Integrative Metabolomic and Transcriptomic Analysis Reveals CYP450-Driven Biosynthesis of Pentacyclic Triterpene Acids in Cecropia angustifolia

Juan Esteban Vivas
Icesi University: Universidad Icesi
Paula Catherine Muñoz
Icesi University: Universidad Icesi
Laura Rodriguez
Icesi University: Universidad Icesi
Adrián Camilo Rodriguez
Icesi University: Universidad Icesi
Guillermo Leon Montoya Pelaez
glmontoya@icesi.edu.co

Icesi University: Universidad Icesi https://orcid.org/0000-0003-3963-9007

Research Article

Keywords: Cecropia angustifolia, Pentacyclic triterpene acids, Cytochrome P450, Metabolomics, Transcriptome analysis

Posted Date: August 22nd, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7197317/v1

Version of Record: A version of this preprint was published at Plant Cell, Tissue and Organ Culture (PCTOC) on December 11th, 2025. See the published version at https://doi.org/10.1007/s11240-025-03285-y.

Abstract

Pentacyclic triterpene acids (PTAs) are specialized plant metabolites known for their anti-inflammatory and antidiabetic preclinical reports. *Cecropia angustifolia*, a prominent species in the Andean montane forests, produces a variety of PTAs in its roots, but little is known about enzymes responsible for their oxidative structural decorations. In this study, we explored the biosynthetic route of PTAs oxidation by cytochrome P450 (P450s) enzymes, comparing wild-collected and in-vitro cultivated root tissues. Chemical profiling was carried out using targeted metabolomics to assess PTAs in both types of extracts. Oxidation patterns differed between the samples, suggesting distinct enzyme activity depending on growth conditions. To support these findings, we assembled a de novo transcriptome based on publicly available Cecropia RNA-seq data and predicted over 26,000 open reading frames. Four P450 families (CYP716A, CYP716C, CYP72A, and CYP71D) are potential contributors to triterpene oxidation, which were identified by homology-based annotation. Gene expression was examined by RT-qPCR, revealing differential transcriptional activity between wild and in-vitro roots. Nevertheless, these differences link the gene expression profile of P450s involved in this pathway to the TPAs metabolic expression. Notably, despite changes in expression, the capacity to generate oxidized PTAs was retained under both conditions. Our results indicate that the oxidative metabolism of triterpenes in *C. angustifolia* is largely conserved and responsive to environmental cues. This study enhances the current understanding of PTA biosynthesis and supports the use of in-vitro systems as a practical strategy to obtain valuable metabolites while minimizing ecological impact.

Key Message

Oxidative biosynthesis of triterpene acids in *Cecropia angustifolia* is conserved across growth conditions, supporting *in-vitro* systems as sustainable platforms for producing bioactive metabolites via CYP450 pathways.

Introduction

Cecropia angustifolia is a dominant tree species in the montane ecosystems of the South American Andes, recognized for its ability to synthesize pentacyclic triterpene acids (PTAs), a class of bioactive compounds with significant therapeutic potential (Sheng and Sun, 2011). These secondary metabolites, derived from oleanane and ursane skeletons, have been extensively studied for their anti-inflammatory and antidiabetic properties (Gutiérrez et al., 2020; Montoya Peláez et al., 2013; Mosquera et al., 2018). Understanding the biosynthetic pathways leading to PTA formation in *C. angustifolia* is essential to unravel the molecular mechanisms underlying their production and to develop sustainable strategies for their utilization.

Cytochrome P450 enzymes (CYP450s) play a pivotal role in the oxidative modification of triterpenoid scaffolds, facilitating the structural diversification of PTAs through site-specific oxidation at key carbon positions, such as C2, C20, C23, and C28. These modifications significantly influence the biological

activity and solubility of PTAs, affecting their pharmacological potential. Although previous studies have identified CYP450 families such as CYP716A, CYP716C, CYP72A, and CYP71D as key enzymes in triterpene oxidation in various plant species, their specific function in *C. angustifolia* remains largely unexplored.

A comprehensive understanding of these enzymatic transformations is crucial for developing biotechnological approaches aimed at sustainable production of PTAs. *In-vitro* culture techniques have emerged as a viable alternative for the controlled production of bioactive metabolites, offering a sustainable solution that reduces pressure on wild populations (Murthy et al., 2014). However, the extent to which *in-vitro* conditions influence CYP450 expression and metabolic fluxes in *C. angustifolia* remains unknown. This study integrates targeted metabolomic and transcriptomic approaches to elucidate the biosynthetic pathways of PTAs in *C. angustifolia* by analyzing both wild and *in-vitro* cultivated root materials. Using de novo transcriptome assembly, identification of homologous CYP450 genes, and gene expression analysis, we aim to correlate enzymatic activity with the structural diversity of PTAs detected through mass spectrometry. This investigation provides novel insights into the enzymatic modifications of PTAs and establishes a foundation for the sustainable biotechnological exploitation of these valuable bioactive compounds.

Materials and Methods

Herbal material and condition for in vitro culture

Roots of *Cecropia angustifolia* were collected in July 2021 from Pance, Valle del Cauca, Colombia (geographic coordinates: 3°20'12.0"N, 76°38'41.6"W; altitude: 1,771 m). The plant material was authenticated by comparison with voucher specimen OD2126 deposited at the herbarium of Universidad Icesi. Access to genetic resources was granted under Agreement No. 180 between Universidad Icesi and the Colombian Ministry of Environment (Addendum No. 10), issued on May 7, 2018. All procedures complied with the Nagoya Protocol.

Young leaves from a wild *Cecropia angustifolia* plant were collected by performing a transverse cut at the leaf adjacent to the shoot apical meristem. The plant material was subjected to a surface disinfection protocol based on the method described by Silva et al., (2015), with slight modifications. Initially, leaves were washed with Extran® MA02 Neutral soap to remove superficial debris. Subsequently, under a horizontal laminar flow cabinet, the samples were immersed in 70% ethanol for 1 minute, followed by two rinses with sterile Type II water. The material was then submerged in a solution containing 2% sodium hypochlorite, 2% Tego 51®, and Tween 20 (4 drops per 100 mL) for at least 5 minutes. This was followed by three rinses with distilled water.

After surface disinfection, 5 mm × 5 mm explants were cultured in solid Murashige and Skoog medium (Cat. No. M519, Phytotech Labs) supplemented with 30 g/L sucrose and 3 g/L CultureGel™ Type-1 gelling agent (Cat. No. G434, Phytotech Labs). To induce rhizogenesis, the medium was enriched with

10.0 mmol indole-3-butyric acid (IBA) (Cat. No. I538, Phytotech Labs), as proposed by Vasquez-Delgado et al., (2023). Cultures were maintained in darkness at 25 ± 2°C. Medium renewal and root propagation were performed every three months until the material was used in subsequent methodological steps.

PTAs fraction preparation

The herbal wild or *in-vitro* material was dried for three consecutive days at 40°C and then shredded using a blade mill. Extraction was carried out using a solid-liquid extractor at bench scale with ethyl acetate (EA) and dichloromethane (DCM) at a 1:10 ratio of herbal material to solvent. The mixture was stirred at 90 rpm at room temperature for 72 hours. The residue was removed by filtration, and the solvent was recovered and evaporated to dryness.

De novo transcriptome assembly

De novo transcriptome from *Cecropia obtusifolia* in Sequence Read Archive (SRA), Id number: SRX7344830 (Cadena-Zamudio et al., 2020) was used to identify mRNAs related to PT's enzymes. FastQC and CutAdapt (Martin, 2011) were used to determine data quality and to remove sequence adapters, respectively. Quality pair-end reads filter was performed by Trimmomatic (Bolger et al., 2014). The parameters used in this tool were SLIDINGWINDOW: 4:20, MINLEN: 50, LEADING: 20. De novo transcriptome assembly method Trinity (Grabherr et al., 2011) was used with default settings (Minimum Contig Length: 200 and K-mer: 7). Contings from Pfam-A.hmm database ("InterPro Pfam-A data," n.d.), proteins database from Urticaceae family in Uniprot (2nd november 2022) and the assembly result using Trinity were used to get ORFs (Open reading Frames) from protein enzyme targets in Transdecoder (Haas et al., 2013).

Annotation and classification of ORFs

ORFs were used for BLAST searches and annotations against NCBI protein database using an E-value cut off 10^{-10} (E-value < 0.000000000010) and an identity percentage greater than 60%. Results were filtered by length, considering only those with a length greater than 100bp for the study. The annotated ORFs were aligned with JalView (Waterhouse et al., 2009) with ClustalW under default parameters. The multiple alignments were used to identify conserved functional domains, determine the homology of protein families, and primers design.

Homologues genes identification and amplification of functional domains

Primers' design

Functional domains sequences were used to design primers. Parameters used were melting temperature (Tm): 65 – 55°C, PCR product size: 150–350 bp, GC content: 50–65%, hetero-dimer energy >-10.0 kcal/mol, homo-dimer energy >-10.0 kcal/mol, hairpin < -10, primer size 17–25 bp. The primers were synthesized and purified by HPLC (**Table S1**). Physicochemical properties of designed primers were verified using OligoAnalyzer™ Tool.

RNA isolation and reverse transcription

Cecropia angustifolia samples, both *in-vitro* and wild type, were processed inside a flow chamber to remove surface moisture using sterile paper towels for each sample. For wild-type samples, DEPC-treated water was used to clean the collected secondary roots. Afterward, the samples were preserved in liquid nitrogen for further use.

The samples were macerated with liquid nitrogen. RNA RNeasy® Power Plant® Kit (50) (Cat No. 13500-50, QIAGEN) was used for RNA isolation using a manufacturing protocol. NanoDrop™ 2000 (Thermo Fisher Sicentific) was used to evaluate RNA concentration and absorbance relation 260/230 y 260/280.

Taking the RNA isolated, the cDNA was synthesized thought RT-PCR using M-MuLV RT (Cat No. M0253L, New England Biolabs) following manufacture recommendations with some modifications: RT-PCR mix with $d(T)_{18}$ at 2.5µM, dNTPs 0.25µM, 10X M-MuLV Reaction Buffer at 0.5X, M-MuLV RT (200 UI/µL) at 5 UI/µL, and RNA template at 1000 ng in 40 uL final volume. cDNA was quantified using a NanoDrop™ 2000 (Thermo Fisher Sicentific) evaluating 260/230 y 260/280 absorbance relation. cDNA synthesis was carried out with the following cycle program: 65°C for 5 min, 42°C for 80 min, and an enzyme inactivation step of 65°C for 20 min.

Conventional PCR was performed using cDNA as a template to evaluate the primer design for target sequences. PCR mix with MgCl at 1.25 mM, dNTPs at 0.2 mM, primers at 0.5 mM, Green Taq® DNA Polymerase (Cat No. E00043, GenScript) at 5UI and template at 300 ng at 50 µl final volume. PCR amplification was carried out with a thermal cycler (C1000 Touch Thermal Cycler, Bio-Rad) with the following cycle program: 95°C for 3 min, 40 cycles of 95°C for 1 min, 50–60°C for 1 min, 72°C for 30 s, and a final extension step of 72°C for 5 min. Amplifications were visualized on a 2% agarose gel electrophoresis stained with HydraGreen™. A GeneRuler™ Low Molecular Weight DNA Ladder (Cat. No. N3233S, New England Biolabs) was used as a molecular size marker (**Fig. S5)**.

PCR products sequencing

Amplification products corresponding to the expected molecular weight were sequenced to confirm their molecular identity. Single-strand Sanger sequencing was performed using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Cat. No. 4337455, Applied Biosystems™) on a 3500 Genetic Analyzer (Thermo Fisher Scientific) following the manufacturer's instructions in Laboratorio de Medicina Genómica - Universidad Icesi. Primers in **Table S1** were used in PCR products sequencing.

Relative quantitative expression by RT-qPCR

QuantiNova™ SYBR Green One step RT-PCR Kit (Cat No. 210212, QIAGEN) was used to evaluate mRNA targets expression. For wild type and *in-vitro* samples, 4 ng/µl was used for each one. Tubulin and EF1a1 were the housekeeping genes used. One step RT-PCR amplification was carried out with a real-time thermal cycler (CFX96 Touch Real-Time PCR Detection System, Bio-Rad) with the following cycle program: 50°C for 10 min, 95°C for 2 min, 2 cycles. 40 cycles of 95°C for 5 s, 60°C for 10 s. $2^{-\Delta Ct}$ method was used to quantify targets expression.

Total Pentacyclic triterpene content and profile confirmation

The production of Serjanic acid reported by our lab in *Cecropia telenitida*, gives us the possibility to use it as internal standard $[M-H]^-$ *m/z* 499 to control the production for quantitative expression, since this PT is not produced by *Cecropia angustifolia*. To simplify the analysis, and considering the fraction contains many isobaric triterpenes, it was decided to select the mass to charge ratio $[M-H]^-$ *m/z* 487 represented by the most oxidized molecules like yarumic, isoyarumic, arjunolic or tormentic acids. $[M-H]^-$ *m/z* 471 represented by molecules like hederagenic and 20-hydroxy-ursolic, maslinic, and corosolic acids. The precursors of the mentioned molecules come from less oxidized oleanyl and ursanyl cations $[M-H]^-$ *m/z* 455 represented by oleanolic and ursolic acids.

Analysis and quantitation were carried out in a UPLC-SQD2 mass spectrometry Waters®. Chromatographic conditions were established previously defining a method for multiple metabolites separation. The chromatographic separation of triterpene acids was carried out on Acquity UPLC® BEH C18 (100 x 2.1 mm, 1.7 µm) and the column was kept at 35 ºC. The mobile phase was Mili-Q water with 5 mM of ammonium acetate (A) and methanol hypergrade (B) at constant flow rate of 0.3 mL min^{-1} with a gradient elution of 0 min at 30% A, 1 min at 20% A, 1.50 min 15% A, 3 min 5% A, 4.50 min 2% A, 6.00 min 2% A, 10 min 30%A.

The mass spectrometer uses an electrospray ionization (ESI) source, kept at 500 ºC and acquired in negative mode. Nitrogen was used as a nebulizer and curtain gas. The optimal values for MS parameters were capillary voltage 3,50 kV, cone voltage 80 V; desolvation temperature 500 ºC and desolvation gas source flux 900 L/hr. The analyzer was operated in single ion recording (SIR) mode, which monitored $[M-H]^-$ ions at *m/z* of 487, 455 and 471, and additionally Serjanic acid as the IS at *m/z* 499. Masslynx MS Software was used for instrument control, data acquisition, as well as data processing.

Statistical analysis

ANOVA test and Tukey multiple comparisons analysis were performed. Kolmogorov-smirnoff (a non-parametric test) was used to evaluate the statistical difference between samples of a single variable with a p-value of < 0.05. GraphPad Prisma version 9.5.1 for Windows was used for graphics.

Results and discussion

Pentacyclic triterpene scaffold modifications by P450s

A comprehensive review of relevant literature was conducted to reconstruct the metabolic pathways involved in the biosynthesis of the molecules analyzed in this study. The review focused on enzymatic reactions that could account for the transformations of less oxidized triterpenes (ursolic and oleanolic acid at *m/z* 455) in carbons C28, C2α, C20, and C23. The identification of the enzymes responsible for these transformations was based on reports from various authors, including (Ghosh, 2017), (Miettinen et al., 2017), (Tamura et al., 2017), (Han et al., 2018), (Yasumoto et al., 2017), (Nakamura et al., 2018),(Prall et al., 2016), (Sandeep et al., 2019), (Li et al., 2022), (Dinday and Ghosh, 2023), and (Yu et al., 2019). Based on the information compiled from these studies, the enzymes CYP716A, CYP716C, CYP72A, and CYP71D were selected for further analysis (Fig. 1). It is noteworthy that the enzymatic mechanisms responsible for the oxidation of 20-hydroxyursolic acid and isoyarumic acid remain unidentified. Nevertheless, enzymes that could catalyze reactions to these carbons were included, as reported in the literature.

De Novo Transcriptome assembly and species selection

Based on previously gathered information, species were selected according to the availability of genetic data in public databases, specifically coding sequences for enzymes of interest, and their taxonomic proximity to *Cecropia angustifolia*. As shown in Fig. 1, approximately 13 species were identified as containing at least one of the four targeted P450 enzymes. These species were distributed across three taxonomic orders: Fabales, Fagales, and Rosales, with the latter encompassing the Cecropia genus. The inclusion of species from three different taxonomic orders introduces notable variability into the analysis. For certain enzymes, such as CYP71D, available data in the NCBI database were restricted to species within the order Fabales; therefore, excluding this order would impede the investigation of homologous enzymes in Cecropia. Additionally, enzymes were reported in species such as Lagerstroemia speciosa, Arabidopsis thaliana, and Beta vulgaris. However, as these species belong to the subclass Malvidae, they were excluded from this study. This review facilitated the creation of a custom database, which was subsequently used for functional annotation via BLAST analysis.

The transcriptome hosted in the SRA database under accession number SRX7344830 was assembled, comprising 19,043,380 sequences with a Phred quality score of 30 or higher. Using these reads and applying a k-mer value of 7, 21,330 isotigs were obtained, with an N50 isotig length of 1,500 bp and an average isotig length of 1,088.67 bp. The number of isotigs contributing to the N50 was 5,227. Statistics results from TrinityStats are shown in Table 1. These metrics indicate a high likelihood of identifying the genes of interest, as their coding sequences (CDS) are approximately 1,400 bp in length (Ghosh, 2017). Notably, the large number of isotigs generated may be attributed to the selection of a low k-mer value; typically, a k-mer of 25 is used for de novo transcriptome assemblies. However, in this study, a lower k-mer value was chosen to enhance the representation of genes involved in secondary metabolism, which

are often expressed at low levels and may exhibit limited coverage (Durai and Schulz, 2016). Furthermore, a total of 26,708 open reading frames (ORFs) were identified from these sequences using the TransDecoder tool. For TransDecoder analysis, an Urticaceae-specific protein database retrieved from UniProt was used (“Urticaceae | Taxonomy | UniProt,” n.d.).

Table 1
Statistics for *De Novo* RNA assembly

Statistics for isotig length	
Minimum isotig length	190 bp
Maximum isotig length	14473 bp
Average isotig length	1088.67 bp
Standard deviation of isotig length	825.82 bp
Median isotig length	900 bp
Isotig length N50	1500 bp
Statistics for number of isotigs	
Number of isotigs	21330
Number of isotigs > = 1kb	9630
Number of isotigs at N50	5227
Statistics for the number of bases in isotigs	
Total number of bases in all isotigs	23221397
Number of bases in isotigs > = 1kb	17034607

ORFs obtained through TransDecoder were used for homology annotation via BLASTp. As shown in Table 2, the analysis yielded a set of isotigs that met the established filtering parameters. Amino acid sequences from the CYP71D and CYP72A enzyme families exhibited lower identity percentages compared to those from the CYP716C and CYP716A families. This discrepancy can be attributed to the high structural conservation of functional domains within this enzyme group across species, coupled with substantial diversification in the overall protein structure. These structural variations are primarily influenced by the taxonomic order of the species, as illustrated in Fig. 1 (Prall et al., 2016).

Table 2
Results of homology annotation of P450s using BlastP.

Contig ID	Sequence count	%Identity	E-value	Alignment Length
CYP716A				
DN120_c0_g1_i1_p1	5	80.8%	6.54e^-109	477 aa
DN120_c0_g1_i2_p1	5	87.2%	<e^-250	164 aa
DN120_c0_g1_i3_p1	5	77.7%	<e^-250	484 aa
CYP716C				
DN265_c0_g1_i1_p1	14	82.0%	<e^-250	480 aa
DN265_c0_g1_i2_p1	14	82.0%	<e^-250	480 aa
DN265_c0_g1_i3_p1	14	82.0%	<e^-250	480 aa
DN265_c0_g1_i4_p1	14	82.0%	<e^-250	480 aa
CYP72A				
DN5056_c0_g1_i1_p1	8	68.1%	<e^-250	523 aa
DN350_c0_g1_i1_p1	4	65.5%	<e^-250	186 aa
CYP71D				
DN5854_c0_g1_i1_p1	2	61.4%	3.99e^-88	197 aa

Structural description of annotated proteins through multiple alignments.

To verify the conservation of the functional domains in the annotated target ORFs, multiple sequence alignments were performed using these sequences and comparing their conservation with reference proteins (**Fig. S1-S4**). This verification is necessary to ensure that ORFs conservation occurs within the characteristic functional domains of this type of protein; otherwise, the functional annotation could yield false positives due to non-specific annotations (Prall et al., 2016).

Figure 2 presents the alignment results for the functional domains of the CYP716C, CYP716A, CYP71D, and CYP72A proteins. Conserved regions were identified in key domains, including the oxygen activation site, the characteristic E-X-X-R and P-E-R-F triads of cytochrome P450 enzymes, and the oxygen-binding region. These findings suggest that homology-based functional annotation was primarily guided by sequence conservation within these functional domains, providing evidence for the accuracy of the enzyme sequence annotation. While significant similarity was observed between the annotated sequence clusters for the same protein and their reference sequences, notable differences were evident within the functional domains across the different protein groups (Ghosh, 2017). For instance, in the oxygen-binding domain with the conserved peptide sequence P-F-X-X-G-X-R-X-C-X-G, the structure is preserved across all P450s analyzed, yet specific sequence variations were observed: P-F-G-G-G-P-R-M-

C-P-G for CYP716C and CYP716A; P-F-G-A-G-K/R-R-I/M-C-P-G for CYP71D; and P-F-G-W/G-G-P-R-I-C-X-G for CYP72A (WANG et al., 2015). These sequence differences define the identity of these proteins and were utilized as the basis for primer design, leveraging their respective coding sequences.

As shown in Fig. 2, structural similarity between the functional domains of CYP716C and CYP716A proteins may be due to the diversification of both families from a common ancestor (Ghosh, 2017). Moreover, their enzymatic functionality is similar, as various studies suggest they can recognize similar substrates and catalyze the same reactions, which may vary among individuals, such as oxidations at carbon C2α or C28 (Pu et al., 2023). Therefore, the primer design approach used sequences adjacent to the functional domains to enable their amplification while ensuring specificity for each protein family.

Genetic and metabolic profiling of Cecropia angustifolia in in-vitro and in-vivo cultures.

Once the primers were designed, the target genes were amplified to verify specificity and to confirm the presence or absence of nonspecific amplifications (**Fig. S5**). These results are condensed in the oxidation pathway propose in Fig. 3A and 3B. Additionally, the identity of the PCR fragments obtained was determined through paired-end Sanger sequencing. It is important to note that amplification of the gene corresponding to the CYP716A enzyme was not successful; therefore, it was not included in subsequent analyses. Once the amplification conditions and identity were verified for all primer pairs, the quantification of target gene expression was carried out. Relative expression levels were determined using tubulin (TUB) and elongation factor 1-alpha (EF1α) as housekeeping genes (HK).

Metabolic expression analyses in *Cecropia angustifolia* are consistent with the previous report, a significant difference in TPA production between *in-vitro* and wild-type individuals (Fig. 4A). In fact, the results agree with those reported by Vásquez Delgado (Vasquez-Delgado et al., 2023) as the wild-type TPA sample showed a higher proportion of molecules with a greater number of oxidations (ions with an *m/z* ratio of 487) compared to the *in-vitro* sample. In contrast, the *in-vitro* sample exhibited a higher accumulation of molecules with fewer oxidations, namely ions with *m/z* ratios of 471 and 455 (Fig. 3A and 3B). Furthermore, similar results were observed in another study involving a species of the same genus, *Cecropia telenitida*, where mass spectrometry identified a predominant signal corresponding to ions with an *m/z* ratio of 487 (Gutiérrez et al., 2021), further confirming the similarity in TPA composition among wild-type individuals.

Target mRNA quantification was performed using the HK genes as a reference. As shown in Fig. 4B, the samples exhibit distinct expression profiles, with all evaluated mRNAs showing statistically significant differences between them. In the *in-vitro* samples, a notably higher expression of CYP71D enzymes was observed. This result is consistent with previous reports, as CYP71D is known to catalyze the conversion of molecules with an *m/z* ratio of 455 to 471, which may explain the increased accumulation of these intermediates (Krokida et al., 2013). Furthermore, the *in-vitro* samples showed reduced expression of CYP72A and CYP716C, enzymes involved in the biosynthesis of compounds with an *m/z* ratio of 487. This downregulation likely contributes to the accumulation of *m/z* 471 compounds, such as corosolic

acid, maslinic acid, or hederagenic acid, and to the absence of further oxidation of these triterpenoid cores.

In contrast, the wild-type samples exhibited higher expression levels of enzymes from the CYP72A family, which are responsible for the biosynthesis of molecules with *m/z* ratios of both 471 and 487. This correlates with the results shown in Fig. 4A, where wild-type samples displayed greater accumulation of these compounds. Similarly, the expression of CYP716C may also contribute to this metabolic diversification. However, among the mRNAs analyzed, CYP72A appears to be the main contributor to the production of molecules such as arjunolic acid, asiatic acid, and isoyarumic acid. It is important to note that, since the transcripts encoding CYP716A could not be quantified, it is not possible to discuss the formation of the precursors in this biosynthetic pathway, leaving this aspect currently unknown for *Cecropia angustifolia* (Dinday and Ghosh, 2023).

Figure 4C-D show the chromatograms obtained from wild-type and *in-vitro* extracts using SIR (Single Ion Recording) acquisition mode. As exemplified, differences in expression profiles not only influence the relative abundance of these molecules in both types but also affect the overall number of detectable compounds in the extracts. For instance, in the case of ions with an *m/z* ratio of 471, approximately four distinct chemical species are detected in the wild-type samples, whereas only two peaks are observed in the *in-vitro* samples. This indicates a greater metabolite diversity in wild-type extracts. These differences in chromatographic profiles do not necessarily imply the complete absence of certain compounds in one condition; rather, some metabolites may be produced at levels below the detection threshold. A similar pattern can be observed for the compounds with an *m/z* ratio of 487, where at least three distinct chromatographic peaks are detected in the *in-vitro* samples, each corresponding to different molecular species (positional isomers). In contrast, wild-type samples display an even greater number of distinct compounds within this *m/z* range, further supporting the higher metabolic diversity observed in wild-type systems.

The metabolic expression analyses of *Cecropia angustifolia* revealed significant differences in TPA oxidation patterns between in-vitro and wild-type individuals (Fig. 4C and 4D), where the latter accumulated a higher proportion of highly oxidized molecules (*m/z* 487) compared to the in-vitro samples, which showed greater abundance of less oxidized forms (*m/z* 471 and 455). This aligns with recent findings from comparative metabolic profiling of wild and in-vitro cultivated *C. angustifolia* roots, where wild samples consistently exhibited higher proportions of oxidized pentacyclic triterpene acids, suggesting conservation of TPA biosynthesis despite cultivation conditions (Díaz-Sánchez et al., 2025). Similar oxidation processes have been attributed to the activity of CYP716A family enzymes, which catalyze sequential oxidations at the C-28 position of pentacyclic triterpenes, as demonstrated in *Medicago truncatula* for β-amyrin to oleanolic acid conversion (Matsuba et al., 2013) and in *Solanum lycopersicum* for α and β amyrins transformation (Yasumoto et al., 2017b). The reduced expression of such oxidizing CYP enzymes in in-vitro cultures could explain the lower oxidation states observed, aligning with reports where gene silencing or lack of CYP expression limits downstream oxidation reactions (Chiyo et al., 2024; Thorat et al., 2022). Therefore, these results support the hypothesis that

oxidation-mediated diversification of TPAs in Cecropia is strongly dependent on the differential expression of CYP genes, which may be repressed under in-vitro conditions, ultimately affecting the biosynthetic flux towards highly oxidized triterpenes.

Beyond the epigenetic changes evaluated in this study, it is plausible that the expression of additional enzymes, either isoforms of those already analyzed or entirely distinct catalytic proteins may contribute to structural diversification by targeting specific carbon positions within the molecular scaffold. In particular, cytochrome P450 enzymes (CYPs) represent a diverse and functionally versatile group that could mediate a broad range of oxidative modifications. For instance, CYP71E family members catalyze oxidation at the 6β position; CYP51H and CYP716S are involved in the epoxidation of the C12–C13 double bond in oleanane-type triterpenes; CYP716A enzymes catalyze C-28 carboxylation, C3 ketone formation, and C16 oxidation; while CYP714E and various CYP72A isoforms mediate oxidations at C23, C21, C22, and C16 (Han et al., 2018; Kim et al., 2018). Although the current study did not explore these potential enzymatic players or the epigenetic mechanisms regulating their expression, future investigations integrating transcriptomic, epigenomic, and functional analyses of alternative P450s could yield a more comprehensive understanding of the biosynthetic complexity underlying this pathway.

Author contribution statement

Juan Esteban Vivas: Investigation, Formal analysis, Data curation, and writing the original draft. **Paula Muñoz**: Formal analysis, Data curation. **Laura Rodriguez**: Conceptualization, Data curation. **Adrian Rodriguez**: Conceptualization, Data curation. **Guillermo Montoya**: Conceptualization, writing the original draft, and Funding acquisition.

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information Data Set files. Should any raw data files be needed in another format they are available from the corresponding author upon reasonable request. Source data are provided with this paper.

Declarations

Author contribution statement

Juan Esteban Vivas: Investigation, Formal analysis, Data curation, and writing the original draft. **Paula Muñoz**: Formal analysis, Data curation. **Laura Rodriguez**: Conceptualization, Data curation. **Adrian Rodriguez**: Conceptualization, Data curation. **Guillermo Montoya**: Conceptualization, writing the original draft, and Funding acquisition.

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information Data Set files. Should any raw data files be needed in another format they are available from the corresponding author upon reasonable request. Source data are provided with this paper.

Funding

Ministry of Science, Technology, and Innovation of Colombian Government. Grants 430-2020 code number 211784467730, and 435-2021. Universidad Icesi grant number CA04130107.

Competing Interests

The authors declare that there is no conflict of interest concerning the publication of this article and have no known competing financial interests or personal relationships that could have appeared to influence the work reported.

References

1. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
2. Cadena-Zamudio JD, Nicasio-Torres P, Monribot-Villanueva JL, Guerrero-Analco JA, Ibarra-Laclette E (2020) Integrated analysis of the transcriptome and metabolome of Cecropia obtusifolia: A plant with high chlorogenic acid content traditionally used to treat diabetes mellitus. Int J Mol Sci 21:1–27. https://doi.org/10.3390/ijms21207572
3. Chiyo N, Seki H, Kanamoto T, Ueda H, Kojoma M, Muranaka T (2024) Glycyrrhizin Production in Licorice Hairy Roots Based on Metabolic Redirection of Triterpenoid Biosynthetic Pathway by Genome Editing. Plant Cell Physiol 65:185–198. https://doi.org/10.1093/pcp/pcad161
4. Díaz-Sánchez LM, Chacón-Patiño ML, Blanco-Tirado C, Combariza MY, Montoya G (2025) Comparative Metabolic Profiling of Wild and In Vitro Cultivated *Cecropia angustifolia* Roots Fraction Reveals Conservation of Pentacyclic Triterpene Acid Biosynthesis. https://doi.org/10.1002/pca.70001. Phytochemical Analysis
5. Dinday S, Ghosh S (2023) Recent advances in triterpenoid pathway elucidation and engineering. https://doi.org/10.1016/j.biotechadv.2023.108214. Biotechnol Adv
6. Durai DA, Schulz MH (2016) Informed kmer selection for de novo transcriptome assembly. Bioinformatics 32:1670–1677. https://doi.org/10.1093/bioinformatics/btw217
7. Ghosh S (2017) Triterpene structural diversification by plant cytochrome P450 enzymes. Front Plant Sci 8:1–15. https://doi.org/10.3389/fpls.2017.01886
8. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Di Palma F, Birren BW, Nusbaum C,

Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652. https://doi.org/10.1038/nbt.1883

9. Gutiérrez G, Giraldo-Dávila D, Combariza MY, Holzgrabe U, Tabares-Guevara JH, Ramírez-Pineda JR, Acín S, Muñoz DL, Montoya G, Balcazar N (2020) Serjanic acid improves immunometabolic markers in a diet-induced obesity mouse model. Molecules 25:1486. https://doi.org/10.3390/molecules25071486
10. Gutiérrez G, Valencia LM, Giraldo-Dávila D, Combariza MY, Galeano E, Balcazar N, Panay AJ, Jerez AM, Montoya G (2021) Pentacyclic triterpene profile and its biosynthetic pathway in Cecropia Telenitida as a prospective dietary supplement. Molecules 26:1–14. https://doi.org/10.3390/molecules26041064
11. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512. https://doi.org/10.1038/nprot.2013.084
12. Han JY, Chun JH, Oh SA, Park SB, Hwang HS, Lee H, Choi YE (2018) Transcriptomic Analysis of Kalopanax septemlobus and Characterization of KsBAS, CYP716A94 and CYP72A397 Genes Involved in Hederagenin Saponin Biosynthesis. Plant Cell Physiol 59:319–330. https://doi.org/10.1093/pcp/pcx188
13. InterPro, Pfam -A data [WWW Document], n.d. URL https://www.ebi.ac.uk/interpro/download/pfam/ (accessed 1.29.24)
14. Kim OT, Um Y, Jin ML, Kim JU, Hegebarth D, Busta L, Racovita RC, Jetter R (2018) A novel multifunctional C-23 Oxidase, CYP714E19, is involved in asiaticoside biosynthesis. Plant Cell Physiol 59:1200–1213. https://doi.org/10.1093/pcp/pcy055
15. Krokida A, Delis C, Geisler K, Garagounis C, Tsikou D, Peña-Rodríguez LM, Katsarou D, Field B, Osbourn AE, Papadopoulou KK (2013) A metabolic gene cluster in Lotus japonicus discloses novel enzyme functions and products in triterpene biosynthesis. New Phytol 200:675–690. https://doi.org/10.1111/nph.12414
16. Li Y, Wang J, Li L, Song W, Li M, Hua X, Wang Y, Yuan J, Xue Z (2022) Natural products of pentacyclic triterpenoids: from discovery to heterologous biosynthesis. Nat Prod Rep. https://doi.org/10.1039/d2np00063f
17. Matsuba Y, Nguyen TTH, Wiegert K, Falara V, Gonzales-Vigil E, Leong B, Schäfer P, Kudrna D, Wing RA, Bolger AM, Usadel B, Tissier A, Fernie AR, Barry CS, Pichersky E 2013. Evolution of a Complex Locus for Terpene Biosynthesis in *Solanum*. Plant Cell 25, 2022–2036. https://doi.org/10.1105/tpc.113.111013
18. Miettinen K, Pollier J, Buyst D, Arendt P, Csuk R, Sommerwerk S, Moses T, Mertens J, Sonawane PD, Pauwels L, Aharoni A, Martins J, Nelson DR, Goossens A (2017) The ancient CYP716 family is a

major contributor to the diversification of eudicot triterpenoid biosynthesis. Nat Commun 8. https://doi.org/10.1038/ncomms14153

19. Montoya Peláez GL, Sierra JA, Alzate F, Holzgrabe U, Ramirez-Pineda JR (2013) Pentacyclic triterpenes from Cecropia telenitida with immunomodulatory activity on dendritic cells. Brazilian J Pharmacognosy 23:754–761. https://doi.org/10.1590/S0102-695X2013000500006
20. Mosquera C, Panay A, Montoya G (2018) Pentacyclic Triterpenes from Cecropia telenitida Can Function as Inhibitors of 11β-Hydroxysteroid Dehydrogenase Type 1. Molecules 23, 1444. https://doi.org/10.3390/molecules23061444
21. Murthy HN, Lee EJ, Paek KY (2014) Production of secondary metabolites from cell and organ cultures: Strategies and approaches for biomass improvement and metabolite accumulation. Plant Cell Tissue Organ Cult 118:1–16. https://doi.org/10.1007/s11240-014-0467-7
22. Nakamura M, Linh TM, Lien LQ, Suzuki, Hayato, Mai NC, Giang VH, Tamura K, van Thanh N, Suzuki, Hideyuki, Misaki R, Muranaka T, Ban NK, Fujiyama K, Seki H (2018) Transcriptome sequencing and identification of cytochrome p450 monooxygenases involved in the biosynthesis of maslinic acid and corosolic acid in Avicennia marina. Plant Biotechnol 35:341–348. https://doi.org/10.5511/plantbiotechnology.18.0810a
23. Prall W, Hendy O, Thornton LE (2016) Utility of a phylogenetic perspective in structural analysis of CYP72A enzymes from flowering plants. PLoS ONE 11. https://doi.org/10.1371/journal.pone.0163024
24. Pu X, Chen M, Lei M, Lin X, Zhang J, Ai Z, He J, Liu Y, Yang S, Wang H, Liao J, Zhang L, Huang Q (2023) Discovery of unique CYP716C oxidase involved in pentacyclic triterpene biosynthesis from Camptotheca acuminata. Plant Physiol Biochem 202. https://doi.org/10.1016/j.plaphy.2023.107929
25. Sandeep, Misra RC, Chanotiya CS, Mukhopadhyay P, Ghosh S (2019) Oxidosqualene cyclase and CYP716 enzymes contribute to triterpene structural diversity in the medicinal tree banaba. New Phytol 222:408–424. https://doi.org/10.1111/nph.15606
26. Selwal N, Rahayu F, Herwati A, Latifah E, Supriyono, Suhara C, Suastika K, Mahayu IB, Wani WM, A.K (2023) Enhancing secondary metabolite production in plants: Exploring traditional and modern strategies. J Agric Food Res 14:100702. https://doi.org/10.1016/j.jafr.2023.100702
27. Sheng H, Sun H (2011) Synthesis, biology and clinical significance of pentacyclic triterpenes: a multi-target approach to prevention and treatment of metabolic and vascular diseases. Nat Prod Rep 28:543–593. https://doi.org/10.1039/c0np00059k
28. Silva JAT, da, Winarto B, Dobránszki J, Zeng S (2015) Disinfection procedures for in vitro propagation of Anthurium. Folia Horticulturae 27:3–14. https://doi.org/10.1515/fhort-2015-0009
29. Tamura K, Seki H, Suzuki H, Kojoma M, Saito K, Muranaka T (2017) CYP716A179 functions as a triterpene C-28 oxidase in tissue-cultured stolons of Glycyrrhiza uralensis. Plant Cell Rep 36:437–445. https://doi.org/10.1007/s00299-016-2092-x
30. Thorat SA, Kaniyassery A, Poojari P, Rangel M, Tantry S, Kiran KR, Joshi MB, Rai PS, Botha A-M, Muthusamy A (2022) Differential Gene Expression and Withanolides Biosynthesis During in vitro and

ex vitro Growth of Withania somnifera (L.) Dunal. Front Plant Sci 13. https://doi.org/10.3389/fpls.2022.917770

31. Urticaceae | Taxonomy | UniProt [WWW Document], n.d. URL https://www.uniprot.org/taxonomy/3499 (accessed 6.10.25)
32. Vasquez-Delgado JS, Vivas-Moncayo JE, Lopez-Cortes JV, Combariza MY, Montoya G (2023) Pharmacokinetic assessment and phytochemical triterpene control from Cecropia angustifolia using plant biotechnology. Phytochem Anal 34. https://doi.org/10.1002/pca.3252
33. WANG Y, HUANG LIUJ, XU B, LI Y-M, HUANG J, L.-F., LIN, J., ZHANG, J., MIN, Q.-H., YANG, W.-M., WANG, X.-Z (2015) Mechanism of alternative splicing and its regulation. Biomed Rep 3:152. https://doi.org/10.3892/BR.2014.407
34. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191. https://doi.org/10.1093/bioinformatics/btp033
35. Yasumoto S, Seki H, Shimizu Y, Fukushima EO, Muranaka T (2017a) Functional characterization of CYP716 family P450 enzymes in triterpenoid biosynthesis in tomato. Front Plant Sci 8. https://doi.org/10.3389/fpls.2017.00021
36. Yasumoto S, Seki H, Shimizu Y, Fukushima EO, Muranaka T (2017b) Functional Characterization of CYP716 Family P450 Enzymes in Triterpenoid Biosynthesis in Tomato. Front Plant Sci 8. https://doi.org/10.3389/fpls.2017.00021
37. Yu Y, Zhang G, Chen Y, Bai Q, Gao C, Zeng L, Li Z, Cheng Y, Chen J, Sun X, Guo L, Xu J, Yan Z (2019) Selection of Reference Genes for qPCR Analyses of Gene Expression in Ramie Leaves and Roots across Eleven Abiotic/Biotic Treatments. Sci Rep 9. https://doi.org/10.1038/s41598-019-56640-3

Figures

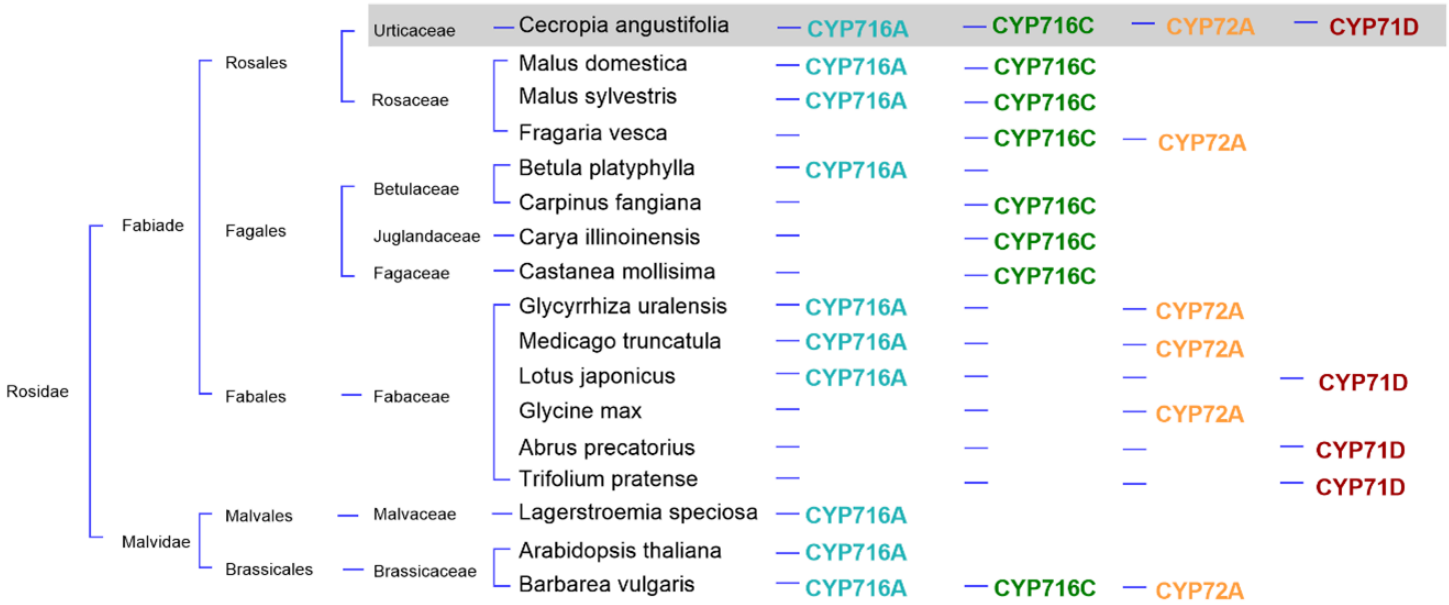


Figure 1

List of species used for data collection in metabolic pathway reconstruction.

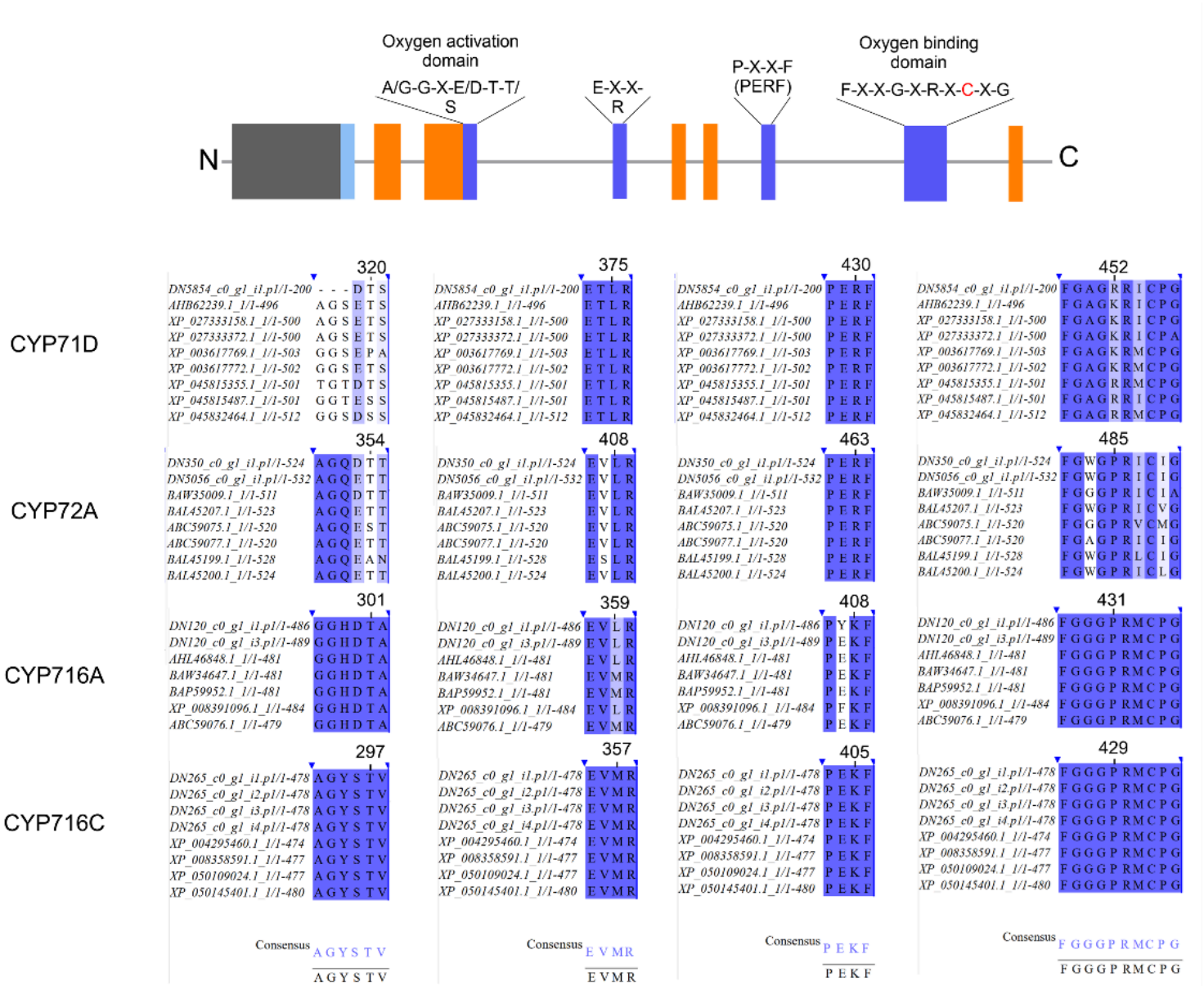
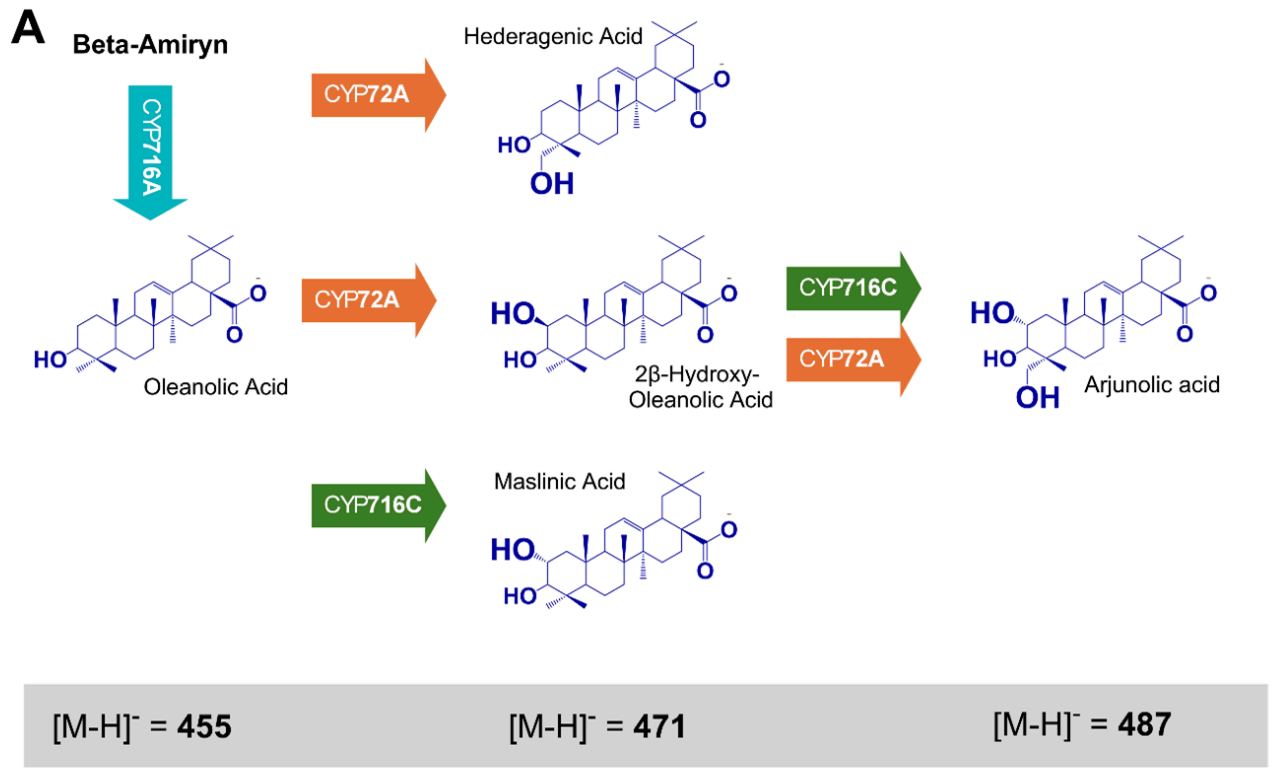


Figure 2

Alignment of the Conservation Degree of Functional Domains of P450 Enzymes. Graph generated using JalView; %Conservation >60%.



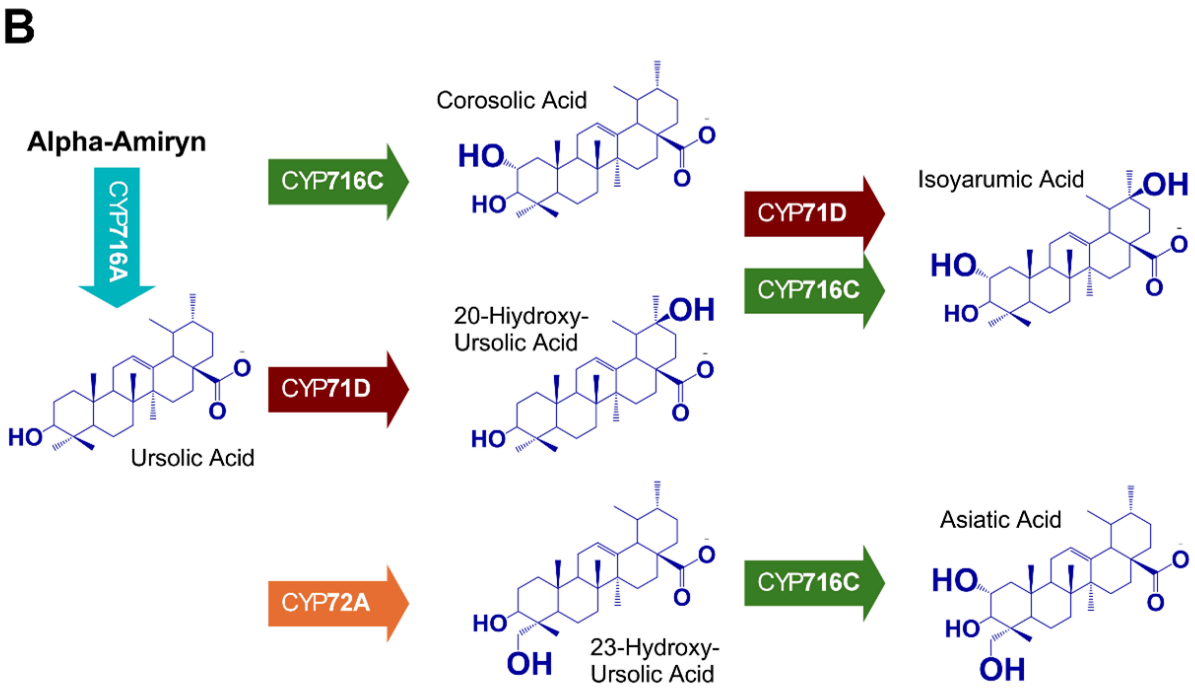


Figure 3

Metabolic pathways associated with TPAs production mediated by cytochrome P450s enzymes, (A) Modification of the oleanane core; (B) Modification of the ursane core.

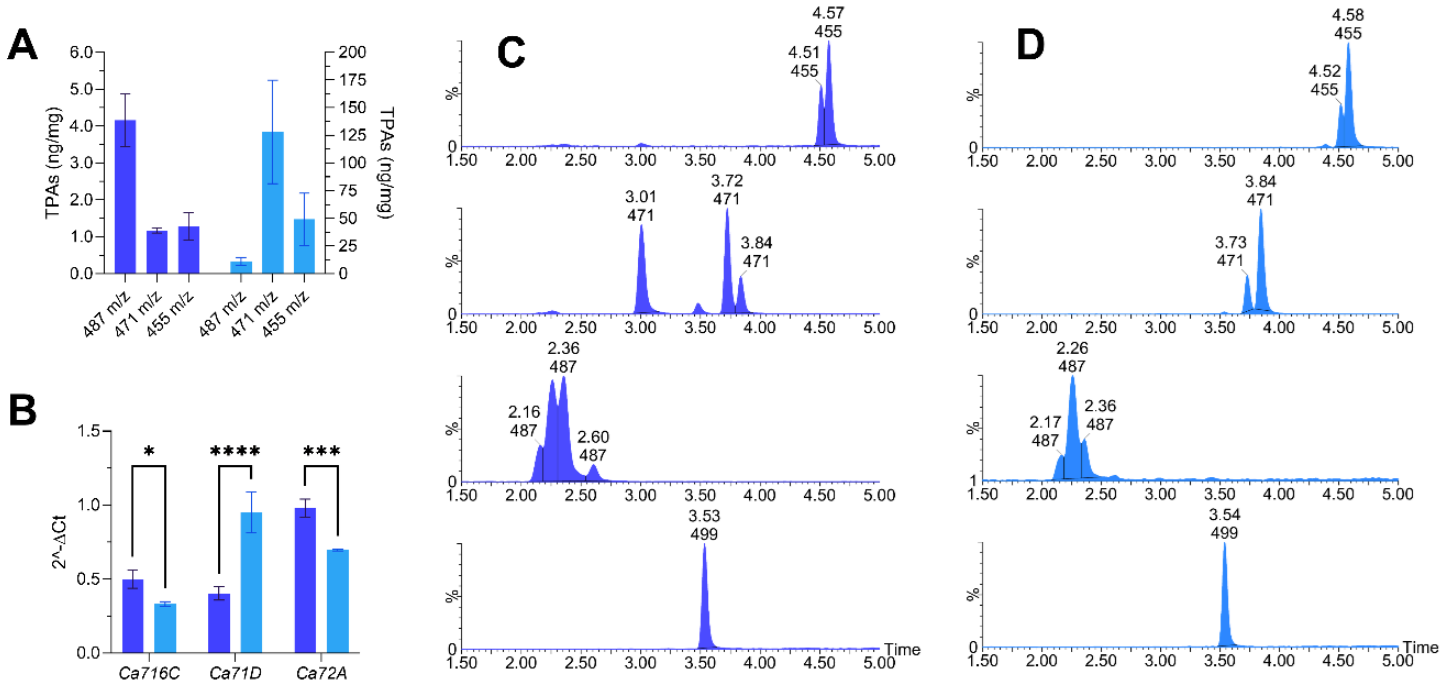


Figure 4

Comparison of gene expression and metabolic profiles between *Cecropia angustifolia in-vitro* (light blue) and wild type (dark blue) individuals. (A) TPA production levels. (B) Expression of target genes. (C and D) Comparison of chromatograms from wild and *in-vitro* tissues, respectively. Results are presented as mean ± SD. A two-way analysis of variance (ANOVA) was performed, followed by Tukey's multiple comparisons test, with a statistical significance level set at $p < 0.05$ (*: $p < 0.001$, ***: $p < 0.01$).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryinformationfile.docx
- GA.png