

Functional diversity and metabolic engineering of plant-specialized metabolites

Shaoqun Zhou¹, Yongshuo Ma^{1,2}, Yi Shang³, Xiaoquan Qi⁴, Sanwen Huang^{1,*}, Jiayang Li⁵

¹Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, Guangdong 518120, China

²Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

³Yunnan Key Laboratory of Potato Biology, The CAAS-YNNU-YINMORE Joint Academy of Potato Sciences, Yunnan Normal University, Kunming, Yunnan 650500, China

⁴Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

⁵State Key Laboratory of Plant Genomics, National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing 100101, China

*Corresponding authors. Jiayang Li, State Key Laboratory of Plant Genomics, National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing 100101, China. E-mail: jyli@genetics.ac.cn; Sanwen Huang, Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, Guangdong 518120, China. E-mail: huangsanwen@caas.cn; Xiaoquan Qi, Key Laboratory of Plant Molecular Physiology, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. E-mail: xqi@ibcas.ac.cn

Abstract

Plants are talented biochemists that produce a broad diversity of small molecules. These so-called specialized metabolites (SMs) play critical roles in the adaptive evolution of plants to defend against biotic and abiotic stresses, attract pollinators, and modulate soil microbiota for their own benefits. Many plant SMs have been used as nutrition and flavor compounds in our daily food, as well as drugs for treatment of human diseases. Current multi-omics tools have significantly accelerated the process of biosynthetic pathway elucidation in plants through correlation analyses, genetic mapping, and *de novo* biosynthetic gene cluster predictions. Understanding the biosynthesis of plant SMs has enabled reconstitution of naturally occurring specialized metabolic pathways in microbial hosts, providing a sustainable supply of these high-value molecules. In this review, we illustrate the general functions of several typical plant SMs in natural ecosystems and for human societies. We then provide an overview of current methods elucidating the biosynthetic pathways of plant SMs, and synthetic biology strategies that optimize the efficiency of heterologous biosynthetic pathways in microbial hosts. Moving forward, dissection of the functions and application of plant SMs by using current multidiscipline approaches would be greatly benefit to the scientific community and human societies.

Keywords: plant specialized metabolites; metabolic engineering

Introduction

Plants are superb biochemists that can produce a large diversity of small molecules by absorbing CO₂, mineral elements, and water from the environment. These diverse metabolites are typically present in lineage-, developmental stage-, and tissue-specific manners, which are hence termed as specialized metabolites (SMs). The total number of structurally distinct SMs found across the plant kingdom has been estimated to be in the range of hundreds of thousands [1]. In contrast to the broadly conserved primary metabolites that are required for plant growth and development under ideal growth conditions, SMs play important roles in the interactions between plants and their environments. SMs add desirable flavors and important health benefits to human nutrition. In the last decades, molecular genetics and physiology studies have revealed biosynthesis of several important plant SMs. These knowledges have enabled synthetic biology approaches to reconstruct biosynthetic pathways of plant SMs in chassis organisms, producing the target compounds on an industrial scale. In this review, we briefly summarize the function of several typical plant SMs in natural ecosystems and for human

consumption. We then delineate current methods in elucidating biosynthetic pathways of SMs in plants, and synthetic biology strategies to enhance target compound production in microbial chassis.

SMs mediate plant–environment interactions

As sessile organisms, plants have adapted to a wide array of hostile environmental conditions. To cope with these challenges, plants produce diverse SMs to protect themselves from environmental stressors, attract animal pollinators for successful cross-fertilization, engineer local soil microbiota to support their own growth, and adapt to abiotic stresses. Examples illustrating these ecological functions of SMs are summarized in this section (Fig. 1).

Protection of plants against insect herbivores and microbial pathogens

Plant SMs were initially appreciated for their protective functions against insect herbivores and pathogenic microbes. Classic

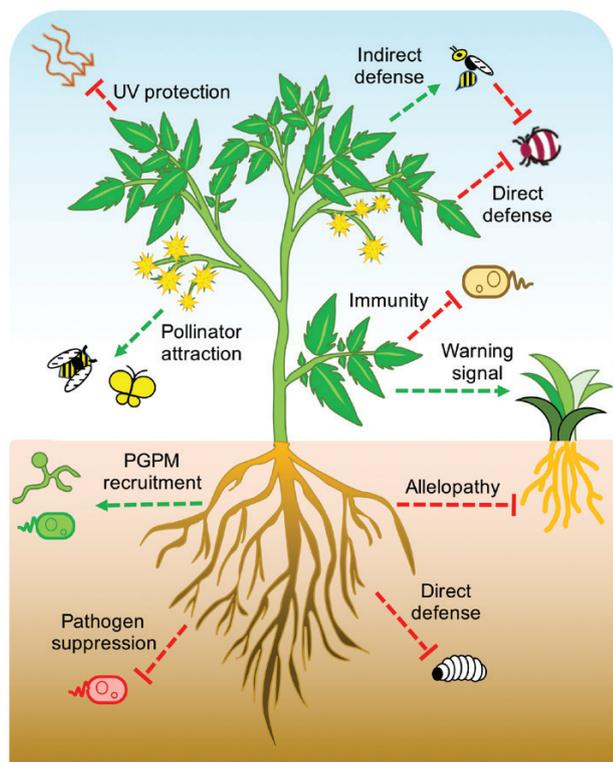


Figure 1 Plant SMs carry out diverse ecological functions. Attractions and repellence/toxicity are represented by green arrows and red T-shaped arrows, respectively. UV, ultraviolet; PGPM, plant growth-promoting microbes.

examples of such compounds include glucosinolate [2], cyanogenic glucosides [3], nicotine [4], and benzoxazinoids [5]. Current mass spectrometry-based high-throughput analytical chemistry platforms and chemo-informatics tools have enabled exploration of more diverse defensive metabolites that present in low abundance in non-model organisms, as recently reviewed by Li and Gaquerel [6]. For example, a natural-history-guided omics approach has been taken to identify a novel compound, caffeoylputrescine-5-(Z)-3-hexenal, from a wild tobacco species, which was directly toxic to *Empoasca* leafhoppers when fed at the physiologically relevant concentration [7]. Interestingly, plants have also evolved exquisite detoxification mechanisms to avoid autotoxicity caused by the SMs they produce [8].

Other than being directly toxic against the intruding organisms, SMs could also protect plants by scavenging cytotoxic reactive oxygen species (ROS), fortifying the physical barriers of plants, altering the behavior of insect herbivores, and serving as environmental signal molecules. For example, many phenolic compounds could promptly remove excessive ROS in plant cells following the rapid ROS burst in response to environmental stimuli to avoid oxidative damage [9, 10]. Some phenolic compounds could also cross-link the cellulose fibers in plant cell walls to resist physical damage inflicted by insect attack and pathogen penetration [11]. Diverse classes of SMs including methyl salicylates [12], fatty acid derivatives [13], and terpenoids [14] have been demonstrated to either repel insect herbivores, or attract their natural enemies to protect the plants under attack. Some volatile compounds could also be perceived by neighboring plants as warning signals [15–21].

While plants are usually protected by the SMs that they produce, co-evolving pests have sometimes evolved measures to take

advantage of plant SMs for their own benefits. For example, the hallmark SMs of maize (*Zea mays*), benzoxazinoids, are used by the specialized underground herbivore western corn rootworm (*Diabrotica virgifera*) as a reliable chemical signal to locate maize roots [22]. These specialized herbivores could not only tolerate the toxicity of benzoxazinoids, but also exploit the metal ion-chelating property of these compounds to supplement iron acquisition in their diets [23], and sequester the toxic compounds to protect themselves from natural enemies [24]. The delivery of plant SMs through complex food chains further extends the ecological significance of these natural products. Indeed, a recent experimental ecology study has demonstrated that allelic variation in a plant-specialized metabolic gene could enhance the persistence of the local food web [25].

Attraction of animal pollinators

The evolutionary success of flowering plants has been associated with insect-mediated cross-pollination. While the fundamental rewards provided to insect pollinators are primary metabolites such as carbohydrates (i.e. nectar) and proteins (i.e. pollen), SMs play critical roles in pollinator attraction by adding colors and fragrances to the floral display [26, 27]. From early field ecology observations to recent molecular genetics dissections, studies in plant-pollinator interactions have transcended multiple biological scales. A classic example was illustrated by a series of studies on *Mimulus* species. Bradshaw [28] identified the *YELLOW UPPER* locus, which influenced the petal color by regulating tissue-specific carotenoid metabolism, and this variation in floral coloration was associated with either bees or hummingbirds as the predominant pollinators in their natural habitats. The authors then hybridized these species to generate an experimental near iso-genic population with the different *YELLOW UPPER* alleles introgressed into the reciprocal genetic background, and demonstrated that the native pollinator visitation rate was negatively influenced by this introgression [29, 30]. More examples demonstrating the function of plant SMs in plant-pollinator interactions have been elegantly synthesized in a recent review by Fattorini and Glover [31]. Interestingly, some pollinator-attracting plant SMs could also exert dual functions as a defensive metabolite against florivorous insects [32].

Modulation of rhizospheric microbiota

As early as the first century AD, it was observed that some plants, such as black walnut (*Juglans nigra*), were capable of inhibiting the growth of neighboring plants, a phenomenon known as allelopathy. It has since been shown that allelopathic interactions between plants are mediated by toxic SMs exuded into the rhizosphere [33]. In the last decades, researchers have demonstrated that root exuded SMs are not limited to toxic compounds, but could rather function as critical chemical messengers that facilitate interactions between plants and soil microbes. Textbook examples on this topic include fabaceous flavones, which induce expression of nodulation genes in *Rhizobium* bacteria [34], and strigolactones, which stimulate branching of germinating arbuscular mycorrhizae fungi [35]. More recently, technological advances in environmental DNA sequencing and microbe culturing have revealed the robust dynamics of plant-microbiota interactions in the rhizosphere. To date, the rhizospheric microbiota-engineering function has been demonstrated for a long list of exuded SMs including benzoxazinoids [23, 36], phenolics [37, 38], and terpenoids [39, 40]. One of the less studied topics in this area is how the SMs in root

exudate are exported on the cellular level. In the model species *Arabidopsis thaliana*, exudation of coumarins was facilitated by an ATP-binding cassette-containing transporter [41]. Recent evidences from our group suggest that a Multidrug and Toxic Compound Extrusion transporter expressed in melon (*Cucumis melo*) roots is responsible for the exudation of a triterpene compound, cucurbitacin B, which can suppress soil-borne fungal pathogens by enriching specific taxa of beneficial bacteria [42].

Adaptation toward abiotic stresses

In addition to protecting plants against biotic stressors, SMs also participate in the plant adaptive processes toward various abiotic stresses. For example, enhanced accumulation of flavonoids in Tibetan barley (*Hordeum vulgare*, qinke) was proposed to be an adaptive mechanism to protect against exacerbated ultraviolet-B irradiation of the Tibetan plateau [43]. Concentrations of various SMs were significantly altered in maize plants grown under low-phosphorus conditions [44]. Most recently, genome-wide association studies followed by candidate validation with genetic mutant analyses have revealed the functionality of a number of maize SMs in drought and salt resistance [45, 46].

SMs used as flavor, nutrition, and medicine

Plants are the primary producers of our ecosphere. While the primary metabolites in plants provide essential nutrients to humans, plant SMs are valued for their flavoring, nutraceutical, and medicinal properties. Current metabolomics methods have enabled researchers to pinpoint and quantify the keystone plant metabolites that are associated with consumer preference or specific therapeutic functions. Some of these examples are delineated in this section (Fig. 2).

Flavors

Sweetness is perhaps one of the most desirable traits for freshly consumed fruits and vegetables (Fig. 2a). Recent studies in tomato (*Solanum lycopersicum*) and blueberry (*Vaccinium corymbosum*) have demonstrated that the sense of sweetness is not only associated with sugar content, but also promoted by various volatile compounds [47, 48]. In contrast, organic acid contents have been reported to be negatively correlated with perceived levels of sweetness in tomato [48], apple (*Malus spp.*) [49–51], and kiwifruit (*Actinidia arguta*) [52, 53]. The complex biochemical composition associated with the sense of sweetness provides the opportunity to improve this important quality trait by targeting the SMs of these crops rather than directly manipulating their sugar metabolism, so as to avoid any potential yield penalty. Furthermore, some plant SMs, such as stevia glycosides and mogrosides, are inherently sweet, and are widely exploited as natural low-caloric sweeteners. Besides sweetness, different fruits and vegetables often have their characteristic tastes and aroma, which are also endowed by their SMs [54]. Distinctive aroma of cinnamon (*Cinnamomum verum*) [55], melon [56], cucumber (*Cucumis sativa*) [57, 58], pumpkin (*Cucurbita spp.*) [59], and kiwifruit [60, 61] are determined by their characteristic volatile phenolics and fatty acid derivatives. A variety of volatile terpenoids and uncommon nitrogen-containing volatiles have also been shown to contribute to the aroma of tomato [62–65].

Since SMs mainly function as defensive compounds for plants in nature, most SMs are toxic and/or have unpleasant taste to humans [66, 67] (Fig. 2a). The presence of these SMs in the edible parts of crop plants has often been negatively selected during

domestication. In cucurbits, a mutation in a fruit-specific transcription factor abolished the biosynthesis of bitter-tasting triterpenoids in fruits [68]. Similar changes in tissue-specific transcription regulation have recently been demonstrated to lower the level of toxic cyanogenic glycosides in domesticated almond kernels [69]. In tomato fruits, a tissue- and developmental stage-specific nitrate/peptide family transporter could export the bitter α -tomatine from vacuole to cytosol during fruit ripening for prompt degradation [70]. On the other hand, some SMs that are nontoxic but have pungent tastes to human are preserved during crop domestication. For example, assorted phenolic acids contribute to bitterness and astringency in tea leaves (*Camellia sinensis*), which contribute to the complex flavor of tea [71, 72]. In olive fruits (*Olea spp.*), the content of bitter-tasting oleuropein needs to be properly controlled during olive breeding for a balanced taste in the fruits [73, 74].

While SMs can directly affect the taste and aroma of freshly consumed fruits and vegetables, food processing for most staple crops has a significant impact on the biochemical composition, and hence the flavor of food. Maillard reaction is one of the most common chemical reactions that occurs during food processing, which produces flavorful volatiles [75]. For example, furans contribute the caramel-like odor of heated carbohydrates, and pyrazines and methional provide the characteristic aroma of baked potatoes [76].

Nutraceuticals and medicine

Besides serving as flavor molecules, some plant SMs, such as vitamins, are essential components of human nutrition. Vitamin C (L-ascorbic acid) is the most classic example of essential nutrient that is exclusively supplied by fresh plants since humans and other primates have lost their own capacity to biosynthesize this vital molecule [77]. Similarly, plants are an important source of diverse B vitamins including B1 (thiamin from grains), B5 (pantothenic acid from sunflower seeds and potatoes), and B9 (folate from spinach) [78–80]. Recent technological advances in genetic engineering and genome editing have inspired a series of bio-fortification efforts to enhance the nutritional quality of staple crops, for more accessible nutrition in under-developed regions. For example, five carotenoid biosynthetic genes have been specifically expressed in maize kernels to promote accumulation of a variety of carotenoids, which can be converted to vitamin A in animals after ingestion [81]. Chickens raised on these bio-fortified feed produced high-carotenoid poultry meat and were more resistant against a protozoan parasite [82]. Most recently, a specific isoform of 7-dehydrocholesterol reductase in tomato was knocked-out to allow accumulation of 7-dehydrocholesterol, which would be simultaneously converted into vitamin D upon ultraviolet irradiation [83].

Plant SMs have been the active ingredients of many traditional medicines for millennia. In the last 2 centuries, detailed analytical chemistry studies have revealed the principal active compounds associated with various medicinal plants (Fig. 2b). For example, the dried aerial parts of the herb *Artemisia annua* have been used in China for centuries to treat fever and malaria [84]. The main active compound, artemisinin, is highly effective against malaria parasites *Plasmodium falciparum*, without being toxic to humans or animals at the same concentration [85]. Cucurbitacins, a class of triterpenoids widely distributed in cucurbit plants, have been exploited for their anti-inflammatory, hepatoprotective, and anti-proliferative effects [86]. Paclitaxel, a diterpene from *Taxus* species, is a well-known chemotherapy agent against various cancers [87].

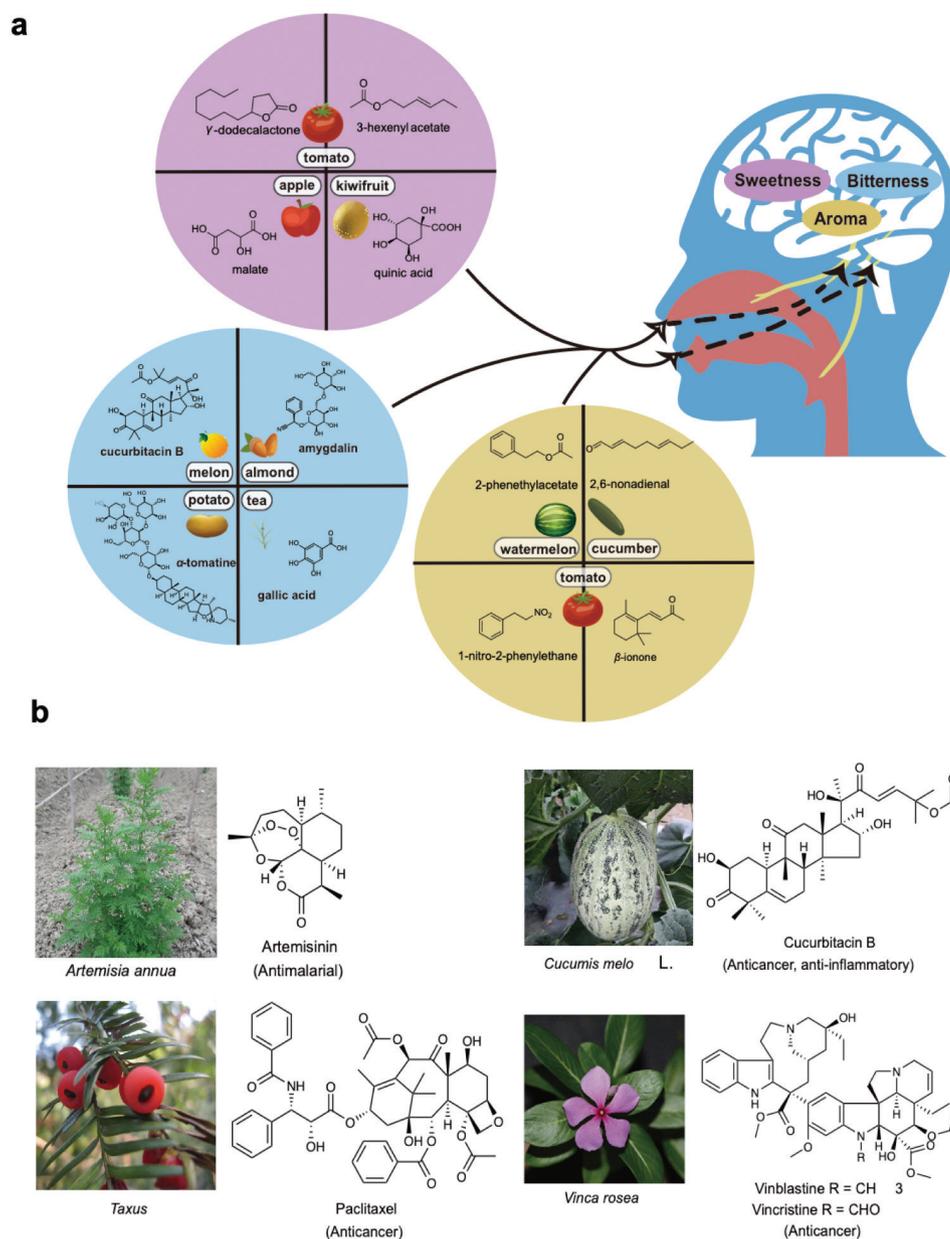


Figure 2 Flavoring and medicinal values of plant SMs for humans. Representative plant SMs and their natural botanical sources are used as flavors (a) and medicines (b).

Other classic chemotherapeutic compounds, vinblastine and vincristine, are naturally sourced from *Vinca rosea* for treatment of Hodgkin's lymphoma and childhood lymphoblastic leukemia [88]. More of the current progress and future perspectives of the use of plant SMs as a source of medicine have been recently reviewed by Jacobowitz and Weng [89].

Elucidating biosynthetic pathways of SMs in plants

Since plant SMs are appreciated for their flavoring and therapeutic functions for human, understanding the biosynthesis of SMs has always been one of the top priorities in plant biology research. Early efforts to clarify plant-specialized metabolism relied on traditional biochemistry techniques such as stable isotope feeding and protein chromatography coupled with *in vitro* enzymatic assays. In the last decade, rapid development in multi-omics

tools have enabled a tremendous leap in specialized metabolic pathway elucidation in plants. This section summarizes studies exemplifying these current methods (Fig. 3).

Metabolite-gene expression correlation

Plant SMs are well-known for their tissue-specific distribution and inducible accumulation. This spatiotemporal heterogeneity allows researchers to identify potential biosynthetic genes by looking for statistical correlation between experimentally determined gene expression and target metabolite abundance (Fig. 3a). Since the 2010s, this simple yet effective approach has been successfully applied to identify key biosynthetic enzymes of a number of renowned plant SMs such as podophyllotoxin [90], vinblastine and vincristine [88], and colchicine [91, 92]. This method also contributed to the recent characterization of root-specific triterpenoid biosynthetic pathways in *Arabidopsis* [39], and

proto-limonoid biosynthesis in rutaceaeous plants [93]. One of the key advantages of this approach is that it only requires a few samples for transcriptomics and targeted metabolomics analyses to identify potential candidate genes. For example, it took 12 RNAseq samples to identify the candidate biosynthetic genes of podophyllotoxin through a co-expression analysis [90]. Also, transcriptome data can be *de novo* assembled for quantitative analyses, which bypasses the requirement to sequence and assemble large plant genomes. These advantages have significantly enhanced the practical viability of correlation analyses in genetic dissection of plant-specialized metabolism, especially in non-model medicinal plant species that lack diverse germplasms and reliable reference genomes. However, it should be noted that some SMs are transported away from their main tissues of biosynthesis and stored elsewhere (e.g. nicotine, cardenolide), which would hence nullify the presumed cross-tissue correlation of the abundance of the SM with the expression of its biosynthetic genes.

Forward genetics

In addition to studies of rare and medicinal SMs in non-model plant species, examination of SMs in staple and economic crops is another fast-developing area in plant research. Here, thanks to the diverse germplasm and genotypic information accumulated from generations of crop breeders, the forward genetics approach is preferred to identify the biosynthetic and/or regulatory genes associated with the target SMs (Fig. 3b). In maize, e.g. classic bi-parental mapping populations are now routinely adopted to locate the genetic loci associated with SMs [94–98]. Though genetic mapping with bi-parental populations alone rarely gets down to single-gene resolution, addition of a clear biochemical hypothesis and gene expression data are usually sufficient to narrow down the list of candidate genes to a range that can be experimentally validated.

Compared to bi-parental populations, the more diverse genome-wide association panels offer much more frequent recombination, and hence much higher mapping resolution. Indeed, studies in rice (*Oryza sativa*) [99], maize [100, 101], and tomato [48, 102] have demonstrated that single-gene level mapping can be achieved with sufficient molecular marker coverage. On the contrary to what was once commonly believed, high-resolution association genetics

revealed that the genetic architecture of plant SMs is not necessarily simpler than that of composite traits such as grain yield [101]. This is probably reflective of the complex *cis* and *trans* regulation of specialized metabolism in plants. Nevertheless, current genetic dissection of SM biosynthetic pathways in crop species have demonstrated the potential to enhance crop quality and to protect crop plants against environmental stresses through genetic engineering [102, 103].

Genome-mining based on operon-like gene clusters

Since the first report of a cluster of five genes (Bx1 through Bx5) that are required for synthesis of antimicrobial compound, 2,4-dihydroxy-1,4-benzoaxin-3-one (DIBOA), at a narrow region on a maize chromosome [104], another metabolic gene cluster for biosynthesis of antimicrobial compounds, avenacins, was defined by genetic and bacterial artificial chromosome analysis in oats [105, 106]. Since then, similar biosynthetic gene clusters (BGCs) have been reported across diverse plant species for the production of many plant SMs [107] (Fig. 3c; Table 1). A potential evolutionary advantage of these BGCs in plant genomes is that the physical proximity between these enzyme-encoding genes could facilitate shared epigenomic modifications and hence coordinated expression regulation [67, 68, 131–133]. For example, tissue-specific *trans* regulators can directly bind to the conserved E-box elements in the promoters of nine clustered cucurbitacin biosynthetic genes, and coordinately activate their transcriptions to produce the target compound [67, 68]. Interestingly, the later studies have demonstrated that BGCs can include not only enzyme-encoding genes, but also genes that encode specific transporters of the SMs produced [134, 135].

As the idea of plant BGCs receives accumulating experimental support, bioinformatic tools that predict BGCs solely based on genomic information have also been developed to annotate previously unknown SM biosynthetic genes [136]. A recent study on wheat (*Triticum aestivum*) and a related model grass species exemplified this reverse chemical genetics approach, where the authors started by annotating potential BGCs in recently published genomes and ended up identifying previously unknown SMs in these species by combining pathogen-inducible gene expression data and functional validation experiments [137].

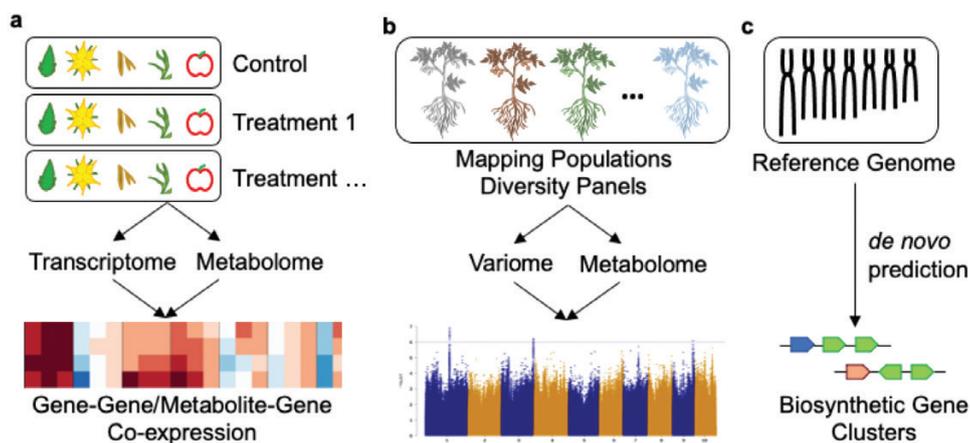


Figure 3 Current methods for deciphering biosynthetic pathways of plant SMs. Schematic representation of (a) correlation analyses based on multi-tissue and multi-treatment transcriptome and metabolome data; (b) forward genetics with *ad hoc* mapping populations and/or genetic diversity panels; (c) *de novo* prediction of operon-like BFCs in reference-grade genomes.

Table 1 Examples of characterized plant BGCs.

Compound(s)/pathway	Class	Plant species	Role in plant	References
Avenacins	Triterpenes	<i>Avena</i> sp.	Antifungal	[8, 105]
Arabidiol/arabidin	Triterpenes	<i>Arabidopsis thaliana</i>	Anti-oomycete, microbiome modulation	[39, 108]
Thalianol/thalianin	Triterpenes	<i>A. thaliana</i> <i>Arabidopsis lyrata</i>	Microbiome modulation	[39, 108, 109]
Marneral	Triterpenes	<i>A. thaliana</i>	Unknown	[108]
Tirucallol	Triterpenes	<i>Capsella rubella</i>	Unknown	[109]
Euphol	Triterpenes	<i>Brassica rapa</i>	Unknown	[109]
Cucurbitacins	Triterpenes	<i>Cucumis sativus</i> <i>Cucumis melo</i> <i>Citrullus lanatus</i>	Antibacterial, antifungal, insecticidal, anti-herbivore	[68, 110]
Yossosides	Triterpenes	<i>Spinacia oleracea</i>	Unknown	[111]
20-hydroxy-betulnic acid	Triterpenes	<i>Lotus japonicus</i>	Unknown	[112]
Momilactones	Diterpenes	<i>Oryza</i> sp. <i>Echinochloa crusgalli</i> <i>Calohyppnum plumiforme</i>	Antibacterial, antifungal, allelopathic	[113, 114]
Phytocassanes/oryzalides	Diterpenes	<i>Oryza sativa</i>	Antibacterial, antifungal	[115]
Casbene diterpenoids	Diterpenes	<i>Ricinus communis</i> <i>Euphorbia peplus</i> <i>Jatropha curcas</i>	Antifungal, antibacterial	[116]
5,10-diketo-casbene	Diterpenes	<i>O. sativa</i>	Antifungal, antibacterial	[117]
Various monoterpenes and diterpenes	Diterpenes/monoterpenes	<i>Solanum</i> sp.	Antibacterial, antifungal	[118]
Lycosantanolol	Diterpenes	<i>Solanum lycopersicum</i>	Unknown	[119]
5- <i>epi</i> -jinkoheremol/ debneyol	Sesquiterpenes	<i>Catharanthus roseus</i>	Fungicidal activity	[120]
α -tomatine	Steroidal glycoalkaloids	<i>S. lycopersicum</i>	Antibacterial, antifungal, insecticidal	[121]
α -solanine α -chaconine	Steroidal glycoalkaloids	<i>Solanum tuberosum</i>	Antibacterial, antifungal, insecticidal	[121]
Noscapine	Benzylisoquinoline alkaloids	<i>Papaver somniferum</i>	Unknown	[122]
Thebaine	Benzylisoquinoline alkaloids	<i>P. somniferum</i>	Unknown	[123]
Hydroxycinnamoyl-tyramine conjugates	Phenolamides	<i>O. sativa</i>	Antibacterial, antifungal	[124]
Feruloylputrescine	Phenylpropanoids	<i>O. sativa</i>	Immunity, cell death	[125]
Dhurrin	Cyanogenic glucosides	<i>Sorghum bicolor</i>	Insecticidal, anti-herbivore	[3]
Linamarin Lotaustralin	Cyanogenic glucosides	<i>L. japonicus</i> <i>Manihot esculenta</i>	Insecticidal, anti-herbivore	[126]
α -/ β -/ γ -Hydroxynitrile glucosides	Hydroxynitrile glucosides	<i>Hordeum vulgare</i>	Unknown	[127]
Falcarindiol	Fatty acids	<i>S. lycopersicum</i>	Antifungal, antibacterial	[128]
β -diketones	Polyketides	<i>H. vulgare</i> <i>Triticum turgidum</i>	Forming physical barrier on leaf surface	[129]
DIBOA/DIMBOA	Benzoxazinoids	<i>Zea mays</i>	Antibacterial, antifungal, insecticidal, allelopathic	[104]
Various acylsugars	Acylsugars	<i>Solanum</i> sp.	Antifungal, insecticidal, anti-herbivore	[130]

It is noteworthy that the functionally related but structurally non-homologous plant BGCs bear close resemblance to prokaryotic operons, since members of both BGCs and operons are tightly linked in the genomes and tend to be co-expressed [138]. That being said, BGCs and operons most likely arise independently during evolution, given the distant phylogenetic separation between plants and bacteria. While prokaryotic operons are well-known to horizontally transfer between co-habituating microbes, plant BGCs tend to arise independently in each phylogenetic lineage which then go on to differential evolutionary trajectories [67, 139–141]. However, recent studies of the classic benzoxazinoid BGCs in maize and wheat suggested that this cluster may have been horizontally transferred between their ancestors, followed by independent evolution of a functionally homologous downstream modification enzyme [142, 143]. It will take further studies to determine whether this proposed horizontal BGC transfer is an isolated case or a harbinger of a more common pattern in plant evolution.

Synthetic biology for sustainable supply of plant SMs

Though plant SMs are widely utilized by human beings, the low abundance of valuable SMs in *planta* and the low efficiency in the extraction process limit their supply from native sources. The expanding inventory of functionally characterized plant enzymes, coupled with advancing synthetic biology tools, provides opportunities to sustainably produce plant SMs in engineered chassis microbes such as *Escherichia coli* and *Saccharomyces cerevisiae* [109, 144–148]. However, heterologous plant SM biosynthetic pathways

in microbes often perturb the innate metabolic balance of the host, resulting in lowered target productivity or even host fatality [149]. Therefore, fine-tuning of the heterologous pathways is essential to unlock the full potential of microbial cell factories for plant SM production. Here we summarize several representative synthetic biology strategies in optimizing the native metabolic pathways of chassis microbes for enhanced production of heterologous compounds (Fig. 4).

Enzyme engineering (design and generation of highly efficient enzymes)

On the most basic molecular level, the efficiency of a metabolic pathway depends on the catalyzing efficiency of the enzymes. Assembly of naturally occurring plant enzymes into heterologous biosynthetic pathways in chassis organisms often suffers from lowered catalytic activity, narrowed substrate specificity, poor protein structure stability, and unexpected allosteric inhibition [150]. Traditionally, these problems had been addressed by increasing the titer of the heterologous enzymes with optimized gene expression levels or gene copy numbers, as well as utilization of superior alternative enzymes from other natural sources [151]. More recently, protein engineering methods such as directed evolution and site-directed mutagenesis have emerged as more efficient approaches to enhance the performance of rate-limiting enzymes [152] (Fig. 4a).

Directed evolution is one of the earliest protein engineering approaches as it does not require detailed knowledge of protein structure. By reiteratively selecting for the most efficient structural variant of the enzyme under study, this process enhances overall enzyme performance through evolution by selection. This

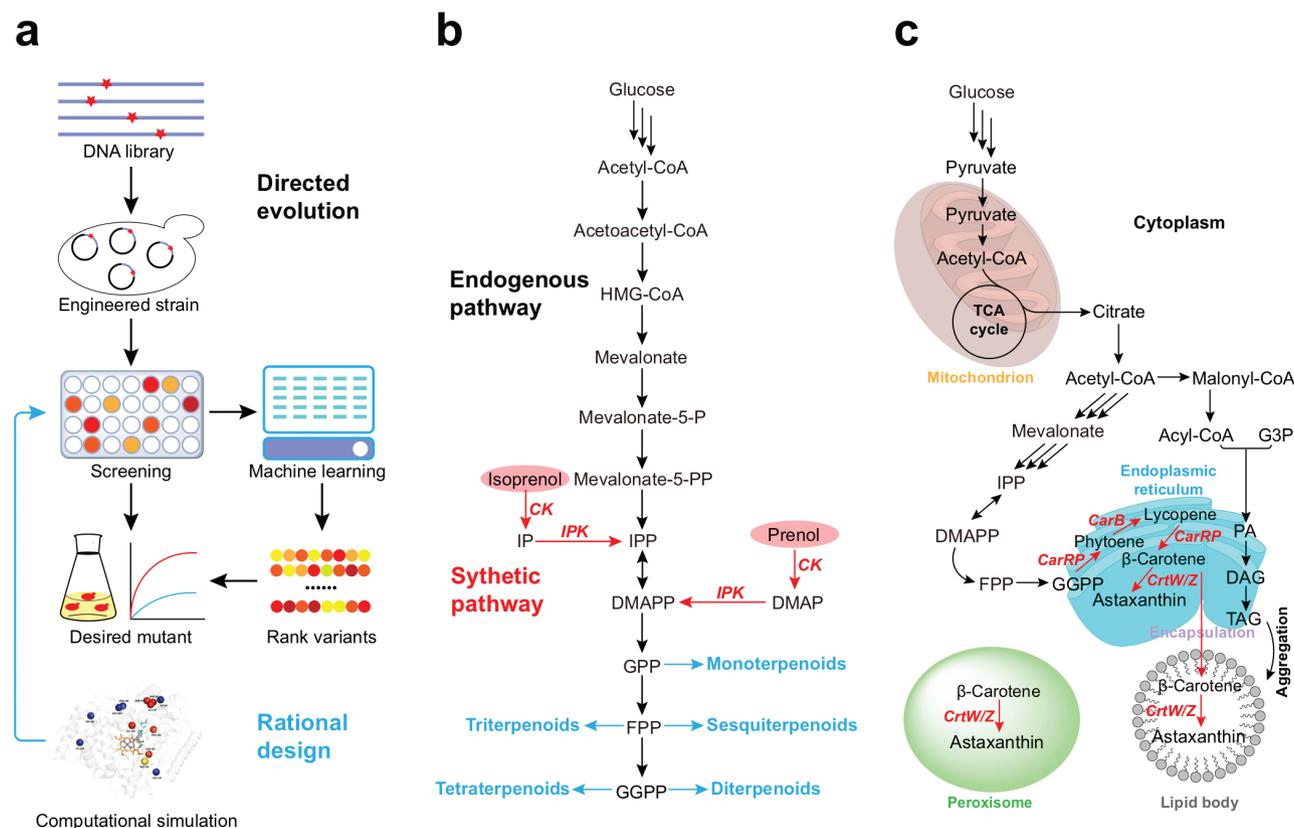


Figure 4 Synthetic biology strategies for enhanced plant SM production in microbial chassis. Schematic representation of (a) enzyme engineering through directed evolution or rational design; (b) introduction of synthetic orthogonal pathways; (c) compartmentalization of synthetic pathways.

approach has been adopted to enhance astaxanthin and various isoprenoids production in yeast and *E. coli* [153–156]. However, it should be noted that successful cases of enzyme engineering through directed evolution are almost exclusively reported for carotenoid-related pathways, as the distinctive pigmentation of the end product provides an easy proxy for high-throughput selection [157]. More recently, researchers have attempted to circumvent this limitation by incorporating machine-learning algorithms into directed evolution schemes to reduce the sequence space that had to be experimentally tested [158, 159].

The rapid development of crystallography-based structural biology has enabled structure-guided protein engineering by computer-assist molecular simulation. Through this so-called site-directed mutagenesis approach, key amino acid residues that affect enzyme activities can be identified by resolving the crystal structure of the target protein and simulating possible molecular interactions. The protein structure can then be optimized by designing proper amino acid substitution schemes [150]. For example, modification of a glycosyltransferase through structure-based rational design achieved a 1800-fold increase in catalytic efficiency for the production of ginsenoside Rh2, a potential anticancer compound [160]. This approach was also applied to improve substrate selectivity and protein stability of enzymes [161]. One of the main bottlenecks for broader application of site-directed mutagenesis of enzymes is the difficulty to obtain precise protein structures experimentally. Recent progress in computational protein structure prediction algorithms (e.g. AlphaFold and RoseTTA-Fold) has significantly promoted the prediction accuracy, reaching the same level of accuracy as the experimentally resolved protein structures in many cases [162, 163]. This combination of prediction accuracy and practical flexibility could enable site-directed improvement of more biosynthetic enzymes of plant SMs.

Synthetic orthogonal pathways (modulation of microbial chassis)

While targeted protein engineering could enhance the metabolic flux within the engineered SM pathways, the metabolic bottlenecks in microbial chassis could also be positioned further upstream, limiting the supply of chemical precursors. In these cases, synthetic orthogonal pathways are built to provide an additional route of precursor supply without perturbing the native pathway to avoid undesirable side-effects [164] (Fig. 4b). This strategy has been widely adopted in the case of heterologous terpenoid production, where the canonical terpene precursors are tightly coupled to sterol biosynthesis essential for the survival chassis organisms. Although heterologous introduction of the entire mevalonate pathway into *E. coli* to replace the native methylerythritol 4-phosphate pathway is now commonly adopted [165–167], both pathways are strongly coupled with the central carbon metabolism and are tightly regulated, thus limiting the availability of the terpenoid precursors, isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). Recently, a synthetic orthogonal pathway was developed to produce IPP and DMAPP from isoprenol or prenol through two steps of phosphorylation [168]. This synthetic Isopentenol Utilization Pathway (IUP) is decoupled from the central carbon metabolism of *E. coli*, and hence can sustain a high-carbon flux toward terpene precursors without disturbing the normal metabolism of the chassis microbe. In addition, IUP only uses ATP as the sole co-factor, further reducing the potential impact of this synthetic pathway on the host metabolism through co-factor competition. A similar

approach of orthogonal IPP and DMAPP production from glycerol has also been reported, which eliminated the need to feed isopentenols in the culturing medium [169]. These orthogonal pathways have managed to enhance heterologous isoprenoid production in various microbial systems [168–171]. In another example, introduction of a heterologous geranyl diphosphate (GPP) synthase using neryl diphosphate as the predominant substrate has enhanced GPP production compared to the native yeast prenyltransferase, which in turn promoted the heterologous monoterpene accumulation [164].

Co-compartmentalization of biosynthetic pathway

In plants, specialized metabolism occurs in specific subcellular locations to promote metabolic efficiency and avoid autotoxicity by bioactive intermediates [89]. Inappropriate localization of a heterologous biosynthetic pathway in the chassis organism could result in a poor target compound yield, side-products accumulation, and/or toxicity to the host [172]. Therefore, current synthetic biology schemes emphasize the proper compartmentalization of heterologous biosynthetic pathways to ensure their proper function in microbial hosts (Fig. 4c).

Localization-aware metabolic engineering in eukaryotic host microbes has enhanced production of isoprenoids [173–177], alkaloids [178, 179], and fatty acid derivatives [180, 181] in various organelles. The value of proper pathway compartmentalization in synthetic biology is probably best exemplified by the recent reconstitution of the tropane alkaloid pathway in baker's yeast [182]. In this systematic engineering scheme, the authors heterologously expressed >20 carefully selected biosynthetic genes from a number of plant and microbe genomes to rebuild the tropane alkaloid pathway in different cellular compartments. During this process, the authors converted the littorine synthase (LS) into an artificial transmembrane protein to bypass the impediment in maturation and trafficking of the natural plant LS protein through yeast *trans*-Golgi network, a process that is required for the final production of the target compounds. To further facilitate the function of this engineered LS in yeasts, the matured protein was targeted to yeast vacuoles, which better mimics the plant tonoplast environment, where this enzyme naturally functions. The tropine substrates were transported into yeast vacuoles by an additional transporter protein encoded by a tobacco gene. In another recent example, the astaxanthin pathway was targeted to three compartments in *Y. lipolytica* to bring astaxanthin biosynthetic enzymes into close proximity with their precursor compound β -carotene [183]. This compartmentalization-facilitated physical proximity accelerated the conversion of β -carotene into astaxanthin, while significantly decreased the accumulation of metabolic intermediates during this process.

Concluding remarks and prospectives

From mediating plant–environment interactions to bringing flavor and health to humans, plant SMs play pivotal roles in our ecosphere and societies. Continued genetic dissection of the biosynthetic pathways and optimization of heterologous production systems are required for sustainable exploration and supply of high-value plant SMs to fulfill the ever-growing demand. Current technological advances in genome editing and high-resolution mass spectrometry-based metabolomics have in turn promoted functional studies of plant SMs in ecological contexts. For instance, accelerated biosynthetic gene identification and

subsequent production of targeted genetic mutants have facilitated examination of the ecological function of specific plant SMs on a single-pathway or single-compound resolution. On a more systematic scale, integration of mass spectra-based molecular network and information theory statistics can be adapted to test classic ecological theories, as recently been demonstrated with wild tobacco populations [184].

With the powerful tools of current multi-omics analytics and synthetic biology, plant biologists are marching into the uncharted biochemical diversity that remains in the enormous number of non-model species, and developing a variety of chassis organisms for efficient production of different plant SMs. We expect these modern technologies to bring new light into the utilization of plant-specialized metabolism for human health and well-beings.

Acknowledgements

We would like to thank the anonymous reviewers for constructive suggestions for improving the manuscript. This work is financially supported by the National Natural Science Foundation of China (31788103 to J.L.), the National Key R&D Program of China (2019YFA0906200 to S.H.), and the National Natural Science Foundation of China (31920103003 to X.Q.).

Author contributions

All authors contributed to the conception and drafting of the manuscript.

Conflict of interest

The authors declare that no conflict of interest exists.

References

- Mithöfer A, Boland W. Plant defense against herbivores: chemical aspects. *Annu Rev Plant Biol* 2012;**63**:431–50.
- Clay NK, Adio AM, Denoux C et al. Glucosinolate metabolites required for an Arabidopsis innate immune response. *Science* 2009;**323**:95–101.
- Gleadow RM, Moller BL. Cyanogenic glycosides: synthesis, physiology, and phenotypic plasticity. *Annu Rev Plant Biol* 2014;**65**:155–85.
- Xu SQ, Brockmoller T, Navarro-Quezada A et al. Wild tobacco genomes reveal the evolution of nicotine biosynthesis. *Proc Natl Acad Sci USA* 2017;**114**:6133–8.
- Zhou S, Richter A, Jander G. Beyond defense: multiple functions of benzoxazinoids in maize metabolism. *Plant Cell Physiology* 2018;**59**:1528–37.
- Li DP, Gaquerel E. Next-generation mass spectrometry metabolomics revives the functional analysis of plant metabolic diversity. *Annu Rev Plant Biol* 2021;**72**:867–91.
- Bai Y, Yang C, Halitschke R et al. Natural history-guided omics reveals plant defensive chemistry against leafhopper pests. *Science* 2022;**375**:eabm2948.
- Li JC, Halitschke R, Li DP et al. Controlled hydroxylations of diterpenoids allow for plant chemical defense without autotoxicity. *Science* 2021;**371**:255–60.
- Apel K, Hirt H. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu Rev Plant Biol* 2004;**55**:373–99.
- Waszczak C, Carmody M, Kangasjarvi J. Reactive oxygen species in plant signaling. *Annu Rev Plant Biol* 2018;**69**:209–36.
- Lee MH, Jeon HS, Kim SH et al. Lignin-based barrier restricts pathogens to the infection site and confers resistance in plants. *EMBO J* 2019;**38**:e101948.
- Hardie J, Isaacs R, Pickett JA et al. Methyl salicylate and (-)-(1R,5S)-myrtenal are plant-derived repellents for black bean aphid, *Aphis fabae* Scop. (Homoptera: Aphididae). *J Chem Ecol* 1994;**20**:2847–55.
- Ameye M, Allmann S, Verwaeren J et al. Green leaf volatile production by plants: a meta-analysis. *New Phytol* 2018;**220**:666–83.
- Wang H, Guo WF, Zhang PJ et al. Experience-induced habituation and preference towards non-host plant odors in ovipositing females of a moth. *J Chem Ecol* 2008;**34**:330–8.
- Rosenkranz M, Chen Y, Zhu P et al. Volatile terpenes - mediators of plant-to-plant communication. *Plant J* 2021;**108**:617–31.
- Zhou S, Jander G. Molecular ecology of plant volatiles in interactions with insect herbivores. *J Exp Bot* 2022;**73**:449–62.
- Hu L. Integration of multiple volatile cues into plant defense responses. *New Phytol* 2022;**233**:618–23.
- Erb M. Volatiles as inducers and suppressors of plant defense and immunity—origins, specificity, perception and signaling. *Curr Opin Plant Biol* 2018;**44**:117–21.
- Hu L, Zhang K, Wu Z et al. Plant volatiles as regulators of plant defense and herbivore immunity: molecular mechanisms and unanswered questions. *Curr Opin Insect Sci* 2021;**44**:82–8.
- Jing T, Du W, Gao T et al. Herbivore-induced DMNT catalyzed by CYP82D47 plays an important role in the induction of JA-dependent herbivore resistance of neighboring tea plants. *Plant Cell Environ* 2021;**44**:1178–91.
- Chen S, Zhang L, Cai X et al. (E)-Nerolidol is a volatile signal that induces defenses against insects and pathogens in tea plants. *Hortic Res* 2020;**7**:52.
- Robert CA, Veyrat N, Glauser G et al. A specialist root herbivore exploits defensive metabolites to locate nutritious tissues. *Ecology Lett* 2012;**15**:55–64.
- Hu L, Mateo P, Ye M et al. Plant iron acquisition strategy exploited by an insect herbivore. *Science* 2018;**361**:694–7.
- Zhang X, van Doan C, Arce CCM et al. Plant defense resistance in natural enemies of a specialist insect herbivore. *Proc Natl Acad Sci USA* 2019;**116**:23174–81.
- Barbour MA, Kliebenstein DJ, Bascombe J. A keystone gene underlies the persistence of an experimental food web. *Science* 2022;**376**:70–3.
- Gang DR. Evolution of flavors and scents. *Annu Rev Plant Biol* 2005;**56**:301–25.
- Grotewold E. The genetics and biochemistry of floral pigments. *Annu Rev Plant Biol* 2006;**57**:761–80.
- Bradshaw HD, Wilbert SM, Otto KG et al. Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). *Nature* 1995;**376**:762–5.
- Bradshaw HD, Schemske DW. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* 2003;**426**:176–8.
- Schemske DW, Bradshaw HD Jr. Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). *Proc Natl Acad Sci USA* 1999;**96**:11910–5.
- Fattorini R, Glover BJ. Molecular mechanisms of pollination biology. *Annu Rev Plant Biol* 2020;**71**:487–515.
- Kessler D, Bing J, Haverkamp A et al. The defensive function of a pollinator-attracting floral volatile. *Funct Ecol* 2019;**33**:1223–32.

33. Bais HP, Weir TL, Perry LG et al. The role of root exudates in rhizosphere interactions with plants and other organisms. *Annu Rev Plant Biol* 2006;**57**:233–66.
34. Redmond JW, Batley M, Djordjevic MA et al. Flavones induce expression of nodulation genes in *Rhizobium*. *Nature* 1986;**323**:632–5.
35. Akiyama K, Matsuzaki K, Hayashi H. Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature* 2005;**435**:824–7.
36. Cotton TEA, Petriacq P, Cameron DD et al. Metabolic regulation of the maize rhizobiome by benzoxazinoids. *ISME J* 2019;**13**:1647–58.
37. Voges M, Bai Y, Schulze-Lefert P et al. Plant-derived coumarins shape the composition of an Arabidopsis synthetic root microbiome. *Proc Natl Acad Sci USA* 2019;**116**:12558–65.
38. Harbort CJ, Hashimoto M, Inoue H et al. Root-secreted coumarins and the microbiota interact to improve iron nutrition in Arabidopsis. *Cell Host & Microbe* 2020;**28**:825–37.e6.
39. Huang AC, Jiang T, Liu YX et al. A specialized metabolic network selectively modulates Arabidopsis root microbiota. *Science* 2019;**364**:eaau6389.
40. Chen Q, Jiang T, Liu YX et al. Recently duplicated sesterterpene (C25) gene clusters in *Arabidopsis thaliana* modulate root microbiota. *Sci China Life Sci* 2019;**62**:947–58.
41. Fourcroy P, Siso-Terraza P, Sudre D et al. Involvement of the ABCG37 transporter in secretion of scopoletin and derivatives by Arabidopsis roots in response to iron deficiency. *New Phytol* 2014;**201**:155–67.
42. Zhong Y, Xun W, Wang X et al. Root-secreted bitter triterpene modulates the rhizosphere microbiota to improve plant fitness. *Nat Plants* 2022;**8**:887–96.
43. Zeng X, Yuan H, Dong X et al. Genome-wide dissection of co-selected UV-B responsive pathways in the UV-B adaptation of Qingke. *Mol Plant* 2020;**13**:112–27.
44. Luo BW, Ma P, Nie Z et al. Metabolite profiling and genome-wide association studies reveal response mechanisms of phosphorus deficiency in maize seedling. *Plant J* 2019;**97**:947–69.
45. Zhang F, Wu JF, Sade N et al. Genomic basis underlying the metabolome-mediated drought adaptation of maize. *Genome Biol* 2021;**22**:260.
46. Liang XY, Liu SY, Wang T et al. Metabolomics-driven gene mining and genetic improvement of tolerance to salt-induced osmotic stress in maize. *New Phytol* 2021;**230**:2355–70.
47. Gilbert JL, Guthart MJ, Gezan SA et al. Identifying breeding priorities for blueberry flavor using biochemical, sensory, and genotype by environment analyses. *PLoS One* 2015;**10**:e0138494.
48. Tieman D, Zhu G, Resende MF Jr et al. A chemical genetic roadmap to improved tomato flavor. *Science* 2017;**355**:391–4.
49. Wege S. Sweet or Sour? Important link between nitrate signaling and malate accumulation identified in Apple. *Plant Physiol* 2020;**183**:439–40.
50. Hu DG, Sun CH, Ma QJ et al. MdMYB1 regulates anthocyanin and malate accumulation by directly facilitating their transport into vacuoles in apples. *Plant Physiol* 2016;**170**:1315–30.
51. Zhang QY, Gu KD, Cheng LL et al. BTB-TAZ domain protein MdBT2 modulates malate accumulation and vacuolar acidification in response to nitrate. *Plant Physiol* 2020;**183**:750–64.
52. Nishyama L, Fukuda T, Shimohashi A et al. Sugar and organic acid composition in the fruit juice of different *Actinidia* varieties. *Food Sci Technol Res* 2008;**14**:67–73.
53. Wang RC, Shu P, Zhang C et al. Integrative analyses of metabolome and genome-wide transcriptome reveal the regulatory network governing flavor formation in kiwifruit (*Actinidia chinensis*). *New Phytol* 2022;**233**:373–89.
54. Schwab W, Davidovich-Rikanati R, Lewinsohn E. Biosynthesis of plant-derived flavor compounds. *Plant J* 2008;**54**:712–32.
55. Atkinson RG. Phenylpropenes: occurrence, distribution, and biosynthesis in fruit. *J Agric Food Chem* 2016;**66**:2259–72.
56. Freilich S, Lev S, Gonda I et al. Systems approach for exploring the intricate associations between sweetness, color and aroma in melon fruits. *BMC Plant Biol* 2015;**15**:71.
57. Buescher RH, Buescher RW. Production and stability of (E,Z)-2,6-nonadienal, the major flavor volatile of cucumbers. *J Food Sci* 2001;**66**:357–61.
58. Wei G, Tian P, Zhang FX et al. Integrative analyses of nontargeted volatile profiling and transcriptome data provide molecular insight into VOC diversity in cucumber plants (*Cucumis sativus*). *Plant Physiol* 2016;**172**:603–18.
59. Maarse H. *Volatile Compounds in Foods and Beverages*. New York: Routledge, 2017, 198–277.
60. Gunther CS, Marsh KB, Winz RA et al. The impact of cold storage and ethylene on volatile ester production and aroma perception in 'Hort16A' kiwifruit. *Food Chem* 2015;**169**:5–12.
61. Zhang A, Zhang Q, Li J et al. Transcriptome co-expression network analysis identifies key genes and regulators of ripening kiwifruit ester biosynthesis. *BMC Plant Biol* 2020;**20**:103.
62. Buttery R. Quantitative and sensory aspects of flavor of tomato and other vegetables and fruits. *Flavor Science: Sensible Principles and Techniques*. Washington DC: American Chemical Society, 1993, 259–286.
63. Baldwin EA, Scott JW, Shewmaker CK et al. Flavor trivia and tomato aroma: biochemistry and possible mechanisms for control of important aroma components. *HortScience* 2000;**35**:1013–22.
64. Liscombe DK, Kamiyoshihara Y, Ghironzi J et al. A flavin-dependent monooxygenase produces nitrogenous tomato aroma volatiles using cysteine as a nitrogen source. *Proc Natl Acad Sci USA* 2022;**119**:e2118676119.
65. Vogel JT, Tieman DM, Sims CA et al. Carotenoid content impacts flavor acceptability in tomato (*Solanum lycopersicum*). *J Sci Food Agric* 2010;**90**:2233–40.
66. Friedman M. Potato glycoalkaloids and metabolites: roles in the plant and in the diet. *J Agric Food Chem* 2006;**54**:8655–81.
67. Zhou Y, Ma YS, Zeng JG et al. Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. *Nat Plants* 2016;**2**:16183.
68. Shang Y, Ma YS, Zhou Y et al. Biosynthesis, regulation, and domestication of bitterness in cucumber. *Science* 2014;**346**:1084–8.
69. Sanchez-Perez R, Pavan S, Mazzeo R et al. Mutation of a bHLH transcription factor allowed almond domestication. *Science* 2019;**364**:1095–8.
70. Kazachkova Y, Zemach I, Panda S et al. The GORKY glycoalkaloid transporter is indispensable for preventing tomato bitterness. *Nat Plants* 2021;**7**:468–80.
71. Xu YQ, Zhang YN, Chen JX et al. Quantitative analyses of the bitterness and astringency of catechins from green tea. *Food Chem* 2018;**258**:16–24.
72. Yang C, Hu ZY, Lu ML et al. Application of metabolomics profiling in the analysis of metabolites and taste quality in different subtypes of white tea. *Food Res Int* 2018;**106**:909–19.
73. Perez AG, Leon L, Sanz C et al. Fruit phenolic profiling: a new selection criterion in olive breeding programs. *Front Plant Sci* 2018;**9**:241.
74. Rodríguez MM, Ópez C, Hong B et al. Two bi-functional cytochrome P450 CYP72 enzymes from olive (*Olea europaea*) catalyze

- the oxidative C-C bond cleavage in the biosynthesis of seco-iridoids – flavor and quality determinants in olive oil. *New Phytol* 2021;**229**:2288–301.
75. Everts S. The Maillard reaction turns 100. *Chem Eng News* 2012;**90**:58–60.
 76. Duckham SC, Dodson AT, Bakker J et al. Effect of cultivar and storage time on the volatile flavor components of baked potato. *J Agric Food Chem* 2002;**50**:5640–8.
 77. Office of Dietary Supplements. Vitamin C - Health Professional Fact Sheet. *Dietary Supplement Fact Sheets* 2021 2021-03-26 [cited 2022 2022-06-01].
 78. Office of Dietary Supplements. Folate - Health Professional Fact Sheet. *Dietary Supplement Fact Sheets* 2021 2021-03-29 [cited 2022 2022-06-01].
 79. Office of Dietary Supplements. Thiamin - Health Professional Fact Sheet. *Dietary Supplement Fact Sheets* 2021 2021-03-26 [cited 2022 2022-06-01].
 80. Office of Dietary Supplements. Pantothenic Acid - Health Professional Fact Sheet. *Dietary Supplement Fact Sheets* 2021 2021-03-26 [cited 2022 2022-06-01].
 81. Zhu C, Naqvi S, Breitenbach J et al. Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. *Proc Natl Acad Sci USA* 2008;**105**:18232–7.
 82. Nogareda C, Moreno JA, Angulo E et al. Carotenoid-enriched transgenic corn delivers bioavailable carotenoids to poultry and protects them against coccidiosis. *Plant Biotechnol J* 2016;**14**:160–8.
 83. Li J, Scarano A, Gonzalez NM et al. Biofortified tomatoes provide a new route to vitamin D sufficiency. *Nat Plants* 2022;**8**:611–6.
 84. Bhakuni RS, Jain DC, Sharma RP et al. Secondary metabolites of *Artemisia annua* and their biological activity. *Current Sci* 2001;**80**:35–48.
 85. Fidock DA, Rosenthal PJ, Croft SL et al. Antimalarial drug discovery: efficacy models for compound screening. *Nat Rev Drug Discovery* 2004;**3**:509–20.
 86. Chen XP, Bao JL, Guo JJ et al. Biological activities and potential molecular targets of cucurbitacins: a focus on cancer. *Anticancer Drugs* 2012;**23**:777–87.
 87. Xiong X, Gou J, Liao Q et al. The *Taxus* genome provides insights into paclitaxel biosynthesis. *Nat Plants* 2021;**7**:1026–36.
 88. Caputi L, Franke J, Farrow SC et al. Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science* 2018;**360**:1235–9.
 89. Jacobowitz JR, Weng JK. Exploring uncharted territories of plant specialized metabolism in the postgenomic era. *Annu Rev Plant Biol* 2020;**71**:631–58.
 90. Lau W, Sattely ES. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science* 2015;**349**:1224–8.
 91. Nett RS, Lau W, Sattely ES. Discovery and engineering of colchicine alkaloid biosynthesis. *Nature* 2020;**584**:148–53.
 92. Nett RS, Sattely ES. Total biosynthesis of the tubulin-binding alkaloid colchicine. *J Am Chem Soc* 2021;**143**:19454–65.
 93. Hodgson H, De La Pena R, Stephenson MJ et al. Identification of key enzymes responsible for protolimonoid biosynthesis in plants: opening the door to azadirachtin production. *Proc Natl Acad Sci USA* 2019;**116**:17096–104.
 94. Meihls LN, Handrick V, Glauser G et al. Natural variation in maize aphid resistance is associated with 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one glucoside methyltransferase activity. *Plant Cell* 2013;**25**:2341–55.
 95. Handrick V, Robert CA, Ahern KR et al. Biosynthesis of 8-O-methylated benzoxazinoid defense compounds in maize. *Plant Cell* 2016;**28**:1682–700.
 96. Zhou S, Kremling KA, Bandillo N et al. Metabolome-scale genome-wide association studies reveal chemical diversity and genetic control of maize specialized metabolites. *Plant Cell* 2019;**31**:937–55.
 97. Ding Y, Weckwerth PR, Poretsky E et al. Genetic elucidation of interconnected antibiotic pathways mediating maize innate immunity. *Nat Plants* 2020;**6**:1375–88.
 98. Förster C, Handrick V, Ding Y et al. Biosynthesis and antifungal activity of fungus-induced O-methylated flavonoids in maize. *Plant Physiol* 2022;**188**:167–90.
 99. Chen W, Gao Y, Xie W et al. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat Genet* 2014;**46**:714–21.
 100. Wen W, Li D, Li X et al. Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat Commun* 2014;**5**:3438.
 101. Zhou S, Zhang YK, Kremling KA et al. Ethylene signaling regulates natural variation in the abundance of antifungal acetylated diferuloylsucroses and *Fusarium graminearum* resistance in maize seedling roots. *New Phytol* 2019;**221**:2096–111.
 102. Zhu G, Wang S, Huang Z et al. Rewiring of the fruit metabolome in tomato breeding. *Cell* 2018;**172**:249–61.e12.
 103. Zhou S, Jander G. Engineering insect resistance using plant specialized metabolites. *Curr Opin Biotechnol* 2021;**70**:115–21.
 104. Frey M, Chomet P, Glawischnig E et al. Analysis of a chemical plant defense mechanism in grasses. *Science* 1997;**277**:696–9.
 105. Qi X, Bakht S, Leggett M et al. A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. *Proc Natl Acad Sci USA* 2004;**101**:8233–8.
 106. Qi X, Bakht S, Qin B et al. A different function for a member of an ancient and highly conserved cytochrome P450 family: From essential sterols to plant defense. *Proc Natl Acad Sci USA* 2006;**103**:18848–53.
 107. Polturak G, Osbourn A. The emerging role of biosynthetic gene clusters in plant defense and plant interactions. *PLoS Pathog* 2021;**17**:e1009698.
 108. Field B, Fiston-Lavier AS, Kemen A et al. Formation of plant metabolic gene clusters within dynamic chromosomal regions. *Proc Natl Acad Sci USA* 2011;**108**:16116–21.
 109. Liu Z, Suarez Duran HG, Harnvanichvech Y et al. Drivers of metabolic diversification: how dynamic genomic neighbourhoods generate new biosynthetic pathways in the Brassicaceae. *New Phytol* 2020;**227**:1109–23.
 110. Zhou Y, Ma Y, Zeng J et al. Convergence and divergence of bitterness biosynthesis and regulation in Cucurbitaceae. *Nat Plants* 2016;**2**:16183.
 111. Jozwiak A, Sonawane PD, Panda S et al. Plant terpenoid metabolism co-opts a component of the cell wall biosynthesis machinery. *Nat Chem Biol* 2020;**16**:740–8.
 112. Krokida A, Delis C, Geisler K et al. A metabolic gene cluster in *Lotus japonicus* discloses novel enzyme functions and products in triterpene biosynthesis. *New Phytol* 2013;**200**:675–90.
 113. Shimura K, Okada A, Okada K et al. Identification of a biosynthetic gene cluster in rice for momilactones. *J Biol Chem* 2007;**282**:34013–8.
 114. Mao L, Kawaide H, Higuchi T et al. Genomic evidence for convergent evolution of gene clusters for momilactone biosynthesis in land plants. *Proc Natl Acad Sci USA* 2020;**117**:12472–80.
 115. Swaminathan S, Morrone D, Wang Q et al. CYP76M7 is an *ent*-cadiene C11 α -hydroxylase defining a second multifunctional

- diterpenoid biosynthetic gene cluster in rice. *Plant Cell* 2009;**21**:3315–25.
116. King AJ, Brown GD, Gilday AD et al. Production of bioactive diterpenoids in the euphorbiaceae depends on evolutionarily conserved gene clusters. *Plant Cell* 2014;**26**:3286–98.
 117. Zhan C, Lei L, Liu Z et al. Selection of a subspecies-specific diterpene gene cluster implicated in rice disease resistance. *Nat Plants* 2020;**6**:1447–54.
 118. Liu Z, Cheema J, Vigouroux M et al. Formation and diversification of a paradigm biosynthetic gene cluster in plants. *Nat Commun* 2020;**11**:5354.
 119. Matsuba Y, Zi J, Jones AD et al. Biosynthesis of the diterpenoid lycosantalanol via nerylneryl diphosphate in *Solanum lycopersicum*. *PLoS One* 2015;**10**:e0119302.
 120. Liang J, An T, Zhu JX et al. Mining of the *Catharanthus roseus* genome leads to identification of a biosynthetic gene cluster for fungicidal sesquiterpenes. *J Nat Prod* 2021;**84**:2709–16.
 121. Itkin M, Heinig U, Tzfadia O et al. Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. *Science* 2013;**341**:175–9.
 122. Winzer T, Gazda V, He Z et al. A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Science* 2012;**336**:1704–8.
 123. Chen X, Hagel JM, Chang L et al. A pathogenesis-related 10 protein catalyzes the final step in thebaine biosynthesis. *Nat Chem Biol* 2018;**14**:738–43.
 124. Shen SQ, Peng M, Fang H et al. An *Oryza*-specific hydroxycinnamoyl tyramine gene cluster contributes to enhanced disease resistance. *Sci Bull* 2021;**66**:2369–80.
 125. Fang H, Shen SQ, Wang D et al. A monocot-specific hydroxycinnamoylputrescine gene cluster contributes to immunity and cell death in rice. *Sci Bull* 2021;**66**:2381–93.
 126. Takos AM, Knudsen C, Lai D et al. Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in *Lotus japonicus* and suggests the repeated evolution of this chemical defence pathway. *Plant J* 2011;**68**:273–86.
 127. Knoch E, Motawie MS, Olsen CE et al. Biosynthesis of the leucine derived α -, β - and γ -hydroxynitrile glucosides in barley (*Hordeum vulgare* L.). *Plant J* 2016;**88**:247–56.
 128. Jeon JE, Kim JG, Fischer CR et al. A pathogen-responsive gene cluster for highly modified fatty acids in tomato. *Cell* 2020;**180**:176–87.e19.
 129. Hen-Avivi S, Savin O, Racovita RC et al. A metabolic gene cluster in the wheat W1 and the barley Cer-cqu loci determines beta-diketone biosynthesis and glaucousness. *Plant Cell* 2016;**28**:1440–60.
 130. Fan P, Wang P, Lou YR et al. Evolution of a plant gene cluster in Solanaceae and emergence of metabolic diversity. *Elife* 2020;**9**:e56717.
 131. Nützmann HW, Doerr D, Ramírez-Colmenero A et al. Active and repressed biosynthetic gene clusters have spatially distinct chromosome states. *Proc Natl Acad Sci USA* 2020;**117**:13800–9.
 132. Yu N, Nützmann HW, MacDonald JT et al. Delineation of metabolic gene clusters in plant genomes by chromatin signatures. *Nucleic Acids Res* 2016;**44**:2255–65.
 133. Cárdenas PD, Sonawane PD, Pollier J et al. GAME9 regulates the biosynthesis of steroidal alkaloids and upstream isoprenoids in the plant mevalonate pathway. *Nat Commun* 2016;**7**:10654.
 134. Kellner F, Kim J, Clavijo BJ et al. Genome-guided investigation of plant natural product biosynthesis. *Plant J* 2015;**82**:680–92.
 135. Darbani B, Motawia MS, Olsen CE et al. The biosynthetic gene cluster for the cyanogenic glucoside dhurrin in *Sorghum bicolor* contains its co-expressed vacuolar MATE transporter. *Sci Rep* 2016;**6**:37079.
 136. Kautsar SA, Suarez Duran HG, Blin K et al. plantSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. *Nucleic Acids Res* 2017;**45**:W55–63.
 137. Polturak G, Dippe M, Stephenson MJ et al. Pathogen-induced biosynthetic pathways encode defense-related molecules in bread wheat. *Proc Natl Acad Sci USA* 2022;**119**:e2123299119.
 138. Chae L, Kim T, Nilo-Poyanco R et al. Genomic signatures of specialized metabolism in plants. *Science* 2014;**344**:510–3.
 139. Nützmann HW, Huang A, Osbourn A. Plant metabolic clusters—from genetics to genomics. *New Phytol* 2016;**211**:771–89.
 140. Kitaoka N, Zhang J, Oyagbenro RK et al. Interdependent evolution of biosynthetic gene clusters for momilactone production in rice. *Plant Cell* 2021;**33**:290–305.
 141. Li R, Zhang J, Li Z et al. Dissecting the labdane-related diterpenoid biosynthetic gene clusters in rice reveals directional cross-cluster phytotoxicity. *New Phytol* 2022;**233**:878–89.
 142. Li B, Förster C, Robert CAM et al. Convergent evolution of a metabolic switch between aphid and caterpillar resistance in cereals. *Sci Adv* 2018;**4**:eaat6797.
 143. Wu D, Jiang B, Ye C-Y et al. Horizontal transfer and evolution of the biosynthetic gene cluster for benzoxazinoids in plants. *Plant Commun* 2022;**3**:100320.
 144. Ajikumar PK, Xiao WH, Tyo KE et al. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* 2010;**330**:70–4.
 145. Galanie S, Thodey K, Trenchard JJ et al. Complete biosynthesis of opioids in yeast. *Science* 2015;**349**:1095–100.
 146. Luo X, Reiter MA, d'Espaux L et al. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* 2019;**567**:123–6.
 147. Paddon CJ, Westfall PJ, Pitera DJ et al. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 2013;**496**:528–32.
 148. Westfall PJ, Pitera DJ, Lenihan JR et al. Production of amorpha-diene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc Natl Acad Sci USA* 2012;**109**:E111–8.
 149. Lu H, Villada JC, Lee PK. Modular metabolic engineering for biobased chemical production. *Trends Biotechnol* 2019;**37**:152–66.
 150. Li C, Zhang R, Wang J et al. Protein engineering for improving and diversifying natural product biosynthesis. *Trends Biotechnol* 2020;**38**:729–44.
 151. Mak WS, Tran S, Marcheschi R et al. Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway. *Nat Commun* 2015;**6**:10005.
 152. Sinha R, Shukla P. Current trends in protein engineering: updates and progress. *Curr Protein Pept Sci* 2019;**20**:398–407.
 153. Zhou P, Li M, Shen B et al. Directed coevolution of β -carotene ketolase and hydroxylase and its application in temperature-regulated biosynthesis of astaxanthin. *J Agric Food Chem* 2019;**67**:1072–80.
 154. Lv X, Gu J, Wang F et al. Combinatorial pathway optimization in *Escherichia coli* by directed co-evolution of rate-limiting enzymes and modular pathway engineering. *Biotechnol Bioeng* 2016;**113**:2661–9.
 155. Chen H, Liu C, Li M et al. Directed evolution of mevalonate kinase in *Escherichia coli* by random mutagenesis for improved lycopene. *RSC Adv* 2018;**8**:15021–8.
 156. Jakočiūnas T, Pedersen LE, Lis AV et al. CasPER, a method for directed evolution in genomic contexts using mutagenesis and CRISPR/Cas9. *Metab Eng* 2018;**48**:288–96.

157. Packer MS, Liu DR. Methods for the directed evolution of proteins. *Nat Rev Genet* 2015;**16**:379–94.
158. Yang KK, Wu Z, Arnold FH. Machine-learning-guided directed evolution for protein engineering. *Nat Methods* 2019;**16**:687–94.
159. Wu Z, Kan SJ, Lewis RD *et al*. Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc Natl Acad Sci USA* 2019;**116**:8852–8.
160. Zhuang Y, Yang GY, Chen X *et al*. Biosynthesis of plant-derived ginsenoside Rh2 in yeast via repurposing a key promiscuous microbial enzyme. *Metab Eng* 2017;**42**:25–32.
161. Li D, Ma Y, Zhou Y *et al*. A structural and data-driven approach to engineering a plant cytochrome P450 enzyme. *Sci China Life Sci* 2019;**62**:873–82.
162. Jumper J, Evans R, Pritzel A *et al*. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;**596**:583–9.
163. Baek M, DiMaio F, Anishchenko I *et al*. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 2021;**373**:871–6.
164. Ignea C, Raadam MH, Motawia MS *et al*. Orthogonal monoterpenoid biosynthesis in yeast constructed on an isomeric substrate. *Nat Commun* 2019;**10**:3799.
165. Liu CL, Bi HR, Bai Z *et al*. Engineering and manipulation of a mevalonate pathway in *Escherichia coli* for isoprene production. *Appl Microbiol Biotechnol* 2019;**103**:239–50.
166. Zada B, Wang C, Park JB *et al*. Metabolic engineering of *Escherichia coli* for production of mixed isoprenoid alcohols and their derivatives. *Biotechnol Biofuels* 2018;**11**:210.
167. Yoshida R, Yoshimura T, Hemmi H. Reconstruction of the “archaeal” mevalonate pathway from the methanogenic archaeon *Methanosarcina mazei* in *Escherichia coli* cells. *Appl Environ Microbiol* 2020;**86**:e02889–19.
168. Chatzivasileiou AO, Ward V, Edgar SM *et al*. Two-step pathway for isoprenoid synthesis. *Proc Natl Acad Sci USA* 2019;**116**:506–11.
169. Clomburg JM, Qian S, Tan Z *et al*. The isoprenoid alcohol pathway, a synthetic route for isoprenoid biosynthesis. *Proc Natl Acad Sci USA* 2019;**116**:12810–5.
170. Lund S, Hall R, Williams GJ. An artificial pathway for isoprenoid biosynthesis decoupled from native hemiterpene metabolism. *ACS Synth Biol* 2019;**8**:232–8.
171. Rico J, Duquesne K, Petit JL *et al*. Exploring natural biodiversity to expand access to microbial terpene synthesis. *Microb Cell Fact* 2019;**18**:23.
172. Dusséaux S, Wajn WT, Liu Y *et al*. Transforming yeast peroxisomes into microfactories for the efficient production of high-value isoprenoids. *Proc Natl Acad Sci USA* 2020;**117**:31789–99.
173. Yee DA, DeNicola AB, Billingsley JM *et al*. Engineered mitochondrial production of monoterpenes in *Saccharomyces cerevisiae*. *Metab Eng* 2019;**55**:76–84.
174. Liu GS, Li T, Zhou W *et al*. The yeast peroxisome: a dynamic storage depot and subcellular factory for squalene overproduction. *Metab Eng* 2020;**57**:151–61.
175. Zhang C, Li M, Zhao GR *et al*. Harnessing yeast peroxisomes and cytosol acetyl-CoA for sesquiterpene α -humulene production. *J Agric Food Chem* 2020;**68**:1382–9.
176. Arendt P, Miettinen K, Pollier J *et al*. An endoplasmic reticulum-engineered yeast platform for overproduction of triterpenoids. *Metab Eng* 2017;**40**:165–75.
177. Kim JE, Jang IS, Son SH *et al*. Tailoring the *Saccharomyces cerevisiae* endoplasmic reticulum for functional assembly of terpene synthesis pathway. *Metab Eng* 2019;**56**:50–9.
178. Grewal PS, Samson JA, Baker JJ *et al*. Peroxisome compartmentalization of a toxic enzyme improves alkaloid production. *Nat Chem Biol* 2021;**17**:96–103.
179. Thodey K, Galanie S, Smolke CD. A microbial biomanufacturing platform for natural and semisynthetic opioids. *Nat Chem Biol* 2014;**10**:837–44.
180. Zhou YJ, Buijs NA, Zhu Z *et al*. Harnessing yeast peroxisomes for biosynthesis of fatty-acid-derived biofuels and chemicals with relieved side-pathway competition. *J Am Chem Soc* 2016;**138**:15368–77.
181. Sheng J, Stevens J, Feng X. Pathway compartmentalization in peroxisome of *Saccharomyces cerevisiae* to produce versatile medium chain fatty alcohols. *Sci Rep* 2016;**6**:26884.
182. Srinivasan P, Smolke CD. Biosynthesis of medicinal tropane alkaloids in yeast. *Nature* 2020;**585**:614–9.
183. Ma Y, Liu N, Greisen P *et al*. Removal of lycopene substrate inhibition enables high carotenoid productivity in *Yarrowia lipolytica*. *Nat Commun* 2022;**13**:572.
184. Li D, Halitschke R, Baldwin IT *et al*. Information theory tests critical predictions of plant defense theory for specialized metabolism. *Sci Adv*, **6**:eaaz0381.