

RESEARCH

Open Access



Transcriptomic analysis of wild *Cannabis sativa*: insights into tissue- and stage-specific expression and secondary metabolic regulation

Jinyuan Hu^{1†}, Zishi Wang^{2†}, He Xu², Zhenlong Wang², Ning Li¹, Rui Feng¹, Jianyu Yin¹, Fangru Liu¹ and Baishi Wang^{1*}

Abstract

Cannabis sativa is a medicinally and economically significant plant known for its production of cannabinoids, terpenoids, and other secondary metabolites. This study presents a transcriptomic analysis to elucidate tissue-specific expression and regulatory mechanisms across leaves, stems, and roots. A total of 2,530 differentially expressed genes (DEGs) were identified, with key genes such as terpene synthase (*TPS*) and phenylalanine ammonia-lyase (*PAL*) exhibiting elevated expression in leaf tissues, emphasizing their roles in terpenoid and phenylpropanoid biosynthesis. Alternative splicing (AS) analysis revealed 8,729 distinct events, dominated by exon skipping, contributing to transcriptomic diversity. Long non-coding RNA (lncRNA) prediction identified 3,245 candidates, many of which displayed tissue-specific expression patterns and co-expression with metabolic genes, suggesting regulatory roles in secondary metabolism. Additionally, 12,314 SNPs and 2,786 INDELs were detected, with notable enrichment in genes associated with secondary metabolite biosynthesis, particularly in leaf tissues. These findings advance the understanding of molecular mechanisms governing secondary metabolism and genetic diversity in *C. sativa*, providing valuable insights for future metabolic engineering and breeding strategies to enhance cannabinoid production.

Highlights

- Transcriptomic profiling of wild *Cannabis sativa* identified 2,530 DEGs with tissue- and stage-specific patterns in secondary metabolism.
- Integrated analysis revealed extensive alternative splicing, lncRNA co-expression, and metabolic gene variants associated with cannabinoid pathways.
- qRT-PCR validation confirmed RNA-seq-based expression trends of key metabolic genes across root, stem, and leaf tissues.

Keywords *Cannabis sativa*, Transcriptomics, Secondary metabolism, Alternative splicing, Genetic variation.

[†]Jinyuan Hu and Zishi Wang contributed equally to this work.

*Correspondence:
Baishi Wang
wangbs1018@symc.edu.cn

¹School of Basic Medicine, Shenyang Key Laboratory for Phenomics, Liaoning Province Key Laboratory for Phenomics of Human Ethnic Specificity and Critical Illness (LPKL-PHESCI), Shenyang Medical College, Shenyang 110034, China

²School of Life Sciences, Zhengzhou University, Zhengzhou 450001, China



Introduction

Scientific and economic relevance of *Cannabis sativa*

C. sativa is a plant of considerable economic, medicinal, and industrial significance, widely recognized for its ability to produce a diverse array of bioactive secondary metabolites, including cannabinoids, terpenoids, and flavonoids. These compounds are extensively applied in pharmaceuticals, nutraceuticals, cosmetics, and industrial products [1, 2]. Cannabinoids, notably tetrahydrocannabinol (THC) and cannabidiol (CBD), have garnered substantial attention due to their therapeutic properties, including analgesic, anti-inflammatory, neuroprotective, and antiepileptic effects [3, 4].

However, the psychoactive nature of THC has led to the classification of *C. sativa* as a controlled substance in many regions, imposing legal constraints on its cultivation and research [5]. To navigate these regulatory challenges while maximizing its pharmaceutical potential, a comprehensive understanding of the molecular mechanisms underlying secondary metabolite biosynthesis is essential [6].

Wild cannabis as a source of genetic diversity

While cultivated varieties of *C. sativa* have been extensively investigated, wild populations from the ecologically diverse Himalayan region offer an unparalleled genetic reservoir for exploring natural metabolic diversity and adaptive traits. These wild accessions often exhibit distinct metabolite profiles, providing valuable insights into how secondary metabolism is regulated in response to environmental factors. By studying the genetic and transcriptomic architecture of wild *C. sativa*, we can uncover novel regulatory elements and metabolic pathways that contribute to adaptive evolution [7, 8].

Biosynthetic pathways of cannabinoids and terpenoids

Cannabinoid and terpenoid biosynthesis in *C. sativa* primarily occur in glandular trichomes, specialized secretory structures located on female floral tissues. The metabolic pathways governing these processes are tightly regulated, exhibiting tissue-specific and developmentally programmed expression patterns [9, 10]. Key enzymes, including terpene synthase (TPS) and phenylalanine ammonia-lyase (PAL), play pivotal roles in terpenoid and phenylpropanoid biosynthesis, providing precursors for cannabinoids such as THC and CBD [11, 12].

Cannabinoid biosynthesis begins with the condensation of olivetolic acid and geranyl pyrophosphate (GPP) to form cannabigerolic acid (CBGA), the central precursor of major cannabinoids. This process involves the concerted action of olivetolic acid cyclase (OAC), geranylpyrophosphate: olivetolate geranyltransferase (GOT), and terminal oxidocyclases, including tetrahydrocannabinolic acid synthase (THCAS), cannabidiolic acid

synthase (CBDAS), and cannabichromenic acid synthase (CBCAS) [13, 14]. These biosynthetic pathways are subject to multiple layers of regulation, including transcriptional, post-transcriptional, and epigenetic mechanisms [15].

Environmental influences on secondary metabolism

Beyond genetic regulation, environmental factors such as light intensity, temperature, and nutrient availability significantly influence cannabinoid and terpenoid biosynthesis in *C. sativa*. Light quality, particularly within specific spectra, has been shown to modulate THC and CBD accumulation by altering the expression of key biosynthetic genes [16]. Similarly, temperature fluctuations can impact secondary metabolite production, with low temperatures often reducing THC and CBD concentrations [17].

Nutrient availability, especially nitrogen supply, also plays a critical role in balancing plant growth and metabolite synthesis. Excessive nitrogen fertilization may reduce cannabinoid yield, while optimal nitrogen levels (ranging from 60 to 210 mg/L) promote both biomass production and metabolite accumulation [18]. Despite these insights, the underlying molecular mechanisms connecting environmental stimuli to secondary metabolism remain insufficiently explored, necessitating further investigation.

Challenges in cannabis research and multi-omics approaches

Research on *C. sativa* faces additional challenges due to its controlled status, limiting large-scale cultivation and experimentation in many regions [19]. Furthermore, the absence of a robust genetic transformation system poses significant hurdles for functional genomics studies [20]. These constraints have resulted in a fragmented understanding of the genetic and regulatory networks governing secondary metabolism.

Multi-omics approaches that integrate transcriptomics, proteomics, metabolomics, and epigenomics have emerged as powerful tools to address these limitations [3, 21]. By applying these methodologies to wild *C. sativa*, researchers can uncover complex regulatory networks that drive tissue-specific secondary metabolism. In particular, transcriptomic analyses can provide insights into differentially expressed genes (DEGs), while investigating alternative splicing (AS) events, long non-coding RNAs (lncRNAs), and genetic variations can reveal additional regulatory layers [22–24].

Study objectives and research significance

Building on the genome-level resources established in our previous study of wild-type *C. sativa* [25], this study extends our research by conducting a comprehensive

transcriptomic analysis across leaves, stems, and roots. By comparing tissue-specific gene expression, alternative splicing (AS) events, long non-coding RNAs (lncRNAs), and genetic variations, we aim to uncover the regulatory mechanisms that drive secondary metabolite biosynthesis in a natural context.

This study not only advances our understanding of the molecular underpinnings of cannabinoid biosynthesis but also provides valuable insights for future metabolic engineering, genetic improvement, and conservation efforts. Ultimately, these findings contribute to the expanding knowledge of cannabis molecular biology, supporting the development of innovative biotechnological approaches for optimizing cannabinoid and other secondary metabolite production.

Materials and methods

Plant materials and sample collection

All plant samples used in this study were collected from wild-growing *C. sativa* populations located in the Kyirong Gully region of the southern Himalayas (Tibet, China), a biodiversity hotspot and ecological gene reservoir. These wild accessions share the same genetic background as the female wild-type individual “JL” previously used for genome sequencing and reference assembly (Gao et al., 2020, *Hortic. Res.*).

Tissue samples were collected from leaves, stems, roots, and whole seedlings at three key developmental stages: vegetative, flowering, and fruiting, with the addition of a seedling stage. In total, 52 samples were collected, comprising four biological replicates for each tissue and stage combination.

In the sample labeling system, the leading capital letter in the sample ID represents a specific combination of tissue type and developmental stage. The assignments are as follows:

- A: Leaf tissue at the flowering stage.
- B: Leaf tissue at the fruiting stage.
- C: Leaf tissue at the vegetative stage.
- D: Flower tissue at the flowering stage.
- E: Fruit tissue at the early fruiting stage.
- F: Fruit tissue at the mature fruiting stage.
- G: Whole-seedling tissue at the seedling stage.
- H: Stem tissue at the fruiting stage.
- I: Stem tissue at the flowering stage.
- J: Stem tissue at the vegetative stage.
- K: Root tissue at the fruiting stage.
- L: Root tissue at the vegetative stage.
- M: Root tissue at the flowering stage.

Each sample was immediately frozen in liquid nitrogen post-harvest and stored at -80 °C to preserve RNA integrity for transcriptomic analysis.

RNA extraction and quality control

Total RNA was extracted from 52 tissue samples of *C. sativa* using the TRIzol reagent (Invitrogen), following the manufacturer's protocol. The samples represented various tissues (leaf, stem, root, and whole-seedling) collected across distinct developmental stages, ensuring comprehensive transcriptomic profiling. RNA purity was assessed using a NanoDrop spectrophotometer (Thermo Fisher Scientific), where OD260/OD280 ratios between 1.8 and 2.1 were considered acceptable. RNA integrity was determined using the Agilent 2100 Bioanalyzer (Agilent Technologies), and only samples with RNA integrity numbers (RIN) ≥ 7.0 were retained for subsequent analyses. Four biological replicates were included per tissue to enhance statistical reliability. Frozen tissue samples were processed promptly under RNase-free conditions to minimize RNA degradation.

RNA-Seq library preparation and sequencing

RNA-seq libraries were prepared using the NEBNext Ultra RNA Library Prep Kit (New England BioLabs), targeting mRNA through poly(A) selection to enrich for protein-coding transcripts. Each library was constructed using 1 µg of total RNA as input, following the manufacturer's instructions. After first-strand and second-strand cDNA synthesis, end repair, adenylation, and adapter ligation were performed, followed by PCR amplification to enrich for successfully ligated fragments.

The final libraries were evaluated for concentration using a Qubit fluorometer (Thermo Fisher Scientific) and for quality using an Agilent 2100 Bioanalyzer (Agilent Technologies). Each library had a final concentration of 20 ng/µL and a volume of 50 µL.

High-quality libraries were sequenced on an Illumina NovaSeq 6000 platform, generating paired-end reads of 150 bp. To ensure sufficient coverage, each sample achieved a sequencing depth of 30–50 million reads. In total, the sequencing generated approximately 1.56 billion raw reads across all samples. The sequencing data have been submitted to the NCBI Sequence Read Archive (SRA) under BioProject ID PRJNA1207709.

Quality control and read processing

Raw sequencing data underwent stringent quality control using FastQC (v0.11.9) to evaluate base quality scores, GC content, and adapter contamination. Adapter sequences and low-quality bases were trimmed using Trimmomatic (v0.39) with the following parameters: sliding window trimming for bases below Q30 and removal of reads shorter than 50 bp. Post-trimming, MultiQC (v1.11) was used to aggregate quality metrics, ensuring that all samples met quality thresholds before proceeding to downstream analyses.

Transcript assembly and quantification

High-quality reads were mapped to the *C. sativa* reference genome (version CS10) using HISAT2 (v2.2.1), with default parameters optimized for spliced read alignment. Alignment statistics, including mapping rates and mismatch frequencies, were calculated to ensure accuracy. Transcript assembly was performed using StringTie (v2.1.7), generating a comprehensive transcript model for each sample. Gene expression levels were normalized to fragments per kilobase of transcript per million mapped reads (FPKM) to account for variations in sequencing depth and transcript length. Alignment results were visualized in the Integrative Genomics Viewer (IGV, v2.12) to verify transcript boundaries and read coverage.

Differential gene expression analysis

Differential expression analysis was conducted using DESeq2 (v1.30.1) to identify significantly upregulated and downregulated genes across tissues and developmental stages. Statistical significance was determined by an adjusted p -value ≤ 0.05 and a fold change threshold of ≥ 2 . Identified DEGs were annotated using functional enrichment tools. GO enrichment focused on biological processes, such as “secondary metabolic processes” and “response to stimuli,” while KEGG pathway analysis highlighted terpenoid and phenylpropanoid biosynthesis. Enrichment visualizations were generated using clusterProfiler (v4.2).

qRT-PCR validation of selected DEGs

To experimentally validate the transcriptomic findings and assess the tissue- and stage-specific expression of key metabolic genes, quantitative real-time PCR (qRT-PCR) was conducted on 10 of representative differentially expressed genes (DEGs) involved in cannabinoid and terpenoid biosynthesis. The selected genes—*CBDAS*, *THCAS*, *TPS1*, *PAL1*, *GPPS*, *OAC*, *AAEL*, *LOX*, *CYP71D*, and *CHS*—exhibited significant differential expression across tissues and developmental stages in the RNA-seq dataset, particularly showing elevated expression in leaf tissues.

Total RNA was extracted from six sample groups representing three tissue types (leaf, stem, root) at both the vegetative and flowering stages. The specific biological replicates used for qRT-PCR were consistent with those selected for transcriptome sequencing, including samples C-444 to C-446 (vegetative leaves), J-470 to J-473 (vegetative stems), L-478 to L-481 (vegetative roots), A-436 to A-438 (flowering leaves), I-466 to I-468 (flowering stems), and M-482 to M-484 (flowering roots). Total RNA was isolated using the TRIzol reagent (Invitrogen, USA), and RNA integrity was verified using the Agilent 2100 Bioanalyzer.

First-strand cDNA was synthesized from 1 μ g of high-quality RNA using the PrimeScript RT reagent kit (Takara, Japan) with oligo(dT) primers. qRT-PCR reactions were performed in 20 μ L volumes using SYBR Green Master Mix (Applied Biosystems, USA) on an ABI StepOnePlus™ Real-Time PCR system. The thermal cycling conditions consisted of an initial denaturation at 95 °C for 3 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 30 s. All reactions were performed in triplicate with three biological replicates per tissue.

Gene-specific primers were designed using Primer3 based on the transcript sequences derived from the assembled RNA-seq data. Primer specificity was confirmed by melting curve analysis and single-band amplification. The *ACTIN* gene was selected as the internal control due to its stable expression across tissue types and developmental stages. Relative gene expression levels were calculated using the $2^{-\Delta\Delta C_t}$ method. The root tissue was designated as the reference group, and expression values for other tissues and stages were normalized accordingly. Expression data were presented as fold changes relative to the root baseline (set to 1.0).

Alternative splicing analysis

Alternative splicing (AS) events were analyzed using rMATS (v4.1.0), enabling the detection of exon skipping, intron retention, alternative 5' splice sites, alternative 3' splice sites, and mutually exclusive exons. Tissue-specific splicing patterns were evaluated, and statistical significance was determined using a false discovery rate (FDR) threshold of 0.05. Genes involved in secondary metabolism exhibited notable splicing variations, particularly in terpene synthase (TPS) and phenylalanine ammonia-lyase (PAL) genes. Exon skipping was the most prevalent AS event across all tissues.

lncRNA prediction and co-expression analysis

Long non-coding RNAs (lncRNAs) were predicted using CPC2 (v2.0) and CNCI (v2.0), integrating structural features and coding potential scores. To ensure reliability, lncRNAs shorter than 200 bp or exhibiting high coding potential were removed.

To investigate lncRNA functional roles, a weighted gene co-expression network analysis (WGCNA, v1.70) was performed to construct co-expression networks. Pairwise Pearson correlation coefficients were calculated between lncRNAs and differentially expressed genes (DEGs), with a soft-threshold power of $\beta = 8$, determined based on scale-free topology model fitting. Modules were identified using dynamic tree cutting with a minimum module size of 30 genes, and hub lncRNAs were selected based on intra-modular connectivity ($k_{ME} > 0.7$).

Functional enrichment analysis of genes co-expressed with lncRNAs was performed using GO (Gene Ontology)

and KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analyses, focusing on pathways related to terpenoid and phenylpropanoid biosynthesis. The significance threshold was set at adjusted $p < 0.05$ (Benjamini-Hochberg correction).

To validate the tissue-specific expression of candidate lncRNAs, FPKM values were visualized using heatmaps, and their regulatory relationships with metabolic genes (e.g., TPS, PAL) were further examined through network visualization using Cytoscape (v3.9).

Variant calling and annotation

SNP and INDEL variants were identified using the GATK HaplotypeCaller (v4.2.0.0) pipeline, following best practices for RNA-seq variant calling. The raw variant dataset underwent a multi-step filtering process to ensure high-confidence variant identification:

Minimum read depth (DP) ≥ 10 to exclude low-coverage regions.

Variant quality score (QUAL) ≥ 30 , ensuring base calling confidence.

Mapping quality (MQ) ≥ 40 , removing poorly aligned reads.

Allele frequency threshold ≥ 0.05 , eliminating ultra-rare variants.

Hard filtering using Fisher Strand Bias (FS < 60) and RMS Mapping Quality (MQRankSum > -12.5).

Variants were annotated using SnpEff (v5.0) to classify synonymous, nonsynonymous, and frameshift mutations, focusing on metabolic genes involved in secondary metabolite biosynthesis. Tissue-specific variant distributions were analyzed, revealing a higher SNP density in leaf samples compared to roots and stems. This pattern suggests potential regulatory variations influencing cannabinoid biosynthesis genes.

Statistical analysis and visualization

All statistical analyses were performed in R (v4.1.2). Data visualization tools, including ggplot2, ComplexHeatmap, and Cytoscape (v3.9), were employed to create expression heatmaps, GO/KEGG enrichment plots, and co-expression network diagrams. Principal component analysis (PCA) and hierarchical clustering were applied to evaluate tissue-specific expression patterns and assess data reproducibility.

Results

Differentially expressed genes (DEGs) identified across tissues and developmental stages

To ensure the accuracy and reliability of subsequent transcriptomic analyses, all samples underwent stringent quality control. Supplementary Table S1 provides a detailed summary of the sequencing data for each sample, including the total clean reads, base pairs, and

alignment rates to the reference genome. The high Q20 and Q30 scores (average $> 94\%$) and GC content consistency among samples underscore the high quality of the sequencing data, which lays a solid foundation for differential gene expression and functional analysis.

The transcriptomic analysis revealed a total of 2,530 differentially expressed genes (DEGs), including 1,482 upregulated genes and 1,048 downregulated genes, highlighting significant tissue-specific expression patterns in *C. sativa*. These DEGs provide insights into the distinct metabolic roles of leaves, stems, and roots, particularly in secondary metabolite biosynthesis and adaptive functions.

In leaf tissues, especially during the flowering and fruiting stages (e.g., samples A-436 and B-1), genes involved in secondary metabolism pathways, such as terpene synthase (TPS) and phenylalanine ammonia-lyase (PAL), were highly expressed. For instance, TPS exhibited an expression level of FPKM = 32.7 (Fig. 1A), emphasizing the critical role of leaves in the production of terpenoids and cannabinoids. This aligns with the metabolic specialization of leaves as primary sites for secondary metabolite biosynthesis. In contrast, root tissues (e.g., K-474 and L-478) displayed lower expression levels of secondary metabolism-related genes but showed significant upregulation of genes associated with primary metabolic processes and signal transduction, reflecting their role in environmental adaptation and nutrient acquisition.

Stem tissues (e.g., H-462 and I-466) exhibited a moderate number of DEGs, primarily linked to structural and transport-related functions. These transcriptional activities align with the primary roles of stems in providing mechanical support and facilitating the movement of water and nutrients. Hierarchical clustering analysis revealed distinct expression profiles across tissues, with terpenoid and phenylpropanoid biosynthesis pathways clustering prominently in leaf samples, stress response pathways dominating root samples, and vascular transport genes characterizing stem samples. These findings highlight the metabolic and functional specialization of tissues in *C. sativa*.

To further dissect the transcriptional differences across tissues and developmental stages, the distribution of DEGs was analyzed (Fig. 1B). The largest number of DEGs was observed in the B-vs-C comparison (1,512 genes), with the majority being upregulated, while the K-vs-L comparison exhibited the fewest DEGs (450 genes). Leaf tissues, particularly during flowering and fruiting stages, contributed to the majority of these DEGs, underscoring their transcriptional dominance in secondary metabolism. Conversely, root and stem tissues demonstrated fewer DEGs, reflecting their more conserved roles in primary metabolism and structural support.

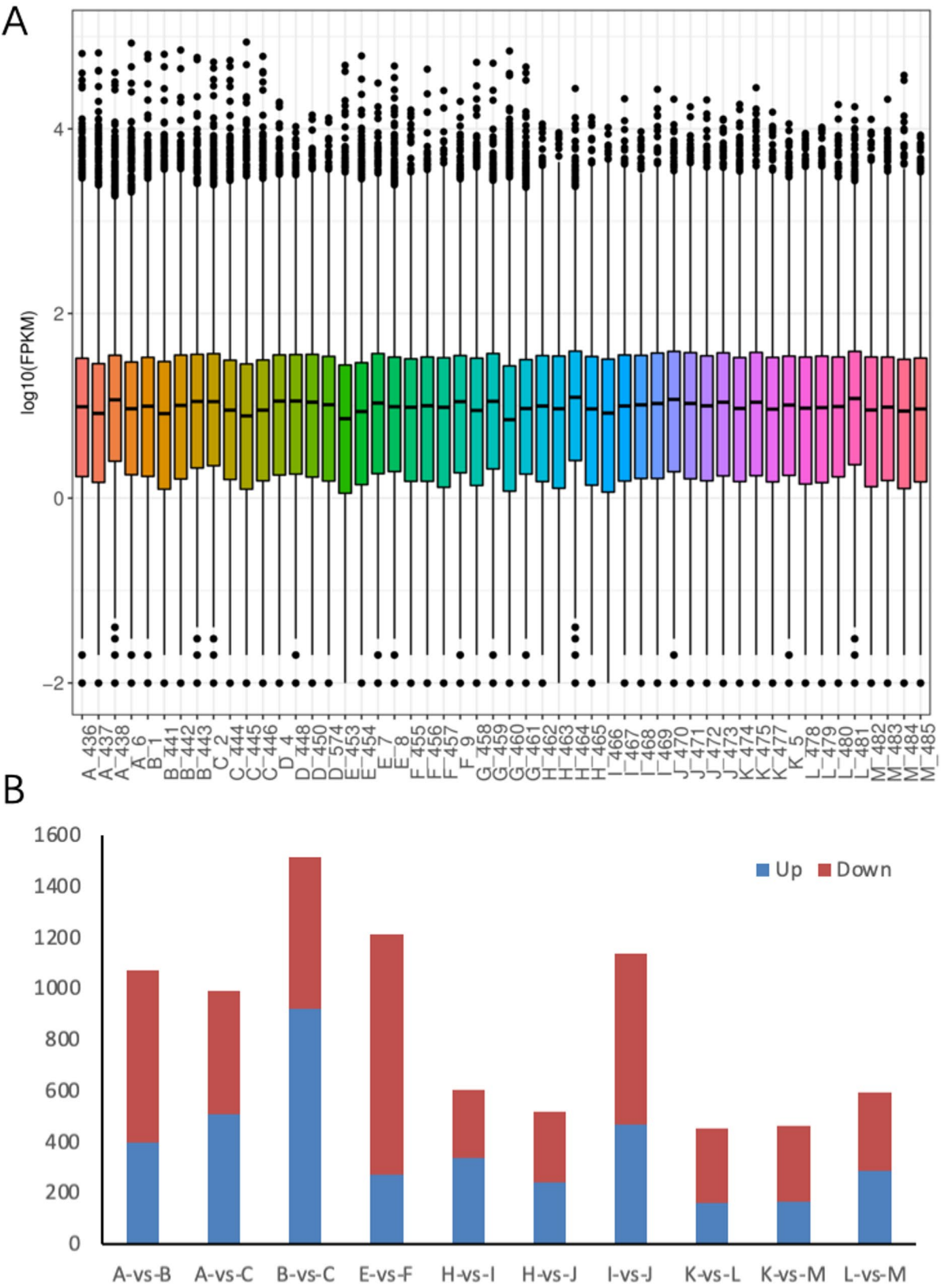


Fig. 1 Transcriptomic Patterns Across Tissues and Developmental Stages in *C. sativa*. **(A)** Distribution of gene and isoform expression levels (FPKM) across analyzed samples. Each box represents the expression distribution within a specific sample, with the median, interquartile range, and extreme values indicated. **(B)** Differentially expressed genes (DEGs) between tissue comparisons. Bars represent total DEGs, with upregulated genes in green and down-regulated genes in red. The x-axis denotes pairwise comparisons, and the y-axis indicates the number of DEGs

The spatial differentiation of DEGs underscores the tissue-specific adaptation of *C. sativa*. The upregulation of key biosynthetic genes, such as TPS and PAL, in leaves reflects their central role in cannabinoid biosynthesis, while the transcriptional activities in roots and stems emphasize their contributions to environmental adaptation and structural integrity. These findings provide a molecular framework for understanding the tissue-specific metabolic functions in *C. sativa* and lay the foundation for future investigations into the regulatory mechanisms driving secondary metabolism in different tissues.

Functional enrichment analysis highlights Tissue-Specific metabolic pathways

To elucidate the biological significance of the differentially expressed genes (DEGs) identified in *C. sativa*, comprehensive Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed. These analyses revealed distinct functional categories and metabolic pathways enriched in different tissues, shedding light on the molecular mechanisms underlying tissue-specific metabolic specialization and secondary metabolism.

GO enrichment analysis identified significant terms associated with secondary metabolic processes, including “secondary metabolic process” (93 genes, $p=0.003$), “oxidoreductase activity” (57 genes, $p=0.012$), and “membrane-associated transport” (48 genes, $p=0.021$). Leaf tissues, particularly those in flowering and fruiting stages (e.g., samples A-436 and B-1), displayed strong enrichment in secondary metabolic processes. Genes such as terpene synthase (TPS) and phenylalanine ammonia-lyase (PAL), pivotal enzymes in terpenoid and phenylpropanoid biosynthesis, were significantly upregulated in these tissues, reflecting their central role in cannabinoid biosynthesis (Fig. 2, Figure S1). This observation aligns with the specialization of leaves as primary sites for secondary metabolite production.

Conversely, DEGs in root tissues (e.g., K-474 and L-478) were enriched in oxidoreductase activity and membrane transport functions, indicating a focus on primary metabolic processes and environmental response. For instance, root-specific DEGs were associated with ion transporters and oxidative stress response genes, emphasizing their adaptive roles in nutrient uptake and stress management. Stems (e.g., samples H-462 and I-466) exhibited moderate enrichment in structural and vascular transport-related terms, underscoring their supportive and transport functions.

KEGG pathway analysis corroborated the tissue-specific roles of DEGs, with significant enrichment in terpenoid backbone biosynthesis (28 genes) and phenylpropanoid metabolism (19 genes). In flowering and

fruiting leaf tissues, terpenoid biosynthesis pathways were prominently upregulated, highlighting the transcriptional activity of TPS in producing precursors for cannabinoids such as THC and CBD. Root tissues, in contrast, showed enrichment in amino acid metabolism pathways, reflecting their focus on primary metabolic functions critical for growth and adaptation to environmental challenges. Stems exhibited enrichment in pathways related to cell wall biosynthesis and lignin metabolism, further supporting their structural roles (Fig. 3, Figure S2).

Hierarchical clustering of GO and KEGG-enriched DEGs provided insights into tissue-specific transcriptional networks. Leaf-specific clusters were dominated by genes involved in secondary metabolism, particularly those associated with terpenoid and phenylpropanoid pathways. Root-specific clusters included genes linked to stress response, signal transduction, and primary metabolic pathways. Stem-specific clusters highlighted genes involved in vascular transport, structural integrity, and cell wall biosynthesis.

These findings underscore the functional diversity and specialization of tissues in the enrichment of secondary metabolism-related pathways in leaves highlights their dominant role in cannabinoid biosynthesis, while the functional adaptations in roots and stems align with their primary roles in environmental response and structural support. Supplementary Figures S1 and S2 further detail the GO and KEGG pathway enrichments for specific tissue and developmental stage comparisons, providing a comprehensive view of the transcriptional and metabolic landscape of *C. sativa*. These insights not only enhance our understanding of the regulatory mechanisms underlying tissue-specific functions but also offer potential targets for future metabolic engineering efforts.

Alternative splicing events in cannabis genes

Alternative splicing (AS) represents a crucial post-transcriptional mechanism that significantly enhances transcriptome diversity by generating multiple mRNA isoforms from a single gene. In this study, 8,729 AS events were identified across all samples, underscoring the regulatory complexity of (Fig. 4). These events were categorized into five primary types, with exon skipping accounting for the largest proportion (42.3%), followed by intron retention (29.7%), alternative 5' splice sites (7.8%), alternative 3' splice sites (6.2%), and mutually exclusive exons (4.0%). This distribution highlights AS as a pivotal regulatory process shaping the transcriptomic landscape of *C. sativa*.

The dominance of exon skipping is consistent with its established role in plants, particularly in metabolically active species, where it contributes to transcript diversity and functional adaptation. Intron retention, as the second

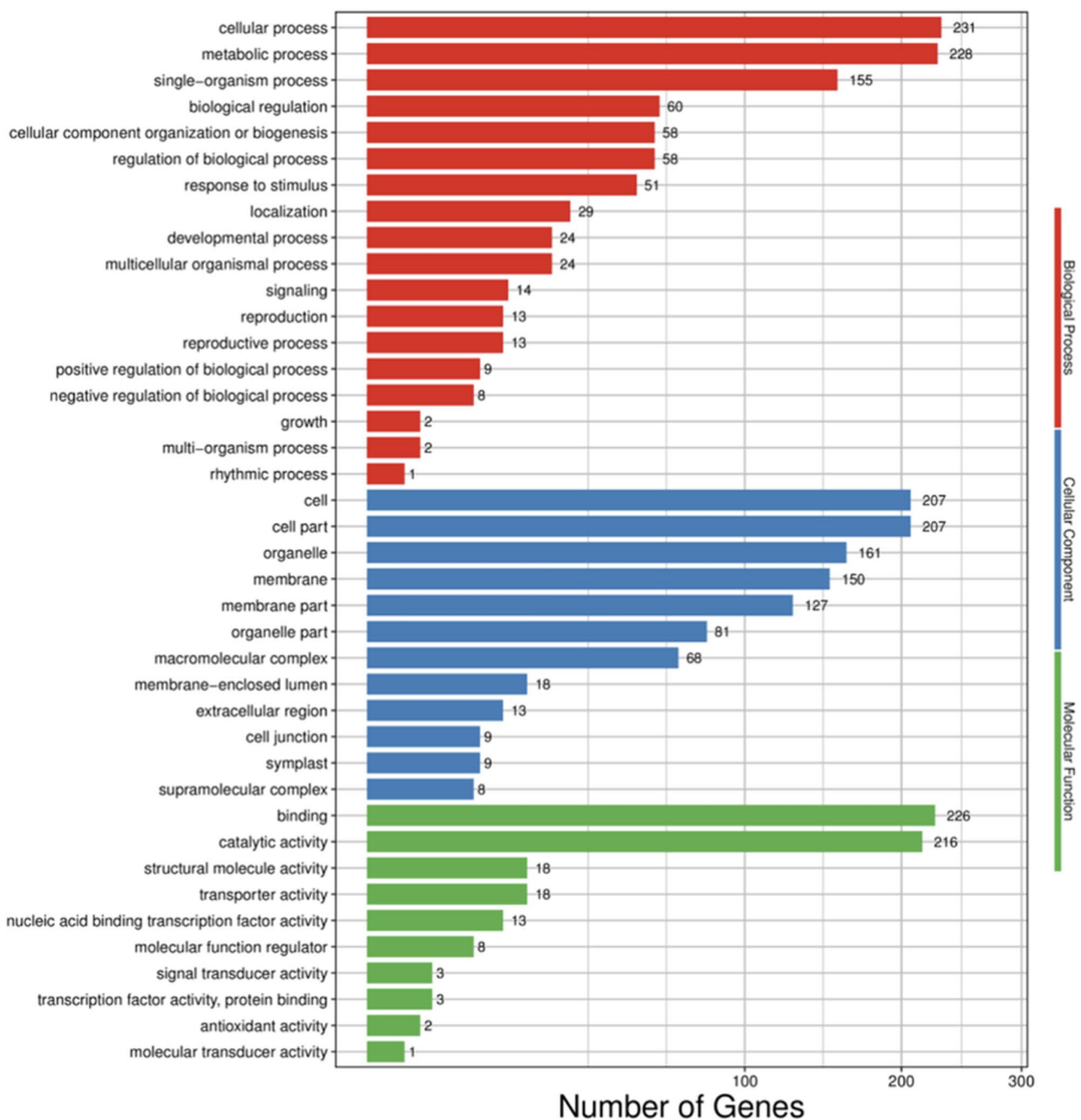


Fig. 2 GO Classification for Differentially Expressed Genes. The y-axis represents the next-level GO terms under the three main GO categories. The x-axis indicates the number of genes annotated to each term, including those associated with its sub-terms. The three classifications represent the primary GO categories: biological process (top), cellular component (middle), and molecular function (bottom)

most prevalent AS event, likely reflects its involvement in fine-tuning transcript stability and translation efficiency. While the dataset primarily provides an overview of AS event distribution, without tissue-specific or stage-specific statistics, the observed trends offer insights into the underlying complexity of transcript regulation.

Certain genes central to secondary metabolic pathways demonstrated notable AS patterns. For example, **terpene**

synthase (TPS), a key enzyme in terpenoid biosynthesis, exhibited multiple isoforms, suggesting a potential link between AS-mediated regulation and the production of secondary metabolites. Similarly, **phenylalanine ammonia-lyase (PAL)**, a critical enzyme in phenylpropanoid metabolism, displayed splicing variations that may influence its functional activity. While the data do not conclusively link these variations to specific tissues, the

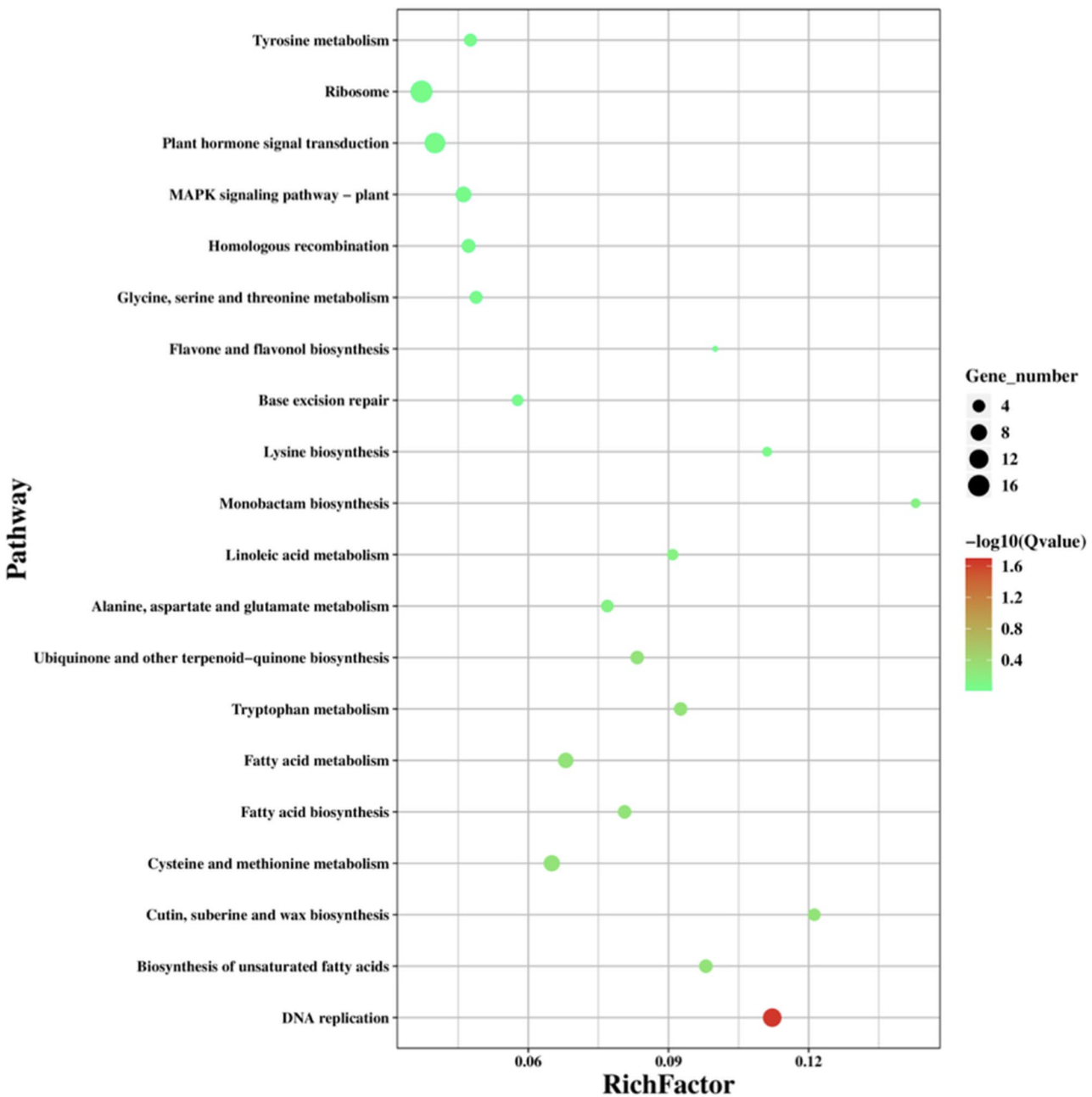


Fig. 3 KEGG Enrichment for Differentially Expressed Genes. The y-axis represents pathway names, sorted by Q-value in ascending order. The x-axis shows the $-\log_{10}(\text{Q-value})$. The size of each dot indicates the number of differentially expressed genes (DEGs) within the pathway, while the color of the dots corresponds to different ranges of the RichFactor

presence of AS in these genes aligns with their roles in pathways essential for secondary metabolite biosynthesis. Collectively, these findings highlight the prominence of exon skipping and intron retention as dominant contributors to AS diversity in *C. sativa*. The regulatory implications of these splicing events underscore their potential roles in modulating metabolic pathways, though further studies are required to delineate their tissue-specific and developmental contributions. This foundational analysis

lays the groundwork for future investigations into the regulatory networks underpinning metabolic specialization in *C. sativa*.

LncRNA prediction and co-expression analysis

Long non-coding RNAs (lncRNAs) are increasingly recognized as crucial regulators of gene expression, influencing transcriptional and post-transcriptional processes. In this study, a comprehensive transcriptomic

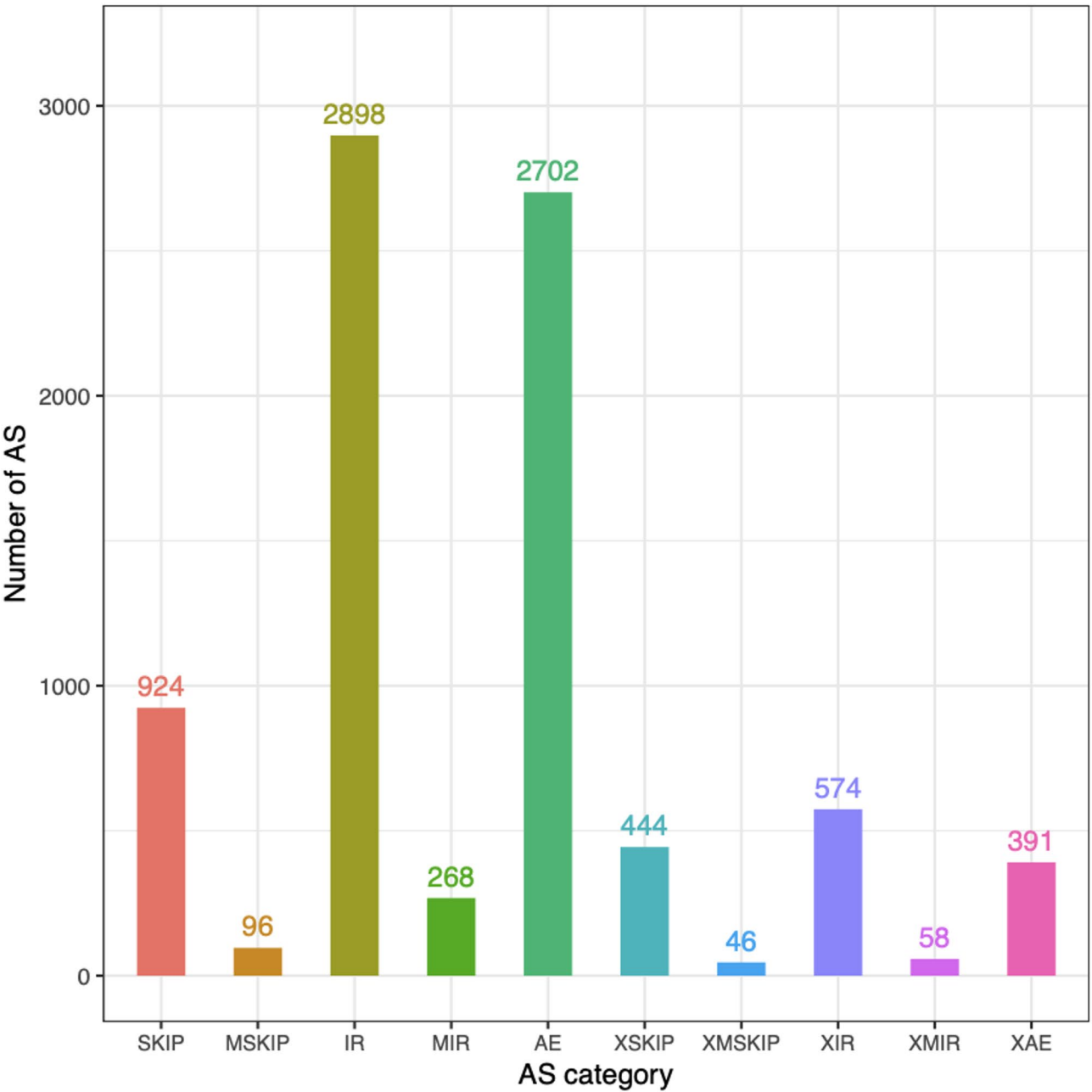


Fig. 4 Classification and Quantification of Alternative Splicing Events. The classification and frequency of alternative splicing (AS) events identified in the *C. sativa* transcriptome. The x-axis represents ten categories of AS events, defined as follows: 1. SKIP: Exon skipping, the omission of an exon during splicing. 2. MSKIP: Mutually exclusive exon skipping, where one of two alternative exons is included. 3. IR: Intron retention, the retention of an intron in the mature mRNA. 4. MIR: Mutually exclusive intron retention, where one of two introns is retained. 5. AE: Alternative exon ends, variability at the 5', 3', or both ends of an exon. 6. XSKIP: Approximate exon skipping, a variant of exon skipping. 7. XMSKIP: Approximate mutually exclusive exon skipping, a variant of mutually exclusive exon skipping. 8. XIR: Approximate intron retention, a variant of intron retention. 9. XMIR: Approximate mutually exclusive intron retention, a variant of mutually exclusive intron retention. 10. XAE: Approximate alternative exon ends, a variant of alternative exon ends. The y-axis indicates the number of AS events in each category. Among these, SKIP (exon skipping) was the most prevalent, followed by IR (intron retention). Less frequent types, such as AE (alternative exon ends) and their approximate variants, also contributed to transcript isoform diversity, albeit to a smaller extent

analysis predicted 3,245 putative lncRNAs in *C. sativa*. To ensure the reliability of these predictions, stringent criteria were applied, differentiating lncRNAs from coding RNAs based on their transcript length, open reading

frame (ORF) size, and statistical scoring metrics such as Fickett and Hexamer scores. Table 1 presents a representative subset of these lncRNAs, illustrating the key attributes that define their non-coding nature. The transcript lengths of these

Table 1 Representative lncRNAs predicted in *C. sativa*

ID	mRNA_size	ORF_size	Fickett_score	Hexamer_score	coding_prob
NOVEL.4.1	577	150	0.7293	0.046036873	0.0973031
NOVEL.5.1	254	105	0.7281	-0.082092359	0.0338590
NOVEL.9.1	274	24	0.6429	-0.679391156	0.0000744
NOVEL.12.1	411	57	1.1312	-0.17482354	0.0093709
NOVEL.13.1	218	84	1.0227	-0.157960125	0.0155264
NOVEL.21.4	616	234	0.5544	0.160131755	0.3260356
NOVEL.23.1	353	84	0.565	-0.200852687	0.0089941
NOVEL.22.1	238	84	0.6565	-0.349590657	0.0024770
NOVEL.28.1	915	99	0.9222	-0.046371259	0.0231512
NOVEL.30.1	238	48	1.0964	-0.252651101	0.0051253
NOVEL.33.1	841	108	0.8332	-0.134724763	0.0114569
NOVEL.34.1	390	45	0.9267	0.033712012	0.0620419
NOVEL.35.1	223	57	0.7534	-0.315509169	0.0030075
NOVEL.40.1	326	159	0.6307	-0.28231007	0.0064097
NOVEL.46.1	206	72	0.5864	-0.116295067	0.0217643
NOVEL.49.1	249	72	1.3293	-0.217914553	0.0080147
NOVEL.55.13	427	63	0.8249	-0.25057987	0.0046423
NOVEL.50.1	403	144	0.6212	0.055228958	0.1204166
NOVEL.54.1	364	114	1.0434	0.014663001	0.0761537
NOVEL.60.1	248	96	0.5752	-0.027052389	0.0537025

lncRNAs exceed 200 base pairs, aligning with the typical classification threshold. Furthermore, their ORFs are relatively short, a characteristic consistent with their non-coding designation. Statistical scores, including Fickett and Hexamer metrics, indicate low coding potential, further corroborated by coding probability values well below established thresholds.

This representative dataset underscores the diverse characteristics of lncRNAs in *C. sativa* and highlights their potential regulatory roles. The high-confidence predictions provide a robust foundation for downstream analyses, such as co-expression studies linking lncRNAs to genes involved in secondary metabolic pathways like terpenoid and phenylpropanoid biosynthesis. By showcasing these representative lncRNAs, Table 1 enhances our understanding of the molecular features that distinguish lncRNAs and sets the stage for functional exploration of their roles in tissue-specific and developmentally regulated metabolic processes.

The overall evaluation of coding potential is comprehensively illustrated in Fig. 5, which integrates multiple analytical perspectives to validate the robustness of lncRNA prediction. The Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves assess the model's performance in distinguishing coding RNAs from non-coding RNAs. The ROC curve demonstrates a high area under the curve (AUC), indicating strong sensitivity and specificity across varying cutoff thresholds. Similarly, the PR curve reflects the model's precision in identifying non-coding RNAs, particularly at low false positive rates.

Additionally, the plot of accuracy versus cutoff values provides insights into the optimal threshold for defining coding probability, balancing sensitivity and specificity to maximize predictive reliability. The two-graph ROC curve further refines this analysis by overlaying sensitivity and specificity curves, allowing the intersection point to serve as the most reliable cutoff. This ensures that the identified lncRNAs are classified with minimal false positives and negatives.

By combining these metrics, Fig. 5 highlights the rigorous approach taken to evaluate and select coding probability thresholds, ensuring that the predicted lncRNAs possess high-confidence non-coding potential. These analyses strengthen the validity of subsequent functional studies and underscore the reliability of the computational pipeline employed in this study.

Among these lncRNAs, a subset exhibited tissue-specific expression patterns, with 1,084 showing significant enrichment in specific tissues. While precise tissue-specific expression was not fully resolved, the identification of functional categories linked to secondary metabolism provides a basis for future investigations into regulatory networks.

Functional co-expression analysis indicated that certain lncRNAs may interact with genes involved in key secondary metabolic pathways, such as terpenoid backbone biosynthesis and phenylpropanoid metabolism. However, the lack of direct evidence for tissue-specific co-expression relationships in the current dataset limits conclusive functional annotations.

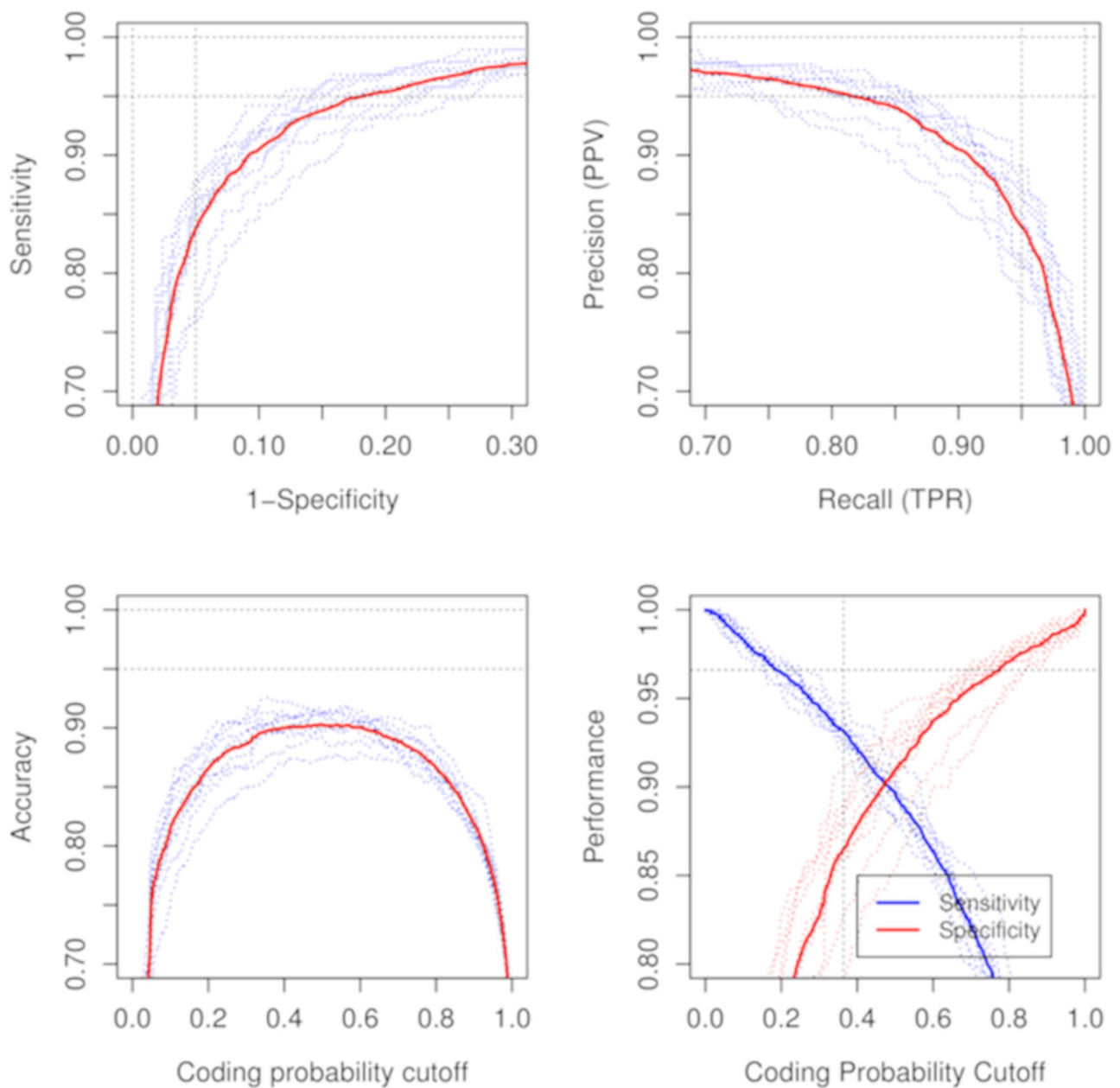


Fig. 5 Coding Potential Evaluation of Predicted lncRNAs. **(A)** ROC curve: Performance of the coding potential classifier under different thresholds. **(B)** PR curve: Precision-Recall relationship for lncRNA identification. **(C)** Accuracy vs. cutoff: Statistical evaluation of classification accuracy. **(D)** Two-graph ROC curve: Determination of the optimal cutoff value for distinguishing coding and non-coding RNAs

The identification of a substantial number of lncRNAs and their predicted regulatory roles emphasizes the potential complexity of transcriptomic regulation in *C. sativa*. These findings lay a robust foundation for further studies into lncRNA functions and their potential applications in metabolic engineering.

SNP and INDEL variants associated with metabolic pathways

Genomic variants, including single nucleotide polymorphisms (SNPs) and insertions/deletions (INDELs), represent critical components of genetic diversity and play essential roles in shaping gene regulation and metabolic specialization in *C. sativa*. This study identified a total of 12,314 SNPs and 2,786 INDELs across all analyzed samples, as detailed in Table 2. These variants, distributed

Table 2 Summary of SNP and INDEL variants across Cannabis tissues

Sample	Total SNP	Homozygosis SNP	Heterozygosis SNP	Total Indel	Homozygosis Indel	Heterozygosis Indel
A_436	86,678	38,777	47,901	10,509	5614	4895
A_437	81,473	38,589	46,287	9828	5107	4721
B_441	109,900	47,563	62,337	13,121	6836	6285
B_442	101,822	43,004	58,818	12,046	6338	6068
C_444	90,897	37,328	53,569	10,942	5434	5508
C_445	100,663	43,139	59,270	12,118	6078	6040
D_448	94,312	41,393	52,919	12,221	6352	5869
D_450	94,912	41,939	52,973	11,481	6021	5460
E_453	116,947	50,387	66,360	14,033	7315	6718
E_454	111,292	50,556	60,736	13,476	7262	6214
F_455	113,752	47,934	75,918	14,702	6931	7771
F_456	117,780	50,093	67,687	14,424	7264	7160
G_458	106,836	45,372	61,464	12,964	6645	6319
G_459	96,459	39,433	59,469	12,610	6645	5965
H_462	128,892	54,825	74,067	15,878	8339	7539
H_463	128,459	56,751	71,708	15,878	8339	7539
I_466	117,844	52,252	64,962	14,509	7748	6761
I_467	120,306	51,962	68,614	14,950	7703	7247
J_470	62,070	47,092	47,047	7109	3509	3600
J_471	110,613	45,187	65,426	13,842	6928	6914
K_474	110,981	47,517	63,464	13,531	7066	6465
K_475	112,825	46,461	58,373	13,664	6899	6577
L_479	111,329	45,747	65,852	13,893	7026	7345
L_480	107,160	44,495	62,665	13,462	6795	6667
M_484	114,659	52,436	62,223	14,227	7735	6492
M_485	109,088	46,421	62,667	13,024	6600	6424

across intergenic, intronic, and coding regions, provide valuable insights into the genetic mechanisms driving tissue-specific metabolic adaptations.

The majority of SNPs and INDELs were located in non-coding regions, such as intronic and intergenic sequences, consistent with their regulatory roles. Coding region variants, although less frequent, included nonsynonymous mutations with the potential to alter protein function. Leaf tissues (e.g., A-436, B-441) exhibited the highest total number of variants, highlighting their metabolic plasticity. This observation aligns with the transcriptional demands required to sustain secondary metabolite biosynthesis during flowering and fruiting stages.

Root tissues (e.g., K-474, L-478) showed a different pattern, with a notable proportion of variants in regulatory regions associated with stress-responsive and primary metabolic pathways. Stem tissues (e.g., H-462, I-466) displayed moderate SNP and INDEL counts, enriched in genes related to structural integrity and transport, reflecting their physiological roles in support and nutrient conduction.

Table 2 summarizes the total number of SNPs and INDELs identified across various tissue samples and developmental stages, offering a detailed perspective on genetic diversity within *C. sativa*. The data provide

insights into the distribution of homozygous and heterozygous variants, revealing tissue-specific genetic compositions. Heterozygous SNPs are more abundant in leaf tissues, reflecting their higher genetic variability and dynamic metabolic activity. Conversely, homozygous SNPs are predominantly observed in root and stem tissues, indicating a more stable genetic structure tailored to their primary functions. Similarly, the distinction between homozygous and heterozygous INDELs highlights comparable patterns, where leaves exhibit greater variability, while roots and stems maintain a more conserved genetic profile. This comprehensive analysis underscores the tissue-specific roles and genetic adaptability of different organs in *C. sativa*.

For instance, leaf samples (e.g., A-436) displayed a higher proportion of heterozygous SNPs and INDELs compared to root (e.g., K-474) and stem (e.g., H-462) samples. This reflects the dynamic genetic activity in leaves, driven by their role as the primary site for secondary metabolism. Conversely, roots and stems demonstrated a greater proportion of homozygous variants, consistent with their conserved metabolic and structural functions.

Functional analysis of coding region variants revealed their involvement in key metabolic pathways. In leaf tissues, nonsynonymous SNPs were enriched in genes such

as terpene synthase (TPS) and phenylalanine ammonia-lyase (PAL), suggesting potential effects on secondary metabolite biosynthesis. Similarly, root variants were linked to amino acid metabolism and hormonal signaling pathways, highlighting their roles in stress adaptation and nutrient uptake. Stem-specific variants were predominantly associated with cell wall biosynthesis and vascular transport.

The distribution and functional enrichment of SNPs and INDELs in *C. sativa* reflect tissue-specific adaptations that align with their distinct metabolic and physiological roles. Variants in key metabolic genes not only enhance our understanding of genetic regulation but also provide valuable markers for breeding programs aimed at optimizing metabolite profiles. Further studies focusing on the functional consequences of these variants are essential to fully unravel their impact on metabolic pathways and plant development.

qRT-PCR validation of key biosynthetic genes

To validate the transcriptomic findings and confirm the tissue- and stage-specific expression trends of candidate metabolic genes, qRT-PCR analysis was performed on ten representative DEGs involved in secondary metabolism, including *CBDAS*, *THCAS*, *TPS1*, *PAL1*, *GPPS*, *OAC*, *AAEI*, *LOX*, *CYP71D*, and *CHS*. These genes were selected based on the following criteria: (i) significant fold change across tissues and stages in RNA-seq analysis; (ii) functional relevance to cannabinoid, terpenoid, or phenylpropanoid biosynthesis; and (iii) enrichment in relevant GO and KEGG pathways.

qRT-PCR was conducted on root, stem, and leaf tissues collected at both the vegetative and flowering stages, using the same biological samples employed in transcriptome sequencing. The *ACTIN* gene was used as an internal control, and expression levels were normalized using the $2^{-\Delta\Delta C_t}$ method, with root tissue designated as the reference baseline (fold change = 1.0).

As shown in Fig. 6, all ten genes displayed strong tissue-specific and developmental stage-dependent expression patterns. At the vegetative stage (Fig. 6A), the majority of genes showed moderate upregulation in leaves (2.6–4.3 fold) relative to roots, with stem tissues exhibiting intermediate expression. In the flowering stage (Fig. 6B), gene expression in leaves was markedly elevated for all targets, particularly for *CBDAS* (8.2-fold) and *THCAS* (7.8-fold), consistent with their roles in active cannabinoid biosynthesis. *PAL1*, *TPS1*, *CHS*, and *GPPS* also showed pronounced upregulation in flowering leaves.

These results were consistent with RNA-seq expression trends and experimentally confirmed the strong activation of secondary metabolic pathways in leaf tissues, especially during the flowering stage. The validated genes provide a robust foundation for future studies on

regulatory mechanisms and genetic improvement in *C. sativa*.

Discussion

Wild-Type *C. sativa* as a valuable genetic resource and preliminary insights

This study presents a preliminary transcriptomic analysis of wild *C. sativa*, providing insights into the potential regulatory mechanisms underlying tissue-specific gene expression and secondary metabolite biosynthesis. The use of wild-type *C. sativa* accessions, previously sequenced at the genome level [25], confers a unique advantage to this study. These accessions exhibit higher heterozygosity and genetic diversity than commercial cultivars, facilitating the discovery of novel regulatory features such as noncanonical splicing events and population-specific variants in key metabolic genes. This reinforces the broader applicability of our findings to natural variation and evolutionary adaptation in *C. sativa*. By integrating differentially expressed genes (DEGs), alternative splicing (AS) events, long non-coding RNA (lncRNA) predictions, and genetic variation analyses, this work advances the understanding of the complex regulatory networks governing metabolic specialization in *C. sativa*.

Tissue-specific metabolic regulation and RT-PCR validation

The identification of 2,530 DEGs across different tissues highlights the tissue-specific roles of secondary metabolism. Leaves, particularly during flowering and fruiting stages, exhibited significant upregulation of key metabolic genes such as terpene synthase (TPS) and phenylalanine ammonia-lyase (PAL), emphasizing their role in the biosynthesis of terpenoids and phenylpropanoids [8]. This aligns with previous studies confirming that leaf tissues serve as primary sites for cannabinoid production [15]. Hierarchical clustering of DEGs further revealed distinct transcriptional signatures across tissues, suggesting spatially regulated metabolic networks [26, 27].

In parallel with transcriptome-wide observations, expression levels of ten genes involved in cannabinoid and terpenoid metabolism were further quantified across tissues and developmental stages. These genes, including *CBDAS*, *THCAS*, *TPS1*, and *PAL1*, exhibited strong transcriptional activation in leaf tissues, particularly during the flowering stage, consistent with their metabolic roles and spatial expression inferred from RNA-seq data. The expression trends confirmed the transcriptional prioritization of cannabinoid biosynthesis pathways in photosynthetically active tissues and aligned with the observed enrichment of secondary metabolism in functional annotation analyses. These results reinforce the reliability of the transcriptomic data and provide additional confidence in the tissue-specific regulatory patterns

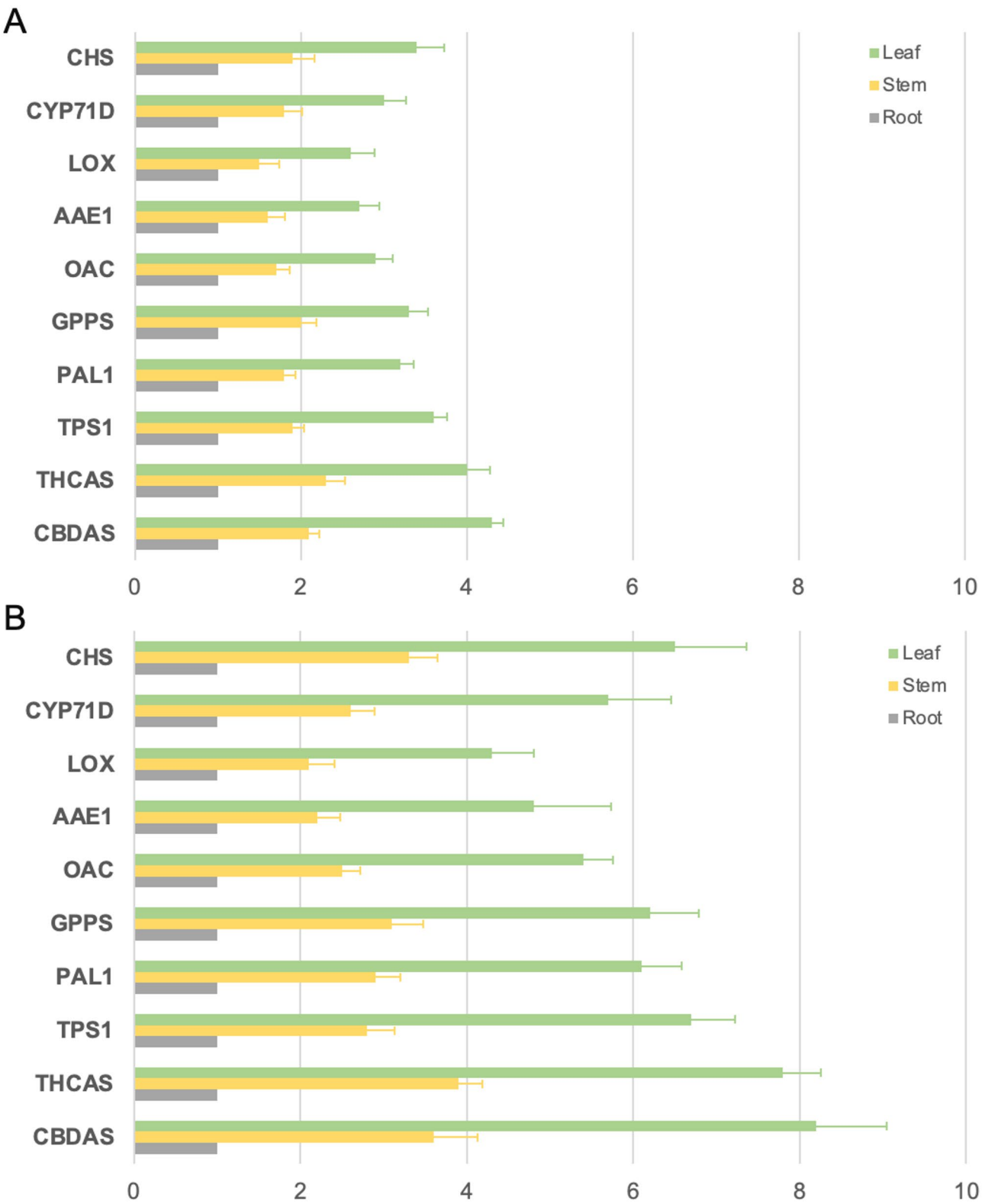


Fig. 6 Validation of key DEGs involved in secondary metabolism by qRT-PCR. **(A)** Expression profiles of ten DEGs across root, stem, and leaf tissues at the vegetative stage. **(B)** Expression profiles of the same DEGs at the flowering stage

proposed for *C. sativa*. However, further investigations using additional independent validation methods, such as proteomics and metabolomics, are necessary to confirm these observations. Additionally, exploring protein activity and metabolite accumulation would provide more direct evidence of functional differences between tissues.

Potential contribution of alternative splicing (AS) to metabolic regulation

This study identified 8,729 distinct alternative splicing (AS) events, with exon skipping being the most prevalent type. AS is a significant contributor to transcriptomic complexity and may play a role in functional diversification of genes, particularly in secondary metabolic pathways [28].

Genes involved in **terpenoid biosynthesis** exhibited a notable enrichment of AS events, suggesting that transcript isoform variation could influence enzyme activity, substrate specificity, or regulatory efficiency [29, 30]. Such transcript diversity may contribute to the fine-tuning of metabolic fluxes, offering a potential mechanism for tissue-specific metabolic regulation.

However, it is important to acknowledge that these observations are predictions based on transcriptomic data and remain speculative without further experimental confirmation. Determining the biochemical consequences of these AS events will require functional validation using protein structural modeling, enzyme kinetic assays, and isoform-specific expression studies. Additionally, integrating proteomics and metabolomics data will provide valuable insights into whether AS-mediated transcript diversity translates into functional differences at the protein and metabolite levels.

Future studies focusing on the functional characterization of AS isoforms, particularly those identified in key biosynthetic genes like TPS and PAL, will be essential to elucidate their roles in secondary metabolism. Such investigations will provide a clearer understanding of how AS contributes to the metabolic adaptability and regulatory flexibility of *C. sativa*.

Putative regulatory roles of lncRNAs in secondary metabolism

lncRNA analysis identified 3,245 putative transcripts, many of which exhibited tissue-specific expression patterns.

Co-expression analysis suggested that specific lncRNAs, particularly those enriched in leaf samples, may fine-tune the transcriptional regulation of metabolic pathways involved in terpenoid and phenylpropanoid biosynthesis [31, 32].

While the involvement of lncRNAs in secondary metabolism has been demonstrated in other medicinal

plants, their precise regulatory roles in *C. sativa* remain to be experimentally validated [33].

Although previous studies have confirmed the participation of lncRNAs in plant secondary metabolism, their specific contributions in *C. sativa* are yet to be determined. The strong correlation between lncRNA expression and the expression of key biosynthetic genes suggests a potential regulatory interaction. However, further functional studies, such as RNA interference (RNAi) or CRISPR/Cas13-based gene knockout experiments, will be essential to confirm their roles.

Potential impact of genetic variations on metabolic diversity

Genetic variation analysis identified 12,314 SNPs and 2,786 INDELs, with a significant number of nonsynonymous SNPs detected in metabolic genes. Studies have shown that genes associated with secondary metabolism tend to exhibit a higher density of nonsynonymous SNPs [34]. Notably, several SNPs located in the coding regions of TPS and PAL genes may contribute to metabolic differences across tissues [35, 36].

Therefore, it can be preliminarily speculated that the variations detected in the coding regions of TPS and PAL genes may lead to differences in enzyme function and substrate specificity. However, these findings require further functional validation to establish their biological significance.

Future studies employing protein structural modeling, enzyme kinetics assays, and comparative metabolomics will be essential to evaluate the impact of these genetic variations. Additionally, integrating genome-wide association studies (GWAS) could further uncover genotype-phenotype relationships, providing deeper insights into the metabolic consequences of these polymorphisms.

These SNPs represent important molecular targets for future structure-function analyses, including protein modeling and enzymatic assays. Clarifying their effects on metabolic flux will enhance our understanding of the natural variation in cannabinoid biosynthesis.

Study limitations and future directions

This study provides preliminary insights into the transcriptomic landscape of wild *C. sativa*, but several limitations remain.

The absence of proteomic and metabolomic data limits our ability to confirm whether transcriptional changes are reflected at the functional level. Integrating proteomic and metabolomic analyses will be essential to validate whether differentially expressed genes (DEGs) and alternative splicing (AS) events translate into functional changes in proteins and metabolites [37]. Future studies employing multi-omics approaches will provide a more

comprehensive understanding of the metabolic regulation in wild *C. sativa*.

Furthermore, while tissue-specific gene expression patterns were observed, this study did not explicitly analyze the effects of environmental factors such as light intensity, temperature, and nutrient availability. These factors are known to significantly influence gene expression and the biosynthesis of secondary metabolites [38–40]. Future research incorporating controlled environment experiments or field-based transcriptomics will be essential to investigate these environmental impacts.

In addition, the functional roles of candidate DEGs, AS events, and lncRNAs remain to be validated using experimental methods such as CRISPR-Cas9 genome editing, RNA interference (RNAi) knockdown, and gene over-expression assays. Expanding the analysis to a broader range of accessions and integrating GWAS may help identify key regulatory factors contributing to metabolic diversity in *C. sativa* [41].

Finally, future studies investigating potential synergistic interactions between cannabinoids and other secondary metabolites, such as terpenoids and flavonoids, may offer new insights into therapeutic applications. A comprehensive understanding of these interactions could facilitate the development of more effective pharmaceutical formulations.

Conclusion

This study provides preliminary insights into the tissue-specific metabolic regulation of wild *C. sativa* through transcriptomic analysis. By identifying differentially expressed genes, alternative splicing events, lncRNAs, and genetic variations, it suggests potential regulatory mechanisms underlying secondary metabolite biosynthesis. The RT-PCR validation of key genes supports the reliability of the transcriptomic data. While further functional and multi-omics validation is necessary, these findings offer a valuable foundation for future research aimed at optimizing cannabinoid production and exploring therapeutic applications.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11697-5>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements

We sincerely acknowledge the Institute of Forensic Science, Ministry of Public Security, and Dr. Shan Gao, Hao Nie, Yongjiu Li, and Ying Zhang for their guidance and assistance in conducting this research. Their guidance and support were crucial for sample collection and ensuring compliance with legal research directions.

Author contributions

Conceptualization, J.H., Z.W., and B.W.; Methodology, J.H. and B.W.; Software, N.L.; Validation, Z.W., B.W., and H.X.; Formal Analysis, N.L.; Investigation, Z.W., J.H., B.W., and R.F.; Resources, J.Y., F.L., H.X., and Z.W.; Data Curation, N.L., Z.W., and R.F.; Writing—Original Draft Preparation, Z.W., N.L., and J.H.; Writing—Review & Editing, B.W., Z.W., and H.X.; Visualization, N.L.; Supervision, B.W.; Project Administration, B.W.; Funding Acquisition, B.W. All authors have read and agreed to the published version of the manuscript.

Funding

This study was financially supported by The Fundamental Research Funds for the Central Public Welfare Research Institutes, 2021JB002; The Key Laboratory of Forensic Genetics, Ministry of Public Security Open Research Project, 2022FGKFKT02; Science and technology plan projects in Liaoning province, 2023-MS-329.

Data availability

The sequencing data for this study have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject ID PRJNA1207709.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 15 February 2025 / Accepted: 9 May 2025

Published online: 26 May 2025

References

1. Mechoulam R, Parker LA, Gallily R, 115–195. Cannabidiol: an overview of some Pharmacological aspects. *J Clin Pharmacol*. 2002;42. <https://doi.org/10.1002/j.1552-4604.2002.tb05998.x>.
2. Silva Sofras FM, Desimone MF. Entourage effect and analytical chemistry: chromatography as a tool in the analysis of the secondary metabolism of Cannabis sativa L. *Curr Pharm Des*. 2023;29:394–406. <https://doi.org/10.2174/1381612829666221103093542>.
3. Andre CM, Hausman JF, Guerriero G. Cannabis sativa: the plant of the thousand and one molecules. *Front Plant Sci*. 2016;7:19. <https://doi.org/10.3389/fpls.2016.00019>.
4. Russo EB, Taming THC. Potential cannabis synergy and phytocannabinoid-terpenoid entourage effects. *Br J Pharmacol*. 2011;163:1344–64. <https://doi.org/10.1111/j.1476-5381.2011.01238.x>.
5. Deguchi M, et al. Metabolic engineering strategies of industrial hemp (*Cannabis sativa* L.): A brief review of the advances and challenges. *Front Plant Sci*. 2020;11:580621. <https://doi.org/10.3389/fpls.2020.580621>.
6. Dimopoulos N, et al. From dawn 'til dusk: daytime progression regulates primary and secondary metabolism in Cannabis glandular trichomes. *J Exp Bot*. 2025;76:134–51. <https://doi.org/10.1093/jxb/erae148>.
7. Barbosa-Xavier K, Pedrosa-Silva F, Almeida-Silva F, Venancio TM. Cannabis Expression Atlas: a comprehensive resource for integrative analysis of *Cannabis sativa* L. gene expression. *bioRxiv*. 2024.2009.2027.615413 (2024). <https://doi.org/10.1101/2024.09.27.615413>.
8. Yeo HC, et al. Comparative transcriptome analysis reveals coordinated transcriptional regulation of central and secondary metabolism in the trichomes of Cannabis cultivars. *Int J Mol Sci*. 2022;23. <https://doi.org/10.3390/ijms23158310>.
9. Booth JK, Bohlmann J. Terpenes in Cannabis sativa - From plant genome to humans. *Plant Sci*. 2019;284:67–72. <https://doi.org/10.1016/j.plantsci.2019.03.022>.
10. Flores-Sanchez IJ, Verpoorte R. Secondary metabolism in cannabis. *Phytochem Rev*. 2008;7:615–39. <https://doi.org/10.1007/s11101-008-9094-4>.
11. Lacerda M, et al. Pharmacokinetics of Non-Psychotropic phytocannabinoids. *Pharmaceutics*. 2025;17. <https://doi.org/10.3390/pharmaceutics17020236>.

12. Alayoubi M, Henry BA, Cahill CM, Cooper ZD. Exploring novel pharmacotherapy candidates for Cannabis use disorder: Uncovering promising agents on the horizon by mechanism of action. *Drugs*. 2024;84:1395–417. <https://doi.org/10.1007/s40265-024-02098-1>.
13. Manzoni OJ, Manduca A, Trezza V. Therapeutic potential of Cannabidiol polypharmacology in neuropsychiatric disorders. *Trends Pharmacol Sci*. 2025;46:145–62. <https://doi.org/10.1016/j.tips.2024.12.005>.
14. Mashabela MD, Kappo AP. Anti-Cancer and Anti-Proliferative potential of Cannabidiol: A cellular and molecular perspective. *Int J Mol Sci*. 2024;25. <https://doi.org/10.3390/ijms25115659>.
15. van Bakel H, et al. The draft genome and transcriptome of Cannabis sativa. *Genome Biol*. 2011;12:R102. <https://doi.org/10.1186/gb-2011-12-10-r102>.
16. Dang M, Arachchige NM, Campbell LG. Optimizing photoperiod switch to maximize floral biomass and cannabinoid yield in Cannabis sativa L.: A Meta-Analytic quantile regression approach. *Front Plant Sci*. 2021;12:797425. <https://doi.org/10.3389/fpls.2021.797425>.
17. Galic A, et al. Effects of cold temperature and acclimation on cold tolerance and cannabinoid profiles of Cannabis sativa L. (Hemp). *Horticulturae*. 2022;8. <https://doi.org/10.3390/horticulturae8060531>.
18. Dilena E, Close DC, Hunt I, Garland SM. Investigating how nitrogen nutrition and pruning impacts on CBD and THC concentration and plant biomass of Cannabis sativa. *Sci Rep*. 2023;13:19533. <https://doi.org/10.1038/s41598-023-46369-5>.
19. Gould J. Cannabis: 4 big questions. *Nature*. 2015;525:18. <https://doi.org/10.1038/525518a>.
20. Schilling S, Melzer R, McCabe PF. Cannabis sativa. *Curr Biol*. 2020;30:R8–9. <https://doi.org/10.1016/j.cub.2019.10.039>.
21. Sirangelo TM, Ludlow RA, Spadafora ND. Multi-Omics approaches to study molecular mechanisms in Cannabis sativa. *Plants (Basel)*. 2022;11. <https://doi.org/10.3390/plants11162182>.
22. Romero P, Peris A, Vergara K, Matus JT. Comprehending and improving cannabis specialized metabolism in the systems biology era. *Plant Sci*. 2020;298:110571. <https://doi.org/10.1016/j.plantsci.2020.110571>.
23. Jiang H, et al. Single-Molecule Real-Time sequencing of Full-Length transcriptome and identification of genes related to male development in Cannabis sativa. *Plants (Basel)*. 2022;11. <https://doi.org/10.3390/plants11243559>.
24. Wu B, et al. Genome-Wide analysis of alternative splicing and Non-Coding RNAs reveal complicated transcriptional regulation in Cannabis sativa L. *Int J Mol Sci*. 2021;22. <https://doi.org/10.3390/ijms22211989>.
25. Gao S, et al. A high-quality reference genome of wild Cannabis sativa. *Hortic Res*. 2020;7:73. <https://doi.org/10.1038/s41438-020-0295-3>.
26. Pimentel H, et al. A dynamic alternative splicing program regulates gene expression during terminal erythropoiesis. *Nucleic Acids Res*. 2014;42:4031–42. <https://doi.org/10.1093/nar/gkt1388>.
27. Caruso M, et al. Comparative transcriptome analysis of Stylar Canal cells identifies novel candidate genes implicated in the self-incompatibility response of Citrus Clementina. *BMC Plant Biol*. 2012;12:20. <https://doi.org/10.1186/1471-2229-12-20>.
28. Le KQ, Prabhakar BS, Hong WJ, Li LC. Alternative splicing as a biomarker and potential target for drug discovery. *Acta Pharmacol Sin*. 2015;36:1212–8. <https://doi.org/10.1038/aps.2015.43>.
29. Greiner S, et al. Chloroplast nucleoids are highly dynamic in ploidy, number, and structure during angiosperm leaf development. *Plant J*. 2020;102:730–46. <https://doi.org/10.1111/tpj.14658>.
30. Borrego EJ, Robertson M, Taylor J, Schultzhaut Z, Espinoza EM. Oxylin biosynthetic gene families of Cannabis sativa. *PLoS ONE*. 2023;18:e0272893. <https://doi.org/10.1371/journal.pone.0272893>.
31. Conneely LJ, Hurgobin B, Ng S, Tamiru-Oli M, Lewsey MG. Characterization of the Cannabis sativa glandular trichome epigenome. *BMC Plant Biol*. 2024;24:1075. <https://doi.org/10.1186/s12870-024-05787-x>.
32. Bhat SA, Najjar MA, Wani AA, Qadir S, John R. The Long-noncoding RNAs: effective players in plant development and stress responses. *J Plant Biochem Biotechnol*. 2024. <https://doi.org/10.1007/s13562-024-00923-y>.
33. Zhao Z, Yang Y, Iqbal A, Wu Q, Zhou L. Biological insights and recent advances in plant long Non-Coding RNA. *Int J Mol Sci*. 2024;25. <https://doi.org/10.3390/ijms252211964>.
34. Yu LX, Zheng P, Bhamidimarri S, Liu XP, Main D. The impact of Genotyping-by-Sequencing pipelines on SNP discovery and identification of markers associated with verticillium wilt resistance in autotetraploid alfalfa (Medicago sativa L). *Front Plant Sci*. 2017;8:89. <https://doi.org/10.3389/fpls.2017.00089>.
35. Pootakham W, et al. Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). *Genomics*. 2015;105:288–95. <https://doi.org/10.1016/j.ygeno.2015.02.002>.
36. Dissanayaka DMSB, Ghahremani M, Siebers M, Wasaki J, Plaxton WC. Recent insights into the metabolic adaptations of phosphorus-deprived plants. *J Exp Bot*. 2020;72:199–223. <https://doi.org/10.1093/jxb/eraa482>.
37. Park SH, et al. Effects of short-term environmental stresses on the onset of cannabinoid production in young immature flowers of industrial hemp (Cannabis sativa L). *J Cannabis Res*. 2022;4:1. <https://doi.org/10.1186/s42238-021-00111-y>.
38. Campbell BJ, Berrada AF, Hudalla C, Amaducci S, McKay JK. Genotype × environment interactions of industrial hemp cultivars highlight diverse responses to environmental factors. *Agrosystems Geosci Environ*. 2019;2:1–11. <https://doi.org/10.2134/age2018.11.0057>.
39. Caplan D, Dixon M, Zheng Y. Increasing inflorescence dry weight and cannabinoid content in medical Cannabis using controlled drought stress. *HortScience*. 2019;54:964–9. <https://doi.org/10.21273/hortsci13510-18>.
40. Danziger N, Bernstein N. Light matters: effect of light spectra on cannabinoid profile and plant development of medical cannabis (Cannabis sativa L). *Ind Crops Prod*. 2021;164:113351. <https://doi.org/10.1016/j.indcrop.2021.113351>.
41. Cai S, et al. CannabisGDB: a comprehensive genomic database for Cannabis Sativa L. *Plant Biotechnol J*. 2021;19:857–9. <https://doi.org/10.1111/pbi.13548>.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.