Cutting-edge plant natural product pathway elucidation

Jianing Han[1,3], Emma Parker Miller[2,3], Sijin Li[1*]

Address:

[1]Robert F. Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY, 14853, USA

[2]Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853, USA

*Corresponding author: Li, Sijin (sijin.li@cornell.edu)

[3]These authors contributed equally to this work.

Corresponding author information:

Sijin Li

sijin.li@cornell.edu

Robert Frederick Smith School of Chemical and Biomolecular Engineering

Cornell University | 214 Olin Hall, Ithaca, NY 14853

(607)255-7713

Author emails:

Jianing Han

jh2728@cornell.edu

Emma Parker Miller

ejm343@cornell.edu

Abstract

Plant natural products (PNPs) play important roles in plant physiology and have been applied across diverse fields of human society. Understanding their biosynthetic pathways informs plant evolution and meanwhile enables sustainable production through metabolic engineering. However, the discovery of PNP biosynthetic pathways remains challenging due to the diversity of enzymes involved and limitations in traditional gene mining approaches. In this review, we will summarize state-of-the-art strategies and recent examples for predicting and characterizing PNP biosynthetic pathways respectively with multi-omics-guided tools and heterologous host systems, and share our perspectives on the systematic pipelines integrating these various bioinformatic and biochemical approaches.

Key words: plant natural products, biosynthetic pathway, multi-omics-guided prediction, heterologous characterization

Introduction

Plants produce more than 200, 000 small molecules [1] with high diversity and complexity, called plant natural products (PNPs). PNPs are important in plant signaling and defense in response to stressors such as herbivores, insects, fungal infection, and allelopathy [2,3]. PNPs also have unique applications useful in various fields including pharmaceuticals and agriculture [4–6]. Unraveling the biosynthetic pathways of PNPs sheds light on understanding how plants evolve to adapt to the changing environment and also paves the way for sustainable production of these valuable compounds via metabolic engineering.

PNPs can be categorized by their different core scaffolds, such as the isoprene-derived terpenoids and nitrogen-containing alkaloids. Scaffold-forming enzymes that catalyze the committed step toward different types of PNP scaffolds are important components of PNP biosynthetic pathways [1]. For example, terpenoids are the largest group of PNPs in nature [7]. Plant terpene synthases (TPSs) catalyze the formation of the core polycyclic carbon-based scaffolds of terpenoids using varied numbers of five-carbon isoprene units synthesized through the mevalonate pathway. Alkaloids, such as benzylisoquinoline alkaloids (BIAs) and monoterpene indole alkaloids (MIAs), typically feature a nitrogen-containing ring derived from amino acids including tyrosine and tryptophan. Amine-aldehyde condensation is the key step for the alkaloid scaffold formation, such as norcoclaurine synthase (NCS) for BIA and strictosidine synthase (STR) for MIA. Based on the core scaffold, a series of tailoring reactions further greatly increases the diversity of PNPs. Typical tailoring reactions involve oxidation by cytochrome P450 enzymes (CYPs), reduction by alcohol dehydrogenases (ADHs), and group transfer by methyltransferases (MTs), acyltransferases (ACTs) and glycosyltransferases (GTs) [1]. Advances in knowledge of the chemical logics of PNP biosynthesis have led to the elucidation of diverse PNP biosynthetic pathways, including the biosynthetic pathways of morphine [8], noscapine [9], and scopolamine [10]. However, these accomplishments represent only a fraction of the diverse PNPs in nature.

Despite fruitful achievements in PNP pathway discovery and metabolic engineering [11–15], the pathway elucidation process is still challenging (Figure 1A). First, the traditional BLAST-based gene mining approach may not predict all the enzymes involved in a very long PNP pathway, which includes complicated chemical logic and various potential substrates, intermediates, and different final products. The wide variety of chemicals possibly involved in the PNP pathway makes it nearly impossible to predict the entire cascaded chemical reactions *a priori*. Second, despite the advances in next-generation sequencing, plant transcriptome information is still limited and hinders efficient gene mining. Enzymes involved in PNP biosynthesis are highly diverse and may be expressed only in certain cultivars, certain plant tissues, at certain growth stages, or when induced by a special stimulus, making it highly possible

that their sequence information might not be detected in general transcriptome analysis. Further complicating the elucidation process is that enzymes catalyzing similar types of reactions may have evolved independently in different species, resulting in low sequence similarity that hinders efficient pathway prediction via gene mining. Characterizing the predicted enzymes in a PNP pathway is also challenging, because different types of enzymes may require different cofactors and sub-cellular environments. In this review, we discuss the state-of-the-art pipelines for PNP pathway prediction and characterization to address these challenges, as well as recent achievements in novel PNP pathway elucidation within the past two years.

Multi-omics-guided pathway prediction

Transcriptome-based approaches, majorly co-expression and comparative analyses, have been the prevalent methods used in PNP pathway prediction, which have been reviewed comprehensively in prior reviews [16–18]. Advances in next-generation sequencing, protein and metabolite quantification, and computational approaches make way for integrated genomics, transcriptomics, proteomics, and metabolomics methods (Figure 1B). Recent advances from the past two years are included in Table 1.

1) Combined transcriptomics and genomics

One way that genomic sequence data has enhanced the elucidation of PNP pathways is through the identification of biosynthetic gene clusters (BGCs). The discovery of BGCs in plants dates back to 1997, where five genes in maize encoding enzymes in 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (DIMBOA) biosynthesis were found to localize closely on the same chromosome [19]. To date, more than 30 plant BGCs have been identified [20,21]. Genomic analysis can predict potential BGCs from sequenced plant genomes. Since genes in a plant BGC are more likely to be in the same pathway, BGC prediction via genomics analysis can significantly narrow down the search space for given pathways. Several computational tools, including plantiSMASH [22], PlantClusterFinder [23], and PhytoClust [24], can predict BGCs with limited biochemical information. However, BGCs in plants typically do not comprise the entire pathway, requiring the identification of additional enzymes to complete the reactions. Transcriptomic co-expression analysis can complement this strategy by searching for the enzymes sharing similar expression patterns with genes in the known BGC. For example, Forman et al. investigated the ginkgolide biosynthesis pathway in ginkgo [25]. Genomic analysis identified a BGC in the ginkgo genome where five *CYP* genes were located physically close (within 2 Mbp) to levopimaradiene synthase (LPS), a diterpene synthase that forms the scaffold of a ginkgo terpenoid. Subsequently, they verified that the five CYPs modified the product of LPS. Transcriptomic analysis further uncovered ten other *CYP*

genes outside the BGC that co-expressed with the BGC genes across different ginkgo tissues. One of them was verified to further catalyze the lactone-ring forming step after the BGC enzymes.

Genomic analysis also allows for gene discovery via comparison across species. Wang et al. identified the biosynthetic pathways producing akuammilan and strychnos alkaloids across the Apocynaceae family through comparison to identified *Catharanthus roseus* genes. The authors initially sequenced the transcriptome of *Alstonia scholaris* to highlight candidates for metabolic pathways of interest. They then used synteny (i.e. shared locus on a particular chromosome) to identify homologous genes in the *Alstonia scholaris* genome based on the *C. roseus* genome and constructed a phylogenetic tree of CYP-encoding genes from MIA-producing plants across the Apocynaceae family to select candidate genes [26]. In another study, Rodríguez-López et al. built a biosynthetic pathway reconstruction algorithm to predict the iridoid biosynthetic mechanism in the Lamiaceae family based on their metabolic profiles and phylogenetic relationships, informed by genomics data. This algorithm predicted a CYP from *Callicarpa americana* that could catalyze bartsioside toward aucubin, which was then identified through transcript expression analysis and biochemically characterized to catalyze the predicted reaction, although neither compound had been previously reported in the genus [27].

2) Combined transcriptomics and metabolomics

Combined transcriptomics with metabolomics is another prevalent multi-omics approach for PNP pathway prediction. PNP pathways exhibit different expression patterns across diverse tissues, growth stages, and environmental conditions, and the accumulation of a metabolic intermediate or product typically correlates well with higher expression of the enzyme catalyzing its production. Experimenters can therefore leverage these differential expression patterns to identify enzymes catalyzing a reaction of interest by investigating correlations between gene expression and metabolite profiles. For instance, Chen et al. found that the external application of methyl jasmonate (MeJA), a hormone associated with plant stress responses, increased saponin production in *Saponaria vaccaria* and upregulated several genes encoding CYPs and glycosyltransferases for saponin biosynthesis [28]. Berman et al. applied matrix-assisted laser desorption/ionization–mass spectrometry imaging (MALDI–MSI), a technique for identification and quantification of metabolites in their spatial context, to map the high-resolution profile of cannabinoid accumulation in different tissues in *Helichrysum umbraculigerum*. Combining this mass spectrometry (MS) technique with tissue-specific transcriptomic analysis, they identified biosynthetic enzymes in the *H. umbraculigerum* cannabinoid pathway, whose expression correlated with higher cannabinoid accumulation in glandular trichomes [29].

3) Other omics approaches

Novel omics-based approaches have emerged to facilitate the discovery of PNP pathways. Stander et al. employed proteomic analysis in the latex exudate of *Rauvolfia tetraphylla* to identify 19 putative ADHs for MIA biosynthesis. Meanwhile, transcriptional co-expression analysis identified 27 putative ADHs, and genomics analysis identified 44 by BGC prediction. Integrating all three prediction results led to one ADH that was involved in yohimbine biosynthesis [30]. Researchers have also leveraged the interactions between proteins in a pathway to profile the protein-protein interactions (PPIs) for novel enzyme discovery. Wu and Liu et al. developed this interactomics-driven prediction approach to identify medium-chain dehydrogenases/reductases (MDRs) involved in MIA biosynthesis from a medicinal plant, *Mitragyna speciosa* (kratom). This approach leveraged post-translational regulation mechanisms in plants to identify novel enzymes that can interact with a known enzyme in MIA biosynthesis, namely strictosidine β-glucosidase (SGD), to form dynamic enzyme complexes [31]. Starting with 20 MDRs that were predicted by transcriptomics and metabolomics analysis, this method selected six MDR candidates that could interact with SGD in yeast and in planta. Four MDRs out of the six candidates were eventually characterized as functional enzymes, leading to the discovery of four different pathway branches, highlighting the opportunity of leveraging dynamic enzyme-enzyme interaction for novel PNP pathway discovery.

Biochemical characterization of putative enzymes

The biochemical characterization of putative enzymes can be carried out in different hosts or platforms, including plants, microorganisms, and in vitro methods (Figure 1C). Most recent studies involve two or more platforms as described in Table 1.

In vitro biochemical assays can identify a predicted enzyme's function in isolation, with optimal pH, temperature, and cofactors. Generally, putative enzymes are overexpressed in a microbial host, predominantly *Escherichia coli*, and purified. Yeast microsomes have also been used instead of *E. coli* for expressing membrane-bound enzymes like CYPs. Furthermore, certain enzymes need to be expressed in a plant host prior to in vitro assays to ensure proper folding and post-translational modifications needed for enzymatic function. For example, Nett et al. investigated the biosynthetic pathway of huperzine A in *Phlegmariurus tetrastichus* and found that the key scaffold-forming reaction catalyzed by α-carbonic anhydrase (CAH) -like (CAL) enzymes took place in the apoplast, the extracellular space between plant cell walls [32]. Attempts to express these enzymes in *E. coli* showed no activity, so the authors turned to expressing the putative CAL enzymes in *N. benthamiana* followed by isolation of apoplast extract and in

vitro enzymatic assay. This led to the characterization of a new type of decarboxylative condensation reaction which had never been previously reported for this class of enzymes.

It remains challenging to characterize enzymes that require specific cellular environments or substrates that are commercially unavailable or unstable in vitro. Characterization in microbe hosts provides advantages over in planta assays due to their short generation time, high scalability, and well-established repertoire of tools for genetic manipulation. Eukaryotic microorganism *Saccharomyces cerevisiae* (baker's yeast) serves as a feasible platform for in vivo enzyme characterization because of its similar cellular structures and organelles to plant cells, including the endoplasmic reticulum (ER), Golgi apparatus, and vacuole. Engineered yeast platforms are also able to efficiently produce PNPs in vivo as the reaction substrates, which might be unstable or inaccessible using traditional chemical synthesis approaches. For example, Deng et al. verified the function of a valencene synthase from *Alpinia oxyphylla* using an efficient terpene-producing yeast. Subsequent construction of a valencene-producing strain provided the appropriate substrate to CYP candidates and led to the functional characterization of three key CYPs in the nootkatone biosynthetic pathway [33]. Similarly, Carroll et al. used yeast strains producing various sterol substrates to identify the function of a CYP from foxglove. This CYP has been characterized to catalyze the side chain cleavage of cholesterol and campesterol, which is the first crucial step for digoxin biosynthesis [34]. Wu et al. used a strigolactone-producing *E. coli*-yeast consortium for rapid gene screening to investigate a CYP enzyme, strigol synthase, in the *Prunus* genus [35]. All these examples highlight the advantages of microbial hosts as an efficient platform for substrate production and putative enzyme expression. Furthermore, microbial hosts can continue to serve as the platforms for valuable PNP biomanufacturing. Kim et al. demonstrated the use of *E. coli* and yeast platforms to characterize novel enzymes and subsequently produce mitragynine, an analgesic candidate from kratom, by reconstructing a pathway composed of genes from kratom, mushroom, and firebush [36].

Heterologous plant hosts can provide a more appropriate cellular context for the efficient expression and characterization of plant enzymes. Agrobacterium-mediated transient transformation in the model plant host *N. benthamiana* provides an alternative to conveniently characterize plant enzymes heterologously. *N. benthamiana* has been commonly used for individual enzyme characterization (see Table 1 for more examples) in combination with characterizations using microbial hosts and in vitro assays. It is noteworthy that a plant enzyme expressed in *N. benthamiana* or yeast might exhibit different product preferences [25,28], indicating the importance of using multiple platforms for PNP pathway characterization.

Moreover, *N. benthamiana* has been widely used for entire PNP pathway reconstruction and characterization. For example, Hong et al. identified a 10-enzyme pathway for strychnine biosynthesis,

using in vitro assays and the *N. benthamiana* platform [37]. More recently, De La Peña et al. elucidated a 22-enzyme pathway including CYPs, ATs, ADHs, and 2-oxoglutarate-dependent dioxygenases (2-ODDs) for limonoid furan biosynthesis purely using *N. benthamiana* [38]. Great efforts have been made to enhance transformation efficiency in *N. benthamiana*. Carlson et al. established experimental regimes that can simultaneously deliver and co-express over twenty genes in *N. benthamiana* leaves [39]. These achievements demonstrate the ability of *N. benthamiana* as a host in characterizing long pathways via efficient co-infiltration and co-expression of multiple enzymes. Notably, Dudley et al. reconstituted the biosynthetic pathway for strictosidine [40], a central intermediate of monoterpene indole alkaloids (MIA), in *N. benthamiana*, showing its future potential for downstream MIA pathway elucidation and biomanufacturing [25,28].

Gene silencing in the native host has been used to characterize enzyme functions, using techniques such as virus-induced gene silencing (VIGS), RNA interference, and CRIPSR-Cas9 knockout lines. Although these methods have drawbacks due to the difficulty in decoupling the enzyme from complicated cross-talk and regulation within the native host, such methods can still complement other validation methods effectively. For example, Palmer et al. used VIGS to confirm the in vivo relevance of several enzymes involved in nepetalactone synthesis in *Nepeta cataria* after those enzymes had already been characterized in vitro [41]. Additionally, Sonawane et al. complemented assays in vitro and in *N. benthamiana* with the generation of CRISPR-Cas9 knockout lines in tomato plants to better understand the biosynthetic pathway for Esculeoside A, a chemical contributing to flavor in the tomato fruit [42]. This example demonstrates the importance of characterizing an enzyme's function in the native host from a breeding perspective.

Recent advances

Notable achievements on PNP pathway discovery in the past two years are summarized in Table 1. Here we highlight two representative examples.

<u>Kratom alkaloid biosynthesis</u>

Kratom (*Mitragyna speciosa*) is a tropical tree producing a variety of MIAs with medicinal uses. Among them, mitragynine and 7-hydroxymitragynine show analgesic effects and potentially have less severe side effects than opioids. The biosynthesis of MIAs involves an amine-aldehyde condensation and a glycoside hydrolysis to form the key intermediate strictosidine aglycone, catalyzed by STR and SGD, respectively. The upstream biosynthetic pathway towards strictosidine aglycone had been elucidated in other MIA-

producing plants, particularly *C. roseus*. Recently, Schotte et al. [49], Kim et al. [36], and Wu and Liu et al. [31] reported the discovery of downstream pathways towards kratom alkaloids. Two medium-chain alcohol dehydrogenases (MsDCS1 and MsDCS2) and one MT (MsEnolMT) were predicted from tissue-specific transcriptomic data and/or sequence alignment with *C. roseus* enzymes, and were verified to participate in mitragynine biosynthesis in vitro, in yeast, and in *N. benthamiana*. Although one hydroxylation step and one methylation step could not be resolved in kratom, Kim et al. achieved the microbial synthesis of mitragynine from tryptamine and secologanin with addition of a MT from *Hamelia patens* (firebush) and a CYP from *Psilocybe cubensis* (a psychedelic mushroom) [36]. Moreover, a variety of other ADHs were identified to catalyze the reduction of different isomers of strictosidine aglycone, initializing the branched pathways of other kratom alkaloids. Wu and Liu et al. also revealed the physical protein-protein interactions between those ADHs and SGD [31], demonstrating the physical interactions of subsequent enzymes in a pathway that can be leveraged for pathway discovery.

Lycopodium alkaloid biosynthesis

Lycopodium alkaloids are produced by plants in the Lycopodiaceae family (clubmosses), including more than 400 molecules. One particularly significant Lycopodium alkaloid is huperzine A (hupA), which can reversely inhibit acetylcholinesterase and impact neural synapse activity. Its bioactive properties make it a potential treatment for neurological diseases including Alzheimer's disease. However, for many years, little was known about its biosynthesis. Researchers had identified the first several enzymes catalyzing the synthesis of two precursors, 4-(2-piperidyl)acetoacetic acid (4PAA) and pelletierine, from lysine [48], but the following scaffold formation and tailoring steps were not reported until 2023. Starting from co-expression analysis in different tissues, Nett et al. [32] selected 131 enzymes sharing similar expression patterns with known enzymes in the pathway. Two short-chain dehygrogenases/reductases (SDRs), one ACT, and one CYP were found to convert pelletierine to its diene derivative in *N. benthamiana*. Previous isotope-labeled metabolomics indicated a step condensing the diene and 4PAA, but no reported enzymes could catalyze this type of chemistry. Tests of the 131 enzymes revealed that the phlegmarane scaffold was synthesized by CALs that are dissimilar to any relevant proteins reported previously in PNP biosynthesis. The two CAL enzymes were individually verified afterwards in *N. benthamiana* and in vitro using *N. benthamiana* apoplast extract. The finding of CALs involved in PNP scaffold formation represents a striking neofunctionalization of this family. Subsequently, two 2-ODDs and one hydroxylase were identified from the 131-enzyme cluster to catalyze the tailoring reactions. Together with previously reported other three tailoring 2-ODDs, the biosynthetic pathway of hupA was completely elucidated.

Conclusions and future perspectives

Recent advances in discovery of PNP pathways, using various multi-omics-guided prediction and biochemical characterization tools, have brought a deeper understanding of the intricate and diverse world of PNPs, shedding light on novel biosynthetic pathways and expanding the possibilities for applications in various fields. However, considering the rapid development of downstream metabolic engineering, including host engineering, enzyme engineering, and expression control, pathway discovery is still the rate-limiting step of PNP biomanufacturing, and thus needs to be further accelerated. Computational tools, including knowledge-based databases and machine learning (ML)-based algorithms, have proved powerful when dealing with extensive multi-omics data, as well as understanding the basic biological and chemical principles. ML-guided tools have played important roles in predicting chemical routes, BGCs, protein structures, enzyme functions, and biomedical activities [57–61], and their further development will significantly facilitate the prediction and characterization of PNP pathways.

Declaration of Interest

The authors declare no conflict of interest.

Acknowledgement

This work was supported by the National Institutes of Health—National Institute on Deafness and Other Communication Disorders under award number R21DC019206 to S. Li, the National Institutes of Health—National Institute of General Medical Sciences under award number T32GM138826, the National Science Foundation under Grants No. MCB-2338009, DBI-2019674, and IOS-2220733, the Schwartz Research Fund, the Academic Venture Fund from Cornell Atkinson Center for Sustainability, and the Research Innovation Fund from Cornell Institute for Digital Agriculture. We thank Dr. Yinan Wu, Dr. Anna Koganitsky, Chang Liu, and Dong Oh Han for their valuable comments on this manuscript.

Figure captions

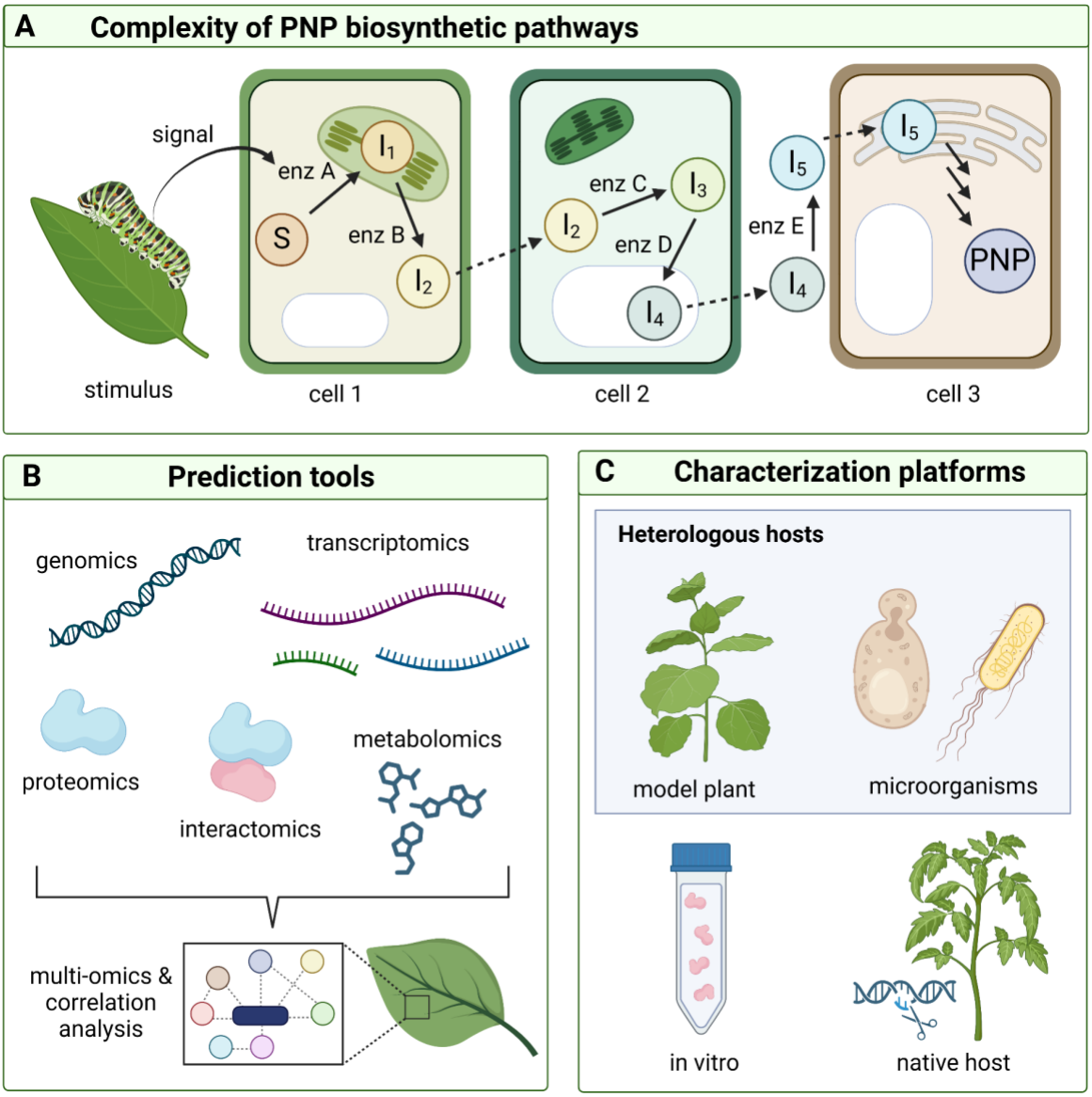

Figure 1. Overview of PNP biosynthetic pathway discovery challenges and elucidation methods. A) An abstract PNP biosynthetic pathway depicts the challenges associated with PNP pathway discovery, which are largely related to the complexity of these pathways. This includes many cascading steps between substrates (S) and intermediates (I), which may require elicitation by a stimulus and may be localized across different plant tissues or subcellular compartments, such as the chloroplast or vacuole. B) Tools for PNP pathway prediction include genomic, transcriptomic, proteomic, interactomic, and metabolomic analysis methods. These methods are frequently used in combination to allow for analysis of the correlated spatial and/or temporal presence of various pathway-associated molecules. C) Once genes

predicted to be involved in PNP biosynthesis have been identified, they can be functionally characterized in heterologous host systems (e.g. model plants such as *Nicotiana benthamiana* or microbes such as *Escherichia coli* or *Saccharomyces cerevisiae*), in vitro, or through gene silencing in the native host plant. Created with BioRender.com.

306 **Table 1. Notable developments in PNP biosynthetic pathway discovery made**
307 **in the past two years.**

Type of PNP	End Product	Host Plant	Enzymes Involved	Prediction method (transcriptomics: T, metabolomics: M, genomics: G, proteomics: P, interactomics: I)	Characterization platform	Ref.
terpenoid	baccatin III	*Taxus* spp.	CYP	T (comparative expression) and M	*N. benthamiana*	[43]
terpenoid	paclitaxel	*Taxus* spp.	CYP, CoA ligase	T (gene co-expression)	*N. benthamiana*	[44]
terpenoid	nootkatone	*Alpinia oxyphylla*	TPS, CYP, cytochrome P450 reductase (CPR), SDR	T and M (tissue-specific co-expression); gene candidates narrowed using phylogenetic analysis	*S. cerevisiae* and in vitro	[33]
terpenoid	ginkgolides	*Ginkgo biloba*	CYP	G (BGC mining) and T (gene co-expression)	*S. cerevisiae* and *N. benthamiana*	[25]
terpenoid	limonoids (kihadalactone A and azadirone)	*Citrus sinensis*, *Melia azedarach*, *Azadirachta indica*	CYP, sterol isomerase (SI), ACT, SDR, aldo-keto reductase (AKR), furan synthase	T and M (tissue-specific co-expression)	*N. benthamiana*	[38]
terpenoid	saponin	*Saponaria vaccaria*	CYP, GT	T (comparative T via elicitation); gene candidates narrowed using phylogenetic analysis	*S. cerevisiae*, *N. benthamiana*, and in vitro (yeast microsomes)	[28]
terpenoid	triptonide	*Tripterygium wilfordii*	CYP	T (BLAST gene mining)	*S. cerevisiae* and *N. benthamiana*	[45]
phenolic	verbascoside	Lamiales spp.	hydroxycinnamoyltransferase (HCT), CYP	T (tissue-specific gene co-expression); gene candidates narrowed using phylogenetic analysis	in vitro (purified enzymes and yeast microsomes), *E. coli*, and *N. benthamiana*	[46]

phenolic	cannabinoids	*Helichrysum umbraculigerum*	acyl-CoA transferase (CoAT), tetraketide synthase (TKS), GT, ACT	G (genome assembly), M, T	in vitro (purified enzymes and yeast microsomes), *E. coli*, *N. benthamiana*, and *S. cerevisiae*	[29]
indole-derived	benzoxazinoids	*Aphelandra squarrosa* & *Lamium galeobdolon*	CYP, Fe(II)/2-oxoglutarate-dependent dioxygenase (2OGD), GT, MT	T and M (tissue-specific co-expression); comparative T (producer vs. non-producer species)	*N. benthamiana* and in vitro (yeast microsomes)	[47]
sesquiterpene alkaloid	huperzine A	*Phlegmariurus tetrastichus*	α-carbonic anhydrase-like protein (CAL; catalyzing condensation & cyclization), 2OGD, α/β hydrolase (ABH)	T (gene co-expression)	*N. benthamiana* and in vitro (*N. benthamiana* extract)	[32]
sesquiterpene alkaloid	huperzine A	*Phlegmariurus tetrastichus*	decarboxylase, copper amine oxidase (CAO), polyketide synthase (PKS), 2OGD	T and M (tissue-specific co-expression)	*N. benthamiana* and in vitro	[48]
monoterpene indole alkaloid (MIA)	mitragynine	*Mitragyna speciosa*	MDR, MT	T (BLAST gene mining)	in vitro, *E. coli*, and *S. cerevisiae*	[36]
MIA	mitragynine and related compounds	*Mitragyna speciosa*	medium-chain alcohol dehydrogenase	T and M (tissue-specific co-expression)	in vitro and *N. benthamiana*	[49]

			(MDR), MT			
MIA	N/A-MIA upstream universal scaffold	*Mitragyna speciosa*	MDR	T and M (tissue-specific co-expression); I	in vitro, *S. cerevisiae*, and *N. benthamiana*	[31]
MIA	strychnine, brucine, diaboline	*Strychnos nux-vomica*	CYP, ABH, MDR, ACT, MT	T and M (tissue-specific co-expression); comparative T (producer vs. non-producer species)	*N. benthamiana* and in vitro (*N. benthamiana* extract)	[37]
MIA	yohimbanes	*Rauvolfia tetraphylla*	MDR	G (genome assembly), M, T, P	*S. cerevisiae* and in vitro	[30]
MIA	akuammiline	*Alstonia scholaris*	CYP, oxidoreductase, ACT	T and G (BLAST gene mining and synteny)	in vitro (yeast microsomes) and *S. cerevisiae*	[26]
terpenoid	astramalabaricosides	*Astragalus membranaceus*	TPS	T (BLAST gene mining)	*S. cerevisiae* and in vitro	[50]
terpenoid	saponin	*Lonicera macranthoides* & *Lonicera japonica*	TPS, GT	comparative G (genome assembly, collinearity analysis), T, M	in vitro and *N. benthamiana*	[51]
flavonoid	melitidin	*Citrus grandis*	GT, ACT, rhamnosyltransferase	G (genome-wide association studies) and M	in vitro and *N. benthamiana*	[52]
flavonoid	hyperoside	*Hypericum monogynum*	CYP, flavonol synthase (FLS), GT	T and M (tissue-specific co-expression)	in vitro, *E. coli*, and *S. cerevisiae*	[53]
aldoxime	phenylacetaldoxime glucoside	*Miconia microphysca*	CYP, GT	comparative T and M via elicitation	in vitro (purified enzymes and yeast microsomes) and *N. benthamiana*	[54]
phenolic	bibenzyls	*Cannabis sativa*	CoA ligase, double-bond reductase, PKS	T (BLAST gene mining)	in vitro	[55]
acyl sugar	N/A—nonspecific analysis	*Solanum lycopersicum* & *Solanum pennellii*	β-ketoacyl reductase (KAR),	comparative G (predicted orthologs, narrowed gene	virus-induced gene silencing in host	[56]

	rubisco small subunit (RBCS)	candidates based on characteristic signs of positive selection), M, T

References

1. Anarat-Cappillino G, Sattely ES: **The chemical logic of plant natural product biosynthesis**. *Curr Opin Plant Biol* 2014, **19**:51–58.

2. Xu M, Galhano R, Wiemann P, Bueno E, Tiernan M, Wu W, Chung IM, Gershenzon J, Tudzynski B, Sesma A, et al.: **Genetic evidence for natural product-mediated plant–plant allelopathy in rice (*Oryza sativa*)**. *New Phytol* 2012, **193**:570–575.

3. Wouters FC, Blanchette B, Gershenzon J, Vassão DG: **Plant defense and herbivore counter-defense: benzoxazinoids and insect herbivores**. *Phytochem Rev* 2016, **15**:1127–1151.

4. Atanasov AG, Zotchev SB, Dirsch VM, Orhan IE, Banach M, Rollinger JM, Barreca D, Weckwerth W, Bauer R, Bayer EA, et al.: **Natural products in drug discovery: advances and opportunities**. *Nat Rev Drug Discov* 2021, **20**:200–216.

5. Yuan H, Ma Q, Ye L, Piao G: **The Traditional Medicine and Modern Medicine from Natural Products**. *Molecules* 2016, **21**.

6. Dayan FE, Cantrell CL, Duke SO: **Natural products in crop protection**. *Bioorg Med Chem* 2009, **17**:4022–4034.

7. Christianson DW: **Structural and Chemical Biology of Terpenoid Cyclases**. *Chem Rev* 2017, **117**:11570–11648.

8. Dastmalchi M, Chen X, Hagel JM, Chang L, Chen R, Ramasamy S, Yeaman S, Facchini PJ: **Neopinone isomerase is involved in codeine and morphine biosynthesis in opium poppy**. *Nat Chem Biol* 2019, **15**:384–390.

9. Li Y, Smolke CD: **Engineering biosynthesis of the anticancer alkaloid noscapine in yeast**. *Nat Commun* 2016, **7**:1–14.

10. Srinivasan P, Smolke CD: **Biosynthesis of medicinal tropane alkaloids in yeast**. *Nature* 2020, **585**:614–619.

11. Courdavault V, O'Connor SE, Jensen MK, Papon N: **Metabolic engineering for plant natural products biosynthesis: new procedures, concrete achievements and remaining limits**. *Nat Prod Rep* 2021, **38**:2145–2153.

12. Birchfield AS, McIntosh CA: **Metabolic engineering and synthetic biology of plant natural products – A minireview**. *Curr Plant Biol* 2020, **24**:100163.

13. Cravens A, Payne J, Smolke CD: **Synthetic biology strategies for microbial biosynthesis of plant natural products**. *Nat Commun* 2019, **10**:1–12.

14. Han J, Wu Y, Zhou Y, Li S: **Engineering *Saccharomyces cerevisiae* to produce plant benzylisoquinoline alkaloids**. *aBIOTECH* 2021, **2**:264–275.

15. Caputi L, Franke J, Farrow SC, Chung K, Payne RME, Nguyen TD, Dang TTT, Teto Carqueijeiro IS, Koudounas K, De Bernonville TD, et al.: **Missing enzymes in the biosynthesis of the anticancer drug vinblastine in *Madagascar periwinkle***. *Science* 2018, **360**:1235–1239.

16. Han R, Rai A, Nakamura M, Suzuki H, Takahashi H, Yamazaki M, Saito K: **De Novo Deep Transcriptome Analysis of Medicinal Plants for Gene Discovery in Biosynthesis of Plant Natural Products**. *Methods Enzymol* 2016, **576**:19–45.

17. Guo J, Huang Z, Sun J, Cui X, Liu Y: **Research Progress and Future Development Trends in Medicinal Plant Transcriptomics**. *Front Plant Sci* 2021, **12**:691838.

18. Tyagi P, Singh D, Mathur S, Singh A, Ranjan R: **Upcoming progress of transcriptomics studies on plants: An overview**. *Front Plant Sci* 2022, **13**.

19. Frey M, Chomet P, Glawischnig E, Stettner C, Grün S, Winklmair A, Eisenreich W, Bacher A, Meeley RB, Briggs SP, et al.: **Analysis of a chemical plant defense mechanism in grasses**. *Science* 1997, **277**:696–699.

20. Polturak G, Osbourni A: **The emerging role of biosynthetic gene clusters in plant defense and plant interactions**. *PLoS Pathog* 2021, **17**:e1009698.

21. Wu Y, Gong FL, Li S: **Leveraging yeast to characterize plant biosynthetic gene clusters**. *Curr Opin Plant Biol* 2023, **71**:102314.

22. Kautsar SA, Suarez Duran HG, Blin K, Osbourn A, Medema MH: **plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters**. *Nucleic Acids Res* 2017, **45**:W55–W63.

23. Schläpfer P, Zhang P, Wang C, Kim T, Banf M, Chae L, Dreher K, Chavali AK, Nilo-Poyanco R, Bernard T, et al.: **Genome-Wide Prediction of Metabolic Enzymes, Pathways, and Gene Clusters in Plants**. *Plant Physiol* 2017, **173**:2041–2059.

24. Töpfer N, Fuchs LM, Aharoni A: **The PhytoClust tool for metabolic gene clusters discovery in plant genomes**. *Nucleic Acids Res* 2017, **45**:7049–7063.

25. Forman V, Luo D, Geu-Flores F, Lemcke R, Nelson DR, Kampranis SC, Staerk D, Møller BL, Pateraki I: **A gene cluster in *Ginkgo biloba* encodes unique multifunctional cytochrome P450s that initiate ginkgolide biosynthesis**. *Nat Commun* 2022, **13**:1–14. **(*)** Novel multifunctional CYPs including scarless C-C bond cleavage and carbon skeleton rearrangement.

26. Wang Z, Xiao Y, Wu S, Chen J, Li A, Tatsis EC: **Deciphering and reprogramming the cyclization regioselectivity in bifurcation of indole alkaloid biosynthesis**. *Chem Sci* 2022, **13**:12389–12395.

27. Rodríguez-López CE, Jiang Y, Kamileen MO, Lichman BR, Hong B, Vaillancourt B, Buell CR, O'Connor SE: **Phylogeny-Aware Chemoinformatic Analysis of Chemical Diversity in Lamiaceae Enables Iridoid Pathway Assembly and Discovery of Aucubin Synthase**. *Mol Biol Evol* 2022, **39**.

28. Chen X, Hudson GA, Mineo C, Amer B, Baidoo EEK, Crowe SA, Liu Y, Keasling JD, Scheller H V.: **Deciphering triterpenoid saponin biosynthesis by leveraging transcriptome response to methyl jasmonate elicitation in *Saponaria vaccaria***. *Nat Commun* 2023, **14**:1–15. **(*)** Different products observed in *N. benthamiana* vs. *S. cerevisiae.*

29. Berman P, de Haro LA, Jozwiak A, Panda S, Pinkas Z, Dong Y, Cveticanin J, Barbole R, Livne R, Scherf T, et al.: **Parallel evolution of cannabinoid biosynthesis**. *Nat Plants 2023 9:5* 2023, **9**:817–831. **(*)** Products observed in yeast/in vitro not seen in *N. benthamiana*.

30. Stander EA, Lehka B, Carqueijeiro I, Cuello C, Hansson FG, Jansen HJ, Dugé De Bernonville T, Birer Williams C, Vergès V, Lezin E, et al.: **The *Rauvolfia tetraphylla* genome suggests multiple distinct biosynthetic routes for yohimbane monoterpene indole alkaloids**. *Commun Biol* 2023, **6**:1–19.

31. Wu Y, Liu C, Koganitsky A, Gong FL, Li S: **Discovering Dynamic Plant Enzyme Complexes in Yeast for Kratom Alkaloid Pathway Identification****. *Angew Chem Int Ed* 2023, **62**:e202307995. **(**)** Established protein-protein interaction (BiFC) assay in yeast for prediction of downstream enzymes in branched pathway.

32. Nett RS, Dho Y, Tsai C, Passow D, Martinez Grundman J, Low YY, Sattely ES: **Plant carbonic anhydrase-like enzymes in neuroactive alkaloid biosynthesis**. *Nature* 2023, **624**:182–191. **(**)** Identification of enzyme localized to apoplast (requires expression in planta for proper function).

33. Deng X, Ye Z, Duan J, Chen F, Zhi Y, Huang M, Huang M, Cheng W, Dou Y, Kuang Z, et al.: **Complete pathway elucidation and heterologous reconstitution of (+)-nootkatone biosynthesis from *Alpinia oxyphylla***. *New Phytol* 2024, **241**:779–792.

34. Carroll E, Ravi Gopal B, Raghavan I, Mukherjee M, Wang ZQ: **A cytochrome P450 CYP87A4 imparts sterol side-chain cleavage in digoxin biosynthesis**. *Nat Commun* 2023, **14**:1–11.

35. Wu S, Zhou A, Hiugano K, Yoda A, Xie X, Yamane K, Miura K, Nomura T, Li Y: **Identification of a Prunus MAX1 homolog as a unique strigol synthase**. *New Phytol* 2023, **239**:1819–1833.

36. Kim K, Shahsavarani M, Garza-García JJO, Carlisle JE, Guo J, De Luca V, Qu Y: **Biosynthesis of kratom opioids**. *New Phytol* 2023, **240**:757–769. **(*)** Achieved microbial biosynthesis of mitragynine with alternative enzymes from firebush and mushroom.

37. Hong B, Grzech D, Caputi L, Sonawane P, López CER, Kamileen MO, Hernández Lozada NJ, Grabe V, O'Connor SE: **Biosynthesis of strychnine**. *Nature* 2022, **607**:617–622.

38. De La Peña R, Hodgson H, Liu JCT, Stephenson MJ, Martin AC, Owen C, Harkess A, Leebens-Mack J, Jimenez LE, Osbourn A, et al.: **Complex scaffold remodeling in plant triterpene biosynthesis**. *Science* 2023, **379**:361–368.

39. Carlson ED, Rajniak J, Sattely ES: **Multiplicity of Agrobacterium infection of *Nicotiana benthamiana* for transient DNA delivery**. *bioRxiv* 2023, doi:10.1101/2023.02.07.527564.

40. Dudley QM, Jo S, Guerrero DAS, Chhetry M, Smedley MA, Harwood WA, Sherden NH, O'Connor SE, Caputi L, Patron NJ: **Reconstitution of monoterpene indole alkaloid biosynthesis in genome engineered Nicotiana benthamiana**. *Commun Biol* 2022, **5**:1–12.

41. Palmer L, Chuang L, Siegmund M, Kunert M, Yamamoto K, Sonawane P, O'Connor SE: **In vivo characterization of key iridoid biosynthesis pathway genes in catnip (*Nepeta cataria*)**. *Planta* 2022, **256**:1–11.

42. Sonawane PD, Gharat SA, Jozwiak A, Barbole R, Heinicke S, Almekias-Siegl E, Meir S, Rogachev I, Connor SEO, Giri AP, et al.: **A BAHD-type acyltransferase concludes the biosynthetic pathway of non-bitter glycoalkaloids in ripe tomato fruit**. *Nat Commun* 2023, **14**:1–15.

43. Jiang B, Gao L, Wang H, Sun Y, Zhang X, Ke H, Liu S, Ma P, Liao Q, Wang Y, et al.: **Characterization and heterologous reconstitution of Taxus biosynthetic enzymes leading to baccatin III**. *Science* 2024, doi:10.1126/SCIENCE.ADJ3484.

44. Zhang Y, Wiese L, Fang H, Alseekh S, Perez de Souza L, Scossa F, Molloy J, Christmann M, Fernie AR: **Synthetic biology identifies the minimal gene set required for paclitaxel biosynthesis in a plant chassis**. *Mol Plant* 2023, **16**:1951–1961.

45. Hansen NL, Kjaerulff L, Heck QK, Forman V, Staerk D, Møller BL, Andersen-Ranberg J: **Tripterygium wilfordii cytochrome P450s catalyze the methyl shift and epoxidations in the biosynthesis of triptonide**. *Nat Commun* 2022, **13**:1–12. **(*)** Different product distributions observed in *N. benthamiana* vs. *S. cerevisiae*.

46. Yang Y, Xi D, Wu Y, Liu T: **Complete biosynthesis of the phenylethanoid glycoside verbascoside**. *Plant Commun* 2023, **4**.

47. Florean M, Luck K, Hong B, Nakamura Y, O'Connor SE, Köllner TG: **Reinventing metabolic pathways: Independent evolution of benzoxazinoids in flowering plants**. *Proc Natl Acad Sci U S A* 2023, **120**:e2307981120.

48. Nett RS, Dho Y, Low YY, Sattely ES: **A metabolic regulon reveals early and late acting enzymes in neuroactive Lycopodium alkaloid biosynthesis**. *Proc Natl Acad Sci U S A* 2021, **118**:e2102949118.

49. Schotte C, Jiang Y, Grzech D, Dang TTT, Laforest LC, León F, Mottinelli M, Nadakuduti SS, McCurdy CR, O'Connor SE: **Directed Biosynthesis of Mitragynine Stereoisomers**. *J Am Chem Soc* 2023, **145**:4957–4963.

50. Song Z, Chen D, Sui S, Wang Y, Cen S, Dai J: **Characterization of a Malabaricane-Type Triterpene Synthase from *Astragalus membranaceus* and Enzymatic Synthesis of Astramalabaricosides**. *J Nat Prod* 2023, **86**:1815–1823.

51. Yin X, Xiang Y, Huang FQ, Chen Y, Ding H, Du J, Chen X, Wang X, Wei X, Cai YY, et al.: **Comparative genomics of the medicinal plants *Lonicera macranthoides* and *L. japonica* provides insight into genus genome evolution and hederagenin-based saponin biosynthesis**. *Plant Biotechnol J* 2023, **21**:2209–2223.

52. Shen S, Wang S, Yang C, Wang C, Zhou Q, Zhou S, Zhang R, Li Y, Wang Z, Dai L, et al.: **Elucidation of the melitidin biosynthesis pathway in pummelo**. *J Integr Plant Biol* 2023, **65**:2505–2518.

53. Yingying W, Zhirong C, Qianqian L, Shuai Z, Yongyi L, Xueyan L, Lingyi K, Jun L: **The parallel biosynthesis routes of hyperoside from naringenin in Hypericum monogynum**. *Hortic Res* 2023, **10**.

54. Müller AT, Nakamura Y, Reichelt M, Luck K, Cosio E, Lackus ND, Gershenzon J, Mithöfer A, Köllner TG: **Biosynthesis, herbivore induction, and defensive role of phenylacetaldoxime glucoside**. *Plant Physiol* 2023, **194**:329–346.

55. Boddington KF, Soubeyrand E, Van Gelder K, Casaretto JA, Perrin C, Forrester TJB, Parry C, Al-Abdul-Wahid MS, Jentsch NG, Magolan J, et al.: **Bibenzyl synthesis in *Cannabis sativa* L**. *Plant J* 2022, **109**:693–707.

56. Ji W, Mandal S, Rezenom YH, McKnight TD: **Specialized metabolism by trichome-enriched Rubisco and fatty acid synthase components**. *Plant Physiol* 2023, **191**:1199–1213.

57. Yuan Y, Shi C, Zhao H: **Machine Learning-Enabled Genome Mining and Bioactivity Prediction of Natural Products**. *ACS Synth Biol* 2023, **12**:2650–2662.

58. Boob AG, Chen J, Zhao H: **Enabling pathway design by multiplex experimentation and machine learning**. *Metab Eng* 2024, **81**:70–87.

59. Yu T, Cui H, Li JC, Luo Y, Jiang G, Zhao H: **Enzyme function prediction using contrastive learning**. *Science* 2023, **379**:1358–1363.

60. Yu H, Deng H, He J, Keasling JD, Luo X: **UniKP: a unified framework for the prediction of enzyme kinetic parameters**. *Nat Commun* 2023, **14**:1–13.

61. Mahood EH, Kruse LH, Moghe GD: **Machine learning: A powerful tool for gene function prediction in plants.** *Appl Plant Sci* 2020 **8**:e11376.