Exploring Uncharted Territories of Plant Specialized Metabolism in the Postgenomic Era

Joseph R. Jacobowitz[1,2] and Jing-Ke Weng[1,2]

[1]Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, 02142, United States, email: wengj@wi.mit.edu

[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, United States

Keywords

specialized metabolism, natural product, biosynthesis, multi-omics, metabolic engineering, metabolon

Abstract

For millennia, humans have harnessed plants for food, raw materials and medicines. It is only within the past two centuries that we have begun to connect particular plant metabolites with their specific properties and utilities. Since the utility of classical molecular genetics is limited beyond model species, the vast specialized metabolic systems present in the Earth's flora remain largely unstudied. With the exploding genomics resources and a rapidly expanding toolbox over the past decade, exploration of plant specialized metabolism in nonmodel species is becoming more feasible than ever before. Here, we review the state-of-the-art tools that enable this rapid progress. We present recent examples of de novo biosynthetic pathway discovery that employs various innovative approaches. We also draw attention to the higher-order organization of plant specialized metabolism at subcellular, cellular, tissue, interorgan and interspecies levels, which will have important implications for future design of comprehensive metabolic engineering strategies.

1. INTRODUCTION

The early establishment and subsequent radiation of plant life on land marked one of the most significant developments in the geological history of Earth (64). To survive the harsh terrestrial environments and defend against coevolving animals and microorganisms, land plants opted for

a unique adaptive strategy by greatly expanding their metabolic systems, which led to the production of a myriad of chemicals not present in their aquatic ancestors. The continuous evolution of these so-called "specialized" metabolites in land plants is so extensive that metabolic profiles vary significantly even between closely related species. Here, we define specialized metabolites as those that are not essential for survival but enhance the reproductive success of the host organism under specific environmental conditions. Collectively, specialized metabolism has played a paramount role in land plants' colonization of nearly all types of terrestrial habitats on planet Earth.

For millennia, humans have tapped into plant specialized metabolism for food, raw materials, and traditional herbal medicines (80). In the past two centuries, the arrival of modern science and technology has revealed the chemical nature of a growing list of plant specialized metabolites, which in turn has expanded their utilities. For example, plant natural products have had pivotal impacts in developing modern medicines to fight against cancer, bacterial infections, parasites, and more (93). The recently emerging synthetic biology approaches further opened new avenues to produce high-value plant natural products in alternative chassis organisms (99). Plant specialized metabolism will continue to serve as an important source of basic scientific discoveries as well as technological innovations that advance human health, renewable energy and sustainable green chemistry.

This review is intended to acquaint readers with the recent development in the field of plant specialized metabolism, whereas some historical aspects and background information are also provided for novices to the field. We discuss the emerging tools that have greatly facilitated the rapid progress in plant specialized metabolism research in recent years. We present examples of de novo biosynthetic pathway elucidation that employ innovative approaches. Finally, we draw attention to the higher-order organization of metabolic systems in native plant hosts, which is key to the plants' ability to achieve remarkable throughput in producing certain metabolites.

1.1. Historical Aspects of Human Exploration of Plant Specialized Metabolism

Since prehistoric times, humans have discovered plants with specific medicinal properties to combat maladies. Around the world, these traditional treatments were passed down through oral traditions and/or compiled into medical texts by ancient physicians. For instance, the classical Chinese medical text *Prescriptions for Emergencies*, authored by Ge Hong (284–346 AD), documented the treatment of fever by cold water extract of sweet wormwood (*Artemisia annua*).

This 1600-year-old text later inspired the discovery of the antimalarial medicine artemisinin from sweet wormwood by Youyou Tu in the 1970s (134), which saved millions of lives and also earned her the 2015 Nobel Prize in Physiology or Medicine. Although the efficacy of most traditional herbal remedies have not been verified through stringent modern-day clinical trials, these traditions were formed based on very long-term trials on real human patients, some of which have already demonstrated enormous value for developing new medicines (80).

With the recognition that discrete plant natural products are linked to specific biological activities, much effort in the twentieth century has been devoted to the field of phytochemistry (103). Significant amounts of knowledge about plant chemistry have been accumulated through the general workflow of compound isolation followed by small molecule structural elucidation. Drug screening using these isolated plant natural products also led to the development of a number of important clinical drugs (93). One prominent example is paclitaxel, which is a phenolic terpene initially isolated from the bark of the pacific yew tree (*Taxus brevifolia*) (137). Paclitaxel is widely used as a highly effective, broad-spectrum first-line chemotherapy medication, but is also one of the most expensive drugs to manufacture. The initial practice of purifying low quantities of paclitaxel from the bark quickly became unsustainable (90). Although total synthesis of paclitaxel was demonstrated by academic labs by the 1990s (49, 50, 95), the reported synthetic schemes were prohibitively costly to scale up for industrial production. An alternative semisynthetic approach was subsequently adopted to produce paclitaxel from 10-deacetylbaccatin, a relatively abundant precursor compound extractable from the needles of European yew (*Taxus baccata*). In recent years, the semisynthetic approach was superseded by *Taxus* cell culture system, which enabled paclitaxel production without the need of raw materials from yew trees. The paclitaxel story is a triumph of modern medicine, but it also illustrates several common bottlenecks that have limited the clinical potential of many otherwise highly promising plant natural products: low purification efficiency and yield, slow-growing or even endangered source plant, infeasible total synthesis, and lack of flexibility to explore additional structural analogs (50, 95).

A biosynthetic approach toward access of valuable plant natural products holds tremendous promise to resolve the aforementioned bottlenecks. However, most plant natural product biosynthetic pathways remain unresolved, whereas de novo elucidation of unknown biosynthetic pathways in plants is no easy task. In the early twentieth century, early pioneers in the field

explored the uncharted biosynthetic routes for several major classes of plant specialized metabolites, mostly through classical radio-labeled tracing experiments and activity-guided enzyme purification from the host plant followed by kinetic characterization (9, 19, 29, 36, 57). Starting in the 1980s, the adoption of several model plants (e.g., *Arabidopsis*, rice and maize) with the power of molecular genetics led to tremendous progress in uncovering the genetic basis of various specialized metabolic pathways, some of which turned out to be conserved throughout land plants, including pathways for lignin, flavonoids, and carotenoids (12, 32, 129). Others are restricted to certain taxa but represented by specific model plants, such as the pathways for benzoxazinoids, thalianol, and glucosinolates (40, 42, 125). However, the molecular genetics approach cannot be readily applied to study nonmodel plants, leaving the lion's share of plant chemodiversity embodied by the entirety of Earth's flora essentially untapped. Over the past decade, this barrier has been significantly lowered thanks to burgeoning techniques from the fields of next-generation sequencing, analytical chemistry, systems biology and synthetic biology. Along with greatly extended and accelerated discoveries of specialized metabolic systems in diverse nonmodel plants, metabolic engineering using knowledge learnt from these systems has begun to realize the vision of accessing high-value plant natural products through total biosynthesis (83, 86, 114).

1.2. Major Classes of Plant Specialized Metabolites

Specialized metabolism branches from various precursors in primary metabolism (142). Major classes of plant specialized metabolites discussed in this review include phenylpropanoids, terpenoids, alkaloids, and peptides (**Figure 1**). Enormous diversity of structure and function exists within each class and additional chemical diversity can also be achieved by joining structural moieties from two or more of these classes (e.g., paclitaxel is a terpene alkaloid with phenylpropanoid acyl moieties).

1.2.1 - Phenylpropanoids

Phenylpropanoid metabolism comprises a major branch of plant specialized metabolism derived from phenylalanine. In addition to the dazzling array of low-molecular-weight phenylpropanoids known to date, some phenylpropanoids also polymerize to yield important plant phenolic polymers, such as lignin and condensed tannin. Many plant-specific enzymes have evolved to produce unique phenylpropanoid scaffolds (46b). For example, type III polyketide

synthases (PKSs) iteratively condense malonyl-CoA and CoA-thioesters of hydroxycinnamic acid derivatives to produce various phenylpropanoid backbone structures, such as flavonoid and curcuminoid scaffolds (39b, 68a, 70a). These phenolic polyketide scaffolds are then further diversified by many additional tailoring enzymes. For instance, icariin, a natural phosphodiesterase 5 inhibitor found in *Epimedium* genus, is a glycosylated, methylated, and prenylated flavonol (31c). Lignans and coumarins are two other important classes of phenylpropanoids. Whereas lignans are dimers of monolignols, coumarins are derived from lactonization of *ortho*-hydroxylated *cis*-hydroxycinnamic acid (31a, 109a, 135a). Finally, phenylpropanoids may themselves be appended to other small molecule scaffolds by enzymes such as BAHD-acyltransferases or serine-carboxypeptidase-like (SCLP) enzymes (121a, 146a).

1.2.2 - Terpenoids

Terpenoids are a structurally diverse class of lipids. Terpenoids are initially formed by synthesizing the isomeric five-carbon dimethylallylpyrophosphate (DMAPP) or isopentenylpyrophosphate (IPP), which are ubiquitous primary metabolites. In plants, DMAPP and IPP are both produced by two evolutionarily distinct pathways: the mevalonic acid (MVA) pathway or the methylerythritol phosphate (MEP) pathway (128a). The MVA pathway is thought to be localized to the cytoplasm, although there is some evidence that suggests it may in fact be peroxisomal (117a). Alternatively, the MEP pathway is localized to plastids (32a, 114b). DMAPP and IPP are then condensed into long, linear prenyl diphosphate chains by a variety of prenyl transferases. The lengths of the prenyl chains come in multiples of five carbons, depending on the number of isoprenoid units. For example, geranyl diphosphate produces C10 terpenes, farnesyl diphosphate produces C15 sesquiterpenes, geranylgeranyl diphosphate produces C20 diterpenes, geranylfarnesyl diphosphate produces C25 sesterterpenes etc. The linear prenyl diphosphates are then converted into a variety of cyclic or acyclic terpenes by terpene synthases.

Terpenes likely represent the richest diversity of scaffold structures among all known plant metabolites. This diversity is produced by the divergent active sites of terpene synthases which guide the cyclization of linear precursors into a huge array of ring structures (143a). One significant example is the diterpene taxadiene, which is produced from geranylgeranyl diphosphate by taxadiene synthase and serves as the precursor for taxol biosynthesis (83b). Linear terpenes without diphosphates may also be cyclized. Perhaps most notably, oxidosqualene

cyclases use 2,3-oxidosqualene as the precursor to produce diverse triterpene backbones, including sterols such as lanosterol and cycloartenol (143b, 127a).

1.2.3 - Alkaloids

Alkaloids are natural products containing nitrogen. They are typically derived from amino acids or nucleotides, and adopt highly variable structures. Important examples of alkaloids include tryptophan-derived monoterpenoid indole alkaloids (MIAs) or bisindole alkaloids (BIAs) and arginine-derived tropane alkaloids (101a, 107a, 125a). Nucleotide-derived alkaloids are also important - most notable is caffeine. Caffeine is derived from guanine and is the most widely consumed psychoactive drug in the world (39a).

Nitrogen is a limiting resource for plant growth. As such, plants that engage in symbiotic relationships with nitrogen-fixing bacteria appear to have relaxed constraints on the amount and diversity of alkaloids they produce. Legumes, which form symbioses with N_2-fixing bacteria, produce a number of interesting alkaloids, such as pyrrolizidine alkaloids (142a, 142b). Nitrogen limitation may be an important consideration when choosing chassis organisms for engineering alkaloid production.

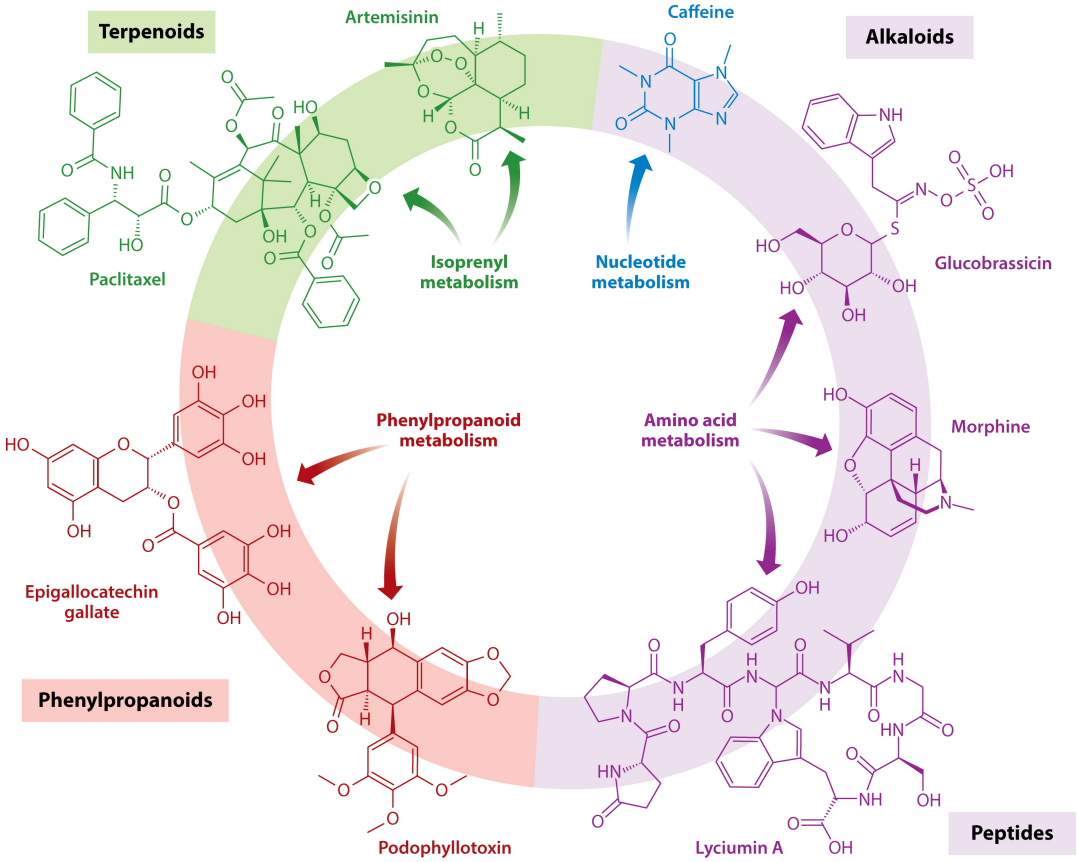
1.2.4 - Peptides

Peptide natural products are small molecules composed of amino acids primarily linked by amide bonds, which are categorized separately from alkaloids by convention. Moreover, peptide natural products are distinct from proteins in that they are often post-translationally modified in ways that are different from those observed for proteins. In recent years, peptides have been increasingly appreciated as signaling molecules in plants, whereas a few structural classes have also been recognized for medicinal properties (99a). In general, peptide natural products are biosynthesized by nonribosomal peptide synthetases (NRPSs) or through ribosome followed by post-translational modifications (109b). All plant peptides studied thus far are ribosomally synthesized and post-translationally modified peptides (RiPPs). Important examples of plant RiPPs include lyciumins, cyclotides, and orbitides (66, 18a, 44a, 114c).

1.2.5 – Others

Many other structural types of plant metabolites also exist that may not fall into the classes discussed above. For examples, fatty-acid-derived oxylipins serve as defense and signal compounds in plants (46a). Modified fatty acids are also found as monomeric components of the protective polymers cutin, suberin, and sporopollenin or deposited onto plant surfaces as wax

(83a). Some plant fatty-acid-derived lipids, such as acetylenic lipids, were reported to contain anticancer and antifungal activities (57a). A few known plant metabolites are halogenated. Halogenation is common among bacterial and marine natural products, but is rare in plants. Acutumine is a chloroalkaloid found in *Menispermum* with reported activity to inhibit T-cell proliferation (129a, 146b). Additionally, chlorinated auxin derivatives have been identified in numerous legumes species (128b). In these cases, the halogenation may originate from the metabolic activities of microbial symbionts. This is indeed the case for maytansine, a macrolide isolated from plants of the genus *Maytenus*, which is biosynthesized by root-associated bacteria and accumulated in plant roots (72). Sulfur-containing metabolites have also been identified in plants. Plants of the *Tagetes* genus exude thiophenes from their roots, which act as potent antibacterials (31b). Another fatty-acid-derived thiophene-containing metabolite, arctinone A, is produced in edible burdock (57a). Thiarubrine A is an organosulfur-heterocycle-containing molecule derived from fatty acids, which displays antibiotic activity and is produced by plants of the *Aspilia* genus (114a).



Jacobowitz JR, Weng J-K. 2020. *Annu. Rev. Plant Biol.* 71:631–58

Figure 1 Major classes of plant specialized metabolites. Plant specialized metabolites fall into several main classes, including terpenoids, phenylpropanoids, alkaloids, and peptides. These are derived from primary metabolic pathways comprising isoprenyl metabolism, phenylpropanoid metabolism, amino acid metabolism, and nucleotide metabolism. Example metabolites are included around the diagram.

2. THE MODERN TOOLBOX FOR STUDYING PLANT SPECIALIZED METABOLISM

Over the past decade, innovations in the fields of genomics, transcriptomics, proteomics, metabolomics, analytical chemistry and synthetic biology have generated many new tools at the disposal of modern-day biologists. Collectively, these tools made de novo biosynthetic pathway discovery in nonmodel species much more feasible than ever before (131). In this section, we review the tools that have had a transformative impact on plant specialized metabolism research in recent years.

2.1. General Tools for Rapid Identification and Functional Testing of *Candida*te Biosynthetic Genes

With one or a group of structurally related plant natural products in mind, researchers typically first hypothesize putative biosynthetic pathway(s) starting from feasible primary metabolite precursors, and through a series of chemical transformations, arriving at the final product(s). Researchers then try to identify the genes encoding enzymes that catalyze the predicted chemical transformations. Here, next-generation sequencing technology provides a major inroad toward pathway discovery. Massively parallel RNA sequencing (RNA-seq) can be used to generate de novo transcriptomes of the producer plants, from which comprehensive gene lists can be derived. If necessary, reference-quality genome sequences of the producer plant can also be obtained with relative ease using new technologies such as PacBio's single molecule, real-time (SMRT) sequencing technology (112) and 10X Genomics' linked-read sequencing technology (101), followed by de novo genome assembly.

With an extensive gene list in hand, researchers select candidate genes based on their biosynthetic proposal(s) (also discussed in section 3). *Candida*te enzymes can be produced heterologously in a variety of microbial hosts, including *Eschericia coli*, *Pichia pastoris*, and

Saccharomyces cerevisiae, for in vitro and in vivo functional characterization (99). In recent years, functional testing of plant biosynthetic genes has also been facilitated by the development of the *Nicotiana benthamiana* transient expression system (115). *N. benthamiana* is a close relative of tobacco. The leaves of *N. benthamiana* can be conveniently infected by syringe infiltration with a culture of Agrobacterium that harbors binary vectors designed to transiently express candidate genes of interest in the leaf tissue (115) (**Figure 2**). This technology has been instrumental to plant biochemists for several reasons. Firstly, *N. benthamiana* provides a cellular context more similar to the native host plant than other microbial hosts for heterologous expression (110). This increases the likelihood that the encoded protein is properly folded and targeted to the correct subcellular compartment. *N. benthamiana* naturally possesses many precursors substrates and cofactors which may be useful for reconstituting the tested enzymatic activity. Second, candidate gene expression is rapid – expression peaks around five days postinfiltration, which is significantly shorter than the turnaround time for generating stable transgenic plants. Third, Agrobacterium cultures can be combined to express multiple genes and coinfiltrated with any substrates that may not be naturally present in *N. benthamiana*. This has enabled pooled testing, in which many candidates can be combined into a single experiment. If the pooled candidates produce the product of interest, the gene list can be further narrowed to identify the minimal gene set required to produce the final product (26). Co-expression of multiple genes has also enabled the reconstitution of an entire biosynthetic pathway in a single, rapid experiment (109).

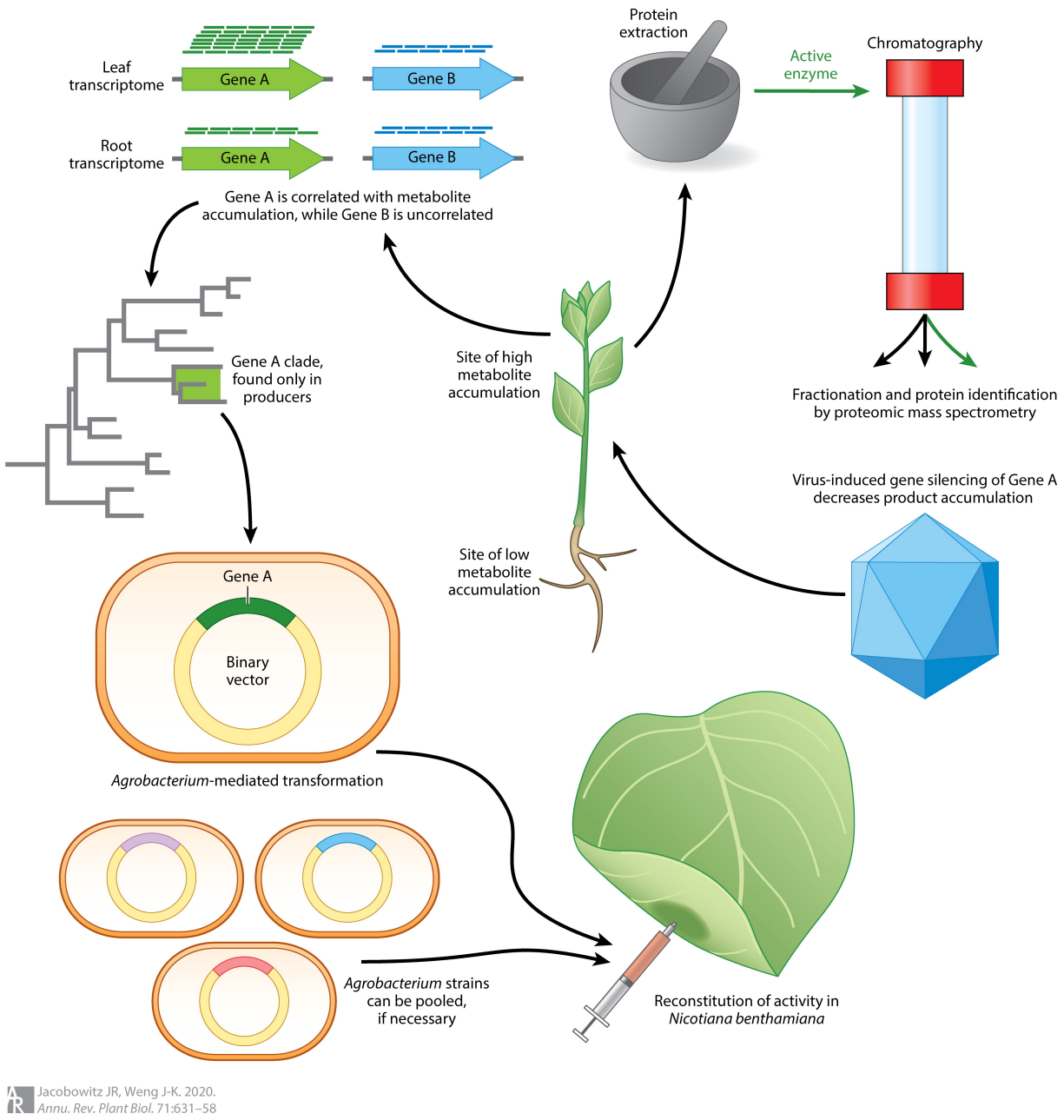


Figure 2 De novo elucidation of specialized metabolic pathway in nonmodel plants by modern approaches. To elucidate an unknown pathway in a nonmodel plant many techniques can be employed. Researchers might correlate gene expression with the accumulation of the natural product of interest. Alternatively, pathway enzymes might be isolated from crude protein extracts and identified by activity-guided fractionation followed by proteomic identification. Genes can be examined with phylogenetic analyses to compare the presence of genes with the presence of the natural product. Enzyme biochemistry can be reconstituted in *N. benthamiana* to test for sufficiency or knock-down in the native plant by VIGS to test for necessity.

To complement the gain-of-function experiment enabled by the *N. benthamiana* transient expression system, one important addition to the toolbox of plant biochemists was the Virus-Induced Gene Silencing (VIGS) method (87). In VIGS, the RNA-based antiviral defense of a plant is used to target the knock down of an endogenous gene (**Figure 2**). VIGS has been used to demonstrate the necessity of enzymes in various natural product biosynthetic pathways in diverse nonmodel plants, including the morphine and *vinca* alkaloid biosynthetic pathways (24, 43, 82, 107, 116).

2.2. Advances in Metabolite Analysis

To identify potential biosynthetic intermediates in the host plant and to characterize candidate enzyme activities, Liquid Chromatography-Mass Spectrometry (LC-MS) has become the most commonly used analytical method. Although the LC-MS technology has existed for decades, recent developments have produced substantially more sensitive, precise and versatile instruments, which greatly enhances our ability to identify and quantify low-abundance compounds even from very complex mixtures. High-resolution MS has also largely replaced the need for radioisotopic labeling. Instead, plants can be fed with stable isotopic precursors, and the isotopes carried by the derived metabolic products can be monitored by high-resolution LC-MS to help delineate unknown pathways and form sensible biosynthetic proposals (27, 140).

Historically, it has been very challenging to gain structural information of small molecules by MS data alone. However, recent advances in both MS hardware (e.g., greatly improved resolution and the new MS^n fragmentation capability) and data analysis software now facilitate MS-based small molecule structure elucidation. Computational fragment analysis algorithms, such as SIRIUS (34), have greatly improved the quality of molecular formula and putative structure prediction from MS spectral data (34). Moreover, several public metabolomics spectral databases have been established over the past few years, which allow researchers to upload, search and compare MS datasets (45, 47, 139). Complementing MS-based analyses, NMR (Nuclear Magnetic Resonance) spectroscopy has served as an essential tried-and-true method for structure elucidation in plant metabolism research. Recent developments in NMR hardware, such as new cryoprobes and stronger 1 – 1.2 GHz magnetic field NMRs have resulted in substantial enhancements to sensitivity and resolution of NMR experiments, which will likely aid NMR-based metabolomics in the future (66a, 107a). Ongoing efforts also seek to integrate LC-MS and NMR into a single pipeline for rapid structure elucidation, thus combining the strengths of these

two technologies (11a). Solid-state NMR technology has also advanced with the development in High Resolution Magic Angle Spinning (HRMAS) NMR. In HRMAS NMR experiments, the sample is rapidly rotated along an axis which is oriented at the "magic angle" with respect to the direction of the magnetic field, which dramatically increases data resolution (48a). HRMAS NMR has been useful for structural elucidation of plant biopolymers such as sporopollenin and cell wall polysaccharides (80a, 140a).

In addition to NMR, the elucidation of the exact stereochemistry of natural products at extremely low quantities has also been facilitated by new developments in the field of analytical chemistry. For instance, the crystalline sponge method bypasses the need for analyte crystallization by using a preformed porous crystalline metal-organic framework that absorbs nanogram quantities of analyte into its cavities prior to X-ray diffraction analysis (55). The crystalline sponge method recently played a key role in defining the functions of two new terpene cyclases from red algae (65) and *Emericella* fungi (89) respectively. It was also used to resolve the absolute structure of a dihydroxycholesterol intermediate that ultimately led to the elucidation of the plant diosgenin biosynthetic pathway (26). In addition to the crystalline sponge method, the recently developed electron cryo-microscopy (cryoEM) method microcrystal electron diffraction (MicroED) enables absolute structure determination from small molecule nanocrystals (composed of femtogram quantity of the analyte) (59). Both methods will serve as powerful tools for natural product structural elucidation, a key component of plant metabolism research.

2.3. Computational Tools that Aid Pathway Discovery

Computational tools play an increasingly important role in organizing massive amounts of plant multi-omics datasets and deriving useful predictive insights from them. Moore et al. recently developed a machine-learning algorithm to predict whether or not a gene may be involved in primary or specialized metabolism based on a variety of factors (92). Relative to primary metabolic enzymes, specialized metabolic genes were found to be more frequently in tandemly duplicated gene clusters and are more likely to be coexpressed with their paralogs. Moreover, specialized metabolic genes tend to exhibit specific spatial expression patterns, less sequence conservation, and low connectivity to gene networks of primary metabolism. Together, machine learning of known metabolic pathways generated a prediction model that can robustly identify specialized metabolic genes in unstudied systems.

It is now appreciated that many plant biosynthetic pathways exhibit some degree of gene clustering in the genome (22, 68). To systematically utilize this information for pathway discovery, several tools have been designed to identify gene clustering and predict specialized metabolic genes based on their positions in plant genomes (5, 41, 63, 119, 130).

3. OVERCOMING HURDLES IN DE NOVO SPECIALIZED METABOLIC PATHWAY ELUCIDATION

Given the existing knowledge base of plant metabolism, common chemical transformations often reappear in diverse pathways and are catalyzed by members of several large enzyme families, including but not limited to cytochrome P450s, iron/2-oxoglutarate-dependent dioxygenases (Fe/2OGs), UDP-sugar glycosyltransferases (UGTs), BAHD and SCPL acyltransferases, methyltransferases, NADH/NADPH-dependent reductases, type III polyketide synthases (PKSs), terpene synthases (TSs), and acyl-CoA ligases (ACLs) (142). In plants, most of these enzyme families have undergone extensive expansion (22). Although genes belonging to these families can be automatically identified by standard transcriptome annotation pipelines, the main challenge is to prioritize candidate genes to test based on other sources of indications. Moreover, some plant natural products contain structural moieties that necessitate unprecedented enzymatic activities. In such cases, a homology-based candidate-gene approach alone may not suffice. Additional hints are therefore necessary to identify genes underlying those novel enzymatic activities. In this section, we discuss recent research that explored innovative approaches to overcome hurdles in de novo elucidation of plant specialized metabolic pathway (**Summarized in Table 1**).

3.1. Taking Advantage of Co-regulated Biosynthetic Gene Networks

Under selection, the expression of genes in a particular plant specialized metabolic pathway sometimes evolved to be highly correlated with each other in a spatial and temporal manner. Moreover, RNA-seq has made it easy to profile global gene expression in multiple tissue types, across developmental stages, or under elicitation by various stimuli (**Figure 2**). As a result, coexpression analysis has become one of the most powerful approaches for candidate biosynthetic gene identification in nonmodel plants. For example, by employing coexpression analysis in Madagascar periwinkle (*Catharanthus roseus*) along with a well-thought-out chemical proposal, Caputi et al. identified four new enzymes, namely precondylocarpine acetate

synthase (PAS), dihydroprecondylocarpine acetate synthase (DPAS), tabersonine synthase (TS), and catharanthine synthase (CS), that work in concert to produce tabersonine and catharanthine scaffolds from stemmadenine acetate, resolving the last missing steps in vinblastine biosynthesis (20). Particularly useful for this analysis was the fact that methyl jasmonate induces vinblastine biosynthetic genes in Madagascar periwinkle hairy root culture, including the ones encoding these four enzymes. Similarly, Lau et al. elucidated podophyllotoxin biosynthesis in mayapple by examining genes that correlate with podophyllotoxin accumulation in the rhizome, stem, and leaf (76). The authors took advantage of the fact that podophyllotoxin biosynthesis is highly inducible – this time upon tissue damage. Among the six podophyllotoxin biosynthetic enzymes identified, a Fe/2OG was found to catalyze closure of the core cyclohexane ring of the aryltetralin scaffold, which is an unprecedented activity for Fe/2OGs. A similar coexpression-based approach was also recently employed to elucidate the biosynthesis of falcarindiol, an acetylenic lipid found in tomato, carrot, and ginseng (57a). Falcarindiol production was induced with different fungal and bacterial pathogens or elicitors and the accumulation of falcarindiol was carefully correlated with transcriptomics data. This analysis uncovered a novel gene cluster controlling falcarindiol biosynthesis and identified ACET1a/b which possesses rare acetylenase activity.

The elucidation of vinca alkaloid, podophyllotoxin, and falcardindiol biosynthesis are three exemplary studies that illustrate the utility of coexpression analysis for de novo biosynthetic pathway discovery in nonmodel plants. These successes depended on biosynthetic genes being correlated with their respective natural product and with each other in planta. Although this is certainly true in many cases, this approach would not be effective if the site of biosynthesis differs from the site of metabolite accumulation. Moreover, this approach may also be limited if intermediates are produced in one part of the plant and transported to a distal part of the plant, where biosynthesis is completed.

3.2. Leveraging Natural Variations to Identify Specialized Metabolic Genes

In addition to coexpression analysis, other correlation-based approaches have also found use in plant specialized metabolism research. Genome-Wide Association Study (GWAS) and Quantitative Trait Locus (QTL) analyses have leveraged the growing genomic resources to uncover associations between genetic polymorphisms and variations in specialized metabolic traits in plants, leading to the discovery of new specialized metabolic genes. Compared to other

types of agricultural traits, metabolic traits tend to be controlled by relatively few loci with large effects, which makes this approach particularly effective for identifying major contributors to the underlying pathways (38). For example, a novel secondary metabolite *N*-malonyl-*D*-alloisoleucine and a *D*-amino acid racemase involved in its biosynthesis were discovered by correlating untargeted metabolomic profiling data with Single Nucleotide Polymorphisms (SNPs) across 440 natural accessions in *Arabidopsis thaliana* (127). Similarly, two spermidine hydroxycinnamoyl transferases were identified by GWAS correlating natural variation of hydroxycinnamoyl spermidine levels with SNPs in 156 accessions of rice (33). It is foreseeable that GWAS approach can be further extended to study specialized metabolic traits in nonmodel plants in the near future. Instead of using SNPs found in reference genome sequences, next-generation shotgun genome or transcriptome sequences can be generated from natural accessions of a nonmodel plant species, from which k-mers derived from the raw sequencing reads can be used to correlate with categorical metabolic traits to arrive at informative genotype-to-phenotype associations (108).

3.3. Detecting Evolutionarily New Biosynthetic Enzymes by Phylogenomics

It is common that correlation-based analyses might not be sufficient to narrow down the candidate pool to a manageable number that can be tested within a reasonable timeframe. This is especially true for extremely large plant enzyme families such as P450s and UGTs. In these cases, phylogenetic analyses can provide additional evolutionary information to further guide candidate gene selection (**Figure 2**). For example, the specific UGT responsible for salidroside biosynthesis in golden root (*Rhodiola rosea*) was recently identified from 113 candidate UGT genes present in the sequenced transcriptomes (133). Many of the candidate UGTs showed correlated gene expression with salidroside accumulation in the roots. To further narrow down the number of candidate UGTs, Torrens-Spence et al. employed a coarse-grain phylogenetic approach, wherein 34 UGTs were selected for functional testing that represent all major clades revealed by the transcriptome-wide UGT phylogenetic analysis. Among these tested, one clade showed the desirable tyrosol:UDP-glucose 8-O-glucosyltransferase (T8GT) activity. Further in-depth characterization of UGTs within this clade identified a single UGT with the maximal specific T8GT activity. Phylogenetic analysis was also recently employed to help identify two styrylpyrone synthases (SPSs), which are neofunctionalized type III PKSs that establish the styrylpyrone scaffold in kavalactone biosynthesis in kava (*Piper methysticum*) (104). The PKS

phylogeny clearly indicates that kava SPSs are evolutionarily derived from chalcone synthases (CHS) through duplication events that occurred only in kava but not in closely related *Piper* species that do not accumulate kavalactones. In a third study, researchers identified CYP71BQ5 and CYP71CD2 as key oxidases in the biosynthesis of azadirachtin, a potent insecticidal compound produced by *Azadirachta indica* (50a). Key to this finding was the use of phylogenetic analysis, which narrowed down these P450s to a clade comprising only *A. indica* sequences, suggesting their involvement in this lineage-specific pathway.

3.4. Activity-guided Approaches to Identify Novel Enzymes

Homology-guided approaches no longer work when a pathway step entails unprecedented biochemistry. In these cases, the classical activity-guided fractionation may be employed to identify enzymes with proposed activities. In an activity-guided fractionation experiment, the enzymatic activity must first be observed in crude protein extract, after which the protein extract is subjected to chromatographic separation step(s) to enrich the target enzymatic activity (**Figure 2**). Although this classical approach might have fallen out of style in the postgenomic era of plant metabolism research, it remains a very powerful approach for identifying novel enzymes in plants. The new transcriptomics and proteomics tools have also made target enzyme identification from active fractions much easier than before. The activity-guided approach was recently used to identify thebaine synthase in opiate alkaloid biosynthesis in opium poppy (*Papaver somniferum*) (24). Thebaine synthase belongs to a poorly understood gene family called pathogenesis-related 10 (PR10) proteins. Although other PR10 members have been identified as enzymes, such as norcoclaurine synthase in benzylisoquinoline alkaloid biosynthesis, it would be nearly impossible to predict the function of thebaine synthase based on sequence alone. A similar approach was taken to identify Papain-like cysteine proteases from *N. benthamiana* that catalyze the N-terminal proteolysis of the cyclotide kalata B1 (kB1) (111). Since many enzyme families are capable of catalyzing hydrolysis chemistry, it would have been very difficult to identify papain-like cysteine proteases involved in kB1 biosynthesis simply by a candidate-based approach.

3.5. Sequence-guided Approaches to Discover New Plant Chemotypes

Thus far, we have discussed the challenge of identifying the genes responsible for unknown biosynthetic steps of a known plant natural product. In some cases, the sequence of the gene

itself gives structural information about the metabolite it synthesizes, which facilitates pathway elucidation. This is best seen in the field of peptide natural products. Kersten et al. recently employed a gene-guided approach to explore the biogenesis of a class of branched cyclic peptides knowns as lyciumins (66). Lyciumins were originally isolated from the roots of goji berry (*Lycium barbarum*), and are the bioactive principles of this Chinese medicinal plant traditionally used to treat hypertension (144). Transcriptome mining of goji berry roots identified a gene that encodes a precursor peptide containing repetitive motifs matching the lyciumin core peptide sequences. Transgenic expression of this gene in *N. benthamiana* was sufficient to reconstitute lyciumin biosynthesis. Additional lyciumin precursor genes that harbor 71 distinct core peptide sequences were identified in 21 of the 116 sequenced plant genomes, suggesting that lyciumin genotypes are widespread and comprise substantial, previously unrecognized chemodiversity in the plant kingdom. Similar workflows have also been applied to the study of plant cyclotides, which feature head-to-tail cyclization and a unique cystine knot topology (141). To date, tremendous diversity of cyclotides has been discovered in Poaceae, Violaceae, Rubiaceae, Cucurbitaceae, Fabaceae, and Solanaceae families (69, 94, 106).

Mechanistic understanding of the structure-function relationships within an enzyme family sometimes also allows prediction of enzymatic product based on signature sequences. For example, Aromatic Amino Acid Decarboxylases (AAADs) are capable of catalyzing either decarboxylation or aldehyde synthase chemistry to produce primary amine or aldehyde products, respectively (132). It was recently determined that a residue on the big catalytic loop, which can either be a tyrosine or phenylalanine, dictates decarboxylase or aldehyde synthase activity, respectively (132). This information was used to resolve the function of an AAAD in salidroside biosynthesis, which was previously thought to be a tyrosine decarboxylate but was instead shown to directly convert tyrosine to 4-hydroxyphenylacetaldehyde (133).

4. HIGHER-ORDER ORGANIZATION OF PLANT SPECIALIZED METABOLISM

Natural selection has propelled many plants to become highly efficient producers of specific natural products. Besides the burgeoning metabolic enzymes that continuously extend new boundaries of plant biosynthetic pathways, metabolic adaptation can also occur at the subcellular, cellular, tissue, interorgan and interspecies levels of biology. Some plants possess specialized tissues or organs, such as glandular trichomes, laticifers, resin ducts, and nectaries,

which are dedicated to produce and/or store specialized metabolites. Our understanding of the higher-order organizations of plant specialized metabolism has not kept pace with our ability to elucidate biochemical pathways, but this information will be useful for devising comprehensive strategies to engineer natural product production in heterologous hosts. With this section, we hope to draw attention to this understudied area of plant specialized metabolism.

4.1. Tissue-specific Metabolic Adaptation

One of the most common approaches employed by metabolic engineers to boost product yield is to adjust native metabolic network of the chassis organism. For instance, researchers may overexpress specific primary metabolic genes to redirect flux toward the engineered heterologous pathway or downregulate genes in other pathways that compete with the heterologous pathway. Evolution has adopted similar approaches to enhance natural product biosynthesis in native producer plants (3). MS technologies such as imaging MS and single-cell MS have enabled metabolic profiling at single-cell resolution in diverse plants, including *Arabidopsis*, Madagascar periwinkle, *Glycyrrhiza glabra*, and *Pelargonium zonale* (79, 84, 91, 128, 145). The emerging single-cell RNA-seq and laser microdissection (LMD) techniques further reveal molecular insights into metabolic specialization in those natural-product-producing cells (21, 122).

Many primary metabolic pathways contain feedback inhibition mechanisms that ensure tight regulation of metabolic flux (71). These mechanisms are key to maintain appropriate steady-state levels of important metabolite pools. Several recent studies show that the ancestral feedback inhibition mechanisms on certain upstream enzymes have been lost to allow accumulation of downstream specialized metabolites at high levels. For example, IPMS3 is a leucine biosynthetic enzyme that is typically subject to feedback inhibition by leucine. Tomato (*Solanum lycopersicum*) possesses a trichome-specific version of IPMS3 that has lost its inhibition domain, and therefore permits the overaccumulation of leucine-derived acylsugars (96). Similarly, Anthranilate synthase α (ASα) is an enzyme involved in tryptophan biosynthesis that converts chorismate to anthranilate and is typically inhibited by tryptophan. Common rue (*Ruta graveolens*) possesses an ASα isoenzyme that is tryptophan-insensitive, allowing common rue to rapidly biosynthesize anthranilate-derived defense compounds specifically upon fungal infection (11). Furthermore, tyrosine is typically synthesized in the plastid via a pathway that is feedback inhibited by tyrosine. However, legumes possess an alternative, cytosolic pathway for tyrosine

biosynthesis that is not subject to feedback inhibition, which is thought to support the biosynthesis of diverse tyrosine-derived natural products (118). Feedback-insensitive 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase from *E. coli* was engineered into *Arabidopsis*, which resulted in higher flux through shikimate metabolism and increased yields of multiple shikimate-derived natural products (135). Future engineering efforts may similarly employ feedback insensitive precursor metabolism to drive desired product accumulation.

4.2. Cellular Compartmentalization of Specialized Metabolism

Compartmentalization is a ubiquitous aspect of eukaryotic cellular metabolism. Many primary metabolic pathways are partitioned to particular organelles, while metabolites can be shuttled between cellular compartments. This is also the case for many specialized metabolic pathways. For example, most core terpenoid biosynthetic enzymes are targeted to the plastid, where the isoprene precursor substrates are produced (70). Plants have also evolved a number of specialized membrane-bound compartments that facilitate biosynthesis, such as tapetosomes or the specialized oil bodies of liverworts (48, 51). Phenyloplasts and tannosomes are two newly described plant metabolic compartments, which are the sites of production and storage for phenolics and condensed tannins, respectively (14, 15). It is likely that additional specialized compartments with metabolic functions are yet to be discovered in plants. Moreover, new tools are being developed to probe organelle biology. An affinity-based method was recently developed to rapidly isolate mitochondria for metabolite profiling before their contents diffuse into the surroundings (23). These methods can be adapted to probe metabolite composition of plant biosynthetic organelles.

Compartmentalization can be leveraged to boost the productivity of engineered organisms. Several groups have examined the effect of different enzyme localizations on metabolite yields. Targeting heterologously expressed patchoulol synthase and farnesyl pyrophosphate synthase to tobacco plastids resulted in 100 - 1000 fold higher accumulation of patchoulol than targeting these enzymes to the cytosol (143). Interestingly, the localization of a promiscuous enzyme can dictate its metabolic output. The bifunctional linalool/nerolidol synthase from strawberry can convert either geranyl diphosphate to linalool or farnesyl diphosphate to nerolidol (2). Targeting linalool/nerolidol synthase to either the plastid or mitochondria favors either linalool or nerolidol production, respectively (62). Plastid engineering and transformation is also emerging as a convenient tool for metabolic engineering. Plastid engineering is advantageous for

a variety of reasons, including the amenability of biosynthetic pathways to be stacked in bacteria-like operons (88). However, plastid engineering has traditionally been restricted to the relatively few plants whose plastid genomes that can be transformed. Recently, however, it was shown that plastid genomes can be inherited horizontally via graft junctions, which greatly expands the potential for plastid engineering (88, 126).

4.3. Physical Interactions of Biosynthetic Enzymes in vivo

The physical interaction and organization of enzymes in the cell is an intriguing phenomenon for some metabolic systems. So far, this has mostly been studied in the context of metabolons, which are complexes of sequential enzymes in a pathway that interact noncovalently (**Figure 3**). The physical organization is thought to channel the product of one enzyme into the active site of a sequential enzyme (53, 146). Numerous primary and specialized metabolic pathways are reported to organize in metabolons in both plants and animals (7). These pathways include phenylpropanoids, lignin, sporopollenin, flavonoids, isoflavonoids, and dhurrin in plants (1, 8, 17, 31, 61, 73). Cytochrome P450s and cytochrome P450 oxidoreductases (PORs) are proposed to serve as major organizers for some metabolons on the cytosolic face of endoplasmic reticulum (ER) (7). For instance in the lignin metabolon, transgenic expression of *P-COUMAROYLSHIKIMATE 3'-HYDROXYLASE* (*C3'H*) relocalizes the cytosolic HYDROXYCINNAMOYL TRANSFERASE (HCT) and 4-COUMAROYL-COA LIGASE to the ER membrane (8). These interactions may also be scaffolded by additional proteins, such as *Arabidopsis* Membrane Steroid Binding Proteins (MSBPs), which were recently found to provide additional scaffold interactions between the P450s C3′H, CINNAMIC ACID 4-HYDROXYLASE, and FERULATE 5-HYDROXYLASE in monolignol biosynthesis (44). Similarly, noncatalytic chalcone isomerases play a role in the β-bitter acid metabolon in hops (*Humulus lupulus*) by binding to and activating chalcone synthase (6, 81). In the dhurrin metabolon, UGT85B1 interacts with CYP79A1 and CYP71E1 to channel flux toward dhurrin biosynthesis (75). Compared to metabolons of primary metabolism, specialized metabolons appear to be characterized by weak protein-protein interactions that make the isolation of metabolons and in vitro reconstitution of metabolon interactions challenging. Metabolons tend to dissociate if conventional detergents are used to solubilize them from membrane fractions. Styrene maleic acid (SMA) polymer has recently been used to isolate and study active metabolons (75). The SMA polymer integrates into membranes and enables the extraction of

intact protein complexes that are surrounded by native lipids (77).

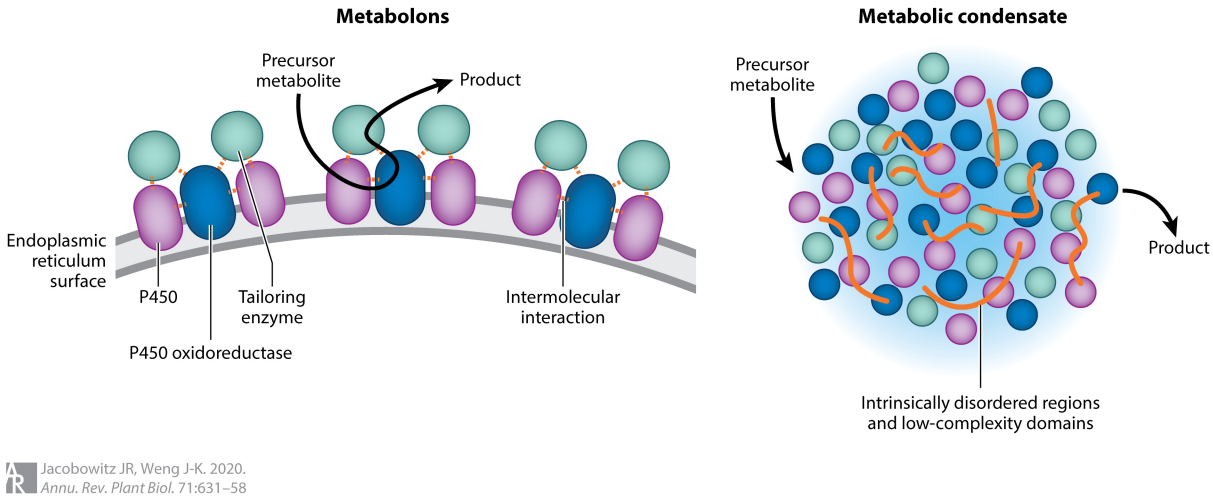


Figure 3 Schematic comparison of metabolons and metabolic condensates. Metabolons and metabolic condensates are two subcellular organizations of plant metabolic pathways. Plant metabolons are characterized by semistable protein-protein interactions which may be organized around P450s, P450 oxidoreductases, or other scaffolding proteins. Metabolic condensates are phase-separated bodies with associations mediated by intrinsically disordered regions or low complexity domains.

Eukaryotic cells contain various membrane-less compartments that display liquid-like properties (54). Classical examples of membrane-less bodies include nucleoli, cajal bodies, and P granules (13, 39). Recent studies have suggested an important role of liquid-liquid phase separation in many cellular processes, including concentrating transcriptional machinery (105), heat shock response (113), and DNA damage response (124). The formation of these phase-separated condensates is driven by multivalent nonspecific interactions between low-complexity domains (LCDs) or intrinsically disordered regions (IDRs) contained within the participating proteins (121). Phase separation also likely plays a role in metabolic processes (**Figure 3**). For example, phase-separated glycolytic bodies (G bodies) form under hypoxic conditions in *Saccharomyces cerevisiae* (58). Like metabolons, the formation of phase-separated metabolic condensates may also facilitate product channeling. While cases of phased-separated metabolic condensates have yet to be clearly defined in plant biology, the principle idea has already been used for metabolic engineering. Researchers recently achieved significantly improved overall yield and product specificity for deoxyviolacein production in yeast using a light-inducible phase separation system which entails fusing deoxyviolacein biosynthetic enzymes to the IDR of the

mammalian RNA binding protein FUS and *Arabidopsis* CRY2 (148).

4.4. Divisions of Biosynthetic Labor and Metabolite Transport Across Cell Types

Although plants may accumulate natural products in a particular tissue or organ, the site of accumulation is not always the site of synthesis. For example, BIAs (e.g., morphine) accumulate in the laticifers of opium poppy (10). However, there is a division of labor among various cells types in the biosynthesis of morphine (**Figure 4**). While most morphine biosynthetic enzymes accumulate in the sieve elements, the genes are transcribed and translated in the neighboring companion cells (117). Thus, the enzymes must be shuttled to the sieve elements. Pathway intermediates are synthesized in the sieve elements before being transported to the laticifers where the final biosynthetic steps are completed to yield final products (100). A similar scheme is employed by Madagascar periwinkle, which divides alkaloid biosynthesis among internal phloem-associated parenchyma, epidermal cells, laticifers, and idioblasts (46). In *Arabidopsis* glucosinolate biosynthesis, glucosinolates mostly accumulate in laticifer-like S cells as well as in epidermal cells, but they are actually synthesized in nearby vasculature parenchyma (97). Moreover, aliphatic and indole glucosinolates exhibit distinct but overlapping sites of biosynthesis. Aliphatic glucosinolate biosynthesis mostly occurs in phloem and xylem parenchyma, while indole glucosinolate biosynthesis is restricted to phloem parenchyma (97).

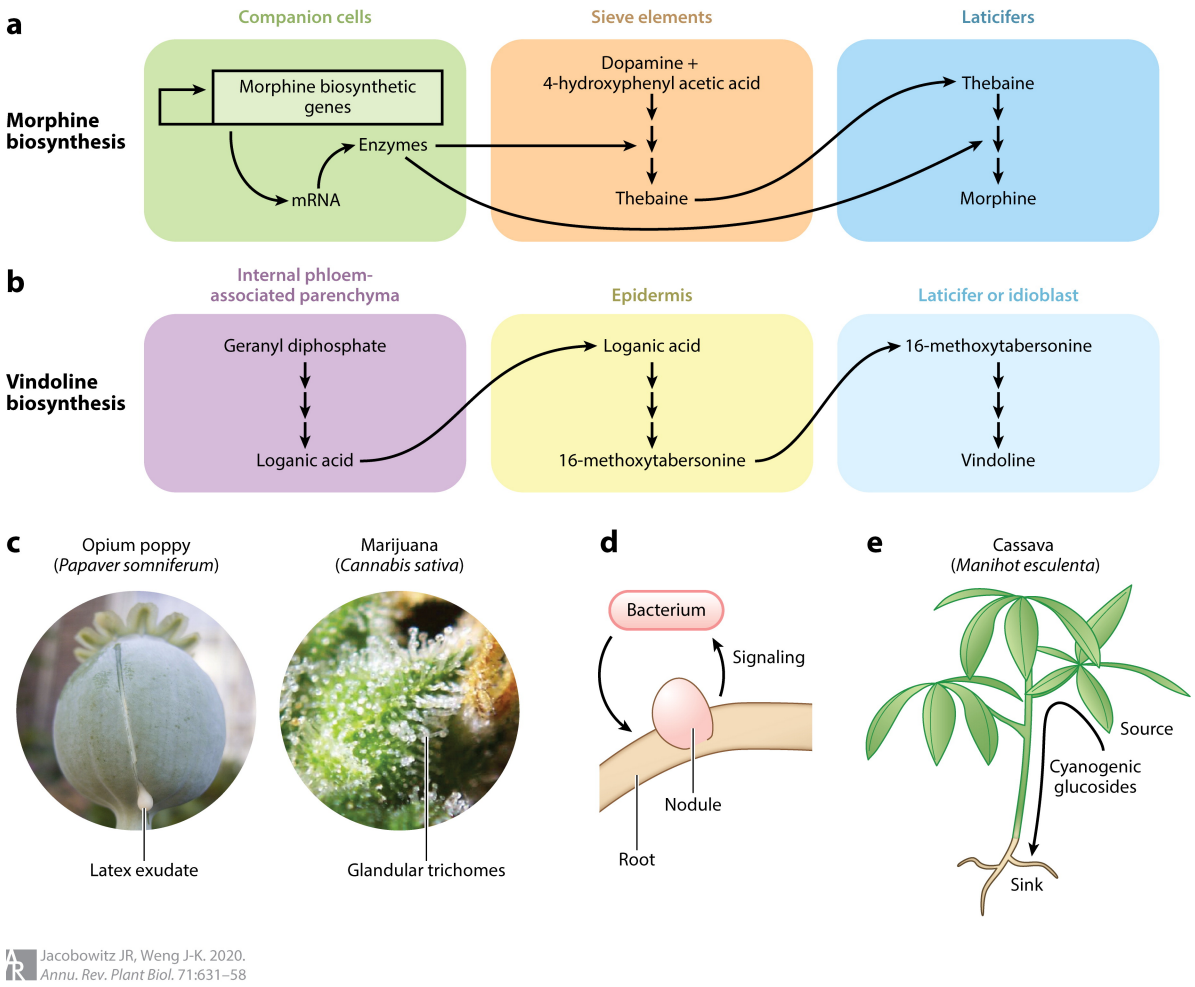


Figure 4 Organization of specialized metabolic machineries across cell types, tissues, and organs. **(A)** Simplified diagram of morphine biosynthesis, which divided across at least three cell types (100). Pathway enzymes are transcribed and translated in companion cells. Enzymes are transported to sieve elements or laticifers. Thebaine is synthesized in sieve elements and transported to laticifers, where morphine biosynthesis is completed. The pathway localization that is shown represents the bulk of morphine biosynthesis. **(B)** Simplified diagram of vindoline biosynthesis, which begins in internal phloem associated parenchyma until loganic acid is formed and transported to the epidermis (28). Loganic acid is converted to 16-methoxytabersonine before being transported to either laticifers or idioblasts, where vindoline biosynthesis is complete. **(C)** Opium poppy plant exuding alkaloid rich latex and glandular trichomes of a *Cannabis* plant. **(D)** Plant-bacterium signaling and symbiosis have implications for plant specialized metabolism. **(E)** Diagram demonstrating transport of cyanogenic glucosides in cassava from their site of synthesis in the leaves to their site of accumulation in roots (60).

In some instances, plant metabolites might be transported long distances from their site of synthesis, i.e., from source to sink. This is perhaps best shown for glucosinolates, which are

produced in leaf and silique tissues before being transported to the seeds (16). Glucosinolate transport depends on two proton-dependent transporters, GLUCOSINOLATE TRANSPORTER 1 (GTR1) and 2 (GTR2) (98). Long-distance transport has also been shown for nicotine, which is produced in the roots and transported to the shoots of tobacco plants (4). Alternatively, cyanogenic glucosides are produced in the leaves of cassava and transported to the roots (60) (**Figure 4**).

Interspecies interactions and metabolite transport may also play important roles in plant specialized metabolism, which have been extensively examined previously. For instance, the Crotolaria legume only synthesizes pyrrolizidine alkaloids in the nodules when engaged in active rhizobial symbiosis (56). It was also recently discovered that symbiotic arbuscular mycorrhizal fungi lack the genes necessary for fatty acid biosynthesis, but are compensated by lipid transfer from the host plant (67). Although this is an example from primary metabolism, it hints that interspecies cooperation might play a role in specialized metabolism as well. This seems to be true for maytansine which accumulates in the roots of plants of the *Putterlickia* genus, but is actually synthesized by endophytic bacteria (72).

As we learn more about where plant metabolism occurs, it is becoming clear that many specialized metabolites are not produced by a single cell type – rather multiple cell types are coordinated in the biosynthesis and storage of metabolites. Although techniques such as in situ hybridization and immunostaining can be readily applied to nonmodel plants to localize gene expression and protein accumulation, respectively, these techniques have not been broadly applied to plant biosynthetic systems. Understanding more about where the biosynthesis is occurring and how the relevant cell types are specialized for their particular functions would greatly benefit engineering exercises. For instance, the division principle of biosynthetic pathways across different cell types in plants may be replicated in heterologous microbial cocultures. High titers of resveratrol production were recently achieved using an *E. coli* coculture system composed of one *p*-coumarate-producing strain and a second strain that converts *p*-coumarate to resveratrol (18). A similar *E. coli-S. cerevisiae* coculture approach was also applied to produce naringenin (147).

4.5. Strategies for Plant Metabolite Storage

Metabolite storage is important for insulating end products from interfering with active biosynthetic machinery and mitigating any cytotoxic effects that may be associated with the

metabolites. Plants have evolved several strategies for metabolite storage, especially for hydrophobic or toxic natural products. Some plants store toxic compounds in a nontoxic form that may be activated at the appropriate time, such as glucosinolates, which are inactive until they are activated by myrosinase upon herbivore attack (52). Moreover, while many hydrophilic metabolites are stored in the vacuole, hydrophobic natural products present a particular challenge. Many plants accumulate large amounts of hydrophobic compounds in the subcuticular space of glandular trichomes (37). For instance, artemisinin accumulates in the subcuticular space in the glandular trichomes of sweet wormwood *(35)*. Oil bodies are also frequently used to store large amounts of hydrophobic compounds, such as *Coleus forskohlii* which stores forskolin-related diterpenoids in specialized oil bodies (102). Alternatively, the hydrophobic extracellular polymer suberin has also been posited as a site for accumulation of triterpenoids in *Tripterygium* roots (74).

Recently, it has been suggested that new solvent systems known as Natural Deep Eutectic Solvents (NADES) may allow plants to store large amounts of hydrophobic compounds (30). NADES are solvents composed of high concentrations of organic acids, salts, and sugars, which occur naturally in cells at very high concentrations. Although NADES have never been conclusively demonstrated to play a role in metabolite storage, many hydrophobic natural products exhibit dramatically increased solubility in NADES. For instance, quercetin is ~400,000 times more soluble in a NADES composed of xylitol, choline, chloride, and water than in water alone (30). Although enzymes are not typically active in NADES, their activity may be recovered upon dilution in water, suggesting that NADES may also play a role in pathway regulation (25). Although intriguing, future work will be required to determine if NADES are indeed used by plants for metabolite storage.

Overall, plants have devised many metabolite-storage strategies that appear to contribute significantly to the fitness of the hosts that evolved to synthesize large quantities of specialized metabolites. Without effective storage mechanisms, high-abundance end products may interfere with normal cellular processes, and therefore limit the extent of production. Similar problems might have confounded the engineering of these pathways in heterologous hosts (138). Thus we may achieve better metabolic engineering by learning more about natural product storage in the native plants and incorporating these strategies into chassis organisms.

5. CONCLUDING REMARKS

While human applications of plant specialized metabolism can be traced back millennia, mechanistic understanding of how plants manufacture natural products of diverse forms and functions has exploded within the past few decades. What began with the assignment of specific observable properties of plants to discrete chemical structures has grown into a blossoming field of research that explores entirely uncharted specialized metabolic pathways from almost any plant of interest. As technologies develop, the list of elucidated pathways will continue to expand, and at a faster pace. We highlight the higher-order organization of specific metabolic processes employed by plants to perfect natural product production. Future advances in this aspect of plant metabolism will instruct the design of more comprehensive metabolic engineering strategies. Harnessing the metabolic ingenuity of plants, arguably the most successful and self-sustaining kingdom of life, is a valuable gateway toward new technologies that ensure a prosperous and sustainable human civilization on the planet Earth.

DISCLOSURE STATEMENT

J.K.W. is a cofounder, a member of the Scientific Advisory Board, and a shareholder of DoubleRainbow Biosciences, which develops biotechnologies related to natural products.

ACKNOWLEDGMENTS

This work was supported by grants from the Keck Foundation, the Family Larsson Rosenquist Foundation, and the National Science Foundation (CHE-1709616). J.R.J. was supported by the National Science Foundation Graduate Research Fellowship (1122374). J.K.W. was supported by the Beckman Young Investigator Program, Pew Scholars Program in the Biomedical Sciences, and the Searle Scholars Program.

LITERATURE CITED

1. Achnine L, Blancaflor EB, Rasmussen S, Dixon RA. 2004. Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. *Plant Cell*. 16(11):3098–109
2. Aharoni A, Giri AP, Deuerlein S, Griepink F, de Kogel W-J, et al. 2003. Terpenoid

metabolism in wild-type and transgenic Arabidopsis plants. *Plant Cell*. 15(12):2866–84

3. Balcke GU, Bennewitz S, Bergau N, Athmer B, Henning A, et al. 2017. Multi-Omics of Tomato Glandular Trichomes Reveals Distinct Features of Central Carbon Metabolism Supporting High Productivity of Specialized Metabolites. *Plant Cell*. 29(5):960–83
4. Baldwin IT. 1989. Mechanism of damage-induced alkaloid production in wild tobacco. *J. Chem. Ecol.* 15(5):1661–80
5. Banf M, Zhao K, Rhee SY. 2019. METACLUSTER—an R package for context-specific expression analysis of metabolic gene clusters. *Bioinformatics* 35(17):3178–80
6. Ban Z, Qin H, Mitchell AJ, Liu B, Zhang F, et al. 2018. Noncatalytic chalcone isomerase-fold proteins in Humulus lupulus are auxiliary components in prenylated flavonoid biosynthesis. *PNAS* 115(22):E5223–32
7. Bassard J-E, Møller BL, Laursen T. 2017. Assembly of Dynamic P450-Mediated Metabolons—Order Versus Chaos. *Curr. Mol. Biol. Rep.* 3(1):37–51
8. Bassard J-E, Richert L, Geerinck J, Renault H, Duval F, et al. 2012. Protein-Protein and Protein-Membrane Associations in the Lignin Pathway. *Plant Cell* 24(11):4465–82
9. Battersby AR, Francis RJ, Ruveda EA, Staunton J. 1965. Biosynthesis of chelidonine and stylopine. *Chemical Communications (London)* 5: 89–91
10. Beaudoin GAW, Facchini PJ. 2014. Benzylisoquinoline alkaloid biosynthesis in opium poppy. *Planta* 240(1):19–32
11. Bohlmann J, Lins T, Martin W, Eilert U. 1996. Anthranilate synthase from Ruta graveolens. Duplicated AS alpha genes encode tryptophan-sensitive and tryptophan-insensitive isoenzymes specific to amino acid and alkaloid biosynthesis. *Plant Physiol*. 111(2):507–14

11a. Boiteau RM, Hoyt DW, Nicora CD, Kinmonth-Schultz HA, Ward JK, Bingol K. 2018. Structure Elucidation of Unknown Metabolites in Metabolomics by Combined NMR and MS/MS Prediction. *Metabolites*. 8(1):E8

12. Bonawitz ND, Chapple C. 2010. The genetics of lignin biosynthesis: connecting genotype to phenotype. *Annu. Rev. Genet.* 44:337–63
13. Brangwynne CP, Eckmann CR, Courson DS, Rybarska A, Hoege C, et al. 2009. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* 324(5935):1729–32
14. Brillouet J-M, Romieu C, Schoefs B, Solymosi K, Cheynier V, et al. 2013. The tannosome is

an organelle forming condensed tannins in the chlorophyllous organs of Tracheophyta. *Ann. Bot.* 112(6):1003–14

15. Brillouet J-M, Verdeil J-L, Odoux E, Lartaud M, Grisoni M, Conéjéro G. 2014. Phenol homeostasis is ensured in vanilla fruit by storage under solid form in a new chloroplast-derived organelle, the phenyloplast. *J. Exp. Bot.* 65(9):2427–35
16. Brown PD, Tokuhisa JG, Reichelt M, Gershenzon J. 2003. Variation of glucosinolate accumulation among different organs and developmental stages of Arabidopsis thaliana. *Phytochemistry* 62(3):471–81
17. Burbulis IE, Winkel-Shirley B. 1999. Interactions among enzymes of the Arabidopsis flavonoid biosynthetic pathway. *PNAS* 96(22):12929–34
18. Camacho-Zaragoza JM, Hernández-Chávez G, Moreno-Avitia F, Ramírez-Iñiguez R, Martínez A, et al. 2016. Engineering of a microbial coculture of Escherichia coli strains for the biosynthesis of resveratrol. *Microb. Cell Fact.* 15(1):163

18a. Camarero JA. 2017. Cyclotides, a versatile ultrastable micro-protein scaffold for biotechnological applications. Bioorg. Med. Chem. Lett. 27(23):5089–99

19. Camm EL, Towers GHN. 1969. Phenylalanine and tyrosine ammonia lyase activity in sporobolomyces roseus. *Phytochemistry* 8(8):1407–13
20. Caputi L, Franke J, Farrow SC, Chung K, Payne RME, et al. 2018. Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science* 360(6394):1235–39
21. Celedon JM, Yuen MMS, Chiang A, Henderson H, Reid KE, Bohlmann J. 2017. Cell-type- and tissue-specific transcriptomes of the white spruce (Picea glauca) bark unmask fine-scale spatial patterns of constitutive and induced conifer defense. *Plant J.* 92(4):710–26
22. Chae L, Kim T, Nilo-Poyanco R, Rhee SY. 2014. Genomic signatures of specialized metabolism in plants. *Science* 344(6183):510–13
23. Chen WW, Freinkman E, Sabatini DM. 2017. Rapid immunopurification of mitochondria for metabolite profiling and absolute quantification of matrix metabolites. *Nat. Protoc.* 12(10):2215–31
24. Chen X, Hagel JM, Chang L, Tucker JE, Shiigi SA, et al. 2018. A pathogenesis-related 10 protein catalyzes the final step in thebaine biosynthesis. *Nat. Chem. Biol.* 14(7):738–43
25. Choi YH, van Spronsen J, Dai Y, Verberne M, Hollmann F, et al. 2011. Are natural deep

eutectic solvents the missing link in understanding cellular metabolism and physiology? *Plant Physiol*. 156(4):1701–5

26. Christ B, Xu C, Xu M, Li F-S, Wada N, et al. 2019. Repeated evolution of cytochrome P450-mediated spiroketal steroid biosynthesis in plants. *Nat. Commun.* 10(1):3206
27. Colón AJM, Morgan JA, Dudareva N, Rhodes D. 2009. Application of Dynamic Flux Analysis in Plant Metabolic Networks. In *Plant Metabolic Networks*, ed. J Schwender, pp. 285–305. New York, NY: Springer New York
28. Courdavault V, Papon N, Clastre M, Giglioli-Guivarc'h N, St-Pierre B, Burlat V. 2014. A look inside an alkaloid multisite plant: the Catharanthus logistics. *Curr. Opin. Plant Biol.* 19:43–50
29. Croteau R, Karp F. 1977. Biosynthesis of monoterpenes: partial purification and characterization of 1,8-cineole synthetase from Salvia officinalis. *Arch. Biochem. Biophys.* 179(1):257–65
30. Dai Y, van Spronsen J, Witkamp G-J, Verpoorte R, Choi YH. 2013. Natural deep eutectic solvents as new potential media for green technology. *Anal. Chim. Acta.* 766:61–68
31. Dastmalchi M, Bernards MA, Dhaubhadel S. 2016. Twin anchors of the soybean isoflavonoid metabolon: evidence for tethering of the complex to the endoplasmic reticulum by IFS and C4H. *Plant J.* 85(6):689–706

31a. Davin LB, Wang HB, Crowell AL, Bedgar DL, Martin DM, et al. 1997. Stereoselective bimolecular phenoxy radical coupling by an auxiliary (dirigent) protein without an active center. *Science.* 275(5298):362–66

31b. Deineka VI, Tret'yakov MY. 2014. Thiophenes of Tagetes flowers and partial purification of xanthophyll esters. *Univers. J. Agr. Res.* 2(3):101-6

31c. Dell'Agli M, Galli GV, Dal Cero E, Belluti F, Matera R, et al. 2008. Potent inhibition of human phosphodiesterase-5 by icariin derivatives. *J. Nat. Prod.* 71(9):1513–17

32. DellaPenna D, Pogson BJ. 2006. VITAMIN SYNTHESIS IN PLANTS: Tocopherols and Carotenoids. *Annu. Rev. Plan.t Biol.* 57:711–38

32a. Disch A, Hemmerlin A, Bach TJ, Rohmer M. 1998. Mevalonate-derived isopentenyl diphosphate is the biosynthetic precursor of ubiquinone prenyl side chain in tobacco BY-2 cells. *Biochem. J.* 331 (Pt 2):615–21

33. Dong X, Gao Y, Chen W, Wang W, Gong L, et al. 2015. Spatiotemporal distribution of

phenolamides and the genetics of natural variation of hydroxycinnamoyl spermidine in rice. *Mol. Plant.* 8(1):111–21

34. Dührkop K, Fleischauer M, Ludwig M, Aksenov AA, Melnik AV, et al. 2019. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat. Methods* 16(4):299–302
35. Duke MV, Paul RN, Elsohly HN, Sturtz G, Duke SO. 1994. Localization of Artemisinin and Artemisitene in Foliar Tissues of Glanded and Glandless Biotypes of Artemisia annua L. *Int. J. Plant Sci.* 155(3):365–72
36. el-Basyouni SZ, Chen D, Ibrahim RK, Neish AC, Towers GHN. 1964. The biosynthesis of hydroxybenzoic acids in higher plants. *Phytochemistry* 3:485–92
37. Fahn A. 1988. Secretory tissues in vascular plants. *New Phytol*. 108(3):229–57
38. Fang C, Luo J. 2019. Metabolic GWAS-based dissection of genetic bases underlying the diversity of plant metabolism. *Plant J.* 97(1):91–100
39. Feric M, Vaidya N, Harmon TS, Mitrea DM, Zhu L, et al. 2016. Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* 165(7):1686–97

39a Ferré S. 2013. Caffeine and Substance Use Disorders. *J. Caffeine Res.* 3(2):57–58.

39b. Ferrer JL, Jez JM, Bowman ME, Dixon RA, Noel JP. 1999. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat. Struct. Biol*. 6(8):775–84

40. Field B, Osbourn AE. 2008. Metabolic diversification–independent assembly of operon-like gene clusters in different plants. *Science* 320(5875):543–47
41. Franke J, Kim J, Hamilton JP, Zhao D, Pham GM, et al. 2019. Gene Discovery in Gelsemium Highlights Conserved Gene Clusters in Monoterpene Indole Alkaloid Biosynthesis. *Chembiochem* 20(1):83–87
42. 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(15–16):1645–51
43. Geu-Flores F, Sherden NH, Courdavault V, Burlat V, Glenn WS, et al. 2012. An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* 492(7427):138–42
44. Gou M, Ran X, Martin DW, Liu C-J. 2018. The scaffold proteins of lignin biosynthetic cytochrome P450 enzymes. *Nat. Plants* 4(5):299–310

44a. Gran L. 1973. On the effect of a polypeptide isolated from “Kalata-Kalata”(Oldenlandia affinis DC) on the oestrogen dominated uterus. Acta Pharmacol. Toxicol. 33(5-6):400–408

45. Guijas C, Montenegro-Burke JR, Domingo-Almenara X, Palermo A, Warth B, et al. 2018. METLIN: A Technology Platform for Identifying Knowns and Unknowns. *Anal. Chem.* 90(5):3156–64

46. Guirimand G, Guihur A, Poutrain P, Héricourt F, Mahroug S, et al. 2011. Spatial organization of the vindoline biosynthetic pathway in Catharanthus roseus. *J. Plant Physiol.* 168(6):549–57

46a. Guo Q et al. 2018. JAZ repressors of metabolic defense promote growth and reproductive fitness in Arabidopsis. *PNAS* 115 (45):E10768–77

46b. Hahlbrock K, Scheel D. 1989. Physiology and Molecular Biology of Phenylpropanoid Metabolism. *Annu. Rev. Plant Physiol. Plant Mol Biol.* 40:347–369.

47. Haug K, Salek RM, Conesa P, Hastings J, de Matos P, et al. 2013. MetaboLights–an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res*. 41(Database issue):D781–86

48. He X, Sun Y, Zhu R-L. 2013. The Oil Bodies of Liverworts: Unique and Important Organelles in Land Plants. *CRC Crit. Rev. Plant Sci.* 32(5):293–302

48a. Hennel JW, Klinowski J. 2005. Magic Angle Spinning: A Historical Perspective. *Top. Curr. Chem.* 246:1-14

49. Holton RA, Kim HB, Somoza C, Liang F, Biediger RJ, et al. 1994. First total synthesis of taxol. 2. Completion of the C and D rings. *J. Am. Chem. Soc.* 116(4):1599-1600

50. Holton RA, Somoza C, Kim HB, Liang F, Biediger RJ, et al. 1994. First total synthesis of taxol. 1. Functionalization of the B ring. *J. Am. Chem. Soc.* 116(4):1597–98

50a. Hodgson H, De La Peña R, Stephenson MJ, Thimmappa R, Vincent JL, et al. 2019. Identification of key enzymes responsible for protolimonoid biosynthesis in plants: Opening the door to azadirachtin production. *PNAS* 116(34):17096–104

51. Hsieh K, Huang AHC. 2007. Tapetosomes in Brassica Tapetum Accumulate Endoplasmic Reticulum-Derived Flavonoids and Alkanes for Delivery to the Pollen Surface. *Plant Cell*. 19(2):582–96

52. Husebye H, Chadchawan S, Winge P, Thangstad OP, Bones AM. 2002. Guard Cell- and Phloem Idioblast-Specific Expression of Thioglucoside Glucohydrolase 1 (Myrosinase) in

Arabidopsis. *Plant Physiol.* 128(4):1180-8

53. Hyde CC, Padlan EA, Ahmed SA, Miles EW, Davies DR. 1987. 3-DIMENSIONAL STRUCTURE OF THE TRYPTOPHAN SYNTHASE ALPHA-2-BETA-2 COMPLEX FROM SALMONELLA AT 2.8 A RESOLUTION. *J. Biol. Chem.* 263(33):17857–71
54. Hyman AA, Weber CA, Jülicher F. 2014. Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* 30:39–58
55. Inokuma Y, Yoshioka S, Ariyoshi J, Arai T, Hitora Y, et al. 2013. X-ray analysis on the nanogram to microgram scale using porous complexes. *Nature* 495(7442):461–66
56. Irmer S, Podzun N, Langel D, Heidemann F, Kaltenegger E, et al. 2015. New aspect of plant-rhizobia interaction: Alkaloid biosynthesis in Crotalaria depends on nodulation. *PNAS* 112(13):4164–69
57. James WO. 1946. Biosynthesis of the Belladonna Alkaloids. *Nature* 159:196–7

57a. Jeon JE, Kim J-G, 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):176–87.e19

58. Jin M, Fuller GG, Han T, Yao Y, Alessi AF, et al. 2017. Glycolytic Enzymes Coalesce in G Bodies under Hypoxic Stress. *Cell Rep*. 20(4):895–908
59. Jones CG, Martynowycz MW, Hattne J, Fulton TJ, Stoltz BM, et al. 2018. The CryoEM Method MicroED as a Powerful Tool for Small Molecule Structure Determination. *ACS Cent Sci*. 4(11):1587–92
60. Jørgensen K, Bak S, Busk PK, Sørensen C, Olsen CE, et al. 2005. Cassava Plants with a Depleted Cyanogenic Glucoside Content in Leaves and Tubers. Distribution of Cyanogenic Glucosides, Their Site of Synthesis and Transport, and Blockage of the Biosynthesis by RNA Interference Technology. *Plant Physiol.* 139(1):363–74
61. Jørgensen K, Rasmussen AV, Morant M, Nielsen AH, Bjarnholt N, et al. 2005. Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Curr. Opin. Plant Biol.* 8(3):280–91
62. Kappers IF, Aharoni A, van Herpen TWJM, Luckerhoff LLP, Dicke M, Bouwmeester HJ. 2005. Genetic engineering of terpenoid metabolism attracts bodyguards to Arabidopsis. *Science* 309(5743):2070–72
63. 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 Res*. 45(W1):W55–63

64. Kenrick P, Crane PR. 1997. The origin and early evolution of plants on land. *Nature* 389:33–9

65. Kersten RD, Lee S, Fujita D, Pluskal T, Kram S, et al. 2017. A Red Algal Bourbonane Sesquiterpene Synthase Defined by Microgram-Scale NMR-Coupled Crystalline Sponge X-ray Diffraction Analysis. *J. Am. Chem. Soc.* 139(46):16838–44

66. Kersten RD, Weng J-K. 2018. Gene-guided discovery and engineering of branched cyclic peptides in plants. *PNAS* 115(46):E10961–69

66a. Keun HC, Beckonert O, Griffin JL, Richter C, Moskau D, Lindon JC, Nicholson JK. 2002. Cryogenic probe ^{13}C NMR spectroscopy of urine for metabonomic studies. *Anal. Chem.* 74:4588–93

67. Keymer A, Pimprikar P, Wewer V, Huber C, Brands M, et al. 2017. Lipid transfer from plants to arbuscular mycorrhiza fungi. *eLife* 6:e29107

68. King AJ, Brown GD, Gilday AD, Larson TR, Graham IA. 2014. Production of bioactive diterpenoids in the euphorbiaceae depends on evolutionarily conserved gene clusters. *Plant Cell*. 26(8):3286–98

68a. Kita T, Imai S, Sawada H, Kumagai H, Seto H. 2008. The biosynthetic pathway of curcuminoid in turmeric (Curcuma longa) as revealed by 13C-labeled precursors. *Biosci. Biotechnol. Biochem*. 72(7):1789–98.

69. Koehbach J, Attah AF, Berger A, Hellinger R, Kutchan TM, et al. 2013. Cyclotide discovery in Gentianales revisited—identification and characterization of cyclic cystine-knot peptides and their phylogenetic distribution in Rubiaceae plants. *Pept. Sci.* 100(5):438–52

70. Köksal M, Jin Y, Coates RM, Croteau R, Christianson DW. 2010. Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature* 469:116

70a. Kreuzaler F, Hahlbrock K. 1975. Enzymic Synthesis of an Aromatic Ring from Acetate Units. Partial Purification and Some Properties of Flavanone Synthase from Cell-Suspension Cultures of Petroselinum Hortense. *Eur J Biochem.* 56(1):205-13

71. Kroll K, Holland CK, Starks CM, Jez JM. 2017. Evolution of allosteric regulation in chorismate mutases from early plants. *Biochem. J.* 474(22):3705–17

72. Kusari S, Lamshöft M, Kusari P, Gottfried S, Zühlke S, et al. 2014. Endophytes are hidden

producers of maytansine in Putterlickia roots. *J. Nat. Prod.* 77(12):2577–84

73. Lallemand B, Erhardt M, Heitz T, Legrand M. 2013. Sporopollenin biosynthetic enzymes interact and constitute a metabolon localized to the endoplasmic reticulum of tapetum cells. *Plant Physiol*. 162(2):616–25
74. Lange BM, Fischedick JT, Lange MF, Srividya N, Šamec D, Poirier BC. 2017. Integrative Approaches for the Identification and Localization of Specialized Metabolites in Tripterygium Roots. *Plant Physiol*. 173(1):456–69
75. Laursen T, Borch J, Knudsen C, Bavishi K, Torta F, et al. 2016. Characterization of a dynamic metabolon producing the defense compound dhurrin in sorghum. *Science* 354(6314):890–93
76. Lau W, Sattely ES. 2015. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science* 349(6253):1224–28
77. Lee SC, Knowles TJ, Postis VLG, Jamshad M, Parslow RA, et al. 2016. A method for detergent-free isolation of membrane proteins in their local lipid environment. *Nat. Protoc.* 11(7):1149–62
78. Liang Y, Eudes A, Yogiswara S, Jing B, Benites VT, et al. 2019. A screening method to identify efficient sgRNAs in Arabidopsis, used in conjunction with cell-specific lignin reduction. *Biotechnol. Biofuels.* 12:130
79. Li B, Bhandari DR, Janfelt C, Römpp A, Spengler B. 2014. Natural products in Glycyrrhiza glabra (licorice) rhizome imaged at the cellular level by atmospheric pressure matrix-assisted laser desorption/ionization tandem mass spectrometry imaging. *Plant J*. 80(1):161–71
80. Li F-S, Weng J-K. 2017. Demystifying traditional herbal medicine with modern approach. *Nat. Plants* 3:17109

80a. Li F-S, Phyo P, Jacobowitz J, Hong M, Weng J-K. 2019. The molecular structure of plant sporopollenin. *Nat. Plants.* 5(1):41–46

81. Li H, Ban Z, Qin H, Ma L, King AJ, Wang G. 2015. A heteromeric membrane-bound prenyltransferase complex from hop catalyzes three sequential aromatic prenylations in the bitter acid pathway. *Plant Physiol*. 167(3):650–59
82. Li J, Schuman MC, Halitschke R, Li X, Guo H, et al. 2018. The decoration of specialized metabolites influences stylar development. *eLife* 7: e38611
83. Li M, Kildegaard KR, Chen Y, Rodriguez A, Borodina I, Nielsen J. 2015. De novo

production of resveratrol from glucose or ethanol by engineered Saccharomyces cerevisiae. *Metab. Eng.* 32:1–11

83a. Li-Beisson, Y et al. 2010. Acyl-Lipid Metabolism. *The Arabidopsis Book* 8:e0133.

83b. Lin X, Hezari M, Koepp AE, Floss HG, Croteau R. 1996. Mechanism of taxadiene synthase, a diterpene cyclase that catalyzes the first step of taxol biosynthesis in Pacific yew. *Biochemistry.* 35(9):2968–77

84. Lorenzo Tejedor M, Mizuno H, Tsuyama N, Harada T, Masujima T. 2012. In situ molecular analysis of plant tissues by live single-cell mass spectrometry. *Anal. Chem.* 84(12):5221–28

85. Luo S, Dong S-H. 2019. Recent Advances in the Discovery and Biosynthetic Study of Eukaryotic RiPP Natural Products. *Molecules* 24(8): E1541

86. Luo X, Reiter MA, d'Espaux L, Wong J, Denby CM, et al. 2019. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* 567(7746):123–26

87. Lu R, et al. 2003. Virus-induced gene silencing in plants. *Methods* 30(4):296–303

88. Lu Y, Stegemann S, Agrawal S, Karcher D, Ruf S, Bock R. 2017. Horizontal Transfer of a Synthetic Metabolic Pathway between Plant Species. *Curr. Biol.* 27(19):3034–41.e3

89. Matsuda Y, Mitsuhashi T, Lee S, Hoshino M, Mori T, et al. 2016. Astellifadiene: Structure Determination by NMR Spectroscopy and Crystalline Sponge Method, and Elucidation of its Biosynthesis. *Angew. Chem. Int. Ed. Engl.* 55(19):5785–88

90. Mayor S. 2011. Tree that provides paclitaxel is put on list of endangered species. *BMJ* 343:d7411

91. Mizuno H, Tsuyama N, Harada T, Masujima T. 2008. Live single-cell video-mass spectrometry for cellular and subcellular molecular detection and cell classification. *J. Mass Spectrom.* 43(12):1692–700.

92. Moore BM, Wang P, Fan P, Leong B, Schenck CA, et al. 2019. Robust predictions of specialized metabolism genes through machine learning. *PNAS* 116(6):2344–53

93. Newman DJ, Cragg GM, Snader KM. 2000. The influence of natural products upon drug discovery (Antiquity to late 1999). *Nat. Prod. Rep.* 17:215–234

94. Nguyen KNT, Nguyen GKT, Nguyen PQT, Ang KH, Dedon PC, Tam JP. 2016. Immunostimulating and Gram-negative-specific antibacterial cyclotides from the butterfly pea (Clitoria ternatea). *FEBS J.* 283(11):2067–90

95. Nicolaou KC, Yang Z, Liu JJ, Ueno H, Nantermet PG, et al. 1994. Total synthesis of taxol.

Nature 367(6464):630–34

96. Ning J, Moghe GD, Leong B, Kim J, Ofner I, et al. 2015. A Feedback-Insensitive Isopropylmalate Synthase Affects Acylsugar Composition in Cultivated and Wild Tomato. *Plant Physiol*. 169(3):1821–35
97. Nintemann SJ, Hunziker P, Andersen TG, Schulz A, Burow M, Halkier BA. 2018. Localization of the glucosinolate biosynthetic enzymes reveals distinct spatial patterns for the biosynthesis of indole and aliphatic glucosinolates. *Physiol. Plant.* 163(2):138–54
98. Nour-Eldin HH, Andersen TG, Burow M, Madsen SR, Jørgensen ME, et al. 2012. NRT/PTR transporters are essential for translocation of glucosinolate defence compounds to seeds. *Nature* 488(7412):531–34
99. O'Connor SE. 2015. Engineering of Secondary Metabolism. *Annu. Rev. Genet.* 49:71–94

99a. Olsson V, Joos L, Zhu S, Gevaert K, Butenko MA, De Smet I. 2019. Look Closely, the Beautiful May Be Small: Precursor-Derived Peptides in Plants. *Annu. Rev. Plant Biol*. 70:153–86

100. Onoyovwe A, Hagel JM, Chen X, Khan MF, Schriemer DC, Facchini PJ. 2013. Morphine biosynthesis in opium poppy involves two cell types: sieve elements and laticifers. *Plant Cell*. 25(10):4110–22
101. Ott A, Schnable JC, Yeh C-T, Wu L, Liu C, et al. 2018. Linked read technology for assembling large complex and polyploid genomes. *BMC Genom*. 19(1):651

101a. Pan, Q., Mustafa, N.R., Tang, K. et al. 2016. Monoterpenoid indole alkaloids biosynthesis and its regulation in Catharanthus roseus: a literature review from genes to metabolites *Phytochem. Rev* 15: 221-50

102. Pateraki I, Andersen-Ranberg J, Hamberger B, Heskes AM, Martens HJ, et al. 2014. Manoyl oxide (13R), the biosynthetic precursor of forskolin, is synthesized in specialized root cork cells in Coleus forskohlii. *Plant Physiol*. 164(3):1222–36
103. Phillipson JD, David Phillipson J. 2001. Phytochemistry and medicinal plants. *Phytochemistry*. 56(3):237–43
104. Pluskal T, Torrens-Spence MP, Fallon TR, De Abreu A, Shi CH, Weng J-K. 2019. The biosynthetic origin of psychoactive kavalactones in kava. *Nat. Plants* 5(8):867–78
105. Plys AJ, Kingston RE. 2018. Dynamic condensates activate transcription. *Science* 361(6400):329–30

106. Porto WF, Miranda VJ, Pinto MFS, Dohms SM, Franco OL. 2016. High-performance computational analysis and peptide screening from databases of cyclotides from poaceae. *Biopolymers* 106(1):109–18

107. Qu Y, Safonova O, De Luca V. 2019. Completion of the canonical pathway for assembly of anticancer drugs vincristine/vinblastine in Catharanthus roseus. *Plant J.* 97(2):257–66

107a. Quinn CM, Wang M, Polenova T. 2018. NMR of Macromolecular Assemblies and Machines at 1 GHz and Beyond: New Transformative Opportunities for Molecular Structural Biology. *Methods Mol. Biol.* 1688:1–35

108. Rahman A, Hallgrímsdóttir I, Eisen M, Pachter L. 2018. Association mapping from sequencing reads using k-mers. *eLife* 7:e32920

109. Rajniak J, Barco B, Clay NK, Sattely ES. 2015. A new cyanogenic metabolite in Arabidopsis required for inducible pathogen defence. *Nature* 525(7569):376–79

109a. Ralph J, Bunzel M, Marita JM, Hatfield RD, Lu F, et al. 2004. Peroxidase-dependent cross-linking reactions of p-hydroxycinnamates in plant cell walls. *Phytochem. Rev.* 3(1):79–96

109b. Recktenwald J, Shawky R, Puk O, Pfennig F, Keller U, et al. 2002. Nonribosomal biosynthesis of vancomycin-type antibiotics: a heptapeptide backbone and eight peptide synthetase modules. Microbiology. 148(Pt 4):1105–18

110. Reed J, Stephenson MJ, Miettinen K, Brouwer B, Leveau A, et al. 2017. A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metab. Eng.* 42:185–93

111. Rehm FBH, Jackson MA, De Geyter E, Yap K, Gilding EK, et al. 2019. Papain-like cysteine proteases prepare plant cyclic peptide precursors for cyclization. *PNAS* 116(16):7831–36

112. Rhoads A, Au KF. 2015. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics* 13(5):278–89

113. Riback JA, Katanski CD, Kear-Scott JL, Pilipenko EV, Rojek AE, et al. 2017. Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response. *Cell* 168(6):1028–40.e19

114. Ro D-K, Paradise EM, Ouellet M, Fisher KJ, Newman KL, et al. 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440(7086):940–43

114a. Rodriguez E, Aregullin M, Nishida T, Uehara S, Wrangham R, et al. 1985. Thiarubrine A, a bioactive constituent ofAspilia (Asteraceae) consumed by wild chimpanzees. *Experientia*. 41(3):419–20

114b. Rohmer M. 1999. The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep.* 16(5):565–74

114c. Saether O, Craik DJ, Campbell ID, Sletten K, Juul J, Norman DG. 1995. Elucidation of the primary and three-dimensional structure of the uterotonic polypeptide kalata B1. Biochemistry. 34(13):4147–58

115. Sainsbury F, Thuenemann EC, Lomonossoff GP. 2009. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* 7(7):682–93

116. Salim V, Yu F, Altarejos J, De Luca V. 2013. Virus-induced gene silencing identifies C atharanthus roseus 7-deoxyloganic acid-7-hydroxylase, a step in iridoid and monoterpene indole alkaloid biosynthesis. *Plant J*. 76(5):754–65

117. Samanani N, Alcantara J, Bourgault R. 2006. The role of phloem sieve elements and laticifers in the biosynthesis and accumulation of alkaloids in opium poppy†. *The Plant J*. 47(4):547–63

117a. Sapir-Mir M, Mett A, Belausov E, Tal-Meshulam S, Frydman A, et al. 2008. Peroxisomal Localization of Arabidopsis Isopentenyl Diphosphate Isomerases Suggests That Part of the Plant Isoprenoid Mevalonic Acid Pathway Is Compartmentalized to Peroxisomes. *Plant Physiol.* 148(3):1219-28

118. Schenck CA, Chen S, Siehl DL, Maeda HA. 2015. Non-plastidic, tyrosine-insensitive prephenate dehydrogenases from legumes. *Nat. Chem. Biol.* 11(1):52–57

119. Schläpfer P, Zhang P, Wang C, Kim T, Banf M, et al. 2017. Genome-Wide Prediction of Metabolic Enzymes, Pathways, and Gene Clusters in Plants. *Plant Physiol*. 173(4):2041–59

120. Shadle G, Chen F, Srinivasa Reddy MS, Jackson L, Nakashima J, Dixon RA. 2007. Down-regulation of hydroxycinnamoyl CoA: Shikimate hydroxycinnamoyl transferase in transgenic alfalfa affects lignification, development and forage quality

121. Shin Y, Brangwynne CP. 2017. Liquid phase condensation in cell physiology and disease. *Science* 357(6357): eaaf4382

121a. Shirley AM, McMichael CM, Chapple C. 2001. The sng2 mutant of Arabidopsis is

defective in the gene encoding the serine carboxypeptidase-like protein sinapoylglucose:choline sinapoyltransferase. *Plant J.* 28(1):83–94

122. Shulse CN, Cole BJ, Ciobanu D, Lin J, Yoshinaga Y, et al. 2019. High-Throughput Single-Cell Transcriptome Profiling of Plant Cell Types. *Cell Rep*. 27(7):2241–47.e4

123. Simmons BA, Loqué D, Ralph J. 2010. Advances in modifying lignin for enhanced biofuel production. *Curr. Opin. Plant Biol.* 13(3):313–20

124. Singatulina AS, Hamon L, Sukhanova MV, Desforges B, Joshi V, et al. 2019. PARP-1 Activation Directs FUS to DNA Damage Sites to Form PARG-Reversible Compartments Enriched in Damaged DNA. *Cell Rep*. 27(6):1809–21.e5

125. Sønderby IE, Geu-Flores F, Halkier BA. 2010. Biosynthesis of glucosinolates - gene discovery and beyond. *Trends Plant Sci*. 15(5):283–90

125a. Srinivasan, P., Smolke, C.D. 2019. Engineering a microbial biosynthesis platform for de novo production of tropane alkaloids. *Nat Commun*. 10:3634

126. Stegemann S, Bock R. 2009. Exchange of genetic material between cells in plant tissue grafts. *Science* 324(5927):649–51

127. Strauch RC, Svedin E, Dilkes B, Chapple C, Li X. 2015. Discovery of a novel amino acid racemase through exploration of natural variation in Arabidopsis thaliana. *PNAS* 112(37):11726–31

127a. Suzuki M, Xiang T, Ohyama K, Seki H, Saito K, et al. 2006. Lanosterol synthase in dicotyledonous plants. *Plant Cell Physiol.* 47(5):565–71

128. Takahashi K, Kozuka T, Anegawa A, Nagatani A, Mimura T. 2015. Development and Application of a High-Resolution Imaging Mass Spectrometer for the Study of Plant Tissues. *Plant Cell Physiol*. 56(7):1329–38

128a. Tholl D, Lee S. 2011. Terpene specialized metabolism in Arabidopsis thaliana. *The Arabidopsis Book*. 9:e0143

128b. Tivendale ND, Davidson SE, Davies NW, Smith JA, Dalmais M, et al. 2012. Biosynthesis of the halogenated auxin, 4-chloroindole-3-acetic acid. *Plant Physiol.* 159(3):1055–63

129. Tohge T, de Souza LP, Fernie AR. 2017. Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. *J. Exp. Bot.* 68(15):4013–28

129a. Tomita M, Okamoto Y, Kikuchi T, Osaki K, Nishikawa M, et al. 1971. Studies on the Alkaloids of Menispermaceous Plants. CCLIX. Alkaloids of Menispermum dauricum DC.

(Suppl. 7). Structures of Acutumine and Acutumidine, Chlorine-Containing Alkaloids with a Novel Skeleton. *Chemical & Pharmaceutical Bulletin.* 19(4):770–91

130. Töpfer N, Fuchs L-M, Aharoni A. 2017. The PhytoClust tool for metabolic gene clusters discovery in plant genomes. *Nucleic Acids Res*. 45(12):7049–63

131. Torrens-Spence MP, Fallon TR, Weng JK. 2016. A Workflow for Studying Specialized Metabolism in Nonmodel Eukaryotic Organisms. *Methods Enzymol*. 576:69–97

132. Torrens-Spence MP, Liu P, Ding H, Harich K, Gillaspy G, Li J. 2013. Biochemical evaluation of the decarboxylation and decarboxylation-deamination activities of plant aromatic amino acid decarboxylases. *J. Biol. Chem.* 288(4):2376–87

133. Torrens-Spence MP, Pluskal T, Li F-S, Carballo V, Weng J-K. 2018. Complete Pathway Elucidation and Heterologous Reconstitution of Rhodiola Salidroside Biosynthesis. *Mol. Plant.* 11(1):205–17

134. Tu Y. 2016. Artemisinin-A Gift from Traditional Chinese Medicine to the World (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* 55(35):10210–26

135. Tzin V, Malitsky S, Zvi MMB, Bedair M, Sumner L, et al. 2012. Expression of a bacterial feedback-insensitive 3-deoxy-d-arabino-heptulosonate 7-phosphate synthase of the shikimate pathway in Arabidopsis elucidates potential metabolic bottlenecks between primary and secondary metabolism. *New Phytol.* 194(2):430–9

135a. Vanholme, R., Sundin, L., Seetso, K.C. et al. 2019. COSY catalyses trans–cis isomerization and lactonization in the biosynthesis of coumarins. *Nat. Plants* 5:1066–75

136. Voelker SL, Lachenbruch B, Meinzer FC, Jourdes M, Ki C, et al. 2010. Antisense Down-Regulation of 4CL Expression Alters Lignification, Tree Growth, and Saccharification Potential of Field-Grown Poplar

137. Wall ME, Wani MC. 1995. Camptothecin and taxol: discovery to clinic—thirteenth Bruce F. Cain Memorial Award Lecture. *Cancer Res*. 55(4):753–60

138. Wang C, Liwei M, Park J-B, Jeong S-H, Wei G, et al. 2018. Microbial Platform for Terpenoid Production: Escherichia coli and Yeast. *Front Microbiol.* 9:2460

139. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, et al. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* 34(8):828–37

140. Wang P, Guo L, Jaini R, Klempien A, McCoy RM, et al. 2018. A 13C isotope labeling

method for the measurement of lignin metabolic flux in Arabidopsis stems. *Plant Methods*. 14:51

140a. Wang T, Park YB, Caporini MA, Rosay M, Zhong L, et al. 2013. Sensitivity-enhanced solid-state NMR detection of expansin's target in plant cell walls. *PNAS* 110(41):16444–49

141. Weidmann J, Craik DJ. 2016. Discovery, structure, function, and applications of cyclotides: circular proteins from plants. *J. Exp. Bot.* 67(16):4801–12

142. Weng J-K. 2014. The evolutionary paths towards complexity: a metabolic perspective. *New Phytol*. 201(4):1141–49

142a. Williams, M., & Molyneux, R. 1987. Occurrence, Concentration, and Toxicity of Pyrrolizidine Alkaloids in Crotalaria Seeds. *Weed Science* 35(4), 476–481

142b. Wink M. 2013. Evolution of secondary metabolites in legumes (Fabaceae). *S. Afr. J. Bot*. 89:164–75

143. Wu S, Schalk M, Clark A, Miles RB, Coates R, Chappell J. 2006. Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants. *Nat. Biotechnol.* 24(11):1441–47

143a. Xu M, Wilderman PR, Peters RJ. 2007. Following evolution's lead to a single residue switch for diterpene synthase product outcome. *PNAS* 104(18):7397–7401

143b. Xu R, Fazio GC, Matsuda SPT. 2004. On the origins of triterpenoid skeletal diversity. *Phytochemistry.* 65(3):261–91

144. Yahara S, Shigeyama C, Nohara T, Okuda H, Wakamatsu K, Yasuhara T. 1989. Structures of anti-ace and -renin peptides from lycii radicis cortex. *Tetrahedron Letters*.30(44):6041–2,

145. Yamamoto K, Takahashi K, Mizuno H, Anegawa A, Ishizaki K, et al. 2016. Cell-specific localization of alkaloids in Catharanthus roseus stem tissue measured with Imaging MS and Single-cell MS. *PNAS* 113(14):3891–96

146. Yanofsky C, Rachmeler M. 1958. The exclusion of free indole as an intermediate in the biosynthesis of tryptophan in Neurospora crassa. *Biochim. Biophys. Acta.* 28(3):640–41

146a. Yonekura-Sakakibara K, Tanaka Y, Fukuchi-Mizutani M, Fujiwara H, Fukui Y, et al. 2000. Molecular and Biochemical Characterization of a Novel Hydroxycinnamoyl-CoA: Anthocyanin 3-O-Glucoside-6″-O-Acyltransferase from Perilla frutescens. *Plant Cell Physiol.* 41(4):495–502

146b. Yu BW, Meng LH, Chen JY, Zhou TX, Cheng KF, et al. 2001. Cytotoxic oxoisoaporphine

alkaloids from Menispermum dauricum. *J. Nat. Prod.* 64(7):968–70

147. Zhang W, Liu H, Li X, Liu D, Dong X-T, et al. 2017. Production of naringenin from D-xylose with co-culture of E. coli and S. cerevisiae. *Eng. Life Sci.* 17(9):1021–29

148. Zhao EM, Suek N, Wilson MZ, Dine E, Pannucci NL, et al. 2019. Light-based control of metabolic flux through assembly of synthetic organelles. *Nat. Chem. Biol.* 15(6):589–97

Table 1 Summary of approaches described in Section 3

Approach	Advantages	Disadvantages	References
Co-expression	-Can be rapidly applied to nonmodel species	-Site of synthesis may differ from site of accumulation -Frequently depends on homology-based predictions	(20, 76, 57a)
QTL/GWAS	-Can be used to correlate unknown metabolites to genes in a homology-independent manner	-Requires large genomic resources	(33, 127)
Phylogenetic	-Identifies signatures of associated with the evolution of novel metabolic traits	-Phylogenies do not always accurately reflect evolutionary history	(104, 133, 50a)
Activity-guided	-Identifies enzymes in a homology independent manner	-Requires large amounts of starting material -Requires a robust activity assay	(24, 111)
Sequence-guided	-Can be used to predict chemistry based on growing genomic and transcriptomic resources	-Mostly limited to peptide natural products	(66, 69, 94)