



Contents lists available at ScienceDirect

Engineering

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

Research
1K Herb Genomes Project—Review

Plant Terpenoid Diversity: From Pathway Elucidation to Ecological Functions

Xiaochen Wang^a, Guodong Wang^{a,b,*}

^a State Key Laboratory of Seed Innovation, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

^b College of Advanced Agricultural Sciences, Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:

Received 18 May 2025

Revised 16 November 2025

Accepted 25 November 2025

Available online xxxx

Keywords:

Plant terpenoid

Multi-omics

Biosynthesis

Functional characterization

ABSTRACT

Terpenoids are the most diverse class of specialized metabolites in plants, despite being derived from two simple precursors, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). Terpene biosynthesis can be broadly divided into two stages: the formation of upstream carbon–hydrogen scaffolds and the downstream modification of these scaffolds. In this review, we first highlight recent advances in the biosynthesis of the upstream carbon–hydrogen backbones. We then discuss the techniques and methodologies used to elucidate the molecular details of the downstream modification reactions, particularly in the context of medicinal plants. Finally, we explore and forecast the future of biological function studies of plant terpenoids, with a specific focus on their roles *in planta*. This review aims to provide a comprehensive overview of current research trends and insights into terpene biosynthesis and their functional characterization in plants.

© 2025 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The study of terpene biosynthesis has a long history, dating back to the early 20th century, when pioneering researchers, including Nobel laureate Otto Wallach, made significant contributions to understanding terpene structural diversity. Major breakthroughs occurred in the 1950s, with the discovery that isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) serve as universal precursors for all terpenoids [1]. This laid the foundation for elucidating terpene biosynthetic pathways, leading to the identification of the mevalonic acid (MVA) and methylerythritol phosphate (MEP) pathways responsible for DMAPP/IPP synthesis [2]. The discovery of bioactive plant-derived terpenoids, such as the anticancer drug paclitaxel, the anti-malarial compound artemisinin, and phytohormones, including gibberellins, abscisic acid (ABA), brassinosteroids (BRs), strigolactones, and cytokinins, which contain one dimethylallyl group, further stimulated interest in terpene biosynthesis.

Roughly, diverse plant-based terpenes can be classified into monoterpenes (C₁₀), sesquiterpenes (C₁₅), diterpenes (C₂₀), sesterterpenes (C₂₅), triterpenes (C₃₀), tetraterpenes (C₄₀), and polyterpe-

nes [3]. The meroterpenes, which are partially derived from terpenoid biosynthetic pathways, are also included in this review [4]. Advances in molecular biology and chemical analysis have enabled the identification and functional characterization of key biosynthetic enzymes, such as prenyltransferases, terpene synthases (TPSs), and cytochrome P450s (Fig. 1). With the advent of high-throughput sequencing and sophisticated analytical technologies in the twentieth century, researchers have gained unprecedented insights into the genetic and biochemical mechanisms underlying terpene biosynthesis and its regulation in response to developmental and environmental cues.

Currently, terpene biosynthesis is a highly dynamic area of research, focusing on pathway elucidation, metabolic engineering, functional genomics, and the exploration of terpenoids for pharmaceutical, agricultural, and industrial applications. In this review, we summarize the latest advancements in plant terpenoid research from the following aspects.

2. New discoveries from five-carbon-atom (C₅) building blocks to terpenoid scaffolds

After the formation of C₅ building blocks (DMAPP and IPP) from both the MEP and MVA pathways, the C₅ building blocks are subsequently catalyzed by isoprenyl diphosphate synthase (IDS, also

* Corresponding author.

E-mail address: gdwang@genetics.ac.cn (G. Wang).

<https://doi.org/10.1016/j.eng.2025.11.021>

2095-8099/© 2025 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

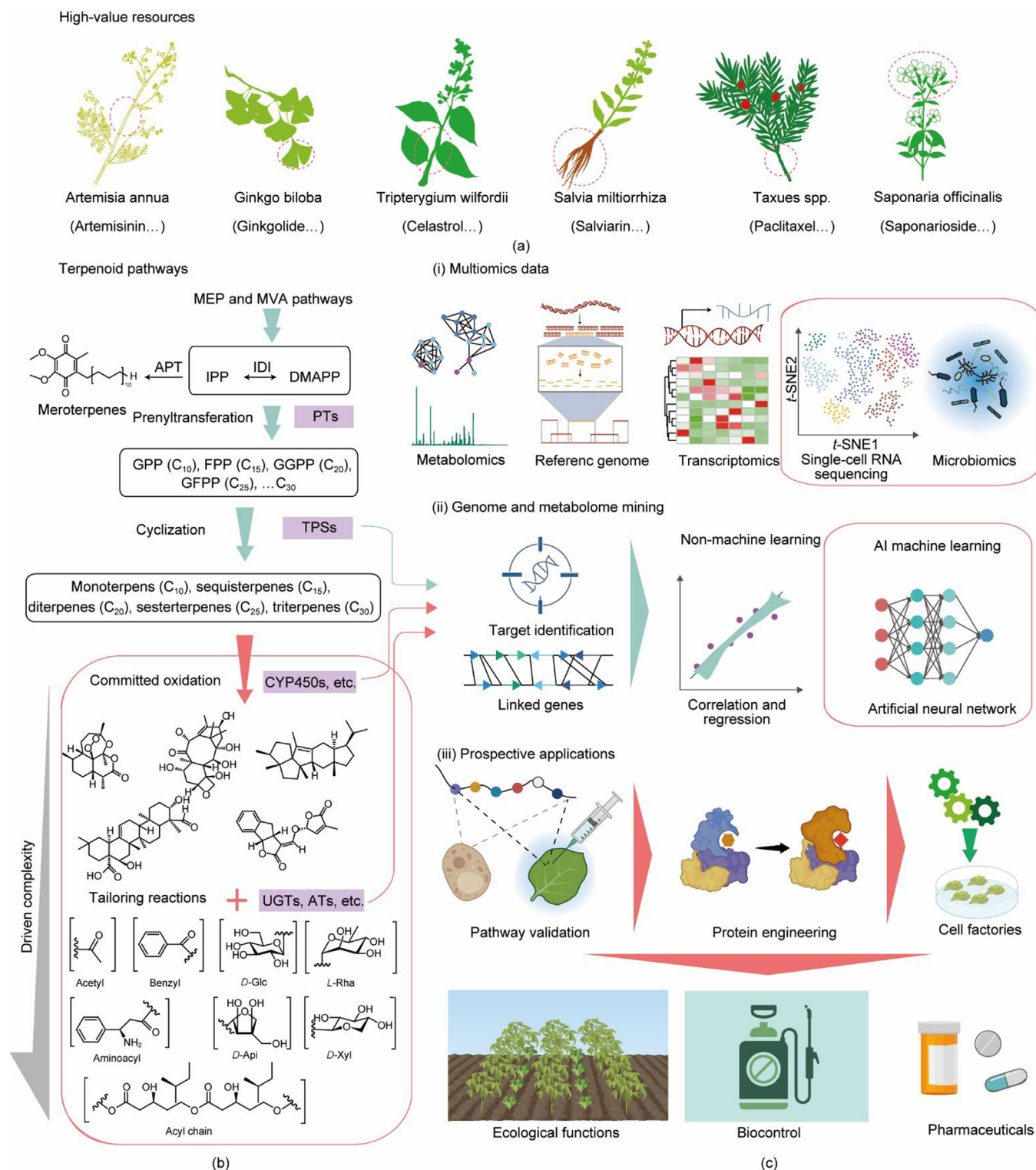


Fig. 1. A comprehensive pipeline for biosynthetic pathway identification and functional elucidation of plant terpenoids. (a) Pharmaceutical components accumulated in specific organs of medicinal plants (indicated with red circles). (b) Biosynthetic pathways of terpenoids start with the DMAPP/IPP from MVA and MEP pathways; the core cyclic skeletons are produced from prenyl diphosphates with varying five-carbon-atom (C₅) units through cyclization by TPSSs followed by oxidation catalyzed by cytochrome P450 monooxygenases (CYP450s), and further tailoring reactions drive the complexity of terpenoids to produce the targeted compounds. (c) Schematic diagram outlining a generalized strategy for elucidating the biosynthetic pathways of targeted terpenes in medicinal plants. (i) Classic analyses typically use metabolomics, genomics, and transcriptomics (even single-cell level omics are considered). The microbiome can aid in the discovery of novel genomic resources for identifying variations in terpenoids. (ii) The identification of biosynthesis-related gene clusters (BGCs) is currently a key method used for the validation of genes encoding enzymes involved in specialized terpenoid pathways. Genome and metabolome mining use nonmachine learning methods such as correlation and regression (for example, linking metabolomic and transcriptomics data) and deep learning, among other methods. (iii) *Nicotiana benthamiana* and yeast are useful platforms for the reconstitution of terpene pathways. Synthetic biology approaches are transforming field production to the generation of gene circuits for the stable expression of complex specialized terpenes in cell factories for pharmaceutical, biocontrol and ecological purposes. The blue arrows indicate well-established pathways of terpene biosynthesis; the red arrows indicate cutting-edge research focusing on the steps of terpene biosynthesis. Api: apiose; Glc: glucose; Rha: rhamnose; Xyl: xylose; AI: artificial intelligence; IDI: isopentenyl diphosphate isomerase; APT: aromatic prenyltransferase; UGT: UDP-dependent glycosyltransferase; AT: acyltransferase; PT: prenyltransferase; t-SNE: t-distributed stochastic neighbor embedding. The figure was created using Adobe Illustrator and BioRender.com.

named prenyltransferase) and TPS to generate a diverse array of terpenoid scaffolds. Both *trans*-IDSs and *cis*-IDS from plants share a similar condensation mechanism of DMAPP and IPP to produce prenyl diphosphates with varying chain lengths, despite being evolutionarily unrelated [5]. Compared with the well-studied farnesyl diphosphate synthases (FPPSs, C₁₅) and geranylgeranyl diphosphate synthases (GGPPSs, C₂₀), geranylgeranyl diphosphate synthases (GGPPSs, C₂₅) have only recently been functionally characterized in plants [6–8]. Comparative structural analyses of *bona fide* GGPPS (At4g36810, C₂₀), one GFPPS (At3g14550, C₂₅), and one polyprenyl diphosphate synthase (PPPS, At3g20160, > C₂₅) from *Arabidopsis thaliana* revealed that the chain-length control mechanism of plant *trans*-IDSs could be explained by a “three floor” model in which each floor is composed of two or three residues [8]. Briefly, the elongation terminates at the first, second, or third floor to yield C₁₅, C₂₀, and C₂₅ diphosphate products, respectively. Longer-chain products (> C₃₀) are formed when elongation proceeds beyond the third floor [8]. Sesterterpene synthases (sesterTPSs) that utilize GFPP as a substrate were subsequently identified in Brassicaceae and Lamiaceae [9–13]. Very recently, a versatile sesterTPS, CbTPS1, was characterized from the medicinal Brassicaceae plant *Capsella bursa-pastoris*, which produced 20 distinct sesterterpenes in an engineered *Escherichia coli* (*E. coli*) system [12]. Although functional sesterTPSs have been reported only in Brassicaceae and Lamiaceae, the widespread occurrence of both native and highly oxidized sesterterpenes in plants suggests that additional sesterTPSs, as well as the enzymes responsible for their oxidative tailoring, remain to be discovered [14,15]. Phylogenetic analyses of plant GGPPS/GFPPSs revealed a substantial evolutionary distance between GFPPSs from Brassicaceae and Lamiaceae, implying that sesterterpene biosynthetic capacity is independently recruited across distinct plant lineages [14].

Recently, Ma et al. [16] provided genetic evidence that GFPPSs are responsible for sesterterpene biosynthesis in *Arabidopsis*. Disruption of the GFPPS step (quadruple mutation for disrupting four *GFPPS* genes) led to increased accumulation of geranylgeranyl diphosphate (GGPP)-derived terpenes, indicating pathway competition. Interestingly, *Arabidopsis* plants overexpressing the *GFPPS* gene exhibited delayed development and leaf yellowing, phenotypes primarily caused by a reduction in endogenous GGPP. Mechanistic studies revealed that GFPPS not only competes for DMAPP/IPP precursors in plastids but also directly suppresses GGPPS activity through protein–protein interactions.

In addition to diterpenes, *bona fide* GGPPS in the plastid can regulate and partition the metabolic flux toward different classes of terpenoids through interactions with distinct partner proteins, including monoterpenes [17] and photosynthesis-related isoprenoids [18]. Among the interacting partners of GGPPS, the functional evolution of homo and heterodimeric geranyl(geranyl) diphosphate synthase (G(G)PPS) enzymes has been well characterized. With the expansion of the *GGPPS* gene family in bryophytes, certain members gradually acquired the ability to synthesize geranyl diphosphate (GPP), ultimately giving rise to functionally specialized GPPS enzymes in gymnosperms. Moreover, other members have evolved into catalytically inactive small subunit proteins. These first appeared as class II small subunit (SSU II) in ferns and gymnosperms, which enhanced the enzymatic activity of GGPPS and later gave rise to class I small subunit (SSU I) in angiosperms, which were capable of redirecting GGPPS (also referred to as LSU, large subunit) activity toward GPP production [18]. Notably, the biochemical outcomes of LSU–SSU interactions exhibit species-specific biochemical characteristics; for example, the major product of LSU–SSU I in *Antirrhinum majus* and *Clarkia breweri* flowers is GPP [19,20], whereas in LSU–SSU I, hop trichomes produce both GPP and GGPP [21], LSU–SSU II in *Arabidopsis* produces GPP and

GGPP [17], and LSU–SSU II in rice predominantly produces GGPP [22].

In addition to short-chain *trans*-IDS, significant progress has been made in understanding long-chain *trans*-IDSs, which catalyze the formation of polyprenyl chains longer than C₄₅. For example, hexaprenyl diphosphate synthase (Coq1) is responsible for synthesizing the isoprenoid side chain (C₄₅ or C₅₀) of coenzyme Q (CoQ, also known as ubiquinone). In land plants, Coq1 dictates the predominant CoQ isoform, namely, CoQ₉ with a C₄₅ tail or CoQ₁₀ with a C₅₀ tail [23]. The benzoquinone ring, another component of CoQ, is derived from aromatic amino acids such as phenylalanine or tyrosine. Recently, Xu et al. [24] have conducted a comprehensive analysis of more than 1000 land plant species to investigate the distribution of CoQ₉ and CoQ₁₀, identifying a critical amino acid substitution at position 240 (methionine or leucine) in Coq1 that governs the chain length of the polyprenyl tail. On the basis of these findings, the team also demonstrated gene editing in rice and wheat to produce CoQ₁₀-enriched crops for potential nutritional applications in humans. In the CoQ biosynthetic pathway, the benzoquinone ring is condensed with a polyprenyl chain (C₄₅ or C₅₀) by a 4-hydroxybenzoate octaprenyltransferase (UbiA) prenyltransferase family member, Coq2, forming the prenylated aromatic intermediate, which is subsequently modified through a series of hydroxylation and methylation steps to generate the final CoQ molecule. In addition to its role in CoQ biosynthesis, the UbiA family of prenyltransferases also catalyzes key prenylation reactions in the biosynthesis of other essential plant metabolites, including plastoquinones, hemes, chlorophylls, and tocopherols (vitamin E) [25]. Moreover, a specialized clade of UbiA prenyltransferases has evolved in plants to catalyze the prenylation of phenolic compounds, typically using DMAPP or GPP as prenyl donors [26]. To date, prenylated phenolics (also referred to as meroterpenoids because of their partial derivation from terpenoid biosynthetic pathways) have been identified in more than 4000 plant species. A recent example is hyperforin, a polycyclic polyprenylated acylphloroglucinol (PPAP) that serves as a chemotaxonomic marker for the Hypericaceae family [27,28]. The complete biosynthetic pathway of hyperforin in *Hypericum perforatum* was recently elucidated by Wu et al. [29], who revealed that its scaffold is derived from the degradation of branched-chain amino acids. Crucially, the addition of isoprenoid units is mediated by four UbiA family prenyltransferases, underscoring the central role of this enzyme family in diversifying plant prenylated phenolics [30].

In addition to canonical plant TPSS, microbial TPS-like (MTPSL) enzymes have recently been identified in plants and are likely acquired by nonseed land plants through horizontal gene transfer from microbes [31,32]. Structurally, plant MTPSLs contain only α -domains, whereas canonical TPSSs are composed of three domains, α , β , and γ , or two domains, α and β . The highly conserved aspartate-rich “DDxxD” motif, which is involved in binding the diphosphate moiety of the substrate and is required for ionization of the allylic diphosphate ester bond to initiate catalysis, is found in domain α of MTPSLs and TPSSs [31]. Thus far, *in vitro* biochemical assays have demonstrated that, similar to canonical TPSSs, MTPSLs can utilize GPP, farnesyl diphosphate (FPP), and GGPP to form monoterpenes, sesquiterpenes, and diterpenes, respectively [33,34]. However, the biological functions of these terpenoid compounds remain unclear and warrant further investigation.

3. Multiomics-driven elucidation of terpenoid scaffold modifications

The vast diversity of scaffold-modifying reactions in terpenes has long posed a central challenge in plant terpenoid research. This difficulty arises because downstream tailoring enzymes, such as

oxidases (e.g., cytochrome P450s), glycosyltransferases, and acyltransferases, typically belong to large, functionally diversified gene families, often comprising scores to hundreds of members, which complicates the assignment of precise enzymatic functions (Fig. 1(b)). Nevertheless, recent advances, particularly rapid improvements in sequencing technologies and the widespread adoption of heterologous pathway reconstruction in *Nicotiana benthamiana* (*N. benthamiana*) and microbial chassis, have substantially accelerated the elucidation of plant terpenoid biosynthetic pathways and their associated functions [35,36]. For candidate gene discovery, gene coexpression analysis remains a workhorse approach. In parallel, expanding genomic resources have provided valuable insights into the evolutionary origins and diversification of terpenoid metabolism, a perennial theme in plant terpenoid research (Fig. 2). Additional drivers include synthetic biology, metabolome genome-wide association studies (mGWASs), genome editing, and single-cell transcriptomics/metabolomics. Representative examples of terpenoid pathway elucidation using multiomics

are summarized in Table 1 [11,12,24,30,37–70]. Below, we provide a detailed discussion of selected examples.

Among recent advances in terpene biosynthesis enabled by multiomics approaches, the elucidation of Taxol (paclitaxel) biosynthesis has emerged as a focal point of interest. Taxol, a highly oxygenated tetracyclic diterpenoid produced at low abundance in yew species, remains a cornerstone in cancer chemotherapy. The publication of reference genomes for two yew species in 2021 has reignited efforts to fully elucidate the Taxol biosynthetic pathway [46,71]. Shortly thereafter, five research groups independently identified key upstream cytochrome P450-catalyzed reactions [43–45,51,72], enabling the reconstruction of a heterologous biosynthetic pathway that produces the crucial intermediate baccatin III, an industrial precursor for Taxol [44,45]. Most recently, Liang et al. [73] leveraged publicly available genomic and transcriptomic data (gene–gene coexpression analysis) to identify the enzymes responsible for the final two steps in Taxol biosynthesis, C2'α-hydroxylation and 3'-N-benzoylation. They

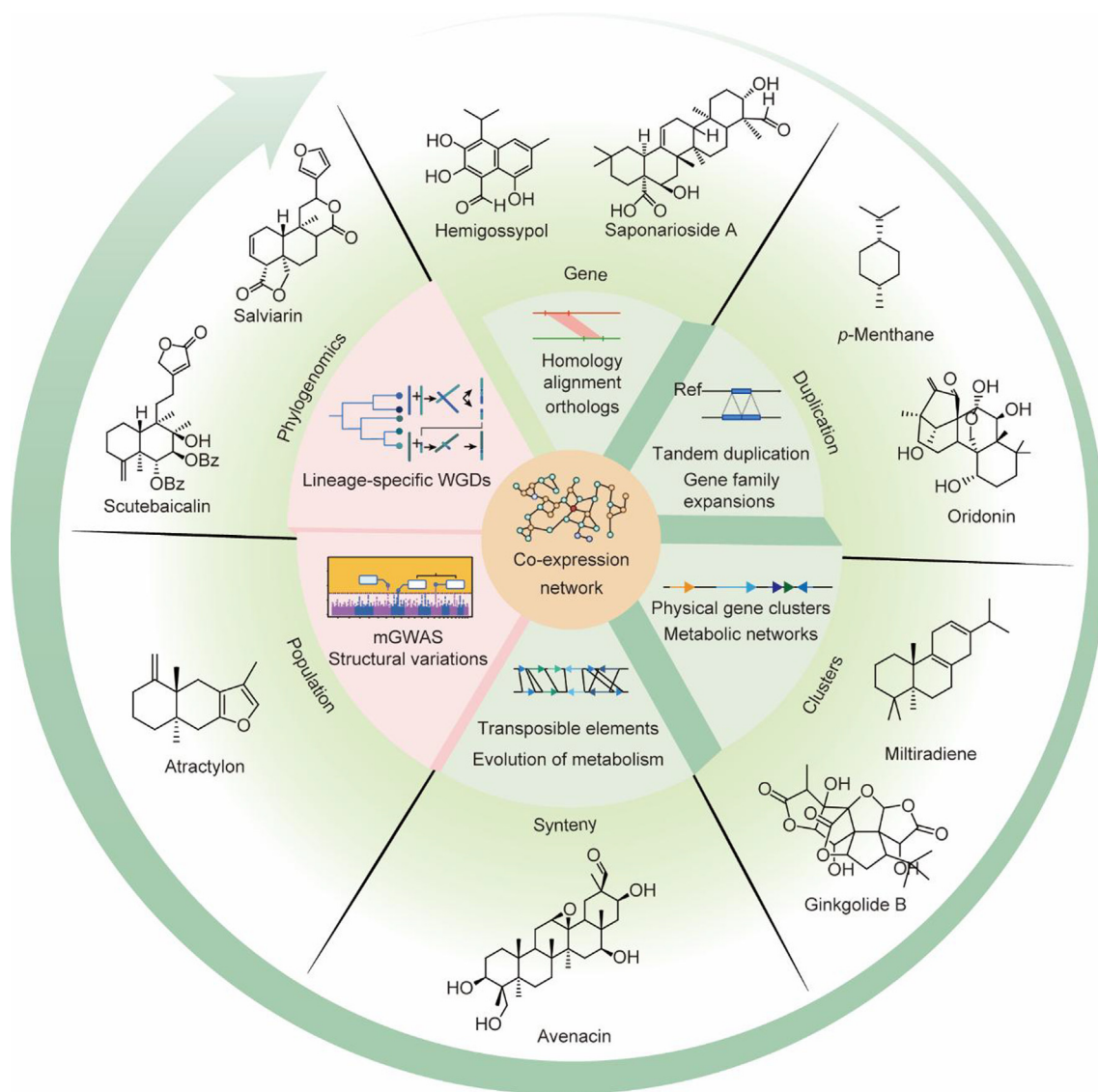


Fig. 2. Genome-enabled insights into plant terpenoid metabolism. High-quality reference genomes provide multiscale evolutionary perspectives on terpenoid diversity, while gene–gene coexpression analyses play a central role in prioritizing candidates for pathway elucidation. Representative terpenoids include hemigossypol (*Gossypium hirsutum* L., Malvaceae), *p*-menthane (*Schizonepeta tenuifolia* L., Lamiaceae), oridonin (*Isodon rubescens* L., Lamiaceae), miltiradiene (*Leonurus japonicus* L., Lamiaceae), ginkgolide (*Ginkgo biloba* L., Ginkgoaceae), avenacin (*Avena strigosa* L., Poaceae), atractylon (*Atractylodes lancea* L., Asteraceae), and salviarin and scutebaicalin (*Scutellaria barbata* L., Lamiaceae). mGWAS: metabolome genome-wide association study; WGD: whole-genome duplication.

Table 1
Notable examples of plant terpenoid biosynthesis and transportation published since 2020.

Type of terpenoids	End product	Host plant	Candidate gene determination	Characterization platform	References
Monoterpene (C ₁₀)/ sesquiterpene (C ₁₅)	Geraniol	<i>Rosa chinensis</i>	Homolog gene searching	<i>N. benthamiana</i>	[37]
	Patchouli	<i>Pogostemon cablin</i>	Genome mining and homolog gene searching	<i>E. coli</i> and <i>in vitro</i> assays	[38]
	Gossypol	<i>Gossypium hirsutum</i>	Homolog gene searching and gene coexpression	<i>E. coli</i> and <i>in vitro</i> assays; CRISPR/Cas9-based knockout	[39,40]
	Atractylon <i>p</i> -Menthane	<i>Atractylodes lancea</i> <i>Schizonepetaten uifolia</i>	Genome mining and mGWAS Genome mining, gene coexpression, and BGC	<i>E. coli</i> and <i>in vitro</i> assays; VIGS <i>E. coli</i> , <i>S. cerevisiae</i> , and <i>in vitro</i> assays	[41] [42]
Diterpene (C ₂₀)	Taxol	<i>Taxus spp</i>	Gene-metabolite coexpression	<i>N. benthamiana</i> and <i>S. cerevisiae</i>	[43–46]
	Momilactone	<i>Oryza sativa</i>	Gene coexpression and BGC	<i>E. coli</i> and <i>in vitro</i> assays; <i>N. benthamiana</i>	[47,48]
	Miltiradiene	<i>Leonotis leonurus</i> , <i>Plectranthus barbatus</i> , and <i>Prunella vulgaris</i>	Comparative genome, gene coexpression, and BGC	<i>N. benthamiana</i>	[49]
	Oridonin	<i>Isodon rubescens</i>	Genome mining, gene coexpression, and BGC	<i>N. benthamiana</i>	[50]
Sesterterpene (C ₂₅)	Hordedane	<i>Hordeum vulgare</i>	Gene coexpression and BGC	<i>S. cerevisiae</i> and <i>N. benthamiana</i>	[51]
	Ginkgolide	<i>Ginkgo biloba</i>	Gene coexpression and BGC	<i>N. benthamiana</i>	[52]
	Diverse sesterterpenes	<i>Arabidopsis thaliana</i>	Homolog gene searching and BGC	<i>E. coli</i> and <i>in vitro</i> assays	[11]
	Diverse sesterterpenes	<i>Capsella bursa-pastoris</i>	Homolog gene searching	<i>E. coli</i> and <i>in vitro</i> assays	[12]
Triterpene (C ₃₀)	QS-7, QS-21	<i>Quillaja saponaria</i>	Gene coexpression	<i>N. benthamiana</i>	[53]
	Avenacin	<i>Avena strigosa</i>	Comparative genome	<i>E. coli</i> and <i>in vitro</i> assays; <i>N. benthamiana</i>	[54]
	Glycyrrhizin	<i>Glycyrrhiza uralensis</i> ,	Gene coexpression and homolog gene searching	<i>S. cerevisiae</i> and <i>in planta</i>	[55]
	Aescin	<i>Aesculus chinensis</i>	Genome mining, gene coexpression, and BGC	<i>N. benthamiana</i>	[56]
Steroids	Saponarioside A	<i>Saponaria officinalis</i>	Gene coexpression	<i>N. benthamiana</i>	[57]
	Ursane-type triterpenes	<i>Eriobotrya japonica</i> and <i>Gillenia trifoliata</i>	Genome mining and gene coexpression	<i>N. benthamiana</i>	[58]
	Triterpenoid saponins	<i>Botrytis cinerea</i>	Mutant screening	<i>N. benthamiana</i> and <i>in planta</i>	[59]
	Steroid glycoalkaloids	<i>Solanum</i> plants	Gene coexpression and BGC	<i>S. cerevisiae</i> , <i>N. benthamiana</i> , and <i>in planta</i>	[60,61]
Tetraterpene (C ₄₀)	Strigolactone	<i>Sorghum bicolor</i>	Gene coexpression	<i>Xenopus laevis</i> oocyte and <i>in planta</i>	[62]
	Strigolactone	<i>Zea mays</i>	Mutant screening	<i>N. benthamiana</i> and <i>in planta</i>	[63]
Polyterpene (> C ₄₀)	Strigolactone	<i>Oryza sativa</i>	Gene coexpression	<i>In planta</i>	[64]
	Rubber	<i>Taraxacum kok-saghyz</i>	GWAS and homolog gene searching	CRISPR/Cas9-based knockout	[65,66]
Meroterpene	Rubber	<i>Lactuca sativa</i> L.	Homolog gene searching	CRISPR/Cas10-based knockout	[67]
	CoQ ₁₀	<i>Oryza sativa</i>	Homolog gene searching	CRISPR/Cas9-based mutagenesis	[24,68]
	Camptothecin	<i>Camptotheca acuminata</i>	Genome mining and gene coexpression	<i>S. cerevisiae</i> and <i>in vitro</i> assays	[69,70]
	Hyperforin	<i>Hypericum perforatum</i>	Genome mining and gene coexpression	<i>N. benthamiana</i>	[30]

BGC: biosynthesis-related gene cluster; CRISPR: clustered regularly interspaced short palindromic repeats; Cas9: CRISPR-associated protein 9; VIGS: virus-induced gene silencing; *S. cerevisiae*: *Saccharomyces cerevisiae*.

further reconstructed the late-stage pathway in yeast, achieving Taxol synthesis from baccatin III. McClune et al. [74] developed a multiplexed perturbation strategy to transcriptionally profile 2901 cell states of *Taxus media*, which was further used for gene-gene coexpression analysis. They discovered seven new genes that were directly involved in Taxol biosynthesis. Interestingly, they reported that a nuclear transport factor 2-like protein plays a key role in taxane production. They successfully reconstructed a 20-enzyme pathway for the *de novo* biosynthesis of 3'-*N*-debenzoyl-2'-deoxy-paclitaxel in *N. benthamiana* plants.

Another illustrative example of multiomics-driven pathway elucidation is that of *Quillaja saponaria* (QS) saponins, which are potent vaccine adjuvants originating from the Chilean soapbark tree (*Quillaja saponaria* (*Q. saponaria*)), including QS-7 and QS-21. The biosynthetic pathway of QS saponins remained elusive until the genome of *Q. saponaria* was sequenced [53]. Using genome mining, Reed et al. identified 16 enzymes responsible for generating intermediates with a quillaic acid scaffold (QA-TriR-F). They further characterized three enzymes, QsACT1, QsUGT73B44, and

QsUGT91AP1, that are required for the biosynthesis of QS-7 [53]. QS-21, which accumulates at higher levels in the soapbark tree and exhibits adjuvant properties comparable to those of QS-7, is regarded as a more promising adjuvant. Structurally, QS-21 contains a branched trisaccharide at C3, a linear tetrasaccharide at C28, and a glycosylated acyl chain at the C18 ester linked to the aglycone. The C18 acyl chain is derived from *L*-isoleucine and converted into 2-methylbutyric acid. Genome mining coupled with coexpression analysis revealed that a carboxyl- coenzyme A (CoA) ligase (CCL1) activates 2-methylbutyric acid to its CoA form, which serves as the starter unit for chain elongation by type III polyketide synthase (PKSIII). With five additional enzymes, the complete 20-step QS-21 pathway was reconstructed in *N. benthamiana* [75]. Liu et al. [76] further developed a biosynthetic platform in *Saccharomyces cerevisiae* that expresses 38 enzymes to produce QS-21 analogs. Similarly, saponins such as saponarioside A and B (SpA and SpB), which are found mainly in *Saponaria officinalis* (*S. officinalis*), share structural features with QS-21. Jo et al. [57] sequenced the *S. officinalis* genome and identified 14

enzymes required for SpB biosynthesis through genome mining and coexpression analysis, followed by pathway reconstitution in tobacco. They also discovered an unusual transglycosidase, SoGH1, which catalyzes the addition of *D*-quinovose to the C28-linked *D*-fucose moiety of SpB. Such studies, which resolve entire biosynthetic pathways spanning more than ten steps in a single effort, were virtually impossible before the advent of multiomics technologies.

As noted above, genomic data analysis not only reveals the breadth of candidate gene families but also offers valuable insights into their chromosomal distribution (Fig. 2). Among these analyses, mining biosynthesis-related gene clusters (BGCs) for targeted terpenoids represents a critical task. Plant BGCs are defined as physically linked genomic regions that contain at least three nonhomologous genes encoding enzymes involved in specialized metabolic pathways [77]. The increasing availability of high-quality plant genome assemblies, coupled with the development of dedicated tools for BGC identification, has greatly facilitated the discovery of BGCs across diverse plant species [49,78–81]. As summarized in Table 1, plant terpene BGCs are not uncommon. Notably, most plant terpene BGCs contain not only target pathway genes but also nonpathway genes. Moreover, it is common for certain intermediate steps to be catalyzed by genes located outside the cluster. A recent example of these can be found in the avenacins, a group of triterpene glucosides, identified in a BGC in oats (*Avena strigosa*). Qiao et al. [54] reported that two new genes, saponin-deficient 4 (*SAD4*) and *L*-phenylalanine ammonia-lyase 2 (*PAL2*), were not clustered with the previously identified avenacin BGC located on chromosome 1. These nonclustered genes were further named cluster auxiliary enzymes (CAEs).

Beyond providing a straightforward view of the chromosomal distribution of metabolic genes, comparative genomics at the family or genus level yields critical insights into the origins of metabolic pathways and the diversification of plant terpenoids. Owing to the rapid expansion of genomic resources for medicinal members of Lamiaceae, our understanding of terpenoid origins in this family has become increasingly refined. Notably, Li et al. [82] generated a high-quality reference genome for the medicinal species *Scutellaria barbata* and elucidated the biosynthetic pathway of anticancer clerodane-type diterpenoids. Building on this, phylogenetic and synteny analyses across 16 Lamiales genomes of class I and class II diterpene synthases (diTPSs) revealed that clerodane-related class I diTPSs in *Scutellaria* form a single clade, whereas those in *Salvia* are derived from distinct lineages; moreover, in contrast to the previously assumed single origin, clerodane-associated class II diTPSs in *Salvia* appear to have evolved independently [82]. Together, these findings sharpen our understanding of the evolution of clerodane biosynthetic pathways in Lamiaceae and provide a foundation for dissecting the downstream steps of clerodane diterpenoid metabolism. In addition, Hu et al. [83] conducted comprehensive metabolomic and transcriptomic profiling of 71 representative *Salvia* species, coupled with ancestral trait and gene reconstructions and evolutionary model building. They reported that tanshinone diterpenoids exhibit a pronounced clade-specific distribution across *Salvia* and that a *Salvia*-specific cytochrome P450 subfamily (CYP76AK) is a key determinant of this lineage-restricted pattern [58]. Collectively, these studies offer an instructive framework for probing the mechanisms that generate chemical diversity in medicinal plants.

4. Physiological and ecological functions of plant terpenoids

The elucidation of biosynthetic pathways of plant terpenoids, particularly those with high value-added potential, has found its foremost application in synthetic biology, including both plant-

based and microbial chassis systems. In addition to their direct commercialization, these naturally scarce terpenoid compounds can also serve as valuable resources for high-throughput screening of additional biological functions. Interested readers are encouraged to consult recent reviews on this topic [84–86]. In addition to pathway elucidation, researchers, particularly plant biologists, are increasingly focusing on the functions of these metabolic genes and compounds within the plant itself, as well as their interactions with the environment, including plant–plant, plant–parasitic plant, and plant–microbe interactions (Fig. 3). Deciphering the physiological and ecological roles of terpenoids will provide a solid foundation for future “intelligent breeding” not only in crops but also in medicinal plants, for example, by increasing biomass yield or increasing the accumulation of valuable target terpenoids.

With the advancement of precise gene editing and synthetic biology tools, researchers are accelerating the elucidation of the biological and physiological functions of each intermediate and gene in metabolic pathways. For example, soyasaponins, oleanane-type triterpenoid saponins that are abundant in soybean (*Glycine max* L.) seeds, include type A soyasaponins characterized by acetylated terminal sugars at the C22 position. Yuan et al. [87] identified a candidate benzylalcohol *O*-acetyltransferase, anthocyanin *O*-hydroxycinnamoyltransferase, *N*-hydroxycinnamoyl/benzoyltransferase, deacetylvindoline 4-*O*-acetyltransferase (BAHD) acyltransferase gene, *GmSSAcT1*, through gene coexpression analysis and biochemically confirmed its role in catalyzing sequential acetylation of type A soyasaponins. Disruption of *GmSSAcT1* via clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9 (CRISPR/Cas9) resulted in the complete loss of type A soyasaponins, the accumulation of nonacetylated forms, and seed germination defects. However, the precise mechanism underlying the phytotoxicity of nonacetylated soyasaponin A remains unclear. Similar self-phytotoxicity of intermediate terpenoids has also been reported in other species [88,89]. In the family Solanaceae, including tomato, potato, and eggplant, steroidal glycoalkaloids and steroidal saponins are two classes of valuable sterol-derived products. Jozwiad et al. [60] identified a novel cellulose synthase-like enzyme, termed GAME15, which acts as a dual-function enzyme and is essential for steroidal glycoalkaloid (SGA) biosynthesis in Solanaceae. GAME15 catalyzes cholesterol 3-*O*-glucuronidation and organizes a metabolon with CYP450s (GAME6 and GAME8) and oxidases in the endoplasmic reticulum (ER) membrane. Silencing *GAME15* in tomato and potato abolished SGAs and caused cholesterol accumulation. Transient reconstitution in *N. benthamiana* confirmed the function of *GAME15* in directing metabolic flux toward SGA biosynthesis. A second study revealed the dual function of *GAME15* in terms of glucosylation and scaffold protein in steroidal SGA biosynthesis. The depletion of *GAME15* increased the pest susceptibility of *Solanum* plants (Fig. 3 (a)) [61]. This *in planta* information lays the foundation for the future development of crop varieties with improved taste, enhanced health benefits, and increased pest/pathogen resistance.

It is well known that plant hormones play crucial roles in regulating plant growth, development, and environmental adaptability. Among them, strigolactones, discovered relatively recently, have been shown to regulate shoot branching and mediate root–rhizosphere interactions, such as symbioses with arbuscular mycorrhizal fungi [90]. Strigolactone, like ABA, is derived from β -carotene, but little is known about its coordination mechanism until the recent characterization of the tillering 20 (*T20*) gene in rice. The *T20* mutant showed pronounced dwarfism and increased tiller number. This gene encodes a ζ -carotene isomerase, which catalyzes a key step in carotenoid biosynthesis [91]. Strigolactone biosynthesis proceeds from carotene through carlactonic acid (CLA) to canonical strigolactones. Using cross-species coexpression analysis of transcriptomic datasets from *Arabidopsis*, tomato, and

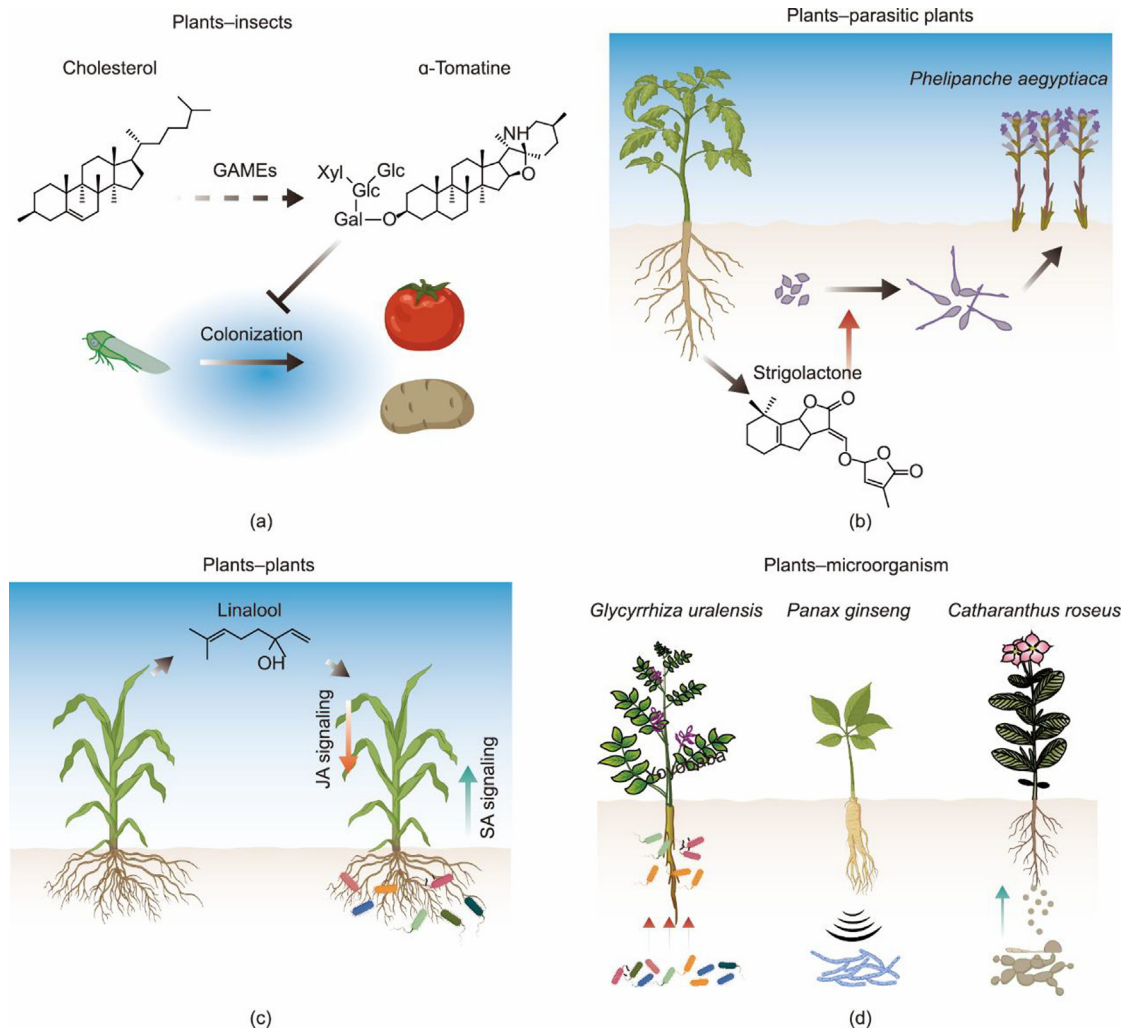


Fig. 3. Multifaceted physiological and ecological roles of plant terpenoids. (a) Steroidal glycoalkaloids (SGAs) contribute to pest resistance. (b) *Striga* germination is induced by strigolactones exuded from the roots of Solanaceae plants into the rhizosphere. (c) Linalool and hormones synergistically mediate high-density defense adaptation in maize populations. (d) Rhizosphere microbes play important roles in regulating the accumulation of bioactive compounds in various medicinal plants. GAME: glycoalkaloid metabolism enzyme.

rice, the Li et al. [64] identified a cytochrome P450, OsCYP706C2, as a key player in 4-oxo-MeCLA production. Loss of OsCYP706C2 impairs 4-oxo-MeCLA biosynthesis and exudation, leading to altered rice root development, including fewer crown roots under control conditions and reduced root biomass under phosphate deficiency. The same group also characterized ZmCYP706C37 in maize, which catalyzes oxidative steps toward zealactol and zealactonoic acid. Reduced ZmCYP706C37 activity shifts the strigolactone composition and decreases *Striga* germination and infection [63].

Recently, Shi et al. [62] identified and characterized two strigolactone transporters in sorghum, SbSLT1 and SbSLT2, which belong to the ATP-binding cassette (ABC) transporter family. These proteins are responsible for exporting synthesized strigolactones into the rhizosphere, where they stimulate the germination of *Striga* seeds. Knockout of SbSLT1 and/or SbSLT2 significantly reduced *Striga* parasitism while maintaining normal plant growth and yield. Similar findings have been reported in tomato, where the ABC transporters SlABCG45 and SlABCG44 play comparable roles (Fig. 3(b)) [92]. All these discoveries offer a novel strategy for developing *Striga*-resistant crop varieties and hold great promise for enhancing global food security.

Recently, an increasing number of researchers have recognized that plant-specific metabolites play a critical role in shaping the composition and function of the rhizosphere microbiome [93,94]. Previous studies have shown that disrupting either the sesterterpene or triterpene biosynthetic pathway in *Arabidopsis* leads to profound changes in the rhizosphere microbial community [95,96]. Current research is shifting from merely describing changes in the microbial composition toward elucidating the functional consequences of these alterations. Zhong et al. [97] reported that both melon and watermelon can exude cucurbitacin B and cucurbitacin E into the surrounding soil. Further analysis of rhizosphere microbial communities using 16S ribosomal RNA (rRNA) amplicon and metagenomic sequencing combined with *in vitro* biochemical assays revealed that root-secreted cucurbitacin B can serve as a carbon source to promote the growth of *Enterobacter* species in the rhizosphere. This, in turn, facilitates the enrichment of antagonistic bacteria such as *Bacillus*, thereby establishing an effective defense mechanism in melon against the soil-borne pathogen *Fusarium*. Guo et al. [98] demonstrated that the “volatile monoterpene (linalool)–hormone (jasmonate/salicylate)–rhizosphere microbe” axis synergistically mediates high-density defense adaptation in a maize population, elucidating the

molecular basis by which plants integrate aboveground signals with belowground processes to optimize survival strategies (Fig. 3(c)).

Given the wide diversity of terpenoid compounds and rhizosphere microbial communities, their interactions are inevitably associated with a wide range of physiological outcomes. Uncovering the underlying patterns and mechanisms of these interactions will require further in-depth investigation. Although constrained by the limitations of transgenic techniques, the field of medicinal plant research has also begun to focus on interactions between rhizosphere metabolites and microbes. For example, Zhong et al. [99] reported that the contents of major bioactive compounds (liquiritin and glycyrrhizic acid) in wild *Glycyrrhiza uralensis* were greater than those in cultivated plants. Multiomics analyses further suggested that rhizosphere microbes may play important roles in regulating the accumulation of liquiritin and glycyrrhizic acid. Sun et al. [100] investigated the effects of different concentrations of root exudates on ginseng (*Panax ginseng* C.A. Mey) growth, ginsenoside accumulation, and the composition and diversity of soil fungal communities. The results revealed that low concentrations of root exudates promoted ginseng growth and development, whereas high concentrations disrupted the rhizosphere fungal community and reduced plant adaptability. These findings provide new insights for the cultivation and breeding of medicinal plants (Fig. 3(d)).

5. Challenges and artificial intelligence (AI)-driven terpenoid research

Over the past decade, AI has increasingly contributed to and accelerated progress across research fields, including the study of plant terpenoid metabolism. In this field, achieving high-throughput acquisition of both qualitative and quantitative information on terpenoids remains a critical and continuously evolving challenge. Although chromatography–mass spectrometry (e.g., liquid chromatography (LC)–tandem mass spectrometry (MS/MS)) has been used for many years and is widely applied to elucidate plant metabolic pathways, accurately extracting qualitative and quantitative information from LC–MS data continues to be a major obstacle (Fig. 1(c)). Existing software tools, such as METLIN and MassBank, enable compound annotation on the basis of high-resolution mass spectrometry data; however, most laboratories still rely heavily on their own limited libraries of authentic standards for reliable analysis. For the *de novo* identification of unknown compounds, structural elucidation remains dependent on traditional methods, including compound isolation, purification, and nuclear magnetic resonance (NMR) spectroscopy.

AI, particularly machine learning (ML), is emerging as a promising direction with the potential to fundamentally transform and overcome these limitations, and there have already been some successful cases in chemical identification/prediction and mass spectrometry imaging [101–103]. A critical challenge, however, lies in the quality and standardization of input data. Mass spectrometry data derived from different sources, such as varying instrument manufacturers, data formats, and resolution settings, can introduce significant variability. Without consistent standards, models trained on such heterogeneous datasets may produce biased or nonreproducible results. Therefore, the availability of high-quality, standardized training datasets represents one of the most significant limiting factors for the successful application of ML, not only in terpenoid research but also across broader areas of biological science. To achieve this goal, several practical suggestions should be considered: ① Community-driven consortia (similar to metabolomics repositories) should be established to collect and share raw LC–MS/MS data with standardized metadata; ② uniform data processing pipelines (peak picking, alignment, and normaliza-

tion) should be implemented to minimize variability before being fed into ML models; and ③ coalescing collaborative validation should be encouraged by integrating authentic standards from multiple laboratories into shared reference libraries.

In more mature application domains, AI has already greatly accelerated the optimization of metabolic engineering, most visibly through deep-learning (DL) systems such as AlphaFold for protein structure prediction. In heterologous hosts, a persistent bottleneck is the suboptimal partitioning of metabolic flux arising from competition between native and introduced enzymes (e.g., TPSs, CYP450s, and UDP-dependent glycosyltransferases (UGTs)) [104,105]. To address this, researchers increasingly pair data-driven ML methods—such as reinforcement learning for strain-design policy search and few-shot active learning—with mechanistic constraint-based modeling (CBM) to map feasible flux states, pinpoint bottlenecks, and guide gene deletions or knockdowns that improve intermediate availability [106,107]. Reliable prediction, however, still hinges on high-quality multiomics measurements and trustworthy kinetic parameters. Protein engineering has seen especially rapid gains from DL: EVOLVEpro combines protein language models with few-shot active learning to increase enzyme activity using minimal experimental data [108], while generative diffusion frameworks such as RFdiffusion and P450Diffusion design enzymes with specified catalytic functions, enabling *de novo* catalysts with high activities [109,110]. Beyond enzyme-centric design, both ML and DL are being increasingly used for genome mining, multiomics network inference, and orchestration of the design–build–test–learn (DBTL) cycle in pathway engineering. These workflows are already well established in microbial chassis [111,112] and are poised to expand rapidly to plant systems for terpenoid production.

6. Conclusions and future perspectives

As summarized above, the integration of genomic resources, multiomics technologies, and advanced functional platforms is reshaping research on terpenoid metabolism in medicinal plants. With high-quality reference genomes, coexpression analysis can efficiently prioritize candidate genes, whereas high-throughput combinatorial expression systems enable rapid pathway reconstruction and validation. AI-driven approaches that combine phylogenetics, structure prediction, and multiomics data are poised to become the standard for elucidating complex terpenoid biosynthetic networks. Consequently, the field has progressed from the stepwise identification of single enzymes to the concurrent dissection of entire pathways, often encompassing more than ten reactions, particularly in medicinal plant systems. The next priority, beyond pathway mapping, is twofold: ① precisely enhancing *in planta* accumulation of target metabolites and ② defining their physiological and ecological functions (Fig. 1(c)). Together, these advances establish the prerequisites for deploying medicinal plants as versatile chassis for synthetic biology. In the future, efforts should extend beyond pathway-centric perspectives to examine how bioactive metabolites and terpenoids shape interactions with environmental microbes and how metabolomic insights can be leveraged to optimize these relationships. Such integrative strategies will improve the environmental adaptability, quality, and resilience of medicinal plants, laying the groundwork for sustainable utilization and intelligent breeding.

CRedit authorship contribution statement

Xiaochen Wang: Writing – review & editing, Writing – original draft. **Guodong Wang:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition.

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.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (32470265 and 32000232) and the National Key Research and Development Program of China (2021YFF1001200).

Owing to space limitations, we regret that some colleagues' work could not be cited, and we sincerely apologize for this.

References

- Ruzicka L. The isoprene rule and the biogenesis of terpenic compounds. *Experientia* 1953;9(10):357–67.
- Rohmer M, Rohmer M. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat Prod Rep* 1999;16(5):565–74.
- Christianson DW. Structural and chemical biology of terpenoid cyclases. *Chem Rev* 2017;117(17):11570–648.
- Matsuda Y, Abe I. Biosynthesis of fungal meroterpenoids. *Nat Prod Rep* 2016;33(1):26–53.
- Jia Q, Chen F. Catalytic functions of the isoprenyl diphosphate synthase superfamily in plants: a growing repertoire. *Mol Plant* 2016;9(2):189–91.
- Nagel R, Bernholz C, Vranova E, Košuth J, Bergau N, Ludwig S, et al. *Arabidopsis thaliana* isoprenyl diphosphate synthases produce the C25 intermediate geranylarnesyl diphosphate. *Plant J* 2015;84(5):847–59.
- Liu Y, Luo SH, Schmidt A, Wang GD, Sun GL, Grant M, et al. A geranylarnesyl diphosphate synthase provides the precursor for sesterterpenoid (C-25) formation in the glandular trichomes of the mint species *Leucoscepttrum canum*. *Plant Cell* 2016;28(3):804–22.
- Wang C, Chen Q, Fan D, Li J, Wang G, Zhang P. Structural analyses of short-chain prenyltransferases identify an evolutionarily conserved GPPS clade in Brassicaceae plants. *Mol Plant* 2016;9(2):195–204.
- Shao J, Chen QW, Lv HJ, He J, Liu ZF, Lu YN, et al. (+)-Thalianatriene and (–)-retigeranin B catalyzed by sesterterpene synthases from *Arabidopsis thaliana*. *Org Lett* 2017;19(7):1816–9.
- Huang AC, Kautsar SA, Hong YJ, Medema MH, Bond AD, Tantillo DJ, et al. Unearthing a sesterterpene biosynthetic repertoire in the Brassicaceae through genome mining reveals convergent evolution. *Proc Natl Acad Sci USA* 2017;114(29):E6005–14.
- Chen Q, Li J, Liu Z, Mitsuhashi T, Zhang Y, Liu H, et al. Molecular basis for sesterterpene diversity produced by plant terpene synthases. *Plant Commun* 2020;1(5):100051.
- Guo K, Tang X, Liu YC, Cheng HZ, Liu H, Fan YZ, et al. From monocyclization to pentacyclization: a versatile plant cyclase produces diverse sesterterpenes with anti-liver fibrosis potential. *Adv Sci* 2025;12(9):e2415370.
- Li DS, Hua J, Luo SH, Liu YC, Chen YG, Ling Y, et al. An extremely promiscuous terpenoid synthase from the Lamiaceae plant *Colquhounia coccinea* var. *mollis* catalyzes the formation of sester-/di-/sesqui-/mono-terpenoids. *Plant Commun* 2021;2(5):100233.
- Chen Q, Li J, Ma Y, Yuan W, Zhang P, Wang G. Occurrence and biosynthesis of plant sesterterpenes (C25), a new addition to terpene diversity. *Plant Commun* 2021;2(5):100184.
- Guo K, Liu Y, Li SH. The untapped potential of plant sesterterpenoids: chemistry, biological activities and biosynthesis. *Nat Prod Rep* 2021;38(12):2293–314.
- Ma YH, Chen QW, Wang YY, Zhang F, Wang C, Wang G. Heteromerization of short-chain prenyltransferase controls precursor allocation within a plastidial terpenoid network. *J Integr Plant Biol* 2023;65(5):1170–82.
- Chen Q, Fan D, Wang G. Heteromeric geranyl(geranyl) diphosphate synthase is involved in monoterpene biosynthesis in *Arabidopsis* flowers. *Mol Plant* 2015;8(9):1434–7.
- Song S, Jin R, Chen Y, He S, Li K, Tang Q, et al. The functional evolution of architecturally different plant geranyl diphosphate synthases from geranylgeranyl diphosphate synthase. *Plant Cell* 2023;35(6):2293–315.
- Tholl D, Kish CM, Orlova I, Sherman D, Gershenzon J, Pichersky E, et al. Formation of monoterpenes in *Antirrhinum majus* and *Clarkia breweri* flowers involves heterodimeric geranyl diphosphate synthases. *Plant Cell* 2004;16(4):977–92.
- Orlova I, Nagegowda DA, Kish CM, Gutensohn M, Maeda H, Varbanova M, et al. The small subunit of snapdragon geranyl diphosphate synthase modifies the chain length specificity of tobacco geranylgeranyl diphosphate synthase in planta. *Plant Cell* 2009;21(12):4002–17.
- Wang G, Dixon RA. Heterodimeric geranyl(geranyl)diphosphate synthase from hop (*Humulus lupulus*) and the evolution of monoterpene biosynthesis. *Proc Natl Acad Sci USA* 2009;106(24):9914–9.
- Zhou F, Wang CY, Gutensohn M, Jiang L, Zhang P, Zhang D, et al. A recruiting protein of geranylgeranyl diphosphate synthase controls metabolic flux toward chlorophyll biosynthesis in rice. *Proc Natl Acad Sci USA* 2017;114(26):6866–71.
- Ducluzeau AL, Wamboldt Y, Elowsky CG, Mackenzie SA, Schuurink RC, Basset GJC. Gene network reconstruction identifies the authentic *trans*-prenyl diphosphate synthase that makes the solanesyl moiety of ubiquinone-9 in *Arabidopsis*. *Plant J* 2012;69(2):366–75.
- Xu JJ, Lei Y, Zhang XF, Li JX, Lin Q, Wu XD, et al. Design of CoQ(10) crops based on evolutionary history. *Cell* 2025;188(7):1941–54.
- Li WK. Bringing bioactive compounds into membranes: the UbiA superfamily of intramembrane aromatic prenyltransferases. *Trends Biochem Sci* 2016;41(4):356–70.
- Munakata R, Yazaki K. How did plants evolve the prenylation of specialized phenolic metabolites by means of UbiA prenyltransferases? *Curr Opin Plant Biol* 2024;81:102601.
- Yang XW, Grossman RB, Xu G. Research progress of polycyclic polyprenylated acylphloroglucinols. *Chem Rev* 2018;118(7):3508–58.
- Richard JA. Chemistry and biology of the polycyclic polyprenylated acylphloroglucinol hyperforin. *Eur J Org Chem* 2013;2014(2):273–99.
- Wu S, Malaco Morotti AL, Wang S, Wang Y, Xu X, Chen J, et al. Convergent gene clusters underpin hyperforin biosynthesis in St John's wort. *New Phytol* 2022;235(2):646–61.
- Wu S, Morotti ALM, Yang J, Wang E, Tatsis EC. Single-cell RNA sequencing facilitates the elucidation of the complete biosynthesis of the antidepressant hyperforin in St. John's wort. *Mol Plant* 2024;17(9):1439–57.
- Jia Q, Kollner TG, Gershenzon J, Chen F. MTPSLs: new terpene synthases in nonseed plants. *Trends Plant Sci* 2018;23(2):121–8.
- Jia QD, Li GL, Kollner TG, Fu J, Chen X, Xiong W, et al. Microbial-type terpene synthase genes occur widely in nonseed land plants, but not in seed plants. *Proc Natl Acad Sci USA* 2016;113(43):12328–33.
- Chen L, Gao R, Wei G, Luo S, Köllner TG, Xu H, et al. Microbial-type terpene synthases significantly contribute to the terpene profile of glandular trichomes of the fern *Dryopteris fragrans* (L.). *Plant J* 2025;121(6):e70079.
- Wei G, Jia QD, Chen XL, Köllner TG, Bhattacharya D, Wong GKS, et al. Terpene biosynthesis in red algae is catalyzed by microbial-type but not typical plant terpene synthases. *Plant Physiol* 2019;179(2):382–90.
- Gharat SA, Tamhane VA, Giri AP, Aharoni A. Navigating the challenges of engineering composite specialized metabolite pathways in plants. *Plant J* 2025;121(6):e70100.
- 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.
- Conart C, Bomzan DP, Huang XQ, Bassard JE, Paramita SN, Saint-Marcoux D, et al. A cytosolic bifunctional geranyl/farnesyl diphosphate synthase provides MVA-derived GPP for geraniol biosynthesis in rose flowers. *Proc Natl Acad Sci USA* 2023;120(19):e2221440120.
- Shen Y, Li W, Zeng Y, Li Z, Chen Y, Zhang J, et al. Chromosome-level and haplotype-resolved genome provides insight into the tetraploid hybrid origin of patchouli. *Nat Commun* 2022;13(1):3511.
- Huang JQ, Fang X, Tian X, Chen P, Lin JL, Guo XX, et al. Aromatization of natural products by a specialized detoxification enzyme. *Nat Chem Biol* 2020;16(3):250–6.
- Lin JL, Fang X, Li JX, Chen ZW, Wu WK, Guo XX, et al. Dirigent gene editing of gossypol enantiomers for toxicity-depleted cotton seeds. *Nat Plants* 2023;9(4):605–15.
- Zhu J, Li QQ, Zhang QY, Liu L, Tang R, He L, et al. Characterisation of early biosynthetic steps of atracylon via an integrative strategy. *Plant Biotechnol J* 2025;pbj.70335.
- 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 by an inverted duplication controls monoterpene biosynthesis in *Schizonepeta tenuifolia*. *Mol Plant* 2023;16(3):533–48.
- Yang C, Wang Y, Su Z, Xiong L, Wang P, Lei W, et al. Biosynthesis of the highly oxygenated tetracyclic core skeleton of Taxol. *Nat Commun* 2024;15(1):2339.
- 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.
- 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.
- 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.
- Kitaoka N, Zhang J, Oyagbenro RK, Brown B, Wu Y, Yang B, et al. Interdependent evolution of biosynthetic gene clusters for momilactone production in rice. *Plant Cell* 2021;33(2):290–305.
- De La Pena R, Sattely ES. Rerouting plant terpene biosynthesis enables momilactone pathway elucidation. *Nat Chem Biol* 2021;17(2):205–12.
- Bryson AE, Lanier ER, Lau KH, Hamilton JP, Vaillancourt B, Mathieu D, et al. Uncovering a multiterpene biosynthetic gene cluster in the Lamiaceae reveals a dynamic evolutionary trajectory. *Nat Commun* 2023;14(1):343.

- [50] Sun Y, Shao J, Liu H, Wang H, Wang G, Li J, 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.
- [51] Liu Y, Esposto D, Mahdi LK, Porzel A, Stark P, Hussain H, et al. Hordedane diterpenoid phytoalexins restrict *Fusarium graminearum* infection but enhance *Bipolaris sorokiniana* colonization of barley roots. *Mol Plant* 2024;17(8):1307–27.
- [52] Forman V, Luo D, Geu-Flores F, Lemcke R, Nelson DR, Kampranis SC, et al. A gene cluster in *Ginkgo biloba* encodes unique multifunctional cytochrome P450s that initiate ginkgolide biosynthesis. *Nat Commun* 2022;13(1):5143.
- [53] Reed J, Orme A, El-Demerdash A, Owen C, Martin LBB, Misra RC, et al. Elucidation of the pathway for biosynthesis of saponin adjuvants from the soapbark tree. *Science* 2023;379(6638):1252–64.
- [54] Qiao X, Houghton A, Reed J, Steuernagel B, Zhang J, Owen C, et al. Comprehensive mutant chemotyping reveals embedding of a lineage-specific biosynthetic gene cluster in wider plant metabolism. *Proc Natl Acad Sci USA* 2025;122(12):e2417588122.
- [55] Chung SY, Seki H, Fujisawa Y, Shimoda Y, Hiraga S, Nomura Y, et al. A cellulose synthase-derived enzyme catalyzes 3-O-glucuronosylation in saponin biosynthesis. *Nat Commun* 2020;11(1):5664.
- [56] Sun W, Yin Q, Wan H, Gao R, Xiong C, Xie C, et al. Characterization of the horse chestnut genome reveals the evolution of aescin and aesculin biosynthesis. *Nat Commun* 2023;14(1):6470.
- [57] Jo S, El-Demerdash A, Owen C, Srivastava V, Wu D, Kikuchi S, et al. Unlocking saponin biosynthesis in soapwort. *Nat Chem Biol* 2025;21(2):215–26.
- [58] Su W, Jing Y, Lin S, Yue Z, Yang X, Xu J, et al. Polyploidy underlies co-option and diversification of biosynthetic triterpene pathways in the apple tribe. *Proc Natl Acad Sci USA* 2021;118(20):e2101767118.
- [59] Escaray FJ, Felipo-Benavent A, Antonelli CJ, Balaguer B, Lopez-Gresa MP, Vera P. Plant triterpenoid saponins function as susceptibility factors to promote the pathogenicity of *Botrytis cinerea*. *Mol Plant* 2024;17(7):1073–89.
- [60] Jozwiak A, Panda S, Akiyama R, Yoneda A, Umemoto N, Saito K, et al. A cellulose synthase-like protein governs the biosynthesis of Solanum alkaloids. *Science* 2024;386(6728):eadq5721.
- [61] Boccia M, Kessler D, Seibt W, Grabe V, Rodríguez López CE, Grzech D, et al. A scaffold protein manages the biosynthesis of steroidal defense metabolites in plants. *Science* 2024;386(6728):1366–72.
- [62] Shi J, Mei C, Ge F, Hu Q, Ban X, Xia R, et al. Resistance to *Striga* parasitism through reduction of strigolactone exudation. *Cell* 2025;188(7):1955–66.
- [63] Li C, Dong L, Durairaj J, Guan JC, Yoshimura M, Quinodoz P, et al. Maize resistance to witchweed through changes in strigolactone biosynthesis. *Science* 2023;379(6627):94–9.
- [64] Li C, Haider I, Wang JY, Quinodoz P, Suarez Duran HG, Méndez LR, et al. OsCYP706C2 diverts rice strigolactone biosynthesis to a noncanonical pathway branch. *Sci Adv* 2024;10(35):eadq3942.
- [65] Lin T, Xu X, Du H, Fan X, Chen Q, Hai C, et al. Extensive sequence divergence between the reference genomes of *Taraxacum kok-saghyz* and *Taraxacum mongolicum*. *Sci China Life Sci* 2022;65(3):515–28.
- [66] Chao J, Wu S, Shi M, Xu X, Gao Q, Du H, et al. Genomic insight into domestication of rubber tree. *Nat Commun* 2023;14(1):4651.
- [67] Kwon M, Hodgins CL, Salama EM, Dias KR, Parikh A, Mackey AV, et al. New insights into natural rubber biosynthesis from rubber-deficient lettuce mutants expressing goldenrod or guayule cis-prenyltransferase. *New Phytol* 2023;239(3):1098–111.
- [68] Xu JJ, Hu M, Yang L, Chen XY. How plants synthesize coenzyme Q. *Plant Commun* 2022;3(5):100341.
- [69] Kang M, Fu R, Zhang P, Lou S, Yang X, Chen Y, et al. A chromosome-level *Camptotheca acuminata* genome assembly provides insights into the evolutionary origin of camptothecin biosynthesis. *Nat Commun* 2021;12(1):3531.
- [70] Dudley QM, Jo S, Guerrero DAS, Chhetry M, Smedley MA, Harwood WA, et al. Reconstitution of monoterpene indole alkaloid biosynthesis in genome engineered *Nicotiana benthamiana*. *Commun Biol* 2022;5(1):949.
- [71] Cheng J, Wang X, Liu X, Zhu X, Li Z, Chu H, et al. Chromosome-level genome of Himalayan yew provides insights into the origin and evolution of the paclitaxel biosynthetic pathway. *Mol Plant* 2021;14(7):1199–209.
- [72] Zhao Y, Liang F, Xie Y, Duan YT, Andeadelli A, Pateraki I, et al. Oxetane ring formation in Taxol biosynthesis is catalyzed by a bifunctional cytochrome P450 enzyme. *J Am Chem Soc* 2024;146(1):801–10.
- [73] Liang F, Xie Y, Zhang C, Zhao Y, Motawia MS, Kampranis SC. Elucidation of the final steps in Taxol biosynthesis and its biotechnological production. *Nat Synth* 2025;4(10):1212–22.
- [74] McClune CJ, Liu JC, Wick C, et al. Discovery of *FoT01* and *Taxol* genes enables biosynthesis of baccatin III. *Nature* 2025;643:582–92.
- [75] Martin LBB, Kikuchi S, Rejzke M, Owen C, Reed J, Orme A, et al. Complete biosynthesis of the potent vaccine adjuvant QS-21. *Nat Chem Biol* 2024;20(4):493–502.
- [76] Liu Y, Zhao X, Gan F, et al. Complete biosynthesis of QS-21 in engineered yeast. *Nature* 2024;629:937–44.
- [77] Nützmann HW, Osbourn A. Gene clustering in plant specialized metabolism. *Curr Opin Biotechnol* 2014;26:91–9.
- [78] Liang X, Wang Y, Shen W, Liao B, Liu X, Yang Z, et al. Genomic and metabolomic insights into the selection and differentiation of bioactive compounds in citrus. *Mol Plant* 2024;17(11):1753–72.
- [79] Zhang H, He Q, Xing L, Wang R, Wang Y, Liu Y, et al. The haplotype-resolved genome assembly of autotetraploid rhubarb *Rheum officinale* provides insights into its genome evolution and massive accumulation of anthraquinones. *Plant Commun* 2024;5(1):100677.
- [80] Zhan CS, Shen SQ, Yang CK, 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.
- [81] Smit SJ, Lichman BR. Plant biosynthetic gene clusters in the context of metabolic evolution. *Nat Prod Rep* 2022;39(7):1465–82.
- [82] Li HX, Wu S, Lin RX, Xiao Y, Malaco Morotti AL, 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.
- [83] Hu JD, Qiu S, Wang FY, Li Q, Xiang CL, Di P, et al. Functional divergence of CYP76AKs shapes the chemodiversity of abietane-type diterpenoids in genus. *Nat Commun* 2023;14(1):14.
- [84] Bureau JA, Oliva ME, Dong YM, Ignea C. Engineering yeast for the production of plant terpenoids using synthetic biology approaches. *Nat Prod Rep* 2023;40(12):1822–48.
- [85] Liu X, Zhang P, Zhao Q, Huang AC. Making small molecules in plants: a chassis for synthetic biology-based production of plant natural products. *J Integr Plant Biol* 2023;65(2):417–43.
- [86] Yin JY, Lai M, Yu XY, Su DD, Xiong XY, Li YL. Comprehensive strategies for paclitaxel production: insights from plant cell culture, endophytic microorganisms, and synthetic biology. *Hortic Res* 2025;12(3):uhae346.
- [87] Yuan J, Ma L, Wang Y, Xu X, Zhang R, Wang C, et al. A recently evolved BAHD acetyltransferase, responsible for bitter soyasaponin A production, is indispensable for soybean seed germination. *J Integr Plant Biol* 2023;65(11):2490–504.
- [88] Itkin M, Rogachev I, Alkan N, Rosenberg T, Malitsky S, Masini L, et al. GLYCOALKALOID METABOLISM1 is required for steroidal alkaloid glycosylation and prevention of phytotoxicity in tomato. *Plant Cell* 2011;23(12):4507–25.
- [89] Kazachkova Y, Zemach I, Panda S, Bocobza S, Vainer A, Rogachev I, et al. The GORKY glycoalkaloid transporter is indispensable for preventing tomato bitterness. *Nat Plants* 2021;7(4):468–80.
- [90] Mashiguchi K, Seto Y, Yamaguchi S. Strigolactone biosynthesis, transport and perception. *Plant J* 2021;105(2):335–50.
- [91] Liu X, Hu Q, Yan J, Sun K, Liang Y, Jia M, et al. Zeta-carotene isomerase suppresses tillering in rice through the coordinated biosynthesis of strigolactone and abscisic acid. *Mol Plant* 2020;13(12):1784–801.
- [92] Ban X, Qin L, Yan J, Wu J, Li Q, Su X, et al. Manipulation of a strigolactone transporter in tomato confers resistance to the parasitic weed broomrape. *Innovation* 2025;6(3):100815.
- [93] Wang X, Zhang J, Lu X, Bai Y, Wang G. Two diversities meet in the rhizosphere: root specialized metabolites and microbiome. *J Genet Genomics* 2024;51(5):467–78.
- [94] Fan X, Ge AH, Qi S, Guan Y, Wang R, Yu N, et al. Root exudates and microbial metabolites: signals and nutrients in plant–microbe interactions. *Sci China Life Sci* 2025;68(8):2290–302.
- [95] Chen QW, Jiang T, Liu YX, Liu H, Zhao T, Liu Z, et al. Recently duplicated sesterterpene (C25) gene clusters in *Arabidopsis thaliana* modulate root microbiota. *Sci China Life Sci* 2019;62(7):947–58.
- [96] Huang AC, Jiang T, Liu YX, Bai YC, Reed J, Qu B, et al. A specialized metabolic network selectively modulates *Arabidopsis* root microbiota. *Science* 2019;364(6440):eaau6389.
- [97] Zhong Y, Xun W, Wang X, Tian S, Zhang Y, Li D, et al. Root-secreted bitter triterpene modulates the rhizosphere microbiota to improve plant fitness. *Nat Plants* 2022;8(8):887–96.
- [98] Guo DS, Liu ZL, Raaijmakers JM, Xu Y, Yang J, Erb M, et al. Linalool-triggered plant–soil feedback drives defense adaptation in dense maize plantings. *Science* 2025;389(6761):389.
- [99] Zhong CF, Chen CY, Gao X, et al. Multi-omics profiling reveals comprehensive microbe–plant–metabolite regulation patterns for medicinal plant *Fisch*. *Plant Biotechnol J* 2022;20:1874–87.
- [100] Sun J, Yang J, Zhao SY, Yu Q, Weng L, Xiao C. Root exudates influence rhizosphere fungi and thereby synergistically regulate yield and quality. *Front Microbiol* 2023;14:14.
- [101] Beck AC, Muhoberac M, Randolph CE, Beveridge CH, Wijewardhane PR, Kenttämaa HI, et al. Recent developments in machine learning for mass spectrometry. *ACS Meas Sci Au* 2024;4(3):233–46.
- [102] Kaskow JA, Hahnert ET, Porter TK, Lu Y, Stanev V, Niu C, et al. Predicting peptide ionization efficiencies for electrospray ionization mass spectrometry using machine learning. *J Am Soc Mass Spectrom* 2024;35(10):2297–307.
- [103] Wadie B, Stuart L, Rath CM, Drotleff B, Mamedov S, Alexandrov T. METASPACE-ML: context-specific metabolite annotation for imaging mass spectrometry using machine learning. *Nat Commun* 2024;15(1):9110.
- [104] Zhang J, Hansen LG, Gudich O, Viehriq K, Lassen LMM, Schrübbars L, et al. A microbial supply chain for production of the anti-cancer drug vinblastine. *Nature* 2022;609(7926):341–7.
- [105] Zhu Y, Yan X, Li W, Qiao J, Zhao GR. Modular metabolic engineering of *saccharomyces cerevisiae* for enhanced production of ursolic acid. *J Agric Food Chem* 2025;73(6):3580–90.
- [106] Koch M, Duigou T, Faulon JL. Reinforcement learning for bioretrosynthesis. *ACS Synth Biol* 2020;9(1):157–68.
- [107] Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox vol 3.0. *Nat Protoc* 2019;14(3):639–702.
- [108] Jiang K, Yan Z, Di Bernardo M, Sgrizzi SR, Villiger L, Kayabolen A, et al. Rapid in silico directed evolution by a protein language model with EVOLVEpro. *Science* 2025;387(6732):eadr6006.

- [109] Watson JL, Juergens D, Bennett NR, Trippe BL, Yim J, Eisenach HE, et al. De novo design of protein structure and function with RFdiffusion. *Nature* 2023;620(7976):1089–100.
- [110] Hu RE, Yu CH, Ng IS. GRACE: generative redesign in artificial computational enzymology. *ACS Synth Biol* 2024;13(12):4154–64.
- [111] Chen R, Wang M, Keasling JD, Hu T, Yin X. Expanding the structural diversity of terpenes by synthetic biology approaches. *Trends Biotechnol* 2024;42(6):699–713.
- [112] Liao L, Xie M, Zheng X, Zhou Z, Deng Z, Gao J. Molecular insights fast-tracked: AI in biosynthetic pathway research. *Nat Prod Rep* 2025;42(5):911–36.