

Pan-genome-driven analysis of forest trees' secondary metabolism: from multi-omics integration to computational pharmacological evaluation

Yuhan Feng¹, Shumei Zhang², Fulei Nie^{2,3}, Rui Cao^{2,3*} and Xiaoming Song^{1,2,3*}

¹ Hebei Key Laboratory for Chronic Diseases, Tangshan Key Laboratory for Preclinical and Basic Research on Chronic Diseases, School of Basic Medical Sciences, North China University of Science and Technology, Tangshan 063210, China

² School of Life Sciences, North China University of Science and Technology, Tangshan 063210, China

³ Key Laboratory for quality of salt alkali resistant TCM of Hebei Administration of TCM, North China University of Science and Technology, Tangshan 063210, China

* Correspondence: caoruiivv@163.com (Cao R); songxm@ncst.edu.cn (Song X)

Abstract

Forest trees produce structurally diverse secondary metabolites, which serve as vital defensive compounds against biotic and abiotic stresses and constitute important resources for natural-product-derived drug leads. Compared with annual herbaceous model plants, forest trees exhibit biological features including highly heterozygous genomes, long generation times, and abundant population-level structural variations. Conventional single-reference genomes cannot fully capture presence-absence variations (PAVs), copy number variations (CNVs), and large-scale structural variations (SVs) that are relevant to secondary metabolism, which restricts the efficiency of mining pharmacologically valuable metabolites from forest trees. Pan-genomics provides a new research framework for deciphering the evolution of trees' secondary metabolism, enabling a systematic dissection of how whole-genome duplication (WGD) and population-level structural variations shape secondary metabolic gene families. Multi-omics integration strategies such as weighted gene co-expression network analysis (WGCNA), metabolite-based Genome-Wide Association Studies (mGWAS), and expression Quantitative Trait Loci (eQTL) facilitate the identification of hub genes regulating metabolite biosynthesis. Computational pharmacological approaches, including molecular docking, molecular dynamics simulations, and Absorption, Distribution, Metabolism, Excretion and Toxicity (ADMET) prediction, support high-throughput virtual screening of candidate compounds. Nevertheless, computational simulations only generate scientific hypotheses and cannot substitute for *in vitro* and *in vivo* pharmacological experimental validation. As a Perspective article, this paper constructs a closed-loop technical pipeline: Pan-genomic dissection of variation → multi-omics identification of hub genes → computational pharmacological evaluation for mining natural products from forest trees. We systematically analyze the technical merits, intrinsic limitations, and practical bottlenecks of each step within this pipeline and propose priority directions for future research, providing references for drug translation research of nontimber forest products.

Keywords: Pan-genome, Multi-omics integration, Secondary metabolism, Gene family evolution, Molecular dynamics, Forest trees, Computational pharmacology, Forest trees, Computational pharmacology

Citation: Feng Y, Zhang S, Nie F, Cao R, Song X. 2026. Pan-genome-driven analysis of forest trees' secondary metabolism: from multi-omics integration to computational pharmacological evaluation. *Forestry Research Advances* 1: e018 <https://doi.org/10.48130/fra-0026-0016>

Introduction

Forest ecosystems harbor abundant, structurally unique secondary metabolites. Multiple tree-derived metabolites possess well-documented pharmacological values, such as paclitaxel from *Taxus brevifolia*, camptothecin from *Camptotheca acuminata*, and betulinic acid from *Betula platyphylla*, which are representative lead molecules for antitumor and anti-inflammatory drug development^[1,2]. As perennial woody plants, forest trees differ substantially from annual herbaceous species, generally characterized by high genomic heterozygosity, long generation cycles, and immature genetic transformation systems^[3]. Under long-term evolutionary selective pressure, whole-genome duplication events, transposon bursts, and population-level structural variations collectively reshape secondary metabolic pathways and form unique chemical defence systems in woody plants^[4].

A single-reference genome only represents genomic information from one individual and loses substantial presence-absence variations (PAVs), copy number variations (CNVs), and structural variations (SVs) closely associated with secondary metabolic traits, making it difficult to interpret metabolic-phenotype divergence

within tree populations. With the maturation of long-read sequencing technologies, pan-genomes for several tree species, including poplar (*Populus* spp.), tea (*Camellia sinensis*)^[5], and oak (*Quercus* spp.), have been successively published, overcoming the inherent drawbacks of single-reference genomes and laying the data foundations for systematically dissecting secondary metabolic evolution in forest trees^[4].

Multi-omics integration has become a mainstream paradigm for dissecting plants' metabolic-regulatory networks. Datasets of pan-genomic variations combined with transcriptomics and metabolomics, via the joint analytical strategies of weighted gene co-expression network analysis (WGCNA) and metabolite-based Genome-Wide Association Studies (mGWAS) combined with expression Quantitative Trait Loci (eQTL), enable identification of hub genes governing secondary metabolism. Coupled with computational biophysical simulations, high-throughput screening systems bridging plants' genomic variations and human drug targets can be established^[6].

This work is presented as a Perspective rather than a comprehensive exhaustive literature review. It aims to sort out the complete technical chain ranging from pan-genomic evolutionary dissection

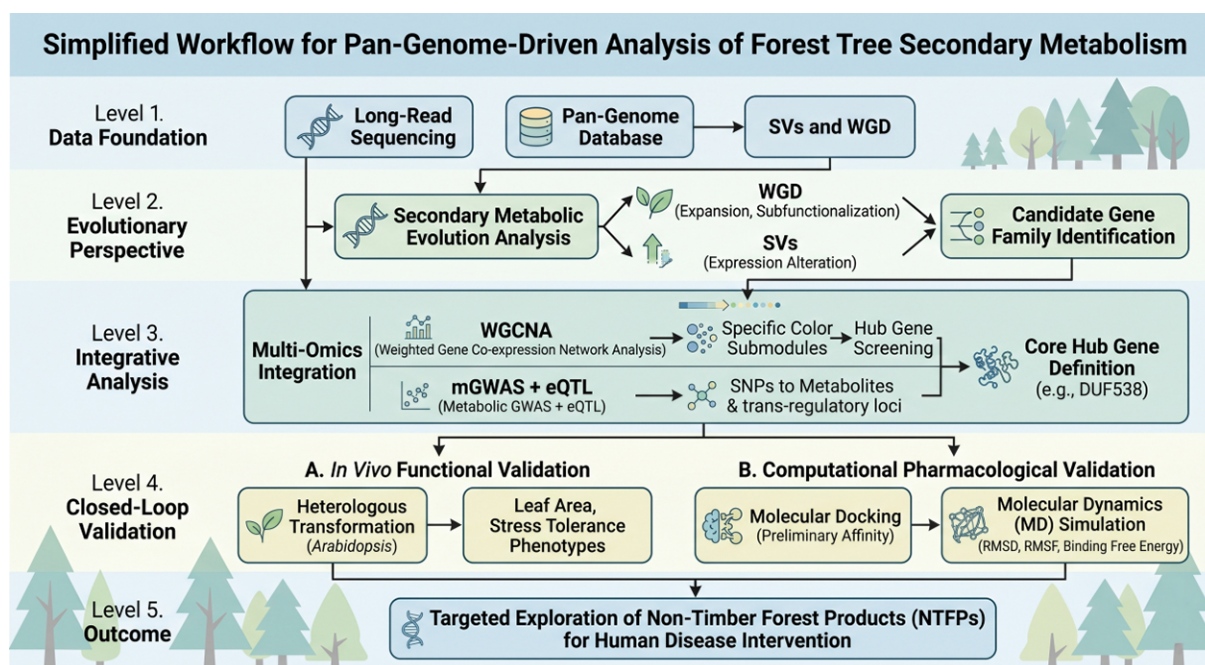


Fig. 1 Schematic diagram of the overall workflow for a pan-genome-driven analysis of secondary metabolism in forest trees. The diagram illustrates the complete technical workflow from pan-genomic data resources, evolutionary analysis, multi-omics integrative analysis, and closed-loop validation to natural product exploitation.

and multi-omics hub gene mining, to computational pharmacological translation of tree-derived metabolites, and to analyze scientific questions and technical bottlenecks at each stage, offering insights for research on trees' secondary metabolism and natural product exploitation. This paper concludes by summarizing the core logics and prospecting follow-up research directions (Fig. 1).

Pan-genomic perspectives on evolutionary drivers of secondary metabolism in forest trees

Whole-genome duplication (WGD) acts as a core evolutionary force driving the large-scale expansion of plant genomes^[7]. Homologous genes generated by duplication achieve spatiotemporal expression divergence through subfunctionalization, and some copies acquire neofunctionalized mutations to remodel secondary metabolic biosynthetic networks^[8]. Near-gapless genome assemblies for *Pseudocarya sinensis* demonstrate that a Maleae-specific WGD event drove the tandem duplication expansion of phenylpropanoid pathway gene families, accompanied by the amplification of flavonoid biosynthesis-related gene families^[9,10]. Phylogenomic analyses of seed plants' pan-genomes show that subfunctionalization contributes to the local adaptive evolution of metabolic pathways; diploidization following polyploidization eliminates large quantities of redundant genes, and retained paralogous genes have endowed forest trees with distinctive chemical defence capabilities.

SV constitutes a core variant category in pan-genome datasets. Compared with single-nucleotide polymorphisms (SNPs), SVs involve large-scale chromosomal rearrangements and can alter *cis*-regulatory elements in promoter regions; transposon insertions reshape the epigenetic modification status of target genes^[11–13]. Pan-genome analyses of 17 poplar (*Populus* spp.) populations

identified 30 *PopPLY* family members. *Ka/Ks* calculations indicate purifying selection maintains structural conservation of abscisic acid signalling. Six discrete SVs were detected at *Populus euphratica* Pyrabactin resistance-like(*PYL*) loci. Transcriptomic data reveal that these SVs alter transcript abundances of *PYL6* and *PYL17* and mediate drought-stress-response divergence in root and leaf tissues^[14].

A unique merit of pan-genome analysis lies in its capability to capture PAVs and CNVs frequently missed by single-reference assemblies. Such variants are particularly critical for secondary metabolic traits, because gene dosage effects directly modulate metabolic flux. PAVs covering complete biosynthetic gene clusters have been identified in multiple forest tree lineages, confirming SVs as important contributors to interindividual metabolic diversity.

Conventional comparative genomics versus pan-genome analysis in forest trees

Conventional comparative genomics mostly relies on a small set of reference genomes and excels at resolving WGD-mediated gene family expansion at species-divergence scales, yet cannot capture intrapopulation PAVs, CNVs, and SVs. Forest trees' pan-genomes enable analyses of intrapopulation genetic variations and address biological questions such as why metabolic phenotypes differ among conspecific individuals. These two approaches are complementary and nonsubstitutable. WGD represents ancient species-scale evolutionary events, whereas PAVs/SVs/CNVs mostly reflect recent population-level adaptive variations, marking their essential distinction in evolutionary timescales^[15].

Section summary and limitations

Current studies on WGD and the pan-genomes of forest trees still exhibit prominent shortcomings. Most forest trees' pan-genomes are built with limited population sample sizes. Numerous secondary-metabolism-related PAVs are only characterized at the sequence level without supporting metabolic phenotype

Table 1. Representative pan-genome–multi-omics case studies for secondary metabolism research in forest trees.

Species	Omics type	Core scientific question	Key findings	Limitations of the study	Ref.
<i>Pseudocystodonia sinensis</i>	PacBio gapless genome	Impacts of WGD on the phenylpropanoid pathway evolution	Maleae-specific WGD drives expansion of phenylpropanoid and flavonoid gene families	Only one individual genome; lacks population-level variant information	[16]
<i>Populus euphratica</i>	Pan-genome + transcriptome	How SVs regulate stress-responsive metabolic pathways	SVs at <i>PYL</i> loci alter gene transcription and modulate drought responses	No combined metabolomic analysis; secondary metabolic–phenotype associations are absent	[17]
<i>Populus tomentosa</i>	Pan-genome + mGWAS + eQTL	Genetic basis for coumarin metabolism	<i>DUF538</i> is identified as a core gene for coumarin biosynthesis	Limited population sample size; no pharmacological translation investigation	[18]

validations. Biological functions for most duplicated paralogous gene copies remain experimentally uncharacterized^[4]. Given such evolutionary understandings, multi-omics approaches are required to identify hub genes governing secondary metabolite biosynthesis from massive genomic variations.

Representative pan-genome–multi-omics case studies for secondary metabolism in forest trees are summarized in Table 1.

Multi-omics integration and identification of core metabolic networks

WGCNA represents a classic clustering algorithm for high-throughput transcriptomic data^[19]. Conventional analytical workflows prioritize high-weight modules with maximum node counts. Such modules are frequently enriched in housekeeping genes for basal metabolism and show limited overlap with secondary metabolic genes. Therefore, soft threshold parameters should be adjusted to mine specific secondary co-expression submodules. Selection of color specific submodules filters expression noise from housekeeping genes; accurately localizes regulatory networks for flavonoid, phenylpropanoid, and other specialized metabolic pathways; and improves efficiency for hub gene screening^[20–22].

The intrinsic limitations of WGCNA include the following. WGCNA only constructs correlation-based co-expression networks and cannot infer causal gene-regulatory relationships. It builds undirected networks from pairwise gene expression correlations and cannot judge regulatory directions or causal flows. Tools such as GENIE3 and GRNBoost2 should be supplemented for causal–regulatory dissection.

mGWAS maps metabolic chemical phenotypes onto genomic variant loci^[23]. Combined with eQTL analysis, it further introduces transcriptional regulatory dimensions. In population-level multi-omics studies of *Populus tomentosa*, mGWAS pinpointed SNP loci associated with coumarin metabolism; eQTL mapped trans-regulatory loci. Coupled with selective sweep analyses, the core gene *DUF538* was finally retrieved. Heterologous transformation experiments validated its function in coordinately modulating coumarin-related metabolic flux.

Compared with standalone analytical methods, the combination of the WGCNA submodule strategy and mGWAS–eQTL improves the specificity of hub gene discovery, while possessing clear application boundaries. Cross-platform data dimension discrepancies reduce the goodness-of-fit of statistical models. Conventional dimensionality-reducing algorithms risk discarding functionally critical yet less abundant metabolic nodes. The high heterozygosity of forest trees' genomes introduces substantial background noise^[3]. Relative to inbred model plants, this method demands larger population sizes and deeper sequencing coverage.

Multi-omics integration should not be restricted to transcriptome–metabolome dual-omics. Proteomics reflects

post-translational regulation and bridges gaps between transcript abundance and metabolic phenotypes^[24]. Single-cell and spatial transcriptomics resolve the cellular heterogeneity underlying secondary metabolism and pinpoint specific cell subpopulations responsible for metabolite biosynthesis^[25,26]. Nevertheless, high lignification in woody tissues impedes the isolation of protoplasts, and the cell wall's interference hinders the large-scale deployment of single-cell and spatial transcriptomics in forest tree research^[27].

Section summary and limitations

Existing multi-omics-integrated studies on forest trees are mostly confined to a small number of model tree species. Unified cohort datasets across different tissues and environmental conditions are scarce. Heterogeneity among data generated by distinct sequencing platforms increases difficulties for multi-omics fusion analyses. Multi-omics approaches can retrieve candidate genes and metabolites related to trees' secondary metabolism. However, the ecological defensive activities of tree metabolites do not equate to human pharmacological activities, and remarkable logical gaps exist between them.

From plant defense to human pharmacology: translational opportunities and logical gaps

Forest trees synthesize abundant secondary metabolites against herbivores, pathogenic microorganisms, and oxidative stresses^[28,29]. Some of these compounds have been developed into clinical drugs: Paclitaxel from *Taxus* (a microtubule stabilizer for antitumor therapy), camptothecin from *Camptotheca acuminata* (a topoisomerase-I inhibitor), and betulinic acid from birch (*Betula* spp.) (antimelanoma and anti-human immunodeficiency virus (HIV) activities).

Critical logical revision: Some previous literature proposed a "mechanistic parallelism" between the ecological defense functions of plant metabolites and human pharmacological activities, which represents an oversimplification. Although partial phenolic and terpenoid compounds function via electrophilic reactivity, redox modulation, and competitive enzyme inhibition, and such biochemical mechanisms are conserved across biological kingdoms, anti-insect or antimicrobial activities in plants cannot be directly equated to therapeutic efficacy in humans. Species-specific target divergence, *in vivo* bioavailability, metabolic degradation pathways, and dosage discrepancies collectively prevent ecological bioactivities in plants from translating into pharmacological effects in humans. Ecological activity merely serves as preliminary screening evidence instead of pharmacological proof^[30].

It should be noted that classic active molecules such as paclitaxel, camptothecin, and betulinic acid were originally discovered via traditional phytochemistry rather than pan-genomics or

multi-omics. The value of pan-genomics and multi-omics lies in resolving their biosynthetic pathways, mining natural population variations, and discovering novel structural analogues.

Three cases of a complete evidence chain

- (1) Betulinic acid from birch (*Betula* spp.): pan-genomics resolved the regulatory effects of PAV variations within triterpene biosynthetic gene clusters on betulinic acid accumulation. Molecular dynamics simulations verified the binding of betulinic acid derivatives to the inflammatory targets COX-2 and 5-LOX, and their anti-inflammatory activities have been validated by *in vitro* cellular assays.
- (2) Flavonoid metabolites in poplar (*Populus* spp.): multi-omics integrative analyses identified *MYB* family hub genes governing flavonoid biosynthesis and accumulation. Computational simulations evaluated their binding potentials to PPAR γ and α -glucosidase and suggested hypoglycemic prospects, which await further wet-lab validations.
- (3) Paclitaxel biosynthetic pathway in *Taxus*: pan-genomics identified SV variations modulating the expression of taxadiene synthase gene families and altering paclitaxels' biosynthetic capacity. Anti-cancer targets of paclitaxel are well supported by abundant experimental evidence.

Representative bioactive secondary metabolites derived from forest tree species are compiled in Table 2.

Forest trees have experienced recurrent WGD events and long-term adaptive evolution, conferring great potential to generate structurally diverse specialized metabolites and rendering them valuable resources for natural-product-based drug discovery^[9].

Molecular dynamics and computational pharmacological evaluation

After the candidate genes have been predicted by bioinformatic analyses, heterologous transformation systems deliver functional evidence at macroscopic phenotypic levels^[35–37]. Nevertheless, molecular-level physical evidence is required for applying tree-derived metabolites in research into human disease interventions.

Molecular docking involves a preliminary assessment of molecular steric hindrance and electrostatic complementarity at the binding interfaces. Molecular dynamics (MD) simulations further resolve the thermodynamic stability of bound complexes. The GROMACS package executes 100-nanosecond trajectory simulations with Optimized Potentials for Liquid Simulations All-Atom (OPLS-AA) force-fields, while the LigParGen server generates small-molecule topological parameters^[38,39]. Outputs including Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF) and Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) binding free energies provide atomistic-scale descriptions for ligand–receptor interactions^[40,41].

Key clarification: MD simulation is a hypothesis-generating predictive tool and is not equivalent to pharmacological experimental validation. Confirmation of pharmacological efficacy necessitates *in vitro* enzyme inhibition assays, cellular function tests, and ultimately *in vivo* animal efficacy models. Computational pharmacology functions to prioritize candidate molecules at high throughput and to narrow down candidate pools for subsequent experimental validations.

Complete *in silico* pharmacological assessment also incorporates ADMET prediction. Machine learning models evaluate oral bioavailability, penetration of the blood–brain barrier, cytochrome-P450 metabolism profiles, and hepatotoxicity risks so that compounds with poor pharmacokinetic properties can be eliminated ahead of resource-intensive wet-lab experiments.

Section summary and limitations

Numerous secondary metabolites from forest trees possess complex stereochemical structures. The lack of well-adapted force-field parameters for some small molecules compromises the reliability of simulations' outputs. Although the complete technical chain combining pan-genomics, multi-omics, and computational pharmacology has been established, multiple technical bottlenecks restrict translational applications of natural products originating from forest trees.

The input, output, advantages, and limitations of core computational–pharmacology workflows applied to tree-derived natural products are listed in Table 3.

Table 2. Representative bioactive secondary metabolites derived from forest tree species.

Tree species	Secondary metabolite	Biosynthetic pathway	Pharmacological target	Evidence level	Ref.
<i>Taxus brevifolia</i>	Paclitaxel	Terpenoid biosynthetic pathway	Tubulin	<i>In vitro</i> + <i>in vivo</i> experimental evidence	[31]
<i>Camptotheca acuminata</i>	Camptothecin	Terpenoid–indole–alkaloid pathway	Topoisomerase I	<i>In vitro</i> + <i>in vivo</i> experimental evidence	[32]
<i>Betula platyphylla</i>	Betulinic acid	Triterpenoid biosynthetic pathway	COX-2, 5-LOX	<i>In vitro</i> cellular assays; computational simulations	[33]
<i>Populus</i> spp.	Flavonoids	Flavonoid biosynthetic pathway	PPAR γ , α -glucosidase	Computational prediction only; experimental validations pending	[34]

Table 3. Input, output, advantages, and limitations of core computational pharmacology evaluation workflows applied to tree-derived natural products.

Workflow	Input	Output	Advantages	Limitations	Ref.
Molecular docking	Proteins' 3D structure, small-molecule structure	Preliminary binding modes and docking scores	High throughput, fast computation speed	Static model; ignores proteins' conformational dynamics; relatively high false positive rates	[41]
Molecular dynamics simulation	Docked complex, force-field parameters	RMSD/RMSF, binding free energies	Atomistic-scale dynamic characterization of molecular interactions	High computational resource consumption; force-field parameter deficits for complex natural products	[42]
ADMET prediction	2D/3D small-molecule structures	Pharmacokinetic and toxicity risk profiles	Early elimination of candidates with poor druglikeness	Performance depends on the training datasets; prediction accuracy is limited for tree-specific metabolites	[43]

2D: 2-dimension; 3D: 3-dimension; RMSD: Root Mean Square Deviation; RMSF: Root Mean Square Fluctuation.

Technical bottlenecks and interdisciplinary prospects

Current major technical bottlenecks

(1) Pan-genomic level: Most forest trees' pan-genomes are constructed with limited population sizes. Many secondary metabolism-related PAVs/CNVs are only sequenced without matched metabolic phenotype datasets, leading to severely insufficient functional annotations^[44].

(2) Multi-omics level: High sequencing costs restrict large-population multi-omics studies on forest trees. Data heterogeneity arising from different experimental platforms creates obstacles for fusion analyses. The characteristics of woody samples hinder the large-scale deployment of single-cell and spatial transcriptomics.

(3) Pharmacological translation level: Cases with complete evidence chains from genomic variations to *in vivo* pharmacological efficacy are rare. Plenty of computationally screened candidate metabolites lack follow-up wet-lab validations.

Practical future research directions

(1) Construct tree-specific pan-genome-metabolome population cohorts, integrating genotype and metabolic phenotype datasets to decipher the functional effects of secondary metabolism-associated PAVs.

(2) Develop multi-omics-integration algorithms adapted to highly heterozygous tree genomes to reduce background noise and improve the accuracy of hub gene identification.

(3) Establish standardized closed-loop workflows: Pan-genome mining → multi-omics screening → *in vitro* functional validation → high-throughput computational pharmacology prioritization → cellular / *in vivo* animal efficacy validation. It must be emphasized that computational simulations cannot replace experimental assays.

(4) Promote applications of deep learning-based artificial intelligence (AI) in SV detection, metabolite structure interpretation, and virtual compound screening, while objectively acknowledging the limitations of AI-driven predictions^[45].

(5) Leverage the strengths of single-cell and spatial transcriptomics to resolve spatiotemporal cellular specificities underlying secondary metabolite biosynthesis.

Section summary and limitations

Overall, research concerning forest trees' secondary metabolism remains at an early interdisciplinary fusion stage. Specialized cohort resources, tailored analytical algorithms, and multilayer experimental validation systems are yet to be further improved. Synthesizing evolutionary dissection, multi-omics mining, and computational pharmacology evaluations, we draw general conclusions for this manuscript.

Conclusions

Long-read sequencing has updated the foundational data infrastructure for forest trees

Long-read-sequencing technologies have pushed forest tree research into the pan-genomic era. WGD events, together with population-level SV/PAV/CNV variations, co-drive the evolutionary remodeling of secondary metabolic pathways in forest trees. Multi-omics-integrated strategies enable the identification of regulatory

hub genes for secondary metabolism from massive genomic variations. Computational pharmacology tools such as molecular dynamics and ADMET prediction support high-throughput assessments of interactions between tree-derived metabolites and human targets and accelerate screening for natural lead compounds.

Nevertheless, this whole technical chain still has prominent drawbacks. Construction of pan-genome cohorts for forest trees is inadequate, and technical obstacles persist for multi-omics-data fusion. Cases with complete experimentally validated evidence chains linking tree genomic variations to bona fide pharmacological activities are scarce. Predictions from computational simulations must be verified via multilayer wet-lab experiments and cannot be directly regarded as evidence for pharmacological efficacy.

From a Perspective viewpoint, this paper outlines a closed-loop workflow for mining natural products from forest trees. We hope to foster interdisciplinary intersections among evolutionary genomics, multi-omics, and computational pharmacology, and promote translational research converting nontimber forest product resources into drug leads^[46].

Author contributions

The authors confirm their contributions to the paper as follows: Study conception and design: Cao R, Song X; data collection: Feng Y, Zhang S, Nie F; analysis and interpretation of the results: Feng Y, Zhang S, Nie F, Cao R, Song X; draft manuscript preparation: Feng Y, Cao R. All authors reviewed the results and approved the final version of the manuscript.

Data availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Acknowledgments

This work was supported by the Central Guidance and Support for Local Science and Technology Development Fund (246Z2509G), the Medical Engineering Integration Project of North China University of Science and Technology (ZD-YG-202411), the National Natural Science Foundation of China (32573037), the S&T Program of Hebei (23372505D), and the Hebei Natural Science Foundation (H2023209084).

Conflict of interest

The authors declare that they have no conflict of interest.

Dates

Received 29 June 2026; Revised 20 August 2026; Accepted 17 September 2026; Published online 15 October 2026

References

- [1] Ou-Yang T, Zhang Y, Luo HZ, Liu Y, Ma SC. 2023. Novel compounds discovery approach based on UPLC-QTOF-MS/MS chemical profile reveals birch bark extract anti-inflammatory, -oxidative, and-proliferative effects. *Journal of Ethnopharmacology* 306:116148
- [2] Wender PA, Gentry ZO, Fanelli DJ, Luu-Nguyen QH, McAteer OD, et al. 2022. Practical synthesis of the therapeutic leads tigilanol tiglate and its analogues. *Nature Chemistry* 14:1421–1426
- [3] Meng X, Fu Y, Qi Y, Jin Z, Li P, et al. 2025. Genome-wide association studies in forestry. *Molecular Biology Reports* 52:848

- [4] Shi T, Zhang X, Hou Y, Jia C, Dan X, et al. 2024. The super-pangenome of *Populus* unveils genomic facets for its adaptation and diversification in widespread forest trees. *Molecular Plant* 17:725–746
- [5] Yadav S, Rahim MS, Devi A, Sharma RK. 2025. Revolutionizing speciality teas: multi-omics prospective to breed anthocyanin-rich tea. *Food Research International* 209:116312
- [6] Rahmani B, Akbari H, Esmaili H, Solaimanian S, Mobasheri M, et al. 2025. Discovering bioactive pharmaceuticals from natural products for type 2 diabetes mellitus using network pharmacology, molecular docking, and molecular dynamics. *Scientific Reports* 15:43205
- [7] Feng X, Chen Q, Wu W, Wang J, Li G, et al. 2024. Genomic evidence for rediploidization and adaptive evolution following the whole-genome triplication. *Nature Communications* 15:1635
- [8] Liu S, Zhou Z, Wang W, Jia C, Li D, et al. 2025. Genomic analyses of 238 seed plants reveal the evolutionary mechanisms driving specialization of F3H, ANS, and FLS in flavonoid biosynthesis. *Frontiers in Plant Science* 16:1703405
- [9] Liu X, Shuai Y, Zhao X, Zhang M, Yan Y, et al. 2025. Genome-wide identification and evolution-profiling analysis of tps gene family in *Camphora longepaniculata* and screening of key TPS genes. *Frontiers in Plant Science* 16:1546000
- [10] Zhang Y, Zhang S, Guo W, Zhang Y, Chen X, et al. 2025. A near-gapless genome assembly of *Pseudocarya sinensis* uncovers unique phenylpropanoid pathways. *Plant Biotechnology Journal* 23:3581–3596
- [11] Mencia R, Arce AL, Houriet C, Xian W, Contreras A, et al. 2025. Transposon-triggered epigenetic chromatin dynamics modulate EFR-related pathogen response. *Nature Structural & Molecular Biology* 32:199–211
- [12] Nguyen VH, Mittelsten Scheid O, Gutzat R. 2025. Heat stress response and transposon control in plant shoot stem cells. *Plant Physiology* 197(4):kiaf110
- [13] Tossolini I, Mencia R, Arce AL, Manavella PA. 2025. The genome awakens: transposon-mediated gene regulation. *Trends in Plant Science* 30:857–871
- [14] Han X, Qiu C, Gai Z, Zhai J, Song J, et al. 2025. Pan-genome-based characterization of the PYL transcription factor family in *Populus*. *Plants* 14(16):2541
- [15] Wang Y, Wu Z, Jia M, Li J, Song T, et al. 2024. The identification and characterization of the *PeGRF* gene family in *Populus euphratica* oliv. heteromorphic leaves provide a theoretical basis for the functional study of *PeGRF9*. *International Journal of Molecular Sciences* 26(1):66
- [16] Liu H, Zhou H, Ye H, Li M, Ma J, et al. 2025. Integrated multi-omics analyses provide new insights into genomic variation landscape and regulatory network candidate genes associated with walnut endocarp. *The Plant Journal* 122:e70113
- [17] Zhang Z, Chen Y, Zhang J, Ma X, Li Y, et al. 2020. Improved genome assembly provides new insights into genome evolution in a desert poplar (*Populus euphratica*). *Molecular Ecology Resources* 20(3):781–794
- [18] Zhang W, Jin Z, Huang R, Huang W, Li L, et al. 2024. Multi-omics analysis reveals genetic architecture and local adaptation of coumarins metabolites in *Populus*. *BMC Plant Biology* 24:1170
- [19] Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559
- [20] Wei M, Hsieh JA, Dang JF, Tong B, Li H, et al. 2026. Integrating scRNA-seq and snRNA-seq with spatial transcriptomics to unlock the xylem puzzle. *Genome Biology* 27(1):109
- [21] Yue Y, Ma S, Niu J, Ren Y, Zhao X, et al. 2025. Integrative transcriptomic and metabolomic analysis reveal the flavonoid biosynthesis of *Physalis* under low temperature stress. *Physiologia Plantarum* 177:e70445
- [22] Zhang G, Zhang D. 2026. Integrated metabolomic and transcriptomic analysis of phenylpropanoid biosynthesis in *Silphium perfoliatum*. *Current Issues in Molecular Biology* 48(2):230
- [23] Chen L, Wang H, Xu T, Liu R, Zhu J, et al. 2025. A telomere-to-telomere gap-free assembly integrating multi-omics uncovers the genetic mechanism of fruit quality and important agronomic trait associations in pomegranate. *Plant Biotechnology Journal* 23:2852–2870
- [24] Xie D, Tian Y, Zheng F, Wu L, Xu G, et al. 2025. Multi-omics and integrative analytics in natural products discovery. *Journal of Visualized Experiments* (225):e69458
- [25] Li G, Cai M, Liao X, Ma GL, Fan X. 2026. Exploring specialized metabolic pathways in medicinal plants with single-cell and spatial omics. *Acta Pharmaceutica Sinica B* 16:2929–2946
- [26] Wang J, Zhou Y, Zhang M, Li X, Liu T, et al. 2025. Resolving floral development dynamics using genome and single-cell temporal transcriptome of *Dendrobium devonianum*. *Plant Biotechnology Journal* 23:2997–3011
- [27] Hsieh JA, Liou PC, Lin CC, Dai X, Hu CW, et al. 2025. Single-cell and spatial multiomics identifies heterogeneous xylem development driven by mechanical stress in *Populus*. *Developmental Cell* 60:3131–3148.E8
- [28] Çiçek SS, Pfeifer Barbosa AL, Wenzel-Storjohann A, Segovia JFO, Bezerra RM, et al. 2022. Chemical and biological evaluation of Amazonian medicinal plant *Vouacapaoua americana* aubl. *Plants* 12(1):99
- [29] Zhang XX, Li H, Xue J, Yu-Wen W, Wang SY, et al. 2026. Dynamic switching of phenylpropanoid metabolic flux mediates reversible growth-defense tradeoffs in *Salix brachista*. *Plant Science* 369:113188
- [30] Alkac FZ, Yamali C. 2026. A review on the development of semisynthetic phytochemicals in the discovery of anticancer drugs. *Current Topics in Medicinal Chemistry* 26:1809–1832
- [31] Weaver BA. 2014. How Taxol/paclitaxel kills cancer cells. *Molecular Biology of the Cell* 25:2677–2681
- [32] Lorence A, Nessler CL. 2004. Camptothecin, over four decades of surprising findings. *Phytochemistry* 65:2735–2749
- [33] Szlasa W, Ślusarczyk S, Nawrot-Hadzik I, Abel R, Zalesińska A, et al. 2023. Betulin and its derivatives reduce inflammation and COX-2 activity in macrophages. *Inflammation* 46:573–583
- [34] Olaokun OO, Manonga SA, Zubair MS, Maulana S, Mkolo NM. 2022. Molecular docking and molecular dynamics studies of antidiabetic phenolic compound isolated from leaf extract of *Englerophytum magalismontanum* (sond.) T.D.Penn. *Molecules* 27(10):3175
- [35] Gao Y, Gong F, Ma X, Zhang Z, Wang Y. 2025. Functional identification of *JrMYBs* gene in Walnut (*Juglans regia*) demonstrates it enhances drought stress tolerance. *Physiology and Molecular Biology of Plants* 31:375–387
- [36] Wu Z, Wang X, Deng J, Wu W, Hu J, et al. 2026. PbrMYB31 enhances abiotic stress tolerance by promoting suberin biosynthesis in *Arabidopsis*. *BMC Plant Biology* 26(1):582
- [37] Zhang X, Chen Y, Zhu S, Liu N, Li J, et al. 2026. Study on the positive regulation of drought tolerance by PdMYB2R032 based on R2R3-MYB gene family analysis in *Populus deltoides*. *BMC Plant Biology* 26(1):409
- [38] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, et al. 2015. Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2:19–25
- [39] Dodda LS, Cabeza de Vaca I, Tirado-Rives J, Jorgensen WL. 2017. LigPar-Gen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Research* 45:W331–W336
- [40] Sarmah P, Konwar P, Saikia J, Borah T, Verma JS, et al. 2024. Screening of potent inhibitor from *Aquilaria malaccensis* Lam. against arachidonic inflammatory enzymes: an insight from molecular docking, ADMET, molecular dynamics simulation and MM-PBSA approaches. *Journal of Biomolecular Structure and Dynamics* 42:12622–12636
- [41] Maravandi Z, Gholamian S, Samadi A, Gross JD. 2026. In silico evaluation of procyanidin as a potential ESR1 inhibitor: docking and MD insights in uterine fibroids and endometriosis. *In Silico Pharmacology* 14:29
- [42] Thakur M, Gurjar BP, Khushboo, Nagarajan U, Gogoi B, et al. 2026. Molecular and pharmacological impacts of phytochemicals on the enoyl-acyl carrier protein reductase: potential therapeutic implications in tuberculosis. *Integrative Biology* 18:zyag007
- [43] Fu L, Shi S, Yi J, Wang N, He Y, et al. 2024. ADMETlab 3.0: an updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support. *Nucleic Acids Research* 52:W422–W431
- [44] Shi J, Tian Z, Lai J, Huang X. 2023. Plant pan-genomics and its applications. *Molecular Plant* 16:168–186
- [45] Si Y, Lee JS, Jun G, Kang HM, Lee JH. 2025. Spatial omics enters the microscopic realm: opportunities and challenges. *Trends in Genetics* 41:774–787
- [46] Wang B, Liu Q, Zhao W, Zhang T, Zhang D, et al. 2026. Revolutionizing drug discovery from natural products: the roles of artificial intelligence and multi-omics in accelerating innovation. *Acta Pharmaceutica Sinica B* 16:4103–4127



Copyright: © 2026 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit <https://creativecommons.org/licenses/by/4.0/>.