Natural products of medicinal plants: biosynthesis and bioengineering in post-genomic era

Li Guo1,*, Hui Yao1,*, Weiwei Chen1, Xumei Wang3, Peng Ye3, Zhichao Xu2, Sisheng Zhang4 and Hong Wu†

1Shandong Laboratory of Advanced Agricultural Sciences at Weifang, Peking University Institute of Advanced Agricultural Sciences, Weifang, Shandong 261000, China
2Key Laboratory of Bioactive Substances and Resources Utilization of Chinese Herbal Medicine, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100193, China
3School of Pharmacy, Xi’an Jiaotong University, Xi’an 710061, China
4State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, Guangdong Laboratory For Lingnan Modern Agriculture, College of Life Sciences, South China Agricultural University, Guangzhou 510642, China
5College of Life Science, Northeast Forestry University, Harbin 150040, China
*Corresponding authors. E-mails: wh@scau.edu.cn; li.guo@pku-iaas.edu.cn
†Contributed equally to this work.

Abstract

Globally, medicinal plant natural products (PNPs) are a major source of substances used in traditional and modern medicine. As we human race face the tremendous public health challenge posed by emerging infectious diseases, antibiotic resistance and surging drug prices etc., harnessing the healing power of medicinal plants gifted from mother nature is more urgent than ever in helping us survive future challenge in a sustainable way. PNP research efforts in the pre-genomic era focus on discovering bioactive molecules with pharmaceutical activities, and identifying individual genes responsible for biosynthesis. Critically, systemic biological, multi- and inter-disciplinary approaches integrating and interrogating all accessible data from genomics, metabolomics, structural biology, and chemical informatics are necessary to accelerate the full characterization of biosynthetic and regulatory circuitry for producing PNPs in medicinal plants. In this review, we attempt to provide a brief update on the current research of PNPs in medicinal plants by focusing on how different state-of-the-art biotechnologies facilitate their discovery, the molecular basis of their biosynthesis, as well as synthetic biology. Finally, we humbly provide a foresight of the research trend for understanding the biology of medicinal plants in the coming decades.

Natural products in medicinal plants: hidden treasures with healing power

Plants have existed on Earth for hundreds of millions of years and evolved ingenious chemical factories to survive exogenous and endogenous stresses [1]. These chemicals known as secondary metabolites or natural products are synthesized by plants to accommodate environmental changes without disrupting much of their cellular and developmental physiological processes [2]. To date more than 100,000 natural products are present in the Kingdom of Plants, primarily involved in plant defense against biotic and abiotic stresses [3]. These powerful substances are also important chemical signals mediating plant communication with symbiotic microorganisms, and attracting pollinators and seed dispersal. Derived from primary metabolites, secondary metabolites accumulate at cellular, tissue and organ levels through diverse biosynthetic pathways [4]. Plant natural products (PNPs) are generally divided into three classes: phenolics, terpenoids, and alkaloids [5] and broadly used as pharmaceuticals, nutraceuticals, cosmetics, and fine chemicals. Phenolics are synthesized from the shikimic acid biosynthetic pathway where the final products are formed after phenylalanine and aromatic amino acids undergo deamination, hydroxylation, and coupling reactions [7] (Fig. 1A). In general, phenolics consist of monophenols such as benzoids, and polyphenols such as flavonoids, stilbenoids, and coumuminoids [6]. Terpenoids are biosynthesized through mevalonic acid and mehylerythritol phosphate pathways from isopentenyl diphosphate (IPP, C10), the precursor and fundamental structural unit of all terpenoids including monoterpenoids (C10), sesquiterpenoids (C15), diterpenoids (C20), and triterpenoids (C30) [8] (Fig. 1B). Alkaloids are a large group of plant nitrogen-containing compounds with a broad range of pharmaceutical activities such as painkilling (e.g. morphine), cough-suppressing (e.g. noscapine), anti-inflammation (e.g. sanguinarine, berberine), and anti-cancer (e.g vinblastine, noscapine) (Fig. 2). Originated from the amino acid and isoamylene biosynthetic pathway, alkaloids are generally classified into sparteine, quinine, mescline, coniine, and aconitine (Fig 2). The biosynthetic pathways of different alkaloids are diversified and often independent. Recently, Lichman [9] has summarized alkaloid pathways as four general steps: (i) accumulation of an amine precursor; (ii) accumulation of an aldehyde precursor; (iii) formation of an iminium cation; and (iv) a manich-like reaction.
PNPs are bioactive substances dispensable to normal plant cellular functions yet vital to biodefense and environmental adaptation of plants with a sessile lifestyle. As a major source of traditional and modern medicine, PNP s have had broad implications for human health as a herb remedy for thousands of years, and changed the course of human civilization and history [9]. The healing power of medicinal herbs has long been recognized and harnessed by our human ancestors and forefathers, who learned to use plant-based folk medicine to cure ailments such as headache, fever, and pains. For example, archaeologists found in a grave from Shanidar – an archaeological site in Iraq – that a Neanderthal man may have used plant-based medicine at around 60,000 BCE, the earliest record of human use of herb medicine [10]. In addition, ancient Europeans had started to cultivate and use different varieties of opium poppy plants at least 5000 years ago [11]. Another example is tobacco, which was historically used to treat various ailments including yaws, syphilis, and black death given its strong antimicrobial activities [12,13].

Much of early human knowledge about medical herbs is documented in ancient scriptures and literature such as Treatise on Cold Pathogenic and Miscellaneous Diseases by Zhang Zhongqing (Eastern Han dynasty) and Compendium of Materia Medica composed by Li Shizhen (Ming dynasty). However, the use of medicinal herbs in traditional medicine has been largely considered as empiricism with little knowledge of the chemical properties of the effective PNP s. The first isolated PNP was morphine, a benzylisoquinoline alkaloid in opium poppy (Papaver somniferum) plants, by Germany pharmacist Friedrich Sertürner at around 1817, marking the birth of modern chemistry of natural products [14]. Later on, another alkaloid quinine isolated from the bark of cinchona tree (Cinchona officinalis) became the first effective medicine against malaria, caused by mosquito-transmitted Plasmodium species [15]. Salicin, the original source of aspirin and identified in the bark of willow tree (Salix babylonica), is another significantly used PNP, used prominently as a pain reliever [16].

To date, hundreds of plant-based bioactive compounds have been identified and many are used as an effective treatment of human diseases such as ginsenoside (anti-tumor), paclitaxel/taxol (anti-tumor), and artemisinin (anti-malaria) etc. The sesquiterpene endoperoxide artemisinin from Artemisia annua is recommended by the WHO as the most effective drug against malaria [17]. The paclitaxel (taxol) isolated from the bark trees of Taxus genus has been approved for the treatment of ovarian, breast, and lung cancer, as well as Kaposi’s sarcoma [18]. These examples have demonstrated that medicinal plants and their powerful natural products have great healing power and have shaped the development of human history.

Throughout history, the human race has battled with many infectious diseases such as tuberculosis, cholera, malaria, black death, influenza, smallpox, etc. The most recent public health challenge comes from the coronavirus pandemic initiated in 2019 (COVID-19) which continuously threatens the world with the non-stop emergence of new variants. These pandemics each and collectively have led to significant progress in human knowledge of medical sciences and the generation of new public health solutions in which medicinal plant biology and chemistry. Here, we attempt to review the current research of PNP s by focusing on how different state-of-the-art multi-discipline technologies facilitate their discovery, the molecular basis of biosynthesis, as well as bioengineering via breeding and synthetic biology. Finally, we highlight a few research trends regarding the exploitation of medicinal plants in the next decades [25].

**Decoding biosynthetic pathways of PNP s is key to their applications**

The challenge of acquiring PNP s en masse stems from the fact that medicinal plants have a tight regulation of producing these chemicals, mostly localized in specific cells and tissues under certain conditions [26]. The biosynthesis of PNP s typically goes through a cascade of enzymatic reactions converting primary metabolites into various structurally diverse secondary metabolites. Although some reactions can occur spontaneously in nature, most steps require a catalysis by enzymes such as cytochrome P450, methyltransferases, O-methyltransferases, deaminases, UDP-glucuronosyltransferases, etc. A major challenge in exploitation of PNP s is to understand their biosynthetic pathways so that the bioengineering approach can be applied to produce them on a massive scale through plant breeding or synthetic biology. Fully resolving the biosynthetic pathways of any PNP is a daunting task, because plants have evolved a complex cellular network of metabolic pathways [27], formed by various enzymes catalyzing a myriad of biochemical reactions, and regulatory proteins fine-tuning the spatial and temporal accumulation of PNP s. Currently, the full biochemical pathways remain unknown for the majority of medicinal PNP s after years of research efforts, highlighting the difficulty of decoding PNP biosynthetic pathways.

Before the genomic era, plant biosynthetic pathway characterization was laborious and time-consuming, either relying on approaches such as isotope labeling and forward genetics, such as by creating random mutants followed by analysing their metabolic profiles, or involving sequence-homology based
Figure 1. Proposed general biosynthetic pathway of plant phenolics (A) and terpenoids (B). DMAPP: dimethylallyl pyrophosphate; IPP: isopentenyl pyrophosphate; MEP: methlyerythritol phosphate; MVA: mevalonic acid.

Figure 2. Common alkaloid types (A) and examples of pharmaceutical alkaloids (B) in medicinal plants.

gene cloning to identify individual biosynthetic enzymes [28]. These studies typically identify genes encoding parts of the pathways through guilt-by-association, as loss- or gain-of-function mutations of true biosynthetic genes can alter metabolic profiles. However, they are usually inept to resolve the complete components and reconstruct the pathway owing to the pleiotropic or promiscuous nature of the biosynthetic enzymes. The majority of our knowledge about PNP biosynthetic pathways so far is derived from such homology-based or transcriptome-based gene mining [29, 30]. With growing volumes of genomic data available for medicinal plants, it is now common to exploit high-throughput data mining combined with experimental validations to untangle the complex biosynthetic pathways and networks underlying PNP accumulation in medicinal herbs.

PNP biosynthetic genes are usually co-expressed and co-regulated in specific tissues and growth stages. Therefore, in the post-genomic era, gene expression profiling techniques such as whole transcriptome sequencing (RNA-seq) enable researchers to quickly narrow down the co-expressed candidate genes encoding specific biosynthetic pathways, typically via comparative transcriptome analysis of plant samples with contrasting levels of metabolic production. Transcriptomic analysis combined with pathway inference based on chemistry logic, and experimental validations in heterologous hosts are routinely used to identify biosynthetic genes for PNPs such as thebaine and noscapine of opium poppy (P. somniferum) [31], sanguinarine and chelerythrine of Macleaya cordata [32], vinblastine of Madagascar periwinkle [33], colchicine of Colchicum autumnale [34] and strychnine of Strychnos nux-vomica [35]. Alternatively, proteomic profiling is also used to identify proteins corresponding to specific biosynthetic pathways, capturing translational and post-translational modifications unseen in transcriptomic data. Candidate genes identified by omic profiling analysis are then validated experimentally to ascertain their biochemical functions such as catalyzing specific reactions, metabolite transport or transcriptional regulation. It typically involves heterologously expressing the candidate gene(s) in microbial cells including bacteria (Escherichia coli [36], Corynebacterium glutamicum [37]), yeasts (Saccharomyces cerevisiae...
Validations [42]. This approach still suffers from the fact that PNPs, thus generating hypothesis for downstream experimental geno-expression network correlated with accumulation of association. Network mining using these big data can reconstruct proteomic with metabolomic profiling in different tissues and mining of multi-dimensional data such as transcriptomic or raphy (GC) coupled with mass spectrometry (MS). Integrated mining of multi-dimensional data such as transcriptomic or proteomic with metabolomic profiling in different tissues and growth stages of medicinal plants enables gene-metabolite association. Network mining using these big data can reconstruct the gene co-expression networks correlated with accumulation of PNPs, thus generating hypothesis for downstream experimental validations [43]. This approach still suffers from the fact that many biosynthetic genes may be cryptic or lowly expressed, leaving them undiscovered by the expression-based methods. Thus, systematic decoding of PNP biosynthetic pathways will require a framework of high-throughput analysis of omic data from plants growing under multiple conditions or developmental stages. For example, two recent studies acquired transcriptomic and metabolomic data of tomato and rice plants across the full spectrum of growing stages, and by integrative data analysis revealed gene modules associated with natural products [44, 45]. It is expected that similar analysis from full growth stages of medicinal plants will facilitate discovery of biosynthetic genes and yield critical insight into the regulatory networks underlying PNP biosynthesis.

Medicinal plant genomes yield insights into composition and evolution of biosynthetic pathways

Genome sequence dictates the foundation of biological functions of all life forms. Despite the progress made by the homology- and transcriptome-based approach, the elucidation of full biosynthetic pathways is often hampered by a lack of reference genome sequences for medicinal plants. A reference genome sets the foundation to identify all protein-coding genes, regulatory DNA elements and importantly their precise genomic locations. Next-generation sequencing (NGS) and third-generation sequencing (TGS) technologies have revolutionized biological sciences by changing the way genomes are decoded. The first human and plant (Arabidopsis thaliana) genome assemblies were initially achieved using Sanger sequencing. Despite the high accuracy, Sanger sequencing is expensive and time-consuming to generate sequencing data for assembly of eukaryotic genomes of mid to large sizes. Since around 2005, genome assembly projects started to adopt high-throughput sequencing platforms like Solexa, Ion Torrent and later Illumina, giving rise to the first reference genomes for many organisms. However, contig-level assemblies using short reads have low contiguity (low N50) with numerous assembly errors due to high repeat content of plant genomes, even when a high coverage and long insert such as mate-pair libraries or linked-reads are used. Scaffold level assemblies are often improved using genetic map data such as GBS (genotyping by sequencing), or optical maps to anchor the contigs to linkage groups. However, a high-resolution genetic map is essential to reducing contig misplacement but often unavailable for non-model plants. TGS technologies developed by Pacific Biosciences (PB) and Oxford Nanopore technology (ONT) produce single-molecule DNA sequencing reads of 20 kb or longer, albeit error-prone (up to 15%). TGS became a game changer for genome assembly because long reads can often span most repetitive regions. Besides, technologies such as chromatin conformation capture sequencing (e.g. Hi-C) and Bionano are now routinely used to anchor contigs to chromosomes and correct misassemblies present in NGS and TGS draft assemblies. As a result, for model organisms and many agricultural organisms, genome assembly quality and contiguity have leaped to much higher levels with a combination of long-read and short-read sequencing data.

Medicinal plant genomes have a wide genome size range, high heterozygosity rates and repeat contents, making them difficult to assemble correctly [46]. Leveraging these different technologies combined with improving bioinformatic algorithms has yielded a growing number of high-quality medicinal plant genome assemblies and annotations. Nearly 100 species of medicinal plants (Table 1) have at least one version of reference genome available [47–139], although the quality of current genome assemblies varies depending on the genome complexity of medicinal plants and choice of sequencing technologies as well as computational tools used in assembly. The herbgenomics initiative, firstly proposed in 2010, has greatly promoted the elucidation of biosynthetic pathways for many medicinal bioactive ingredients [48]. Under this initiative and many independent genome projects, several medicinal plants have had reference genomes assembled even at chromosome level, such as P. somniferum [49], Camptotheca acuminata [50], Scutellaria baicalensis [51], Panax notoginseng [52], Tripterygium wilfordii [53], Salvia miltiorrhiza [54], Taxus wallichiana [55], and Erigeron breviscapus [56]. Lately, a Chinese consortium of the 1 K Herb Genomes Project has been officially launched to produce high-quality genome sequences for 1000 high-value TCM in order to promote the study and exploitation of their PNP.

The chromosome-level genome assemblies are instrumental to highly robust downstream genomic analyses such as comparative genome analysis, chromosome evolution analysis, and gene cluster characterization etc. For example, chromosome-level assemblies of P. somniferum has allowed Guo et al. to discover a gene cluster (BIA gene cluster) that encodes biosynthetic pathways for two morphinans: morphine and noscapine [49]. Two additional assemblies of Papaver species produced by Yang et al. have enabled them to reconstruct the evolutionary history of Papaver karyotypes, showing morphinian biosynthetic pathways underwent punctuated evolution pattern [57]. Tu et al. [53] assembled a high quality chromosomal-scale Tripterygium wilfordii genome and found the recent duplication of triptolide biosynthetic pathway genes. Then multiple omics methods were integrated to construct gene-to-metabolite network, and finally a CYP72B70 that participated in triptolide biosynthesis was identified. The genome assemblies for C. acuminata [50] and Catharanthus roseus [33] also provide comprehensive genomic resources for the analysis of camptothecin and vinblastine biosynthesis pathway, which share common upstream to produce loganic acid, and then flux into two independent branches, respectively. The C. roseus genome, coupled with chemical investigations, enabled the discovery of the last two enzymes of precamptothecine acetate synthase and dihydroprecamptothecine synthase responsible for vinblastine biosynthesis [33, 58], resolving a long-standing question of how vinblastine/vincristine is synthesized, making their heterologous production possible. The C. acuminata genome found two secoliganic acid synthases that converted the loganic acid flux into camptothecin production, and the downstream candidate genes set the foundation to fully discover camptothecin biosynthetic mechanism [50]. Furthermore, high quality genomes are also essential in genome-wide association studies (GWAS) to identify quantitative trait loci
**Table 1.** Sequenced genomes of medicinal plants and their primary medicinal ingredients.

| Species                     | Genome size | Contig N50 | No. of annotated genes | Assembly Levela | Release time | Primary medicinal ingredients                                                                 | Reference |
|-----------------------------|-------------|------------|------------------------|-----------------|--------------|------------------------------------------------------------------------------------------------|-----------|
| Apium graveolens            | 3.33 Gb     | 790.6 kb   | 31,326                 | C               | 2020         | Apigenin and luteolin                                                                          | [60]      |
| Aralia elata                | 1.05 Gb     | 1.20 Mb    | 35,042                 | C               | 2022         | Oleaneane-type triterpenoids                                                                   | [61]      |
| Coriandrum sativum          | 2.12 Gb     | 604.1 kb   | 40,747                 | C               | 2020         | Mannitol, furlfural and linalool                                                               | [62]      |
| Panax ginseng               | 3.41 Gb     | 19.75 Mb   | 65,913                 | C               | 2020         | Dammarane-type saponins                                                                       | [63]      |
| Panax japonicus             | 2.09 Gb     | 1.22 Mb    | 74,307                 | C               | 2022         | Dammarane-type saponins                                                                       | [63]      |
| Panax notoginseng           | 2.66 Gb     | 1.12 Mb    | 37,606                 | C               | 2020         | Dammarane-type ginsenoside                                                                    | [52]      |
| Panax quinquefolius         | 3.60 Gb     | 0.87 Mb    | 64,247                 | C               | 2022         | Dammarane-type saponins                                                                       | [63]      |
| Panax stipuleanatus         | 2.15 Gb     | 2.88 Mb    | 41,224                 | C               | 2022         | Dammarane-type saponins                                                                       | [63]      |
| Allium sativum              | 16.24 Gb    | 194 kb     | 57,561                 | C               | 2020         | Allin                                            | [64]      |
| Dendrobium huoshanense      | 1.28 Gb     | 598 kb     | 21,070                 | C               | 2020         | Polysaccharides and alkaloids                                                                  | [65]      |
| Dendrobium officinalis      | 1.23 Gb     | 1.44 Mb    | 27,631                 | C               | 2021         | Polysaccharides, alkaloids and flavonoids                                                       | [66]      |
| Gastrodia elata             | 1.04 Gb     | 9.18 Mb    | 18,844                 | C               | 2022         | Gastrodin, phloxybenzyl alcohol and parision                                                   | [67]      |
| Arctium lappa               | 1.73 Gb     | 74.69 Mb   | 47,055                 | C               | 2022         | Arctigenin and arctinin                                                                        | [68]      |
| Artemisia annua             | 1.55 Gb     | 262.1 kb   | 54,347                 | C, H            | 2022         | Artemisinin                                                                                   | [69]      |
| Artemisia argyi             | 8.03 Gb     | 6.25 Mb    | 279,294                | C, H            | 2022         | Hispidulin, jacesidin, and eupatini                                                            | [70]      |
| Carthamus tinctorius        | 1.06 Gb     | 21.23 Mb   | 33,343                 | C               | 2021         | Linoleic acid and hydroxysafflor yellow A                                                       | [71]      |
| Chrysanthemum nankingense   | 2.53 Gb     | 130.7 kb   | 56,870                 | S               | 2018         | Flavonoids and terpenes                                                                       | [72]      |
| Erigeron breviscapus        | 1.43 Gb     | 140.95 kb  | 43,514                 | C               | 2020         | Scutellarin                                                                                   | [56]      |
| Smallanthus sonchifolius    | 2.72 Gb     | 87.39 Mb   | 89,960                 | C               | 2022         | Saccharides (inulin)                                                                           | [68]      |
| Platycodon grandiflorus     | 622.86 Mb   | 29.94 Mb   | 22,358                 | C               | 2022         | Triterpenoid saponins                                                                          | [73]      |
| Isatis indigotica           | 293.88 Mb   | 1.18 Mb    | 30,323                 | C               | 2020         | Indole alkaloids, 7-sitosterol and daucosterol                                                  | [74]      |
| Lepidium meyenii            | 743 Mb      | 81.78 kb   | 96,417                 | S               | 2016         | Macaridine, macamides and macaene                                                             | [75]      |
| Polygonum cuspidatum        | 2.56 Gb     | 2.77 kb    | 55,075                 | S               | 2019         | Stibenes and quinones                                                                          | [76]      |
| Tripterygium wilfordii      | 348.38 Mb   | 4.36 Mb    | 28,321                 | C               | 2020         | Triptolide                                                                                    | [53]      |
| Ceratophyllum demersum      | 860.5 Mb    | 2.57 Mb    | 30,138                 | C               | 2020         | Plastocyanin and ferredoxin                                                                   | [77]      |
| Camptotheca acuminata       | 414.95 Mb   | 1.47 Mb    | 27,940                 | C               | 2021         | Camptothecin                                                                                  | [50]      |
| Benincasa hispida           | 912.95 Mb   | 145 kb     | 27,467                 | C               | 2019         | Vitamins and flavonoids                                                                        | [78]      |
| Gynostemma pentphyllum      | 518.46 Mb   | 6.67 Mb    | 25,285                 | C               | 2021         | Gypenosides                                                                                   | [79]      |
| Luffa cylindrica            | 656.19 Mb   | 8.80 Mb    | 25,508                 | C               | 2020         | Sapogenins                                                                                     | [80]      |
| Monodora chromaria          | 293.6 Mb    | 3.3 Mb     | 26,427                 | C               | 2020         | Cucurbitacins and cucurbitane glycosides                                                       | [81]      |
| Siraitia grosvenorii        | 467.07 Mb   | 433.68 kb  | 30,565                 | S               | 2018         | Mogrosides                                                                                    | [82]      |
| Lonicerapponica             | 843.2 Mb    | 2.1 Mb     | 33,939                 | C               | 2020         | Luzeinol and chlorogenic acid                                                                  | [83]      |
| Diospyros lotus             | 907 Mb      | 1.06 Mb    | 40,532                 | C               | 2020         | Tannin                                                                                       | [84]      |
| Camellia sinensis           | 3.02 Gb     | 19.96 Mb   | 36,951                 | C               | 2017         | Caffeine and theanine                                                                          | [85]      |
| Glycyrrhiza uralensis       | 378.86 Mb   | 7.32 kb    | 38,135                 | S               | 2017         | Glycyrrhizin and liquiritin                                                                   | [86]      |
| Senna tora                  | 526.36 Mb   | 4.03 Mb    | 45,268                 | C               | 2020         | Anthraquinones                                                                                | [87]      |
| Spatholobus suberecutus     | 798.47 Mb   | 2.05 Mb    | 31,634                 | C               | 2019         | Eriodictyol and dihydroquercetin                                                               | [88]      |
| Eucommia ulmoides           | 947.84 Mb   | 13.16 Mb   | 26,001                 | C               | 2020         | Pinoresin diglucoside and aucubin                                                             | [89]      |
| Calotropis gigantea         | 157.28 Mb   | 48.58 Mb   | 18,197                 | S               | 2018         | Cardenolides                                                                                  | [90]      |

Continued
| Species                  | Genome size | Contig N50 | No. of annotated genes | Assembly Level¹ | Release time | Primary medicinal ingredients                       | Reference |
|-------------------------|-------------|------------|------------------------|-----------------|--------------|-----------------------------------------------------|-----------|
| Catharanthus roseus     | 523 Mb      | 26.2 kb    | 33,829                 | S               | 2015         | Vinblastine and vincristine                         | [33]      |
| Gelsemium elegans       | 335.13 Mb   | 10.23 Mb   | 26,768                 | C               | 2020         | Gelsemine and koumine alkaloids                    | [91]      |
| Gardenia jasminoides    | 535.68 Mb   | 1.03 Mb    | 35,967                 | C               | 2020         | Genipin and crocins                                 | [92]      |
| Morinda officinalis     | 484.85 Mb   | 4.21 Mb    | 27,698                 | C               | 2021         | Antraquinones and monotropein                      | [93]      |
| Ophiopogon pumila       | 439.90 Mb   | 18.49 Mb   | 32,389                 | C               | 2021         | Camptothecin                                       | [94]      |
| Andrographis paniculata | 284.52 Mb   | 5.15 Mb    | 24,015                 | C               | 2020         | Neoaandrographolide                                 | [95]      |
| Mentha longifolia       | 353 Mb      | 4.47 kb    | 35,597                 | S               | 2017         | Menthol and menthol                                | [96]      |
| Ocimum basilicum        | 2.06 Gb     | 48.30 Kb   | 78,990                 | S               | 2020         | Linalool and cineole                               | [97]      |
| Origanum vulgare        | 630.04 Mb   | 26.28 kb   | 32,623                 | S               | 2020         | Carvacrol, thymol and ocimene                      | [97]      |
| Perilla frutescens      | 1.24 Gb     | 3.21 Mb    | 95,008                 | C               | 2021         | Perillaldehyde, perillalcohol, limonene             | [98]      |
| Pogostemon cablin       | 1.94 Gb     | 7.97 Mb    | 109,696                | C, H            | 2022         | Patchouli alcohol                                  | [99]      |
| Rosmarinus officinalis  | 1.01 Gb     | 21.82 kb   | 51,389                 | S               | 2020         | Rosmarinic acid, camphor and carnosol              | [97]      |
| Salvia boustoniana      | 462.44 Mb   | 1.18 Mb    | 44,044                 | C               | 2021         | Tanshinone and salvianolic acid                     | [100]     |
| Salvia miltiorrhiza     | 594.75 Mb   | 2.7 Mb     | 32,483                 | S               | 2020         | Phenolic acids and tanshinones                     | [54]      |
| Salvia splendens        | 809.16 Mb   | 3.77 Mb    | 88,489                 | C               | 2021         | Anticoagulant                                      | [101]     |
| Scutellaria baicalensis | 377.0 Mb    | 2.10 Mb    | 33,414                 | C               | 2020         | Baicalin, scutellarein, norwogonin, wogonin        | [102]     |
| Scutellaria baicalensis | 353.0 Mb    | 2.50 Mb    | 41,697                 | C               | 2020         | Baicalin, scutellarein, norwogonin, wogonin        | [102]     |
| Forsythia suspensa      | 737.47 Mb   | 7.33 Mb    | 33,062                 | S               | 2020         | Pionesinol and phillyrin                           | [103]     |
| Antirrhinum magnus      | 520 Mb      | 7.30 Mb    | 37,714                 | C               | 2019         | Antirhinoside and linolenic acid                    | [104]     |
| Chinomnthus praecox     | 695.36 Mb   | 2.19 Mb    | 25,591                 | C               | 2020         | Essential oil                                      | [105]     |
| Cinnamomum kaneharae   | 730.42 Mb   | 498.9 kb   | 27,899                 | C               | 2019         | Camphor                                            | [106]     |
| Paris polyphylla        | 70.18 Gb    | 1.81 kb    | 34,257                 | S               | 2020         | Polyphenylins                                      | [107]     |
| Ricinus communis        | 325.5 Mb    | 21.1 kb    | 31,237                 | S               | 2010         | Ricin                                              | [108]     |
| Hypericum perforatum    | 373.65 Mb   | 1.41 Mb    | 29,150                 | S               | 2021         | Hypercin, hyperforin, and melatonin                | [109]     |
| Linum usitatissimum     | 318.25 Mb   | 20.1 kb    | 43,484                 | S               | 2012         | Unsaturated fatty acids and secoisolaricresinol diglucoside | [110]     |
| Passiflora edulis       | 1.34 Gb     | 6.40 Mb    | 23,171                 | C               | 2021         | C-glycosyl flavonoids                              | [111]     |
| Aquilaria sinensis      | 783.8 Mb    | 60.21 Kb   | 35,965                 | C               | 2020         | Sesquiterpenes and flindersiacromone                | [112]     |
| Punica granatum         | 328.38 Mb   | 66.97 kb   | 29,229                 | C               | 2017         | Punicalagin                                        | [113]     |
| Euryale ferox           | 725.2 Mb    | 4.75 Mb    | 40,297                 | C               | 2020         | Sterols                                            | [77]      |
| Nymphaea colorata       | 409 Mb      | 2.1 Mb     | 31,580                 | C               | 2020         | Alkaloids                                          | [114]     |
| Aristolochia contorta  | 210.53 Mb   | 2.63 Mb    | 18,311                 | C               | 2022         | Benzyloisoquinoline alkaloids and aristolochic acids | [115]     |
| Piper nigrum            | 761.22 Mb   | NA         | 63,466                 | C               | 2019         | Morphine and codeine                               | [49]      |
| Coix lacryma-jobi       | 1.73 Gb     | 3.19 Mb    | 44,845                 | C               | 2020         | Coixol                                             | [116]     |
| Nelumbo nucifera        | 821.29 Mb   | 484.3 kb   | 32,124                 | C               | 2020         | Neferine, liensinine, isoliensinine, and nuciferine | [117]     |
| Macleaya cordata        | 378 Mb      | 25 kb      | 22,328                 | S               | 2017         | Sanguinarine and chelerythine                      | [119]     |
| Papaver somniferum      | 2.72 Gb     | 1.77 Mb    | 51,213                 | C               | 2018         | Morphine and codeine                               | [49]      |
| Coptis chinensis        | 936.60 Mb   | 806.6 kb   | 41,004                 | C               | 2021         | Protoberberine-type alkaloids                      | [120]     |
Table 1. Continued

| Species                  | Genome size | Contig N50 | No. of annotated genes | Assembly Levela | Release time | Primary medicinal ingredients | Reference |
|--------------------------|-------------|------------|------------------------|-----------------|--------------|-------------------------------|-----------|
| Cannabis sativa          | 876.15 Mb   | 1.96 Mb    | 33 674                 | C               | 2018         | Delta-9-tetrahydrocannabinolic acid and cannabidiolic acid [121] |
| Broussonetia papyrifera  | 386.83 Mb   | 171.2 kb   | 30 512                 | C               | 2019         | Saponin, vitamin B and oil [122] |
| Morus notabilis          | 335.39 Mb   | 5.72 Mb    | 26 010                 | C               | 2020         | Phenolic acids, flavonoids and alkaloids [123] |
| Ziziphus jujuba          | 437.65 Mb   | 34.0 kb    | 32 808                 | C               | 2014         | Ziziphus saponin [124] |
| Eriobotrya japonica      | 760.10 Mb   | 5.02 Mb    | 45 743                 | C               | 2020         | Volatile oil, amygda lin and ursolic acid [125] |
| Rosa chinensis           | 513.85 Mb   | 22.2 Mb    | 36 377                 | C               | 2018         | Gallic acid [126] |
| Boehmeria nivea          | 344.62 Mb   | 24.5 kb    | 30 237                 | S               | 2018         | Ramie acid [127] |
| Pistacia vera            | 671.28 Mb   | 714.2 kb   | 31 784                 | S               | 2019         | Phenolic compounds and vitamins [128] |
| Acer truncatum           | 633.28 Mb   | 773.17 kb  | 28 438                 | S               | 2020         | Unsaturated fatty acids (Nervonic acid) [129] |
| Xanthoceras sorbilolium  | 439.97 Mb   | 645.45 kb  | 21 059                 | C               | 2019         | Unsaturated fatty acids (Nervonic acid) [130] |
| Rhodiola crenulata       | 344.5 Mb    | 25.4 kb    | 31 517                 | S               | 2017         | Salidroside and tyrosol [131] |
| Ipomoea nil              | 734.8 Mb    | 1.87 Mb    | 42 783                 | C               | 2016         | Diterpenoids [132] |
| Datura stramonium        | 1.29 Gb     | 58.2 kb    | 30 934                 | S               | 2021         | Scopolamine and atropine [133] |
| Wurfalnia villosa        | 2.80 Gb     | 9.13 Mb    | 42 588                 | C               | 2022         | Bornyl acetate, borneol, and camphor [134] |
| Zingiber officinale      | 1.53 Gb     | 4.68 Mb    | 39 217                 | C, H            | 2021         | Gingerols, ginderdiols, zingerone, paradols, and shogao [135] |
| Selaginella tamariscina  | 300.73 Mb   | 2.14 kb    | 27 761                 | S               | 2018         | Selaginellins and amentoflavone [136] |
| Ginkgo biloba            | 9.87 Gb     | 1.58 Mb    | 27 832                 | C               | 2021         | Ginkgolide and bilobalide [137] |
| Gnetum montanum          | 4.07 Gb     | 475.2 kb   | 27 491                 | S               | 2018         | Stilbenoids [138] |
| Taxus chinensis          | 10.23 Gb    | 2.44 Mb    | 42 746                 | C               | 2021         | Paclitaxel [139] |
| Taxus wushiiiana         | 10.9 Gb     | 8.6 Mb     | 44 008                 | C               | 2021         | Paclitaxel [55] |

*Genome assembly level. C, chromosome level; S, scaffold level; H, haplotype-resolved genome assembly; NA: not reported

(QTLs) associated with PNP content. For example, Fan et al. assembled a chromosome-level genome of medicinal plant Panax notoginseng and performed GWAS on 240 cultivars to identify genes linked to dry root weight and stem thickness [59]. The availability of high-quality reference genomes provides critical resources to the elucidation of biosynthetic pathways of high-value PNPs in medicinal plants.

**Discovery of metabolic gene clusters through genomic mining**

Microbial genes controlling secondary metabolite biosynthesis are typically clustered in certain genomic regions, known as metabolic gene clusters (MGCs). Unlike microbes, plant biosynthesis genes are usually dispersed throughout the genome and MGCs have long been considered rare in plants. However, MGCs have recently been identified in several plants including Arabidopsis [140, 141], rice [142], maize [143], and several medicinal plants such as opium poppy [49] and Taxus [55, 139]. The MGCs contain at least three non-homologous genes, ranging from tens to several hundred kilobases in total length. To date, over 30 MGCs in plants have been reported to encode PNP biosynthetic pathways based on experimental evidence [143], encoding terpenoids [140–142], alkaloids [49], steroidal glycoalkaloids [144], and fatty acids [145]. Some of these PNPs encoded by gene clusters have medicinal values such as morphine and noscapine, while many have known roles in antimicrobial, allelopathic activity, and plant defense against herbivores and pathogens. Many of these functionally characterized MGCs have been discovered even before a reference genome is available. In the post-genomic era, the availability of reference genomes expedited identification of MGCs in plants using genomic mining. Recently, computer algorithms such as Plantismash [146], Phytoclust [147] and PlantClusterFinder [148] have been developed to predict MGCs encoding potential biosynthetic pathways of secondary metabolites. A large number of potential MGCs have been found in plants encoding unknown biosynthetic pathways, highlighting the limitation of our knowledge of what and how plants can produce chemically. For example, genome mining of P. somniferum genome has revealed 84 MGCs, among which one cluster encoding the pathway for morphinan and noscapine has been functionally validated [49, 149]. Tomato (Solanum lycopersicum) has 47 predicted MGCs, four of which have been associated with alpha-tomatine [144], lycosantalonol [150], fatty acids [145], and hydroxycinnamic acid amide [151] biosynthesis. A six-gene cluster for taxadiene biosynthesis was recently identified in the Taxus genome, involving in the first two biosynthetic steps, which helps to decode the complete taxol biosynthesis in the
future [139]. Wheat genome mining combined with transcriptomic analysis has recently identified six pathogen-induced biosynthetic pathways encoded by MGCs, producing flavonoids and terpenes that could potentially be used as phytoalexins in disease control [152]. Combining transcriptomic and metabolic profiling data would be useful to link MGCs with particular metabolites, such as the targeting of a four-gene cluster with the falcarindiol biosynthesis in tomato [145], although such strategy faces the challenge of lacking co-expression in many MGCs.

Despite the genomic discovery of hundreds of plant MGCs, questions remain to be answered regarding this specialized genetic architecture. First, how did plants gain gene clusters during evolution? MGCs in bacterial and fungi are commonly formed through gene duplication, translocation, and horizontal gene transfer (HGT). Despite a few exceptions [153, 154], it is uncommon for plants to undergo HGT and there must be special mechanisms for MGC formation in plants. Recent genomic analyses [57, 155] have shown that structural variation events such as whole genome duplications, chromosome fission and fusion, gene duplication, translocation and loss have been implicated in the birth and evolution of biosynthesis gene clusters, supporting the theory of punctuated evolution in forming plant specialized metabolites. Overall, investigation of plant MGC evolution remains at the infant stage, requiring comparative analysis of a large number of plant genomes. It will help us understand the major driving forces of PNP evolution and its ecological impact in nature.

Second, how do many of the MGCs actually contribute to PNP biosynthesis? As more candidate MGCs continue to be identified from plant genomes via an in silico approach, it is critical to functionally characterize predicted MGCs of unknown function and associate them with potential natural products. Expressing the candidate MGCs in a heterologous host such as yeast or E. coli will be informative to determine the synthesized product, as shown by several recent examples [151, 156]. For instance, Kong et al. used yeast as a chassis to investigate the function of a tomato gene cluster, and discovered a novel naringenin chalcone synthase responsible for the production of dihydro-coumarolyl anthranilate amide [151]. The caveat of this approach is that most predicted plant MGCs are quite large (up to hundreds of kb), presenting a huge challenge to cloning them into bacterial or yeast expression vectors. The plant MGCs successfully cloned and expressed in microbial hosts are mostly mini gene clusters of several kb containing only a few open reading frames. In addition, expressing plant proteins in yeast or E. coli cells does not always work due to different codon usage and post-translational modifications in eukaryotic versus prokaryotic cells. Alternatively, validating using plant host such as Nicotiana benthamiana enables expression of candidate MGCs via Agrobacterium infiltration of plant leaves or cells, followed by metabolomic detection, although it faces the same problem of delivering and expressing long MGC fragment into tobacco cells. Recently, a platform for high-throughput secondary metabolite discovery has been developed for filamentous fungi by cloning and expressing fragmented genomic DNAs containing MGCs into fungal artificial chromosomes followed by metabolomic profiling [157]. Application of this or a similar approach has not been reported in plants considering the low genomic fraction of plant MGCs and lack of a proper artificial chromosome cloning system. A high-throughput MGC validation platform will expedite the identification and utilization of novel PNP.

Bioengineering of PNP through plant biotechnology

In nature, PNPs are accumulated at low abundance and only in specific tissues and developmental stages of medicinal plants for two major reasons. Firstly, biosynthesis of these compounds consumes energy and competes with normal plant vegetative growth and reproduction. Secondly, most PNPs have cell toxicity from which plants have to protect themselves by detoxification, storing them in compartments, or only producing the toxic chemicals when and where needed. Evolutionarily, PNPs have probably undergone natural as well as human selection. For example, P. somniferum accumulates high levels of painkilling morphines in capsules instead of other tissues, and the amount of morphine produced differed among cultivars [158]. By contrast, its close relative Papaver rhaoeas only produces a trace amount of morphine [57]. This suggests that the ability to produce morphine has been under natural selection in poppy plants, and likely selected by domestication and breeding process.

The naturally low content of PNPs in medicinal herbs renders a major bottleneck in drug developments and clinical therapeutics. Structures of PNPs are often too complex for a profitable production by total chemical synthesis. Therefore, plant extraction remains the primary commercial source of most PNP for pharmaceuticals, causing over-exploitation of natural resources and instability to the Earth’s ecosystem. For instance, taxol, a well-known anticancer drug ingredient derived from the bark of yew tree once put the yew on the verge of extinction due to exhaustive exploitation. The demand of medical PNPs thus stimulates the breeding of superior germplasm resources for sustainable use of medicinal herbs. There are many challenges in breeding medicinal plants for high PNP yield, including limited understanding of how PNPs are exactly made and regulated by plants, lack of high-quality genome sequence, annotation and molecular markers, long breeding cycles as well as the difficulty of genetic transformation. Herb genomic research has accelerated the identification of functional genes and genome-wide molecular markers, linked molecular markers with desired characters, and improved breeding medicinal herbs. Many efforts have been made to increase PNP yield through plant breeding and biotechnological improvement, including artemisinin in A. annua [159, 160], morphine in P. somniferum [161], THC (tetrahydrocannabinol) in Cannabis sativa [162], etc. A. annua has been a primary source of the anti-malaria drug artemisinin. A. annua transcriptome sequencing enabled construction of genetic linkage map and identification of quantitative trait loci (QTL) that control artemisinin yield [160], providing genetic resources for molecular breeding. Phenotype selection coupled with molecular breeding in the past decade has led to the production of Artemisia F1 Seed (https://www.artemisiaf1seed.org), increasing the artemisinin yield from 5 kg per hectare to 55 kg per hectare with a 1.44% of dry weight [159]. To date, A. annua remains the sole source of artemisinin globally, although artemisinin metabolic engineering has been reported [163].

The rise of modern biotechnology provides a novel strategy for precise and expedited medicinal plant breeding. A key breakthrough is the revolutionary genome editing technology, most notably the CRISPR-cas9 system, that allows precise genome bases to obtain traits of interest at an unprecedented pace [164]. This biotechnology has powered the next generation of plant breeding to improve crop yield and quality such as PNP content. Unlike model and crop plants, the genome editing of medical herbs is still at the infant stage, hindered by the lack
of genomic information and a reliable genetic transformation system. Nevertheless, CRISPR-Cas9 based genome editing has been reported in several medicinal plants towards optimizing production of pharmacological components in *P. somniferum* [165], *S. miltiorrhiza* [166], *Dendrobium officinale* [167] and *Camelina sativa* [168]. Notably, genome editing successfully targeted three FAD2 (fatty acid desaturase 2) genes in allelohexaploid *Camelina sativa* and enhanced seed fatty acid levels [168]. CRISPR-Cas9 mediated gene deletion significantly decreased the benzyloxyquinoline alkaloid flux in transgenic opium poppies [165]. Plant genetic engineering offers clear advantages in improving the yield of PNPs [169] as plant chassis naturally carries many fundamental plant biosynthetic gene circuits, making them natural cell factories to produce PNPs of interest. A paradigmatic case is the Golden Rice [170], where the whole β-carotene biosynthetic pathway was introduced into rice endosperm using *Agrobacterium*-mediated co-transformation to generate rice plants with carotenoid content up to 1.6 mg/g in the endosperm. Moreover, a high-efficiency vector system was developed for transgene stacking to engineer anthocyanin biosynthesis in rice endosperm [171]. In addition, Zhu et al. [172] developed a novel method called combinatorial nuclear transformation to generate multiplex-transgenic plants allowing five carotenogenic genes to be simultaneously transferred into a white maize through biolistic transformation, resulting in transgenic plants with elevated levels of β-carotene.

Besides modifying existing metabolic pathways, genetically engineered plant chassis offers a cheap and sustainable source to produce high-value PNPs. The *de novo* production of PNPs in model organisms such as *Arabidopsis*, tobacco, tomato, and moss has progressed rapidly recently. For example, Fuentes et al. [173] transferred the entire artemisinic acid metabolic pathway from *A. annua* to tobacco chloroplast genome using combinatorial supertransformation of transplastomic recipient lines (COSTREL). Plants with high artemisinic acid levels were then isolated through screening large populations of transplastomic lines. In addition, strategies to increase terpenoids including overexpression of rate-limiting enzymes, chloroplast-compartmentalized engineering, and integration of transcription factors have been applied to the production of momilactone [174] and taxadiene [175] in tobacco. Momilactones are a group of diterpenes predominantly found in rice with an allelopathic activity. Through changing the subcellular localization of prenyltransferase and diterpene synthases, the diterpene biosynthesis was rerouted from chloroplast to cytosolic MEP pathway, significantly promoting the production of momilactone in *N. benthamiana* [175]. Noteworthy, this strategy also enabled the discovery of missing steps in momilactone B pathway, providing insights into pathway reconstitution and elucidation for desired products.

Metabolic engineering of an *in vitro* plant tissue culture, such as suspension cell culture and hairy root culture, is another efficient approach to yield valuable phytochemicals. Hairy roots are induced by *Agrobacterium rhizogenes* mediated transformation, which can be applied as high-capacity bioreactor to produce PNPs without the need of light and hormones. Hairy root cultures were successfully induced to overproduce cannabinoids in *C. sativa* [176], phytoestrogen and ginsenosides in *Panax ginseng* [177], and curcumin in *Atropa belladonna* [178]. Suspension cell culture has also made great advances to produce valuable PNPs with high yields. Plant Cell Fermentation (PCF® Technology [https://phytonbiotech.com/]) could produce natural taxol directly from plant cells of *Taxus chinensis* var. *maren*, while the metabolic engineered grapevine cells were able to produce resveratrol derivatives when elicited with MeJA and methylated cyclodextrins [179]. In the suspension cell culture system, the addition of heterologous elicitors induces the biosynthetic gene expression and increases the production of PNPs. Glandular trichomes are hairy structures differentiated from epidermal cells, featured by their enormous capacity to synthesize, store and secrete large quantities of metabolites with distinct types, making them an excellent platform for decoding the biosynthesis pathway of PNPs, and efficient phytochemical factories to produce PNPs [180]. Kortbeek et al. [181] engineered tomato glandular trichomes where a farnesyl diphosphate synthase was overexpressed, resulting in a decline of monoterpene production in the trichomes.

**Synthetic biology: a green revolution for PNP bioengineering and industrialization**

Although plant breeding can produce cultivars that accumulate higher level of metabolites than others do, it is still inefficient and environmentally unsustainable for massive production for commercial uses in most cases. Alternatively, synthetic biology, where microbial (e.g. yeast and bacteria) chassis are used to massively produce PNPs offers a more effective and environment-friendly alternative. Synthetic biology aims to design microbial cell factories carrying genetic circuits made of biosynthetic genes for heterologous production of PNPs. It has several advantages over plant-based extraction, such as circumventing the requirement of growing plants, rapid production and little interference from natural environment [182]. With the development of synthetic biology tools and knowledge of biosynthetic pathways, PNPs such as artemisinic acid [183], amorpha diene [163], taxadiene [36], cannabinoids [184], morphine [38], and noscapine [185] has been successfully produced using engineered microbial cells. A prominent example is the semi-synthesis of artemisinin [183], where an engineered amorpha diene-producing yeast [38] produces artemisinic acid with the titer of 25 g/L, later converted to artemisinin by chemical synthesis. This opens a route to the industrial production of artemisinin against the urgent demand of anti-malarial drugs. Parallelly, Luo et al. [184] partitioned the cannabinoid metabolic pathway into three modules: an engineered *S. cerevisiae* MEP pathway to make more flux into geranyl pyrophosphate, a hexanoyl-CoA biosynthetic pathway and several Cannabis genes to accumulate more olivetolic acid, as well as a heterologous downstream pathway to form the corresponding cannabinoids. The engineered yeast strains yielded 1.6 mg/L of cannabinoid from the simple sugar galactose, laying a foundation for the large-scale production of cannabinoids.

Despite progress, it remains challenging to use microbe as a chassis to synthesize high-value PNPs [182]. First, decoding biosynthetic pathways is the prerequisite for successful heterologous production but remains a challenge for vast majority of PNPs. For instance, previous attempts to reconstruct taxol pathways in microbe without knowing the steps converting downstream taxadiene to final taxol ended up with a production of taxadiene, the first committed intermediate [185]. The combination of bioinformatics and downstream functional validations in heterologous hosts to resolve the biosynthetic pathways of PNPs from sequencing data plays a key role in synthetic biology today (Fig 3). For example, the *Taxus* genome analysis identified a functional grouping of CYP725As and a taxadiene gene cluster, which will facilitate the future elucidation of taxol biosynthesis [55, 146].
Second, choosing and optimizing a microbial host is essential to maximize yield of PNP. E. coli and S. cerevisiae are two of the most widely used microorganisms to engineer biosynthetic pathways, given their fast growth, well-known genetic background and well-established genetic manipulation methods [36, 38, 163, 183–185]. However, it is challenging to express functional plant-derived cytochrome P450 genes in prokaryotic E. coli which lacks an intracellular organelle system, post-translational modification and electron transfer machinery, thus limiting its application in heterologous production of many PNP [36]. For example, E. coli was engineered to produce artemisinic acid with high titer of 20 g/L amorphadiene, achieving only 1 g/L artemisinic acid [186], much lower than the aforementioned 25 g/L artemisinic acid in S. cerevisiae. Apart from the commonly used E. coli and S. cerevisiae, nonconventional chassis cells are also used such as the industrial production of 4-hydroxybenzoic acid by C. glutamicum [187] and (+)-nootkatone by Pichia pastoris [39].

Third, reconstituting an efficient module is key to heterogeneous production of PNP. To construct heterogeneous expression cassettes, codon-optimization of plant genes is often required for microbial expression [36, 38]. Partition of complex biosynthetic pathways into several modules [36, 184] and the strains are optimized by synthetic biology tools.

Finally, the balance between microbial growth and PNP yield should be considered as many PNP and their intermediates, such as alkaloids and phenols, are toxic to the microorganisms. Taken together, with the rapid development of technologies including genome sequencing, multi-omic data mining, structure biology, directed evolution of proteins, and design of novel proteins, it is expected that the efficiency in synthesizing PNP in microorganisms will be greatly improved in the near future.

Opportunities and challenges for future PNP research

It is an exciting time to study PNP in this golden era of chemical biology. New technologies in both experimental and computational sciences are emerging at an unprecedented pace. These
### Table 2. A summary of transcription factors (TFs) regulating biosynthesis of plant natural products (PNP).

| PNP types   | Species                          | TFs                                                                 | Regulated genes                              | PNPAs                                      |
|-------------|---------------------------------|----------------------------------------------------------------------|-----------------------------------------------|--------------------------------------------|
| **Flavonoids** | Citrus reticulata cv. suavissima | CtrERF32 [188], CtrERF33 [188]                                       | CitCHL1                                      | Naringenin chalcone, (2S)-Naringenin       |
|             | Artemisia annua |                           |                                               | Red anthocyanin                            |
|             | Gentiana triflora |                           |                                               | Gentiodaphycin                             |
|             | Petunia hybrida |                           |                                               | Anthocyanin                                |
|             | Myrica rubra |                           |                                               | Anthocyanin                                |
|             | Epimedium sagittatum |                           |                                               | Anthocyanin, Kaempferol, Quercetin, MgA     |
|             | C. reticulata |                           |                                               | Andrographolactone                          |
|             | Ginkgo biloba |                           |                                               | Phenolic acids, Tanshinones                 |
|             | P. hybrid |                           |                                               | (2S)-Naringenin                            |
|             | M. rubra |                           |                                               | Red anthocyanin                            |
|             | A. majus |                           |                                               | Gentiodaphycin                             |
|             | Gentiana triflora |                           |                                               | Anthocyanin                                |
|             | Ipomoea nil |                           |                                               | Proanthocyanidin, Benzoic acids, Flavan-3-ols |
|             | Medicago truncatula |                           |                                               | Anthocyanin                                |
|             | *i. nil* |                           |                                               | Artemisinin                                |
| **Terpenoids** | Artemisia annua |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | Panax ginseng |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | A. annua |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | A. annua |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | A. annua |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | P. ginseng |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | Glycyrrhiza uralensis |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | Betula platyphylla |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | M. truncatula |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | A. annua |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | A. annua |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | P. ginseng |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | Mentha spicata |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | Salvia miltiorrhiza |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | Glycine max |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | Pogostemon cablin |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | P. cablin |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | Salvia miltiorrhiza |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |
|             | M. spicata |                           |                                               | Artemisinin, Artemisinic acid, Dihydroartemisinic acid, Ginsenoside Rb1, Ginsenoside Rg1, Ginsenoside Re | Artemisinin |

Continued
technologies have enabled the untangling the complex biosynthetic pathways and networks using systems biology approach, providing insight into the mechanistic and evolutionary mechanisms behind the chemical and structural diversity of natural products. Accurate and complete assembly of medicinal plant genomes is the key to understanding genome function and evolutionary patterns and improving plant traits through breeding and synthetic biology. Horticultural plant genomes are notoriously difficult to assemble, due to their wide range of genome sizes (up to hundreds of Gb), various ploidy levels, high repeat content and heterozygosity. Yet with genome technologies advancing quickly in recent years, chromosome-scale assembly, once a rarity, is now a realistic target for most genome projects. A typical genome project now adopts a combination of different technologies including long-read and short-read sequencing data, 10x Genomics data, Hi-C sequencing data or Bionano optical genome maps. Bioinformatic algorithms are continuing to be developed to solve complex problems of genome assembly and annotation leveraging these different technologies [249]. As a result, a growing number of medicinal plants now have chromosome-level genome assembly and some offer obvious improvement over previous versions of assembly and annotation [244, 73].

Despite the progress, a long journey is ahead to resolve the complete genome sequence of medicinal plants. Twenty years after the first plant reference genome was produced, the majority of plant reference genomes initially assembled using short reads remain unimproved. Recently, a telomere to telomere (T2T) genome assembly has been produced for a human cell line CHM13, resolving the nearly complete haploid genome sequence of human being [245]. The achievement is largely reliant on the use of high fidelity PacBio long-read sequencing data (HiFi reads), in addition to the ultra-long ONT reads and Hi-C reads. This is not just a milestone of human genomics, but also has a profound impact on animal and plant genomic research by kicking off an ambitious journey to resolve the complete genome sequence of all known living organisms on this planet. In fact, shortly after the publication of the human T2T genome assembly, the nearly complete genome sequence of A. thaliana [246, 247], and the gap-free reference genomes of rice [248] and watermelon [249] were reported. The nearly complete genome sequences of plants reveal tandem repeats of satellite regions located in centromeres, and find new genes that weren’t accessible in previous versions of assemblies. T2T plant genome sequences will make huge impact on plant biology allowing researchers to grasp a full complement of genetic elements associated with various traits including growth, development, and physiology. Another technical challenge for plant genome assembly is, instead of generating a collapsed diploid assembly, the ability to produce a haploid-resolved assembly that separates the parental and maternal haplotypes, also known as genome phasing. The phased assembly allows understanding of mechanisms behind heterosis and allele-specific gene regulation that contributes to many plant biological processes. To date there are only a handful of plant genomes that have been assembled and phased, including tea [250], lychee [251], and pear [252]. Commonly, genome phasing is conducted through trio-binning sequencing where genomes of parents are used to untangle the two haploid genomes of a child, although the information of parents is often unavailable for medicinal plants. In such cases, recently genome assembly methods such as hifiasm [253] allow for the resolution of haploids by using HiFi reads and Hi-C data without relying on sequencing data of family trios.

Unraveling the regulatory mechanisms underlying PNP biosynthesis is beneficial to improving the quality of traditional Chinese medicine through genetic breeding and metabolic engineering. Genomics and metabolomics of plant tissues have shown that the genes of biosynthetic pathways for PNFs are often co-expressed in specific tissues [49, 139]. So far it remains elusive how PNP biosynthetic enzymes and pathways are regulated in such a spatial and temporal fine-tuned manner. Studies regarding the molecular mechanisms of transcriptional regulation related to biosynthetic

### Table 2. Continued

| PNP types | Species | TFs | Regulated genes | PNs |
|-----------|---------|-----|-----------------|-----|
| Alkaloids | Eschscholzia californica | 20 EcAP2/ERF TFs [226] ORCA1/2 [227–229], CrERF5 [230], CrCR1 [231] | Ec6OMT, EcCYP719A5 STR, ASa, TDC, CPR, DXS, D4H, SLS, GES, SLS1, SGS, Redox1, SAT, PRX1, HLI1, G10H, DAT | Benzy1 isoquinoline alkaloids Vinodine, Catharanthine, Vinblastine, Vincristine, Secologanin, Ajmalicine, Anhydrovinblastine, Serpentine |
| Catharanthus roseus | NnRAV1 [232] Nn7OMT2 | NnTYDC1, NnCYP80G, Nn7OMT2 | NnCS1 TYDC, NCS, 60MT, CNMT, CYP80B2, 400MT, BBE, SMT, CYP719A1 | Benzy1 isoquinoline alkaloid isoquinoline alkaloid |
| Nelumbo nucifera | Coptis japonica | NhHLH1 [232] CjHLH1 [233] | NnTYDC1, NnCS1 TYDC, NCS, 60MT, CNMT, CYP80B2, 400MT, BBE, SMT, CYP719A1 | Terpenoid indole alkaloid Steroidal glycoalkaloids |
| C. roseus | C. roseus | CrMYC1, CrMYC2, CrBIS1, CrBIS2, CrBIS3 [234–236] | TDC, STR | CrBIS2, CrBIS3 [234–236] | Terpenoid indole alkaloid Terpenoid indole alkaloid |
| C. roseus | ZCT1 [237], ZCT2 [237], ZCT3 [237] | SSR2, C5-SD, HMGR, CAS | NnNCS1 | Berberine |
| C. roseus | Coptis japonica | CjWRKY1 [238] | TYDC, NCS, 60MT, CNMT, CYP80B2, 400MT, BBE, CYP719A1, SMT | Terpenoid indole alkaloid Terpenoid indole alkaloid |
| C. roseus | C. roseus | CrWRKY1 [239] | TDC, AS, DXS, SLS, SGD | CrWRKY1 [239] | Terpenoid indole alkaloid Terpenoid indole alkaloid |
| C. roseus | C. roseus | CrGBF1 [240], CrGBF2 [240] | STR | CrGBF1 [240], CrGBF2 [240] | Terpenoid indole alkaloid Terpenoid indole alkaloid |
| C. roseus | CrGATA1 [241] | T16H2, D4H, DAT, T3O, T3R | CrGATA1 [241] | Terpenoid indole alkaloid Terpenoid indole alkaloid |
| C. roseus | CrBIS1 [242] | D4H, DAT | CrBIS1 [242] | Terpenoid indole alkaloid Terpenoid indole alkaloid |
| N. nucifera | NnMYB8 [233], NnMYB12 [233], NnMYB113 [232] | NnTDC1, NnNCS1 | NnTDC1, NnNCS1 | Terpenoid indole alkaloid Terpenoid indole alkaloid |
| N. nucifera | NnRAV1 [232] | NnTDC1, NnNCS1 | NnTDC1, NnNCS1 | Terpenoid indole alkaloid Terpenoid indole alkaloid |
pathways of PNP s are still limited, most of which being focused on the specialized medicinal plants with well-known PNP s, high-quality genomic data, and an established transgenic system. The expression of PNP biosynthetic pathways is controlled by epigenetic mechanisms such as chromatin topology dynamics as shown by Nützmann et al. in model plant Arabidopsis [254]. Thus, it will be exciting to reveal what roles 3D genomic architecture and organization play in regulating biosynthesis of flavonoids, terpenoids, and alkaloids in medicinal plants. In addition, several transcription factors have been identified to regulate the process over the years (Table 2), offering targets to enhance PNP accumulation potentially via overexpression (activator) or silencing (repressor). However, it remains a challenge to stably transform most medicinal plants and obtaining transgenic plants often takes years even if it does show promise of significantly improving the PNP level. Another caveat is that expressing an excess amount of any epigenetic or transcriptional regulators, usually entangled in complex regulatory networks, can potentially lead to undesired traits. Therefore, a systems biology approach is essential to tease apart the PNP-specific gene circuits for precision modulation of target traits. Moreover, the picture is still a blur in cell heterogeneity that accounts for the cell-type specific expression of biosynthetic enzymes, transporters, and gene regulators. Recently, cutting-edge technologies such as single-cell transcriptomics (scRNA-seq) and spatial-transcriptomics have been widely applied to mammalian and plant tissues to generate a cell atlas and identify cell types within these tissues. The use of such technologies in PNP research has yet to be reported as of the date of this review being written, although spatial metabolomics has been reported to investigate the distribution of plant metabolites [255, 256]. It will be very interesting to identify the specific cell types expressing PNP biosynthetic genes using the scRNA-seq and spatio-transcriptomics, guiding a precise design and bioengineering of gene circuits for PNP improvement. The key challenge of the application of single-cell genomics in medicinal plants includes lack of high-quality reference genome and gene annotations, single-cell preparations, tissue and cell-type specific markers, and robust methods to integrative anlaysis of omic profiles at a single-cell level. Protocols and methods developed for human and mammalian samples are well in place but remain to be tested and optimized for medicinal plant studies. 

Last but not least, plant and microbe interactions also have a big influence on the profiles and abundance of PNP s. The importance of location and environment for growing TCM to their medicinal values and pharmaceutical properties has long been recognized and documented in traditional medicine scriptures. The difference could be down to a combination of factors such as the ecological environment, weather and perhaps the microbiota inhabiting in the soils where the TCM grows. Although the exact formula of microbial communities involved in modulating the PNP, as well as the mechanisms of regulation remain elusive, the association between microbiota and medicinal properties in plants has been suggested in several recent studies. An outstanding example of such association is reported in model plant A. thaliana which produces specific types of triterpenoids to electively modulate root bacteria in root microbiome [257]. Interestingly, isolation and re-inoculation of these bacteria in the roots can induce an increased production of these very PNP s. It remains to be studied how different types of microbes form microbial communities and signaling networks to interact with plants, either internally (endophytes) or externally (e.g. surface of roots and leaves), contribute to the accumulation of specific PNP s. Equally interesting is the regulation of plant microbiota by plant metabolites during plant–microbe interactions. Compared to model and crop plants, microbiome studies of medicinal plants have been quite limited. A combination of culture-based and culture-free microbiome analysis with functional metabolomics in medicinal plant rhizosphere and endophytes will help identify the association between plant microbiota and specialized metabolites. With this knowledge, it will be possible to modulate the production of special natural products in controlled settings such as greenhouse and plant factories in the future, much more efficient and effective than that which could be harvested from traditional authentic herb medicines.

Acknowledgements

This project is supported by the National Natural Science Foundation of China (Grant No. 31970317 and 32070368), Key Realm R&D Program of Guangdong Province, China (No. 2020B02022100), and the Natural Science Foundation of Shaanxi Province, China (No. 2020JZ-05). L. G. is also supported by a faculty startup package from Peking University Institute of Advanced Agricultural Sciences, and Yuandu Scholarship from municipal government of Weifang. The authors would like to thank anonymous reviewers for their comments and suggestions to improve the manuscript.

Conflict of interests

The authors declare that they have no conflict of interest.

References

1. Zaynab M, Fatima M, Abbas S et al. Role of secondary metabolites in plant defense against pathogens. Microb Pathog. 2018;124:198–202.
2. Adeleji AA, Babalola OO. Secondary metabolites as plant defensive strategy: a large role for small molecules in the near root region. Planta. 2020;252:61.
3. Li YQ, Kong D, Fu Y et al. The effect of developmental and environmental factors on secondary metabolites in medicinal plants. Plant Physiol Biochem. 2020;148:80–9.
4. Yang L, Wen KS, Ruan X et al. Response of plant secondary metabolites to environmental factors. Molecules. 2018;23:762.
5. Wang SC, Alseckh S, Fernie AR et al. The structure and function of major plant metabolite modifications. Mol Plant. 2019;12:899–919.
6. Marchioli R, dos Santos WD, Constantin RP et al. Biosynthesis and metabolic actions of simple phenolic acids in plants. Phytochem Rev. 2020;19:865–906.
7. Koeduka T. The phenylpropene synthase pathway and its applications in the engineering of volatile phenylpropanoids in plants. Plant Biotechnol (Tokyo). 2014;31:401–7.
8. Ye P, Liang S, Wang X et al. Transcriptome analysis and targeted metabolic profiling for pathway elucidation and identification of a geraniol synthase involved in iridoid biosynthesis from Gardenia jasminoides. Ind Crop Prod. 2019;132:48–58.
9. Šantić Ž et al. The historical use of medicinal plants in traditional and scientific medicine. Psychiatr Danub. 2017;29:787–92.
10. Lietava J et al. Medicinal plants in a middle Paleolithic grave Shanidar IV? J Ethnopharmacol. 1992;35:263–6.
11. Norn S, Kruse FR, Kruse E. History of opium poppy and morphine. Dan Medicinist Arbo. 2005;33:171–84.
12. Sanchez-Ramos JR et al. The rise and fall of tobacco as a botanical medicine. J Herb Med. 2020;22:100374.
13. Warner KE, MacKay J. The global tobacco disease pandemic: nature, causes, and cures. Glob Public Health. 2006;1:65–86.
14. Jurna I. Sertuerm und Morphin? Eine historische vignette. Schmerz. 2003;17:280–3.
15. Achan J, Talisuna AO, Erhart A et al. Quinine, an old antimarial drug in a modern world: role in the treatment of malaria. Malar J. 2011;10:144.
16. Desborough MJ, Keeling DM. The aspirin story - from willow to wonder drug. Br J Haematol. 2017;177:674–83.
17. World Health Organization. World Malaria Report 2010. Geneva: WHO; 2010.
18. Weaver BA. How taxol/paclitaxel kills cancer cells. Mol Biol Cell. 2014;25:2677–81.
19. Barnette KG, Gordon MS, Rodriguez D et al. Oral sabizabulin for high-risk, hospitalized adults with Covid-19: interim analysis. NEJM Evid. 2022;1:EVIDoa2200145.
20. Jorgensen SCJ, Kebriaei R, Dresser LD. Remdesivir: review of pharmacology, pre-clinical data, and emerging clinical experience for COVID-19. Pharmacotherapy 2020;40:659–71.
21. Mitjà O, Clotet B. Use of antiviral drugs to reduce COVID-19 transmission. Lancet Glob Health 2020;8:e639–40.
22. Wu CR, Liu Y, Yang Y et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. Acta Pharm Sin B. 2020;10:766–88.
23. Liang SJ et al. Study on flavonoid and bioactivity features of the pericarp of Citrus Reticulatae ‘Chachi’ during storage. Arab J Chem. 2021;15:103653.
24. Hu K, Guan WJ, Bi Y et al. Efficacy and safety of Lianhuaiqinweng capsules, a repurposed Chinese herb, in patients with coronavirus disease 2019: a multicenter, prospective, randomized controlled trial. Phytotherapy. 2021;85:153242.
25. Lichman BR. The scaffold-forming steps of plant alkaloid biosynthesis. Nat Pro Rep. 2021;38:103–29.
26. Yang CQ, Fang X, Wu XM et al. Transcriptional regulation of plant secondary metabolism. J Integr Plant Biol. 2012;54:703–12.
27. Benidir MAB, Kang KB, Genta-Jouve G et al. Advances in decomposing complex metabolite mixtures using substructure-and network-based computational metabolomics approaches. Nat Prod Rep. 2021;38:1967–93.
28. Bohlmann J, Meyer-Gaune G, Croteau R. Plant terpenoid synthases: molecular biology and phylogenetic analysis. PNAS. 1998;95:4126–33.
29. Attieh D, Djiana R, Koonjul P et al. Cloning and functional expression of two plant thiol ethyltransferases: a new class of enzymes involved in the biosynthesis of sulfur volatiles. Plant Mol Biol. 2002;50:511–21.
30. Oumaroon A, Decker G, Schmidt J et al. (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;36:808–19.
31. Isabel DP. Integration of deep transcriptome and proteome analyses reveals the components of alkaloid metabolism in opium poppy cell cultures. BMC Plant Biol. 2010;10:252.
32. Zeng JG, Liu Y, Liu W et al. Integration of transcriptome, proteome and metabolism data reveals the alkaloids biosynthesis in Macleaya cordata and Macleaya microcarpa. PLoS One. 2013;8:e53409.
33. Kellner F, Kim J, Clavijo BJ et al. Genome-guided investigation of plant natural product biosynthesis. Plant J. 2015;82:680–92.
34. Nett RS, Lau W, Sattely ES. Discovery and engineering of colchicine alkaloid biosynthesis. Nature. 2020;584:148–53.
62. Song X, Wang J, Li N et al. High quality genome of Erigeron breviscapus provides a reference for herbal plants in Asteraceae. Mol Ecol Resour. 2021;21:153–69.

57. Yang X, Gao S, Guo L et al. Three chromosome-scale Pupaver genomes reveal punctuated patchwork evolution of the morphinan and noscapine biosynthesis pathway. Nat Commun. 2021;12:6030.

58. Caputi L, Franke J, Farrow SC et al. Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. Science. 2018;360:1235–9.

59. Fan G, Liu X, Sun S et al. The chromosome level genome and genome-wide association study for the agronomic traits of Panax Notoginseng. iScience. 2020;23:101538.

60. Song X, Sun P, Yuan J et al. The celery genome sequence reveals sequential paleopolyploidizations, karyotype evolution and resistance gene reduction in apiales. Plant Biotechnol J. 2021;19:731–44.

61. Wang Y et al. Deletion and tandem duplications of biosynthetic genes drive the diversity of triterpenoids in Arula elata. Nat Commun. 2022;13:1–16.

62. Song X, Wang J, Li N et al. Deciphering the high-quality genome sequence of coriander that causes controversial feelings. Plant Biotechnol J. 2020;18:1444–56.

63. Wang Z et al. Reshuffling of the ancestral core-eudicot genome shaped chromatin topology and epigenetic modification in Panax. Nat Commun. 2022;13:1–12.

64. Sun X, Zhu S, Li N et al. A chromosome-level genome assembly of garlic (Allium sativum) provides insights into genome evolution and allicin biosynthesis. Mol Plant. 2020;13:1328–39.

65. Han B, Jing Y, Dai J et al. A chromosome-level genome assembly of dendrobium huoshanense using long reads and hi-C data. Genome Biol Evol. 2020;12:2486–90.

66. Niu Z, Zhu F, Fan Y et al. The chromosome-level reference genome assembly for Dendrobium officinale and its utility of functional genomics research and molecular breeding study. Acta Pharm Sin B. 2021;11:2080–92.

67. Bae E, An C, Kang MJ et al. Chromosome-level genome assembly of the fully mycorrhizotrophic orchid Gastrodia elata. G3. 2022;12:jkap433.

68. Fan W, Wang S, Wang H et al. The genomes of chicory, endive, great burdock and yacron provide insights into Asteraceae palaeopolyploidyization history and plant inulin production. Mol Ecol Resour. 2022;22:3124–40.

69. Liao B, Shen X, Xiang L et al. Allele-aware chromosome-level genome assembly of Artemisia annua reveals the correlation between ADS expansion and artemisinin yield. Mol Plant. 2022;15:1310–28.

70. Miao Y, Luo D, Zhao T et al. Genome sequencing reveals chromosome fusion and extensive expansion of genes related to secondary metabolism in Artemisia argyi. Plant Biotechnol J. 2022;20:1902–15.

71. Wu Z, Liu H, Zhan W et al. The chromosome-scale reference genome of safflower (Carthamus tinctorius) provides insights into linoleic acid and flavonoid biosynthesis. Plant Biotechnol J. 2021;19:1725–42.

72. Song C, Liu Y, Song A et al. The chrysanthenum nankingense genome provides insights into the evolution and diversification of chrysanthemum flowers and medicinal traits. Mol Plant. 2018;11:1482–91.

73. Jia Y, Chen S, Chen W et al. A chromosome-level reference genome of Chinese balloon flower (Platycodon grandiflorus). Front Genet. 2022;13:869784.

74. Kang M, Wu H, Yang Q et al. A chromosome-scale genome assembly of Isatis indigotica, an important medicinal plant used in traditional Chinese medicine. Hortic Res. 2020;7:18.

75. Zhang J, Tian Y, Yan L et al. Genome of plant maca (Lepidium meyenii) illuminates genomic basis for high-altitude adaptation in the Central Andes. Mol Plant. 2016;9:1066–77.

76. Li C et al. Assembly and annotation of a draft genome of the medicinal plant Polygonum cuspidatum. Front Plant Sci. 2019;10:1274.

77. Yang Y, Sun P, Lv L et al. Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. Nat Plants. 2020;6:215–22.

78. Xie D, Xu Y, Wang J et al. The wax gourd genomes offer insights into the genetic diversity and ancestral cucurbit karyotype. Nat Commun. 2019;10:5158.

79. Huang D, Ming R, Xu S et al. Chromosome-level genome assembly of Gynostemma pentaphyllum provides insights into gypenoside biosynthesis. DNA Res. 2021;28:dsab018.

80. Wu H, Zhao G, Gong H et al. A high-quality sponge gourd (Luffa cylindrica) genome. Hortic Res. 2020;7:128.

81. Cui J, Yang Y, Luo S et al. Whole-genome sequencing provides insights into the genetic diversity and domestication of bitter gourd (Momordica spp.). Hortic Res. 2020;7:85.

82. Xia M et al. Improved de novo genome assembly and analysis of the Chinese cucurbit Sitrata grosvenori, also known as monk fruit or luo-han-guo. Gigascience. 2018;7:gjy067.

83. Pu X, Li Z, Tian Y et al. The honeysuckle genome provides insight into the molecular mechanism of carotenoid metabolism underlying dynamic flower coloration. New Phytol. 2020;227:930–43.

84. Akagi T, Shirasawa K, Nagasaki H et al. The persimmon genome reveals clues to the evolution of a lineage-specific sex determination system in plants. PLoS Genet. 2020;16:e1008566.

85. Xia E, Zhang HB, Sheng J et al. The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. Mol Plant. 2017;10:866–77.

86. Mochida K, Sakurai T, Seki H et al. Draft genome assembly and annotation of Glycyrrhiza uralensis, a medicinal legume. Plant J. 2017;89:181–94.

87. Kang S et al. Genome-enabled discovery of anthraquinone biosynthesis in Semen tara. Nat Commun. 2020;11:1–11.

88. Qin S, Wu L, Wei K et al. A draft genome for Spatholobus suberectus. Sci Data. 2019;6:1–9.

89. Li Y, Wei H, Yang J et al. High-quality de novo assembly of the Eucommia ulmoides haploid genome provides new insights into evolution and rubber biosynthesis. Hortic Res. 2020;7:183.

90. Ooopes GM, Hamilton JP, Kim J et al. Genome assembly and annotation of the medicinal PlantCalotropis gigantea, a producer of anticancer and antimarial cardenolides. G3. 2018;8:385–91.

91. Liu Y, Tang Q, Cheng F et al. Whole-genome sequencing and analysis of the Chinese herbal plant Gelsemium elegans. Acta Pharm Sin B. 2020;10:574–82.

92. Xu Z et al. Tandem gene duplications drive divergent evolution of caffeine and crocin biosynthetic pathways in plants. BMC Biol. 2020;18:1–14.

93. Wang J, Xu S, Mei Y et al. A high-quality genome assembly of Morinda officinalis, a famous native southern herb in the Lingnan region of southern China. Hortic Res. 2021;8:135.

94. Rai A et al. Chromosome-level genome assembly of Ophiirrhiza pumila reveals the evolution of camptothecin biosynthesis. Nat Commun. 2021;12:1–19.
95. Liang Y, Chen S, Wei K et al. Chromosome level genome assembly of Andrographis paniculata. Front Genet. 2020;11:701.
96. Vining KJ, Johnson SR, Akhami A et al. Draft genome sequence of Mentha longifolia and development of resources for mint cultivar improvement. Mol Plant. 2017;10:323–39.
97. Bornowski N, Hamilton JP, Liao P et al. Genome sequencing of four culinary herbs reveals tetraploid genes underlying chemodiversity in the Nepetoideae. DNA Res. 2020;27:dsaa016.
98. Zhang Y et al. Incipient diploidization of the medicinal plant Perilla within 10,000 years. Nat Commun. 2021;12:1–13.
99. Shen Y, Li W, Zeng Y et al. Chromosome-level and haplotype-resolved genome provides insights into the tetraploid hybrid origin of patchouli. Nat Commun. 2022;13:1–19.
100. Zheng X, Chen D, Chen B et al. Insights into salvinian acid B biosynthesis from chromosome-scale assembly of the salvia boleyana genome. Acta Bot Sin. 2021;63:1309–23.
101. Jia K, Liu H, Zhang RG et al. Comparative genome analysis of Scutellaria baicalensis and Scutellaria barbata reveals the evolution of active flavonoid biosynthesis. Genom Proteom Bioinform 2020;18:230–40.
102. Li L, Cushman SA, He YX et al. Genome sequencing and population genomics modeling provide insights into the local adaptation of weeping wosyntia. Hortic Res. 2020;7:130.
103. Li M, Zhang D, Gao Q et al. Genome structure and evolution of Antirrhinum majus L. Nat Plants. 2019;5:174–83.
104. Shang J et al. The chromosome-level wintersweet (Chimonanthus praecox) genome provides insights into floral scent biosynthesis and flowering in winter. Genome Biol. 2020;21:1–28.
105. Chaw SM, Liu YC, Wu YW et al. Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution. Nat Plants. 2019;5:63–73.
106. Li J, Lv M, Du L et al. An enormous Paris polyphylla genome sheds light on genome size evolution and polyphenyl biogenesis. BioRxiv 2020. preprint: not peer reviewed.
107. Chan AP, Crabtree J, Zhao Q et al. Draft genome sequence of the oilseed species Ricinus communis. Nat Biotechnol. 2010;28:951–6.
108. Zhou W et al. Whole-genome sequence data of Hypericum perforatum and functional characterization of melatonin biosynthesis by N-acetylserotonin O-methyltransferase. J Pinel Res. 2021;70:e12709.
109. Wang Z, Hobson N, Galindo L et al. The genome of flax (Linum usitatissimum) assembled de novo from short shotgun sequence reads. Plant J. 2012;72:461–73.
110. Xia Z, Huang D, Zhang S et al. Chromosome-scale genome assembly provides insights into the evolution and flavor synthesis of passion fruit (Passiflora edulis Sims). Hortic Res. 2021;8:14.
111. Nong W, Law STS, Wong AYP et al. Chromosomal-level reference genome of the incense tree Aquilaria sinensis. Mol Ecol Resour. 2020;20:971–9.
112. Qin G, Xu C, Ming R et al. The pomegranate (Punica granatum L.) genome and the genomics of punicalagin biosynthesis. Plant J. 2017;91:1108–28.
113. Zhang L, Chen F, Zhang X et al. The water lily genome and the early evolution of flowering plants. Nature. 2020;577:79–84.
114. Cui X, Meng F, Fan X et al. Chromosome-level genome assembly of Aristolochia contorta provides insights into the biosynthesis of benzylisoquinoline alkaloids and aristolochic acids. Hortic Res. 2022;9:uhac005.
115. Hu L, Xu Z, Wang M et al. The chromosome-scale reference genome of black pepper provides insight into piperine biosynthesis. Nat Commun. 2019;10:4702.
116. Guo C, Wang Y, Yang A et al. The Coix genome provides insights into Panicoideae evolution and papery hull domestication. Mol Plant. 2020;13:399–413.
117. Shi T, Rahmani RS, Gugger PF et al. Distinct expression and methylation patterns for genes with different fates following a single whole-genome duplication in flowering plants. Mol Biol Evol. 2020;37:2394–413.
118. Liu X, Liu Y, Huang P et al. The genome of medicinal plant Macleaya cordata provides new insights into benzylisoquinoline alkaloids metabolism. Mol Plant. 2017;10:975–89.
119. Liu Y et al. Analysis of the Coptis chinensis genome reveals the diversification of protoberberine-type alkaloids. Nat Commun. 2021;12:1–13.
120. Nong W, Law STS, Wong AYP et al. Genome sequencing and analysis of the Japanese morning glory Ipomoea nil. Mol Plant. 2019;12:661–77.
121. Grasa CJ, Wenger JP, Dabney C et al. A complete cannabis chromosome assembly and adaptive admixture for elevated cannabidiol (CBD) content. New Phytol. 2021;230:1665–79.
122. Peng X, Liu H, Chen P et al. A chromosome-scale genome assembly of paper mulberry (Broussonetia papyrifera) provides new insights into its forage and papermaking usage. Mol Plant. 2019;12:661–77.
123. Xuan Y, Ma B, Li D et al. Chromosome restructuring and number change during the evolution of Morus notabilis and Morus alba. Hortic Res. 2022;9:uhab030.
124. Liu M, Zhao J, Cai QL et al. The complex jujube genome provides insights into fruit tree biology. Nat Commun. 2014;5:5315.
125. Jiang S, An H, Xu F et al. Chromosome-level genome assembly and annotation of the loquat (Eriobotrya japonica) genome. GigaScience. 2020;9:giaa015.
126. Raymond O, Gouzy J, Just J et al. The Rosa genome provides new insights into the domestication of modern roses. Nat Genet. 2018;50:772–7.
127. Luan M, Jian JB, Chen P et al. Draft genome sequence of ramie, Boehmeria nivea (L.) Gauchid. Mol Eco Resour. 2018;18:639–45.
128. Zeng L, Tu XL, Dai H et al. Whole genomes and transcriptomes reveal adaptation and domestication of pistachio. Genome Biol. 2019;20:79.
129. Ma Q, Sun T, Li S et al. The Acer truncatum genome provides insights into nervonic acid biosynthesis. Plant J. 2020;104:662–78.
130. Liang Q, Li H, Li S et al. The genome assembly and annotation of yellowhorn (Kantheroceras sorbifolium Bunge). GigaScience. 2019;8:giz071.
131. Fu Y, Li L, Hao S et al. Draft genome sequence of the Tibetan medicinal herb Rhodiola crenulata. GigaScience. 2017;6:1–5.
132. Meekal A, Jayakumar V, Nitasaka E et al. Genome sequence and analysis of the Japanese morning glory Ipomoea nil. Nat Commun. 2016;7:13295.
133. De-la-Cruz IM, Hallab A, Olives-Pinto U et al. Genomic signatures of the evolution of defence against its natural enemies in the poisonous and medicinal plant Datura stramonium (Solanaceae). Sci Rep. 2021;11:882.
134. Yang P, Zhao HY, Wei JS et al. Chromosome-level genome assembly and functional characterization of terpene synthases provide insights into the volatile terpenoid biosynthesis of Wurfbainia villosa. Plant J. 2022.
135. Li H, Wu L, Dong Z et al. Haplotypel-resolved genome of diploid ginger (Zingiber officinale) and its unique gingerol biosynthetic pathway. Hortic Res. 2021;8:189.
136. Xu Z, Xin T, Bartels D et al. Genome analysis of the ancient tracheophyte Selaginella tamariscina reveals evolutionary features relevant to the acquisition of desiccation tolerance. Mol Plant. 2018;11:983–94.

137. Liu H, Wang X, Wang G et al. The nearly complete genome of Ginkgo biloba illuminates gymnosperm evolution. Nat Plants. 2021;7:748–56.

138. Wan T, Liu ZM, Li LF et al. A genome for gnetophytes and early evolution of seed plants. Nat Plants. 2018;4:82–9.

139. Xiong X, Gou J, Liao Q et al. The Taxus genome provides insights into paclitaxel biosynthesis. Nat Plants. 2021;7:1026–36.

140. Field B, Osbourn AE. Metabolic diversification-independent assembly of operon-like gene clusters in different plants. Science. 2008;320:543–7.

141. Field B, Fiston-Lavier AS, Kemen A et al. Formation of plant metabolic gene clusters within dynamic chromosomal regions. PNAS. 2011;108:16116–21.

142. Mao LF, Kawaide H, Higuchi T et al. Functional genomics for convergent evolution of gene clusters for moramilactone biosynthesis in land plants. PNAS. 2020;117:12472–80.

143. Frey M, Chomet P, Glawischnig E et al. Analysis of a chemical plant defense mechanism in grasses. Science. 1997;277:696–9.

144. Itkin M, Heinig U, Tzfadia O et al. Biosynthesis of antinutritional alkaloids in Solanaceae crops is mediated by clustered genes. Science. 2013;341:175–9.

145. Jeon JE, Kim JG, Fischer CR et al. A pathogen-responsive gene cluster for highly modified fatty acids in tomato. Cell. 2020;180:176–187.e19.

146. Kautsar SA, Suarez Duran HG, Blin K et al. PlantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters. Nucleic Acids Res. 2017;45:W55–63.

147. Töpfer N, Fuchs LM, Aharoni A. The PhytoClust tool for metabolic gene clusters discovery in plant genomes. Nucleic Acids Res. 2017;45:7049–63.

148. Schläpfer P, Zhang P, Wang C et al. Genome-wide prediction of metabolic enzymes, pathways, and gene clusters in plants. Plant Physiol. 2017;173:2041–59.

149. Winzer T, Gazda V, He Z et al. A Papaver somniferum 10-gene cluster for synthesis of the anticancer alkaloid noscapine. Science. 2012;336:1704–8.

150. Matsuba Y, Zl J, Jones AD et al. Biosynthesis of the diterpenoid lycosantalanol via nerylerylnyl diphasate in Solanum lycopersicum. PloS One. 2015;10:e0119302.

151. Kong D, Li S, Smolke CD. Discovery of a previously unknown biosynthetic capacity of norariginen chinale synthase by heterologous expression of a tomato gene cluster in yeast. Sci Adv. 2020;6:eaab1143.

152. Polturak G, Dippe M, Stephenson MJ et al. Pathogen-induced biosynthetic pathways encode defense-related molecules in bread wheat. PNAS. 2022;119:e2123299119.

153. Wu D, Jiang B, Ye CY et al. Horizontal transfer and evolution of the biosynthetic gene cluster for benzoazinoids in plants. Plant Commun. 2022;3:100320.

154. Ma J, Wang S, Zhu X et al. Major episodes of horizontal gene transfer drove the evolution of land plants. Mol Plant. 2022;15:857–71.

155. Fan P, Wang P, Lou YR et al. Evolution of a plant gene cluster in Solanaceae and emergence of metabolic diversity. elife. 2020;9:e56717.

156. Kotopka BJ, Smolke CD. Production of the cyanogenic glycoside dhurrin in yeast. Metab Eng Commun. 2019;9:e00092.

157. Cleverenger KD, Bok JW, Ye R et al. A scalable platform to identify fungal secondary metabolites and their gene clusters. Nat Chem Biol. 2017;13:895–901.

158. Li Q, Ramasamy S, Singh P et al. Gene clustering and copy number variation in alkaloid metabolic pathways of opium poppy. Nat Commun. 2020;11:190.

159. Wetzstein HY, Porter JA, Janick J et al. Selection and clonal propagation of high artemisinin genotypes of Artemisia annua. Front Plant Sci. 2018;9:358.

160. Graham IA, Besser K, Blumer S et al. Molecular genetic diversity and association mapping of morphine content and agronomic traits in Turkish opium poppy (Papaver somniferum) germplasm. Mol Breeding. 2016;36:66.

161. Naim-Feil E et al. The characterization of key physiological traits of medicinal cannabis (Cannabis sativa L.) as a tool for precision breeding. BMC Plant Biol. 2021;21:1–15.

162. Westfall PJ, Pitera DJ, Lenihan JR et al. Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. PNAS. 2012;109:E111–8.

163. Celik I, Camci H, Kose A et al. Molecular genetic diversity and association mapping of morphine content and agronomic traits in Turkish opium poppy (Papaver somniferum) germplasm. Mol Breeding. 2016;36:66.

164. Gao CX. Genome engineering for crop improvement and future agriculture. Cell. 2021;184:1621–35.

165. Alagoy Y et al. Manipulating the biosynthesis of bioactive compound alkaloids for next-generation metabolic engineering in opium poppy using CRISPR-Cas 9 genome editing technology. Sci Rep. 2016;6:1–9.

166. Li B, Cui G, Shen G et al. Targeted mutagenesis in the medicinal plant salvia miltiorrhiza. Sci Rep. 2017;7:1–9.

167. Kui L, Chen H, Zhang W et al. Building a genetic manipulation tool box for orchid biology: identification of constitutive promoters and application of CRISPR/Cas9 in the orchid, Dendrobium officinale. Front Plant Sci. 2017;8:2036.

168. Jiang WZ, Henry IM, Lynagh PG et al. Significant enhancement of fatty acid composition in seeds of the alloxehaploid, Camelina sativa, using CRISPR/Cas9 gene editing. Plant Biotechnol J. 2017;15:648–57.

169. Zhu X, Liu X, Liu T et al. Synthetic biology of plant natural products: from pathway elucidation to engineered biosynthesis in plant cells. Plant Commun. 2021;2:100229.

170. Ye XD, al-Bahili S, Klöti A et al. Engineering the provitamin a (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science. 2000;287:303–5.

171. Zhu Q, Yu S, Zeng D et al. Development of “purple endosperm rice” by engineering anthocyanin biosynthesis in the endosperm with a high-efficiency transgene stacking system. Mol Plant. 2017;10:918–29.

172. Zhu C, Naqvi S, Breitenbach J et al. Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. PNAS. 2008;105:18232–7.

173. Fuentes P, Zhou F, Erban A et al. A new synthetic biology approach allows transfer of an entire metabolic pathway from a medicinal plant to a biomass crop. elife. 2016;5:e13664.

174. De La Peña R, Sattely ES. Rerouting plant terpene biosynthesis enables morphilate pathway elucidation. Nat Chem Biol. 2021;17:205–12.

175. Li J, Mutanda I, Wang K et al. Chloroplastic metabolic engineering coupled with isoprenoid pool enhancement for committed taxanes biosynthesis in Nicotiana benthamiana. Nat Commun. 2019;10:4850.
Horticulture Research, 2022, 9: uhac223

176. Farag S, Kayser O. Cannabinoids production by hairy root cultures of Cannabis sativa L. J Plant Sci USA. 2015;6:1874–84.

177. Kim YK, Kim YB, Uddin MR et al. Enhanced triterpene accumulation in Panax ginseng hairy roots overexpressing mevalonate-5-pyrophosphate decarboxylase and farnesyl pyrophosphate synthase. ACS Synth Biol. 2014;3:773–9.

178. Singh S, Pandey P, Akhtar MQ et al. A new synthetic biology approach for the production of curcumin and its glucoside in Atropa belladonna hairy roots. J Biotechnol. 2021;328:23–33.

179. Martínez-Márquez A, Morante-Carriel JA, Ramírez-Estrada K et al. Production of highly bioactive resveratrol analogues pterostilbene and piceatannol in metabolically engineered grapevine cell cultures. Plant Biotechnol J. 2016;14:1813–25.

180. Schuurink R, Tissier A. Glandular trichomes: micro-organs with model status? Neu Physiol. 2020;225:2251–66.

181. Kortbeek RW, Xu J, Ramírez A et al. Engineering of tomato glandular trichomes for the production of specialized metabolites. Method Enzymol. 2016;576:305–31.

182. Xu X, Liu Y, Du G et al. Microbial chassis development for natural product biosynthesis. Trends Biotechnol. 2020;38:779–86.

183. Paddon CJ, Westfall PJ, Pitera DJ et al. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature. 2013;496:528–32.

184. Luo X, Reiter MA, d’Espaux L et al. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. Nature. 2019;567:123–6.

185. Li Y, Li S, Thodey K et al. Complete biosynthesis of noscapine and halogenated alkaloids in yeast. PNAS. 2018;115:E3922–31.

186. Tsuruta H, Paddon CJ, Eng D et al. High-level production of amorpha-4,11-diene, a precursor of the antimalarial agent artemisinin, in Escherichia coli. Plant Physiol. 2009;149:4489.

187. Osorio-Montalvo P, Sáenz-Carbonell L, de-la-Peña C. 5-Azacytidine: a promoter of epigenetic changes in the quest to improve plant somatic embryogenesis. Int J Mol Sci. 2018;19:3182.

188. Zhao C, Liu X, Gong Q et al. Three AP2/ERF family members modulate flavonoid synthesis by regulating type IV chalcone isomerase in citrus. Plant Biotechnol J. 2021;19:671–88.

189. Goodrich J, Carpenter R, Coen ES. A common gene regulates pigmentation pattern in diverse plant species. Cell. 1992;68:955–64.

190. Nakatsuka T, Haruta KS, Pitaksutheepong C et al. Identification and characterization of R2R3-MYB and bHLH transcription factors regulating anthocyanin biosynthesis in gentian flowers. Plant Cell Physiol. 2008;49:1818–29.

191. Spell C, Quattrocorcio F, Mol JNM et al. anthocyanin1of petunia encodes a basic helix-loop-helix protein that directly activates transcription of structural anthocyanin genes. Plant Cell. 2000;12:1619–31.

192. Liu X, Yin XR, Allan AC et al. The role of MrbHLH1 and MrMYB1 in regulating anthocyanin biosynthetic genes in tobacco and Chinese bayberry (Myrica rubra) during anthocyanin biosynthesis. Plant Cell Tiss Org. 2013;115:285–98.

193. Huang W et al. A R2R3-MYB transcription factor regulates the flavonol biosynthetic pathway in a traditional Chinese medicinal plant, Epimedium sagittatum. Front Plant Sci. 2016;7:1089.

194. Liu C, Long J, Zhu K et al. Characterization of a citrus R2R3-MYB transcription factor that regulates the flavonol and hydroxycinnamic acid biosynthesis. Sci Rep. 2016;6:2535.

195. Xu F, Ning Y, Zhang W et al. An R2R3-MYB transcription factor as a negative regulator of the flavonoid biosynthesis pathway in Ginkgo biloba. Funct Integr Genomics. 2014;14:177–89.
228. van der Fits L, Memelink J. ORCA3, a jasmonate-responsive transcription factor in Arabidopsis thaliana. Plant J. 2017;90:520–34.

229. Liu T, Luo T, Guo X et al. PgMYB2, a MeJA-responsive transcription factor, positively regulates the dammarenediol synthase gene expression in Panax ginseng. Int J Mol Sci. 2019;20:2219.

230. Chotel G, Montiel G, Pré M et al. CrMYC1, a Catharanthus roseus elicitor- and jasmonate-responsive bHLH transcription factor that binds the G-box element of the strictosidines synthase gene promoter. J Exp Bot. 2003;54:2587–8.

231. Zhang J, Zhou L, Zheng X et al. Overexpression of SmMYB9b enhances tanshinone concentration in Salvia miltiorrhiza hairy roots. Plant Cell Rep. 2010;29:112672.

232. Zhou ML, Hou HL, Zhu XM et al. Soybean transcription factor GmMYBZ2 represses catharanthine biosynthesis in hairy roots of Catharanthus roseus. Appl Microbiol Biot. 2011;91:1095–105.

233. Chen X, Li J, Liu Y et al. PatSWC4, a methyl jasmonate-responsive MYB (v-myb avian myeloblastosis viral oncogene homolog)-related transcription factor, enhances tanshinone accumulation and decreases phenolic acid content in Salvia miltiorrhiza hairy roots. Sci Rep. 2017;7:5104.

234. Zhang J, Zhou L, Zheng X et al. Overexpression of SmMYB9b enhances tanshinone concentration in Salvia miltiorrhiza hairy roots. Plant Cell Rep. 2017;36:1297–309.

235. Zhao ML, Hou HL, Zhu XM et al. Soybean transcription factor GmMYBZ2 represses catharanthine biosynthesis in hairy roots of Catharanthus roseus. Appl Microbiol Biotechnol. 2011;91:1095–105.

236. Chen X, Li J, Liu Y et al. PatSWC4, a methyl jasmonate-responsive MYB (v-myb avian myeloblastosis viral oncogene homolog)-related transcription factor, positively regulates patchouliol biosynthesis in Pogostemon cablin. Ind Crop Prod. 2020;154112672.

237. Li J, Chen X, Zhou X et al. Identification of trihelix transcription factors in Pogostemon cablin reveals PatGT-1 negatively regulates patchouliol biosynthesis. Ind Crop Prod. 2021;161:113182.

238. Shi M, Zhou W, Zhang J et al. Methyl jasmonate induction of tanshinone biosynthesis in Salvia miltiorrhiza hairy roots is mediated by jasmonate zim-domain repressor proteins. Sci Rep. 2016;6:20919.

239. Zhou Y, Zhou X, Li Q et al. Molecular cloning, bioinformatics analysis, and transcriptional profiling of JA21 and JA22 from Salvia miltiorrhiza. Biotechnol Appl Bio. 2017;64:27–34.

240. Wang Q, Reddy VA, Panicker D et al. Metabolic engineering of terpene biosynthesis in plants using a trichome-specific transcription factor MsYABBYs from spearmint (Mentha spicata). Plant Biotechnol J. 2016;14:1619–32.

241. Yamada Y, Kokubu Y, Chaki K et al. Isoquinoline alkaloid biosynthesis is regulated by a unique bHLH-type transcription factor in Coptis japonica. Plant Cell Physiol. 2011;52:1131–41.

242. Chatel G, Montiel G, Pré M et al. CrMYC1, a Catharanthus roseus elicitor- and jasmonate-responsive bHLH transcription factor that binds the G-box element of the strictosidines synthase gene promoter. J Exp Bot. 2003;54:2587–8.

243. Zhang H, Hedhili S, Montiel G et al. The basic helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the ORCA genes that regulate alkaloid biosynthesis in Catharanthus roseus. Plant J. 2011;67:61–71.

244. Singh SK et al. bHLH iridoid synthesis 3 is a member of a bHLH gene cluster regulating terpenoid indole alkaloid biosynthesis in Catharanthus roseus. Plant Direct. 2021;5:e00305.

245. Pauw B, Hilliou FAO, Martin VS et al. Zinc finger proteins act as transcriptional repressors of alkaloid biosynthesis genes in Catharanthus roseus. J Biol Chem. 2004;279:52940–8.

246. Kato N, Dubouzet E, Kokubu Y et al. Identification of a WRKY protein as a transcriptional regulator of benzylisoquinoline alkaloid biosynthesis in Coptis japonica. Plant Cell Physiol. 2007;48:8–18.

247. Suttipanta N, Pattanaik S, Kulshrestha M et al. The transcription factor CrWRKY1 positively regulates the terpenoid indole alkaloid biosynthesis in Catharanthus roseus. Plant Physiol. 2011;157:1081–93.

248. Sibéri Y, Benhamron S, Memelink J et al. Catharanthus roseus G-box binding factors 1 and 2 act as repressors of strictosidine synthase gene expression in cell cultures. Plant Mol Biol. 2001;45:477–88.

249. Liu Y, Patra B, Pattanaik S et al. GATA and PIF transcription factors regulate light-induced biosynthesis of vindoline in Catharanthus roseus. Plant Physiol. 2011;157:1081–93.

250. Feher Y, Benhamron S, Memelink J et al. Catharanthus roseus G-box binding factors 1 and 2 act as repressors of strictosidine synthase gene expression in cell cultures. Plant Mol Biol. 2001;45:477–88.

251. Li Y, Patra B, Pattanaik S et al. GATA and PIF transcription factors regulate light-induced biosynthesis of vindoline in Catharanthus roseus. Plant Physiol. 2011;157:1081–93.

252. Li Y, Patra B, Pattanaik S et al. GATA and PIF transcription factors regulate light-induced biosynthesis of vindoline in Catharanthus roseus. Plant Physiol. 2011;157:1081–93.

253. Rice ES, Green RE. New approaches for genome assembly and scaffolding. Annu Rev Anim Biosci. 2019;7:17–40.

254. Wang J, Li J, Li Z et al. Genomic insights into longan evolution and domestication. Annu Rev Plant. 2022;9:uhac021.

255. Miga KH, Koren S, Rhie A et al. Telomere-to-telomere assembly of a complete human X chromosome. Nature. 2020;585:79–84.

256. Wang B, Yang X, Jia Y et al. High-quality Arabidopsis thaliana genome assembly with Nanopore and HiFi long reads. Genom Proteom Bioinf. 2022;20:4–13.

257. Naish M, Alonge M, Wlodzimierz P et al. Catharanthus roseus CrWRKY1 positively regulates the terpenoid indole alkaloid biosynthesis in Catharanthus roseus. Plant Physiol. 2011;52:1131–41.

258. Singh SK et al. bHLH iridoid synthesis 3 is a member of a bHLH gene cluster regulating terpenoid indole alkaloid biosynthesis in Catharanthus roseus. Plant Direct. 2021;5:e00305.
248. Song JM, Xie WZ, Wang S et al. Two gap-free reference genomes and a global view of the centromere architecture in rice. Mol Plant. 2021;14:1757–67.

249. Deng Y, Liu S, Zhangi Y et al. A telomere-to-telomere gap-free reference genome of watermelon and its mutation library provide important resources for gene discovery and breeding. Mol Plant. 2022;15:1268–84.

250. Zhou Z, Tan H, Li Q et al. Trichome and ARTEMISININ regulator 2positively regulates trichome development and artemisinin biosynthesis in Artemisia annua. New Phytol. 2020;228:932–45.

251. Chalvin C, Drevensek S, Dron M et al. Genetic control of glandular trichome development. Trends Plant Sci. 2020;25:477–87.

252. Li Y, Kong D, Bai M et al. Correlation of the temporal and spatial expression patterns of HQT with the biosynthesis and accumulation of chlorogenic acid in Lonicera japonica flowers. Hortic Res. 2019;6:73.

253. Cheng H, Concepcion GT, Feng X et al. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18:170–5.

254. Nützmann HW, Doerr D, Ramírez-Colmenero A et al. Active and repressed biosynthetic gene clusters have spatially distinct chromosome states. Proc Natl Acad Sci U S A. 2020;117:13800–9.

255. Susniak K, Krysa M, Gieroba B et al. Recent developments of MALDI MSI application in plant tissues analysis. Acta Biochim Pol. 2020;67:277–81.

256. Sun C, Ma S, Li L et al. Visualizing the distributions and spatiotemporal changes of metabolites in Panax notoginseng by MALDI mass spectrometry imaging. J Ginseng Res. 2021;45:726–33.

257. Huang AC, Jiang T, Liu YX et al. A specialized metabolic network selectively modulates Arabidopsis root microbiota. Science. 2019;364:eaau6389.