Chinese Herbal Medicines 18 (2026) 3–10

Contents lists available at ScienceDirect

Chinese Herbal Medicines

journal homepage: www.elsevier.com/locate/chmed

Perspective

Herbgenomics: Unraveling natural product biosynthesis in traditional Chinese medicine

Jing Wang [a], Jingyuan Song [b], Shilin Chen [c,*], Zhichao Xu [a,*]

[a] Key Laboratory of National Forestry and Grassland Administration on Chinese Herbal Medicine, College of Life Science, Northeast Forestry University, Harbin 150040, China
[b] Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100193, China
[c] Institute of Herbgenomics, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China

ARTICLE INFO

Article history:
Received 27 September 2025
Revised 3 November 2025
Accepted 18 November 2025
Available online 19 November 2025

Keywords:
biosynthesis
herbgenomics
medicinal plants
natural products
traditional Chinese medicine

ABSTRACT

Medicinal plants synthesize structurally diverse pharmacologically active natural products through conserved precursor pathways, which are further diversified by lineage-specific tailoring enzymes. Although the biosynthetic pathways of well-studied compounds (e.g., berberine, paclitaxel, and ginsenosides) have been extensively characterized, most bioactive components in Traditional Chinese Medicine (TCM) remain poorly understood. The emerging discipline of Herbgenomics has significantly advanced biosynthetic gene discovery and pathway elucidation. By employing genomics-driven strategies such as biosynthetic gene cluster mining, co-expression analysis, and integrated transcriptome-metabolome profiling, Herbgenomics enables systematic identification of key biosynthetic enzymes and uncovers evolutionary mechanisms (e.g., whole-genome and tandem gene duplications) driving metabolic innovation. Furthermore, it provides a foundation for drug development by leveraging gene-encoded natural diverse components and genome-wide pan-receptor platforms (e.g., pan-GPCR). Through the integration of pathway analysis, regulatory mechanisms, molecular breeding, and synthetic biology, Herbgenomics establishes a comprehensive framework for exploring and engineering TCM natural product biosynthesis, offering a sustainable pathway to the discovery and production of bioactive compounds.
© 2025 Tianjin Press of Chinese Herbal Medicines. Published by ELSEVIER B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Medicinal plants represent a treasured repository of bioactive plant natural products (PNPs), forming the material basis of traditional Chinese medicine (TCM). The therapeutic efficacy of TCM herbs primarily stems from their complex specialized metabolites, such as alkaloids, terpenoids, and flavonoids. Notable examples include artemisinin (from *Artemisia annua* L.) for malaria, paclitaxel (originally from *Taxus* species) and vinblastine [from *Catharanthus roseus* (L.) G. Don] for cancer, strychnine (from *Strychnos nux-vomica* L.) as a pesticide, and anti-inflammatory agents like berberine (from *Coptis chinensis* Franch.) (Guo et al., 2022; Qin et al., 2025). These PNPs are typically secondary metabolites synthesized through conserved and lineage-specific biosynthetic pathways during the plant's evolution (Wang, Zhang, Li, & Zhao, 2024).

However, the sustainable use and large-scale production of these PNPs face significant challenges. Currently, most TCM active compounds rely on direct extraction from medicinal plants, which is limited by the low natural abundance of these compounds in plants and the technical complexity of their total chemical synthesis (Li, Li, Li, Gong, & Zhao, 2025). Understanding the biosynthetic pathways of these valuable PNPs is essential for addressing these challenges. Elucidating the biosynthetic genes and enzymes enables targeted metabolic engineering of PNPs in microorganisms or plant chassis (Qin et al., 2025). Herbgenomics provides a systematic and high-throughput approach to deciphering the genetic basis of PNP biosynthesis (Xin et al., 2019). By leveraging multi-omics integration (e.g., genomics, transcriptomics, and metabolomics), it facilitates the rapid discovery of candidate biosynthetic pathways and accelerates the functional characterization of key enzymatic steps (Chen et al., 2025; Sun, Xu, Song, & Chen, 2022; Xiao et al, 2025).

2. Progress of medicinal plant natural product biosynthesis

The structural diversity of pharmacologically active natural products in medicinal plants, such as terpenoids, flavonoids, alkaloids, and phenylpropanoids, originates from conserved upstream pathways that branch into species-specific modification networks

* Corresponding authors.
E-mail addresses: slchen@cdutcm.edu.cn (S. Chen), zcxu@nefu.edu.cn (Z. Xu).

https://doi.org/10.1016/j.chmed.2025.11.008
1674-6384/© 2025 Tianjin Press of Chinese Herbal Medicines. Published by ELSEVIER B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

(Fig. 1). Although core biosynthetic routes are increasingly mapped, the full enzymatic complexity (e.g. modifing enzymes) and regulatory mechanisms behind most compounds remain poorly understood.

Terpenoid biosynthesis begins with the methylerythritol phosphate (MEP) or mevalonate (MVA) pathways to produce basic isoprenoid precursors. Skeletal diversity is introduced by terpene synthase (TPS) and oxidosqualene cyclase (OSC) genes, while further structural modifications, like those observed in ginsenosides, astragalosides, and tanshinones, are mediated by enzymes including cytochrome P450s (CYP450s), UDP-glycosyltransferases (UGTs), methyltransferases (MTs), and acyltransferases (Lin et al., 2023). The examples include artemisinin (a sesquiterpene lactone from *A. annua*), which is synthesized from amorpha-4,11-diene via amorpha-4,11-diene synthase (ADS) and oxidized by CYP71AV1 and artemisinic aldehyde Δ11 (13)-double bond reductase 2 (DBR2) to form artemisinic acid before non-enzymatic conversion to artemisinin (Shen et al., 2018). Paclitaxel, an anticancer diterpenoid from *Taxus chinensis* (Pilg.) Rehder, is produced through a complex pathway involving over 20 enzymes, such as taxadiene synthase (TDS), CYP450s, and acyltransferases (Jiang et al., 2024). Ganoderic acids (triterpenoids from *Ganoderma lingzhi*) are synthesized via the key steps catalyzed by squalene synthase (SQS), lanosterol synthase (LSS), and CYP512A3 (Du

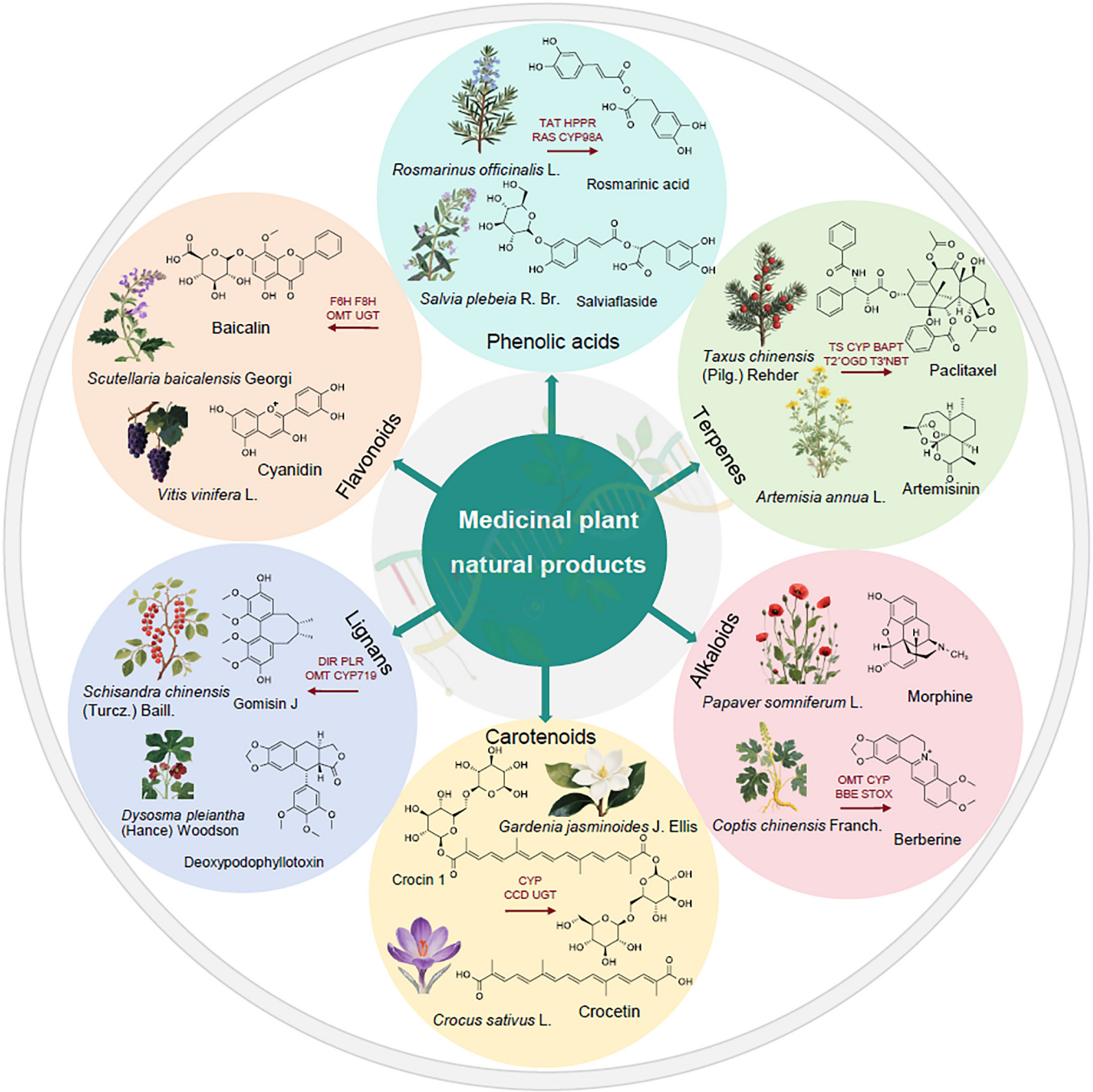

Fig. 1. Representative medicinal plants and their characteristic natural products. The schematic illustrates medicinal plant species and their key bioactive compounds, highlighting examples of structurally diverse natural products derived from specialized metabolic pathways.

et al., 2025). These diverse pathways highlight how enzymatic diversification gives rise to structurally and functionally distinct terpenoids with significant medicinal utility.

Flavonoid and lignan biosynthesis, two major branches of plant specialized metabolism, originate from the phenylpropanoid pathway and exhibit remarkable structural diversification through lineage-specific enzymatic modifications. The flavonoid pathway begins with the formation of naringenin catalyzed by key enzymes including chalcone synthase (CHS) and chalcone isomerase (CHI), followed by species-specific modifications (Wang, Zhang, Li, & Zhao, 2024). In *Scutellaria baicalensis* Georgi, flavone synthase II (FNSII), CYP82D, along with specific UGTs and *O*-methyltransferaseas (OMTs), collaboratively produce bioactive compounds such as baicalein and wogonin (Zhao et al., 2019; Gao et al., 2022). On the other hand, the lignan biosynthesis involves dirigent protein-mediated dimerization of monolignols and subsequent tailoring steps. In *Podophyllum* species, pinoresinol-lariciresinol reductase (PLR) and CYP719A23 are key to the formation of podophyllotoxin, a key precursor to the chemotherapeutic agent etoposide (Lau & Sattely, 2015). Likewise, in *Schisandra chinensis* (Turcz.) Baill., the biosynthesis of dibenzocyclooctadiene lignans depends on laccases (LACs), CYP719, and OMTs to introduce bioactivity-determining structural features (Liu et al., 2025). The evolution of new enzyme functions, coupled with the ability of a single enzyme to process multiple substrates (substrate promiscuity), drives the extensive structural diversity of these compounds.

Alkaloid biosynthesis exhibits exceptional structural heterogeneity, drawing from aromatic amino acids (e.g., tyrosine, tryptophan) or terpenoid precursors. Key enzymes include putrescine *N*-methyltransferase (PMT) and CYP80F1 in tropane alkaloid biosynthesis (e.g., hyoscyamine and scopolamine in nightshades) (Srinivasan & Smolke, 2020). Monoterpene indole alkaloid formation (e.g., vinblastine in *C. roseus*) requires strictosidine synthase (STR) and strictosidine *β-D*-glucosidase (SGD) (Grzech, Hong, Caputi, Sonawane, & O'Connor, 2023). Benzylisoquinoline alkaloids (e.g., morphine, berberine) involve norcoclaurine synthase (NCS), norcoclaurine 6-*O*-methyltransferase (6OMT/NOMT), oxidoreductases (STORR, [(*S*)- to (*R*)-reticuline]) and CYP719A enzymes formation (Tian et al., 2024; Xu et al., 2024). The evolution of fusion enzymes like STORR underscores the dynamic genomic mechanisms enabling chemical innovation (Catania et al., 2022). Similarly, colchicine synthesis in *Gloriosa superba* L. utilizes autumnaline synthase (CYP75B109), while lycopodium alkaloids derive from lysine via lysine decarboxylase (LDC) (Nett, Lau, & Sattely, 2020). The evolution of unique enzymes in different plants is a key mechanism for generating the vast structural diversity of alkaloids. This innovation is often driven by gene duplication and natural selection.

Across all major classes of specialized metabolites, including terpenoids, flavonoids, alkaloids, and lignans, diversification is driven by expansions within shared enzyme families: CYPs introduce oxidative modifications in all categories, UGTs mediate glycosylation of terpenoids, flavonoids, and alkaloids, and MTs methylate alkaloids and polyphenols. Although the core pathways forming basic scaffolds for terpenoids and alkaloids are established, the majority of putative UGTs and CYPs remain functionally uncharacterized. These enzymes are essential for determining the site-specific functionalization patterns that ultimately govern pharmacological activity (Song et al., 2024; Liu et al., 2024). As of now, biosynthetic pathways remain fragmented for over 90% of medicinal plant natural products. Future research should therefore focus on deciphering pathway regulatory networks and exploring the vast diversity of uncharacterized enzymes in medicinal plants.

3. Unraveling the key enzymes in natural product biosynthesis

The quest to decipher biosynthetic pathways in medicinal plants has undergone transformative technological evolution. Early efforts relied on biochemical approaches like radioisotope tracing (e.g., strychnine in *S. nux-vomica,* taxol in *T. chinensis*) and enzyme purification (Eisenreich, Menhard, Hylands, Zenk, & Bacher, 1996; Heimberger & Scott, 1973; Woodward, 1948). Subsequent advances incorporated molecular biology techniques for targeted gene cloning, exemplified by the isolation of key artemisinin pathway genes ADS and CYP71AV1 in *A. annua* via heterologous expression (Newman et al., 2006; Teoh, Polichuk, Reed, Nowak, & Covello, 2006). With the rise of Herbgenomics and comparative genomics, the discovery of biosynthetic genes has accelerated considerably (Chen et al., 2015). Genomics-driven strategies, like biosynthetic gene cluster (BGC) mining, co-expression analysis, homology-based screening, and integrated transcriptomic and metabolomic profiling, have been widely adopted to prioritize candidate genes.

BGC mining leverages chromosome-level genome assemblies to identify co-localized enzymatic modules, though such clusters remain relatively rare in medicinal plants. Notable examples include the noscapine gene cluster in *Papaver somniferum* L. (Winzer et al., 2012), taxadiene gene cluster in *Taxus* species (Zhang, Scossa, & Fernie, 2021); reticuline biosynthesis gene cluster in Ranunculales and Magnoliids (Shan et al., 2025); the vinblastine/vincristine pathway in *C. roseus* (Caputi et al., 2018), triterpenoid saponins biosynthesis gene clusters in *Aesculus chinensis* Bunge (Sun et al., 2023), and the tanshinone biosynthesis cluster in *Salvia miltiorrhiza* Bunge (Song et al., 2020). These BGCs typically comprise at least three non-homologous genes and span from tens to hundreds of kilobases. Historically, plant biosynthetic genes were thought to be dispersed throughout the genome, making BGCs uncommon, however, the availability of reference genomes has facilitated more systematic discovery of these clusters in the post-genomic era.

Transcriptomic co-expression analysis has dramatically accelerated gene discovery by correlating the expression of enzyme-encoding genes with metabolite accumulation across different tissues or developmental stages. This allows rapid prioritization of candidates for functional validation (Surendran, Ranjisha, Aswati Nair, & Pillai, 2022). Successful applications include identifying paclitaxel biosynthetic genes in *Taxus* species (Liang et al., 2025), ipecac alkaloid pathways in *Carapichea ipecacuanha* (Brot.) L. Andersson (Colinas et al., 2025), sinoacutine biosynthetic genes in *Menispermum dauricum* DC. (An et al., 2024), strychnine biosynthetic genes in *S. nux-vomica* (Hong et al., 2022), and astragaloside biosynthetic genes in *Astragalus membranaceus* (Fisch.) Bunge (Xu et al., 2024). Homology-based screening provides another powerful strategy, leveraging sequence conservation through alignment with functionally characterized genes or phylogenetic reconstruction using evolutionary models. For instance, ibogamine 10-hydroxylase was identified in *Tabernanthe iboga* Baill. based on its homology to tabersonine 16-hydroxylase from *C. roseus*, both catalyzing hydroxylation at the same position on the indole ring (Farrow et al., 2018).

As an increasing number of candidate genes are uncovered, functional validation becomes essential to fully elucidate biosynthetic pathways. Common approaches include heterologous expression in microbial systems (e.g., *Escherichia coli*, *Saccharomyces cerevisiae*) or plant hosts (e.g., *Nicotiana benthamiana* Domin) for pathway reconstruction, as well as genetic techniques such as clustered regularly interspaced short palindromic repeats (CRISPR), stable transformation, or virus-induced gene silencing (VIGS), followed by metabolomic profiling to confirm gene function (Lu, Martin-Hernandez, Peart, Malcuit, & Baulcombe, 2003;

Reed & Osbourn, 2018). Nevertheless, genetic manipulation remains challenging for many medicinal plants due to the lack of efficient transformation and regeneration protocols (Ferrer-Miralles & Garcia-Fruitós, 2024; Lambertz et al., 2014).

4. Advantages of Herbgenomics in elucidating natural product biosynthesis

Herbgenomics employs integrated multi-omics strategies to systematically decipher the complexity of biosynthetic pathways in medicinal plants. A key advantage lies in its ability to not only identify biosynthetic genes but also reveal the evolutionary mechanisms underpinning metabolic innovation. The remarkable diversity of plant-derived medicinal compounds derived is largely from genomic evolution processes, particularly expansion of enzymatic genes and functional diversition driven by whole-genome duplication (WGD), tandem gene duplication, and lineage-specific evolutionary events.

Phylogenomic analyses enable researchers to trace the origin and evolution of metabolic pathways across different clades, shedding light on how novel enzymes and compounds arise. WGD and tandem duplication events, which are widespread in medicinal plant genomes, provide extra copies of genes that could evolve new functions (Fig. 2). For instance, the genome of *A. annua* shows expanded sesquiterpene synthase families resulting from WGD, while tandem duplication of *CYP76AH* genes in *Salvia* species contributes to terpenoid structural variety (Chen et al., 2023; Lin et al., 2025). Convergent evolution offers another fascinating mechanism, whereby distantly related species independently evolve similar biosynthetic traits. For example, cannabinoid production occurs in both *Helichrysum umbraculigerum* Less. (Asterales) and *Cannabis sativa* L. (Rosales), berberine biosynthesis in *C. chinensis* (Ranunculales) and *Phellodendron amurense* Rupr. (Sapindales), as well as independent origins of ipecac alkaloid pathways in *C. ipecacuanha* (Gentianales) and *Alangium salviifolium* (L.f.) Wangerin (Cornales) (Berman et al., 2023; Colinas et al., 2025; Xu et al., 2024). The divergent evolution drives lineage-specific expansions. Tandem gene duplications drove the divergent evolution of caffeine and crocin biosynthetic pathways in coffee and *Gardenia jasminoides* J. Ellis (Xu et al., 2020).

Gene duplication often leads to functional specialization via subfunctionalization or neofunctionalization. An example of such diversification is observed in *Houttuynia cordata* Thunb., where duplicated genes including (*S*)-*N*-methylcoclaurine 5-*N*-methyltransferase (*NMT5*) and *CYP80G2* have undergone neo- or sub-functionalization. The neofunctionalized NMT5 specifically catalyzes the production of di-methylated coclaurine, while CYP80G2 produces isoboldine (Shan et al., 2025). In *Menispermum canadense* L., the dechloroacutumine halogenase (*DAH*) evolved via tandem duplication of a flavonol synthase (*FLS*) gene, followed by neofunctionalization and gene loss events. DAH acquired 2OG-dependent chlorinase activity to convert dechloroacutumine to acutumine, while retaining weak ancestral FLS activity toward dihydrokaempferol. FLS lacks oxidase or halogenase activity on dechloroacutumine (Kim et al., 2025). Duplication and divergence of oxidosqualene cyclase (*OSC*) genes allow plants to generate hundreds of distinct triterpene scaffolds, which serve as precursors for a vast array of specialized metabolites (Ji, Osbourn, & Liu, 2024). These evolutionary insights provide a foundation for designing novel bioactive compounds.

Herbgenomics also greatly accelerates the mining of BGCs, which are genomic loci where biosynthetic genes are physically clustered. Well-characterized BGCs include those for saponin pathways (*BAS-CYP716A-UGT87A*) in Sapindaceae (Mai et al., 2025), coumarin biosynthesis (*C2′H-PT-OC*) in Apiaceae (Huang et al., 2024), morphinan biosynthetic gene cluster (*STORR-SALSYN-SALAT-SALR-THS*) in *P. somniferum* (Yang et al., 2021), and tanshinones biosynthetic gene cluster (*CPS-KSL-CYP76AH*) in

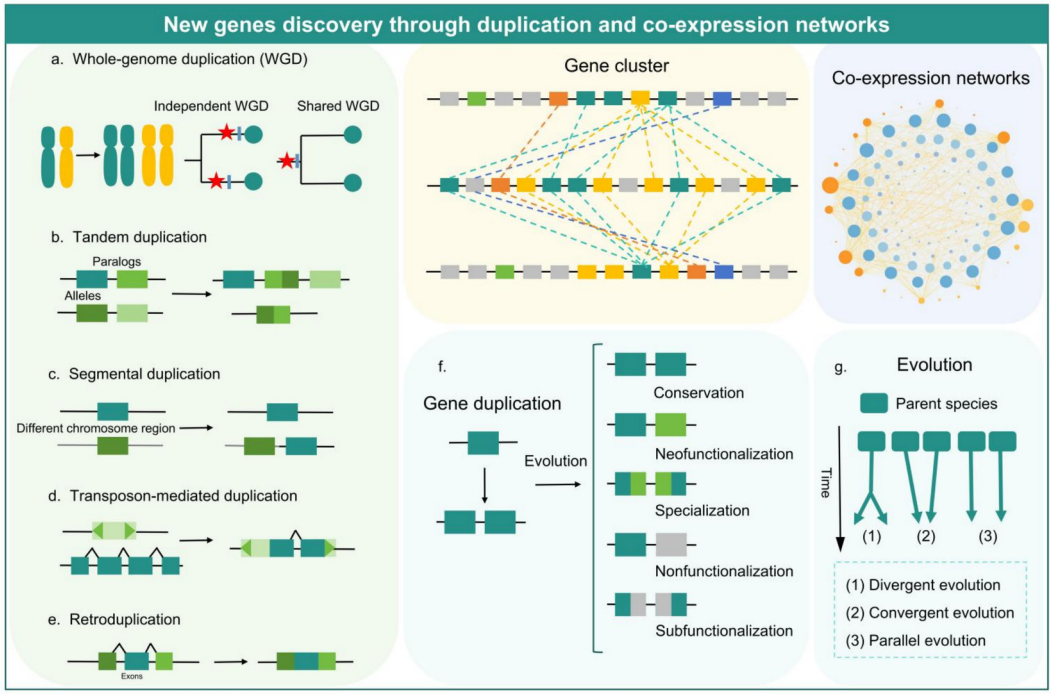

Fig. 2. News genes discovery through duplication and co-expression networks. Mechanisms of gene duplication and their roles in the evolution of new genes. (a) WGD or polyploidy to produce duplicate genes. (b) Tandem duplication: duplication of a gene via unequal crossing-over between similar alleles. (c) Segmental duplication: multiple genes through polyploidy followed by chromosome rearrangements. (d) Transposon-mediated duplication, in which a DNA transposon acquires fragments of genes. (e) Retroduplication, which is then randomly inserted into a chromosome to form a new duplication gene. (f) Schematic representation of the evolutionary fates of genes following duplication events. (g) Illustration of the evolutionary trajectories that shape metabolic pathway diversity in descendant species.

S. miltiorrhiza (Song et al., 2020). Such clusters likely arise through gene duplication, neofunctionalization, and genomic reorganization. Additionally, spatio-temporal omics approaches further enhance resolution by uncovering compartmentalized biosynthesis. For example, multi-tissue transcriptomics and single-cell analyses have elucidated cell-specific gene expression associated with monoterpenoid indole alkaloid accumulation in *C. roseus* and ginsenoside production in *Panax ginseng* C. A. Mey leaves (Dangwal, Suri, & Kaur, 2025; Wang et al., 2025). Similarly, integrated genomics, single-cell transcriptomics, and mass spectrometry imaging have revealed the spatial distribution and biosynthesis of ganoderic acids in *Ganoderma lingzhi* (Du et al., 2025).

The rapid advancement in medicinal plant genomics provides an unprecedented resource for evolutionary and functional studies. To date, over 400 medicinal plant genomes have been sequenced, such as *S. miltiorrhiza*, *Panax notoginseng* (Burkill) F. H. Chen, *Andrographis paniculata* (Burm.f.) Nees, *Lonicera caerulea* L., and *Dendrobium devonianum* Paxton (Sun et al., 2019; Jiang et al., 2021; Song et al., 2020; Wang, et al., 2025). Large-scale initiatives like the 1 K Herb Plant Genome Project and artificial intelligence (AI)-powered repositories such as the Gene-encoded Natural Diverse Components Repository (GNDC) are systematically expanding genomic and metabolomic data across medicinal species, enabling deeper insights into genetic encoding mechanisms and natural chemodiversity (Su et al., 2022; Chen et al., 2025). In summary, Herbgenomics has transformed the paradigm of pathway elucidation from fragmented gene discovery to systems-level and evolutionary-aware investigation. Through comparative genomics, phylogenomics, and multi-omics integration, researchers can now reconstruct the evolutionary trajectory of metabolic pathways and uncover lineage-specific adaptations. This holistic approach not only deciphers biosynthetic complexity but also bridges traditional knowledge with modern predictive research, ushering in a new era for natural product discovery and synthetic biology.

5. Future prospects and conclusions

Despite transformative advances, Herbgenomics continues to face significant challenges that hinder the comprehensive elucidation of natural product biosynthesis. A major limitation lies in the prevalence of incomplete genome assemblies and functional annotations. High-quality genomic resources remain scarce for most medicinal plants, with a pronounced taxonomic bias: economically important crops are well-represented, while non-model medicinal

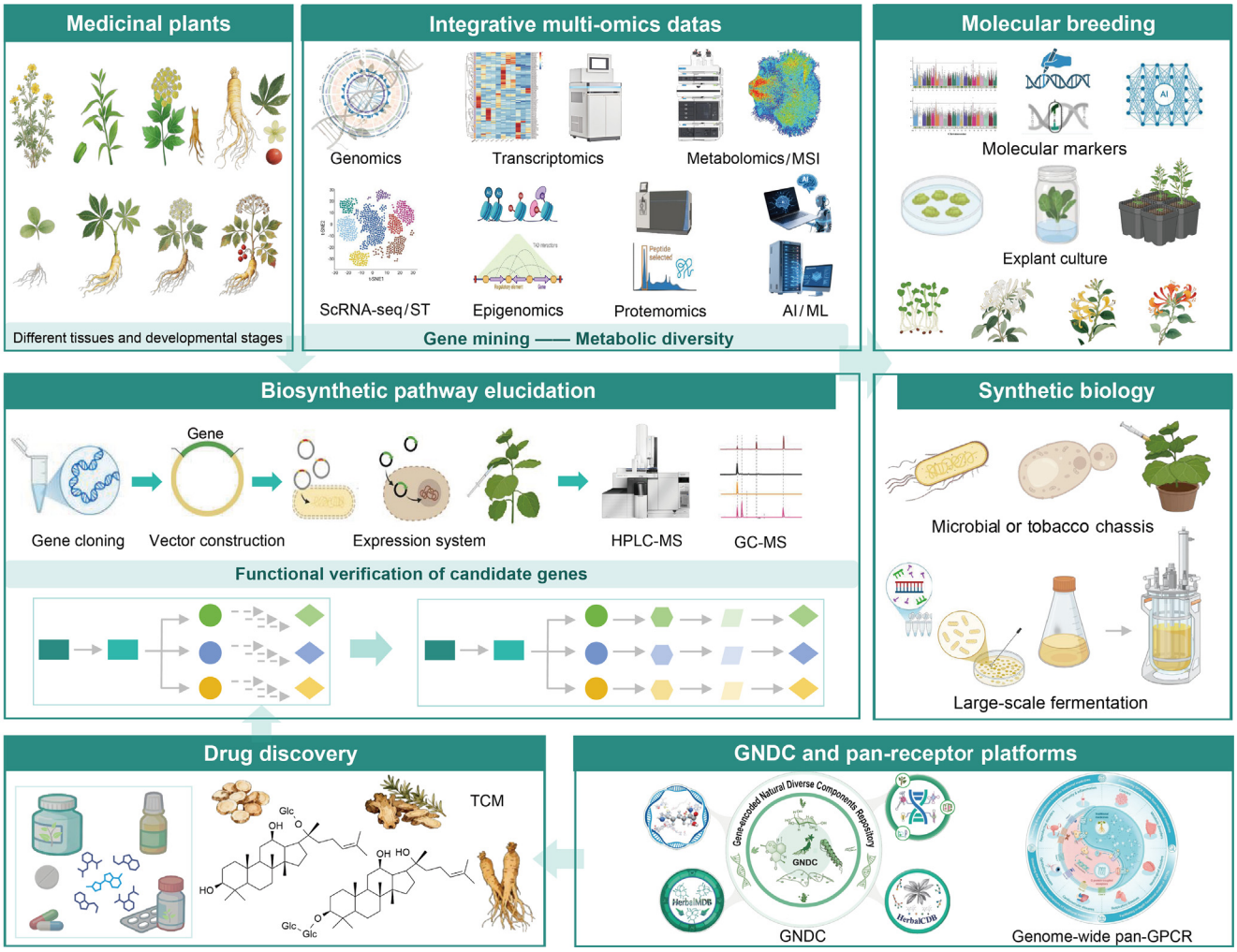

Fig. 3. An integrative multi-omics framework for gene discovery, pathway elucidation, and biotechnological application in medicinal plants. The framework integrates multi-omics data, including genomics, transcriptomics, and metabolomics, obtained from diverse tissues and developmental stages. AI/ML methods are employed to analyze these data, facilitating gene mining and metabolic diversity assessment to elucidate biosynthetic pathways. Candidate genes are functionally validated via gene cloning, heterologous expression, and metabolite profiling. The resulting insights support applications in molecular breeding and synthetic biology, enabling sustainable production of bioactive natural products. The integration of GNDC and pan-GPCR platforms further enhances drug discovery by enabling systematic screening of bioactive compounds and their targets.

species are severely neglected (Cheng, Wang, Zhu, Ye, & Ye, 2025). Furthermore, over 600 species often possess colossal genomes (>20 Gb/1C), such as *Paris polyphylla* var. *yunnanensis* (Franch.) Hand.-Mazz (54.58 Gb), *Lilium davidii* var. *unicolor* (Hoog) Cotton (35.60 Gb) and *Fritillaria cirrhosa* D. Don (42 Gb) (Zeng et al., 2025; Li et al., 2020; Xu et al., 2024). However, integration with multi-omics data, such as transcriptomes from various tissues and developmental stages, single-cell transcriptomics, mass spectrometry imaging, proteomics, and epigenomics, will accelerate the discovery of novel enzymes, regulators, and transporters (Cho et al., 2025; Sanches, de Melo, Porcari, & de Carvalho, 2024). This comprehensive approach enables the mapping of detailed biosynthetic pathways, providing precise blueprints for synthetic biology applications.

AI is revolutionizing biosynthetic enzyme annotation and pathway prediction (Reinhardt, Craft, & Weng, 2025). AI accelerates compound screening and improves molecular property prediction accuracy. Meanwhile, machine learning (ML) techniques and advanced computational tools empower researchers to overcome traditional barriers and explore PNPs resources more efficiently (Hannigan et al., 2019; Niranjan, Uttarkar, Kaul, & Varghese, 2023). For example, the deep learning for biosynthetic gene cluster prediction and network models for gene-metabolite association, are revolutionizing enzyme annotation and pathway prediction. AlphaFold-guided molecular docking recently identified critical genes for bioactive compound biosynthesis, bypassing laborious mutagenesis screens (Lyu et al., 2024). Recent advances in multi-omics approaches have enabled deeper understanding of the intricate networks governing plant metabolism and biosynthesis, which integrates data from genomics, transcriptomics, and metabolomics, analyzes them via AI/ML to facilitate gene discovery and pathway elucidation, and ultimately enables metabolic engineering and molecular breeding (Fig. 3). These genomic insights are now being translated into sustainable production platforms through synthetic biology approaches (Rai, Mishra, Mishra, Rai, & Singh, 2025). For compounds dependent on plant-specific posttranslational modifications, CRISPR-enhanced plant cell cultures, e.g., by editing regulatory genes to enhance pathway flux, provide viable production alternatives (Kumar & Jain, 2015).

Notably, AI and multi-omics technologies have revolutionized natural product and drug discovery. For instance, the GNDC catalogs over 234 million natural components from thousands of herb genomic datasets, establishing a big-data-driven paradigm for drug discovery (Chen et al., 2025). Additionally, genome-wide pan-receptor drug discovery platforms (e.g., pan-GPCR) offer an innovative strategy to explore comprehensive interactions between traditional medicines and the GPCRome, facilitating the elucidation of the multi-component/multi-target characteristics of TCM (Bi et al., 2025; Yang et al., 2024). The integration of GNDC and pan-GPCR platforms enables the identification of novel, high-potency natural compounds and guides the green production of drug molecules. Collectively, Herbgenomics has built a robust bridge linking traditional medicine with modern life sciences, providing a data-driven framework that facilitates natural product research and accelerates the discovery and development of herbal medicine.

CRediT authorship contribution statement

Jing Wang: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. **Jingyuan Song:** Conceptualization, Supervision, Writing – review & editing. **Shilin Chen:** Conceptualization, Writing – review & editing. **Zhichao Xu:** Conceptualization, Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by Innovation Team and Talents Cultivation Program of National Administration of Traditional Chinese Medicine (No. ZYYCXTD-D-202408).

References

An, Z. J., Gao, R. R., Chen, S. S., Tian, Y., Li, Q., Tian, L. X., et al. (2024). Lineage-specific CYP80 expansion and benzylisoquinoline alkaloid diversity in early-diverging eudicots. *Advanced Science, 11*(19), e2309990.

Berman, P., de Haro, L. A., Jozwiak, A., Panda, S., Pinkas, Z., Dong, Y., et al. (2023). Parallel evolution of cannabinoid biosynthesis. *Nature Plants, 9*(5), 817–831.

Bi, Z. H., Li, H., Liang, Y. T., Sun, D., Liu, S. X., Chen, W., et al. (2025). Emerging paradigms for target discovery of traditional medicines: A genome-wide pan-GPCR perspective. *The Innovation, 6*(3), 100774.

Caputi, L., Franke, J., Farrow, S. C., Chung, K., Payne, R. M. E., Nguyen, T. D., et al. (2018). Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science, 360*(6394), 1235–1239.

Catania, T., Li, Y., Winzer, T., Harvey, D., Meade, F., Caridi, A., et al. (2022). A functionally conserved STORR gene fusion in *Papaver* species that diverged 16.8 million years ago. *Nature Communications, 13*, 3150.

Chen, H. Y., Guo, M. X., Dong, S. T., Wu, X. L., Zhang, G. B., He, L., et al. (2023). A chromosome-scale genome assembly of *Artemisia argyi* reveals unbiased subgenome evolution and key contributions of gene duplication to volatile terpenoid diversity. *Plant Communications, 4*(3), 100516.

Chen, S. L., Song, J. Y., Sun, C., Xu, J., Zhu, Y. J., Verpoorte, R., et al. (2015). Herbal genomics: Examining the biology of traditional medicines. *Science, 347*(6219), S27–S29.

Chen, W., Yu, Z. Y., Leng, L., Sun, D., Liu, H., Gong, R. Z., et al. (2025). Artificial intelligence-curated repository of gene-encoded natural diverse components from herbal medicines. *The Innovation, 101011*.

Cheng, L. T., Wang, Z. L., Zhu, Q. H., Ye, M., & Ye, C. Y. (2025). A long road ahead to reliable and complete medicinal plant genomes. *Nature Communications, 16*(1), 2150.

Cho, Y., Kadam, U., Park, B., Amariillis, S., Nguyen, K. T., Can, M. T., et al. (2025). Recent progress in single-cell transcriptomic studies in plants. *Plant Biotechnology Reports, 19*(2), 91–103.

Colinas, M., Morweiser, C., Dittberner, O., Chioca, B., Alam, R., Leucke, H., et al. (2025). Ipecac alkaloid biosynthesis in two evolutionarily distant plants. *Nature Chemical Biology, 21*(11), 1794–1805.

Dangwal, M., Suri, G. S., & Kaur, G. (2025). Single-cell multi-omics in the medicinal plant *Catharanthus roseus*: A new era of next generation therapeutics. *Journal of Plant Biochemistry and Biotechnology, 34*(1), 6–12.

Du, Y. P., Peng, S., Chen, H. G., Li, J., Huang, F. Y., Chen, W. X., et al. (2025). Unveiling the spatiotemporal landscape of *Ganoderma lingzhi*: Insights into ganoderic acid distribution and biosynthesis. *Engineering*. https://doi.org/10.1016/j.eng.2025.03.030.

Eisenreich, W., Menhard, B., Hylands, P. J., Zenk, M. H., & Bacher, A. (1996). Studies on the biosynthesis of taxol: The taxane carbon skeleton is not of mevalonoid origin. *Proceedings of the National Academy of Sciences of the United States of America, 93*(13), 6431–6436.

Farrow, S. C., Kamileen, M. O., Meades, J., Ameyaw, B., Xiao, Y. L., & O'Connor, S. E. (2018). Cytochrome P450 and *O*-methyltransferase catalyze the final steps in the biosynthesis of the anti-addictive alkaloid ibogaine from *Tabernanthe iboga*. *Journal of Biological Chemistry, 293*(36), 13821–13833.

Ferrer-Miralles, N., & Garcia-Fruitós, E. (2024). Heterologous expression of difficult to produce proteins in bacterial systems. *International Journal of Molecular Sciences, 25*(2), 822.

Gao, R. R., Lou, Q., Hao, L. J., Qi, G. H., Tian, Y., Pu, X. D., et al. (2022). Comparative genomics reveal the convergent evolution of CYP82D and CYP706X members related to flavone biosynthesis in Lamiaceae and Asteraceae. *Plant Journal, 109*(5), 1305–1318.

Grzech, D., Hong, B. K., Caputi, L., Sonawane, P. D., & O'Connor, S. E. (2023). Engineering the biosynthesis of late-stage vinblastine precursors precondylocarpine acetate, catharanthine, tabersonine in *Nicotiana benthamiana*. *ACS Synthetic Biology, 12*(1), 27–34.

Guo, L., Yao, H., Chen, W. K., Wang, X. M., Ye, P., Xu, Z. C., et al. (2022). Natural products of medicinal plants: Biosynthesis and bioengineering in post-genomic era. *Horticulture Research, 9*, uhac223.

Hannigan, G. D., Prihoda, D., Palicka, A., Soukup, J., Klempir, O., Rampula, L., et al. (2019). A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Research, 47*(18), e110.

Heimberger, S. I., & Scott, A. I. (1973). Biosynthesis of strychnine. *Journal of the Chemical Society, Chemical Communications, 6*, 217.

Hong, B. K., Grzech, D., Caputi, L., Sonawane, P., López, C. E. R., Kamileen, M. O., et al. (2022). Biosynthesis of strychnine. *Nature, 607*(7919), 617–622.
Huang, X. C., Tang, H. Y., Wei, X. F., He, Y. D., Hu, S. Y., Wu, J. Y., et al. (2024). The gradual establishment of complex coumarin biosynthetic pathway in Apiaceae. *Nature Communications, 15*, 6864.
Ji, W. J., Osbourn, A., & Liu, Z. H. (2024). Understanding metabolic diversification in plants: Branchpoints in the evolution of specialized metabolism. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 379*(1914), 20230359.
Jiang, B., Gao, L., Wang, H. J., Sun, Y. P., Zhang, X. L., Ke, H., et al. (2024). Characterization and heterologous reconstitution of *Taxus* biosynthetic enzymes leading to baccatin III. *Science, 383*(6683), 622–629.
Jiang, Z. Q., Tu, L. C., Yang, W. F., Zhang, Y. F., Hu, T. Y., Ma, B. W., et al. (2021). The chromosome-level reference genome assembly for *Panax notoginseng* and insights into ginsenoside biosynthesis. *Plant Communications, 2*(1), 100113.
Kim, C. Y., Kastner, D. W., Mitchell, A. J., Gutierrez, M. A., Yao, J. S., Neumann, E. N., et al. (2025). Tracing the stepwise Darwinian evolution of a plant halogenase. *Science Advances, 11*(33), eadv6898.
Kumar, V., & Jain, M. (2015). The CRISPR-Cas system for plant genome editing: Advances and opportunities. *Journal of Experimental Botany, 66*(1), 47–57.
Lambertz, C., Garvey, M., Klinger, J., Heesel, D., Klose, H., Fischer, R., et al. (2014). Challenges and advances in the heterologous expression of cellulolytic enzymes: A review. *Biotechnology for Biofuels, 7*(1), 135.
Lau, W., & Sattely, E. S. (2015). Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science, 349*(6253), 1224–1228.
Li, J., Lv, M. Q., Du, L., Yunga, A., Hao, S. J., & Zhang, Y. L., et al. (2020). An enormous *Paris polyphylla* genome sheds light on genome size evolution and polyphyllin biogenesis. *bioRxiv*, 2020.06.01.126920.
Li, L., Li, Q., Li, Y. X., Gong, D. D., & Zhao, B. N. (2025). Research progress and challenges of molecular recognition techniques in the screening of active ingredients in traditional Chinese medicine. *Journal of Pharmaceutical Analysis, 15*(9), 101243.
Liang, F. Y., Xie, Y. M., Zhang, C., Zhao, Y., Motawia, M. S., & Kampranis, S. C. (2025). Elucidation of the final steps in Taxol biosynthesis and its biotechnological production. *Nature Synthesis, 4*(10), 1212–1222.
Lin, J. J., Yin, X., Zeng, Y. R., Hong, X. Y., Zhang, S. C., Cui, B. M., et al. (2023). Progress and prospect: Biosynthesis of plant natural products based on plant chassis. *Biotechnology Advances, 69*, 108266.
Lin, R. X., Li, H. X., Xiao, Y. R., Wang, Z., Liu, L. C., Saalbach, G., et al. (2025). Three cytochrome P450 enzymes consecutively catalyze the biosynthesis of furanoclerodane precursors in *Salvia* species. *Plant Communications, 6*(5), 101286.
Liu, J. S., Xiong, R. L., Liu, L. J., Zhou, T. P., Xue, J. Y., Sun, D., et al. (2025). Insights into angiosperm evolution and lineage-specialized lignan biosynthesis from the early-diverging *Schisandra* genome. *Science Advances, 11*(33), eadw0486.
Liu, R., Wu, X. Y., Jiang, Z. Q., Liu, X., Zhang, Y. F., Zhao, H., et al. (2024). Characterization of a xylosyltransferase from *Panax notoginseng* catalyzing ginsenoside 2′-*O* glycosylation in the biosynthesis of notoginsenosides. *Journal of Natural Products, 87*(9), 2160–2169.
Lu, R., Martin-Hernandez, A. M., Peart, J. R., Malcuit, I., & Baulcombe, D. C. (2003). Virus-induced gene silencing in plants. *Methods, 30*(4), 296–303.
Lyu, J. K., Kapolka, N., Gumpper, R., Alon, A., Wang, L., Jain, M. K., et al. (2024). AlphaFold2 structures guide prospective ligand discovery. *Science, 384*(6702), eadn6354.
Mai, Y. X., Hu, H. M., Ji, W. J., Xiao, Y. X., Zhou, H., Zeng, Z. H., et al. (2025). Evolution and functional characterization of a biosynthetic gene cluster for saponin biosynthesis in Sapindaceae. *Molecular Plant, 18*(7), 1089–1093.
Nett, R. S., Lau, W., & Sattely, E. S. (2020). Discovery and engineering of colchicine alkaloid biosynthesis. *Nature, 584*(7819), 148–153.
Newman, J. D., Marshall, J., Chang, M., Nowroozi, F., Paradise, E., Pitera, D., et al. (2006). High-level production of amorpha-4, 11-diene in a two-phase partitioning bioreactor of metabolically engineered *Escherichia coli*. *Biotechnology and Bioengineering, 95*(4), 684–691.
Niranjan, V., Uttarkar, A., Kaul, A., & Varghese, M. (2023). A machine learning-based approach using multi-omics data to predict metabolic pathways. *Methods in Molecular Biology, 2553*, 441–452.
Qin, K. Z., Liu, F., Zhang, C. B., Deng, R., Fernie, A. R., & Zhang, Y. J. (2025). Systems and synthetic biology for plant natural product pathway elucidation. *Cell Reports, 44*(6), 115715.
Rai, K. K., Mishra, J., Mishra, A. K., Rai, A. C., & Singh, P. K. (2025). Synthetic and systems biology strategies for enhanced plant secondary metabolite production. *Environmental Sustainability, 8*(2), 203–222.
Reed, J., & Osbourn, A. (2018). Engineering terpenoid production through transient expression in *Nicotiana benthamiana*. *Plant Cell Reports, 37*(10), 1431–1441.
Reinhardt, J. K., Craft, D., & Weng, J. K. (2025). Toward an integrated omics approach for plant biosynthetic pathway discovery in the age of AI. *Trends in Biochemical Sciences, 50*(4), 311–321.
Sanches, P. H. G., de Melo, N. C., Porcari, A. M., & de Carvalho, L. M. (2024). Integrating molecular perspectives: Strategies for comprehensive multi-omics integrative data analysis and machine learning applications in transcriptomics, proteomics, and metabolomics. *Biology, 13*(11), 848.
Shan, C. H., Zhou, X., Zhu, J. J., Rashid, A., Wang, J. H., An, N., et al. (2025). Gene duplication and clustering underlie the conservation and diversification of benzylisoquinoline alkaloid biosynthesis in plants. *Nature Communications, 16*, 7669.
Shen, Q., Zhang, L. D., Liao, Z. H., Wang, S. Y., Yan, T. X., Shi, P., et al. (2018). The genome of *Artemisia annua* provides insight into the evolution of Asteraceae family and artemisinin biosynthesis. *Molecular Plant, 11*(6), 776–788.
Song, Y. T., Zhang, Y. T., Wang, X., Yu, X. K., Liao, Y., Zhang, H., et al. (2024). Telomere-to-telomere reference genome for *Panax ginseng* highlights the evolution of saponin biosynthesis. *Horticulture Research, 11*(6), uhae107.
Song, Z. Q., Lin, C. C., Xing, P. Y., Fen, Y. Y., Jin, H., Zhou, C. H., et al. (2020). A high-quality reference genome sequence of *Salvia miltiorrhiza* provides insights into tanshinone synthesis in its red rhizomes. *The Plant Genome, 13*(3), e20041.
Srinivasan, P., & Smolke, C. D. (2020). Biosynthesis of medicinal tropane alkaloids in yeast. *Nature, 585*(7826), 614–619.
Su, X. J., Yang, L. L., Wang, D. L., Shu, Z. Q., Yang, Y. C., Chen, S. L., et al. (2022). 1 K medicinal plant genome database: An integrated database combining genomes and metabolites of medicinal plants. *Horticulture Research, 9*, uhac075.
Sun, W., Leng, L., Yin, Q. G., Xu, M. M., Huang, M. K., Xu, Z. C., et al. (2019). The genome of the medicinal plant *Andrographis paniculata* provides insight into the biosynthesis of the bioactive diterpenoid neoandrographolide. *Plant Journal, 97*(5), 841–857.
Sun, W., Xu, Z. C., Song, C., & Chen, S. L. (2022). Herbgenomics: Decipher molecular genetics of medicinal plants. *The Innovation, 3*(6), 100322.
Sun, W., Yin, Q. G., Wan, H. H., Gao, R. R., Xiong, C., Xie, C., et al. (2023). Characterization of the horse chestnut genome reveals the evolution of aescin and aesculin biosynthesis. *Nature Communications, 14*, 6470.
Surendran, K., Ranjisha, K. R., Aswati Nair, R., & Pillai, P. P. (2022). Genomics and metabolomics: A strategy for elucidation of metabolic pathways in medicinal plants. In *Phytochemical Genomics: Plant Metabolomics and Medicinal Plant Genomics* (pp. 343–360). Singapore: Springer Nature Singapore.
Teoh, K. H., Polichuk, D. R., Reed, D. W., Nowak, G., & Covello, P. S. (2006). *Artemisia annua* L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. *FEBS Letters, 580*(5), 1411–1416.
Tian, Y., Kong, L. Z., Li, Q., Wang, Y. F., Wang, Y. M., An, Z. J., et al. (2024). Structural diversity, evolutionary origin, and metabolic engineering of plant specialized benzylisoquinoline alkaloids. *Natural Product Reports, 41*(11), 1787–1810.
Wang, J., Wang, X. T., Ma, Y. W., Gao, R. R., Wang, Y. M., An, Z. J., et al. (2025). *Lonicera caerulea* genome reveals molecular mechanisms of freezing tolerance and anthocyanin biosynthesis. *Journal of Advanced Research, 76*, 293–305.
Wang, J., Zhou, Y., Zhang, M. C., Li, X. P., Liu, T. X., Liu, Y. L., et al. (2025). Resolving floral development dynamics using genome and single-cell temporal transcriptome of *Dendrobium devonianum*. *Plant Biotechnology Journal, 23*(8), 2997–3011.
Wang, M. C., Zhang, S. Q., Li, R., & Zhao, Q. (2024). Unraveling the specialized metabolic pathways in medicinal plant genomes: A review. *Frontiers in Plant Science, 15*, 1459533.
Wang, R. B., Zhang, L. Y., Li, X., Sun, W. S., Wang, Y. K., Huang, L. Q., et al. (2025). Single-cell sequencing and mass spectrometry imaging reveal the multicellular compartmentalization map of plant triterpenes: A ginsenoside in *Panax ginseng* example. *Plant Biotechnology Journal, 23*(10), 4461–4476.
Winzer, T., Gazda, V., He, Z. S., Kaminski, F., Kern, M., Larson, T. R., et al. (2012). A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Science, 336*(6089), 1704–1708.
Woodward, R. B. (1948). Biogenesis of the *Strychnos* alkaloids. *Nature, 162*(4108), 155.
Xiao, S. M., Chu, Y., Liao, B. S., Cai, S. Y., Liao, X. J., Cao, X. F., et al. (2025). Application of herbgenomics in evaluation and control of quality uniformity of Chinese medicinal materials. *Chinese Traditional and Herbal Drugs, 56*(9), 3291–3304.
Xin, T. Y., Zhang, Y., Pu, X. D., Gao, R. R., Xu, Z. C., & Song, J. Y. (2019). Trends in herbgenomics. *Science China Life Sciences, 62*(3), 288–308.
Xu, B. Y., Huang, J. P., Peng, G. Q., Cao, W. Y., Liu, Z., Chen, Y., et al. (2024). Total biosynthesis of the medicinal triterpenoid saponin astragalosides. *Nature Plants, 10*(11), 1826–1837.
Xu, S. J., Chen, R. Z., Zhang, X. Q., Wu, Y. F., Yang, L. Y., Sun, Z. Y., et al. (2024). The evolutionary tale of lilies: Giant genomes derived from transposon insertions and polyploidization. *The Innovation, 5*(6), 100726.
Xu, Z. C., Pu, X. D., Gao, R. R., Demurtas, O. C., Fleck, S. J., Richter, M., et al. (2020). Tandem gene duplications drive divergent evolution of caffeine and crocin biosynthetic pathways in plants. *BMC Biology, 18*(1), 63.
Xu, Z. C., Tian, Y., Wang, J., Ma, Y. W., Li, Q., Zhou, Y. Z., et al. (2024). Convergent evolution of berberine biosynthesis. *Science Advances, 10*(48), eads3596.
Yang, H. T., Wang, Y. F., Liu, W., He, T. P., Liao, J. Y., Qian, Z. Z., et al. (2024). Genome-wide pan-GPCR cell libraries accelerate drug discovery. *Acta Pharmaceutica Sinica B, 14*(10), 4296–4311.
Yang, X. F., Gao, S. H., Guo, L., Wang, B., Jia, Y. Y., Zhou, J., et al. (2021). Three chromosome-scale *Papaver* genomes reveal punctuated patchwork evolution of the morphinan and noscapine biosynthesis pathway. *Nature Communications, 12*, 6030.
Zeng, P., Zong, H., Han, Y. W., Zhang, W. X., Tian, Z. Z., Zhou, B. T., et al. (2025). Two Melanthiaceae genomes with dramatic size difference provide insights into giant genome evolution and maintenance. *Nature Plants, 11*(8), 1500–1513.
Zhang, Y. J., Scossa, F., & Fernie, A. R. (2021). The genomes of *Taxus* species unveil novel candidates in the biosynthesis of taxoids. *Molecular Plant, 14*(11), 1773–1775.
Zhao, Q., Yang, J., Cui, M. Y., Liu, J., Fang, Y. M., Yan, M. X., et al. (2019). The reference genome sequence of *Scutellaria baicalensis* provides insights into the evolution of wogonin biosynthesis. *Molecular Plant, 12*(7), 935–950.

Shilin Chen is the Chief Professor at Chengdu University of Traditional Chinese Medicine. He proposed the discipline of "Herbgenomics" and initiated the 1K Herb Genomics Project. His research focuses on the AI- and genomics-based drug discovery, biosynthesis of natural products and molecular-assisted breeding of medicinal plants. In 2023, he was elected an Academician of the Chinese Academy of Engineering.