CHAPTER

1

Genomics and Evolution of Medicinal Plants

OUTLINE

1.1 INTRODUCTION

There are more than 300,000 species of extant seed plants around the globe (Jiao et al., 2011a,b). About 60% of plants have medicinal use in post-Neolithic human history. Nowadays, people collect plants for medicinal use from not only wild environments but also artificial cultivation, which is an indispensable part of human civilization. There are over 10,000 medicinal plant species in China, accounting for c. 87% of the Chinese materia medica (CMM) (Chen et al., 2010). Medicinal plants are also essential raw materials of many chemical drugs, for example, the blockbuster drugs for antimalarial and anticancer therapies. Currently more than one-third of clinical drugs are from botanical extracts and/or their derivatives. Unfortunately, most medicinal plants have not been domesticated, and currently there is no toolkit

Ranunculales Medicinal Plants. http://dx.doi.org/10.1016/B978-0-12-814232-5.00001-0

FIGURE 1.1 ***Taxus cuspidata* var. nana.** *Source: Taken in Dalian Jiaotong University, China.*

to improve their medicinal attributes for better clinical efficacy. Immoderate harvesting has led to a supply crisis of phytomedicine, exemplifying in taxane-producing *Taxus* plants (Hao et al., 2012a) (Fig. 1.1). On the other hand, successful domestication and improvement are not realistic without deeper insights into the evolutionary pattern of medicinal plant genomes. Artificial selection can be regarded as an accelerated and targeted natural selection. Studies of medicinal plant genome evolution are crucial to not only the ubiquitous mechanisms of plant evolution and phylogeny but also plant-based drug discovery and development, as well as the sustainable utilization of plant pharmaceutical resources. This chapter presents a preliminary examination of the recent developments in medicinal plant genome evolution research and summarizes the benefits, gaps, and prospects of the current research topics.

1.2 EVOLUTION OF GENOME, GENE, AND GENOTYPE

1.2.1 Genome Sequencing

The genomic studies of medicinal plants lag behind those of model plants and important crop plants. The genome sequences encompass essential information of plant origin, evolution, development, physiology, inheritable traits, epigenomic regulation, etc. These elements are the premise and foundation of deciphering genome diversity and chemodiversity (especially various secondary metabolites with potential bioactivities) at the molecular level. The high-throughput sequencing of medicinal plants could not only shed light on the biosynthetic pathways of medicinal compounds, especially secondary metabolites (Boutanaev et al., 2015), and their regulation mechanisms but also play a major role in the molecular breeding of high-yield medicinal cultivars and molecular farming of transgenic medicinal strains.

A few principles should be considered when selecting medicinal plants for whole genome sequencing projects. First, the source plants of well-known and expensive CMMs or important chemical drugs that are in heavy demand have priority, for example, *Panax ginseng* (Chen et al., 2011; Zhao et al., 2015) and *Artemisia annua* (Moses et al., 2015) (Fig. 1.2); second, the representative plants whose pharmaceutical components are relatively unambiguous and that have typical secondary metabolism pathways, for example, *Salvia* medicinal plants (Hao et al., 2015a,b); third, the characteristic plants that are in the large medicinal genus/family, such as *Glycyrrhiza uralensis* (Chinese liquorice; *Fabaceae*) (Hao et al., 2012b, 2015c)

FIGURE 1.2 ***Artemisia annua*** **of Asteraceae.** *Source: Taken in Tashilunpo Monastery, Tibet, China.*

and *Lycium chinense* (Chinese boxthorn; *Solanaceae*) (Yao et al., 2011); fourth, the medicinal plants that are potential model plants and have considerable biological data; and last, the medicinal plants whose genetic backgrounds are known, with reasonably small diploid genomes and relatively straightforward genome structures, are preferred.

As there is a lack of comprehensive molecular genetic studies for most medicinal plants, it is vital to have some preliminary genome evaluations before the whole genome sequencing. First, DNA barcoding techniques (Hao et al., 2012c) could be used to authenticate the candidate species; second, karyotypes should be determined by observing metaphase chromosomes; and last, flow cytometry and pulsed field gel electrophoresis (PFGE) (Hao et al., 2011b, 2015b) could be chosen to determine the ploidy level and genome size. For example, flow cytometry was used to determine the genome size of four *Panax* species (Pan et al., 2014), with *Oryza sativa* as the internal standard. *Panax notoginseng* (San Qi in traditional Chinese medicine (TCM)) has the largest genome (2454.38 Mb), followed by *P. pseudoginseng* (2432.72 Mb), *P. vietnamensis* (2018.02 Mb), and *P. stipuleanatus* (1947.06 Mb), but their genomes are smaller than the *Pa. ginseng* genome (~3.2 Gb). A more reliable approach for species without the reference genome is the genome survey via the whole genome shotgun sequencing (Polashock et al., 2014). Such nondeep sequencing (30 × coverage), followed by the bioinformatics analysis, is highly valuable in assessing the genome size, heterozygosity, repeat sequence, guanine/cytosine (GC) content, etc., facilitating the decision-making of the whole genome sequencing approaches. In addition, RAD-Seq (restriction-site associated DNA sequencing; Fig. 1.3) (Rubin et al., 2012) could be chosen to construct a RAD library and perform the low-coverage genome sequencing of reduced representation, which is an effective approach for assessing the heterozygosity of the candidate genome.

The whole genome sequencing platform is chosen based on the budgetary resources and the preliminary evaluation of candidate genomes (Chen et al., 2010). GS FLX or Illumina HiSeq 2500 platforms might be suitable for the small simple genome. However, the majority of the plant genomes belong to the complex genome, which refers to the diploid/polyploidy genome, with >50% repeat sequences and >0.5% heterozygosity. Two or more sequencing platforms could be combined for shotgun and paired-end sequencing, while large insert libraries, for example, BAC (bacterial artificial chromosome) (Hao et al., 2015b), yeast artificial chromosome (Noskov et al., 2011), and Fosmid (Hao et al., 2011b), can be constructed for sequencing, then

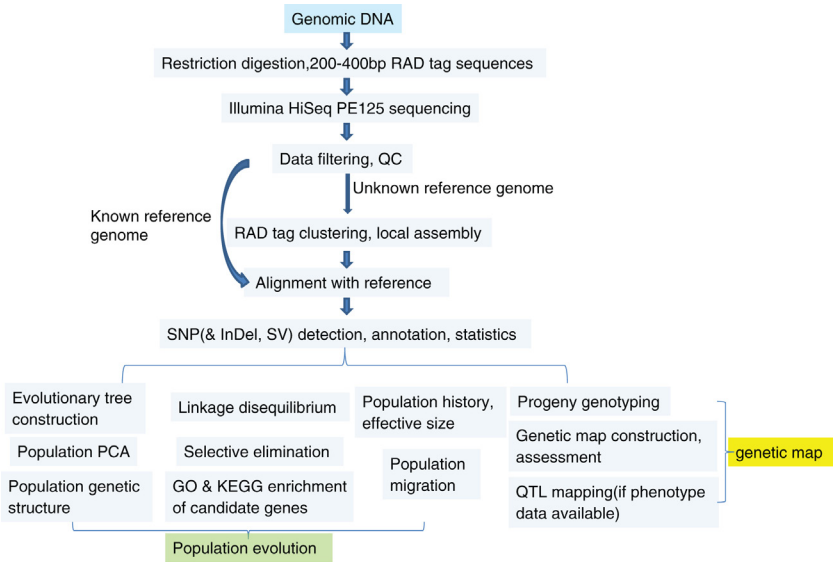

FIGURE 1.3 Technology roadmap of RAD-Seq and its utility in population evolution and genetic map. *InDel*, Insertion and deletion; *PCA*, principal component analysis; *PE*, paired end; *QC*, quality control; *QTL*, quantitative trait loci; *SV*, splice variant.

the sophisticated bioinformatics softwares (Cai et al., 2015; Chalhoub et al., 2014; Denoeud et al., 2014; Kim et al., 2014; Qin et al., 2014) can be used for sequence quality control and assembly. For instance, GS FLX and shotgun sequencing can be used for the initial genome assembly to generate 454 contigs, then the paired-end sequencing data from the Illumina HiSeq or SOLiD platform can be used to determine the order and orientation of 454 contigs, thus generating scaffolds. Next, Illumina HiSeq or SOLiD data are used to fill the gaps between some contigs. These steps streamline the genome sequencing pipeline as a whole.

The genetic map and physical map are fundamental tools for the assembly of the complex plant genome and functional genomics research. The genetic linkage map of *Bupleurum chinense* (Bei Chai Hu in TCM) was constructed using 28 ISSR (intersimple sequence repeat) and 44 SSR (microsatellite) markers (Zhan et al., 2010); 29 ISSRs and 170 SRAPs (sequence-related amplified polymorphisms) were mapped to 25 linkage groups of *Siraitia grosvenorii* (Luo Han Guo in TCM) (Liu et al., 2011). These preliminary results are useful in metabolic gene mapping, map-based cloning, and marker-assisted selection of medicinal traits. The high-throughput physical map could be anchored via the BAC-pool sequencing (Cviková et al., 2015), which, along with its integration with high-density genetic maps, could benefit from next generation sequencing (NGS) and high-throughput array platforms (Ariyadasa and Stein, 2012). The development of dense genetic maps of medicinal plants is still challenging, as the parental lines and their progenies with the unambiguous genetic link are not available for most medicinal plants.

1.2.2 Chloroplast Genome Evolution

Chloroplast (cp) is responsible for photosynthesis, and its genome sequences have versatile utility in evolution, adaptation, and robust growth of most medicinal plants. The substitution

rate of the cp nucleotide sequence is three to four times faster than that of the mitochondria (mt) sequence (Zhao et al., 2015), implicating more uses of the former in inferring both interspecific and intraspecific evolutionary relationships (Ma et al., 2014; Malé et al., 2012; Qian et al., 2013; Su et al., 2014; Wu et al., 2013; Xu et al., 2012; Zhao et al., 2015).

Pa. ginseng is a "crown" TCM plant and frequently used in health-promoting food and clinical therapy. NGS technology provides insight into the evolution and polymorphism of *Pa. ginseng* cp genome (Zhao et al., 2015). The cp genome length of Chinese *Pa. ginseng* cultivars Damaya (DMY), Ermaya (EMY), and Gaolishen (GLS) was 156,354 bp, while the genome length was 156,355 bp in wild ginseng (YSS), which is smaller than Omani lime (*C. aurantiifolia*; 159,893 bp) (Su et al., 2014) and 12 *Gossypium* cp genomes (159,959–160,433 bp) (Xu et al., 2012) but bigger than *Rhazya stricta* cp genome (154,841 bp) (Park et al., 2014). Gene content, GC content, and gene order in DMY are quite similar to other strains, and nucleotide sequence diversity of inverted repeat region (IR) is lower than that of large single-copy region (LSC) and small single-copy region (SSC). The high-resolution reads were mapped to the genome sequences to investigate the differences of the minor allele, which showed that the cp genome is heterogeneous during domestication; 208 minor allele sites with minor allele frequencies of $\geq$0.05 were identified. The polymorphism site numbers per kb cp genome of DMY, EMY, GLS, and YSS were 0.74, 0.59, 0.97, and 1.23, respectively. All minor allele sites were in LSC and IR regions, and the four strains showed the same variation types (substitution base or indel) at all identified polymorphism sites. The minor allele sites of the cp genome underwent purifying selection to adapt to the changing environment during domestication. The study of medicinal plant cp genomes with particular focus on minor allele sites would be valuable in probing the dynamics of the cp genomes and authenticating different strains and cultivars.

The genus *Citrus* contains many economically important fruits that are grown worldwide for their high nutritional and medicinal value. Due to frequent hybridizations among species and cultivars, the exact number of natural species and the evolutionary relationships within this genus are blurred. It is essential to compare the *Citrus* cp genomes and to develop suitable genetic markers for both basic research and practical use. A reference-assisted approach was adopted to assemble the complete cp genome of Omani lime (Su et al., 2014), whose organization and gene content are similar to most rosid lineages characterized to date. Compared with the sweet orange (*Ci. sinensis*), three intergenic regions and 94 simple sequence repeats (SSRs) were identified as potentially informative markers for resolving interspecific relationships, which can be harnessed to better understand the origin of domesticated *Citrus* and foster the germplasm conservation. A comparison among 72 species belonging to 10 families of representative rosid lineages also provides new insights into their cp genome evolution.

The monocot family Orchidaceae, evolutionarily more ancient than asterids and rosids, is one of the largest angiosperm families, including many medicinal, horticultural, and ornamental species. Orchid phytometabolites display antinociceptive (Morales-Sánchez et al., 2014), antiangiogenic (Basavarajappa et al., 2014), and antimycobacterial (Ponnuchamy et al., 2014) activities, etc. In South Asia, orchid bulb is used for the treatment of asthma, bronchitis, throat infections, and dermatological infections and also used as a blood purifier (Nagananda and Satishchandra, 2013). Sequencing the complete cp genomes of the medicinal plant *Dendrobium officinale* (Tie Pi Shi Hu in TCM) and the ornamental orchid *Cypripedium macranthos* reveals their gene content and order, as well as potential RNA-editing sites (Luo et al., 2014). The cp genomes of these two species and five known photosynthetic orchids are

similar in structure as well as gene order and content, but the organization of the IR/SSC junction and ndh gene is distinct. IRs flanking the SSC region underwent expansion or contraction in different *Orchidaceae* species. Fifteen highly divergent protein-coding genes were identified and are useful in phylogenetic inference of orchids. Cp phylogenomic analysis can be used to resolve the interspecific relationship that cannot be inferred by a few cp markers. Bamboo leaves are used as a component in TCM for the antiinflammatory function (Koide et al., 2011). Medicinal bamboo cupping therapy is applied to reduce fibromyalgia symptoms (Cao et al., 2011). Bamboo extracts exhibit antioxidant effects (Jiao et al., 2011a,b) and are used to treat chronic fever and infectious diseases (Wang et al., 2012). The whole cp genome data sets of 22 temperate bamboos considerably increased resolution along the backbone of tribe *Arundinarieae* (temperate woody bamboo) and afforded solid support for most relationships regardless of the very short internodes and long branches in the tree (Ma et al., 2014). An additional cp phylogenomic study, involving the full cp genome sequences of eight *Olyreae* (herbaceous bamboo) and 10 *Arundinarieae* species, strengthened the soundness of the above study and recovered monophyletic relationship between *Bambuseae* (tropical woody bamboo) and *Olyreae* (Wysocki et al., 2015).

The monocot genus *Fritillaria* (*Liliaceae*) has about 140 species of bulbous perennial plants that embraces taxa of both horticultural and medicinal importance. The bulbs of plants belonging to the *Fritillaria cirrhosa* group have been used as antitussive and expectorant herbs in TCM for thousands of years (Wu et al., 2015). The anticancer activity and cardiovascular effects of *Fritillaria* phytometabolites are well documented (Hao et al., 2015c). *Fritillaria* species have attracted attention also because of their remarkably large genome sizes, with all values recorded to date above 30 Gb (Day et al., 2014). A phylogenetic reconstruction, including the most currently recognized species diversity of the genus, was performed (Day et al., 2014). Three regions of the cp genome were sequenced in 92 species (c. 66% of the genus) and in representatives of nine other genera of *Liliaceae*. Eleven low-copy nuclear genes were screened in selected species, but they had limited utility in phylogenetic reconstruction. Phylogenetic analysis of a combined plastid data set supported the monophyly of the majority of presently identified subgenera. However, the subgenus *Fritillaria*, which is by far the largest subgenera and includes the most important species used in TCM, is found to be polyphyletic. Clade, containing the source plants of Chuan Bei Mu, Hubei Bei Mu, and Anhui Bei Mu, might be treated as a separate subgenus (Hao et al., 2013a). The Japanese endemic subgenus *Japonica*, which contains the species with the largest recorded genome size for any diploid plant, is sister to the largely Middle Eastern and Central Asian subgenus *Rhinopetalum*, which is significantly incongruent with the nuclear internal transcribed spacer (ITS) tree. Convergent or parallel evolution of phenotypic traits may be a common cause of incongruence between morphology-based classifications and the results of molecular phylogeny. While relationships between most major *Fritillaria* lineages can be resolved, these results also highlight the need for data from more independently evolving loci, which is pretty perplexing given the huge nuclear genomes found in these plants.

Medicinal plant diversity, comprised of genetic diversity, medicinal species diversity, ecological system diversity, etc. (Hao et al., 2014a), results from the intricate interactions between medicinal plant and environment, and thus is profoundly influenced by the ecological complex and the relevant versatile ecological processes. The effects of the evolutionary processes have to be taken into full consideration when explaining the link between climatic/ecological factors and medicinal plant diversity, especially where there is strong, uneven differentiation of species. A distinguished example is the "sky islands" of Southwest

China (He and Jiang, 2014), where the extraordinarily rich resources of medicinal plants rose and thrived during the Quaternary Period. To date, many medicinal tribes and genera, for example, *Pedicularis* (Eaton and Ree, 2013) (Fig. 1.4), *Clematis* (Hao et al., 2013b), *Aconitum* (Hao et al., 2013c, 2015d,e), and *Delphinium* (Jabbour and Renner, 2012), are still in the process of rapid radiation and dynamic differentiation. The cp genome sequence can be regarded as the super-barcode of the organelle scale and thus can be used to probe the intraspecific variations (Whittall et al., 2010) and phylogeographic patterns of the same species in the disparate geographic locations (e.g., geoherb or Daodi medicinal materials) (Zhao et al., 2012). The application of the cp genome sequence at the population level may provide clues for the timing and degree of the intraspecific differentiation. Distilling the interpopulation relationship from the cp data set can be considered a more detailed phylogenetic reconstruction.

1.2.3 Mitochondria (mt) Genome Evolution

Some fundamental evolution concepts, such as lateral gene transfer, are bolstered by the inquiry of the origin of mt, while plants are especially useful inelucidating the mechanisms of cytonuclear coevolution. Although the gene order of the mt genome might evolve relatively faster in land plants, the substitution rate of its nucleotide sequence is merely 1/100 that of its animal sequence (Hao et al., 2014a). Therefore, the mt genome sequence is less useful than the cp one in inferring the phylogenetic relationship of medicinal species (Henriquez et al., 2014). Notwithstanding, analyzing genome sequences contributes knowledge about the evolution of the mt genome. Moreover, the terpene synthase has been found in mt (Hsu et al., 2012), highlighting its utility in secondary metabolism.

R. stricta (*Apocynaceae*) is native to arid regions in South Asia and the Middle East and is used extensively in folk medicine. Analyses of the complete cp and mt genomes and a nuclear (nr) transcriptome of *Rhazya* shed light on intercompartmental transfers between genomes and the patterns of evolution among eight asterid mt genomes (Park et al., 2014). The *Rhazya* genome is

FIGURE 1.4 ***Pedicularis longiflora* of Scrophulariaceae.** *Source: Taken in Dingri County, Tibet, China.*

highly conserved with gene content and order identical to the ancestral organization of angiosperms. The 548,608 bp mt genome contains recombination-derived repeats that generate a compound organization; transferred DNA from the cp and nr genomes as well as bidirectional DNA transfers between the mt and the nucleus are also disclosed. The mt genes sdh3 and rps14 have been transferred to the nucleus and have acquired targeting transit peptides. Two copies of rps14 are present in the nucleus; only one has the mt targeting transit peptide and may be functional. Phylogenetic analyses suggest that *Rhazya* has experienced a single transfer of this gene to the nucleus, followed by a duplication event. The phylogenetic distribution of gene losses and the high level of sequence divergence in targeting transit peptides suggest multiple independent transfers of both sdh3 and rps14 across asterids. Comparative analyses of mt genomes of eight asterids indicates a complicated evolutionary history in this thriving eudicot clade, with substantial diversity in genome organization and size, repeat, gene and intron content, and amount of alien DNA from the cp and nr genomes. The genomic data enable a rigorous inspection of the gene transfer events.

1.2.4 Nuclear Genome Evolution

1.2.4.1 Monocots

The whole cp genome data set is not enough to elucidate the phylogenetic relationship of groups undergoing rapid radiation, for example, Zingiberales (Barrett et al., 2014). The cp genome is equivalent to one gene locus, thus it only represents one fulfillment to the coalescent random processes and cannot be used with confidence to reconstruct the evolutionary history of the populations. The most genetic history of any medicinal plant hides in the nr genome.

High-throughput sequencing and the relevant bioinformatics advances have revolutionized contemporary thinking on nuclear genome/transcriptome evolution and provided basic data for further breeding endeavors. *Coix* (*Poaceae*), a closely related genus of *Sorghum* and *Zea*, has 9–11 species with different ploidy levels. The exclusively cultivated *C. lacryma-jobi* ($2n = 20$) is widely used in East and Southeast Asia as food and traditional medicine. *C. aquatica* has three fertile cytotypes ($2n = 10$, 20, and 40) and one sterile cytotype ($2n = 30$), *C. aquatica* HG, which is found in Guangxi of China (Cai et al., 2014). Low coverage genome sequencing (genome survey) showed that around 76% of the *C. lacryma-jobi* genome and 73% of the *C. aquatica* HG genome are repetitive sequences, among which the long terminal repeat (LTR) retrotransposable elements dominate, but the proportions of many repeat sequences vary greatly between the two species, suggesting their evolutionary divergence. A novel 102 bp variant of centromeric satellite repeat CentX and two other satellites are exclusively found in *C. aquatica* HG. The FISH analysis and fine karyotyping showed that *C. lacryma-jobi* is likely a diploidized paleotetraploid species, and *C. aquatica* HG is possibly from a recent hybridization. These *Coix* taxa share more coexisting repeat families and higher sequence similarity with *Sorghum* than with *Zea*, which agrees with the phylogenetic relationship.

The heterozygous genome sequences of the tropical epiphytic orchid *Phalaenopsis equestris* provide insights into the unique crassulacean acid metabolism (CAM) (Cai et al., 2015). The assembled genome contains 29,431 predicted protein-coding genes and is rich in genes that might be involved in self-incompatibility pathways, which ensure the genetic diversity and enhance the fitness and survival. An orchid-specific paleopolyploidy event is disclosed,

which preceded the radiation of most orchid clades, and gene duplication might have contributed to the evolution of CAM photosynthesis in *Ph. equestris*. The expanded and diversified families of MADS-box C/D-class, B-class AP3, and AGL6-class genes might contribute to the highly specialized morphology of orchid flowers. LTRs are the most abundant transposable element (Fig. 1.5), followed by long interspersed nuclear elements (LINEs).

1.2.4.2 Basal Eudicots

The *Macleaya cordata* (*Papaveraceae*, Ranunculales) genome covering 378 Mb encodes 22,328 predicted protein-coding genes, with 43.5% being transposable elements (Liu et al., 2017). As a member of basal eudicots, this genome lacks the paleohexaploidy event that occurred in almost all eudicots. From the genomics data, all 16 metabolic genes for sanguinarine and chelerythrine biosynthesis were retrieved, and the biochemical activities of 14 genes were validated. These genomics and metabolic data show the conserved benzylisoquinoline alkaloid (BIA) metabolic pathways in *M. cordata* and provide the knowledge base for future productions of BIAs by crop improvement or microbial pathway reconstruction.

1.2.4.3 Eudicots: Asterids

Whole genome sequencing has been implemented in the representative species of some plant families/genera (Fig. 1.6), for example, *Capsicum annuum* (Kim et al., 2014; Qin et al., 2014), *Coffea canephora* (Denoeud et al., 2014), *Brassica napus* (Chalhoub et al., 2014), and *Ph. equestris* (Cai et al., 2015). The genome sequences of the cultivated pepper Zunla-1 (*Cap. annuum*) and its wild progenitor Chiltepin (*Cap. annuum* var. *glabriusculum*) were compared to provide insights into *Capsicum* domestication and specialization. The pepper genome expanded ~0.3 Mya by a rapid amplification of retrotransposon elements, resulting in a genome containing ~81% repetitive sequences and 34,476 protein-coding genes. Comparison of cultivated and wild pepper genomes with 20 resequencing accessions revealed

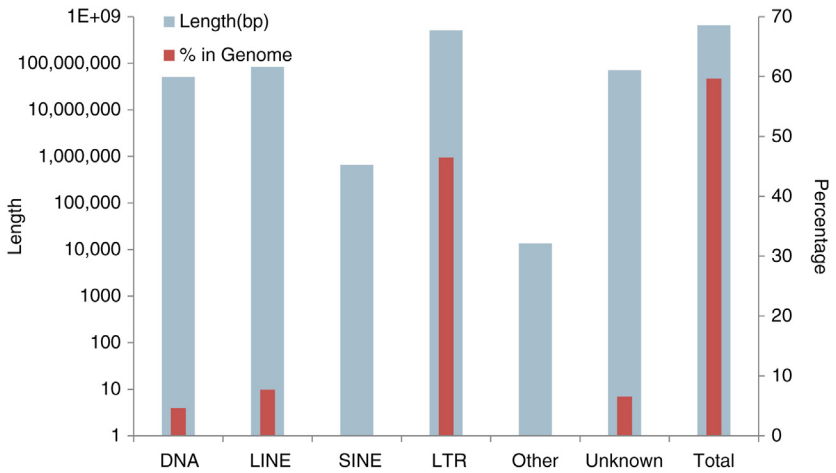

FIGURE 1.5 Categories of transposable elements predicted in the orchid genome (Cai et al., 2015). *DNA*, DNA transposon; *LINE*, long interspersed element (retrotransposon); *LTR*, long terminal repeat (retrotransposon); *SINE*, short interspersed nuclear element (retrotransposon).

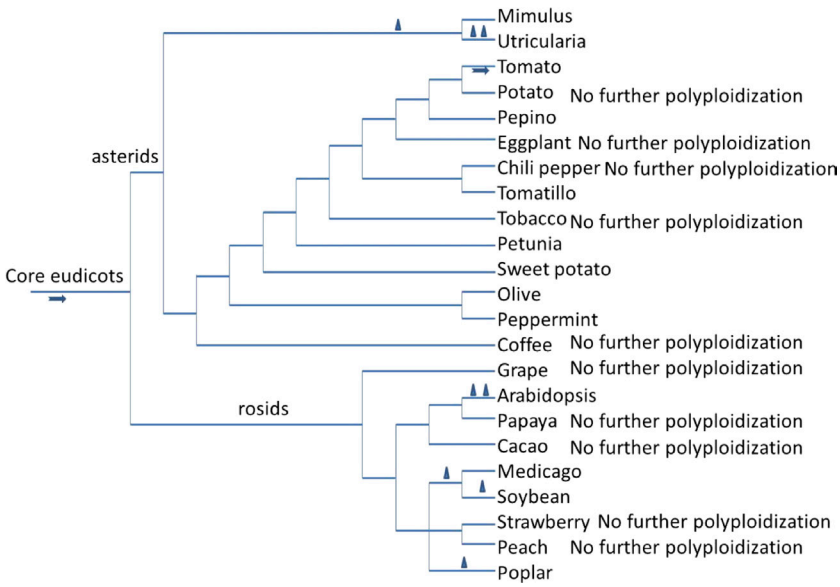

FIGURE 1.6 Examples of the phylogeny and genome duplication history of core eudicots. *Arrowheads* indicate hexaploidization; *triangles* indicate tetraploidization. The current evidence does not suggest further polyploidization after speciation in the genomes of potato, eggplant, chili pepper, tobacco, coffee, grape, papaya, cacao, strawberry, and peach. Few genome data are available in pepino, tomatillo, and many other species.

molecular signatures of artificial selection, providing a list of candidate domestication genes (Qin et al., 2014). Dosage compensation effect of tandem duplication genes might contribute to the pungency divergence in pepper (Qin et al., 2014). The *Capsicum* reference genome, along with tomato and potato genomes, provides critical information for the study of the evolution of other *Solanaceae* species, including the well-known *Atropa* medicinal plants.

The highly heterozygous *Salvia miltiorrhiza* (Danshen, *Lamiaceae*) genome was assembled with the help of 395× raw read coverage using Illumina technologies and about 10× raw read coverage using single molecular sequencing technology (Zhang et al., 2015). The final draft genome is approximately 641 Mb, with a contig N50 size of 82.8 kb and a scaffold N50 size of 1.2 Mb. Further analyses predicted 34,598 protein-coding genes and 1644 unique gene families in the Danshen genome, which provides a valuable resource for the investigation of novel bioactive compounds in this traditional Chinese herb.

One of the milestone breakthroughs is the successful sequencing and assembly of the complex heterozygous genome. The heterozygous genome of *Co. canephora* has been deciphered (Denoeud et al., 2014), and it displays a conserved chromosomal gene order among asterid angiosperms. Although it shows no sign of the whole-genome triplication identified in *Solanaceae* species, the genome includes several species-specific gene family expansions, for example, *N*-methyltransferases (NMTs) involved in caffeine biosynthesis, defense-related genes, and alkaloid and flavonoid enzymes involved in secondary metabolite production. Caffeine NMTs expanded through sequential tandem duplications independently and are distinct from those of cacao and tea, suggesting that caffeine in eudicots is of polyphyletic origin and its biosynthesis underwent convergent evolution.

The 3.5-Gb genome of *Pa. ginseng* contains more than 60% repeats and encodes 42,006 predicted genes (Xu et al., 2017). Twenty-two transcriptome data sets and mass spectrometry images of ginseng roots were used to precisely quantify the functional genes. Thirty-one genes were identified to be involved in the ginsenoside biosynthetic mevalonic acid pathway, eight of which were 3-hydroxy-3-methylglutaryl-CoA reductases. A total of 225 UDP-glycosyltransferases (UGTs) were identified, which constitute one of the largest gene families of ginseng. Tandem repeats contributed to the duplication and divergence of UGTs. Molecular modeling of UGTs in the 71st, 74th, and 94th families revealed a regiospecific conserved motif at the *N*-terminus, which captures ginsenoside precursors. The ginseng genome represents a valuable resource for understanding and improving the breeding, cultivation, and synthetic biology of this king of TCM.

Pan. notoginseng experienced a series of genome evolution events that created the unique medicinal properties of this famous medicinal plant (Chen et al., 2017; Zhang et al., 2017). For example, a recent polyploidy event occurred about 26 million years ago, and there were a large number of specific duplications of triterpenoid biosynthesis-related gene families; genes related to triterpenoid saponin biosynthesis formed many gene clusters. Comparative genomics, transcriptomics, and comparative phytochemistry further confirmed that the rapid functional variation and evolution of genomes determines the chemodiversity, which is closely related to the therapeutic efficacy of *Pan. notoginseng*. Most genes associated with the saponin biosynthesis are mainly expressed in flowers and leaves; after synthesis, saponins could be transported and stored in the roots. This discovery subverts the view that the *Pan. notoginseng* saponins are synthesized in the roots.

The highly heterozygous *Erigeron breviscapus* genome was assembled using a combination of PacBio, single-molecular, real-time sequencing and next-generation sequencing methods on the Illumina HiSeq platform (Yang et al., 2017). The final draft genome is around 1.2 Gb, with contig and scaffold N50 sizes of 18.8 kb and 31.5 kb, respectively. Further analyses predicted 37,504 protein-coding genes in the *E. breviscapus* genome and 8172 shared gene families among the *Compositae* species.

1.2.4.4 Gene Family

More than 40 plant genomes have been sequenced, representing a diverse set of taxa of agricultural, energy, medicinal, and ecological importance (Cai et al., 2015; Chalhoub et al., 2014; Denoeud et al., 2014; Kim et al., 2014; Qin et al., 2014). Gene family members are often inferred from DNA sequence homology, but deeper insights into evolutionary processes contributing to gene family dynamics are imperative. In a comparative genomics framework, multiple lines of evidence can be generated by gene synteny, sequence homology, and protein-based hidden Markov modeling (HMM) to extract homologous super-clusters composed of multi-domain resistance (R)-proteins of the NB-LRR (nucleotide binding-leucine rich repeat) type, which are involved in plant innate immunity (Hofberger et al., 2014). Twelve eudicot plant genomes were screened to assess the intra- and interspecific diversity of R-proteins, where 2363 NB-LRR genes were found. Half of the R-proteins have tandem duplicates, and 22% of gene copies are left from ancient polyploidy events (ohnologs, whole genome duplication duplicates). The positive Darwinian selection and major differences in molecular evolution rates (K_a/K_s) were detected among tandem (mean = 1.59), ohnolog (1.36), and singleton (1.22) R-gene duplicates. The distribution pattern of all 140 NB-LRR genes present in the model

plant *Arabidopsis* is species specific, and four distinct clusters of NB-LRR "gatekeeper" loci sharing syntenic orthologs across all analyzed genomes were identified and could be useful for the gene-edited plant breeding. The near-complete set of multidomain R-protein clusters in a eudicot-wide scale could shed light on evolutionary dynamics underlying diversification of the plant innate immune system. More functional NB-LRR genes could be identified from more sequenced plant species.

The estimated upper limit of extant plants is c. 450,000, indicating the potentially enormous biological space. The multiple and recurrent genome duplications during plant evolution result in the generation of novel biosynthetic pathways of diverse medicinal compounds, which are frequently involved in plant defense and disease resistance and, more importantly, create huge chemical space for drug discovery and development. The duplicated gene copies could explain the diversification processes of the multigene secondary metabolism pathways, such as those involved in the biosynthesis of terpenoids (Boutanaev et al., 2015), benzoxazinoid (Dutartre et al., 2012), steroidal glycoalkaloid (Manrique-Carpintero et al., 2013), and glucosinolates (Hofberger et al., 2013). More than 200,000 secondary metabolites have been found in angiosperms, many of which could stem from the genome duplication-based rapid innovation of complex traits (Hao et al., 2014a).

1.2.4.5 Single Copy Gene

Single copy genes are common across angiosperm genomes. Based on 29 sufficiently high-quality sequenced genomes, the large-scale identification and evolutionary characterization of single copy genes among multiple species is possible (Han et al., 2014). A significant negative correlation was found between the number of duplicate blocks and the number of single copy genes. Only 17% of single copy genes are located in organelles, most of which are involved in binding and catalytic activity. Most single copy genes are in nuclear genomes. Single copy genes have a stronger codon bias than nonsingle copy genes in eudicots (Han et al., 2014). The relatively high expression level of single copy genes was partially confirmed by the RNA-Seq (transcriptome sequencing) data. Unlike in most other species, there is a strong negative correlation between N_c (effective number of codons) and GC3 (G + C content at third codon position) of single copy genes in grass genomes. Compared to nonsingle copy genes, single copy genes are of more conservation, as indicated by K_a and K_s values. Selective constraints on alternative splicing are weaker in single copy genes than in low-copy family genes (1–10 paralogs) and stronger than high-copy family genes (>10 paralogs). Using concatenated, shared, single copy genes, a well-resolved phylogenetic tree can be obtained. Addition of intron sequences improved the branch support, but striking incongruences are also obvious. Inclusion of intron sequences might be more appropriate for the phylogenetic reconstruction at lower taxonomic levels. Evolutionary constraints between single copy genes and nonsingle copy genes are distinct and are somewhat species specific, especially between eudicots and monocots.

1.2.5 Transcriptome

The high cost of the whole genome sequencing is still formidable. The accurate sequence assembly is still challenging, especially when the genome is of a high proportion of repeat sequences, high heterozygosity, and nondiploid. RNA-Seq is a powerful tool for the

assessment of gene expression and the identification and characterization of molecular markers in nonmodel organisms (Hao et al., 2012b). Unlike genome sequences, the intron sequences are not included in the RNA-Seq data set, and the Unigene (contig) assembly is not disturbed by the repeat sequences and the ploidy level. The global view of the ethnomedicine resources and the accurate delimitation of the novel medicinal taxa cannot be achieved without the molecular phylogeny based on the complete taxon sampling of the relevant tribes/genera. Due to the plummeting cost of RNA-Seq, the dense taxon sampling is now possible in the phylotranscriptomic studies. It is obvious that the large-scale comparative transcriptome studies, including those of medicinal plants, are more feasible than comparative genomics based on whole genome sequencing. As shown in National Centre for Biotechnology Information (NCBI) PubMed, sequence read archive (SRA), and Gene Expression Omnibus (GEO) databases, transcriptomes of hundreds of medicinal plants have been sequenced, for example, *Caryophyllales* (Yang et al., 2015) (Fig. 1.7), *Fabaceae* (Cannon et al., 2015), *Oenothera* (*Onagraceae*) (Hollister et al., 2015), *Rhodiola algida* (*Crassulaceae*) (Zhang et al., 2014), *S. sclarea* (*Lamiaceae*) (Hao et al., 2015a), *Polygonum cuspidatum* (*Polygonaceae*) (Hao et al., 2012d), *Taxus mairei* (*Taxaceae*) (Hao et al., 2011a), etc. The single copy orthologous gene sequences could be extracted from the UniGene data sets of multiple medicinal plants (Hao et al., 2012c), which can be used in the phylogenetic reconstruction and evolutionary analyses (Hao et al., 2012d; Yang et al., 2015). The information uncovered in transcriptome studies could serve in the characterization of important traits related to secondary metabolite formation and in probing the relevant molecular mechanisms (Hao et al., 2011a, 2012c, 2015a; Zhang et al., 2014).

Reconstructing the origin and evolution of land plants and their algal relatives is a vital problem in plant phylogenetics and is essential for understanding how novel adaptive traits, for example, secondary metabolites, arose. Despite advances in molecular systematics, some evolutionary relationships remain poorly resolved. Inferring deep phylogenies with rapid diversification is often tricky, and genome-scale data significantly increase the number of informative characters for analyses. Since the sparse taxon sampling could result in inconsistent results, transcriptome data of 92 streptophyte taxa were generated and analyzed along with 11 published plant genome sequences (Wickett et al., 2014). Phylogenetic reconstructions were conducted using 852 nuclear genes and 1,701,170 aligned sites. Robust support for a sister-group relationship between land plants and one streptophyte green algae, the *Zygnematophyceae*, was obtained. Strong and robust support for a clade comprising liverworts and mosses contradicts the widely accepted view of early land plant evolution. Phylogenetic

FIGURE 1.7 ***Stellaria chinensis*** **of Caryophyllaceae.** *Source: Taken in Taibai Mountain, Shaanxi, China.*

hypotheses could be tested using phylotranscriptomic approach to give deeper insights and novel arguments into the evolution of fundamental plant traits, including the fascinating chemodiversity.

The transcriptome sequencing also sheds light on other untapped issues of plant evolution. Arbuscular mycorrhizal (AM) are symbiotic systems in nature and have great significance in promoting the growth and stress resistance of medicinal plants (Zeng et al., 2013). AM has multifaceted effects on the active ingredients of TCM plants. The transcriptomes of nine phylogenetically divergent non-AM symbiosis plants were analyzed to reveal the correlation between the loss of AM symbiosis and the loss of many symbiotic genes (Delaux et al., 2014), which was found in four additional plant lineages besides the *Arabidopsis* lineage (Brassicales), implicating the convergent evolution. RNA-Seq was used to outline gene sequence and expression discrepancy between cultivated tomato and five allied wild species (Koenig et al., 2013). Human handling of the genome has profoundly altered the tomato transcriptome via directed admixture and by secondarily choosing nonsynonymous over synonymous substitutions. A hitherto unidentified paleopolyploidy event that arose 20–40 million years ago was uncovered based on the transcriptomes of 11 *Linum* species (Sveinsson et al., 2014), which is specific to a clade enclosing cultivated flax (*L. usitatissimum*) and other mainly blue-flowered species.

1.2.6 Evolution and Population Genetics/Genomics

SSRs play a major role as molecular markers for genome analysis and plant breeding. The microsatellites existing in the complete genome sequences would have a direct role in the genome organization, recombination, gene regulation, quantitative genetic variation, and evolution of genes. Microsatellite markers have been characterized for many medicinal plant families and genera, for example, Acanthaceae family (Kaliswamy et al., 2015), *Artemisia* genus (Karimi et al., 2015), *Camellia* genus (Li et al., 2015), Chinese jujube (Wang et al., 2014), etc. For instance, 11 nuclear SSR loci were used to reveal the relative low genetic diversity of three *Camellia taliensis* (Da Li tea) populations, three *Ci. sinensis* var. *assamica* (Pu Er tea in TCM) populations, and two transitional populations of *Ca. taliensis*. A momentous genetic differentiation was found between *Ci. sinensis* var. *assamica* and *Ca. taliensis* populations. The transitional populations of *Ca. taliensis* stemmed mainly from *Ca. taliensis* and underwent genetic differentiation during domestication. Gene introgression was spotted in the cultivated *Ci. sinensis* var. *assamica* and *Ca. taliensis* of the same tea garden, and genetic material of *Ca. taliensis* seemingly intruded into *Ci. sinensis* var. *assamica*, suggesting that the former was genetically involved in the domestication of the latter. These results are useful for protecting the genetic resources of ancient tea plants. The whole nucleotide sequences—for example, the genomic sequences (Hao et al., 2015b) or the transcriptomic sequences (Hao et al., 2012b, 2015a)—of plant species can be obtained from NCBI databases and screened for the presence of SSRs.

Both ISSR (Hao et al., 2010) and SRAP markers were suitable for discriminating among the studied individuals, and the SRAP markers were more efficient and preferable (Karimi et al., 2015). Multiple regression analysis revealed statistically significant associations between rust resistance and some molecular markers; this can provide clues for identification of the individuals with higher rust resistance. RAPD (randomly amplified polymorphic DNA) and

ISSR markers were used to characterize *Schisandra chinensis* with white fruit (Li et al., 2014). The molecular marker-based study of genetic diversity helps in assessing the studied germplasm, which would be a valuable genetic resource for future breeding. Based on such a study, *in situ* conservation measures or other methods could be recommended to preserve the valuable medicinal plant genetic resources.

Sinopodophyllum hexandrum is an endangered *Berberidaceae* (Ranunculales) medicinal plant, and its genetic diversity must be protected against habitat loss and anthropogenic factors. The Qinling Mountains are an *Si. hexandrum* distribution area, where unique environmental features highly affect the evolution of the species. ISSR analysis of 32 natural populations revealed the genetic diversity and population structure of *Si. hexandrum* in Qinling and provides reference data for evolutionary and conservation studies (Liu et al., 2014). The 32 populations fell into three major groups, where analysis of molecular variance confirmed significant variation among populations. The high genetic differentiation may be attributed to the limited gene flow within the species. The spatial pattern and geographic locations of different populations are not correlated with one another. In light of the low within-population genetic diversity, high differentiation among populations, and the increasing anthropogenic pressure on the species, *in situ* conservation is proposed to preserve *Si. hexandrum* in Qinling, and other populations must be sampled to maintain genetic diversity of the species for the *ex situ* preservation.

SNPs (single nucleotide polymorphisms) are much more abundant than SSRs in most species (Clevenger et al., 2015), including medicinal plants. The mutation rate of SNPs (10^{-9} per locus per generation) is much lower than that of SSRs (10^{-3}–10^{-4}) (Guichoux et al., 2011). Generally there are only two alleles in each SNP site, while more than 10 alleles can be in each SSR. The highly polymorphic SSRs are especially suitable for detecting the hybridization between closely related species and studying the gene flow/introgression (Wee et al., 2015). SSRs are of lower ascertainment bias and are also good for studying the recent population structure. Mining suitable SSR sites via transcriptome sequencing data sets is fast and affordable; for example, 3446 microsatellites are identified from 2718 Unigenes (16.8% of 16,142 assembled sequences) of the *S. sclarea* transcriptome (Fig. 1.8). Trinucleotide (1883) is the predominant microsatellite, followed by dinucleotide (1144) and mononucleotide (315), indicating that many microsatellites are in translated regions of the expressed genes. CCG/CGG is the predominant trinucleotide SSR, followed by AAG/CTT and AGC/GCT. AG/CT is the most common dinucleotide SSR. Of the identified repeats, 601 (19.2%) have sufficient flanking sequence information to allow for PCR primer design. Intriguingly, many SSR motifs are linked with unique sequences encoding enzymes involved in phenylpropanoid/terpenoid metabolism. For instance, SSRs were detected in phenylalanine ammonia-lyase, 4-coumarate-CoA ligase, hydroxyphenylpyruvate dioxygenase, flavonoid 3′-hydroxylase, cinnamyl alcohol dehydrogenase, and lignan glycosyltransferase sequences, which belong to the phenylpropanoid pathway; SSRs were also found in 2-*C*-methyl-D-erythritol 4-phosphate pathway genes (1-deoxy-D-xylulose 5-phosphate synthase, 1-deoxy-D-xylulose 5-phosphate reductoisomerase, 2-*C*-methyl-D-erythritol 4-phosphate cytidylyltransferase, 4-hydroxy-3-methylbut-2-enyl diphosphate synthase), mevalonate pathway genes (mevalonate pyrophosphate decarboxylase) and other terpenoid biosynthesis genes (isopentenyl diphosphate isomerase, cytochrome P450 71D18, pinene synthase, squalene synthase, and squalene monooxygenase). These SSRs might be useful in future breeding and ecological studies. One of the

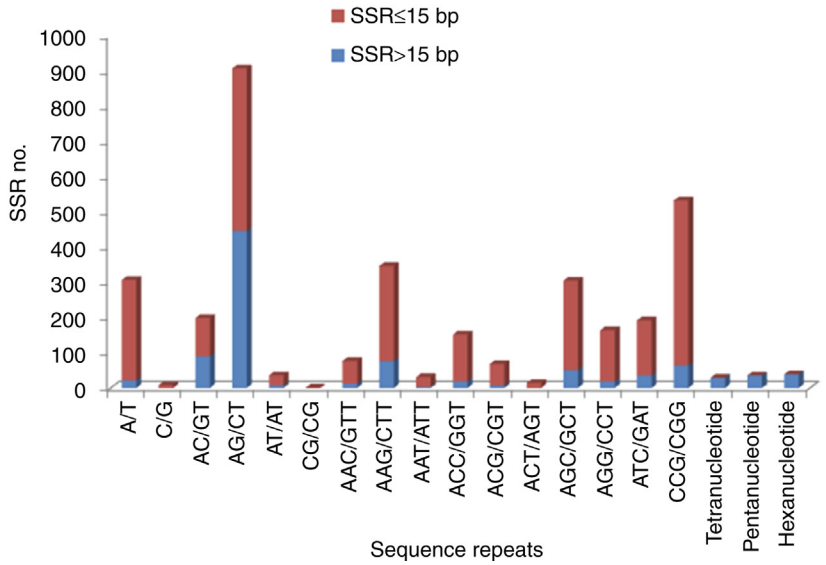

FIGURE 1.8 SSRs predicted from the *Salvia sclarea* transcriptome data set (Hao et al., 2012b). Msatcommander (http://code.google.com/p/msatcommander/) was used to annotate SSRs. BatchPrimer3 (You et al., 2008) was employed to design PCR primers in the flanking regions of the detected SSRs, setting a minimum product size of 100 bp, a minimum primer length of 18 bp, a minimum GC content of 30%, a melting temperature between 50°C and 70°C, and a maximum melting temperature difference between primers of 8°C.

major drawbacks of SSRs is the low universality and poor transferability; that is, usually the species-specific SSR primers have to be developed. The other disadvantage of SSRs is their uncertain mutation model, which is often simplified to be the stepwise mutation model, but the actual mutation pattern might be more complicated.

NGS tool kits could provide grist for the medicinal plant phylogeography mill. Sufficiently abundant SNPs could be identified directly from the genome sequences of the model plants. Most medicinal plants lack the genomic data; therefore, two alternative strategies can be adopted. The faster and cheaper one is mining suitable SNP sites via transcriptome sequencing data sets (Hao et al., 2012b). However, the subsequent PCR primer design might not be successful, as no information about the intergenic sequences and the introns are available from the RNA-Seq data. On the other hand, large amounts of SNPs can be obtained by the simplified genome sequencing, mainly referring to RAD-Seq (Eaton and Ree, 2013) and genotyping-by-sequencing (GBS) (He et al., 2014a); although, their reproducibility and reliability warrant further improvements.

Plants of various evolutionary levels, not only higher plants, are harnessed in TCM and worldwide ethnomedicine. The caterpillar fungus *Ophiocordyceps sinensis* (Dong Chong Xia Cao in TCM) is one of the most valuable medicinal fungi in the world, and host insects of family Hepialidae (Lepidoptera) are a must to complete its life cycle. The genetic diversity and phylogeographic structures of the host insects are characterized using mtCOI (cytochrome oxidase subunit I) sequences (Quan et al., 2014). Abundant haplotype and nucleotide diversity were mainly found in the east edge of the Qinghai-Tibet Plateau (QTP), which is the diversity

center or microrefuge of the host insects. The genetic variation of the host insects is negligible among 72.1% of all *O. sinensis* populations. All host insects are monophyletic except for those of four *O. sinensis* populations around Qinghai Lake. Significant phylogeographic structure was revealed for the monophyletic host insects, and the three major phylogenetic groups corresponded to specific geographical areas. The divergence of most host insects might occur at c. 3.7 Ma, shortly before the rapid uplift of the QTP. The geographical distribution and star-like network of the haplotypes implied that most host insects were derived from the relicts of a once-widespread host that subsequently became fragmented. Most host insects underwent recent demographic expansions that began c. 0.118 Ma in the late Pleistocene, suggesting that the genetic diversity and distribution of the present-day insects could be ascribed to effects of the QTP uplift and glacial advance/retreat cycles during the Quaternary ice age. These results provide valuable reference to the conservation and sustainable use of both host insects and *O. sinensis*.

Population genetics can be upgraded to population genomics using the large data set of transcriptomes from multiple species (Hollister et al., 2015). The dearth of extant asexual species might be partially caused by buildup of harmful mutations and intensified elimination risk linked with repressed recombination and segregation in these species, which was tested with a data set of 62 transcriptomes of 29 *Oenothera* species (*Onagraceae*; Hollister et al., 2015). The nonsynonymous polymorphism is more abundant than the synonymous variation within asexual species, implying relaxed purifying selection. Asexual species also displayed more transcripts with premature stop codons. The increased proportion of nonsynonymous mutations was positively associated with divergence time between sexual and asexual species. These results suggest that sex enables selection against deleterious alleles.

1.3 MECHANISMS OF SPECIES EVOLUTION AND DIVERSIFICATION

The incidence of polyploidy in land plant evolution has led to an acceleration of genome variations compared with other crown eukaryotes and is connected with key innovations in plant evolution (Cannon et al., 2015). Increasing genome resources facilitates linking genomic alterations to the origins of novel phytochemical and physiological features of medicinal plants. Ancestral gene contents for key nodes of the plant family tree are inferred (Jiao and Paterson, 2014). The ancestral WGDs (whole genome duplications) concentrating c. 319–192 million years ago expedited the diversification of regulatory genes vital to seed and flower development and were responsible for key innovations followed by the upsurge and ultimate supremacy of seed plants and flowering plants. Widespread polyploidy in angiosperms might be the major factor generating novel genes and expanding some gene families (Hofberger et al., 2013). However, most gene families lose the majority of duplicated copies in an early neutral process, and a few families are actively selected for single-copy status. It is challenging to link genome modifications to speciation, diversification, and the phytochemical and/or physiological innovations that jointly comprise biodiversity and chemodiversity. Ongoing evolutionary genomics investigations may greatly improve the resolution, enabling the identification of specific genes responsible for particular innovations. More concise understanding of plant evolution may enrich the fundamental knowledge of botanical diversity, including medicinally important traits that sustain humanity.

Case studies are important to illustrate the correlation between WGD and the diversification of secondary metabolism pathways. WGD and the tandem duplication facilitated glucosinolate pathway diversification in the mustard family (Brassicaceae) (Hofberger et al., 2013) (Fig. 1.9). In *Arabidopsis thaliana*, at least 52 biosynthetic and regulatory genes are involved in the glucosinolate biosynthesis. *Aethionema arabicum*, basal to other *Brassicales* species, harbors 67glucosinolate biosynthesis genes, most of which have the ortholog in *Ar. thaliana*, displaying the syntenic relationship. In *Ar. thaliana*, 45% of the protein-coding genes have more than one copy, while 95% of *Ar. thaliana* and 97% of *Aethionema* glucosinolate pathway genes possess multiple copies, suggesting the particular diversification of this defense pathway. The sequence alignment and phylogenetic analysis showed that the significant duplications of glucosinolate pathway genes occurred during the last common WGD event. The tandem duplication and the subsequent subfunctionalization and neofunctionalization further increase the genetic diversity and chemodiversity of the glucosinolate secondary metabolites, thus enhancing the phenotypic plasticity and adaptation. More importantly, the chemical space of the diverse secondary metabolites has great potential in drug discovery. Duplicated gene copies also explain the diversification process of terpenoids (Boutanaev et al., 2015), the largest class of plant natural products. Tracing the roots of terpene biosynthesis and diversification in plants reveals that distinct genomic mechanisms of pathway assembly have evolved in eudicots and monocots.

Besides polyploidy, allopatric divergence, climatic oscillation-based divergence, hybridization and introgression, and pollination-mediated isolation are also highlighted as the mechanisms of medicinal species evolution, especially in the hot spot areas of biodiversity, such as QTP (Wen et al., 2014) (Fig. 1.10). Rapid species diversification followed the extensive uplift of QTP and brought about numerous morphologically and phytochemically distinct species. Both morphological and metabolic phenotype innovations are apparently ecologically adaptive, and the underlying molecular mechanisms are still elusive.

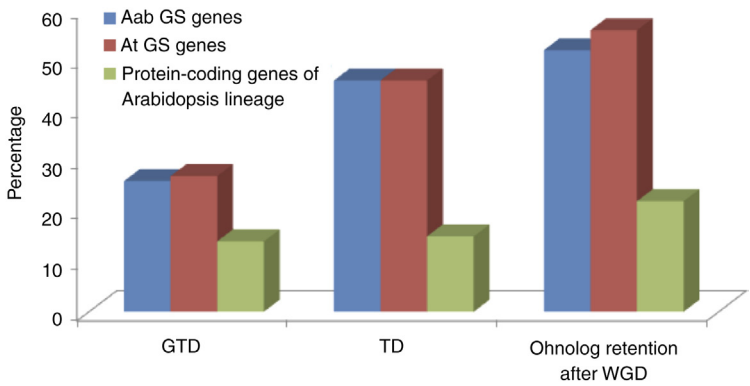

FIGURE 1.9 Duplicate distribution among *Arabidopsis thaliana* (At) protein-coding genes compared with AtGS (glucosinolate) and *Aethionema arabicum* (Aab) GS loci, according to Hofberger et al., 2013. Percentage of genes with retained ohnolog (clusters of dose-sensitive genes organized in functional modules), tandem duplicate (TD), and gene transposition duplicate (GTD) are shown. GS metabolic plasticity during lineage evolution arose from a combination of increased ohnolog retention and TD rates.

FIGURE 1.10 **A representative landscape of QTP.** *Source: Taken in July 2017, near Ningjin Kangsha snow mountain, Tibet, China.*

1.4 PHENOTYPE EVOLUTION AND ECOLOGY

Medicinal plants synthesize an arsenal of protective molecules, most of which are secondary metabolites and can be ingested by animals and humans, and then help them antagonize against disadvantageous environmental conditions (Sternberg et al., 2015). The epidemiological (parasite prevalence and virulence) and environmental (medicinal plant toxicity and abundance) conditions that predict the evolution of genetically fixed versus phenotypically plastic forms of animal medication can be identified using the tritrophic interaction between the monarch butterfly, its protozoan parasite, and its food plant *Asclepias* spp. as a test case (Choisy and de Roode, 2014). Analogously, in folk medicine practice people accumulate knowledge about the relative benefits (the antiparasitic/antimicrobial properties of medicinal plants) and costs (side effects of phytomedicine, the costs of searching for medicine) in ethnomedicine practice, which determine whether medication is for therapeutic use or preventive use.

Numerous botanical compounds, as the integral part of plant defense mechanisms, also bind and modify fundamental regulators of animal physiological processes in ways that enhance animal adaptation to the ever-changing environments (Kennedy, 2014). The underlying mechanism might be that animals and fungi, as heterotrophs, are capable of sensing chemical signals produced by plants and responding actively to the biotic/abiotic stress (xenohormesis) (Howitz and Sinclair, 2008). These plant-derived cues offer early warning about fading ecological circumstances, permitting the heterotrophs to get ready for misfortune when conditions are still auspicious. Plant secondary metabolites could activate the evolutionarily conserved cellular stress response and subsequently enhance the cellular adaptation to adversity in both plants themselves and animals that consume them. Xenohormesis could explain TCM pharmacological effects from an evolutionary and ecological perspective (Qi et al., 2013). Medicinal herb, microbial, and human cellular signal transduction pathways have many conserved similarities, enabling beneficial effects of botanical metabolites in humans via a process of "cross-kingdom" signaling (Kennedy, 2014).

Daodi medicinal material (geoherb) is produced in particular geographic regions, where there is defined ecological environment and cultivation pipelines (Zhao et al., 2012). The clinical efficacy of geoherb is superior to that of the same medicinal plant growing in other regions. The special medicinal features of a plant are determined by its genome, while the proper ecological conditions have major effects on the formation of geoherb. For instance, Zhejiang of China is the best production area of the geoherb Bai Shao (*Paeonia lactiflora*), where the paeoniflorin content of *P. lactiflora* roots was positively correlated with soil pH and rhizosphere bacterial diversity (Yuan et al., 2014) but negatively correlated with the organic matter content of the rhizosphere. The rhizosphere soil properties have a close relationship with the geoherbalism of *F. thunbergii* (Zhe Bei Mu in TCM) (Shi et al., 2011) and *Pa. ginseng* (Ying et al., 2012).

The section *Moutan* of the genus *Paeonia* consists of eight species that are distributed in a particular area of China from which various secondary metabolites, including monoterpenoid glucosides, flavonoids, tannins, stilbenes, triterpenoids, steroids, paeonols, and phenols, have been found. The metabolic phenotype evolves in the differentiated niche and in response to the plant-insect and plant-microbe interactions (Yuan et al., 2014), which can be used for the chemotaxonomy of the section *Moutan* (He et al., 2014). Forty-three metabolites were identified from eight species by high-performance liquid chromatography-quadrupole time of flight-mass spectrometry, including 17 monoterpenoid glucosides, 11 galloyl glucoses, five flavonoids, six paeonols, and four phenols. PCA (principal component analysis) and HCA (hierarchical cluster analysis) showed a clear separation between the species based on metabolomic similarities, and four groups were identified, which coincides with conventional classification based on the morphological and geographical distributions. *P. decomposita*, from the geoherb production region of Sichuan, China, was found to be a transition species between two subsections. According to the metabolic fingerprints, *P. ostii* (Feng Dan in TCM) and *P. suffruticosa* (Mu Dan) could be the same species. The metabolic profiles of *P. delavayi* (wild Mu Dan) were highly variable, and no significant difference was found between *P. delavayi* and *P. ludlowii* (yellow Mu Dan), implying that they either have a close evolutionary relationship or underwent the convergent evolution of the specialized metabolism. The combination of metabolomics and multivariate analyses has great potential for guiding chemotaxonomic studies of other medicinal plants (Hao et al., 2012a).

With the surge of NGS technology, it is becoming common to perform the phylogenetic study based on genomic data. However, for most medicinal plants it is not realistic to rely on the whole genome sequencing data. RAD-Seq is easily applied to nonmodel plants for which no reference genome is available (Eaton and Ree, 2013; Fig. 1.3), and it is promising for reconstructing phylogenetic relationships in evolutionarily younger clades in which sufficient numbers of orthologous restriction sites are retained across species (Rubin et al., 2012). Coincidentally, the younger clades are more likely to harbor a wider variety of secondary metabolites, as the chemodiversity often accompanies the rapid radiation and diversification. The evolutionarily young *Pedicularis* section *Cyathophora* is a systematically refractory clade of the broomrape family (Orobanchaceae). The phylogenetic inferences were performed based on the data sets of 40,000 RAD loci (Eaton and Ree, 2013). The maximum likelihood and Bayesian methods generated similar trees that had two major clades: a "rex-thamnophila" clade, comprised of two species and several subspecies with relatively low floral diversity and geographically widespread distributions at lower elevations, and a "superb" clade, comprised of three species with high floral diversity and isolated geographic distributions at

higher altitudes. Levels of molecular divergence between subspecies in the rex-thamnophila clade are similar to those between species in the superba clade. The significant introgression among nearly all taxa in the rex-thamnophila clade was identified, while no gene flow was detected between clades or among taxa within the superba clade. The geographic isolation, following the uplift of QTP in the Quaternary Period and the emergence of "sky islands" (He and Jiang, 2014), might be crucial in the advent of species barriers by enabling local adaptation and differentiation without the influence of homogenizing gene flow. *Pedicularis* plants are traditionally used in folk medicine. It is intriguing to study its chemotaxonomy and treat the chemodiversity and biodiversity data in a holistic approach for drug discovery and development.

Understanding which factors determine chemical diversity has the potential to shed light on plant defenses against herbivores and diseases and accelerate drug discovery. Traditionally, *Cinchona* alkaloids were the primary treatment for malaria. The genetic profiles of *Cinchona calisaya* leaf samples were generated from four plastid and ITS regions of 22 *Cin. calisaya* stands in the Yungas region of Bolivia (Maldonado et al., 2017). Climatic and soil parameters were characterized and bark samples were analyzed for content of four major alkaloids to explore the utility of evolutionary history (phylogeny) in determining variation within species under natural conditions. A significant phylogenetic signal was found for the content of quinine and cinchonidine and total alkaloid content. Climatic parameters, primarily driven by changing altitude, predicted 20.2% of the overall alkaloid variation, and geographical separation accounted for a further 9.7%. A clade of high alkaloid–producing trees spanned a narrow range between 1100 and 1350 m. The climate expressed by altitude was not a significant driver when accounting for phylogeny, suggesting that the chemical diversity is primarily driven by phylogeny. Comparisons of the relative effects of both environmental/ecological and genetic variability in determining plant chemical diversity should be performed at the genotypic level if the extensive genotypic variation in plant biochemistry is to be fully understood.

1.5 PHARMACOPHYLOGENY VS. PHARMACOPHYLOGENOMICS

1.5.1 Concept

Diverse new terms are emerging in the genomic era, such as phylogenomics, pharmacophylogenomics, and phylotranscriptomics, which are somewhat overlapping with pharmaphylogeny (pharmacophylogeny/pharmacophylogenetics) (Hao et al., 2014a; Hao and Xiao, 2017). Phylogenomics is the crossing of evolutionary biology and genomics, in which genome data are utilized for evolutionary reconstructions. Pharmacophylogeny, advocated by Pei-gen Xiao since the 1980s (Xiao, 1980; Peng et al., 2006), focuses on the phylogenetic relationship of medicinal plants and aims to foster the sustainable utilization of TCM resources and is thus nurtured by molecular phylogeny, chemotaxonomy, ethnopharmacology, and bioactivity studies (Fig. 1.11). Phylogenomics can be integrated into the pipeline of drug discovery and development and can extend the field of pharmaphylogeny at the omic level, thus the concept of pharmacophylogenomics, initially emphasizing the genomic analysis of the evolutionary history of drug targets (Searls, 2003), could be redefined as an upgraded version of pharmacophylogeny.

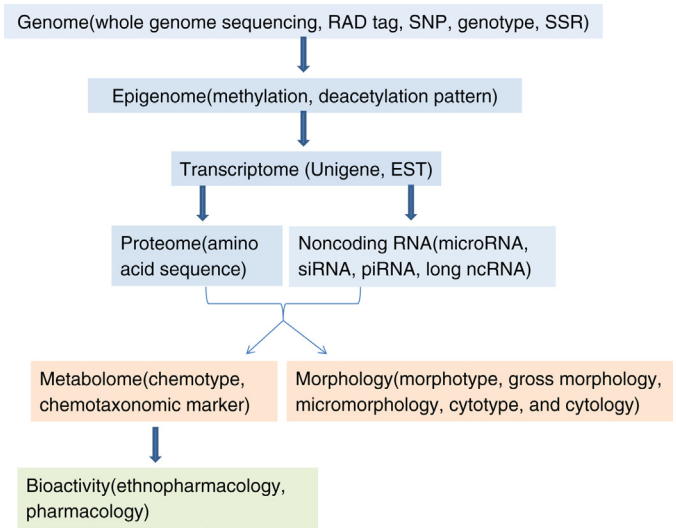

FIGURE 1.11 Omics data that could be used in the pharmacophylogeny inference. *EST*, Expressed sequence tag; *RAD*, restriction site associated DNA; *SNP*, single nucleotide polymorphism; *SSR*, simple sequence repeat.

The new conceptual framework of pharmacophylogenomics highlights the comprehensive analysis of the evolutionary history of medicinal organisms (especially the predominant medicinal plants), in particular the congruence and conflict between molecular phylogeny and chemotaxonomy (Day et al., 2014; Hao et al., 2013a–c; He et al., 2014), the orthology and paralogy relationships (Yang et al., 2015; Yang and Smith, 2014), the degree and landscape of evolutionary transformation they have undergone, and the involved evolving metabolic pathways and regulatory networks. More specifically, first, the tree of life of different scales can be constructed based on the genomic information to determine the phylogenomic relationship of medicinal plant groups, for example, the relationship between geoherb (higher content of medicinal compounds and better therapeutic efficacy) (Zhao et al., 2012) and nongeoherb populations; second, the genomic data, in particular those from the RAD-Seq or GBS, can be exploited to estimate the divergence time, reconstruct the geographic distribution, and infer the origin and the spatial distribution pattern of extant medicinal plants/geoherbs (Hao et al., 2014a, 2015d,e) (Fig. 1.3); third, within the context of the temporal tree, the ecological factors, environmental attributes, and evolutionarily innovative traits can be combined to dissect the diversification process and mechanism of medicinal plants; fourth, the origin and structure of the phylogenetic diversity of medicinal plants can be revealed; the diversity of medicinal compounds can be dissected based on biodiversity to promote drug discovery via high-throughput screening (Hao et al., 2015d,e); and last but not least, the dynamic alteration of the medicinal plant diversity can be predicted, and then the appropriate conservation and development strategies can be developed.

1.5.2 Molecular Phylogeny and Therapeutic Utility

During evolution, plants develop tactics of chemical defenses, leading to the evolution of specialized metabolites with diverse potencies. A correlation between phylogeny and biosynthetic pathways could offer a predictive approach enabling more efficient selection of

alternative and/or complementary plants for guaranteeing clinical use and novel lead discovery. This relationship has been rigorously tested, and the potential predictive power is subsequently validated (Rønsted et al., 2012). A phylogenetic hypothesis was proposed for the medicinal plant subfamily Amaryllidoideae (Amaryllidaceae) based on parsimony and Bayesian analysis of nuclear, cp, and mt DNA sequences of more than 100 species (Rønsted et al., 2012). It is intriguing to test whether alkaloid diversity and activity in bioassays related to the central nervous system are significantly correlated with molecular phylogeny. Evidence for a significant phylogenetic signal in these traits is found, albeit the effect is not strong. Several genera are nonmonophyletic, highlighting the importance of using phylogeny for understanding character distribution. Lack of congruence between specialized metabolism and molecular phylogeny is not unusual (Day et al., 2014; Hao et al., 2013a–c, 2015c; He et al., 2014), and the prominent factor is convergent evolution.

At least 20,654 phytochemicals from 16,102 plants are associated with 1592 human disease phenotypes. Only 8% of 36,932 phytochemicals are localized in certain parts of the taxonomy (Jensen et al., 2014). For example, the genus *Lens* (*Fabaceae*, Fabales), which includes lentils, and *Citrus* (*Rutaceae*, Sapindales), which includes orange, contain 60 out of 562 compounds and 42 out of 214 compounds, respectively, that are not found anywhere else on the taxonomy. On the other hand, compounds such as β-sitosterol, palmitic acid, and catechin are spread all over the taxonomy. It is possible that the synthesis of small compounds in plants is mainly defined by short-term regulatory rather than long-term evolutionary adaptation to the environment.

Alkaloid diversity and *in vitro* inhibition of acetylcholinesterase and binding to the serotonin reuptake transporter are significantly correlated with phylogeny (Rønsted et al., 2012), illustrating the validity of pharmacophylogeny, which has implications for the use of molecular phylogenies to interpret chemical evolution and biosynthetic pathways to select candidate taxa for lead discovery and to make policies regarding therapeutic use and conservation priorities. The phylogenetic classification was also taken into account in evaluating colchicine and related phenethylisoquinoline alkaloids of the family Colchicaceae (Larsson and Rønsted, 2014). The evolutionary reasoning can be utilized for inferring mechanisms in, for example, drug resistance in cancer and infections, which could exemplify how thinking about evolution influences the plant selection in drug lead discovery and how phylogeny knowledge may be used to evaluate predicted biosynthetic pathways.

The common practice of grouping medicinal plant uses into standardized categories, in terms of systems of the human body, may restrict the relevance of phylogenetic predictions (Ernst et al., 2016), as they only poorly reflect biological responses to the botanical drug. Medicinal plant uses should be interpreted from a perspective of the biological response, revealing different phylogenetic patterns of presumed underlying bioactivity. In the cosmopolitan and pharmaceutically highly relevant genus *Euphorbia* (Fig. 1.12), identifying anti-inflammatory uses highlighted a greater phylogenetic diversity and number of potentially promising species than standardized categories, which allow for a more targeted approach for future phylogeny-guided drug discovery at an early screening stage, possibly resulting in higher discovery rates of novel chemistry with functional bioactivity.

The correlation between the plant molecular phylogeny and therapeutic utility has been suggested (Grace et al., 2015; Leonti et al., 2013; Saslis-Lagoudakis et al., 2012). For instance, bulky, juicy leaves representative of medicinal aloes (Aloeaceae, Liliales) rose during the most recent expansion ~10 million years ago and are powerfully associated with

FIGURE 1.12 ***Euphorbia stracheyi*** **of Euphorbiaceae.** *Source: Taken in Shangri-La Alpine Botanical Garden, Yunnan, China.*

the molecular phylogeny and correlated to the probability of a species being used for therapy (Grace et al., 2015). A noteworthy, though feeble, phylogenetic hint is apparent in the remedial uses of aloes, signifying that their pharmaceutical properties do not arise stochastically across the clades of the evolutionary tree. The taxonomic clades included in native pharmacopoeias are indeed associated with certain disease groups, and ecology and angiosperm phylogeny, which could be the alternative and/or complementary for chemical kinship and convergence, to a certain extent explain the observed preference of the therapeutic use. For instance, evolutionarily related plants from New Zealand, Nepal, and the Cape of South Africa are used to combat diseases of the same therapeutic spaces (Saslis-Lagoudakis et al., 2012), which powerfully shows the self-determining discovery of the botanical value. A considerably greater fraction of recognized medicinal plants is present in these phylogenetic groups than in haphazard samples, suggesting that screening work be focused on a subgroup of traditionally used plants that are more affluent in medicinal molecules. The phylogenetic/phylogenomic cross-cultural evaluations would invigorate the use of old-fashioned knowledge in bioprospecting. Statistical analysis of the ethnopharmacology data based on Chinese medicinal plants of Magnoliidae (Xiao et al., 1986), Hamamelidae, and Caryophyllidae (Xiao et al., 1989) has been performed to summarize the distribution pattern of ethnomedicine uses across three subclasses. These nearly extinct traditional knowledges, collected nationwide during the TCM resources survey, lay a foundation for further quantitative correlation studies of molecular phylogeny and therapeutic efficacy.

1.5.3 Chinese Medicinal Plants

Chinese medicinal material resource is the foundation of the development of TCM. In the study of sustainable utilization of TCM resources, adopting innovative theories and methods to find new TCM resources is one of the hot spots and is always highlighted (Hao et al., 2015d,e). Pharmacophylogeny interrogates the phylogenetic relationship of medicinal organisms (especially medicinal plants), as well as the intrinsic correlation of morphological taxonomy, molecular phylogeny, chemical constituents, and therapeutic efficacy

(ethnopharmacology and pharmacological activity). This new discipline may have the power to change the way we utilize medicinal plant resources and develop plant-based drugs. Phylogenomics can be integrated into the flowchart of drug discovery and development and extends the field of pharmacophylogeny at the omic level. Analogously, phyloproteomics can be used in the proteome-based phylogeny study (Villar et al., 2013); phyloepigenomics could be used to examine the evolutionary relationship at the epigenomic level (Martin et al., 2011); and phylometagenomics is applicable in the exploration of medicinal plant-associated microbiota (Brindefalk et al., 2011).

The theory of TCM's property (Yao Xing in Chinese) is the core part of TCM theory. Meridian (Gui Jing in Chinese) theory is an important part of TCM's property theory. The medicine is selected according to the meridian to which it belongs, which improves the accuracy and pertinence of clinical drug use and is of great significance for guiding the clinical prescription of Chinese medicine. To study the association and distribution of TCMs with different meridian tropism on the phylogenetic tree, and to provide a basis for the interpretation and evaluation of TCM meridian tropism, 2435 herbs and a related 3044 species were screened (Li et al., 2017). Among species of the Viridiplantae, up to 1151 species belong to the liver meridian; among species of the Spermatophyta, up to 1109 species are classified into the liver meridian; among monocots, up to 110 species belong to the lung meridian. The association rules for the same meridian tropism were distributed on the same branch or nearby branch of the tree. For example, *Taxus* is related with the kidney meridian, *Caprifoliaceae* and *Rubia* have a close relationship with the liver meridian, and *Punica* is related to the colorectal meridian. There is a close relationship between Yao Xing and the phylogeny. TCMs with close phylogenetic relationships may have the same meridian tropism, which provides a new index and reference for prediction and evaluation of Yao Xing, selection and compatibility, and clinical application of new TCMs. Marine Chinese medicine (MCM) is one important part of TCM. The exploration of marine organism resources is a good base for the development of MCM, and the evaluation of Yao Xing of the new MCM resource is a key issue of the clinic application of MCM. A total of 613 MCMs and related 1091 marine species were screened (Fu et al., 2015). The MCMs of similar Yao Xing cluster on the phylogenetic tree. The MCMs from the same taxonomic family are more likely to have the same Yao Xing. For example, the marine plantae Chlorophyta, Florideophyceae, and Phaeohpyceaeare related with cold nature (Han Xing in Chinese), while the marine animalia Decapoda, Malacostraca, and Arthropoda are closely related with hot nature (Re Xing). The neutral nature (Ping Xing) was shown in Squamata. These results implied the close relationship between Yao Xing and the phylogeny, which can be used in predicting and evaluating Yao Xing of new MCM.

Many medicinally important tribes and genera, such as *Clematis* (Hao et al., 2013b), *Pulsatilla*, *Anemone*, *Cimicifugeae* (Hao et al., 2013d), *Nigella*, *Delphinieae* (Jabbour and Renner 2012) (Fig. 1.13), *Adonideae*, *Aquilegia*, *Thalictrum* (Zhu and Xiao 1991), and *Coptis*, belong to Ranunculaceae family. Chemical components of this family include several representative groups: BIA, ranunculin, triterpenoid saponin and diterpene alkaloid, etc. Ranunculin and magnoflorine were found to coexist in some genera. Other medicinal compounds also show some intriguing distribution patterns in 5 subfamilies and 10 tribes (Hao et al., 2015d,e; Wang et al., 2009). Compared with other plant families, Ranunculaceae has the most species recorded in China Pharmacopoeia (CP) version 2015. However, many *Ranunculaceae* species, for example, those that are closely related to CP species, as well as those endemic to China,

FIGURE 1.13 ***Delphinium yuanum*** **of Ranunculaceae.** *Source: Taken in Shangri-La Alpine Botanical Garden, Yunnan, China.*

have not been investigated in depth (Hao et al., 2015d,e), and their phylogenetic relationship and potential in medicinal use remain elusive. As such, it is proposed to select Ranunculaceae to exemplify the utility of pharmacophylogenomics and to elaborate the new concept empirically. It is argued that phylogenetic and evolutionary relationships of medicinally important tribes and genera within Ranunculaceae could be elucidated at the genomic, transcriptomic, and metabolomic levels, from which the intrinsic correlation between medicinal plant genotype and metabolic phenotype, and between genetic diversity and chemodiversity of closely related taxa, could be revealed. This proof-of-concept study would enrich the intension and spread the extension of pharmacophylogeny, promote the development of TCM genomics, and boost the sustainable development of Chinese medicinal plant resources.

1.5.4 Aconitum

Aconitum (Delphinieae, Ranunculaceae) has more than 300 species in the temperate regions of the Northern Hemisphere, over half of which are distributed in China. This genus has two subgenera, *Lycoctonum* and *Aconitum* (Hao et al., 2014a) (Fig. 1.14). The southwest China, particularly Hengduan Mountains, is the most important center of origin and diversity of the genus. Many *Aconitum* species are used as poisonous and medicinal plants. Their anticancer activity, cardioactive effect, analgesic activity, antiinflammatory activity, effect on energy metabolism, and antimicrobial and pesticidal activities—mainly due to the abundant diterpenoid alkaloids—are well-archived (Hao et al., 2013c). The correlation between molecular phylogeny, chemical components, and medicinal uses in *Aconitum* is notable (Hao et al., 2013c; Xiao et al., 2006). Diterpenoid alkaloids belong to four skeletal types: C_{18}, C_{19}, C_{20} and bisditerpenoid alkaloids. The subgenera *Lycoctonum* contain mainly the C_{18} (lappaconine-type and ranaconine-type) and C19 (lycoctonine-type). Roots of *Lycoctonum* plants exhibit a relatively lower toxicity and have been used to combat against rheumatism, pains, irregular menstruation, etc. This subgenus worths a more detailed phytochemical investigation for the new lead discovery and development. The Chinese taxa of section *Aconitum* (predominant in subgenera *Aconitum*) are morphologically divided into 11 series.

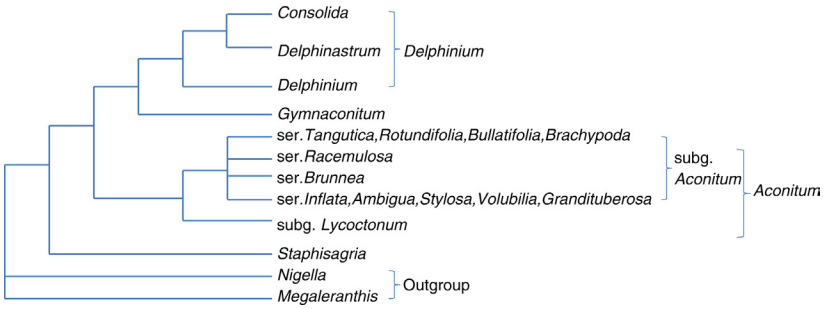

FIGURE 1.14 Cladogram of the Ranunculaceae tribe Delphinieae, according to Hao et al., 2013c and Wang et al., 2013. *Gymnaconitum* and *Staphisagria* were regarded as the subgenus of *Aconitum* and *Delphinium* respectively. *Consolida*, usually treated as an independent genus, could belong to the genus *Delphinium*.

The series *Tangutica* and *Rotundifolia* have abundant lactone-type C_{19}-diterpenoid alkaloids (Xiao et al., 2006), which can be considered as the chemical markers of these two series. The toxicity of their roots is much lower than those of the series *Bullatifolia* and *Brachypoda*, and the whole plants are traditionally used in Western China for high fever. The highly toxic aconitine-type diester C_{19} dominate in the series *Stylosa* (Da Wu Tou in TCM). The series *Ambigua* contains mainly the aconitine-type C_{19} with anisoyloxy residues, indicating its close affinity to series *Stylosa*. Several species of the series *Volubilia* have the highly advanced 15-hydroxyl aconitine-type C_{19}, indicating their possible kinship to series *Inflata*, which harbors two of the most widely used TCM/CP aconite species *A. carmichaeli* (Wu Tou in TCM) and *A. kusnezoffii* (Bei Wu Tou in TCM). *A. hemsleyanum* of the series *Volubilia*, as well as many other *Aconitum* herbs, is morphologically polymorphic and displays a substantial interpopulational phytochemical variation. The series *Grandituberosa* is more toxic than series *Inflata*, *Volubilia*, and *Ambigua*.

The morphology-based 11-series classification of section *Aconitum*, subgenus *Aconitum*, is not supported by chemotaxonomy and molecular phylogeny. Molecular phylogeny based on nr and cp DNA sequences divided the nine morphologically similar series into two clusters, which is bolstered by the chemotaxonomic data. Series *Rotundifolia* and *Brachypoda*, as well as *Tangutica* and *Bullatifolia*, are not monophyletic groups and cluster together. The series *Ambigua*, *Stylosa*, *Volubilia*, and *Inflata* are also not monophyletic but are intermingled on the phylogenetic tree (Hao et al., 2013c, 2015c). Series *Grandituberosa* is closer to *Volubilia* than to other series. *A. brunneum* and *A. racemulosum* are distinct in both molecular phylogeny and chemotaxonomy. *Gymnaconitum*, previously regarded as a subgenus of *Aconitum* but distinct phytochemically from *Aconitum*, is between *Aconitum* and the genus *Delphinium* in molecular phylogeny (Fig. 1.7) and now treated as an independent genus (Wang et al., 2013). The high possibility of deriving novel chemical entities from untapped species in traditionally used drug-productive genera/families has been suggested (Zhu et al., 2012). New genomic technologies that discover hidden gene clusters (Boutanaev et al., 2015), pathways, and interspecific crosstalk allow the unearthing of innovative natural products (Zhu et al., 2011). It is critical to assimilate the omic platforms into *Aconitum* studies for both the sustainable utilization of *Aconitum* pharmaceutical resources and finding novel compounds with potential clinical utility and less toxicity.

1.6 CONCLUSION AND PROSPECTS

The trend of integrating genomics and evolution into studies of medicinal plants is perceivable, and therefore, it is time to summarize the current progress in the relevant fields to make full use of evolutionary biology/genomics and revolutionize the roadmap of medicinal plant inquiries. Plants included in the same node of a phylogeny commonly have similar food and medicinal uses, which is called "ethnobotanical convergence" (Garnatje et al., 2017). This phylogenetic approach, together with the "omics" revolution, shows how combining modern technologies with traditional ethnobotanical knowledge could be used to identify potential new applications of plants. This chapter gives a brief analysis of the association and the distinguished features of the multifaceted medicinal plant evolution and genomics studies in the context of the plant-based drug discovery and the sustainable utilization of traditional pharmaceutical resources. A phylogenetic approach along with transcriptomics and other omics has value for understanding the evolution of medicinal plants, and a stronger case for the utility of these methods for future identification of useful genes and/or taxa for medicinal use is warranted.

The welfare of the global human population rests on provisioning services delivered by 12% of the Earth's ~400,000 plant species. People preferentially use large, widespread species rather than small, narrow-ranged species (Cámara-Leret al., 2017), but the latter potentially contain medicinally important compounds. Relying on plant size and availability may prevent the optimal realization of wild-plant services, since ecologically rare but chemically important clades cannot be overlooked. The research paradigms of medicinal plant genome and evolution are evolving, and the use of omics techniques is reshaping the landscape of this dynamic field. Genomics, transcriptomics, proteomics, metabolomics, and other omics platforms generate formidably big data, which cannot be used efficiently in probing plant genome and evolution without the aid of advancing bioinformatics. Medicinal plants evolve new traits to adapt to the changing environments and pave the road for themselves to a better life, while both hypothesis-driven and big data-driven, studies integrate herbal technology, biotechnology, and information technology to pave the road for human to a healthier life.

References

Ariyadasa, R., Stein, N., 2012. Advances in BAC-based physical mapping and map integration strategies in plants. J. Biomed. Biotechnol., 184854.

Barrett, C.F., Specht, C.D., Leebens-Mack, J., et al., 2014. Resolving ancient radiations: can complete plastid gene sets elucidate deep relationships among the tropical gingers (Zingiberales)? Ann. Bot. 113, 119–133.

Basavarajappa, H.D., Lee, B., Fei, X., et al., 2014. Synthesis and mechanistic studies of a novel homoisoflavanone inhibitor of endothelial cell growth. PLoS One 9 (4), e95694.

Boutanaev, A.M., Moses, T., Zi, J., 2015. Investigation of terpene diversification across multiple sequenced plant genomes. Proc. Natl. Acad. Sci. USA 112 (1), E81–E88.

Brindefalk, B., Ettema, T.J., Viklund, J., et al., 2011. A phylometagenomic exploration of oceanic alphaproteobacteria reveals mitochondrial relatives unrelated to the SAR11 clade. PLoS One 6, e24457.

Cai, Z., Liu, H., He, Q., et al., 2014. Differential genome evolution and speciation of *Coixlacryma-jobi* L. and *Coix aquatica* Roxb. hybrid guangxi revealed by repetitive sequence analysis and fine karyotyping. BMC Genomics (15), 1025.

Cai, J., Liu, X., Vanneste, K., et al., 2015. The genome sequence of the orchid *Phalaenopsis equestris*. Nat. Genet. 47 (1), 65–72.

Cámara-Leret, R., Faurby, S., Macía, M.J., et al., 2017. Fundamental species traits explain provisioning services of tropical American palms. Nat. Plants 3, 16220.
Cannon, S.B., McKain, M.R., Harkess, A., et al., 2015. Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. Mol. Biol. Evol. 32 (1), 193–210.
Cao, H., Hu, H., Colagiuri, B., et al., 2011. Medicinal cupping therapy in 30 patients with fibromyalgia: a case series observation. Forsch. Komplementmed. 18 (3), 122–126.
Chalhoub, B., Denoeud, F., Liu, S., et al., 2014. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. Science 345 (6199), 950–953.
Chen, S.L., Sun, Y., Xu, J., et al., 2010. Strategies of the study on Herb Genome Program. Acta Pharm. Sin. 45 (7), 807–812.
Chen, S.L., Luo, H., Li, Y., et al., 2011. 454 EST analysis detects genes putatively involved in ginsenoside biosynthesis in *Panax ginseng*. Plant Cell. Rep. 30 (9), 1593–1601.
Chen, W., Kui, L., Zhang, G., et al., 2017. Whole-genome sequencing and analysis of the Chinese herbal plant *Panax notoginseng*. Mol. Plant. 10 (6), 899–902.
Choisy, M., de Roode, J.C., 2014. The ecology and evolution of animal medication: genetically fixed response versus phenotypic plasticity. Am. Nat. 184(S1), S31–S46.
Clevenger, J., Chavarro, C., Pearl, S.A., et al., 2015. Single nucleotide polymorphism identification in polyploids: a review, example, and recommendations. Mol. Plant. 8 (6), 831–846.
Cviková, K., Cattonaro, F., Alaux, M., et al., 2015. High-throughput physical map anchoring via BAC-pool sequencing. BMC Plant Biol. 15 (1), 99.
Day, P.D., Berger, M., Hill, L., et al., 2014. Evolutionary relationships in the medicinally important genus *Fritillaria* L. (Liliaceae). Mol. Phylogenet. Evol. 80, 11–19.
Delaux, P.M., Varala, K., Edger, P.P., et al., 2014. Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. PLoS Genet. 10 (7), e1004487.
Denoeud, F., Carretero-Paulet, L., Dereeper, A., et al., 2014. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science 345 (6201), 1181–1184.
Dutartre, L., Hilliou, F., Feyereisen, R., 2012. Phylogenomics of the benzoxazinoid biosynthetic pathway of Poaceae: gene duplications and origin of the Bx cluster. BMC Evol. Biol. 12, 64.
Eaton, D.A., Ree, R.H., 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (Pedicularis: Orobanchaceae). Syst. Biol. 62 (5), 689–706.
Ernst, M., Saslis-Lagoudakis, C.H., Grace, O.M., et al., 2016. Evolutionary prediction of medicinal properties in the genus *Euphorbia* L. Sci. Rep. 6, 30531.
Fu, X.J., Wang, Z.G., Wang, C.Y., et al., 2015. The distribution and association relationships of marine Chinese medicine with different nature in the phylogenetic tree of marine organisms. World Sci. Technol./Mode Trad. Chin. Med. Mate Med. 17, 2189–2196.
Garnatje, T., Peñuelas, J., Vallès, J., 2017. Ethnobotany, phylogeny, and 'omics' for human health and food security. Trends Plant. Sci. 22 (3), 187–191.
Grace, O.M., Buerki, S., Symonds, M.R., et al., 2015. Evolutionary history and leaf succulence as explanations for medicinal use in aloes and the global popularity of aloe vera. BMC Evol. Biol. 15, 29.
Guichoux, E., Lagache, L., Wagner, S., et al., 2011. Current trends in microsatellite genotyping. Mol. Ecol. Resour. 11 (4), 591–611.
Han, F.M., Peng, Y., Xu, L., et al., 2014. Identification, characterization, and utilization of single copy genes in 29 angiosperm genomes. BMC Genomics 15, 504.
Hao, D.C., Xiao, P.G., 2017. An Introduction of Plant Pharmacophylogeny. Chemical Industry Press, Beijing.
Hao, D.C., Chen, S.L., Xiao, P.G., et al., 2010. Authentication of medicinal plants by DNA-based markers and genomics. Chin. Herb. Med. 2 (4), 250–261.
Hao D C., Ge, G., Xiao, P.G., et al., 2011a. The first insight into the tissue specific taxus transcriptome via Illumina second generation sequencing. PLoS One 6 (6), e21220.
Hao, D.C., Yang, L., Xiao, P.G., 2011b. The first insight into the Taxus genome via fosmid library construction and end sequencing. Mol. Genet. Genomics 285 (3), 197–205.
Hao, D.C., Xiao, P.G., Ge, G.B., et al., 2012a. Biological, chemical, and omics research of *Taxus* medicinal resources. Drug Dev. Res. 73, 477–486.
Hao, D.C., Chen, S.L., Xiao, P.G., et al., 2012b. Application of high-throughput sequencing in medicinal plant transcriptome studies. Drug Dev. Res. 73, 487–498.

Hao, D.C., Ma, P., Mu, J., et al., 2012c. De novo characterization of the root transcriptome of a traditional Chinese medicinal plant *Polygonum cuspidatum*. Sci. China Life Sci. 55 (5), 452–466.
Hao, D.C., Xiao, P.G., Peng, Y., et al., 2012d. Evaluation of the chloroplast barcoding markers by mean and smallest interspecific distances. Pak. J. Bot. 44 (4), 1271–1274.
Hao, D.C., Gu, X.J., Xiao, P.G., et al., 2013a. Phytochemical and biological research of *Fritillaria* medicine resources. Chin. J. Nat. Med. 11 (4), 330–344.
Hao, D.C., Gu, X.J., Xiao, P.G., 2013b. Chemical and biological research of *Clematis* medicinal resources. Chin. Sci. Bull. 58, 1120–1129.
Hao, D.C., Gu, X.J., Xiao, P.G., 2013c. Recent advances in the chemical and biological studies of *Aconitum* pharmaceutical resources. J. Chin. Pharm. Sci. 22 (3), 209–221.
Hao, D.C., Gu, X.J., Xiao, P.G., 2013d. Recent advance in chemical and biological studies on *Cimicifugeae* pharmaceutical resources. Chin. Herb. Med. 5 (2), 81–95.
Hao, D.C., Xiao, P.G., Liu, M., et al., 2014a. Pharmaphylogeny vs. pharmacophylogenomics: molecular phylogeny, evolution and drug discovery. Yao Xue Xue Bao 49 (10), 1387–1394.
Hao, D.C., Chen, S.L., Osbourn, A., et al., 2015a. Temporal transcriptome changes induced by methyl jasmonate in *Salvia sclarea*. Gene 558 (1), 41–53.
Hao, D.C., Vautrin, S., Song, C., et al., 2015b. The first insight into the *Salvia* (Lamiaceae) genome via BAC library construction and high-throughput sequencing of target BAC clones. Pak. J. Bot. 47 (4), 1347–1357.
Hao, D.C., Gu, X.J., Xiao, P.G., 2015c. Medicinal Plants: Chemistry, Biology and Omics, first ed. Elsevier-Woodhead, Oxford, ISBN 9780081000854.
Hao, D.C., Xiao, P.G., Liu, L.W., et al., 2015d. Essentials of pharmacophylogeny: knowledge pedigree, epistemology and paradigm shift. China J. Chin. Mat. Med. 40 (13), 1–8.
Hao, D.C., Xiao, P.G., Ma, H.Y., et al., 2015e. Mining chemodiversity from biodiversity: pharmacophylogeny of medicinal plants of the *Ranunculaceae*. Chin. J. Nat. Med. 13 (7), 507–520.
He, K., Jiang, X.L., 2014. Sky islands of southwest China I. An overview of phylogeographic patterns. Chin. Sci. Bull. 59, 585–597.
He, C.N., Peng, B., Dan, Y., et al., 2014. Chemical taxonomy of tree peony species from China based on root cortex metabolic fingerprinting. Phytochemistry 107, 69–79.
He, J., Zhao, X., Laroche, A., et al., 2014a. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Front. Plant Sci. 5, 484.
Henriquez, C.L., Arias, T., Pires, J.C., et al., 2014. Phylogenomics of the plant family Araceae. Mol. Phylogenet Evol. 75, 91–102.
Hofberger, J.A., Lyons, E., Edger, P.P., et al., 2013. Whole genome and tandem duplicate retention facilitated glucosinolate pathway diversification in the mustard family. Genome Biol. Evol. 5, 2155–2173.
Hofberger, J.A., Zhou, B., Tang, H., et al., 2014. A novel approach for multi-domain and multi-gene family identification provides insights into evolutionary dynamics of disease resistance genes in core eudicot plants. BMC Genomics 15, 966.
Hollister, J.D., Greiner, S., Wang, W., et al., 2015. Recurrent loss of sex is associated with accumulation of deleterious mutations in *Oenothera*. Mol. Biol. Evol. 32 (4), 896–905.
Howitz, K.T., Sinclair, D.A., 2008. Xenohormesis: sensing the chemical cues of other species. Cell 133 (3), 387–391.
Hsu, C.Y., Huang, P.L., Chen, C.M., et al., 2012. Tangy scent in *Toona sinensis* (Meliaceae) leaflets: isolation, functional characterization, and regulation of TsTPS1 and TsTPS2, two key terpene synthase genes in the biosynthesis of the scent compound. Curr. Pharm. Biotechnol. 13 (15), 2721–2732.
Jabbour, F., Renner, S.S., 2012. A phylogeny of *Delphinieae* (*Ranunculaceae*) shows that *Aconitum* is nested within *Delphinium* and that Late Miocene transitions to long life cycles in the Himalayas and Southwest China coincide with bursts in diversification. Mol. Phylogenet. Evol. 62 (3), 928–942.
Jensen, K., Panagiotou, G., Kouskoumvekaki, I., 2014. Integrated text mining and chemoinformatics analysis associates diet to health benefit at molecular level. PLoS Comput. Biol. 10 (1), e1003432.
Jiao, Y., Paterson, A.H., 2014. Polyploidy-associated genome modifications during land plant evolution. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 369 (1648)doi: 10.1098/rstb.2013.0355.
Jiao, J., Lü, G., Liu, X., et al., 2011a. Reduction of blood lead levels in lead-exposed mice by dietary supplements and natural antioxidants. J. Sci. Food Agric. 91 (3), 485–491.
Jiao, Y., Wickett, N.J., Ayyampalayam, S., et al., 2011b. Ancestral polyploidy in seed plants and angiosperms. Nature 473 (7345), 97–100.

Kaliswamy, P., Vellingiri, S., Nathan, B., et al., 2015. Microsatellite analysis in the genome of *Acanthaceae*: an *in silico* approach. Pharmacogn. Mag. 11 (41), 152–156.
Karimi A, Hadian, J., Farzaneh, M., et al., 2015. Evaluation of genetic variability, rust resistance and marker-detection in cultivated *Artemisia dracunculus* from Iran. Gene 554 (2), 224–232.
Kennedy, D.O., 2014. Polyphenols and the human brain: plant "secondary metabolite" ecologic roles and endogenous signaling functions drive benefits. Adv. Nutr. 5 (5), 515–533.
Kim, S., Park, M., Yeom, S.I., et al., 2014. Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. Nat. Genet. 46 (3), 270–278.
Koenig, D., Jiménez-Gómez, J.M., Kimura, S., et al., 2013. Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. Proc. Natl. Acad. Sci. USA 110 (28), E2655–E2662.
Koide, C.L., Collier, A.C., Berry, M.J., et al., 2011. The effect of bamboo extract on hepatic biotransforming enzymes—findings from anobese-diabetic mouse model. J. Ethnopharmacol. 133 (1), 37–45.
Larsson, S., Rønsted, N., 2014. Reviewing Colchicaceae alkaloids—perspectives of evolution on medicinal chemistry. Curr. Top. Med. Chem. 214 (2), 274–289.
Leonti, M., Cabras, S., Castellanos, M.E., et al., 2013. Bioprospecting: evolutionary implications from a post-olmec pharmacopoeia and the relevance of widespread taxa. J. Ethnopharmacol. 147 (1), 92–107.
Li, X.K., Wang, B., Zheng, Y.C., et al., 2014. Molecular characters of *Schisandra chinensis* with white fruit by RAPD and ISSR makers. Zhong Yao Cai. 37 (4), 568–572.
Li, M.M., Kasun, M., Yan, L., et al., 2015. Genetic involvement of *Camellia taliensis* in the domestication of *C. sinensis* var. *assamica* (Assimica Tea) revealed by nuclear microsatellite markers. Plant Diversity Resour. 37 (1), 29–37.
Li, J.Y., Fu, X.J., Li, X.B., et al., 2017. Association relationships of traditional Chinese medicine with different meridian tropism in phylogenetic tree. Chin. J. Expe. Trad. Med. Form 23 (12), 194–200.
Liu, L., Ma, X., Wei, J., et al., 2011. The first genetic linkage map of Luohanguo (*Siraitia grosvenorii*) based on ISSR and SRAP markers. Genome 54 (1), 19–25.
Liu W, Yin, D., Liu, J., et al., 2014. Genetic diversity and structure of *Sinopodophyllum hexandrum* (Royle) Ying in the Qinling Mountains, China. PLoS One 9 (10), e110500.
Liu, X., Liu, Y., Huang, P., et al., 2017. The genome of medicinal plant *Macleaya cordata* provides new insights into benzylisoquinoline alkaloids metabolism. Mol. Plant 10 (7), 975–989.
Luo, J., Hou, B.W., Niu, Z.T., et al., 2014. Comparative chloroplast genomes of photosynthetic orchids: insights into evolution of the Orchidaceae and development of molecular markers for phylogenetic applications. PLoS One 9 (6), e99016.
Ma, P.F., Zhang, Y.X., Zeng, C.X., et al., 2014. Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (poaceae). Syst. Biol. 63 (6), 933–950.
Maldonado, C., Barnes, C.J., Cornett, C., et al., 2017. Phylogeny Predicts the Quantity of Antimalarial Alkaloids within the Iconic Yellow Cinchona Bark (Rubiaceae: *Cinchona calisaya*). Front Plant Sci. 8, 391.
Malé, P.J., Bardon, L., Besnard, G., et al., 2012. Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. Proc. Natl. Acad. Sci. USA 109, 17519–17524.
Manrique-Carpintero, N.C., Tokuhisa, J.G., Ginzberg, I., et al., 2013. Sequence diversity in coding regions of candidate genes in the glycoalkaloid biosynthetic pathway of wild potato species. G3 (Bethesda) 3, 1467–1479.
Martin, D.I., Singer, M., Dhahbi, J., et al., 2011. Phyloepigenomic comparison of great apes reveals a correlation between somatic and germline methylation states. Genome Res. 21, 2049–2057.
Morales-Sánchez, V., Rivero-Cruz, I., Laguna-Hernández, G., et al., 2014. Chemical composition, potential toxicity, and quality control procedures of the crude drug of *Cyrtopodium macrobulbon*. J. Ethnopharmacol. 154 (3), 790–797.
Moses, T., Pollier, J., Shen, Q., et al., 2015. OSC2 and CYP716A14v2 catalyze the biosynthesis of triterpenoids for the cuticle of aerial organs of *Artemisia annua*. Plant Cell. 27 (1), 286–301.
Nagananda, G.S., Satishchandra, N., 2013. Antimicrobial activity of cold and hot successive pseudobulb extracts of *Flickingerianodosa* (Dalz) Seidenf. Pak. J. Biol. Sci. 16 (20), 1189–1193.
Noskov, V.N., Chuang, R.Y., Gibson, D.G., et al., 2011. Isolation of circular yeast artificial chromosomes for synthetic biology and functional genomics studies. Nat. Protoc. 6 (1), 89–96.
Pan, Y.Z., Zhang, Y.C., Gong, X., et al., 2014. Estimation of genome size of four *Panax* species by flow cytometry. Plant Diversity Resour. 36 (2), 233–236.
Park, S., Ruhlman, T.A., Sabir, J.S., et al., 2014. Complete sequences of organelle genomes from the medicinal plant *Rhazya stricta* (Apocynaceae) and contrasting patterns of mitochondrial genome evolution acrossasterids. BMC Genomics 15, 405.

Peng, Y., Chen, S.B., Liu, Y., et al., 2006. A pharmacophylogenetic study of the Berberidaceae (s.l). Acta Phytotaxo. Sin. 44 (3), 241–257.
Polashock, J., Zelzion, E., Fajardo, D., et al., 2014. The American cranberry: first insights into the whole genome of a species adapted to bog habitat. BMC Plant Biol. 14, 165.
Ponnuchamy, S., Kanchithalaivan, S., Ranjith Kumar, R., et al., 2014. Antimycobacterial evaluation of novel hybrid arylidene thiazolidine-2,4-diones. Bioorg. Med. Chem. Lett. 24 (4), 1089–1093.
Qi, H.Y., Li, L., Yu, J., 2013. Xenohormesis: understanding biological effects of traditional Chinese medicine from an evolutionary and ecological perspective. Zhongguo Zhong Yao Za Zhi 38 (19), 3388–3394.
Qian, J., Song, J., Gao, H., et al., 2013. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. PLoS One 8 (2), e57607.
Qin, C., Yu, C., Shen, Y., et al., 2014. Whole-genome sequencing of cultivated and wild peppers provides insights into Capsicum domestication and specialization. Proc. Natl. Acad. Sci. USA 111 (14), 5135–5140.
Quan, Q.M., Chen, L.L., Wang, X., et al., 2014. Genetic diversity and distribution patterns of host insects of caterpillar fungus *Ophiocordyceps sinensis* in the Qinghai-Tibet Plateau. PLoS One 9 (3), e92293.
Rønsted, N., Symonds, M.R., Birkholm, T., et al., 2012. Can phylogeny predict chemical diversity and potential medicinal activity of plants? A case study of Amaryllidaceae. BMC Evol. Biol. 12, 182.
Rubin, B.E., Ree, R.H., Moreau, C.S., 2012. Inferring phylogenies from RAD sequence data. PLoS One 7 (4), e33394.
Saslis-Lagoudakis, C.H., Savolainen, V., Williamson, E.M., et al., 2012. Phylogenies reveal predictive power of traditional medicine in bioprospecting. Proc. Natl. Acad. Sci. USA 109 (39), 15835–15840.
Searls, D.B., 2003. Pharmacophylogenomics: genes, evolution and drug targets. Nat. Rev. Drug Discov. 2 (8), 613–623.
Shi, J.Y., Yuan, X.F., Lin, H.R., et al., 2011. Differences in soil properties and bacterial communities between the rhizosphere and bulk soil and among different production areas of the medicinal plant *Fritillaria thunbergii*. Int. J. Mol. Sci. 12 (6), 3770–3785.
Sternberg, E.D., de Roode, J.C., Hunter, M.D., 2015. Trans-generational parasite protection associated with paternal diet. J. Anim. Ecol. 84 (1), 310–321.
Su, H.J., Hogenhout, S.A., Al-Sadi, A.M., et al., 2014. Complete chloroplast genome sequence of Omani lime (*Citrus aurantiifolia*) and comparative analysis within the rosids. PLoS One 9 (11), e113049.
Sveinsson, S., McDill, J., Wong, G.K., et al., 2014. Phylogenetic pinpointing of a paleopolyploidy event within the flax genus (*Linum*) using transcriptomics. Ann. Bot. 113 (5), 753–761.
Villar, M., Popara, M., Mangold, A.J., et al., 2013. Comparative proteomics for the characterization of the most relevant Amblyomma tick species as vectors of zoonotic pathogens worldwide. J. Proteomics 105, 204–216.
Wang, W., Lu, A.M., Ren, Y., et al., 2009. Phylogeny and classification of Ranunculales evidence from four molecular loci and morphological data. Persp. Plant. Ecol. Evol. Syst. 11, 81–110.
Wang, J., Yue, Y.D., Tang, F., et al., 2012. Screening and analysis of the potential bioactive components in rabbit plasma after oral administration of hot-water extracts from leaves of *Bambusa textilis* McClure. Molecules 17 (8), 8872–8885.
Wang, W., Liu, Y., Yu, S.X., et al., 2013. *Gymnaconitum*, a new genus of Ranunculaceae endemic to the Qinghai-Tibetan Plateau. Taxon 62, 713–722.
Wang, S., Liu, Y., Ma, L., et al., 2014. Isolation and characterization of microsatellite markers and analysis of genetic diversity in Chinese jujube (*Ziziphus jujuba* Mill). PLoS One 9 (6), e99842.
Wee, A.K., Takayama, K., Chua, J.L., et al., 2015. Genetic differentiation and phylogeography of partially sympatric species complex *Rhizophoramucronata* Lam. and *R. stylosa* Griff. using SSR markers. BMC Evol. Biol. 15, 57.
Wen, J., Zhang, J.Q., Nie, Z.L., et al., 2014. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. Front. Genet. 5, 4.
Whittall, J.B., Syring, J., Parks, M., et al., 2010. Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. Mol. Ecol. 19(S1), 100–114.
Wickett, N.J., Mirarab, S., Nguyen, N., et al., 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci. USA 111 (45), E4859–E4868.
Wu, C.S., Chaw, S.M., Huang, Y.Y., 2013. Chloroplast phylogenomics indicates that *Ginkgo biloba* is sister to cycads. Genome Biol. Evol. 5, 243–254.
Wu, K., Mo, C., Xiao, H., et al., 2015. Imperialine and verticinone from bulbs of *Fritillaria wabuensis* inhibit pro-inflammatory mediators in LPS-stimulated RAW 264.7 macrophages. Planta Med. 81 (10), 821–829.
Wysocki, W.P., Clark, L.G., Attigala, L., et al., 2015. Evolution of the bamboos (Bambusoideae; Poaceae): a full plastomephylogenomic analysis. BMC Evol. Biol. 15, 50.

Xiao, P.G., 1980. A preliminary study of the correlation between phylogeny, chemical constituents and pharmaceutical aspects in the taxa of Chinese Ranunculaceae. Acta Phytotaxo Sin 18 (2), 142–153.
Xiao, P.G., Wang, L.W., Lv, S.J., 1986. Statistical analysis of the ethnopharmacologic data based on Chinese medicinal plants by electronic computer I Magnoliidae. Chin. J. Integ. Trad. West Med. 6 (4), 253–256.
Xiao, P.G., Wang, L.W., Qiu, G.S., et al., 1989. Statistical analysis of the ethnopharmacologic data based on Chinese medicinal plants by electronic computer II *Hamamelidae* and *Caryophyllidae*. Chin. J. Integ. Trad. West. Med. 9 (7), 429–432.
Xiao, P.G., Wang, F.P., Gao, F., et al., 2006. A pharmacophylogenetic study of *Aconitum* L (Ranunculaceae) from China. Acta Phytotaxon Sin 44 (1), 1–46.
Xu, Q., Xiong, G., Li, P., et al., 2012. Analysis of complete nucleotide sequences of 12 Gossypium chloroplast genomes: origin and evolution of allotetraploids. PLoS One 7 (8), e37128.
Xu, J., Chu, Y., Liao, B., et al., 2017. *Panax ginseng* genome examination for ginsenoside biosynthesis. Gigascience 6 (11), 1–15.
Yang, Y., Smith, S.A., 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. Mol. Biol. Evol. 31 (11), 3081–3092.
Yang, Y., Moore, M.J., Brockington, S.F., et al., 2015. Dissecting molecular evolution in the highly diverse plant clade Caryophyllales using transcriptome sequencing. Mol. Biol. Evol. 32 (8), 2001–2014.
Yang, J., Zhang, G., Zhang, J., et al., 2017. Hybrid de novo genome assembly of the Chinese herbal fleabane *Erigeron breviscapus*. Gigascience 6 (6), 1–7.
Yao, X., Peng, Y., Xu, L.J., et al., 2011. Phytochemical and biological studies of *Lycium* medicinal plants. Chem. Biodivers. 8 (6), 976–1010.
Ying, Y.X., Ding, W.L., Li, Y., 2012. Characterization of soil bacterial communities in rhizospheric and nonrhizospheric soil of *Panax ginseng*. Biochem. Genet. 50 (11–12), 848–859.
You, F.M., Huo, N., Gu, Y.Q., et al., 2008. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics 9 (253).
Yuan, X.F., Peng, S.M., Wang, B.L., et al., 2014. Effects of growth years of *Paeonia lactiflora* on bacterial community in rhizosphere soil and paeoniflorin content. Zhongguo Zhong Yao Za Zhi 39 (15), 2886–2892.
Zeng, Y., Guo, L.P., Chen, B.D., et al., 2013. Arbuscular mycorrhizal symbiosis for sustainable cultivation of Chinese medicinal plants: a promising research direction. Am. J. Chin. Med. 41 (6), 1199–1221.
Zhan, Q.Q., Sui, C., Wei, J.H., et al., 2010. Construction of genetic linkage map of *Bupleurum chinense* DC. using ISSR and SSR markers. Yao Xue Xue Bao 45 (4), 517–523.
Zhang, F., Gao, Q., Khan, G., et al., 2014. Comparative transcriptome analysis of aboveground and underground tissues of *Rhodiolaalgida*, an important ethno-medicinal herb endemic to the Qinghai-Tibetan Plateau. Gene 553 (2), 90–97.
Zhang, G., Tian, Y., Zhang, J., 2015. Hybrid de novo genome assembly of the Chinese herbal plant danshen (*Salvia miltiorrhiza* Bunge). Gigascience 4, 62.
Zhang, D., Li, W., Xia, E.H., et al., 2017. The medicinal herb *Panax notoginseng* genome provides insights into ginsenoside biosynthesis and genome evolution. Mol. Plant. 10 (6), 903–907.
Zhao, Z.Z., Guo, P., Brand, E., 2012. The formation of daodi medicinal materials. J. Ethnopharmacol. 140 (3), 476–481.
Zhao, Y., Yin, J., Guo, H., et al., 2015. The complete chloroplast genome provides insight into the evolution and polymorphism of *Panax ginseng*. Front. Plant. Sci. 5, 696.
Zhu, M., Xiao, P.G., 1991. Chemosystematic studies on *Thalictrum* L. in China. Acta Phytotaxon Sin 29 (4), 358–369.
Zhu, F., Qin, C., Tao, L., et al., 2011. Clustered patterns of species origins of nature-derived drugs and clues for future bioprospecting. Proc. Natl. Acad. Sci. USA 108 (31), 12943–12948.
Zhu, F., Ma, X.H., Qin, C., et al., 2012. Drug discovery prospect from untapped species: indications from approved natural product drugs. PLoS One 7 (7), e39782.