

An Updated Snapshot of Recent Advances in Transcriptomics and Genomics of Phytomedicinals

Biswapriya B. Misra

Department of Biology, Genetics Institute, Plant Molecular and Cellular Biology Program, University of Florida, Gainesville, FL 32610, USA
Email: biswapriyamisra@ufl.edu

Abstract

Medicinal plants have been of great importance to human health care since the advent of medicine. A huge array of molecules has been obtained from these phytopharmaceutical-yielding species that have influenced human lives since the beginning of plant-based life-saving medicines. Some of these molecules have taken the form of taxol, aspirin, and artemisinin. With the flourishing era of high-throughput next generation sequencing technologies, a hot pursuit for sequencing the genomes and transcriptomes of these life-saving plants is underway. Although few genomes have been sequenced or are currently being addressed, the number of transcriptomes sequenced has sky-rocketed in the last couple of years and continues to surge forward with immense pace, covering all important genera of medicinal plants. I have attempted to provide the current status, progress, opportunities, and challenges of these sequencing endeavors in this comprehensive and updated review. It is my hope that this information will provide both specialists and non-specialists with the current trends and future directions of this interesting category of plants.

Keywords: medicinal plant, metabolic pathway, genome, next generation sequencing, transcriptome

Introduction

Metabolites are small chemical entities present in living organisms with a molecular weight of less than 1000 Da. The collection of such small molecules in organelles, cells, tissues, organs and bio-fluids at a particular time are known as the metabolome. The plant metabolome is estimated to be at a size of 200,000 Da or more (Goodacre *et al.* 2004). Furthermore, a statistical analysis predicted the number of plant metabolites to fall in a range between 200,000 and 1,000,000 compounds (Afendi *et al.* 2012). Historically, the metabolomes of these plants that yield medicinal compounds, known as phytomedicinals, are a result of the 'secondary' and 'specialized' metabolism operative contained within them.

The World Health Organization (WHO) recorded that 80% of the world's developing population depends primarily on herbal medicine for basic healthcare (Vines, 2004). These phytomedicine-yielding plant species that are used in herbal

medicine are generally known as medicinal plants. They belong to typical taxonomic families, mostly dicot angiosperms (i.e., Lamiaceae, Asteraceae, Apocyanaceae, Scrophulariaceae, and Solanaceae). Plant-based drugs have traditionally been extracted and isolated in small or industrial scale quantities using the original plant materials. However, this method of harvesting these drugs is drastically changing, with the recombinant production of phytomedicinals using microbial models emerging. Yeast, *Escherichia coli*, and a host of other models are the microbes being used to extract taxol, artemisinin and other drugs (Xiao *et al.*, 2013). In order to obtain a detailed understanding of the specialized metabolism operative in these plants, metabolite profiling, biochemical approaches, natural products chemistry, gene-cloning, protein purification, and gene expression approaches were used. These techniques yielded information on specific proteins, novel transcripts, and chemical

processes being undergone in these plants. The technological breakthroughs achieved in the past couple of decades have enabled high throughput studies of plants' genomes, transcriptomes, proteomes, and metabolomes with a fair amount of time, effort, and costs. Out of ~ 370,000 known plant species only ~80,000 species are represented by at least a single GenBank entry (Carlos *et al.*, 2012).

Plant sequencing efforts have been centered mostly on angiosperms; more specifically, the eudicots that include economically important crops. For example, expressed sequence tag (EST) sequencing and mRNA profiling using either microarrays or serial analysis of gene expression (SAGE), and digital gene expression (DGE) now allow for a comprehensive analysis of the transcriptome (Sumner *et al.*, 2003). Additionally, the advent of high throughput methods, like next generation sequencing platforms, has enabled whole-genome sequencing (WGS) advances. Unfortunately, hardly any compilations have been attempted to gauge the current progress of medicinal plant genomes; although transcriptomes have been discussed recently (Xiao *et al.*, 2013). In fact, the compilation of Chen *et al.* 2011 provides a comprehensive overview of the development of sequencing technologies, summarizes the sequenced plant genomes, discusses the approaches in sample preparation for sequencing, genome assembly, annotation, and gene family mining, but largely fails to identify the advances in medicinal plant genomics in a realistic sense. On the other hand, most genome sequencing attempts have focused on model plants i.e., dicot model *Arabidopsis thaliana*, monocot model rice (*Oryza sativa*), unicellular viridiplantae *Chlamydomonas*, woody plant model *Populus trichocarpa* and mostly other commercially beneficial plants such as coffee, soybean, castor bean, strawberry, maize, sorghum, papaya, grape, cotton, ginseng, apple, rubber, cucumber and others. Thus medicinal plant genomics has been a less-focused field of investigation, apart from the sequenced genomes of neem and hemp

Owing to the large amounts of specialized metabolites that each of the green plant species produce, they all can be labeled as medicinal plants. For convenience, I have labeled and addressed only those species which are cultivated or exploited for their medicinal properties, either ethnopharmacologically or more recently as a source of new medicine. In this review, I attempt to update both the specialized and non-specialized readers on the present status of medicinal plant genomics and transcriptomics with exhaustive literature search and data mining.

Significant progress to date

The list of sequenced plant genomes include 55-61 dicots, 16 monocots, 18 algae, and one each of pteridophyte, bryophyte, and gymnosperm [CoGePedia: http://genomeevolution.org/wiki/index.php/Sequenced_plant_genomes and List of sequenced plant genomes: http://en.wikipedia.org/wiki/List_of_sequenced_plant_genomes]. Most of these plants were sequenced using Sanger sequencing or hybrid multi-platform approaches, such as 454, SOLiD and Illumina. The most efficient sequencing platforms, such as the Illumina HiSeq 2000, sequence ~100 Gbp/ day, capable of sequencing a 3 Gbp genome at high coverage for >US\$10,000 (Schatz *et al.*, 2012). However, the efforts in sequencing the medicinal plant genomes are very limited. The only medicinal plant genomes sequenced are that of the Neem tree (*Azadirachta indica*) by GANIT Labs, which reveals a 364 Mbp, with 20,000 genes supported by transcriptomes obtained from shoot, root, leaf, flower, and seeds (Krishnan *et al.*, 2012), and hemp (*Cannabis sativa*), that has a 820 Mbp genome and ~30,000 genes (van Bakel *et al.*, 2011). More recently, the chloroplast genomes of several medicinal plants were reported. These genomes were shown to facilitate population, phylogenetic, and chloroplast-based genetic engineering studies of these medicinal plants, such as that of *Salvia miltiorrhiza* (Qian *et al.*, 2013), Mongolia medicine *Artemisia frigida* (Liu *et al.*, 2013), and *Mahonia bealei*

(Berberidaceae) (Ma *et al.*, 2013). All these efforts in genome sequencing could lead to development of stronger comparative genomics resources for furthering the development of medicinal plants.

The launch of Traditional Chinese Medicine (TCM) modernization projects in China and elsewhere in recent years has made remarkable progress of herbal medicine in laboratory development, ingredient profiling, and manufacture production (Zhang *et al.*, 2014). Expectedly, transcriptomics studies in TCM-yielding plants have been the most extensively studied (Hao *et al.*, 2012). For example, using 454 pyrosequencing, Chen *et al.*, have obtained transcriptomic data sets for 10 medicinal plants, i.e., yew (*Taxus cuspidate*), Yin Xing (*Ginkgo biloba*), Qian Ceng Ta (*Huperzia serrata*), *Phlegmarius carinatus*, American ginseng (*Panax quinquefolius*), Chinese sage/ Dan Shen (*Salvia miltiorrhiza*), San Qi/ Tian Qi (*P. notoginseng*), Happy tree (*Camptotheca acuminata*), Chinese liquorice (*Glycyrrhiza uralensis*) and *P. ginseng*, mostly yielding wealthy information of cytochrome P450 (CYP 450s) and glucosyltransferases. Transcriptomic analysis of the medicinal plant *Catharanthus roseus* aided in the construction of a detailed metabolic pathway database, 'CathaCyc' that contains 390 pathways (Moerkercke *et al.*, 2013). Similarly, the transcriptome analysis aided with untargeted metabolic profiling in anti-cancer alkaloid (camptothecin and anthraquinone) producing medicinal plant *Ophiorrhiza pumila* yielded a wealth of information on this precious genomic resource. Additionally, the transcriptome and metabolome data sets were compared for the alkaloid-producing hairy root with a non-alkaloid producing cell suspension culture (Yamazaki *et al.*, 2013). Furthermore, *Glycyrrhiza* plants produce several bioactive phytochemicals. Glycyrrhizin, for example, exhibits pluripotent pharmacological activities. A detailed transcriptome analysis in *G. uralensis*, provided interesting information on the roles of CYP450s and vacuolar transporters in the biosynthesis of glycyrrhizin (Ramilowski *et al.*, 2013).

The need for genome-based approaches for the authentication of medicinal plants has been realized (Sucher and Carles, 2008). This is done by using genome-based features such as traditional PCR-sequencing based methods, molecular marker development, DNA microarray, genotyping, and bar coding. The authors provide a very exhaustive list of 511 medicinal plants and a list of publications on the application of genome-based methods for the authentication of medicinal plants.

Hao *et al.* 2012, have provided an elaborate account of application of high-throughput sequencing in medicinal plant transcriptome studies, where they have clearly categorized the transcriptomic studies on those used for (i) metabolic pathway studies, and (ii) molecular markers. They also have identified simple sequence repeats (SSRs), single nucleotide polymorphisms (SNPs) and digital gene expression (DGE) as popular ways for molecular marker mining studies. A large-scale and integrative transcriptome sequence and expression profiling study for three species of Asterids that produce medicinally important monoterpenoid indole alkaloids (MIAs) was conducted across a wide range of developmental tissues and in cultured cells and roots, following elicitor treatments in *C. roseus*. The MIAs are the major source of therapeutics for treatment of cancer (vinblastine, vincristine, and camptothecin), hypertension (reserpine, ajmalicine), malaria (quinine), and analgesics (7-hydroxymitragynine). In this study, exhaustive transcriptomics resources were established in various developmental stages of cells and tissues: *in vitro* and *in vivo*, mutant and wild type, and elicitor treated seedlings in *C. roseus*. This provided a unique and huge resource for the understanding of MIA biosynthesis and a holistic understanding into the metabolism of this important medicinal. The coexpression analysis of MIA biosynthetic pathway genes approach, followed in this comparative study of the three MIA producing species, would enable elucidation

of the most chemically prolific, but difficult, species of plants (Góngora-Castillo *et al.*, 2012).

Furthermore, seven candidate DNA barcodes (*psbA-trnH*, *matK*, *rbcL*, *rpoC1*, *ycf5*, *ITS2*, and *ITS*) from medicinal plant species were investigated for their PCR amplification efficiency, differential intra- and inter-specific divergences, and the DNA bar-coding gap (Chen *et al.*, 2013). Results indicated that the second internal transcribed spacer (ITS2) of nuclear ribosomal DNA was rendered most efficient for DNA barcoding applications that was reflected from the ability to successfully discriminate ITS2 in more than 6600 plant samples belonging to 4800 species from 753 diverse genera, a 92.7% success at the species level (Chen *et al.*, 2013). In addition, an integrated web medicinal materials DNA database: MMDBD (Medicinal Materials DNA Barcode Database) [<http://137.189.42.34/mherbsdb/>] that provides a medicinal materials DNA barcode database and several bioinformatics tools for data storage, analysis and exchange for promotion of the identification of medicinal materials (Lou *et al.*, 2010). These interesting and robust developments may impact the way medicinal plants could be identified and archived in years to come.

Global players with formidable steps ahead

The 1000 plants (oneKP or 1KP) initiative is an international multi-disciplinary consortium generating large-scale gene sequencing data for over 1000 species of plants. Major supporters include Alberta Enterprise and Advanced Education, Musea Ventures (Somekh Family Foundation), Beijing Genomics Institute in Shenzhen (BGI-Shenzhen), Alberta Innovates Technology Futures (AITF-iCORE Strategic Chair), iPlant Tree-of-Life (iPToL) Grand Challenge, and WestGrid Compute-Calcul. They intend to sequence representatives of nearly all of the 415 known angiosperm families. OneKP Capstone Wiki enlists the highlights of the activities and success of the project. The medicinal plant transcriptomes that have been sequenced

include the following: *Argemone mexicana*, *Eschscholzia californica*, *Papaver bracteatum*, *Papaver rhoeas*, *Papaver setigerum*, *Papaver somniferum*, *Psychotria ipecacuanha*, *Larrea tridentate*, *Camptotheca acuminata*, *Allamanda cathartica*, *Catharanthus roseus*, *Dioscorea villosa*, *Colchicum autumnale*, *Gloriosa superba*, *Taxus baccata*, and *Taxus cuspidate*. The genome sequences for these plants are currently in the process of being sequenced. The SOL-100 sequencing project (<http://solgenomics.net/organism/sol100/view>) describes the list of 100 species that have been included for sequencing, of which 25 are currently being sequenced, mostly belonging to the genera *Solanum* and *Nicotiana*. This list includes several medicinally important solenoid such as *Withania somnifera* and *Physalis peruviana*.

Medicinal Plant Consortium (MPC) and the Medicinal Plant/Human Health Consortium (MP/HHC) are two separate but similar projects that were funded by grants totaling about \$9 million, awarded by the National Institute of General Medical Sciences at NIH and made possible by the American Recovery and Reinvestment Act A research consortium. Both started in 2009 and focused on 'Transcriptome characterization sequencing and assembly of medicinal plants relevant to human health' [Link: <http://www.uic.edu/pharmacy/MedPITranscriptome/>]. A collaborative network of several universities and research institutions spans across the globe for MPC, led by Washington State University (WSU). These two consortia are focused on medicinal plants such as ginseng (*Panax spp.*), andrographis (*Andrographis paniculata*), marijuana (*Cannabis sativa*), hoodia (*Hoodia gordonii*), ginkgo (*Ginkgo biloba*), foxglove (*Digitalis purpurea*), *C. roseus*, opium poppy (*Papaver somniferum*), Himalayan mayapple (*Podophyllum hexandrum*), and camptotheca (*Camptotheca acuminata*). The Medicinal Plant Metabolomics Resource website (<http://medicinalplantgenomics.msu.edu/> & http://metnetdb.org/mpmr_public/) is an enormously resourceful link to record the current status of medicinal plant genomics. For

the major 14 enlisted medicinal plants on MPC, i.e., [*Atropa belladonna*, *Camptotheca accuminata*, *Cannabis sativa*, *Catharanthus roseus*, *Digitalis purpurea*, *Dioscorea villosa*, *Echinaceae purpurea*, *Ginkgo biloba*, *Hoodia gordonii*, *Hypericum perforatum*, *Panax quinquefolius*, *Rauwolfia serpentine*, *Rosmarinus officinalis*, *Valeriana officinalis*], their common names, taxonomic information, drug uses, chemical structures of major chemical principles, results from chemical trials, major metabolites, enzymes involved in their biosynthesis, entries in GenBank, assembly information, Pubmed literature records, and external links to Wikipedia are provided in a very comprehensive manner. Similarly, the transcriptomes of all the 14 enlisted medicinal plants are freely downloadable. More recently, a co-analysis of metabolomics and transcriptomics data for several species have become publicly available for medicinal plant comparative 'omics' data-mining.

Furthermore, the PhytoMetaSyn (www.phytometacyn.ca) project investigates metabolic pathways of selected important medicinal plants which yield high value metabolites. The final aim is to reconstruct these pathways and model heterologous systems such as the yeast cells to make the commercial scale production of such compounds feasible. Thus, it is an enormous public resource of metabo-genomic information for about 75 plants that are the source of high-value natural products. The annotation results for the transcriptomes of the 20 non-model plant species generated using 454 GS-FLX Titanium and Illumina GA sequencing platforms was provided by a follow-up publication from Xiao *et al.*, 2013. Although, Xiao *et al.*, 2013 have started to investigate specialized metabolite biosynthesis in non-model plants to be able to establish a data-mining framework applying next-generation sequencing and computational prowess for 75 medicinal plants belonging to 31 plant families, they reported the results obtained from the first twenty species as a proof of concept. These

species included *Arnica montana*, *Centella asiatica*, *Catharanthus roseus*, *Glaucium flavum*, *Hypericum perforatum*, *Lippia dulcis*, *Papaver bracteatum*, *Sanguinaria canadensis*, *Silene vulgaris*, *Tabernamontana elegans*, *Valeriana officinalis*, and *Xanthium strumarium*. The transcriptomes were obtained from diverse tissues such as leaf, roots, stem, rhizome, glands, and flowers. These transcriptome projects were centered on six sub-groups (sesquiterpenes, diterpenes, triterpenes, monoterpenoid indole alkaloids, benzyloquinoline alkaloids, and polyketides) of natural products belonging to three general categories: terpenoids (10 plants), alkaloids (9 plants), and polyketides. The authors clearly maintain that the selection of plant species was based on (a) availability of tissues, (b) commercial weight of key metabolites, (c) extent of understanding biochemical diversity across species and (d) availability of biochemical resources for specialized metabolites (Xiao *et al.*, 2013). A BLAST portal is publicly available for searching against desirable medicinal plant species for further downstream analyses of gene families and functional genomics studies.

Practical challenges and inherent bottlenecks

Needless to say, there have been transcriptomics investigations that have provided more generalized insights, without actually focusing on the specialized metabolism or carrying out detailed investigations into the biology of medicinal plants. Regardless of the constraints presented towards sequencing of medicinal plants' genomes, such as large and polyploidy genomes, the progress has been inconsistent and a lot remains to be addressed and done. The Table-1 also reiterates equivocally the mechanistic approaches into transcriptomics investigations for many instances, without any concrete biological insights as well as limited attempts into WGS studies. For example, there are no studies underscoring the effects of domestication on phytomedicinal production by the species of interest, the results of genome duplication events on improvement of

metabolically active gene families, reflections in complexity of metabolic networks, the relationship between phenotype to the genotypes, or a correlation of metabolic uniqueness with genomic specialty features of particular medicinal plants. There have been limited attempts in sequencing more and more valuable medicinal plant genomes, owing to reasons that are probably unknown.

The challenges posed by the plant genome sequencing efforts are two-fold. First, the inherent features of plant genomes themselves present big challenges. For example, the genome size, higher levels of ploidy, the complex nature of the genome, higher repetitive features (~80%), transposable elements, heterozygosity, expansive gene families, and huge arrays of non-coding RNAs all add to the complexity of these genomes. Additionally, problems arise using the genome sequencing technologies; things like DNA contamination, sample preparation, sequencing errors, or difficulties in *de novo* assembly (Carlos *et al.*, 2012). Apart from these, some challenges associated with medicinal plant genome sequencing are (a) higher contents of secondary metabolites which impede good quality DNA isolations, (b) nonavailability of genomic resources from the taxa under study for annotation and assembly, (c) uniqueness of transcripts that may not have known orthologues in other sequenced and model plant genomes, and (d) unique gene families and uncharacterized genetic elements. Nevertheless, the need for the expansion of sequencing efforts beyond the commercially important crops to include other non-commodity and non-model species has been recently realized (Carlos *et al.*, 2012).

We are still far away from push-button sequencing and assembly of plant genomes into completely finished genomes at lower costs (Schatz *et al.*, 2012). Approximately two-thirds of the 50,000 different phytomedicinals in use are collected from the wild and 4000- 10,000 of them are possibly endangered (Edwards, 2004). It is important to note that an insignificant number of reports exist on molecular marker-

based approaches towards medicinal plant improvement (Canter *et al.*, 2005).

Thus, large-scale bioprospection approaches in phytomedicinal genomics and transcriptomics is still lacking when it comes to exploration of novel plant species.

Concluding remarks and future prospects

The functional genomics approach aided by targeted-metabolite and transcriptome profiling approaches would lead to an efficient way of deciphering novel gene functions involved in specialized metabolic pathways in phytomedicinals. The development of next-generation sequencing, coupled with the advancement of computational methods, has allowed researchers to access the transcriptomes of recalcitrant genomes, such as those of medicinal plant species (Góngora-Castillo *et al.*, 2012). The transcriptomics study concerning medicinal plants has become the most active area in medicinal plant genome research. Hence, the application status of transcriptomics in medicinal plants based on emergence, development, and methodology of transcriptomics have been extensively exploited. The fact that they belong to unique taxonomic families, have limited genomic resources, and are labeled as 'non-model', has hurt the chances of their genomes being sequenced to a greater effect. However, transcriptome sequencing has enjoyed major success. These endeavors in genomics and transcriptomics would enable a robust understanding of metabolic pathways that yield important phytomedicinals and facilitate highly valued metabolic engineering approaches for some rare medicinals.

The optimism is that the '-omics' revolution that is spreading from model species to those with more complex genomes (i.e., 'muddle' species) will influence research and exploitation of medicinal species in a significant manner (Canter *et al.*, 2005).

Sl. No	Plant Species	Transcriptome	Interesting and Major Findings	References
Genomes				
1	Neem, <i>Azadirachta indica</i> L.	Genome; Root, leaf, stem and flower transcriptomes	364 Mbp genome, 20,000 genes ; exclusive & enhanced expression of genes involved in terpenoid biosynthesis pathways compared to other sequenced angiosperms	Krishnan <i>et al.</i> , 2012
2	Hemp, <i>Cannabis sativa</i> L.	Genome; Flower transcriptome	820Mbp, ~30,000 genes, copy numbers of transcripts for cannabinoid pathway enzymes http://www.medicinalgenomics.com/	Bakel <i>et al.</i> , 2011
3	Dwarf birch, <i>Betula nana</i>	Genome	450 Mbp, assembled restriction-site-associated DNA (RAD) markers	Wang <i>et al.</i> , 2013
4	Common Ash, <i>Fraxinus excelsior</i> L.	Genome	880 Mbp, underway; British Ash Tree Genome Project: http://www.ashgenome.org/	Eurofins MWG Operon and University of London
Transcriptomes				
1	American ginseng, <i>Panax quinquefolius</i> L.	Root	MeJA induced transcriptome studied, 150 CYP450s and 235 glycosyltransferases (GTs), transcripts of ginsenoside biosynthesis	Sun <i>et al.</i> , 2010
2	Konjac, <i>Amorphophallus konjac</i> and Voodoo lily, <i>Amorphophallus bulbifer</i>	-	10,754 SSR markers identified in <i>Amorphophallus</i> using transcriptome, 177 polymorphic markers validated in 25 individuals	Zheng <i>et al.</i> , 2013
3	Ashwagandha, <i>Withania somnifera</i>	Leaf and root	Differential gene expression of CYP450s, GTs and methyltransferase gene families in leaf and root indicating tissue-specific biosynthesis of withanolides	Gupta <i>et al.</i> , 2013
4	Black pepper, <i>Piper nigrum</i> L.	Roots	Predicted proteome (4472) with 52% homology with <i>Arabidopsis</i> .	Gordo <i>et al.</i> , 2012
5	Chai Hu, <i>Bupleurum chinense</i> DC.	Roots	ESTs for saikosaponin biosynthesis transcripts encode enzymes that catalyze the formation of the saikosaponin backbone; 246 CYP450s and 102 GTs	Sui <i>et al.</i> , 2011
6	Cape gooseberry, <i>Physalis peruviana</i>	Leaf, root, and phyto-pathosystems	10,000 molecular markers (SSRs and SNPs) from EST assembly	Meneses <i>et al.</i> 2013
7	Cape gooseberry, <i>Physalis peruviana</i>	Leaf	9,436 transcripts with multiple-exon models and conserved intron positions; 5,971 SSR markers from assembled ESTs	Garzón-Martínez <i>et al.</i> , 2012
8	Asiatic pennywort, <i>Centella asiatica</i> (L.) Urban	Leaf	General conclusions	Sangwan <i>et al.</i> , 2013
9	Safed Musli, <i>Chlorophytum borivilium</i>	Leaf	Transcripts represent all enzymes involved in saponin and flavonoid biosynthesis, few represented alkaloid biosynthesis, and plant	Kalra <i>et al.</i> , 2013

			defense response	
10	Edible chrysanthemum, <i>Chrysanthemum nankingense</i> (Nakai) Tzvel	Stems and leaves	General conclusions	Wang <i>et al.</i> , 2013
11	Pink Rock-Rose, <i>Cistus creticus</i> subsp. <i>creticus</i>	Trichomes from young leaves	8% sequences were secondary metabolism-related and involved mostly flavonoid and terpenoid biosynthesis; 38 % of the ESTs showed no significant similarity to DNA in public databases	Falara <i>et al.</i> , 2008
12	Common foxglove, <i>Digitalis purpurea</i> L.	Leaves, stems, flowers and roots	140 unigenes involved in cardiac glycoside biosynthesis, grouped into 30 families, of which 29 were identified for the first time.	Wu <i>et al.</i> , 2012
13	Berberine vine, <i>Coscinium fenestratum</i> (Gaertn.) Colebr.	Stem	Benzoisoquinoline alkaloid biosynthesis pathway transcripts identified.	Parthasarathy and Shrishya, 2013
14	T'ieh-p'i Shih-hu, <i>Dendrobium officinale</i> Kimura et Migo	Stem	69 unique sequences representing 25 genes involved in alkaloid backbone biosynthesis, 1061 SSR markers detected from 36,407 unigenes	Guo <i>et al.</i> , 2013
15	Barrenwort, <i>Epimedium sagittatum</i> (Sieb. Et Zucc.) Maxim	Leaf	Flavonoid biosynthesis pathway genes identified, SSRs recorded	Zeng <i>et al.</i> , 2010
16	Lang-Du, <i>Euphorbia fischeriana</i>	Root	24 and 9 candidate transcripts involved in the terpenoid and diterpenoid biosynthesis pathways, respectively	Barrero <i>et al.</i> , 2011
17	Fenugreek, <i>Trigonella foenum-graecum</i> L.	Seedlings	6775 transcripts mapped to plant biochemical pathways including the diosgenin biosynthesis pathway, 11 key enzymes identified that control the pathway route to diosgenin biosynthesis	Vaidya <i>et al.</i> , 2012
18	Jiaogulan, <i>Gynostemma pentaphyllum</i> (Thunb.)	Roots and leaves	Transcripts identified that are involved in biosynthesis of mono-, di-, tri- and sesquiterpenoids and gynosaponins (gypenosides)	Subramaniyam <i>et al.</i> , 2011
19	Happy Tree, <i>Camptotheca acuminata</i>	Young leaves	521 ESTs representing 20 enzymes genes involved in the backbone synthesis of the camptothecin biosynthetic pathway	Sun <i>et al.</i> , 2011
20	Himalayan Mayapple, <i>Podophyllum hexandrum</i> Royle	Cell Culture	Schematic model pathways for phenylpropanoid and podophyllotoxin biosynthesis developed	Bhattacharyya <i>et al.</i> , 2013
21	Firmoss, <i>Huperzia serrata</i> and Keeled Tassel-fern, <i>Phlegmariurus carinatus</i>	Roots, stems and leaves	115 <i>H. serrata</i> and 98 <i>P. carinatus</i> unique putative transcripts associated with the biosynthesis of triterpenoids, lycopodium alkaloids and flavones/flavonoids	Luo <i>et al.</i> , 2010

22	Insulin plant, <i>Costus pictus</i> D. Don	Leaf	Transcripts involved in bixin, geraniol and geranial biosynthesis identified, validation of a putative norbixin methyltransferase, a precursor of bixin biosynthesis	Annadurai <i>et al.</i> , 2012
23	Woad, <i>Isatis indigotica</i>	Flowers, leaves, stems, and roots	104 unigenes encoding enzymes of biosynthetic pathways of indole, terpenoid, and phenylpropanoid identified	Chen <i>et al.</i> , 2013
24	Japanese honeysuckle, <i>Lonicera japonica</i> Thunb.	Flower corollas, petals	Expression profiling of transcripts involved in biosynthesis of phenolic acids, flavonoids, terpenoids, and fatty acids	Yuan <i>et al.</i> , 2012
25	Japanese honeysuckle, <i>Lonicera japonica</i> Thunb.	Buds and leaves	Transcripts involved in the biosynthesis of chlorogenic acid and luteoloside	He <i>et al.</i> , 2013
26	<i>Macleaya cordata</i> and <i>Macleaya microcarpa</i> ,	Roots, leaves and fruits at different developmental phases of the plant	Differential alkaloid biosynthesis, storage and transport in <i>M. cordata</i> and <i>M. microcarpa</i> before flowering and after fruiting for roots and leaves noted	Zeng <i>et al.</i> , 2013
27	Mountain pepper, <i>Litsea cubeba</i> (Lour.) Pers	Flower buds, full open flowers, young leaves, leaf buds and fruits	285 unigenes involved in terpenoid pathways, including 14 terpene synthase genes	Han <i>et al.</i> , 2013
28	Opium poppy, <i>Papaver somniferum</i>	Cell cultures	Most sanguinarine alkaloid biosynthetic enzymes identified, complemented with proteomics	Desgagné-Penix <i>et al.</i> , 2010
29	Ginseng, <i>Panax ginseng</i> C.A. Meyer	Roots	9 genes involved in biosynthesis of ginsenoside skeletons, including 133 CYP450s and 235 GTs	Chen <i>et al.</i> , 2011
30	Ginseng, <i>Panax notoginseng</i> (Burk) F.H. Chen	Root	41 unique sequences representing 11 genes involved in triterpene saponin backbone biosynthesis, higher expression of dammarenediol synthase (DS)-the first committed enzyme in the biosynthetic pathway of major triterpene saponins; 174 CYP450s and 242 GTs	Luo <i>et al.</i> , 2011
31	Love apple, <i>Paris polyphylla</i> var. <i>yunnanensis</i>	Embryo developmental stages	11 phytohormone-related genes and 5 other genes show differential expression patterns in the embryo and endosperm during seed stratification	Qi <i>et al.</i> , 2013
32	Kutki, <i>Picrorhiza kurrooa</i>	Leaf	Differential expression of CYP450s, GTs and transcripts of picroside biosynthesis in response to 2 temperature regimes	Gahlan <i>et al.</i> , 2012
33	Himalayan mayapple,	Rhizome	Regio-specific methylenedioxy bridge-forming CYP450s identified that	Marques <i>et al.</i> , 2012

	<i>Podophyllum hexandrum</i> and Mayapple, <i>Podophyllum peltatum</i>		catalyze pluviatolide formation	
34	Japanese Knotweed, <i>Polygonum cuspidatum</i>	Root	Anthraquinone and resveratrol biosynthesis pathway genes, 18 UDP-GTs identified as candidates in glucoside biosynthesis	Hao <i>et al.</i> , 2012
35	Radish, <i>Raphanus sativus</i> L.	Tuberous roots at early and late seedling stage	Starch, sucrose, phytohormone and phenylpropanoid biosynthesis pathway transcripts dominant in tuberous root development	Wang <i>et al.</i> , 2012
36	Safflower, <i>Carthamus tinctorius</i> L.	Tubular flower tissue	Conserved flavonoid and unsaturated fatty acids pathway genes	Lulin <i>et al.</i> , 2012
37	Red sage, <i>Salvia miltiorrhiza</i> Bunge	Root	27 unigenes (encoding 15 enzymes) involved in tanshinones biosynthesis, and 29 unigenes (encoding 11 enzymes) involved in phenolic acids biosynthesis.	Li <i>et al.</i> , 2010
38	Sea buckthorn, <i>Hippophae rhamnoides</i> L.	Mature seeds	Primary metabolism (protein>nucleic acid>carbohydrate>lipid), fatty acid and lipid biosynthetic pathways were highly represented	Fatima <i>et al.</i> , 2012
39	St. John's wort, <i>Hypericum perforatum</i> L.	Roots, stems, leaves, and flowers	260 unigenes involved in the production of hypericin, hyperforin, and melatonin; another 2,291 unigenes classified as potential Type III polyketide synthases	He <i>et al.</i> , 2012
40	Sweetleaf, <i>Stevia rebaudiana</i>	Diverse tissue types	5,876 SSRs identified in 4,879 transcripts	Vallejo <i>et al.</i> , 2014
41	Sweet Wormwood, <i>Artemisia annua</i> L.	Glandular and filamentous trichomes	MEP, MVA, and lipid biosynthesis pathways up regulated in glandular trichomes; significant differential expression between apical and sub-apical cells of glandular trichomes.	Soetaert <i>et al.</i> , 2013
42	Japanese yew, <i>Taxus cuspidata</i>	Needles	753 simple sequence repeat motifs, putative genes of Taxol biosynthesis	Wu <i>et al.</i> , 2011
43	Chinese yew, <i>Taxus mairei</i>	Leaf, stem, and root	Higher expression of taxane biosynthetic genes in roots than in leaf and the stem, also supported by metabolomics datasets	Hao <i>et al.</i> , 2011
44	Tea Plant, <i>Camellia sinensis</i>	Flowers at the big bud stage [petals, pistils and stamens]	SSR-based linkage map was constructed that covered 1,156.9 cM with 237 SSR markers distributed in 15 linkage groups	Tan <i>et al.</i> , 2013
45	Valerian, <i>Valeriana officinalis</i>	Callus, young leaves, mature leaves, lateral roots, and stem/petioles	Valerena-1,10-diene synthase identified from the transcriptome	Yeo <i>et al.</i> , 2013
46	Wax Gourd, <i>Benicasa hispida</i>	Shoot tips, leaves, flowers, fruits and stems	6, 242 SSRs detected as potential molecular markers	Biao <i>et al.</i> , 2013

Table 1. Over-view of the current progress and advances in genomics and transcriptomics of medicinal plant resources.

Thus, functional and comparative genomics approaches would enable drug discovery, drug development, and large-scale production of phytomedicinals in the coming future. In combination with widely targeted and untargeted metabolomics, proteomics and transcriptome and genome-based approaches are likely to provide the essential 'healing' touch to the systems biology understanding of phytomedicinal species for researchers in conservation, phytopharmaceuticals, the drug industry, systematic study, and the forensic and herbal industry.

Acknowledgements

The author acknowledges a Postdoctoral Research Associate Fellowship being availed at University of Florida and the facilities providing training and literature resources for compilation of this review, and a supportive Principal Investigator. The author also thanks the anonymous reviewers for their critical insights in improvising the manuscript to its current form.

References

1. Afendi, F.M., Okada, T., Yamazaki, M., *et al.* KNApSACk family databases: integrated metabolite-plant species databases for multifaceted plant research. *Plant and Cell Physiology* 53, (2012): e1.
2. Ângelo, P.C.S., Nunes-Silva C.G., Brígido, M.M. *et al.* Guarana (*Paullinia cupana* var. *sorbilis*), an anciently consumed stimulant from the Amazon rain forest: the seeded-fruit transcriptome. *Plant Cell Reports* 27, (2008): 117.
3. Annadurai, R.S., Jayakumar, V., Mugasimangalam, R.C., *et al.* Next generation sequencing and *de novo* transcriptome analysis of *Costus pictus* D. Don, a non-model plant with potent anti-diabetic properties. *BMC Genomics* 13, (2012): 663.
4. Bajgain, P., Richardson, B., Price, J. *et al.* Transcriptome characterization and polymorphism detection between subspecies of big sagebrush (*Artemisia tridentata*). *BMC Genomics* 12, (2011): 370.
5. Barrero M. and Stella, L. Advances on genetics and genomics in Cape gooseberry (*Physalis peruviana*). In *Plant and Animal Genome XXI Conference*. Plant and Animal Genome, 2013.
6. Barrero, R.A., Chapman, B., Yang, Y., *et al.* *De novo* assembly of *Euphorbia fischeriana* root transcriptome identifies prostratin pathway related genes. *BMC Genomics* 12, (2011): 600.
7. Bhattacharyya, D., Sinha, R., Hazra, S. *et al.* *De novo* transcriptome analysis using 454 pyrosequencing of the Himalayan Mayapple, *Podophyllum hexandrum*. *BMC Genomics* 14, (2013): 748.
8. Canter, P. H., Thomas, H. and Ernst, E. Bringing medicinal plants into cultivation: opportunities and challenges for biotechnology. *Trends in Biotechnology* 23, (2005): 180.
9. Chen, J., Dong, X., Li, Q. *et al.* Biosynthesis of the active compounds of *Isatis indigotica* based on transcriptome sequencing and metabolites profiling. *BMC Genomics* 14, (2013): 857
10. Chen, S., Luo, H., Li Y. *et al.* 454 EST analysis detects genes putatively involved in ginsenoside biosynthesis in *Panax ginseng*. *Plant Cell Reports* 30, (2011): 1593.
11. Chen, S., Yao, H., Han, J. *et al.* Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS One* 5, (2010): e8613.
12. Chen, S., Xiang L., Guo, X. *et al.* An introduction to the medicinal plant genome project. *Frontiers of Medicine* 5, (2011): 178.
13. Claros, M.G., Bautista, R., Guerrero-Fernández, D. *et al.* Why assembling plant genome sequences is so challenging. *Biology* 1, (2012): 439-459.
14. Desgagné-Penix, I., Khan, M., Schriemer, D. *et al.* Integration of deep transcriptome and proteome analyses reveals the components of alkaloid metabolism in opium poppy cell cultures. *BMC Plant Biology* 10, (2010): 252.
15. Edwards, R. No remedy in sight for herbal ransack. *New Scientist* 181, no. 2429 (2004): 10.
16. Facchini, P.J., Bohlmann, J., Covello, P.S. *et al.* Synthetic biosystems for the production of high-

- value plant metabolites. *Trends Biotechnology* 30 (2012):127.
- 17.Falara, V., Fotopoulos, V., Margaritis, T. *et al.* Transcriptome analysis approaches for the isolation of trichome-specific genes from the medicinal plant *Cistus creticus* subsp. *creticus*. *Plant Molecular Biology* 68, (2008): 633.
- 18.Fatima, T., Snyder, C.L., Schroeder, W.R. *et al.* Fatty acid composition of developing sea buckthorn (*Hippophae rhamnoides* L.) berry and the transcriptome of the mature seed. *PLoS One* 7, (2012): e34099.
- 19.Gahlan, P., Singh, H.R., Shankar, R. *et al.* *De novo* sequencing and characterization of *Picrorhiza kurrooa* transcriptome at two temperatures showed major transcriptome adjustments. *BMC Genomics* 12, (2011): 636.
- 20.Garzón-Martínez, G. A., Zhu, Z.I., Landsman, D. *et al.* The *Physalis peruviana* leaf transcriptome: assembly, annotation and gene model prediction. *BMC Genomics* 13, (2012): 151.
- 21.Ge, G.B., Xiao, P.G., Zhang, Y.Y. and Yang, L. The first insight into the tissue specific *Taxus* transcriptome via Illumina second generation sequencing. *PLoS One* 6, (2011): e21220.
- 22.Ghangal, R., Chaudhary, S., Jain, M. *et al.* Optimization of *de novo* short read assembly of seabuckthorn (*Hippophae rhamnoides* L.) transcriptome. *PLoS One* 8, (2013): e72516.
23. Góngora-Castillo, E., Childs, K.L., Fedewa, G. *et al.* Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. *PLoS One* 7, (2012): e52506.
- 23.Góngora-Castillo, E., Fedewa, G., Yeo, Y., Chappell, J. *et al.* Genomic approaches for interrogating the biochemistry of medicinal plant species. *Methods in Enzymology* 517, (2011): 139.
- 24.Goodacre, R., Vaidyanathan, S., Dunn, W.B. *et al.* Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnology* 22, (2004): 245.
- 25.Gordo, S., Pinheiro, D., Moreira, E. *et al.* High-throughput sequencing of black pepper root transcriptome. *BMC Plant Biology* 12, (2012): 168.
- 26.Guo, X., Li, Y., Li, C. *et al.* Analysis of the *Dendrobium officinale* transcriptome reveals putative alkaloid biosynthetic genes and genetic markers. *Gene* (2013).
- 27.Gupta, P., Goel, R., Pathak, S. *et al.* *De novo* assembly, functional annotation and comparative analysis of *Withania somnifera* leaf and root transcriptomes to identify putative genes involved in the withanolides biosynthesis. *PLoS One* 8, (2013): e62714.
- 28.Han, X.J., Wang, Y.D., Chen, Y.C. *et al.* Transcriptome sequencing and expression analysis of terpenoid biosynthesis genes in *Litsea cubeba*. *PLoS One* 8, (2013): e76890.
- 29.Hao, D.C., Ma, P., Mu, J. *et al.* *De novo* characterization of the root transcriptome of a traditional Chinese medicinal plant *Polygonum cuspidatum*. *Science China Life Sciences* 55, (2012): 452.
- 30.Hao, D.C., Chen, S.L., Xiao, P.G. *et al.* Application of high-throughput sequencing in medicinal plant transcriptome studies. *Drug Development Research* 73, (2012): 487.
- 31.He, L., Xu, X., Li, Y. *et al.* Transcriptome analysis of buds and leaves using 454 pyrosequencing to discover genes associated with the biosynthesis of active ingredients in *Lonicera japonica* Thunb. *PLoS One* 8, (2013): e62922.
- 32.He, M., Wang, Y., Hua, W. *et al.* *De novo* sequencing of *Hypericum perforatum* transcriptome to identify potential genes involved in the biosynthesis of active metabolites. *PLoS One* 7, (2012): e42081.
- 33.Jiang, B., Xie, D., Liu, W. *et al.* *De novo* assembly and characterization of the transcriptome, and development of SSR markers in wax gourd (*Benicasa hispida*). *PLoS One* 8, (2013).
- 34.Kalra, S., Puniya, B.L., Kulshreshtha, D. *et al.* *De novo* transcriptome sequencing reveals important molecular networks and metabolic pathways of the plant, *Chlorophytum borivillianum*. *PLoS One* 8, (2013): e83336.
- 35.Krishnan, N.M., Pattnaik, S., Deepak, S.A. *et al.* *De novo* sequencing and assembly of *Azadirachta indica* fruit transcriptome. *Current Science* 101, (2011): 1553.

36. Krishnan, N., Pattnaik, S., Jain, P. *et al.* A draft of the genome and four transcriptomes of a medicinal and pesticidal angiosperm *Azadirachta indica*. *BMC Genomics* 13, (2012): 464.
37. Li, Y., Sun, C., Luo, H.M. *et al.* Transcriptome characterization for *Salvia miltiorrhiza* using 454 GS FLX. *Yao xue xue bao= Acta pharmaceutica Sinica* 45, (2010): 524.
38. Liu, Y., Huo, N., Dong, L. *et al.* Complete chloroplast genome sequences of Mongolia medicine *Artemisia frigida* and phylogenetic relationships with other plants. *PLoS One* 8, (2013): e57533.
39. Lou, S.K., Wong, K.L., Li, M. *et al.* An integrated web medicinal materials DNA database: MMDBD (Medicinal Materials DNA Barcode Database). *BMC Genomics* 11, (2010): 402.
40. Lulin, H., Xiao, Y., Pei, S. *et al.* The first Illumina-based de novo transcriptome sequencing and analysis of safflower flowers. *PLoS One* 7, (2012): e38653.
41. Luo, H., Sun, C., Sun, Y. *et al.* Analysis of the transcriptome of *Panax notoginseng* root uncovers putative triterpene saponin-biosynthetic genes and genetic markers. *BMC Genomics* 12, (2011): S5.
42. Luo, H., Li, Y., Sun, C. *et al.* Comparison of 454-ESTs from *Huperzia serrata* and *Phlegmariurus carinatus* reveals putative genes involved in lycopodium alkaloid biosynthesis and developmental regulation. *BMC Plant Biology* 10, (2010): 209.
43. Ma, J., Yang, B., Zhu, W. *et al.* The complete chloroplast genome sequence of *Mahonia bealei* (Berberidaceae) reveals a significant expansion of the inverted repeat and phylogenetic relationship with other angiosperms. *Gene* 528, (2013): 120.
44. Marques, J.V., Kim, K.W., Lee, C. *et al.* Next generation sequencing in predicting gene function in podophyllotoxin biosynthesis. *Journal of Biological Chemistry* 288, (2013): 466.
45. Muranaka, T., and Saito, K. Phytochemical genomics on the way. *Plant and Cell Physiology* 54, (2013): 645.
46. Parthasarathy, S. RNA-seq analysis of *Coscinium fenestratum* (Gaertn.) Colebr a critically endangered medicinal plant from Western Ghats of India. In *Plant and Animal Genome XXII Conference*. Plant and Animal Genome, 2014.
47. Pyle, B.W., Tran, H.T., Pickel, B. *et al.* Enzymatic synthesis of valerenenolide, a diene by a unique sesquiterpene synthase from the valerian plant (*Valeriana officinalis*). *FEBS Journal* 279, (2012): 3136.
48. Qi, J., Zheng, N., Zhang, B. *et al.* Mining genes involved in the stratification of *Paris polyphylla* seeds using high-throughput embryo transcriptome sequencing. *BMC Genomics* 14, (2013): 358.
49. Qian, J., Song, J., Gao, H., *et al.* The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS One* 8, (2013): e57607.
50. Ramilowski, J.A., Sawai, S., Seki, H. *et al.* *Glycyrrhiza uralensis* transcriptome landscape and study of phytochemicals. *Plant and Cell Physiology* 54, (2013): 697.
51. Sangwan, R.S., Tripathi, S., Singh, J. *et al.* De novo sequencing and assembly of *Centella asiatica* leaf transcriptome for mapping of structural, functional and regulatory genes with special reference to secondary metabolism. *Gene* (2013).
52. Schatz, M.C., Witkowski, J. and McCombie, W.R. Current challenges in de novo plant genome sequencing and assembly. *Genome Biology* 13, (2012): 243.
53. Sharma, P.C., Jain, A. and Chaudhary, S. Transcriptome Analysis in Seabuckthorn (*Hippophae rhamnoides* L.), a medicinally important plant. *International Conference on Environmental and Biological Sciences Conference*, (2012) Bangkok, Thailand
54. Shukla, A.K., Shasany, A.K., Gupta, M.M. *et al.* Transcriptome analysis in *Catharanthus roseus* leaves and roots for comparative terpenoid indole alkaloid profiles. *Journal of Experimental Botany* 57, (2006): 3921.
55. Soetaert, S., Van Neste, C.M.F., Vandewoestyne, M.L., *et al.* Differential transcriptome analysis of glandular and filamentous trichomes in *Artemisia annua*. *BMC Plant Biology* 13, (2013): 220.

56. Subramaniyam, S., Mathiyalagan, R. Gyo, I.J. *et al.* Transcriptome profiling and *in silico* analysis of *Gynostemma pentaphyllum* using a next generation sequencer. *Plant Cell Reports* 30, (2011): 2075.
57. Sucher, N. J., Hennell, J.R., and Carles, M.C. Genomic and transcriptomic profiling: Tools for the quality production of plant-based medicines. In *Biotechnology for Medicinal Plants*, pp. 439-455. Springer Berlin Heidelberg, 2013.
58. Sui, C., Zhang, J., Wei, J. *et al.* Transcriptome analysis of *Bupleurum chinense* focusing on genes involved in the biosynthesis of saikosaponins. *BMC Genomics* 12, (2011): 539.
59. Sumner, L.W., Mendes, P. and Dixon, R.A. Plant metabolomics: Large-scale phytochemistry in the functional genomics era. *Phytochemistry* 62, (2003): 817.
60. Sun, C., Li, Y., Wu, Q. *et al.* *De novo* sequencing and analysis of the American ginseng root transcriptome using a GS FLX Titanium platform to discover putative genes involved in ginsenoside biosynthesis. *BMC Genomics* 11, (2010): 262.
61. Sun, P., Song, S., Zhou, L. *et al.* Transcriptome analysis reveals putative genes involved in iridoid biosynthesis in *Rehmannia glutinosa*. *International Journal of Molecular Sciences* 13, (2012): 13748.
62. Sun, Y., Luo, H., Li, Y. *et al.* Pyrosequencing of the *Camptotheca acuminata* transcriptome reveals putative genes involved in camptothecin biosynthesis and transport. *BMC Genomics* 12, (2011): 533.
63. Tan, L.Q., Wang, L.Y., Wei, K. *et al.* Floral transcriptome sequencing for SSR marker development and linkage map construction in the tea plant (*Camellia sinensis*). *PLoS One* 8, (2013): e81611.
64. Vaidya, K., Ghosh, A., Kumar, V. *et al.* *De novo* transcriptome sequencing in *Trigonella foenum-graecum* L. to identify genes involved in the biosynthesis of diosgenin. *The Plant Genome* 6, (2013).
65. Vallejo, V. Development of *De novo* reference transcriptome assembly and large-scale identification of simple sequence repeat markers in *Stevia rebaudiana*. In *Plant and Animal Genome XXII Conference*. Plant and Animal Genome, 2014.
66. van Bakel, H., Stout, J.M., Cote, A.G. *et al.* The draft genome and transcriptome of *Cannabis sativa*. *Genome Biology* 12, (2011): R102.
67. van Moerkercke, A., Fabris, M., Pollier, J. *et al.* CathaCyc, a metabolic pathway database built from *Catharanthus roseus* RNA-Seq data. *Plant and Cell Physiology* 54, (2013): 673.
68. Vines, G. Herbal harvests with a future: towards sustainable sources for medicinal plants. *Plantlife International*, 2004.
69. Wang, N., Thomson, M., Bodles, W. *et al.* Genome sequence of dwarf birch (*Betula nana*) and cross-species RAD markers. *Molecular Ecology* 22, (2013): 3098.
70. Wang, H., Jiang, J., Chen, S. *et al.* Next-generation sequencing of the *Chrysanthemum nankingense* (Asteraceae) transcriptome permits large-scale unigene assembly and SSR marker discovery. *PLoS One* 8, (2013): e62293.
71. Wang, S., Wang, X., He, Q. *et al.* Transcriptome analysis of the roots at early and late seedling stages using Illumina paired-end sequencing and development of EST-SSR markers in radish. *Plant Cell Reports* 31, (2012): 1437.
72. Wu, B., Li, Y., Yan, H. *et al.* Comprehensive transcriptome analysis reveals novel genes involved in cardiac glycoside biosynthesis and miRNAs associated with secondary metabolism and stress response in *Digitalis purpurea*. *BMC Genomics* 13, (2012): 15.
73. Wu, Q., Sun, C., Luo, H. *et al.* Transcriptome analysis of *Taxus cuspidata* needles based on 454 pyrosequencing. *Planta Medica* 77, (2011): 394.
74. Wurtele, E.S., Chappell, J., Jones, A.D. *et al.* Medicinal plants: A public resource for metabolomics and hypothesis development. *Metabolites* 2, (2012): 1031.
75. Xiao, M., Zhang, Y., Chen, X. *et al.* Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J. Biotechnol.* 166, (2013): 122.
76. Yamazaki, M., Mochida, K., Asano, T. *et al.* Coupling deep transcriptome analysis with untargeted metabolic profiling in *Ophiorrhiza*

pumila to further the understanding of the biosynthesis of the anti-cancer alkaloid camptothecin and anthraquinones. *Plant and Cell Physiology* 54, (2013): 686.

77.Yeo, Y. S., Nybo, S.E., Chittiboyina, A.G. *et al.* Functional identification of valerena-1, 10-diene synthase, a terpene synthase catalyzing a unique chemical cascade in the biosynthesis of biologically active sesquiterpenes in *Valeriana officinalis*. *Journal of Biological Chemistry* 288, (2013): 3163.

78.Yuan, Y., Song, L., Li, M. *et al.* Genetic variation and metabolic pathway intricacy govern the active compound content and quality of the Chinese medicinal plant *Lonicera japonica* Thunb. *BMC Genomics* 13, (2012): 195.

79.Zeng, J., Liu, Y., Liu, W. *et al.* Integration of transcriptome, proteome and metabolism data reveals the alkaloids biosynthesis in *Macleaya cordata* and *Macleaya microcarpa*. *PloS One* 8, (2013): e53409.

80.Zeng, S., Xiao, G., Guo, J. *et al.* Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) Maxim. *BMC Genomics* 11, (2010): 94.

81.Zhang, H.P., Pan, J.B., Zhang, C. *et al.* Network understanding of herb medicine via rapid identification of ingredient-target interactions. *Scientific Reports* 4, (2014).

82.Zheng, M., Tian, J., Yang, G. *et al.* Transcriptome sequencing, annotation and expression analysis of *Nannochloropsis sp.* at different growth phases. *Gene* (2013).

83.Zheng, X., Pan, C., Diao, Y. *et al.* Development of microsatellite markers by transcriptome sequencing in two species of *Amorphophallus* (Araceae). *BMC Genomics* 14, (2013): 490.

84.Zhou, M., Liu, Y., Chen, T. *et al.* Components of the peptidome and transcriptome persist in lin wa pi: The dried skin of the Heilongjiang brown frog (*Rana amurensis*) as used in traditional Chinese medicine. *Peptides* 27, (2006): 2688.