

The Integration of Metabolomics and Next-Generation Sequencing Data to Elucidate the Pathways of Natural Product Metabolism in Medicinal Plants

Authors

Federico Scossa^{1,2}, Maria Benina³, Saleh Alseekh¹, Youjun Zhang^{1,3}, Alisdair R. Fernie^{1,3}

Affiliations

- 1 Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany
- 2 Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Rome, Italy
- 3 Center of Plant Systems Biology and Biotechnology, Plovdiv, Bulgaria

Key words

Papaver somniferum, Papaveraceae, *Catharanthus roseus*, Apocynaceae, alkaloids, caffeine, cannabinoids, ginseng, artemisinin, taxol

received February 8, 2018

revised April 6, 2018

accepted May 8, 2018

Bibliography

DOI <https://doi.org/10.1055/a-0630-1899>

Published online May 29, 2018 | *Planta Med* 2018; 84: 855–873 © Georg Thieme Verlag KG Stuttgart · New York | ISSN 0032-0943

Correspondence

Dr. Alisdair R. Fernie

Max Planck Institute of Molecular Plant Physiology
Am Mühlenberg 1, 14476 Potsdam-Golm, Germany
Phone: + 49 33 15 67 80, Fax: + 49 33 15 67 84 08
fernied@mpimp-golm.mpg.de

Correspondence

Dr. Federico Scossa

Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria
Via di Fioranello 52, 00134 Rome, Italy
Phone: + 39 06 79 34 81 11, Fax: + 39 06 79 34 01 58
federico.scossa@crea.gov.it

ABSTRACT

Plants have always been used as medicines since ancient times to treat diseases. The knowledge around the active components of herbal preparations has remained nevertheless fragmentary: the biosynthetic pathways of many secondary metabolites of pharmacological importance have been clarified only in a few species, while the chemodiversity present in many medicinal plants has remained largely unexplored. Despite the advancements of synthetic biology for production of medicinal compounds in heterologous hosts, the native plant species are often the most reliable and economic source for their production. It thus becomes fundamental to investigate the metabolic composition of medicinal plants to characterize their natural metabolic diversity and to define the biosynthetic routes *in planta* of important compounds to develop strategies to further increase their content. We present here a number of case studies for selected classes of secondary metabolites and we review their health benefits and the historical developments in their structural elucidation and characterization of biosynthetic genes. We cover the cases of benzoisoquinoline and monoterpene indole alkaloids, cannabinoids, caffeine, ginsenosides, withanolides, artemisinin, and taxol; we show how the “early” biochemical or the more recent integrative approaches—based on omics-analyses—have helped to elucidate their metabolic pathways and cellular compartmentation. We also summarize how the knowledge generated about their biosynthesis has been used to develop metabolic engineering strategies in heterologous and native hosts. We conclude that following the advent of novel, high-throughput and cost-effective analytical technologies, the secondary metabolism of medicinal plants can now be examined under the lens of systems biology.

Introduction

The first archeological evidences of the use of herbal remedies date back to prehistory: Neanderthals, for example, who have been long considered mainly meat-eaters, had instead already a good knowledge of the surrounding vegetation and adopted so-

phisticated diets: their dental plaques contained residues of several herbs, indicating the early consumption of plants, perhaps already for self-medication purposes [1].

Written records of the use of medicinal plants—including recipes for preparing decocts and extracts—were also common in Ancient Egypt, Greece, Rome, China, and in the Middle East.

During the modern era, a more rational approach began to be applied to the study of herbal medicines, as was the case with the discovery of the properties of foxglove (*Digitalis purpurea* L., family Plantaginaceae) to treat edema and heart failures [2, 3].

Since the beginning of the 19th century, in parallel with the development of the pharmaceutical industry, there was an impetus in the isolation of new compounds possessing a therapeutic or commercial potential. In 1805, morphine was isolated from the latex of opium poppies and went immediately into commercial production in Europe and the United States, where it soon reached widespread popularity as a pain relief medication [4]. After the discovery of morphine, many other compounds with therapeutic effects were isolated and purified from plants.

The botanical drugs we use today, like the ancient herbal remedies, are all examples of complex mixtures enriched in plant secondary metabolites. Along evolution, plants have in fact developed a vast array of chemical defenses to stand up against their enemies (herbivores, fungi) to attract pollinators or to disseminate various chemical signals in their surrounding environment. Secondary metabolites are present in all higher plants but display a large structural diversity: different taxa usually accumulate different classes of secondary metabolites, reflecting the adaptations to the various ecological niches plants colonized on Earth [5]. This is in contrast to the current knowledge about the role and distribution of primary metabolites (amino acids, organic acids, carbohydrates, etc.). Primary metabolites represent the intermediates of those metabolic pathways related to the basic processes of plant growth and development (e.g., glycolysis, TCA cycle, ATP (adenosine triphosphate) synthesis, Calvin-Benson cycle, etc.); as such, their presence is not confined to specific taxa, and the key metabolic steps for their biosynthesis and degradation are mostly conserved across the green lineage. So although most of the primary metabolic pathways have been well described in plants, both at the genetic and biochemical level, the elucidation of the pathways of secondary metabolites has lagged behind, due to their confined taxonomic distribution and inherent difficulties in purifying them from natural sources (due both to their low amounts and chemical complexity). The study of plant secondary metabolism is thus of interest not only for answering basic research questions—such as the evolution of metabolic pathways, the extent of natural metabolic diversity, and pathway regulation in relation to the environmental conditions—but also from an applied perspective, given that most of the natural products of medicinal importance are actually secondary metabolites.

Thus, although nearly 400,000 flowering plants have been classified so far, only a fraction of these, around 20,000, has been used since ancient times for medicinal purposes [5], and only a minority of these has been studied in detail with regard to the metabolic composition and biological effects of their crude extracts [6]. The Dictionary of Natural Products, for example, which is a curated database of various chemical entities isolated from plants and microbes, contains around 160,000 entries; this number is, however, considered a round-down approximation of the extant diversity of secondary metabolites in higher plants [7].

Today, almost 30% of the new chemical entities released by the FDA (Food and Drug Administration) are either (entirely) natural products, botanical drugs, or semisynthetic derivatives of a natu-

ral product [8]. The pharmaceutical industry has been, however, rather reluctant in investing in large-scale screening of natural products for drug discovery [6]. One of the reasons limiting the screening of small molecules in plants has been the inherent difficulties in the purification of known compounds in adequate yields, but also, as we have mentioned above, the incomplete knowledge of many of the biosynthetic pathways of secondary metabolites [9–11]. The full knowledge of the pathways of plant secondary metabolites is of course essential to develop alternative strategies of production in heterologous hosts (yeast, bacteria) for pharmaceutical applications [12].

The advances in a number of systems-biology disciplines (genomics, transcriptomics, metabolomics, and computation biology), however, fueled by the decreasing costs for generating large-scale molecular data, are revolutionizing our research approaches also in the field of medicinal plants.

In the present review, we will present examples where the application of traditional biochemical and omic-based approaches contributed to new discoveries in the pathways of some secondary metabolites of medicinal importance. We will not cover in detail the knowledge acquired so far on the chemistry of natural products (but we refer the reader to recent excellent reviews on the subject: [13] for benzyloquinoline alkaloids, [14] and [15] for monoterpene indole alkaloids (MIAs), [16] for cannabinoids, [17] for xanthine alkaloids, [18] for ginsenosides, [19] for withanolides, [20] for artemisinin, and [21] for taxol), and we instead focus on the historical developments and the advances made recently in completing the missing parts of the puzzle in the biosynthesis of some important natural products. In the first part of this review, we chose to focus on the cases of benzyloquinoline alkaloids (BIAs), MIAs, cannabinoids, and caffeine, as they represent exemplary cases of how the application of several approaches, based on the integration of genomics and metabolomics, has helped clarify specific biochemical steps or entire pathway branches that had remained elusive. In the second part of this review, we will summarize the knowledge acquired so far on the biosynthesis of specific compounds (ginsenosides, withanolides, artemisinin, and taxol) from other important medicinal plants where we believe integrative approaches could help further the elucidation of their secondary metabolism with a view on the discovery of novel metabolites of medicinal importance. For each presented case study, we survey the health-related benefits and current medicinal use of these compounds and how traditional “reductionist” and integrative approaches are accelerating the development of metabolic engineering strategies (in heterologous and native hosts) for the production of secondary metabolites of pharmaceutical interest.

Approaches for Pathway Discovery

Traditionally, the first steps in the elucidation of plant metabolic pathways were based on the identification of a rather limited number of primary metabolites and on the use of radioactive labels to follow their fate. These were essentially the approaches that brought to the discovery of the reactions of the path of carbon in photosynthesis: the strategy was based on exposing a green algae to a stream of ^{14}C -labeled CO_2 , followed by extrac-

tion, separation, and identification of metabolites with paper chromatography. The gradual decrease of the exposure time to labeled CO₂ allowed, for example, the identification of the product immediately downstream of the CO₂ fixation reaction (phosphoglyceric acid [22, 23]). Similarly, the remaining intermediates of the various reactions were identified, increasing the exposure time to labeled CO₂ [24]. With the advent of recombinant DNA technology, these initial labeling approaches were combined with the isolation of the respective genes and with the synthesis and purification of candidate enzymes. The advent of these new technologies was also accompanied by an increasing interest toward secondary metabolites, which were initially considered only as “waste” products of primary metabolites, with no physiological or ecological role [25]. The use of molecular biology techniques (i.e., molecular cloning and heterologous expressions systems) along with classical protein biochemistry allowed, for example, to assess *in vitro* the catalytic properties, substrate specificities, and identity of the products for a large number of enzymes involved in secondary metabolism (and several examples from these early, targeted approaches for pathway discovery of medicinally important phytochemicals are reported in this review). In recent years, the leap of genomic technologies, with the relative ease in collecting large-scale sequence data, has bred new life into metabolism research [26]. The increasing number of available genome sequence is now frequently integrated with high-resolution/deep-coverage metabolomics approaches [27] not only to uncover structural and regulatory aspects of pathways of secondary metabolism, but also to go deeper into the evolution of metabolism across the diversification of land plants (and landmark examples in this area are the recent reconstructions of the synthesis of nicotine and caffeine [28, 29]). The case studies presented here thus represent successful examples of how targeted molecular approaches or, more recently, the combination of next-generation genomics with metabolic profiling are revolutionizing the field of medicinal plants with new knowledge concerning the synthesis of natural products.

Benzoisoquinoline Alkaloids

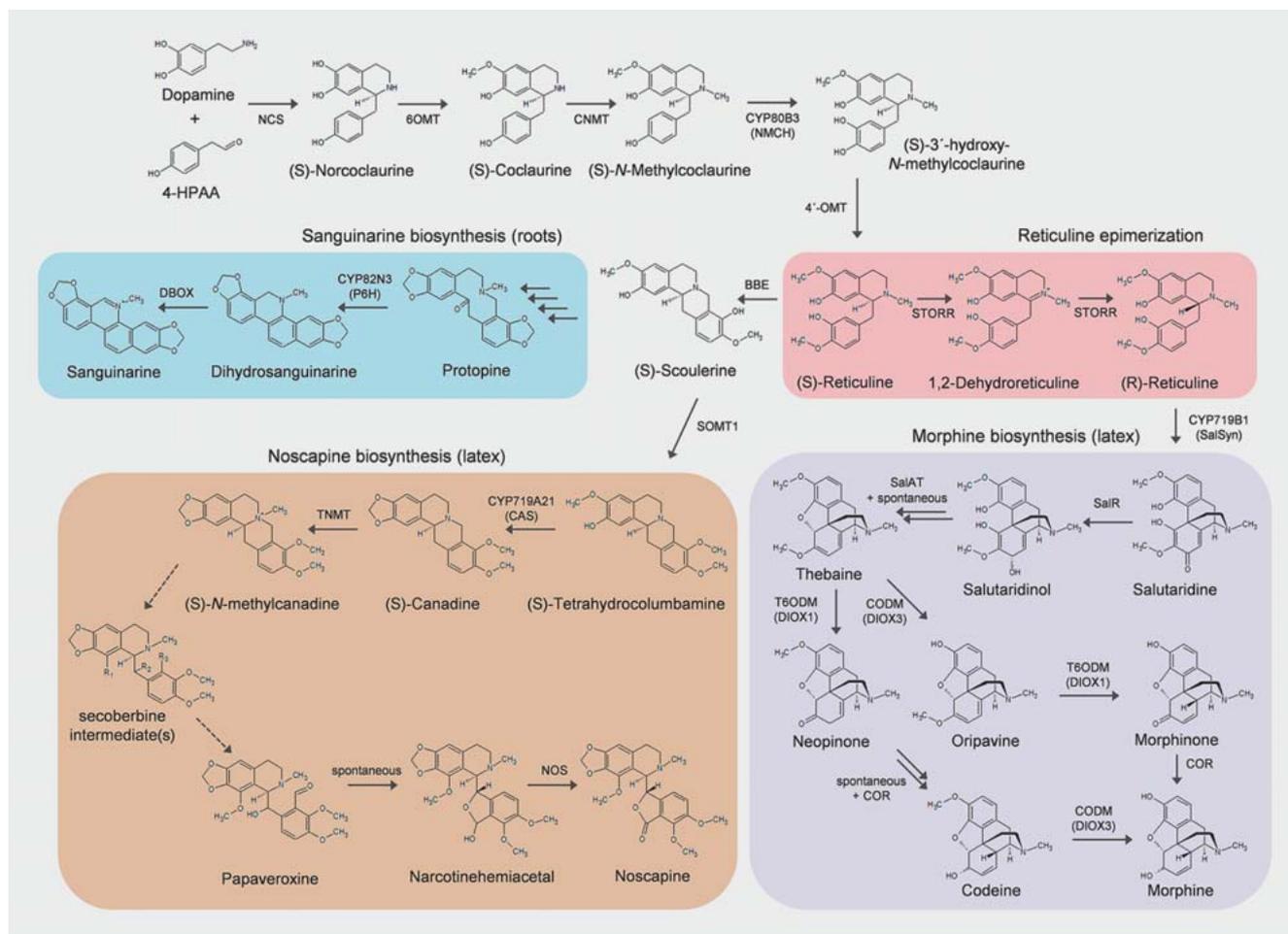
BIAs represent perhaps the oldest medicines humans have used to treat pain. These alkaloids belong to a large family with over 2500 known structures; they are mostly restricted to members of the order Ranunculales (in particular, they are present in the families Papaveraceae and Berberidaceae), Magnoliales, and Laurales. Among the Papaveraceae, opium poppy (*Papaver somniferum* L.) has emerged as the model species to study metabolism of important BIAs, as this plant accumulates large amounts of different subgroups of these alkaloids [30]. The most abundant BIAs in roots of opium poppy are those of the benzophenanthridine-type (e.g., sanguinarine, a potent anti-inflammatory agent that has also showed antitumor properties [31, 32]), while the latex preferentially accumulates varying amounts of morphine and codeine (“morphinans”). Although the increasing use of opioid drugs (natural morphinans and their semisynthetic derivatives like oxycodone) in clinical practice is now raising concerns given their history of abuse, there is no doubt that morphine and codeine represent effective analgesics in the treatment of severe pain, at least in

the short-term following an acute trauma [33]. The initial isolation of morphine from the latex of opium poppy stimulated further research into the elucidation of its biosynthesis in plants. The first studies were based on radiolabel incorporation of a few candidate substrates and established tyrosine and its derivatives as the precursors of morphine [34, 35]. We now know, after decades of research that have seen the application of more detailed tracer studies and biochemical characterization of the related enzymes, that the biosynthesis of BIAs involves a highly branched network of chemical transformations starting from two tyrosine derivatives, dopamine and 4-hydroxyphenylacetaldehyde (4-HPAA) [13]. These two metabolites condense to give rise to (*S*)-norcoclaurine, which is in turn modified by a number of *O*-, *N*-methyltransferases and oxidoreductases to produce (*S*)-reticuline, the precursor of almost all subgroups of BIAs. From (*S*)-reticuline, the pathway diverges into different branches, which may be active only in some species or tissues, resulting in the wide structural diversity of BIAs subgroups that has been observed in plants (► Fig. 1).

The early efforts in the elucidation of BIA biosynthesis were based on the purification of the putative enzymes and on the screening of cDNA libraries for the isolation of the corresponding genes; these initial studies allowed, for example, the characterization of norcoclaurine synthase, the enzyme responsible for the condensation of dopamine and 4-HPAA, producing (*S*)-norcoclaurine [36, 37]. Similar approaches have been followed in the elucidation of the remaining early steps of the BIA pathway: the synthesis of (*S*)-coclaurine, for example, by the action of a 6-*O*-methyltransferase (nococlaurine 6-*O*-methyltransferase, *6OMT*) [38, 39] or, analogously, the synthesis of (*S*)-*N*-methylcoclaurine by a *N*-methyltransferase (coclaurine *N*-methyltransferase, *CNMT*, [40]). The late steps of morphinan biosynthesis remained instead uncharacterized until the development of global gene expression resources for opium poppy. After screening a number of varieties and mutants differing in their accumulation of morphine, two candidate genes were eventually proposed on the basis of the correlation of their expression with the accumulation profiles of morphinans. The discovery of these two genes, thebaine-6-*O*-demethylase (*DIOX1*) and codeine-*O*-demethylase (*DIOX3*), was thus made possible by the development of cDNA microarrays from an opium poppy EST (expressed sequence tag) database [41].

The advent of these “early” global gene expression resources in *P. somniferum* (ESTs collections, microarray) heralded a new era in the study of BIA metabolism. Additional gene expression resources—based on next-generation sequencing—were developed and integrated with metabolomics and proteomics studies in order to identify novel gene candidates [42]. As an example of this approach, known cytochrome genes of the CYP80B3 and CYP82N3 subfamilies, responsible for hydroxylating (*S*)-*N*-methylcoclaurine and protopine, respectively, were used as queries in a co-expression analysis to discover additional BIA biosynthetic genes in several accessions of opium poppy [43].

More recently, integrative approaches based on the combination of gene expression analyses and metabolic profiling were also fundamental in unveiling the nature of a biochemical step in BIA biosynthesis that had remained elusive for a long time. The first step of the morphinan branch is the conversion of (*S*)-reticuline



► **Fig. 1** BIAs biosynthetic pathways of *P. somniferum* (opium poppy) discussed in the text. All BIAs derive from (S)-norcoclaurine, the product of the condensation of two tyrosine derivatives, dopamine and 4-HPAA. After a series of O-, N-methyltransferase and hydroxylation reactions, (S)-norcoclaurine is converted into (S)-reticuline, the central precursor of all BIAs biosynthetic branches. NCS: norcoclaurine synthase; NMCH: (S)-N-methylcoclaurine 3'-hydroxylase; 4'-OMT 3'-hydroxy-N-methylcoclaurine 4'-hydroxylase; STORR: (S)-to-(R) reticuline (aka REPI, reticuline epimerase); P6H: protopine 6-hydroxylase; DBOX: dihydrobenzophenanthridine oxidase; SalSyn: salutaridine synthase; SalR: salutaridine reductase; SalAT: salutaridinol 7-O-acetyltransferase; T6ODM: thebaine 6-O-demethylase; CODM: codeine O-demethylase; COR: codeinone reductase; SOMT1: scoulerine 9-O-methyltransferase; CAS: canadine synthase; TNMT: tetrahydroprotoberberine N-methyltransferase; NOS: noscapine synthase. Dashed arrows indicate multiple steps.

into its *R* stereoisomer; the reaction is a two-step process involving the oxidation of (S)-reticuline to 1,2-dehydroreticuline and the subsequent reduction to (R)-reticuline. Although the reaction was supposed to be catalyzed by two different genes, in agreement with the reaction being a two-step process, screening of transcriptome libraries from opium poppy identified instead a single fused gene composed of two domains. This gene, named *STORR* (from *S*- to *R*-reticuline), encodes a unique bifunctional protein containing a P450 monooxygenase at the N-terminus and an oxidoreductase at the C-terminus [44, 45]. The genetic analysis of opium mutants with impaired synthesis of morphine and high accumulation of reticuline confirmed *STORR* as the causal locus for the epimerization of *S*- to *R*-reticuline. Bifunctional genes like *STORR*, including monooxygenases fused with various additional domains (hydrolase, dioxygenase), have been found also in secondary metabolic pathways of other organisms (e.g., fungi,

[46]); it is thus possible that the occurrence of these genes could represent a sort of metabolic channeling of higher efficiency, in which highly unstable intermediates—like those formed during an epimerization reaction—are converted into final products by the action of a single protein rather than by a multienzymatic assembly [44].

Another example of the application of integrative approaches to the metabolism of BIAs lies in the elucidation of the biosynthesis of noscapine. This alkaloid belongs to the phtalideisoquinoline subgroup of MIAs; it was already widely used for its antitussive properties but has recently been demonstrated to possess antitumor activity given its ability to bind tubulin and arrest cell division in a number of cancer cell lines [47]. It was later showed that noscapine specifically targets the NF- κ B signaling pathway in tumor cells, repressing proteins involved in cell invasion and tumor proliferation [48]. Early radiolabeling experiments in the 1960s

traced back the origin of noscapine to (*S*)-scoulerine [49], which is produced starting from (*S*)-reticuline by the action of a FAD-linked (FAD: flavin adenine dinucleotide) oxidoreductase (BBE, berberine-bridge enzyme). From (*S*)-scoulerine, the synthesis of noscapine requires at least six additional biosynthetic steps, including *O*- and *N*-methylations and several oxidations, but only recently could the complete pathway to noscapine be elucidated in detail. The clarification of the pathway was made possible thanks to the availability of opium poppy varieties accumulating different amounts of noscapine and morphinans. Stems and capsules of these varieties were subjected to RNA sequencing and metabolic profiling to identify genes specifically expressed by the high-noscapine variety (HN1). A number of *O*- and *N*-methyltransferases, along with several cytochrome P450s, were found to be highly expressed only in the high-noscapine variety. Genomic analysis showed that these genes were actually exclusive of HN1. A mQTL (metabolic quantitative trait loci) analysis for noscapine content in an F2 population identified a single locus that was found to be strongly linked to the high-noscapine phenotype in the segregating generation. The locus contained a cluster of 10 genes spanning 220 kbp; the clustered genes corresponded to those previously identified as being exclusively present in HN1 [50]. The reconstruction of the pathway was also supported by virus-induced silencing of the cluster genes, thus allowing to confirm the role of each gene and measuring the accumulation of the various intermediates [50]. The occurrence of the high-noscapine cluster was not a feature unique to BIA metabolism: cluster organization is in fact a recurrent feature in the genomic organization of pathway genes of secondary metabolism [51].

The knowledge acquired so far on the biosynthesis of medically important BIAs has of course allowed the transfer of partial or entire pathways into non-plant hosts. Strategies for chemical synthesis of morphinans (e.g., morphine, codeine, aka “opiates”) have in fact been demonstrated not to be economically feasible; therefore, the licit cultivation of opium poppy is the only source of opiates, from which several semisynthetic derivatives (“opioids”) can be also obtained through semisynthesis (e.g., hydrocodone, [52]). Synthesis of thebaine and hydrocodone, for example, has been obtained in yeast starting from common precursors of primary metabolism. This has required the (over)expression of over 20 genes from yeast itself, plant (*P. somniferum* and *Papaver bracteatum* Lindl.), bacteria, and mammals. Many of the genes transformed into yeast were specifically engineered to increase their activity and stability (e.g., through site-specific mutagenesis to make the enzymes less sensitive to feedback inhibition or to modify their glycosylation patterns); although the fermentation titers for the production of thebaine and hydrocodone remained nevertheless low, especially when compared with the yields obtained with direct purification from opium or semisynthesis, the results obtained so far represent a starting point for further optimization of an alternative strategy for opioids production [53,54]. Similar strategies have also been followed in *Escherichia coli* [55] and in yeast for the synthesis of dihydrosanguinarine, a BIA showing antitumor activity [56].

Monoterpenoid Indole Alkaloids

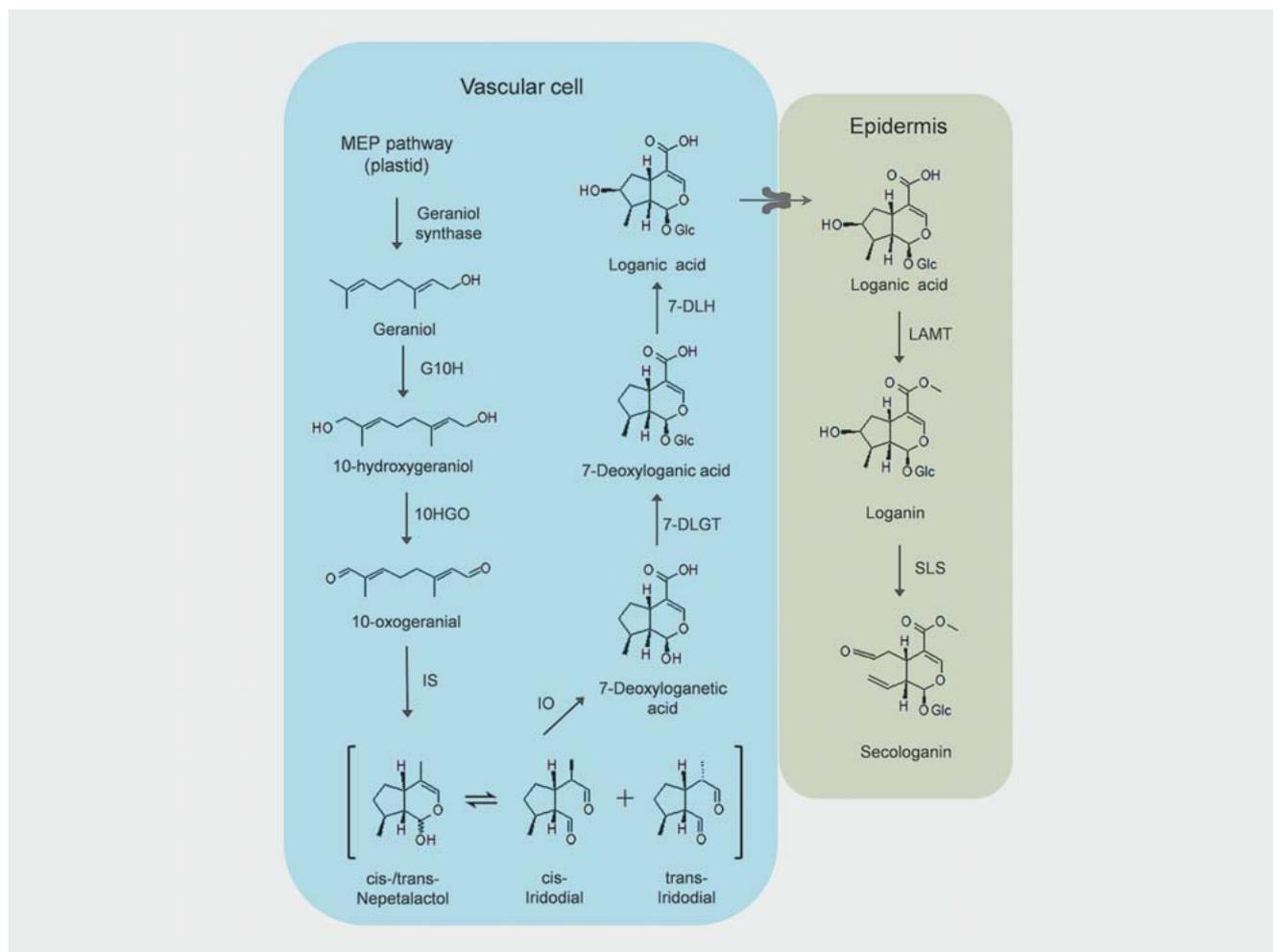
MIAAs represent another class of important alkaloids whose biosynthesis has been studied in detail due to their diverse pharmacological effects. Vinblastine and vincristine, for example, two MIAAs of the bisindole-type isolated from the plant *Catharanthus roseus* (L.) G. Don (Madagascar periwinkle, family Apocynaceae), show toxicity to white blood cells of mammals and are used today as effective medications to treat tumors like lymphoma and myeloma [57]. Other important MIAAs include camptothecin, an inhibitor of DNA topoisomerase I isolated from the tree *Camptotheca acuminata* Decne (Nyssaceae) (irinotecan, a semisynthetic derivative of camptothecin, is one of the most diffused chemotherapeutic in the treatment of colon cancer) and quinine, an antimalarial isolated from the bark of the *Cinchona* trees (*Cinchona* spp.). Quinine is still in use today, although it has been replaced by artemisinin as the recommended first-line treatment for malaria.

MIAAs constitute a large family, with over 3000 structures identified to date. They are mostly confined to plants of the order Gentianales, in the family of Apocynaceae, Loganiaceae, and Rubiaceae. The species *C. roseus*, which synthesizes over 150 different MIAAs, has emerged in this case as the model plant for studying the biosynthesis and regulation of this important class of alkaloids [30].

The biosynthetic pathway of MIAAs is complex. As an example, the complete biosynthesis of vinblastine in *C. roseus* proceeds through at least 30 enzymatic steps, which take place in several different tissues (phloem-associated parenchyma, epidermis, mesophyll, laticifer) and subcellular compartments (plastid, nucleus, ER [endoplasmic reticulum], and vacuole) [58–61]. The chemical complexity of most of the active MIAAs hampered developments in chemical synthesis; this factor, combined with the general low number of MIAAs recovered from plant sources, drove efforts toward the elucidation of biochemical pathways as a necessary step to develop metabolic engineering strategies. We will thus first summarize here the main branches of the MIAA biosynthetic pathway to later focus on the recent discoveries made in the elucidation of the steps that were previously poorly characterized.

As the name suggests, all MIAAs contain a terpenoid and an indole moiety. The terpenoid moiety derives from secologanin, a cyclic monoterpene formed from geraniol. The indole moiety of MIAAs is instead coming from tryptamine, as a result of the decarboxylation of tryptophan. Tryptamine and secologanin then condense to give rise to strictosidine, the precursor of all MIAAs. The whole pathway thus consists of four main parts:

- The first part is the synthesis of geraniol, through the plastid MEP (methylerythritol 4-phosphate) pathway. Although two different routes exist in plants for the synthesis of terpenoid precursors (the cytosolic mevalonate and the plastidic MEP pathway [62]), early labeling studies supported the origin of the terpene moiety of MIAAs from the MEP pathway [63].
- The second part is the conversion of geraniol into secologanin in a series of eight steps that have been elucidated recently (iridoid pathway [64–68]) (► **Fig. 2**).
- The “mid-pathway” then involves the formation of strictosidine starting from secologanin and tryptamine [69], its degly-



► **Fig. 2** MIA biosynthesis (iridoid pathway) in *C. roseus*. The entire pathway is composed by eight steps converting geraniol into secologanin. Geraniol is mainly derived from the plastidal MEP pathway. The early steps in the pathway, up to the synthesis of loganic acid, take place in the phloem-associated parenchyma (vascular cells), while the last two genes in the pathway have been localized to the epidermal cells. The gene responsible for transporting loganic acid across the two cell types has not been identified yet. 10HGO: 10-hydroxygeraniol oxidoreductase; IS: iridoid synthase; IO: iridoid oxidase; 7-DLGT: 7-deoxyloganic acid glucosyltransferase; 7-DLH: 7-deoxyloganic acid hydroxylase; LAMT: loganic acid methyltransferase; SLS: secologanin synthase.

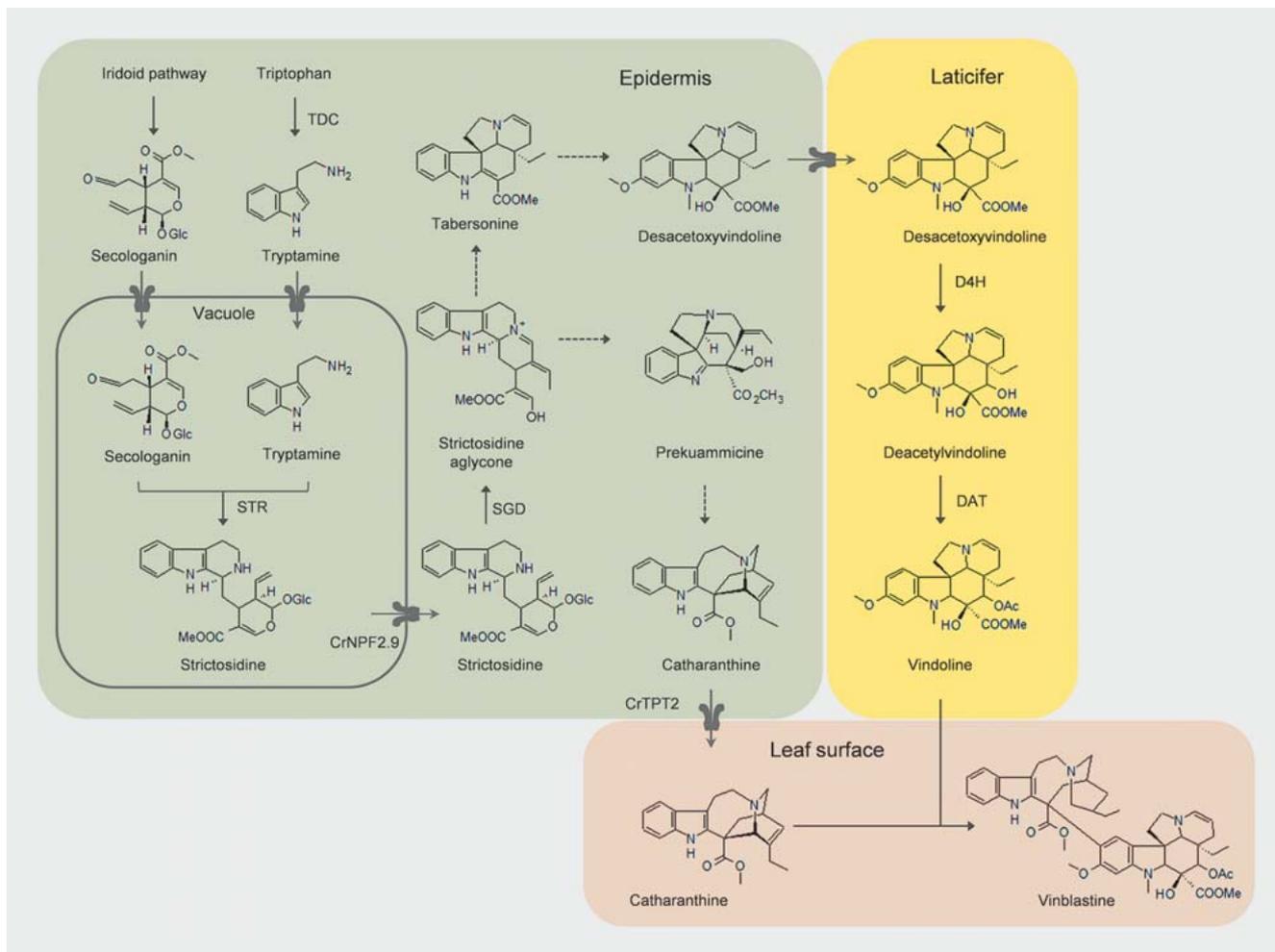
cosylation [70], and a series of downstream transformations whose steps have been clarified, in part, only recently [71, 72] (► **Fig. 3**).

- Finally, the “late-pathway” converts tabersonine, a downstream product of strictosidine, into vindoline, the immediate precursor of vinblastine [73–75] (► **Fig. 3**).

The first approaches in the elucidation of the steps of MIA biosynthesis were mostly based on conventional strategies starting from the purification of the single enzymes, analysis of their AA (amino acid) sequences, and cloning of full-length clones from cDNA libraries using degenerate primers. This was the approach followed, for example, for the identification of geraniol-10-hydroxylase (*G10H*), the enzyme responsible for the hydroxylation of geraniol, the first step of the iridoid pathway [76, 77] and for the purification and cloning of strictosidine beta-glucosidase [70, 78]. More recently, several transcriptome resources and databases

have been developed in *C. roseus*, and these have been used for initial selection of candidate genes of MIA biosynthesis [79–85].

As an example of this approach, transcriptome datasets from several tissues of a *C. roseus* plant [68, 86] have been screened to identify the gene responsible for an elusive step in iridoid biosynthesis, the cyclization reaction of 10-oxogeraniol into iridodial (iridoid synthase). Since the reaction was known to occur in the presence of NADH (nicotinamide adenine dinucleotide [reduced])/NADPH (nicotinamide adenine dinucleotide phosphate [reduced]), the genes using these two cofactors were first selected from the entire transcriptome dataset; then only the transcripts showing a similar expression profile to that of *G10H* (an upstream gene in the same pathway) were retained and considered as candidates for iridoid synthase. The transcript showing the highest correlation to *G10H* was selected for functional validation. The expression of the enzyme in *E. coli* showed that it was able to convert 10-oxogeraniol into *cis-trans* nepetalactol (which is in equilibrium with *cis-trans* iridodial), and VIGS (virus-induced gene silencing) of



► **Fig. 3** “Mid” and “late” pathway steps in the biosynthesis of MIAs in *C. roseus*. The first step is the condensation between secologanin (end product of the Iridoid biosynthesis) and tryptamine to form strictosidine in the vacuole of epidermal cells. Strictosidine is then exported from the vacuole into the cytosol through a transporter of the nitrate/peptide family (CrNPF2.9). The deglycosylated form of strictosidine (strictosidine aglycone) is the central biosynthetic intermediate of many MIAs types. Vindoline, for example, derives from tabersonine and accumulates in laticifers; prekuammicine is instead the precursor of catharanthine, which is then exported to the leaf surface via another transporter, CrTPT2. Leaf damage or herbivory can cause cell disruption, allowing catharanthine and vindoline to react together and form the dimeric MIA vinblastine. TDC: tryptophan decarboxylase; STR: strictosidine synthase; SGD: strictosidine beta-glucosidase; D4H: desacetoxyvindoline 4-hydroxylase; DAT: deacetylvindoline 4-O-acetyltransferase. Dashed arrows indicate multiple steps.

the candidate gene in *C. roseus* confirmed downregulation of the transcript and the lower accumulation of several MIAs downstream of iridoid synthase (e.g., vindoline and catharanthine) [68]. Mining the expression databases from *C. roseus* and analysis of coregulation with additional known genes, proved to be useful also for the discovery of other genes involved in the remaining steps of iridoid biosynthesis [64].

One of the most interesting features of MIA biosynthesis is the spatial distribution of its enzymes. The various parts of the pathway operate in fact in different cell types: (i) the MEP reactions and the early reactions of iridoid biosynthesis occur in the phloem-associated parenchyma; (ii) the remaining steps of the iridoid pathway and the “mid” reactions take place in the epidermis, while (iii) the reactions of the late pathway occur in laticifers [75]. Adding to this complexity, the reactions taking place in the leaf epidermis are also compartmentalized at the subcellular level:

the condensation of tryptamine and secologanin to form strictosidine occur in fact in the vacuole, while the downstream transformations of strictosidine occur in the nucleus and in the cytosol [87]. In particular, the physical separation between the synthesis of strictosidine (vacuole) and its immediate successive step, deglycosylation (nucleus), implies the existence of an export system from the vacuole. Transporter genes have long remained elusive in MIA biosynthesis, with only two systems characterized to date: the export of catharanthine (the immediate precursor of vinblastine) to the leaf surface [88] and the sequestration of vindoline inside the vacuole of mesophyll cells [89]. Also in this case, however, the recent developments of transcriptome resources, combined with functional studies *in planta*, allowed the elucidation of a transporter gene responsible for the export of strictosidine from the vacuole to the cytosol [90]. In order to identify transporter genes, self-organizing maps (SOMs) have been used to cluster all

transcript contigs according to the similarity of their expression profiles across a wide range of tissues and developmental stages. The high-quality nodes of the SOMs that contained known MIA biosynthetic genes were then retained and inspected for the presence of putative transporter genes. This led to the identification of a candidate transporter of the NPF (nitrate/peptide family) family (*CrNPF2.9*). Further analysis confirmed the role of this gene in the export of strictosidine from the vacuole. For example, transient silencing of *CrNPF2.9* in leaf of *C. roseus* led to a necrotic phenotype, probably as a result of the increase in the vacuolar accumulation of strictosidine [90].

As in the case of BIAs, several strategies have been attempted also for production of MIAs in microbial hosts. The commercial production of vincristine and vinblastine, for example, which are powerful therapeutic agents for the treatment of several forms of blood cancer, relies entirely on extraction from plant sources. Most of the active MIAs, however, including vincristine and vinblastine, are produced in extremely low amounts, so their extraction from plant tissues is uneconomical and laborious for commercial production. The first attempt to produce MIAs in microbial hosts focused on the production of strictosidine in yeast. Strictosidine represents in fact the central precursor for a number of MIAs of medical importance (vincristine, vinblastine, quinine, strychnine, ajmalicine). Reconstitution of the pathway in *Saccharomyces cerevisiae* required the integration of a total of 21 genes; of these, 15 represented the entire known plant MIA pathway, while the remaining six were either duplication of yeast endogenous genes or animal-derived sequences. The transformed yeast strain also contained targeted deletions of endogenous genes to decrease the flux into competing routes. As reported already for opiate production in yeast, also in this case the final yields of strictosidine remained nevertheless low (around 0.5 mg/L) for commercial production [91]; the production of this yeast strain represents in any case the basis for further optimization of the flux toward strictosidine or as a starting point for the synthesis of non-natural products [92].

Cannabinoids

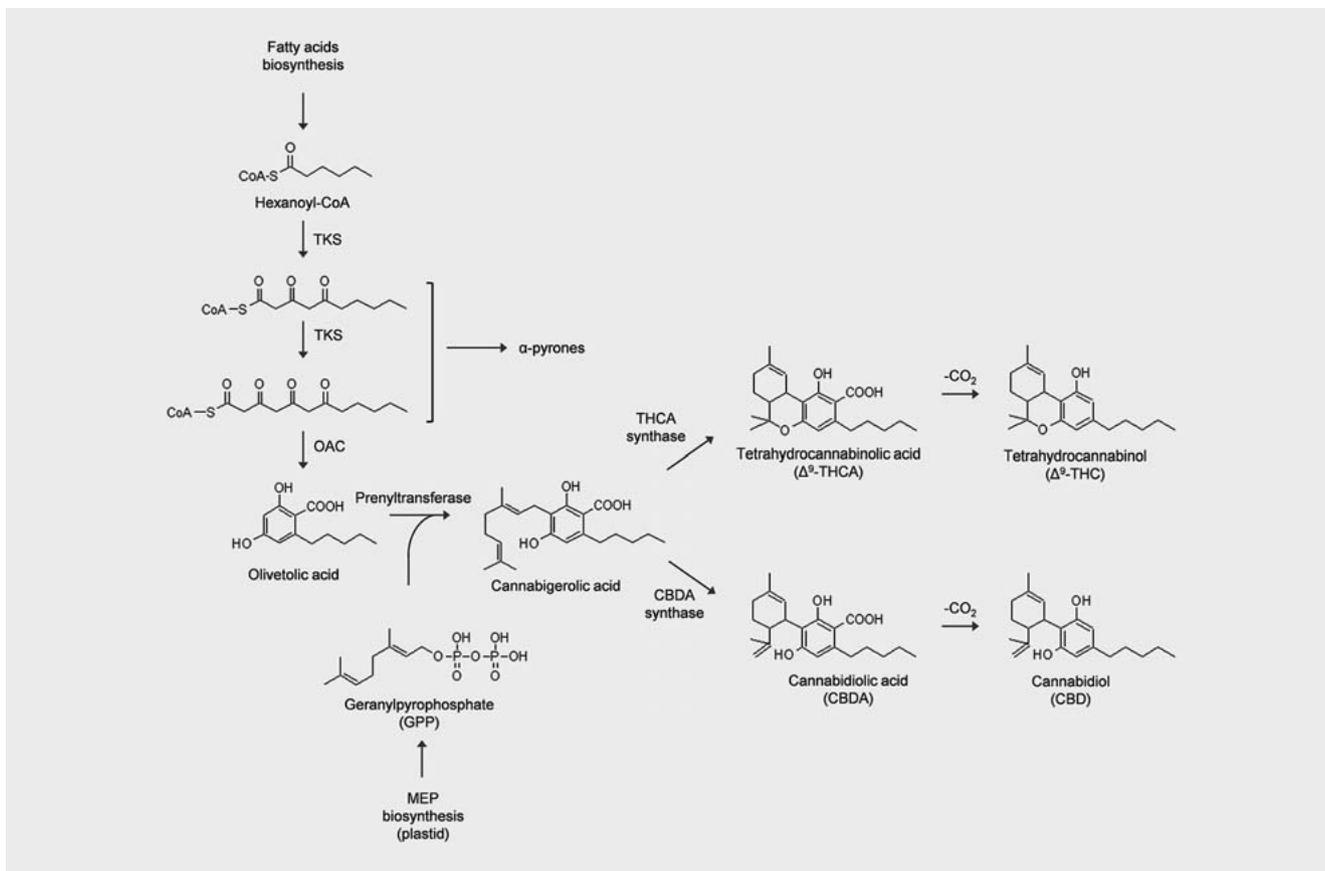
Cannabinoids constitute a group of terpenic alkylresorcinols found in *Cannabis sativa* L., a dioecious plant of the Cannabaceae family. They accumulate in the glandular cavity of specific types of trichomes (capitate sessile or stalked trichomes), which are particularly abundant in female flowers and, to a lesser extent, in other parts of the plant (e.g., leaves, shoots). More than 120 different cannabinoids have been isolated to date [16], although the study of their medical and pharmacological effects focused on the most abundant ones, tetrahydrocannabinol (THC) and cannabidiol (CBD) [93, 94]. Scientific studies on the medical effects of cannabinoids were stimulated by anecdotes reported by people who used to smoke cannabis to relieve pain or to treat a number of conditions (loss of appetite, insomnia). Cannabis in fact represents one of the first plants used for medicinal purposes since ancient times. The first reports of its medical use date back to 2700 BC, when teas and other infuses were already prepared in China to relieve symptoms of rheumatism and arthritis. Also, archeological evidences from a burial cave near Jerusalem, dating back to

390 BC, document the use of smoked cannabis to relieve pain. In addition to its use as a medicine, cannabis has always been used as a source of textile fibers (“bast” fibers) and as a recreational psychoactive drug to achieve a status of mental high. Zoroastrian priests and shamans (~ 500 BC), for example, used cannabis to reach ecstasy during their religious ceremonies [95]. Today, fiber-type cannabis plants continue to be used as a fiber in the textile and bioplastic industries [96], while marijuana-type cannabis represents one of the most highly consumed recreational drugs in the world. Despite the strict regulations around cannabis research, several cannabinoid preparations have been tested in controlled trials for relieving symptoms associated to cancer or HIV [97].

The isolation and structural elucidation of cannabinoids began in the 1940s with the isolation of cannabiol and cannabidiol [98, 99], but it was not until 1964 that the structure of Δ^9 -THC—the main psychoactive component—was reported [100]. In a series of papers from the 1990s, it was found that THC exerts its effects through binding to two different receptors in the human body: CB1, which is present in the brain [101, 102], and CB2, which is instead mainly located in the immune system [103]. The characterization of these receptors led to the discovery of additional substances produced by the human body that also target the cannabinoid receptors [104]. These endogenous ligands were named endocannabinoids to distinguish them from the phytocannabinoids produced in the trichomes of the cannabis plant. We now know that the interaction between endocannabinoids and CB1/CB2 constitutes the “endocannabinoid system”, a central regulator of homeostasis in the human body. Typical responses mediated by this system include pain perception, memory, appetite, immunity, and, of course, the neurological responses induced by the psychoactive Δ^9 -THC [105].

Although more than 120 phytocannabinoids have been reported in the literature, their biosynthesis has been fully described only for the most abundant components, THCA (tetrahydrocannabinolic acid) and CBDA (cannabidiolic acid) (► Fig. 4). THCA is the most abundant cannabinoid in marijuana-type plants, while CBDA, which does not possess psychoactive properties, is instead the most abundant in hemp (fiber-type plants). We will present here some examples to show the advances made in the elucidation of the steps in the core cannabinoid pathway. While the first steps to be defined, historically, were based on classical enzyme purification approaches and homology-based cloning of the corresponding genes, more recently the development of genomics and transcriptomics resources in cannabis have helped to clarify additional biosynthetic steps [106–108]. Also, at least initially, the elucidation of the cannabinoid pathway was made difficult by the low incorporation of the label [109] and by the fact that cannabinoids occur *in vivo* as carboxylic acids but are then decarboxylated to neutral (active) forms during heating or smoking.

All phytocannabinoids are formed by an alkylresorcinol (phenolic) moiety coupled to a monoterpene (► Fig. 4). Labeling studies using ^{13}C -glucose showed that the monoterpene moiety derived from the plastidial MEP pathway, while the alkylresorcinol was produced through the polyketide pathway [110]. The first step in the synthesis of THCA and CBDA is the condensation of olivetolic acid (OA, an alkylresorcinol) with geranylpyrophosphate (GPP), leading to cannabigerolic acid (CBGA), the immediate pre-



► **Fig. 4** Biosynthetic pathways of the major phytocannabinoids, Δ^9 -THC and CBD. The alkyresorcinol (phenolic lipid) moiety of cannabinoids derive from the polyketide pathway, in which hexanoyl-CoA is first condensed with three molecules of malonyl-CoA by the action of TKS and then cyclizes to form OA in a reaction catalyzed by OA cyclase (OAC). The addition of GPP, from the plastidial MEP pathway, then generates CBGA, the immediate precursor of Δ^9 -THCA and CBDA. Δ^9 -THCA (and its decarboxylated form, delta9-THC) represent the psychoactive compounds of marijuana-type plants. The most abundant cannabinoid in hemp (fiber-type cannabis) is instead the non-psychoactive CBDA.

cursor of THCA and CBDA. The reaction is catalyzed by an aromatic prenyltransferase (geranyl pyrophosphate: olivetolate geranyltransferase, *GOT*), which was isolated in 1998 [111]. The gene (*CsPT*) was later cloned and shown to be expressed in leaves, flowers, and trichomes [112, 113].

CBGA is then the substrate of two different FAD oxidases: the tetrahydrocannabinolic acid synthases (*THCA synthase*) and the cannabidiolic acid synthase (*CBDA synthase*), which produce, respectively, THCA and CBDA. The two genes, which share 84% similarity, are encoded by different loci [114]. Both THCA and CBDA synthase were purified through enzymatic assays from crude extracts and their respective genes cloned using degenerate PCR primers (THAS: [115]; CBDA: [116, 117]).

The steps leading to the synthesis of the alkyresorcinol precursor of cannabinoids, OA, have, however, remained elusive, and it was not until recently that these biosynthetic steps have been clarified. OA was long supposed to be synthesized starting from hexanoyl-CoA through successive condensations with three molecules of malonyl-CoA, in a series of steps catalyzed by a type III polyketide synthase (*PKS*, [118, 119]). A type III *PKS* cloned from cannabis leaves (named tetraketide synthase, *TKS*), however, did not produce OA and was instead shown to accumulate, among other by-

products, α -pyrones [120]. These metabolites were typical downstream products of polyketide pathways in bacteria lacking polyketide cyclase activity [121]. On the basis of this, candidates with structural similarity to polyketide cyclases were selected from an EST library of cannabis trichomes, leading to the identification of a member of the dimeric $\alpha\beta$ barrel protein superfamily (DABB superfamily). This gene, which was distantly related to type II polyketide cyclases of bacteria (*Streptomyces*), was able to convert, in the presence of TKS, hexanoyl-CoA and malonyl-CoA into OA, acting effectively as a noncanonical polyketide cyclase [107]. A similar approach, based on mining the same EST database from cannabis trichomes, was also used to identify the acyl-activating enzyme responsible for the synthesis of hexanoyl-CoA, the first step of the polyketide pathway in cannabinoid biosynthesis [108].

The elucidation of the steps in the biosynthesis of the main phytocannabinoids opened the possibility to transfer the pathway to heterologous hosts for commercial production of THCA/THC and CBDA/CBD. These two cannabinoids have in fact several pharmacological effects. THC, the neutral psychoactive form of THCA, targets mainly the CB1 receptor in the central nervous system and has analgesic and antispastic activities. Its consumption is, however, associated to well-known side effects (memory loss, de-

creased coordination, and, in some individuals, anxiety, [122]). CBD, on the other hand, may reduce the side effects of THC and has shown pharmacological potential to reduce inflammation and symptoms of epilepsy [123]. Sativex, which is the only cannabinoid-based drug approved so far in 27 countries, is a mouth spray of THC and CBD. This drug is used today to treat the spasticity associated to multiple sclerosis [94]. Given the potential shown by THC and CBD, various strategies have been attempted in metabolic engineering of cannabinoids. Cell cultures of cannabis, even in the presence of elicitors, have resulted in limited yields, probably due to the lack of compartmentalization required by the high toxicity of cannabinoids [124, 125]. A more promising approach might be represented by the production of THCA synthase in *Pichia pastoris* and its use in a cell-free two-liquid phase reactor to drive the synthesis of THCA. Also, this system, however, achieved relatively low yields ($0.121 \text{ g} \cdot \text{L}^{-1} \cdot \text{h}^{-1}$ of THCA), probably as a consequence of the sensitivity of THCA synthase to be inhibited by its substrate [126, 127].

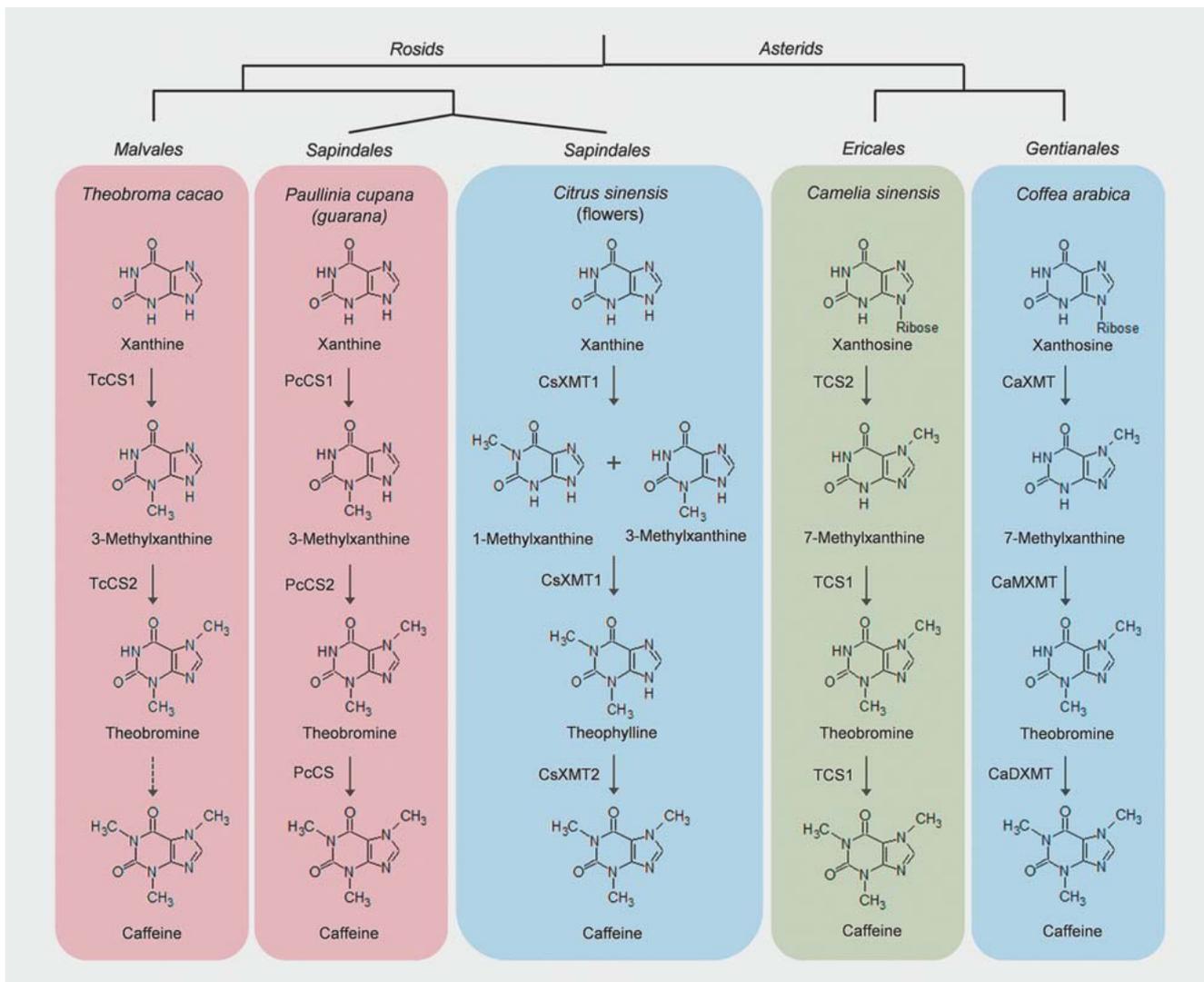
Today, the regulations around the use of cannabis, and the research around it, are becoming less strict. Several European countries and the United States have exemptions for the medical use of marijuana; other U.S. states have legalized cannabis consumption, in moderate amounts, for personal use. Canada and Israel have funding bodies and programs specific for cannabis research. As the regulations in cannabis research will ease, we anticipate the development of additional genomic and metabolomics resources in cannabis. The integration of these resources will aid the elucidation of the full biosynthetic pathways of cannabinoids, opening the way to the discovery of novel compounds of potential medicinal importance.

Caffeine

Caffeine (1,3,7-trimethylxanthine) is a xanthine (purine) alkaloid found in guarana, yerba maté, cacao, and several species used to make tea. Traditionally, it is called guaranine when it comes from the guarana plant (*Paullinia cupana* Kunth, family Sapindaceae), theine when it comes from the tea plant (*Camellia sinensis* (L.) Kuntze, family Theaceae), and mateine in mate infusions; however, they all are the same compound. In addition, cacao, which accumulates only trace amounts of caffeine, contains the similar compound theobromine, which has similar, albeit less potent, bioactivities to caffeine. Of the species listed above, genome sequences for coffee [128], tea [129], and cacao [130] have been published indicating that at least three metabolic pathways for caffeine biosynthesis evolved independently co-opting genes from different gene families. The appearance of at least three pathways for caffeine biosynthesis in higher plants is thus an example of recurrent convergent evolution: the presence of caffeine *per se* in species from multiple plant orders (Malvales, Sapindales, Ericales, and Gentianales) did not always imply the recruitment of homologous genes [29] (► Fig. 5). Intriguingly, this study, which relied on sequence information from five flowering species, revealed that caffeine biosynthesis was characterized by an even greater degree of convergent evolution than was previously thought, with citrus, chocolate, and guarana plants containing two previously unknown pathways of caffeine synthesis using ei-

ther caffeine synthase or xanthine methyltransferase-like enzymes. Moreover, ancestral sequence reconstruction revealed that these pathways would have arisen rapidly since the ancestral enzymes were co-opted from their previous biochemical roles to those of caffeine biosynthesis. As such, this seminal paper provides a fantastic blueprint for studies into the evolution of natural product biosynthesis.

Caffeine's exact function *in planta* is unclear, and two main roles, which are by no means mutually exclusive, have been proposed. In the first of these, sometimes called the chemical defense theory, caffeine is believed to protect young leaves and fruit from predators [131, 132]. In keeping with this, Uefuji et al. [133] demonstrated that leaves of transgenic tobacco (*Nicotiana tabacum* L. [Solanaceae]) plants, engineered to produce caffeine, were less susceptible to insect feeding than leaves that did not contain caffeine. In the second, sometimes known as the allelopathic theory, caffeine is believed to be released by the seed coat to prevent germination of other seeds [134]. Evaluation of the cacao genome, the first of the three caffeine-containing species to be sequenced, suggested that cacao harbors a rich repertoire of homologs of secondary metabolism-associated genes, including pathways for oils, storage lipids, flavonoids, and terpenes as well as the alkaloid class to which caffeine belongs. The analysis of multiple metabolomics studies of this species suggests that functional prediction of the gene repertoire mentioned above was indeed largely correct [135]. The evolution of caffeine and indeed its metabolic precursor theobromine was, however, looked at in more detail following publication of the coffee and tea genomes [128, 129]. Intriguingly, coffee was characterized to contain several species-specific gene family expansions including that of the xanthine *N*-methyltransferases (XMTs) involved in caffeine production and revealed that these genes expanded through sequential tandem duplications independently of genes from cacao and tea. As for cacao, a large number of metabolomics studies have been performed on coffee and tea identifying high contents of caffeine, quinate, and chlorogenic acid in the former [136–138] and catechins, terpenes, and caffeine in the latter [139–141]. Since there is also an increasing amount of transcriptomics data available for these species [142–148], it would appear likely that evaluating the dynamic behavior of transcripts related to caffeine biosynthesis in comparison to other unknown genes (and to the levels of the metabolites themselves) will greatly enhance our understanding as to how these pathways are controlled. One study of particular interest is the long read sequencing of the coffee bean transcriptome since this provided more and longer transcript variants specifically allowing the identification of a further 10 transcripts likely to encode key enzyme isoforms of caffeine biosynthesis [142]. This information thus greatly extends the number of candidate genes that are potentially important determinants of the final caffeine level within plant cells, and their study will thus prove instrumental in allowing rational design of metabolic engineering strategies aimed at modifying caffeine content. In addition, two other studies, this time in tea, have been highly informative in analyzing the regulation of caffeine biosynthesis. The first of these built gene regulatory networks for secondary metabolism of a wide range of tea tissues implicating a large number of transcription factors in the regulation of caffeine



► **Fig. 5** Biosynthetic pathways of caffeine biosynthesis. The synthesis of caffeine evolved independently in several orders of eudicots. Two different gene families have been recruited to synthesize caffeine: (i) caffeine synthases (CS), which sequentially methylate xanthine (in cacao and guarana) or xanthosine (in *C. sinensis*) to eventually produce caffeine; (ii) XMTs, which are instead active in the flowers of *C. sinensis* and in coffee (*C. arabica*). Different substrate specificities of CS and XMT enzymes gave rise to at least three main pathways in caffeine-accumulating plants. The first pathway represents the CS lineage and is the route present in cacao and guarana (red); the second pathway is the synthesis of caffeine operated by the XMT genes (*C. sinensis* and *C. arabica*, blue); *C. sinensis* has instead recruited the genes in the CS lineage but synthesizes caffeine through the same sequence of intermediates detected in *C. arabica* (green). Guarana and *Citrus sinensis*, although both members of the Sapindales, have converged on caffeine synthesis co-opting different genes. CS: caffeine synthase.

biosynthesis [149]. The second article used a comparative transcriptomic and metabolomics analysis of tea and oil tea that does not produce caffeine, indicating higher expression of the key phenylpropanoid enzymes flavanone-3-hydroxylase, dihydroflavonol reductase, and anthocyanidin reductase in tea but lower levels of phenylalanine ammonia-lyase and chalcone isomerase; however, the exact link between this and the levels of caffeine is not apparent from this study [150]. Thus, these studies offer hints to the regulation; however, due to the genetic recalcitrance of the species, it will likely be several years before these can be confirmed at the molecular level.

Caffeine is a compound whose medicinal properties are at least in part offset by its addictive properties [151, 152], and as such, it

remains very much debated as to how healthy it actually is. That said, a lot of the idea that coffee is dangerous springs from work in the 1970s and 1980s in which its consumption was linked to higher incidence of cancer and heart disease [153, 154]; however, much of this early research should be disregarded since it did not take into account other health-detrimental habits in the cohorts such as cigarette smoking. More recent analyses evaluating health and diet data of a cohort of 400,000 adults over a period of 13 years revealed no evidence that coffee consumption increased death from either these diseases or indeed any others with anything but a minor drop in mortality rate among regular coffee drinkers [155]. Coffee has additionally been linked to lower rates of type 2 diabetes [155], reduced risk for some cancers [156], and

protection against Parkinson's disease [157], as well as inhibiting propagation of hepatitis C virus [158]; however, as we detail below, at least some of these proposed functions remain very much under debate. By contrast, caffeine has been suggested to inhibit lipid anabolism and thereby have a contributory role in metabolic syndrome [159]. In addition, coffee consumption has been linked to diversity of gut bacteria and is often added to painkillers in the belief that it aids in analgesic efficiency [160]. Largely on the basis of its properties as a stimulant, overconsumption of caffeine has a number of (short-term) health-negative effects including paranoia, restlessness, anxiety, high blood pressure, very fast and abnormal heart rate, vomiting, and confusion [161].

However, given the richness in terms of metabolic diversity of all species accumulating caffeine and the specific medicinal implications of any one of their constituents, it is clearly very hard to disentangle, as is the case of all food-based bioactives, the health-positive effects of one from another.

That said, interestingly, several studies have shown that decaffeinated coffee has the same health properties, suggesting—although by no means proving due to the small amounts of residual caffeine in such beverages—that caffeine itself is not the bioactive ingredient in such instances. This fact aside, the current consensus appears to be that there are relatively few health-negative effects of caffeine (with the exception of those following extreme consumption). Although the purported health-positive effects remain somewhat contentious, it is likely that in the coming years they will be exposed to severe scrutiny, and only then we will be in a position to categorically state the case that caffeine is effective against any one ailment or the other.

Ginsenosides

Ginsenosides constitute a group of triterpenoid saponins that are exclusively produced in plants of the *Panax* genus (family Araliaceae). The name “Panax” comes from Greek, meaning “all-healing,” and refers to the medicinal properties of these plants. Of the nine existing *Panax* species, three in particular have been studied in relation to their pharmacological activities: *Panax ginseng* C. A. Mey. (Chinese ginseng), *Panax quinquefolium* L. (American ginseng), and *Panax notoginseng* (Burkill) F.H.Chen [162]. These species have been—and still are—widely used in Chinese traditional medicine to treat a number of ailments, including fatigue, anemia, rheumatisms, and cardiac disorders. The use of ginseng as a herbal remedy dates back to about 100 AD, when it was believed that the dry root powder of this plant possessed miraculous healing effects [163].

Ginsenosides accumulate during the normal development of the ginseng plant. The total amount of ginsenosides has been shown to be higher in leaves of one-year-old seedlings and mature roots [18]. The accumulation and composition of ginsenosides is regulated during growth, but the exact mechanism of how this occurs still remains not clear [164]. At least 150 naturally occurring ginsenosides have been described so far [165], and a number of multiple benefits on human health has been reported, such as strong anti-oxidative, antitumoral, and anti-inflammatory activities.

Ginsenosides have been classified according to their chemical skeleton in two different types: dammarane- and oleanane-type ginsenosides. Based on the glycosides attached, the dammarane ginsenosides are further divided into three different subgroups: PPD-type (protopanaxadiol), PPT-type (protopanaxatriol), and ocotillol-type (► Fig. 6).

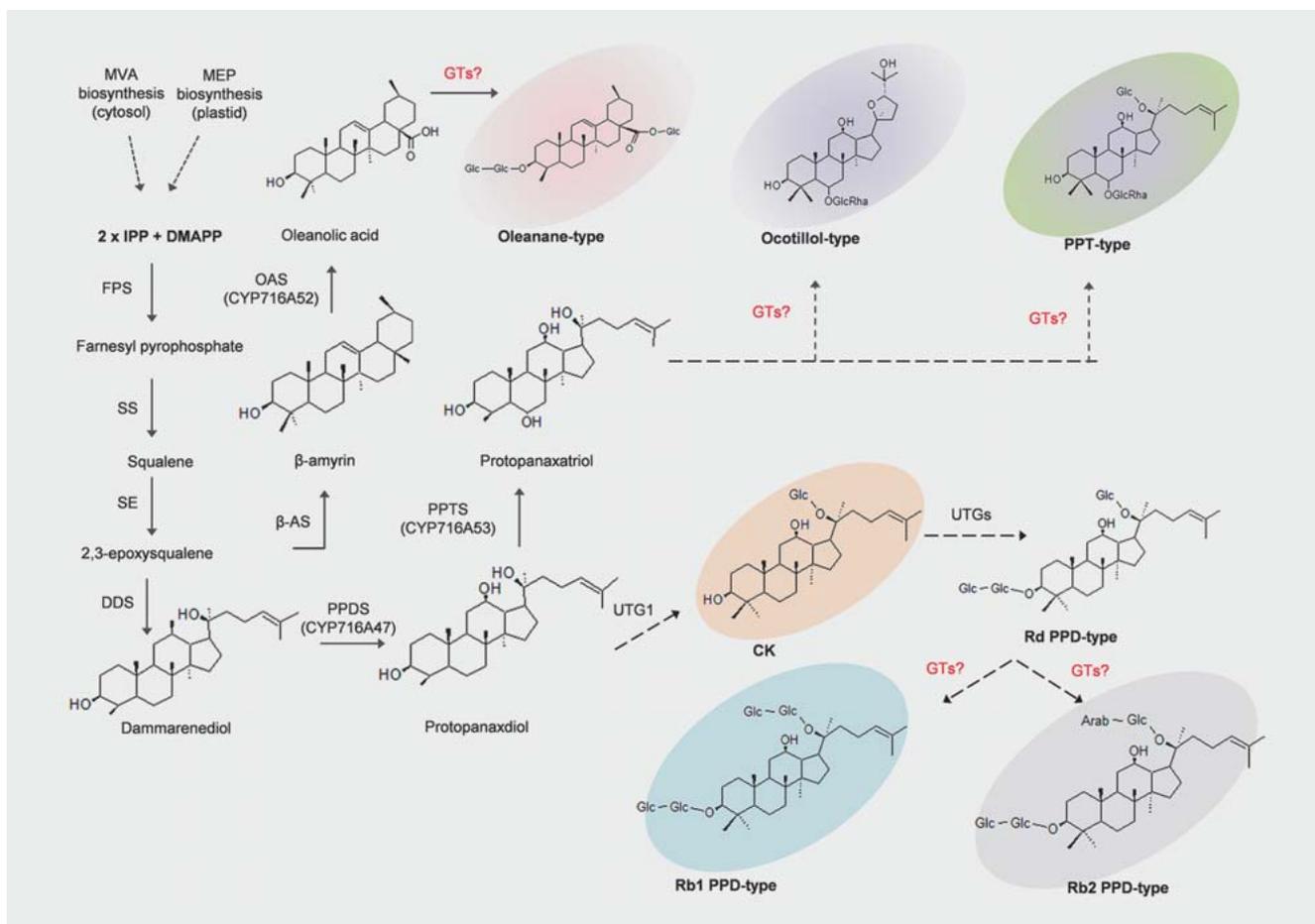
Recent studies showed that the molecular structure of the ginsenosides is important in defining their medical properties. The anticancer activities of these saponins depend on the number of sugar molecules and on their attachment position [162]. Protopanaxadiol and protopanaxatriol ginsenosides with no sugar residues or PPT and PPD ginsenosides containing up to three sugar residues inhibited different types of cancer, while others containing a higher amount of sugar residue showed none or very weakly antiproliferative effects [166–168]. Furthermore, it has been shown that the biological response of different types of ginsenosides is also related to the number and positions of the hydroxyl groups, which reflects the polarity of these molecules and thus facilitates the interaction with the cell membrane [169–172]. Also, differences in stereochemistry were demonstrated to produce different pharmacological effects [173].

The biosynthetic pathway of ginsenosides is not entirely characterized and many steps still need to be elucidated. The studies so far show that the main precursor used for the triterpene ginsenosides is squalene, which is formed from the condensation of two farnesyl pyrophosphate (FPP) molecule. The synthesis of each FPP requires the condensation of one dimethylallyl pyrophosphate (DMAPP) with two molecules of isopentenyl pyrophosphate (IPP). IPP can be produced in the cytosol through the mevalonic acid (MVA) pathway or in the chloroplast from the methylerythritol (MEP) pathway [62]. The role of the plastidial IPP is still unclear since ginsenoside biosynthesis mainly relies on the pool of cytosolic IPP [174], although a certain degree of compensation was observed in case of inhibition of either MEV or MEP [175].

The crucial steps in the generation of ginsenoside diversity are the cyclization of 2,3-oxidosqualene by oxidosqualene cyclases (OSCs) and the subsequent hydroxylations and glycosylations [176, 177] (► Fig. 6). Dammareniol synthase (DDS) is a member of the family of OSCs, which is specifically found only in *Panax* species [18]. Its encoding gene has been characterized as the very first step in ginsenoside biosynthesis [178, 179].

The product of this enzymatic conversion is dammareniol, which is the precursor of three of the four types of ginsenosides: PPD-, PPT-, and ocotillol-type. In the next subsequent reactions, the dammareniol is hydroxylated in two consecutive reactions to protopanaxadiol and protopanaxatriol by protopanaxadiol and protopanaxatriol synthases (PPDS and PPTS, members of the cytochrome P450 family). Both protopanaxadiol and protopanaxatriol are further glycosylated by uridine diphosphate (UDP)-dependent glycosyltransferases (UGTs), whose genes remain to be identified. Extensive additional glycosyl decorations give rise to the diversity of all detected ginsenosides [180]. Recent studies provided a better understanding of a part of PPT-type biosynthetic pathway by characterization of four *P. ginseng* UGTs catalyzing protopanaxatriol glycosylations [181].

The biosynthesis of the oleanane-type ginsenosides starts always from 2,3-oxidosqualene, which is then cyclized to β -amyrin



► **Fig. 6** Ginsenoside biosynthesis. The crucial step in the generation of ginsenoside diversity is the cyclization of 2,3-epoxysqualene. One of the cyclization reactions leads to the production of β -amyrin, which is precursor of the oleanane-type ginsenosides. An alternative cyclization of 2,3-epoxysqualene, catalyzed by DDS, leads to the formation of dammarenediol, which is then the precursor of ocotillol-, PPT-, and PPD-type ginsenosides. Compound K is a dammarenediol-type ginsenoside isolated from human blood after oral administration of *P. ginseng* and has not been detected so far in *Panax* plants. Many of the enzymatic steps in the ginsenoside biosynthesis have not been well characterized, but two gene families play key roles in generating ginsenoside diversity: the CYPs and the UGTs. SE: squalene epoxidase; β -AS: β -amyrin synthase; OAS: oleanane acid synthase; GT glycosyltransferase; UGT UDP-glycosyltransferase. Reactions with genes marked in red indicate hypothetical steps. Dashed arrows indicate multiple steps.

by β -amyrin synthase and converted to oleanolic acid by the action of oleanane acid synthase, member of the cytochrome P450s family. The remaining reactions, leading to glycosylated oleanane-type ginsenosides, are catalyzed by additional UGT genes that have not been identified so far (► **Fig. 6**).

In the last years, a novel dammarenediol-type ginsenoside (compound K) has been isolated from human blood after oral administration of ginseng [182]. Interestingly, compound K has been never detected in *Panax* plants. The authors suggested that this novel ginsenoside could actually represent a minor component whose biosynthesis may actually occur in *Panax* plants, since the transcripts encoding two of the fundamental enzymes (CYP716A47 and UGT1) responsible for its conversion are present in *P. ginseng* tissues. Compound K could possess a number of beneficial effects for human health, given its anticancer, anti-diabetes, and anti-inflammatory properties tested *in vitro* [183, 184]. Currently, compound K is synthesized from deglycosylation of PPD-type ginsenosides [185].

Given the medicinal importance of ginsenosides, a number of bioengineering strategies have been developed in order to increase their production and to compensate the time required for field cultivation, which generally involves four to six years. Four different main strategies have been undertaken to synthesize ginsenosides in native and heterologous hosts: (i) developing cell and tissue culture methods [186]; (ii) adventitious root cultures [187]; (iii) transgenic plants [188]; and (iv) engineered yeast systems [189].

The first tissue culture of ginseng was reported in 1964 [190], and many other successful studies followed afterward [191, 192]. The effects of different growth regulators on the final product formation have been evaluated, including sucrose (used as the most common carbon source in ginseng cultures), phosphate, copper, and nitrate. These investigations showed that the rate of biomass growth and the respective ginsenoside content correlated directly with the medium sugar concentration (up to 60 g L⁻¹). Higher sugar concentrations inhibited cell growth and had a negative im-

pact on ginsenoside production [193]. Phosphate, copper, and nitrates in different concentrations improved the ginsenoside yield and thus stimulated ginsenoside production in cell cultures [194, 195].

An example of the tissue culture approach is using adventitious roots as high biomass producers and studying the effect of different treatments or chemical elicitors [189, 196, 197]. As the major physiological role of the ginsenosides is related to plant defense [198, 199], stress-inducible factors have been used in order to improve their production. Treatments with methyl jasmonate and salicylic acid generally induced oxidative stress and increased ginsenoside content, as well as gamma-irradiation, which enhanced the final product up to 16-fold [200, 201].

In addition to the cell and tissue culture methods, genetic engineering methods have been used successfully to up- and down-regulate key genes involved in ginsenoside biosynthesis, such as 3-hydroxy-3-methylglutaryl coenzyme A, squalene synthase (SS), cytochrome p450 (CYPs), and DDS. Transgenic plants over-expressing these genes showed an increased amount of ginsenosides [188, 202–204].

Successful achievements of producing PPD, PPT, oleanolic acid, and compound K have been also made by using engineered yeast strains [185, 203, 205].

All these works provide an insight into the complex mechanisms of ginsenoside biosynthesis and explore new methods for large-scale production of these important pharmacological compounds. Nevertheless, many efforts still need to be done in order to further elucidate the biochemical pathways leading to ginsenoside formation, as well as to clarify the events responsible for their diversification in *Panax* species. Further studies are needed to improve the current available platforms and resources, as well as to advance the knowledge about their clinical applications.

Withanolides

Withanolides are a group of naturally occurring C-28 oxygenated steroidal lactone triterpenoids that have been found in at least 15 genera of Solanaceae (e.g., *Withania*, *Tubocapsicum*, *Lycium*, *Datura*, to mention few). Their presence has been reported also in Fabaceae (legumes) and Lamiaceae (the family to which most aromatic plants belong) [206]. Within Solanaceae, the shrub *Withania somnifera* (L.) Dunal (“Indian ginseng” or “Ashwagandha”) has been the focus of several pharmacological studies, given its wide use in Ayurveda (the major system of Indian traditional medicine) as a general tonic to increase vigor and memory and lessen the symptoms associated to rheumatism, fatigue, and dehydration [207]. On the basis of the anecdotal reports from the Ayurvedic practices, *W. somnifera* extracts were subjected to intense pharmacological scrutiny and showed to possess promising antitumor and anti-inflammation properties [208–210].

Despite the growing relevance of withanolides in medical research (which we will cover in detail further below), information about their biosynthetic routes and pathway regulation *in planta* remain scarce. Over the past years, more than 200 different withanolides have been isolated from roots, berries, and leaves of *W. somnifera* [19]; the focus of most of the pharmacological research was placed, however, almost exclusively on Withaferin A

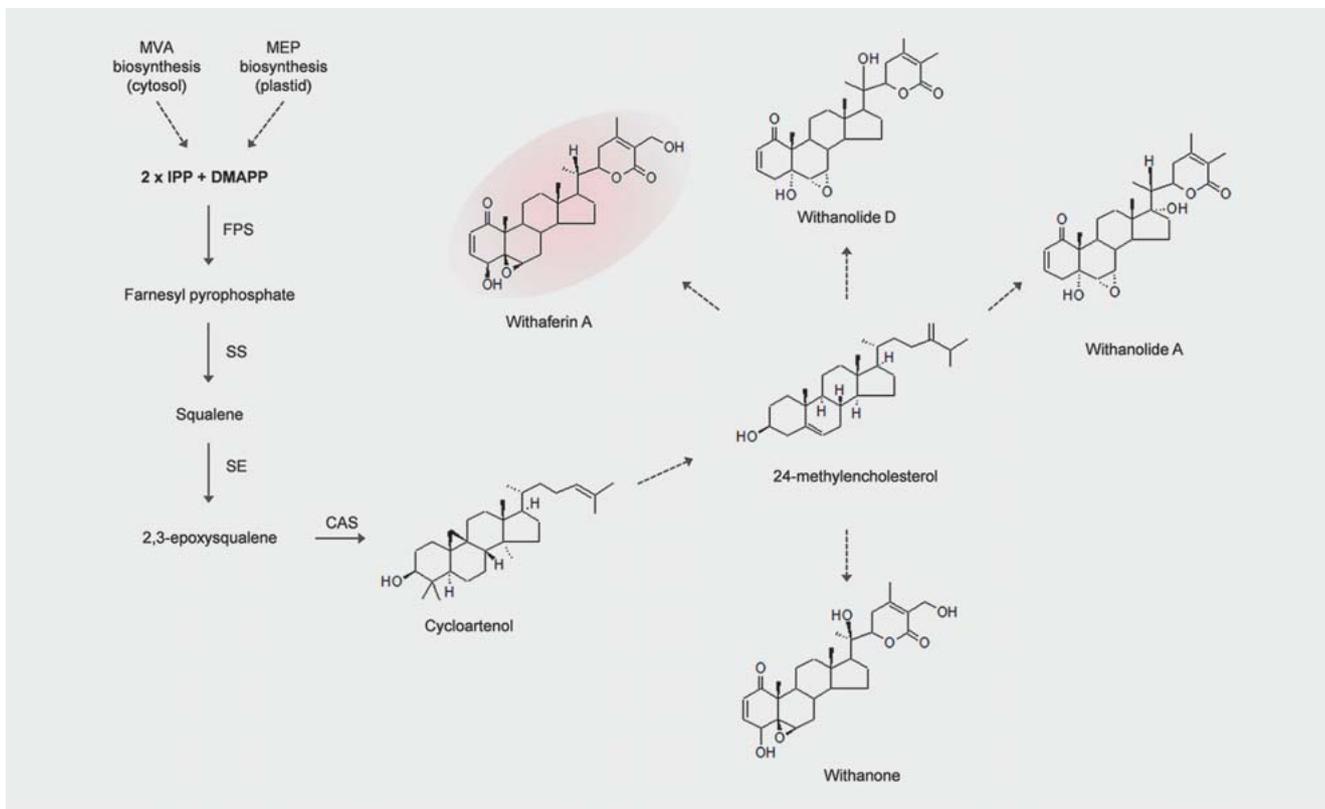
(► Fig. 7), the first withanolide to be isolated from *W. somnifera* [211]. In general, we now know that the C28-steroidal lactones are biosynthesized from the C5-terpenoid precursors IPP and DMAPP. As in the case of ginsenosides, the key step in the synthesis of withanolides is the cyclization of 2,3-oxidosqualene. In the biosynthesis of withanolides, the product of this reaction is cycloartenol, which is then converted to 24-methylenecholesterol, the precursor of all withanolides (► Fig. 7). Methylenecholesterol is then subjected to a series of hydroxylations, elongations, glycosylations of the carbocyclic skeleton, and further cyclization of its side chain, resulting in compounds with complex structural features [212–216]. According to the difference in the substituted groups of C-17 side chain, withanolides can be divided into two types; type A with a δ -lactone or δ -lactol and type B with γ -lactone or γ -lactol side chain [217]. Some recent investigations have identified putative regulatory and structural genes involved in withanolide biosynthesis [218–221].

As we have already mentioned, in the past few decades, withanolides attracted considerable research attention, and several studies were carried out to investigate the pharmacological and biological activities of this class of metabolites and their role in human medicine. Withanolide extracts from *W. somnifera* showed to possess anti-inflammatory, cytotoxic and antitumor activities [222]; there are also indications that the administration of *Withania* extracts improved memory retention in rats [223] and cognitive functions in humans [224, 225]. Withanolide A, withanolide B, withaferin A, and withanone, in particular, showed protective effect on the neuronal tissues of frontal cortex and corpus striatum in rats and prevented increase of lipid peroxidation [226, 227]. These early investigations on the effects of *Withania* extracts in attenuating cerebral functional deficits led to more targeted studies on the potential beneficial effects of withanolides in neurodegenerative diseases. Recent studies showed, for example, that a root extract of *W. somnifera* was effective in decreasing the accumulation of β -amyloid peptides in the brains of rats affected by Alzheimer's disease [228]. Also, a crude *Withania* extract relieved significantly the symptoms of drug-induced parkinsonism (tremor, rigidity) in model rats [229].

Withanolides have also shown promising antitumor activities. Withanolide A and Withaferin A are two of the best studied withanolides for their capacity to significantly reduce the survival of various cancer cell lines and decrease the size of breast tumors implanted in rats [230–232]. The effect of Withaferin A, in particular, seems related to its capacity to interfere in the pathways of protein degradation and recycling (which are highly active in cancer cells), through inhibition of tubulin polymerization: this inhibition would prevent the formation of autophagy-related structures, which are essential for protein recycling [233].

Also, other withanolides (e.g., withanolide D, 17 α -hydroxywithanolide D, physagulines) were extracted from stems, roots, and leaves of *Tubocapsicum anomalum* (Franch. & Sav.) Makino (Solanaceae) and *Physalis angulata* L. (Solanaceae), and all exhibited high and significant cytotoxicity against several human cancer cell lines [234–236].

Despite the increasing evidence concerning the beneficial effects of these compounds, there are still many areas that remain to be investigated, especially regarding the biosynthesis and reg-



► **Fig. 7** Overview of withanolide biosynthesis. The precursor of all withanolides is 24-methylenecholesterol, which undergoes a series of hydroxylations and further modifications of the side chain in a series of steps not yet completely elucidated. Methylenecholesterol is a downstream product of cycloartenol, which is in turn derived from the cyclization of 2,3-epoxysqualene. Withaferin A (red) was the first withanolide to be isolated from *W. somnifera* and is today the best characterized in terms of pharmacological effects. Abbreviations: SE: squalene epoxidase; CAS: cycloartenol synthase. Dashed arrows indicate multiple steps.

ulation of the withanolide pathway. *W. somnifera* is an important and highly valued plant in traditional medicine and showed promising effects in small-scale clinical trials [237, 238]. In the future, the full elucidation of withanolide biosynthesis will help to transfer the pathway to heterologous hosts for cost-effective biosynthesis of the active components; on the other hand, the development of biotechnology protocols for *Withania* spp. will guide future efforts for functional studies in this important genus and will provide the genetic materials for targeted breeding and commercial exploitation.

Artemisinin

Artemisinin is a sesquiterpene lactone isolated from the Chinese herb *Artemisia annua* L. (Asteraceae), known as qinghaosu (sweet worm-wood) in traditional medicine, and mainly used for its anti-malarial effect. In addition to that, recent studies showed promising anticancer, antiviral, and anti-inflammatory activities [239].

The first report on the healing properties of *A. annua* extracts dates back to 340 AD by Ge Hong in his book *Zhou Hou Bei Ji Fang* (*A Handbook of Prescriptions for Emergencies*). It was only in 1971, however, that the active compound was isolated and characterized, due to the work of the Chinese chemist Youyou Tu [240,

241], who was later awarded the Nobel prize for medicine in 2015 for her discovery of artemisinin.

Artemisinin became essential in the treatment of uncomplicated malaria caused by the parasite *Plasmodium falciparum* and has established itself as the most potent of all antimalarial drugs [242]. Although the mechanism of action is still not completely understood, the use of artemisinin and its derivatives in combined therapies contributed significantly to the reduction in malaria mortality [243]. Artemisinin is currently the first-line treatment against malaria [244, 245], despite the emergence in recent years of cases of resistance in Southeast Asia. Recent studies showed that the resistance is mainly due to the K13 mutation in *P. falciparum* parasites [246, 247].

Given the complex structure of natural artemisinin, the main commercial source for this compound so far is the natural plant. Artemisinin is produced by the glandular trichomes of *A. annua*, but its accumulation *in planta* is low (0.01–1.4% dry weight) and highly dependent on the plant variety [248]. Based on this, the extraction of artemisinin is relatively expensive and its production cannot meet the global demand.

In order to face these fundamental problems, many efforts to increase artemisinin production have been attempted. Significant results in this direction were obtained in the field of molecular biology, synthetic biology, and genetic and metabolic engineer-

ing. All these achievements would have not been possible without the characterization of the genes and enzymes related to artemisinin biosynthesis. In the early studies, radioactive-isotope labeling has been used to show that artemisinin derives from IPP and DMAPP, which are synthesized both from the cytosolic mevalonate (MVA) and from the plastidial 2-c-methyl-d-erythritol 4-phosphate (MEP) pathway [249–252]. The condensation of two molecules of IPP with one molecule of DMAPP forms FPP, which is then converted to amorpha-4,11-diene by amorphadiene synthase (ADS) [253]. Amorphadiene is subsequently oxidized, first, to artemisinic alcohol and then to artemisinic aldehyde by a CYP71AV1 and its redox partner cytochrome P450 reductase (CPR) [254, 255]. Artemisinic aldehyde is then converted to dihydroartemisinic aldehyde by the enzyme DBR2 (artemisinic aldehyde Δ 11(13) reductase) and oxidized to dihydroartemisinic acid (DHAA) by aldehyde dehydrogenase (ALDH1) [256, 257]. The export of DHAA to the trichome and its photooxidation then yields artemisinin (► Fig. 8).

The elucidation of the artemisinin biosynthetic pathway has been a fundamental step in exploring and developing the bioengineering tools used to enhance its production. Different directions have been undertaken in order to improve the artemisinin biosynthesis in the same *A. annua* species or in different host organisms.

Germplasm selection and breeding have been used for creating superior cultivars [258]. The studies reported so far describe a number of cultivars with increased artemisinin content from 1 to 2.4% (DW), but due to instable artemisinin production, these lines have not been considered as a valuable commercial source [259, 260].

Transgenic *A. annua* plants have also been produced with the aim of increasing the amount of artemisinin. In general, two main strategies have been used: the first one based on the overexpression of structural or regulatory genes [261–263], and the second one based on the inhibition of competing pathways, such as, for example, the squalene pathway [264].

Overexpression of several genes responsible for key steps of artemisinin biosynthesis, such as farnesyl pyrophosphate synthase (FPS), ADS, CYP71AV1, CPR, and DBR2 led to approximately a double increase of artemisinin production [265–267].

Based on these conclusions, many research groups focused their interest in co-overexpressing two or more genes in *A. annua* to further increase the amount of artemisinin [262, 263]. For example, co-overexpression of FPS, CYP71AV1, and CPR genes increased the artemisinin content by 3.6 fold (2.9 mg/g fw) in comparison with control plants [267], and the simultaneous overexpression of ADS, CYP71AV1, and CPR resulted in 2.4-fold increase of artemisinin (15.1 mg/g DW) compared to control plants [268].

Recently, several transcription factors of different families, including WRKY, bHLH, NAC, and MYC have been isolated and characterized in *A. annua*. The overexpression of these genes also increased the final amount of total artemisinin [261, 266, 269–271].

The other approach used to enhance the artemisinin amount is to block the key enzymatic steps in competitive pathways to divert the flow predominantly into artemisinin biosynthesis [262]. Inhibition of the expression of the SS gene, which uses farnesyl

pyrophosphate as a substrate and catalyzes the first step of the sterol pathway, increasing the artemisinin content up to 31.4 mg/g (a three-fold increase with respect to control plants) [264].

In order to explore the metabolic engineering approaches for alternative artemisinin production, several heterologous hosts have been tested. The steps leading to the synthesis of amorphadiene have been engineered in *E. coli* by introducing the MVA pathway from yeast (*S. cerevisiae*) and a synthetic ADS gene [272]. The results obtained reached a titer of 300 mg/L amorphadiene [273].

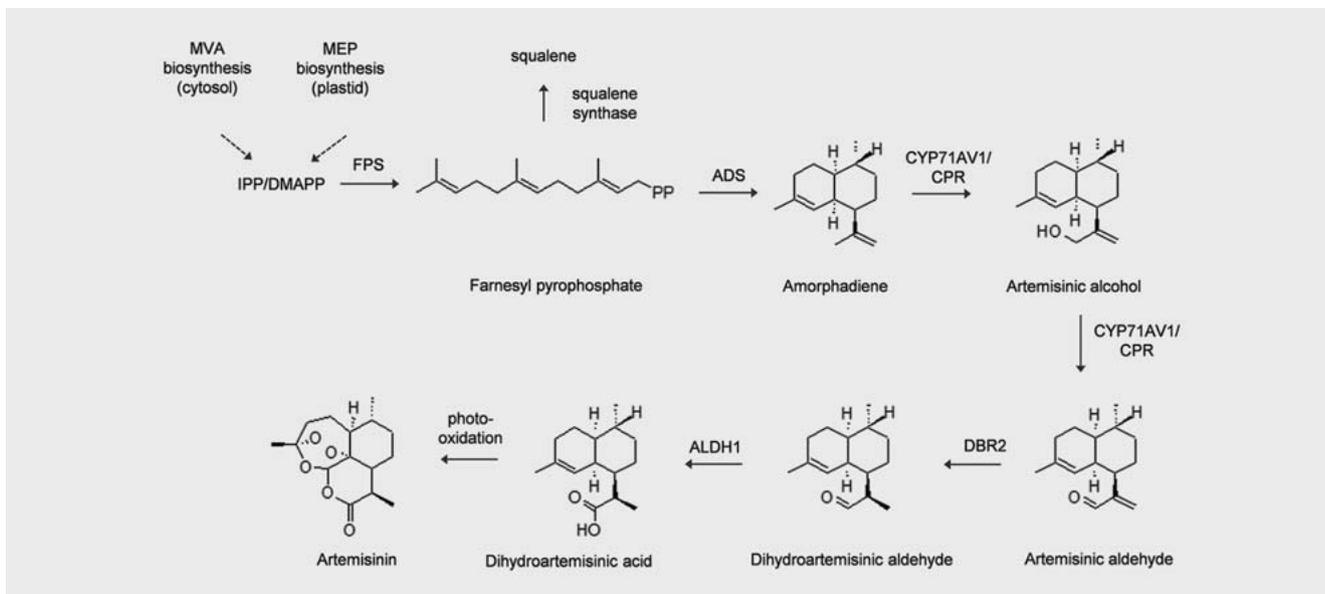
Another attempt to enhance artemisinin production has been made in plant hosts. *Nicotiana* species have been selected as potentially the most suitable ones because of their favorable characteristics (rapid growth and high biomass) [263]. An innovative approach consisted in the insertion of biosynthetic genes in both the nucleus and chloroplast genomes, leading to a final yield of 120 μ g/g artemisinic acid [274]. Despite these efforts, however, the production levels in *Nicotiana* remained low and therefore not suitable for commercial production.

To date, the most prominent achievement in the field of metabolic engineering is the production of artemisinic acid in yeast. In this case, the MVA pathway has been introduced into *S. cerevisiae* along with ADS and CYP71AV1, allowing the conversion of amorphadiene to artemisinic acid in three oxidation steps. As a result, around 100 mg/L of artemisinic acid have been obtained [254]. The system was further improved by the introduction of two additional enzymes, a plant dehydrogenase (ADH1) and a second cytochrome (CYB5), which were both positive regulators of artemisinin biosynthesis. The process reached titers up to 25 mg/L of artemisinic acid, which is the maximum amount achieved so far [275]; this improved yeast system has, however, found modest market impact due to the lower costs associated to the direct extraction of artemisinin from plants [276].

Taxol

Taxol (paclitaxel) is a complex diterpenoid extracted from the bark of the pacific yew (*Taxus brevifolia* Nutt., family Taxaceae), a tree native to the west coastal region of North America. In 1960, taxol was discovered during a large phytochemical screening aimed at the identification of cytotoxic natural products from plants. This effort was jointly conducted by the National Cancer Institute and the U. S. Department of Agriculture [277, 278]. Taxol belongs to a large family of taxoids (taxane diterpenoids) that accumulate in *Taxus* species, where they play an important role in plant defense. Taxoids deter the feeding activities of mammals and insects and protect the plants from fungi colonization [279].

Taxol is formed by a tetracyclic oxahaptadecane skeleton decorated with eight functional oxygen groups, two acyl groups, and a benzyl group [280]. After the elucidation of its structure in 1971 [277], several clinical trials led to its approval by the FDA as an anticancer drug for the treatment of a wide range of cancers (ovarian, breast, lung, Kaposi's sarcoma, cervical, and pancreatic) [281]. Since then, taxol has become a leading anticancer drug, whose total sales exceed several billion U.S. dollars per year [282]. The mechanism of action of taxol is based on its capacity



► **Fig. 8** Metabolic pathway of artemisinin biosynthesis. The first step of artemisinin synthesis is the condensation of IPP/DMAPP into farnesyl-pyrophosphate (FPP). FPP is then cyclized to amorphaadiene by ADS and further oxidized to artemisinic alcohol and artemisinic aldehyde by CYP71AV1 and its redox partner CPR. Artemisinic aldehyde is converted to dihydroartemisinic aldehyde by DBR2, and then to DHAA by ALDH1. Artemisinin is produced by spontaneous photo-oxidation of DHAA.

to interfere with the function of microtubules during cell division, causing their polymerization even at low temperatures. This property renders taxol highly cytotoxic to cancer cells [281].

The amount of taxol that can be extracted from the bark of the adult trees of *T. brevifolia* is however, extremely low. Around 12 kg of bark material yield only 0.5 g of purified taxol [278]; therefore, alternatives sources or methods for taxol production must be developed to avoid the need to rely on destructive bark harvesting [283].

In addition to that, the knowledge of the pathway of taxol biosynthesis remains incomplete. Of the 20 hypothesized enzymatic steps, only 14 have been well characterized [280,284,285] (► **Fig. 9**). The current understanding of the taxol biosynthetic pathway includes at least eight oxidation steps, five acetyl/aryl transferase steps, a C4 β ,C20-epoxidation reaction, a phenylalanine aminomutase step, *N*-benzoylation, and two CoA esterifications [282]. The presence of several putative enzymes in the pathway was recently suggested by analyzing the transcripts of *Taxus baccata* L. cells elicited with methyl jasmonate [285].

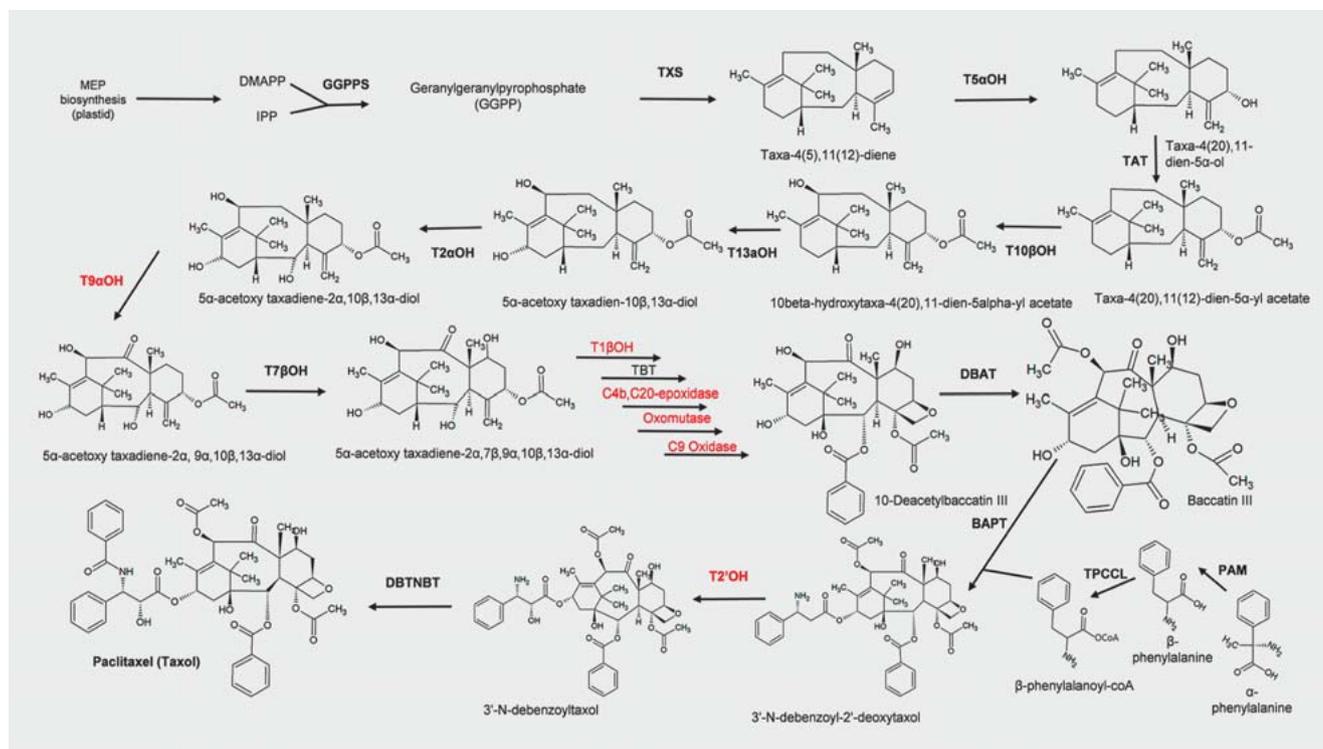
The precursors of taxol are IPP and DMAPP from the plastidial MEP pathway. Geranylgeranyl pyrophosphate synthase catalyzes the condensation of three molecules of IPP and one of DMAPP into geranylgeranyl pyrophosphate (GGPP), which is then cyclized by taxadiene synthase into taxa-4(5),11(12)-diene (taxadiene). Taxadiene is then the central precursor from which all taxane diterpenoids originate. In the branch leading to taxol biosynthesis, taxadiene is hydroxylated by different P450 hydroxylases. The order of the reactions and some of the genes responsible for these subsequent catalytic steps are, however, not clear yet: from the isolation of the putative intermediates, several hydroxylations should occur at positions C1, C2, C4, C7, and C9, as well as a further oxidation at C9 and a C4 β ,C20 epoxidation. The product of

this series of poorly characterized steps is baccatin III, a key intermediate that can be also extracted from the needles of *T. brevifolia* and constitutes the starting substrate for semisynthesis of taxol and other taxane diterpenoids [280]. Baccatin III is then esterified on C13 with a β -phenylalanyl moiety yielding 3'-*N*-debenzoyl-2'-deoxy-taxol, in a reaction catalyzed by baccatin III: 3-amino,13-phenylpropanoyltransferase (► **Fig. 9**).

From 3'-*N*-debenzoyl-2'-deoxy-taxol, the last two steps of the biosynthesis leading to taxol require the hydroxylation and terminal *N*-benzoylation of the β -phenylalanine side chain by a yet uncharacterized taxane-2' α -hydroxylase and a *N*-benzoyl transferase (DBTNBT) [21,285].

Today, the supply of taxol for medical use cannot be achieved from natural sources. As a consequence of the initial overharvesting of the bark for taxol extraction, *T. brevifolia* is now in a near threatened state [286]. On the other hand, total chemical synthesis of taxol, which was achieved in 1994 [287], has never been considered as an economically feasible alternative, due to the high complexity of the process. The current standard for taxol production is now semisynthesis, starting from the isolation of the intermediates baccatin III or 10-deacetyl baccatin III from *Taxus* cell cultures. Taxol can also be produced entirely from *Taxus* cell suspension cultures. The whole process, after decades of optimization based on the use of chemical elicitors (e.g., methyl jasmonate) and improvement of growth conditions, has now reached yields in the range of several hundred mg per liter of culture [282,288].

A partial alternative to *Taxus* cell culture was represented by the transfer of the known part of the pathway—up to taxadiene—to *E. coli*. Bacteria (and yeast) offer in fact a higher growth rate with respect to plant cell cultures and are generally easier to manipulate. The insertion of two pathway modules into *E. coli* (the



► **Fig. 9** Overview of taxol biosynthesis. The pathway leading to taxol is composed by at least 20 enzymatic steps; of these, only 14 have been characterized (enzymes in red indicate hypothetical steps). TXS: taxadiene synthase; T5 α OH: taxane 5 α -hydroxylase; TAT: taxadiene-5 α -ol-*O*-acetyl transferase; T10 β OH: taxane 10 β -hydroxylase; T13 α OH: taxane 13 α -hydroxylase; T2 α OH: taxane 2 α -hydroxylase; T9 α OH: taxane 9 α -hydroxylase; T7 β OH: taxane 7 β -hydroxylase; T1 β OH: taxane 1 β -hydroxylase; TBT: taxane-2 α -*O*-benzoyltransferase; DBAT: 10-deacetyl baccatin III-10-*O*-acetyltransferase; T2'OH: taxane 2' α -hydroxylase; PAM: phenylalanineaminomutase; TBPCCL: β -phenylalanine coenzyme A ligase. Figure modified from [280].

MEP pathway and the GGPP synthase/taxadiene synthase pathway) resulted in final yield of around 1 g/L of taxadiene. Although taxadiene is a distant precursor of baccatin III (and thus several steps—some of which still unknown—separate taxadiene from taxol), the metabolic engineering of *E. coli* was an important achievement for the future full transfer of this important pathway to a microbial host [289].

Bioinformatic Resources for Medicinal Plants

In recent years, the decreasing costs associated with sequencing and assembly of genomic data led to the release of a high number of whole-plant genome sequences, including several from medicinal plants [290]. In some cases, as we detail below, this was accompanied by the development of several communal bioinformatics resources that integrated various types of omics datasets. Clearly, given the complexity of secondary metabolism of medicinal plants with respect to crops and model plants species, these resources offer the opportunity to mine specifically the metabolic pathways of medicinal plants and correlate, for example, the number of specific metabolites with the genomic data (e.g., gene expression, sequence polymorphisms). We provide below a sur-

vey of the main genomic databases that have been recently developed for some of the most studied medicinal plants.

Medicinal Plant Genomics Resource [291] is an example of a large, collaborative effort between several research institutions containing genome and metabolome data of 14 taxonomically diverse medicinal species, including *Atropa belladonna* L. (family Solanaceae), *C. sativa*, *C. roseus*, *Panax quinquefolius* L. The website offers an easy-to-use interface for a BLAST (basic local alignment search tool) search against the sequenced species and provides access to the various genome browsers of medicinal plants. The files related to the genome and transcript assemblies are also available for download. *C. acuminata* (the “happy tree” of Chinese traditional medicine, [292]), *Calotropis gigantea* (L.) W. T. Aiton (a shrub of the Apocynaceae family growing in Southeast Asia, which is known for producing cardiac glycosides [293]), and a new variety of *C. roseus* are the latest medicinal plants whose genomic and transcriptomic data have been added to the database. The database also contains metabolic profiling data (mainly acquired through LC-MS), collected from several tissues of medicinal plants.

Another example of a resource offering a range of tools for visualization and analysis of metabolic networks and ‘omics’ data is CathaCyc, a metabolic pathway database built from metabolic and RNA-seq data of the plant *C. roseus* [82]. CathaCyc is a repository for genes, enzymes, reactions, and pathways of primary and

secondary metabolism; it contains 390 pathways with more than 1300 enzymes. The database also integrates the draft genome data of *C. roseus* [74]. The enzymes in CathaCyc have also been linked to ORCAE [294], a genome annotation resource, allowing the users to validate and edit gene annotations [295].

In 2011, a consortium of U.S. research organizations, funded by NIH, launched the project Transcriptome Characterization, Sequencing, and Assembly of Medicinal Plants Relevant to Human Health [296]. Currently, the database contains transcriptome data related to 31 species of medicinal importance, including, among others, *Cinchona pubescens* Vahl (the quinine tree, family Rubiaceae), *Colchicum autumnale* L. (family Colchicaceae, the source of colchicine), *Datura stramonium* L. (family Solanaceae), and *Podophyllum peltatum* L. (family Berberidaceae) (mayapple; the roots of *Podophyllum* accumulate podophylotoxin, the precursor of the chemotherapeutic etoposide [297]).

Recently, another database has been established within the Phytometasyn project (www.phytometasyn.ca). It contains *de novo* transcript assemblies of around 20 medicinal plants including the plant *Eschscholzia californica* Cham. (California poppy, a member of Papaveraceae accumulating several active BIAs, mainly those of the pavine-type, e.g., eschscholtzidine [83]).

Future Prospects

For centuries, plants have always been used as remedies to treat a great number of symptoms. Even today, a large part of the world population relies on herbal medicines as a major source of health care, especially in Asia, Africa, and Latin America. In some rural areas, traditional medicines based on herbal drugs are the only source of health care. Almost 30% of the modern drugs we use today are actually derived from natural products; an ever-increasing number of these, coming from plants, are now in the process of being approved for market either as main active ingredients or as supplements. Several clinical trials of herbal medicines are now underway in the United States for the treatment of food allergies, asthma, and gastric inflammation [298].

We are now at the beginning of a new phase in which integrative approaches of genomics and metabolomics are applied to the study of the metabolism of medicinal plants. These approaches have begun to revolutionize our understanding of at least two main aspects of herbal medicines: (i) the biosynthesis, and pathway regulation, of many plant secondary metabolites of medicinal importance [290]; (ii) the mechanism of action of many of these plant herbal components on human metabolism and health [299, 300]. We see in this avalanche of knowledge both challenges and avenues for further research. We think there is a urgent need to develop faster, more informative and comprehensive analytical approaches for profiling and characterizing a larger number of metabolites; these challenges can be overcome also with the development of computational metabolomics strategies for metabolite annotation [301,302], *de novo* pathway reconstruction [303], and analysis of natural variation [304]. We clearly recognize the long history and the potential of traditional medicines as a source of well-being, but we also reason that a more intense scrutiny should be conducted on herbal drugs—including rigorous studies on their chemical composition and clinical trials—before

claims could be made in relation to their therapeutic efficacy. This new knowledge could then be used—as we have seen in the case studies presented here (especially in the case of artemisinin)—to set up platforms for metabolic engineering and enable sustainable production of medicinal phytochemicals. Finding alternative ways for production of these compounds—outside of their respective native plant hosts—is also relevant to preserve natural resources in their native habitats, as the case of taxol has shown during the initial overharvesting of *T. brevifolia*. Scientists and policy makers need to find a better balance to promote a sustainable use of genetic resources, especially from the hot spots of world biodiversity (e.g., the Amazonian forest). A new equilibrium need to be established between ecological conservation and bioprospecting for novel drug discoveries from plants [11].

Acknowledgements

The authors thank the German Federal Ministry of Education and Research, project Plant-INNO, and the European Union's Horizon 2020 research and innovation program, project PlantaSYST (SGA-CSA No. 739582 under FPA No. 664620).

Conflict of Interest

The authors declare no conflicts of interest.

References

- [1] Hardy K, Buckley S, Collins MJ, Estalrich A, Brothwell D, Copeland L, Garcia-Taberner A, Garcia-Vargas S, de la Rasilla M, Lalueza-Fox C, Huguet R, Bastir M, Santamaria D, Madella M, Wilson J, Cortes AF, Rosas A. Neanderthal medics? Evidence for food, cooking, and medicinal plants entrapped in dental calculus. *Naturwissenschaften* 2012; 99: 617–626
- [2] Krikler DM. The foxglove, “The old woman from Shropshire” and William Withering. *J Am Coll Cardiol* 1985; 5: A3–A9
- [3] Somberg J, Greenfield B, Tepper D. Digitalis – historical development in clinical medicine. *J Clin Pharmacol* 1985; 25: 484–489
- [4] Brook K, Bennett J, Desai SP. The chemical history of morphine: an 8000-year journey, from resin to *de-novo* synthesis. *J Anesth Hist* 2017; 3: 50–55
- [5] Wurtzel ET, Kutchan TM. Plant metabolism, the diverse chemistry set of the future. *Science* 2016; 353: 1232–1236
- [6] Harvey AL, Edrada-Ebel R, Quinn RJ. The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov* 2015; 14: 111–129
- [7] Buckingham J. *Dictionary of natural Products*. Abingdon: Taylor & Francis; 1993
- [8] Newman DJ, Cragg GM. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J Nat Prod* 2012; 75: 311–335
- [9] Rishton GM. Natural products as a robust source of new drugs and drug leads: past successes and present day issues. *Am J Cardio* 2008; 101: 43D–49D
- [10] Barker A, Kettle JG, Nowak T, Pease JE. Expanding medicinal chemistry space. *Drug Discov Today* 2013; 18: 298–304
- [11] Skircyc A, Kierszniowska S, Meret M, Willmitzer L, Tzotzos G. Medicinal bioprospecting of the Amazon rainforest: a modern Eldorado? *Trends Biotechnol* 2016; 34: 781–790
- [12] O'Connor SE. Engineering of secondary metabolism. *Annu Rev Genet* 2015; 49: 71–94

- [13] Dastmalchi M, Park MR, Morris JS, Facchini P. Family portraits: the enzymes behind benzyloquinoline alkaloid diversity. *Phytochem Rev* 2018; 17: 249–277
- [14] Pan Q, Mustafa NR, Tang K, Choi YH, Verpoorte R. Monoterpenoid indole alkaloids biosynthesis and its regulation in *Catharanthus roseus*: a literature review from genes to metabolites. *Phytochem Rev* 2016; 15: 221–250
- [15] Thamm AMK, Qu Y, De Luca V. Discovery and metabolic engineering of iridoid/secoiridoid and monoterpenoid indole alkaloid biosynthesis. *Phytochem Rev* 2016; 15: 339–361
- [16] Radwan MM, Wanas AS, Chandra S, ElSohly MA. Natural Cannabinoids of Cannabis and Methods of Analysis. In: Chandra S, Lata H, ElSohly MA, eds. *Cannabis sativa* L. – Botany and Biotechnology. Cham: Springer International Publishing; 2017: 161–182
- [17] Ashihara H, Mizuno K, Yokota T, Crozier A. Xanthine Alkaloids: Occurrence, Biosynthesis, and Function in Plants. In: Kinghorn AD, Falk H, Gibbons S, Kobayashi JI, eds. *Progress in the Chemistry of organic natural Products* 105. Cham: Springer International Publishing; 2017: 1–88
- [18] Kim YJ, Zhang DB, Yang DC. Biosynthesis and biotechnological production of ginsenosides. *Biotechnol Adv* 2015; 33: 717–735
- [19] Yang BY, Xia YG, Pan J, Liu Y, Wang QH, Kuang HX. Phytochemistry and biosynthesis of δ -lactone withanolides. *Phytochem Rev* 2016; 15: 771–797
- [20] Xie DY, Ma DM, Judd R, Jones AL. Artemisinin biosynthesis in *Artemisia annua* and metabolic engineering: questions, challenges, and perspectives. *Phytochem Rev* 2016; 15: 1093–1114
- [21] Croteau R, Ketchum REB, Long RM, Kaspera R, Wildung MR. Taxol biosynthesis and molecular genetics. *Phytochem Rev* 2006; 5: 75–97
- [22] Calvin M, Benson AA. The path of carbon in photosynthesis. *Science* 1948; 107: 476–480
- [23] Benson A, Calvin M. The dark reductions of photosynthesis. *Science* 1947; 105: 648–649
- [24] Benson AA, Calvin M. Carbon dioxide fixation by green plants. *Annu Rev Plant Phys* 1950; 1: 25–42
- [25] Hartmann T. From waste products to ecochemicals: fifty years research of plant secondary metabolism. *Phytochemistry* 2007; 68: 2831–2846
- [26] Kim J, Buell CR. A revolution in plant metabolism: genome-enabled pathway discovery. *Plant Physiol* 2015; 169: 1532–1539
- [27] Fernie AR, Pichersky E. Focus issue on metabolism: metabolites, metabolites everywhere. *Plant Physiol* 2015; 169: 1421–1423
- [28] Xu SQ, Brockmoller T, Navarro-Quezada A, Kuhl H, Gase K, Ling ZH, Zhou WW, Kreitzer C, Stanke M, Tang HB, Lyons E, Pandey P, Pandey SP, Timmermann B, Gaquerel E, Baldwin IT. Wild tobacco genomes reveal the evolution of nicotine biosynthesis. *Proc Natl Acad Sci U S A* 2017; 114: 6133–6138
- [29] Huang RQ, O'Donnell AJ, Barboline JJ, Barkman TJ. Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes. *Proc Natl Acad Sci U S A* 2016; 113: 10613–10618
- [30] Facchini PJ, De Luca V. Opium poppy and Madagascar periwinkle: model non-model systems to investigate alkaloid biosynthesis in plants. *Plant J* 2008; 54: 763–784
- [31] Sun M, Lou W, Chun JY, Cho DS, Nadiminty N, Evans CP, Chen J, Yue J, Zhou Q, Gao AC. Sanguinarine suppresses prostate tumor growth and inhibits survivin expression. *Genes Cancer* 2010; 1: 283–292
- [32] Sun M, Chun J, Lou W, Gao A. Sanguinarine is a novel inhibitor of survivin and inhibits the growth of prostate cancer cells. *J Urology* 2010; 183: E459
- [33] Ballantyne JC, Shin NS. Efficacy of opioids for chronic pain: a review of the evidence. *Clin Journal Pain* 2008; 24: 469–478
- [34] Battersby AR, Harper BJT. Rate study on the incorporation of tyrosine into morphine, codeine and thebaine. *Tetrahedron Lett* 1960; 1: 21–24
- [35] Leete E. Biosynthesis of alkaloids. *Science* 1965; 147: 1000–1006
- [36] Rueffer M, Elshagi H, Nagakura N, Zenk MH. (S)-Norlaudanoline synthase – the 1st enzyme in the benzyloquinoline biosynthetic-pathway. *FEBS Lett* 1981; 129: 5–9
- [37] Samanani N, Liscombe DK, Facchini PJ. Molecular cloning and characterization of norcoclaurine synthase, an enzyme catalyzing the first committed step in benzyloquinoline alkaloid biosynthesis. *Plant J* 2004; 40: 302–313
- [38] Morishige T, Tsujita T, Yamada Y, Sato F. Molecular characterization of the S-adenosyl-L-methionine: 3'-hydroxy-N-methylcoclaurine 4'-O-methyltransferase involved in isoquinoline alkaloid biosynthesis in *Coptis japonica*. *J Biol Chem* 2000; 275: 23398–23405
- [39] Ounaron A, Decker G, Schmidt J, Lottspeich F, Kutchan TM. (R,S)-Reticuline 7-O-methyltransferase and (R,S)-norcoclaurine 6-O-methyltransferase of *Papaver somniferum* – cDNA cloning and characterization of methyl transfer enzymes of alkaloid biosynthesis in opium poppy. *Plant J* 2003; 36: 808–819
- [40] Choi KB, Morishige T, Shitan N, Yazaki K, Sato F. Molecular cloning and characterization of coclaurine N-methyltransferase from cultured cells of *Coptis japonica*. *J Biol Chem* 2002; 277: 830–835
- [41] Hagel JM, Facchini PJ. Dioxygenases catalyze the O-demethylation steps of morphine biosynthesis in opium poppy. *Nat Chem Biol* 2010; 6: 273–275
- [42] Desgagne-Penix I, Khan MF, Schriemer DC, Cram D, Nowak J, Facchini PJ. Integration of deep transcriptome and proteome analyses reveals the components of alkaloid metabolism in opium poppy cell cultures. *BMC Plant Biol* 2010; 10: 252
- [43] Desgagne-Penix I, Farrow SC, Cram D, Nowak J, Facchini PJ. Integration of deep transcript and targeted metabolite profiles for eight cultivars of opium poppy. *Plant Mol Biol* 2012; 79: 295–313
- [44] Winzer T, Kern M, King AJ, Larson TR, Teodor RI, Donninger SL, Li Y, Dowle AA, Cartwright J, Bates R, Ashford D, Thomas J, Walker C, Bowser TA, Graham IA. Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. *Science* 2015; 349: 309–312
- [45] Farrow SC, Hagel JM, Beaudoin GAW, Burns DC, Facchini PJ. Stereochemical inversion of (S)-reticuline by a cytochrome P450 fusion in opium poppy. *Nat Chem Biol* 2015; 11: 728–732
- [46] Hoffmann I, Oliw EH. Discovery of a linoleate 9S-dioxygenase and an allene oxide synthase in a fusion protein of *Fusarium oxysporum*. *J Lipid Res* 2013; 54: 3471–3480
- [47] Ye KQ, Ke Y, Keshava N, Shanks J, Kapp JA, Tekmal RR, Petros J, Joshi HC. Opium alkaloid noscapine is an antitumor agent that arrests metaphase and induces apoptosis in dividing cells. *Proc Natl Acad Sci U S A* 1998; 95: 1601–1606
- [48] Sung B, Ahn KS, Aggarwal BB. Noscapine, a benzyloquinoline alkaloid, sensitizes leukemic cells to chemotherapeutic agents and cytokines by modulating the NF-kappa B signaling pathway. *Cancer Res* 2010; 70: 3259–3268
- [49] Battersby AR, Foulkes DM, Hirst M, Parry GV, Staunton J. Alkaloid biosynthesis. Part XI. Studies related to formation and oxidation of reticuline in morphine biosynthesis. *J Chem Soc C* 1968: 210–216
- [50] Winzer T, Gazda V, He Z, Kaminski F, Kern M, Larson TR, Li Y, Meade F, Teodor R, Vaistij FE, Walker C, Bowser TA, Graham IA. A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Science* 2012; 336: 1704–1708
- [51] Nutzmann HW, Huang AC, Osbourn A. Plant metabolic clusters – from genetics to genomics. *New Phytol* 2016; 211: 771–789
- [52] Hudlicky T. Recent advances in process development for opiate-derived pharmaceutical agents. *Can J Chem* 2015; 93: 492–501
- [53] Thodey K, Galanie S, Smolke CD. A microbial biomanufacturing platform for natural and semisynthetic opioids. *Nat Chem Biol* 2014; 10: 837–844
- [54] Galanie S, Thodey K, Trenchard JJ, Interrante MF, Smolke CD. Complete biosynthesis of opioids in yeast. *Science* 2015; 349: 1095–1100

- [55] Nakagawa A, Matsumura E, Koyanagi T, Katayama T, Kawano N, Yoshimatsu K, Yamamoto K, Kumagai H, Sato F, Minami H. Total biosynthesis of opiates by stepwise fermentation using engineered *Escherichia coli*. *Nat Commun* 2016; 7: 10390
- [56] Fossati E, Ekins A, Narcross L, Zhu Y, Falgoutyret JP, Beaudoin GAW, Facchini PJ, Martin VJJ. Reconstitution of a 10-gene pathway for synthesis of the plant alkaloid dihydrosanguinarine in *Saccharomyces cerevisiae*. *Nat Commun* 2014; 5: 3283
- [57] Bates D, Eastman A. Microtubule destabilising agents: far more than just antimetabolic anticancer drugs. *Brit J Clin Pharmacol* 2017; 83: 255–268
- [58] Roepke J, Salim V, Wu M, Thamm AMK, Murata J, Ploss K, Boland W, De Luca V. Vinca drug components accumulate exclusively in leaf exudates of Madagascar periwinkle. *Proc Natl Acad Sci U S A* 2010; 107: 15287–15292
- [59] St-Pierre B, Vazquez-Flota FA, De Luca V. Multicellular compartmentation of *Catharanthus roseus* alkaloid biosynthesis predicts intercellular translocation of a pathway intermediate. *Plant Cell* 1999; 11: 887–900
- [60] Guirimand G, Guihur A, Ginis O, Poutrain P, Hericourt F, Oudin A, Lanoue A, St-Pierre B, Burlat V, Courdavault V. The subcellular organization of strictosidine biosynthesis in *Catharanthus roseus* epidermis highlights several trans-tonoplast translocations of intermediate metabolites. *FEBS J* 2011; 278: 749–763
- [61] Guirimand G, Guihur A, Poutrain P, Hericourt F, Mahroug S, St-Pierre B, Burlat V, Courdavault V. Spatial organization of the vindoline biosynthetic pathway in *Catharanthus roseus*. *J Plant Physiol* 2011; 168: 549–557
- [62] Vranova E, Coman D, Gruissem W. Network analysis of the MVA and MEP pathways for isoprenoid synthesis. *Annu Rev Plant Biol* 2013; 64: 665–700
- [63] Money T, Wright IG, Mccapra F, Hall ES, Scott AI. Biosynthesis of indole alkaloids. Vindoline. *J Am Chem Soc* 1968; 90: 4144–4150
- [64] Miettinen K, Dong LM, Navrot N, Schneider T, Burlat V, Pollier J, Woittiez L, van der Krol S, Lukan R, Ilc T, Verpoorte R, Oksman-Caldentey KM, Martinoia E, Bouwmeester H, Goossens A, Memelink J, Werck-Reichhart D. The seco-iridoid pathway from *Catharanthus roseus*. *Nat Commun* 2014; 5: 3608
- [65] Salim V, Yu F, Altarejos J, De Luca V. Virus-induced gene silencing identifies *Catharanthus roseus* 7-deoxyloganic acid-7-hydroxylase, a step in iridoid and monoterpene indole alkaloid biosynthesis. *Plant J* 2013; 76: 754–765
- [66] Asada K, Salim V, Masada-Atsumi S, Edmunds E, Nagatoshi M, Terasaka K, Mizukami H, De Luca V. A 7-Deoxyloganic acid glucosyltransferase contributes a key step in secologanin biosynthesis in Madagascar periwinkle. *Plant Cell* 2013; 25: 4123–4134
- [67] Salim V, Wiens B, Masada-Atsumi S, Yu F, De Luca V. 7-Deoxyloganic acid synthase catalyzes a key 3 step oxidation to form 7-deoxyloganic acid in *Catharanthus roseus* iridoid biosynthesis. *Phytochemistry* 2014; 101: 23–31
- [68] Geu-Flores F, Sherden NH, Courdavault V, Burlat V, Glenn WS, Wu C, Nims E, Cui YH, O'Connor SE. An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature* 2012; 492: 138–142
- [69] Mcknight TD, Roessner CA, Devagupta R, Scott AI, Nessler CL. Nucleotide-sequence of a cDNA-encoding the vacuolar protein strictosidine synthase from *Catharanthus roseus*. *Nucleic Acids Res* 1990; 18: 4939
- [70] Geerlings A, Ibanez MML, Memelink J, van der Heijden R, Verpoorte R. Molecular cloning and analysis of strictosidine beta-d-glucosidase, an enzyme in terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *J Biol Chem* 2000; 275: 3051–3056
- [71] Tatsis EC, Carqueijeiro I, de Bernonville TD, Franke J, Dang TTT, Oudin A, Lanoue A, Lafontaine F, Stavrinides AK, Clastre M, Courdavault V, O'Connor SE. A three enzyme system to generate the Strychnos alkaloid scaffold from a central biosynthetic intermediate. *Nat Commun* 2017; 8: 316
- [72] Stavrinides A, Tatsis Evangelos C, Foureau E, Caputi L, Kellner F, Courdavault V, O'Connor SE. Unlocking the diversity of alkaloids in *Catharanthus roseus*: nuclear localization suggests metabolic channeling in secondary metabolism. *Chem Biol* 2015; 22: 336–341
- [73] Besseau S, Kellner F, Lanoue A, Thamm AMK, Salim V, Schneider B, Geu-Flores F, Hofer R, Guirimand G, Guihur A, Oudin A, Glevarec G, Foureau E, Papon N, Clastre M, Giglioli-Guivarc'h N, St-Pierre B, Werck-Reichhart D, Burlat V, De Luca V, O'Connor SE, Courdavault V. A pair of tabersonine 16-hydroxylases initiates the synthesis of vindoline in an organ-dependent manner in *Catharanthus roseus*. *Plant Physiol* 2013; 163: 1792–1803
- [74] Kellner F, Kim J, Clavijo BJ, Hamilton JP, Childs KL, Vaillancourt B, Cepela J, Habermann M, Steuernagel B, Clissold L, McLay K, Buell CR, O'Connor SE. Genome-guided investigation of plant natural product biosynthesis. *Plant J* 2015; 82: 680–692
- [75] Qu Y, Easson ML, Froese J, Simionescu R, Hudlicky T, De Luca V. Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proc Natl Acad Sci U S A* 2015; 112: 6224–6229
- [76] Meijer AH, Dewaal A, Verpoorte R. Purification of the cytochrome-P-450 enzyme geraniol 10-hydroxylase from cell-cultures of *Catharanthus roseus*. *J Chromatogr* 1993; 635: 237–249
- [77] Collu G, Unver N, Peltenburg-Looman AMG, van der Heijden R, Verpoorte R, Memelink J. Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett* 2001; 508: 215–220
- [78] Luijendijk TJC, Stevens LH, Verpoorte R. Purification and characterisation of strictosidine beta-D-glucosidase from *Catharanthus roseus* cell suspension cultures. *Plant Physiol Bioch* 1998; 36: 419–425
- [79] Rischer H, Oresic M, Seppanen-Laakso T, Katajamaa M, Lammertyn F, Ardiles-Diaz W, Van Montagu MCE, Inze D, Oksman-Caldentey KM, Goossens A. Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* cells. *Proc Natl Acad Sci U S A* 2006; 103: 5614–5619
- [80] Murata J, Roepke J, Gordon H, De Luca V. The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *Plant Cell* 2008; 20: 524–542
- [81] Gongora-Castillo E, Childs KL, Fedewa G, Hamilton JP, Liscombe DK, Magallanes-Lundback M, Mandadi KK, Nims E, Runguphan W, Vaillancourt B, Varbanova-Herde M, DellaPenna D, McKnight TD, O'Connor S, Buell CR. Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. *PLoS One* 2012; 7: e52506
- [82] Van Moerkercke A, Fabris M, Pollier J, Baart GJE, Rombauts S, Hasnain G, Rischer H, Memelink J, Oksman-Caldentey KM, Goossens A. CathaCyc, a metabolic pathway database built from *Catharanthus roseus* RNA-seq data. *Plant Cell Physiol* 2013; 54: 673–685
- [83] Xiao M, Zhang Y, Chen X, Lee EJ, Barber CJS, Chakrabarty R, Desgagne-Penix I, Haslam TM, Kim YB, Liu EW, MacNevin G, Masada-Atsumi S, Reed DW, Stout JM, Zerbe P, Zhang YS, Bohlmann J, Covello PS, De Luca V, Page JE, Ro DK, Martin VJJ, Facchini PJ, Sensen CW. Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J Biotechnol* 2013; 166: 122–134
- [84] Verma M, Chhangal R, Sharma R, Sinha AK, Jain M. Transcriptome analysis of *Catharanthus roseus* for gene discovery and expression profiling. *PLoS One* 2014; 9: e103583
- [85] de Bernonville TD, Foureau E, Parage C, Lanoue A, Clastre M, Londono MA, Oudin A, Houille B, Papon N, Besseau S, Glevarec G, Atehortua L, Giglioli-Guivarc'h N, St-Pierre B, De Luca V, O'Connor SE, Courdavault V. Characterization of a second secologanin synthase isoform producing both secologanin and secoxyloganin allows enhanced *de novo* assembly of a *Catharanthus roseus* transcriptome. *BMC Genomics* 2015; 16: 619

- [86] Giddings LA, Liscombe DK, Hamilton JP, Childs KL, DellaPenna D, Buell CR, O'Connor SE. A stereoselective hydroxylation step of alkaloid biosynthesis by a unique cytochrome P450 in *Catharanthus roseus*. *J Biol Chem* 2011; 286: 16751–16757
- [87] Guirimand G, Courdavault V, Lanoue A, Mahroug S, Guihur A, Blanc N, Giglioli-Guivarc'h N, St-Pierre B, Burlat V. Strictosidine activation in Apocynaceae: towards a “nuclear time bomb”? *BMC Plant Biol* 2010; 10: 182
- [88] Yu F, De Luca V. ATP-binding cassette transporter controls leaf surface secretion of anticancer drug components in *Catharanthus roseus*. *Proc Natl Acad Sci U S A* 2013; 110: 15830–15835
- [89] Carqueijeiro I, Noronha H, Duarte P, Geros H, Sottomayor M. Vacuolar transport of the medicinal alkaloids from *Catharanthus roseus* is mediated by a proton-driven antiport. *Plant Physiol* 2013; 162: 1486–1496
- [90] Payne RME, Xu DY, Foureau E, Teto Carqueijeiro MI, Oudin A, de Bernonville TD, Novak V, Burrow M, Olsen CE, Jones DM, Tatsis EC, Pendle A, Halkier BA, Geu-Flores F, Courdavault V, Nour-Eldin HH, O'Connor SE. An NPF transporter exports a central monoterpene indole alkaloid intermediate from the vacuole. *Nat Plants* 2017; 3: 16208
- [91] Brown S, Clastre M, Courdavault V, O'Connor SE. *De novo* production of the plant-derived alkaloid strictosidine in yeast. *Proc Natl Acad Sci U S A* 2015; 112: 3205–3210
- [92] Kennedy J. Mutasynthesis, chemobiosynthesis, and back to semi-synthesis: combining synthetic chemistry and biosynthetic engineering for diversifying natural products. *Nat Prod Rep* 2008; 25: 25–34
- [93] Mechoulam R, Hanus LO, Pertwee R, Howlett AC. Early phytocannabinoid chemistry to endocannabinoids and beyond. *Nat Rev Neurosci* 2014; 15: 757–764
- [94] Owens B. Drug development: the treasure chest. *Nature* 2015; 525: S6
- [95] Pain S. A potted history. *Nature* 2015; 525: S10
- [96] Andre CM, Hausman JF, Guerriero G. *Cannabis sativa*: the plant of the thousand and one molecules. *Front Plant Sci* 2016; 7: 19
- [97] Grotenhermen F, Muller-Vahl K. Medicinal uses of marijuana and cannabinoids. *Crit Rev Plant Sci* 2016; 35: 378–405
- [98] Todd AR. Hashish. *Experientia* 1946; 2: 55–60
- [99] Mechoulam R, Shvo Y. Hashish. 1. Structure of cannabidiol. *Tetrahedron* 1963; 19: 2073–2078
- [100] Gaoni Y, Mechoulam R. Isolation, structure, and partial synthesis of an active constituent of hashish. *J Am Chem Soc* 1964; 86: 1646–1647
- [101] Devane WA, Dysarz FA, Johnson MR, Melvin LS, Howlett AC. Determination and characterization of a cannabinoid receptor in rat-brain. *Mol Pharmacol* 1988; 34: 605–613
- [102] Matsuda LA, Lolait SJ, Brownstein MJ, Young AC, Bonner TI. Structure of a cannabinoid receptor and functional expression of the cloned cDNA. *Nature* 1990; 346: 561–564
- [103] Munro S, Thomas KL, Abushaar M. Molecular characterization of a peripheral receptor for cannabinoids. *Nature* 1993; 365: 61–65
- [104] Devane WA, Hanus L, Breuer A, Pertwee RG, Stevenson LA, Griffin G, Gibson D, Mandelbaum A, Etinger A, Mechoulam R. Isolation and structure of a brain constituent that binds to the cannabinoid receptor. *Science* 1992; 258: 1946–1949
- [105] Russo EB. Beyond cannabis: plants and the endocannabinoid system. *Trends Pharmacol Sci* 2016; 37: 594–605
- [106] van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, Page JE. The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol* 2011; 12: R102
- [107] Gagne SJ, Stout JM, Liu EW, Boubakir Z, Clark SM, Page JE. Identification of olivetolic acid cyclase from *Cannabis sativa* reveals a unique catalytic route to plant polyketides. *Proc Natl Acad Sci U S A* 2012; 109: 12811–12816
- [108] Stout JM, Boubakir Z, Ambrose SJ, Purves RW, Page JE. The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in *Cannabis sativa* trichomes. *Plant J* 2012; 71: 353–365
- [109] Kajima M, Piraux M. The biogenesis of cannabinoids in *Cannabis sativa*. *Phytochemistry* 1982; 21: 67–69
- [110] Fellermeier M, Eisenreich W, Bacher A, Zenk MH. Biosynthesis of cannabinoids – incorporation experiments with C-13-labeled glucoses. *Eur J Biochem* 2001; 268: 1596–1604
- [111] Fellermeier M, Zenk MH. Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol. *FEBS Lett* 1998; 427: 283–285
- [112] Page JE, Boubakir Z. Aromatic prenyltransferase from cannabis. US: National Research Council of Canada, University of Saskatchewan. US patent US20150128301 A1; 2014
- [113] Sirikantaramas S, Taura F. Cannabinoids: Biosynthesis and biotechnological Applications. In: Chandra S, Lata H, ElSohly MA, eds. *Cannabis sativa* L. – Botany and Biotechnology. Cham: Springer International Publishing; 2017: 183–206
- [114] Weiblen GD, Wenger JP, Craft KJ, ElSohly MA, Mehmedic Z, Treiber EL, Marks MD. Gene duplication and divergence affecting drug content in *Cannabis sativa*. *New Phytol* 2015; 208: 1241–1250
- [115] Sirikantaramas S, Morimoto S, Shoyama Y, Ishikawa Y, Wada Y, Shoyama Y, Taura F. The gene controlling marijuana psychoactivity – molecular cloning and heterologous expression of delta(1)-tetrahydrocannabinolic acid synthase from *Cannabis sativa* L. *J Biol Chem* 2004; 279: 39767–39774
- [116] Taura F, Morimoto S, Shoyama Y. Purification and characterization of cannabidiolic-acid synthase from *Cannabis sativa* L. Biochemical analysis of a novel enzyme that catalyzes the oxidocyclization of cannabigerolic acid to cannabidiolic acid. *J Biol Chem* 1996; 271: 17411–17416
- [117] Taura F, Sirikantaramas S, Shoyama Y, Yoshikai K, Shoyama Y, Morimoto S. Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type *Cannabis sativa*. *FEBS Lett* 2007; 581: 2929–2934
- [118] Shimizu Y, Ogata H, Goto S. Type III polyketide synthases: functional classification and phylogenomics. *Chembiochem* 2017; 18: 50–65
- [119] Marks MD, Tian L, Wenger JP, Omburo SN, Soto-Fuentes W, He J, Gang DR, Weiblen GD, Dixon RA. Identification of candidate genes affecting delta(9)-tetrahydrocannabinol biosynthesis in *Cannabis sativa*. *J Exp Bot* 2009; 60: 3715–3726
- [120] Taura F, Tanaka S, Taguchi C, Fukamizu T, Tanaka H, Shoyama Y, Morimoto S. Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway. *FEBS Lett* 2009; 583: 2061–2066
- [121] Shen YM, Yoon P, Yu TW, Floss HG, Hopwood D, Moore BS. Ectopic expression of the minimal whiE polyketide synthase generates a library of aromatic polyketides of diverse sizes and shapes. *Proc Natl Acad Sci U S A* 1999; 96: 3622–3627
- [122] MacCallum CA, Russo EB. Practical considerations in medical cannabis administration and dosing. *Eur J Intern Med* 2018; 49: 12–19
- [123] Burstein S. Cannabidiol (CBD) and its analogs: a review of their effects on inflammation. *Bioorgan Med Chem* 2015; 23: 1377–1385
- [124] Flores-Sanchez JJ, Pec J, Fei JN, Choi YH, Dusek J, Verpoorte R. Elicitation studies in cell suspension cultures of *Cannabis sativa* L. *J Biotechnol* 2009; 143: 157–168
- [125] Farag S, Kayser O. Cannabinoids production by hairy root cultures of *Cannabis sativa* L. *Am J Plant Sci* 2015; 6: 1874–1884
- [126] Lange K, Schmid A, Julsing MK. Delta(9)-tetrahydrocannabinolic acid synthase production in *Pichia pastoris* enables chemical synthesis of cannabinoids. *J Biotechnol* 2015; 211: 68–76

- [127] Lange K, Schmid A, Julsing MK. Delta 9-tetrahydrocannabinolic acid synthase: the application of a plant secondary metabolite enzyme in biocatalytic chemical synthesis. *J Biotechnol* 2016; 233: 42–48
- [128] Denoëud F, Carretero-Paulet L, Dereeper A, Droc G, Guyot R, Pietrella M, Zheng CF, Alberti A, Anthony F, Aprea G, Aury JM, Bento P, Bernard M, Bocs S, Campa C, Cenci A, Combes MC, Crouzillat D, Da Silva C, Daddiego L, De Bellis F, Dussert S, Garsmeur O, Gayraud T, Guignon V, Jahn K, Jamilloux V, Joet T, Labadie K, Lan TY, Leclercq J, Lepelley M, Leroy T, Li LT, Librado P, Lopez L, Munoz A, Noel B, Pallavicini A, Perrotta G, Poncet V, Pot D, Priyono, Rigoreau M, Rouard M, Rozas J, Tranchant-Dubreuil C, VanBuren R, Zhang Q, Andrade AC, Argout X, Bertrand B, de Kochko A, Graziosi G, Henry RJ, Jayarama, Ming R, Nagai C, Rounsley S, Sankoff D, Giuliano G, Albert VA, Wincker P, Lashermes P. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* 2014; 345: 1181–1184
- [129] Xia EH, Zhang HB, Sheng J, Li K, Zhang QJ, Kim C, Zhang Y, Liu Y, Zhu T, Li W, Huang H, Tong Y, Nan H, Shi C, Shi C, Jiang JJ, Mao SY, Jiao JY, Zhang D, Zhao Y, Zhao YJ, Zhang LP, Liu YL, Liu BY, Yu Y, Shao SF, Ni DJ, Eichler EE, Gao LZ. The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. *Mol Plant* 2017; 10: 866–877
- [130] Argout X, Salse J, Aury JM, Guiltinan MJ, Droc G, Gouzy J, Allegre M, Chaparro C, Legavre T, Maximova SN, Abrouk M, Murat F, Fouet O, Poulain J, Ruiz M, Roguet Y, Rodier-Goud M, Barbosa-Neto JF, Sabot F, Kudrna D, Ammiraju JSS, Schuster SC, Carlson JE, Sallet E, Schiex T, Dievart A, Kramer M, Gelley L, Shi Z, Berard A, Viot C, Boccara M, Risterucci AM, Guignon V, Sabau X, Axtell MJ, Ma ZR, Zhang YF, Brown S, Bourge M, Golser W, Song XA, Clement D, Rivallan R, Tahi M, Akaza JM, Pitollat B, Gramacho K, D'Hont A, Brunel D, Infante D, Kebe I, Costet P, Wing R, McCombie WR, Guiderdoni E, Quetier F, Panaud O, Wincker P, Bocs S, Lanaud C. The genome of *Theobroma cacao*. *Nat Genet* 2011; 43: 101–108
- [131] Frischknecht PM, Baumann TW. Stress-induced formation of purine alkaloids in plant-tissue culture of *Coffea arabica*. *Phytochemistry* 1985; 24: 2255–2257
- [132] Hollingsworth RG, Armstrong JW, Campbell E. Pest control: caffeine as a repellent for slugs and snails. *Nature* 2002; 417: 915–916
- [133] Uefuji H, Tatsumi Y, Morimoto M, Kaothien-Nakayama P, Ogita S, Sano H. Caffeine production in tobacco plants by simultaneous expression of three coffee N-methyltransferases and its potential as a pest repellent. *Plant Mol Biol* 2005; 59: 221–227
- [134] Friedman J, Waller GR. Caffeine hazards and their prevention in germinating-seeds of coffee (*Coffea arabica* L.). *J Chem Ecol* 1983; 9: 1099–1106
- [135] Wang L, Nagele T, Doerfler H, Fagner L, Chaturvedi P, Nukarinen E, Bellaire A, Huber W, Weiszmann J, Engelmeier D, Ramsak Z, Gruden K, Weckwerth W. System level analysis of cacao seed ripening reveals a sequential interplay of primary and secondary metabolism leading to polyphenol accumulation and preparation of stress resistance. *Plant J* 2016; 87: 318–332
- [136] García-Flores M, Juárez-Colunga S, García-Casarrubias A, Trachsel S, Winkler R, Tiessen A. Metabolic profiling of plant extracts using direct-injection electrospray ionization mass spectrometry allows for high-throughput phenotypic characterization according to genetic and environmental effects. *J Agric Food Chem* 2015; 63: 1042–1052
- [137] Chang KL, Ho PC. Gas chromatography time-of-flight mass spectrometry (GC-TOF-MS)-based metabolomics for comparison of caffeinated and decaffeinated coffee and its implications for Alzheimer's disease. *PLoS One* 2014; 9: e104621
- [138] Wei F, Furihata K, Koda M, Hu F, Miyakawa T, Tanokura M. Roasting process of coffee beans as studied by nuclear magnetic resonance: time course of changes in composition. *J Agric Food Chem* 2012; 60: 1005–1012
- [139] Kellogg JJ, Wallace ED, Graf TN, Oberlies NH, Cech NB. Conventional and accelerated-solvent extractions of green tea (*Camellia sinensis*) for metabolomics-based chemometrics. *J Pharmaceut Biomed* 2017; 145: 604–610
- [140] Li P, Dai W, Lu M, Xie D, Tan J, Yang C, Zhu Y, Lv H, Peng Q, Zhang Y, Guo L, Ni D, Lin Z. Metabolomic analysis reveals the composition differences in 13 Chinese tea cultivars of different manufacturing suitability. *J Sci Food Agric* 2018; 98: 1153–1161
- [141] Kellogg JJ, Graf TN, Paine MF, McCune JS, Kvalheim OM, Oberlies NH, Cech NB. Comparison of metabolomics approaches for evaluating the variability of complex botanical preparations: green tea (*Camellia sinensis*) as a case study. *J Nat Prod* 2017; 80: 1457–1466
- [142] Cheng B, Furtado A, Henry RJ. Long-read sequencing of the coffee bean transcriptome reveals the diversity of full-length transcripts. *Gigascience* 2017; 6: 1–13
- [143] Zhang QF, Liu MY, Ruan JY. Integrated transcriptome and metabolic analyses reveals novel insights into free amino acid metabolism in *Huangjinya* tea cultivar. *Front Plant Sci* 2017; 8: 291
- [144] Mukhopadhyay M, Mondal TK, Chand PK. Biotechnological advances in tea (*Camellia sinensis* [L.] O. Kuntze): a review. *Plant Cell Rep* 2016; 35: 255–287
- [145] da Hora Junior BT, Poloni Jde F, Lopes MA, Dias CV, Gramacho KP, Schuster I, Sabau X, Cascardo JC, Mauro SM, Gesteira Ada S, Bonatto D, Micheli F. Transcriptomics and systems biology analysis in identification of specific pathways involved in cacao resistance and susceptibility to witches' broom disease. *Mol Biosyst* 2012; 8: 1507–1519
- [146] Dereeper A, Bocs S, Rouard M, Guignon V, Ravel S, Tranchant-Dubreuil C, Poncet V, Garsmeur O, Lashermes P, Droc G. The coffee genome hub: a resource for coffee genomes. *Nucleic Acids Res* 2015; 43: D1028–D1035
- [147] Campos NA, Panis B, Carpentier SC. Somatic embryogenesis in coffee: the evolution of biotechnology and the integration of omics technologies offer great opportunities. *Front Plant Sci* 2017; 8: 1460
- [148] Weinreb O, Amit T, Youdim MBH. A novel approach of proteomics and transcriptomics to study the mechanism of action of the antioxidant-iron chelator green tea polyphenol (–)-epigallocatechin-3-gallate. *Free Radic Biol Med* 2007; 43: 546–556
- [149] Li CF, Zhu Y, Yu Y, Zhao QY, Wang SJ, Wang XC, Yao MZ, Luo D, Li X, Chen L, Yang YJ. Global transcriptome and gene regulation network for secondary metabolite biosynthesis of tea plant (*Camellia sinensis*). *BMC Genomics* 2015; 16: 560
- [150] Tai Y, Wei C, Yang H, Zhang L, Chen Q, Deng W, Wei S, Zhang J, Fang C, Ho C, Wan X. Transcriptomic and phytochemical analysis of the biosynthesis of characteristic constituents in tea (*Camellia sinensis*) compared with oil tea (*Camellia oleifera*). *BMC Plant Biol* 2015; 15: 190
- [151] Taylor AE, Davey Smith G, Munafò MR. Associations of coffee genetic risk scores with consumption of coffee, tea and other beverages in the UK Biobank. *Addiction* 2018; 113: 148–157
- [152] James JE, Stirling KP. Caffeine – a survey of some of the known and suspected deleterious effects of habitual use. *Brit J Addict* 1983; 78: 251–258
- [153] Program BCDS. Coffee drinking and acute myocardial infarction. *Lancet* 1972; 300: 1278–1281
- [154] Jick H, Miettinen OS, Neff RK, Shapiro S, Heinonen OP, Slone D. Coffee and myocardial-infarction. *N Engl J Med* 1973; 289: 63–67
- [155] Freedman ND, Park Y, Abnet CC, Hollenbeck AR, Sinha R. Association of coffee drinking with total and cause-specific mortality. *N Engl J Med* 2012; 366: 1891–1904
- [156] Alicandro G, Tavani A, La Vecchia C. Coffee and cancer risk: a summary overview. *Eur J Cancer Prev* 2017; 26: 424–432
- [157] Ahmed I, Lee PC, Lill CM, Nielsen SS, Artaud F, Gallagher LG, Lorient MA, Mulot C, Nacfer M, Liu T, Biernacka JM, Armasu S, Anderson K, Farin FM, Lassen CF, Hansen J, Olsen JH, Bertram L, Maraganore DM, Checkoway H, Ritz B, Elbaz A. Lack of replication of the GRIN2A-by-coffee interaction in Parkinson disease. *PLoS Genet* 2014; 10: e1004788

- [158] Tanida I, Shirasago Y, Suzuki R, Abe R, Wakita T, Hanada K, Fukasawa M. Inhibitory effects of caffeic acid, a coffee-related organic acid, on the propagation of hepatitis C virus. *Jpn J Infect Dis* 2015; 68: 268–275
- [159] Platt DE, Ghassibe-Sabbagh M, Salameh P, Salloum AK, Haber M, Mouzaya F, Gauguier D, Al-Sarraj Y, El-Shanti H, Zalloua PA, Abchee AB. Caffeine impact on metabolic syndrome components is modulated by a CYP1A2 variant. *Ann Nutr Metab* 2016; 68: 1–11
- [160] Jaquet M, Rochat I, Moulin J, Cavin C, Bibiloni R. Impact of coffee consumption on the gut microbiota: a human volunteer study. *Int J Food Microbiol* 2009; 130: 117–121
- [161] Crocq MA. Alcohol, nicotine, caffeine, and mental disorders. *Dialogues Clin Neurosci* 2003; 5: 175–185
- [162] Nag SA, Qin JJ, Wang W, Wang MH, Wang H, Zhang RW. Ginsenosides as anticancer agents: *in vitro* and *in vivo* activities, structure-activity relationships, and molecular mechanisms of action. *Front Pharmacol* 2012; 3: 25
- [163] Hemmerly TE. Ginseng farm in Lawrence-County, Tennessee. *Econ Bot* 1977; 31: 160–162
- [164] Oh JY, Kim YJ, Jang MG, Joo SC, Kwon WS, Kim SY, Jung SK, Yang DC. Investigation of ginsenosides in different tissues after elicitor treatment in *Panax ginseng*. *J Ginseng Res* 2014; 38: 270–277
- [165] Christensen LP. Ginsenosides: Chemistry, Biosynthesis, Analysis, and potential Health Effects. In: Taylor S, ed. *Advances in Food and Nutrition Research*, vol. 55. Cambridge: Academic Press; 2008: 1–99
- [166] Wang CZ, Aung HH, Ni M, Wu JA, Tong RB, Wicks S, He TC, Yuan CS. Red American ginseng: ginsenoside constituents and antiproliferative activities of heat-processed *Panax quinquefolius* roots. *Planta Med* 2007; 73: 669–674
- [167] Yue PYK, Wong DYL, Wu PK, Leung PY, Mak NK, Yeung HW, Liu L, Cai Z, Jiang ZH, Fan TPD, Wong RNS. The angiostatic effects of 20(R)-ginsenoside Rg3. *Biochem Pharmacol* 2006; 72: 437–445
- [168] Yu Y, Zhou Q, Hang Y, Bu X, Jia W. Antiestrogenic effect of 20S-protopanaxadiol and its synergy with tamoxifen on breast cancer cells. *Cancer* 2007; 109: 2374–2382
- [169] Wang W, Zhao Y, Rayburn ER, Hill DL, Wang H, Zhang R. *In vitro* anti-cancer activity and structure-activity relationships of natural products isolated from fruits of *Panax ginseng*. *Cancer Chemother Pharmacol* 2007; 59: 589–601
- [170] Wang W, Rayburn ER, Hao M, Zhao Y, Hill DL, Zhang RW, Wang H. Experimental therapy of prostate cancer with novel natural product anti-cancer ginsenosides. *Prostate* 2008; 68: 809–819
- [171] Wang W, Wang H, Rayburn ER, Zhao Y, Hill DL, Zhang R. 20(S)-25-methoxyl-dammarene-3 beta,12 beta,20-triol, a novel natural product for prostate cancer therapy: activity *in vitro* and *in vivo* and mechanisms of action. *Br J Cancer* 2008; 98: 792–802
- [172] Wang W, Rayburn ER, Hang J, Zhao Y, Wang H, Zhang R. Anti-lung cancer effects of novel ginsenoside 25-OCH3-PPD. *Lung Cancer* 2009; 65: 306–311
- [173] Kang KS, Kim HY, Yamabe N, Yokozawa T. Stereospecificity in hydroxyl radical scavenging activities of four ginsenosides produced by heat processing. *Bioorg Med Chem Lett* 2006; 16: 5028–5031
- [174] Schramek N, Huber C, Schmidt S, Dvorski SEM, Knispel N, Ostrozhenkova E, Pena-Rodriguez LM, Cusido RM, Wischmann G, Eisenreich W. Biosynthesis of ginsenosides in field-grown *Panax ginseng*. *JSM Biotechnol Bioeng* 2014; 2: 1033
- [175] Zhao S, Wang L, Liu L, Liang Y, Sun Y, Wu J. Both the mevalonate and the non-mevalonate pathways are involved in ginsenoside biosynthesis. *Plant Cell Rep* 2014; 33: 393–400
- [176] Wang J, Gao WY, Zhang J, Zuo BM, Zhang LM, Huang LQ. Advances in study of ginsenoside biosynthesis pathway in *Panax ginseng* C.A. Meyer. *Acta Physiol Plant* 2012; 34: 397–403
- [177] Xu R, Fazio GC, Matsuda SPT. On the origins of triterpenoid skeletal diversity. *Phytochemistry* 2004; 65: 261–291
- [178] Wang L, Zhao SJ, Cao HJ, Sun Y. The isolation and characterization of dammarenediol synthase gene from *Panax quinquefolius* and its heterologous co-expression with cytochrome P450 gene PqD12H in yeast. *Funct Integr Genomics* 2014; 14: 545–557
- [179] Hu W, Liu N, Tian Y, Zhang L. Molecular cloning, expression, purification, and functional characterization of dammarenediol synthase from *Panax ginseng*. *Biomed Res Int* 2013; 2013: 7
- [180] Jung SC, Kim W, Park SC, Jeong J, Park MK, Lim S, Lee Y, Im WT, Lee JH, Choi G, Kim SC. Two ginseng UDP-glycosyltransferases synthesize ginsenoside Rg(3) and Rd. *Plant Cell Physiol* 2014; 55: 2177–2188
- [181] Wei W, Wang PP, Wei YJ, Liu QF, Yang CS, Zhao GP, Yue JM, Yan X, Zhou ZH. Characterization of *Panax ginseng* UDP-glycosyltransferases catalyzing protopanaxatriol and biosyntheses of bioactive ginsenosides F1 and Rh1 in metabolically engineered yeasts. *Mol Plant* 2015; 8: 1412–1424
- [182] Hasegawa H. Proof of the mysterious efficacy of ginseng: basic and clinical trials: metabolic activation of ginsenoside: deglycosylation by intestinal bacteria and esterification with fatty acid. *J Pharmacol Sci* 2004; 95: 153–157
- [183] Zheng ZZ, Ming YL, Chen LH, Zheng GH, Liu SS, Chen QX. Compound K-induced apoptosis of human hepatocellular carcinoma MHCC97-H cells *in vitro*. *Oncol Rep* 2014; 32: 325–331
- [184] Lee S, Kwon MC, Jang JP, Sohng JK, Jung HJ. The ginsenoside metabolite compound K inhibits growth, migration and stemness of glioblastoma cells. *Int J Oncol* 2017; 51: 414–424
- [185] Yan X, Fan Y, Wei W, Wang PP, Liu QF, Wei YJ, Zhang L, Zhao GP, Yue JM, Zhou ZH. Production of bioactive ginsenoside compound K in metabolically engineered yeast. *Cell Res* 2014; 24: 770–773
- [186] Wu JY, Zhong JJ. Production of ginseng and its bioactive components in plant cell culture: current technological and applied aspects. *J Biotechnol* 1999; 68: 89–99
- [187] Baque MA, Moh SH, Lee EJ, Zhong JJ, Paek KY. Production of biomass and useful compounds from adventitious roots of high-value added medicinal plants using bioreactor. *Biotechnol Adv* 2012; 30: 1255–1267
- [188] Lee MH, Jeong JH, Seo JW, Shin CG, Kim YS, In JG, Yang DC, Yi JS, Choi YE. Enhanced triterpene and phytosterol biosynthesis in *Panax ginseng* overexpressing squalene synthase gene. *Plant Cell Physiol* 2004; 45: 976–984
- [189] Sivakumar G, Yu KW, Paek KY. Biosafe ginseng: a novel source for human well-being. *Eng Life Sci* 2005; 5: 527–533
- [190] Luo S. Callus culture of *Panax ginseng*. *Plant Physiol Commun* 1964; 2: 26–38
- [191] Furuya T, Yoshikawa T, Orihara Y, Oda H. Studies on plant-tissue cultures. 39. Saponin production in cell-suspension cultures of *Panax ginseng*. *Planta Med* 1983; 48: 83–87
- [192] Inomata S, Yokoyama M, Gozu Y, Shimizu T, Yanagi M. Growth-pattern and ginsenoside production of agrobacterium-transformed *Panax ginseng* roots. *Plant Cell Rep* 1993; 12: 681–686
- [193] Choi KT, Ahn IO, Park JC. Production of ginseng saponin in tissue-culture of ginseng (*Panax ginseng* Ca Meyer). *Russ J Plant Physiol* 1994; 41: 784–788
- [194] Zhang YH, Zhong JJ, Yu JY. Enhancement of ginseng saponin production in suspension cultures of *Panax notoginseng*: manipulation of medium sucrose. *J Biotechnol* 1996; 51: 49–56
- [195] Zhong JJ, Zhu QX. Effect of initial phosphate concentration on cell growth and ginsenoside saponin production by suspended cultures of *Panax notoginseng*. *Biotechnol Appl Biochem* 1995; 55: 241–247
- [196] Paek KY, Murthy HN, Hahn EJ, Zhong JJ. Large Scale Culture of Ginseng adventitious Roots for Production of Ginsenosides. In: Zhong JJ, Bai FW, Zhang W, eds. *Biotechnology in China I: From Bioreaction to Bio-separation and Bioremediation*. Berlin, Heidelberg: Springer; 2009: 151–176

- [197] Gao X, Zhu C, Jia W, Gao W, Qiu M, Zhang Y, Xiao P. Induction and characterization of adventitious roots directly from the explants of *Panax notoginseng*. *Biotechnol Lett* 2005; 27: 1771–1775
- [198] Osbourn AE. Preformed antimicrobial compounds and plant defense against fungal attack. *Plant Cell* 1996; 8: 1821–1831
- [199] Augustin JM, Kuzina V, Andersen SB, Bak S. Molecular activities, biosynthesis and evolution of triterpenoid saponins. *Phytochemistry* 2011; 72: 435–457
- [200] Ali MB, Yu KW, Hahn EJ, Paek KY. Methyl jasmonate and salicylic acid elicitation induces ginsenosides accumulation, enzymatic and non-enzymatic antioxidant in suspension culture *Panax ginseng* roots in bioreactors. *Plant Cell Rep* 2006; 25: 613–620
- [201] Kim DS, Song M, Kim SH, Jang DS, Kim JB, Ha BK, Kim SH, Lee KJ, Kang SY, Jeong IY. The improvement of ginsenoside accumulation in *Panax ginseng* as a result of gamma-irradiation. *J Ginseng Res* 2013; 37: 332–340
- [202] Kim YJ, Lee OR, Oh JY, Jang MG, Yang DC. Functional analysis of 3-hydroxy-3-methylglutaryl coenzyme a reductase encoding genes in triterpene saponin-producing ginseng. *Plant Physiol* 2014; 165: 373–387
- [203] Han JY, Kim HJ, Kwon YS, Choi YE. The Cyt P450 enzyme CYP716A47 catalyzes the formation of protopanaxadiol from dammarenediol-II during ginsenoside biosynthesis in *Panax ginseng*. *Plant Cell Physiol* 2011; 52: 2062–2073
- [204] Han JY, Kwon YS, Yang DC, Jung YR, Choi YE. Expression and RNA interference-induced silencing of the dammarenediol synthase gene in *Panax ginseng*. *Plant Cell Physiol* 2006; 47: 1653–1662
- [205] Dai Z, Wang B, Liu Y, Shi M, Wang D, Zhang X, Liu T, Huang L, Zhang X. Producing aglycons of ginsenosides in bakers' yeast. *Sci Rep* 2014; 4: 3698
- [206] Mirjalili MH, Moyano E, Bonfill M, Cusido RM, Palazon J. Steroidal lactones from *Withania somnifera*, an ancient plant for novel medicine. *Molecules* 2009; 14: 2373–2393
- [207] Dar NJ, Hamid A, Ahmad M. Pharmacologic overview of *Withania somnifera*, the Indian ginseng. *Cell Mol Life Sci* 2015; 72: 4445–4460
- [208] Rai M, Jogee PS, Agarkar G, dos Santos CA. Anticancer activities of *Withania somnifera*: current research, formulations, and future perspectives. *Pharm Biol* 2016; 54: 189–197
- [209] Falkenberg KD, Jakobs A, Matern JC, Dörner W, Uttarkar S, Trentmann A, Steinmann S, Coulibaly A, Schomburg C, Mootz HD, Schmidt TJ, Klempnauer KH. Withaferin A, a natural compound with anti-tumor activity, is a potent inhibitor of transcription factor C/EBP β . *Biochim Biophys Acta* 2017; 1864: 1349–1358
- [210] Mohan R, Hammers H, Bargagna-Mohan P, Zhan X, Herbstritt C, Ruiz A, Zhang L, Hanson A, Conner B, Rougas J, Pribluda V. Withaferin A is a potent inhibitor of angiogenesis. *Angiogenesis* 2004; 7: 115–122
- [211] Lavie D, Glotter E, Shvo Y. Constituents of *Withania somnifera* Dun. III. The side chain of Withaferin A¹. *J Org Chem* 1965; 30: 1774–1778
- [212] Bhat WW, Lattoo SK, Razda S, Dhar N, Rana S, Dhar RS, Khan S, Vishwakarma RA. Molecular cloning, bacterial expression and promoter analysis of squalene synthase from *Withania somnifera* (L.) Dunal. *Gene* 2012; 499: 25–36
- [213] Razdan S, Bhat WW, Rana S, Dhar N, Lattoo SK, Dhar RS, Vishwakarma RA. Molecular characterization and promoter analysis of squalene epoxidase gene from *Withania somnifera* (L.) Dunal. *Mol Biol Rep* 2013; 40: 905–916
- [214] Senthil K, Wasnik NG, Kim YJ, Yang DC. Generation and analysis of expressed sequence tags from leaf and root of *Withania somnifera* (Ashwagandha). *Mol Biol Rep* 2010; 37: 893–902
- [215] Dhar N, Rana S, Bhat WW, Razdan S, Pandith SA, Khan S, Dutt P, Dhar RS, Vaishnavi S, Vishwakarma R, Lattoo SK. Dynamics of withanolide biosynthesis in relation to temporal expression pattern of metabolic genes in *Withania somnifera* (L.) Dunal: a comparative study in two morpho-chemovariants. *Mol Biol Rep* 2013; 40: 7007–7016
- [216] Sabir F, Mishra S, Sangwan RS, Jadaun JS, Sangwan NS. Qualitative and quantitative variations in withanolides and expression of some pathway genes during different stages of morphogenesis in *Withania somnifera* Dunal. *Protoplasma* 2013; 250: 539–549
- [217] Chen LX, He H, Qiu F. Natural withanolides: an overview. *Nat Prod Rep* 2011; 28: 705–740
- [218] Rana S, Lattoo SK, Dhar N, Razdan S, Bhat WW, Dhar RS, Vishwakarma R. NADPH-cytochrome P450 reductase: molecular cloning and functional characterization of two paralogs from *Withania somnifera* (L.) Dunal. *PLoS One* 2013; 8: e57068
- [219] Mishra S, Bansal S, Mishra B, Sangwan RS, Jadaun AJS, Sangwan NS. RNAi and homologous over-expression based functional approaches reveal triterpenoid synthase gene-cycloartenol synthase is involved in downstream withanolide biosynthesis in *Withania somnifera*. *PLoS One* 2016; 11: e0149691
- [220] Agarwal AV, Gupta P, Singh D, Dhar YV, Chandra D, Trivedi PK. Comprehensive assessment of the genes involved in withanolide biosynthesis from *Withania somnifera*: chemotype-specific and elicitor-responsive expression. *Funct Integr Genomics* 2017; 17: 477–490
- [221] Agarwal AV, Singh D, Dhar YV, Michael R, Gupta P, Chandra D, Trivedi PK. Virus-induced silencing of key genes leads to differential impact on withanolide biosynthesis in the medicinal plant, *Withania somnifera*. *Plant Cell Physiol* 2018; 59: 262–274
- [222] Singh N, Bhalla M, de Jager P, Gilca M. An overview on ashwagandha: a Rasayana (rejuvenator) of Ayurveda. *Afr J Tradit Complement Altern Med* 2011; 8: 208–213
- [223] Ghosal S, Lal J, Srivastava R, Bhattacharya SK, Upadhyay SN, Jaiswal AK, Chattopadhyay U. Immunomodulatory and CNS effects of sitoindosides IX and X, two new glycowithanolides from *Withania somnifera*. *Phytother Res* 1989; 3: 201–206
- [224] Chengappa KN, Bowie C, Schlicht PJ, Fleet D, Brar JS, Jindal R. Randomized placebo-controlled adjunctive study of an extract of *Withania somnifera* for cognitive dysfunction in bipolar disorder. *J Clin Psychiatry* 2013; 74: 1076–1083
- [225] Choudhary D, Bhattacharyya S, Bose S. Efficacy and safety of ashwagandha (*Withania somnifera* (L.) Dunal) root extract in improving memory and cognitive functions. *J Diet Suppl* 2017; 14: 599–612
- [226] Dhuley JN. Effect of ashwagandha on lipid peroxidation in stress-induced animals. *Ethnopharmacol* 1998; 60: 173–178
- [227] Durg S, Dhadde SB, Vandal R, Shivakumar BS, Charan CS. *Withania somnifera* (Ashwagandha) in neurobehavioural disorders induced by brain oxidative stress in rodents: a systematic review and meta-analysis. *J Pharm Pharmacol* 2015; 67: 879–899
- [228] Sehgal N, Gupta A, Valli RK, Joshi SD, Mills JT, Hamel E, Khanna P, Jain SC, Thakur SS, Ravindranath V. *Withania somnifera* reverses Alzheimer's disease pathology by enhancing low-density lipoprotein receptor-related protein in liver. *Proc Natl Acad Sci U S A* 2012; 109: 3510–3515
- [229] Ahmad M, Saleem S, Ahmad AS, Ansari MA, Yousuf S, Hoda MN, Islam F. Neuroprotective effects of *Withania somnifera* on 6-hydroxydopamine induced Parkinsonism in rats. *Hum Exp Toxicol* 2005; 24: 137–147
- [230] Devi PU, Akagi K, Ostapenko V, Tanaka Y, Sugahara T. Withaferin A: A new radiosensitizer from the Indian medicinal plant *Withania somnifera*. *Int J Radiat Biol* 1996; 69: 193–197
- [231] Prakash J, Gupta SK, Dinda AK. *Withania somnifera* root extract prevents DMBA-induced squamous cell carcinoma of skin in Swiss albino mice. *Nutr Cancer* 2002; 42: 91–97
- [232] Stan SD, Hahm ER, Warin R, Singh SV. Withaferin A causes FOXO3a- and bim-dependent apoptosis and inhibits growth of human breast cancer cells *in vivo*. *Cancer Res* 2008; 68: 7661–7669
- [233] Ghosh K, De S, Mukherjee S, Das S, Ghosh AN, Sengupta S. Withaferin A induced impaired autophagy and unfolded protein response in hu-

- man breast cancer cell-lines MCF-7 and MDA-MB-231. *Toxicol In Vitro* 2017; 44: 330–338
- [234] Hsieh PW, Huang ZY, Chen JH, Chang FR, Wu CC, Yang YL, Chiang MY, Yen MH, Chen SL, Yen HF, Lübken T, Hung WC, Wu YC. Cytotoxic withanolides from *Tubocapsicum anomalum*. *J Nat Prod* 2007; 70: 747–753
- [235] Abdeljebbar LH, Benjouad A, Morjani H, Merghoub N, El Haddar S, Humam M, Christen P, Hostettmann K, Bekkouche K, Amzazi S. Antiproliferative effects of withanolides from *Withania adpressa*. *Therapie* 2009; 64: 121–127
- [236] Lee SW, Pan MH, Chen CM, Chen ZT. Withangulatin I, a new cytotoxic withanolide from *Physalis angulata*. *Chem Pharm Bull (Tokyo)* 2008; 56: 234–236
- [237] Biswal BM, Sulaiman AM, Ismail HC, Zakaria H, Jalil Abdul MI, Muhammad KI. AOS14 Phase II clinical study of combination chemotherapy with herb *Withania somnifera* (ashwagandha) in breast cancer. *Eur J Cancer* 2012; 48: S8–S9
- [238] Biswal BM, Sulaiman SA, Ismail HC, Zakaria H, Musa KI. Effect of *Withania somnifera* (ashwagandha) on the development of chemotherapy-induced fatigue and quality of life in breast cancer patients. *Integr Cancer Ther* 2013; 12: 312–322
- [239] Ho WE, Peh HY, Chan TK, Wong WSF. Artemisinins: pharmacological actions beyond anti-malarial. *Pharmacol Ther* 2014; 142: 126–139
- [240] Tu Y. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. *Nat Med* 2011; 17: 1217
- [241] White NJ. Qinghaosu (artemisinin): the price of success. *Science* 2008; 320: 330
- [242] Nosten F, White NJ. Artemisinin-based combination treatment of falciparum malaria. *Am J Trop Med Hyg* 2007; 77: 181–192
- [243] World Health Organization. Global Health Observatory. Available at <http://www.who.int/gho/malaria/en/>. Accessed May 23, 2018
- [244] Banek K, Lalani M, Staedke SG, Chandramohan D. Adherence to artemisinin-based combination therapy for the treatment of malaria: a systematic review of the evidence. *Malar J* 2014; 13: 7
- [245] Laloo DG, Shingadia D, Bell DJ, Beeching NJ, Whitty CJM, Chiodini PL. UK malaria treatment guidelines 2016. *J Infect* 2016; 72: 635–649
- [246] Straimer J, Gnädig NF, Witkowski B, Amaratunga C, Duru V, Ramadan AP, Dacheux M, Khim N, Zhang L, Lam S, Gregory PD, Urnov FD, Mercereau-Puijalon O, Benoit-Vical F, Fairhurst RM, Ménard D, Fidock DA. K13-propeller mutations confer artemisinin resistance in *Plasmodium falciparum* clinical isolates. *Science* 2015; 347: 428
- [247] Mbengue A, Bhattacharjee S, Pandharkar T, Liu H, Estiu G, Stahelin RV, Rizk SS, Njimoh DL, Ryan Y, Chotivanich K, Nguon C, Ghorbal M, Lopez-Rubio JJ, Pfender M, Emrich S, Mohandas N, Dondorp AM, Wiest O, Haldar K. A molecular mechanism of artemisinin resistance in *Plasmodium falciparum* malaria. *Nature* 2015; 520: 683
- [248] Duke MV, Paul RN, Elsohly HN, Sturtz G, Duke SO. Localization of artemisinin and artemisitene in foliar tissues of glanded and glandless biotypes of *Artemisia annua* L. *Int J Plant Sci* 1994; 155: 365–372
- [249] Akhila A, Thakur RS, Popli SP. Biosynthesis of artemisinin in *Artemisia annua*. *Phytochemistry* 1987; 26: 1927–1930
- [250] Nair MSR, Basile DV. Bioconversion of arteannuin B to artemisinin. *J Nat Prod* 1993; 56: 1559–1566
- [251] Sangwan RS, Agarwal K, Luthra R, Thakur RS, Singh-Sangwan N. Bio-transformation of arteannuin acid into arteannuin-B and artemisinin in *Artemisia annua*. *Phytochemistry* 1993; 34: 1301–1302
- [252] Ma DM, Wang Z, Wang L, Alejos-Gonzales F, Sun MA, Xie DY. A genome-wide scenario of terpene pathways in self-pollinated *Artemisia annua*. *Mol Plant* 2015; 8: 1580–1598
- [253] Bouwmeester HJ, Wallaart TE, Janssen MHA, van Loo B, Jansen BJM, Posthumus MA, Schmidt CO, De Kraker JW, König WA, Franssen MCR. Amorpho-4, 11-diene synthase catalyses the first probable step in artemisinin biosynthesis. *Phytochemistry* 1999; 52: 843–854
- [254] Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, Chang MCY, Withers ST, Shiba Y, Sarpong R, Keasling JD. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 2006; 440: 940
- [255] Teoh KH, Polichuk DR, Reed DW, Nowak G, Covello PS. *Artemisia annua* L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. *FEBS Lett* 2006; 580: 1411–1416
- [256] Teoh KH, Polichuk DR, Reed DW, Covello PS. Molecular cloning of an aldehyde dehydrogenase implicated in artemisinin biosynthesis in *Artemisia annua*. *Botany* 2009; 87: 635–642
- [257] Zhang Y, Teoh KH, Reed DW, Maes L, Goossens A, Olson DJH, Ross ARS, Covello PS. The molecular cloning of artemisinic aldehyde $\Delta 11(13)$ reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in *Artemisia annua*. *J Biol Chem* 2008; 283: 21501–21508
- [258] Simonnet X, Quennoz M, Carlen C. New *Artemisia annua* hybrids with high artemisinin content. *Acta Horticulturae* 2008; 769: 371–373
- [259] Nicolas D, Xavier S, Myriam G. The genetics of artemisinin content in *Artemisia annua* L. and the breeding of high yielding cultivars. *Curr Med Chem* 2001; 8: 1795–1801
- [260] Cockram J, Hill C, Burns C, Arroo RRJ, Woolley JG, Flockart I, Robinson T, Atkinson CJ, Davies MJ, Dungey N, Greenland AJ, Smith LLMJ, Bentley S. Screening a diverse collection of *Artemisia annua* germplasm accessions for the antimalarial compound, artemisinin. *Plant Genet Resour* 2012; 10: 152–154
- [261] Chen M, Yan T, Shen Q, Lu X, Pan Q, Huang Y, Tang Y, Fu X, Liu M, Jiang W, Lv Z, Shi P, Ma YN, Hao X, Zhang L, Li L, Tang K. Glandular trichome-specific WRKY 1 promotes artemisinin biosynthesis in *Artemisia annua*. *New Phytol* 2017; 214: 304–316
- [262] Tang K, Shen Q, Yan T, Fu X. Transgenic approach to increase artemisinin content in *Artemisia annua* L. *Plant Cell Rep* 2014; 33: 605–615
- [263] Ikram NKBK, Simonsen HT. A review of biotechnological artemisinin production in plants. *Front Plant Sci* 2017; 8: 1966
- [264] Zhang L, Jing F, Li F, Li M, Wang Y, Wang G, Sun X, Tang K. Development of transgenic *Artemisia annua* (Chinese wormwood) plants with an enhanced content of artemisinin, an effective anti-malarial drug, by hairpin-RNA-mediated gene silencing. *Biotechnol Appl Biochem* 2009; 52: 199–207
- [265] Banyai W, Kirdmanee C, Mii M, Supaibulwatana K. Overexpression of farnesyl pyrophosphate synthase (FPS) gene affected artemisinin content and growth of *Artemisia annua* L. *Plant Cell Tissue Organ Cult* 2010; 103: 255–265
- [266] Ma C, Wang H, Lu X, Wang H, Xu G, Liu B. Terpenoid metabolic profiling analysis of transgenic *Artemisia annua* L. by comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry. *Metabolomics* 2009; 5: 497–506
- [267] Chen Y, Shen Q, Wang Y, Wang T, Wu S, Zhang L, Lu X, Zhang F, Jiang W, Qiu B, Gao E, Sun X, Tang K. The stacked over-expression of FPS, CYP71AV1 and CPR genes leads to the increase of artemisinin level in *Artemisia annua* L. *Plant Biotechnol Rep* 2013; 7: 287–295
- [268] Lu X, Shen Q, Zhang L, Zhang F, Jiang W, Lv Z, Yan T, Fu X, Wang G, Tang K. Promotion of artemisinin biosynthesis in transgenic *Artemisia annua* by overexpressing ADS, CYP71AV1 and CPR genes. *Ind Crops Prod* 2013; 49: 380–385
- [269] Shen Q, Yan T, Fu X, Tang K. Transcriptional regulation of artemisinin biosynthesis in *Artemisia annua* L. *Sci Bull* 2016; 61: 18–25
- [270] Yu ZX, Li JX, Yang CQ, Hu WL, Wang LJ, Chen XY. The jasmonate-responsive AP2/ERF transcription factors AaERF1 and AaERF2 positively

- regulate artemisinin biosynthesis in *Artemisia annua* L. *Mol Plant* 2012; 5: 353–365
- [271] Lv Z, Wang S, Zhang F, Chen L, Hao X, Pan Q, Fu X, Li L, Sun X, Tang K. Overexpression of a novel NAC domain-containing transcription factor gene (AaNAC1) enhances the content of artemisinin and increases tolerance to drought and *Botrytis cinerea* in *Artemisia annua*. *Plant Cell Physiol* 2016; 57: 1961–1971
- [272] Martin VJJ, Pitera DJ, Withers ST, Newman JD, Keasling JD. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat Biotechnol* 2003; 21: 796
- [273] Anthony JR, Anthony LC, Nowroozi F, Kwon G, Newman JD, Keasling JD. Optimization of the mevalonate-based isoprenoid biosynthetic pathway in *Escherichia coli* for production of the anti-malarial drug precursor amorpha-4, 11-diene. *Metab Eng* 2009; 11: 13–19
- [274] Fuentes P, Zhou F, Erban A, Karcher D, Kopka J, Bock R. A new synthetic biology approach allows transfer of an entire metabolic pathway from a medicinal plant to a biomass crop. *Elife* 2016; 5: e13664
- [275] Paddon CJ, Westfall PJ, Pitera DJ, Benjamin K, Fisher K, McPhee D, Leavell MD, Tai A, Main A, Eng D, Polichuk DR, Teoh KH, Reed DW, Treynor T, Lenihan J, Jiang H, Fleck M, Bajad S, Dang G, Dengrove D, Diola D, Dorin G, Ellens KW, Fickes S, Galazzo J, Gaucher SP, Geistlinger T, Henry R, Hepp M, Horning T, Iqbal T, Kizer L, Lieu B, Melis D, Moss N, Regentin R, Secrest S, Tsuruta H, Vazquez R, Westblade LF, Xu L, Yu M, Zhang Y, Zhao L, Lievens J, Covello PS, Keasling JD, Reiling KK, Renninger NS, Newman JD. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 2013; 496: 528
- [276] Peplow M. Synthetic biology's first malaria drug meets market resistance. *Nature* 2016; 530: 389–390
- [277] Wani MC, Taylor HL, Wall ME, Coggon P, McPhail AT. Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *J Am Chem Soc* 1971; 93: 2325–2327
- [278] Wani MC, Horwitz SB. Nature as a remarkable chemist: a personal story of the discovery and development of taxol. *Anticancer Drugs* 2014; 25: 482
- [279] Ogden L. *Taxus* (yews) – a highly toxic plant. *Vet Hum Toxicol* 1988; 30: 563–564
- [280] McElroy C, Jennewein S. Taxol® biosynthesis and production: from forests to fermenters. In: Schwab W, Lange B, Wust M, eds. *Biotechnology of Natural Products*. Cham: Springer; 2018: 145–185
- [281] Weaver BA. How taxol/paclitaxel kills cancer cells. *Mol Biol Cell* 2014; 25: 2677–2681
- [282] Onrubia M, Cusido R, Ramirez K, Hernandez-Vazquez L, Moyano E, Bonfill M, Palazon J. Bioprocessing of plant *in vitro* systems for the mass production of pharmaceutically important metabolites: paclitaxel and its derivatives. *Curr Med Chem* 2013; 20: 880–891
- [283] Gersmann H, Aldred J. Medicinal tree used in chemotherapy drug faces extinction. Available at <https://www.theguardian.com/environment/2011/nov/10/iucn-red-list-tree-chemotherapy>. Accessed May 23, 2018
- [284] Onrubia M, Pollier J, Vanden Bossche R, Goethals M, Gevaert K, Moyano E, Vidal-Limon H, Cusido RM, Palazon J, Goossens A. Taximin, a conserved plant-specific peptide is involved in the modulation of plant-specialized metabolism. *Plant Biotechnol J* 2014; 12: 971–983
- [285] Ramirez-Estrada K, Altabella T, Onrubia M, Moyano E, Notredame C, Osuna L, Vanden Bossche R, Goossens A, Cusido RM, Palazon J. Transcript profiling of jasmonate-elicited *Taxus* cells reveals a β -phenylalanine-CoA ligase. *Plant Biotechnol J* 2016; 14: 85–96
- [286] The IUCN Red List of Threatened Species, *Taxus brevifolia*. Available at <http://dx.doi.org/10.2305/IUCN.UK.2013-1.RLTS.T34041A2841142.en>. Accessed May 23, 2018
- [287] Nicolaou KC, Yang Z, Liu JJ, Ueno H, Nantermet PG, Guy RK, Claiborne CF, Renaud J, Couladouros EA, Paulvannan K, Sorensen EJ. Total synthesis of taxol. *Nature* 1994; 367: 630
- [288] Yukimune Y, Tabata H, Higashi Y, Hara Y. Methyl jasmonate-induced overproduction of paclitaxel and baccatin III in *Taxus* cell suspension cultures. *Nat Biotechnol* 1996; 14: 1129–1132
- [289] Ajikumar PK, Xiao WH, Tyo KEJ, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G. Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. *Science* 2010; 330: 70–74
- [290] Rai A, Saito K, Yamazaki M. Integrated omics analysis of specialized metabolism in medicinal plants. *Plant J* 2017; 90: 764–787
- [291] Medicinal Plant Genomics Resource. Available at <http://medicinalplantgenomics.msu.edu/>. Accessed May 23, 2018
- [292] Zhao DY, Hamilton JP, Pham GM, Crisovan E, Wiegert-Rininger K, Vaillancourt B, DellaPenna D, Buell CR. *De novo* genome assembly of *Camptotheca acuminata*, a natural source of the anti-cancer compound camptothecin. *Gigascience* 2017; 6: 1–7
- [293] Hoopes GM, Hamilton JP, Kim J, Zhao D, Wiegert-Rininger K, Crisovan E, Buell CR. Genome assembly and annotation of the medicinal plant *Calotropis gigantea*, a producer of anti-cancer and anti-malarial cardenolides. *G3 (Bethesda)* 2018; 8: 385–391
- [294] Orcae. Online Resource for Community Annotation of Eukaryotes. Available at <http://bioinformatics.psb.ugent.be/orcae/>. Accessed May 23, 2018
- [295] Sterck L, Billiau K, Abeel T, Rouze P, van de Peer Y. ORCAE: online resource for community annotation of eukaryotes. *Nat Methods* 2012; 9: 1041
- [296] Transcriptome Characterization, Sequencing, and Assembly of Medicinal Plants Relevant to Human Health. Available at <https://apps.pharmacy.uic.edu/depts/pcrps/MedTranscriptomePlants/>. Accessed May 23, 2018
- [297] Lau W, Sattely ES. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science* 2015; 349: 1224–1228
- [298] Li FS, Weng JK. Demystifying traditional herbal medicine with modern approaches. *Nat Plants* 2017; 3: 17109
- [299] Huang W, Manglik A, Venkatakrishnan AJ, Laeremans T, Feinberg EN, Sanborn AL, Kato HE, Livingston KE, Thorsen TS, Kling RC, Granier S, Gmeiner P, Husbands SM, Traynor JR, Weis WI, Steyaert J, Dror RO, Kobilka BK. Structural insights into μ -opioid receptor activation. *Nature* 2015; 524: 315
- [300] Gigant B, Wang CG, Ravelli RBG, Roussi F, Steinmetz MO, Curmi PA, Sobel A, Knossow M. Structural basis for the regulation of tubulin by vinblastine. *Nature* 2005; 435: 519–522
- [301] Gaquerel E, Kuhl C, Neumann S. Computational annotation of plant metabolomics profiles via a novel network-assisted approach. *Metabolomics* 2013; 9: 904–918
- [302] Naake T, Gaquerel E. MetCirc: navigating mass spectral similarity in high-resolution MS/MS metabolomics data. *Bioinformatics* 2017; 33: 2419–2420
- [303] Morreel K, Saeys Y, Dima O, Lu FC, Van de Peer Y, Vanholme R, Ralph J, Vanholme B, Boerjan W. Systematic structural characterization of metabolites in arabidopsis via candidate substrate-product pair networks. *Plant Cell* 2014; 26: 929–945
- [304] Li DP, Baldwin IT, Gaquerel E. Navigating natural variation in herbivory-induced secondary metabolism in coyote tobacco populations using MS/MS structural analysis. *Proc Natl Acad Sci U S A* 2015; 112: E4147–E4155