-expression pattern with the cannabinoid pathway. This expression pattern spatially coincides with the accumulation of monoterpenes and cannabinoids, implicating their involvement in specialized metabolism.

Subcellular localization analysis further confirmed that CsGPPSsu1, CsGPPSsu3, and CsGPPSlu were all targeted to the plastids/chloroplasts (Fig. 3B). Despite the pivotal role of GPP precursors in cannabinoid and monoterpene production, the specific enzymatic architecture and functional mechanisms of CsGPPSs have not been experimentally characterized. To definitively decipher the biochemical roles of these candidates, we performed a functional combinatorial assay in an *E. coli* system lacking plant-specific GPPS large subunits. As shown in Fig. 3C, the expression of CsGPPSsu3 alone (with a reporter CsTPS6DK) resulted in significant monoterpene production, demonstrating that CsGPPSsu3 functions as an autonomous homodimeric GPPS. In contrast, CsGPPSsu1 alone showed no catalytic activity. Subsequent LCA results confirmed that CsGPPSsu1 physically interacts with the CsGPPSlu (Fig. 3D), suggesting it acts as an obligate regulatory subunit by recruiting the large subunit to form a heterodimeric complex to enhance GPP flux in specialized reproductive tissues.

3.4. Haplotype-resolved TPS gene family identification and their expression patterns

A genome-wide search of the two DK haplotypes genomes identified 56 putative *CsTPS* genes, comprising 44 and 48 members in each haplotype (Table S11). After removing incomplete or truncated entries, 41 high-confidence *CsTPSDK* genes remained. Phylogeny with 102 representative TPSs from nine additional cannabis cultivars and 40 other plant species (Table S9) placed the 41 CsTPSDK proteins into five subfamilies: TPS-a (14), TPS-b (18), TPS-c (1), TPS-e/f (4), and TPS-g (4) (Fig. 4A and Supporting Information Fig. S4). The TPS-a clade clustered with sesquiterpene synthases (*e.g.*, humulene and germacrene synthases). Among the TPS-b members, 13 were further allocated to TPS-b2 and five to TPS-b1, both of which were implicated in the biosynthesis of monoterpenes and hemiterpenes. The TPS-g subfamily clustered with acyclic terpene synthases (*e.g.*, nerolidol and linalool), and TPS-c/TPS-e/f with diterpene synthases, converged with CPPS and *ent*-kaurene synthase, respectively.

CsTPS genes (*CsTPS1DK-CsTPS56DK*) spanned eight pseudochromosomes and one unanchored scaffold (Fig. 4B). Extensive proximal/tandem duplication formed subfamily-specific clusters

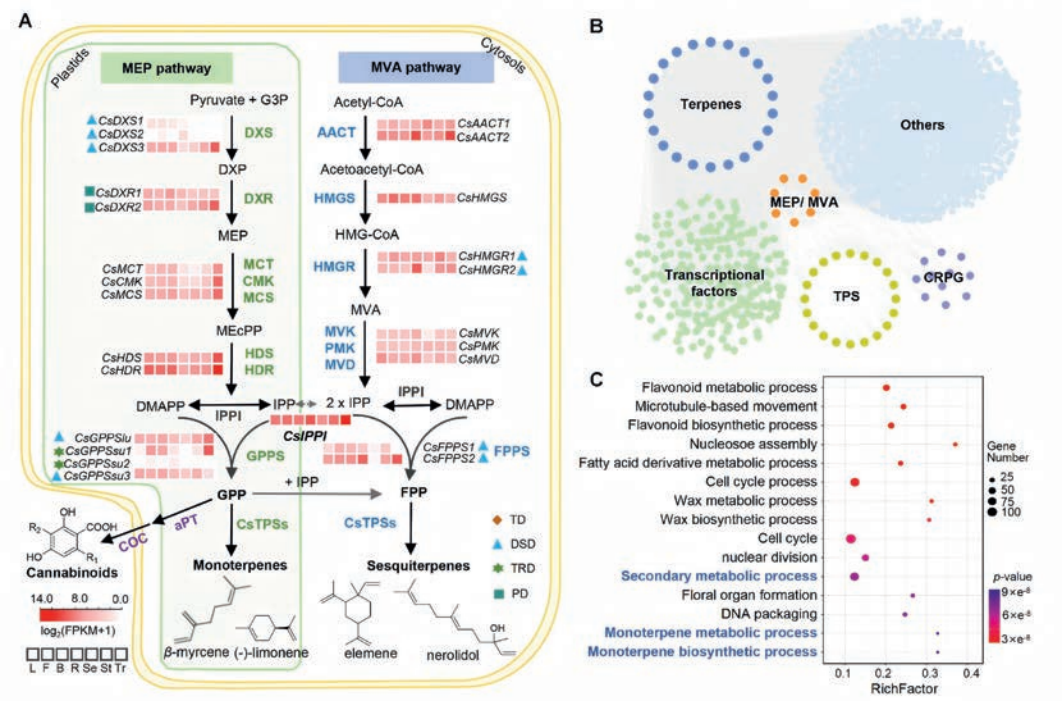

Figure 2 The correlated gene expression patterns of cannabis terpenes. (A) Tissue-specific expression profiles of genes involved in terpene biosynthesis. The biosynthesis of terpenes involves two pathways: the MVA and MEP pathways. aPT, aromatic prenyltransferases; COC, cannabinoid oxidocyclase. The expression profiles for six tissues of cv DK, including leaves (L), flowers (F), bracts (B), roots (R), seeds (Se), and stems (St), as well as previously published transcriptome data for cannabis stalk glandular trichomes. These data were analyzed using the same method. Brown diamond, blue triangle, green star, and cyan square indicated tandem duplication (TD), dispersed duplication (DSD), transposed duplication (TRD), and proximal duplication (PD), respectively. (B) The co-expression gene network of 22 terpenes. A total of 3496 genes were identified that were correlated with at least one terpene, with a high correlation ($r \geq 0.9$) as the threshold. These genes included 18 TPS, seven terpene-upstream genes, 159 transcriptional factors, and 11 cannabinoid-related pathway genes (CRPG). (C) Gene ontology analysis of highly correlated genes from B.

including a 12-gene TPS-b2 cluster within 480-kb on haplotype-2 chromosome 1. This region co-localized with tailoring enzymes such as CYP450 and GTs, suggesting coordinated biosynthetic potential. The homologous region on haplotype-1 exhibited structural variation, including fragment insertions, translocations, and duplications, whereas a cluster on chromosome 2 was conserved across haplotypes (Fig. 4C, Supporting Information Fig. S5). A detailed collinearity analysis revealed significant copy number variations and arrangement differences between the two DK haplotypes. For example, in the TPS-b2 cluster on Chromosome 1, haplotype-2 harbors additional tandem duplications that are entirely absent in haplotype-1. Such structural imbalance between haplotypes suggests that the terpene biosynthetic potential is differentially inherited and deployed within a single heterozygous individual, a layer of regulation masked in previous consensus-based pangenome studies.

To investigate the transcriptional dynamics of these genes, we analyzed *CsTPSDK* expression profiles across the developmental stage and multi-cultivar RNA-seq sets (Fig. 4D and Supporting Information Fig. S6). Of the 41 annotated genes, 30 showed expression levels exceeding 5 Fragments Per Kilobase of transcript per Million mapped reads (FPKM) in at least one tissue. Fifteen exhibited strong, consistent expression in flowers and bracts across cultivars, indicating a conserved core program for inflorescence terpenes (Fig. S6). Except for *CsTPS19DK*, the homologs of the other 14 have been functionally validated in other previously studied cannabis cultivars, supporting their putative roles. At the same time, several *CsTPS* genes displayed pronounced cultivar-specific divergence. For instance, *CsTPS52-2DK* was over 100-fold higher in DK/RP than in PK/TI/GG, whereas *CsTPS53-2DK* exhibited the opposite trend (Fig. S6). These findings highlighted both the conserved transcriptional programs associated with core terpene biosynthesis and the substantial genotype-dependent regulation that can contribute to cultivar-specific terpene profiles in cannabis.

To investigate whether the two haplotypes contribute equally to terpene biosynthesis, we analyzed allele-specific expression patterns for genes in the MEP/MVA pathways across all sampled tissues (Fig. 4E). By summing expression levels (FPKM) from eachhaplotype across the spatiotemporal series, we identified genes with significant allelic bias ($|\log_2\text{fold-change}| > 1$, representing $>$ 2-fold expression difference between haplotypes). The majority of terpene pathway genes showed balanced expression from both alleles, indicating functional conservation for core biosynthetic machinery. However, eight *CsTPS* genes and CsGPPSsu3 exhibited pronounced haplotype-specific expression bias. The haplotype-specific deployment of certain TPSs suggests that cultivar phenotypes result from additive or complementary contributions of allelic variants. These results highlight the importance of haplotype-resolved genomic approaches for dissecting the genetic architecture of complex metabolic traits in heterozygous crop species like cannabis.

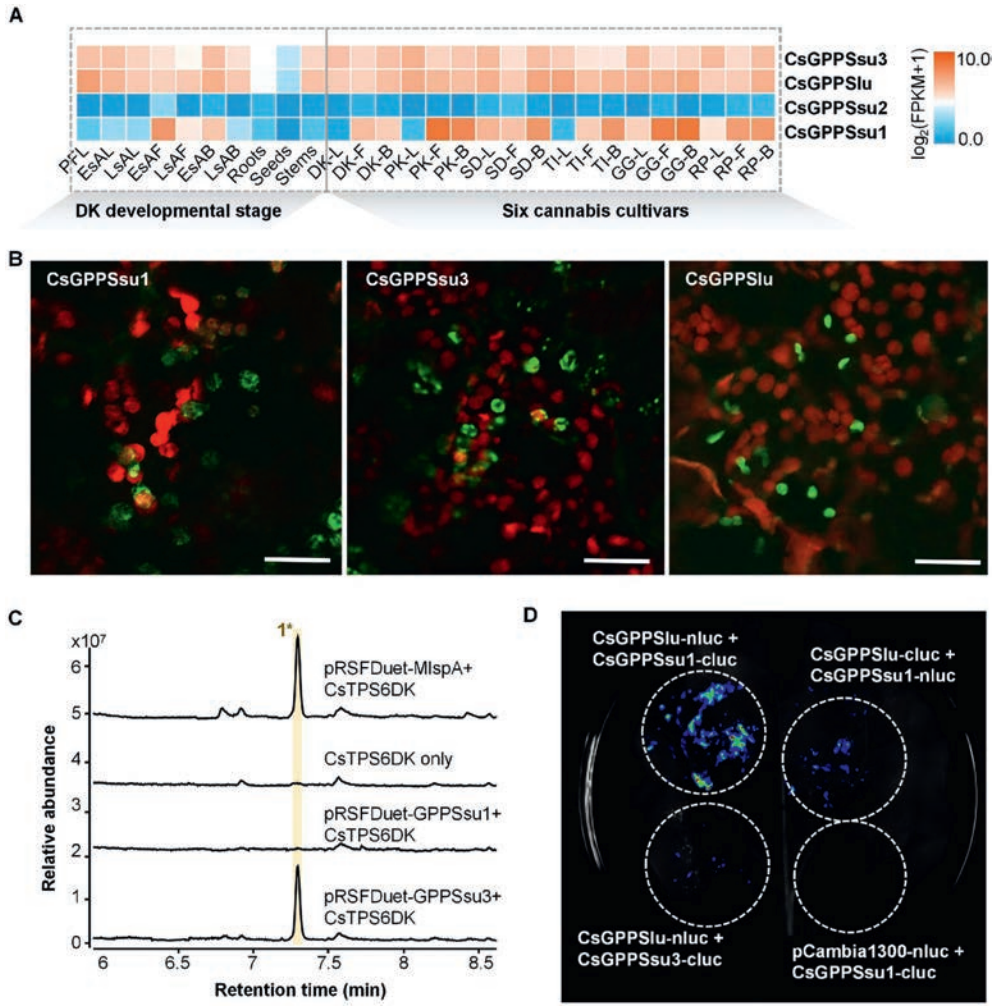

Figure 3 Functional characterization and interaction of GPPS subunits in *C. sativa*. (A) Expression profiles of *CsGPPSsu1-3* and *CsGPPSlu* across diverse tissues, highlighting the flower and bract-specific expression of *CsGPPSsu1*. (B) Subcellular localization of the CsGPPS subunits in *N. benthamiana* leaves, confirming their exclusive targeting to the plastids. GFP fluorescence indicated the location of each fusion protein (green), and the location of chloroplasts was determined by chlorophyll autofluorescence (red). Pictures shown were the merged channel. Scale bar, 20 μm. (C) GC−MS chromatograms of GPP-derived monoterpenes produced by recombinant CsGPPS subunits. CsGPPSsu3 functions as an autonomous homodimeric GPPS, whereas CsGPPSsu1 shows no independent catalytic activity. (D) *In vivo* verification of physical interactions between CsGPPSlu and CsGPPSsu1 using Luciferase Complementation Assays (LCA).

3.5. Functional characterization of CsTPSDK enzymes

To determine whether CsTPSDKs have been functionally characterized in other cannabis cultivars, we performed a BLAST analysis using a conservative threshold of 95% amino acid identity to assign sequences to the same CsTPS identifier. Orthology mapping identified 20 *CsTPSDK* genes without prior functional characterization and noted eight previously characterized *CsTPSs* absent from DK. We attempted to clone full-length CDSs for the 20 uncharacterized *CsTPSDKs* and successfully obtained six, including *CsTPS19DK*, *CsTPS23DK*, *CsTPS26DK*, *CsTPS27DK*, *CsTPS28DK*, and *CsTPS37DK* (Table S5). Of the 14 remaining genes that could not be cloned, ten exhibited low or no expression in DK tissues, largely explaining the cloning failures. Expression patterns of the six cloned *CsTPSDKs* were tissue-biased: *CsTPS27DK* was exclusive to juvenile inflorescences, *CsTPS19DK*/*CsTPS23DK*/*CsTPS26DK* were enriched in late bloom; *CsTPS28DK* was leaf-biased; and *CsTPS37DK* was root-specific (Fig. 4D).

To assess their enzymatic functions, the six *CsTPSDK* genes were heterologously expressed in *Escherichia coli* strains engineered for high-yield production of GPP and FPP. The volatile products were analyzed by GC/MS and identified by comparison with authentic standards and the NIST library (Fig. 4, Supporting Information Fig. S7, and Tables S12 and S13). As validation, we re-characterized two reported TPSs: TPS3PK-PK24618.1 and TPS33PK-KY624371. Their DK homologs were CsTPS6DK (Cs_C01H1G002170; 99.82% aa identity) and CsTPS3DK (Cs_C01H1G001840, 98.87% identity). Consistent with prior work and phylogenetic placement, CsTPSDK6 yielded β-myrcene **(1)** as the major product[23] (Fig. 5A). In contrast, we could not confirm the previously reported limonene synthase activity for CsTPS3DK. To elucidate the basis for this inactivity, we performed a rigorous sequence alignment of the cloned cDNA. This analysis revealed the absence of the conserved NSE/DTE motif essential for Mg^{2+} coordination, indicating that CsTPS3DK is a non-functional gene rather than an active variant at least in one

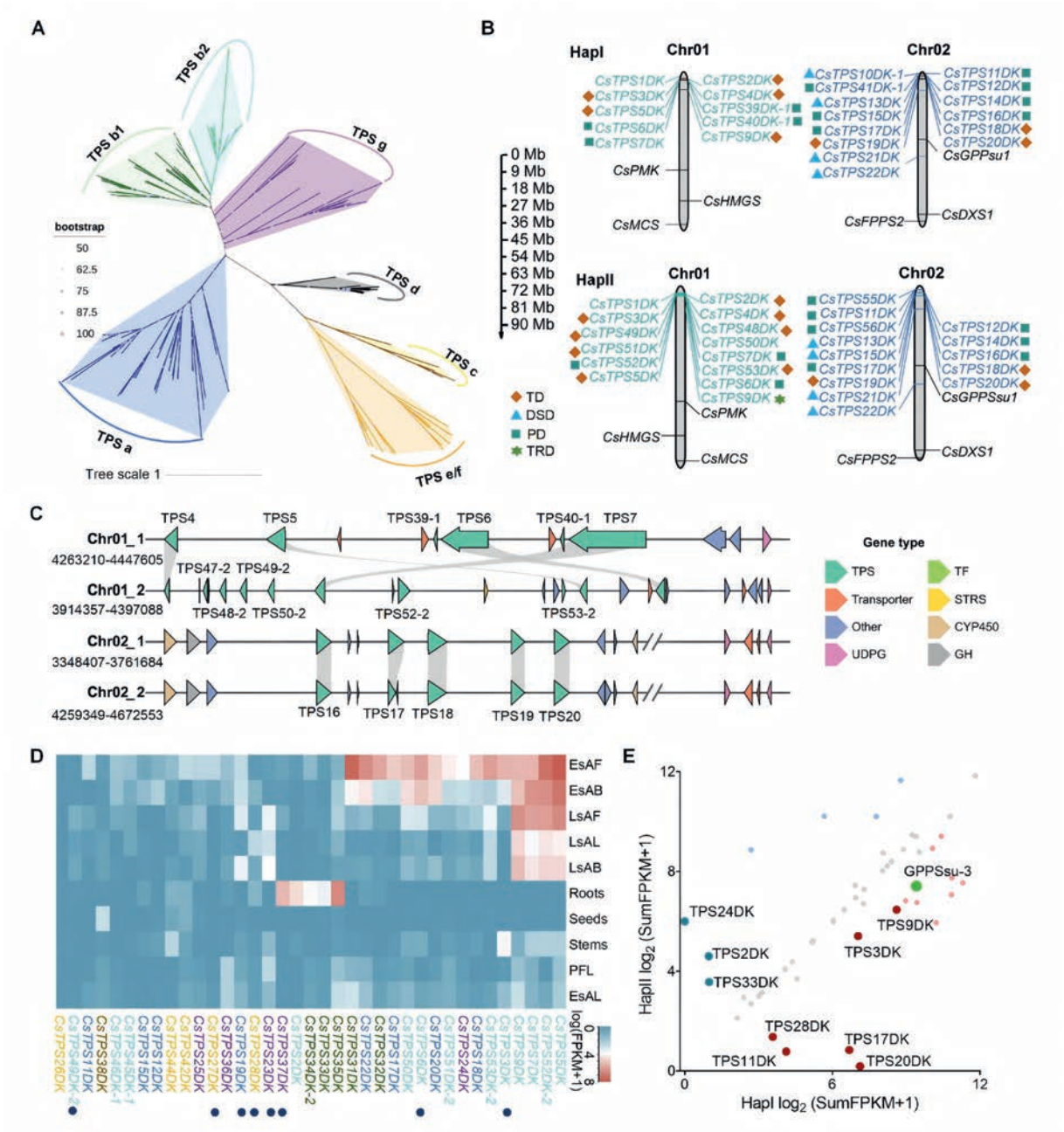

Figure 4 Haplotype-resolved genomic architecture and allele-specific expression of the *Cannabis sativa* TPS gene family. (A) The phylogenetic classification of the CsTPS gene family. A total of 41 *CsTPSDK* genes were identified, categorized into five subfamilies: TPS-a (14), TPS-b (18), TPS-c (1), TPS-e/f (4), and TPS-g (4). The phylogenetic analysis included 102 representative TPSs from nine other cannabis cultivars and 40 different plant species. (B) Chromosomal localization of CsTPSs on the haplotype-resolved DK genome. Genes are colored by subfamily. Symbols denote gene duplication types: tandem (TD, brown diamond), dispersed (DSD, blue triangle), transposed (TRD, green star), and proximal (PD, cyan square). (C) Synteny analysis showing collinearity of TPS clusters between the two haplotypes. Homologous gene pairs are linked, with gene types indicated by colors. (D) Spatiotemporal expression pattern of 36 expressed CsTPSs across different tissues and different developmental stages. CsTPS subfamilies are represented by different colors. Blue dots highlight genes functionally characterized in this study. (E) Allele-specific expression analysis of terpene pathway genes. The scatter plot compares transcript abundance ($\log_2$SumFPKM+1) between HapI (*X*-axis) and HapII (*Y*-axis). Red and blue dots represent genes with significant allelic bias ($|\log_2 FC| > 1$, representing $>$ 2-fold difference), biased towards HapI and HapII, respectively. Grey dots indicate genes with balanced expression ($|\log_2 FC| \leq 1$). Biased TPSs and GPPSsu3 are labeled.

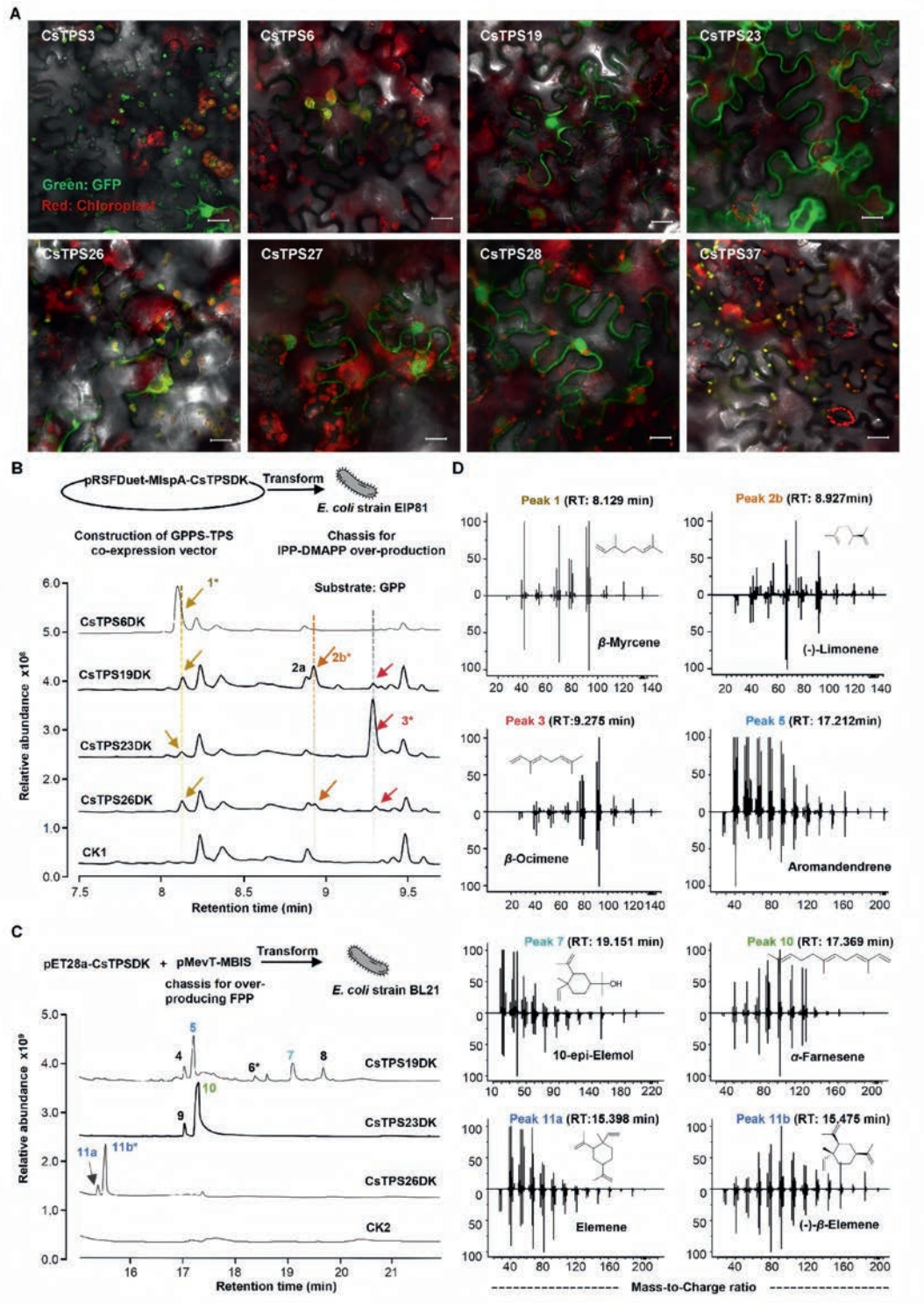

Figure 5 Functional verification of recombinant candidate CsTPSDK. (A) The subcellular localization of eight functional CsTPSDKs in *N. benthamiana* leaves. GFP fluorescence indicated the location of each fusion protein (green), and the location of chloroplasts was determined by chlorophyll autofluorescence (red). (B) *In vitro* enzymatic activities using GPP as the substrate. CK1 represented the transformants containing only the pRSFDuet-MlspA vector, and the products were analyzed using SPEM-GC−MS. Products **1**−**3** were identified as β-myrcene, limonene, and β-ocimene, respectively. Products **2a** and **2b** corresponded to limonene in different chiralities. The arrows indicated the corresponding products. (C) Enzymatic activities of candidate CsTPSDKs with FPP as the substrate. CK2 represented the co-transformants containing the empty vector and pMevT-MBIS. The experiments were conducted at least three times. Products **4**−**11** were predicted to be valencene, aromandendrene, nerolidol, 10-*epi*-elemol, D-nerolidol, (*Z*,*E*)-α-farnesene, α-farnesene, and β-elemene. Products **11a** and **11b** represented β-elemene in different chiralities. (D) The representative head-to-tail GC−MS/MS spectrum of the products and their corresponding references in the NIST database. The stars in B and C represented that the products were confirmed using standards.

haplotype in this cultivar (Supporting Information Fig. S8). This finding provides a definitive molecular example of haplotype-specific allelic non-functionalization. Despite high transcript levels in inflorescences, the output of specific terpenes can be effectively silenced or shifted by the presence of such non-functional alleles in one haplotype, highlighting the necessity of haplotype resolution for accurately predicting metabolic phenotypes based on genomic and transcriptomic data.

With GPP, CsTPS19DK produced (−)-limonene (**2b**) and β-myrcene (**1**) as major products, plus minor D-limonene (**2a**) and β-ocimene (**3**); CsTPS23DK primarily generated β-ocimene (**3**) with trace β-myrcene (**1**); CsTPS26DK yielded a similar but low-abundance monoterpene spectrum to CsTPS19DK; and CsTPS27DK/CsTPS28DK/CsTPS37DK produced no detectable monoterpenes under these conditions.

With FPP, CsTPS19DK, CsTPS23DK, CsTPS26DK, and CsTPS37DK produced significant but indistinguishable sesquiterpenes under splitless GC injection conditions (Fig. S7). Using a 5:1 split injection, CsTPS19DK yielded aromandendrene (**5**) as the major product, along with several minor sesquiterpenes (compounds **4**, **6**–**8, 11a**, and **16**–**18**) (Fig. 5B). Although phylogenetically within a mono-TPS clade, CsTPS23DK unexpected generated α-farnesene and maaliene in multiple chiral forms (**9** and **10**, **19** and **20**). CsTPS26DK primarily produced (−)-β-elemene (**11b)** with trace β-elemene (**11a**). TPS-g CsTPS37DK catalyzed the production of nerolidol (**12** and **15**) and farnesol (**13**, **14**) (Fig. S7). The TPS-e/f clade in general contained two deep branches that diverged before the split of the gymnosperm/angiosperm lineages (Figs. 4A and 6), one encoding a plastidic KS and one encoding a cytosolic geranyllinalool synthase (GLS). Unlike tomato and Arabidopsis, we identified a single KS gene (CsTPS44DK) and three GLS genes (CsTPS27-29DK); we have not yet tested their diterpene synthase activities. Notably, CsTPS27DK and CsTPS28DK both produced trace amounts of nerolidol when supplied with FPP, suggesting latent cross-substrate within these lineages. Across six cultivars, the in-planta occurrence of major enzymatic products mirrored *in vitro* activities for several CsTPSDK enzymes (Supporting Information Fig. S9), indicating that differential deployment of shared TPSs explains much of the chemotypic variation.

Following the *in vitro* functional characterization, we further investigated the subcellular localization of these TPSs to corroborate their biological roles *in vivo*. Transient expression of GFP-fusion constructs in tobacco leaves revealed distinct localization patterns. Confocal microscopy analysis confirmed that CsTPS3, CsTPS6, CsTPS26, and CsTPS37 were explicitly targeted to the chloroplasts. In contrast, CsTPS19, CsTPS23, CsTPS27, and CsTPS28 were observed in both the nucleus and cytosol. Collectively, these findings demonstrate a spatial separation of terpene biosynthesis in *C. sativa*, ensuring that these enzymes are compartmentalized with their respective substrates for efficient metabolic flux.

3.6. Lineage-specific diversification and mechanistic routes to sesquiterpene diversity

To elucidate the evolutionary relationships of CsTPSs, we expanded our phylogenetic analysis to inclue closely related species from the Rosales order: *H. lupulus* (hop, Cannabaceae) and *M. notabilis* (mulberry, Moraceae), alongside distinct outgroups (Arabidopsis and tomato) (Supporting Information Fig. S10A and S10B). This inter-species comparison revealed a dramatic lineage-specific expansion within the Cannabaceae family. While the mulbery genome contains a contracted TPS family (only 12 genes), both cannabis (∼41 genes) and hop (49 genes) exhibit a massive expansion, particularly within the TPS-b and TPS-g subfamilies (Fig. S10C). The TPS-b2 clade was conserved across all three Rosales species, while the sheer number of TPS-b members in cannabis and hop indicates that extensive gene duplication events drove the diversification of monoterpene synthases after the divergence of Cannabaceae from Moraceae.

To further dissect the functional implications of this expansion, we constructed a detailed phylogenetic and functional map incorporating all previously characterized cannabis TPSs alongside 32 *Arabidopsis* and 29 tomato TPSs (Fig. 6 and Supporting Information Table S14)[12]. Across the tree, the three species shared a single common TPS-a ancestor, with each lineage expanding TPS-a independently. In contrast, cannabis exhibited a marked expansion of TPS-b: almost all cannabis TPS-b genes formed a distinct clade relative to tomato and *Arabidopsis*, reflecting an elevated monoterpene production capacity. This clade encompassed a diverged cytosol-localized branch with both monoterpene and sesquiterpene synthase activities, indicating functional diversification. Cannabis also harbored more TPS-b genes (18) than tomato (9) or *Arabidopsis* (6), all encoding monoterpene synthases, and all but one (CsTPS23DK) were phylogenetically distinct from non-cannabis orthologs.

By integrating this phylogenetic framework with a multi-dimensional functional summary covering gene expression, subcellular localization, and enzymatic properties, as well as haplotype-level structural variation described above, these data suggested that cannabis TPS evolution has created a broad yet specialized enzymatic platform. The expansion of the TPS-b subfamily in Cannabaceae, coupled with diverse spatiotemporal regulation and catalytic promiscuity, underpins the rich terpene chemodiversity observed in *C. sativa*.

3.7. Biosynthetic framework and metabolite detection patterns reveal complex regulation of sesquiterpene diversity

By integrating characterized CsTPS product spectra with our comprehensive metabolite profiling dataset (Fig. 7), we sought to validate the biological relevance of *in vitro* functional assignments and understand how TPS diversity translated to metabolic phenotypes. We presented a biosynthetic framework organizing the structural diversity of cannabis sesquiterpenes according to proposed carbocation rearrangement pathways. This classification, derived from TPS product profiles and established terpene biosynthetic principles[22,38,39], reveals that the CsTPSDK enzymes access at least 12 distinct sesquiterpene skeletal classes through alternative initial cyclizations modes (1,6-/1,7-/1,10-/1,11-ring closures) and subsequent rearrangements (Fig. 7A). Despite phylogenetic clustering within TPS-a, CsTPS19DK and CsTPS26DK utilize distinct initial cyclization mechanisms from the same FPP substrate, demonstrating subfamily-level catalytic promiscuity. This rationalizes how a moderately-sized TPS family generates extensive sesquiterpene diversity, with individual enzymes capable of producing multiple skeletal types through alternative carbocation rearrangements.

To evaluate whether the enzymatic products occur in planta, we cross-referenced all enzymatic products of DK TPS enzymes against our spatiotemporal metabolite dataset spanning 28 samples (Fig. 7B, Fig. S9). The majority of sesquiterpene products were detected in at least one cultivar or developmental stage,

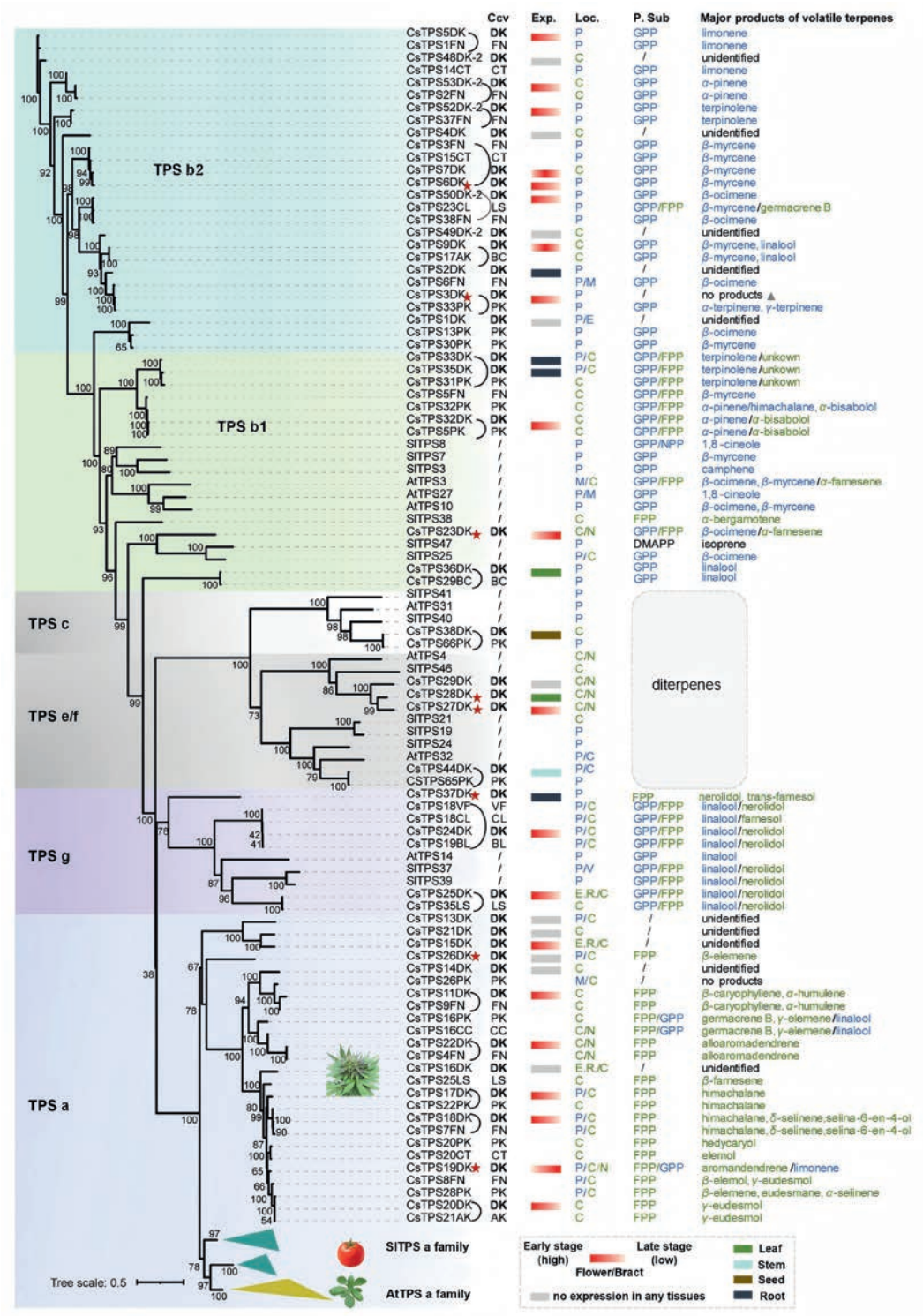

Figure 6 Phylogenetic analysis and functional summary of the cannabis TPS gene family. The left panel presents a maximum-likelihood phylogenetic tree of the TPS proteins. The tree includes TPSs identified in the DK cannabis genome and other cannabis cultivars, as well as TPSs from tomato and Arabidopsis. TPS clades are colored as follows: TPS-a (blue), TPS-b1 (cyan), TPS-b2 (green), TPS-c (grey), TPS-e/f (grey), and TPS-g (purple). The five panels on the right summarize, for each gene, the cannabis cultivar examined (Ccv), expression profile (Exp.), subcellular localization (Loc.), preferred substrate (P. Sub), and major products. Preferred substrates are assigned based on current experimental data and established biological roles in cannabis. Notably, functional characterization in this study prioritized the most biologically relevant precursors (GPP and FPP) responsible for the volatile terpenoid landscape; while potential activity towards other precursors cannot be entirely ruled out, they were not the primary focus of this volatile-centric atlas. Red stars mark genes functionally validated in this study, while genes enclosed by the black boundary were considered the same gene based on ≥95% amino-acid sequence identity. Cultivar abbreviations (*e.g.*,

confirming the in planta relevance of these enzymes. These specific distribution pattern demonstrated that the volatile profiles are shaped by both the available enzymatic repertoire and its differential deployment across genotypes and developmental contexts. However, metabolite distribution patterns revealed complexity beyond simple expression-metabolite correlations. The product distributions of characterized CsTPSDKs did not simply align with their transcript profiles across cultivars and developmental stages. This discrepancy suggests that terpene accumulation is governed by complex regulatory mechanisms beyond transcript abundance. These potentially include precursor partitioning, the product promiscuity and functional redundancy inherent to the TPS family, and rapid downstream modification of the primary products.

Collectively, these analyses demonstrate that while TPS catalytic diversity provides the biosynthetic foundation for sesquiterpene chemodiversity, the translation from enzymatic potential to metabolic phenotype involves multi-layered regulation. This complexity underscores the need for integrated approaches combining genomics, transcriptomics, and metabolomics to understand and manipulate terpene biosynthesis in cannabis.

4. Discussion

This study integrated a haplotype-resolved genomic survey with metabolomics, transcriptomics, and enzyme assays to expand and consolidate the functional coverage of the cannabis TPS family, linking TPS activity with cultivar- and stage-specific CVT profiles. By combining structural genomics with expression and functional data, we delineated how genomic architecture, spatiotemporal expression, and catalytic versatility together shaped chemodiversity of terpenes in *C. sativa*.

4.1. Functional diversification and genomic organization of CsTPS genes

Our haplotype-resolved counts and subfamily composition were consistent with earlier reports of a mid-sized but chemically versatile cannabis TPS family, although the number of CsTPS loci identified in DK ($n = 56$) slightly exceeded previous estimates ($n = 30-50$)[20,21]. Extensive lineage-specific proximal and tandem duplications within TPS-a and TPS-b clusters suggested genomic hotspots for terpene biosynthetic diversification, paralleling independent TPS expansions in other terpene-rich taxa[40] such as *Oryza sativa*[41,42] and *Lanxangia tsaoko*[43]. Comparative analysis with *H. lupulus* and *M. notabilis* further contextualized this expansion within the Rosales order. The parallel expansion of the TPS-b subdamily in cannabis and hop highlights a shared evolutionary strategy closely linked to the development of glandular trichomes and high-resin production. While the TPS-b2 clade is structurally conserved across these lineages, its massive copy number expansion in Cannabaceae suggests that lineage-specific duplications drove the chemical divergence from non-resin-producing relatives.

Haplotype comparisons further revealed structural divergence within syntenic TPS clusters, which may underpin allelic variation in terpene profiles among cultivars. This structural and functional asymmetry between genomic haplotypes serves as a fundamental but overlooked dimension of cannabis specialized metabolism. Beyond mere gene presence, our data demonstrate that haplotype-specific architecture characterized by significant copy number variations (CNVs) and allelic divergence, provides the molecular substrate for chemical diversification. A definitive example is the allelic non-functionalization of CsTPS3DK identified in our study. Despite its close homology to functional synthases, the haplotype-specific deletion of the conserved NSE/DTE motif (Fig. S7 and S8) results in catalytic silencing. Such fine-scale genomic heterogeneity creates a metabolic reservoir, where functional and non-functional alleles are differentially inherited to establish the biosynthetic potential of specific cultivars[20,44]. In the absence of targeted knock-down models to definitively uncouple *in vivo* metabolic fluxes, we propose that haplotype asymmetry functions as a foundational biosynthetic blueprint. It defines the presence or absence of functional alleles, forming a structural constraint that governs the observed metabolic landscape. In addition, TPS genes frequently occurred in cluster with tailoring enzyme genes (*e.g.*, CYPs and GTs), a conserved genomic feature of plant specialized metabolism observed beyond cannabis[45,46].

4.2. Spatiotemporal regulation of terpene biosynthesis and genetic basis of cultivar-specific terpene profiles

Metabolomic profiling across developmental stages and tissues demonstrated strong enrichment of CVTs in flowers, bracts, and leaves, aligning with the spatial distribution of glandular trichomes and their ecological roles in pollinator attraction and defense. Early- and late-flowering stages exhibited distinct terpene profiles, reflecting developmental phase-specific shifts in metabolic allocation. Transcriptomic data corroborated these trends, revealing co-expression of MEP precursor genes, most TPSs, and cannabinoid-pathway genes in glandular trichome-rich tissues, supporting a shared metabolic foundation favoring monoterpene production. This highly coordinated regulation is anchored by both the co-expression networks and the subcellular compartmentalization. Our integrated analysis of GPPSsu exemplifies this synergy. While the autonomous CsGPPSsu3 drives baseline GPP supply, the regulatory CsGPPSsu1 acts as a central node in the trichome-specific network. By confirming the exclusively plastidial targeting of these subunits (Fig. 2D) and their physical assembly into Lu−Su complexes *via* LCA interaction assays (Fig. 2E), we elucidate how *C. sativa* achieves high-flux precursor partitioning. The physical overlap of these precursor factories in the same organelle as the downstream TPS-b and cannabinoid enzymes ensures efficient metabolic channeling in secretory cells. This pattern aligned with previous studies identifying glandular

DK, PK) match those in the main text and Supporting Information In the gene expression column, colors denote the primary tissue of expression (red, flowers/bracts with gradients indicating developmental stages; green, leaves; cyan, stems; brown, seeds; dark blue, roots). Subcellular localization was predicted using TargetP, DeepLoc, and WoLF PSORT: plastid (P; primarily chloroplast), mitochondria (M), cytoplasm (C), nucleus (N), endoplasmic reticulum (E.R.), and vacuole (V). Major products highlighted in green and blue correspond to FPP- and GPP-derived products, respectively. Grey triangles indicate no major product; minor products are listed in Supporting Information Table S13. The unidentified indicates that almost all of these genes were not expressed in any tissues, thereby causing the failure of gene cloning.

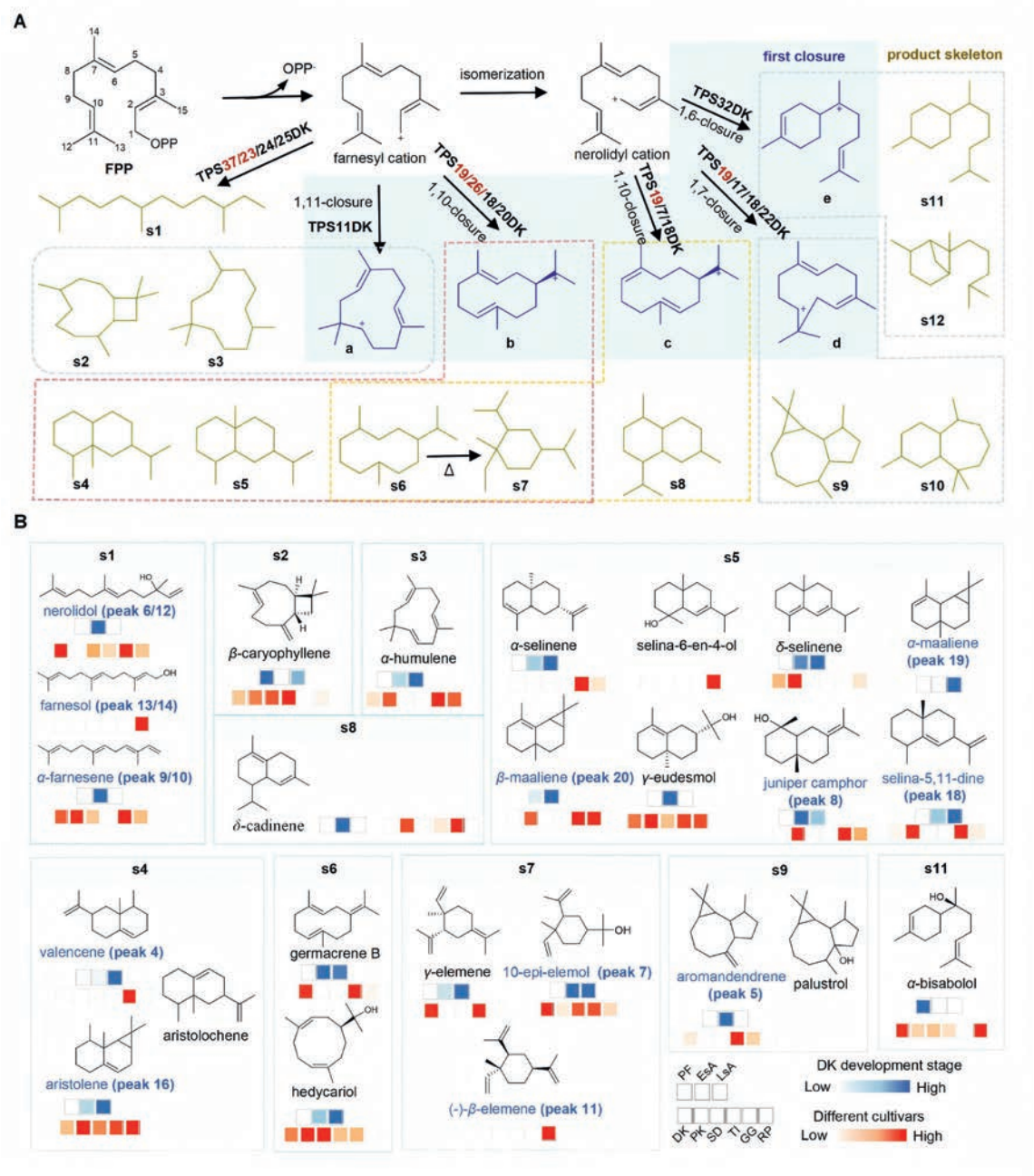

Figure 7 Biosynthetic framework for cannabis sesquiterpene diversity and spatiotemporal distribution patterns of TPS-derived metabolites. (A) Proposed biosynthetic framework for sesquiterpene skeletal diversity. Hypothetical carbocation rearrangement pathways rationalize the structural diversity of sesquiterpenes in Cannabis, inferred from characterized TPS product profiles (this study and literature) and established terpene biosynthetic mechanisms (Degenhardt et al., 2009). Light blue boxes (a−e) represent proposed intermediate carbocations following initial FPP cyclization: (a) humulyl cation, (b−c) germacradienyl cation, (d) cycloheptenyl cation, (e) bisabolyl cation. Colored boxes (s1−s12) denote distinct sesquiterpene skeletal classes (*e.g.*, s1, farnesane; s2, caryophyllene; s3, humulane; s4, valencane; s5, eudesmane; s6, germacrane; s7, elemane; s8, cadinane; s9, aromadendrane; s10, himachalane; s11, bisabolane; s12, bergamotane). TPSs functionally characterized in this study are highlighted in red. (B) Spatiotemporal distribution patterns of TPS-derived metabolites across cultivars and developmental stages. Representative sesquiterpene products from each skeletal class (Panel A) are mapped against their in planta detection patterns across developmental stages and cultivars. Products verified through recombinant enzyme assays in this study are highlighted in blue; peak numbers in parentheses correspond to structures in Fig. 5B and Fig. S7. The dual-tier mini-heatmap illustrate the metabolic landscape: the upper row represents the developmental progression in the DK cultivar, while the lower row shows the relative abundance across six commercial cultivars. Single-row heatmaps indicate detection restricted to specific groups, while absent heatmaps denote compounds not identified within our 227-terpene dataset.

trichomes as specialized "biofactories" whose metabolic output tracked organ identity and developmental stage[5,23,47,48]. By mapping CsTPS expression onto these anatomical and developmental gradients, our study provided a molecular framework for both conserved "core" terpene bouquets and cultivar-specific chemical signatures.

Comparisons among cultivars revealed a shared core terpene set alongside distinct cultivar-specific profiles. PCA and hierarchical clustering grouped samples by both tissue type and genetic background, while differential CsTPS expression, such as the high abundance of *CsTPS52-2DK* transcript levels in DK and RP versus CsTPS53-2DK enrichment in PK, TI, and GG, provided a mechanistic basis for these chemotypic differences. Similar patterns have been reported in other species, where allelic variation and promoter divergence in TPS genes drive chemotypic diversity[49−52]. The integration of metabolite and transcript networks here underscored transcriptional regulation as a major determinant of cultivar-specific terpene composition, with potential for targeted selection in breeding programs.

4.3. Consolidated functional landscape of CsTPS and catalytic diversification

By combining functional assays for six previously uncharacterized CsTPSDKs with curated activities from prior cannabis TPS studies, we expanded the functional coverage of the current cannabis TPS family. Newly assayed enzymes broadened the known mono- and sesquiterpene product spectrum and revealed substrate use across subfamily boundaries. Several CsTPSs accepted both GPP and FPP, and even TPSs from the diterpene pathway produced sesquiterpenes, consistent with the widespread multi-substrate capacity of plant TPSs[22,53,54]. A parallel case in *Solanum habrochaites*, where a TPS-e/f β-phellandrene synthase produces two sesquiterpenes from FPP[55], highlights the evolutionary conservation of TPS catalytic flexibility. These findings link the evolutionary expansion described earlier with biochemical versatility, reinforcing TPS as a driver of terpene diversity.

Functionally annotated, cluster-aware TPS catalogs enabled marker-assisted selection for aroma traits and hypothesized entourage effects. They also informed metabolic engineering strategies targeting precursor partitioning (*e.g.*, GPPS/prenyltransferase selection) and TPS expression to tune volatile output. Substrate-flexible enzymes widened the design possibilities: manipulating subcellular precursor pools or TPS targeting can shift product spectra without altering coding sequences. Given the dual economic and pharmacological value of terpenes as both aroma determinants and modulators of cannabinoid effects, these insights offer tangible tools for cultivar improvement.

While we established the functional landscape of cannabis TPSs using an *in vivo* microbial platform, testing these enzymes against a broader panel of precursors and *in vitro* kinetic parameters were not determined. Our biochemical attempts revealed that while these enzymes function efficiently within the engineered host, they exhibit significant instability upon isolation. Consequently, the relative production levels reported here reflect the effective biocatalytic capability in a cellular context, offering practical insights for metabolic engineering. Future studies using optimized expression system or native enzyme purification from cannabis tissues may overcome these limitations. Additionally, functional assays covered only part of the uncharacterized TPS repertoire; extending this to low-expression or tissue-restricted expression in planta[56], as well as demonstrating how specific haplotype combinations produce distinct metabolite profiles, will help complete the biosynthetic map. The interplay between terpene and cannabinoid pathways, particularly regarding precursor competition and coordinated regulation under environmental stress, remains underexplored. Emerging techniques such as single-cell transcriptomics[57] and spatial metabolomics of glandular trichomes promise finer resolution of terpene biosynthesis at the cellular level, enabling more precise pathway modeling and rational aroma tailoring.

In conclusion, by bridging the gap between haplotype-resolved genomics, exhaustive spatiotemporal profiling, and mechanistic biochemical assays, this study transcends the simple cataloging of genes. We deliver a comprehensive functional map that reveals how the interplay between genomic structural asymmetry and catalytic divergence shapes the volatile terpenoid landscape. These insights not only resolve the long-standing ambiguity in heterozygous cannabis genomes but also provide a high-resolution framework for the rational metabolic design of aroma and pharmacological traits.

Accession numbers

Assembled and annotated haplotype-resolved DK genomes and RNA-seq reads of cannabis, including developmental stages and cultivars have been deposited in National Genomics Data Center (NGDC, China) under the BioProject accession: PRJCA029174. The nucleotide sequences for the functionally characterized *CsTPSDK* genes in this study have been deposited in NCBI under accession numbers PX306361−PX306368.

Data availability statement

All data supporting the findings of this study are available within the paper and its the Supporting Information Due to an ongoing manuscript regarding the genome assembly, the assembled and annotated DK genome are currently under embargo.

Acknowledgments

This work was supported by the China Postdoctoral Science Foundation (2024M760392), the National Natural Science Foundation of China (Grant No.32300319), the National Key R&D Program of China (2023YFC3504800), and the Natural Science Foundation of Guangdong Province (2023A1515010516, China). We are grateful for the instrument support provided by the large-scale instrument sharing platform of Institute of Chinese Material Medica, China Academy of Chinese Medical Sciences.

Author contributions

Yaolei Mi: writing-original draft and conceptualization; Mofan Zhang: data curation and validation; Kaiquan Zhang: resources and investigation; Jun Li and Song Yan: methodology and data curation; Xuewen Zhu, Liwei Wang, and Yupeng Du: Data curation and visualization; Sifan Wang and Xue Cao: methodology and funding acquisition; Shilin Chen, Weisun Sun, and Zhichao Xu: supervision, writing-review & editing, and funding.

Conflicts of interest

The authors declare no conflicts of interest.

Appendix A. Supporting information

Supporting Information to this article can be found online at https://doi.org/10.1016/j.apsb.2026.03.035.

References

1. Pichersky E, Raguso RA. Why do plants produce so many terpenoid compounds? *New Phytol* 2018;**220**:692−702.
2. Gershenzon J, Dudareva N. The function of terpene natural products in the natural world. *Nat Chem Biol* 2007;**3**:408−14.
3. Boncan DAT, Tsang SSK, Li C, Lee IHT, Lam H-M, Chan TF, et al. Terpenes and terpenoids in plants: interactions with environment and insects. *Int J Mol Sci* 2020;**21**:7382.
4. Bouwmeester H, Schuurink RC, Bleeker PM, Schiestl F. The role of volatiles in plant communication. *Plant J* 2019;**100**:892−907.
5. Huchelmann A, Boutry M, Hachez C. Plant glandular trichomes: natural cell factories of high biotechnological interest. *Plant Physiol* 2017;**175**:6−22.
6. Vranová E, Coman D, Gruissem W. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu Rev Plant Biol* 2013;**64**: 665−700.
7. Zager JJ, Lange I, Srividya N, Smith A, Lange BM. Gene networks underlying cannabinoid and terpenoid accumulation in Cannabis. *Plant Physiol* 2019;**180**:1877−97.
8. Liao B, Shen X, Xiang L, Guo S, Chen S, Meng Y, et al. Allele-aware chromosome-level genome assembly of Artemisia annua reveals the correlation between ADS expansion and artemisinin yield. *Mol Plant* 2022;**15**:1310−28.
9. Wang S, Liang M, Chen W, Wan H, Meng X, Zhu X, et al. Completing the biosynthesis of the clinically important diterpenoid andrographolide in *Andrographis Paniculata*. *Angew Chem Int Ed* 2025;**64**:e202425303.
10. Xiong X, Gou J, Liao Q, Li Y, Zhou Q, Bi G, et al. The Taxus genome provides insights into paclitaxel biosynthesis. *Nat Plants* 2021;**7**: 1026−36.
11. Pichersky E, Noel JP, Dudareva N. Biosynthesis of plant volatiles: nature's diversity and ingenuity. *Science* 2006;**311**:808−11.
12. Zhou F, Pichersky E. The complete functional characterisation of the terpene synthase family in tomato. *New Phytol* 2020;**226**:1341−60.
13. Boutanaev AM, Moses T, Zi J, Nelson DR, Mugford ST, Peters RJ, et al. Investigation of terpene diversification across multiple sequenced plant genomes. *Proc Natl Acad Sci USA* 2015;**112**:E81−8.
14. Chen F, Tholl D, Bohlmann J, Pichersky E. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J* 2011;**66**:212−29.
15. Booth JK, Bohlmann J. Terpenes in *Cannabis sativa*—from plant genome to humans. *Plant Sci* 2019;**284**:67−72.
16. Xie Z, Mi Y, Kong L, Gao M, Chen S, Chen W, et al. *Cannabis sativa* : origin and history, glandular trichome development, and cannabinoid biosynthesis. *Hortic Res* 2023;**10**. uhad150.
17. Kovalchuk I, Pellino M, Rigault P, Van Velzen R, Ebersbach J, Ashnest JR, et al. The genomics of *Cannabis* and its close relatives. *Annu Rev Plant Biol* 2020;**71**:713−39.
18. Bautista JL, Yu S, Tian L. Flavonoids in *Cannabis sativa*: biosynthesis, bioactivities, and biotechnology. *ACS Omega* 2021;**6**:5119−23.
19. Russo EB. Taming THC: potential cannabis synergy and phytocannabinoid-terpenoid entourage effects. *Br J Pharmacol* 2011; **163**:1344−64.
20. Booth JK, Page JE, Bohlmann J. Terpene synthases from *Cannabis sativa*. *PLoS One* 2017;**12**:e0173911.
21. Allen KD, McKernan K, Pauli C, Roe J, Torres A, Gaudino R. Genomic characterization of the complete terpene synthase gene family from *Cannabis sativa*. *PLoS One* 2019;**14**:e0222363.
22. Booth JK, Yuen MMS, Jancsik S, Madilao LL, Page JE, Bohlmann J. Terpene synthases and terpene variation in *Cannabis sativa*. *Plant Physiol* 2020;**184**:130−47.
23. Livingston SJ, Quilichini TD, Booth JK, Wong DCJ, Rensing KH, Laflamme-Yonkman J, et al. *Cannabis glandular* trichomes alter morphology and metabolite content during flower maturation. *Plant J* 2020;**101**:37−56.
24. Wiles D, Roest J, Shanbhag B, Vivian J, Beddoe T. Integrated platform for structural and functional analysis of terpene synthases of *Cannabis sativa*. *PeerJ* 2025;**13**:e19723.
25. Fulvio F, Pieracci Y, Ascrizzi R, Bassolino L, Flamini G, Paris R. Insights into terpenes profiling and transcriptional analyses during flowering of different *Cannabis sativa* L. chemotypes. *Phytochemistry* 2024:114294.
26. Aizpurua-Olaizola O, Soydaner U, Öztürk E, Schibano D, Simsir Y, Navarro P, et al. Evolution of the cannabinoid and terpene content during the growth of *Cannabis sativa* plants from different chemotypes. *J Nat Prod* 2016;**79**:324−31.
27. Mi Y, Cao X, Zhu X, Chen W, Meng X, Wan H, et al. Characterization and co-expression analysis of ATP-binding cassette transporters provide insight into genes related to cannabinoid transport in *Cannabis sativa* L. *Int J Biol Macromol* 2023;**242**:124934.
28. Gao M, Chen S, Kong L, Wang L, Meng X, Xie Z, et al. Genome-wide identification and characterization of CONSTANS-like transcription factors reveal that three CsCOLs regulate the cannabinoid biosynthesis in Cannabis. *Plant Physiol Bioch* 2025;**224**:109942.
29. Woźniczka K, Konieczyński P, Plenis A, Bączek T, Roszkowska A. SPME as a green sample-preparation technique for the monitoring of phytocannabinoids and endocannabinoids in complex matrices. *J Pharm Anal* 2023;**13**:1117−34.
30. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;**32**:268−74.
31. Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, et al. Gene duplication and evolution in recurring polyploidization−diploidization cycles in plants. *Genome Biol* 2019;**20**:38.
32. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant* 2020;**13**:1194−202.
33. Lei D, Qiu Z, Wu J, Qiao B, Qiao J, Zhao G-R. Combining metabolic and monoterpene synthase engineering for *de novo* production of monoterpene alcohols in *Escherichia coli*. *ACS Synth Biol* 2021;**10**: 1531−44.
34. Martin VJJ, Pitera DJ, Withers ST, Newman JD, Keasling JD. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat Biotechnol* 2003;**21**:796−802.
35. Gülck T, Møller BL. Phytocannabinoids: origins and biosynthesis. *Trends Plant Sci* 2020;**25**:985−1004.
36. Lynch RC, Padgitt-Cobb LK, Garfinkel AR, Knaus BJ, Hartwick NT, Allsing N, et al. Domesticated cannabinoid synthases amid a wild mosaic cannabis pangenome. *Nature* 2025;**643**:1001−10.
37. Braich S, Baillie RC, Jewell LS, Spangenberg GC, Cogan NOI. Generation of a comprehensive transcriptome atlas and transcriptome dynamics in medicinal Cannabis. *Sci Rep* 2019;**9**:16583.
38. Ma J, Li D, Xu M, Xia Y, Fang G, Wei Q, et al. Multifunctional sesquiterpene synthase from *Curcuma wenyujin* reveals the biosynthetic mechanism of sesquiterpenes with diverse skeletons. *Int J Biol Macromol* 2025;**318**:145152.
39. Degenhardt J, Köllner TG, Gershenzon J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* 2009;**70**:1621−37.
40. Jiang S-Y, Jin J, Sarojam R, Ramachandran S. A comprehensive survey on the terpene synthase gene family provides new insight into its evolutionary patterns. *Genome Biol Evol* 2019;**11**:2078−98.
41. Luck K, Chen X, Norris AM, Chen F, Gershenzon J, Köllner TG. The reconstruction and biochemical characterization of ancestral genes furnish insights into the evolution of terpene synthase function in the Poaceae. *Plant Mol Biol* 2020;**104**:203−15.
42. Chen H, Köllner TG, Li G, Wei G, Chen X, Zeng D, et al. Combinatorial evolution of a terpene synthase gene cluster explains eerpene variations in Oryza. *Plant Physiol* 2020;**182**:480−92.

43. Chen S, Zhang M, Ding S, Xu Z, Wang S, Meng X, et al. Comprehensive characterization of volatile terpenoids and terpene synthases in *Lanxangia tsaoko*. *Mol Hortic* 2025;**5**:20.
44. Cheah LC, Liu L, Plan MR, Peng B, Lu Z, Schenk G, et al. Product profiles of promiscuous enzymes can be altered by controlling *in vivo* spatial organization. *Adv Sci (Weinh)* 2023;**10**:e2303415.
45. Liu C, Smit SJ, Dang J, Zhou P, Godden GT, Jiang Z, et al. A chromosome-level genome assembly reveals that a bipartite gene cluster formed *via* an inverted duplication controls monoterpenoid biosynthesis in *Schizonepeta tenuifolia*. *Mol Plant* 2023;**16**:533−48.
46. Frey M, Gohr ST, Köllner TG, Bathe U, Lackus ND, Padilla-Gonzalez F, et al. Biosynthesis of biologically active terpenoids in the mint family (Lamiaceae). *Nat Prod Rep* 2025;**42**:1887−908.
47. Livingston SJ, Rensing KH, Page JE, Samuels AL. A polarized supercell produces specialized metabolites in cannabis trichomes. *Curr Biol* 2022;**32**:4040−4047.e4.
48. Kortbeek RWJ, Xu J, Ramirez A, Spyropoulou E, Diergaarde P, Otten-Bruggeman I, et al. Engineering of tomato glandular trichomes for the production of specialized metabolites. *Methods Enzymol* 2016;**576**: 305−31.
49. Bao T, Kimani K, Li Y, Li H, Yang Y, Zhang J, et al. Allelic variation of terpene synthases drives terpene diversity in the wild species of the Freesia genus. *Plant Physiol* 2023;**192**:2419−35.
50. Köllner TG, Schnee C, Gershenzon J, Degenhardt J. The variability of sesquiterpenes emitted from two Zea mays cultivars is controlled by allelic variation of two terpene synthase genes encoding stereoselective multiple product enzymes. *Plant Cell* 2004;**16**:1115−31.
51. Liu S, Yang S, Su P. Chemo-enzymatic synthesis of bioactive compounds from traditional Chinese medicine and medicinal plants. *Sci Tradit Chin Med* 2024;**2**:95−103.
52. Xu S, Kreitzer C, McGale E, Lackus ND, Guo H, Köllner TG, et al. Allelic differences of clustered terpene synthases contribute to correlated intraspecific variation of floral and herbivory-induced volatiles in a wild tobacco. *New Phytol* 2020;**228**:1083−96.
53. Bleeker PM, Mirabella R, Diergaarde PJ, VanDoorn A, Tissier A, Kant MR, et al. Improved herbivore resistance in cultivated tomato with the sesquiterpene biosynthetic pathway from a wild relative. *Proc Natl Acad Sci U S A* 2012;**109**:20124−9.
54. Pazouki L, Niinemets Ü. Multi-substrate terpene synthases: their occurrence and physiological significance. *Front Plant Sci* 2016;**7**: 1019.
55. Sallaud C, Rontein D, Onillon S, Jabès F, Duffé P, Giacalone C, et al. A novel pathway for sesquiterpene biosynthesis from Z,Z-farnesyl pyrophosphate in the wild tomato Solanum habrochaites. *Plant Cell* 2009;**21**:301−17.
56. Cao X, Qin Z, Fan P, Wang S, Meng X, Wan H, et al. A simple widely applicable hairy root transformation method for gene function studies in medicinal plants. *Acta Pharm Sin B* 2025;**15**:4300−5.
57. Chen J, Wang Y, Ko J. Single-cell and spatially resolved omics: advances and limitations. *J Pharm Anal* 2023;**13**:833−5.