Review Articles

Acupuncture and Herbal Medicine

OPEN

Genome sequencing provides potential strategies for drug discovery and synthesis

Chunsheng Zhao[1], Ziwei Zhang[1], Linlin Sun[1], Ronglu Bai[1], Lizhi Wang[1,*], Shilin Chen[2,*]

Abstract

Medicinal plants are renowned for their abundant production of secondary metabolites, which exhibit notable pharmacological activities and great potential for drug development. The biosynthesis of secondary metabolites is highly intricate and influenced by various intrinsic and extrinsic factors, resulting in substantial species diversity and content variation. Consequently, precise regulation of secondary metabolite synthesis is of utmost importance. In recent years, genome sequencing has emerged as a valuable tool for investigating the synthesis and regulation of secondary metabolites in medicinal plants, facilitated by the widespread use of high-throughput sequencing technologies. This review highlights the latest advancements in genome sequencing within this field and presents several strategies for studying secondary metabolites. Specifically, the article elucidates how genome sequencing can unravel the pathways for secondary metabolite synthesis in medicinal plants, offering insights into the functions and regulatory mechanisms of participating enzymes. Comparative analyses of plant genomes allow identification of shared pathways of metabolite synthesis among species, thereby providing novel avenues for obtaining cost-effective biosynthetic intermediates. By examining individual genomic variations, genes or gene clusters associated with the synthesis of specific compounds can be discovered, indicating potential targets and directions for drug development and the exploration of alternative compound sources. Moreover, the advent of gene-editing technology has enabled the precise modifications of medicinal plant genomes. Optimization of specific secondary metabolite synthesis pathways becomes thus feasible, enabling the precise editing of target genes to regulate secondary metabolite production within cells. These findings serve as valuable references and lessons for future drug development endeavors, conservation of rare resources, and the exploration of new resources.

Keywords: Biosynthetic pathways, Gene editing, Genome sequencing, Medicinal plants, Secondary metabolites

Graphical abstract: http://links.lww.com/AHM/A67

Introduction

Medicinal plants have played a crucial role in the advancement of human health and medicine throughout history. They represent a valuable source of natural medicines, harboring numerous biologically active compounds that have been included in applications in traditional medicine as well as modern drug development. Genome sequencing is a comprehensive technology that enables the analysis and sequencing of an organism's entire genome[1]. This technology can supply a wealth of information pertaining to an organism's genomic sequence, including the DNA nucleotide sequence and details regarding gene location, structure, and function[2–3]. A considerable number of plant species have been sequenced over the past two decades[4], providing a comprehensive overview of research progress in plant genome sequencing. Their analysis encompasses the completion of genome sequencing for 788 plant species and publication of 1,031 reference genomes. Notably, the number of sequenced species continues to grow, producing an ever-increasing list of sequenced plant genomes that are made publicly available. Medicinal plants are important sources of medicinal compounds and investigating the pharmacological activities and synthetic pathways of drug components in medicinal plants has become a prominent research area. The rapid development of second-generation sequencing technology[5], coupled with cost reductions, has facilitated the unveiling of an increasing number of genome sequences of medicinal plants, including *Panax ginseng* and *Catharanthus roseus*[6–7]. These sequences represent valuable resources that contribute to the comprehensive elucidation of medicinal plant genomes. Sequencing the complete genomes of medicinal plants yields valuable information about their genetic makeup, including size, structure, and gene count. This knowledge forms the foundation for in-depth investigations of gene function and metabolic pathways of medicinal plants. Genome sequencing enables the identification and annotation of genes associated with secondary metabolism, such as the tanshinone synthase genes involved in tanshinolate synthesis[8]. Secondary metabolites in medicinal plants

[1] School of Chinese Materia Medica, Tianjin University of Traditional Chinese Medicine, Tianjin, China; [2] School of Chinese Materia Medica, Chengdu University of Traditional Chinese Medicine, Chengdu, China

*Corresponding author: Shilin Chen, School of Chinese Materia Medica, Chengdu University of Traditional Chinese Medicine, Chengdu, China, E-mail: slchen@icmm.ac.cn; Lizhi Wang, School of Chinese Materia Medica, Tianjin University of Traditional Chinese Medicine, Tianjin, China, E-mail: lzhwang_2009@163.com.

Acupuncture and Herbal Medicine (2023) 3:4

Received 28 June 2023 / Accepted 22 August 2023

http://dx.doi.org/10.1097/HM9.0000000000000076

often exhibit important pharmacological activities and their synthesis usually occurs through complex metabolic pathways involving multiple enzyme-catalyzed reactions. Understanding the specific mechanisms of these metabolic pathways helps to understand the molecular basis of drug synthesis in plants and thus targets studies to increase yield. Genome sequencing also serves as a basis for studying gene function in medicinal plants. By comparing and annotating genome sequences, various genes related to enzymes, transporter proteins, and transcription factors can be predicted and identified[9–10]. Subsequent functional studies, including gene expression analyses and knockout experiments[11], have shed light on the roles of these genes in plant growth, development, and secondary metabolism. The synthesis of secondary metabolites can be regulated by increasing or decreasing the expression of specific enzymes through transgenic or gene-editing techniques. In addition, a deeper understanding of plant responses to internal and external environments can lead to the discovery of regulatory networks for drug synthesis in plants and identification of regulatory nodes that can be used to efficiently increase the yield of medicinal plants. Furthermore, genome sequencing plays a vital role in establishing a germplasm resource database for medicinal plants[12]. Many medicinal plants, such as *Taxus chinensis*, face a shortage of resources during wild collection or cultivation; establishing alternative plant resources can address this issue. Through research and screening of other plants, species with similar pharmacological activities can be identified. These plants can grow and produce the desired medicinal compounds in a relatively short period, thus reducing dependence on the original medicinal plants. In addition, the search for alternative plant resources can enrich the sources of medicines and provide more choices, thereby facilitating the conservation and sustainable utilization of these valuable botanical resources. These investigations provide a solid scientific foundation for unraveling the intricate mechanisms underlying the synthesis of medicinal compounds in plants, advancing drug development endeavors and enhancing the quality and yield of medicinal plants.

Genome sequencing illuminates biosynthesis pathways of secondary metabolites in medicinal plants

Unraveling secondary metabolic pathways in medicinal plants

The present gene sequencing technology has undergone three notable advancements, profoundly enhancing the investigation of molecular breeding and biosynthesis of medicinal plants. Ongoing research has primarily concentrated on the analysis of bioactive constituents found in medicinal plants, encompassing the study of biosynthetic pathways[13]. The development of novel gene mining techniques has also facilitated the endeavors of biologists engaged in exploring medicinal plants, enabling the discovery of new genes and subsequent identification of secondary metabolites. By examining the genetic homology between *T. chinensis* and other species, a distinctive gene associated with the synthesis of paclitaxel was detected. This gene catalyzes the conversion of diterpenoid precursors to paclitaxel, indicating a specialized evolutionary pathway for paclitaxel production. While paclitaxel is predominantly sourced from *T. chinensis*, the scarcity of species resources and subsequent low yield fail to meet market demand[14]. Whole-genome duplication facilitates the rapid evolution of angiosperms, whereby the duplication of genetic material in medicinal plants results in new gene functions, representations, or metabolic pathways[15]. In addition to *T. chinensis*, the sequencing of *P. ginseng*, *Salvia miltiorrhiza*, *Ganoderma lucidum*, and *C. roseus* genomes is nearing completion. Consequently, scientists can acquire genomic information on these plants, enabling further investigation of their genetic characteristics, gene functions, and metabolic pathways. As shown in Figure 1 this endeavor holds significant implications for uncovering the pharmacological properties of medicinal plants, developing novel drugs, and increasing yield. These studies are anticipated to propel the advancement, utilization, and conservation of medicinal plants, thereby satisfying market and medical demands more effectively.

Increasing the yield of important compounds from medicinal plant secondary metabolites by heterologous expression

Contemporary genomics research has predominantly focused on microorganisms, animals, and crops, while investigations concerning medicinal plants have received comparatively less attention. Consequently, there remains a dearth of comprehensive research avenues for uncovering high-quality secondary metabolites and corresponding synthesis genes in medicinal plants. Conversely, heterologous expression systems offer controlled and reproducible conditions for analyzing the activities of proteins or enzymes[16]. Within life sciences, heterologous protein expression technologies have progressively permeated various domains, owing to advancements in genomic endeavors. Examination of gene functionalities and their interactions centers on employing protein expression systems to express the target gene, adapting the expression profile to specific requirements. Accordingly, the modification of target gene expression is facilitated. *Escherichia coli* serves as the principal heterologous host for recombinant protein expression. Nonetheless, variations in *E. coli* expression have been documented; thus, the choice of different *E. coli* strains can yield divergent expression outcomes and subsequently distinct biosynthesis profiles[17]. The discovery of the mechanism of ginsenoside synthesis, informed by modern bioinformatics analysis and heterologous expression, imparts novel perspectives for the production of valuable saponins. Furthermore, exploration of the genetic repertoire has enabled the cloning and expression of two novel diterpene synthases that hold potential for the synthetic generation of diterpenoids in *Salvia sclarea*[18]. Concomitantly, genomic investigations have led to the identification of a previously undiscovered lectin, Bfl-II, in *Zingiber officinale,* which exhibits distinct dissimilarities from known lectin gene sequences. Intriguingly, heterologous expression of this gene intensified the inhibitory effect of the lectin on cancer cell proliferation, thereby demonstrating noteworthy anticancer activity[19].

Genome sequencing to obtain the same intermediate products and synthesize downstream products

Application of genome sequencing in terpenoid synthesis

Genome sequencing entails a comprehensive analysis of an organism's DNA sequence, enabling the determination of both gene and non-coding DNA sequences[20]. This technique offers valuable insights into an organism's genetic information, encompassing aspects such as gene number, composition, arrangement, gene–gene interactions, and gene–phenotype relationships. Terpenoids, a ubiquitous class of plant biochemicals, play crucial physiological and ecological roles[21]. Genome sequencing is a powerful tool for identifying key enzymes and genes involved in pathways of terpenoid synthesis across different plant species. The widespread use of genome sequencing technology allows examination of terpenoid synthesis in diverse plant species. Genome sequencing, coupled with bioinformatics analysis, has revealed striking genomic-level similarities in terpene synthesis pathways among various plants, highlighting shared genes and pathways. This observation suggests a high degree of conservation and co-evolution in the biosynthesis of these compounds. Researchers have harnessed genome sequencing techniques to identify and analyze genes and pathways associated with terpene synthesis in different plants, while comparing their genomic-level similarities. Furthermore, the identification of intermediate compounds in terpene synthesis, such as germacrene A, a precursor to substances, especially β-elemene, and artemisinin, has been achieved[22]. The advancement of genome sequencing technology offers a refined and comprehensive approach to analyzing the synthesis of these intermediates, thereby unraveling the molecular mechanisms underlying terpene synthesis pathways for plant secondary metabolites.

Obtaining the same intermediates and their downstream pathways from different plants

Genome sequencing techniques have been employed to identify and analyze genes and pathways associated with biosynthesis in medicinal plants, enabling comparisons of genomic-level similarities across different species. Notably, several common intermediates have been discovered through whole-genome identification and expression analysis in plants, including *Strobilanthes cusia*, *Rauvolfia verticillate*, *Camptotheca acuminata*, *Camellia sinensis*, *Brassica napus*, and *Dendrobium candidum*[23–29]. Among these intermediates, strictosidine holds particular significance as a pivotal compound in the metabolism of terpenoid indole alkaloids. As shown in Figure 2, strictosidine serves as both a coupling product of the iridoid and indole synthesis pathways and as a precursor compound for monoterpene indole alkaloids; these include camptothecin, quinine, vindoline, and vinblastine, which have important roles, including against tumors, and in the treatment of malaria, control of blood sugar, and lowering of blood pressure These findings underscore the crucial role of strictosidine in the metabolism of monoterpene indole alkaloids.

Li[30] used single-cell RNA sequencing (scRNA-seq) to acquire leaf-specific gene expression profiles of *C. roseus* at the cellular level. Through descending and clustering analyses, distinct cell populations within *C. roseus* leaves were grouped and effectively identified. Notably, this study elucidated the spatial and temporal expression patterns of genes involved in the indole alkaloid synthesis pathway, offering unprecedented insights into *C. roseus* leaf biology. Moreover, the investigation successfully resolved several pathways of vinblastine biosynthesis, enhancing our understanding of the intricate metabolic processes in this plant.

The iridoid synthesis pathway involves a series of sequential reactions, as illustrated in Figure 3. Initially, geranyl diphosphate (GPP) undergoes hydrolysis by geraniol synthase (GES), resulting in the formation of geraniol. Subsequently, geraniol is oxidized to 8-oxogeranial and undergoes a series of methylation reactions, leading to the formation of loganin. Finally, secologanin synthase (SLS) catalyzes the ring-opening of loganin, ultimately yielding secologanin[31–37]. In contrast, indole synthesis involves a distinct set of enzymatic steps. Initially, anthranilate is synthesized from chorismate through enzymatic catalysis. It is followed by the multi-step enzymatic synthesis of indole. Additionally, indole and serine are subjected to enzymatic catalysis, resulting in the formation of L-tryptophan, which undergoes decarboxylation to produce tryptamine[38]. Finally, strictosidine synthase couples secologanin and tryptamine, to produce strictosidine[39–40]. The biosynthesis of various terpenoid indole alkaloids relies on strictosidine as the precursor molecule. As shown in Figure 3 for camptothecin, strictosidine or strictosidinic acid can undergo enzymatic or non-enzymatic reactions to yield strictosamide. Subsequently, strictosamide serves as a substrate and is recognized by cytochrome monooxygenase P450, resulting in a cascade of oxidation, ring-opening, and rearrangement reactions, leading to the formation of pumiloside, which serves as the foundational skeleton for camptothecin and its derivatives. Pumiloside is further subjected to a multi-step reduction process that converts the carbonyl group to a double bond. This intermediate then undergoes glycosylation to generate a glycosylated camptothecin precursor, which is ultimately hydrolyzed and oxidized by glycosidase to yield camptothecin[41–42]. Regarding amarin, strictosidine is enzymatically hydrolyzed to produce strictosidine aglycone. Strictosidine aglycone is an active but unstable intermediate that spontaneously reacts and undergoes demethylation under the influence of various enzymes, leading to the formation of the final product, amarin[43–48]. For vendoline and catharanthine synthesis, strictosidine aglycone is enzymatically reduced to geissoschizine. Subsequent reactions result in the formation of dihydehydroprecondylocarpine acetate, an unstable intermediate. Finally, through cycloaddition reactions catalyzed by either catharanthine synthase (CS) or tabersonine synthase (TS), this intermediate is transformed into catharanthine and tabersonine[49–52], respectively. Tabersonine then undergoes catalysis by several enzymes to yield vindoline[53–55]. In conclusion, genome sequencing has provided insights into alternative plant sources that are cost-effective and readily

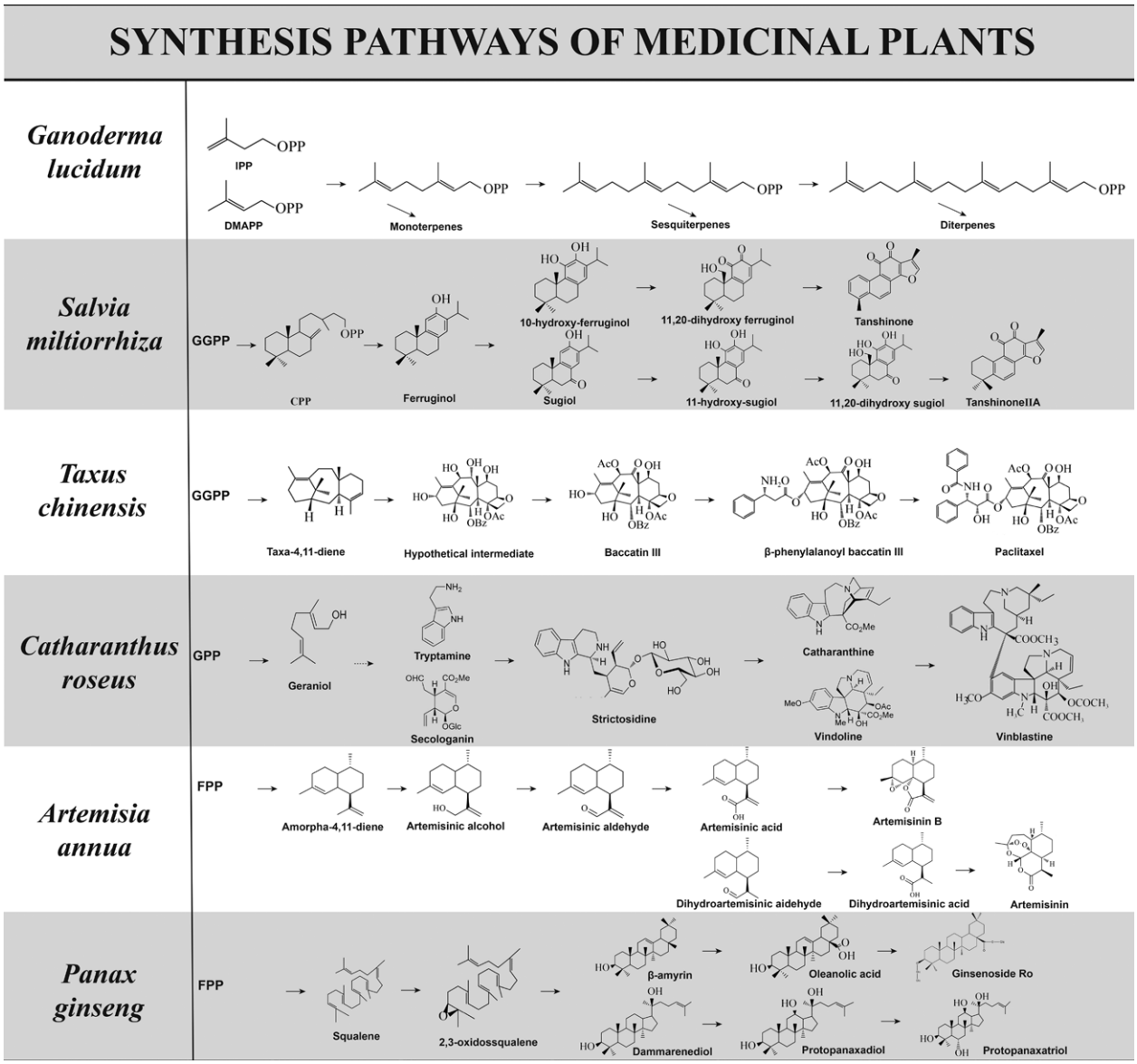

Figure 1. The figure shows six medicinal plants that have been sequenced, namely *Ganoderma lucidum*, *Salvia miltiorrhiza*, *Taxus chinensis*, *Catharanthus roseus*, *Artemisia annua*, and *Panax ginseng*, and outlines the biosynthetic pathways of their major secondary metabolites. Arrows do not represent direct synthesis. CPP: copalyl diphosphate; DMAPP: Dimethylallyl diphosphate; FPP: Farnesyl diphosphate; GGPP: Geranylgeranyl diphosphate; GPP: Geranyl diphosphate; IPP: Isopentenyl diphosphate; OPP: Diphosphate.

available for large-scale production, thereby reducing extraction costs. The biosynthesis of secondary metabolites in plants is typically a complex process involving multiple enzyme-catalyzed reactions, often requiring the coordination of various genes. Therefore, application of genome-wide approaches based on sequencing technologies is crucial for in-depth exploration of the mechanisms of secondary metabolite biosynthesis in diverse plant species, thereby contributing to future scientific research and production endeavors.

Discovery of novel compound resources through genome sequencing

Exploration of compound resources from plant endophytic fungi cc

Secondary plant metabolites exhibit diverse pharmacological activities and play crucial biological roles. However, their natural production levels are often insufficient to meet increasing pharmaceutical demand, necessitating the exploration of novel sources. Endophytic fungi represent vital components of plant microecosystems. Recent investigations have highlighted four major categories of endophytic fungal secondary metabolites: polyketides, nonribosomal peptides, alkaloids, and terpenes[56]. By establishing long-standing mutually beneficial symbiotic relationships with plants, endophytic fungi profoundly influence plant metabolic processes and physiological activities, thus serving as a significant resource to overcome the scarcity of secondary metabolites[57].

Artemisinin, a highly potent antimalarial drug, is a sesquiterpene lactone compound obtained from *Artemisia annua*[58]. However, the extraction of artemisinin from natural plant sources suffers from challenges such as low yield, high cost, and limited efficiency, which restrict its availability for therapeutic use. Consequently, exploring alternative approaches to obtain artemisinin in a more accessible manner becomes pressing. Notably, microbial

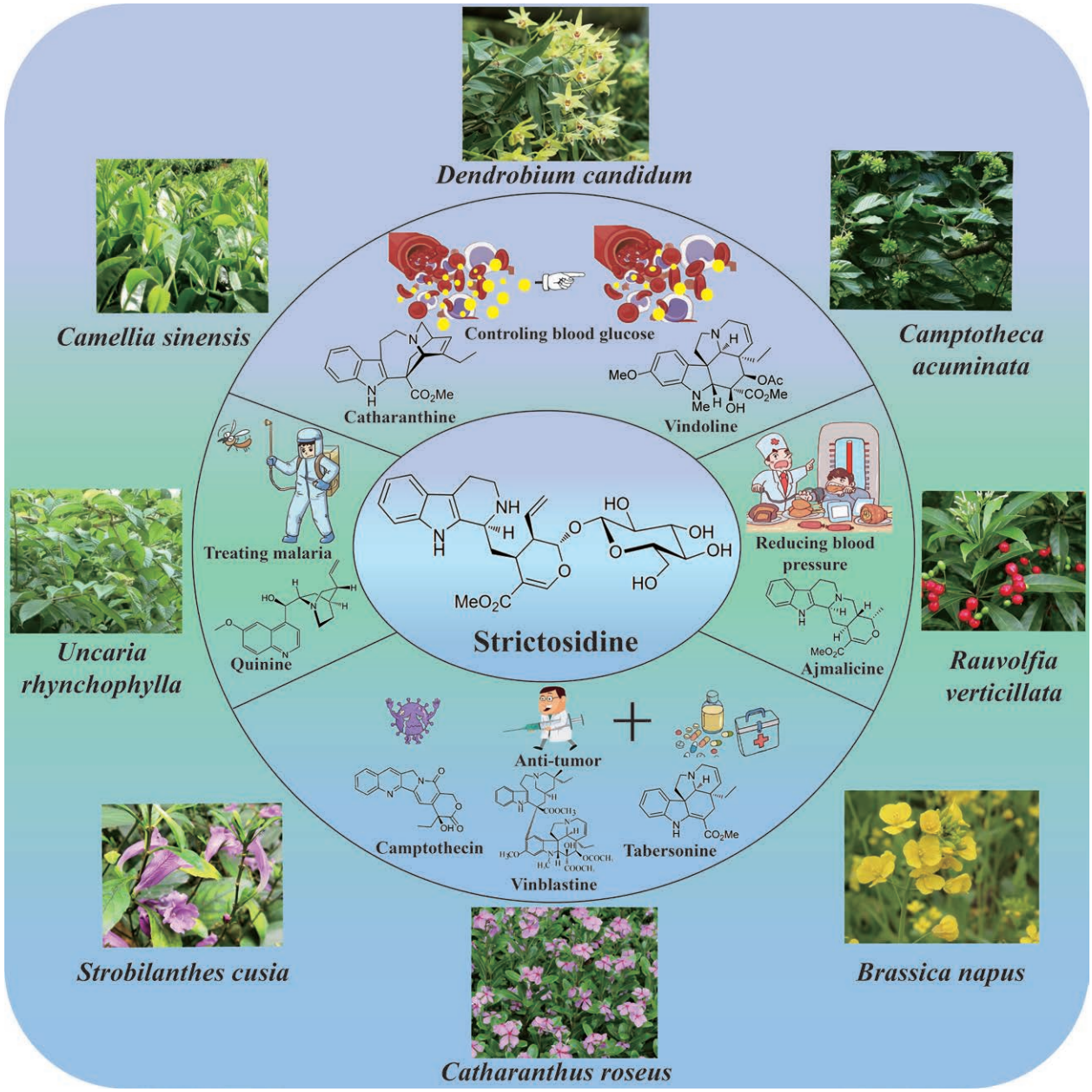

Figure 2. Chemical structure of strictosidine and compound sources in several medicinal plants—*Strobilanthes cusia*, *Rauvolfia verticillata*, *Camptotheca acuminata*, *Camellia sinensis*, *Catharanthus roseus*, *Dendrobium candidum*, *Uncaria rhynchophylla*, and *Brassica napus*. Strictosidine is the intermediate for the synthesis of several final products. Their pharmacological effects are also illustrated.

synthesis of artemisinin represents a promising strategy for its production. As shown in Figure 4, artemisinic acid, which, in *A. annua*, is present in markedly higher quantities than artemisinin, serves as a precursor for artemisinin synthesis. Tian[59] isolated the endophytic fungal strain B4 from wild *A. annua* and identified it as *Penicillium oxalicum* B4 through comparative analysis. The symbiotic relationship between *P. oxalicum* B4 and its host was confirmed through root infestation and colonization experiments in *A. annua* seedlings. After 23 days of incubation in an Murashige and Skoog (MS) culture system, B4-infected tissue culture seedlings exhibited a remarkable 42.86% increase in artemisinin content compared with the control group. This finding demonstrates the efficacy of this method in enhancing artemisinin yield and supports the widespread exploration of artemisinin production.

Exploration of alternative plant sources for novel compound resources

The vast array of pharmacologically active secondary metabolites found in plants has diverse applications. However, the limited availability and complex extraction processes of active ingredients in certain medicinal plants pose challenges to their large-scale production, necessitating the exploration of alternative resources.

One such secondary metabolite is paclitaxel, derived from the bark of *T. chinensis*. Notably, paclitaxel exhibits potent antitumor effects[60] and has been extensively employed in clinical settings for treating various cancers, including breast, ovarian, and lung cancers[61–63]. In addition, recent studies have highlighted the potential of paclitaxel in managing chronic inflammatory diseases[64]. Unfortunately, *T. chinensis*, the main source of paclitaxel, is in danger of depletion, and China has

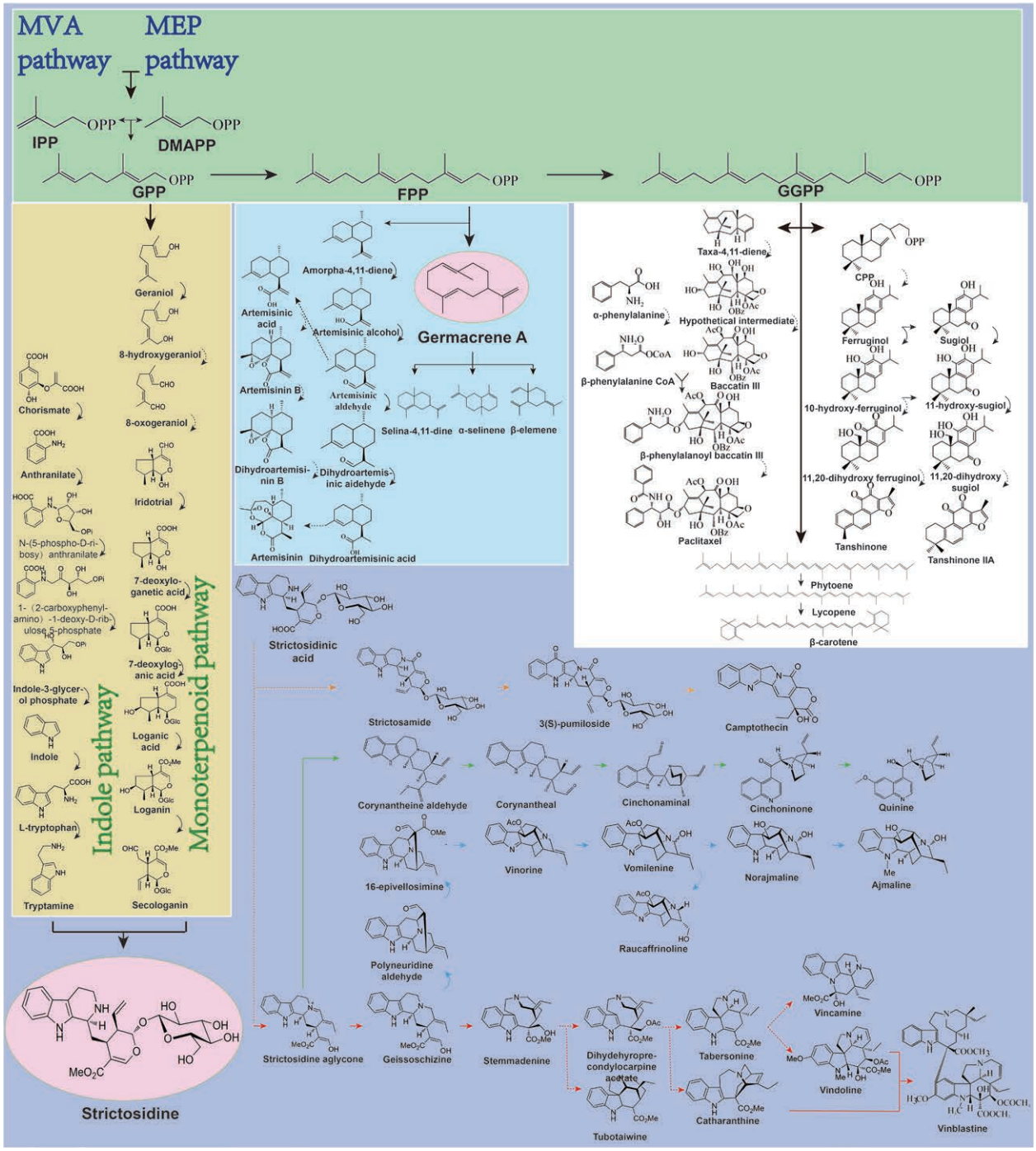

Figure 3. The diagram shows the pathways for the synthesis of strictosidine (yellow rectangle), camptothecin (orange arrows), quinine (green arrows), ajmaline (blue arrows), and vinblastine and vincamine (red arrows) using strictosidine as the precursor. The pathways for the synthesis of artemisinin (blue rectangle), β-elemene (blue rectangle), paclitaxel (white rectangle), lycopene (white rectangle), tanshinone (white rectangle) are also included. Dashed lines indicate multi-step reactions. DMAPP: Dimethylallyl diphosphate; FPP: Farnesyl diphosphate; GGPP: Geranylgeranyl diphosphate; GPP: Geranyl diphosphate; IPP: Isopentenyl diphosphate; MEP: 2-C-Methyl-D-erythritol 4-phosphate; MVA: Mevalonate; OPP: Diphosphate.

classified it as a national-level–protected wild plant to protect resources. The goal of scientists in related fields is to identify alternatives or develop new production strategies to reduce dependence on *T. chinensis* resources. Such studies play a crucial role in conserving *T. chinensis* resources, meeting the escalating demand for paclitaxel, and exploring the untapped potential of other plant resources. To search for alternative resources, the first step is to conduct genome sequencing of *T. chinensis* and *Corylus avellana* to obtain DNA sequence data. These will then be compared with the genomic data of known species to identify similar DNA sequence segments. The identified DNA sequence segments will undergo gene prediction and functional annotation to determine the genes they represent and their functions. The genomes of *T. chinensis* and *C. avellana* will be compared to identify shared and unique genes and to understand their genetic similarities and differences. Functional analyses will be conducted to understand the biological roles of shared genes or genes with similar functions. If certain functions are found to play important roles in both species, *C. avellana* can potentially serve as an alternative

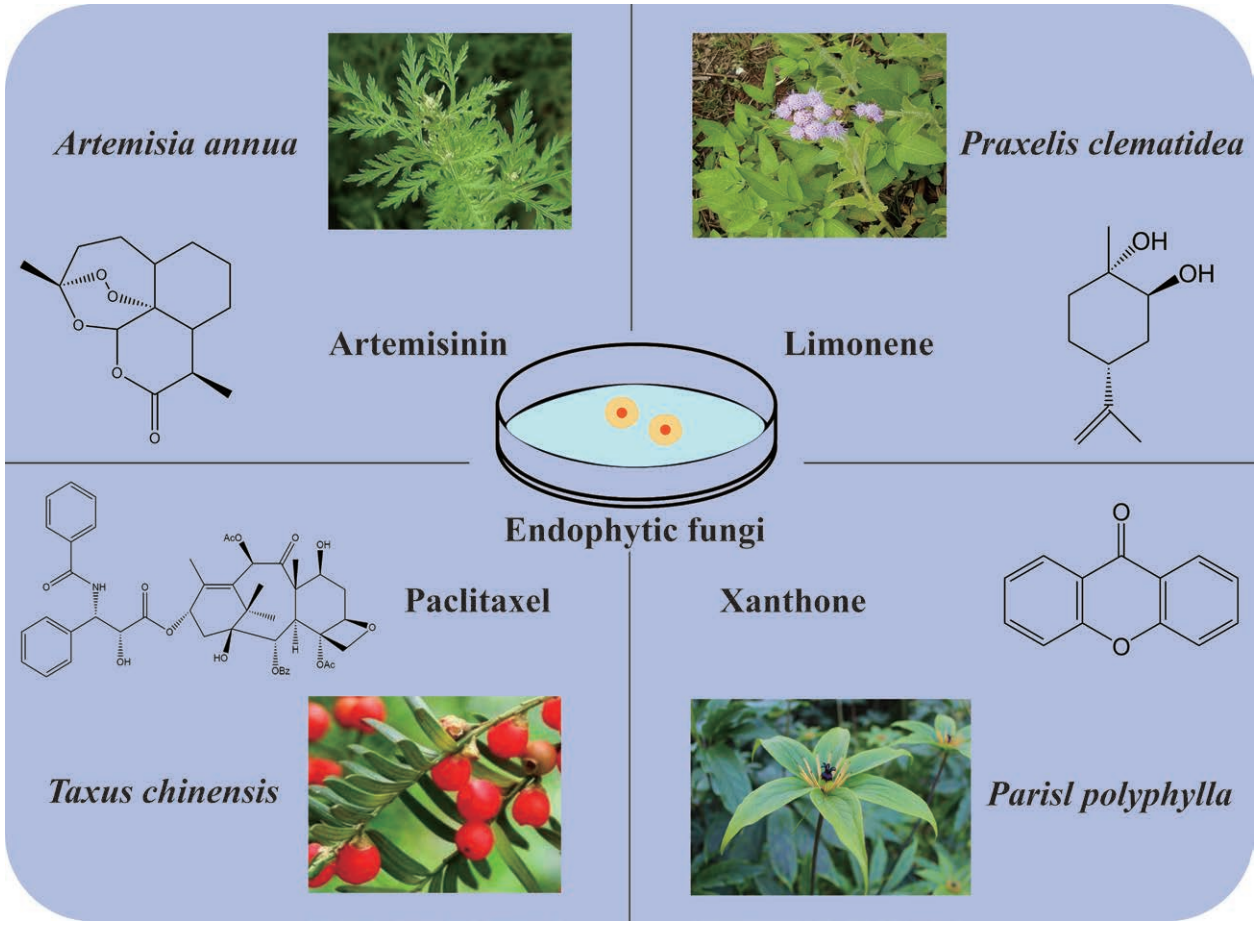

Figure 4. The figure shows that artemisinin, paclitaxel, xanthone, limonene resources can be obtained from endophytic fungi.

resource for *T. chinensis*. Paclitaxel was initially detected in *C. avellana* extracts through the application of High Pressure/Performance Liquid Chromatography-Mass Spectrometer (HPLC-MS) techniques in 2000 by a research team led by Professor Angela Hoffman[65]. The study focused on hazelnut varieties that exhibited resistance to Eastern blight in the United States. Hoffman and her team proposed that the paclitaxel found in *C. avellana* extracts originated from symbiotic endophytic fungi. Building upon Hoffman's experimental conditions and the paclitaxel spectra observed in hazelnuts, as shown in Figure 4, Zhang et al.[66] conducted a study on *Corylus heterophylla* in the Anshan area, confirming the presence of paclitaxel. Compared with Taxus plants, hazelnuts offer abundant, productive, and fast-growing characteristics, regarded as a dependable alternative resource. To ensure its pharmacological activity, first, the compound must be purified and structurally identified to confirm that the obtained biological activity is related to the target compound. Appropriate rat models were used to simulate human diseases, and the test compound was administered to the experimental animals to observe its effects on the disease model, including inflammation levels, lesion sizes, and behavioral changes. To obtain accurate results, appropriate control groups should be used for comparison and the experiments should be repeated. Of note, the results of the pharmacological activity validation are only preliminary evaluations, and further research and validation are crucial to ensuring its medicinal potential.

Genome sequencing and gene editing for precise intracellular regulation

Regulation of intracellular processes through the activation or inhibition of key gene expression

The regulation of protein expression in cells operates at the gene level and activation or inhibition of crucial genes plays a considerable role in determining the final product content. Certain secondary metabolites, including monoterpenes, sesquiterpenes, and phenolic compounds, possess noteworthy physiological functions, such as antibacterial, antiviral, antioxidant, analgesic, and anti-inflammatory properties[67–69]. The yield of these metabolites may be influenced by the regulation of key enzymes. For instance, *A. annua*, renowned for its abundance of artemisinin, demonstrated an increase in artemisinin content through key enzyme induction, leading to heightened expression of the cytochrome P450 monooxygenase (*CYP71AV1*) gene—an essential enzyme involved in artemisinin biosynthesis[70]. Different radiation spectra affect metabolite morphology, gene expression, quality, and quantity in different ways. Blue light, in particular, has a positive impact on secondary metabolism and morphology in various anatomical aspects, especially regarding anatomical features related to leaf petiole area, trichome frequency, epidermal and mesophyll differentiation, and thickness. When plants are cultivated under blue light, the expression of the *ADS* gene markedly increases, leading to higher artemisinin levels. Both red and blue light enhance artemisinin content and

the expression of the *ADS* and *CYP71AV1* genes, which are key enzymes involved in artemisinin biosynthesis. The increased expression of these key enzymes leads to an increase in downstream artemisinin production.

Andrographis paniculata is a medicinal plant traditionally used as an anti-inflammation and antibacterial herb. Andrographolide, the major active component of *A. paniculata*, exhibits diverse pharmacological activities, including anti-inflammation, anticancer, antiobesity, antidiabetes, and other activities[71].

WRKY transcription factors can specifically bind to elements in the promoter region of target genes. One study analyzed the promoter sequences of key enzyme genes involved in the andrographolide biosynthesis pathway, and five binding sites were identified, with WRKY1OS having all the promoter sequences of key enzyme genes. Different genes have different binding sites in their promoters. The results suggest that WRKYs directly bind to promoter elements to regulate the expression of key enzyme genes, thus influencing the biosynthesis of andrographolide[72].

Integration of genome sequencing and gene editing for precise intracellular regulation

With the advancement of the CRISPR-Cas system, third-generation gene-editing technology has emerged as the most extensively employed tool due to its cost-effectiveness and simplicity. Initially discovered in prokaryotes, the CRISPR-Cas system serves as a prokaryotic adaptive immune system wherein RNA molecules containing spacer sequences assist Cas proteins in recognizing, binding, and cleaving exogenous nucleic acids. Among the various types, the type II CRISPR/Cas9 system is most commonly used[73]. Single-guide RNAs (sgRNAs), formed by the fusion of trans-activated Cas9 nucleic acid endonuclease and Cas9 nucleic acid endonuclease complexes, are the prevalent forms of chimeric RNAs employed. The CRISPR-Cas9 system plays a pivotal role in gene editing by enabling the design of specific sgRNAs that direct Cas9 proteins to precise locations within the target genome, facilitating gene knockdown, insertion, or modification. The simplicity and efficacy of this technology render it the preferred choice for a wide array of applications in biological research, gene therapy, and other fields.

P. ginseng is rich in pentacyclic triterpenoid ginsenosides, and CYP716A47 and CYP716A53v2 are the key enzymes in their biosynthesis. Ginsenoside content can be selectively increased by the precise regulation of these two key enzymes. Genome sequencing coupled with gene editing is involved in the intracellular regulation of the medicinal plant *P. ginseng*. *P. ginseng* (Renshen) roots contain pharmacologically active ginsenosides, which are classified as protopanaxadiol (PPD)- or protopanaxatriol (PPT)-type saponins according to their glycosidic element structure. Regarding the biosynthetic pathway of ginsenosides, as shown in Figure 5, FPP is the direct precursor, and squalene synthase (ERG9) catalyzes the formation of squalene from two FPP molecules; squalene epoxidase (ERG1) catalyzes squalene to produce 2,3-oxidosqualene, which, in turn, is catalyzed by β-amyrin synthase (β-AS) and dammarenediol-II synthase (DS) to form the pentacyclic triterpene β-amyrin and the tetracyclic triterpene dammarenediol-II (DM), respectively. Oleanolic acid is formed by oleanolic acid synthase (OAS), followed by formation of oleanolic glycosides by glycosyltransferase; DM is catalyzed by protopanaxadiol synthase (PPDS) and protopanaxatriol synthase (PPTS) to form PPD and PPT in turn. Subsequently, UDP-glycosyltransferase (UGT) catalyzes the formation of dammarane-type ginsenosides. PPD ginsenosides include Rbl, Rb2, Rc, Rd, F2, Rg3, and Rh2, and PPT ginsenosides include Re, Rgl, Rg2, Rf, and Rhl[74]. The initial sequencing of the *P. ginseng* genome was followed by a thorough analysis of the genomic data to identify crucial genes involved in the biosynthesis of ginsenosides in *P. ginseng*, specifically PPD and PPT. This analysis involved comparisons of databases that contained known pathways for ginsenoside synthesis. By employing the CRISPR-Cas9 system, targeted mutagenesis of the PPT synthase gene in *P. ginseng* has been successfully performed. To accomplish this, Choi et al.[75] developed specific sgRNAs for targeted mutagenesis. Our findings revealed a marked reduction in PPT-type ginsenosides in the roots, accompanied by an increased accumulation of PPD-type ginsenosides. It suggests the presence of competing pathways for the production of PPD- and PPT-type ginsenosides. Notably, a see-saw effect based on the content of the intermediate product dammarendiol-II was observed, although, when the overall impact is considered, the total production of ginsenosides may not have been markedly altered. The CRISPR/Cas9 system demonstrated remarkable efficiency in entirely eliminating PPT-type ginsenosides from *P. ginseng* through the complete knockout of the target gene.

Tanshinones, lipophilic diterpenes isolated from the rhizome of *S. miltiorrhiza*, have diverse pharmacological activities against human ailments, including neurological diseases[76].

The rosmarinic acid synthase gene, SmRAS, which is involved in phenolic acid biosynthesis, and the diterpene synthase gene, SmCPS1, which is involved in tanshinone biosynthesis, can be precisely and effectively knocked out in *S. miltiorrhiza* using CRISPR/Cas9. In this study, the sgRNA of SmMYB98 was inserted into the modified pCAMBIA1300 vector to obtain the CRISPR-Cas9-KO vector. The recombinant vector pCAMBIA1300-CRISPR/Cas9-SmMYB98sgRNA was then introduced, using the pCAMBIA1300 empty vector as a control. HPLC was used to detect tanshinones in the hairy roots of SmMYB98-OE and SmMYB98-KO plants. The results showed that SmMYB98 positively regulates tanshinone biosynthesis in the hairy roots of SmMYB98 lines[77].

Summary and outlook

Research on genome sequencing in medicinal plants has made considerable advancements, yielding crucial insights into the genomic characteristics, secondary metabolic pathways, and regulatory mechanisms[78–80]. Through the analysis of whole-genome expression characteristics, a very important intermediate in plants such as *C. acuminata* and *C. roseus*, known as strictosidine, was discovered. Strictosidine serves as a precursor compound for terpenoid indole alkaloids such as camptothecin and vinblastine. When comparing the genetic homology of

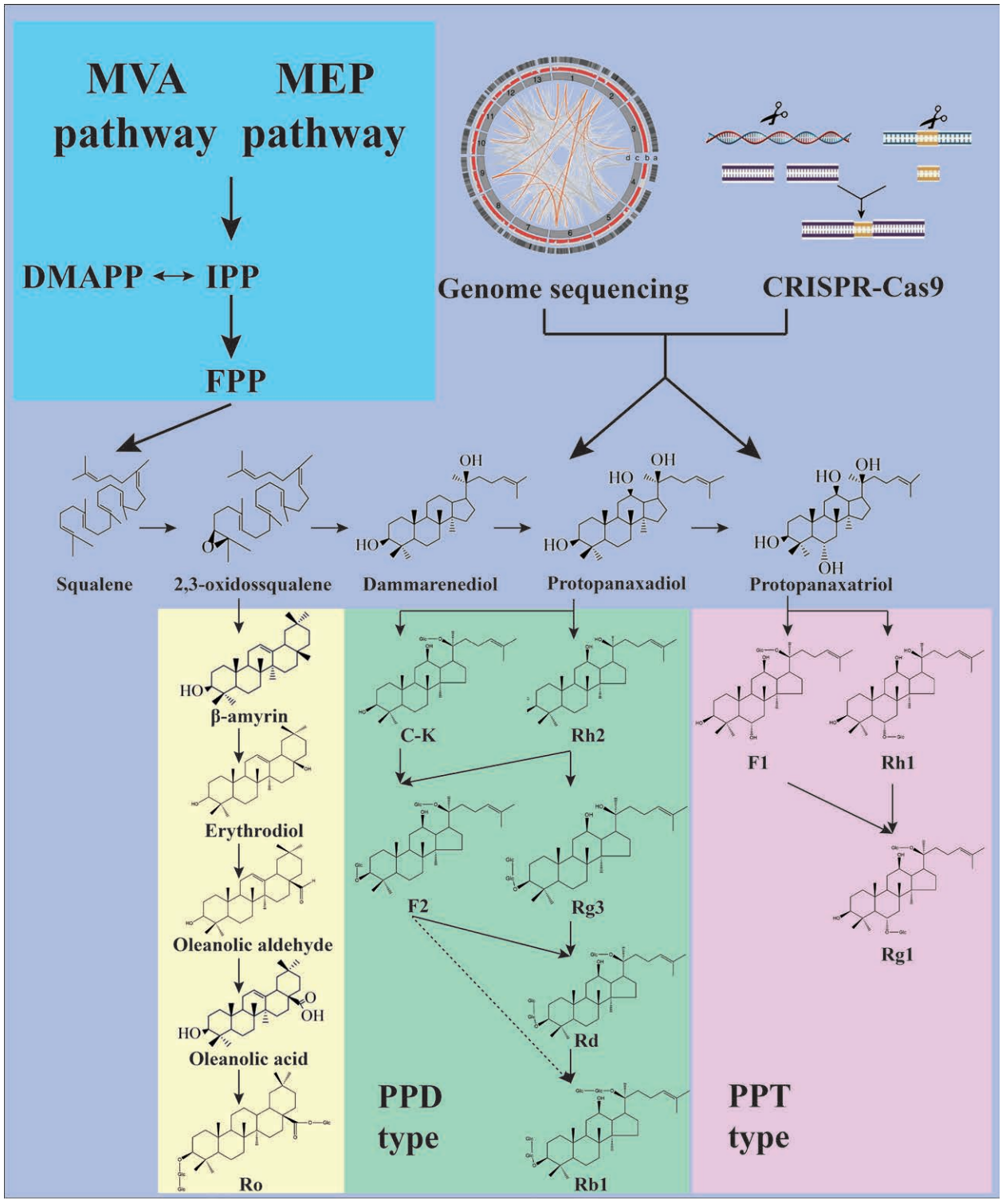

Figure 5. Ginsenoside R_o and PPD- and PPT-type ginsenoside synthesis pathways. DMAPP: Dimethylallyl diphosphate; FPP: Farnesyl diphosphate; IPP: Isopentenyl diphosphate; MEP: 2-C-Methyl-D-erythritol 4-phosphate; MVA: Mevalonate; PPD: Protopanaxadiol; PPT: Protriol.

T. chinensis with other species, unique genes related to paclitaxel biosynthesis were found in the former. These genes catalyze the production of geranylgeranyl diphosphate, which specialized during evolution to form the pathway for paclitaxel synthesis. Professor Angela Hoffman's research team first discovered the presence of paclitaxel in the extract of the hazel plant. By regulating the cytochrome P450 monooxygenase in *A. annua*, the expression of cytochrome P450 monooxygenase genes and artemisinin content were increased. In the past few decades, genome sequencing has successfully deciphered the genome sequences of numerous medicinal plants, including *S. miltiorrhiza*, *C. roseus*, *Coptis chinensis*, *P. ginseng*, and *Papaver*[81–83]. The disclosure and sharing of these genomic data have provided valuable resources and tools for comprehensive investigations of medicinal plant genomes. However, several challenges persist in genome sequencing. First, generation of vast amounts of data necessitates sophisticated data processing and analysis methods. Issues such as data accuracy, quality

control, sequence comparison, and variant detection pose ongoing challenges[84]. Moreover, many genomes contain abundant repetitive sequences such as transposable elements, gene duplications, and tandem repeats[85]. These repetitive sequences complicate genome sequencing and assembly, making it difficult to accurately obtain or assemble sequences in certain genomic regions[86]. Although second-generation sequencing technologies produce numerous short-read sequences, they are insufficient for obtaining the long-read fragments required for accurate genome assembly, variant detection, and comprehensive annotation of gene structure and function[87]. Therefore, development of long-read sequencing technologies remains a priority. In contrast, single-cell transcriptomics offers the ability to examine the transcriptomes of individual cells, revealing cellular heterogeneity[88] and differences in transcript expression among subpopulations[89]. Concurrently, single-cell genome sequencing provides insights into genomic information at the individual cell level, enabling the exploration of genetic heterogeneity among cells[90], of transcriptomic changes during cell development, and of cellular differences among individuals[91]. The advancement of single-cell genome sequencing represents a current research frontier and holds promise for personalized medicine, developmental biology, and disease research. Future genome sequencing studies will focus on single-cell sequencing, long-read sequencing, multi-omics data integration, and genome function annotation, deepening our understanding of genomes and expanding the prospects of life science research and applications. The continual development of genome sequencing will open new avenues for studying and applying medicinal plants, further advancing the modernization of Chinese medicine and personalized drug development, making significant contributions to human health and the field of medicine.

Conflicts of interest statement

Shilin Chen is editorial board member of this journal. None of the other authors declare any conflicts of interest.

Funding

This research was funded by the National Natural Science Foundation of China, grant number 81603221.

Author contributions

Chunsheng Zhao, Lizhi Wang, and Shilin Chen designed the manuscript; Chunsheng Zhao wrote the manuscript; Linlin Sun, Ronglu Bai, and Ziwei Zhang helped to revise the manuscript. All authors read and approved the manuscript.

Ethical approval of studies and informed consent

Not applicable.

Acknowledgments

Thanks for the 7890B-7000D (Agilent Technologies) instrument support of College of Pharmaceutical Engineering of Traditional Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin 301617, China.

Data availability

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

References

[1] Dorado G, Gálvez S, Rosales TE, et al. Analyzing modern biomolecules: the revolution of nucleic-acid sequencing—review. *Biomolecules* 2021;11(8):1111.
[2] Searle B, Müller M, Carell T, et al. Third-generation sequencing of epigenetic DNA. *Angew Chem Int Ed Engl* 2023;62(14):e202215704.
[3] Payne AC, Chiang ZD, Reginato PL, et al. In situ genome sequencing resolves DNA sequence and structure in intact biological samples. *Science* 2021;371(6532):eaay3446.
[4] Sun Y, Shang L, Zhu QH, et al. Twenty years of plant genome sequencing: achievements and challenges. *Trends Plant Sci* 2022;27(4):391–401.
[5] van Dijk EL, Auger H, Jaszczyszyn Y, et al. Ten years of next-generation sequencing technology. *Trends Genet* 2014;30(9):418–426.
[6] Jayakodi M, Choi BS, Lee SC, et al. Ginseng Genome Database: an open-access platform for genomics of Panax ginseng. *BMC Plant Biol* 2018;18(1):62.
[7] Yu X, Wang W, Yang H, et al. Transcriptome and comparative chloroplast genome analysis of Vincetoxicum versicolor: insights into molecular evolution and phylogenetic implication. *Front Genet* 2021;12:602528.
[8] Xu Z, Peters RJ, Weirather J, et al. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of Salvia miltiorrhiza and tanshinone biosynthesis. *Plant J* 2015;82(6):951–961.
[9] Reuscher S, Akiyama M, Yasuda T, et al. The sugar transporter inventory of tomato: genome-wide identification and expression analysis. *Plant Cell Physiol* 2014;55(6):1123–1141.
[10] Li Z, Schulz MH, Look T, et al. Identification of transcription factor binding sites using ATAC-seq. *Genome Biol* 2019;20(1):45.
[11] Ma Y, Zhang L, Huang X. Genome modification by CRISPR/Cas9. *FEBS J* 2014;281(23):5186–5193.
[12] Yang Z, Wang Z, Wang W, et al. ggComp enables dissection of germplasm resources and construction of a multiscale germplasm network in wheat. *Plant Physiol* 2022;188(4):1950–1965.
[13] Alami MM, Ouyang Z, Zhang Y, et al. The current developments in medicinal plant genomics enabled the diversification of secondary metabolites' biosynthesis. *Int J Mol Sci* 2022;23(24):15932.
[14] Xiong X, Gou J, Liao Q, et al. The Taxus genome provides insights into paclitaxel biosynthesis. *Nat Plants* 2021;7(8):1026–1036.
[15] Guo C, Luo Y, Gao L-M, et al. Phylogenomics and the flowering plant tree of life. *J Integr Plant Biol* 2023;65:299–323.
[16] Kumondai M, Hishinuma E, Gutiérrez Rico EM, et al. Heterologous expression of high-activity cytochrome P450 in mammalian cells. *Sci Rep* 2020;10(1):14193.
[17] Shang T, Fang CM, Ong CE, et al. Heterologous expression of recombinant human cytochrome P450 (CYP) in Escherichia coli: N-terminal modification, expression, isolation, purification, and reconstitution. *BioTech (Basel)* 2023;12(1):17.
[18] Caniard A, Zerbe P, Legrand S, et al. Discovery and functional characterization of two diterpene synthases for sclareol biosynthesis in Salvia sclarea (L.) and their relevance for perfume manufacture. *BMC Plant Biol* 2012;12:119.
[19] da Silva PL, Cardoso G, Kremer FS, et al. Heterologous expression and characterization of a new galactose-binding lectin from Bauhinia forficata with antiproliferative activity. *Int J Biol Macromol* 2019;128:877–884.
[20] Levy SE, Boone BE. Next-Generation sequencing strategies. *Cold Spring Harb Perspect Med* 2019;9(7):a025791.
[21] Bergman ME, Davis B, Phillips MA. Medically useful plant terpenoids: biosynthesis, occurrence, and mechanism of action. *Molecules* 2019;24(21):3961.
[22] Chang WC, Song H, Liu HW, et al. Current development in isoprenoid precursor biosynthesis and regulation. *Curr Opin Chem Biol* 2013;17(4):571–579.

[23] Rather GA, Sharma A, Jeelani SM, et al. Metabolic and transcriptional analyses in response to potent inhibitors establish MEP pathway as major route for camptothecin biosynthesis in Nothapodytes nimmoniana (Graham) Mabb. *BMC Plant Biol* 2019;19(1):301.

[24] Sharma A, Verma P, Mathur A, et al. Overexpression of tryptophan decarboxylase and strictosidine synthase enhanced terpenoid indole alkaloid pathway activity and antineoplastic vinblastine biosynthesis in Catharanthus roseus. *Protoplasma* 2018;255(5):1281–1294.

[25] Ma J, Zhang X, Tian Y, et al. Molecular cloning and heterologous of putative strictosidine synthases from Arabidopsis thaliana. *Chin J Appl Environ Biol* 2013;19(2):224–230.

[26] Wang X, Liang L, Li B, et al. Analysis on STR sequence structure and expression pattern of strictosidine synthetase gene in Dendrobium officinale. *Biotic Resources* 2020;42(4):404–413.

[27] Xu R, Cao F, Peng J, et al. Cloning and anaiysis of strictosidine synthase in Rauvolfia yunnanensis. *J Cent South Univ Forestry Technol* 2012;32(6):128–131.

[28] Zhou Q, Li Y, Guo W, et al. Cloning and expression analysis of CsSTR1 gene in tea plant. *Southwest China J Agricultural Sci* 2022;35(10):2296–2302.

[29] Wang M, Zhou Y, Wang M, et al. Cloning and sequence analysis of strictosidine synthase cDNA in Brassica napus. *J Sichuan Univ* 2008;185(5):1277–1280.

[30] Li Y. Single-cell RNA sequencing reveals spatial distribution of biosynthetic pathway of monoterpene indole alkaloid in Catharanthus roseus leaves. Beijing: Peking Union Medical College; 2021.

[31] Miettinen K, Dong L, Navrot N, et al. The seco-iridoid pathway from Catharanthus roseus [published correction appears in Nat Commun 2014;5:4175]. *Nat Commun* 2014;5:3606.

[32] Okada K, Saito T, Nakagawa T, et al. Five geranylgeranyl diphosphate synthases expressed in different organs are localized into three subcellular compartments in Arabidopsis. *Plant Physiol* 2000;122(4):1045–1056.

[33] Iijima Y, Gang DR, Fridman E, et al. Characterization of geraniol synthase from the peltate glands of sweet basil. *Plant Physiol* 2004;134(1):370–379.

[34] Collu G, Unver N, Peltenburg-Looman AM, et al. Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett* 2001;508(2):215–220.

[35] Asada K, Salim V, Masada-Atsumi S, et al. A 7-deoxyloganetic acid glucosyltransferase contributes a key step in secologanin biosynthesis in Madagascar periwinkle. *Plant Cell* 2013;25(10):4123–4134.

[36] Irmler S, Schröder G, St-Pierre B, et al. Indole alkaloid biosynthesis in Catharanthus roseus: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *Plant J* 2000;24(6):797–804.

[37] Murata J, Roepke J, Gordon H, et al. The leaf epidermome of Catharanthus roseus reveals its biochemical specialization. *Plant Cell* 2008;20(3):524–542.

[38] Noé W, Mollenschott C, Berlin J. Tryptophan decarboxylase from Catharanthus roseus cell suspension cultures: purification, molecular and kinetic data of the homogenous protein. *Plant Mol Biol* 1984;3(5):281–288.

[39] Kutchan TM, Hampp N, Lottspeich F, et al. The cDNA clone for strictosidine synthase from Rauvolfia serpentina. DNA sequence determination and expression in Escherichia coli. *FEBS Lett* 1988;237(1-2):40–44.

[40] de Waal A, Meijer AH, Verpoorte R. Strictosidine synthase from Catharanthus roseus: purification and characterization of multiple forms. *Biochem J* 1995;306(Pt 2):571–580.

[41] Sadre R, Magallanes-Lundback M, Pradhan S, et al. Metabolite diversity in alkaloid biosynthesis: a multilane (Diastereomer) highway for camptothecin synthesis in camptotheca acuminata. *Plant Cell* 2016;28(8):1926–1944.

[42] Salim V, Jones AD, DellaPenna D. Camptotheca acuminata 10-hydroxycamptothecin O-methyltransferase: an alkaloid biosynthetic enzyme co-opted from flavonoid metabolism. *Plant J* 2018;95(1):112–125.

[43] Dang TT, Franke J, Carqueijeiro IST, et al. Sarpagan bridge enzyme has substrate-controlled cyclization and aromatization modes. *Nat Chem Biol* 2018;14(8):760–763.

[44] Dogru E, Warzecha H, Seibel F, et al. The gene encoding polyneuridine aldehyde esterase of monoterpenoid indole alkaloid biosynthesis in plants is an ortholog of the alpha/betahydrolase super family. *Eur J Biochem* 2000;267(5):1397–1406.

[45] Bayer A, Ma X, Stöckigt J. Acetyltransfer in natural product biosynthesis--functional cloning and molecular analysis of vinorine synthase. *Bioorg Med Chem* 2004;12(10):2787–2795.

[46] Geissler M, Burghard M, Volk J, et al. A novel cinnamyl alcohol dehydrogenase (CAD)-like reductase contributes to the structural diversity of monoterpenoid indole alkaloids in Rauvolfia. *Planta* 2016;243(3):813–824.

[47] Ruppert M, Woll J, Giritch A, et al. Functional expression of an ajmaline pathway-specific esterase from Rauvolfia in a novel plant-virus expression system. *Planta* 2005;222(5):888–898.

[48] Cázares-Flores P, Levac D, De Luca V. Rauvolfia serpentina N-methyltransferases involved in ajmaline and Nβ-methylajmaline biosynthesis belong to a gene family derived from γ-tocopherol C-methyltransferase. *Plant J* 2016;87(4):335–342.

[49] Qu Y, Easson MEAM, Simionescu R, et al. Solution of the multistep pathway for assembly of corynanthean, strychnos, iboga, and aspidosperma monoterpenoid indole alkaloids from 19E-geissoschizine. *Proc Natl Acad Sci U S A* 2018;115(12):3180–3185.

[50] Caputi L, Franke J, Farrow SC, et al. Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science* 2018;360(6394):1235–1239.

[51] Tatsis EC, Carqueijeiro I, Dugé de Bernonville T, et al. A three enzyme system to generate the Strychnos alkaloid scaffold from a central biosynthetic intermediate. *Nat Commun* 2017;8(1):316.

[52] Carqueijeiro I, Dugé de Bernonville T, Lanoue A, et al. A BAHD acyltransferase catalyzing 19-O-acetylation of tabersonine derivatives in roots of Catharanthus roseus enables combinatorial synthesis of monoterpene indole alkaloids. *Plant J* 2018;94(3):469–484.

[53] Besseau S, Kellner F, Lanoue A, et al. A pair of tabersonine 16-hydroxylases initiates the synthesis of vindoline in an organ-dependent manner in Catharanthus roseus. *Plant Physiol* 2013;163(4):1792–1803.

[54] Liscombe DK, Usera AR, O'Connor SE. Homolog of tocopherol C methyltransferases catalyzes N methylation in anticancer alkaloid biosynthesis. *Proc Natl Acad Sci USA* 2010;107(44):18793–18798.

[55] Qu Y, Easson ML, Froese J, et al. Completion of the seven-step pathway from tabersonine to the anticancer drug precursor vindoline and its assembly in yeast. *Proc Natl Acad Sci USA* 2015;112(19):6224–6229.

[56] Liu J, Liu G. Analysis of secondary metabolites from plant endophytic fungi. *Methods Mol Biol* 2018;1848:25–38.

[57] Ancheeva E, Daletos G, Proksch P. Bioactive Secondary Metabolites from Endophytic Fungi. *Curr Med Chem* 2020;27(11):1836–1854.

[58] Fröhlich T, Çapcı Karagöz A, Reiter C, et al. Artemisinin-derived dimers: potent antimalarial and anticancer agents. *J Med Chem* 2016;59(16):7360–7388.

[59] Tian H. The stimulation biosynthesis, the bioconversion of artemisinin acid and biodegradation of triclosan by an endophytic Penicillium oxalicum B4 from Artemisia annua L. Suzhou: Suzhou University; 2019.

[60] Zhu L, Chen L. Progress in research on paclitaxel and tumor immunotherapy. *Cell Mol Biol Lett* 2019;24:40.

[61] Abu Samaan TM, Samec M, Liskova A, et al. Paclitaxel's mechanistic and clinical effects on breast cancer. *Biomolecules* 2019;9(12):789.

[62] Lau TS, Chan LKY, Man GCW, et al. Paclitaxel induces immunogenic cell death in ovarian cancer via TLR4/IKK2/SNARE-dependent exocytosis. *Cancer Immunol Res* 2020;8(8):1099–1111.

[63] Blair HA, Deeks ED. Albumin-bound paclitaxel: a review in non-small cell lung cancer. *Drugs* 2015;75(17):2017–2024.

[64] Zhang D, Yang R, Wang S, et al. new uses for an old drug. *Drug Des Devel Ther* 2014;8:279–284.

[65] Service RF. Hazel trees offer new source of cancer drug. *Science* 2000;288(5463):27–28.

[66] Zhang L, Gao L, Lin B. Extraction of taxol from the skin of hazel in anshan area. *Fujian Anal Test* 2003;12(2):29–30, 42.

[67] Zaynab M, Fatima M, Abbas S, et al. Role of secondary metabolites in plant defense against pathogens. *Microb Pathog* 2018;124:198–202.

[68] Shen N, Wang T, Gan Q, et al. Plant flavonoids: classification, distribution, biosynthesis, and antioxidant activity. *Food Chem* 2022;383:132531.

[69] Fiorito S, Epifano F, Taddeo VA, et al. Recent acquisitions on oxyprenylated secondary metabolites as anti-inflammatory agents. *Eur J Med Chem* 2018;153:116–122.

[70] Lopes EM, Guimarães-Dias F, Gama TDSS, et al. Artemisia annua L. and photoresponse: from artemisinin accumulation, volatile profile and anatomical modifications to gene expression. *Plant Cell Rep* 2020;39(1):101–117.

[71] Dai Y, Chen SR, Chai L, et al. Overview of pharmacological activities of Andrographis paniculata and its major compound andrographolide. *Crit Rev Food Sci Nutr* 2019;59(suppl 1):S17–S29.
[72] Zhang R, Chen Z, Zhang L, et al. Genomic characterization of WRKY transcription factors related to andrographolide biosynthesis in andrographis paniculata. *Front Genet* 2021;11:601689.
[73] Gupta D, Bhattacharjee O, Mandal D, et al. CRISPR-Cas9 system: a new-fangled dawn in gene editing. *Life Sci* 2019;232:116636.
[74] Kim YJ, Zhang D, Yang DC. Biosynthesis and biotechnological production of ginsenosides. *Biotechnol Adv* 2015;33(6 Pt 1):717–735.
[75] Choi HS, Koo HB, Jeon SW, et al. Modification of ginsenoside saponin composition via the CRISPR/Cas9-mediated knockout of protopanaxadiol 6-hydroxylase gene in Panax ginseng. *J Ginseng Res* 2022;46(4):505–514.
[76] Subedi L, Gaire BP. Tanshinone IIA: a phytochemical as a promising drug candidate for neurodegenerative diseases. *Pharmacol Res* 2021;169:105661.
[77] Li Q, Fang X, Zhao Y, et al. The SmMYB36-SmERF6/SmERF115 module regulates the biosynthesis of tanshinones and phenolic acids in salvia miltiorrhiza hairy roots. *Hortic Res* 2022;10(1):uhac238.
[78] Yoshinaga Y, Daum C, He G, et al. Genome sequencing. *Methods Mol Biol* 2018;1775:37–52.
[79] Singh D, Singh CK, Taunk J, et al. Linking genome wide RNA sequencing with physio-biochemical and cytological responses to catalogue key genes and metabolic pathways for alkalinity stress tolerance in lentil (Lens culinaris Medikus). *BMC Plant Biol* 2022;22(1):99.
[80] Hou Y, Guo H, Cao C, et al. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Res* 2016;26(3):304–319.
[81] Wang S, Li Q. Genome-wide identification of the Salvia miltiorrhiza SmCIPK gene family and revealing the salt resistance characteristic of SmCIPK13. *Int J Mol Sci* 2022;23(12):6861.
[82] Zhong F, Ke W, Li Y, et al. Comprehensive analysis of the complete mitochondrial genomes of three Coptis species (C. chinensis, C. deltoidea and C. omeiensis): the important medicinal plants in China. *Front Plant Sci* 2023;14:1166420.
[83] Xu J, Chu Y, Liao B, et al. Panax ginseng genome examination for ginsenoside biosynthesis. *Giga Science* 2017;6(11):1–15.
[84] Pei L, Wang B, Ye J, et al. Genome and transcriptome of Papaver somniferum Chinese landrace CHM indicates that massive genome expansion contributes to high benzylisoquinoline alkaloid biosynthesis. *Hortic Res* 2021;8(1):5.
[85] Chen C, Xing D, Tan L, et al. Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science* 2017;356(6334):189–194.
[86] Yuan J, Zhang X, Li F, et al. Genome sequencing and assembly strategies and a comparative analysis of the genomic characteristics in Penaeid Shrimp species. *Front Genet* 2021;12:658619.
[87] Zhang T, Zhou J, Gao W, et al. Complex genome assembly based on long-read sequencing. *Brief Bioinform* 2022;23(5):bbac305.
[88] Wang Z, Chai C, Wang R, et al. Single-cell transcriptome atlas of human mesenchymal stem cells exploring cellular heterogeneity. *Clin Transl Med* 2021;11(12):e650.
[89] Zhang M, Hu S, Min M, et al. Dissecting transcriptional heterogeneity in primary gastric adenocarcinoma by single cell RNA sequencing. *Gut* 2021;70(3):464–475.
[90] Wen L, Tang F. Single-cell sequencing in stem cell biology. *Genome Biol* 2016;17:71.
[91] Shan B, Barker CS, Shao M, et al. Multilayered omics reveal sex- and depot-dependent adipose progenitor cell heterogeneity. *Cell Metab* 2022;34(5):783–799.e7.

How to cite this article: Zhao CS, Zhang ZW, Sun LL, Bai RL, Wang LZ, Chen SL. Genome sequencing provides potential strategies for drug discovery and synthesis. Acupunct Herb Med 2023;3(4):244–255. doi: 10.1097/HM9.0000000000000076