Natural Product Reports

Plant-based engineering for production of high-valued natural products

Journal:	*Natural Product Reports*
Manuscript ID	NP-REV-03-2022-000017.R1
Article Type:	Review Article
Date Submitted by the Author:	10-May-2022
Complete List of Authors:	Sirirungruang, Sasilada; University of California Berkeley Markel, Kasey; University of California Berkeley Shih, Patrick; University of California Davis

Plant-based engineering for production of high-valued natural products

Sasilada Sirirungruang[1,2,3,4], Kasey Markel[1,2,3], Patrick M. Shih[1,2,3,5]

[1] Department of Plant and Microbial Biology, University of California, Berkeley, CA
[2] Feedstocks Division, Joint BioEnergy Institute, Emeryville, CA
[3] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA
[4] Center for Biomolecular Structure, Function and Application, Suranaree University of Technology, Nakhon Ratchasima, Thailand
[5] Innovative Genomics Institute, University of California, Berkeley, CA

Abstract

Plants are a unique source of complex specialized metabolites, many of which play significant roles in human society. In many cases, however, the availability of these metabolites from naturally occurring sources fails to meet current demands. Thus, there is much interest in expanding the production capacity of target plant molecules. Traditionally, plant breeding, chemical synthesis, and microbial fermentation are considered the primary routes towards large scale production of natural products. Here, we explore the advances, challenges, and future of plant engineering as a complementary path. Although plants are an integral part of our food and agricultural systems and sustain an extensive array of chemical constituents, their complex genetics and physiology have prevented the optimal exploitation of plants as a production chassis.. We highlight emerging engineering tools and scientific advances developed in recent years that have improved the prospects of using plants as a sustainable and scalable production platform. We also discuss technological limitations and overall economic outlook of plant-based production of natural products.

Table of content

1. Introduction

Plant specialized metabolites are structurally complex and diverse, and many are of high value to society as flavors, fragrances, nutraceuticals, and pharmaceuticals (**Figure 1; Table 1**).[1–4] As many plant natural products are biosynthesized at low levels by native producing species and natural supplies are limited, there is a strong interest to expand their production. Selective breeding programs and wide-spread cultivation are the traditional approaches to increasing supply although these avenues can encounter various limitations. For example, many plants of interest are slow-growing or lack the genetic resources required for modern breeding approaches. Chemical synthesis can provide an alternative approach to bulk production of high-value natural products. However, this approach becomes expensive for complex metabolites, particularly those including many fused rings and stereocenters.[5] With the relatively recent rise of biotechnology, microbial fermentation of natural products has gathered much attention.[6,7] This process uses engineered unicellular organisms that express the enzymatic pathways of interest to assemble target molecules, their precursors, and their new-to-nature analogues. However, after many years, few processes have become economically viable, in part due to the relatively low titers achieved by cross-kingdom metabolic engineering and the higher costs associated with microbial fermentation compared to agriculture. Ultimately, in many cases, field production of native-producer plants remains the principal commercial source of plant natural products decades after their discovery, such as opium poppy for opiates and cannabis for cannabinoids.

Recently, plant synthetic biology has seen many technological advances that have expanded the phenotypic opportunity space of engineered plants. Although the field is still in its infancy, an argument can be made for engineering whole plants for large-scale production of plant natural products. Synthetic biology applies engineering principles – such as modularity and standardization – to biological systems in pursuit of the extraordinary gains in efficiency seen in mechanical and electrical engineering after the adoption of standards in those disciplines. It aims to advance the ability to design, construct, and characterize biological complexity in a predictable and consistent manner through the iterative design-build-test-learn cycle.[8] While most advances in synthetic biology have been focused on microbes, plant synthetic biology offers unique opportunities complementary to microbial engineering.

In this article, we explore the possibility of using whole plants as hosts for large scale production of high-value plant specialized metabolites. Realization of the full potential of plants as hosts for effective production of selected target molecules requires a detailed understanding of their complexity at the organism level as well as an appropriate toolbox to engineer them. We seek to demonstrate that, with a better understanding of plant physiology and metabolism and an increasing number of plant synthetic biology tools becoming available, whole plants could serve as viable production hosts for high-value natural products.

1.1 Whole plants as a chassis for production of natural products

Whole plants may serve as ideal production chassis for plant natural products because they are naturally metabolically and physiologically adapted for biosynthesis of complex molecules. Recent estimates suggest that between 200,000 and 1 million metabolites are found in the plant kingdom.[9] These include many broad classes of compounds such as sugars, fatty acids, phenylpropanoids, polyketides, and terpenes. In addition to the large number of compound families, much of plant chemical diversity is also derived from variations of substituents on the backbone carbon scaffolds. Several plant metabolites have been explored for their bioactivity; however, the number is minuscule relative to the full expanse of

chemical diversity of the kingdom. Thus, plants intrinsically harbor a complex metabolite pool and represent a rich source of molecular structures to be explored and engineered for new functions.

Most high-value molecules from plants are produced through processes known collectively as specialized metabolism, by complex enzymatic machinery native to plants. Biosynthetic enzyme families that are highly expanded in the plant kingdom include glycosyltransferases, terpene synthases, and redox enzymes, many of which are rarely found in bacteria and fungi.[10] Although some biosynthetic enzyme families such as phenylalanine ammonia lyases and chalcone synthases are found in virtually all land plants, others (e.g., flavone synthase I) are specific to certain phylogenetic clades, enabling the enormous chemical diversity within the kingdom.[11,12] The large repertoire of plant biosynthetic enzymes suggests that plants harbor cellular or subcellular environments compatible with all their functions. This capacity offers many opportunities for and may simplify engineering.

Plant cells have unique features that do not exist in microbial cells, which enable the compartmentalization of various biosynthetic enzymes/pathways. Compartmentalization offers a tailored microenvironment for pathway enzymes and sequesters the effects of bioactive metabolites on the producing cells. Plastids, vacuole, and ER bodies are examples of subcellular structures that take part in secondary metabolism. Subcellular compartments may also allow orthogonal control and utilization of separate metabolite pools. For example, terpenoid biosynthesis in plants occurs in two compartments by two largely independent pathways: the mevalonate pathway occurring in the cytosol and the methylerythritol phosphate pathway occurring in the plastid. Both pathways make the same isoprenoid precursor isopentenyl diphosphate but are separately regulated and specialize in subsets of terpenes.[13,14]

Moreover, many plants also contain specialized organs that are devoted to biosynthesis, storage, and/or secretion of secondary metabolites. One major site of secondary metabolism is the trichomes, which are epidermal structures found in the aerial tissues of a wide variety of plant species.[15,16] They can be unicellular and simply act as storage units, or multicellular and represent dedicated cell factories to produce specialized metabolites. Plants have evolved many other specialized tissues for accumulation of natural products, including S-cells and myrosin cells,[17] laticifers,[18] idioblasts,[19] oil cells,[20] and traumatic resin ducts.[21] Many dedicated organs and cellular structures are equipped with transport and storage mechanisms that allow specialized metabolites to accumulate to very high levels, in some cases exceeding 25% dry weight (DW).[22] Metabolites contained in these structures can condense into liquid droplets, further sequestering them from the general cytoplasmic milleiou and mitigating toxicity.[22,23]

Plants are cultivated and consumed at gigaton scale worldwide; fruits, vegetables, herbs, spices, and medicinal plants are an integral part of many cultures and traditional knowledge on ways to cultivate them for optimal production output exists for various climates and terrains. Thus, engineered plants would be able to be directly plugged into current agricultural infrastructure and frameworks compared to heterologous production of microbial fermentation or chemical synthesis factories.

1.2 The biotechnology journey of artemisinin

There are few plant natural products that have traversed the worlds of plant cultivation and engineering to large-scale production in heterologous microbial hosts. As one of the few case studies, we briefly outline the history of artemisinin production to highlight the potential and challenges of synthetic biology applied

to increasing access to plant natural products. Artemisinin is a natural sesquiterpene lactone found in *Artemisia annua* (sweet wormwood) plants and an effective agent against malaria. In 2002, the World Health Organization (WHO) recommended artemisinin-based combination therapy (ACT), the administration of artemisinin with another effective antimalarial drug, as the first-line treatment for uncomplicated malaria. After 20 years, ACT is still the first-line treatment of malaria infection in most countries due to its high efficacy.[24]

Following the WHO recommendation, demands for artemisinin skyrocketed. Artemisinin is traditionally harvested from the leaf tissue of *A. annua*, which generally produces the molecule at around 1% of its dry DW.[25,26] Thus, the supply of artemisinin is scarce and at times cannot meet global demand, leading to large price fluctuation from year to year.[27] The price fluctuations in turn reduce incentives for farmers who rely on steady incomes to grow more *A. annua* and prevent effective planning in the international fight against malaria. In addition to the low concentration, *A. annua* is relatively slow growing and self-incompatible, which has prevented the development of consistently yielding inbred lines.[26] These shortcomings have fueled efforts to find complementary paths to large quantities of cost competitive artemisinin. Many chemical syntheses of artemisinin had been reported over the years.[28,29] Most synthetic routes, however, involve long reaction sequences due to the structural complexity of the target molecule, resulting in high costs and low yields. Although these syntheses led to various artemisinin analogues, many of which have shown promising antimalarial activity, none led to the reliable supply of a low-cost malarial treatment the world needs.

The discovery of amorphadiene synthase, a key artemisinin biosynthetic enzyme, in 2002 offered the opportunity to engineer artemisinin precursor biosynthesis in a heterologous host.[30] In what became known as the semi-synthetic artemisinin project, the enzyme was overexpressed in *Escherichia coli* along with the mevalonate pathway to produce amorpha-4,11-diene.[31] However, *E. coli* was unable to optimally support the function of the following step in the biosynthesis, which involves cytochrome P450 monooxygenase CYP71AV1. Thus, a late-stage intermediate, artemisinic acid, was produced by engineered yeast strains expressing amorphadiene synthase and CYP71AV1.[32] Through various optimization steps, yields of up to 25 g/L were achieved.[33] Artemisinic acid could then be chemically converted to artemisinin.[34]

The semi-synthetic artemisinin from microbial fermentation was intended to supplement plant-sourced artemisinin in meeting the global demands. The high-yielding engineered yeast production platform was scaled-up,[35] and the semi-synthetic artemisinin hit the market in 2014. Although it was originally considered a success story for synthetic biology, and many hoped it would offer an adequate, reliable supply of the much-needed treatment, large scale production of semi-synthetic artemisinin stopped just after one year due to business considerations, despite continued technological advances that increased the yield and lowered the costs of semi-synthetic artemisinin production.[35,36]

In parallel, artemisinin biosynthesis has also been engineered in plant-based systems including in the model plant *Nicotiana benthamiana*, in cell suspension culture of *A. annua*, and in whole *A. annua* plants. The plant-based heterologous approaches were able to achieve direct production of artemisinin, rather than its precursors; however, yields in these systems are much lower than in *A. annua* plant at up to 0.68% DW of *N. benthamiana* leaves[37] and 0.11g/L in cell culture.[38] Metabolic engineering of *A. annua*

plants has been extensively explored,[39] and production yield has been increased to 3% DW,[40] higher than the yields achieved by breeding.

Despite engineering the non-model organism *A. annua* being more challenging than *N. benthamiana, S. cerevisiae, and E. coli* , transgenic *A. annua* lines so far provide the highest yield of the desired product, and global supply currently remains mainly from *A. annua* cultivation. It remains possible that further research and development will further decrease the costs of the microbial fermentation process making semi-synthetic artemisinin economically competitive, especially in the case of crop disruptions due to natural disaster or geopolitical events. Nevertheless, the journey of semi-synthetic artemisinin begs a reconsideration of optimal commercial scale production host for plant natural products and demonstrates that whole plants are difficult to outperform for production at scale. As the choice of production route affects production outcomes, it demands a careful consideration on a molecule-by-molecule basis.

2. Advances in understanding of plant metabolic organization

The past decade saw a boom in the discovery of plant natural products biosynthetic genes and pathways (Figure 2). These discoveries present the opportunity to understand and engineer plant metabolite production. With gene players known, synthetic biology can be applied to export pathways for heterologous production, alter enzyme specificity to produce analogues, and overexpress genes to increase yields, for example. In addition, the discovery of regulatory mechanisms of pathway gene expression patterns may also allow us to exploit the natural process to gain control of production dynamics.

An increase in the number of characterized natural product pathways provides the foundation to a better understanding of the function and organization of their genes and enzymes (Figure 3). Such insight offers a glimpse into the basis of efficiency achieved by nature that so far has proven difficult to replicate. As such, metabolic engineering efforts may benefit from nature’s guide to optimizing production yield. Here, we describe the organizational themes found in plant natural product biosynthetic modules and their relevance in engineering.

2.1 Biosynthetic gene and pathway discovery

Identification of the enzymes necessary to produce a target molecule is the initial step in engineering the production of high-value natural products. In recent years, pathway discovery has been greatly accelerated by the decreasing costs of sequencing technologies.[41–44] As few plant biosynthetic enzymes are experimentally characterized and as plant genomes often contain many genes of the same enzyme family, it is often not possible to deduce enzyme functions from primary sequence information alone. Thus, knowledge on the gene expression pattern is valuable, based on the rationale that enzymes functioning in a common biosynthetic pathway often share a common expression pattern. Moreover, comparison of candidate genes to their homologs or considerations of their synteny in other related species that also produce or lack the target molecule has proven to be informative.[32,45,46] Indeed, many recent successful discoveries have relied on the analysis of tissue-specific transcriptomes of the producing species and comparative genomics with related non-producing or producing species.[47–51]

With an expanding number of pathways characterized, a picture emerges that a relatively small number of enzyme families make up most plant biosynthetic pathways.[10] These families include glycosyltransferases, methyltransferases, ketosynthases, terpene synthases, and redox enzymes. Often, a pathway consists of a major synthase responsible for the formation of the main scaffold, accompanied by various tailoring enzymes that then add or alter substituents.[52] These findings are consistent with the plant natural product landscape that much of the diversity arises from variations of the substituents on the main carbon scaffolds. In fact, expansion of tailoring enzymes in a lineage-specific manner due to tandem gene duplication and neofunctionalization are thought to contribute to plant chemical diversity. For example, BAHD acetyltransferase has been found greatly expanded in Populus and linked to the diversity of phenylpropanoids of the family.[53] For engineering purposes, functional characterization of these specific classes of enzymes will enable future efforts to produce a range of high-value plant natural products.

These themes, along with the increasing number of characterized natural product biosynthetic gene clusters, were central to tools invented to assist plant biosynthetic pathway discovery such as PlantiSMASH,[10] Phytoclust,[54] and PlantClusterFinder.[52] The platforms consider the combination of enzyme functions, genomic loci, and when available, gene expression data to allow for optimal accuracy. These tools have already facilitated identification of key enzymes in many metabolic pathways, laying the groundwork for metabolic engineering efforts.[45,55]

Another notable finding from recent discoveries is that pathways can share enzymes.[56] In fact, evidence is emerging that plant biosynthetic pathways form co-expression networks, with individual enzymes as links among them.[57] This finding may be worth considering in engineering: overexpression, deletion, or heterologous expression of a pathway may interfere with or be interfered with by off-pathway metabolism. Indeed, it has been shown many times that target molecules or pathway intermediates can get glycosylated by background glycosyltransferases.[58,59] Orthogonally functioning modules are usually valued in synthetic biology but may require careful attention to achieve in the complex metabolic network of plants.

In addition to enzymes, knowledge on the regulatory inner workings of a biosynthetic pathway is highly valuable for engineering the native producer. Secondary metabolic pathways are thought to be controlled mainly at the transcription level by transcription factors in the MYB, bHLH, WRKY, AP2/ERF, bZIP, and NAC families and many were discovered in the past decade.[60]. Engineering the expression level of relevant transcription factors have shown to improve yields of vinca alkaloids,[61,62] and artemisinin[63], among others. Altering the regulatory mechanism of a pathway may also increase product yield by relieving negative feedback mechanisms, uncoupling production from necessary natural signals, or expanding the tissue types in which the target molecule is produced, without the knowledge of the biosynthetic enzymes.[64]

2.2 Biosynthetic gene clusters in plants

Recent increases in available plant genomes allow characterized biosynthetic genes to be mapped to their chromosomal locations. Although plant biosynthetic pathway genes are not organized into operons as their bacterial counterparts, they are often found in loose clusters.[45,65–68] Clusters vary in size from a few to hundreds of kilobases.[57] The current working definition of a cluster is the presence of at least three enzymes from different enzyme families in genomic proximity.[57]

Because plants are primarily restricted to vertical genetic transfer, cluster formations are thought to be driven by negative selection against incomplete cluster inheritance that results in the loss of metabolite production and/or buildup of toxic intermediates.[69] However, often not all pathway enzymes are clustered. In other words, clusters are not required. This suggests that plant biosynthetic gene clusters may be flexible to lose and regain enzyme functions from other parts of the genome in inheritance. This balance between maintaining function and evolving novel pathway functionality helps explain the chemical diversity of plants.

Discoveries of phylogenetically related biosynthetic gene clusters provide a better opportunity to understand how clusters emerge and evolve. Orthologous clusters seem to utilize different members of the same gene families to elaborate related metabolites.[66,70] These enzymes seem to have been "selected" from a pool of enzymes by nature. This combinatorial picture of many biosynthetic gene clusters suggests that enzymes could be mixed and matched based on their family and function to produce analogous products. In fact, this principle had been applied to synthesize a suite of new-to-nature products.[71,72] On the other hand, analogous clusters that produce the same molecules but evolved independently were also found.[73,74] This shows that plants often find multiple ways to produce the same molecules and suggests that in some cases, many pathway options may be available to choose from for production of a certain product.

2.3 Metabolons in natural product biosynthetic pathways

As natural products are formed by the actions of a long cascade of enzymes, these enzymes must work together to ensure successful formation of their target molecule in a precise and reliable manner without interference with other cellular processes. Thus, enzymes in the same pathway often co-localize with one another (Figure 3).[75–77] Co-localization of cascading enzymes provides a microenvironment optimal for their functions. It serves to increase the effective concentration of enzymes and substrates and allows channeling of intermediates between sequential enzymes, reducing transit time and preventing unfavorable crosstalk. Fitting for the branching nature of plant secondary metabolite production, it also allows for enzymes to exchange interaction partners depending on the microenvironment.[75] Co-localization may also simplify regulation, as all enzyme players experience similar changes in the environment and can respond in a concerted manner.

Modes of interactions include formation of multi-enzyme complexes, co-encapsulation within membrane-bound compartments, and formation into protein-based microcompartments. Metabalons are one form of enzyme co-localization observed in plant natural product biosynthetic pathways, in which transient protein complexes form from cascading enzymes of a pathway.[78] Metabolons in secondary metabolism are thought to be anchored to the cytosolic side of the ER membrane by a cytochrome P450 enzyme and its redox partner, which are membrane proteins.[76,79] Other pathway proteins that are otherwise soluble then interact with the P450 to form a complex. The interactions between natural product biosynthetic enzymes suggest that co-localization gives advantage to these pathways. As such, protein-protein interactions may be engineered to increase efficiency of biosynthetic pathways.[80] This concept has been utilized to increase production yield of dhurrin[81] and taxol precursors[82] in *N. benthamiana*.

Furthermore, metabolons suggest that pathway enzymes from the same organism have been evolutionarily optimized for their collective function, not just the function of the individual enzymes. This

concept may be of consideration in engineering. Protein complex formation may be an advantage of utilizing all enzymes from the same organism, instead of mixing and matching enzymes from different origins. It also suggests that specific molar ratios of pathway enzymes may be required for optimal throughput. Moreover, metabolon formation may require specific physiological features of the cell of native producer that may or may not exist in a heterologous host.

On the other hand, metabolon formation might present an obstacle to engineering. Metabolons suggest that enzymes that otherwise appear cytosolic may in fact have preferred binding partners.[83] Thus, attempts to repurpose enzymes in a synthetic pathway may be affected by their association with native partners. Furthermore, biosynthetic enzyme complexes may be subject to undesired regulatory mechanisms, which may be sidestepped by substituting with enzymes from another species.

3. Production of high-value metabolites in heterologous plant hosts

One route to produce high-value plant natural products is to express their biosynthetic pathways in a heterologous plant host. Heterologous production in model plants benefits from established methods of cultivation, pathway construction, transformation, and product extraction. In addition, a heterologous host uncouples production from geographical and legal considerations that may be tied to native producers.[84] For example, some native producers can only be grown under certain climate conditions or are protected due to their endangered status. Heterologous production may also be especially suitable for metabolites whose native producers grow slowly, have long regeneration time, require a very particular climate to thrive, or are protected or endangered. **Figure 2** summarizes some axes of considerations, which may lend production of natural products favorable in native host species or by heterologous expression.

To perform heterologous expression, the knowledge of all enzyme activities involved is required. Historically, the identity of biosynthetic enzymes has been the major barrier to engineering specialized metabolite production. However, as more biosynthetic pathways are discovered and more biosynthetic enzymes are characterized in detail, an increasing number of products will be applicable to heterologous production. In addition, in cases where not all enzymes from the native producer have been identified, enzymes from different organisms may substitute for the required transformation to yield the target product.

3.1 Transient expression in *Nicotiana benthamiana*

The most used model plant for pathway overexpression is *Nicotiana benthamiana*. *N. benthamiana* grows rapidly, has well-characterized metabolism, is readily transformable, and amenable to laboratory experiments, including transient expression experiments. Transient expression experiments have a much shorter experimental time frame than constructing stable transformants. It allows faster iterations through various pathway designs and provides potential access to products that are urgently needed. In fact, transient expression in *N. benthamiana* is the most widely used method to characterize genetic parts and enzyme functions. Thus, natural product production in *N. benthamiana* benefits from a large array of well-characterized tools that allow sophisticated pathway designs and potentially lead to more reliable outcomes. Here, we highlight some recent transient expression studies that successfully increased production yields by incorporating enzyme targeting, metabolon formation, or synthetic organelle formation into their pathway designs.

N. benthamiana allows experimentation with engineering in different cellular compartments. Plants synthesize terpenoids in both the cytosol and the chloroplast, allowing engineered terpenoid biosynthesis to be targeted to either compartment. Wu *et al.* were the first to increase production titer of sesquiterpene patchoulolfrom 0.01 to 0.03% DW when they targeted patchoulol synthase and farnesyl diphosphate synthase to the chloroplast.[85] Similarly, Dong *et al.* found that the production yield of monoterpene geraniol was higher when biosynthetic enzymes were targeted to the chloroplast than to the cytosol and mitochondria.[59] In producing diterpene taxadiene-5α-ol, Li *et el.* found that while taxadiene synthase localized to the chloroplast, taxadiene-5α-hydroxylase localized to the cytosolic side of the ER membrane and thus could not work together to make the product.[82] Rerouting the enzymes, the authors found that chloroplast targeting yielded 4.5-fold higher product than cytosolic targeting.[82] On the other hand, De la Pena and Sattely were able to elucidate the biosynthesis of and produce momilactone B by rerouting diterpene synthases OsCPS4 and OsKSL4 from the chloroplast to the cytosol, after their initial attempt was unsuccessful with plastidic expression of the enzymes.[86] These studies emphasize the importance of enzyme targeting and co-localization and suggest that different pathways may require different subcellular environments to achieve optimal yield. Thus, the ability to target pathways to the cytosol, chloroplast, mitochondria, or other subcellular compartments in *N. benthamiana* offers a unique advantage and opportunity to increase product yield.

Synthetic metabolon was demonstrated in *N. benthamiana* to increase the production of dhurrin by 5-fold. Gnanasekaran *et al.* targeted the dhurrin biosynthetic enzymes, two membrane-bound cytochrome P450 enzymes and a soluble glucosyltransferase, to the chloroplast with the rationale that ferredoxin, reduced through the photosynthetic electron transport chain, could efficiently donate electrons to the P450s.[87] However, in doing so, the dhurrin biosynthetic metabolon could not properly form. De Jesus *et al.* improved pathway efficiency by constructing synthetic metabolons on the thylakoid membrane. The authors fused dhurrin biosynthetic enzymes to TatB and TatC proteins, which form complexes comprising several copies of TatBC. Fusion enzymes improved dhurrin yield from 0.1% to 0.5% fresh weight and decreased off-pathway intermediates.[81] Synthetic metabolons have not been utilized widely for engineering biosynthetic pathways in plants, but this example demonstrates that it can be an effective strategy to improve production yield.

Synthetic organelles may also be constructed in conjunction with biosynthetic pathway overexpression to sequester products, especially volatile organics. To demonstrate this concept, Delatte *et al*. co-expressed acylglycerol acyltransferase, transcription factor WRINKL1, and oil body interacting protein oleosin, which together promote the formation of liquid droplets, with α-bisabolol synthase and truncated 3-hydroxyl-3-methyl-glutaryl-coenzyme A reductase (tHMGR).[88] The authors observed an increase of α-bisabolol production from 0.035% to 0.07% DW 7 days post infiltration when liquid droplets were formed. Localization of α-bisabolol to the lipid body was confirmed by microscopy. A similar 2-fold increase in production yield was observed for sesquiterpenes (E)-b-caryophyllene and α-barbatene. Similarly, Sadre *et al.* were able to sequester abietadiene to the liquid droplet fraction by co-expressing a microalgal lipid droplet surface protein with biosynthetic pathway of the diterpenoid product although total production yield did not change.[89]

In addition, transient expression in *N. benthamiana* is highly compatible with combinatorial biosynthesis. The platform enables co-expression of genes by co-infiltrating different combinations of agrobacterium strains without having to make multi-gene constructs. Reed *et al.* used the transient expression platform to

biosynthesize 41 triterpenes by co-infiltrating β-amyrin synthase and tHMGR with a single or pair-wise combinations of five cytochrome P450 enzymes.[90] Davis *et al*. produced halogenated tryptophan and auxin analogues by infiltrating *N. benthamiana* with bacterial tryptophan halogenases or co-infiltrating with bacterial tryptophan halogenases, aminotransferase, and flavin-dependent monooxygenase respectively.[91] Andersen-Ranberg *et al.* co-infiltrated combinations of ten class-I and ten class-II diterpene synthases and produced 57 labdane- and clerodane-type skeletons, most of which were new-to-nature.[71] These studies demonstrate the power of leveraging transient expression in *N. benthamiana.*

Some products produced through transient expression in *N. benthamiana* have been scaled up.[90,92] However, product yield achieved through this method is generally low. This is presumably because the physiological and chemical background of *N. benthamiana* may not match that optimal for the target molecule production. High titers may be hampered by the limited availability of precursors, the instability of pathway enzymes in the heterologous host, imbalanced metabolic fluxes, or the toxicity of intermediates and products. Moreover, engineered pathways often consider only biosynthetic enzyme functions and expression levels. Studies highlighted above are among few examples of when cellular physiology was utilized to benefit production.[81,88,89] Thus, considering localization, transport, storage, and detoxification of the product compound at the cellular and tissue levels may provide a path to increasing production yield. In this regard, further studies on the physiology and biochemical makeup of *N. benthamiana* leaf mesophyll cells may allow for more optimal utilization of the platform for production.

3.2 Other potential heterologous host plants

Production in *N. benthamiana* may be difficult to scale up beyond laboratory scale as the plant is small, fragile, and not agriculturally grown. In this regard, other well-characterized heterologous hosts that may be suitable for heterologous production of high-value metabolites are *Nicotiana tabacum* and *Physcomitrella patens*. Unlike N. benthamiana,, engineering in these hosts is generally accomplished through the generation of stable lines rather than transient expression. Multi-step pathway engineering in stable lines is more difficult than transient expression of multiple enzymatic steps by separate strains of agrobacterium, but results in transgenic plants more suitable for field-scale production.

N. tabacum may be a good alternative host for field level production as it grows larger than *N. benthamiana* and is still as easily transformable. It also has a long history of large-scale cultivation, which enables high-quality estimates of biomass yield in a particular region.[93,94] This allows for calculation of the "production parity titer", the point at which a hectare of land can produce the same amount of target molecule per year as growth of the native host. Since *N. tabacum* is widely cultivated, it will likely be easier and cheaper to grow than most native-producer plants. Thus, any stable transgenic lines achieving the production parity titer are likely to be economically viable to grow in conjunction with production in the native host. At substantially higher titers, transgenic hosts may replace native host production. **Table 1** shows approximate production parity titers for several plant natural products. Due to compounding uncertainties in yield per unit biomass and biomass harvested per hectare, the values given are estimates, and in most cases averaged from many sources. **Figure 1** plots selected data from **Table 1**, drawing particular attention to the relationship between concentration in the native host, productivity of native host, and cost per gram of target molecule.

The close evolutionary relationship between *N. benthamiana* and *N. tabacum* enable pathway constructs to be largely optimized and interchangeable between the two species. Thus, transient expression experiments in *N. benthamiana* can be used for prototyping, followed by subsequent transformation into *N. tabacum* for field scale production. Similar to in *N. benthamiana*, ease of experimentation with *N. tabacum* allows experimenting with different strategies to increase production yield, some of which are highlighted below.

Recent examples of high-yielding pathway design include production of proanthocyanidins and dhurrin. By overexpressing foreign enzymes and regulatory elements, metabolites undetectable in the wild-type plants could be generated to over 0.1% DW. Fresquet-Corrales *et al.* were able to produce up to 0.348% DW of total proanthocyanidins, up from undetectable level in wild-type plant, by expressing two transcription factors from *Antirrhinum majus* that upregulate the anthocyanin pathway, and anthocyanidin reductase and leucoanthocyanidin reductase from of *Medicago truncatula*.[95] Similarly, Gnanasekaran *et al*. produced dhurrin at yields 0.1-0.2% DW by integrating dhurrin biosynthetic pathway to the chloroplast genome.[87]

N. tabacum is also amenable to compartmentalization strategies to improve titers. By inducing synthetic droplets in the chloroplasts, Zhao *et al.* increased the production of squalene to 0.26% fresh weight. The hydrophobic domain of oleosin was co-expressed with pathway enzymes, and all proteins were targeted to the chloroplast using transit peptide. Raman scattering microscopy showed presence of lipid droplets in the chloroplasts when the hydrophobic domain of oleosin was co-expressed.[96] In another example, Zhang *et al.* were able to increase the titer of diterpenoid cembratrien-ol when driving the expression of cembratrien-ol synthase[97] or 1-Deoxy-D-xylulose 5-phosphate reductoisomerase[98] with a trichome-specific promoter. Notably, a trichome-specific promoter yielded a higher amount of product than the CaMV 35S promoter and was associated with an increase in the diameter of trichome secretory cells and upregulation of the MEP pathway.[97,98]

Beside *N. tabacum*, the moss *Physcomitrella patens* has recently gathered much attention as an alternative metabolic engineering host as it is easily transformable, amenable to genome editing, and has a fast growth rate. *P. patens* is haploid in its dominant growth phase and has an outstandingly high rate of homologous recombination, making it suitable for precise gene targeting.[99] Studies related to *P. patens* are also facilitated by its completely sequenced genome and saturated mutant collection. Various promoters have been developed for gene expression in *P. patens*, including chemical and light inducible ones.[100,101,102] Moreover, King *et al.* reported transformation assisted assembly in *P. patens* protoplasts that let to the desired phenotypes at 10-22% efficiency after selection.[103] As such, *P. patens* has been engineered to produce plant natural products such as artemisinin at yields up to 0.02% DW,[104] patchoulol at 0.003% DW,[105] and sclareol at 0.28% DW.[106] However, *P. patens* are grown in photobioreactors, which may present a challenge when scaling. Although as differentiated tissues, *P. patens* are more stable to maintain than plant suspension cell cultures, their photobioreactor costs may be more expensive than those associated with microbial fermentation.

The underlying causes of differing levels of efficiency for the production of heterologous metabolites between differing hosts are poorly understood. Among many possibilities are differing levels of initial substrate availability and differing levels of tolerance to toxic intermediates. As more metabolomics data becomes available for bulk plant tissues, it may be possible to determine which host plants have the most

initial substrate available, one key consideration in host plant choice. The question of which host plantswill best tolerate toxic intermediates and products is less amenable to advance prediction, but one may consider a detoxification/sequestration mechanism as an integral part of production pathway design to mitigate such issue.

Host plants used for heterologous expression of metabolic pathways tend to be selected from the pool of well-understood model organisms. As a consequence, host plants are all relatively easy to grow and have established methodologies for growth at large scale, with the partial exception of *P. patens.* Using data on known agricultural yields, estimating the production parity titer should be used when choosing appropriate host systems to engineer.

3.3 Production of taxadiene in heterologous plant hosts

The major contributing factors dictating yield in each heterologous plant host are not well understood, making it challenging to determine the "optimal" host ahead of experiments. To compare the metabolic potential of heterologous plant hosts, we look at the production of a taxol precursor, taxadiene, as a case study. Taxol is an effective anti-cancer agent found in the bark of yew trees. Due to the low titer found in yew bark (0.015% DW) and the slow-growing and sparsely distributed nature of the species, heterologous production of taxol has been explored extensively. Taxadiene is the first committed intermediate in taxol synthesis and is cyclized from geranylgeranyl diphosphate by taxadiene synthase (TS). While taxadiene production has been achieved in microbes *E. coli* and *S. cerevisiae*, downstream transformation of taxadiene to taxol requires cytochrome P450 enzymes, which have been shown to be challenging to reconstitute in microbes. Thus, plant-based production is favored. By directly comparing yields, we compare the differences in the production potential of five transgenic species for taxadiene.

Many attempts have been made to produce taxadiene in *N. benthamiana.* Hasan *et al*. introduced TS into *N. benthamiana* under the control of the CaMV 35S promoter and produced taxadiene at 0.0025% DW in a homozygous line. Silencing phytoene desaturase, which competes for geranylgeranyl diphosphate substrate further increased the yield to 0.0048% DW, but the silenced plant suffered a growth defect.[107] Later on, Fu *et al.* compared taxadiene yield from *N. benthamiana* stable lines expressing TS in the cytosol, in the chloroplast from chloroplast genome, in both compartments, and in the chloroplast from nuclear genome, and found them to be 0.000015%, 0.0007%, 0.0003%, 0.0087% DW respectively in 7-week old plants.[108]

Taxadiene production was also achieved in other transformable plant species. Besumbes *et al*. expressed TS under the control of CaMV 35S promoter in *Arabidopsis thaliana*. The authors found that the homozygous transgenic lines produced taxadiene at yields up to 0.0000025% DW. Production was also associated with developmental defects.[109] Kovacs *et al*. produced taxadiene in mature tomato fruit by expressing TS under the control of CaMV 35S promoter or ripening-specific polygalacturonase promoter. Taxadiene yields were 0.0279% DW under 35S and 0.0381% under polygalacturonase promoter, but transgenic plants suffer from severely reduced seed formation.[110] Moreover, Li *et al.* transformed TS under the control of CaMV 35S promoter into *Artemisia annua* and was able to produce taxadiene at yields up to 0.0075% DW in 2.5-month-old plants and 0.0129% DW in 6.5-month-old plants.[111] Lastly, Anterola *et al.* transformed *P. patens* protoplast with a construct expressing TS under the control of a

ubiquitin promoter and detected taxadiene at 0.05% fresh weight, without noticeable phenotypic defects.[112]

The yields achieved in tomato fruits and *A. annua* are noticeably higher than those detected in *N. benthamiana* and *A. thaliana* although they were achieved after a longer time frame. Yield in *P. patens* also surpassed those found in all vascular plant species although it was achieved in a bioreactor rather than in soil. Some transgenic host plants were reported to suffer phenotypic defects associated with production while others did not. This demonstrates that different heterologous hosts have their advantages and drawbacks that should be accounted for when choosing a production system. Moreover, the underlying reasons behind the drastic difference in yields of over three orders of magnitudes observed in different hosts are not well understood. A systematic comparison between different scalable plant hosts may benefit future efforts towards high-value natural product production. Overall, we have very little understanding of the contributing factors limiting production in specific plant hosts. More basic knowledge of the underlying plant metabolism and physiology of various plants is required to make plants a more engineerable system.

4. Engineering native plant producers of high-value natural products

Another possible route to scale up production of high-value natural products is to directly engineer the native producers. Engineering native plant producers potentially allows for increased yields without developing a functional heterologous expression system. Engineering may also be possible without the knowledge of all the genes involved in biosynthesis. Native plant producers may be engineered to increase yield by upregulating pathway expression or expand the range of tissue types in which biosynthesis occurs. Some native producers may inherently be more amenable to cultivation within an agricultural system that they are already farmed for extraction of their natural products. Engineering native producers may also alleviate pressure on the species in the wild put on by unsustainable gathering.

Genetic engineering of native plant producers has many advantages over traditional breeding. Traditional breeding relies on existing natural variations and the ability to cross plants. Plants that propagate primarily through vegetative propagation lack genetic diversity, making it difficult to improve phenotypes. Breeding may also be difficult for some species with complex genomes such as polyploids. Moreover, it can take multiple generations to arrive at the desired phenotypes through breeding, making engineering much more feasible especially for plants with long life cycles. Engineering may also allow introducing traits that are not found among natural genetic variation, including production of new-to-nature compound analogs.

However, native producers of high-value natural products are often not model plants; thus, engineering tools applicable to them are less available, making the tasks more challenging. Recent technological developments have lowered barriers to engineering non-model plants due to advances in genetic material delivery, genome editing, and plant regeneration. In this section, we review recent examples of successful engineering efforts in non-model plants, relevant technological advances, and remaining challenges.

4.1 Engineering *Catharanthus roseus* for production of vinca alkaloids

As an example of engineering native production hosts to improve yield, we look at engineering strategies applied in whole *Catharanthus roseus* plant to increase the titers of vindoline, a precursor of effective anticancer agent vinblastine. Vindoline is a monoterpene indole alkaloid (MIA) that is synthesized from monoterpene geraniol and tryptamine precursors. It is coupled with another MIA catharanthine to produce vinblastine. In 2012, an efficient transformation and regeneration protocol for *C. roseus* was developed,[113] making it possible to engineer transgenic lines with increased MIA formation.

One strategy that has been utilized is overexpression of upstream enzymes. Wang *et al.* created transgenic plants overexpressing an upstream enzyme deacetylvindoline-4-O-acetyltransferase with CaMV 35S promoter and increased production of vindoline up to 2.4-fold production from 0.11% to up to 0.272% DW.[113] Similarly, Kumar *et al.* showed that C. *roseus* overexpressing geranyl diphosphate synthase and geraniol synthase enhanced vindoline accumulation up to 2.5-fold.[114] The largest increase observed to date came from a study of Sharma *et la.*, which reported that overexpression of tryptophan decarboxylase and strictosidine synthase increased yield from 0.04% to 0.47% DW.[115]

Knocking out a competing pathway to vindoline biosynthesis has also been attempted. Kumar *et al.* explored the role of geranylgeranyl diphosphate synthase CrGGPPS2 in the production of MIAs. Despite its supposed role in a competing biosynthetic pathway drawing from the same geranyl diphosphate pool, the authors found that downregulation of CrGGPPS2 had the opposite of the expected effect. Instead of increasing vindoline yield, CrGGPPS22 down regulation significantly decreased the expression of transcription factors and pathway genes related to MIA biosynthesis, resulting in reduced vindoline production. Overexpression of CrGGPPS2, however, enhanced vindoline accumulation up to 2.5-fold.[116] This study did not effectively knock down competing pathways but demonstrated that engineering regulatory mechanisms could have beneficial effects on production titers.

Other attempts to engineer the regulatory mechanism of vindoline biosynthetic pathway in *C. roseus* involved altering expression levels of transcription factors. Pan *et al*. showed that overexpression of transcription factor ORCA3 or ORCA3 and geraniol 10-hydroxylase with CaMV 35S promoter, increased vindoline accumulation from 0.07% to up to 0.283 % and 0.3% DW respectively.[61] Similarly, Lie *et al.* identified transcription factor CR1 to negatively regulate MIA biosynthetic genes. The authors reported that accumulation of vindoline in the leaves of CR1-silenced increased roughly 3.3-fold.[62]

Combinations of approaches presented here may improve *C. roseus* engineering prospects to further increase yield or enable analogue synthesis.[117] However, to introduce multiple changes to the genome of a non-model plant is a labor-intensive process that usually takes at least three months to experiment, and may or may not deliver the desired phenotype. Moreover, some genomic changes that increase specialized metabolite production may interfere with regeneration.[95] Nonetheless, development of advanced methods to facilitate engineering non-model plants would greatly benefit production of plant natural products, given the already existing cultivation practices and agricultural infrastructure already in place for some medicinal plants.

4.2 Advances in transformation and regeneration of non-model plants

Engineering native producers of high-value natural products is challenging because they are not model species, and well-characterized tools applicable to them are lacking. The conventional method of plant

transformation is Agrobacterium-mediated transformation, which randomly integrates parts of the Ti plasmid into plant chromosomes. Following transformation, transgenic plant cells are regenerated into whole plants in a process which involves callus formation, and shoot and root induction. Although the method has successfully created many transgenic plant species, it suffers from extreme genotype-specificity and low efficiency. A successful transformation and regeneration protocol requires genotype-specific optimization of the amount of plant hormones auxins and cytokinins along with other tissue culture parameters in a lengthy, laborious process, which often fails. Here, we highlight some challenges and recent technological developments that facilitate transformation and regeneration in non-model plants.

The ectopic expression of developmental regulators to increase transformation and regeneration efficiency has dramatically altered the hurdles associated with plant transformation. Lowe *et al.* demonstrated that the transformation frequency of immature embryos, mature seed embryo sections, and seedling-derived leaf segments of recalcitrant maize inbred lines could be dramatically increased when *BABYBOOM* and *WUSCHEL* were expressed following Agrobacterium infection. The authors were able to induce formation of transgenic calli in 33 of 50 maize inbred lines experimented. In some genotypes, the authors reported increasing callus transformation efficiency from under 2% to over 25%. Subsequent removal of *BABYBOOM* and *WUSCHEL* through CRE-mediated excision led to healthy, fertile T_0 transgenic plants. The authors showed that the method was also effective on sorghum, rice, and sugarcane.[118]

Similarly, Maher *et al*. demonstrated that co-expression of developmental regulators *WUSCHEL* and *SHOOT MERISTEMLESS* led to *de novo* formation of meristem following Agrobacterium-mediated transformation of seedlings, cuttings, and soil-grown plants. These meristem structures could be transferred to the rooting medium and subsequently to soil. Optimization of regulator combinations led the authors to successfully induce meristem formation in *A. thaliana*, *N. benthamiana*, tomato, grape, and potato. In addition, the authors were able to introduce genetic changes into the *de novo* meristem, allowing construction of transgenic plants through this method. Although the method can result in mutants with phenotypic abnormalities, it bypasses tissue culture optimization, which is the most time-consuming step of plant regeneration.[119]

In order to reduce phenotypic abnormalities in transgenic plants as a result of ectopic expression of developmental regulators, other developmental genes were explored. Kong *et al.* and Debernardi *et al.* found that stable expression of GROWTH REGULATING FACTORS (GRF) improved transformation and regeneration without causing pleiotropic effects. Kong *et al.* reported that sugar beet calli transformed with Agrobacterium expressing *A. thaliana* GRF5 under CaMV 35S promoter showed enhanced transgenic shoot formation.[120] The authors also found expression of GRF5 and homologs to improve callus formation in canola, and shoot production in soybean and sunflower. Similarly, Debernardi *et al.* demonstrated that a fusion protein combining wheat GRF4 and its cofactor GRF-INTERACTING FACTOR 1 (GIF1) increased transformation efficiency of wheat immature embryos by an average of 7.8-fold and of fresh rice seeds by 2.1-fold.[121] Notably, the fusion protein allowed the authors to regenerate wheat shoots in culture medium without exogenous cytokinin. A similar fusion protein based on citrus and grape GRF-GIF pairs increased regeneration frequency in Carrizo citrange seeds up to 7.4-fold. Furthermore, the fusion protein construct was combined with CRISPR/Cas editing machinery to generate genome-edited progeny. These technologies drastically improve the transformation and regeneration

efficiency of genotypes and species previously presumed to be recalcitrant and have great promise in applications in producers of high-value natural products.

Another challenge in plant engineering is precise transgene-free genome editing. As Agrobacterium-mediated transformation results in random integration of genetic material, methods to create site-specific changes are desired. Since the discovery of the CRISPR/Cas system, an ever-expanding toolbox of gene-editing and other Cas protein-based methods have greatly facilitated plant engineering.[122] One unique challenge facing plant genome editing is polyploidy. Editing efficiency is lower in polyploid organisms than in diploid organisms as multiple alleles must be simultaneously edited. Nevertheless, it has been demonstrated in many polyploid plants using high expression levels of highly active Cas proteins.[123] Another barrier to implementing CRISPR/Cas-based genome editing in a larger variety of plant species has been in the delivery of the editing machinery to plant cells. PEG-mediated direct delivery of CRISPR/Cas9 ribonucleoproteins into protoplasts have been shown to produce transgene-free edited cells,[124,125] but to deliver genetic materials into intact plant cells for transgene-free editing has been challenging due to the rigidity of plant cell walls.

A promising technology is nanoparticle-mediated transformation. Recently, it was discovered that nanoparticle-mediated protocols allow delivery of genetic materials to intact plant cells without damaging them. Various types of nanoparticles including mesoporous silica nanoparticles,[126] DNA nanostructures,[127] and carbon nanotubes[128] have been shown to successfully penetrate plant cell walls in a genotype-independent manner. For instance, Demirer *et al.* demonstrated that double-stranded plasmid DNA can be passively delivered into intact plant cells, resulting in transient gene expression in leaves of *N. benthamiana*, arugula, wheat and cotton.[128] Nanoparticle-mediated delivery allows transient expression and gene silencing; however, it has not been demonstrated to make heritable, permanent genetic changes.

Applications of recently developed technologies in native production hosts of high-value natural products largely remain unexplored. Streamlined methods to edit, transform, and regenerate medicinal plants may unlock unexplored biosynthetic capacities of these metabolically rich species. It will facilitate molecule discovery, enzyme discovery, production and may enable exploration of new-to-nature analogues.

5. Implications and prospects

As an expanding number of plant specialized metabolite biosynthetic pathways are characterized, more opportunities are available for engineered biosynthesis to contribute to the production of plant natural products. Biosynthesis may be upregulated in the native producing species or reconstituted in an engineered heterologous plant or microbial host to improve access to these high-value small molecules. In determining the preferred production route, many factors may be taken into consideration. For instance, the structural complexity of the target molecule determines the length of the synthetic route, the type of chemistry involved, and enzymes required to achieve production. The nature of the producing species plays an important role in ultimately determining the costs and availability of raw materials. Moreover, the selling price of the product determines the minimum yield required for a production scheme to be economically viable.

Although microbial engineering has been the center of synthetic biology, plants offer many unique opportunities regarding specialized metabolite production due to their unique metabolic capacity, cellular

biology, and overall physiology. The most notable example is cytochrome P450 enzymes that function well in plant cells but have proven challenging to reconstitute in bacterial or yeast cells. This family of enzymes is central to plant specialized metabolism and is currently the bottle neck in microbial fermentation of plant natural products. Furthermore, plants can accumulate products to much higher levels than ever observed in microbial fermentation due to dedicated structures such as the trichomes that limit the bioactivity of the product on producing cells. Such compartmentalization also plays a major role in optimizing biosynthetic enzyme functions. These advantages make plants an attractive host for high-value plant natural product production.

However, complex plant physiology also gives rise to challenges that the emerging technologies seek to overcome. Functional genetic parts and ways to assemble large DNA constructs that allow flexible pathway designs were not available until recently. Efficient, reliable transformation and regeneration strategies necessary to create transgenic plants are also still being developed for genotype-independent engineering of non-model plants. It was recently discovered that co-expression of developmental regulators increases transformation and regeneration efficiency in many recalcitrant genotypes. Transient expression of CRISPR/Cas machinery had been achieved in protoplasts, achieving transgene-free editing.

Selecting which host to engineer – native producer or heterologous plant host – can dramatically alter and constrain various facets of metabolic engineering and production. Currently, heterologous production in another plant host has not been able to achieve yields as high as found in the native producers. Thus, in cases where the native producer is accessible and amenable to engineering, it is likely that cultivation of the native host will be more economical. This has been observed in the case of artemisinin and vinca alkaloids. However, in cases where the native producer of the metabolite of interest is not amenable to industrial scale cultivation, heterologous production may present a good alternative route. Heterologous production, especially in *N. benthamiana*, which allows testing combinatorial designs through transient co-infiltration, may also be preferred for production of new-to-nature analogues.

The causes of lower yields in heterologous expression systems are often not well understood. Reasons may involve low or imbalanced metabolic flux, the instability of pathway enzymes in the heterologous host, product toxicity, or cellular environment. New strategies to increase yields such as enzyme colocalization and product compartmentalization are currently being explored. Research into natural strategies used by high-producing genotypes may inform future engineering efforts. Ultimately, addressing these major knowledge gaps will translate into massive improvements in our ability to engineer plant metabolism, in turn greatly improving access to high-value plant molecules.

6. Conflicts of interest

There are no conflicts to declare.

7. Acknowledgements

We thank the DOE Joint BioEnergy Institute, which is supported by the Office of Science, Office of Biological and Environmental Research, the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

8. References

1. R. Subramani, M. Narayanasamy, and K.-D. Feussner, *3 Biotech*., 2017, **7**, 172.
2. D. Zhang and A. Kanakkanthara, *Cancers (Basel)*., 2020, **12**, 1721.
3. A. G. Atanasov, S. B. Zotchev, V. M. Dirsch, International Natural Product Sciences Taskforce, and C. T. Supuran, *Nat. Rev. Drug Discov.*, 2021, **20**, 200–216.
4. C. R. Barnum, B. J. Endelman, and P. M. Shih, *Front. Plant Sci.*, 2021, **12**, 691462.
5. C. F. Stratton, D. J. Newman, and D. S. Tan, *Bioorg. Med. Chem. Lett.*, 2015, **25**, 4802–4807.
6. J. V. Pham, M. A. Yilma, A. Feliz, M. T. Majid, N. Maffetone, J. R. Walker, E. Kim, H. J. Cho, J. M. Reynolds, M. C. Song, S. R. Park, and Y. J. Yoon, *Front. Microbiol.*, 2019, **10**, 1404.
7. A. Cravens, J. Payne, and C. D. Smolke, *Nat. Commun.*, 2019, **10**, 2142.
8. N. J. Patron, *New Phytol.*, 2020, **227**, 295–310.
9. S. Wang, S. Alseekh, A. R. Fernie, and J. Luo, *Mol. Plant*., 2019, **12**, 899–919.
10. S. A. Kautsar, H. G. Suarez Duran, K. Blin, A. Osbourn, and M. H. Medema, *Nucleic Acids Res.*, 2017, **45**, W55–W63.
11. M. Wink, *Nat. Prod. Commun.*, 2008, **3**, 1934578X0800300.
12. Q.-Z. Wang, S. R. Downie, and Z.-X. Chen, *Genome*., 2018, **61**, 103–109.
13. M. Rodríguez-Concepción and A. Boronat, *Curr. Opin. Plant Biol.*, 2015, **25**, 17–22.
14. L. Zhao, W. Chang, Y. Xiao, H. Liu, and P. Liu, *Annu. Rev. Biochem.*, 2013, **82**, 497–530.
15. X. Wang, C. Shen, P. Meng, G. Tan, and L. Lv, *BMC Plant Biol.*, 2021, **21**, 70.
16. R. Schuurink and A. Tissier, *New Phytol.*, 2020, **225**, 2251–2266.
17. S. Chhajed, B. B. Misra, N. Tello, and S. Chen, *Front. Plant Sci.*, 2019, **10**, 618.
18. M. V. Ramos, D. Demarco, I. C. da Costa Souza, and C. D. T. de Freitas, *Trends Plant Sci.*, 2019, **24**, 553–567.
19. J. Roepke, V. Salim, M. Wu, A. M. K. Thamm, J. Murata, K. Ploss, W. Boland, and V. De Luca, *Proc Natl Acad Sci USA*., 2010, **107**, 15287–15292.
20. E. Lewinsohn, *Ann. Bot.*, 1998, **81**, 35–39.
21. D. Martin, D. Tholl, J. Gershenzon, and J. Bohlmann, *Plant Physiol.*, 2002, **129**, 1003–1018.
22. C. Knudsen, N. J. Gallage, C. C. Hansen, B. L. Møller, and T. Laursen, *Nat. Prod. Rep.*, 2018, **35**, 1140–1155.
23. T. Laursen, B. L. Møller, and J.-E. Bassard, *Trends Plant Sci.*, 2015, **20**, 20–32.
24. N. J. White, *Science*., 2008, **320**, 330–334.
25. T. R. Larson, C. Branigan, D. Harvey, T. Penfield, D. Bowles, and I. A. Graham, *Industrial Crops and Products*., 2013, **45**, 1–6.
26. I. A. Graham, K. Besser, S. Blumer, C. A. Branigan, T. Czechowski, L. Elias, I. Guterman, D. Harvey, P. G. Isaac, A. M. Khan, T. R. Larson, Y. Li, T. Pawson, T. Penfield, A. M. Rae, D. A. Rathbone, S. Reid, J. Ross, M. F. Smallwood, V. Segura, and D. Bowles, *Science*., 2010, **327**, 328–331.
27. B. Kazaz, S. Webster, and P. Yadav, *Prod. Oper. Manag.*, 2016, **25**, 1576–1600.
28. M. K. Tiwari and S. Chaudhary, *Med. Res. Rev.*, 2020, **40**, 1220–1275.
29. A. Kumari, M. Karnatak, D. Singh, R. Shankar, J. L. Jat, S. Sharma, D. Yadav, R. Shrivastava, and V. P. Verma, *Eur. J. Med. Chem.*, 2019, **163**, 804–829.
30. P. Mercke, M. Bengtsson, H. J. Bouwmeester, M. A. Posthumus, and P. E. Brodelius, *Arch. Biochem. Biophys.*, 2000, **381**, 173–180.
31. V. J. J. Martin, D. J. Pitera, S. T. Withers, J. D. Newman, and J. D. Keasling, *Nat. Biotechnol.*, 2003, **21**, 796–802.

32. D.-K. Ro, E. M. Paradise, M. Ouellet, K. J. Fisher, K. L. Newman, J. M. Ndungu, K. A. Ho, R. A. Eachus, T. S. Ham, J. Kirby, M. C. Y. Chang, S. T. Withers, Y. Shiba, R. Sarpong, and J. D. Keasling, *Nature*., 2006, **440**, 940–943.
33. C. J. Paddon, P. J. Westfall, D. J. Pitera, K. Benjamin, K. Fisher, D. McPhee, M. D. Leavell, A. Tai, A. Main, D. Eng, D. R. Polichuk, K. H. Teoh, D. W. Reed, T. Treynor, J. Lenihan, M. Fleck, S. Bajad, G. Dang, D. Dengrove, D. Diola, and J. D. Newman, *Nature*., 2013, **496**, 528–532.
34. C. J. Paddon and J. D. Keasling, *Nat. Rev. Microbiol.*, 2014, **12**, 355–367.
35. M. Peplow, *Nature*., 2016, **530**, 389–390.
36. S. H. Kung, S. Lund, A. Murarka, D. McPhee, and C. J. Paddon, *Front. Plant Sci.*, 2018, **9**, 87.
37. M. Farhi, E. Marhevka, J. Ben-Ari, A. Algamas-Dimantov, Z. Liang, V. Zeevi, O. Edelbaum, B. Spitzer-Rimon, H. Abeliovich, B. Schwartz, T. Tzfira, and A. Vainstein, *Nat. Biotechnol.*, 2011, **29**, 1072–1074.
38. A. Baldi and V. K. Dixit, *Bioresour. Technol.*, 2008, **99**, 4609–4614.
39. N. Ikram and H. T. Simonsen, *Front. Plant Sci.*, 2017, **8**, 1966.
40. L. Zhang, F. Jing, F. Li, M. Li, Y. Wang, G. Wang, X. Sun, and K. Tang, *Biotechnol. Appl. Biochem.*, 2009, **52**, 199–207.
41. J. R. Jacobowitz and J.-K. Weng, *Annu. Rev. Plant Biol.*, 2020, **71**, 631–658.
42. M. Xiao, Y. Zhang, X. Chen, E.-J. Lee, C. J. S. Barber, R. Chakrabarty, I. Desgagné-Penix, T. M. Haslam, Y.-B. Kim, E. Liu, G. MacNevin, S. Masada-Atsumi, D. W. Reed, J. M. Stout, P. Zerbe, Y. Zhang, J. Bohlmann, P. S. Covello, V. De Luca, J. E. Page, and C. W. Sensen, *J. Biotechnol.*, 2013, **166**, 122–134.
43. N. Matasci, L.-H. Hung, Z. Yan, E. J. Carpenter, N. J. Wickett, S. Mirarab, N. Nguyen, T. Warnow, S. Ayyampalayam, M. Barker, J. G. Burleigh, M. A. Gitzendanner, E. Wafula, J. P. Der, C. W. dePamphilis, B. Roure, H. Philippe, B. R. Ruhfel, N. W. Miles, S. W. Graham, and G. K.-S. Wong, *Gigascience*., 2014, **3**, 17.
44. One Thousand Plant Transcriptomes Initiative, *Nature*., 2019, **574**, 679–685.
45. L. Guo, T. Winzer, X. Yang, Y. Li, Z. Ning, Z. He, R. Teodor, Y. Lu, T. A. Bowser, I. A. Graham, and K. Ye, *Science*., 2018, **362**, 343–347.
46. X. Xiong, J. Gou, Q. Liao, Y. Li, Q. Zhou, G. Bi, C. Li, R. Du, X. Wang, T. Sun, L. Guo, H. Liang, P. Lu, Y. Wu, Z. Zhang, D.-K. Ro, Y. Shang, S. Huang, and J. Yan, *Nat. Plants*., 2021, **7**, 1026–1036.
47. W. Lau and E. S. Sattely, *Science*., 2015, **349**, 1224–1228.
48. L. Caputi, J. Franke, S. C. Farrow, K. Chung, R. M. E. Payne, T.-D. Nguyen, T.-T. T. Dang, I. Soares Teto Carqueijeiro, K. Koudounas, T. Dugé de Bernonville, B. Ameyaw, D. M. Jones, I. J. C. Vieira, V. Courdavault, and S. E. O'Connor, *Science*., 2018, **360**, 1235–1239.
49. S. C. Farrow, M. O. Kamileen, L. Caputi, K. Bussey, J. E. A. Mundy, R. C. McAtee, C. R. J. Stephenson, and S. E. O'Connor, *J. Am. Chem. Soc.*, 2019, **141**, 12979–12983.
50. H. Hodgson, R. De La Peña, M. J. Stephenson, R. Thimmappa, J. L. Vincent, E. S. Sattely, and A. Osbourn, *Proc Natl Acad Sci USA*., 2019, **116**, 17096–17104.
51. S. Galanie, K. Thodey, I. J. Trenchard, M. Filsinger Interrante, and C. D. Smolke, *Science*., 2015, **349**, 1095–1100.
52. P. Schläpfer, P. Zhang, C. Wang, T. Kim, M. Banf, L. Chae, K. Dreher, A. K. Chavali, R. Nilo-Poyanco, T. Bernard, D. Kahn, and S. Y. Rhee, *Plant Physiol.*, 2017, **173**, 2041–2059.
53. L. K. Tuominen, V. E. Johnson, and C.-J. Tsai, *BMC Genomics*., 2011, **12**, 236.

54. N. Töpfer, L.-M. Fuchs, and A. Aharoni, *Nucleic Acids Res.*, 2017, **45**, 7049–7063.
55. L. Mao, H. Kawaide, T. Higuchi, M. Chen, K. Miyamoto, Y. Hirata, H. Kimura, S. Miyazaki, M. Teruya, K. Fujiwara, K. Tomita, H. Yamane, K.-I. Hayashi, H. Nojiri, L. Jia, J. Qiu, C. Ye, M. P. Timko, L. Fan, and K. Okada, *Proc Natl Acad Sci USA*., 2020, **117**, 12472–12480.
56. N. Kitaoka, J. Zhang, R. K. Oyagbenro, B. Brown, Y. Wu, B. Yang, Z. Li, and R. J. Peters, *Plant Cell*., 2021, **33**, 290–305.
57. M. H. Medema and A. Osbourn, *Nat. Prod. Rep.*, 2016, **33**, 951–962.
58. T. W. J. M. van Herpen, K. Cankar, M. Nogueira, D. Bosch, H. J. Bouwmeester, and J. Beekwilder, *PLoS ONE*., 2010, **5**, e14222.
59. L. Dong, E. Jongedijk, H. Bouwmeester, and A. Van Der Krol, *New Phytol.*, 2016, **209**, 679–690.
60. R. Jan, S. Asaf, M. Numan, Lubna, and K.-M. Kim, *Agronomy*., 2021, **11**, 968.
61. Q. Pan, Q. Wang, F. Yuan, S. Xing, J. Zhao, Y. H. Choi, R. Verpoorte, Y. Tian, G. Wang, and K. Tang, *PLoS ONE*., 2012, **7**, e43038.
62. J. Liu, F. Gao, J. Ren, X. Lu, G. Ren, and R. Wang, *Front. Plant Sci.*, 2017, **8**, 2082.
63. Z. Zhou, H. Tan, Q. Li, Q. Li, Y. Wang, Q. Bu, Y. Li, Y. Wu, W. Chen, and L. Zhang, *New Phytol.*, 2020, **228**, 932–945.
64. R. P. Barone, D. K. Knittel, J. K. Ooka, L. N. Porter, N. T. Smith, and D. K. Owens, *Current Plant Biology*., 2020, **24**, 100121.
65. G. Polturak, Z. Liu, and A. Osbourn, *Current Opinion in Green and Sustainable Chemistry*., 2021, 100568.
66. Z. Liu, J. Cheema, M. Vigouroux, L. Hill, J. Reed, P. Paajanen, L. Yant, and A. Osbourn, *Nat. Commun.*, 2020, **11**, 5354.
67. J. Franke, J. Kim, J. P. Hamilton, D. Zhao, G. M. Pham, K. Wiegert-Rininger, E. Crisovan, L. Newton, B. Vaillancourt, E. Tatsis, C. R. Buell, and S. E. O'Connor, *Chembiochem*., 2019, **20**, 83–87.
68. M. Frey, P. Chomet, E. Glawischnig, C. Stettner, S. Grün, A. Winklmair, W. Eisenreich, A. Bacher, R. B. Meeley, S. P. Briggs, K. Simcox, and A. Gierl, *Science*., 1997, **277**, 696–699.
69. R. J. Peters, *New Phytol.*, 2020, **227**, 992–994.
70. J. Zhang and R. J. Peters, *Proc Natl Acad Sci USA*., 2020, **117**, 13867–13869.
71. J. Andersen-Ranberg, K. T. Kongstad, M. T. Nielsen, N. B. Jensen, I. Pateraki, S. S. Bach, B. Hamberger, P. Zerbe, D. Staerk, J. Bohlmann, B. L. Møller, and B. Hamberger, *Angew Chem Int Ed Engl*., 2016, **55**, 2142–2146.
72. A. Calgaro-Kozina, K. M. Vuu, J. D. Keasling, D. Loqué, E. S. Sattely, and P. M. Shih, *ACS Cent. Sci.*, 2020, **6**, 1394–1400.
73. L. Guo, J. Qiu, C. Ye, G. Jin, L. Mao, H. Zhang, X. Yang, Q. Peng, Y. Wang, L. Jia, Z. Lin, G. Li, F. Fu, C. Liu, L. Chen, E. Shen, W. Wang, Q. Chu, D. Wu, S. Wu, and L. Fan, *Nat. Commun.*, 2017, **8**, 1031.
74. E. Pichersky and E. Lewinsohn, *Annu. Rev. Plant Biol.*, 2011, **62**, 549–566.
75. K. Jørgensen, A. V. Rasmussen, M. Morant, A. H. Nielsen, N. Bjarnholt, M. Zagrobelny, S. Bak, and B. L. Møller, *Curr. Opin. Plant Biol.*, 2005, **8**, 280–291.
76. K. A. Nielsen, D. B. Tattersall, P. R. Jones, and B. L. Møller, *Phytochemistry*., 2008, **69**, 88–98.
77. L. J. Sweetlove and A. R. Fernie, *Nat. Commun.*, 2018, **9**, 2136.
78. S. Mucha, S. Heinzlmeir, V. Kriechbaumer, B. Strickland, C. Kirchhelle, M. Choudhary, N. Kowalski, R. Eichmann, R. Hueckelhoven, E. Grill, B. Kuster, and E. Glawischnig, *Plant Cell*.,

2019, **31**, 2697–2710.
79. J.-E. Bassard, B. L. Møller, and T. Laursen, *Curr. Mol. Biol. Rep.*, 2017, **3**, 37–51.
80. J. B. Y. H. Behrendorff, G. Borràs-Gas, and M. Pribil, *Trends Biotechnol.*, 2020, **38**, 432–446.
81. M. P. R. Henriques de Jesus, A. Zygadlo Nielsen, S. Busck Mellor, A. Matthes, M. Burow, C. Robinson, and P. Erik Jensen, *Metab. Eng.*, 2017, **44**, 108–116.
82. J. Li, I. Mutanda, K. Wang, L. Yang, J. Wang, and Y. Wang, *Nat. Commun.*, 2019, **10**, 4850.
83. T. Laursen, J. Borch, C. Knudsen, K. Bavishi, F. Torta, H. J. Martens, D. Silvestro, N. S. Hatzakis, M. R. Wenk, T. R. Dafforn, C. E. Olsen, M. S. Motawia, B. Hamberger, B. L. Møller, and J.-E. Bassard, *Science*., 2016, **354**, 890–893.
84. V. Amirkia and M. Heinrich, *Phytochem. Lett.*, 2014, **10**, xlviii–liii.
85. S. Wu, M. Schalk, A. Clark, R. B. Miles, R. Coates, and J. Chappell, *Nat. Biotechnol.*, 2006, **24**, 1441–1447.
86. D. Feike, A. V. Korolev, E. Soumpourou, E. Murakami, D. Reid, A. Breakspear, C. Rogers, S. Radutoiu, J. Stougaard, W. A. Harwood, G. E. D. Oldroyd, and J. B. Miller, *Plant Biotechnol. J.*, 2019, **17**, 2234–2245.
87. T. Gnanasekaran, D. Karcher, A. Z. Nielsen, H. J. Martens, S. Ruf, X. Kroop, C. E. Olsen, M. S. Motawie, M. Pribil, B. L. Møller, R. Bock, and P. E. Jensen, *J. Exp. Bot.*, 2016, **67**, 2495–2506.
88. T. L. Delatte, G. Scaiola, J. Molenaar, K. de Sousa Farias, L. Alves Gomes Albertti, J. Busscher, F. Verstappen, C. Carollo, H. Bouwmeester, and J. Beekwilder, *Plant Biotechnol. J.*, 2018, **16**, 1997–2006.
89. R. Sadre, P. Kuo, J. Chen, Y. Yang, A. Banerjee, C. Benning, and B. Hamberger, *Nat. Commun.*, 2019, **10**, 853.
90. J. Reed, M. J. Stephenson, K. Miettinen, B. Brouwer, A. Leveau, P. Brett, R. J. M. Goss, A. Goossens, M. A. O'Connell, and A. Osbourn, *Metab. Eng.*, 2017, **42**, 185–193.
91. K. Davis, D. S. Gkotsi, D. R. M. Smith, R. J. M. Goss, L. Caputi, and S. E. O'Connor, *Front. Plant Sci.*, 2020, **11**, 581675.
92. B. J. Schultz, S. Y. Kim, W. Lau, and E. S. Sattely, *J. Am. Chem. Soc.*, 2019, **141**, 19231–19235.
93. USDA,. https://www.nass.usda.gov/Charts_and_Maps/Crops_County/tfc-yi.php, .
94. CEIC. https://www.ceicdata.com/en/china/yield-per-hectare-farm-crops-flue-cured-tobacco/cn-farm-crops-flue-cured-tobacco-yield-per-hectare-yh, .
95. S. Fresquet-Corrales, E. Roque, A. Sarrión-Perdigones, M. Rochina, M. P. López-Gresa, H. M. Díaz-Mula, J. M. Bellés, F. Tomás-Barberán, J. P. Beltrán, and L. A. Cañas, *PLoS ONE*., 2017, **12**, e0184839.
96. C. Zhao, Y. Kim, Y. Zeng, M. Li, X. Wang, C. Hu, C. Gorman, S. Y. Dai, S.-Y. Ding, and J. S. Yuan, *ACS Synth. Biol.*, 2018, **7**, 774–781.
97. H. Zhang, S. Zhang, Y. Yang, H. Jia, and H. Cui, *Plant Cell Physiol.*, 2018, **59**, 566–574.
98. H. Zhang, D. Niu, J. Wang, S. Zhang, Y. Yang, H. Jia, and H. Cui, *Plant Cell Physiol.*, 2015, **56**, 2125–2138.
99. R. Reski, H. Bae, and H. T. Simonsen, *Plant Cell Rep.*, 2018, **37**, 1409–1417.
100. K. Müller, D. Siegel, F. Rodriguez Jahnke, K. Gerrer, S. Wend, E. L. Decker, R. Reski, W. Weber, and M. D. Zurbriggen, *Mol. Biosyst.*, 2014, **10**, 1679–1688.
101. M. Kubo, A. Imai, T. Nishiyama, M. Ishikawa, Y. Sato, T. Kurata, Y. Hiwatashi, R. Reski, and M. Hasebe, *PLoS ONE*., 2013, **8**, e77356.
102. A. Peramuna, H. Bae, E. K. Rasmussen, B. Dueholm, T. Waibel, J. H. Critchley, K. Brzezek, M.

Roberts, and H. T. Simonsen, *Biochem. Biophys. Res. Commun.*, 2018, **500**, 418–422.
103. B. C. King, K. Vavitsas, N. K. B. K. Ikram, J. Schrøder, L. B. Scharff, J.-É. Bassard, B. Hamberger, P. E. Jensen, and H. T. Simonsen, *Sci. Rep.*, 2016, **6**, 25030.
104. N. K. B. Khairul Ikram, A. Beyraghdar Kashkooli, A. V. Peramuna, A. R. van der Krol, H. Bouwmeester, and H. T. Simonsen, *Front. Bioeng. Biotechnol.*, 2017, **5**, 47.
105. A. Peramuna, H. Bae, C. Quiñonero López, A. Fromberg, B. Petersen, and H. T. Simonsen, *PLoS ONE*., 2020, **15**, e0243620.
106. X.-W. Pan, L. Han, Y.-H. Zhang, D.-F. Chen, and H. T. Simonsen, *Plant Biotechnol. Rep.*, 2015, **9**, 149–159.
107. M. M. Hasan, H.-S. Kim, J.-H. Jeon, S. H. Kim, B. Moon, J.-Y. Song, S. H. Shim, and K.-H. Baek, *Plant Cell Rep.*, 2014, **33**, 895–904.
108. J. Fu, W. Xu, W. Huang, B. Wang, S. Li, J. Zhang, and L. Chang, *Planta*., 2021, **253**, 107.
109. O. Besumbes, S. Sauret-Güeto, M. A. Phillips, S. Imperial, M. Rodríguez-Concepción, and A. Boronat, *Biotechnol. Bioeng.*, 2004, **88**, 168–175.
110. K. Kovacs, L. Zhang, R. S. T. Linforth, B. Whittaker, C. J. Hayes, and R. G. Fray, *Transgenic Res.*, 2007, **16**, 121–126.
111. M. Li, F. Jiang, X. Yu, and Z. Miao, *Biomed Res. Int.*, 2015, **2015**, 504932.
112. A. Anterola, E. Shanle, P.-F. Perroud, and R. Quatrano, *Transgenic Res.*, 2009, **18**, 655–660.
113. Q. Wang, S. Xing, Q. Pan, F. Yuan, J. Zhao, Y. Tian, Y. Chen, G. Wang, and K. Tang, *BMC Biotechnol.*, 2012, **12**, 34.
114. S. R. Kumar, H. B. Shilpashree, and D. A. Nagegowda, *Front. Plant Sci.*, 2018, **9**, 942.
115. A. Sharma, P. Verma, A. Mathur, and A. K. Mathur, *Protoplasma*., 2018, **255**, 1281–1294.
116. S. R. Kumar, A. Rai, D. P. Bomzan, K. Kumar, A. Hemmerlin, V. Dwivedi, R. C. Godbole, V. Barvkar, K. Shanker, H. B. Shilpashree, A. Bhattacharya, A. R. Smitha, N. Hegde, and D. A. Nagegowda, *Plant J.*, 2020, **103**, 248–265.
117. W. Runguphan, X. Qu, and S. E. O'Connor, *Nature*., 2010, **468**, 461–464.
118. K. Lowe, E. Wu, N. Wang, G. Hoerster, C. Hastings, M.-J. Cho, C. Scelonge, B. Lenderts, M. Chamberlin, J. Cushatt, L. Wang, L. Ryan, T. Khan, J. Chow-Yiu, W. Hua, M. Yu, J. Banh, Z. Bao, K. Brink, E. Igo, and W. Gordon-Kamm, *Plant Cell*., 2016, **28**, 1998–2015.
119. M. F. Maher, R. A. Nasti, M. Vollbrecht, C. G. Starker, M. D. Clark, and D. F. Voytas, *Nat. Biotechnol.*, 2020, **38**, 84–89.
120. J. Kong, S. Martin-Ortigosa, J. Finer, N. Orchard, A. Gunadi, L. A. Batts, D. Thakare, B. Rush, O. Schmitz, M. Stuiver, P. Olhoft, and D. Pacheco-Villalobos, *Front. Plant Sci.*, 2020, **11**, 572319.
121. J. M. Debernardi, D. M. Tricoli, M. F. Ercoli, S. Hayta, P. Ronald, J. F. Palatnik, and J. Dubcovsky, *Nat. Biotechnol.*, 2020, **38**, 1274–1279.
122. Y. Zhang, A. A. Malzahn, S. Sretenovic, and Y. Qi, *Nat. Plants*., 2019, **5**, 778–794.
123. Z. Liang, K. Chen, T. Li, Y. Zhang, Y. Wang, Q. Zhao, J. Liu, H. Zhang, C. Liu, Y. Ran, and C. Gao, *Nat. Commun.*, 2017, **8**, 14261.
124. M. Malnoy, R. Viola, M.-H. Jung, O.-J. Koo, S. Kim, J.-S. Kim, R. Velasco, and C. Nagamangala Kanchiswamy, *Front. Plant Sci.*, 2016, **7**, 1904.
125. S. Subburaj, S. J. Chung, C. Lee, S.-M. Ryu, D. H. Kim, J.-S. Kim, S. Bae, and G.-J. Lee, *Plant Cell Rep.*, 2016, **35**, 1535–1544.
126. Z. Hajiahmadi, R. Shirzadian-Khorramabad, M. Kazemzad, and M. M. Sohani, *Sci. Hortic.*, 2019, **243**, 367–375.

127. H. Zhang, G. S. Demirer, H. Zhang, T. Ye, N. S. Goh, A. J. Aditham, F. J. Cunningham, C. Fan, and M. P. Landry, *Proc Natl Acad Sci USA*., 2019, **116**, 7543–7548.
128. G. S. Demirer, H. Zhang, J. L. Matos, N. S. Goh, F. J. Cunningham, Y. Sung, R. Chang, A. J. Aditham, L. Chio, M.-J. Cho, B. Staskawicz, and M. P. Landry, *Nat. Nanotechnol.*, 2019, **14**, 456–464.

Tables and figures

Table 1: Cost, concentration in native producer, yield of native producer in kg product / Ha / yr, and "parity titer" to achieve comparable production in *N. tabacum* for several plant natural products. Values are estimates derived from averaging many sources. Pricing information is intended to reflect consumer price, and is principally derived from the Drugbank Online database, taking the cheaper estimates for larger bulk consumer prices DrugBank (https://go.drugbank.com/) and GoodRx (https://www.goodrx.com/).

Drug	**Class of natural product**	**Native producer**	**Cost (USD per gram)**	**Concentration in native**	**Yield kg / ha/yr**	**Titer required in *N. tabaccum* to achieve same production / ha / yr (% DW)**
Reserpine	monoterpene indole alkaloid	*Rauvolfia serpentina* or Indian snakeroot is an evergreen shrub growing to 1 m tall, found in east and southeast asia. Roots can be harvested for extraction from plants 2-3 years old.	8300	0.05 %	1	0.05
Artemisinin	sesquiterpene	*Artemisia annua* or sweet wormwood is an annual plant growing up to 1 m tall. Plant is native to temperate Asia and prefers warm and sunny conditions. Leaves are harvested at the beginning of flowering 190-240 days after sowing.	1	2%	10	0.5
Digoxin	triterpene glycoside	*Digitalis lanata* or Grecian Foxglove is an evergreen biennial or short-lived perennial growing to 0.6 m by 0.3 m. It is native to the Balkan regions. Leaves can be harvested for extraction after 2 years.	450	0.1%	15	0.75
Pilocarpine	alkaloid	*Pilocarpus microphyllus* or Maranham Jaborandi is a small evergreen tree, growing up to 7 m tall and native to northern Brazil. Leaves are harvested for extraction. Plant is listed as	120	0.5%	16	0.8

		endangered since the 1990s.				
Sennosides	anthraquinone glycosides	*Cassia angustifolia* or Indian senna is a perennial shrub growing to 1 m. It is native to Egypt and grown commercially throughout the world. Leaves are harvested 3,5, and 7 months after sowing.	10	3.5%	24	1.2
Taxol	tetracyclic diterpene	*Taxus brevifolia* or pacific yew is a dioecious evergreen tree growing to 15 m tall and with a trunk up to 0.5 m diameter. Bark is harvested for extraction. Plant is currently listed as near threatened	4900	0.03 %	0.015	0.00075
Vinblastine	monoterpene indole alkaloid	*Catharanthus roseus* or Madagascar periwinkle is a small perennial plant native to Madagascar. It is grown in tropical and subtropical regions as an ornamental plant. Leaves are harvested for extraction	5500	0.00 4%	0.08	0.004

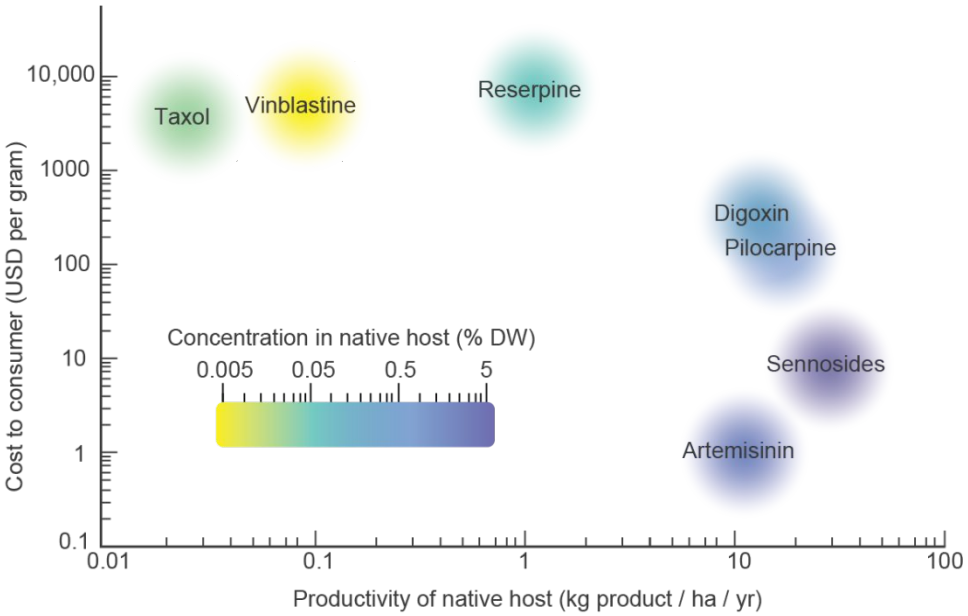

Figure 1: Relationship between price, concentration in native host, and productivity in native host for selected plant natural products with major markets. All values are approximations, as denoted by the large and faded circles for each drug. Costs to consumers were estimated by searching for the cheapest drug formulations online.

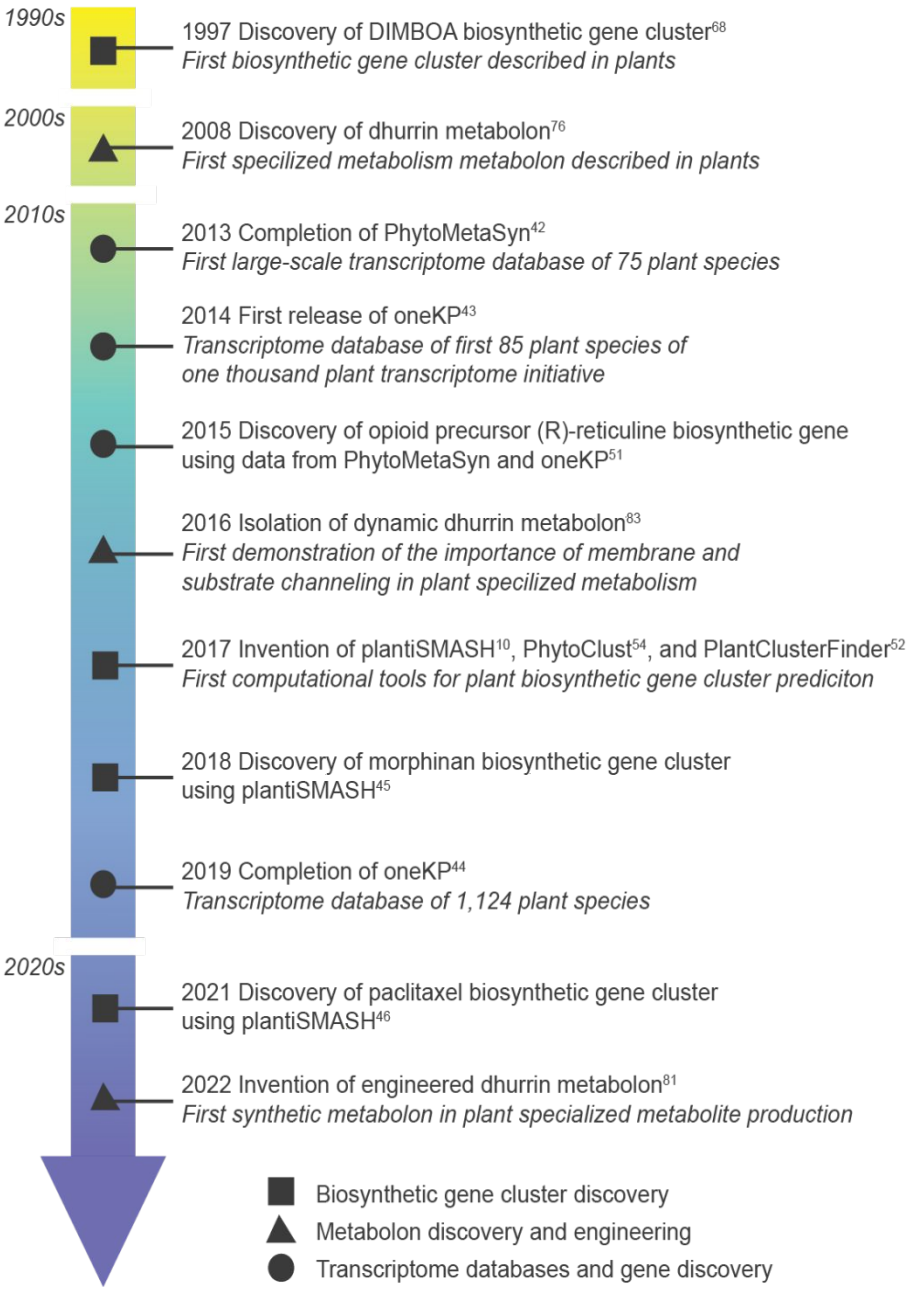

Figure 2: Recent inventions of enabling technologies and databases, and related landmark discoveries in plant biosynthetic pathway organization. Select events related to biosynthetic gene cluster discovery (□), metabolon discovery and engineering (▲), and transcriptome databases and gene discovery (□) are depicted.

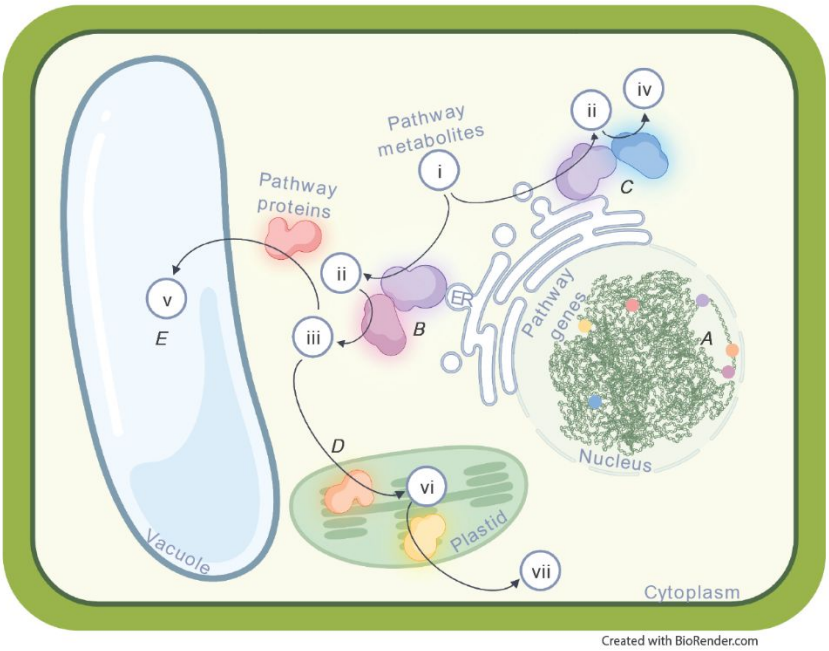

Figure 3: Metabolic organization of biosynthetic genes and enzymes of plant high-value natural products. A) Pathway genes may be clustered. B) Cascading pathway proteins may form protein complexes and co-localize to a particular subcellular structure. C) A pathway protein may have multiple interaction partners, leading to pathway branching and multiple products. D) Pathway intermediates may be shuttled to different cellular compartments, leading to different products. E) Product may be detoxified and sequestered in a dedicated compartment.

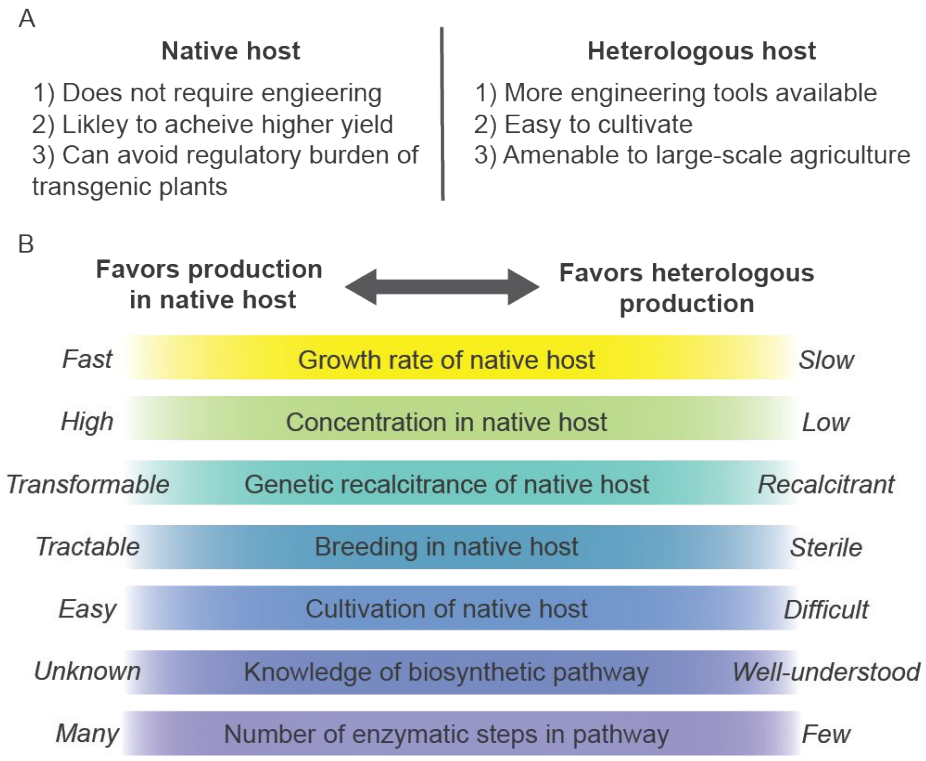

Figure 4: Tradeoffs between production of plant natural products in native host versus heterologous system. A) Summary of primary benefits of the two whole-plant production methods of plant natural products. B) Framework to assist decision-making for the production method of a particular target molecule. Each target molecule can be imagined in a many-dimensional space along many axes. For each axis, one extreme favors growth of the native producer plant whereas the other favors engineering of a heterologous host.

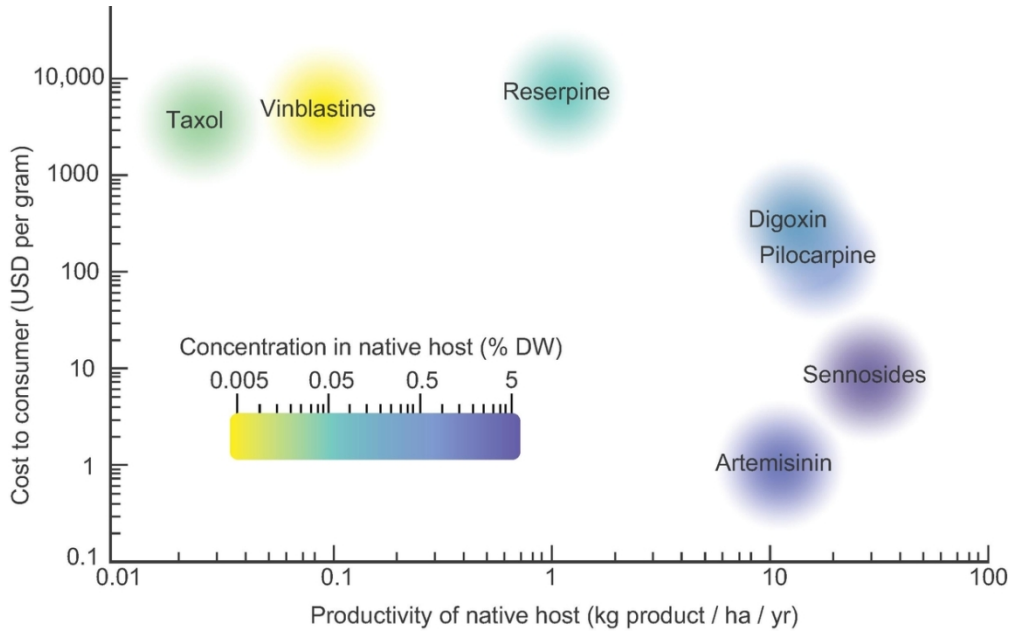

Figure 1: Relationship between price, concentration in native host, and productivity in native host for selected plant natural products with major markets. All values are approximations, as denoted by the large and faded circles for each drug. Costs to consumers were estimated by searching for the cheapest drug formulations online.

104x65mm (300 x 300 DPI)

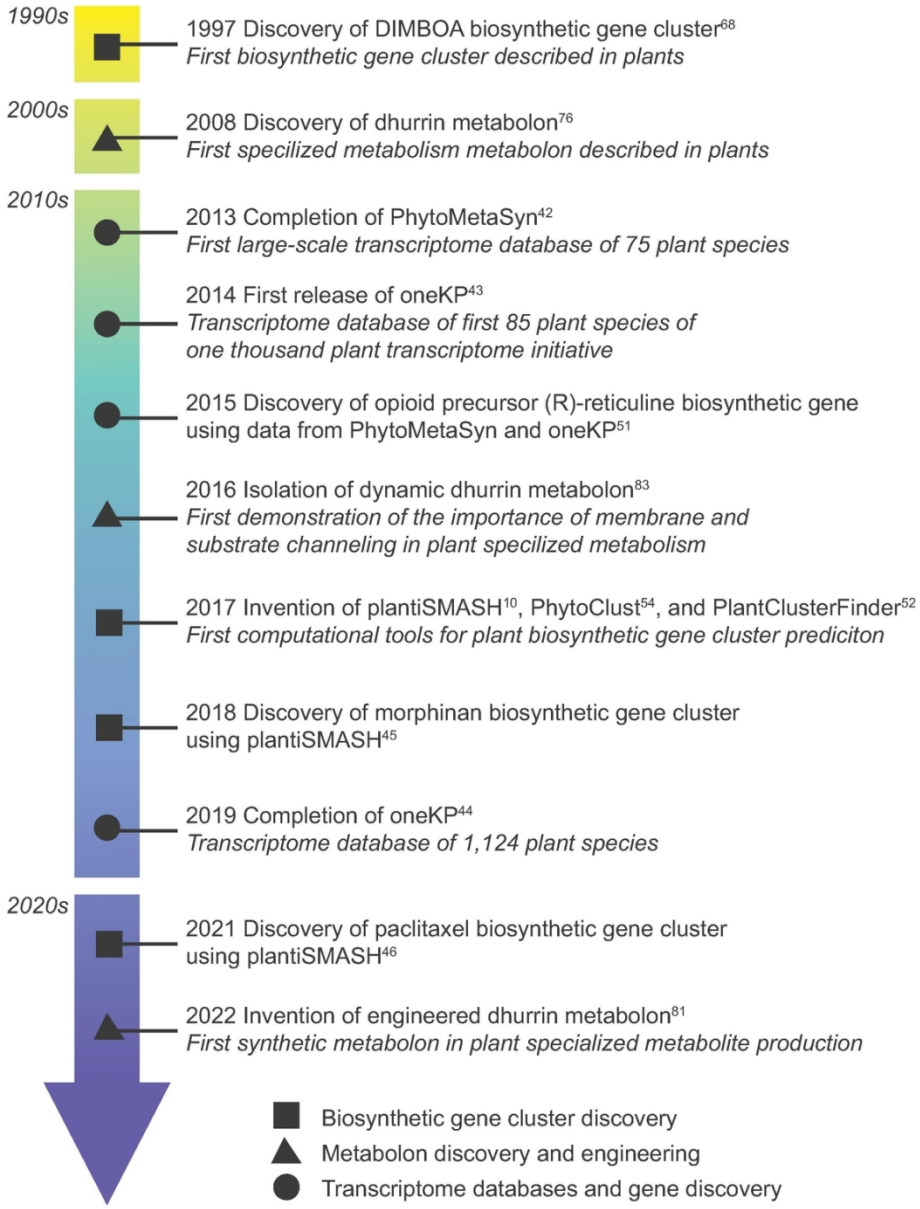

Figure 2: Recent inventions of enabling technologies and databases, and related landmark discoveries in plant biosynthetic pathway organization. Select events related to biosynthetic gene cluster discovery (□), metabolon discovery and engineering (▲), and transcriptome databases and gene discovery (□) are depicted.

98x129mm (300 x 300 DPI)

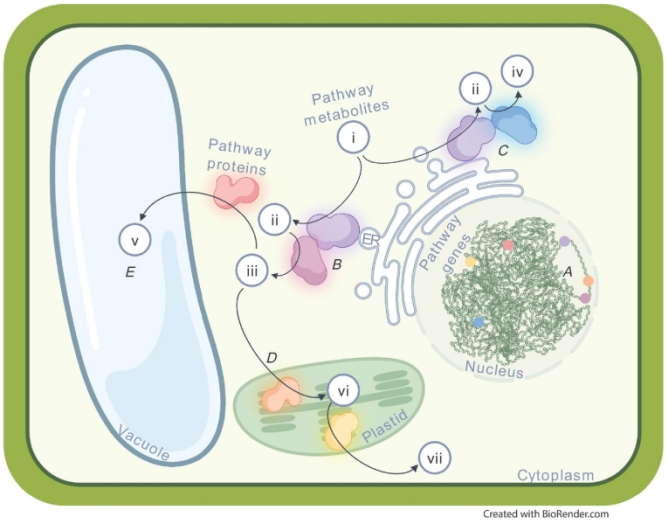

Figure 3: Metabolic organization of biosynthetic genes and enzymes of plant high-value natural products. A) Pathway genes may be clustered. B) Cascading pathway proteins may form protein complexes and co-localize to a particular subcellular structure. C) A pathway protein may have multiple interaction partners, leading to pathway branching and multiple products. D) Pathway intermediates may be shuttled to different cellular compartments, leading to different products. E) Product may be detoxified and sequestered in a dedicated compartment.

165x116mm (300 x 300 DPI)

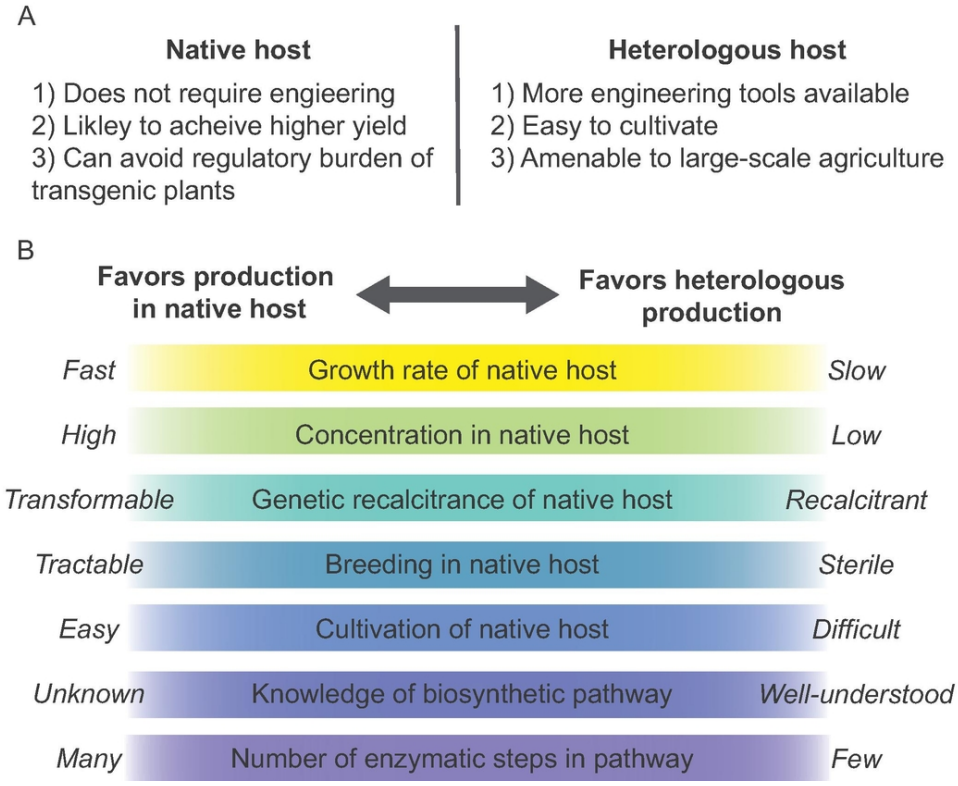

Figure 4: Tradeoffs between production of plant natural products in native host versus heterologous system. A) Summary of primary benefits of the two whole-plant production methods of plant natural products. B) Framework to assist decision-making for the production method of a particular target molecule. Each target molecule can be imagined in a many-dimensional space along many axes. For each axis, one extreme favors growth of the native producer plant whereas the other favors engineering of a heterologous host.

91x71mm (300 x 300 DPI)