Omics strategies for plant natural product biosynthesis

Si Wan[1,2,3], Peter J. Schaap[2*], Maria Suarez-Diez[2*] and Wei Li[1*]

[1] *Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China*
[2] *Laboratory of Systems and Synthetic Biology, Wageningen University & Research, 6708WE Wageningen, The Netherlands*
[3] *Guangxi Key Laboratory of Medicinal Resources Protection and Genetic Improvement/ Guangxi Engineering Research Center of TCM Resource Intelligent Creation, National Center for TCM Inheritance and Innovation, Guangxi Botanical Garden of Medicinal Plants, Nanning 530023, China*
* Corresponding authors, E-mail: peter.schaap@wur.nl; maria.suarezdiez@wur.nl; liwei11@caas.cn

Abstract

Plant natural products play a crucial role in ecological balance, human health, industrial applications, and biodiversity conservation, making them invaluable across various fields. Elucidation of their biosynthetic pathways is important for further synthetic biology applications. Through gene cluster, co-expression, and population association assays, researchers leverage extensive genomics, transcriptomics, and metabolomics data produced by multi-omic technologies to uncover metabolic genes involved in biosynthetic pathways. New techniques such as single-cell sequencing, MS imaging, and machine learning have shown their potential. Here we reviewed multiple omics studies of natural product biosynthesis and discussed the promise and potential of developing techniques for this task.

Introduction

Many valuable plant secondary metabolites are unique to certain plant species or groups of closely related species involved in complex biosynthetic pathways, and different kinds of catalytic enzymes. Large-scale production of natural plant products is hampered by the long cultivation periods, requirements for specific cultivation conditions, reliance on climatological conditions, seasonally dependent growth, and competition for arable land for food production[1]. In addition, extraction of these secondary metabolites is often quite costly and has a large environmental impact due to the low concentrations of the products[2]. As an example, the production of 1 kg of flavonoids requires the processing of 0.25−33 tons of dry fruits or vegetables[3], but the yearly global need for flavonoid extracts exceeds 3,000 tons (estimated by market size). As a result, microbial production of value-added, plant-derived compounds increasingly attracts commercial interest in the food and pharmaceutical industries. Synthetic biology strategies necessitate heterologous expression of plant-derived genes to reconstruct the biosynthetic pathway within the target microbial host. This demands detailed knowledge of these pathways and several studies have focused on elucidating the biosynthetic pathways.

Along with the development of high-throughput technologies, biologists can comprehensively measure genome constitution, gene expression, and metabolism: Next Generation Sequencing (NGS) technologies have pushed the development of plant genomics, enabling comprehensive investigation of genome composition, structure, functional elements, and evolutionary dynamics. Transcriptomics, which also leverages NGS, examines global gene expression patterns by quantifying RNA abundance per gene. Metabolomics profiles the complete set of metabolites using chromatographic techniques coupled with mass spectrometry, and nuclear magnetic resonance (NMR) spectroscopy. These methods collectively regarded as 'omics' generate vast amounts of data requiring bioinformatic analysis. Along with the development of multi-omics techniques, reverse genetics started to flourish in the discovery of pathways for the biosynthesis of plant natural products

in recent years. *De novo* genome assembly, followed by structural and functional annotation identifies genes and their functions based on homology; biosynthetic gene clusters (BGC) are regularly found by taking advantage of the similarity of structural arrangements of genes in common biosynthetic pathways; co-expression of function elucidated genes from transcriptome studies provides convincing inferences of candidate catalytic genes; metabolome measurement, and second-generation sequencing of plant populations enable metabolome genome-wide association studies (mGWAS) to identify variation in Single Nucleotide polymorphisms (SNPs) related to secondary metabolism; comparative genomics and analysis of the pan-genome of a taxon are useful in detecting large structure variations for related metabolic phenotypes; comparison of omics datasets also dictate the evolutionary trajectory of plant natural product biosynthesis pathways[4,5].

This review highlights recent advancements in plant natural product biosynthesis enabled by multi-omics approaches (Fig. 1) and discusses future perspectives facilitated by emerging technological innovations. These omics strategies, including *de novo* genome assembly, transcriptome coexpression, mGWAS, and pan-genome, have substantially advanced the identification of catalytic enzymes, characterization of metabolic gene clusters, and elucidation of evolutionary trajectories. We also examined emerging technologies, such as single-cell sequencing, spatial transcriptomics, mass spectrometry imaging, and, notably, machine learning, a transformative computational tool with broad interdisciplinary applications These advanced techniques are expected to push plant natural product research to the next level.

Plant natural products and their biosynthetic pathways

Plants are sessile organisms. As they cannot escape from environmental threats and biotic stressors they developed defense strategies by generating various natural products to combat plant diseases, herbivory insects, abiotic stresses, and in addition, to attract beneficial
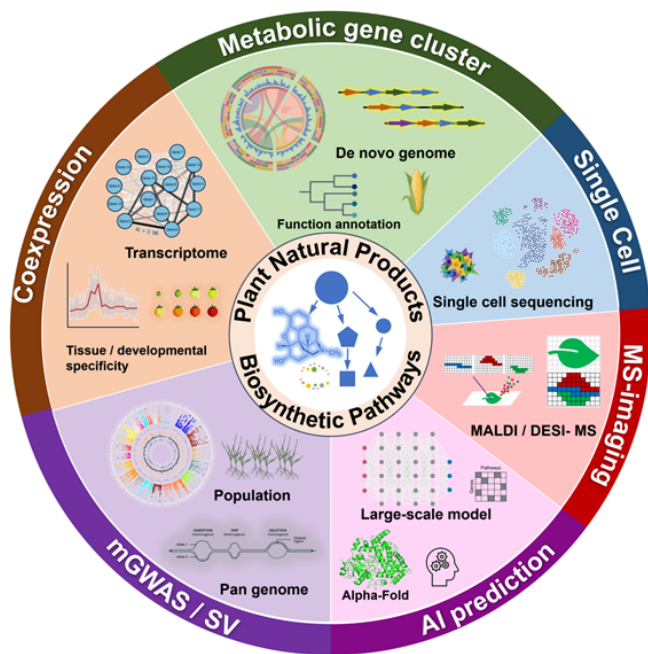
Fig. 1 Overview of methods to elucidate natural biosynthetic pathways for plant natural products. Computational methods are combined and integrated with genomics, transcriptomics, and metabolomics measurements to identify relevant genes.

organisms like pollinators[4,6]. These products have also proven valuable for humans and plant natural products can be used in a range of applications. For instance, pyrethrins from pyrethrum and benzoxazinoids from maize show effects of repelling and killing insects[7,8], hence they are considered natural pesticides. Other plant natural products, such as polyamines and phenolics participate in plant growth and development such as flowering and fruit set[9,10]. Furthermore, a broad range of natural products including terpenoids and fatty acid derivatives from tomato and cucumber fruits are appreciated as food additives due to their contribution to distinctive and complex flavor profiles[11–13]. In addition, some of these metabolites possess pharmacological activities like anti-inflammatory and anti-pathogenic activities and are therefore used as plant-derived drugs or as traditional medicines for curing human diseases[4–6,14]. Examples are artemisia for malaria treatment[15,16], paclitaxel for cancer chemotherapy[17], and tropane alkaloids as anticholinergics[18].

Plant natural products are regulated by various factors. The regulation of gene expression in biosynthetic pathways is usually accomplished by transcription factors (TFs) directing downstream gene expression[19,20]. Under drought stress, baicalin and wogonoside levels in *Scutellaria baicalensis* initially showed a slight decrease followed by a significant increase at a later stage. Transcriptomic analysis revealed that TF SbWRKY34 negatively regulated this drought response[21]. In rice, a receptor-like kinase (OsRLCK160) was shown to interact with a bZIP family TF (OsbZIP48) and promoted flavonoid accumulation to enhance UV-B tolerance[22]. Phytochrome-interacting factors (PIFs), members of the basic helix-loop-helix (bHLH) transcription factor family, directly associate with phytochromes, the red/far-red light photoreceptors[23]. In tomato melatonin biosynthesis, SlPIF4 acts as a negative regulator by suppressing the expression of the key biosynthetic gene *SlCOMT2*. However, under red light conditions, SlphyB2 (phytochrome B2) promotes the degradation of SlPIF4, thereby relieving the repression and enhancing melatonin accumulation[24]. Integrating transcriptome and spatio-temporal metabolome data resulted in a metabolic network of tomato, which revealed the regulation of TFs, such as *SlMYB75* and *SlERF.G3-like* in flavonoids biosynthesis, and *SlGAME9* and *SlbHLH14* in steroidal glycoalkaloids (SGA)[25]. In cucumber, GWAS revealed a BGC of nine genes involved in cucurbitacins biosynthesis and two TFs regulating this BGC in leaves and fruits, named *Bl* (*Bitter leaf*) and *Bt* (*Bitter fruit*)[26]. Regulations may be achieved by epigenetic modification as well[20]. For instance, the ubiquitin-like protein SILENCING DEFECTIVE 2 (SDE2) interacted with HETEROCHROMATIN PROTEIN 1 (LHP1) and together increased the H3K27me3 level to repress the anthocyanin biosynthesis in *Arabidopsis thaliana*[27]. A study on telomere-to-telomere (T2T) genome assemblies of two Chili pepper species revealed the placenta-specific biosynthesis of capsaicinoid was coordinately regulated by the low methylation level at the open chromatin regions (OCRs) with placenta specificity[28].

Plant-specialized metabolites are generally synthesized from common primary metabolites[29]. Depending on the type of precursors and sharing structures, they are divided into several groups, commonly known as terpenoids, alkaloids, phenylpropanoids, polyketides, etc.[5]. Synthesis of these plant natural products usually requires complex pathways formed by enzymes from multiple gene families. Several of the enzyme families involved have been thoroughly researched. Notably, enzymes like cytochrome P450s (CYP450s) and uridine diphosphate (UDP)-glycosyltransferases (UGTs) are part of large families comprising hundreds of members[20,21,30,31]. The sequence diversity within biosynthetic pathways poses a challenge for comprehensive candidate gene identification, as it is impractical to functionally assess all possibilities simultaneously. However, the use of reverse genetics enables the selection of the most probable candidates for functional tests to relieve the experimental work intensity.

De novo genome assembly and annotation, and identification of metabolic gene clusters

Plants have complex genomes with large sizes and bulk of tandem repeats, which make them difficult to assemble. However, the development of the 3rd generation ultra-long sequencing techniques allows plant scientists to obtain plant *de novo* genomes at the chromosome-scale, including gap-free and telomere-to-telomere (T2T) level resolutions[32]. This development also facilitates the discovery of biosynthetic gene clusters (BGCs) which are groups of genes involved in a specific biosynthetic pathway and closely located on a chromosome[33–38]. Although genes from BGCs of eukaryotic plants cannot be controlled by unique promoters, they often participate in the same metabolic pathway and experience shared regulation[39], such as the seed-specific BGC for aescin and aesculin biosynthesis in *Aesculus chinensis*[40], and the pathogen-defensive BGCs in bread wheat[41].

BGCs might arise as a result of gene duplication, genome reorganization, or whole genome duplication, and might acquire new functions for natural product biosynthesis later during evolution[40,42–44]. BGCs across different plant species can evolve analogous functions through mechanisms of convergent or parallel evolution[45,46]. The first biosynthetic gene cluster (BGC) identified in plants was discovered in maize in 1997, comprising a tryptophan synthase gene (*Bx1*) and four cytochrome P450 genes (*Bx2–Bx5*) located on chromosome 4. This BGC is responsible for the biosynthesis of benzoxazinoid, a defensive metabolite against pathogens[47]. Subsequently, two UDP-glycosyltransferases (UGT) *Bx8* and *Bx9* were identified as additional components of this cluster[48]. Since then, an increasing number of BGCs have been identified in plants, including

a cluster comprising 10 synthetic genes for noscapine biosynthesis in *Papaver somniferum* reported in 2012[49]. However, it is important to note that not all genes involved in natural product biosynthesis are physically clustered. For example, genes responsible for the biosynthesis of berberine in *Coptis chinensis* and saponarioside in soapwort (*Saponaria officinalis*) are dispersed throughout the genome[50,51].

As finding BGCs is often an important part of genome annotation and analysis, bioinformatic tools have been developed specifically for BGC identification. PlantiSMASH is a widely used online tool that efficiently identifies genomic loci encoding multiple (sub)families of specialized metabolic enzymes. To perform this task, plantiSMASH integrates a curated library of enzyme families associated with plant biosynthetic pathways, in conjunction with CD-HIT-based clustering of predicted protein sequences[52].

The identification of BGCs in newly assembled plant genomes informed our understanding of these specialized metabolic pathways, a progress largely driven by advances in genomics. The assembled genome of *Taxus chinensis* var. *mairei* helped identify a BGC for paclitaxel, an important anti-cancer compound. Within this BGC, members of the CYP725A family (T5αH, T13αH, T2αH, T7βH, T9αH1, TOT1) catalyze specific hydroxylation reactions, which are critical steps in forming this complex natural product[53,54]. A mirror-structured BGC for *p*-menthane monoterpenoids biosynthesis was identified in the assembled genome of *Schizonepeta tenuifolia*[55]. A gene cluster for ginkgolide biosynthesis containing five *CYP450*s was discovered by mining the published *Gingko biloba* genome[56]. Hyperforin in St John's wort has been shown to involve two distinct BGCs[57].

In field crops and horticultural plants, typical BGCs associated with plant defense have been identified. Examples include BGCs responsible for phytoalexin production in wheat, the avenacin biosynthetic cluster in oat, and the BGC for the biosynthesis of falcarindiol a highly modified fatty acid, in tomato[14,41,58]. Additionally, a fatty acid metabolic gene cluster conserved across the Poaceae family has been shown to control male fertility rice thereby influencing rice yield[39]. The genomes of *Isodon rubescens,* and *Tripterygium wilfordii* contained tandem-duplicated *CYP706V* and *CYP82D* clusters for oxidative modification of biosynthesis oridonin and triptolide, respectively, both of which are diterpenoid with anti-inflammatory and anticancer applications primarily isolated from plants and used in traditional Chinese medicine[59,60]. It is worth noting that members of the CYP706 family have been found in various other plant species and are associated with the biosynthesis of compounds that contribute to plant defense and resistance. Examples are CYP706A3 involved in terpenoid biosynthesis cluster responsible for flower defense and herbicide resistance in *A. thaliana*, CYP706B1 from cotton, and CYP706M1 from *Alaska cedar* both of which were involved in the biosynthesis of anti-herbivory sesquiterpenes. Furthermore, CYP706C55 from *Eucalyptus cladocalyx* was found to be involved in the biosynthesis of cyanogenic glucosides, a class of chemicals defending plants against herbivores[61–65]. These BGC discoveries indicated the importance of genome architecture in informing our understanding of plant natural product biosynthesis.

Population genetics, Pan-genome, and mGWAS

In plant species with numerous germplasm resources, pan-genome analysis integrates genomic information across different germplasms and populations. This population-based genomic approach enables efficient identification of key genetic loci and facilitates the evolutionary tracing of natural product biosynthesis. Third-generation

sequencing further enables the precise identification of genomic locations with large structural variations associated with metabolic synthesis and its regulation[66]. Since the accumulation of a specific secondary metabolite can be considered a measurable plant phenotype, population genetics approaches traditionally used in morphological and physiological trait analysis can likewise be applied to investigate natural product biosynthesis. Metabolome genome-wide association study (mGWAS) integrates metabolic traits or metabolome data to the genome to identify SNPs in candidate loci via large-scale correlation[67,68].

Analysis of genomes could reveal the evolution of typical genes and gene families involved in natural product biosynthesis as well as the evolution of the whole pathways. A typical example is found in tomatoes: Integration of pan-genome data with mGWAS revealed a loss of flavor chemicals in commercial tomato varieties and pointed out a negative correlation between fruit size and sugar content[11,12,69,70]. By analyzing 980 metabolites from 442 lines, mGWAS identified 3,526 SNPs suggesting that fruit weight might indirectly change metabolite accumulation by gene linkage[11]. Pan-genomes of 100 tomato genomes uncovered 238,490 structural variants, further confirming gene loss in domesticated varieties, and showcasing the impact of genome structure variations on phenotypic traits[69,70]. Other examples include a genomic analysis of 204 selected domesticated and bred watermelons, which led to the identification of 29 candidate genes associated with 20 metabolites. The findings suggest that flavor was enhanced through a decrease in flavonoids and cucurbitacins during the domestication process[71]. Carotenoids are nutritional substances that have become significant selection criteria in carrots, resequencing of 630 carrot accessions revealed selection on carotenoid-associated loci[72]. Resequencing analysis of 390 peanut accessions revealed that genes associated with enhanced oil content have undergone positive selection during the breeding process[73].

GWAS has been introduced to study rice agronomic traits, including the biosynthesis and diversity of rice grain oil[74]. Pan-genome analysis of citrus species uncovered that metabolic gene families involved in flavonoid biosynthesis expanded during evolution, and identified a key gene involved in citric acid biosynthesis[75]. mGWAS identified a gene cluster in the pomelo genome associated with the content of melitidin, a potential anti-cholesterol flavonoid, and successfully uncovered the biosynthetic pathway of melitidin biosynthesis[76]. Population genetics analysis identified a gene cluster in citrus associated with polymethoxyflavone (PMF) accumulation, which originated through tandem duplication events followed by neofunctionalization[77]. From 29 assembled T2T reference genomes and the resequencing data of 466 grapevine cultivars, a variation map with 9,105,787 short variations and 236,449 structural variations (SVs) was obtained. 32 related candidate loci were enriched based on eight metabolic phenotypes providing scientific insight into wine flavor development[78].

Population genomics has also been used for natural products other than edible substances. Genome assembly and resequencing of rubber tree genome accessions proved artificial selection on higher latex production and identified domesticated genes[79]. A multi-omics study revealed that jasmonic acid signaling, triggered by leafhopper infestation, regulates pest resistance in *Nicotiana attenuata*. The analysis further identified a caffeoylputrescine–green leaf volatile conjugate as the key metabolite conferring resistance to leafhoppers[80]. It is anticipated that the exponential increase in plant genomic data will significantly deepen our understanding of plant BGCs.

Transcriptomics and co-expression analysis

Plant natural products usually accumulate in specific tissues. They can also be induced under certain conditions[81]. Genes from the same biosynthetic pathway are often co-regulated, thus reducing energy expenditure and limiting the accumulation of toxic products. Regulation of expression of genes in biosynthetic pathways is essential for optimizing metabolic processes in response to environmental demands. Genes in these pathways are often regulated at the transcription level. Therefore associations between metabolite accumulation and changes in gene expression have been found in selected organs or tissues, or at certain developmental stages, or only after stimulation (e.g. herbivore bait or microbe infection)[82−84]. Whole-cell RNA extraction followed by sequencing (RNA-seq) allows measuring expression levels of all genes in the genome in different plant tissues under different treatments. These data can be used for gene identification. Genes of known function in the pathway can be used as a 'bait' for co-expression analysis for downstream genes, and specific accumulation patterns of metabolites can be used for correlation analysis and gene selection[85,86]. Commonly used methods include Pearson's correlation and network approaches. For instance, weighted gene co-expression network analysis (WGCNA) identified co-expressed genes jointly participating in the biosynthetic pathway[85]. Methods to generate and analyze transcriptomics data keep being updated and optimized, still biosynthetic gene identification strategies still rely on co-expression and phenotype association analysis[87] as the following examples related to medicinal plants illustrate.

Many key genes in natural product biosynthetic pathways have been identified by co-expression analysis combined with genomic analysis. Analysis of differentially expressed genes (DEG) analysis is frequently used when RNA-seq data is available. For tissue-specific plant natural products, DEG analysis is effective in discovering candidate genes as it can identify genes with high expression levels in the accumulating tissue. This analysis could be expanded to investigate differences between plant species. The bioactive components ginsenosides (triterpene saponins) from traditional Chinese herbal medicine *Panax notoginseng* mainly accumulate in the root and rhizome, and five UGTs involved in the ginsenoside biosynthetic pathway were identified by tissue-specific transcriptomic analysis[88]. Similarly, the identification of the genes associated to the biosynthesis of the anti-obesity agent celastrol partly depended on the differential expression of *CYP450*s in different tissues of *Tripterygium wilfordii*[89]. Differences in the expression of *CYP450* in plant tissue was also key to identify genes related to the biosynthesis of painkiller mitragynine in *Mitragyna speciosa*[90]. Similarly, differential expression of α-carbonic anhydrases (CAHs) contributed to the discovery of their novel functional roles in the biosynthesis of Lycopodium alkaloids[91]. The identification of strychnine biosynthetic genes was achieved by considering both tissue-specific gene expression and differences in the gene expression levels between plant species producing strychnine or not[92].

Identification of biosynthetic genes usually comprehensively considers the correlation between gene and 'bait' gene expression, and the correlation between gene expression and natural product accumulation. Three key CYP450 genes for baccatin III, an intermediate in paclitaxel biosynthesis, were revealed by analysing co-expression patterns and associations between pathway intermediates and gene expression[54,93,94]. Biosynthetic genes for the anti-cancer natural product camptothecin were identified by analyzing published transcriptomics data using as a reference a newly assembled genome All camptothecin biosynthesis-related genes were located in the same module of the gene co-expression network reconstructed using WGCNA[95]. In our previous study, we identified

two dehydrogenases, two CYP450s, and one Nudix phosphatase required in the biosynthesis of pyrethrins, the natural pesticide derived from the Pyrethrum plant. These genes were identified by analyzing transcriptomics data obtained at five developmental stages of the flower as well as vegetative tissues leaf, stem, and root; and the key strategy to identify these candidates was investigating their co-expression with the gene *TcGLIP*[96−99]. Similarly, another investigation led to the discovery of 22 enzymes for limonoid biosynthesis[100], indicating the effectiveness of co-expression analysis based on transcriptomics.

Research in biosynthetic pathways often combines metabolome and gene expression data. A Lycopodium alkaloid biosynthetic regulon was detected by paired transcriptomic and metabolomic analyses[101]. The *O*-methyltransferase (OMT) and CYP450s for colchicine alkaloid biosynthesis, and the CYP450s for triptolide biosynthesis were also discovered by combining these data types[102,103]. Identification of BAHD acyltransferases for 3β-tigloyloxytropane, an important intermediate in calystegine biosynthesis, employed tissue-specific DEG analysis, metabolite and gene expression association analysis, phylogenetic and subcellular localization[104]. QS saponin from *Quillaja saponaria* relied on genome mining to find out the required biosynthetic enzymes but also implemented RNA-seq and co-expression analysis for assistance[105]. Similar approaches were later used for the elucidation of saponariosides B biosynthesis in S*aponaria officinalis* by the same research group[50]. Although transcriptomic analysis sometimes only plays a small part in the evolutionary study[106−108], it strongly supports the elucidation of natural product biosynthetic pathways[24,109−111].

Innovation and combination of multi-omics strategies

Conventional omics approaches are limited, therefore they are often combined with other approaches or with upgrades in technology. Identification of metabolic signals usually relies on the abundance of the metabolites, the availability of standards, and on databases for metabolite identification. Plant metabolites are highly diverse, but often their abundance is very low. Mass spectrum techniques facilitate obtaining metabolic information from plant tissues or at developmental stages. Mass imaging visualizes the distribution of metabolites on plant tissue to direct the best strategy choice for synthetic pathway elucidation[112,113]. For example, *in situ* MS imaging of the stem of *Isodon rubescens* showed that ordonin, an *ent*-kaurene-type diterpenoid, was synthesized in apices, but gradually diluted in young leaves, helping pathway elucidation in subsequent research[59]. In another example, MALDI-MSI imaging of the cross-section of horse chestnut pericarp showed the specific accumulation of Escin (barrigenol-type triterpenoid saponins) in cotyledon[40].

Despite the tissue-specific accumulation, biosynthesis of some natural products is highly specific to a certain small range of cells such as trichomes or root hairs. Conventional tissue specific RNA-seq may normalize and weaken the expression signals of target genes. One solution is subdividing the samples with more criteria. Research on oridonin biosynthesis introduced growth as a parameter: beyond normal tissue samples like roots, stems, flowers, buds, and leaves, researchers left half of the leaves opposite to the sampled leaves and took samples of them 14 d later[59]. The results suggested shoot apex as the actual location for oridonin biosynthesis. Amaryllidaceae alkaloids (AmAs) accumulate in most parts of the Amaryllidoideae plant. To find out the particular part producing AmAs, the long leaves of daffodils were cut into short sections for sampling[114].

Single-cell sequencing is a novel and effective tool to recognize

gene expression and find their coexistence among cell types[115–117]. Research on cotton applying this technique confirmed the glandular trichome-specific expression of genes related to gossypol-type terpenoids and volatile terpenoid biosynthesis and identified two novel transcription factors for terpenoid biosynthesis in secretory glandular cells[118]. A 'hyper' cell type was identified in St. John's wort by single-cell RNA sequencing, where hyperforin biosynthesis *de novo* takes place, and four prenyltransferases were identified for the complete pathway by gene coexpression among single cells[119].

The above two solutions, sample selection, and single-cell sequencing, may be combined. Recent work on paclitaxel biosynthesis developed a combinatorial method named multiplexed perturbation x single nuclei (mpXsn) to combine both approaches. In mpXsn researchers prepared a series of samples treated with different conditions for increasing paclitaxel biosynthesis, and pooled all samples for single-cell RNA sequencing[120]. This innovative method discovered eight new genes, proving a promising outlook of both omics strategies and research on natural product biosynthesis.

Combination of various approaches complements the limitation of single methods. Multi-omics for tomato revealed new metabolic genes and pathways by combining metabolome with population genome and transcriptome for mGWAS, expression quantitative trait locus (eQTL), and correlation relationship analysis[11]. This research first collected 610 tomato accessions and generated genomic, metabolomic and transcriptomic datasets; the multi-omic datasets provides, general information on SNPs, metabolite composition and DEGs; the correlation in these datasets was investigated by mGWAS and eQTL, and a multi-omic network was established using the results; focusing on the domesticated traits of tomato, the researchers finally discovered candidate genes involved in tomato quality domestication. Combinations of multiple methods have also advanced research stagnated for a long time. As an example, research on the biosynthesis of the natural hallucinogen mescaline had been stuck for decades, but a complex method combining genomics, transcriptomics, metabolomics, and molecular modeling finally elucidated the biosynthetic pathway in peyote[121]. In this work, researchers first analyzed the composition of peyote metabolites and the specific part of mescalin accumulation by LC-MS/MS and MALDI-MSI to infer the possible intermediates and pathway; then a genome was assembled and annotated based on pair-end transcriptomes; depending on the predicted pathway, enzymes belonging to CYP450, methyltransferase (MT), L-tyrosine/L-DOPA decarboxylase (TyDC) and polyphenol oxidase (PPO) families were selected as candidates; after function characterization, one TyDC, one CYP450, and two MTs were identified, and the pathway was reconstituted in yeast and tobacco leaves. The above cases indicate the importance and effectiveness of combinatorial multi-omic methods.

Machine learning and AI-assisted exploration

Recent advances have driven the application of machine learning and artificial intelligence (AI) approaches in almost every scientific area. In agriculture, the combination of machine learning and various cameras and sensors has led to a new discipline called plant phenomics[122]. Advanced cameras such as thermal infrared camera and 3D time-of-flight cameras, and diverse sensors including chlorophyll fluorescence sensors, laser distance sensors, and RGB sensors have been used to collect crop phenotypic data including plant height, leaf area index, leaf color, tiller density, grain yield, moisture content, and pathogen infection[122]. Utilization of machine learning algorithms in image processing has enabled analysis of such large data. For example, deep learning techniques such as Convolutional Neural Networks (CNN)

allowed quick detection of rice disease in the early stages from pictures; Support Vector Machine (SVM) and Gaussian Processes Classifier (GPC) were used to detect moisture deficit by analyzing thermal images of canopies; finally, analysis of light detection and ranging (LiDAR) technology could predict canopy geometry and yield of apple tree[122–125]. These applications required large amounts of data to build models of concerned features.

Since omics strategies have been widely applied in biological research, gene prediction, and identification based on machine learning are introduced into natural product biosynthetic research. Optimized statistics and machine learning have been used to analyze RNA-seq data and predict gene function[87,126]. The integration of large-scale omics datasets, including genomic, transcriptomic, and metabolomic information, has significantly advanced our understanding of plant-specialized metabolism. These datasets provide detailed insights into gene sequences, structural features, genomic loci, BGCs, phylogenetic relationships, and gene regulatory networks. Importantly, computational modeling and machine learning are emerging as powerful tools to synthesize these complex datasets, enabling data-driven predictions and hypothesis generation. As multi-omics resources continue to expand, their synergistic application holds great promise for elucidating biosynthetic pathways and guiding metabolic engineering efforts in plants[127]. For instance, a machine learning–based genome scan was analyzed to predict introgression regions for grape populations[128]. In crop breeding, machine learning has been used to build genomic prediction model for parental selection. These models offered a more comprehensive and deeper analysis of the complex interactions among vast datasets, and their prediction could be adjusted according to the target of breeding by assigning weights to different traits[122,129].

AI also holds significant potential for identifying genes involved in biosynthetic pathways. The introduction of the AlphaFold model marked a breakthrough in structural biology by enabling highly accurate prediction of protein structures. Subsequent models, such as RoseTTAFoldNA and AlphaFold 3, have further expanded these capabilities to include the prediction of protein–ligand interactions, encompassing metal ions, nucleic acids, and post-translationally modified residues[130–132]. In combination with molecular docking that predicts interaction patterns between enzyme and substrate molecules, these AI-powered models help in the functional verification of enzymes and contributed to downstream research such as protein design and virtual screening. A recent example was the identification of a UGT involved in salidroside biosynthesis by predicting protein structure with RoseTTAFold and virtual screening with AutoDock Vina, a tool for molecular docking[133]. Large language models also help develop more advanced BGC detection algorithms, such as DeepBGC[134], and self-supervised training BiGCARP[135]. These algorithms have achieved promising results for BGCs in microbial genomes by leveraging large-scale pre-trained language models to embed Pfam domains and training with masked language model objectives, and therefore effectively capturing higher-order dependencies within sequences.

The application of AI methods has also significantly enhanced the efficiency and accuracy of data analysis in response to the rapid expansion of multi-omics technologies. For instance, the multi-omics analysis platform Compounds And Transcripts Bridge (CAT Bridge) integrates various statistical approaches alongside an AI agent to improve the interpretation of transcriptome–metabolome associations. This platform has demonstrated strong performance, particularly in the analysis of longitudinal omics datasets[136]. For more complex datasets like single-cell transcriptome data,

generative AI models such as scGPT showed good performance in cell type annotation, perturbation prediction, multi-batch data integration, and gene network inference[137]. The modelling capacity was also used for optimizing natural product metabolism for higher yield. The integration of design of experiment (DoE) strategies with machine learning approaches has been successfully applied to optimize metabolic pathways, exemplified by the enhanced production of p-coumaric acid in engineered yeast strains[138–140]. The deep learning–based model BioNavi-NP has demonstrated notable potential in predicting plausible biosynthetic pathways directly from the chemical structures of natural products. Through extensive training on known biosynthetic reactions, BioNavi-NP can infer enzyme-catalyzed transformations and pathway logic, providing valuable insights for pathway elucidation and metabolic engineering. Its application not only accelerates the discovery of unknown biosynthetic routes but also supports the rational design of synthetic biology strategies[141].

Gene function validation

Finding candidate genes and gene clusters represents only the initial step in elucidating biosynthetic pathways. The next essential task is to validate the functions of the candidate genes. There are various methods developed to test gene function, and could be primarily divided into *in vitro* and *in vivo*. *In vitro* experiments give direct results by eliminating the interference of the internal environment of plants, while *in vivo* experiments can show the real situation about how candidate gene function in plants.

The most straightforward method for *in vitro* validation involves incubating purified enzymes with their respective substrates in a buffered reaction system. Target enzymes are usually obtained through protein expression in bacteria cells followed by tag-based purification. *E. coli* is the most common chassis for *in vitro* protein expression. In mescaline biosynthesis, for example, TyDC candidates and MT candidates were heterologously expressed in *E. coli* and the lysate was used for *in vitro* enzyme assays[121]. The 3*β*-tigloyloxytropane synthase (TS) from *Atropa belladonna* was expressed in *E. coli* and purified via MBP-tag[104]. Similarly, ApCPS2 and ApGGPPS from *Andrographis paniculata* were characterized by an *in vitro* assay using *E. coli* expressed and purified protein[142,143]. However, there are enzymes like membrane proteins that usually require eukaryotic organelles for proper translation, folding, modification, and functioning, making them difficult to express in prokaryotes[144]. Membrane proteins like CYP450s were usually insoluble when directly expressed in *E. coli* and therefore required extra modification[144]. Hence for these enzymes, eukaryotic chassis were introduced. Yeast, especially *Saccharomyces cerevisiae*, is commonly used in enzyme expression as it provides a sufficient environment including endoplasmic reticulum (ER) for membrane enzyme expression[144]. For secreted protein, inducible expression, and purification protocols are generally comparable to those used in prokaryotics. For membrane enzymes, a better alternative might be microsomes as it maintains the conformation of enzymes. Several studies, such as flavonoid diversification in *Scutellaria*, aporphine alkaloids biosynthesis, colchicine alkaloid, and huperzine A biosynthesis, have employed microsome for *in vitro* enzyme assay to characterize functions of CYP450s[91,102,145,146].

Exogenous expression is another method to validate gene function. Substrates could be directly supplied to the engineered microbial strains or produced by the chassis. The CYP450s involved in aescin and aesculin biosynthesis were characterized in engineered yeast[40], so as the CYP450s catalyzing oxidative rearrangement for xanthanolide[147]. In the study of tripterifordin and neotripterifordin

biosynthesis, yeast strains were engineered to characterize CYP450 monooxygenase C20ox[148]. The model plant, tobacco *Nicotiana benthamiana* is also a popular chassis for enzyme function validation, especially in plant natural product biosynthesis, as it provides a more suitable environment for plant gene expression. Characterization of biosynthetic enzymes via transient expression in *N. benthamiana* included CYP725A subfamily enzymes for baccatin III[54,94], a gene set involved in paclitaxel biosynthesis[93], UGTs for saponins in soapbark tree and soapwort[50,105], complete pathway genes for pyrethric acid[97], a CYP450 gene cluster involved in ginkgolide biosynthesis[56]. Other model plants, such as *A. thaliana*[149], *Oryza sativa* (rice)[150], *Solanum lycopersicum* (tomato), and *Cucumis sativus* (cucumber)[151] were also used as chassis in natural product biosynthesis studies. In addition, some studies employed insect cells for exogenous expression in gene function validation[54]. These methods were often combined in the same study to eliminate the probable side effects of endogenous metabolism of chassis[54,56,91,94,102,121,148].

Compared to exogenous expression, *in vivo* methods typically validate gene function in the original plants. Gene knock-outs and overexpression are widely used strategies to verify gene functions. Popular approaches include transgenic technology and genome editing that create plant lines with stable inheritance.

However, stable genetic transformation systems are underdeveloped or unavailable for many plant species, particularly medicinal plants. Consequently, transient gene expression is an effective alternative method for functional gene analysis and pathway reconstruction. Hairy roots are abnormal growth induced by *Agrobacterium rhizogenes* infection but are similar to normal roots in anatomy and metabolism[152]. *A. rhizogenes* can harbor target genes inserted in its T-DNA region that can then be integrated into the plant genome. This allows gene function validation with the hairy root system[152,153]. The hairy root transformation system has been established in many medicinal plants to enhance natural product biosynthesis and verify gene function, the promotion of saikosaponins in *Bupleurum chinense* with overexpression of *BcERF3* in hairy root as an example[154,155]. Crops also employed hairy root systems for rapid and efficient transformation. A reported hairy root system for soybean required only 16 d for the whole workflow of transformation, and allowed genome editing for gene function characterization[153].

Virus-induced gene silencing (VIGS) is another *in vivo* validation method utilizing plant defense mechanisms against virus infection. This technique suppresses the expression of the target gene in plants by infecting plants with a virus conveying fragment of the target gene and inducing post-transcriptional gene silencing, and the system has been applied in tobacco and tomato[156]. As a transient transformation method, VIGS avoids the obstacle in establishing stable transformation system, showing a significant advantage in research on medicinal plants. VIGS at the cotyledon stage of *Catharanthus roseus* showed successful silencing of several transcription factors regulating terpenoid indole alkaloid (TIA) biosynthesis and resulted in a decrease of TIA intermediates[157]. Another study on *C. roseus* used VIGS to inhibit *CrZIP* transcription in function characterization[158]. Suppression of *ApCPS2* via VIGS drastically decreased the content of andrographolide and weakened the defense of *A. paniculata* against herbivores[159]. In *Ginkgo biloba* and *Lycoris chinensis*, VIGS application was also reported[160,161].

Conclusions and perspectives

Whatever strategies are used in plant natural product biosynthesis research, the final goal is the industrial production of natural products.

Compared to plants, microbes showed multiple advantages for industrial production, such as rapid growth, a smaller genome, plenty of engineering tools, and a mature fermentation industry. A few natural products have been produced by engineered microbial strains and the yield achieved the potential for industrial production. The plant-derived antimalarial drug, artemisinin, was produced by hemi-biosynthesis[162]. An engineered yeast (*S. cerevisiae*) strain produced the precursor artemisinic acid at an optimized level of 25 g/L, and artemisinic acid was chemically catalyzed to artemisinin[15,162]. Similarly, vindoline and catharanthine were produced by engineered yeast (*S. cerevisiae*) and the anti-cancer phytochemical vinblastine was synthesized by *in vitro* chemical coupling[163]. Another study using yeast *Pichia pastoris* to produce catharanthine reached a titre 2.57 mg/L, indicating that *P. pastoris* was a potential alternative to microbial chassis[164]. *Yarrowia lipolytica* is another yeast species. Yields of engineered *Yarrowia lipolytica* strains for gastrodin production achieved over 13 g/L[165,166], and resveratrol production reached 22.5 g/L[167]. These outcomes suggested the promising potential for producing plant natural products by microbes.

Along with the rapid development of omics techniques, a vast amount of research applied *de novo* genome assembly, transcriptome co-expression, and population genomics (mGWAS and pan-genome) to investigate the biosynthesis of natural products. New strategies including single-cell RNA sequencing, mass imaging, and machine learning provide a deeper and more comprehensive understanding of natural product biosynthesis, whether in plants or microbes, and have started to play important roles. These novel techniques effectively complement classic techniques and are expected to further support the discovery of unknown biosynthetic pathways, such as the whole pathway of paclitaxel biosynthesis which has puzzled scientists for a long time, or uncover more biosynthetic pathways of lesser-studied plant natural products. For natural products that have been successfully synthesized de novo in microbial chassis but still suffer from limited yields, such as tropane alkaloids[168] and vaccine adjuvant QS-21[169], the technological innovation is expected to accelerate optimization. Ultimately, the rapid development of multi-omic strategies continues to deepen our understanding of plant natural product biosynthesis and paves the way for the efficient industrial production of plant-derived therapeutics.

Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Li W; draft manuscript preparation: Wan S, Li W, Schaap PJ, Suarez-Diez M. All authors reviewed the results and approved the final version of the manuscript.

Data availability

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

Acknowledgments

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Chen SL, Yu H, Luo HM, Wu Q, Li CF, et al. 2016. Conservation and sustainable use of medicinal plants: problems, progress, and prospects. *Chinese Medicine* 11(1):37
2. Gil-Martín E, Forbes-Hernández T, Romero A, Cianciosi D, Giampieri F, et al. 2022. Influence of the extraction method on the recovery of bioactive phenolic compounds from food industry by-products. *Food Chemistry* 378:131918
3. Rodriguez A, Strucko T, Stahlhut SG, Kristensen M, Svenssen DK, et al. 2017. Metabolic engineering of yeast for fermentative production of flavonoids. *Bioresource Technology* 245:1645−54
4. Zhou X, Liu Z. 2022. Unlocking plant metabolic diversity: A (pan)-genomic view. *Plant Communications* 3(2):100300
5. Jacobowitz JR, Weng JK. 2020. Exploring uncharted territories of plant specialized metabolism in the postgenomic era. *Annual Review of Plant Biology* 71(1):631−58
6. Pichersky E, Raguso RA. 2018. Why do plants produce so many terpenoid compounds? *New Phytologist* 220(3):692−702
7. Lybrand DB, Xu H, Last RL, Pichersky E. 2020. How Plants Synthesize Pyrethrins: Safe and Biodegradable Insecticides. *Trends in Plant Science* 25(12):1240−51
8. Frey M, Schullehner K, Dick R, Fiesselmann A, Gierl A. 2009. Benzoxazinoid biosynthesis, a model for evolution of secondary metabolic pathways in plants. *Phytochemistry* 70(15):1645−51
9. Gutzeit HO, Ludwig-Mueller J. 2014. Function of natural substances in plants. *Plant Natural Products: Synthesis, Biological Functions and Practical Applications*. Weinheim, Germany: Wiley-VCH.
10. Chakraborty A, Chaudhury R, Dutta S, Basak M, Dey S, et al. 2022. Role of metabolites in flower development and discovery of compounds controlling flowering time. *Plant Physiology and Biochemistry* 190:109−18
11. Zhu G, Wang S, Huang Z, Zhang S, Liao Q, et al. 2018. Rewiring of the fruit metabolome in tomato breeding. *Cell* 172(1-2):249−261.e12
12. Tieman D, Zhu G, Resende MFR Jr, Lin T, Nguyen C, et al. 2017. A chemical genetic roadmap to improved tomato flavor. *Science* 355:391−94
13. Forss DA, Dunstone EA, Ramshaw EH, Stark W. 1962. The flavor of cucumbers. *Journal of Food Science* 27(1):90−93
14. Jeon JE, Kim JG, Fischer CR, Mehta N, Dufour-Schroif C, et al. 2020. A pathogen-responsive gene cluster for highly modified fatty acids in tomato. *Cell* 180(1):176−187.e19
15. Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, et al. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440:940−43
16. Smith AB, Chekan JR. 2023. Engineering yeast for industrial-level production of the antimalarial drug artemisinin. *Trends in Biotechnology* 41(3):267−69
17. Weaver BA. 2014. How Taxol/paclitaxel kills cancer cells. *Molecular Biology of the Cell* 25(18):2677−81
18. Kohnen-Johannsen KL, Kayser O. 2019. Tropane Alkaloids: Chemistry, Pharmacology, Biosynthesis and Production. *Molecules* 24(4):796
19. Jahan T, Huda MdN, Zhang K, He Y, Lai D, et al. 2025. Plant secondary metabolites against biotic stresses for sustainable crop protection. *Biotechnology Advances* 79:108520
20. Li Q, Duncan S, Li Y, Huang S, Luo M. 2024. Decoding plant specialized metabolism: new mechanistic insights. *Trends in Plant Science* 29(5):535−45
21. Zhang T, Zhang C, Wang W, Hu S, Tian Q, et al. 2025. Effects of drought stress on the secondary metabolism of *Scutellaria baicalensis* Georgi and the function of *SbWRKY34* in drought resistance. *Plant Physiology and Biochemistry* 219:109362
22. Zhang F, Huang J, Guo H, Yang C, Li Y, et al. 2022. OsRLCK160 contributes to flavonoid accumulation and UV-B tolerance by regulating OsbZIP48 in rice. *Science China Life Sciences* 65(7):1380−94

23. Castillon A, Shen H, Huq E. 2007. Phytochrome Interacting Factors: central players in phytochrome-mediated light signaling networks. *Trends in Plant Science* 12(11):514−21

24. Zhang Z, Zhang X, Chen Y, Jiang W, Zhang J, et al. 2023. Understanding the mechanism of red light-induced melatonin biosynthesis facilitates the engineering of melatonin-enriched tomatoes. *Nature Communications* 14:5525

25. Li Y, Chen Y, Zhou L, You S, Deng H, et al. 2020. MicroTom metabolic network: rewiring tomato metabolic regulatory network throughout the growth cycle. *Molecular Plant* 13(8):1203−18

26. Shang Y, Ma Y, Zhou Y, Zhang H, Duan L, et al. 2014. Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* 346:1084−88

27. Zhang Z, Liang C, Ren Y, Lv Z, Huang J. 2024. Interaction of ubiquitin-like protein SILENCING DEFECTIVE 2 with LIKE HETEROCHROMATIN PROTEIN 1 is required for regulation of anthocyanin biosynthesis in *Arabidopsis thaliana* in response to sucrose. *New Phytologist* 243(4):1374−86

28. Chen W, Wang X, Sun J, Wang X, Zhu Z, et al. 2024. Two telomere-to-telomere gapless genomes reveal insights into *Capsicum* evolution and capsaicinoid biosynthesis. *Nature Communications* 15(1):4295

29. Weng JK, Philippe RN, Noel JP. 2012. The rise of chemodiversity in plants. *Science* 336:1667−70

30. Hansen CC, Nelson DR, Møller BL, Werck-Reichhart D. 2021. Plant cytochrome P450 plasticity and evolution. *Molecular Plant* 14(8):1244−65

31. Zhang P, Zhang Z, Zhang L, Wang J, Wu C. 2020. Glycosyltransferase GT1 family: Phylogenetic distribution, substrates coverage, and representative structural features. *Computational and Structural Biotechnology Journal* 18:1383−90

32. Kong W, Wang Y, Zhang S, Yu J, Zhang X. 2023. Recent advances in assembly of complex plant genomes. *Genomics, Proteomics & Bioinformatics* 21(3):427−39

33. Su W, Jing Y, Lin S, Yue Z, Yang X, et al. 2021. Polyploidy underlies co-option and diversification of biosynthetic triterpene pathways in the apple tribe. *Proceedings of the National Academy of Sciences of the United States of America* 118(20):e2101767118

34. Han X, Zhang J, Han S, Chong SL, Meng G, et al. 2022. The chromosome-scale genome of *Phoebe bournei* reveals contrasting fates of terpene synthase (TPS)-a and TPS-b subfamilies. *Plant Communications* 3(6):100410

35. Yang J, Wu Y, Zhang P, Ma J, Yao YJ, et al. 2023. Multiple independent losses of the biosynthetic pathway for two tropane alkaloids in the Solanaceae family. *Nature Communications* 14(1):8457

36. Li P, Yan MX, Liu P, Yang DJ, He ZK, et al. 2024. Multiomics analyses of two *Leonurus* species illuminate leonurine biosynthesis and its evolution. *Molecular Plant* 17(1):158−77

37. Qin L, Hu Y, Wang J, Wang X, Zhao R, et al. 2021. Insights into angiosperm evolution, floral development and chemical biosynthesis from the *Aristolochia fimbriata* genome. *Nature Plants* 7(9):1239−53

38. Zhan C, Shen S, Yang C, Liu Z, Fernie AR, et al. 2022. Plant metabolic gene clusters in the multi-omics era. *Trends in Plant Science* 27(10):981−1001

39. Yang C, Shen S, Zhan C, Li Y, Zhang R, et al. 2024. Variation in a Poaceae-conserved fatty acid metabolic gene cluster controls rice yield by regulating male fertility. *Nature Communications* 15(1):6663

40. Sun W, Yin Q, Wan H, Gao R, Xiong C, et al. 2023. Characterization of the horse chestnut genome reveals the evolution of aescin and aesculin biosynthesis. *Nature Communications* 14(1):6470

41. Polturak G, Dippe M, Stephenson MJ, Chandra Misra R, Owen C, et al. 2022. Pathogen-induced biosynthetic pathways encode defense-related molecules in bread wheat. *Proceedings of the National Academy of Sciences of the United States of Americ* 119(16):e2123299119

42. Field B, Osbourn AE. 2008. Metabolic diversification—independent assembly of operon-like gene clusters in different plants. *Science* 320:543−47

43. Fan P, Wang P, Lou YR, Leong BJ, Moore BM, et al. 2020. Evolution of a plant gene cluster in Solanaceae and emergence of metabolic diversity. *eLife* 9:e56717

44. Liu Z, Cheema J, Vigouroux M, Hill L, Reed J, et al. 2020. Formation and diversification of a paradigm biosynthetic gene cluster in plants. *Nature Communications* 11:5354

45. Mao L, Kawaide H, Higuchi T, Chen M, Miyamoto K, et al. 2020. Genomic evidence for convergent evolution of gene clusters for momilactone biosynthesis in land plants. *PProceedings of the National Academy of Sciences of the United States of America* 117(22):12472−80

46. Berman P, De Haro LA, Jozwiak A, Panda S, Pinkas Z, et al. 2023. Parallel evolution of cannabinoid biosynthesis. *Nature Plants* 9(5):817−31

47. Frey M, Chomet P, Glawischnig E, Stettner C, Grün S, et al. 1997. Analysis of a chemical plant defense mechanism in grasses. *Science* 277:696−99

48. Von Rad U, Hüttl R, Lottspeich F, Gierl A, Frey M. 2001. Two glucosyltransferases are involved in detoxification of benzoxazinoids in maize. *The Plant Journal* 28(6):633−42

49. Winzer T, Gazda V, He Z, Kaminski F, Kern M, et al. 2012. A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Science* 336(6089):1704−8

50. Jo S, El-Demerdash A, Owen C, Srivastava V, Wu D, et al. 2024. Unlocking saponin biosynthesis in soapwort. *Nature Chemical Biology* 21:215−26

51. Liu Y, Wang B, Shu S, Li Z, Song C, et al. 2021. Analysis of the *Coptis chinensis* genome reveals the diversification of protoberberine-type alkaloids. *Nature Communications* 12:3276

52. Kautsar SA, Suarez Duran HG, Blin K, Osbourn A, Medema MH. 2017. plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Research* 45(W1):W55−W63

53. Xiong X, Gou J, Liao Q, Li Y, Zhou Q, et al. 2021. The *Taxus* genome provides insights into paclitaxel biosynthesis. *Nature Plants* 7(8):1026−36

54. Jiang B, Gao L, Wang H, Sun Y, Zhang X, et al. 2024. Characterization and heterologous reconstitution of *Taxus* biosynthetic enzymes leading to baccatin III. *Science* 383:622−29

55. Liu C, Smit SJ, Dang J, Zhou P, Godden GT, et al. 2023. A chromosome-level genome assembly reveals that a bipartite gene cluster formed via an inverted duplication controls monoterpenoid biosynthesis in *Schizonepeta tenuifolia*. *Molecular Plant* 16(3):533−48

56. Forman V, Luo D, Geu-Flores F, Lemcke R, Nelson DR, et al. 2022. A gene cluster in *Ginkgo biloba* encodes unique multifunctional cytochrome P450s that initiate ginkgolide biosynthesis. *Nature Communications* 13(1):5143

57. Wu S, Malaco Morotti AL, Wang S, Wang Y, Xu X, et al. 2022. Convergent gene clusters underpin hyperforin biosynthesis in St John's wort. *New Phytologist* 235(2):646−61

58. Li Y, Leveau A, Zhao Q, Feng Q, Lu H, et al. 2021. Subtelomeric assembly of a multi-gene pathway for antimicrobial defense compounds in cereals. *Nature Communications* 12(1):2563

59. Sun Y, Shao J, Liu H, Wang H, Wang G, et al. 2023. A chromosome-level genome assembly reveals that tandem-duplicated CYP706V oxidase genes control oridonin biosynthesis in the shoot apex of *Isodon rubescens*. *Molecular Plant* 16(3):517−32

60. Zhang Y, Gao J, Ma L, Tu L, Hu T, et al. 2023. Tandemly duplicated CYP82Ds catalyze 14-hydroxylation in triptolide biosynthesis and precursor production in *Saccharomyces cerevisiae*. *Nature Communications* 14(1):875

61. Boachon B, Burdloff Y, Ruan JX, Rojo R, Junker RR, et al. 2019. A promiscuous CYP706A3 reduces terpene volatile emission from Arabidopsis flowers, affecting florivores and the floral microbiome. *The Plant Cell* 31(12):2947−72

62. Abdollahi F, Alebrahim MT, Ngov C, Lallemand E, Zheng Y, et al. 2021. Innate promiscuity of the CYP706 family of P450 enzymes provides a suitable context for the evolution of dinitroaniline resistance in weed. *New Phytologist* 229(6):3253−68

63. Cankar K, van Houwelingen A, Goedbloed M, Renirie R, de Jong RM, et al. 2014. Valencene oxidase CYP706M1 from Alaska cedar (*Callitropsis nootkatensis*). *FEBS Letters* 588(6):1001−7

64. Luo P, Wang YH, Wang GD, Essenberg M, Chen XY. 2001. Molecular cloning and functional identification of (+)-δ-cadinene-8-hydroxylase,

a cytochrome P450 mono-oxygenase (CYP706B1) of cotton sesquiterpene biosynthesis. *The Plant Journal* 28(1):95–104

65. Hansen CC, Sørensen M, Veiga TAM, Zibrandtsen JFS, Heskes AM, et al. 2018. Reconfigured cyanogenic glucoside biosynthesis in *Eucalyptus cladocalyx* involves a cytochrome P450 CYP706C55. *Plant Physiology* 178(3):1081–95

66. Shi J, Tian Z, Lai J, Huang X. 2023. Plant pan-genomics and its applications. *Molecular Plant* 16(1):168–86

67. Luo J. 2015. Metabolite-based genome-wide association studies in plants. *Current Opinion in Plant Biology* 24:31–38

68. Tibbs Cortes L, Zhang Z, Yu J. 2021. Status and prospects of genome-wide association studies in plants. *The Plant Genome* 14(1):e20077

69. Gao L, Gonda I, Sun H, Ma Q, Bao K, et al. 2019. The tomato pangenome uncovers new genes and a rare allele regulating fruit flavor. *Nature Genetics* 51(6):1044–51

70. Alonge M, Wang X, Benoit M, Soyk S, Pereira L, et al. 2020. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* 182(1):145–161.e23

71. Yuan P, Xu C, He N, Lu X, Zhang X, et al. 2023. Watermelon domestication was shaped by stepwise selection and regulation of the metabolome. *Science China Life Sciences* 66(3):579–94

72. Coe K, Bostan H, Rolling W, Turner-Hissong S, Macko-Podgórni A, et al. 2023. Population genomics identifies genetic signatures of carrot domestication and improvement and uncovers the origin of high-carotenoid orange carrots. *Nature Plants* 9(10):1643–58

73. Lu Q, Huang L, Liu H, Garg V, Gangurde SS, et al. 2024. A genomic variation map provides insights into peanut diversity in China and associations with 28 agronomic traits. *Nature Genetics* 56(3):530–40

74. Zhou H, Xia D, Li P, Ao Y, Xu X, et al. 2021. Genetic architecture and key genes controlling the diversity of oil composition in rice grains. *Molecular Plant* 14(3):456–69

75. Huang Y, He J, Xu Y, Zheng W, Wang S, et al. 2023. Pangenome analysis provides insight into the evolution of the orange subfamily and a key gene for citric acid accumulation in citrus fruits. *Nature Genetics* 55(11):1964–75

76. Shen S, Wang S, Yang C, Wang C, Zhou Q, et al. 2023. Elucidation of the melitidin biosynthesis pathway in pummelo. *Journal of Integrative Plant Biology* 65(11):2505–18

77. Peng Z, Song L, Chen M, Liu Z, Yuan Z, et al. 2024. Neofunctionalization of an *OMT* cluster dominates polymethoxyflavone biosynthesis associated with the domestication of citrus. *Proceedings of the National Academy of Sciences of the United States of America* 121(14):e2321615121

78. Liu Z, Wang N, Su Y, Long Q, Peng Y, et al. 2024. Grapevine pangenome facilitates trait genetics and genomic breeding. *Nature Genetics* 56:2804–14

79. Chao J, Wu S, Shi M, Xu X, Gao Q, et al. 2023. Genomic insight into domestication of rubber tree. *Nature Communications* 14:4651

80. Bai Y, Yang C, Halitschke R, Paetz C, Kessler D, et al. 2022. Natural history–guided omics reveals plant defensive chemistry against leafhopper pests. *Science* 375:eabm2948

81. Huang XQ, Dudareva N. 2023. Plant specialized metabolism. *Current Biology* 33(11):R473–R478

82. Schilmiller AL, Last RL, Pichersky E. 2008. Harnessing plant trichome biochemistry for the production of useful compounds. *The Plant journal* 54:4702–11

83. Li D, Heiling S, Baldwin IT, Gaquerel E. 2016. Illuminating a plant's tissue-specific metabolic diversity using computational metabolomics and information theory. *Proceedings of the National Academy of Sciences of the United States of America* 113(47):E7610–E7618

84. Creelman RA, Mullet JE. 1997. Biosynthesis and action of jasmonates in plants. *Annual Review of Plant Physiology* 48(1):355–81

85. Omranian N, Kleessen S, Tohge T, Klie S, Basler G, et al. 2015. Differential metabolic and coexpression networks of plant metabolism. *Trends in Plant Science* 20(5):266–68

86. Saito K, Matsuda F. 2010. Metabolomics for functional genomics, systems biology, and biotechnology. *Annual Review of Plant Biology* 61(1):463–89

87. Wang P, Moore BM, Uygun S, Lehti-Shiu MD, Barry CS, et al. 2021. Optimising the use of gene expression data to predict plant metabolic pathway memberships. *New Phytologist* 231(1):475–89

88. Jiang Z, Tu L, Yang W, Zhang Y, Hu T, et al. 2021. The chromosome-level reference genome assembly for *Panax notoginseng* and insights into ginsenoside biosynthesis. *Plant Communications* 2(1):100113

89. Zhao Y, Hansen NL, Duan YT, Prasad M, Motawia MS, et al. 2023. Biosynthesis and biotechnological production of the anti-obesity agent celastrol. *Nature Chemistry* 15(9):1236–46

90. Schotte C, Jiang Y, Grzech D, Dang TT, Laforest LC, et al. 2023. Directed biosynthesis of mitragynine stereoisomers. *Journal of the American Chemical Society* 145(9):4957–63

91. Nett RS, Dho Y, Tsai C, Passow D, Martinez Grundman J, Low Y-Y, Sattely ES. 2023. Plant carbonic anhydrase-like enzymes in neuroactive alkaloid biosynthesis. *Nature*

92. Hong B, Grzech D, Caputi L, Sonawane P, López CER, et al. 2022. Biosynthesis of strychnine. *Nature* 607:617–22

93. Zhang Y, Wiese L, Fang H, Alseekh S, Perez de Souza L, et al. 2023. Synthetic biology identifies the minimal gene set required for paclitaxel biosynthesis in a plant chassis. *Molecular Plant* 16(12):1951–61

94. Zhao Y, Liang F, Xie Y, Duan YT, Andeadelli A, et al. 2024. Oxetane ring formation in taxol biosynthesis is catalyzed by a bifunctional cytochrome P450 enzyme. *Journal of the American Chemical Society* 146(1):801–10

95. Kang M, Fu R, Zhang P, Lou S, Yang X, et al. 2021. A chromosome-level *Camptotheca acuminata* genome assembly provides insights into the evolutionary origin of camptothecin biosynthesis. *Nature Communications* 12:3531

96. Li W, Lybrand DB, Zhou F, Last RL, Pichersky E. 2019. Pyrethrin biosynthesis: the cytochrome P450 oxidoreductase CYP82Q3 converts jasmolone to pyrethrolone. *Plant Physiology* 181(3):934–44

97. Xu H, Li W, Schilmiller AL, Van Eekelen H, De Vos RCH, et al. 2019. Pyrethric acid of natural pyrethrin insecticide: complete pathway elucidation and reconstitution in *Nicotiana benthamiana*. *New Phytologist* 223(2):751–65

98. Li W, Zhou F, Pichersky E. 2018. Jasmone hydroxylase, a key enzyme in the synthesis of the alcohol moiety of pyrethrin insecticides. *Plant Physiology* 177(4):1498–509

99. Li W, Lybrand DB, Xu H, Zhou F, Last RL, et al. 2020. A trichome-specific, plastid-localized *Tanacetum cinerariifolium* nudix protein hydrolyzes the natural pyrethrin pesticide biosynthetic intermediate *trans*-chrysanthemyl diphosphate. *Frontiers in Plant Science* 11:482

100. De La Peña R, Hodgson H, Liu JC-T, Stephenson MJ, Martin AC, Owen C, Harkess A, Leebens-Mack J, Jimenez LE, Osbourn A, Sattely ES. 2023. Complex scaffold remodeling in plant triterpene biosynthesis. *Science* 379:361–68

101. Nett RS, Dho Y, Low YY, Sattely ES. 2021. A metabolic regulon reveals early and late acting enzymes in neuroactive Lycopodium alkaloid biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* 118(24):e2102949118

102. Nett RS, Lau W, Sattely ES. 2020. Discovery and engineering of colchicine alkaloid biosynthesis. *Nature* 584:148–53

103. Tu L, Su P, Zhang Z, Gao L, Wang J, et al. 2020. Genome of *Tripterygium wilfordii* and identification of cytochrome P450 involved in triptolide biosynthesis. *Nature Communications* 11:971

104. Zeng J, Liu X, Dong Z, Zhang F, Qiu F, et al. 2024. Discovering a mitochondrion-localized BAHD acyltransferase involved in calystegine biosynthesis and engineering the production of 3β-tigloyloxytropane. *Nature Communications* 15:3623

105. Reed J, Orme A, El-Demerdash A, Owen C, Martin LBB, et al. 2023. Elucidation of the pathway for biosynthesis of saponin adjuvants from the soapbark tree. *Science* 379:1252–64

106. Hu J, Qiu S, Wang F, Li Q, Xiang CL, et al. 2023. Functional divergence of CYP76AKs shapes the chemodiversity of abietane-type diterpenoids in genus *Salvia*. *Nature Communications* 14:4696

107. Florean M, Luck K, Hong B, Nakamura Y, O'Connor SE, et al. 2023. Reinventing metabolic pathways: independent evolution of benzoxazinoids in flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* 120(42):e2307981120

108. Wang HT, Wang ZL, Chen K, Yao MJ, Zhang M, et al. 2023. Insights into the missing apiosylation step in flavonoid apiosides biosynthesis of Leguminosae plants. *Nature Communications* 14:6658

109. Chavez BG, Srinivasan P, Glockzin K, Kim N, Montero Estrada O, et al. 2022. Elucidation of tropane alkaloid biosynthesis in *Erythroxylum coca* using a microbial pathway discovery platform. *Proceedings of the National Academy of Sciences* 119(49):e2215372119

110. Deng X, Ye Z, Duan J, Chen F, Zhi Y, et al. 2024. Complete pathway elucidation and heterologous reconstitution of (+)-nootkatone biosynthesis from *Alpinia oxyphylla*. *New Phytologist* 241:779−92

111. Edwards A, Njaci I, Sarkar A, Jiang Z, Kaithakottil GG, et al. 2023. Genomics and biochemical analyses reveal a metabolon key to β-L-ODAP biosynthesis in Lathyrus sativus. *Nature Communications* 14(1):876

112. Bhandari DR, Wang Q, Friedt W, Spengler B, Gottwald S, et al. 2015. High resolution mass spectrometry imaging of plant tissues: towards a plant metabolite atlas. *Analyst* 140(22):7696−709

113. Horn PJ, Chapman KD. 2024. Imaging plant metabolism *in situ*. *Journal of Experimental Botany* 75(6):1654−70

114. Mehta N, Meng Y, Zare R, Kamenetsky-Goldstein R, Sattely E. 2024. A developmental gradient reveals biosynthetic pathways to eukaryotic toxins in monocot geophytes. *Cell* 187(20):5620−5637.e10

115. Liu Z, Zhou Y, Guo J, Li J, Tian Z, et al. 2020. Global dynamic molecular profiling of stomatal lineage cell development by single-cell RNA sequencing. *Molecular Plant* 13(8):1178−93

116. Lopez-Anido CB, Vatén A, Smoot NK, Sharma N, Guo V, et al. 2021. Single-cell resolution of lineage trajectories in the Arabidopsis stomatal lineage and developing leaf. *Developmental Cell* 56(7):1043−1055.e4

117. Dai Y, Zhang S, Guan J, Wang S, Zhang H, et al. 2024. Single-cell transcriptomic analysis of flowering regulation and vernalization in Chinese cabbage shoot apex. *Horticulture Research* 11:uhae214

118. Lin JL, Chen L, Wu WK, Guo XX, Yu CH, et al. 2023. Single-cell RNA sequencing reveals a hierarchical transcriptional regulatory network of terpenoid biosynthesis in cotton secretory glandular cells. *Molecular Plant* 16(12):1990−2003

119. Wu S, Morotti ALM, Yang J, Wang E, Tatsis EC. 2024. Single-cell RNA sequencing facilitates the elucidation of the complete biosynthesis of the antidepressant hyperforin in St. John's wort. *Molecular Plant* 17:1439−57

120. McClune CJ, Liu JCT, Wick C, De La Peña R, Lange BM, et al. 2024. Multiplexed perturbation of yew reveals cryptic proteins that enable a total biosynthesis of baccatin III and Taxol precursors. *bioRxiv* Preprint

121. Berman P, de Haro LA, Cavaco AR, Panda S, Dong Y, et al. 2024. The biosynthetic pathway of the hallucinogen mescaline and its heterologous reconstruction. *Molecular Plant* 17(7):1129−50

122. Farooq MA, Gao S, Hassan MA, Huang Z, Rasheed A, et al. 2024. Artificial intelligence in plant breeding. *Trends in Genetics* 40(10):891−908

123. Raza SEA, Smith HK, Clarkson GJJ, Taylor G, Thompson AJ, et al. 2014. Automatic detection of regions in spinach canopies responding to soil moisture deficit using combined visible and thermal imagery. *PLoS ONE* 9(6):e97612

124. Gené-Mola J, Gregorio E, Auat Cheein F, Guevara J, Llorens J, et al. 2020. Fruit detection, yield prediction and canopy geometric characterization using LiDAR with forced air flow. *Computers and Electronics in Agriculture* 168:105121

125. Ritharson PI, Raimond K, Mary XA, Robert JE, J A. 2024. DeepRice: a deep learning and deep feature based classification of rice leaf disease subtypes. *Artificial Intelligence in Agriculture* 11:34−49

126. Moore BM, Wang P, Fan P, Leong B, Schenck CA, et al. 2019. Robust predictions of specialized metabolism genes through machine learning. *Proceedings of the National Academy of Sciences of the United States of America* 116(6):2344−53

127. Wang P, Schumacher AM, Shiu SH. 2022. Computational prediction of plant metabolic pathways. *Current Opinion in Plant Biology* 66:102171

128. Xiao H, Liu Z, Wang N, Long Q, Cao S, et al. 2023. Adaptive and maladaptive introgression in grapevine domestication. *Proceedings of the National Academy of Sciences of the United States of America* 120(24):e2222041120

129. Feng W, Gao P, Wang X. 2024. AI breeder: Genomic predictions for crop breeding. *New Crops* 1:100010

130. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583−89

131. Abramson J, Adler J, Dunger J, Evans R, Green T, et al. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630:493−500

132. Baek M, McHugh R, Anishchenko I, Jiang H, Baker D, et al. 2024. Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA. *Nature Methods* 21(1):117−21

133. Yao Y, Chen F, Wu C, Chang X, Cheng W, et al. 2025. Structure-based virtual screening aids the identification of glycosyltransferases in the biosynthesis of salidroside. *Plant Biotechnology Journal* 23(5):1725−35

134. Liu M, Li Y, Li H. 2022. Deep learning to predict the biosynthetic gene clusters in bacterial genomes. *Journal of Molecular Biology* 434(15):167597

135. Rios-Martinez C, Bhattacharya N, Amini AP, Crawford L, Yang KK. 2023. Deep self-supervised learning for biosynthetic gene cluster detection and product classification. *PLOS Computational Biology* 19(5):e1011162

136. Yang B, Meng T, Wang X, Li J, Zhao S, et al. 2024. CAT Bridge: an efficient toolkit for gene–metabolite association mining from multiomics data. *GigaScience* 132:giae083

137. Cui H, Wang C, Maan H, Pang K, Luo F, et al. 2024. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. *Nature Methods* 21(8):1470−80

138. Moreno-Paz S, van der Hoek R, Eliana E, Zwartjens P, Gosiewska S, et al. 2024. Machine learning-guided optimization of p-Coumaric acid production in yeast. *ACS Synthetic Biology* 13:1312−22

139. Moreno-Paz S, Schmitz J, Suarez-Diez M. 2024. *In silico* analysis of design of experiment methods for metabolic pathway optimization. *Computational and Structural Biotechnology Journal* 23:1959−67

140. Moreno-Paz S, van der Hoek R, Eliana E, Martins dos Santos VAP, Schmitz J, et al. 2024. Combinatorial optimization of pathway, process and media for the production of *p*-coumaric acid by Saccharomyces cerevisiae. *Microbial Biotechnology* 17(3):e14424

141. Zheng S, Zeng T, Li C, Chen B, Coley CW, et al. 2022. Deep learning driven biosynthetic pathways navigation for natural products with BioNavi-NP. *Nature Communications* 13:3342

142. Misra RC, Garg A, Roy S, Chanotiya CS, Vasudev PG, et al. 2015. Involvement of an ent-copalyl diphosphate synthase in tissue-specific accumulation of specialized diterpenes in *Andrographis paniculata*. *Plant Science* 240:50−64

143. Wang J, Lin HX, Su P, Chen T, Guo J, et al. 2019. Molecular cloning and functional characterization of multiple geranylgeranyl pyrophosphate synthases (ApGGPPS) from Andrographis paniculata. *Plant Cell Reports* 38:117−28

144. Durairaj P, Li S. 2022. Functional expression and regulation of eukaryotic cytochrome P450 enzymes in surrogate microbial cell factories. *Engineering Microbiology* 2:100011

145. Qiu S, Wang J, Pei T, Gao R, Xiang C, et al. 2025. Functional evolution and diversification of CYP82D subfamily members have shaped flavonoid diversification in the genus *Scutellaria*. *Plant Communications* 6:101134

146. Li Q, Jiao X, Li X, Shi W, Ma Y, et al. 2024. Identification of the cytochrome P450s responsible for the biosynthesis of two types of aporphine alkaloids and their *de novo* biosynthesis in yeast. *Journal of Integrative Plant Biology* 66(8):1703−17

147. Li C, Li Y, Wang J, Lu F, Zheng L, et al. 2025. An independent biosynthetic route to frame a xanthanolide-type sesquiterpene lactone in Asteraceae. *The Plant Journal* 121(2):e17199

148. Wang J, Xie Q, Wang X, Long M, Chen Y, et al. 2025. Discovery of key cytochrome P450 monooxygenase (C20ox) enables the complete synthesis of tripterifordin and neotripterifordin. *ACS Catalysis* 15(3):2690−702

149. Li Y, Xu J, Li G, Wan S, Batistič O, et al. 2019. Protein S-acyl transferase 15 is involved in seed triacylglycerol catabolism during early seedling growth in Arabidopsis. *Journal of Experimental Botany* 70(19):5205−16

150. Zhu Q, Yu S, Zeng D, Liu H, Wang H, et al. 2017. Development of "Purple Endosperm Rice" by engineering anthocyanin biosynthesis in the endosperm with a high-efficiency transgene stacking system. *Molecular Plant* 10(7):918−29

151. Liao J, Liu T, Xie L, Mo C, Qiao J, et al. 2023. Heterologous mogrosides biosynthesis in cucumber and tomato by genetic manipulation. *Communications Biology* 6:191

152. Irigoyen S, Ramasamy M, Pant S, Niraula P, Bedre R, et al. 2020. Plant hairy roots enable high throughput identification of antimicrobials against Candidatus Liberibacter spp. *Nature Communications* 11:5802

153. Cheng Y, Wang X, Cao L, Ji J, Liu T, et al. 2021. Highly efficient *Agrobacterium rhizogenes*-mediated hairy root transformation for gene functional and gene editing analysis in soybean. *Plant Methods* 17:73

154. Zhu Y, Zhu X, Wen Y, Wang L, Wang Y, et al. 2024. Plant hairy roots: Induction, applications, limitations and prospects. *Industrial Crops and Products* 219:119104

155. Han W, Xu J, Wan H, Zhou L, Wu B, et al. 2022. Overexpression of BcERF3 increases the biosynthesis of saikosaponins in *Bupleurum chinense*. *FEBS Open Bio* 12(7):1344−52

156. Burch-Smith TM, Anderson JC, Martin GB, Dinesh-Kumar SP. 2004. Applications and advantages of virus-induced gene silencing for gene function studies in plants. *The Plant Journal* 39(5):734−46

157. Liu Y, Lyu R, Singleton JJ, Patra B, Pattanaik S, et al. 2024. A Cotyledon-based Virus-Induced Gene Silencing (Cotyledon-VIGS) approach to study specialized metabolism in medicinal plants. *Plant Methods* 20(1):26

158. Yadav S, Badajena S, Khare P, Sundaresan V, Shanker K, Mani DN, Shukla AK. 2025. Transcriptomic insight into zinc dependency of vindoline accumulation in Catharanthus roseus leaves: relevance and potential role of a CrZIP. *Plant Cell Reports* 44(2):43

159. Garg A, Srivastava P, Verma PC, Ghosh S. 2024. ApCPS2 contributes to medicinal diterpenoid biosynthesis and defense against insect herbivore in *Andrographis paniculata*. *Plant Science* 342:112046

160. Liu S, Zhang H, Meng Z, Jia Z, Fu F, Jin B, Cao F, Wang L. 2025. The LncNAT11−MYB11−F3'H/FLS module mediates flavonol biosynthesis to regulate salt stress tolerance in Ginkgo biloba. *Journal of Experimental Botany* 76(4):1179−201

161. Cheng G, Shu X, Wang Z, Wang N, Zhang F. 2023. Establishing a Virus-Induced Gene Silencing System in Lycoris chinensis. *Plants* 12(13):2458

162. Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, et al. 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 496:528−32

163. Zhang J, Hansen LG, Gudich O, Viehrig K, Lassen LMM, et al. 2022. A microbial supply chain for production of the anti-cancer drug vinblastine. *Nature* 609:341−47

164. Gao J, Zuo Y, Xiao F, Wang Y, Li D, et al. 2023. Biosynthesis of catharanthine in engineered Pichia pastoris. *Nature Synthesis* 2(3):231−42

165. Gu Y, Jiang Y, Li C, Zhu J, Lu X, et al. 2024. High titer production of gastrodin enabled by systematic refactoring of yeast genome and an antisense-transcriptional regulation toolkit. *Metabolic Engineering* 82:250−61

166. Wu Y, Li S, Sun B, Guo J, Zheng M, et al. 2024. Enhancing gastrodin production in *Yarrowia lipolytica* by metabolic engineering. *ACS Synthetic Biology* 13(4):1332−42

167. Liu M, Wang C, Ren X, Gao S, Yu S, et al. 2022. Remodelling metabolism for high-level resveratrol production in *Yarrowia lipolytica*. *Bioresource Technology* 365:128178

168. Srinivasan P, Smolke CD. 2020. Biosynthesis of medicinal tropane alkaloids in yeast. *Nature* 585:614−19

169. Liu Y, Zhao X, Gan F, Chen X, Deng K, et al. 2024. Complete biosynthesis of QS-21 in engineered yeast. *Nature* 629:937−44