SCIENCE CHINA
Life Sciences

SPECIAL TOPIC: Herbal small RNA, novel precision medicine from Mother Nature
•REVIEW•

March 2019 Vol.62 No.3: 288–308
https://doi.org/10.1007/s11427-018-9352-7

Trends in herbgenomics

Tianyi Xin[1], Yu Zhang[1], Xiangdong Pu[1], Ranran Gao[1], Zhichao Xu[1*] & Jingyuan Song[1,2*]

[1]*Key Lab of Chinese Medicine Resources Conservation, State Administration of Traditional Chinese Medicine of the People's Republic of China, Institute of Medicinal Plant Development, Peking Union Medical College & Chinese Academy of Medical Sciences, Beijing 100193, China;*
[2]*Yunnan Branch, Institute of Medicinal Plant Development, Peking Union Medical College & Chinese Academy of Medical Sciences, Jinghong 666100, China*

Received March 16, 2018; accepted May 3, 2018; published online August 14, 2018

From Shen Nong's Herbal Classic (Shennong Bencao Jing) to the Compendium of Materia Medica (Bencao Gangmu) and the first scientific Nobel Prize for the mainland of China, each milestone in the historical process of the development of traditional Chinese medicine (TCM) involves screening, testing and integrating. After thousands of years of inheritance and development, herbgenomics (bencaogenomics) has bridged the gap between TCM and international advanced omics studies, promoting the application of frontier technologies in TCM. It is a discipline that uncovers the genetic information and regulatory networks of herbs to clarify their molecular mechanism in the prevention and treatment of human diseases. The main theoretical system includes genomics, functional genomics, proteomics, transcriptomics, metabolomics, epigenomics, metagenomics, synthetic biology, pharmacogenomics of TCM, and bioinformatics, among other fields. Herbgenomics is mainly applicable to the study of medicinal model plants, genomic-assisted breeding, herbal synthetic biology, protection and utilization of gene resources, TCM quality evaluation and control, and TCM drug development. Such studies will accelerate the application of cutting-edge technologies, revitalize herbal research, and strongly promote the development and modernization of TCM.

herbgenomics (bencaogenomics), traditional Chinese medicine, omics, practical application

Citation: Xin, T., Zhang, Y., Pu, X., Gao, R., Xu, Z., and Song, J. (2019). Trends in herbgenomics. Sci China Life Sci 62, 288–308. https://doi.org/10.1007/s11427-018-9352-7

Introduction

Traditional Chinese medicine (TCM) is not only an important part of excellent traditional Chinese culture but also the condensation of profound philosophical human ancestral wisdom. Reflecting thousands of years of experience in health preservation, TCM truly makes a great contribution to human life and health (Tu, 2011). With the development of the chemical and medicinal industries, the exploration of efficient components, monomer components and patent medicines of TCM has become a focus of TCM modernization. The remarkable achievement of Prof. Youyou Tu led her to win the 2015 Nobel Prize in Physiology or Medicine, an event that has become a milestone in TCM research. Currently, the recyclable and sustainable exploration and utilization of TCM have turned into leading issues, and re-investigating the biology of genuine TCM species will promote TCM research (Chen et al., 2015).

As a newly developing discipline, herbgenomics is a principle that uncovers genetic information and regulatory networks regarding herbs and clarifies their molecular mechanisms in the prevention and treatment of human diseases. Herbgenomics focuses on the study of medicinal organisms and their interactions in humans. This discipline includes herbal genomics, transcriptomics, functional genomics,

*Corresponding authors (Zhichao Xu, email: zcxu@implad.ac.cn; Jingyuan Song, email: songjingyuan@hotmail.com)

proteomics, metabolomics, epigenomics, metagenomics, and many other aspects (Figure 1). Herbgenomics systematically explores genes that are related to the synthesis of active compounds and have good agronomic characteristics, laying a foundation for the varietal improvement of "authentic" TCM and the conservation of gene resources (Chen et al., 2015). Establishment of a genome research system for herb species that contain important active ingredients will provide a theoretical basis for the recognition of TCM properties. The molecular basis of "authentic" TCM has been elucidated at the genome level, thus serving as an impetus for innovative TCM research and development, and offering technological support for secondary metabolite biosynthesis and metabolic engineering. In addition, herbgenomics will facilitate the elucidation of targets of herbal action and mechanisms in disease treatment and provide support for personalized precision medicine (Chen, 2017). In an amazing recent finding, a miRNA from honeysuckle (*Lonicera japonica*) could directly target influenza virus (Zhou et al., 2015), and a miRNA derived from *Rhodiola crenulata* could decrease the expression of fibrotic hallmark genes and proteins (Du et al., 2017). These findings indicate that active ingredients in TCM are not limited to small-molecule compounds, such as terpenoids, flavonoids, and alkaloids, but also include organic macromolecules, such as peptides and nucleic acids (Du et al., 2017; Mehta et al., 2002; Zhou et al., 2015).

In general, herbgenomics will greatly promote the application of frontier life science technologies in medicinal plants and TCM. Research regarding medicinal organisms and TCM will develop rapidly, eventually coming to the forefront of life science research.

Theoretical system of herbgenomics

Genomics

Eukaryotic genomes include three components: nuclear, chloroplast, and mitochondrial genomes. Genomics focuses on gene numbers, the linear distribution and distance of each gene on the chromosome, and the coding and structures of intergenic spacer genes. Genomics research is concentrated on the establishment of genetic maps, physical maps, sequence maps, and gene maps. For medicinal organisms, three main components of genomics research exist: protein-coding regions, non-coding regions, and repetitive sequences. A typical characteristic of medicinal organisms is that many belong to the same family, and even the same genus, which leads to their having similar medicinal functions. Comparing the genome structures of these organisms helps researchers identify similarities and dissimilarities among different gene families, understand the evolutionary history of species, and deduce gene functions. In the past 20 years, many methods of classifying gene families have been established based on nucleotide sequences. For instance, according to known sequence characteristics, such as motifs and domains, and clustering methods based on pairwise alignment, a phylogenetic tree can be constructed to deduce a gene family (Xu et al., 2016a).

Nuclear genomic sequences contain genetic information regarding species origin, evolution, growth and development, synthetic pathways for active compounds, and more. Since the first higher plant whole genome, *Arabidopsis thaliana*, was published in 2000 (Arabidopsis Genome Initiative, 2000), hundreds of land plants and algae have been or are being sequenced. The whole genomes of several

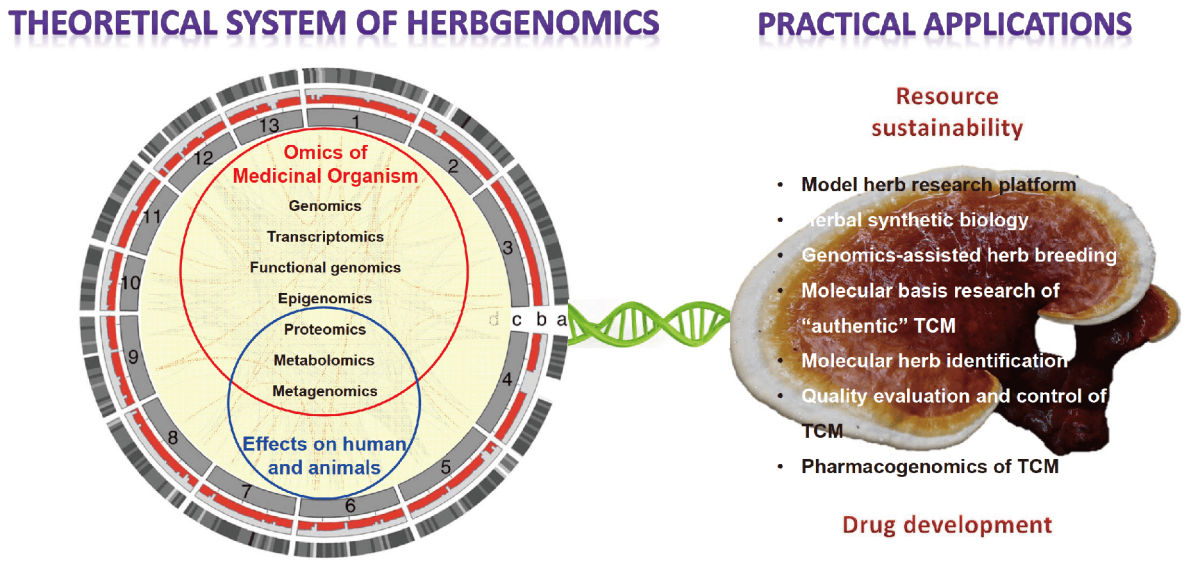

Figure 1 Theoretical system and practical applications of herbgenomics. The ideogram of genomic feature and the image of the herb *Ganoderma lucidum* are from Chen et al., 2012.

medicinal organisms, including *Ganoderma lucidum* (Chen et al., 2012), *Salvia miltiorrhiza* (Xu H. et al., 2016), *Panax notoginseng* (Chen G. et al., 2017; Zhang D. et al., 2017), *Panax ginseng* (Xu et al., 2017; Jayakodi et al., 2018), *Artemisia annua* (Shen et al., 2018), *Selaginella tamariscina* (Xu et al., 2018), and *Rosa chinensis* (Raymond et al., 2018) have been published (Table 1). Although these data are great resources for medicinal organism research, we still need to address the flaws in genome and transcriptome annotation, such as increasing the number of genes that can be annotated and remembering that the absence of gene annotation does not mean the absence of gene function. An automated annotation pipeline and a recipe and reference chart outlining the typical steps in gene annotation using publicly available resources was provided by Bolger et al. (Bolger et al., 2018).

The acquisition of whole genome information has greatly promoted basic research on medicinal organisms amd provided a foundation for researching the synthetic pathways of secondary metabolites and plant genome evolution. For instance, the elucidation of the *Ganoderma lucidum* genome made it a potential model organism for the research of secondary metabolic pathways and their regulation in medicinal fungi (Chen et al., 2012). Analysis of the *Salvia miltiorrhiza* genome sequence revealed four copalyl diphosphate synthase/cytochromes P450 (*CPS/CYP*) gene clusters along with other genes potentially encoding biosynthetic enzymes of tanshinone and further expected to aid molecular breeding (Xu H. et al., 2016). The *Panax* genomes provide knowledge on the biosynthesis of ginsenoside and ample genetic resources for identifying novel drug candidates in closely related *Panax* species (Chen W. et al., 2017; Zhang D. et al., 2017; Xu et al., 2017; Jayakodi et al., 2018). The expansion and functional diversification of genes in *Artemisia annua* encoding enzymes involved in terpene biosynthesis are consistent with the evolution of the artemisinin biosynthetic pathway (Shen et al., 2018). Genomic analyses highlight the unique evolution and complex regulation of the desiccation response in *Selaginella tamariscina* and provide significant insights into the desiccation tolerance mechanism and the evolution of resurrection plants (Xu et al., 2018). A genomic segment of Chinese ancestry in the *Rosa chinensis* genome identified new candidate genes for recurrent blooming and provided new insights into the domestication of modern roses (Raymond et al., 2018).

The diversity of active compounds in medicinal plants is derived from the complexity of plant genome evolution. In addition to biosynthetic pathways, whole genome duplication (WGD), alternative splicing, alternative untranslated regions (UTRs), and epigenetic control are also important parts of the evolutionary process. WGD has long been recognized as an important evolutionary force, especially in plants. Research conducted on *Panax notoginseng* showed that one round of WGD to have occurred ~26.15 million years ago and that this recent WGD event distinctly occurred in *Panax notoginseng* after the *Panax notoginseng*-kiwifruit divergence (Zhang D. et al., 2017). Alternative splicing is a regulated process during gene expression that can result in a single gene coding for multiple proteins. The use of SMRT long-read sequencing offered the ability to examine alternative splicing in *Salvia miltiorrhiza* root, which was found to occur in approximately 40% of the detected gene loci, including several involved in isoprenoid (terpenoid) metabolism (Xu et al., 2015). Alternative 3′ UTRs facilitate the formation of alternative protein complexes, which can perform alternative protein functions (Mayr, 2016). In addition, epigenetic pathways can regulate genome-wide changes in gene expression and then affect the cell-fate transition. Certain epigenetic factors may be recruited by transcription factors to relocate to new loci and regulate gene expression. Cross talk among hormone signaling, transcription factors and epigenetic factors is involved in different types of plant regeneration, suggesting that elegant and complex regulatory mechanisms control which type of regeneration is triggered in plants under different circumstances (Xu and Huang, 2014).

Chloroplasts are plastids at the center of photosynthesis, and they are generally considered to originate from ancient cyanobacteria. Chloroplast genomes are rarely linear and usually comprise circular DNA molecules, exhibiting multiple copies in cells with typical lengths of 120~160 kb and accounting for approximately 10%~20% of the total DNA in leaves. Chloroplast genomes are highly conserved and can be divided into four parts: large single copy (LSC), small single copy (SSC), and two inverted repeat (IR) sequences. By the end of September 2017, 1,823 chloroplast genomes, ranging in size from 19.4 to 610.06 kb, had been submitted to the National Center for Biotechnology Information (NCBI) database, 1,591 of which belong to land plants (https://www.ncbi.nlm.nih.gov/genome/browse/). Based on the characteristics of chloroplast genome structures and sequential conservatism, taxonomists widely use chloroplast genomes to study plant phylogenetics. Compared with traditional molecular markers, chloroplast genomes have longer sequence lengths and higher identification efficiencies and thus have the potential to become super DNA barcodes (Guo et al., 2017). The *de novo* assembly of the chloroplast genomes of three *Fritillaria* species using a single molecule, real-time (SMRT) circular consensus sequencing strategy and analysis of the phylogenetic relationship of Liliales, provided a useful exploration of super DNA barcode research (Li et al., 2014). The transformation of chloroplast genomes offers multiple advantages over that of nuclear genomes, showing great potential in terms of species improvement and biological agent production, among others (Adem et al., 2017). The availability of sequenced medicinal plant chloroplast genomes has enhanced our understanding of chloroplast biology,

Table 1 Published whole genomes of medicinal organisms

Kingdom	Phylum	Class	Order	Family	Species	Assembled genome size (Mb)	Sequencing platform	Reference
Fungi	Basidiomycota	Agaricomycetes	Polyporales	Ganodermataceae	*Ganoderma lucidum*	43.3	Roche 454 GS FLX, Illumina GAII	Chen et al., 2012
			Polyporales	Ganodermataceae	*Ganoderma sinensis*	48.96	Roche 454 GS FLX, Illumina HiSeq 2000	Zhu et al., 2015
Algae	Phaeophyta	Phaeophyceae	Laminariales	Laminariaceae	*Saccharina japonica*	537	Illumina HiSeq 2000, Roche 454 GS FLX, PacBio RS	Ye et al., 2015
Plantae	Pteridophyta	Lycophytina	Selaginellales	Selaginellaceae	*Selaginella tamariscina*	301	PacBio RSII platform	Xu et al., 2018
Plantae	Gymnospermae	Ginkgopsida	Ginkgoales	Ginkgoaceae	*Ginkgo biloba*	10,610	Illumina HiSeq 2000	Guan et al., 2016
	Angiospermae	Dicotyledoneae	Cucurbitales	Cucurbitaceae	*Momordica charantia*	339	Illumina MiSeq, HiSeq 2500	Urasaki et al., 2017
			Geraniales	Euphorbiaceae	*Ricinus communis*	350.6	ABI3730xl	Chan et al., 2010
			Myrtiflorae	Punicaceae	*Punica granatum*	328.38	Illumina HiSeq 2000	Qin et al., 2017
			Papaverales	Brassicaceae	*Lepidium meyenii*	743	Illumina HiSeq 2500	Zhang J. et al., 2016
			Papaverales	Brassicaceae	*Raphanus sativus*	402	Illumina HiSeq 2000	Kitashiba et al., 2014
			Papaverales	Papaveraceae	*Macleaya cordata*	378	Illumina HiSeq 2000	Liu et al., 2017
			Ranunculales	Nymphaeaceae	*Nelumbo nucifera*	929/792	Roche 454 GS FLX, Illumina HiSeq 2000	Ming et al., 2013; Wang et al., 2013
			Rhamnales	Rhamnaceae	*Ziziphus jujube*	437.65	Illumina HiSeq 2000	Liu et al., 2014
			Violales	Cucurbitaceae	*Cucumis melo*	375	Illumina GAIIx, Roche 454 GS FLX, and Sanger	Garcia-Mas et al., 2012
			Violales	Cucurbitaceae	*Citrullus lanatus*	353.5	Illumina GAII	Guo S. et al., 2013
			Rosales	Crassulaceae	*Rhodiola crenulata*	344.5	Illumina HiSeq 2000/ 4000	Fu et al., 2017
			Rosales	Rosaceae	*Prunus persica*	224.6	Sanger	Verde et al., 2013
			Rosales	Rosaceae	*Rosa chinensis*	503	Illumina HiSeq 2000, PacBio RSII platform	Raymond et al., 2018
			Fabales	Fabaceae	*Glycyrrhiza uralensis*	379	Illumina HiSeq 2000, PacBio RS II	Mochida et al., 2017
			Fabales	Fabaceae	*Glycine max*	950	Sanger	Schmutz et al., 2010
			Fabales	Fabaceae	*Vigna angularis*	466.7	Illumina HiSeq 2000	Yang K. et al., 2015
			Theineae	Theaceae	*Camellia sinensis*	3,020	Illumina HiSeq 2000	Xia et al., 2017
			Umbelliflorae	Araliaceae	*Panax ginseng*	3,430 /~3,000	Illumina HiSeq X-Ten Illumina HiSeq	Xu et al., 2017 Jayakodi et al., 2018
			Umbelliflorae	Araliaceae	*Panax notoginseng*	1,850/2,390	Illumina HiSeq 2000	Chen W. et al., 2017; Zhang D. et al., 2017
			Umbelliflorae	Nyssaceae	*Camptotheca acuminata*	384.5	Illumina HiSeq 2000	Zhao et al., 2017
			Urticales	Moraceae	*Cannabis sativa*	534	Roche 454 GS FLX, Illumina HiSeq	van Bakel et al., 2011
			Urticales	Moraceae	*Morus notabilis*	357.4	Illumina	He et al., 2013
			Geraniales	Linaceae	*Linum usitatissimum*	302	Illumina GAII	Wang et al., 2012
			Campanulales	Compositae	*Erigeron breviscapus*	1,200	Illumina HiSeq 2500, PacBio RS II	Yang et al., 2017
			Gentianales	Apocynaceae	*Catharanthas roseus*	523	Illumina HiSeq	Kellner et al., 2015

(*To be continued on the next page*)

(*Continued*)

Kingdom	Phylum	Class	Order	Family	Species	Assembled genome size (Mb)	Sequencing platform	Reference
			Solanales	Convolvulaceae	*Ipomoea nil*	734.8	Illumina HiSeq 2000/ 2500, and PacBio RS II platform	Hoshino et al., 2016
			Tubiflorae	Labiatae	*Mentha longifolia*	353	Illumina HiSeq 2000, PacBio RS platform	Vining et al., 2017
			Tubiflorae	Labiatae	*Pogostemon cablin*	1,150	Illumina HiSeq 2500	He et al., 2016
			Tubiflorae	Labiatae	*Salvia miltiorrhiza*	538	Roche 454, Illumina HiSeq 2000, PacBio RS	Xu H. et al., 2016
			Tubiflorae	Pedaliaceae	*Sesamum indicum*	274	Illumina HiSeq 2000	Wang et al., 2014
			Campanulales	Asteraceae	*Artemisia annua*	1,780	Roche 454 GS FLX, Illumina HiSeq 2500 and MiSeq, PacBio RSII platform	Shen et al., 2018
		Monocotyledoneae	Graminales	Gramineae	*Setaria italica*	396.7	Illumina Genome Analyzer II, Roche 454 GS FLX, and Sanger ABI 3730	Bennetzen et al., 2012
			Graminales	Gramineae	*Hordeum vulgare*	5,100	Illumina GAIIx and Hiseq 2000, Roche 454 GS FLX, and Sanger ABI 3730	Mayer et al., 2012
			Microspermae	Orchidaceae	*Dendrobium officinale*	1,350	Illumina HiSeq 2000, PacBio RS	Yan et al., 2015
			Microspermae	Orchidaceae	*Gastrodia elata*	1,060	Illumina HiSeq 2500 platform	Yuan et al., 2018
Animalia	Mollusca	Bivalvia	Ostreoida	Ostreidae	*Crassostrea gigas*	559	Illumina	Zhang G. et al., 2012
	Annelida	Clitellata	Rhynchobdellida	Glossiphoniidae	*Helobdella robusta*	228	Sanger	Simakov et al., 2013

intracellular gene transfer, diversity, conservation, and the genetic basis by which chloroplast transgenes can be engineered to enhance medicinal plant agronomic traits or produce high-value biomedical products. A new synthetic biological approach allowed the entire artemisinic acid biosynthesis pathway to be transferred to the tobacco chloroplast genome, ultimately creating plants that produce more than 120 mg of artemisinic acid per kg biomass. This provided efficient strategies for engineering complex biochemical pathways in plants and optimizing metabolic output (Fuentes et al., 2016).

The mitochondrial genome plays key roles in animal and plant growth and development, adaptation to environmental conditions, and reproduction. According to the endosymbiotic hypothesis, mitochondria appeared more than 1.5 billion years ago. The mammalian mitochondrial genome comprises covalently closed, double-stranded DNA molecules that mostly range from 15 to 17 kb (Gualberto et al., 2014). Compared with nuclear genomes, mitochondrial genomes have the following characteristics: (i) simple genetic structure, strict maternal lineage, no recombination, and less genetic resetting phenomena; (ii) fast evolution rate, allowing highly sensitive genetic analysis; and (iii) no tissue specificity, celluar abundance, and more than 1,000~10,000 copies per cell. Some mitochondrial genomes of medicinal animals including *Cervus nippon hortulorum* (Liu Y. et al., 2016), *Cervus elaphus songaricus* (Li Y. et al., 2016), and *Sepiella maindroni* (Zheng et al., 2016), have already been published. The mitochondrial genomes of modern plants have retained many features of their ancestral bacterial genomes, including double-stranded DNA, organelle replication and expression, and high copy numbers per cell in many photosynthetic eukaryotes. The sizes of plant mitochondrial genomes range from hundreds of kb to dozens of Mb (Gualberto et al., 2014). In contrast to the highly conserved animal mitochondrial genomes, plant mitochondrial genomes exhibit complex structures. The known molecular configurations of plant mitochondrial genomes recently included Y-type, H-type, multivariate linear, single linear, circle and circle and linear coexisting. Higher plant mitochondria frequently recombine, which is mediated by repeat sequences. The complete mitochondrial genomes of some medicinal plants and fungi, such as *Rhazya stricta* (Park et al., 2014), *Ganoderma lucidum* (Li J. et al., 2013), and *Ophiocordyceps sinensis* (Li et al., 2015a), are now available. These studies provide valuable information for

improving our understanding of mitochondrial genome evolution in angiosperms and fungi. Genomic data have enabled a rigorous examination of gene transfer events and provided valuable genomic resources for utilizing important medicinal plants in biotechnology applications.

Transcriptomics

Herbal transcriptomics is focused on the structure, function, and regulation of gene transcription in specific tissues or cells in selected herb growth stages at the overall level. Transcriptome studies of the expression of genes in cells or tissues at the whole-mRNA level can provide expression characteristics for all genes and possible functional information for proteins. Transcriptomes are a dynamic combination of genomes and external physical influences, and they reflect complete gene expression data in cells of an individual organism in a particular organ or tissue, or at a specific developmental or physiological stage. Transcriptomes can be used to compare gene expression differences in different tissues or physiological conditions, to discover genes associated with specific physiological functions, and to deduce the function of unknown genes. Transcriptomics studies have changed the research model of single genes and brought genomics research into the postgenomic era.

Transcriptome research on TCM provides effective data for parsing genetic information regarding TCM at the transcriptional level and has significant advantages and applications in discovering key enzyme genes in secondary metabolism biosynthetic pathways and clarifying secondary metabolism pathways and regulatory mechanisms (Table 2). Transcriptomics has many applications in TCM studies using different sequencing platforms, including RNA-seq by second generation sequencing and isoform sequencing (Iso-seq) by SMRT sequencing. First, transcriptomics provides a foundation for analysis of the metabolism pathways of active TCM compounds, such as ginsenosides (Cao et al., 2015; Lee et al., 2017; Li C. et al., 2013; Wang et al., 2016a; Zhang et al., 2017a), triterpene saponin (Liu M. et al., 2015; Luo et al., 2011), and tanshinone (Cui et al., 2015; Guo et al., 2016; Xu et al., 2015; Zhang X., et al., 2015; Zhou et al., 2017), and elucidates candidate gene function by *in vitro* enzymatic catalysis. For example, based on RNA-seq results, six candidate CYP genes were selected and successfully demonstrated that *CYP76AH1* could catalyze a unique four-electron oxidation cascade on miltiradiene to produce ferruginol both *in vitro* and *in vivo* (Guo J. et al., 2013). Obtained by a combination of NGS and SMRT sequencing, the full-length transcriptome sequences and splice variants of *Salvia miltiorrhiza* provided further insight into tanshinone biosynthesis (Xu et al., 2015). Second, transcriptomics offers basic data for identifying the functions of genes related to medicinal plant growth, development, disease resistance, antiadversity, and other excellent characteristics. For instance, RNAi knockdown of *SmCPS5* related to gibberellin biosynthesis resulted in dwarf transgenic plants with significantly shorter pinnately compound leaves, shorter and narrower top leaves, and smaller flowers compared with wild-type *Salvia miltiorrhiza* (Cui et al., 2015). In addition, transcriptomics supplies a research foundation for studying genetic diversity and molecular marker development. For example, 20 new microsatellite markers were developed to investigate the species' population genetics based on the *de novo* assembly and characterization of the leaf, bud, and fruit transcriptome of *Juglans mandshurica* (Hu et al., 2016).

Functional genomics

Functional genomics, also known as postgenomics, attempts to utilize data from genomic and transcriptomic projects to describe the functions of genes. Compared with genomics, functional genomics is a dynamic process that concentrates on gene transcription, translation, expression regulation and protein interactions. Based on transcriptomics, proteomics and metabolomics, functional genomics analyses of Chinese herbal medicines are used to identify functional genes. Functional genomics mainly focuses on the construction of model medicinal plant platforms and analysis of the synthetic and regulatory mechanisms of secondary metabolites. In addition, the resistance to stress and the genetic mechanisms underlying other excellent agronomic traits are also studied. The descriptions described herein mainly focus on the two aspects of biosynthetic pathways and their regulation, as described below.

The first aspect of herbal functional genomics involves researching the biosynthesis of active compounds (Table S1 in Supporting Information). Natural medicinal plant products are diverse and complex, and most have common upstream metabolic pathways that have been analyzed thoroughly. Currently, functional characterization of specific natural products mainly focuses on analysis of downstream metabolic pathways. Isopentenyl pyrophosphate (IPP), the common terpenoid precursor, is produced by the MVA (mevalonate) and MEP (2-C-methyl-*D*-erythritol-4-phosphate) pathways (Xu et al., 2016b). A study on the biosynthesis of artemisinin found that farnesyl diphosphate (FPP) could be catalyzed into artemisinic acid by amorpha-4,11-dienesynthase (*ADS*), cytochrome P450 reductase (*CPR1*), *CYP71AV1*, and alcohol dehydrogenase 1 (*ALDH1*). Synthetic biology research has elucidated an efficient biosynthetic route to artemisinic acid in *Saccharomyces cerevisiae*, achieving fermentation titers of 25 g L^{-1} (Paddon et al., 2013). Tanshinones are highly oxidized natural products from the roots/rhizomes of the medicinal model plant *Salvia miltiorrhiza*. Functional genomic studies have indicated that

Table 2 Several medicinal plant transcriptomes reported in the last three years

Species	Tissue	Data	Focused gene	Biosynthesis pathway	Sequencing platform	Reference
Andrographis paniculata	Leaf	83,800 transcripts	Transcription factor, *CYP450s*	Terpenoid	Illumina HiSeq 2000	Cherukupalli et al., 2016
Anemone flaccida	Rhizome, leaf	36,617 genes		Triterpenoid saponin	Illumina	Zhan et al., 2016
Anoectochilus roxburghii	Dry and third-stage seed	173,781 unigenes			Illumina HiSeq 4000	Liu S. et al., 2015
Artemisia annua	Seedling	185,653 unigenes	Transcription factor	Artemisinin	Illumina HiSeq 4000	Hao et al., 2017
Catharanthus roseus	Hairy root	30,281 unigenes	Anthranilate synthase	Terpenoid indole alkaloid (TIA)	Illumina HiSeq 2000	Sun J. et al., 2016
Cistanche deserticola	Succulent stem	95,787 transcripts	Phenylalanine ammonia-lyase (*PAL*)	lignin, phenylethanoid glycosides (PhGs)	Illumina HiSeq 2000	Li et al., 2015b
Clitoria ternatea	Root, flower	137,334/92,595 contigs			Illumina HiSeq 2000	Nguyen et al., 2016
Dioscorea nipponica	Young rhizome	153,924 unigenes		Dioscin	Illumina HiSeq 2000	Sun et al., 2017
Gardenia jasminoides	Leaf, green fruit, red fruits	141,665 unigenes	*CCD*, *ALDH*, *UGT*	Crocin	Illumina HiSeq 2000	Ji et al., 2017
Gnetum parvifolium	Leaf	94,816 unigenes		Phenylpropanoid, flavonoid, stilbenoid	Illumina HiSeq 2000	Deng et al., 2016
Hypericum perforatum	Flower	33,860 isotigs			Roche 454 GS FLX	Galla et al., 2015
Lonicera japonica	shoot apex, stem, leaf-1 (youngest leaf near shoot apex), leaf-2 (second leaf), leaf-3 (mature leaf), green floral bud, white floral bud, white flower, yellow flower	243,185 unigenes		Chlorogenic acid (CGA), luteolosides, secoiridoid	Illumina HiSeq 2000	Rai et al., 2017
Lonicera japonica	Leaf, flower	150,523 unigenes	Transcription factor		Illumina HiSeq 2000	Zhang L. et al., 2016
Paeonia Ostii	Root, stem, leaf	77,704 unigenes	Transcription factor		Illumina HiSeq 2000	Wang et al., 2016b
Panax ginseng	Root	182,881 unigenes		Triterpenoid saponins	Illumina HiSeq 4000	Zhang et al., 2017a
Panax ginseng	Main root body, lateral root, rhizome	232,702 contigs		Ginsenosides	Illumina HiSeq 2000	Jayakodi et al., 2015
Periploca sepium	Root, adventitious root, calli, leaf	71,629 unigenes		Terpene, steroid	Illumina HiSeq 2500	Zhang et al., 2017b
Pinellia ternata	tuber	89,068 unigenes	Transcription factor	Benzoic acid, ephedrine	Illumina HiSeq 2000	Zhang G. et al., 2016
Platycodon grandiflorus	Root, leaf, petal	40,420 unigenes	*CYP716A141*	Triterpenoid saponin	Illumina HiSeq 2000	Tamura et al., 2017
Platycodon grandiflorus	Root	50,408 transcripts	*CYP450s*	Triterpenoid saponin	Illumina HiSeq 2000	Ma et al., 2016
Pueraria lobata	Root, stem, leaf	81,508 contigs		Isoflavonoid	Illumina HiSeq 1000	Han et al., 2015a
Rehmannia glutinosa	Root, leaf	66,906 unigenes			Illumina HiSeq 2500	Li M. et al., 2017
Rehmannia glutinosa	Root	99,708 unigenes			Illumina HiSeq 2000	Yang Y. et al., 2015
Salvia miltiorrhiza	Root, stem, leaf, flower	160,468 unigenes	*CYP450s*, laccases	Salvianolic acids, tanshinone	PacBio RS II, Illumina HiSeq 2000/2500	Xu Z. et al., 2016; Xu et al., 2015
Sophora flavescens	Stem, leaf, flower, young bud, mature bud, bud right before blossom, pedicel while bud stage, pedicel while blossom, callus	83,325 contigs		Isoflavonoid, quinolizidine alkaloids	Illumina	Han et al., 2015b
Swertia japonica	Root, aerial	81,729 unigenes	Glycosyl transferase	Swertiamarin, amarogentin, mangiferin	Illumina HiSeq 2000	Rai et al., 2016
Tripterygium wilfordii	Root	105,556 transcripts	*TPS*	Terpenoid	Illumina HiSeq 2500	Hansen et al., 2017
Tulipa edulis	Stolon	74,006 unigenes			Illumina HiSeq 2500	Miao et al., 2016
Viscum coloratum	Leaf	39,226,078 reads			Illumina HiSeq 2000	Kim et al., 2017a

the diterpenoid synthases, *CPS* and kaurene synthase (*KS*) catalyze *E*,*E*,*E*-geranylgeranyl diphosphate (GGPP) into miltiradiene, which is the backbone of tanshinones. Then, miltiradiene is transformed into a series of tanshinone intermediates by the downstream modification of *CYP76AH1*, *CYP76AH3*, *CYP76AK1* and *2OGD5* (2-oxoglutarate-dependent dioxygenase) (Guo et al., 2016; Guo J. et al., 2013; Xu and Song, 2017). Another kind of active compound in *Salvia miltiorrhiza* is water-soluble phenolic acids, such as rosmarinic acid (RA) and lithospermic acid B (LAB). Although the upstream pathways of phenolic acid biosynthesis in *Salvia miltiorrhiza* have been clearly reported, researchers continue to search for gene homologs encoding key enzymes (Xu et al., 2016b). For instance, isotopic labeling using [ring-^{13}C]-phenylalanine combined with UPLC/Q-TOF test showed that 4-coumaroyl-CoA and 3,4-dihydroxyphenyllactic acid (DHPL) are coupled by the ester-forming enzyme RA synthase (*SmRAS*) to form 4-coumaroyl-3′,4′-dihydroxyphenyllactic acid (4C-DHPL) and then hydroxylated by *SmCYP98A14* to form RA (Di et al., 2013). Based on hybrid sequencing (Illumina and SMRT) of the *Salvia miltiorrhiza* root transcriptome, 15 full-length transcripts and four alternative splicing events of enzyme-coding genes were identified as involved in the biosynthesis of RA. In addition, the results demonstrated that LAB accumulates in the phloem and xylem of roots, in accordance with the expression patterns of the identified key genes related to RA biosynthesis (Xu et al., 2016c).

Most alkaloids, mainly distributed in plants with very diverse structures, come from amino acids, and only a few are formed by amination. Vinblastine and vincristine are monoterpenoid indole alkaloids (MIAs) whose biosynthesis pathways mainly include iridoid, indole and vindoline and which involve more than 30 key enzymes. The metabolic pathways of (seco)iridoids in *Catharanthus roseus* were fully elucidated via analyses of 8-hydroxygeraniol oxidoreductase, iridoid oxidase, 7-deoxyloganetic acid glucosyltransferase, and 7-deoxyloganic acid hydroxylase (Miettinen et al., 2014). Noscapine is a benzylisoquinoline alkaloid (BIA) derived from scoulerine in *Papaver somniferum*. Based on transcriptomic analysis and bacterial artificial chromosome (BAC) sequencing, a gene cluster comprising ten genes was found to exclusively express HN1, a high noscapine-producing poppy variety. Functional characterization showed that these genes related to the biosynthesis of noscapine, which demonstrated the order of catalysis in the noscapine metabolic pathway and provided a novel biosynthetic pathway for noscapine (Winzer et al., 2012).

Flavonoids are characteristic secondary metabolites in plants and can be classified into six types according to structural differences. Phenolic-CoA is the initial precursor of flavonoids, and 2s-flavanones are the main precursors of five other flavonoids. For instance, during the catalysis of chalcone synthase (*CHS*) and chalcone isomerase (*CHI*), phenolic-CoA can be transformed into 2s-flavanones, and 2s-flavanones can be further converted to flavones, isoflavones, flavonols, flavanols, and anthocyanidins by flavone synthase (*FNS*), isoflavone synthase (*IFS*), flavanone 3-hydroxylase (*F3H*), flavonol synthase (*FLS*), dihydroflavonol 4-reductase (*DFR*), leucoanthocyanidin reductase (*LAR*) and anthocyanidin synthase (*ANS*) (Guo et al., 2017). Herbgenomics studies have rapidly led research on the biosynthesis and regulation of natural products into the “golden age”.

The second aspect of herbal functional genomics is research on the regulation of active compound biosynthesis. Transcription factors play key roles in the regulation of active compound biosynthesis. With the evolution of medicinal plants, active compounds have been shown to constitute a series of complex regulatory mechanisms that aid their survival in severe environments by enhancing or inhibiting the expression of other genes. Two transcription factors in *Artemisia annua*, *AaERF1* and *AaERF2*, can be strongly induced by jasmonate and have been demonstrated to positively enhance the activity of *AaADS* and *AaCYP71AV1* (Yu et al., 2012). In addition, another transcription factor, *AaORA*, can also enhance the expression levels of *AaADS*, *AaCYP71AV1*, *AaDBR2* and *AaERF1* (Lu et al., 2013). These three regulators are enormously valuable for the biosynthesis of artemisinin. As a major regulator of the expression of *ORCA3*, a methyl jasmonate-responsive gene, *CrMYC2* is a *Catharanthus roseus* transcription factor that plays an important role in regulating expression of the alkaloid biosynthetic gene indirectly. Thus, *CrMYC2* is considered the key point in the production of alkaloids (Zhang et al., 2011). According to transcriptomic data, the basic leucine zipper (*bZIP*) genes were highly expressed in the periderm of *Salvia miltiorrhiza*. It has been reported that *SmbZIP7* and *SmbZIP20* potentially participate in the regulation of tanshinone biosynthesis (Zhang et al., 2018). In conclusion, discovering transcription factors is important for research on functional genes, and information gained from these discoveries is used to breed medicinal plants and yield active ingredients.

Proteomics

Proteomics is the study of proteins in organisms, tissues, and cells at the overall level and includes not only the identification and quantification of proteins but also the determination of their localization, modifications, interactions, activities, and functions. Herbal proteomics is the application of proteomics in TCM and includes two research aspects. The first aspect involves comparing differences in the protein groups of different herbs or different organs in the same herb and can be used to evaluate correlations between active TCM ingredients and changes in the protein group, which reveals

the molecular mechanism underlying TCM activity. The second aspect of herbal proteomics involves seeking proteins related to TCM targets by comparing differences in the protein expression profiles of animal cells or tissues before and after TCM administration.

With the development of postgenomics, studies on proteins in TCM have received increasing attention. A comparison study of wild and cultivated ginseng (*Panax ginseng*) revealed that the contents of 14 amino acids were higher in wild ginseng than in cultivated ginseng (Sun H. et al., 2016). Moreover, accumulations of amino acid metabolism-related enzymes and their derivatives, such as glutamate decarboxylase and *S*-adenosylmethionine, were also higher in wild ginseng. These results fully validate the significant medicinal value of wild ginseng and have reference value for investigations on ginseng medicinal functions (Sun H. et al., 2016). In addition, combined with mass spectrometry (MS)-based proteomics, many proteins related to the biosynthesis of artemisinin were found to be highly expressed in trichome-enriched regions rather than in leaves of *Artemisia annua*. These proteins can be divided into two categories: proteins involved in the artemisinin metabolic pathway and other highly abundant proteins. These data indicated that additional artemisinin biosynthetic pathways may exist within trichomes (Bryant et al., 2015).

Studies on proteins are very important for analyzing the mechanism of TCM and mainly focus on single compounds, multiple compounds, and TCM formulas. Tanshinone IIA is an active ingredient in *Salvia miltiorrhiza*, and proteomic analysis of CaSki cells induced by tanshinone IIA revealed that tanshinone IIA can be used to treat cervical cancer via activating ER stress pathways and ameliorating C/EBP-homologous proteins and apoptosis signal-regulating kinase 1, ultimately leading to apoptosis. This study demonstrated that tanshinone IIA may be a useful drug for advanced cervical carcinoma (Pan et al., 2013).

The different global protein profiles between normal HepG2 cells and HepG2 cells treated with polysaccharides isolated from *Phellinus linteus* (PL), *Ganoderma lucidum* (GL) and *Auricularia auricula* (AA) demonstrated that PL/GL/AA polysaccharides inhibit the proliferation of HepG2 hepatoma cell lines by inducing cell apoptosis and blocking cancer cells in the G_1 and/or S phases. This result further enhances the value of medicinal fungi in traditional oriental medicine systems (Chai et al., 2016). In addition, by triple-omics (transcriptomics, proteomics and metabolomics) profiling, research of the Bufei Yishen formula (BYF) showed that many genes, proteins and metabolites are involved in long-term therapeutic actions during chronic obstructive pulmonary disease (COPD), providing a basis for understanding the mechanisms underlying BYF at a system level (Li et al., 2016a).

Metabolomics

Metabolomics is an emerging omics technology following genomics, transcriptomics and proteomics. The metabolomics concept, first proposed in 1999 (Nicholson et al., 1999), considers the composition of and dynamic changes in small molecular metabolites in organisms, tissues, and even individual cells. Herbal metabolomics specifically relates to Chinese herbal medicines and utilizes various analytical methods to qualitatively and quantitatively measure gene or environmental influences on the small-molecule metabolites of Chinese herbal medicines. Herbal metabolomics can analyze metabolite biosynthetic pathways, metabolic networks and regulatory mechanisms by measuring metabolic profiles. Herbal metabolomics research includes mainly the identification and quality evaluation of Chinese herbal medicines; breeding and stress tolerance analysis of Chinese herbal medicines; and analysis of primary and secondary metabolic pathways, metabolic networks, metabolic engineering and synthetic biology research, among others. In addition, metabolomics can lay foundations for Chinese herbal medicine breeding, innovative drug development and quality safety evaluations.

An increasing number of studies on the applications of metabolomics to Chinese herbal medicines have been reported. The biosynthesis and regulation of tanshinone, one of the most important active ingredients in *Salvia miltiorrhiza*, have drawn the attention of many researchers. While exploring the biological functions of *SmCPSs* and *SmKSLs* (kaurene synthase-like cyclases), 21 potential intermediates identified as diterpenoids were revealed by metabolomics analyses, indicating a complex network for tanshinone metabolism defined by certain key biosynthetic steps (Cui et al., 2015). Using RNA interference (RNAi), quantitative reverse transcription polymerase chain reaction (qRT-PCR), ultra-pure liquid chromatography (UPLC), and liquid chromatography-mass spectrometry (LC-MS), *2OGD5* was confirmed to play an important role in the biosynthesis of tanshinone (Xu and Song, 2017).

Agarwood (*Aquilaria sinensis*) is widely used in medicine, incense and perfumes, and the primary metabolites, secondary metabolites and DNA markers of different agarwood cultivars were identified using ^{1}hydrogen nuclear magnetic resonance (^{1}H-NMR), gas chromatography-mass spectrophotometry (GC-MS) and DNA-based techniques. This research not only provided a powerful method for the identification of agarwood but also contributed to the reliability and security of agarwood trading markets (Nguyen et al., 2017).

Due to the increasing demand of Danggui (DG, *Angelica* sinensis), it is usually adulterated with European Danggui (EDG, *Levisticum officinale*). Researchers found the potential regulatory mechanism underlying the blood concentra-

tion of DG and EDG on acetylphenyl hydrazine (APH) that induces blood deficiency using NMR metabolic analysis combined with comprehensive drug efficacy assessment methods. This result suggests that metabolomics in combination with efficacy and similarity analyses may reveal subtle biological differences between DG and EDG and highlights the potential for metabolomics to quantify the pharmacological effects of Chinese herbal medicines (Zhang Z. et al., 2016).

Epigenomics

Herbal epigenomics aims to investigate the epigenetic variation in medicinal plants of important economic value and model medicinal plants involved in different secondary metabolite biosynthetic pathways. Herbal epigenomics includes DNA methylation, covalent protein modification, chromatin remodeling, and non-coding RNA regulation (Booth and Brunet, 2016).

Accumulating evidence has shown that epigenetics is associated with the formation of "authentic" TCMs. In the case of Chinese ginseng, morphological and physiological differences were observed between wild and cultivated ginseng. Epigenetic analyses demonstrated that cultivated ginseng possesses significantly lower levels of cytosine methylation (especially for CHG methylation sites) and shows distinct methylation patterns compared with those of wild ginseng (Li M. et al., 2015). These results suggested that domestication has modulated the epigenetic modifications of cultivated ginseng, and some cytosine methylation changes might be associated with phenotypic divergences between cultivated and wild ginseng. The relationship between epigenetics and "authentic" TCMs provides a basis for the quality assessment and identification of TCMs.

In eukaryotes, miRNAs are ubiquitous non-coding RNA species that target complementary mRNA sequences and result in either translational repression or target degradation. Previous studies have revealed that miRNAs regulate secondary metabolites, abiotic stress, and growth and developmental processes in medicinal plants. *Picrorhiza kurroa* is a medicinal herb of industrial value due to the presence of the secondary metabolites picroside-I and picroside-II. Computational miRNA identification was conducted in six *Picrorhiza kurroa* transcriptomes generated from root, shoot, and stolon organs varying in their growth, development, and culture conditions. Validation of miRNA and expression analyses by qRT-PCR and 5′ rapid amplification of DNA ends (RACE) revealed that miRNA-4995 plays a regulatory role in terpenoid biosynthesis, ultimately affecting the production of picroside-I. miR-5532 and miR-5368 had negligible expression in field-grown samples compared to samples cultured *in vitro*, suggesting their role in regulating *Picrorhiza kurroa* growth under cultured conditions (Vashisht et al., 2015). These results provide insight into the possible role of epigenetics in regulating the biology of organs and metabolite biosynthesis.

In summary, studies on the relationship of epigenetics and TCM can be used to analyze gene-environment interactions, which could help improve the understanding of the epigenetic mechanisms of "authentic" herb formation. Moreover, herbal epigenomics provides an effective platform to detect the possible roles of epigenetic variations in controlling various biological processes. The potential regulatory role of epigenetic modification provides new insights into epigenetic research in medicinal plants.

Metagenomics

Herbal metagenomics is a newly developing discipline that applies molecular biological techniques to analyze the microbial diversity, population structure, evolutionary relationships, functional activities, and interactions of Chinese herbal medicines and microorganisms in their surrounding environment.

Finding diverse microbial communities in plants has important implications in herbal medicine. Juzen-taiho-to (JTT) is an immune-boosting formulation of ten medicinal herbs used clinically to boost the immune functions of humans in East Asia. The presence of lipopolysaccharide (LPS) was examined in JTT by a series of chemical and biochemical studies (Montenegro et al., 2015), and the findings indicated that JTT might contain immune-boosting bacteria. To characterize the bacteria in JTT, 16S ribosomal RNA sequencing was carried out for *Angelica sinensis* (dried root), one of the most potent immunostimulatory herbs in JTT, revealing a total of 519 bacteria genera in *Angelica sinensis*. The most abundant genus was *Rahnella*, which is widely distributed in water and plants, and the abundance of *Rahnella* appeared to correlate with the immunostimulatory activity of *Angelica sinensis*. These results provide evidence of a bacterial contribution in medicinal herbs.

Microbial diversity and community composition significantly affect agricultural soil productivity, plant growth, and crop quality. Notoginseng (*Panax notoginseng*), a valuable herbal medicine, inflicts high death rates in continuous cropping systems, and variations in the soil microbial community are considered the primary cause of notoginseng mortality (Dong et al., 2016). Variations in the diversity and composition of soil microbial communities were assessed by high-throughput sequencing methods, revealing that Burkholderiales, Syntrophobacteraceae, *Myrmecridium*, *Phaeosphaeria*, *Fusarium*, and *Phoma* were better adapted to the colonization of diseased plants. By contrast, Thermogemmatisporaceae, Actinosynnemataceae, Hydnodontaceae, Herpotrichiellaceae, and *Coniosporium* might be pathogen antagonists. These findings provide a clear scope for

screening beneficial notoginseng microbes and pathogens. Currently, continuous cropping, for which the rhizosphere soil microenvironment plays an important role in formatting, serves as an obstacle for the cultivation of medicinal plants. Therefore, soil metagenomics provides new insights into removing continuous cropping obstacles in medicinal plants.

After the effective compositions of TCM are metabolized or biotransformed by human intestinal bacteria, their metabolites can be absorbed more easily. Ginseng polysaccharides and ginsenosides in a ginseng decoction were investigated in an over-fatigue and acute cold stress model (Zhou et al., 2016), revealing that ginseng polysaccharides improved the intestinal metabolism and absorption of certain ginsenosides. Meanwhile, perturbed holistic gut microbiota were reinstated, and the growth of two major metabolic ginsenoside bacteria was particularly enhanced. These results suggested that TCMs might regulate the growth of human intestinal flora.

TCM can regulate the composition of intestinal flora and restore the homeostasis of intestinal microecology. Isoliquiritigenin (ISL), a flavonoid extracted from licorice, reduced tumor incidence rates *in vivo*, and the effects of ISL on colitis-associated colorectal cancer (CAC) development and gut microbiota were evaluated using an azoxymethane and dextran sulfate sodium (AOM/DSS)-induced mouse model of CAC (CACM) (Wu M. et al., 2016). High-throughput sequencing and terminal restriction fragment length polymorphism (T-RFLP) studies on the bacterial 16S rRNA gene indicated that the gut microbial community structure shifted significantly following AOM/DSS treatment. ISL reversed this imbalance and altered the familial constituents of gut microbiota. At the genus level, ISL reduced the abundance of opportunistic pathogens (*Escherichia* and *Enterococcus*), and increased the levels of probiotics, particularly butyrate-producing bacteria (*Butyricicoccus*, *Clostridium*, and *Ruminococcus*). Therefore, ISL and gut microbiota may have synergistic anticancer effects. Metagenomic analyses of the gut microbiota provide a basis for analyzing the effect of the gut microbiota on diseases.

In general, herbal metagenomics technology can be applied in two aspects, to screen for functional genes and exploit proteins with new functions and to determine interactions among various microorganisms or between microorganisms and their surrounding environment. As the application of metagenomic technology expands, studies on library construction, screening, sequencing and analysis drive metagenomics toward even more practical values.

Key technologies in herbgenomics

Herbgenomics involves both a theoretical basis and technology for genomics, functional genomics, proteomics, transcriptomics, metabonomics, epigenomics, metagenomics, synthetic biology, pharmacogenomics of TCM, and bioinformatics. The experimental methods of herbgenomics include high-throughput sequencing, genetic mapping, physical mapping, the generation of mutation libraries, tissue culture and genetic transformation, protein purification and identification, spectroscopy technology, and genome editing. Here, we simply introduce the key technologies employed for high-throughput sequencing, bioinformatic analysis, and biosynthetic pathway analysis of natural products.

The strategies applied in genomics research include genome or transcriptome sequencing, genome assembly at the scaffold or chromosome level, genome annotation and evolutionary analysis, conducted via a series of high-throughput sequencing and bioinformatic analyses. High-throughput sequencing, or next-generation sequencing, was selected by *Nature Methods* as the method of the year in 2007 (Chi, 2008; Schuster, 2008) and has provided fascinating opportunities in life sciences. With the development of sequencing techniques, Illumina sequencing currently serves as the industry standard in this field and can be used for whole-genome sequencing, target sequencing, transcriptome sequencing, ChIP-Seq, DNA methylation analysis, and gene expression profile analysis, among others. However, next-generation sequencing platforms present some limitations, such as short read lengths and PCR bias introduced by amplification. SMRT sequencing, represented by PacBio and Oxford Nanopore Technologies sequencing platforms, has been developed to address these limitations. The generation of long reads (20 kb) or ultra-long reads (>100 kb) will greatly improve *de novo* genome assembly, the identification of full-length transcripts, target sequencing, epigenomics, and metagenomics. The high-throughput data and diverse features generated by different sequencing platforms pose major challenges in bioinformatic analyses. Recently, many bioinformatic software platforms have been developed to achieve *de novo* genome assembly of highly heterozygous and repetitive DNA. For example, platforms such as SOAPdenovo, AbySS, ALLPATH-LG, and Velvet have been used to assemble short reads; CCS, Quiver, Pbdagcon, PacBioToCA, Proovread, LoRDEC, and LSC have been used to correct read errors resulting from SMRT sequencing via clustering of low-quality long reads or alignment of high-quality short reads; and FALCON, HGAP, Miniasm, CANU, and SMARTdenovo have been employed for the *de novo* assembly of long reads from SMRT sequencing. The assembled scaffolds or contigs are then mounted onto long scaffolds or chromosomes using optical mapping or Hi-C sequencing. The construction of genetic maps enables a drafted genome to be assembled at the chromosome level, providing support for researching whole-herb genomes in depth. Genetic mapping can also be used to study quantitative trait locus (QTL) gene numbers and relative chromo-

some locations. For example, *Artemisia annua* serves as the foundation for genomics-assisted herb breeding (Graham et al., 2010). Thus, the target genome will be annotated for mRNA loci, protein sequences, and repetitive DNA, and the annotated mRNA will be employed to evaluate differences in gene expression or alternative splicing events via RNA-seq or Iso-seq (Chen G. et al., 2017). The completeness of genome assembly and annotation was further evaluated by multiple bioinformatic programs or experimental methods, such as BUSCO (assessing genome assembly and annotation completeness with single-copy orthologs) (Simão et al., 2015), REAPR (a universal tool for genome assembly evaluation) (Hunt et al., 2013), and GAGE (a critical evaluation of genome assemblies and assembly algorithms) (Salzberg et al., 2012). The genome sequencing of medicinal plants will provide abundant information for gene selection and cloning related to biosynthetic pathways, plant development, and response to stress.

The analysis of biosynthetic pathways will provide support for natural product synthetic biology in herbgenomics. Here, the crucial factor is how to select and identify candidate genes related to the biosynthesis of natural products. Co-expression, differential gene expression, and gene cluster analyses are the most commonly used strategies for gene selection related to biosynthesis based on RNA-seq data and genome references. Knock-down or knock-out technologies, such as VIGS, RNAi, and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9, have been applied to identify the functions of candidate genes *in vivo*. For instance, the CRISPR/Cas9 system has been greatly applied for efficient and versatile genome engineering in various organisms, especially model microbes, animals and plants (Cong et al., 2013; Shan et al., 2014; Huang et al., 2017; Shen et al., 2017). Recently, CRISPR/Cas9 has also been successfully used as a knock-out system in a few medicinal plants, such as the 3′-hydroxy-*N*-methylcoclaurine 4′-*O*-methyltransferase (*4′OMT2*) gene, which regulates the biosynthesis of BIAs in *Papaver somniferum* (Alagoz et al., 2016); *C3H*, *C4H*, *4CL*, *CCR*, and *IRX*, which belong to the lignocellulose biosynthesis pathway in *Dendrobium officinale* (Kui et al., 2016); the committed diterpene synthase gene (*SmCPS1*) involved in tanshinone biosynthesis (Li B. et al., 2017); and the *RAS* involved in water-soluble phenolic acid biosynthesis in *Salvia miltiorrhiza* (Zhou et al., 2018). However, application in medicinal plants is still restricted by the lack of reference genomes and transgenetic technology.

In the metabolomes of transgenic lines, the reduction and accumulation of metabolites are analyzed via LC-MS technologies. Then, candidate genes are further functionally identified through catalytic activity analysis *in vitro*. Genetic transformation technology refers to the introduction of artificially separated or exogenously modified genes to medicinal biology using recombinant DNA techniques, which change genetic traits and improve quality, insect resistance, disease resistance, and stress resistance. Tissue culture technology makes the rapid propagation of medicinal plants a reality and is the foundation of the biosynthesis and biotransformation of phytomedicines. Mutant library construction technology provides the basic material for functional gene analyses in herbgenomics.

Practical applications of herbgenomics

Model herb research platform

Model species are those intentionally selected to study specific biological phenomena and are based on thousands of studies and general acceptance. Model species enable researchers to understand universal biological concepts from complex phenomena and properties of biological systems. For example, *Arabidopsis thaliana*, *Danio rerio*, *Drosophila melanogaster*, *Escherichia coli*, *Mus musculus*, *Oryza sativa*, *Saccharomyces cerevisiae*, and *Zea mays* are main model species in different kinds of organisms. Model species for medicinal organisms are much more complex. Secondary metabolites are the main effector compounds of medicinal organisms, and they are used to study the biosynthetic pathways and the regulatary and control mechanisms of these organisms. To date, mature model species research systems are lacking in medicinal organisms. For example, the secondary metabolites of medicinal plants include terpenes, alkaloids, fatty acids, and phenylpropanoids, while their biosynthetic pathways comprise the malonic acid, shikimic acid, mevalonic acid, and amino acid pathways. The same compound can be synthesized by different single pathways or by a combination of multiple pathways, and the kinds of compounds and synthetic capacities of different organisms vary. Therefore, with reference to different biosynthetic pathways, different model herbs should be selected.

Selection criteria for model herbs should include characteristics similar to those of model organisms, such as a short generation cycle, more progeny, easy to culture and genetically transform, not harmful to human beings or the environment, and stable phenotype. In addition, relatively small genome sizes and easy whole-genome sequencing are important new standards for sifting model species. With the full-speed development of biotechnology, several species, including *Ganoderma lucidum* and *Salvia miltiorrhiza*, have emerged as valuable model herbs for studying metabolic activities and genetics (Chen et al., 2015). *Ganoderma lucidum* meets the criteria for an ideal model herb and involves multiple secondary metabolite biosynthetic pathways, making it an ideal model to study the biosynthesis of secondary metabolites and their regulation. Recently, elucidation of the *Ganoderma lucidum* genome at the chromosomal level has

laid a solid foundation for its application as a medicinal model organism. This organism supposedly plays important roles in studies on the diversity of secondary metabolites, the development of medicinal fungi and the synthetic biology of natural medicines (Sun et al., 2013). *Salvia miltiorrhiza* possesses important medicinal and economic value, and its genome and transcriptome have been significantly reported. Construction of a mutant library for *Salvia miltiorrhiza*, a genome map with high quality, and a functional genome will be investigated (Song et al., 2013). Investigations of model organisms have also led to the development of experimental biological techniques. The model medicinal organism system will enhance our understanding of the genetic basis of synthetic metabolites and accelerate the improvement of medicinal species research (Xu et al., 2014).

Herbal synthetic biology

Based on herbgenomics research, herbal synthetic biology involves the discovery of natural product synthesis-related elements, which are then designed and standardized by engineering principles. Through the assembly and integration of reconstructed biosynthetic pathways and metabolic networks in chassis cells, researchers elucidated the orientation and efficient heterogeneous synthesis of active compounds, which solved a series of major issues in natural medicine research, development, and production.

TCM contains various kinds of medicinal active compounds, which are related to complex secondary metabolism pathways. However, in addition to the metabolic pathways of paclitaxel and artemisinin, our understanding of most biosynthetic pathways of medicinally active compounds is very scarce. The development of herbgenomics provides a basis for herbal synthetic biology. To complete the challenging task of achieving sufficient target product output for industrialization, an optimized strategy for the natural product synthetic biology system has been proposed. Integrated use of the optimizations of single elements, exogenous metabolic pathways, chassis systems and fermentation conditions can be used to modify host cells, artificially precisely regulate the expression of exogenous genes, and elucidate the combined and coordinated expression of multiple genes. This will maximize target production output and provide new lead compounds for natural medicine research (Awan et al., 2016).

Due to requiring fewer synthetic steps and its huge commercial value, the artemisinin biosynthetic pathway has become one of the most popular pathways studied. From 2003 to 2013, Keasling and colleagues engineered a *Saccharomyces cerevisiae* strain and produced artemisinic acid at 25 g L^{-1} (Dietrich et al., 2009; Martin et al., 2003; Paddon et al., 2013; Westfall et al., 2012). The successful synthesis and recent entry of semi-synthetic artemisinin into commercial production is the first demonstration of the use of synthetic biology for the development and production of pharmaceutical agents (Paddon and Keasling, 2014). This not only demonstrates the incredible potential of available technologies but also illuminates how this work can be applied to the production of other pharmaceutical agents (Paddon and Keasling, 2014).

A better understanding of biosynthetic pathways will lead to better metabolic engineering approaches, which are the ultimate goal of herbal synthetic biology. Furthermore, the structures of natural products could be diversified by combining and introducing biosynthetic pathways into other organisms, such as bacteria or yeast. The conventional practice in herbal synthetic biology is to introduce heterologous biosynthetic pathways into expression systems capable of producing products. However, a different approach for the large-scale production of pure compounds could be engineered by an entirely novel synthetic genome. For example, Gibson et al. created *Mycoplasma mycoides* cells controlled by a chemically synthesized genome that had the expected phenotypic properties and were capable of continuous self-replication (Gibson et al., 2010). Two key enzymes, flavonoid-7-*O*-glucuronosyltransferase and flavone-6-hydroxylase, were identified in the biosynthetic pathway from the *Erigeron breviscapus* genome and engineered into yeast to produce breviscapine from glucose. Following the metabolic engineering and optimization of fed-batch fermentation, the concentrations of scutellarin and apigenin-7-*O*-glucuronide, two major active ingredients of breviscapine, reached 108 and 185 mg L^{-1}, respectively (Liu et al., 2018).

Genomics-assisted herb breeding

Genomics-assisted herb breeding is the product of molecular breeding in the high-throughput sequencing era, and molecular biology and biotechnology in combination would improve the selection efficiency and breeding process. This method would involve a combination of genotype- and phenotype-related information and the selection and design of multiple characteristics simultaneously, thus enabling molecular marker-assisted breeding. Genome and transcriptome sequence data provide large amounts of simple sequence repeat (SSR) and single-nucleotide polymorphism (SNP) markers, which are beneficial to the construction of high-density genetic and physical maps. Genomic-assisted breeding for complex traits in medicinal organisms requires understanding and manipulating many factors that influence growth, development, and responses to an array of biotic and abiotic stressors. The ultimate goal of the whole-genome strategy is to obtain optimal genotypes/genes, alleles or haplotypes, and linkage disequilibrium (LD) blocks, optimizing gene networks and specific genomic regions into breeding products with desirable phenotypes (Xu et al.,

2012). In addition, using transcriptome sequences to design SSR markers could improve the accuracy of genetic diversity and molecular-assisted breeding research. GeneChip is a high-throughput, rapid and parallel nucleic acid sequencing and quantitative analysis technology that operates on the principles of other nucleic acid molecular hybridization methods. GeneChip is widely used in applications such as the detection of specific and new genes, gene expression level tests, gene mutation and polymorphic analyses, and genome sequencing and promotes the development of genetic-assisted breeding science and the production of new varieties.

Since they are considered minor crops, herbs have experienced limited genomics-assisted improvements due to high costs; however, next-generation sequencing and its increasing affordability have dramatically accelerated marker selection breeding programs through the sequencing of wild varieties and different herbal cultivars that represent a valuable reservoir of genetic diversity. Graham et al. identified genes and markers for fast-track breeding based on a detailed genetic map and found 14 traits that could affect artemisinin yield. In addition, they confirmed that the genotype has a strong influence on both glasshouse-grown and field-grown material, and a significant association of positive artemisinin yield QTLs in parents that produced hybrids with an increased artemisinin yield was observed (Graham et al., 2010). A high-density genetic map of *Sesamum indicum* was constructed based on 3,804 pairs of new DNA markers that were developed by restriction site-associated DNA sequencing (RAD-seq) in combination with SSR markers. The identification of yield-related QTLs in this study laid a preliminary foundation for marker-assisted selection toward yield traits in sesame (Wu et al., 2014). A genome-wide association study (GWAS) focused on genes with large effects on flowering time and the content of medicinally important α-bisabolol in chamomile (*Matricaria recutita*). GWAS data paved the way for the identification of marker-trait associations and the development of reliable markers for practical breeding (Otto et al., 2017).

Molecular identification of herbs

With the development of molecular biotechnology, the detection of biological genetic diversity and the classification and identification of different species at the DNA molecular level have become a reality. Common molecular identification methods include amplified fragment length polymorphism (AFLP) (Oliveira et al., 2016; Tripathi et al., 2012), direct amplification length polymorphism (DALP) (Ma et al., 2008), sequence-related amplified polymorphism (SRAP) (Wang et al., 2017; Zheng et al., 2017), target region-amplified polymorphism (TRAP) (Liu F. et al., 2016), random amplification of polymorphic DNA (RAPD) (Zhang C. et al., 2015), inter-simple sequence repeat (ISSR) (Giri et al., 2017; Saki et al., 2016; Xing et al., 2016), and SSR analyses (Kim Y. et al., 2017; Kim et al., 2017b; Tian et al., 2016). These methods are mainly used in genetic relationship analyses, biological diversity analyses, and species identification. Medicinal organisms, such as *Panax ginseng* (Diao et al., 2009), *Xanthium sibiricum* (Tomasello and Heubl, 2017), and *Apis mellifera* (Wu et al., 2017), have been identified by PCR-RFLP analyses, and *Aristolochia manshuriensis* (Wu L. et al., 2016) and many other herbal medicines have been identified by loop-mediated isothermal amplification (LAMP) (Li et al., 2016b). The potential power of DNA barcoding combined with high-resolution melting (Bar-HRM) technology was introduced for herbal medicine identification (Sun W. et al., 2016).

In recent years, DNA barcoding has been rapidly developed and utilized in medicinal species identification, drug quality control, and market supervision, which promoted the standardization of TCM identification. To date, this technology has been recorded in British and Chinese Pharmacopoeias (State Pharmacopoeia Committee, 2015; The British Pharmacopoeia Commission, 2015). An international authoritative DNA barcoding system for herbal materials based on the combination of ITS2+*psbA-trnH* regions was constructed in 2014 (Chen et al., 2014). The system has been widely used in traditional herbal medicine enterprises. Researchers have successfully used this system to identify some single-herb products and herbs from markets (Han et al., 2016; Xin et al., 2013). The first online DNA barcode identification system has been reported, which includes standard barcode sequences for approximately 95% of the species recorded in the Japanese Pharmacopoeia (JP, 16th edition) (Chen X. et al., 2017). This tool provides users with basic information on each crude drug recorded in the JP, DNA barcoding identification of herbal material, and standard operating procedures (SOPs) from sampling to data analysis. A primary study on biological quality control for traditional patent medicine (TPM) was performed in which SMRT sequencing and DNA barcoding were used to detect Yimu Wan, which provided an affordable way to monitor the legality and safety of TPM (Jia et al., 2017). Based on DNA metabarcoding and SMRT sequencing, the biomonitoring method for traditional herbal medicinal products (THMPs) showed favorable repeatability, reliability, and sensitivity for the detection of herbal ingredients. The error in SMRT sequencing did not affect the method's ability to identify multiple prescribed species and several adulterants or contaminants (Xin et al., 2018a). Besides, shotgun metagenomic sequencing was shown to be a complementary method for the precise species detection of traditional Chinese patent medicines (Xin et al., 2018b). In addition, the nucleotide signatures and SNP double peaks for specific herbal species could also detect adulterants and substitutions in Chinese

patent medicines (Gao et al., 2017).

Pharmacogenomics of TCM

Pharmacogenomics of TCM is important for clarifying herbal drug mechanisms and a Gordian knot of TCM modernization. Although both herbal and chemical drugs are active compounds used to treat diseases, herbal drugs are complex biological systems that are utterly different from chemical systems. Human body responses to herbal drugs are multidimensional and exhibit nonlinear complex effects. Through long-term clinical practice, the curative effects of most herbal drugs have been confirmed; however, the active compounds that enter the human body and have therapeutic effects and their metabolic processes remain unclear. Studies on pharmacogenomics of TCM can help us understand the *in vivo* metabolic pathways, existing forms, influential factors and efficacy material bases of TCM. The basic link between *in vivo* metabolism and therapeutic effects is the direct or indirect interaction of drug and organism molecules, which inflict structural and functional changes at multiple levels ranging from genetic information to overall function implementation, and the bases for determining the structures and functions of these levels are genes. Therefore, research on gene expression, regulation, and modification together with multi-component, multi-link and multi-target *in vivo* research will help clarify the *in vivo* metabolic process and molecular mechanisms of TCM.

Pharmacogenomics of TCM is based on the genetic polymorphisms of drug interactions, such as the polymorphisms of drug metabolism enzymes, receptors, and targets. These polymorphisms may cause individual differences in therapeutic effects and adverse reactions, while *in vivo* TCM metabolism may be more complex. The *in vivo* reactions and metabolism of TCM involve the interaction of multiple genes, while genetic polymorphisms lead to the diversity of *in vivo* metabolism reactions, thus providing a foundation for *in vivo* TCM metabolism and drug reaction research at the genomic level. Compared with pharmacogenetics, which focuses on the pharmacokinetics and pharmacodynamics of a single or few genes, pharmacogenomics of TCM has a broader scope that includes all genes that determine the TCM effect in the whole genome. Thus, pharmacogenomics of TCM systematically evaluates gene interactions and how they affect disease susceptibility, pharmacological function, drug disposition, and response to treatment and guides the development of new TCM drugs and individualized clinical treatment based on this platform. For example, exogenous plant microRNAs in food can regulate the expression of target genes in mammals. One of the most highly enriched exogenous plant microRNAs in the sera of Chinese subjects is mIR168a, which is primarily acquired orally through food intake. *In vitro* and *in vivo* function studies have demonstrated that mIR168a could bind the mRNA of human/mouse low-density lipoprotein receptor adapter protein 1 (LDLRAP1), inhibit LDLRAP1 expression in the liver, and consequently decrease LDL removal from mouse plasma (Zhang L. et al., 2012). A honeysuckle-encoded atypical microRNA, MIR2911, was reported as the first active component identified in TCM to directly target various influenza A viruses (IAVs) and may represent a novel type of natural product that effectively suppresses viral infection (Zhou et al., 2015). The microRNA HJT-Srna-m7 derived from *Rhodiola crenulata* can effectively reduce the expression levels of fibrotic hallmark genes and proteins both in alveoli *in vitro* and in mouse lung tissues *in vivo*, which provides an innovative treatment strategy for the oral delivery of microRNAs as therapeutic medications (Du et al., 2017).

In addition, the aggregate genomes of human microbiota and the diverse metabolic activities that they encode have the potential to revolutionize modern therapeutics (Haiser and Turnbaugh, 2012). Members of the gut microbiota can also influence xenobiotic metabolism by altering host gene expression and producing compounds that interfere with metabolism outside of the gut (Haiser and Turnbaugh, 2012). A randomized clinical study of specifically designed isoenergetic diets, combined with fecal shotgun metagenomics, was conducted and demonstrated that gut bacteria can be selectively promoted by dietary fibers and alleviate type 2 diabetes (Zhao et al., 2018). In addition, TCM may restore the homeostasis of gut microbiota and consequently promote the systemic exposure of concomitant small molecules in the decoction. For instance, a water extract of *Ganoderma lucidum* mycelium not only reverses high-fat diet (HFD)-induced gut dysbiosis, as indicated by the decreased Firmicutes-to-Bacteroidetes ratios and endotoxin-bearing Proteobacteria levels, but also maintains intestinal barrier integrity and reduces metabolic endotoxemia (Chang et al., 2015). Through modulation of the gut microbiota composition, water extracts of *Ganoderma lucidum* could also have anti-obesogenic and antidiabetic effects (Martel et al., 2017).

Conclusion and perspectives

We summarized the trends in herbgenomics, including the theoretical system, key technologies, and practical applications. Herbgenomics research achievements will help to elucidate the synthesis and regulatory pathways of active compounds of medicinal organisms, and then promote natural drug screening and biosynthetic research. In the meantime, it will also speed up the breeding of good varieties of medicinal plants and promote the commercial development of organic TCM agriculture. Many innovative technologies including the SMRT and nanopore sequencing, chromosome

organization and genome editing of CRISPR/Cas9 have been involved in herbgenomics research. These technologies will benefit the high-quality genome assembly, elucidation of candidate gene function related to catalysis of active compounds, growth and development, disease resistance, and response to biotic and abiotic stress of medicinal plants. Pharmacogenomics of TCM is one of the most important focuses of herbgenomics. A clear understanding of the mechanism underlying the human body response to herbal drugs and the metabolic process of active compounds in the human body will help us recognize the effects of gut microbiota and discover new active substances. In conclusion, herbgenomics will open a novel field for TCM research and provide a historical opportunity to improve the independent innovation ability of drug research and design in China.

Compliance and ethics *The author(s) declare that they have no conflict of interest.*

Acknowledgements *This work was supported by the National Natural Science Foundation of China (81573398) and CAMS Innovation Fund for Medical Sciences (CIFMS, 2016-I2M-3-016).*

Adem, M., Beyene, D., and Feyissa, T. (2017). Recent achievements obtained by chloroplast transformation. Plant Methods 13, 30.
Alagoz, Y., Gurkok, T., Zhang, B., and Unver, T. (2016). Manipulating the biosynthesis of bioactive compound alkaloids for next-generation metabolic engineering in opium poppy using CRISPR-Cas 9 genome editing technology. Sci Rep 6, 30910.
Awan, A.R., Shaw, W.M., and Ellis, T. (2016). Biosynthesis of therapeutic natural products using synthetic biology. Adv Drug Deliver Rev 105, 96–106.
Arabidopsis Genome Initiative, (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796–815.
Bennetzen, J.L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A.C., Estep, M., Feng, L., Vaughn, J.N., Grimwood, J., et al. (2012). Reference genome sequence of the model plant *Setaria*. Nat Biotechnol 30, 555–561.
Bolger, M.E., Arsova, B., and Usadel, B. (2018). Plant genome and transcriptome annotations: from misconceptions to simple solutions. Brief Bioinform 19, 437–449.
Booth, L.N., and Brunet, A. (2016). The aging epigenome. Mol Cell 62, 728–744.
Bryant, L., Flatley, B., Patole, C., Brown, G.D., and Cramer, R. (2015). Proteomic analysis of *Artemisia annua*—towards elucidating the biosynthetic pathways of the antimalarial pro-drug artemisinin. BMC Plant Biol 15, 175.
Cao, H., Nuruzzaman, M., Xiu, H., Huang, J., Wu, K., Chen, X., Li, J., Wang, L., Jeong, J.H., Park, S.J., et al. (2015). Transcriptome analysis of methyl jasmonate-elicited *Panax ginseng* adventitious roots to discover putative ginsenoside biosynthesis and transport genes. Int J Mol Sci 16, 3035–3057.
Chai, Y., Wang, G., Fan, L., and Zhao, M. (2016). A proteomic analysis of mushroom polysaccharide-treated HepG2 cells. Sci Rep 6, 23565.
Chan, A.P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiu, D., Melake-Berhan, A., Jones, K.M., Redman, J., Chen, G., et al. (2010). Draft genome sequence of the oilseed species *Ricinus communis*. Nat Biotechnol 28, 951–956.
Chang, C.J., Lin, C.S., Lu, C.C., Martel, J., Ko, Y.F., Ojcius, D.M., Tseng, S.F., Wu, T.R., Chen, Y.Y.M., Young, J.D., et al. (2015). *Ganoderma lucidum* reduces obesity in mice by modulating the composition of the gut microbiota. Nat Commun 6, 7489.
Chen, G., Shi, T., and Shi, L. (2017). Characterizing and annotating the genome using RNA-seq data. Sci China Life Sci 60, 116–125.
Chen, S. (2017). I19 Major achievements of traditional medicine in treating epidemic and chronic diseases. Biochem Pharmacol 139, 110.
Chen, S., Xu, J., Liu, C., Zhu, Y., Nelson, D.R., Zhou, S., Li, C., Wang, L., Guo, X., Sun, Y., et al. (2012). Genome sequence of the model medicinal mushroom *Ganoderma lucidum*. Nat Commun 3, 913.
Chen, S., Pang, X., Song, J., Shi, L., Yao, H., Han, J., and Leon, C. (2014). A renaissance in herbal medicine identification: from morphology to DNA. Biotech Adv 32, 1237–1244.
Chen, S., Song, J., Sun, C., Xu, J., Zhu, Y., Verpoorte, R., and Fan, T. (2015). Herbal genomics: examining the biology of traditional medicines. Science 347, S27–S29.
Chen, W., Kui, L., Zhang, G., Zhu, S., Zhang, J., Wang, X., Yang, M., Huang, H., Liu, Y., Wang, Y., et al. (2017). Whole-genome sequencing and analysis of the Chinese herbal plant *Panax notoginseng*. Mol Plant 10, 899–902.
Chen, X., Xiang, L., Shi, L., Li, G., Yao, H., Han, J., Lin, Y., Song, J., and Chen, S. (2017). Identification of crude drugs in the Japanese pharmacopoeia using a DNA barcoding system. Sci Rep 7, 42325.
Cherukupalli, N., Divate, M., Mittapelli, S.R., Khareedu, V.R., and Vudem, D.R. (2016). *De novo* assembly of leaf transcriptome in the medicinal plant *Andrographis paniculata*. Front Plant Sci 7, 1203.
Chi, K.R. (2008). The year of sequencing. Nat Methods 5, 11–14.
Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823.
Cui, G., Duan, L., Jin, B., Qian, J., Xue, Z., Shen, G., Snyder, J.H., Song, J., Chen, S., Huang, L., et al. (2015). Functional divergence of diterpene syntheses in the medicinal plant *Salvia miltiorrhiza*. Plant Physiol 169, 1607–1618.
Deng, N., Chang, E., Li, M., Ji, J., Yao, X., Bartish, I.V., Liu, J., Ma, J., Chen, L., Jiang, Z., et al. (2016). Transcriptome characterization of *Gnetum parvifolium* reveals candidate genes involved in important secondary metabolic pathways of flavonoids and stilbenoids. Front Plant Sci 7, 174.
Di, P., Zhang, L., Chen, J., Tan, H., Xiao, Y., Dong, X., Zhou, X., and Chen, W. (2013). ^{13}C tracer reveals phenolic acids biosynthesis in hairy root cultures of *Salvia miltiorrhiza*. ACS Chem Biol 8, 1537–1548.
Diao, Y., Lin, X.M., Liao, C.L., Tang, C.Z., Chen, Z.J., and Hu, Z.L. (2009). Authentication of *Panax ginseng* from its adulterants by PCR-RFLP and ARMS. Planta Med 75, 557–560.
Dietrich, J.A., Yoshikuni, Y., Fisher, K.J., Woolard, F.X., Ockey, D., McPhee, D.J., Renninger, N.S., Chang, M.C.Y., Baker, D., and Keasling, J. D. (2009). A novel semi-biosynthetic route for artemisinin production using engineered substrate-promiscuous $P450_{BM3}$. ACS Chem Biol 4, 261–267.
Dong, L., Xu, J., Feng, G., Li, X., and Chen, S. (2016). Soil bacterial and fungal community dynamics in relation to *Panax notoginseng* death rate in a continuous cropping system. Sci Rep 6, 31802.
Du, J., Liang, Z., Xu, J., Zhao, Y., Li, X., Zhang, Y., Zhao, D., Chen, R., Liu, Y., Joshi, T., et al. (2018). Plant-derived phosphocholine facilitates cellular uptake of anti-pulmonary fibrotic HJT-sRNA-m7. Sci China Life Sci 61, in press doi: 10.1007/s11427-017-9026-7.
Fu, Y., Li, L., Hao, S., Guan, R., Fan, G., Shi, C., Wan, H., Chen, W., Zhang, H., Liu, G., et al. (2017). Draft genome sequence of the Tibetan medicinal herb *Rhodiola crenulata*. Gigascience 6, 1–5.
Fuentes, P., Zhou, F., Erban, A., Karcher, D., Kopka, J., and Bock, R. (2016). A new synthetic biology approach allows transfer of an entire metabolic pathway from a medicinal plant to a biomass crop. Elife 5, pii: e13664.
Galla, G., Vogel, H., Sharbel, T.F., and Barcaccia, G. (2015). *De novo* sequencing of the *Hypericum perforatum* L. flower transcriptome to identify potential genes that are related to plant reproduction sensu lato. BMC Genomics 16, 254.
Gao, Z., Liu, Y., Wang, X., Song, J., Chen, S., Ragupathy, S., Han, J., and Newmaster, S.G. (2017). Derivative technology of DNA barcoding (-

nucleotide signature and SNP double peak methods) detects adulterants and substitution in Chinese patent medicines. Sci Rep 7, 5858.
Garcia-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, G., González, V.M., Hénaff, E., Câmara, F., Cozzuto, L., Lowy, E., et al. (2012). The genome of melon (*Cucumis melo* L.). Proc Natl Acad Sci USA 109, 11872–11877.
Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.Y., Algire, M.A., Benders, G.A., Montague, M.G., Ma, L., Moodie, M.M., et al. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329, 52–56.
Giri, L., Jugran, A.K., Bahukhandi, A., Dhyani, P., Bhatt, I.D., Rawal, R.S., Nandi, S.K., and Dhar, U. (2017). Population genetic structure and marker trait associations using morphological, phytochemical and molecular parameters in *Habenaria edgeworthii*—a threatened medicinal orchid of West Himalaya, India. Appl Biochem Biotechnol 181, 267–282.
Graham, I.A., Besser, K., Blumer, S., Branigan, C.A., Czechowski, T., Elias, L., Guterman, I., Harvey, D., Isaac, P.G., Khan, A.M., et al. (2010). The genetic map of *Artemisia annua* L. identifies loci affecting yield of the antimalarial drug artemisinin. Science 327, 328–331.
Gualberto, J.M., Mileshina, D., Wallet, C., Niazi, A.K., Weber-Lotfi, F., and Dietrich, A. (2014). The plant mitochondrial genome: dynamics and maintenance. Biochimie 100, 107–120.
Guan, R., Zhao, Y., Zhang, H., Fan, G., Liu, X., Zhou, W., Shi, C., Wang, J., Liu, W., Liang, X., et al. (2016). Draft genome of the living fossil *Ginkgo biloba*. Gigascience 5, 49.
Guo, D., Xue, Y., Li, D., He, B., Jia, X., Dong, X., and Guo, M. (2017). Overexpression of *CtCHS1* increases accumulation of quinochalcone in safflower. Front Plant Sci 8, 1409.
Guo, H., Liu, J., Luo, L., Wei, X., Zhang, J., Qi, Y., Zhang, B., Liu, H., and Xiao, P. (2017). Complete chloroplast genome sequences of *Schisandra chinensis*: genome structure, comparative analysis, and phylogenetic relationship of basal angiosperms. Sci China Life Sci 60, 1286–1290.
Guo, J., Ma, X., Cai, Y., Ma, Y., Zhan, Z., Zhou, Y.J., Liu, W., Guan, M., Yang, J., Cui, G., et al. (2016). Cytochrome P450 promiscuity leads to a bifurcating biosynthetic pathway for tanshinones. New Phytol 210, 5-25–534.
Guo, J., Zhou, Y.J., Hillwig, M.L., Shen, Y., Yang, L., Wang, Y., Zhang, X., Liu, W., Peters, R.J., Chen, X., et al. (2013). CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts. Proc Natl Acad Sci USA 110, 12108–12113.
Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W.J., Zhang, H., Zheng, Y., Mao, L., Ren, Y., Wang, Z., et al. (2013). The draft genome of watermelon (*Citrullus lanatus*) and resequencing of 20 diverse accessions. Nat Genet 45, 51–58.
Haiser, H.J., and Turnbaugh, P.J. (2012). Is it time for a metagenomic basis of therapeutics? Science 336, 1253–1255.
Han, J., Pang, X., Liao, B., Yao, H., Song, J., and Chen, S. (2016). An authenticity survey of herbal medicines from markets in China using DNA barcoding. Sci Rep 6, 18723.
Han, R., Takahashi, H., Nakamura, M., Yoshimoto, N., Suzuki, H., Shibata, D., Yamazaki, M., and Saito, K. (2015a). Transcriptomic landscape of *Pueraria lobata* demonstrates potential for phytochemical study. Front Plant Sci 6, 426.
Han, R., Takahashi, H., Nakamura, M., Bunsupa, S., Yoshimoto, N., Yamamoto, H., Suzuki, H., Shibata, D., Yamazaki, M., and Saito, K. (2-015b). Transcriptome analysis of nine tissues to discover genes involved in the biosynthesis of active ingredients in *Sophora flavescens*. Biol Pharmaceut Bull 38, 876–883.
Hansen, N.L., Heskes, A.M., Hamberger, B., Olsen, C.E., Hallström, B.M., Andersen-Ranberg, J., and Hamberger, B. (2017). The terpene synthase gene family in *Tripterygium wilfordii* harbors a labdane-type diterpene synthase among the monoterpene synthase TPS-b subfamily. Plant J 89, 429–441.
Hao, X., Zhong, Y., Fu, X., Lv, Z., Shen, Q., Yan, T., Shi, P., Ma, Y., Chen, M., Lv, X., et al. (2017). Transcriptome analysis of genes associated with the artemisinin biosynthesis by jasmonic acid treatment under the light in *Artemisia annua*. Front Plant Sci 8, 971.
He, N., Zhang, C., Qi, X., Zhao, S., Tao, Y., Yang, G., Lee, T.H., Wang, X., Cai, Q., Li, D., et al. (2013). Draft genome sequence of the mulberry tree *Morus notabilis*. Nat Commun 4, 2445.
He, Y., Xiao, H., Deng, C., Xiong, L., Nie, H., and Peng, C. (2016). Survey of the genome of *Pogostemon cablin* provides insights into its evolutionary history and sesquiterpenoid biosynthesis. Sci Rep 6, 26405.
Hoshino, A., Jayakumar, V., Nitasaka, E., Toyoda, A., Noguchi, H., Itoh, T., Shin-I, T., Minakuchi, Y., Koda, Y., Nagano, A.J., et al. (2016). Genome sequence and analysis of the Japanese morning glory *Ipomoea nil*. Nat Commun 7, 13295.
Hu, Z., Zhang, T., Gao, X.X., Wang, Y., Zhang, Q., Zhou, H.J., Zhao, G.F., Wang, M.L., Woeste, K.E., and Zhao, P. (2016). *De novo* assembly and characterization of the leaf, bud, and fruit transcriptome from the vulnerable tree *Juglans mandshurica* for the development of 20 new microsatellite markers using Illumina sequencing. Mol Genet Genomics 291, 849–862.
Huang, X., Zhou, G., Wu, W., Duan, Y., Ma, G., Song, J., Xiao, R., Vandenberghe, L., Zhang, F., D'Amore, P.A., et al. (2017). Genome editing abrogates angiogenesis *in vivo*. Nat Commun 8, 112.
Hunt, M., Kikuchi, T., Sanders, M., Newbold, C., Berriman, M., and Otto, T.D. (2013). REAPR: a universal tool for genome assembly evaluation. Genome Biol 14, R47.
Jayakodi, M., Choi, B.S., Lee, S.C., Kim, N.H., Park, J.Y., Jang, W., Lakshmanan, M., Mohan, S.V.G., Lee, D.Y., and Yang, T.J. (2018). Ginseng Genome Database: an open-access platform for genomics of *Panax ginseng*. BMC Plant Biol 18, 62.
Jayakodi, M., Lee, S.C., Lee, Y.S., Park, H.S., Kim, N.H., Jang, W., Lee, H. O., Joh, H.J., and Yang, T.J. (2015). Comprehensive analysis of *Panax ginseng* root transcriptomes. BMC Plant Biol 15, 138.
Ji, A., Jia, J., Xu, Z., Li, Y., Bi, W., Ren, F., He, C., Liu, J., Hu, K., and Song, J. (2017). Transcriptome-guided mining of genes involved in crocin biosynthesis. Front Plant Sci 8, 518.
Jia, J., Xu, Z., Xin, T., Shi, L., and Song, J. (2017). Quality control of the traditional patent medicine Yimu Wan based on SMRT sequencing and DNA barcoding. Front Plant Sci 8, 926.
Kellner, F., Kim, J., Clavijo, B.J., Hamilton, J.P., Childs, K.L., Vaillancourt, B., Cepela, J., Habermann, M., Steuernagel, B., Clissold, L., et al. (2015). Genome-guided investigation of plant natural product biosynthesis. Plant J 82, 680–692.
Kim, B.Y., Park, H.S., Kim, S., and Kim, Y.D. (2017a). Development of microsatellite markers for *Viscum coloratum* (Santalaceae) and their application to wild populations. Appl Plant Sci 5, pii: apps.1600102.
Kim, B.Y., Park, H.S., Lee, J.H., Kwak, M., and Kim, Y.D. (2017b). Development of microsatellite markers based on expressed sequence tags in *Asparagus cochinchinensis* (Asparagaceae). Appl Plant Sci 5, pii: apps. 1700021.
Kim, Y., Kim, A.Y., Jo, A., Choi, H., Cho, S.S., and Choi, C. (2017). Development of user-friendly method to distinguish subspecies of the Korean medicinal herb *Perilla frutescens* using multiplex-PCR. Molecules 22, 665.
Kitashiba, H., Li, F., Hirakawa, H., Kawanabe, T., Zou, Z., Hasegawa, Y., Tonosaki, K., Shirasawa, S., Fukushima, A., Yokoi, S., et al. (2014). Draft sequences of the radish (*Raphanus sativus* L.) genome. DNA Res 21, 481–490.
Kui, L., Chen, H., Zhang, W., He, S., Xiong, Z., Zhang, Y., Yan, L., Zhong, C., He, F., Chen, J., et al. (2016). 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 7, 2036.
Lee, Y.S., Park, H.S., Lee, D.K., Jayakodi, M., Kim, N.H., Koo, H.J., Lee, S.C., Kim, Y.J., Kwon, S.W., and Yang, T.J. (2017). Integrated transcriptomic and metabolomic analysis of five *Panax ginseng* cultivars reveals the dynamics of ginsenoside biosynthesis. Front Plant Sci 8, 1048.
Li, B., Cui, G., Shen, G., Zhan, Z., Huang, L., Chen, J., and Qi, X. (2017).

Targeted mutagenesis in the medicinal plant *Salvia miltiorrhiza*. Sci Rep 7, 43320.
Li, C., Zhu, Y., Guo, X., Sun, C., Luo, H., Song, J., Li, Y., Wang, L., Qian, J., and Chen, S. (2013). Transcriptome analysis reveals ginsenosides biosynthetic genes, microRNAs and simple sequence repeats in *Panax ginseng* C. A. Meyer. BMC Genomics 14, 245.
Li, J., Zhang, J., Chen, H., Chen, X., Lan, J., and Liu, C. (2013). Complete mitochondrial genome of the medicinal mushroom *Ganoderma lucidum*. PLoS ONE 8, e72038.
Li, J., Zhao, P., Yang, L., Li, Y., Tian, Y., and Li, S. (2016a). System biology analysis of long-term effect and mechanism of Bufei Yishen on COPD revealed by system pharmacology and 3-omics profiling. Sci Rep 6, 25492.
Li, J., Xiong, C., Liu, Y., Liang, J., and Zhou, X. (2016b). Loop-mediated isothermal amplification (LAMP): emergence as an alternative technology for herbal medicine identification. Front Plant Sci 7, 1956.
Li, M., Yang, Y., Feng, F., Zhang, B., Chen, S., Yang, C., Gu, L., Wang, F., Zhang, J., Chen, A., et al. (2017). Differential proteomic analysis of replanted *Rehmannia glutinosa* roots by iTRAQ reveals molecular mechanisms for formation of replant disease. BMC Plant Biol 17, 116.
Li, M., Shi, F., Zhou, Y., Li, Y., Wang, X., Zhang, C., Wang, X., Liu, B., Xiao, H., and Li, L. (2015). Genetic and epigenetic diversities shed light on domestication of cultivated ginseng (*Panax ginseng*). Mol Plant 8, 1612–1622.
Li, Q., Li, Y., Song, J., Xu, H., Xu, J., Zhu, Y., Li, X., Gao, H., Dong, L., Qian, J., et al. (2014). High-accuracy *de novo* assembly and SNP detection of chloroplast genomes using a SMRT circular consensus sequencing strategy. New Phytol 204, 1041–1049.
Li, Y., Ba, H., and Yang, F. (2016). Complete mitochondrial genome of *Cervus elaphus songaricus* (Cetartiodactyla: Cervinae) and a phylogenetic analysis with related species. Mitochondrial DNA 27, 620–621.
Li, Y., Hu, X.D., Yang, R.H., Hsiang, T., Wang, K., Liang, D.Q., Liang, F., Cao, D.M., Zhou, F., Wen, G., et al. (2015a). Complete mitochondrial genome of the medicinal fungus *Ophiocordyceps sinensis*. Sci Rep 5, 13892.
Li, Y., Wang, X., Chen, T., Yao, F., Li, C., Tang, Q., Sun, M., Sun, G., Hu, S., Yu, J., and Song, S. (2015b). RNA-Seq based *de novo* transcriptome assembly and gene discovery of *Cistanche deserticola* fleshy stem. PLoS ONE 10, e125722.
Liu, F., Guo, Q.S., Shi, H.Z., Cheng, B.X., Lu, Y.X., Gou, L., Wang, J., Shen, W.B., Yan, S.M., and Wu, M.J. (2016). Genetic variation in *Whitmania pigra*, *Hirudo nipponica* and *Poecilobdella manillensis*, three endemic and endangered species in China using SSR and TRAP markers. Gene 579, 172–182.
Liu, M.H., Yang, B.R., Cheung, W.F., Yang, K.Y., Zhou, H.F., Kwok, J.S. L., Liu, G.C., Li, X.F., Zhong, S., Lee, S.M.Y., et al. (2015). Transcriptome analysis of leaves, roots and flowers of *Panax notoginseng* identifies genes involved in ginsenoside and alkaloid biosynthesis. BMC Genomics 16, 265.
Liu, M.J., Zhao, J., Cai, Q.L., Liu, G.C., Wang, J.R., Zhao, Z.H., Liu, P., Dai, L., Yan, G., Wang, W.J., et al. (2014). The complex jujube genome provides insights into fruit tree biology. Nat Commun 5, 5315.
Liu, S., Chen, J., Li, S.C., Zeng, X., Meng, Z., and Guo, S. (2015). Comparative transcriptome analysis of genes involved in GA-GID1-DELLA regulatory module in symbiotic and asymbiotic seed germination of *Anoectochilus roxburghii* (Wall.) Lindl. (Orchidaceae). Int J Mol Sci 16, 30190–30203.
Liu, X., Cheng, J., Zhang, G., Ding, W., Duan, L., Yang, J., Kui, L., Cheng, X., Ruan, J., Fan, W., et al. (2018). Engineering yeast for the production of breviscapine by genomic analysis and synthetic biology approaches. Nat Commun 9, 448.
Liu, X., Liu, Y., Huang, P., Ma, Y., Qing, Z., Tang, Q., Cao, H., Cheng, P., Zheng, Y., Yuan, Z., et al. (2017). The genome of medicinal plant *Macleaya cordata* provides new insights into benzylisoquinoline alkaloids metabolism. Mol Plant 10, 975–989.
Liu, Y., Liu, X., and Zhang, M. (2016). The complete mitochondrial genome of Sika deer *Cervus nippon hortulorum* (Artiodactyla: Cervidae) and phylogenetic studies. Mitochondrial DNA A DNA Mapp Seq Anal 27, 2967–2968.
Lu, X., Zhang, L., Zhang, F., Jiang, W., Shen, Q., Zhang, L., Lv, Z., Wang, G., and Tang, K. (2013). *AaORA*, a trichome-specific AP2/ERF transcription factor of *Artemisia annua*, is a positive regulator in the artemisinin biosynthetic pathway and in disease resistance to *Botrytis cinerea*. New Phytol 198, 1191–1202.
Luo, H., Sun, C., Sun, Y., Wu, Q., Li, Y., Song, J., Niu, Y., Cheng, X., Xu, H., Li, C., et al. (2011). Analysis of the transcriptome of *Panax notoginseng* root uncovers putative triterpene saponin-biosynthetic genes and genetic markers. BMC Genomics 12, S5.
Ma, C., Gao, Z., Zhang, J., Zhang, W., Shao, J., Hai, M., Chen, J., Yang, S., and Zhang, G. (2016). Candidate genes involved in the biosynthesis of triterpenoid saponins in *Platycodon grandiflorum* identified by transcriptome analysis. Front Plant Sci 7, 673.
Ma, Y.S., Yu, H., Li, Y.Y., Yan, H., and Cheng, X. (2008). A study of genetic structure of *Stephania yunnanensis* (Menispermaceae) by DALP. Biochem Genet 46, 227–240.
Martel, J., Ojcius, D.M., Chang, C.J., Lin, C.S., Lu, C.C., Ko, Y.F., Tseng, S.F., Lai, H.C., and Young, J.D. (2017). Anti-obesogenic and antidiabetic effects of plants and mushrooms. Nat Rev Endocrinol 13, 149–160.
Martin, V.J.J., Pitera, D.J., Withers, S.T., Newman, J.D., and Keasling, J.D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nat Biotechnol 21, 796–802.
Mayer, K.F.X., Waugh, R., Brown, J.W.S., Schulman, A., Langridge, P., Platzer, M., Fincher, G.B., Muehlbauer, G.J., Sato, K., et al. (2012). A physical, genetic and functional sequence assembly of the barley genome. Nature 491, 711–716.
Mayr, C. (2016). Evolution and biological roles of alternative 3′UTRs. Trends Cell Biol 26, 227–237.
Mehta, S.R., Eikelboom, J.W., Rupprecht, H.J., Lewis, B.S., Natarajan, M. K., Yi, C., Pogue, J., and Yusuf, S. (2002). Efficacy of hirudin in reducing cardiovascular events in patients with acute coronary syndrome undergoing early percutaneous coronary intervention. Eur Heart J 23, 117–123.
Miao, Y., Zhu, Z., Guo, Q., Zhu, Y., Yang, X., and Sun, Y. (2016). Transcriptome analysis of differentially expressed genes provides insight into stolon formation in *Tulipa edulis*. Front Plant Sci 7, 409.
Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Pollier, J., Woittiez, L., van der Krol, S., Lugan, R., Ilc, T., et al. (2014). The seco-iridoid pathway from *Catharanthus roseus*. Nat Commun 5, 3606.
Ming, R., VanBuren, R., Liu, Y., Yang, M., Han, Y., Li, L.T., Zhang, Q., Kim, M.J., Schatz, M.C., Campbell, M., et al. (2013). Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). Genome Biol 14, R41.
Mochida, K., Sakurai, T., Seki, H., Yoshida, T., Takahagi, K., Sawai, S., Uchiyama, H., Muranaka, T., and Saito, K. (2017). Draft genome assembly and annotation of *Glycyrrhiza uralensis*, a medicinal legume. Plant J 89, 181–194.
Montenegro, D., Kalpana, K., Chrissian, C., Sharma, A., Takaoka, A., Iacovidou, M., Soll, C.E., Aminova, O., Heguy, A., Cohen, L., et al. (2015). Uncovering potential 'herbal probiotics' in Juzen-taiho-to through the study of associated bacterial populations. Bioorg Med Chem Lett 25, 466–469.
Nguyen, H.T., Min, J.E., Long, N.P., Thanh, M.C., Le, T.H.V., Lee, J., Park, J.H., and Kwon, S.W. (2017). Multi-platform metabolomics and a genetic approach support the authentication of agarwood produced by *Aquilaria crassna* and *Aquilaria malaccensis*. J Pharmaceut Biomed Anal 142, 136–144.
Nguyen, K.N.T., Nguyen, G.K.T., Nguyen, P.Q.T., Ang, K.H., Dedon, P.C., and Tam, J.P. (2016). Immunostimulating and Gram-negative-specific antibacterial cyclotides from the butterfly pea (*Clitoria ternatea*). FEBS J 283, 2067–2090.
Nicholson, J.K., Lindon, J.C., and Holmes, E. (1999). "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR

spectroscopic data. Xenobiotica 29, 1181–1189.
Oliveira, T.G., Pereira, A.M., Coppede, J.S., Franca, S.C., Ming, L.C., and Bertoni, B.W. (2016). Genetic diversity analysis of *Croton antisyphiliticus* Mart. using AFLP molecular markers. Genet Mol Res 15, doi: 10.4238/gmr.15017461.
Otto, L.G., Mondal, P., Brassac, J., Preiss, S., Degenhardt, J., He, S., Reif, J.C., and Sharbel, T.F. (2017). Use of genotyping-by-sequencing to determine the genetic structure in the medicinal plant chamomile, and to identify flowering time and alpha-bisabolol associated SNP-loci by genome-wide association mapping. BMC Genomics 18, 599.
Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D., et al. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 496, 528–532.
Paddon, C.J., and Keasling, J.D. (2014). Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. Nat Rev Micro 12, 355–367.
Pan, T.L., Wang, P.W., Hung, Y.C., Huang, C.H., and Rau, K.M. (2013). Proteomic analysis reveals tanshinone IIA enhances apoptosis of advanced cervix carcinoma CaSki cells through mitochondria intrinsic and endoplasmic reticulum stress pathways. Proteomics 13, 3411–3423.
Park, S., Ruhlman, T.A., Sabir, J.S.M., Mutwakil, M.H.Z., Baeshen, M.N., Sabir, M.J., Baeshen, N.A., and Jansen, R.K. (2014). Complete sequences of organelle genomes from the medicinal plant *Rhazya stricta* (Apocynaceae) and contrasting patterns of mitochondrial genome evolution across asterids. BMC Genomics 15, 405.
Qin, G., Xu, C., Ming, R., Tang, H., Guyot, R., Kramer, E.M., Hu, Y., Yi, X., Qi, Y., Xu, X., et al. (2017). The pomegranate (*Punica granatum* L.) genome and the genomics of punicalagin biosynthesis. Plant J 91, 1108–1128.
Rai, A., Kamochi, H., Suzuki, H., Nakamura, M., Takahashi, H., Hatada, T., Saito, K., and Yamazaki, M. (2017). *De novo* transcriptome assembly and characterization of nine tissues of *Lonicera japonica* to identify potential candidate genes involved in chlorogenic acid, luteolosides, and secoiridoid biosynthesis pathways. J Nat Med 71, 1–15.
Rai, A., Nakamura, M., Takahashi, H., Suzuki, H., Saito, K., and Yamazaki, M. (2016). High-throughput sequencing and *de novo* transcriptome assembly of *Swertia japonica* to identify genes involved in the biosynthesis of therapeutic metabolites. Plant Cell Rep 35, 2091–2111.
Raymond, O., Gouzy, J., Just, J., Badouin, H., Verdenaud, M., Lemainque, A., Vergne, P., Moja, S., Choisne, N., Pont, C., et al. (2018). The *Rosa* genome provides new insights into the domestication of modern roses. Nat Genet 50, 772–777.
Saki, S., Bagheri, H., Deljou, A., and Zeinalabedini, M. (2016). Evaluation of genetic diversity amongst *Descurainia sophia* L. genotypes by inter-simple sequence repeat (ISSR) marker. Physiol Mol Biol Plants 22, 97–105.
Salzberg, S.L., Phillippy, A.M., Zimin, A., Puiu, D., Magoc, T., Koren, S., Treangen, T.J., Schatz, M.C., Delcher, A.L., Roberts, M., et al. (2012). GAGE: a critical evaluation of genome assemblies and assembly algorithms. Genome Res 22, 557–567.
Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., Hyten, D.L., Song, Q., Thelen, J.J., Cheng, J., et al. (2010). Genome sequence of the palaeopolyploid soybean. Nature 463, 178–183.
Schuster, S.C. (2008). Next-generation sequencing transforms today's biology. Nat Methods 5, 16–18.
Shan, Q., Wang, Y., Li, J., and Gao, C. (2014). Genome editing in rice and wheat using the CRISPR/Cas system. Nat Protoc 9, 2395–2410.
Shen, L., Hua, Y., Fu, Y., Li, J., Liu, Q., Jiao, X., Xin, G., Wang, J., Wang, X., Yan, C., et al. (2017). Rapid generation of genetic diversity by multiplex CRISPR/Cas9 genome editing in rice. Sci China Life Sci 60, 506–515.
Shen, Q., Zhang, L., Liao, Z., Wang, S., Yan, T., Shi, P., Liu, M., Fu, X., Pan, Q., Wang, Y., et al. (2018). The genome of *Artemisia annua* provides insight into the evolution of asteraceae family and artemisinin biosynthesis. Mol Plant 11, 776–788.
Simakov, O., Marletaz, F., Cho, S.J., Edsinger-Gonzales, E., Havlak, P., Hellsten, U., Kuo, D.H., Larsson, T., Lv, J., Arendt, D., et al. (2013). Insights into bilaterian evolution from three spiralian genomes. Nature 493, 526–531.
Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212.
Song, J., Luo, H., Li, C., Sun, C., Xu, J., and Chen, S. (2013). *Salvia miltiorrhiza* as medicinal model plant. Acta Pharmaceut Sin 48, 1099–1106.
State Pharmacopoeia Committee. (2015). Pharmacopoeia of the People's Republic of China Part IV. 383–385.
Sun, C., Hu, Y.L., Xu, J., Luo, H.M., Li, C.F., Song, J.Y., Guo, H.W., and Chen, S.L. (2013). *Ganoderma lucidum*: an emerging medicinal model fungus for study of the biosynthesis of natural medicines (in Chinese). Sci Sin Vitae 43, 447–457.
Sun, H., Liu, F., Sun, L., Liu, J., Wang, M., Chen, X., Xu, X., Ma, R., Feng, K., and Jiang, R. (2016). Proteomic analysis of amino acid metabolism differences between wild and cultivated *Panax ginseng*. J Ginseng Res 40, 113–120.
Sun, J., Manmathan, H., Sun, C., and Peebles, C.A.M. (2016). Examining the transcriptional response of overexpressing anthranilate synthase in the hairy roots of an important medicinal plant *Catharanthus roseus* by RNA-seq. BMC Plant Biol 16, 108.
Sun, W., Li, J.J., Xiong, C., Zhao, B., and Chen, S.L. (2016). The potential power of bar-HRM technology in herbal medicine identification. Front Plant Sci 7, 367.
Sun, W., Wang, B., Yang, J., Wang, W., Liu, A., Leng, L., Xiang, L., Song, C., and Chen, S. (2017). Weighted gene co-expression network analysis of the dioscin rich medicinal plant *Dioscorea nipponica*. Front Plant Sci 8, 789.
Tamura, K., Teranishi, Y., Ueda, S., Suzuki, H., Kawano, N., Yoshimatsu, K., Saito, K., Kawahara, N., Muranaka, T., and Seki, H. (2017). Cytochrome P450 monooxygenase CYP716A141 is a unique β-amyrin C-16β oxidase involved in triterpenoid saponin biosynthesis in *Platycodon grandiflorus*. Plant Cell Physiol 58, 874–884.
The British Pharmacopoeia Commission. (2015). British Pharmacopoeia Appendix XI V. Deoxyribonucleic Acid (DNA) Based Identification Techniques for Herbal Drugs.
Tian, Z., Zhang, F., Liu, H., Gao, Q., and Chen, S. (2016). Development of SSR markers for a Tibetan medicinal plant, *Lancea tibetica* (Phrymaceae), based on RAD sequencing. Appl Plant Sci 4, pii: apps. 1600076.
Tomasello, S., and Heubl, G. (2017). Phylogenetic analysis and molecular characterization of *Xanthium sibiricum* using DNA barcoding, PCR-RFLP, and specific primers. Planta Med 83, 946–953.
Tripathi, N., Saini, N., Nair, P., and Tiwari, S. (2012). Lack of genetic diversity of a critically endangered important medicinal plant *Chlorophytum borivilianum* in Central India revealed by AFLP markers. Physiol Mol Biol Plants 18, 161–167.
Tu, Y. (2011). The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. Nat Med 17, 1217–1220.
Urasaki, N., Takagi, H., Natsume, S., Uemura, A., Taniai, N., Miyagi, N., Fukushima, M., Suzuki, S., Tarora, K., Tamaki, M., et al. (2017). Draft genome sequence of bitter gourd (*Momordica charantia*), a vegetable and medicinal plant in tropical and subtropical regions. DNA Res 24, 51–58.
van Bakel, H., Stout, J.M., Cote, A.G., Tallon, C.M., Sharpe, A.G., Hughes, T.R., and Page, J.E. (2011). The draft genome and transcriptome of *Cannabis sativa*. Genome Biol 12, R102.
Vashisht, I., Mishra, P., Pal, T., Chanumolu, S., Singh, T.R., and Chauhan, R.S. (2015). Mining NGS transcriptomes for miRNAs and dissecting their role in regulating growth, development, and secondary metabolites production in different organs of a medicinal herb, *Picrorhiza kurroa*. Planta 241, 1255–1268.
Verde, I., Abbott, A.G., Scalabrin, S., Jung, S., Shu, S., Marroni, F., Zhebentyayeva, T., Dettori, M.T., Grimwood, J., Cattonaro, F., et al. (2013). The high-quality draft genome of peach (*Prunus persica*) identifies

unique patterns of genetic diversity, domestication and genome evolution. Nat Genet 45, 487–494.
Vining, K.J., Johnson, S.R., Ahkami, A., Lange, I., Parrish, A.N., Trapp, S. C., Croteau, R.B., Straub, S.C.K., Pandelova, I., and Lange, B.M. (2017). Draft genome sequence of *Mentha longifolia* and development of resources for mint cultivar improvement. Mol Plant 10, 323–339.
Wang, J., Li, J., Li, J., Liu, S., Wu, X., Li, J., and Gao, W. (2016a). Transcriptome profiling shows gene regulation patterns in ginsenoside pathway in response to methyl jasmonate in *Panax quinquefolium* adventitious root. Sci Rep 6, 37263.
Wang, L., Xiao, A.H., Ma, L.Y., Chen, F.J., Sang, Z.Y., and Duan, J. (2017). Identification of *Magnolia wufengensis* (Magnoliaceae) cultivars using phenotypic traits, SSR and SRAP markers: insights into breeding and conservation. Genet Mol Res in press doi: 10.4238/gmr16019473.
Wang, L., Yu, S., Tong, C., Zhao, Y., Liu, Y., Song, C., Zhang, Y., Zhang, X., Wang, Y., Hua, W., et al. (2014). Genome sequencing of the high oil crop sesame provides insight into oil biosynthesis. Genome Biol 15, R39.
Wang, Y., Dong, C., Xue, Z., Jin, Q., and Xu, Y. (2016b). *De novo* transcriptome sequencing and discovery of genes related to copper tolerance in *Paeonia ostii*. Gene 576, 126–135.
Wang, Y., Fan, G., Liu, Y., Sun, F., Shi, C., Liu, X., Peng, J., Chen, W., Huang, X., Cheng, S., et al. (2013). The sacred lotus genome provides insights into the evolution of flowering plants. Plant J 76, 557–567.
Wang, Z., Hobson, N., Galindo, L., Zhu, S., Shi, D., McDill, J., Yang, L., Hawkins, S., Neutelings, G., Datla, R., et al. (2012). The genome of flax (*Linum usitatissimum*) assembled *de novo* from short shotgun sequence reads. Plant J 72, 461–473.
Westfall, P.J., Pitera, D.J., Lenihan, J.R., Eng, D., Woolard, F.X., Regentin, R., Horning, T., Tsuruta, H., Melis, D.J., Owens, A., et al. (2012). Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. Proc Natl Acad Sci USA 109, E111–E118.
Winzer, T., Gazda, V., He, Z., Kaminski, F., Kern, M., Larson, T.R., Li, Y., Meade, F., Teodor, R., Vaistij, F.E., et al. (2012). A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noscapine. Science 336, 1704–1708.
Wu, K., Liu, H., Yang, M., Tao, Y., Ma, H., Wu, W., Zuo, Y., and Zhao, Y. (2014). High-density genetic map construction and QTLs analysis of grain yield-related traits in sesame (*Sesamum indicum* L.) based on RAD-Seq techonology. BMC Plant Biol 14, 274.
Wu, L., Wang, B., Zhao, M., Liu, W., Zhang, P., Shi, Y., Xiong, C., Wang, P., Sun, W., and Chen, S. (2016). Rapid identification of Officinal Akebiae Caulis and its toxic adulterant Aristolochiae Manshuriensis Caulis (*Aristolochia manshuriensis*) by loop-mediated isothermal amplification. Front Plant Sci 7, 887.
Wu, M., Wu, Y., Deng, B., Li, J., Cao, H., Qu, Y., Qian, X., and Zhong, G. (2016). Isoliquiritigenin decreases the incidence of colitis-associated colorectal cancer by modulating the intestinal microbiota. Oncotarget 7, 85318–85331.
Wu, M.C., Lu, T.H., and Lu, K.H. (2017). PCR-RFLP of mitochondrial DNA reveals two origins of *Apis mellifera* in Taiwan. Saudi J Biol Sci 24, 1069–1074.
Xia, E.H., Zhang, H.B., Sheng, J., Li, K., Zhang, Q.J., Kim, C., Zhang, Y., Liu, Y., Zhu, T., Li, W., et al. (2017). The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. Mol Plant 10, 866–877.
Xin, T., Su, C., Lin, Y., Wang, S., Xu, Z., and Song, J. (2018b). Precise species detection of traditional Chinese patent medicine by shotgun metagenomic sequencing. Phytomedicine 47, 40–47.
Xin, T., Xu, Z., Jia, J., Leon, C., Hu, S., Lin, Y., Ragupathy, S., Song, J., and Newmaster, S.G. (2018a). Biomonitoring for traditional herbal medicinal products using DNA metabarcoding and single molecule, real-time sequencing. Acta Pharmaceut Sin B 8, 488–497.
Xin, T., Yao, H., Gao, H., Zhou, X., Ma, X., Xu, C., Chen, J., Han, J., Pang, X., Xu, R., et al. (2013). Super food *Lycium barbarum* (Solanaceae) traceability via an internal transcribed spacer 2 barcode. Food Res Int 54, 1699–1704.
Xing, P., Liu, T., Song, Z., and Li, X. (2016). Genetic diversity of *Toona sinensis* Roem in China revealed by ISSR and SRAP markers. Genet Mol Res 15, 15038387.
Xu, H., Song, J., Luo, H., Zhang, Y., Li, Q., Zhu, Y., Xu, J., Li, Y., Song, C., Wang, B., et al. (2016). Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza*. Mol Plant 9, 949–952.
Xu, J., Chu, Y., Liao, B., Xiao, S., Yin, Q., Bai, R., Su, H., Dong, L., Li, X., Qian, J., et al. (2017). *Panax ginseng* genome examination for ginsenoside biosynthesis. Gigascience 6, 1–15.
Xu, J., Sun, C., Xu, Z.C., Ji, A.J., Hu, Y.L., Sun, W., Wang, L.Z., Wang, B., Yang, P., Zhang, X., et al. (2014). Research strategy for model medicinal species (in Chinese). Chin Sci Bull (Chin Ver) 59, 733–742.
Xu, L., and Huang, H. (2014). Genetic and epigenetic controls of plant regeneration. Curr Top Dev Biol 108, 1–33.
Xu, Y., Lu, Y., Xie, C., Gao, S., Wan, J., and Prasanna, B.M. (2012). Whole-genome strategies for marker-assisted plant breeding. Mol Breeding 29, 833–854.
Xu, Z., Ji, A., Song, J., and Chen, S. (2016a). Genome-wide analysis of auxin response factor gene family members in medicinal model plant *Salvia miltiorrhiza*. Biol Open 5, 848–857.
Xu, Z., Ji, A., Zhang, X., Song, J., and Chen, S. (2016b). Biosynthesis and regulation of active compounds in medicinal model plant *Salvia miltiorrhiza*. Chin Herbal Meds 8, 3–11.
Xu, Z., Luo, H., Ji, A., Zhang, X., Song, J., and Chen, S. (2016c). Global identification of the full-length transcripts and alternative splicing related to phenolic acid biosynthetic genes in *Salvia miltiorrhiza*. Front Plant Sci 7, 100.
Xu, Z., Peters, R.J., Weirather, J., Luo, H., Liao, B., Zhang, X., Zhu, Y., Ji, A., Zhang, B., Hu, S., et al. (2015). Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. Plant J 82, 951–961.
Xu, Z., and Song, J. (2017). The 2-oxoglutarate-dependent dioxygenase superfamily participates in tanshinone production in *Salvia miltiorrhiza*. J Exp Bot 68, 2299–2308.
Xu, Z., Xin, T., Bartels, D., Li, Y., Gu, W., Yao, H., Liu, S., Yu, H., Pu, X., Zhou, J., et al. (2018). Genome analysis of the ancient tracheophyte *Selaginella tamariscina* reveals evolutionary features relevant to the acquisition of desiccation tolerance. Mol Plant 11, 983–994.
Yan, L., Wang, X., Liu, H., Tian, Y., Lian, J., Yang, R., Hao, S., Wang, X., Yang, S., Li, Q., et al. (2015). The genome of *Dendrobium officinale* illuminates the biology of the important traditional Chinese orchid herb. Mol Plant 8, 922–934.
Yang, J., Zhang, G., Zhang, J., Liu, H., Chen, W., Wang, X., Li, Y., Dong, Y., and Yang, S. (2017). Hybrid *de novo* genome assembly of the Chinese herbal fleabane *Erigeron breviscapus*. Gigascience 6, 1–7.
Yang, K., Tian, Z., Chen, C., Luo, L., Zhao, B., Wang, Z., Yu, L., Li, Y., Sun, Y., Li, W., et al. (2015). Genome sequencing of adzuki bean (*Vigna angularis*) provides insight into high starch and low fat accumulation and domestication. Proc Natl Acad Sci USA 112, 13213–13218.
Yang, Y., Li, M., Li, X., Chen, X., Lin, W., and Zhang, Z. (2015). Transcriptome-wide identification of the genes responding to replanting disease in *Rehmannia glutinosa* L. roots. Mol Biol Rep 42, 881–892.
Ye, N., Zhang, X., Miao, M., Fan, X., Zheng, Y., Xu, D., Wang, J., Zhou, L., Wang, D., Gao, Y., et al. (2015). *Saccharina* genomes provide novel insight into kelp biology. Nat Commun 6, 6986.
Yu, Z.X., Li, J.X., Yang, C.Q., Hu, W.L., Wang, L.J., and Chen, X.Y. (2012). The jasmonate-responsive AP2/ERF transcription factors *AaERF1* and *AaERF2* positively regulate artemisinin biosynthesis in *Artemisia annua* L.. Mol Plant 5, 353–365.
Yuan, Y., Jin, X., Liu, J., Zhao, X., Zhou, J., Wang, X., Wang, D., Lai, C., Xu, W., Huang, J., et al. (2018). The *Gastrodia elata* genome provides insights into plant adaptation to heterotrophy. Nat Commun 9, 1615.
Zhan, C., Li, X., Zhao, Z., Yang, T., Wang, X., Luo, B., Zhang, Q., Hu, Y., and Hu, X. (2016). Comprehensive analysis of the triterpenoid saponins biosynthetic pathway in *Anemone flaccida* by transcriptome and prot-

eome profiling. Front Plant Sci 7, 1094.
Zhang, C., Mei, Z., Cheng, J., He, Y., Khan, M.A., Luo, P., Imani, S., and Fu, J. (2015). Development of SCAR markers based on improved R-APD amplification fragments and molecular cloning for authentication of herbal medicines *Angelica sinensis*, *Angelica acutiloba* and *Levisticum officinale*. Nat Prod Commun 10, 1743–1747.
Zhang, D., Li, W., Xia, E.H., Zhang, Q.J., Liu, Y., Zhang, Y., Tong, Y., Zhao, Y., Niu, Y.C., Xu, J.H., et al. (2017). The medicinal herb *Panax notoginseng* genome provides insights into ginsenoside biosynthesis and genome evolution. Mol Plant 10, 903–907.
Zhang, G., Fang, X., Guo, X., Li, L., Luo, R., Xu, F., Yang, P., Zhang, L., Wang, X., Qi, H., et al. (2012). The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490, 49–54.
Zhang, G., Jiang, N., Song, W., Ma, C., Yang, S., and Chen, J. (2016). *De novo* sequencing and transcriptome analysis of *Pinellia ternata* identify the candidate genes involved in the biosynthesis of benzoic acid and ephedrine. Front Plant Sci 7, 1209.
Zhang, H., Hedhili, S., Montiel, G., Zhang, Y., Chatel, G., Pré, M., Gantet, P., and Memelink, J. (2011). 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 67, 61–71.
Zhang, J., Li, X., Lu, F., Wang, S., An, Y., Su, X., Li, X., Ma, L., and Han, G. (2017b). *De novo* sequencing and transcriptome analysis reveal key genes regulating steroid metabolism in leaves, roots, adventitious roots and calli of *Periploca sepium* Bunge. Front Plant Sci 8, 594.
Zhang, J., Tian, Y., Yan, L., Zhang, G., Wang, X., Zeng, Y., Zhang, J., Ma, X., Tan, Y., Long, N., et al. (2016). Genome of plant maca (*Lepidium meyenii*) illuminates genomic basis for high-altitude adaptation in the Central Andes. Mol Plant 9, 1066–1077.
Zhang, J., Su, H., Zhang, L., Liao, B., Xiao, S., Dong, L., Hu, Z., Wang, P., Li, X., Huang, Z., et al. (2017a). Comprehensive characterization for ginsenosides biosynthesis in ginseng root by integration analysis of chemical and transcriptome. Molecules 22, 889.
Zhang, L., Hou, D., Chen, X., Li, D., Zhu, L., Zhang, Y., Li, J., Bian, Z., Liang, X., Cai, X., et al. (2012). Exogenous plant *MIR168a* specifically targets mammalian *LDLRAP1*: evidence of cross-kingdom regulation by microRNA. Cell Res 22, 107–126.
Zhang, L., Long, Y., Fu, C., Xiang, J., Gan, J., Wu, G., Jia, H., Yu, L., and Li, M. (2016). Different gene expression patterns between leaves and flowers in *Lonicera japonica* revealed by transcriptome analysis. Front Plant Sci 7, 637.
Zhang, X., Luo, H., Xu, Z., Zhu, Y., Ji, A., Song, J., and Chen, S. (2015). Genome-wide characterisation and analysis of *bHLH* transcription factors related to tanshinone biosynthesis in *Salvia miltiorrhiza*. Sci Rep 5, 11244.
Zhang, Y., Xu, Z., Ji, A., Luo, H., and Song, J. (2018). Genomic survey of bZIP transcription factor genes related to tanshinone biosynthesis in *Salvia miltiorrhiza*. Acta Pharmaceut Sin B 8, 295–305.
Zhang, Z., Fan, M., Hao, X., Qin, X., and Li, Z. (2016). Integrative drug efficacy assessment of Danggui and European Danggui using NMR-based metabolomics. J Pharmaceut Biomed Anal 120, 1–9.
Zhao, D., Hamilton, J.P., Pham, G.M., Crisovan, E., Wiegert-Rininger, K., Vaillancourt, B., Dellapenna, D., and Buell, C.R. (2017). *De novo* genome assembly of *Camptotheca acuminata*, a natural source of the anti-cancer compound camptothecin. Gigascience 6, 1–7.
Zhao, L., Zhang, F., Ding, X., Wu, G., Lam, Y.Y., Wang, X., Fu, H., Xue, X., Lu, C., Ma, J., et al. (2018). Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. Science 359, 1151–1156.
Zheng, F., Wang, B., and Wang, Q. (2016). Complete mitochondrial genome of the *Sepiella maindroni* (Sepioidea: Sepiidae). Mitochondrial DNA A DNA Mapp Seq Anal 27, 3766–3767.
Zheng, Y., Xu, S., Liu, J., Zhao, Y., and Liu, J. (2017). Genetic diversity and population structure of Chinese natural bermudagrass (*Cynodon dactylon* (L.) Pers.) germplasm based on SRAP markers. PLoS ONE 12, e177508.
Zhou, S.S., Xu, J., Zhu, H., Wu, J., Xu, J.D., Yan, R., Li, X.Y., Liu, H.H., Duan, S.M., Wang, Z., et al. (2016). Gut microbiota-involved mechanisms in enhancing systemic exposure of ginsenosides by coexisting polysaccharides in ginseng decoction. Sci Rep 6, 22474.
Zhou, W., Huang, Q., Wu, X., Zhou, Z., Ding, M., Shi, M., Huang, F., Li, S., Wang, Y., and Kai, G. (2017). Comprehensive transcriptome profiling of *Salvia miltiorrhiza* for discovery of genes associated with the biosynthesis of tanshinones and phenolic acids. Sci Rep 7, 10554.
Zhou, Z., Li, X., Liu, J., Dong, L., Chen, Q., Liu, J., Kong, H., Zhang, Q., Qi, X., Hou, D., et al. (2015). Honeysuckle-encoded atypical microRNA2911 directly targets influenza A viruses. Cell Res 25, 39–49.
Zhou, Z., Tan, H., Li, Q., Chen, J., Gao, S., Wang, Y., Chen, W., and Zhang, L. (2018). CRISPR/Cas9-mediated efficient targeted mutagenesis of *RAS* in *Salvia miltiorrhiza*. Phytochemistry 148, 63–70.
Zhu, Y., Xu, J., Sun, C., Zhou, S., Xu, H., Nelson, D.R., Qian, J., Song, J., Luo, H., Xiang, L., et al. (2015). Chromosome-level genome map provides insights into diverse defense mechanisms in the medicinal fungus *Ganoderma sinense*. Sci Rep 5, 11087.

SUPPORTING INFORMATION

Table S1 Several key enzymes involved in the biosynthetic pathway of specific active compounds

The supporting information is available online at http://life.scichina.com and https://link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.