Specialty Journal of Pharmacognosy, Phytochemistry, and Biotechnology
ISSN: 3062-441X
2025, Volume 5, Issue 2, Page No: 1-10

Available online at: www.galaxypub.co/page/journals

Galaxy Publication

Inferring Medicinal-Plant Biosynthetic Pathways Without a Single Omics Layer Dominating the Answer: A Multi-Omics Evidence Architecture for Gene–Metabolite Assignment

Wouter De Smet[1]*, Liesbeth Van Dam[1], Pieter Janssens[2]

[1] Department of Multi-Omics Biosynthetic Pathway Inference, Faculty of Pharmaceutical Sciences, KU Leuven, Leuven, Belgium.

[2] Department of Gene–Metabolite Assignment, Faculty of Pharmacy, Ghent University, Ghent, Belgium.

***E-mail** ✉ wouter.desmet@kuleuven.be

Received: 14 April 2025; **Revised:** 21 July 2025; **Accepted:** 25 July 2025

ABSTRACT

Medicinal-plant pathway discovery increasingly combines genome assemblies, gene-expression profiles, metabolite measurements, spatial information, and functional assays, yet the presence of multiple data layers does not itself resolve which gene produces, modifies, transports, or regulates a given metabolite. This article develops an original, evidence-bounded multi-omics architecture for gene–metabolite assignment in specialized metabolism. The analysis separates candidate nomination from biological and chemical constraint, conflict diagnosis, and functional resolution. It argues that genomics is most informative for candidate space, locus organization, duplication, and evolutionary context; transcriptomics for conditional coordination and cellular context; metabolomics for chemical occurrence and transformation plausibility; and spatial or paired-cell measurements for localization and cross-modality alignment. Agreement among layers strengthens an assignment only when the observations are sufficiently independent and mechanistically compatible. Conversely, disagreement should not automatically invalidate a candidate because transport, compartmentation, paralogy, pathway branching, analytical incompleteness, and reference-genome defects can generate legitimate discordance. The proposed architecture therefore treats confidence as claim-specific and conditional on evidence quality, orthogonality, compatibility, and the degree of functional testing, rather than on the number of omics layers supporting a candidate. Its intended output is not a universal score but an explicit, revisable statement of what is supported, what remains alternative, and what validation would be most discriminating. The framework is limited by uneven medicinal-plant resources, incomplete metabolite identification, context-dependent measurements, and the absence of prospective benchmarking across diverse pathways.

Keywords: Medicinal plants, Specialized metabolism, Multi-omics, Biosynthetic pathways, Gene–metabolite assignment, Evidence integration

How to Cite This Article: De Smet W, Van Dam L, Janssens P. Inferring Medicinal-Plant Biosynthetic Pathways Without a Single Omics Layer Dominating the Answer: A Multi-Omics Evidence Architecture for Gene–Metabolite Assignment. Spec J Pharmacogn Phytochem Biotechnol. 2025;5(2):1-10. https://doi.org/10.51847/Zhr4qwpUq8

Introduction

Medicinal plants produce structurally diverse specialized metabolites through pathways that can be distributed across duplicated enzyme families, tissues, cell types, and subcellular compartments. Pathway discovery has accordingly moved beyond isolated gene annotation toward combinations of genomics, transcriptomics, metabolomics, and computational or paired-omics strategies. Current methodological syntheses converge on a central point: these layers provide complementary rather than interchangeable evidence for pathway nomination and assignment [1–3]. The practical difficulty is therefore no longer simply obtaining more data. It is deciding what each observation is entitled to mean when several observations concern the same candidate relationship.

This distinction matters because a familiar high-confidence-looking pattern can still be evidentially narrow. Coexpression may strongly prioritize genes whose transcription follows a pathway, but it does not by itself

establish substrate specificity, reaction order, transport, or direct catalytic participation [4]. Likewise, genome proximity, metabolite covariance, or shared tissue enrichment can all be biologically informative without answering the same biochemical question. Treating these signals as interchangeable risks allowing the most abundant or statistically convenient layer to dominate an inference that actually requires several different kinds of constraint.

Medicinal-plant single-cell multi-omics illustrates both the opportunity and the problem. In *Catharanthus roseus*, cell-resolved evidence helps localize pathway-associated genes and interpret pathway organization, yet the same system shows that transport and intercellular partitioning can separate the site of transcript accumulation from the location of downstream metabolites [5]. Apparent noncolocalization may therefore be informative rather than contradictory. More generally, the interpretation of agreement depends on whether the observations concern the same biochemical event, whereas the interpretation of disagreement depends on whether plausible spatial, temporal, evolutionary, or analytical explanations remain.

This article proposes a multi-omics evidence architecture for gene–metabolite assignment. Its contribution is conceptual rather than a new prediction algorithm: candidate relationships are evaluated according to the evidential role of each observation, the quality and dependence of supporting data, the biological explanations available for concordance or conflict, and the strength of subsequent functional resolution. No omics layer is granted default priority. Instead, the architecture asks whether genomic, transcriptional, chemical, spatial, evolutionary, and functional evidence constrain different parts of the same biosynthetic claim and whether remaining alternatives are explicit enough to guide discriminating validation. The objective is consequently not to maximize multimodal agreement, but to make the logic connecting observation, assignment, uncertainty, and validation inspectable.

Why gene–metabolite assignment remains uncertain

The most scalable discovery signals are often associative. Coexpression networks can connect genes to specialized-metabolic programs, while plant biosynthetic-gene-cluster analyses and genome-mining tools can identify loci whose organization is consistent with pathway function. These approaches are valuable for narrowing candidate space, but correlation or predicted cluster membership alone does not constitute functional assignment [6–8]. Their appropriate evidential role is nomination: they identify relationships worth investigating without establishing that a particular enzyme catalyzes a particular transformation.

Sequence- and annotation-based inference introduces a second uncertainty. Genome-wide predictions can assign putative enzyme classes, pathways, or clusters across plant genomes, but the resulting candidate space inherits uncertainty from annotation quality, paralogy, and the extent to which existing sequence–function knowledge represents the focal chemistry [9]. This distinction is particularly important in specialized metabolism, where closely related enzymes may differ in substrate range, reaction position, stereochemical outcome, or product spectrum.

Genome organization is also less stable than a simple contiguous-cluster model implies. Comparative work in Lamiaceae shows that a validated specialized-metabolism cluster can follow a dynamic evolutionary trajectory, so physical proximity should be interpreted together with phylogenetic and functional evidence rather than as an invariant pathway signature [10]. Conversely, functionally connected genes can occupy noncanonical arrangements: a bipartite cluster in *Schizonepeta tenuifolia* demonstrates that a fixed cluster template can miss a valid biosynthetic organization [11].

Duplicated enzymes add gene-level ambiguity. In *Isodon rubescens*, tandem CYP706V genes, tissue context, and functional evidence were needed to resolve oxidase roles in oridonin biosynthesis; duplication or high expression alone would not justify assigning every paralog to the same reaction [12]. At a broader evolutionary scale, medicinal skullcap genomes show that related clerodane diterpene biosynthesis can have polyphyletic origins [13]. Homology therefore supplies plausibility rather than an automatic transfer of pathway function. Uncertainty is consequently structural: several candidates may satisfy one evidential criterion while remaining distinguishable under another.

What each omics layer contributes

Genomics establishes the candidate landscape within which other evidence is interpreted. The opium-poppy genome connected large-scale genome organization and evolutionary events with morphinan biosynthesis, illustrating how locus structure can clarify the history and arrangement of medicinal-alkaloid genes [14]. In *Scutellaria baicalensis*, combining the reference genome with expression and pathway evidence refined

understanding of lineage-specific wogonin biosynthesis beyond annotation alone [15]. *Coptis chinensis* similarly connected gene-family diversification with protoberberine alkaloid diversity while leaving individual paralog function as a more specific evidentiary question [16]. The *Taxus* genome narrowed candidate space for paclitaxel biosynthesis without making every reaction step functionally resolved [17]. Genomic evidence is therefore especially informative about what genes are available, how they are organized, and how plausible candidates arose.

Transcriptomics contributes conditional coordination. Tissue-resolved *Taxus* analysis connected a phloem-specific transcriptional regulator with paclitaxel-pathway regulation, demonstrating why tissue context can be more informative than bulk abundance; nevertheless, transcript level is not equivalent to metabolic flux [18]. Transcript evidence is consequently strongest when it constrains where, when, or under which biological conditions candidate genes participate in a coordinated program.

Metabolomic and spatial observations address different uncertainties. Integrated mass-spectrometry imaging and single-cell transcriptomics in *Taxus mairei* associated taxoid distributions with cell states and transport contexts, showing that spatial mismatch can reflect metabolite movement rather than erroneous candidate selection [19]. Chemical-tag-based semi-annotated metabolomics further demonstrates that informative chemical features can facilitate gene discovery before complete metabolite identification, provided chemical-annotation uncertainty remains explicit [20]. Spatial multi-omics in medicinal plants can consequently add tissue and cellular constraints while remaining sensitive to resolution and modality matching [21].

Even direct pairing of modalities does not erase the inferential boundary. Single-cell metabolome and RNA-seq multiplexing reduces cross-sample alignment uncertainty by measuring transcripts and metabolites within the same plant cell [22]. Yet same-cell association remains distinct from direct catalysis. In the proposed architecture, genomics primarily constrains possibility and organization, transcriptomics constrains conditional coordination, metabolomics constrains chemical occurrence and transformation plausibility, and spatial or paired measurements constrain localization and cross-modality alignment. Their value lies in different questions answered, not in competing for a single position in an evidence hierarchy.

Agreement and conflict between evidence types

Agreement among evidence types is meaningful only after the quality of each contributing observation is considered. A gap-free *Scutellaria baicalensis* genome exposed gene-family and pathway information that can be obscured by fragmented assemblies, showing that reference completeness changes the candidate evidence available for downstream inference [23]. A telomere-to-telomere *Chaenomeles speciosa* genome independently demonstrates how reducing structural blind spots can strengthen candidate discovery while leaving biochemical assignment as a separate evidentiary step [24]. Agreement built on incomplete genomic representation should therefore not be treated as equivalent to agreement built on a well-resolved locus.

Conflict likewise has more than one interpretation. Regulation, subcellular organization, transport, and pathway plasticity can produce genuine nonconcordance among transcript, protein, and metabolite patterns [25]. Technical effects can generate superficially similar patterns. A metabolite detected away from a candidate transcript may have moved; an apparently absent transcript may derive from dilution of a rare cell type; a candidate gene may be absent because the reference is incomplete; and chemically unresolved features may incorrectly appear to disagree with a genomic hypothesis. Down-weighting the discordant evidence before diagnosing these alternatives would therefore confuse inconsistency with falsification.

A useful organizing distinction is to ask what question each observation actually answers. Modern specialized-metabolism research increasingly separates chemical identity, spatial location, biosynthetic mechanism, and biological context rather than expecting one modality to resolve all four [26]. In the architecture proposed here, agreement means that evidence streams make compatible claims at their appropriate analytical levels. Conflict means either that claims genuinely compete or that observations have been compared across different levels, compartments, conditions, or resolutions. The distinction prevents agreement from becoming a vote count and prevents disagreement from becoming an automatic rejection criterion.

Table 1 organizes these states by evidential function and shows where apparently similar support carries different interpretive limits.

Table 1. Evidence roles, concordance states, and interpretation boundaries in medicinal-plant gene–metabolite assignment

Evidence state	Principal support	Main unresolved ambiguity	Role in the proposed architecture
Genomic plausibility	Candidate repertoire, locus organization, duplication, evolutionary context	Annotation error; paralog ambiguity	Nomination and structural constraint
Transcriptional coordination	Condition-, tissue-, or cell-linked pathway participation	Regulation without direct catalysis	Contextual support
Chemical association	Metabolite occurrence and plausible transformations	Incomplete identity; indirect coupling	Chemical constraint
Spatial alignment	Colocalization or interpretable separation	Transport; compartmentation; resolution mismatch	Localization constraint
Cross-layer concordance	Compatible observations from distinct evidential roles	Shared bias or dependence	Strengthens an assignment when dependencies remain explicit
Cross-layer discordance	Competing or level-mismatched observations	Biological separation versus measurement error	Requires conflict diagnosis before rejection
Functional resolution	Reaction or pathway behavior under tested conditions	Native necessity, regulation, transport, or flux may remain	Stronger assignment state without universal certainty

The resulting logic is better represented as a multilevel architecture than as a linear sequence, because uncertainty sources can alter several evidence types simultaneously and because disagreement can trigger reinterpretation rather than simple candidate elimination. **Figure 1** therefore depicts the manuscript's proposed separation of evidence contribution, reconciliation, confidence interpretation, and pathway-reconstruction boundaries.

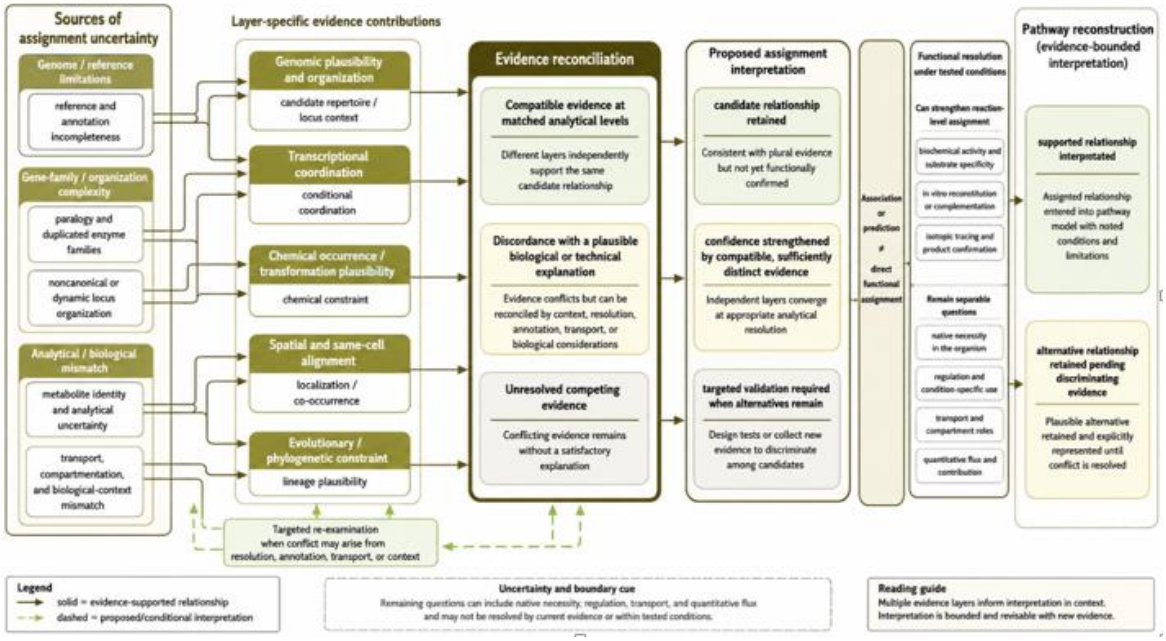

Figure 1. From layer-specific observations and discordance to evidence-bounded gene–metabolite assignment

Integrating evidence without treating correlation as causation

The key transition in pathway assignment is from an observation that nominates a relationship to evidence that tests the biochemical proposition embedded in it. In *Catharanthus roseus*, discovery of previously missing vinblastine-pathway enzymes combined candidate identification with experimental demonstration of the required transformations, providing reaction-level evidence that coexpression or genomic association alone could not supply [27]. The evidential gain lies not simply in adding another data type, but in asking a different question: whether the encoded protein can perform the chemically required reaction under defined conditions.

Independent completion of the canonical vincristine/vinblastine pathway reinforces this distinction. Chemically coherent intermediates and experimentally demonstrated enzyme functions can resolve pathway steps that remain ambiguous when candidates are ranked primarily by association [28]. Functional evidence is nevertheless conditional. Activity in an assay or reconstructed system establishes what occurred under the tested conditions; it need not establish native quantitative flux, regulatory importance, or exclusivity of a route in every tissue or state. Recent paclitaxel work makes this boundary especially clear. Synthetic reconstruction has been used to identify a minimal gene set, reconstitution has exposed networked early-pathway chemistry, and heterologous characterization has resolved enzymes leading toward baccatin III [29–31]. These experiments can test sufficiency, reaction order, product formation, and pathway branching in ways that observational co-variation cannot. Their strength should not be diluted by calling them merely "another omics layer," nor inflated into proof of every native-system property.

The subsequent discovery of additional Taxol-pathway genes followed by reconstructed baccatin III biosynthesis further illustrates that orthogonal nomination becomes more persuasive when candidate genes survive functional and chemical testing [32]. The proposed architecture therefore does not convert multimodal correlation into causality by accumulation. It uses associative evidence to narrow alternatives, mechanistic compatibility to constrain them, and functional experiments to resolve the specific claims those experiments actually test.

This separation also prevents a common category error between necessity, sufficiency, and capability. An enzyme may catalyze a reaction in vitro yet not be necessary in planta because another paralog compensates; a reconstructed gene set may be sufficient to produce an intermediate in a chassis without reproducing the native route; and loss of a candidate may alter a metabolite indirectly through regulation or transport. The proposed interpretation therefore attaches causal language only to the intervention and endpoint actually examined. Stronger evidence changes the scope of an assignment, not merely its position on a single linear confidence scale.

A multi-omics assignment architecture

The proposed architecture begins by separating evidence according to function rather than technology. A candidate may enter through genomic proximity, sequence-function prediction, coexpression, chemical covariance, spatial association, phylogenetic conservation, or another justified observation. At entry, the evidence is tagged conceptually by its provenance, analytical level, and main vulnerability. Curated experimentally supported enzyme knowledge can provide a useful prior for interpreting a candidate, but such reference knowledge remains different from validation in the focal medicinal plant [33]. The architecture therefore treats prior annotation as context, not inherited certainty.

Candidate evidence then passes through three linked assessments. First, a quality assessment asks whether the underlying genome, annotation, metabolite identity, spatial resolution, or expression measurement is adequate for the intended claim. Second, a compatibility assessment asks whether observations support the same biochemical proposition at comparable biological scales. Third, a dependence assessment asks whether apparently independent signals could share a common source—for example, whether two transcript-derived analyses are effectively repeating one association. These assessments do not produce numerical weights. They identify why support should be strengthened, qualified, or re-examined.

A further safeguard is claim decomposition. "Gene X belongs to pathway Y" is too broad to evaluate as one proposition because sequence identity, expression, localization, catalytic activity, pathway necessity, and contribution to a final metabolite are different claims. The proposed architecture therefore decomposes a candidate assignment into the smallest biologically meaningful propositions supported by the available evidence. Evidence can strengthen one proposition while leaving another open. This prevents a validated catalytic step from silently validating localization or flux, and prevents a strong genomic prior from being reported as an established biochemical role.

Evidence reconciliation follows. Compatible observations from different roles can narrow plausible assignments; explainable discordance can trigger a targeted reinterpretation; unresolved conflict preserves competing candidates. Systems and synthetic biology increasingly connect candidate discovery with iterative functional testing rather than a one-way ranking cascade [34]. The architecture adopts that iterative logic as a proposed evidential loop: validation can resolve a candidate, reveal an incorrect premise, expose pathway branching, or return the analysis to upstream evidence.

The distinction between data pairing and functional resolution remains explicit. Cell-resolved medicinal-plant multi-omics can add powerful localization and pathway context [5]. Same-cell transcript–metabolite measurement

can reduce alignment uncertainty [22]. Neither observation is made equivalent to a demonstrated biochemical transformation. Conversely, pathway reconstruction can establish strong reaction-level evidence without answering every question about native regulation, transport, or necessity [32]. Thus, no evidence type receives an automatic veto or priority. The architecture asks what uncertainty it removes and what uncertainty it leaves.

Table 2 compares the proposed interpretive states that emerge from this architecture without converting them into a pseudo-quantitative score.

Table 2. Proposed analytical states for integrating gene–metabolite evidence without layer dominance

Analytical state	Evidence pattern	Interpretation	Next evidential action
Nomination	One or more plausible associative signals	Candidate worth retaining	Seek an independent constraint
Constrained candidate	Compatible genomic, chemical, spatial, or evolutionary context	Alternative space narrowed	Test remaining discriminators
Explainable discordance	Mismatch with plausible transport, compartment, context, or measurement explanation	Candidate not rejected automatically	Test the explanation
Unresolved conflict	Competing observations at matched levels	Assignment remains provisional	Acquire discriminating evidence
Functionally resolved step	Required reaction demonstrated under defined conditions	Strong reaction-level support	Evaluate native context where necessary
Reconstruction-ready relation	Compatible evidence plus adequate functional resolution for the stated reconstruction purpose	Incorporate with explicit boundaries	Retain unresolved alternatives separately

Figure 2 makes the architecture operational as a nested evidence ledger in which support, conflict, dependence, and validation remain visible rather than collapsing into a single confidence value.

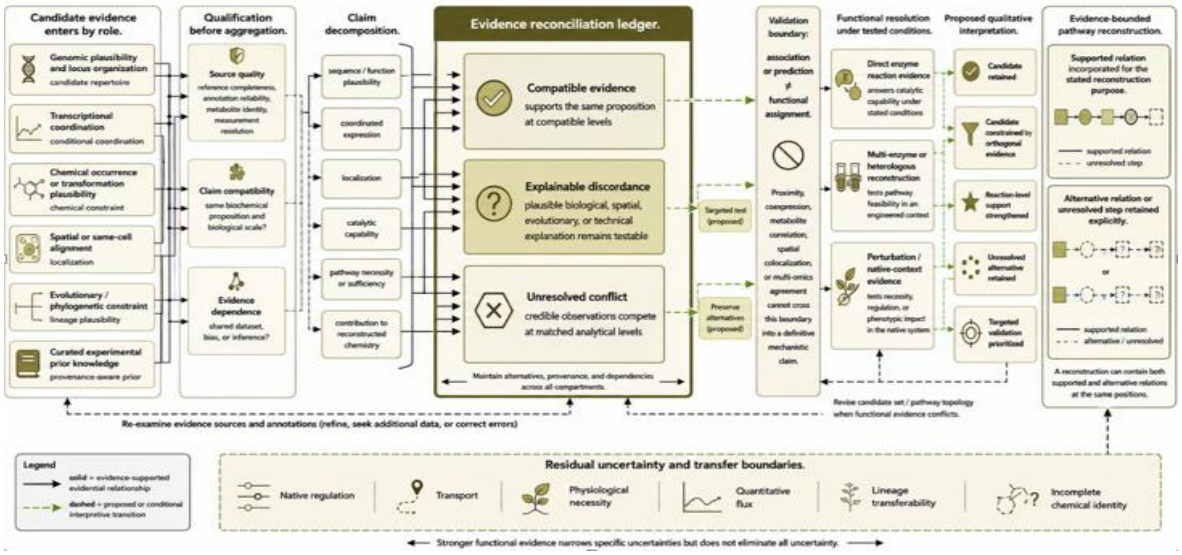

Figure 2. A proposed evidence ledger from candidate nomination to bounded pathway reconstruction

Confidence in candidate biosynthetic relationships

Confidence in this framework is claim-specific. A study may provide strong evidence that a plant contains a chemical scaffold, persuasive evidence for candidate biosynthetic genes, and separate evidence for biological activity, yet those endpoints are not interchangeable. Genome-enabled mining of cepharanthine analogs in *Stephania* exemplifies why biosynthetic and pharmacological observations must remain analytically distinct [35]. Bioactivity can increase the importance of resolving a pathway, but it does not validate which gene generated the metabolite.

The framework therefore defines confidence qualitatively through the conjunction of evidence quality, mechanistic compatibility, independence, and functional resolution. Comparative work on CYP80 expansion and benzylisoquinoline-alkaloid diversity shows how phylogenetic, chemical, and enzyme-functional observations can converge around a biosynthetic interpretation [36]. Such convergence is more informative than repeated

observations derived from the same underlying signal, but it still does not establish a universal numerical threshold for assignment.

Functional evidence changes confidence because it tests a different proposition. Direct enzyme characterization can support a specified reaction [27]. Chemically resolved heterologous reconstruction can test multi-enzyme pathway behavior [31]. Discovery followed by reconstructed pathway chemistry can further reduce uncertainty around missing components [32]. The proposed confidence language should therefore state what has been shown—association, contextual compatibility, catalytic capability, reconstructed sufficiency, or another defined property—rather than labeling a candidate simply "high confidence." When evidence streams are indirect, dependent, or vulnerable to the same failure mode, apparent convergence should be qualified. When functional evidence is strong but native-system questions remain, confidence should be narrow rather than globally downgraded or inflated.

Confidence is consequently multidimensional rather than ordinal. A relationship can be strong with respect to chemical plausibility but weak with respect to native localization, or strong for catalytic capability but unresolved for physiological necessity. The architecture therefore favors short claim-specific confidence statements over labels that merge these dimensions. It also treats contradictory evidence asymmetrically: a technically credible observation that directly contradicts the proposed reaction may deserve greater attention than several weak supporting correlations, whereas a spatial mismatch with a plausible transport explanation may call for targeted localization or perturbation rather than immediate rejection. This proposed logic gives conflict diagnostic value without allowing one discordant modality to become an automatic veto.

Implications for pathway reconstruction

A pathway reconstructed from multi-omics evidence should be represented as an evidence-bounded model, not as a finalized sequence merely because every step has a plausible candidate. Comparative phylogenomics can constrain which pathway histories and gene–metabolite relationships are evolutionarily plausible, while remaining distinct from direct demonstration of enzyme activity in each extant species [37]. Evolutionary evidence is therefore useful for pruning implausible assignments and identifying where lineage-specific alternatives deserve explicit representation.

Duplication and clustering further argue against forcing one-to-one pathway diagrams when the evidence supports diversification. Recent analysis of benzylisoquinoline-alkaloid biosynthesis shows how gene duplication and clustering can contribute both conservation and innovation across plant lineages [38]. In reconstruction, paralogs should therefore remain distinguishable until evidence warrants collapsing them into equivalent roles, and clustered organization should be treated as contextual support rather than a requirement for pathway membership.

The proposed architecture also makes uncertainty productive. Dynamic cluster evolution can justify retaining structurally different candidate arrangements rather than privileging the most contiguous locus [10]. Polyphyletic pathway origins can justify lineage-specific alternatives even when metabolites are chemically related [13]. Evidence for biological activity or another downstream endpoint should not be allowed to resolve a biosynthetic ambiguity that it did not test [35]. A mature reconstruction may consequently contain a supported core alongside explicitly provisional branches, alternative enzymes, or unresolved transport steps. This representation is less visually definitive than a single pathway map, but it is more faithful to the evidence and more useful for deciding which experiment would most efficiently discriminate among remaining possibilities.

The same principle applies to pathway topology. Omics associations often encourage a single ordered chain because a linear diagram is easy to communicate, yet evidence may instead support parallel oxidation sequences, promiscuous enzymes, transport-dependent handoffs, or several candidates for one transformation. The proposed architecture therefore distinguishes a reconstruction hypothesis from a validated reaction sequence. Edges can be retained because they are chemically and biologically compatible while still carrying an explicit need for reaction-level testing. This makes unresolved structure visible instead of burying it in narrative caveats and prevents downstream metabolic engineering or comparative interpretation from inheriting an unjustified certainty.

Validation priorities and limitations

Validation should target the uncertainty most capable of changing the assignment. For medicinal plants, that process can begin upstream: incomplete assemblies and unreliable annotations may remove, fragment, or misplace candidate genes before any integrative analysis begins [39]. Improving reference quality is therefore a prerequisite for trustworthy genomic evidence, not a substitute for biochemical validation.

Analytical uncertainty also requires explicit escalation. Semi-annotated metabolomic features can be valuable for gene discovery, but their structural uncertainty should remain visible until the chemical claim is supported at the necessary level [20]. Similarly, same-cell transcript–metabolite pairing reduces alignment ambiguity but remains observational; causal assignment may require perturbation, enzyme characterization, or another discriminating functional test [22].

Functional reconstruction is powerful but bounded. Reconstructed baccatin III biosynthesis demonstrates how candidate genes can be tested jointly and chemically resolved [32]. Yet native necessity, tissue-specific regulation, transport, and quantitative flux may remain open, while the quality of the medicinal-plant reference genome still constrains what candidates were considered initially [39]. The proposed architecture has not been prospectively benchmarked across pathway classes, and its qualitative confidence states are not calibrated probabilities. Its value therefore lies in transparent reasoning and validation prioritization, not claimed predictive accuracy.

Future evaluation should therefore test the architecture prospectively rather than judging it only against already solved pathways. Useful designs would begin with partially resolved medicinal-plant pathways, register candidate relationships and alternative explanations before functional testing, and then examine whether the architecture identifies the experiments that most reduce assignment uncertainty. Validation should also be deliberately orthogonal: perturbation should be paired, where feasible, with chemically specific metabolite measurements; localization claims should be tested at compatible spatial resolution; and heterologous reconstruction should be interpreted alongside native expression, transport, or genetic evidence when those properties are part of the claim. Such studies could challenge as well as refine the proposed architecture.

Conclusion

Inferring a medicinal-plant biosynthetic pathway is not equivalent to finding the omics layer with the strongest correlation, the densest candidate cluster, or the largest number of supporting observations. Each evidence type removes a different kind of uncertainty while retaining others. Genomic evidence can define candidate repertoire and evolutionary context; transcriptional evidence can identify conditional coordination; chemical and spatial evidence can constrain metabolite identity, occurrence, and localization; and functional experiments can test catalytic or reconstructed pathway propositions. None of these contributions becomes universally decisive merely by being technologically advanced or statistically prominent.

The multi-omics evidence architecture proposed here therefore organizes gene–metabolite assignment around evidential roles, source quality, compatibility, dependence, conflict diagnosis, and claim-specific functional resolution. Agreement strengthens an assignment only when the contributing observations are relevant to the same proposition and are not redundant manifestations of one underlying signal. Discordance can weaken a hypothesis, but it can also reveal transport, compartmentation, paralogy, evolutionary divergence, incomplete annotation, or analytical mismatch. The appropriate response is consequently diagnosis before aggregation.

This architecture is intended to make pathway reconstruction more explicit about what is supported, what remains alternative, and which experiment could most efficiently change the interpretation. It does not provide a universal score, validated prediction rule, or substitute for biochemical and in planta testing. Its central practical implication is narrower: confidence should follow the question that the evidence actually answers. A pathway map becomes more informative when its certainty is bounded as carefully as its chemistry, and when remaining uncertainty is used to design the next discriminating experiment rather than concealed by multimodal agreement.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Zhao K, Rhee SY. Omics-guided metabolic pathway discovery in plants: Resources, approaches, and opportunities. Curr Opin Plant Biol. 2022;67:102222. doi:10.1016/j.pbi.2022.102222

2. Singh KS, van der Hooft JJJ, van Wees SCM, Medema MH. Integrative omics approaches for biosynthetic pathway discovery in plants. Nat Prod Rep. 2022;39(9):1876-96. doi:10.1039/D2NP00032F
3. Wolters FC, Del Pup E, Singh KS, Bouwmeester K, Schranz ME, van der Hooft JJJ, et al. Pairing omics to decode the diversity of plant specialized metabolism. Curr Opin Plant Biol. 2024;82:102657. doi:10.1016/j.pbi.2024.102657
4. Swamidatta SH, Lichman BR. Beyond co-expression: Pathway discovery for plant pharmaceuticals. Curr Opin Biotechnol. 2024;88:103147. doi:10.1016/j.copbio.2024.103147
5. Li C, Wood JC, Vu AH, Hamilton JP, Rodriguez Lopez CE, Payne RME, et al. Single-cell multi-omics in the medicinal plant Catharanthus roseus. Nat Chem Biol. 2023;19(8):1031-41. doi:10.1038/s41589-023-01327-0
6. Wisecaver JH, Borowsky AT, Tzin V, Jander G, Kliebenstein DJ, Rokas A. A global coexpression network approach for connecting genes to specialized metabolic pathways in plants. Plant Cell. 2017;29(5):944-59. doi:10.1105/tpc.17.00009
7. Zhan C, Shen S, Yang C, Liu Z, Fernie AR, Graham IA, et al. Plant metabolic gene clusters in the multi-omics era. Trends Plant Sci. 2022;27(10):981-1001. doi:10.1016/j.tplants.2022.03.002
8. 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-W63. doi:10.1093/nar/gkx305
9. Schläpfer P, Zhang P, Wang C, Kim T, Banf M, Chae L, et al. Genome-wide prediction of metabolic enzymes, pathways, and gene clusters in plants. Plant Physiol. 2017;173(4):2041-59. doi:10.1104/pp.16.01942
10. Bryson AE, Lanier ER, Lau KH, Hamilton JP, Vaillancourt B, Mathieu D, et al. Uncovering a miltiradiene biosynthetic gene cluster in the Lamiaceae reveals a dynamic evolutionary trajectory. Nat Commun. 2023;14(1):343. doi:10.1038/s41467-023-35845-1
11. Liu C, Smit SJ, Dang J, Zhou P, Godden GT, Jiang Z, et al. A chromosome-level genome assembly reveals that a bipartite gene cluster formed via an inverted duplication controls monoterpenoid biosynthesis in Schizonepeta tenuifolia. Mol Plant. 2023;16(3):533-48. doi:10.1016/j.molp.2023.01.004
12. Sun Y, Shao J, Liu HL, Wang H, Wang GY, Li JH, et al. A chromosome-level genome assembly reveals that tandem-duplicated CYP706V oxidase genes control oridonin biosynthesis in the shoot apex of Isodon rubescens. Mol Plant. 2023;16(3):517-32. doi:10.1016/j.molp.2022.12.007
13. Li H, Wu S, Lin R, Xiao Y, Morotti ALM, Wang Y, et al. The genomes of medicinal skullcaps reveal the polyphyletic origins of clerodane diterpene biosynthesis in the family Lamiaceae. Mol Plant. 2023;16(3):549-70. doi:10.1016/j.molp.2023.01.006
14. Guo L, Winzer T, Yang X, Li Y, Ning Z, He Z, et al. The opium poppy genome and morphinan production. Science. 2018;362(6412):343-7. doi:10.1126/science.aat4096
15. Zhao Q, Yang J, Cui MY, Liu J, Fang Y, Yan M, et al. The reference genome sequence of Scutellaria baicalensis provides insights into the evolution of wogonin biosynthesis. Mol Plant. 2019;12(7):935-50. doi:10.1016/j.molp.2019.04.002
16. Liu Y, Wang B, Shu S, Li Z, Song C, Liu D, et al. Analysis of the Coptis chinensis genome reveals the diversification of protoberberine-type alkaloids. Nat Commun. 2021;12(1):3276. doi:10.1038/s41467-021-23611-0
17. Xiong X, Gou J, Liao Q, Li Y, Zhou Q, Bi G, et al. The Taxus genome provides insights into paclitaxel biosynthesis. Nat Plants. 2021;7(8):1026-36. doi:10.1038/s41477-021-00963-5
18. Yu C, Luo X, Zhang C, Xu X, Huang J, Chen Y, et al. Tissue-specific study across the stem of Taxus media identifies a phloem-specific TmMYB3 involved in the transcriptional regulation of paclitaxel biosynthesis. Plant J. 2020;103(1):95-110. doi:10.1111/tpj.14710
19. Yu C, Hou K, Zhang H, Liang X, Chen C, Wang Z, et al. Integrated mass spectrometry imaging and single-cell transcriptome atlas strategies provide novel insights into taxoid biosynthesis and transport in Taxus mairei stems. Plant J. 2023;115(5):1243-60. doi:10.1111/tpj.16315
20. Zhu A, Liu M, Tian Z, Liu W, Hu X, Ao M, et al. Chemical-tag-based semi-annotated metabolomics facilitates gene identification and specialized metabolic pathway elucidation in wheat. Plant Cell. 2024;36(3):540-58. doi:10.1093/plcell/koad286
21. Chen C, Zhang X, Yue M. Spatial multi-omics in medicinal plants: From biosynthesis pathways to industrial applications. Trends Plant Sci. 2024;29(5):510-3. doi:10.1016/j.tplants.2024.02.008

22. Kang M, Vu AH, Casper AL, Kim R, Wurlitzer J, Heinicke S, et al. Single-cell metabolome and RNA-seq multiplexing on single plant cells. Proc Natl Acad Sci U S A. 2025;122(43):e2512828122. doi:10.1073/pnas.2512828122
23. Pei T, Zhu S, Liao W, Fang Y, Liu J, Kong Y, et al. Gap-free genome assembly and CYP450 gene family analysis reveal the biosynthesis of anthocyanins in Scutellaria baicalensis. Hortic Res. 2023;10(12):uhad235. doi:10.1093/hr/uhad235
24. He S, Weng D, Zhang Y, Kong Q, Wang K, Jing N, et al. A telomere-to-telomere reference genome provides genetic insight into the pentacyclic triterpenoid biosynthesis in Chaenomeles speciosa. Hortic Res. 2023;10(10):uhad183. doi:10.1093/hr/uhad183
25. Li Q, Duncan S, Li Y, Huang S, Luo M. Decoding plant specialized metabolism: New mechanistic insights. Trends Plant Sci. 2024;29(5):535-45. doi:10.1016/j.tplants.2023.11.015
26. Dixon RA, Dickinson AJ. A century of studying plant secondary metabolism—From "what?" to "where, how, and why?". Plant Physiol. 2024;195(1):48-66. doi:10.1093/plphys/kiad596
27. Caputi L, Franke J, Farrow SC, Chung K, Payne RME, Nguyen TD, et al. Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. Science. 2018;360(6394):1235-9. doi:10.1126/science.aat4100
28. Qu Y, Safonova O, De Luca V. Completion of the canonical pathway for assembly of anticancer drugs vincristine/vinblastine in Catharanthus roseus. Plant J. 2019;97(2):257-66. doi:10.1111/tpj.14111
29. Zhang Y, Wiese L, Fang H, Alseekh S, Perez de Souza L, Scossa F, et al. Synthetic biology identifies the minimal gene set required for paclitaxel biosynthesis in a plant chassis. Mol Plant. 2023;16(12):1951-61. doi:10.1016/j.molp.2023.10.016
30. Liu JCT, De La Peña R, Tocol C, Sattely ES. Reconstitution of early paclitaxel biosynthetic network. Nat Commun. 2024;15(1):1419. doi:10.1038/s41467-024-45574-8
31. Jiang B, Gao L, Wang H, Sun Y, Zhang X, Ke H, et al. Characterization and heterologous reconstitution of Taxus biosynthetic enzymes leading to baccatin III. Science. 2024;383(6683):622-9. doi:10.1126/science.adj3484
32. McClune CJ, Liu JCT, Wick C, De La Peña R, Lange BM, Fordyce PM, et al. Discovery of FoTO1 and Taxol genes enables biosynthesis of baccatin III. Nature. 2025;643(8071):582-92. doi:10.1038/s41586-025-09090-z
33. Hawkins C, Xue B, Yasmin F, Wyatt G, Zerbe P, Rhee SY. Plant Metabolic Network 16: Expansion of underrepresented plant groups and experimentally supported enzyme data. Nucleic Acids Res. 2025;53(D1):D1606-D1613. doi:10.1093/nar/gkae991
34. Qin K, Liu F, Zhang C, Deng R, Fernie AR, Zhang Y. Systems and synthetic biology for plant natural product pathway elucidation. Cell Rep. 2025;44(6):115715. doi:10.1016/j.celrep.2025.115715
35. Leng L, Xu Z, Hong B, Zhao B, Tian Y, Wang C, et al. Cepharanthine analogs mining and genomes of Stephania accelerate anti-coronavirus drug discovery. Nat Commun. 2024;15(1):1537. doi:10.1038/s41467-024-45690-5
36. An Z, Gao R, Chen S, Tian Y, Li Q, Tian L, et al. Lineage-specific CYP80 expansion and benzylisoquinoline alkaloid diversity in early-diverging eudicots. Adv Sci (Weinh). 2024;11(19):e2309990. doi:10.1002/advs.202309990
37. Hu Y, Wang J, Liu L, Yi X, Wang X, Wang J, et al. Evolutionary history of magnoliid genomes and benzylisoquinoline alkaloid biosynthesis. Nat Commun. 2025;16(1):4039. doi:10.1038/s41467-025-59343-8
38. Shan C, Zhou X, Zhu J, Rashid A, Wang J, An N, et al. Gene duplication and clustering underlie the conservation and diversification of benzylisoquinoline alkaloid biosynthesis in plants. Nat Commun. 2025;16(1):7669. doi:10.1038/s41467-025-63175-x
39. Cheng LT, Wang ZL, Zhu QH, Ye M, Ye CY. A long road ahead to reliable and complete medicinal plant genomes. Nat Commun. 2025;16(1):2150. doi:10.1038/s41467-025-57448-8