

Tissue-specific genome-scale metabolic models of *Erythroxylum novogranatense* reveal distinct leaf-root metabolism beyond cocaine

Johan Nicolas Anzola

Universidad de Los Andes

Diana Marcela Bernal Franco

Universidad de Los Andes

Andrés Fernando González Barrios

Universidad de Los Andes

María Francisca Villegas Torres

m.f.villegastorres@uniandes.edu.co

Universidad de Los Andes

Research Article

Keywords: Tissue-specific GEMs, *Erythroxylum novogranatense*, specialized metabolism, untargeted metabolomics, tropane alkaloids, flux balance analysis

Posted Date: June 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10012027/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background:

Erythroxylum novogranatense produces specialized metabolites in a tissue-dependent manner, but its metabolic network remains poorly described beyond cocaine biosynthesis. Here, we present, the first tissue-specific genome-scale metabolic models for leaf and root of *E. novogranatense*, reconstructed from a chromosome-level genome assembly and contextualized with untargeted LC–QTOF–MS metabolomics.

Results:

The final leaf GEM contained 1,764 reactions, 1,550 metabolites and 1,428 genes, while the root GEM contained 1,768 reactions, 1,565 metabolites and 1,426 genes. Under tissue-specific media, the models reached feasible optimal biomass objective values of 10.11 for leaf and 14.53 for root. Metabolomic profiling produced 122 tissue-level metabolite annotations (57 in leaf, 65 in root), 48 metabolites were shared between tissues, whereas 9 were leaf-specific and 17 were root-specific. Tissue-associated markers incorporated into the GEMs included umbelliferone and cinnamoylcocaine in leaf, and anisodine and palmatine in root. Alkaloids were prevalent in both tissues, while flavonoids and phenylpropanoids were relatively more represented in leaf and terpenoids in root. Constraint-based simulations under a common 90% biomass objective constraint showed that the models share a highly similar structural core but differ in active reaction usage, flux organization and production potential for selected shared metabolites.

Conclusions:

These tissue-specific GEMs provide a quantitative framework to study leaf–root metabolic differentiation in *E. novogranatense*. By integrating genome annotation, manual curation and tissue-resolved metabolomics, the models extend coca metabolism beyond cocaine and enable flux-based prediction of tissue-specific metabolic capacities and candidate specialized pathways.

Background

Plants produce a vast repertoire of specialized metabolites that mediate ecological interactions, modulate growth and development, and contribute to adaptation under biotic and abiotic stress. These compounds, traditionally referred to as secondary metabolites and now more commonly termed specialized metabolites, include alkaloids, phenylpropanoids, flavonoids, terpenoids and lipid derivatives, many of which display antioxidant, antimicrobial, anti-inflammatory, neuroactive or cytoprotective activities of interest for medicine, agriculture, food systems and biotechnology^{1,2}. In plants, the biosynthesis and accumulation of specialized metabolites are not uniformly distributed across the

organism. Instead, they are shaped by tissue and cell identity, developmental stage and physiological context, with many specialized metabolites being produced or accumulated preferentially in particular cells, tissues or organs^{3,4}.

The genus *Erythroxylum* illustrates this dual ecological and pharmacological relevance. Species within this genus synthesize specialized metabolites such as tropane alkaloids, which function as chemical defenses against herbivorous insects, pathogens, and other biotic stress factors, thus influencing ecological interactions within their native habitats. Beyond their contribution to plant fitness, tropane alkaloids have remarkable pharmacological relevance⁵. This relevance is also reflected in the traditional use of several *Erythroxylum* species, particularly *Erythroxylum coca* and *Erythroxylum novogranatense*, collectively known as coca by Indigenous communities from the Andes and Amazon regions⁵. In these contexts, coca leaves are consumed for their stimulant and medicinal properties and are embedded within complex cultural, social and ritual practices⁶. Beyond their notoriety as the botanical source of cocaine, *Erythroxylum* species synthesize a chemically rich mixture of tropane alkaloids, polyphenols, flavonoids, terpenoids and fatty-acid derivatives with reported antioxidant, antimicrobial and cytoprotective activities^{7,8}. Phytochemical studies have confirmed the bioactivity of these compounds, along with phenolic compounds and flavonoids of potential interest for pharmacological and nutraceutical applications⁸, while detailed biochemical studies have elucidated the key steps of the cocaine biosynthetic pathway⁹. However, existing research focuses on specific pathways and a limited number of metabolites, leaving unresolved questions about how different tissues allocate metabolic resources among the various compounds.

A central aspect of this problem is tissue specificity. In plants, photosynthetic organs, storage tissues, and belowground organs perform distinct physiological functions and therefore exhibit different metabolic demands and defensive priorities. Leaves typically allocate carbon toward phenylpropanoids, flavonoids, and other photoprotective metabolites, whereas roots often accumulate defense compounds targeting soil-borne herbivores and pathogens⁴. In the case of alkaloids, tissue localization of biosynthesis can vary among taxa: for example, in *Erythroxylum coca*, tropane alkaloids are synthesized and stored in aboveground green tissues, whereas in members of the *Solanaceae*, tropane alkaloids are produced primarily in root tissues and subsequently translocated throughout the plant⁵. Increasingly, omics studies demonstrate that transcriptomes, proteomes, and metabolomes are strongly tissue-resolved, with many metabolites confined to specific organs or even cell types. Capturing this spatial metabolic specialization within a mechanistic framework requires models that move beyond generic “whole-plant” reconstructions to explicitly represent organ or tissue-specific metabolic states.

Genome-scale metabolic models (GEMs) provide a framework for addressing this complexity by linking genes, enzymes, reactions and metabolites into a stoichiometric network that can be analyzed with constraint-based methods^{3,10}. In plants, GEMs have evolved from generic reconstructions centered mainly on primary metabolism toward multi-compartment and multi-tissue models that represent organ-level metabolic specialization, carbon and nitrogen partitioning, and the integration of specialized

metabolism with growth^{10–13}. Foundational work in *Arabidopsis thaliana* established strategies to represent whole-plant and tissue-associated metabolism, while recent models in species such as *Vitis vinifera* have shown how tissue-specific constraints can be used to explore flux allocation into phenylpropanoid, terpenoid and other specialized pathways^{10,12,13}. These studies demonstrate that tissue-specific GEMs can help connect genome annotation, metabolic network structure and organ-level chemical phenotypes.

In parallel, methodological advances have made it increasingly feasible to embed omics evidence directly into metabolic models. Generic plant GEMs can be contextualized into condition or tissue-specific variants by imposing constraints derived from transcriptomics, proteomics or metabolomics data, thereby pruning reactions that are inactive in a given context and emphasizing pathways that are supported by experimental evidence¹³. In particular, metabolomics-guided approaches use detected metabolites, relative abundances and tissue-specific presence/absence patterns to refine network topology, adjust exchange and transport constraints, and prioritize specialized metabolic branches¹³. This is especially attractive for medicinal and aromatic plants, where the main interest lies in dissecting the tissue-resolved biosynthesis of bioactive compounds rather than only maximizing growth^{4,12}.

Despite recent advances in genomic and phytochemical characterization of *E. novogranatense*, including the availability of a chromosome-scale genome assembly, no genome-scale metabolic model has yet been developed for this species. Furthermore, a tissue-specific framework linking gene content to the distribution of specialized metabolites in leaves and roots is still lacking^{7,9}. As a result, fundamental questions remain about how coca plants balance carbon and nitrogen allocation between growth and defense and how the biosynthesis of tropane alkaloids, flavonoids, and other bioactive metabolites is regulated across tissues.

Here, we address this gap by reconstructing and curating two tissue-specific genome-scale metabolic models of *E. novogranatense*, one for leaf and one for root. The models are anchored in a shared genome-scale network and differentiated using untargeted liquid chromatography–quadrupole time-of-flight mass spectrometry (LC–QTOF–MS) metabolomics. By integrating genomic annotation, biochemical database information, manual curation of specialized metabolism and tissue-resolved metabolomic evidence, this study provides a mechanistic framework for exploring leaf–root metabolic differentiation in *E. novogranatense*. This framework extends the study of coca metabolism beyond cocaine biosynthesis and enables quantitative analysis of how this non-model plant channels its metabolic capacity into a chemically diverse repertoire of specialized metabolites.

Methods

Summary of the methodological pipeline for model reconstruction, contextualization, and simulation

To provide an overview of the reconstruction and analysis workflow, the main methodological steps were summarized in a schematic diagram (Fig. 1). Starting from a chromosome-level genome assembly obtained from NCBI, we performed functional annotation using complementary tools and mapped predicted enzymes to curated biochemical pathway databases. These annotations were used to generate an initial draft genome-scale metabolic model, which was subsequently refined through manual and semi-automatic curation steps. Untargeted LC–QTOF–MS metabolomics of leaves and roots was then integrated to contextualize the model at the tissue level, generating leaf- and root-specific GEMs for downstream constraint-based analyses.

Selection of the reference genome, assembly characteristics, and sequencing

The reference genome of *Erythroxylum novogranatense* (GenBank assembly accession GCA_029891385.1) was obtained from the NCBI Assembly database and corresponds to a chromosome-level assembly reported by Wang et al.¹⁴. This assembly spans approximately 1.3 Gb in total ungapped length and is organized into 12 nuclear chromosomes and one organellar sequence. According to the NCBI assembly report, it comprises 262 scaffolds, with a scaffold N50 of 102.9 Mb and L50 of 6, and 593 contigs, with a contig N50 of 7.6 Mb and L50 of 55. The genome has a GC content of 36% and an estimated sequencing depth of approximately 106×. The assembly was generated using Oxford Nanopore PromethION long reads, Illumina NovaSeq short reads and Hi-C data, and was assembled with SMARTdenovo. Genomic sequences and corresponding annotation files were downloaded in FASTA and GenBank/GFF3 formats and used as the basis for downstream gene prediction, functional annotation and genome-scale metabolic model reconstruction.

Functional annotation of metabolic genes and enzymes

Functional genome annotation was carried out to assign biological roles to predicted protein-coding genes, with particular emphasis on enzymes involved in primary and specialized metabolism. To increase both coverage and confidence in enzyme function assignments, three complementary tools were combined. RAST was used to assign functions based on curated metabolic subsystems from the SEED database and to propose enzyme functions and Enzyme Commission numbers^{15,16}. plantiSMASH was used to detect biosynthetic gene clusters related to specialized metabolism, including alkaloid-, terpene- and hybrid-type clusters¹⁷. eggNOG-mapper was used to transfer functional information from orthologous groups to the predicted protein sequences, providing additional functional descriptions, EC numbers, KEGG orthology identifiers and pathway assignments^{18,19}. The combination of these tools generated a catalog of metabolic genes and enzymatic functions for incorporation into the GEM.

Mapping and curation of metabolic pathways

Annotated enzymes were mapped onto curated biochemical pathway databases, including KEGG, MetaCyc/PlantCyc and BRENDA, to reconstruct the metabolic network of *E. novogranatense*^{20–22}. KEGG and MetaCyc were used to link EC numbers and enzyme names to canonical reactions and pathways, providing the backbone of primary metabolism. PlantCyc complemented this information with plant-specific pathways, particularly those associated with specialized metabolism, while BRENDA supplied additional enzyme properties and reaction evidence. Particular emphasis was placed on pathways associated with tropane alkaloids, flavonoids, phenylpropanoids and lipid metabolism, which were reconstructed by tracing key enzymatic steps across these resources and the available literature.

Automatic reconstruction of the draft metabolic model

To obtain an initial draft of the metabolic network, the functional annotations were imported into the DOE Systems Biology Knowledgebase (KBase), an open-source platform for integrative omics analysis and metabolic modeling²³. Within KBase, the ModelSEED reconstruction pipeline was used to assemble a draft genome-scale metabolic model by mapping genes to reactions in the ModelSEED Biochemistry Database²⁴. The resulting draft model provided preliminary gene–protein–reaction associations, reaction reversibility assignments and subcellular compartmentalization. This automatically generated network served as the structural backbone of the *E. novogranatense* GEM and was exported in SBML format (Systems Biology Markup Language) for subsequent manual curation and refinement.

Manual curation and consistency refinement of the metabolic network

The draft model was manually curated using COBRApy²⁵, focusing on the incorporation and validation of specialized metabolic pathways not adequately represented in the automated reconstruction. Reactions and metabolites associated with tropane alkaloids, flavonoids and lipid metabolism were added and refined based on curated biochemical databases and literature-supported pathway information^{20–22}. For each pathway, enzyme annotations and EC numbers were reviewed, and gene–protein–reaction (GPR) associations were assigned when genomic evidence was available. Reaction stoichiometry, charge and proton balance, reversibility and compartmentalization were checked to ensure physicochemical consistency.

Additionally, network connectivity was evaluated to avoid the inclusion of isolated reactions. Disconnected metabolites and pathway segments unable to carry flux were identified and corrected when possible. This process resulted in a curated GEM in which specialized metabolic pathways are explicitly represented and functionally integrated with the central metabolism of *E. novogranatense*.

Identification of missing reactions and healing of the metabolic network through gap-filling

Although the curated draft model captured most central and specialized metabolic functions, several pathways remained incomplete and could not support biomass production under defined tissue-specific conditions. To address this limitation, a gap-filling procedure was applied to identify a parsimonious set of additional reactions capable of restoring network connectivity and enabling the model to satisfy a predefined functional objective^{26–28}. Candidate reactions were evaluated from curated biochemical reference pools, taking as a reference curated metabolic models of model plants, in this case *Arabidopsis thaliana* and retained reactions were reviewed to prioritize reactions consistent with plant metabolism or supported by biochemical evidence.

Mathematically, gap-filling was formulated as an optimization problem in which the model seeks to minimize the cost of adding candidate reactions while satisfying stoichiometric and functional constraints. The objective can be expressed as:

$$\min \sum c_i z_i$$

1

subject to:

$$Sv = 0$$

2

under conditions:

$$v^* \geq t$$

$$l_i \leq v_i \leq u_i$$

$$v_i \leq u_i z_i$$

where (S) is the stoichiometric matrix, (v) is the flux vector, (z_i) is a binary variable indicating whether candidate reaction (i) is included in the model, (c_i) is the cost associated with adding that reaction, and (l_i) and (u_i) are the lower and upper bounds of reaction (i), respectively. The steady-state constraint (Sv = 0) ensures mass conservation for internal metabolites, while the constraint v* ≥ t requires the model to achieve a minimum value of the defined objective, in this case biomass production under tissue-specific medium conditions. In biological terms, this procedure identifies a minimal set of reactions that restores network connectivity without overexpanding the reconstruction.

Untargeted LC–QTOF–MS metabolomics of leaf and root for model contextualization

Leaf and root samples of *E. novogranatense* were collected in Caldono, Cauca, Colombia, at approximately 2.7° N, 76.48° W and 1,600 m above sea level. The sampling site corresponds to a domestic garden in a rural Andean setting. Collections were carried out in July 2025. Specimens were taxonomically identified using standard botanical keys and comparison with herbarium material. Voucher

specimens were deposited in the ANDES Herbarium of the Museo de Historia Natural C. J. Marinkelle, Universidad de los Andes, under collection number RNC:158. All collections were conducted under permit PR.6.2025.11467, issued by the Colombian Ministry of Environment, in compliance with national regulations on access to biological resources.

All samples were analyzed at the MetCore facility, Universidad de los Andes, Bogotá, Colombia, using reversed-phase liquid chromatography coupled to quadrupole time-of-flight mass spectrometry (LC–QTOF–MS) in positive electrospray ionization mode. Feature detection and chromatographic alignment were performed in Agilent MassHunter Profinder 10.0 using the Batch Recursive Feature Extraction workflow. Putative metabolite annotation followed a multi-tool strategy that combined MS-DIAL for peak detection, MS/MS deconvolution and alignment²⁹, CEU Mass Mediator 3.0 as a meta-search engine against multiple metabolomics databases³⁰, and RefMetaPlant for plant-focused metabolite matching³¹. High-confidence molecular formulas and candidate structures were further refined using SIRIUS 4 and CSI:FingerID based on fragmentation patterns³², while spectral library matching and molecular networking in GNPS facilitated annotation transfer within chemically related clusters³³. Chemical identifiers, including exact mass, InChIKey and ChEBI identifiers, were harmonized using PubChem, ChEBI and KNApSACk^{34–36}. Metabolite identification confidence was assigned according to the Metabolomics Standards Initiative, distinguishing confirmed identifications, putative spectral matches, tentative structural annotations and MS1-based annotations³⁷.

Constraint-based simulations for functional comparison of leaf and root metabolic models

The leaf and root genome-scale metabolic models were evaluated using constraint-based modeling and flux balance analysis (FBA)³⁸. All simulations assumed steady state for internal metabolites (Eq. 2), each reaction flux was constrained by lower and upper bounds:

$$lb_j \leq v_j \leq ub_j$$

which define reaction reversibility, uptake capacity, secretion capacity and transport limits. For each tissue model, FBA was formulated as a linear optimization problem in which the biomass reaction was maximized:

$$\max Z = v_{biomasa}$$

3

subject to (Eq. 2) and ($lb_j \leq v_j \leq ub_j, \forall j \in J$) for all reactions (j). Biomass objective values were interpreted as internal model objective values under the defined constraints, not as experimentally calibrated growth rates.

Tissue-specific media were applied under an uptake-controlled scheme to avoid artificial growth caused by open exchange reactions. Before applying each medium, boundary reactions not included in the

corresponding medium or in a predefined list were closed by setting ($lb = ub = 0$). Then, only the exchange reactions specified for each tissue medium were reopened with their corresponding bounds. Under COBRA convention, negative lower bounds allow uptake from the medium, whereas positive upper bounds allow secretion^{25,39,40}. The leaf model was simulated under an autotrophic medium, allowing inorganic sources such as CO₂ and mineral nutrients, whereas the root model was simulated under a heterotrophic medium allowing organic carbon uptake, together with the corresponding inorganic nutrients and cofactors. Model viability was verified under each tissue-specific medium before comparative analyses.

For each tissue (k), the optimal biomass objective value was calculated under conditions:

$S \cdot v = 0 \quad l \leq v \leq u$ as:

$$B_{opt}^k = \max_{biomass}^k$$

4

Under its corresponding medium. To compare leaf and root models under equivalent relative functional demand, subsequent analyses imposed a minimum biomass constraint equal to 90% of the tissue-specific optimum:

$$v_{biomass}^k \geq 0.9 \cdot B_{opt}^k$$

5

This constraint avoided comparing models operating at non-equivalent physiological states and ensured that both tissues were analyzed under high biomass demand relative to their own feasible optimum.

Structural and functional similarities between the leaf and root GEMs were quantified using the Jaccard index⁴¹:

$$J(A, B) = \frac{|(A \cap B)|}{|(A \cup B)|}$$

6

Structural similarity was calculated using the sets of internal reactions in each model, excluding boundary reactions such as exchanges, sinks and demands. Functional similarity was calculated using the sets of active reactions under the tissue-specific medium. A reaction was considered active when its absolute flux satisfied $|v_j| > 10^{-6}$. This two-level comparison separated network similarity based on reaction presence from functional similarity based on reaction usage under the simulated conditions.

To evaluate whether leaf and root models differ globally in their flux organization, flux sampling was performed under the tissue-specific media and the common biomass constraint subject to (Eq. 2) and $l \leq v \leq u$ and $v_{biomass}^k \geq 0.9 \cdot B_{opt}^k$. Each sampled solution represented a feasible flux vector

satisfying stoichiometric balance, reaction bounds, medium constraints and the biomass constraint⁴². The resulting leaf and root flux matrices were aligned using only reactions shared by both models. This avoided trivial separation caused by reactions present in only one tissue-specific reconstruction.

Before principal component analysis (PCA), reactions with negligible variability across the combined sample set were removed using a standard deviation threshold of 10^{-9} . The remaining shared reaction fluxes were standardized using z-scores:

$$z_{ij} = \frac{v_{ij} - \mu_j}{\sigma_j}$$

7

where x_{ij} is the flux of reaction j in sample i , μ_j and σ_j are the mean and standard deviation across all samples. PCA was then performed on the standardized matrix Z to identify directions of maximum variance, and sample scores were calculated as ($T = ZW$) where W contains the principal component loadings⁴³. The explained variance of the first two components was used to summarize global differences between tissues. To complement PCA, a heatmap of standardized flux differences was generated from the same sampling data. For each selected reaction, z-scores were averaged across samples within each tissue to represent relative flux usage in leaf and root.

Finally, the production potential of selected metabolites was evaluated by maximizing their corresponding demand reactions under the same tissue-specific media and biomass constraint ($v_{\text{biomass}} \geq 0.9 \cdot B_{\text{opt}}$). This ensured that metabolite production was assessed under comparable growth conditions, allowing direct comparison of biosynthetic capacity between leaf and root models.

Results

Reconstruction and properties of tissue-specific GEMs

To determine the metabolic potential of *Erythroxylum novogranatense*, we first used the chromosome-level genome assembly as the basis for functional annotation and draft model reconstruction. The combination of Plant RAST, eggNOG-mapper and plantiSMASH provided complementary evidence for metabolic gene annotation and pathway reconstruction. Plant RAST recovered a broad set of enzymatic functions associated with plant primary metabolism, while eggNOG-mapper assigned functional descriptions, Enzyme Commission numbers and KEGG orthology identifiers to a large proportion of the predicted proteins. In parallel, plantiSMASH identified biosynthetic gene clusters associated with specialized metabolism, including alkaloid, terpene and hybrid cluster classes (Table 1).

Table 1
Summary of functional annotation resources applied to the *E. novogranatense* genome.

Tool / resource	Scope (input)	Key metabolic annotation statistics
1) Plant RAST (PlantSEED)	28,604 protein sequences	1,971 enzymes (97%) of plant primary metabolism; 498 distinct enzymatic functions
2) eggNOG-mapper	27,528 protein queries	25,649 proteins with functional descriptions (93.2%); 5,780 proteins with ≥ 1 EC (1,299 ECs); 13,317 with ≥ 1 KEGG KO (4,075 KOs); 6,333 proteins (23.0%) in COG categories (C, E, F, G, H, I, P, Q); 8,197 proteins (29.8%) in category S (function unknown)
3) plantiSMASH	Whole nuclear genome	50 predicted biosynthetic gene clusters (BGC); major classes: 20 saccharide, 8 putative, 5 alkaloid, 4 terpene, plus multiple hybrid clusters (e.g. fatty acid–polyketide–alkaloid)

Together, these annotation resources supported the reconstruction of a draft metabolic network containing central metabolic pathways and a preliminary representation of specialized metabolism. The initial model comprised 1,066 reactions and 1,134 metabolites. As expected for an automatic reconstruction of a non-model plant species, primary metabolism was better represented than specialized metabolism, whereas alkaloid, flavonoid and lipid-related pathways were only partially represented or remained fragmented.

Manual curation was performed to expand and connect specialized metabolic routes relevant to *E. novogranatense*, tropane alkaloids, flavonoids and lipid-derived metabolism. In the tropane alkaloid module, 21 reactions and 46 metabolites were incorporated to connect central and polyamine-related precursors with intermediates such as putrescine, N-methylputrescine and littorine, and with characteristic alkaloids such as cocaine and benzoylecgonine. This step improved the connectivity of the tropane pathway within the broader metabolic network.

The flavonoid module was expanded with 12 reactions and 14 metabolites, covering the formation of chalcones and flavanones, their conversion into flavonols, and the production of glycosylated derivatives such as kaempferol glucoside and luteolin-related compounds. The lipid-derived module incorporated 26 reactions and 27 metabolites, representing reactions linked to acetyl-CoA and malonyl-CoA metabolism and the formation of medium- and long-chain lipid derivatives. Overall, manual curation expanded the draft reconstruction with 59 specialized-metabolism reactions and generated a more connected representation of the main pathway families selected for tissue-specific contextualization.

The integration of leaf- and root-detected metabolites further refined the models and allowed their tissue-specific contextualization. In the leaf GEM, metabolomics-guided contextualization incorporated 59 metabolites connected through 114 reactions within the tissue-specific subnetwork. Representative leaf-associated compounds included umbelliferone, cinnamoylcocaine, chlorogenic acid, norcocaine and oxygenated or glycosylated flavonoids, including luteolin glucoside. These additions enriched the representation of coumarin, phenylpropanoid, flavonoid and acylated tropane metabolism in the leaf model.

In the root GEM, metabolomics-guided contextualization incorporated 66 metabolites connected through 116 reactions. Root-associated compounds included alkaloids and terpenoid-related metabolites such as palmatine, ancistrocogolensine, anisodine, comosivine, convolvamine and norcorytuberine. These additions expanded the representation of tropane-related and isoquinoline-type alkaloid metabolism and reinforced the chemical differentiation between leaf and root models.

After manual curation and metabolomics-guided contextualization, the final leaf model contained 1,764 reactions, 1,550 metabolites and 1,428 genes, while the root model contained 1,768 reactions, 1,565 metabolites and 1,426 genes. Both models share a common core metabolism but differ in the inclusion of tissue-associated specialized pathways, consistent with the metabolomic profiles of leaf and root.

To illustrate how these additions connect specialized metabolism with the core network, four representative biosynthetic pathways to tissue-specific metabolites are shown (Fig. 2). The figure presents schematic routes from aromatic or polyamine precursors to umbelliferone and cinnamoylcocaine in leaf, and to anisodine and palmatine in root, highlighting key enzymatic steps. This representation captures the types of pathways incorporated during curation and metabolomic contextualization, and provides a reference for interpreting the metabolic differences between leaf and root.

Overall, the transition from the initial model to tissue-specific GEMs contextualized by metabolomics resulted in a substantial increase in the coverage of specialized pathways and the explicit incorporation of experimentally detected metabolites in each tissue, while maintaining a coherent central core of reactions and metabolites on which these differences are built.

Leaf and root metabolomic profile: chemical evidence of tissue specialization

Untargeted LC–QTOF–MS metabolomic profiling detected numerous molecular features across leaf and root tissues of *Erythroxylum novogranatense*. From these features, a total of 122 metabolites were annotated and used as the experimental metabolomic basis for tissue-specific model contextualization (Fig. 3A). According to the Metabolomics Standards Initiative (MSI), most annotations corresponded to Level 2, representing putative structures supported by spectral, formula or database matching. This group included 21 metabolites in leaves and 19 in roots. A smaller subset reached Level 1, with 3 confirmed metabolites in leaves and 2 in roots, whereas 33 leaf metabolites and 44 root metabolites were classified as Level 3 tentative annotations (Table S1). Although many detected molecular features remained structurally unassigned, the annotated fraction captured a representative portion of the chemical diversity present in both tissues.

Among the annotated metabolites, 48 were shared between leaf and root tissues. In addition, 9 metabolites were uniquely annotated in leaves, whereas 17 metabolites were uniquely annotated in roots (Fig. 3B). This distribution shows that both tissues share a common annotated metabolite core, but also

contain tissue-associated compounds that were later used as contextual markers for the leaf and root GEMs.

The classification of annotated metabolites by chemical family showed that alkaloids were the dominant group in both tissues (Fig. 4). In leaves, alkaloids represented 33.33% of annotated metabolites, followed by amino acids and derivatives, flavonoids, phenylpropanoids and polyketides, and smaller contributions from terpenoids, lipids and other chemical classes. In roots, alkaloids represented 38.46% of annotated metabolites, followed by amino acids and derivatives, terpenoids, flavonoids, lipids and lipid-like molecules, and phenylpropanoids and polyketides. Therefore, although alkaloids were prevalent in both tissues, leaves showed a higher relative representation of flavonoids and phenylpropanoid-related compounds, whereas roots showed a higher relative representation of terpenoids and a broader distribution of nitrogenous and lipid-associated metabolites.

Metabolites uniquely detected in each tissue reinforced this chemical differentiation (Table 2). Leaf-associated markers included tropane-related alkaloids such as benzoylecgonine, cinnamoylcocaine and norcocaine, together with flavonoids and phenylpropanoid-related compounds including kaempferol glucoside, luteolin glucoside, tetrahydroxyflavone, umbelliferone, chlorogenic acid and dihydroferuloyltyramine. These compounds are consistent with polyamine-derived alkaloid metabolism, phenylalanine-derived phenylpropanoid metabolism and flavonoid biosynthesis.

In root, tissue-associated markers included alkaloids such as ancistrocogolensine, anisodine, comosivine, convolvamine, daphnezomine A, norcorytuberine, palmatine, retrorsine N-oxide and senecionine. Root-exclusive annotations also included the lipid-derived compound 3-hydroxyhexadec-9-enoyl-phenylalanine, the phenylpropanoid-related compound 1-(3,4-dimethoxycinnamoyl)piperidine, and terpenoids such as costunolide, darutigenol, farnesol, prespatane and dehydrocostus lactone. Together, these metabolites reflect a more heterogeneous chemical profile in root, spanning multiple classes of specialized metabolism.

Table 2

Leaf- and root-specific annotated metabolites used as tissue-specific contextual markers in the GEM.

Leaf		Root	
Alkaloids	• Benzoylecgonine • Cinnamoylcocaine • Norcocaine	Alkaloids	• Ancistrocongolensine • Anisodine • Comosivine • Convolvamine • Daphnezomine A • Norcorytuberine • Palmatine • Retrorsine N-oxide • Senecionine
Flavonoids	• Kaempferol glucoside • Luteolin glucoside • Tetrahydroxyflavone	Flavonoids	-
Lipids and lipid-derived molecules	-	Lipids and lipid-derived molecules	• 3-hydroxyhexadec-9-enoyl-phenylalanine
Phenylpropanoids and polyketides	• Umbelliferone • Chlorogenic acid • Dihydroferuloyltyramine	Phenylpropanoids and polyketides	• 1-(3,4-Dimethoxycinnamoyl)piperidine
Terpenoids	-	Terpenoids	• Costunolide • Darutigenol • Farnesol • Prespatane • Dehydrocostus lactone

Overall, the metabolomic profiles showed that leaf and root tissues share a common set of annotated compounds but diverge in specific branches of specialized metabolism. This pattern reflects tissue-associated chemical differentiation, with each organ displaying a distinct composition of metabolites consistent with its functional role and chemical environment.

Constraint-based simulations reveal tissue-specific flux organization

Under defined tissue-specific media, both leaf and root GEMs reached feasible optimal solutions by maximizing the biomass objective. The leaf model reached an optimal biomass objective value, B_{opt} , of 10.11, whereas the root model reached a B_{opt} of 14.53. These values represent model objective values under the defined constraints and were used for comparative simulations rather than experimentally calibrated growth rates.

At these optimal values, the number of active reactions was very similar between tissues (482 in leaf, 486 in root), indicating that both models use a comparable fraction of their network under the same conditions. However, the overlap between models changed when structural similarity was compared with functional similarity. The structural Jaccard index calculated from internal reactions was 0.94, showing that both models share a broad metabolic core. In contrast, the functional Jaccard index calculated from active internal reactions decreased to 0.73, indicating that leaf and root models used partially different subsets of reactions under biomass-optimal conditions.

Principal component analysis based on shared reaction fluxes clearly separated the leaf and root samples. Taken together, PC1 and PC2 explained 0.509 of variance (explained variance by PC1 = 0.3527, by PC2 = 0.1578). The projections onto the PC1–PC2 plane formed two well-defined groups, with no overlap between tissues, indicating that the feasible solutions for each GEM occupy distinct regions of flux space even under the common biomass constraint. Within each group, subclusters were observed, consistent with the existence of alternative pathways or local redundancies but always preserving a clear tissue identity. Overall, the PCA provides global evidence that the functional organization of flows differs between leaves and roots, despite the common structural core described in the previous sections (Fig. 5).

A heatmap of standardized fluxes was used to examine the reactions underlying the separation observed in PCA (Fig. 6). The analysis was based on shared internal reactions and represents average z-score values for leaf and root flux samples, the z-score summarizes how much the flux of each reaction deviates from the combined leaf-root average, with more extreme values indicating reactions with greater discriminatory capacity between tissues. Several reactions showed higher relative flux values in the root model, including those associated with long-chain isoprenoid biosynthesis, one-carbon metabolism, nucleotide metabolism, amino acid metabolism and transport processes, while others were enriched in the leaf model. This pattern indicates that the PCA separation arises from coordinated differences across multiple parts of the shared metabolic network, rather than from a single pathway.

Finally, the maximum production potential of selected metabolites shared by both tissue models was evaluated (Fig. 7). For most metabolites analyzed, the root model showed higher production capacity than the leaf model. This included central cofactors such as nicotinamide and niacin, phenolic intermediates such as sinapaldehyde and cyanidin, and flavonoid-related compounds such as kaempferol and naringenin. These results show that, despite sharing a large structural core and sustaining comparable relative biomass demand, the two GEMs differ in their predicted capacity to route flux toward selected shared metabolites.

The simulation results confirm that the leaf and root GEMs are both viable under their defined tissue media and share a broad internal reaction core, with a similar number of active reactions at their respective biomass optima. Despite this structural similarity, the partial overlap in active reactions, the clear separation in flux space observed by PCA, the differences in standardized flux distributions, and the distinct metabolite production potentials consistently indicate that each tissue model redistributes flux through the shared network in a characteristic way. These findings suggest that structural similarity between tissues does not necessarily translate into functional equivalence.

Discussion

Genome-scale metabolic modeling of non-model plant species has remained limited, particularly for species with diverse specialized metabolism that extends well beyond the pathways represented in general biochemical databases. The tissue-specific GEMs presented here address this gap for *Erythroxylum novogranatense*, a species whose metabolic biology has been studied almost exclusively through the lens of cocaine biosynthesis. These are the first leaf and root GEMs developed for this species, and their construction required integrating genome annotation, targeted pathway curation and tissue-resolved metabolomics to capture a chemical diversity.

This is important because most studies on *Erythroxylum* have focused on cocaine or on specific groups of tropane alkaloids, while the broader organization of specialized metabolism across tissues has remained poorly represented at the systems level^{6,7}. By explicitly separating leaf and root metabolism, the models provide a framework to study *E. novogranatense* as a chemically diverse plant rather than as a system centered on a single metabolite.

The genome-scale models described in the literature vary in size depending on whether they focus on primary metabolism or include specialized metabolism, the level of compartmentalization (organelles/compartments), and if they are multicellular or even dynamic (e.g., daytime/growth phase). In *Arabidopsis thaliana*, this historical “evolution” is clear from the first GSM reported (~ 1,253 metabolites and ~ 1,406 reactions) to dynamic variants that scale up by orders of magnitude in reactions when they explicitly formalize growth and C/N couplings over time^{11,44,45}. In rice (*Oryza sativa*), the literature ranges from “core” models (hundreds of metabolites/reactions) to leaf models focused on light-dependent metabolism (e.g., ~ 1,484 metabolites; ~1,736 reactions)⁴⁶. In C4 crops such as maize (*Zea mays*), the leap in complexity is even more pronounced when the mesophyll–bundle sheath interaction and multiple tissues are explicitly modeled: there are models ranging from C4GEM (~ 1,755 metabolites; ~1,588 reactions) to C4 leaf reconstructions with thousands of metabolites, reflecting that model “size” is not just about quality but also about how explicitly it represents the biology (tissues/cells/conditions)⁴⁵.

Within this context, our leaf and root models of *E. novogranatense* fall into the range of moderately large GEMs and are comparable to leaf tissue reconstructions when the goal is to capture physiology and functional differences without artificially inflating the network through multi-tissue “duplication”; for example, the *Setaria italica* model reports ~ 1690 metabolites and ~ 1515 reactions, whereas moving to

species with larger reconstructions such as *Solanum lycopersicum* (~ 1998 metabolites; ~2143 reactions) increases the counts due to expanded coverage and tissue framework design^{13,47,48}. In other words, across the model/non-model plant spectrum, our GEMs exhibit structures comparable to, and consistent with reference tissue reconstructions (same compartmentalization logic and coverage), while maintaining a manageable size for mechanistic analyses (e.g., flux PCA and contrasts in metabolite production) without losing the common core expected among tissues of the same species.

The available evidence suggests that the leaf–root contrasts we observed are consistent with phytochemical patterns described at the genus and family level, and with the functional logic of tissue partitioning of specialized metabolism. In recent studies on *Erythroxylum*, 383 compounds reported in 67 species were compiled, with tropane alkaloids highlighted as the most representative group (including an update to 197 reported tropane alkaloids), followed by terpenoids (including diterpenoids/triterpenoids) and phenolic families such as flavonoids and tannins⁷.

In this context, our tissue markers are consistent with the reported literature: in leaves, derivatives associated with the cocaine pathway and flavonoids (e.g., benzoylecgonine, cinnamoylcocaine, flavone/flavonol glycosides, and coumarins such as umbelliferone) are present, while most diterpenes are prevalent in the roots. Our results also agree with the literature that the leaf tissues in the genus typically exhibit a high load of phenolic compounds (total phenols, tannins, and flavonoids). Additionally, tissue-specific gene expression and protein levels indicate tropane alkaloid biosynthesis in the *Erythroxylaceae* occurs in the above-ground portions of the plant^{5,7}. Together, the latter evidence reinforces that the relative enrichment of alkaloids and flavonoids in leaves we observed, is not a metabolite identification artifact but a local manifestation of a widely described trend in this family.

Although alkaloids were represented in both tissues, roots showed a greater relative representation of terpenoids, whereas leaves showed a greater representation of flavonoids. This pattern is consistent with the literature from two perspectives: first, ecology and tissue function, where flavonoids are preferentially associated with photosynthetic organs due to their role in photoprotection and signaling, while in roots, chemical defense pathways and environmental interaction tend to be more prominent; and second, the biosynthetic architecture of alkaloids, in which different plant systems exhibit a specific tissue organization of specialized nitrogen pathways and derivative transport⁴⁹.

In our case, this partitioning is expressed not only as a general difference in "alkaloids" but also as a difference between alkaloid branches: leaves are dominated by compounds associated with the tropane alkaloid pathway, while roots are enriched with isoquinoline/protoberberine alkaloids. In the specific case of the cocaine pathway, there is evidence in *Erythroxylum coca* that the final acylation step is catalyzed by a BAHD acyltransferase (cocaine synthase), experimentally confirming a key segment of the pathway and its immediate intermediates⁴⁹. Furthermore, reviews of the genus's chemical profile show that tropane alkaloids constitute the most distinctive chemical core of *Erythroxylum* and summarize the diversity of skeletons and derivatives reported in various organs⁷.

The enrichment of alkaloids, phenols, and flavonoids in leaves, and of terpenoids and lipids in roots, as indicated by both our markers and metabolomic chemical-class distributions, is consistent with previous reports. This concordance provides a robust basis for interpreting the functional differences observed in GEM analyses (flux production capacity and metabolic reconfiguration) as reflecting genuine tissue-specific metabolic partitioning rather than methodological bias^{6,7,49}.

The simulation results highlight the distinction between structural similarity and functional differentiation in the two GEMs. The higher Bopt obtained for the root model was not interpreted as evidence of a higher absolute growth rate, since biomass maximization in FBA depends on the biomass equation, the imposed bounds and the limiting resources defined in the model^{38,50}. Instead, Bopt was used as an internal reference to compare both models under the same relative biomass requirement.

Under this condition, leaf and root retained a highly similar reaction structure but differed in the set of active reactions, indicating differences in flux usage despite a shared metabolic network. This pattern is consistent with previous plant multi-tissue GEM studies, where tissue identity emerges not only from reaction content but from how flux is distributed across shared pathways^{10,12}.

The PCA analysis supports the same conclusion from a multivariate perspective. Because the analysis was restricted to reactions shared by both models, the separation between leaf and root was not a trivial consequence of tissue-exclusive reactions. Instead, it showed that the common metabolic backbone can sustain distinct flux states depending on tissue-specific reconstruction and constraints. The presence of subgroups within each tissue is compatible with alternative feasible solutions, a known feature of constraint-based models in which multiple flux distributions can satisfy the same optimal or near-optimal objective^{51,52}. However, these alternatives remained organized by tissue, supporting a consistent leaf–root difference in flux organization.

The production potential analysis adds a biochemical output to this comparison. Maximum demand fluxes do not represent metabolite abundance, but they indicate the capacity of each GEM to route flux toward selected compounds while maintaining the 90% Bopt constraint. Therefore, differences in predicted production potential are best understood as differences in feasible biosynthetic capacity, not as direct substitutes for LC–MS intensities. This distinction is relevant because specialized metabolite accumulation depends on biosynthesis, conversion, transport and storage, and is not always directly coupled to biomass or to maximum pathway capacity⁵³. In this context, the higher production potential predicted for several shared metabolites in the root model identifies candidate metabolic capacities that can be tested experimentally, rather than final claims about tissue concentration.

Overall, the simulations indicate that leaf and root GEMs of *E. novogranatense* share a broad metabolic core but differ in how that core is used. This functional differentiation is consistent with the metabolomic evidence and with the expected organization of plant metabolism, where tissue identity arises from both specialized metabolite composition and differential use of central metabolic pathways. Together, these models provide a reproducible framework to examine how *E. novogranatense* allocates metabolic capacity between growth and specialized metabolite biosynthesis beyond cocaine.

Conclusion

This work provides the first integrated representation of the metabolism of *E. novogranatense* at the genomic scale and with tissue resolution, in the form of two leaf and root GEMs that combine a shared primary core with an explicit layer of chemical specialization. Based on genomic annotation, the targeted curation of tropane, flavonoid, terpenoid, and lipid modules, and the selective incorporation of metabolites annotated by metabolomics, the models not only capture the cocaine pathway but also a broader repertoire of compounds consistent with the chemical diversity described for the genus. Biomass optimization analyses, structural similarity, and separation in the flux space show that both tissues operate with a very similar metabolic architecture, but they redistribute the use of the pathway and biosynthetic capacity in a differentiated manner, in accordance with the tissue-specific patterns observed in the metabolomic annotations (abundance of alkaloids/phenols/flavonoids in leaves and alkaloids/lipids/terpenoids in roots). In this way, the models are not limited to reproducing lists of reactions but rather provide a quantitative framework to explore how the allocation of resources is balanced between growth and the production of metabolites of interest in each tissue. While inherent limitations of plant GEM models persist, such as the absence of explicit regulation, the presence of unannotated features, and the need to refine experimental constraints, the presented reconstructions constitute a reproducible starting framework upon which mechanistic hypotheses can be formulated and validation experiments designed. Several of the metabolites represented in these tissue-resolved models have reported bioactivities, which may motivate future studies aimed at evaluating tissue-specific extracts or metabolite-guided biotechnological applications.

Abbreviations

BGC

biosynthetic gene cluster

Bopt

optimal biomass objective value

EC

Enzyme Commission number

FBA

flux balance analysis

GEM

genome-scale metabolic model

GPR

gene–protein–reaction association

KEGG

Kyoto Encyclopedia of Genes and Genomes

KO

KEGG Orthology

LC–QTOF–MS

liquid chromatography–quadrupole time-of-flight mass spectrometry

MSI

Metabolomics Standards Initiative

PCA

principal component analysis

RAST

Rapid Annotation using Subsystem Technology

SBML

Systems Biology Markup Language

COG (Cluster of Orthologous Genes):

- **[C]** Energy production and conversion
- **[E]** Amino acid transport and metabolism
- **[F]** Nucleotide transport and metabolism
- **[G]** Carbohydrate transport and metabolism
- **[H]** Coenzyme transport and metabolism
- **[I]** Lipid transport and metabolism
- **[P]** Inorganic ion transport and metabolism
- **[Q]** Secondary metabolites biosynthesis, transport, and catabolism

Declarations

Ethics approval and consent to participate

This study did not involve human participants, human data, human tissue, animals or animal-derived material. Plant material was collected under permit PR.6.2025.11467 issued by the Colombian Ministry of Environment, in compliance with national regulations on access to biological resources. Voucher specimens were deposited in the ANDES Herbarium of the Museo de Historia Natural C. J. Marinkelle, Universidad de los Andes, under collection number RNC:158.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was financed by the Colombian Ministry of Sciences, Technology and Innovation, grant No. 112721-170-2024, project code 109055, and grant No. 032-2025, project code 112320. Additional support was provided by the Assistant Professorship Funds from Universidad de los Andes, code INV-2023-158-2701. The funders had no role in the study design, data collection, data analysis, interpretation of results, manuscript preparation or decision to submit the work for publication.

Author Contribution

J.N.A. conceived the study, designed the workflow, built the GEMs, performed the analyses, and interpreted the data. D.M.B.F. contributed to data analysis and interpretation. A.F.G.B. conceived the study, designed the workflow, and supervised the modeling analyses. M.F.V.T. contributed to data interpretation and supervised the biological and metabolomic components of the study. All authors reviewed and approved the manuscript.

Acknowledgement

The authors acknowledge institutional support from Universidad de los Andes. The authors also thank the metCORE facility and the Center for the Study on Security and Drugs (CESED) at Universidad de los Andes for their support during the development of this study.

Data Availability

The leaf and root genome-scale metabolic models, together with files required to reproduce the main analyses reported in this study, are available in the GitHub repository of the Grupo de Diseño de Productos y Procesos under the project “GEMs E. novogranatense”: <https://github.com/Grupo-de-Disenio-de-Productos-y-Procesos/GEMs-E.-novogranatense/tree/main/ModelsScripts> used for model verification and for reproducing the principal analyses are available at: <https://github.com/Grupo-de-Disenio-de-Productos-y-Procesos/GEMs-E.-novogranatense/tree/main/ScriptsMetabolomics-derived> tables supporting the tissue-specific contextualization of the models are provided in the supplementary information files.

References

1. Wink M. Plant Secondary Metabolism: Diversity, Function and its Evolution. *Nat Prod Commun.* 2008;3(8):1934578X0800300801. 10.1177/1934578X0800300801.
2. Elshafie HS, Camele I, Mohamed AA. A Comprehensive Review on the Biological, Agricultural and Pharmaceutical Properties of Secondary Metabolites Based-Plant Origin. *Int J Mol Sci.* 2023;24(4):3266. 10.3390/ijms24043266.

3. Gerlin L, Frainay C, Jourdan F, Baroukh C, Prigent S. Plant genome-scale metabolic networks. In: *Advances in Botanical Research*. 2020. 10.1016/bs.abr.2020.09.021
4. Hurgobin B, Lewsey MG. Applications of cell- and tissue-specific 'omics to improve plant productivity. *Emerg Top Life Sci*. 2022;6(2):163–73. 10.1042/ETLS20210286.
5. Chavez BG, Leite Dias S, D'Auria JC. The evolution of tropane alkaloids: Coca does it differently. *Curr Opin Plant Biol*. 2024;81:102606. 10.1016/j.pbi.2024.102606.
6. Restrepo DA, Saenz E, Jara-Muñoz OA, et al. Erythroxylum in Focus: An Interdisciplinary Review of an Overlooked Genus. *Molecules*. 2019;24(20):3788. 10.3390/molecules24203788.
7. Lv Y, Tian T, Wang YJ, Huang JP, Huang SX. Advances in chemistry and bioactivity of the genus Erythroxylum. *Nat Prod Bioprospecting*. 2022;12(1):15. 10.1007/s13659-022-00338-z.
8. Ochoa VG, R CF. Metabolitos en las hojas de Erythroxylum coca Lam y Erithroxylum novogranatense (Morris) Vieron y evaluacion de sus propiedades biologicas mediante bioensayos. *Rev Peru Med Integrativa*. 2017;2(4):828–34. 10.26722/rpmi.2017.24.70.
9. Wang YJ, Huang JP, Tian T, et al. Discovery and Engineering of the Cocaine Biosynthetic Pathway. *J Am Chem Soc*. 2022;144(48):22000–7. 10.1021/jacs.2c09091.
10. Gomes deO, Dal'Molin C, Quek LE, Saa PA, Nielsen LK. A multi-tissue genome-scale metabolic modeling framework for the analysis of whole plant systems. *Front Plant Sci*. 2015;6. 10.3389/fpls.2015.00004.
11. de Dal'Molin O, Quek CG, Palfreyman LE, Brumbley RW, Nielsen SM. AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. *Plant Physiol*. 2010;152(2):579–89. 10.1104/pp.109.148817.
12. Sampaio M, Rocha M, Dias O. A diel multi-tissue genome-scale metabolic model of Vitis vinifera. *PLOS Comput Biol*. 2024;20(10):e1012506. 10.1371/journal.pcbi.1012506.
13. Rao X, Liu W. A Guide to Metabolic Network Modeling for Plant Biology. *Plants*. 2025;14(3):484. 10.3390/plants14030484.
14. Wang YJ, Tain T, Yu JY, et al. Genomic and structural basis for evolution of tropane alkaloid biosynthesis. *Proc Natl Acad Sci*. 2023;120(17):e2302448120. 10.1073/pnas.2302448120.
15. Aziz RK, Bartels D, Best AA, et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*. 2008;9:75. 10.1186/1471-2164-9-75.
16. Overbeek R, Olson R, Pusch GD, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res*. 2014;42(Database issue):D206–214. 10.1093/nar/gkt1226.
17. Kautsar SA, Suarez Duran HG, Blin K, Osbourn A, Medema MH. plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res*. 2017;45(W1):W55–63. 10.1093/nar/gkx305.
18. Huerta-Cepas J, Forslund K, Coelho LP, et al. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol Biol Evol*. 2017;34(8):2115–22. 10.1093/molbev/msx148.

19. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol.* 2021;38(12):5825–9. 10.1093/molbev/msab293.
20. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 2000;28(1):27–30. 10.1093/nar/28.1.27.
21. Caspi R, Billington R, Ferrer L, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 2016;44(D1):D471–480. 10.1093/nar/gkv1164.
22. Jeske L, Placzek S, Schomburg I, Chang A, Schomburg D. BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res.* 2019;47(D1):D542–9. 10.1093/nar/gky1048.
23. Arkin AP, Cottingham RW, Henry CS, et al. KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol.* 2018;36(7):566–9. 10.1038/nbt.4163.
24. Seaver SMD, Liu F, Zhang Q, et al. The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Res.* 2021;49(D1):D575–88. 10.1093/nar/gkaa746.
25. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COstraints-Based Reconstruction and Analysis for Python. *BMC Syst Biol.* 2013;7(1):74. 10.1186/1752-0509-7-74.
26. Pan S, Reed JL. Advances in gap-filling genome-scale metabolic models and model-driven experiments lead to novel metabolic discoveries. *Curr Opin Biotechnol.* 2018;51:103–8. 10.1016/j.copbio.2017.12.012.
27. Kumar VS, Maranas CD. GrowMatch: An Automated Method for Reconciling In Silico/In Vivo Growth Predictions. *PLOS Comput Biol.* 2009;5(3):e1000308. 10.1371/journal.pcbi.1000308.
28. Reed JL, Patel TR, Chen KH, et al. Systems approach to refining genome annotation. *Proc Natl Acad Sci.* 2006;103(46):17480–4. 10.1073/pnas.0603364103.
29. Tsugawa H, Cajka T, Kind T, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods.* 2015;12(6):523–6. 10.1038/nmeth.3393.
30. Gil-de-la-Fuente A, Godzien J, Saugar S, et al. CEU Mass Mediator 3.0: A Metabolite Annotation Tool. *J Proteome Res.* 2019;18(2):797–802. 10.1021/acs.jproteome.8b00720.
31. Shi H, Wu X, Zhu Y, et al. RefMetaPlant: a reference metabolome database for plants across five major phyla. *Nucleic Acids Res.* 2024;52(D1):D1614–28. 10.1093/nar/gkad980.
32. Dührkop K, Fleischauer M, Ludwig M, et al. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat Methods.* 2019;16(4):299–302. 10.1038/s41592-019-0344-8.
33. Wang M, Carver JJ, Phelan VV, et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol.* 2016;34(8):828–37. 10.1038/nbt.3597.
34. Shinbo Y, Nakamura Y, Altaf-Ul-Amin M, et al. KNApSack: A Comprehensive Species-Metabolite Relationship Database. In: Saito K, Dixon RA, Willmitzer L, editors. *Plant Metabolomics*. Springer;

2006. pp. 165–81. 10.1007/3-540-29782-0_13.
35. Hastings J, Owen G, Dekker A, et al. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.* 2016;44(D1):D1214–9. 10.1093/nar/gkv1031.
36. Kim S, Chen J, Cheng T, et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* 2021;49(D1):D1388–95. 10.1093/nar/gkaa971.
37. Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics Off J Metabolomic Soc.* 2007;3(3):211–21. 10.1007/s11306-007-0082-2.
38. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol.* 2010;28(3):245–8. 10.1038/nbt.1614.
39. Schellenberger J, Que R, Fleming RMT, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc.* 2011;6(9):1290–307. 10.1038/nprot.2011.308.
40. Heirendt L, Arreckx S, Pfau T, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat Protoc.* 2019;14(3):639–702. 10.1038/s41596-018-0098-2.
41. Jaccard P. Étude comparative de la distribution florale dans une portion des Alpes et du Jura. Published online January. 1901;1. 10.5169/SEALS-266450.
42. Megchelenbrink W, Huynen M, Marchiori E, optGpSampler. An Improved Tool for Uniformly Sampling the Solution-Space of Genome-Scale Metabolic Networks. *PLoS ONE.* 2014;9(2):e86587. 10.1371/journal.pone.0086587.
43. Jolliffe IT. Principal Component Analysis. Springer-; 2002. 10.1007/b98835.
44. Poolman MG, Miguet L, Sweetlove LJ, Fell DA. A genome-scale metabolic model of Arabidopsis and some of its properties. *Plant Physiol.* 2009;151(3):1570–81. 10.1104/pp.109.141267.
45. Cheung CYM, Poolman MG, Fell DA, Ratcliffe RG, Sweetlove LJ. A Diel Flux Balance Model Captures Interactions between Light and Dark Metabolism during Day-Night Cycles in C3 and Crassulacean Acid Metabolism Leaves. *Plant Physiol.* 2014;165(2):917–29. 10.1104/pp.113.234468.
46. de Dal’Molin O, Quek CG, Palfreyman LE, Brumbley RW, Nielsen SM. C4GEM, a Genome-Scale Metabolic Model to Study C4 Plant Metabolism. *Plant Physiol.* 2010;154(4):1871–85. 10.1104/pp.110.166488.
47. Yuan H, Cheung CYM, Poolman MG, Hilbers PAJ, van Riel NAW. A genome-scale metabolic network reconstruction of tomato (*Solanum lycopersicum* L.) and its application to photorespiratory metabolism. *Plant J Cell Mol Biol.* 2016;85(2):289–304. 10.1111/tpj.13075.
48. de Dal’Molin O, Orellana CG, Gebbie C. Metabolic Reconstruction of *Setaria italica*: A Systems Biology Approach for Integrating Tissue-Specific Omics and Pathway Analysis of Bioenergy Grasses. *Front Plant Sci.* 2016;7. 10.3389/fpls.2016.01138.
49. Schmidt GW, Jirschitzka J, Porta T, et al. The last step in cocaine biosynthesis is catalyzed by a BAHD acyltransferase. *Plant Physiol.* 2015;167(1):89–101. 10.1104/pp.114.248187.

50. Costas d, Maranas. Ali r. zomorodi. *Optimization Methods in Metabolic Networks*. Wiley. Accessed May 22, 2026. <https://www.wiley.com/en-us/Optimization+Methods+in+Metabolic+Networks-p-9781119028499>
51. Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng*. 2003;5(4):264–76. 10.1016/j.ymben.2003.09.002.
52. Bordel S, Agren R, Nielsen J. Sampling the Solution Space in Genome-Scale Metabolic Networks Reveals Transcriptional Regulation in Key Enzymes. *PLOS Comput Biol*. 2010;6(7):e1000859. 10.1371/journal.pcbi.1000859.
53. Paponov M, Antonyan M, Slimestad R, Paponov IA. Decoupling of Plant Growth and Accumulation of Biologically Active Compounds in Leaves, Roots, and Root Exudates of *Hypericum perforatum* L. by the Combination of Jasmonate and Far-Red Lighting. *Biomolecules*. 2021;11(9):1283. 10.3390/biom11091283.

Figures

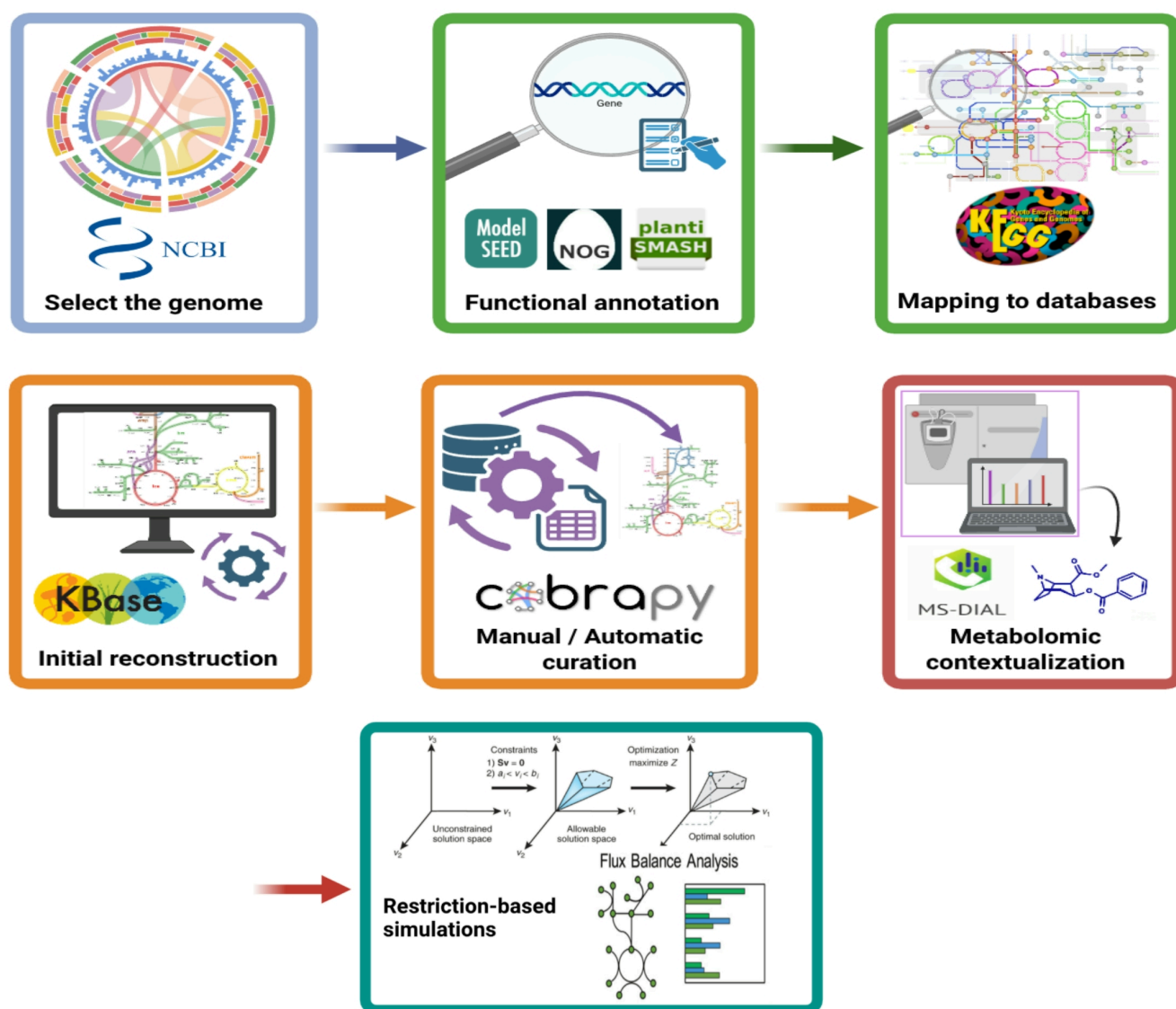


Figure 1

workflow for genome-scale reconstruction, curation and tissue-specific analysis of *E. novogranatense* GEMs

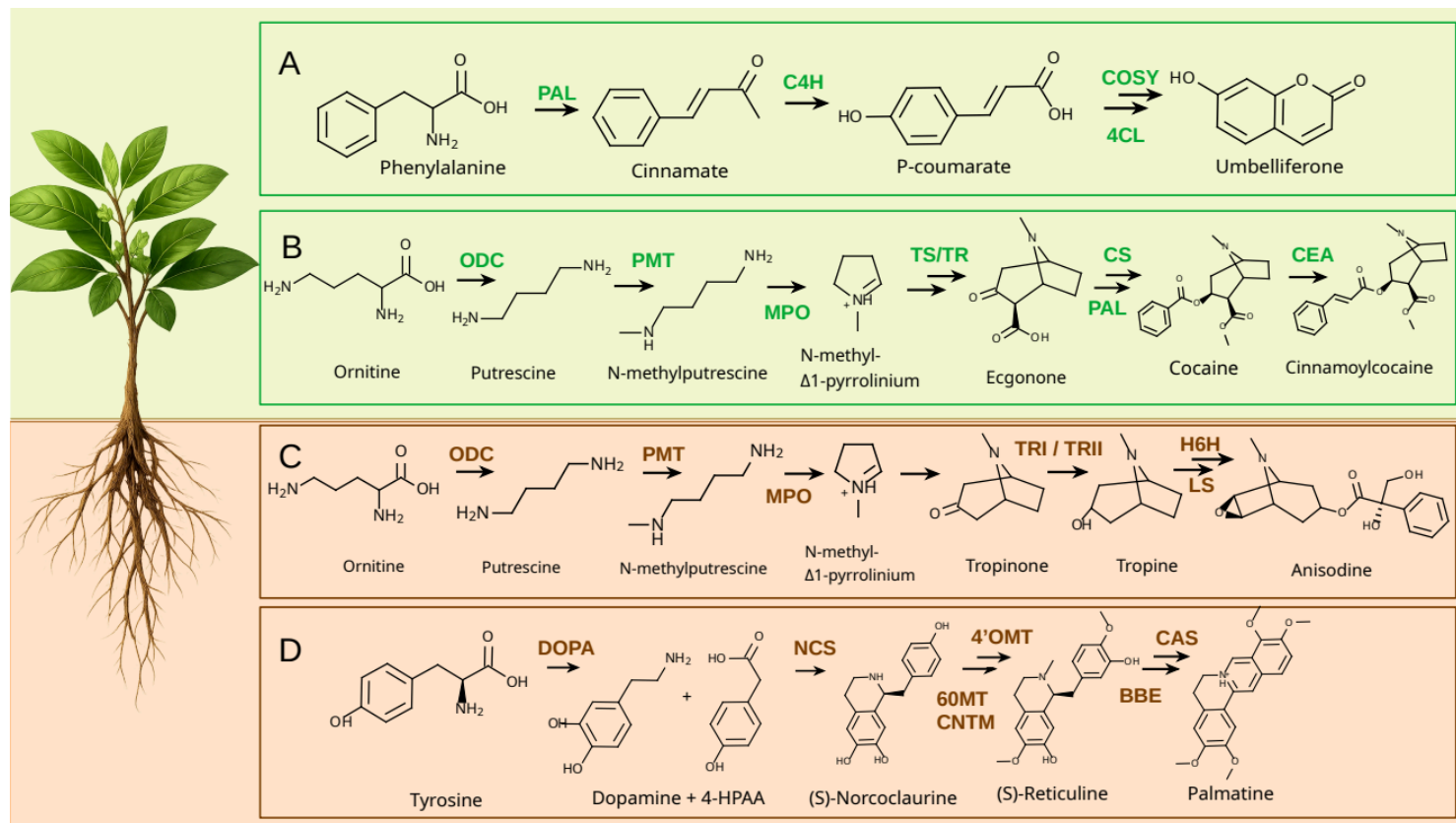


Figure 2

Representative biosynthetic routes to tissue-specific metabolites in leaf and root: Illustrative schematic biosynthetic pathways to differential metabolites in leaf and root: Schematic pathways (panels A–D) that contextualize differential metabolites used as tissue markers and their conceptual connection to central/specialized metabolism: (A) Umbelliferone (schematic / simplified route) (phenylpropanoids/coumarins; leaf), (B) Cinnamoylcocaine (acylated tropanes; leaf), (C) Anisodine (tropanes; root), and (D) Palmatine (conceptual/schematic pathway) (protoberberine-type isoquinoline alkaloids; root); Abbreviations of enzymes/reactions in the scheme: PAL (phenylalanine ammonia-lyase), C4H (cinnamate 4-hydroxylase), 4CL (4-coumarate:CoA ligase), COSY (coumarin synthase); ODC (ornithine decarboxylase), PMT (putrescine N-methyltransferase), MPO (N-methylputrescine oxidase), TS/TR (Tropinone synthase /Tropinone reductase), CS (Cocaine synthase), CEA (Cinnamoyl-CoA:ecgonine acyltransferase), TRI/TRII (tropinone reductase I / II), H6H (hyoscyamine 6 β -hydroxylase), LS (BAHD acyltransferase), DOPA (step to dopamine; decarboxylation in the dopaminergic branch), NCS (norcoclaurine synthase), 4'OMT (3'-hydroxy-N-methylcoclaurine 4'-O-methyltransferase), 6OMT (Norcoclaurine 6-O-methyltransferase), CNMT (Coclaurine N-methyltransferase), BBE (berberine bridge enzyme), CAS (Canadine Synthase / CYP719A).

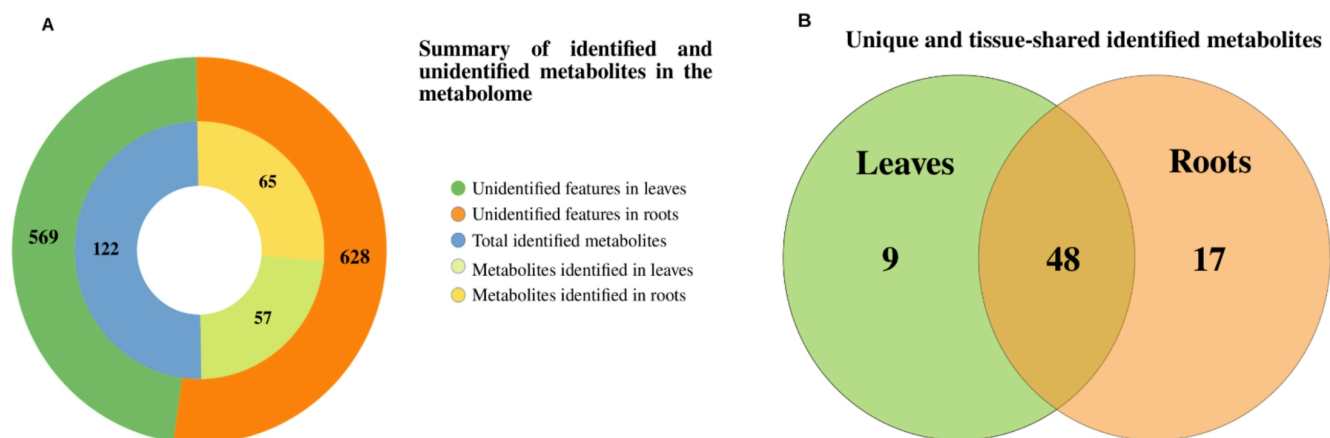


Figure 3

Annotated metabolite overview and tissue-specific distribution in *E. novogranatense*: Overview of annotated metabolites and tissue-specific metabolic distribution in *E. novogranatense*, A: Summary of identified and unidentified metabolites in leaves and roots, B: Tissue-specific and shared identified metabolites between leaves and roots

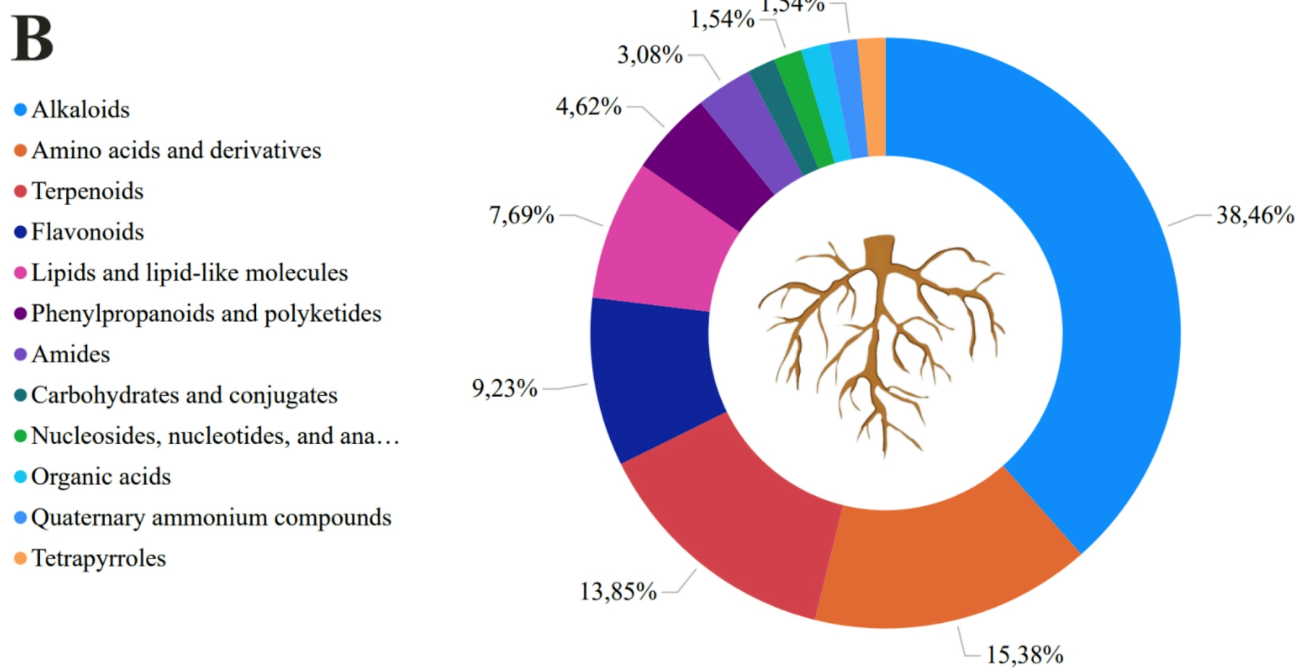
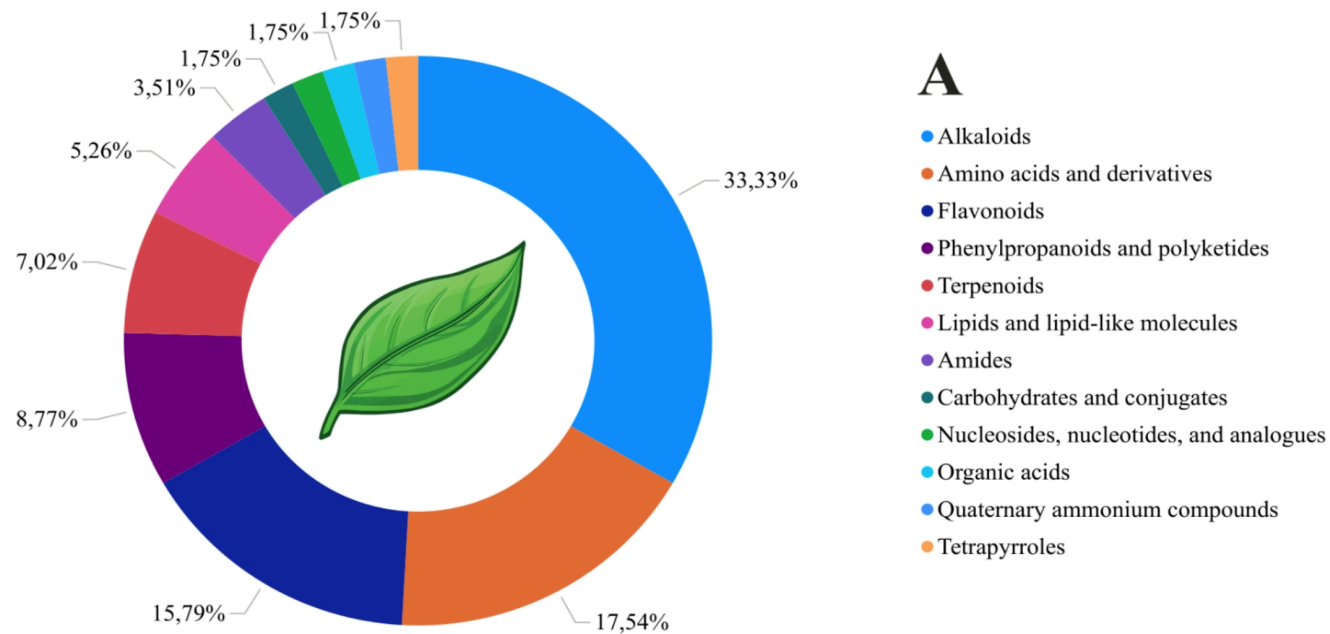


Figure 4

Chemical classification of annotated metabolites in leaf and root tissues: A) Metabolites annotated in leaves, B) Annotated in roots.

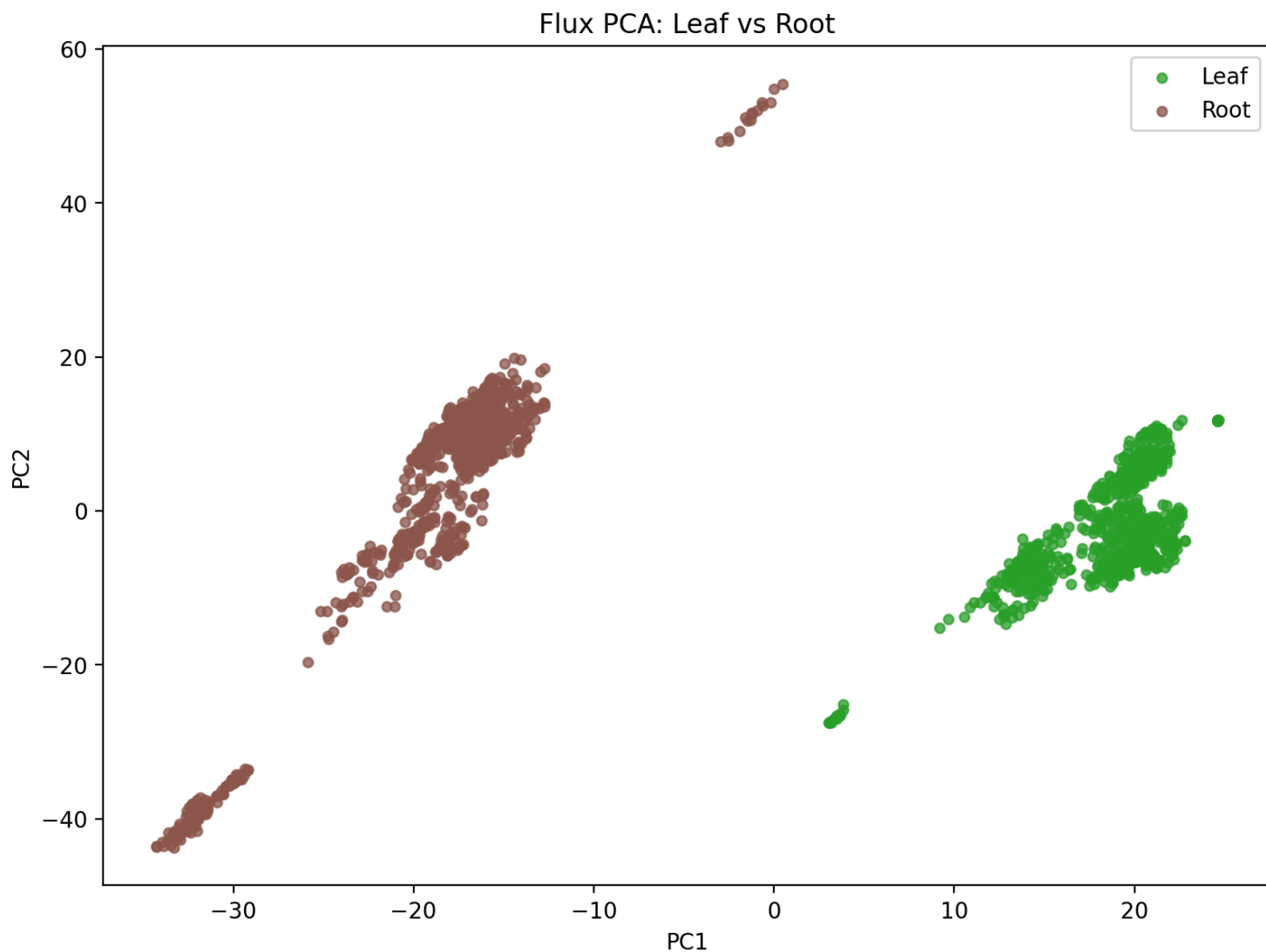


Figure 5

Flux-based separation of leaf and root GEMs: Principal component analysis of flux distributions in the leaf and root genome-scale metabolic models. Each point corresponds to a flux sample obtained under tissue-specific medium conditions and a common relative biomass constraint equivalent to 90% of the tissue-specific Bopt.

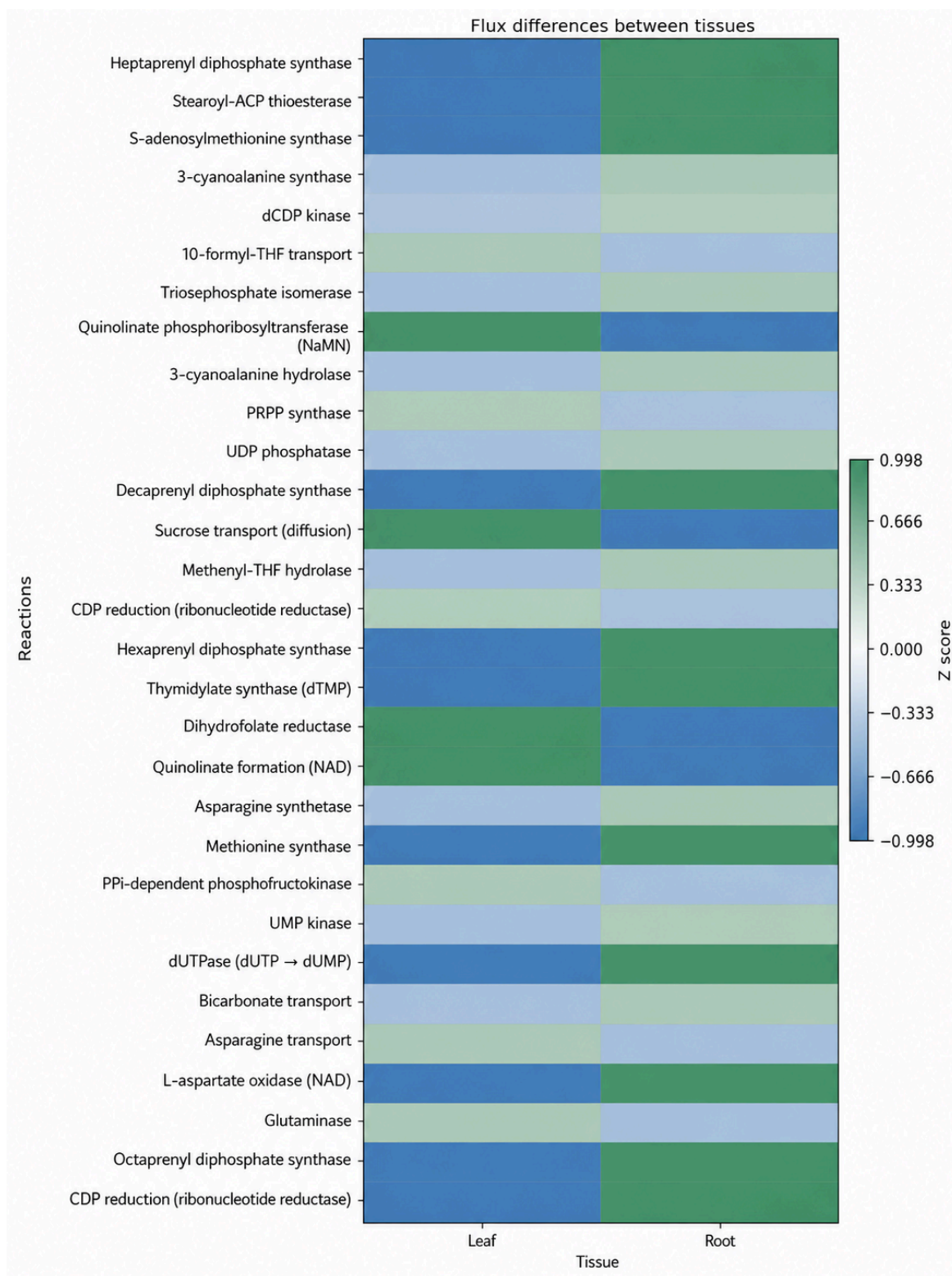


Figure 6

Standardized flux differences between leaf and root reactions in GEMs: Heatmap of mean reaction z-scores for a subset of shared internal reactions, revealing substantial differences in flux behavior between leaf and root tissues. Rows correspond to reactions, and columns correspond to the leaf and root models.

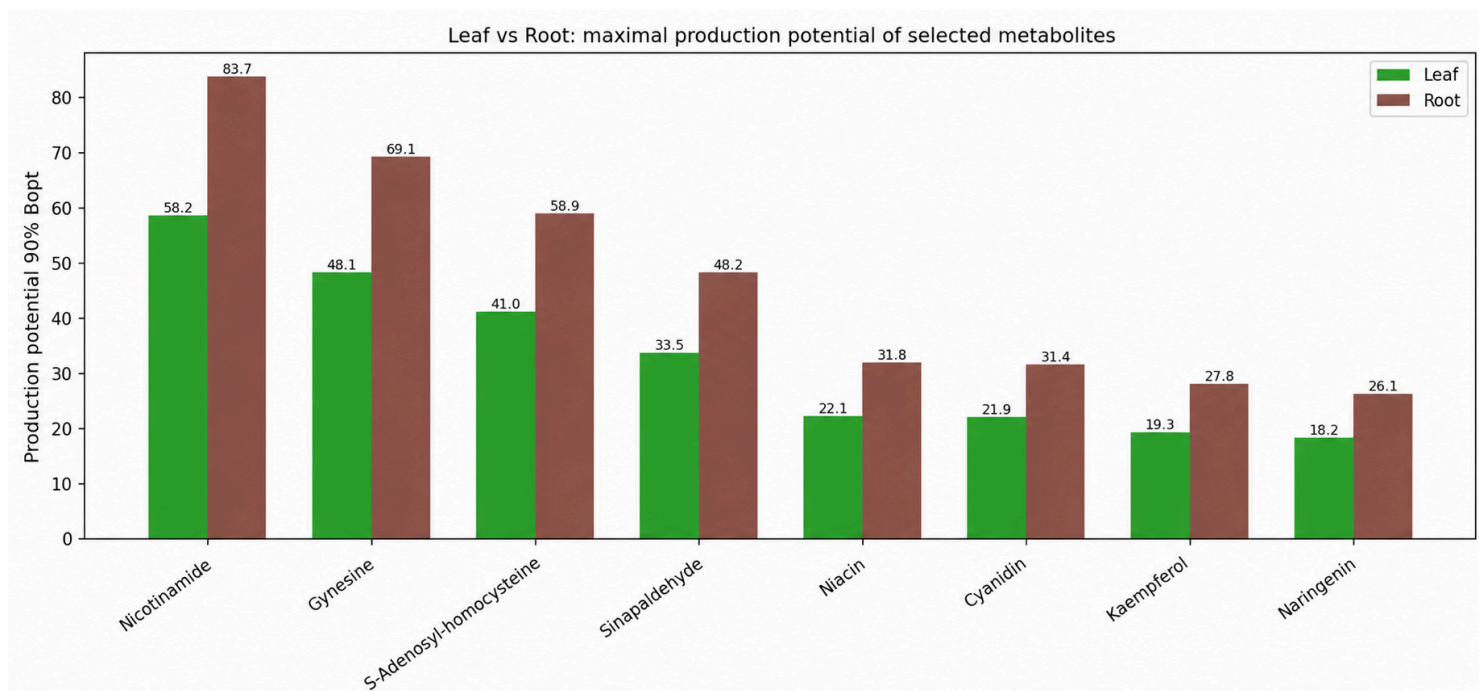


Figure 7

Production potential of selected shared metabolites. Maximum flux through demand reactions for selected metabolites represented in both tissue models, calculated under tissue-specific media while constraining biomass to 90% of the tissue-specific Bopt.

Supplementary Files

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

- [TableS1.xlsx](#)