

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <http://orca.cf.ac.uk/136691/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Chaudhary, Saurabh and Sharma, Prakash C. 2016. Next generation sequencing-based exploration of genomes and transcriptomes of medicinal plants. *Indian Journal of Plant Physiology* 21 (4) , 489–503. [10.1007/s40502-016-0258-8](https://doi.org/10.1007/s40502-016-0258-8) file

Publishers page: <http://dx.doi.org/10.1007/s40502-016-0258-8> <<http://dx.doi.org/10.1007/s40502-016-0258-8>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Next generation sequencing-based exploration of genomes and transcriptomes of medicinal plants

Saurabh Chaudhary,¹

Prakash C. Sharma,^{1,*}

Email prof.pcsharma@gmail.com

¹ University School of Biotechnology, Guru Gobind Singh Indraprastha University, Dwarka 16C, New Delhi, 110078 India

Abstract

Medicinal plants are known for their therapeutic potential and have been associated with human history for their use in traditional medicine systems in different countries. Recent advances in next generation sequencing (NGS) technologies have accelerated research on medicinal plants with reduced cost and efforts. NGS technologies not only provide opportunity for high throughput whole genome sequencing, they also facilitate direct RNA sequencing. The sequence data-sets generated can further be explored for application in various areas of research such as comparative genomics, data mining for small and long non-coding RNAs, mining of molecular markers, gene discovery, etc. Continuous efforts are being made by commercial sequencing service providers in improving technology to overcome bioinformatics challenges in NGS data analysis. In recent past, genome sequence projects on various medicinal plants have been successfully accomplished and few others are in pipeline. Similarly, enormous NGS-based transcriptome data have been generated in a number of medicinal plants. In the present review, we have briefly attempted to address advancement in NGS technology,

genomic and transcriptomics studies on medicinal plants with special emphasis on seabuckthorn (*Hippophae* sp.), a medicinally important plant of Indian Himalayas. Moreover, the scope of implementation of NGS based research on medicinal plants have been explored for the selection of candidate genes involved in particular biosynthesis pathways. The identified genes can be exploited for engineering medicinal plants for producing improved quality biologically active phytochemicals.

Keywords

Medicinal plants
Next generation sequencing
Genome
Transcriptome
Seabuckthorn

Introduction

Many plants are used for therapeutic, medicinal and nutritional purposes from ancient time in different countries worldwide. In recent past, research on medicinal plants has significantly grown because of acceptance of the use of phytomedicines over synthetic drugs for treatment of various diseases. Plants produce thousands of metabolites having diverse medicinal uses. Over 12,000 phytochemicals having biological activity have been identified and isolated from different plants and their parts. Further, this number represents roughly 10% of the projected useful compounds, suggesting the need of exploring more medicinal plants for their pharmacological benefits (Lai and Roy 2004). The use of medicinal plants to treat various diseases is predated human history and widely used in non-industrialized societies. However, the increased global consumption of phytomedicines to fight against various infectious and other diseases has rightly stimulated extensive research on medicinal plants aiming to facilitate production of improved quality phytochemicals and making it

important activity of pharmacological and pharmaceutical sector (Sumner et al. 2015). Moreover, research advancement in the molecular biology techniques specially sequencing technique has revolutionized research on medicinal plants. Studies have been conducted on medicinal plants at genome and transcriptome level allowing identification of key genes involved in the production of bioactive phytochemicals. The emergence of next generation sequencing (NGS) techniques led to the deep understanding of the biosynthesis pathways of biologically active phytochemicals after availability of enormous genomic/transcriptomics sequence data (Sharma and Shrivastava 2016). Comparative genomics, expression analysis, homology search, functional annotation, and phylogenetic analysis further facilitate the identification of genes involved in the production of secondary metabolites and biochemical pathways in medicinal plants.

The advent of NGS has driven the genomic/transcriptomics based inquiries in biology where high throughput sequence reads are generated with less labor and low cost. NGS not only facilitates the genome sequencing, but recent advancements have led to the developments allowing genome resequencing, transcriptome sequencing (RNA-Seq), Chip sequencing (DNA–protein interactions analysis), and epigenome analysis (de Magalhaes et al. 2010). Various NGS technologies include single molecule real-time sequencing (Pacific Biosciences), ion semiconductor (Ion Torrent Sequencing), pyrosequencing (Roche 454), sequencing by synthesis (Illumina), and sequencing by ligation (SOLiD sequencing) (Shendure and Ji 2008; Maclean et al. 2009; Metzker 2010; Mardis 2013). Among all NGS technologies, Pacific Biosciences platform provides the longest read length followed by Roche-454 platform which further has been widely used in transcriptome (RNA-seq) of various non-model plant species. Availability of relatively long reads substantially reduce the de novo assembly challenges as compared to other sequencing platforms (Sun et al. 2010; Guo et al. 2010). Illumina is another highly utilized platform used for transcriptome analysis of various non-model organisms due to its potential for high sequence yield, though, efforts to increase its read length

in underway, its short reads creates challenges for various bioinformatics analysis.

Remarkable progress has been made to explore medicinal plants at genomic/transcriptomic level for the ultimate goal of identify genes involved in biologically active phytochemicals and related pathways through NGS technologies. In the current article an attempt is made to review the application and contribution of NGS technologies in exploring medicinal plants with focus on seabuckthorn (*Hippophae* sp.), an important plant species of Indian Himalayas, having enormous medicinal, nutritional and therapeutic potential.

Next generation sequencing technologies

The first generation sequencing (FGS) technology was developed originally by two groups, Sanger et al. (1977) used chain-terminating inhibitors and Maxam and Gilbert (1977) used chemical degradation strategy. Later, advancement in the sequencing technologies led to the advent of second and third generation sequencing technologies which are collectively called as NGS technologies. The main focus of advancement in sequencing strategies demand for fast and low cost high throughput sequencing. In NGS large amount of sequencing data with millions of reads can be generated in parallel. NGS technologies may be useful for a variety of analysis including genome annotation, discovery of alternative isoforms, and gene expression profiling. Moreover, the technology has also helped researchers to gain deep insight into the mechanisms underlying processes in secondary metabolites biosynthesis, and development of genomic resources for diversity analysis, evolutionary analysis and marker assisted improvement of non model but important plant species (Shendure and Ji 2008; Maclean et al. 2009; Metzker 2010; Mardis 2013; Unamba et al. 2015). Different NGS platforms are based on variable chemistries and methods. The commonly used platforms of NGS and Sanger Sequencing are compared on the basis of chemistry used, read length, accuracy along with advantages and disadvantages in Table 2.

Second generation sequencing

The first commercially available NGS system 454-Roche uses pyrosequencing technology, which is based on detection of pyrophosphate released during nucleotide incorporation (Berka et al. 2010). Initially, the read length generated in 454-Roche was 100-150 bp, with data generated around 20 MB per run from >200,000 reads. The performance of 454-Roche was drastically enhanced by the introduction of 454 GS FLX Titanium system in 2008. The read length increased up to 700–1000 bp with more accuracy and approximately 0.7–14 GB data per run (Liu et al. 2012). The biggest advantage of 454-Roche has been its speed, however, the high cost of reagent and instrument pose discouragement for 454-Roche system. The other NGS platform frequently used for various genome/transcriptome analyses is Illumina, earlier known as Solexa Genome Analyzer (GA). Solexa released its first instrument in 2006 with name Genome Analyzer (GA) which was later taken over by Illumina in 2007. The chemistry used by Illumina exploits sequencing by synthesis (SBS) approach. Initially, the output of GA was 1 GB per run which reached up to 200 GB per run with introduction of HiSeq 2000 platform by Illumina. Though continuous efforts were made to increase the read length in Illumina from 75 to 300 bp, still lesser than 454-Roche. However, the other advantages which makes Illumina platforms as a choice for transcriptome analysis is its potential to generate large amount of data, and lesser cost of reagents and instruments (Quail et al. 2012). The third generation NGS platform named as SOLiD was introduced by Applied Biosystems in 2006. SOLiD exploits Synthesis by Oligo Ligation Detection (SOLiD) chemistry. Initially the read length in ABI SOLiD System was 35 bp with 3 GB data per run and high accuracy of 99.85% (Mardis 2008). Though later in 2010 to 2013, ABI SOLiD released five upgraded platforms from SOLiD to SOLiD 5500xl with improved read lengths up to 85 bp, amount of data generated 30 GB per run, and accuracy of 99.99%. However, the short read lengths and expensive computational analysis continue to be a major drawback as compared to other NGS platforms (Liu et al. 2012; Quail et al. 2012). Ion Torrent Systems Inc. of

Life Technologies has developed semiconductor detection based system, another widely used NGS platform. The read length generated in Ion Torrent is 200–400 bp with 98% accuracy. Though its low cost and high speed is the major advantage, detection of unwanted electronic signals due to homopolymer errors remain its major disadvantage (Rusk 2011).

Third generation sequencing

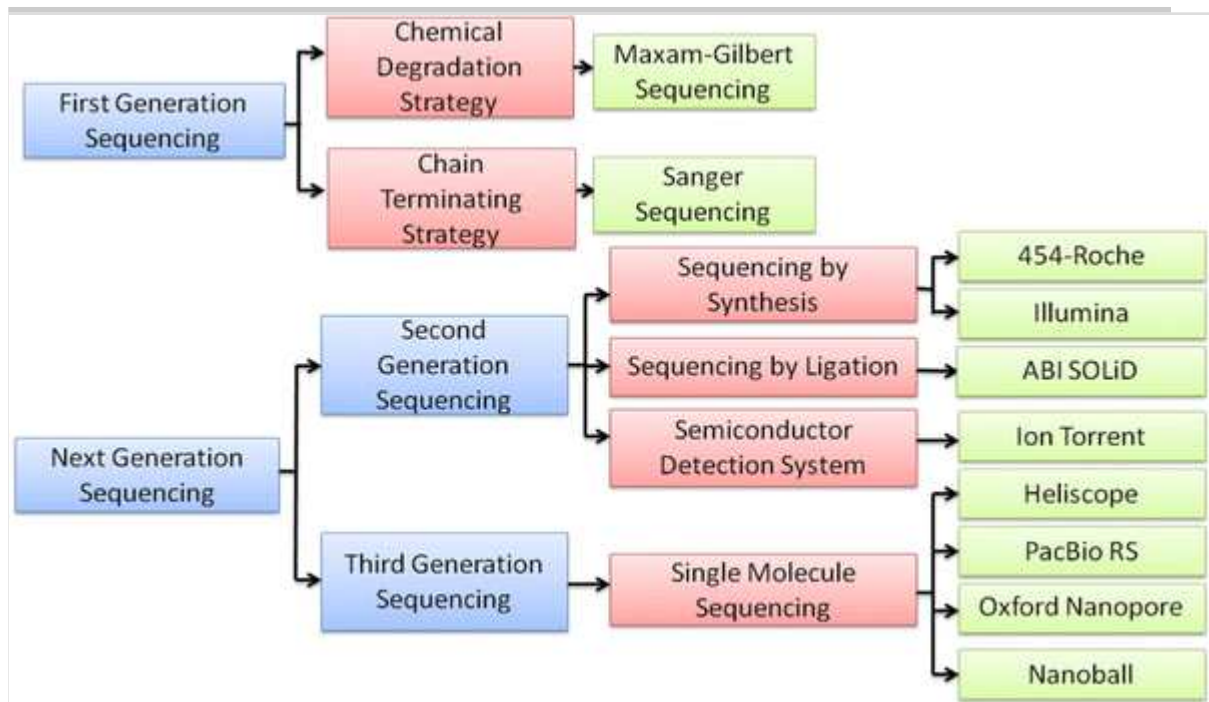
The recent advancements in the field of NGS include technologies developed within last 5 years with the generation of single molecule sequencing (SMS) systems. These are considered as third generation sequencing technologies and include: (i) Heliscope single molecule sequencing by Helicos Biosciences, (ii) Single molecule real time (SMRT) sequencing by Pacific Biosciences, (iii) Nanoball sequencing by Complete Genomics, and (iv) Nanopore sequencing by Oxford Nanopore Technologies. The concept used in third generation sequencing is to sequence and detect single molecule at a time with generation of longer reads proved to be beneficial for NGS technologies, as it magnificently enhanced data generation (Schadt et al. 2010; Thompson and Milos 2011). The first commercially used sequencing platform for SMS technology was Heliscope of Helicos Biosciences Subsequent introduction of PacBio RS platform by Pacific Biosciences, detects the fluorescent label dNTPs in real time manner and the technology has been named as single molecule real time (SMRT) sequencing. The technology further got revolutionized by nanopore sequencing of Oxford Nanopore Technologies where signal detection based upon SMS approach with nano sized pores (Schneider and Dekker 2012). Though the use of third generation sequencing technologies provides number of benefits in terms of potential, read length, amount of data generated, low cost over first and second generation sequencing. Continuous efforts are being made by commercial sequencing providers to overcome the drawbacks faced while using third generation sequencing such as moderate throughput, and bioinformatic challenges to analyze large amount of data (Schadt et al. 2010; Thompson and Milos 2011).

Emerging trends in sequencing technology from chain termination based

sequencing to single molecule sequencing (Fig. 1) have facilitated diverse genomic/transcriptomic research on medicinal plants. Sequence data generated from these technologies provide deep insight of genome/transcriptome of medicinal plants which can be subsequently exploited for the identification of key genes, and understanding processes and pathways for biosynthesis of biologically active phytochemicals.

Fig. 1

DNA sequencing technologies



NGS based research on medicinal plants

Being rich in biologically active compounds, medicinal plants need to be first investigated at genomic/transcriptomic level. Subsequently, informative sequences and expression datasets generated can be used to characterize genes and biochemical pathways involved in biosynthesis of biologically active phytochemicals (Góngora-Castillo and Buell 2013). Research on medicinal plant genome/transcriptome level is also crucial for discovery of phytochemicals and development of plant pharmaceutical

resources. Keeping the importance of genomic/transcriptomic data of medicinal plants, various studies have been conducted in recent past. In the following text, various genomic and transcriptomic projects on medicinal plants are briefly summarized with special emphasis on seabuckthorn, an important medicinal plant of high altitude regions of Eurasia.

NGS based whole genome sequencing (WGS) studies in medicinal plants

In the last 5 years, various NGS technology based whole genome sequencing projects on medicinal plants have been successfully accomplished. In China, a major project called “Herb Genome Project” was initiated with an aim to sequence whole genome and study secondary metabolite pathways of various medicinal plants (Chen et al. 2011). Similarly, National Institute of Health (NIH) initiated the research on medicinal plants genome/transcriptome at Michigan State University (Medicinal Plant Genomics Resource) to identify the genes involved in secondary metabolite pathways (<http://medicinalplantgenomics.msu.edu/index.shtml>). Whole genome sequencing (WGS) studies provided deep insight into genetic and evolutionary frame work of medicinal plants. The sequence datasets generated may further be used for mining key genes, pathways, and molecular markers involved in biosynthesis of biologically active phytochemicals. Though genomic information is not available or little known for majority of medicinal plants, continuous efforts are made to accumulate more and more genomic resources for these non-model but important plant species. Studies carried out in recent years on different organeller genomes in medicinal plants are recently reviewed by Sharma and Shrivastava (2016). Some implications of NGS-based nuclear genomic studies on some traditionally used medicinal plants are summarized in Table 3.

The first NGS based nuclear genome information was provided in *Cannabis sativa*, a plant known for its medicinal and intoxicating properties. The genes and pathways related to cannabinoid biosynthesis

were explored with generation of 534 Mb draft genome of *C. sativa* (van Bakel et al. 2011). Similarly in the succeeding year, genomes of various traditionally used medicinal plants including *Azadirachta indica*, *Capsicum annum*, *Dendrobium catenatum*, *Elaeis guineensis*, *Juglans regia*, *Nelumba nucifera*, *Ocimum tenuiflorum*, *Prunus mume*, *Pyrus communis*, *Ziziphus jujube* were sequenced (Table 3). The first genomic sequence draft of *Azadirachta indica* (Neem), the most commonly used medicinal plant was reported by Krishnan et al. (2012). Unlike other medicinal plants, the genome of *A. indica* is less complex comprising of 20,000 genes representing 364 Mb genome, in comparison to the other medicinal plants having complex and larger genomes. For example, *Ziziphus jujube* genome sequence contains large number of repetitive elements constituting complex genome of 437.65 Mb and 32,808 genes (Liu et al. 2014a, b). Thus, despite of many challenges in sequencing and assembling complex genomes, many whole genome sequencing project have been accomplished and many are underway, owing to revolutionary advances in sequencing technology and informatics tools. Nevertheless, immediate implications of NGS need to be important in the area of genome characterization, gene identification, comparative genomics, molecular markers mining, and more for development of value added phytomedicines (Sharma and Shrivastava 2016).

NGS-based transcriptome sequencing studies in medicinal plants

Most of the medicinal plants are considered as non model species but have some key elements with certain unique characteristics. The transcriptome research on medicinal plant has accelerated with the advancement in NGS technologies, particularly due to development of RNA sequencing either directly or through cDNA. Implementation of NGS technology in recent past has enabled identification of genes encoding certain unique characteristics in medicinal plants. Moreover, it also provides deep understanding of gene expression and underlying mechanism of secondary metabolites synthesis (Hao et al. 2012). Further, as the reference genome

sequence information is not available for most of the medicinal plants, de novo transcriptome assembly of short sequence reads has served as an alternative for gene discovery, comparative analysis, expression analysis and identification of sequence variants (Góngora-Castillo and Buell 2013). The development of RNA-seq can further facilitate research through identification of novel transcripts involved in bioactive compound metabolism and in finding alternate splice forms of genes (Wang et al. 2009). A list of NGS based transcriptome studies in medicinal plants implemented in the last five years is summarized in Table 4.

Transcriptome studies to identify genes involved in biosynthesis pathways

Primary aim of most NGS based transcriptomic studies in medicinal plants is to identify genes involved in biosynthesis of secondary metabolites or biologically active compounds. For example, transcriptome analysis of *Allium sativum*, a well known medicinal plant, aimed to identify genes involved in biosynthesis of organic sulfur (Sun et al. 2012a, b). Another case of *Curcuma longa* transcriptome profiling revealed abundance of transcripts related to ethanol metabolism and other important anticancer and anti-malarial terpenoids (Annadurai et al. 2013). Transcriptome analysis of *Azadirachta indica*, a most widely used medicinal plant, targeted the characterization of metabolic pathways involved in synthesis of bioactive compounds, comparative evolutionary studies and understanding of molecular pathways involved in the azadirachtin synthesis, the key component responsible for its medicinal value (Krishnan et al. 2012). *Withania somnifera*, is another example of this kind of research where NGS based transcriptome studies were performed to elucidate the key genes involved in tissue specific withanolide biosynthesis (Gupta et al. 2013, 2015). Transcriptome of *Catharanthus roseus*, was explored using NGS to develop metabolic pathways database “CathaCyc” (Van Moerkercke et al. 2013). It provides information about metabolic pathways, enzymes, proteins and genes involved in biosynthesis of various bioactive compounds. This type of database may have utility as

a reference for the study of primary and secondary metabolism in medicinal plants including *C. roseus*. Other medicinal plants where genes involved in biosynthesis of certain bioactive compounds were targeted are listed in Table 4. In some transcriptome studies, sequence data-sets were combined with metabolomic data-sets to reveal the complete picture of secondary metabolites. This approach termed as Phytochemical Genomics Approach has been used in various medicinal plants (Muranaka and Saito 2013; Sumner et al. 2015).

Transcriptome sequence based comparative analysis in medicinal plants

Digital gene expression (DGE) analysis is another process used in various transcriptome studies. Such type of studies in medicinal plants are performed for the identification of differentially expressed genes involved in secondary metabolites biosynthesis at different developmental stages or environmental conditions, as regulation of biosynthesis varies with species, developmental stages, environmental conditions and are tissue specific. For example, in case of *Ocimum* sp., a comparative analysis of two species *O. sanctum* and *O. basilicum*, was performed using NGS based transcriptome profiling (Rastogi et al. 2014). The study revealed significant difference in the differential pattern of genes related to secondary metabolites biosynthesis in two species. The data generated in this study may be useful in characterizing genes related to secondary metabolism, and for breeding special chemotypes in *Ocimum* sp. Other examples of transcriptome sequence based comparative analysis in medicinal plants include *Digitalis purpurea*, *Ephedra sinica* and *Scutellaria baicalensis* (Table 4).

Mining of small RNAs and molecular markers in transcriptome data

Transcriptome sequences generated by studies on many medicinal plants were also used for the mining of conserved and novel small RNAs and long non-coding RNAs, and development of molecular markers having

applications in molecular breeding, marker assisted selection and QTL mapping. Transcriptome of *Panax ginseng* has been explored for the identification of novel and conserved microRNAs (miRNAs) (Wu et al. 2012).

The important data generated on transcriptomics of medicinal plants in recent past is likely to provide important leads for modern pharmaceutical industries towards development of herbal based medicines.

NGS based transcriptome studies on seabuckthorn (*Hippophae* sp.)

Seabuckthorn (*Hippophae* sp.), is a hardy, deciduous shrub that belongs to the family Elaeagnaceae. Seabuckthorn, a medicinally and economically important shrub of cold arid zones of high altitudes of Europe and Asia, particularly of Himalayas is also distributed in many other countries of Asia, Europe, South and North America. Since ancient ages, people of many countries of Asia and Europe have been using various parts of the seabuckthorn plant as a source of traditional medicine for effective treatment of various ailments. Extracts of seabuckthorn berries are commonly used for the treatment of skin diseases, jaundice, asthma, and for gastro-intestinal disorders (Singh et al. 2006). Medicinal and nutritional value of seabuckthorn is attributed to the presence of a variety of bioactive compounds in different parts of seabuckthorn i.e. berries, seeds, bark and leaves. Some of the important clinically proven medicinal properties include immunomodulatory, anti-inflammatory, and anti-cancer activity, hepatoprotective activity, anti-stress and adaptogenic activity, modulation of hypoxia-induced transvascular leakage, cardioprotective and anti-atherogenic effects, healing effect on acute and chronic wounds, anti-bacterial and anti-viral effects, anti-radiation effects (Geetha et al. 2008; Suryakumar and Gupta 2011).

In last decade, due to its high nutritional and medicinal value, seabuckthorn has attracted attention of many researchers worldwide. Moreover, the advancement in the field of NGS technology enhanced the

research on transcriptomic level of this multipurpose medicinal plant. Several transcriptome studies were carried out on seabuckthorn using NGS with different aims and objectives, in recent years (Table 1). Fatima et al. (2012) studied the fatty acid composition of Canadian-grown cultivars (ssp. *mongolica*) and explored seabuckthorn seed transcriptome using the 454 GS FLX sequencing technology. They reported the presence of high content of linoleic and α -linolenic acids in seed oil, while the pulp oil contained palmitoleic acid in abundance. Further, they employed 454 GS FLX sequencing on seabuckthorn mature seeds which yielded 500,392 sequence reads, and reported 89,141 putative unigenes. Functional annotation and computational prediction of metabolic pathways indicated that primary metabolism and fatty acid and lipid biosynthesis pathways were highly represented categories in their data-sets.

Table 1

NGS based transcriptome studies in seabuckthorn (*Hiipophae* sp.)

Aim	Tissue used	NGS platform	Data generated	Major finding	Referen
Fatty acid composition in seed; identification of genes and enzymes involved in biosynthesis of essential oils	Seed	454 GS FLX	89,141 putative unigenes	Seed oil of Canadian-grown sea buckthorn cultivars contains high levels of linoleic acid and α -linolenic acid	Fatima e al. (201
Complete transcriptome profiling	Leaf and root	Illumina Hiseq 2000	53 Mb transcriptome; 88,297 unigenes	Comprehensive view of seabuckthorn transcriptome	Ghangal al. (201
Mining of microsatellites	Leaf and root	Illumina Hiseq 2000	7.69% unigenes harbored microsatellite repeats	Microsatellite dynamics in seabuckthorn transcriptome	Jain et a (2014)

Aim	Tissue used	NGS platform	Data generated	Major finding	Referen
Differential expression of genes under cold and freeze stress	Leaf	Illumina Hiseq 2000	428 Differentially expressed genes	Differentially expressed genes (DEGs) under cold and freeze stress	Chaudh and Sharma (2015)

Table 2

Comparative analysis of different NGS platforms and Sanger sequencing technology (

NGS platform	Chemistry	Read length (bp)	Reads generated per run	Accurac
454-Roche	Pyrosequencing	700–1000	1 million	99.9%
Illumina	Sequencing by synthesis (SBS)	75–300	From 1 million to 3 billion, depending upon the sequencer model	99.9%
ABI SOLiD	Sequencing by ligation	35–85	1.2–1.4 million	99.9%
Pacific biosciences	Single-molecule real-time sequencing	10,000–15,000	500–1000 megabases	87%
Ion torrent	Ion semiconductor	up to 400	70–80 million	98%
Sanger sequencing	Chain termination	400–900	–	99.9%

NGS platform	Chemistry	Read length (bp)	Reads generated per run	Accuracy

Table 3

Summary of NGS-based nuclear genome studies in medicinal plants

Medicinal plant	Traditional medicine uses	NGS platform	Aim	Genome size
<i>Azadirachta indica</i>	Anthelmintic, antifungal, antidiabetic, antibacterial, antiviral, contraceptive and sedative	Illumina; pacific biosciences SMRT	Terpenoid biosynthesis pathway	364 Mb
<i>Cannabis sativa</i>	Hallucinogenic, hypnotic, sedative, analgesic, and anti-inflammatory	Illumina, 454-Rosche	Cannabinoid biosynthesis	534 Mb
<i>Capsicum annum</i>	Use to treat flatulence, tympanitis, paralysis, acute diphtheria, malignant scarlet fever	Illumina	Evolution, domestication, and specialization studies	~3.48 Gb
<i>Dendrobium catenatum</i>	Anti-Inflammatory, immuno-enhancing, antioxidant and anti-glycation activities	Illumina	Polysaccharide synthase, flower development and evolution studies	1.01 Gb
<i>Elaeis guineensis</i>	Used as laxative and diuretic, poison antidote, cure for gonorrhea, menorrhagia, bronchitis, headaches, rheumatism, healing of fresh wounds and treat skin infections	Illumina	Oil biosynthesis genes	1.535 Gb
<i>Juglans regia</i>	Antioxidant and antiproliferative, used to treat diabetes mellitus	Illumina	Genes coding for the biosynthesis of	667 Mb

Medicinal plant	Traditional medicine uses	NGS platform	Aim	Genome size
	symptoms		non-structural polyphenols	
<i>Nelumba nucifera</i>	Anticancer, antidepressant, cure insomnia, and diarrhea	Illumina, 454-Roche	Genome characterization and evolution	929 Mb
<i>Ocimum tenuiflorum</i>	Antibacterial, antiallergic, antidepressant	Illumina	Genes behind strong medicinal properties	374 Mb
<i>Prunus mume</i>	Antimicrobial, prevent gastritis and gastric ulcers, enhance the oxidative capacity	Illumina	Genome characterization and evolution	280 Mb
<i>Pyrus communis</i>	Antimicrobial, antiseptic	454-Roche	Genome characterization	577 Mb
<i>Ziziphus jujuba</i>	Alleviate stress, anti-fungal, anti-bacterial, anti-ulcer, anti-inflammatory, antispastic, antifertility/contraception, hypotensive and antinephritic, cardiogenic, antioxidant, immunostimulant, and wound healing properties	Illumina	Vitamin C content, genomic resource	437.6 Mb

Table 4

NGS-based transcriptome studies on medicinal plants

Medicinal plant	Therapeutic potential	Aim of study	References
<i>Allium sativum</i>	Antimicrobial, antithrombotic, hypolipidemic, antiarthritic, hypoglycemic and antitumor activity	De novo assembly and characterization of transcriptome; genes involved in biosynthesis of organic sulfur	Sun et al. (2012a, b)

Medicinal plant	Therapeutic potential	Aim of study	References
<i>Amorphophallus konjac</i> ; <i>Amorphophallus bulbifer</i>	Treatment of asthma, cough, hernia, breast pain, burns as well as hematological and skin disorders	development of transcriptome data and microsatellite markers	Zheng et al. (2013)
<i>Andrographis paniculata</i>	Used in treatment of cancer, diabetes, high blood pressure, ulcer, leprosy, bronchitis, skin diseases, flatulence, colic, influenza, dysentery, dyspepsia and malaria	Identification of novel genes for Ent-labdane-related diterpene (ent-LRD) biosynthetic pathway	Garg et al. (2015)
<i>Asparagus racemosus</i>	Used as a uterine tonic, to improve breast milk, in hyperacidity, and as a general health tonic	Identification of novel genes involved in steroidal sapogenin biosynthesis	Upadhyay et al. (2014)
<i>Astragalus chrysochlorus</i>	Act as adaptogen, treatment of anemia, cold, influenza, diabetes, fatigue or lack of appetite from chemotherapy, heart diseases, hepatitis, kidney disease	Transcriptome profiling; to study selenium accumulation and tolerance mechanisms	Cakir et al. (2015)
<i>Astragalus membranaceus</i>	Used to treat common cold, respiratory infections, allergies, fibromyalgia, anemia, HIV/AIDS, and to strengthen immunity	Characterization of transcriptome	Liu et al. (2014a, b)
<i>Atractylodes lancea</i>	Anticancer, anti-inflammatory, antimicrobial, antipyretic, and activities on central nervous, cardiovascular, and gastrointestinal systems	gene identification involved in sesquiterpenoid biosynthesis	Ahmed et al. (2016)

Medicinal plant	Therapeutic potential	Aim of study	References
<i>Azadirachta indica</i>	Anthelmintic, antifungal, antidibitic, antibacterial, antiviral, contraceptive, sedative	Characterization of metabolic pathways involved in synthesis of bioactive compounds; comparative evolutionary studies and understanding of molecular pathways involved in the azadirachtin synthesis	Krishnan et al. (2012)
<i>Calotropis gigantea</i>	Used for digestive disorders including diarrhea, constipation and stomach ulcers	Transcriptome study; identification of genes in secondary metabolites biosynthesis	Muriira et al. (2015)
<i>Carthamus tinctorius</i>	Used to treat dysmenorrhea, amenorrhea, postpartum abdominal pain and mass, trauma and pain of joints	De novo transcriptome assembly	Lulin et al. (2012)
<i>Cassia angustifolia</i>	Treatment of constipation, irritable bowel syndrome, hemorrhoids, and weight loss	Identification of genes involved in various secondary metabolite pathways; especially those related to the synthesis of sennosides	Reddy et al. (2015)
<i>Catha edulis</i>	To treat obesity, dysphoria and sedation	Identification of gene candidates potentially involved in amphetamine-type alkaloid	Groves et al. (2015)

Medicinal plant	Therapeutic potential	Aim of study	References
		biosynthesis	
<i>Catharanthus roseus</i>	Lowers blood pressure, anti-diabetic, anticancer, antiviral, antibacterial	Development of CathaCyc, a metabolic pathway database	Van Moerkercke et al. (2013)
<i>Centella asiatica</i>	Treatment of varicos veins, psoriasis, chronic venus insufficiency, minor wounds	Identification of genes of terpenoid and ROS metabolism; mining out secondary-metabolism genes; discovery of novel genes for biosynthetic pathways	Sangwan et al. (2013)
<i>Chlorophytum borivilianum</i>	Used as adaptogen	Study of molecular network and important pathways	Kalra et al. (2013)
<i>Codonopsis pilosula</i>	To improve appetite and energy	Elucidation of genes involved in biosynthetic pathways for codonopsis polysaccharides	Gao et al. (2015)
<i>Curcuma longa</i>	Used to treat a variety of internal disorders, such as indigestion, throat infections, common colds, or liver ailments, as well as topically to cleanse wounds or treat skin sores	De novo transcriptome assembly; identification of novel transcripts related to anticancer and antimalarial terpenoids	Annadurai et al. (2013)
<i>Cymbopogon flexuosus</i>	Treatment of digestive tract spasms, stomachache, high blood pressure, convulsions, pain,	De novo transcriptome assembly; identification of genes involved in	Meena et al. (2016)

Medicinal plant	Therapeutic potential	Aim of study	References
	vomiting, cough, achy joints (rheumatism), fever, the common cold, and exhaustion	essential oils	
<i>Dendrobium officinale</i>	Tonic for longevity, used to boost physical and athletic performance	Identification of alkaloid biosynthesis genes; genetic markers development	Guo et al. (2013)
<i>Digitalis purpurea</i>	Used for congestive heart failure, relieving associated fluid retention (edema), irregular heartbeats, asthma, epilepsy, tuberculosis, constipation, and headache	study of comparative analysis of alternate splicing	Wu et al. (2014)
<i>Eleutherococcus senticosus</i>	To treat bone marrow suppression angina, hypercholesterolemia, and neurasthenia with headache, insomnia, and poor appetite	Transcriptome profiling; identification of genes involved in somatic embryogenesis	Tao et al. (2016)
<i>Ephedra sinica</i>	Treat symptoms of bronchial asthma, colds, influenza, allergies, and hives in teas or tinctures	De novo transcriptome assembly; comparative transcriptome analysis for aerial roots and stems	Groves et al. (2015)
<i>Erigeron breviscapus</i>	Treatment of cerebrovascular diseases	Generation of transcriptome sequence resource; effect of methyl jasmonate (MeJA) on scutellarin biosynthesis	Chen et al. (2015)
<i>Gentiana macrophylla</i>	Used in treatment of jaundice, hepatitis,	Identification of genes involved in	Hua et al. (2014)

Medicinal plant	Therapeutic potential	Aim of study	References
	and stomachic and choleric ailments	secoiridoid biosynthesis	
<i>Gentiana rigescens</i>	Treatment of fever, hypertension, muscle spasms, parasitic worms, wounds, cancer, sinusitis, and malaria	De novo transcriptome assembly; identification of genes involved in terpenoid biosynthesis	Zhang et al. (2015a, b)
<i>Glycyrrhiza uralensis</i>	Treatment of arrhythmia, ulcers, inhibits gastric acid secretion, relieves gastrointestinal smooth muscle spasm, eases pain, promotes the secretion of pancreatic juice, relieves cough significantly, prevents asthma, antibacterial, antiviral, anti-inflammatory, anti-allergic	Elucidation of biosynthetic pathways of secondary metabolites	Ramilowski et al. (2013)
<i>Helwingia willd</i>	Antibacterial, anti-inflammatory and blood lipid reducing effects	Transcriptomic exploration and phylogenetic analysis	Sun et al. (2014)
<i>Hypericum perforatum</i>	Antidepressant and anticancer activities	Identification of potential genes that are related to plant reproduction	Galla et al. (2015)
<i>Hypericum perforatum</i>	Antidepressant, anti-inflammatory	Identification of potential genes involved in the biosynthesis of active metabolites	He et al. (2012)
<i>Lonicera japonica</i>	Used to treat fever, headache, cough, thirst and sore throat	De novo transcriptome assembly; identify genes involved in	Yuan et al. (2012); Rai et al. (2016)

Medicinal plant	Therapeutic potential	Aim of study	References
		chlorogenic acid, luteolosides, and secoiridoid biosynthesis pathways	
<i>Lotus corniculatus</i>	Antiinflammatory, antispasmodic, cardiotoxic, carminative, febrifuge, hypoglycaemic, restorative, sedative, tonic, vermifuge	Identification of genes for key enzymes and potential transcription factors related to the flavonoid biosynthesis pathway	Wang et al. (2013)
<i>Lycium chinense</i>	Supplementing the kidney and liver, benefiting the eyes, enhancing immune functions, and protecting against adverse impact of oxidation	Identification of phenylpropanoid biosynthetic genes and phenylpropanoid accumulation; de novo transcriptome assembly; essential genes involved in the carotenoid biosynthesis pathway	Zhao et al. (2013); Wang et al. (2015a, b)
<i>Magnolia sprengeri</i>	To treat menstrual cramps, abdominal pain, abdominal bloating and gas, nausea, and indigestion	Identification of structural and regulatory genes encoding the enzymes involved in the determination of flower color	Shi et al. (2014)
<i>Ocimum sp.</i>	Anti inflammatory, effective in skin rashes, insect bites and itching, effective in fever, cough, bronchitis, acts as a cardiac tonic and purifies blood, very	De novo transcriptome assembly; comparative analysis	Rastogi et al. (2014)

Medicinal plant	Therapeutic potential	Aim of study	References
	effective in reducing blood sugar and blood cholesterol		
<i>Panax ginseng</i>	Used in treatment for Alzheimer's disease, lung disease, flu, erectile dysfunction, multiple sclerosis-related fatigue, improves mental function	Identification of novel and conserved miRNAs; development of root transcriptome; genes for secondary metabolites biosynthesis	Wu et al. (2012); Li et al. (2013); Jayakodi et al. (2015); Subramaniyam et al. (2014)
<i>Panax japonicus</i>	Used as a tonic, anti-inflammatory and haemostatic agent	Generation of transcriptome data; identification of genes involved in triterpenoid saponin backbone biosynthesis	Zhang et al. (2015)
<i>Panax notoginseng</i>	Lower cholesterol; improves cardiovascular function	Establishment of global transcriptome dataset; candidate genes involved in ginsenoside and alkaloid biosynthesis	Liu et al. (2015a, b)
<i>Panax quinquefolius</i>	Improves neurocognitive function, memory	Identification of inenoside biosynthesis genes; to provide transcriptome sequences for seed dormancy	Wu et al. (2013); Qi et al. (2015)
<i>Panax vietnamensis</i>	protect stress-induced antinociception, gastric lesions, pentobarbital sleep	Identification of putative genes involved in triterpenoid saponins biosynthesis	Zhang et al. (2015)

Medicinal plant	Therapeutic potential	Aim of study	References
		pathway; SSR markers development	
<i>Paphiopedilum concolor</i>	Treatment to reduce inflammation and acesodyne, dysmenorrheal, colic, and cystitis	De novo transcriptome assembly; development of microsatellite markers	Li et al. (2015)
<i>Phyllanthus amarus</i>	Use to treat bladder infections, colds, diabetes, hepatitis B, kidney disorders, ulcers, urinary tract infections	De novo transcriptome assembly; decipher various secondary metabolite pathways	Mazumdar and Chattopadhyay (2015)
<i>Picrorhiza kurroa</i>	Treatment of digestive problems, asthma, liver damage, wound healing, vitiligo	De novo transcriptome assembly; identification of genes involved in picrosides biosynthesis	Gahlan et al. (2012)
<i>Pinellia ternata</i>	Removing dampness-phlegm, thermogenesis, treatment of flu	Comprehensive transcriptome analysis; identification of genes involved in the biosynthesis of benzoic acid and ephedrine	Huang et al. (2016); Zhnag et al. (2016a, b)
<i>Plantago ovata</i>	Used to improve gastrointestinal functions, weight loss, prevention and treatment of hypertension and heart diseases	De novo transcriptome assembly; characterization of mucilage biosynthesis pathway	Kotwal et al. (2016)
<i>Platycodon grandiflorum</i>	Lung-heat-clearing, antitussive, expectorant	Transcriptome profiling; gene discovery; marker development;	Ma et al. (2016); Kim et al. (2015)

Medicinal plant	Therapeutic potential	Aim of study	References
		identification of candidate genes involved in the biosynthesis of triterpenoid saponins	
<i>Polygala tenuifolia</i>	Used as sedative, antipsychotic, cognition improving, neuroprotective and antidepressant	Identification of genes involved in various secondary metabolites biosynthesis pathways	Tian et al. (2015)
<i>Polygonum multiflorum</i>	Treatment of liver injury, cancer, diabetes, alopecia, atherosclerosis, and neurodegenerative diseases	Study of mechanisms of MeJA-mediated stilbenoid biosynthesis	Liu et al. (2016)
<i>Polyporus umbellatus</i>	Immunostimulating, anticancer, anti-inflammatory, and hepatoprotective properties	Transcriptome analysis of genes involved in defence response with <i>Armillaria mellea</i> infection	Liu et al. (2015a, b)
<i>Rehmannia glutinosa</i>	Treatment of tinnitus, hair loss, scavenging of free radicals	Identification of the genes responding to replanting disease	Yang et al. (2015)
<i>Rehmannia glutinosa</i>	Use to treat diabetes, anemia, fever, osteoporosis, allergies	Identification of genes involved in iridoid biosynthesis	Sun et al. (2012a, b)
<i>Rhodiola algida</i>	Treating cancer, diabetes, tuberculosis, cold, flu, liver damage, aging, improves hearing and strengthen immunity	Characterization of important traits related to secondary metabolite formation associated molecular mechanisms	Zhang et al. (2014)

Medicinal plant	Therapeutic potential	Aim of study	References
<i>Salvia miltiorrhiza</i>	Treatment of chronic renal failure, coronary heart disease, diabetes	Insight into tanshinone biosynthesis; alternate splicing	Yang et al. (2013); Xu et al. (2015)
<i>Scutellaria baicalensis</i>	To treat respiratory infections, hay fever and fever	EST sequencing and gene expression profiling	Park et al. (2014)
<i>Veratrilla baillonii</i>	Treatment of liver-related diseases	De novo transcriptome assembly; SSR marker discovery	Wang et al. (2015a, b)
<i>Veratrum californicum</i>	treatment for various kinds of cancers	Elucidation of steroid alkaloid biosynthesis	Augustin et al. (2015)
<i>Withania somnifera</i>	Treatment for tumors, tubercular glands, carbuncles, ulcers, burns and wounds	Analysis of the salicylic acid induced leaf transcriptome; analysis of withanolide specific genes to elucidate chemotype as well as tissue specific withanolide biosynthesis	Dasgupta et al. (2014); Gupta et al. (2013), (2015)
<i>Zantedeschia aethiopica</i>	Antibacterial	Characterization of the molecular and physiological defense aspects	de Souza Cândido et al. (2014)

In another study, a comprehensive view of seabuckthorn transcriptome from leaf and root tissues has been provided by Ghangal et al. (2013). They employed next generation massive parallel sequencing technology (Illumina Hiseq 2000) and de novo assembly to gain a comprehensive view of the seabuckthorn transcriptome. The optimization of de novo assembly of short reads were also performed with six assembly tools with two

different approaches to assembled 86,253,874 high quality short reads. Short read assembly yielded 88,297 transcripts (≥ 100 bp), representing about 53 Mb of seabuckthorn transcriptome. The average length of transcripts was 610 bp, N50 length 1198 bp and 91% of the short reads uniquely mapped back to seabuckthorn transcriptome. Among various strategy using different software for de novo assembly, removal of redundancy in short read dataset prior to assembly by short read assembler ABySS following additive k-mer approach was found suitable. Subsequent assembly by long read assembler TGICL yielded the optimum assembly. On the basis of sequence homology significant sequence similarity was shown between *Vitis vinifera* (grapes) and seabuckthorn. Further, BLAST2GO tool was used to assign gene ontology (GO) terms. In all, 38,830 transcripts had at least one GO term associated with each of them representing 43.9% of the seabuckthorn transcriptome. In the category of biological process, the largest groups were “primary metabolic process”, “cellular metabolic process” and “biosynthetic process” suggesting the abundance of secondary metabolite and other biologically active compounds metabolism pathways in seabuckthorn. Moreover, presence of groups like “cellular response to stimulus”, “response to external stimulus”, “response to biotic stimulus”, response to endogenous stimulus”, “response to abiotic stimulus” and “response to stress” in the dataset indicated that a large number of transcripts are expressed in response to environmental stresses (Fig. 2) In the category of molecular function, transcripts with “catalytic activity”, “protein binding” and “nucleotide binding” formed the largest groups. The seabuckthorn transcriptome generated in this study was also screened for the presence of transcription factors and a total of 7421 putative seabuckthorn transcription factor genes, distributed in at least 80 different families, were identified representing 8.4% of seabuckthorn transcripts. The most frequent transcription factors represented C3H, MADS, bHLH, NAC, and FAR1 families. Next, microsatellites or simple sequence repeats (SSR) were mined from NGS based transcriptome data of seabuckthorn (Jain et al. 2014). The study reported 7.69% of 88,297 transcripts harbored

microsatellite repeats with an average of one microsatellite per 6.704 Kb of seabuckthorn transcriptome. Among various repeats dinucleotide were found most abundant followed by trinucleotide repeats. Further, in coding region microsatellites were densely populated followed by 3' and 5' untranslated regions (UTRs). AG and AAG type repeats most frequently occurred in seabuckthorn transcriptome. On the basis of GO terms, 48.81% microsatellites possess unigenes assigned with at least one GO term to assess associations between microsatellite and biological role of known genes (Fig. 3). In “biological process” category, majority of the unigenes were involved in metabolic processes followed by biosynthetic processes, and responses to various stresses (Fig. 3a), again supported the presence of abundant secondary metabolites in seabuckthorn. However, other important processes like transcription and transport were also recorded. For the cellular components category, majority of the unigenes were found to be related to plastid followed by protein complex and cytoplasm (Fig. 3b). When concentrating on molecular function category, majority of unigenes were assigned to protein binding followed by nucleotide binding (Fig. 3c). Among all 25 unigene specific microsatellites were assessed on the basis of polymorphism(s) detected in 18 seabuckthorn collections from Leh (India). The major findings of the study include the microsatellite data-set which can be used in future breeding and molecular biology research projects in seabuckthorn and other medicinal plants.

Fig. 2

Gene ontology classification of seabuckthorn transcripts on the basis of their role in biological processes

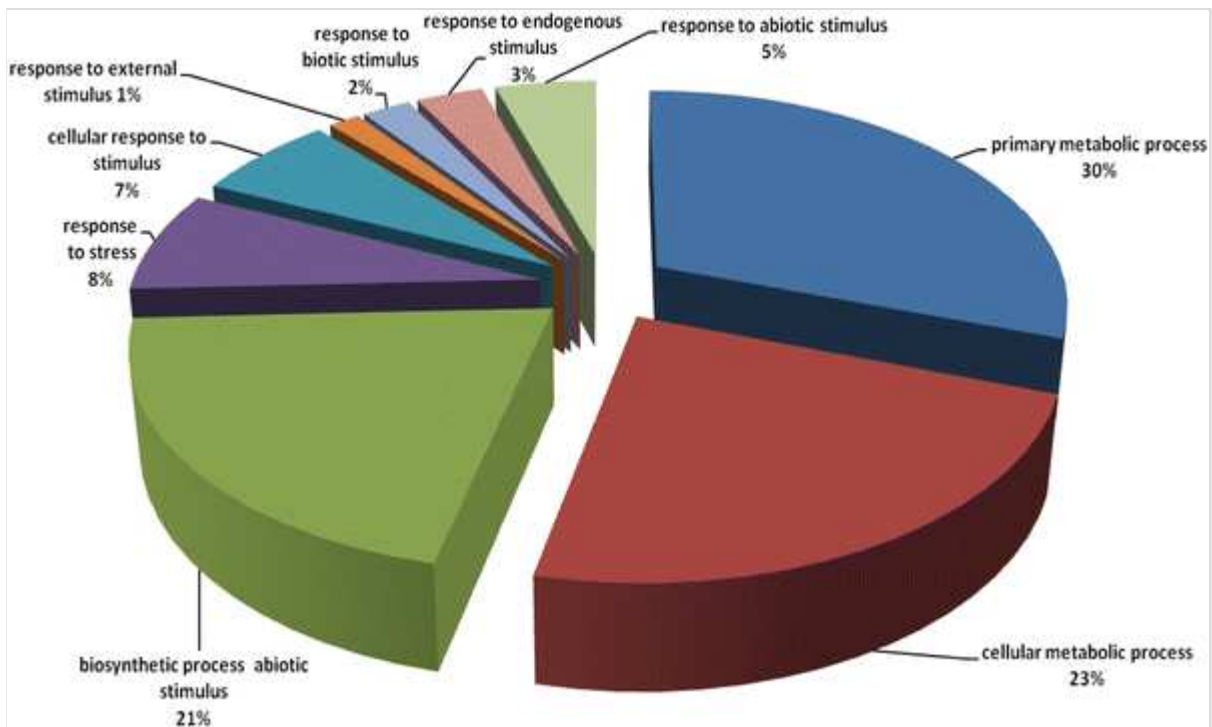
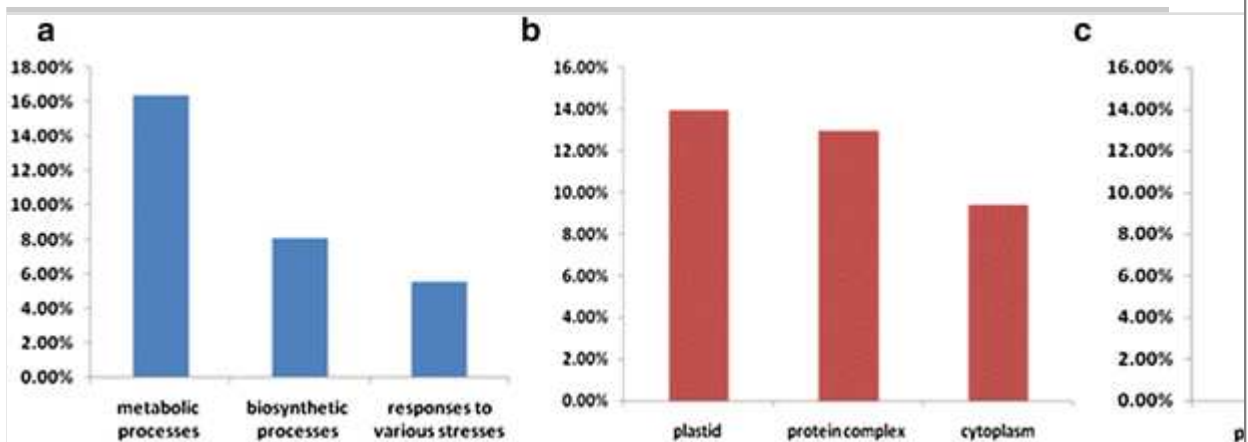


Fig. 3

Gene ontology annotation of microsatellite carrying seabuckthorn unigenes **a** Biological process; **b** cellular component, and **c** molecular function



AQ1

Seabuckthorn besides having medicinal and nutritional value also possesses ecological importance. It has the ability to sustain growth in harsh environmental conditions such as extreme temperatures, drought and

salinity. We employed DeepSAGE, a tag based NGS approach using Illumina Hiseq 2000 platform, to identify differentially expressed genes under cold and freeze stress in seabuckthorn (Chaudhary and Sharma 2015). The study reported 11,922 differentially expressed genes (DEGs) including 6539 up regulated and 5383 down regulated genes in response to cold and freeze stress in seabuckthorn. DEGs generated in the study were further mapped back to existing seabuckthorn transcriptome having 88,297 putative unigenes (Ghangal et al. 2013), and identified 428 cold and freeze stress responsive DEGs. Moreover, expression of 22 DEGs selected from DeepSAGE datasets were validated using qRT-PCR in support of DeepSAGE results. The major finding of the study provided comprehensive global gene expression profile of seabuckthorn under cold and freeze stresses. The data generated could also serve as a valuable resource to develop abiotic stress tolerant transgenic plants.

Conclusion

Medicinal plants are known to human history for their therapeutic potentials. Safer use of herbal based medicine than the synthetic drugs attracts the masses towards using more and more phytomedicines. Recent advances in NGS technologies have offered cost and labor efficient opportunities to explore medicinal plants at genomic and transcriptomic levels. Enormous genome/transcriptome sequence data-sets of various medicinal plants have been generated in recent past. Subsequently, these data-sets could serve as valuable resources for the selection of candidate genes involved in biosynthesis of biologically active phytochemicals. Such genes may ultimately be explored for the development of transgenic medicinal plants with enhanced therapeutic potentials. Though the NGS based research on important medicinal plants has already accelerated, implementation of NGS technologies is required for other plants which have therapeutic potentials and whose genome/transcriptome are not explored yet.

Acknowledgements

The authors acknowledge the financial support provided by the Department of Biotechnology (DBT), Government of India.

References

- Ahmed, S., Zhan, C., Yang, Y., Wang, X., Yang, T., Zhao, Z., et al. (2016). The transcript profile of a traditional Chinese medicine, *Atractylodes lancea*, revealing its sesquiterpenoid biosynthesis of the major active components. *PLoS One*, *11*(3), e0151975.
- Annadurai, R. S., Neethiraj, R., Jayakumar, V., Damodaran, A. C., Rao, S. N., Katta, M. A., et al. (2013). De novo transcriptome assembly (NGS) of *Curcuma longa* L. rhizome reveals novel transcripts related to anticancer and antimalarial terpenoids. *PLoS One*, *8*(2), e56217.
- Augustin, M. M., Ruzicka, D. R., Shukla, A. K., Augustin, J. M., Starks, C. M., O'Neil-Johnson, M., et al. (2015). Elucidating steroid alkaloid biosynthesis in *Veratrum californicum*: Production of verazine in Sf9 cells. *The Plant Journal*, *82*(6), 991–1003.
- Berka, J., Chen, Y. J., Leamon, J. H., Lefkowitz, S., Lohman, K. L., Makhijani, V. B., et al. (2010). Bead emulsion nucleic acid amplification. U.S. Patent No. 7,842,457. Washington, DC: U.S. Patent and Trademark Office.
- Cakır, Ö., Turgut-Kara, N., Arı, Ş., & Zhang, B. (2015). De novo transcriptome assembly and comparative analysis elucidate complicated mechanism regulating *Astragalus chrysochlorus* response to selenium stimuli. *PLoS One*, *10*(10), e0135677.
- Chagné, D., Crowhurst, R. N., Pindo, M., Thrimawithana, A., Deng, C., Ireland, H., et al. (2014). The draft genome sequence of European pear (*Pyrus communis* L. 'Bartlett'). *PLoS One*, *9*(4), e92644.

Chaudhary, S., & Sharma, P. C. (2015). DeepSAGE based differential gene expression analysis under cold and freeze stress in seabuckthorn (*Hippophae rhamnoides* L.). *PLoS One*, *10*(3), e0121982.

Chen, R. B., Liu, J. H., Xiao, Y., Zhang, F., Chen, J. F., Ji, Q., et al. (2015). Deep sequencing reveals the effect of MeJA on scutellarin biosynthesis in *Erigeron breviscapus*. *PLoS One*, *10*(12), e0143881.

Chen, S., Xiang, L., Guo, X., & Li, Q. (2011). An introduction to the medicinal plant genome project. *Frontiers of Medicine*, *5*(2), 178–184.

Dasgupta, M. G., George, B. S., Bhatia, A., & Sidhu, O. P. (2014). Characterization of *Withania somnifera* leaf transcriptome and expression analysis of pathogenesis-related genes during salicylic acid signaling. *PLoS One*, *9*(4), e94803.

de Magalhaes, J. P., Finch, C. E., & Janssens, G. (2010). Next-generation sequencing in aging research: emerging applications, problems, pitfalls and possible solutions. *Ageing Research Reviews*, *9*(3), 315–323.

de Souza Ca[^]ndido, E., da Rocha Fernandes, G., de Alencar, S. A., de Freitas Lima, S. M., de Jesus Miranda, V., Porto, W. F., et al. (2014). Shedding some light over the floral metabolism by arum lily (*Zantedeschia aethiopica*) spathe de novo transcriptome assembly. *PLoS One*, *9*(3), e90487.

Fatima, T., Snyder, C. L., Schroeder, W. R., Cram, D., Datla, R., Wishart, D., et al. (2012). Fatty acid composition of developing sea buckthorn (*Hippophae rhamnoides* L.) berry and the transcriptome of the mature seed. *PLoS One*, *7*(4), e34099.

Gahlan, P., Singh, H. R., Shankar, R., Sharma, N., Kumari, A., Chawla, V., et al. (2012). De novo sequencing and characterization of *Picrorhiza*

kurrooa transcriptome at two temperatures showed major transcriptome adjustments. *BMC Genomics*, 13(1), 1–21.

Galla, G., Vogel, H., Sharbel, T. F., & Barcaccia, G. (2015). De novo sequencing of the *Hypericum perforatum* L. flower transcriptome to identify potential genes that are related to plant reproduction sensu lato. *BMC Genomics*, 16(1), 1–22.

Gao, J. P., Wang, D., Cao, L. Y., & Sun, H. F. (2015). Transcriptome sequencing of *Codonopsis pilosula* and identification of candidate genes involved in polysaccharide biosynthesis. *PLoS One*, 10(2), e0117342.

Garg, A., Agrawal, L., Misra, R. C., Sharma, S., & Ghosh, S. (2015). *Andrographis paniculata* transcriptome provides molecular insights into tissue-specific accumulation of medicinal diterpenes. *BMC Genomics*, 16(1), 1–16.

Geetha, S., Jayamurthy, P., Pal, K., Pandey, S., Kumar, R., & Sawhney, R. C. (2008). Hepatoprotective effects of sea buckthorn (*Hippophae rhamnoides* L.) against carbon tetrachloride induced liver injury in rats. *Journal of the Science of Food and Agriculture*, 88(9), 1592–1597.

Ghangal, R., Chaudhary, S., Jain, M., Purty, R. S., & Sharma, P. C. (2013). Optimization of de novo short read assembly of seabuckthorn (*Hippophae rhamnoides* L.) transcriptome. *PLoS One*, 8(8), e72516.

Góngora-Castillo, E., & Buell, C. R. (2013). Bioinformatics challenges in de novo transcriptome assembly using short read sequences in the absence of a reference genome sequence. *Natural Product Reports*, 30(4), 490–500.

Groves, R. A., Hagel, J. M., Zhang, Y., Kilpatrick, K., Levy, A., Marsolais, F., et al. (2015). Transcriptome profiling of khat (*Catha*

edulis) and *Ephedra sinica* reveals gene candidates potentially involved in amphetamine-type alkaloid biosynthesis. *PLoS One*, 10(3), e0119701.

Guo, X., Li, Y., Li, C., Luo, H., Wang, L., Qian, J., et al. (2013). Analysis of the *Dendrobium officinale* transcriptome reveals putative alkaloid biosynthetic genes and genetic markers. *Gene*, 527(1), 131–138.

Guo, S., Zheng, Y., Joung, J. G., Liu, S., Zhang, Z., Crasta, O. R., et al. (2010). Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types. *BMC Genomics*, 11(1), 1–13.

Gupta, P., Goel, R., Agarwal, A. V., Asif, M. H., Sangwan, N. S., Sangwan, R. S., et al. (2015). Comparative transcriptome analysis of different chemotypes elucidates withanolide biosynthesis pathway from medicinal plant *Withania somnifera*. *Scientific Reports*, 5, 1–13.

Gupta, P., Goel, R., Pathak, S., Srivastava, A., Singh, S. P., Sangwan, R. S., et al. (2013). 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(5), e62714.

Hao, D. C., Chen, S. L., Xiao, P. G., & Liu, M. (2012). Application of high-throughput sequencing in medicinal plant transcriptome studies. *Drug Development Research*, 73(8), 487–498.

He, M., Wang, Y., Hua, W., Zhang, Y., & Wang, Z. (2012). De novo sequencing of *Hypericum perforatum* transcriptome to identify potential genes involved in the biosynthesis of active metabolites. *PLoS One*, 7(7), e42081.

Hua, W., Zheng, P., He, Y., Cui, L., Kong, W., & Wang, Z. (2014). An insight into the genes involved in secoiridoid biosynthesis in *Gentiana macrophylla* by RNA-seq. *Molecular Biology Reports*, *41*(7), 4817–4825.

Huang, X., Jing, Y., Liu, D. J., Yang, B. Y., Chen, H., & Li, M. (2016). Whole-transcriptome sequencing of *Pinellia ternata* using the Illumina platform. *Genetics and Molecular Research*, *15*(2), 1–14.

Jain, A., Chaudhary, S., & Sharma, P. C. (2014). Mining of microsatellites using next generation sequencing of seabuckthorn (*Hippophae rhamnoides* L.) transcriptome. *Physiology and Molecular Biology of Plants*, *20*(1), 115–123.

Jayakodi, M., Lee, S. C., Lee, Y. S., Park, H. S., Kim, N. H., Jang, W., et al. (2015). Comprehensive analysis of *Panax ginseng* root transcriptomes. *BMC Plant Biology*, *15*(1), 1–12.

Kalra, S., Puniya, B. L., Kulshreshtha, D., Kumar, S., Kaur, J., Ramachandran, S., et al. (2013). De novo transcriptome sequencing reveals important molecular networks and metabolic pathways of the plant, *Chlorophytum borivilium*. *PLoS One*, *8*(12), e83336.

Kim, H. J., Jung, J., Kim, M. S., Lee, J. M., Choi, D., & Yeam, I. (2015). Molecular marker development and genetic diversity exploration by RNA-seq in *Platycodon grandiflorum*. *Genome*, *58*(10), 441–451.

Kotwal, S., Kaul, S., Sharma, P., Gupta, M., Shankar, R., Jain, M., et al. (2016). De novo transcriptome analysis of medicinally important *Plantago ovata* using RNA-Seq. *PLoS One*, *11*(3), e0150273.

Krishnan, N. M., Pattnaik, S., Jain, P., Gaur, P., Choudhary, R., Vaidyanathan, S., et al. (2012). A draft of the genome and four

transcriptomes of a medicinal and pesticidal angiosperm *Azadirachta indica*. *BMC Genomics*, 13(1), 1–13.

Lai, P. K., & Roy, J. (2004). Antimicrobial and chemopreventive properties of herbs and spices. *Current Medicinal Chemistry*, 11(11), 1451–1460.

Li, D. M., Zhao, C. Y., Liu, X. R., Liu, X. F., Lin, Y. J., Liu, J. W., et al. (2015). De novo assembly and characterization of the root transcriptome and development of simple sequence repeat markers in *Paphiopedilum concolor*. *Genetics and Molecular Research*, 14(2), 6189–6201.

Li, C., Zhu, Y., Guo, X., Sun, C., Luo, H., Song, J., et al. (2013). Transcriptome analysis reveals ginsenosides biosynthetic genes, microRNAs and simple sequence repeats in *Panax ginseng* C. A. Meyer. *BMC Genomics*, 14(1), 1–11.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., et al. (2012). Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, 2012(2012), 1–11.

Liu, X. B., Ma, L., Zhang, A. H., Zhang, Y. H., Jiang, J., Ma, W., et al. (2014a). High-throughput analysis and characterization of *Astragalus membranaceus* transcriptome using 454 GS FLX. *PLoS One*, 9(5), e95831.

Liu, H., Wu, W., Hou, K., Chen, J., & Zhao, Z. (2016). Deep sequencing reveals transcriptome re-programming of *Polygonum multiflorum* thunb. roots to the elicitation with methyl jasmonate. *Molecular Genetics and Genomics*, 291(1), 337–348.

Liu, M. M., Xing, Y. M., Zhang, D. W., & Guo, S. X. (2015a). Transcriptome analysis of genes involved in defence response in

Polyporus umbellatus with *Armillaria mellea* infection. *Scientific Reports*, 5, 1–13.

Liu, M. H., Yang, B. R., Cheung, W. F., Yang, K. Y., Zhou, H. F., Kwok, J. S. L., et al. (2015b). Transcriptome analysis of leaves, roots and flowers of *Panax notoginseng* identifies genes involved in ginsenoside and alkaloid biosynthesis. *BMC Genomics*, 16(1), 1–12.

Liu, M. J., Zhao, J., Cai, Q. L., Liu, G. C., Wang, J. R., Zhao, Z. H., et al. (2014b). The complex jujube genome provides insights into fruit tree biology. *Nature Communications*, 5, 1–12.

Lulin, H., Xiao, Y., Pei, S., Wen, T., & Shangqin, H. (2012). The first Illumina-based de novo transcriptome sequencing and analysis of safflower flowers. *PLoS One*, 7(6), e38653.

Ma, C. H., Gao, Z. J., Zhang, J. J., Zhang, W., Shao, J. H., Hai, M. R., et al. (2016). Candidate genes involved in the biosynthesis of triterpenoid saponins in *Platycodon grandiflorum* identified by transcriptome analysis. *Frontiers in plant science*, 7, 1–14.

MacLean, D., Jones, J. D., & Studholme, D. J. (2009). Application of ‘next-generation’ sequencing technologies to microbial genetics. *Nature Reviews Microbiology*, 7(4), 287–296.

Mardis, E. R. (2008). Next-generation DNA sequencing methods. *Annual Reviews of Genomics and Human Genetics*, 9, 387–402.

Mardis, E. R. (2013). Next-generation sequencing platforms. *Annual Review of Analytical Chemistry*, 6, 287–303.

Martínez-García, P. J., Crepeau, M. W., Puiu, D., Gonzalez-Ibeas, D., Whalen, J., Stevens, K. A., et al. (2016). The walnut (*Juglans regia*) genome sequence reveals diversity in genes coding for the biosynthesis

of nonstructural polyphenols. *The Plant Journal*, 87(5), 507–532.

Maxam, A. M., & Gilbert, W. (1977). A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, 74(2), 560–564.

Mazumdar, A. B., & Chattopadhyay, S. (2015). Sequencing, de novo assembly, functional annotation and analysis of *Phyllanthus amarus* leaf transcriptome using the Illumina platform. *Frontiers in plant science*, 6, 1–22.

Meena, S., Kumar, S. R., Rao, D. V., Dwivedi, V., Shilpashree, H. B., Rastogi, S., et al. (2016). De novo sequencing and analysis of lemongrass transcriptome provide first insights into the essential oil biosynthesis of aromatic grasses. *Frontiers in Plant Science*, 7, 1–15.

Metzker, M. L. (2010). Sequencing technologies—the next generation. *Nature Reviews Genetics*, 11(1), 31–46.

Ming, R., VanBuren, R., Liu, Y., Yang, M., Han, Y., Li, L. T., et al. (2013). Genome of the long-living sacred lotus (*Nelumbo nucifera* Gaertn.). *Genome Biology*, 14(5), 1–11.

Muranaka, T., & Saito, K. (2013). Phytochemical genomics on the way. *Plant and Cell Physiology*, 54(5), 645–646.

Muriira, N. G., Xu, W., Muchugi, A., Xu, J., & Liu, A. (2015). De novo sequencing and assembly analysis of transcriptome in the Sodom apple (*Calotropis gigantea*). *BMC Genomics*, 16(1), 1–15.

Park, N. I., Choi, I. Y., Choi, B. S., Kim, Y. S., Lee, M. Y., & Park, S. U. (2014). EST sequencing and gene expression profiling in *Scutellaria baicalensis*. *EXCLI Journal*, 13, 392–401.

- Qi, J., Sun, P., Liao, D., Sun, T., Zhu, J., & Li, X. (2015). Transcriptomic analysis of American ginseng seeds during the dormancy release process by RNA-Seq. *PLoS One*, *10*(3), e0118558.
- Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., et al. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum* domestication and specialization. *Proceedings of the National Academy of Sciences*, *111*(14), 5135–5140.
- Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., et al. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, *13*(1), 1–13.
- Rai, A., Kamochi, H., Suzuki, H., Nakamura, M., Takahashi, H., Hatada, T., et al. (2016). De novo transcriptome assembly and characterization of nine tissues of *Lonicera japonica* to identify potential candidate genes involved in chlorogenic acid, luteolosides, and secoiridoid biosynthesis pathways. *Journal of Natural Medicines*, 1–15.
- Ramilowski, J. A., Sawai, S., Seki, H., Mochida, K., Yoshida, T., Sakurai, T., et al. (2013). *Glycyrrhiza uralensis* transcriptome landscape and study of phytochemicals. *Plant and Cell Physiology*, *54*(5), 697–710.
- Rastogi, S., Meena, S., Bhattacharya, A., Ghosh, S., Shukla, R. K., Sangwan, N. S., et al. (2014). De novo sequencing and comparative analysis of holy and sweet basil transcriptomes. *BMC Genomics*, *15*(1), 1–18.
- Reddy, N. R. R., Mehta, R. H., Soni, P. H., Makasana, J., Gajbhiye, N. A., Ponnuchamy, M., et al. (2015). Next generation sequencing and transcriptome analysis predicts biosynthetic pathway of sennosides

from Senna (*Cassia angustifolia* Vahl.), a non-model plant with potent laxative properties. *PLoS One*, 10(6), e0129422.

Rusk, N. (2011). Torrents of sequence. *Nature Methods*, 8(1), 44.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463–5467.

Sangwan, R. S., Tripathi, S., Singh, J., Narnoliya, L. K., & Sangwan, N. S. (2013). 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*, 525(1), 58–76.

Schadt, E. E., Turner, S., & Kasarskis, A. (2010). A window into third-generation sequencing. *Human Molecular Genetics*, 19(R2), R227–R240.

Schneider, G. F., & Dekker, C. (2012). DNA sequencing with nanopores. *Nature Biotechnology*, 30(4), 326–328.

Sharma, S., & Shrivastava, N. (2016). Renaissance in phytomedicines: promising implications of NGS technologies. *Planta*, 244(1), 19–38.

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, 26(10), 1135–1145.

Shi, S. G., Yang, M., Zhang, M., Wang, P., Kang, Y. X., & Liu, J. J. (2014). Genome-wide transcriptome analysis of genes involved in flavonoid biosynthesis between red and white strains of *Magnolia sprengeri* pamp. *BMC Genomics*, 15(1), 1–11.

Singh, V., Mörsel, J. T., Singh, V., Yang, B., Kallio, H., Bala, M., et al. (2006). Development and commercialization of seabuckthorn: a

German experience. Seabuckthorn (*Hippophae* L.). *A multipurpose wonder plant. Vol II: Biochemistry and Pharmacology*, 576–584.

Singh, R., Ong-Abdullah, M., Low, E. T. L., Manaf, M. A. A., Rosli, R., Nookiah, R., et al. (2013). Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature*, 500(7462), 335–339.

Subramaniam, S., Mathiyalagan, R., Natarajan, S., Kim, Y. J., Jang, M. G., Park, J. H., et al. (2014). Transcript expression profiling for adventitious roots of *Panax ginseng* Meyer. *Gene*, 546(1), 89–96.

Sumner, L. W., Lei, Z., Nikolau, B. J., & Saito, K. (2015). Modern plant metabolomics: Advanced natural product gene discoveries, improved technologies, and future prospects. *Natural Product Reports*, 32(2), 212–229.

Sun, C., Li, Y., Wu, Q., Luo, H., Sun, Y., Song, J., et al. (2010). 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(1), 1–12.

Sun, P., Song, S., Zhou, L., Zhang, B., Qi, J., & Li, X. (2012a). Transcriptome analysis reveals putative genes involved in iridoid biosynthesis in *Rehmannia glutinosa*. *International Journal of Molecular Sciences*, 13(10), 13748–13763.

Sun, C., Yu, G., Bao, M., Zheng, B., & Ning, G. (2014). Biological pattern and transcriptomic exploration and phylogenetic analysis in the odd floral architecture tree: *Helwingia willd.* *BMC Research Notes*, 7(1), 1–12.

Sun, X., Zhou, S., Meng, F., & Liu, S. (2012b). *De novo* assembly and

characterization of the garlic (*Allium sativum*) bud transcriptome by Illumina sequencing. *Plant Cell Reports*, 31(10), 1823–1828.

Suryakumar, G., & Gupta, A. (2011). Medicinal and therapeutic potential of Sea buckthorn (*Hippophae rhamnoides* L.). *Journal of Ethnopharmacology*, 138(2), 268–278.

Tao, L., Zhao, Y., Wu, Y., Wang, Q., Yuan, H., Zhao, L., et al. (2016). Transcriptome profiling and digital gene expression by deep sequencing in early somatic embryogenesis of endangered medicinal *Eleutherococcus senticosus* Maxim. *Gene*, 578(1), 17–24.

Thompson, J. F., & Milos, P. M. (2011). The properties and applications of single-molecule DNA sequencing. *Genome Biology*, 12(2), 1–10.

Tian, H., Xu, X., Zhang, F., Wang, Y., Guo, S., Qin, X., et al. (2015). Analysis of *Polygala tenuifolia* transcriptome and description of secondary metabolite biosynthetic pathways by Illumina sequencing. *International Journal of Genomics*, 2015, 1–11.

Unamba, C. I., Nag, A., & Sharma, R. K. (2015). Next generation sequencing technologies: The doorway to the unexplored genomics of non-model plants. *Frontiers in Plant Science*, 6, 1–16.

Upadhyay, A. K., Chacko, A. R., Gandhimathi, A., Ghosh, P., Harini, K., Joseph, A. P., et al. (2015). Genome sequencing of herb Tulsi (*Ocimum tenuiflorum*) unravels key genes behind its strong medicinal properties. *BMC Plant Biology*, 15(1), 1–20.

Upadhyay, S., Phukan, U. J., Mishra, S., & Shukla, R. K. (2014). De novo leaf and root transcriptome analysis identified novel genes involved in steroidal sapogenin biosynthesis in *Asparagus racemosus*. *BMC Genomics*, 15(1), 1–13.

van Bakel, H., Stout, J. M., Cote, A. G., Tallon, C. M., Sharpe, A. G., Hughes, T. R., et al. (2011). The draft genome and transcriptome of *Cannabis sativa*. *Genome Biology*, *12*(10), 1–13.

Van Moerkercke, A., Fabris, M., Pollier, J., Baart, G. J., Rombauts, S., Hasnain, G., et al. (2013). CathaCyc, a metabolic pathway database built from *Catharanthus roseus* RNA-Seq data. *Plant and Cell Physiology*, *54*(5), 673–685.

Wang, G., Du, X., Ji, J., Guan, C., Li, Z., & Josine, T. L. (2015a). De novo characterization of the *Lycium chinense* Mill. leaf transcriptome and analysis of candidate genes involved in carotenoid biosynthesis. *Gene*, *555*(2), 458–463.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57–63.

Wang, Y., Hua, W., Wang, J., Hannoufa, A., Xu, Z., & Wang, Z. (2013). Deep sequencing of *Lotus corniculatus* L. reveals key enzymes and potential transcription factors related to the flavonoid biosynthesis pathway. *Molecular Genetics and Genomics*, *288*(3–4), 131–139.

Wang, L., Wang, Z., Chen, J., Liu, C., Zhu, W., Wang, L., et al. (2015b). De novo transcriptome assembly and development of novel microsatellite markers for the traditional chinese medicinal herb, *Veratilla baillonii* Franch (Gentianaceae). *Evolutionary Bioinformatics Online*, *11*(39), 1–7.

Wu, D., Austin, R. S., Zhou, S., & Brown, D. (2013). The root transcriptome for North American ginseng assembled and profiled across seasonal development. *BMC Genomics*, *14*(1), 1–14.

Wu, B., Suo, F., Lei, W., & Gu, L. (2014). Comprehensive analysis of

alternative splicing in *Digitalis purpurea* by strand-specific RNA-Seq. *PLoS One*, 9(8), e106001.

Wu, B., Wang, M., Ma, Y., Yuan, L., & Lu, S. (2012). High-throughput sequencing and characterization of the small RNA transcriptome reveal features of novel and conserved microRNAs in *Panax ginseng*. *PLoS One*, 7(9), e44385.

Xu, Z., Peters, R. J., Weirather, J., Luo, H., Liao, B., Zhang, X., et al. (2015). Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis. *The Plant Journal*, 82(6), 951–961.

Yang, L., Ding, G., Lin, H., Cheng, H., Kong, Y., Wei, Y., et al. (2013). Transcriptome analysis of medicinal plant *Salvia miltiorrhiza* and identification of genes related to tanshinone biosynthesis. *PLoS One*, 8(11), e80464.

Yang, Y. H., Li, M. J., Li, X. Y., Chen, X. J., Lin, W. X., & Zhang, Z. Y. (2015). Transcriptome-wide identification of the genes responding to replanting disease in *Rehmannia glutinosa* L. roots. *Molecular Biology Reports*, 42(5), 881–892.

Yuan, Y., Song, L., Li, M., Liu, G., Chu, Y., Ma, L., et al. (2012). Genetic variation and metabolic pathway intricacy govern the active compound content and quality of the Chinese medicinal plant *Lonicera japonica* thunb. *BMC Genomics*, 13(1), 1–20.

Zhang, X., Allan, A. C., Li, C., Wang, Y., & Yao, Q. (2015a). De novo assembly and characterization of the transcriptome of the Chinese medicinal herb, *Gentiana rigescens*. *International Journal of Molecular Sciences*, 16(5), 11550–11573.

Zhang, Q., Chen, W., Sun, L., Zhao, F., Huang, B., Yang, W., et al. (2012). The genome of *Prunus mume*. *Nature Communications*, 3, 1318–1325.

Zhang, F., Gao, Q., Khan, G., Luo, K., & Chen, S. (2014). Comparative transcriptome analysis of aboveground and underground tissues of *Rhodiola algida*, an important ethno-medicinal herb endemic to the Qinghai-Tibetan Plateau. *Gene*, 553(2), 90–97.

Zhang, G. H., Jiang, N. H., Song, W. L., Ma, C. H., Yang, S. C., & Chen, J. W. (2016a). De novo sequencing and transcriptome analysis of *Pinellia ternata* identify the candidate genes involved in the biosynthesis of benzoic acid and ephedrine. *Frontiers in Plant Science*, 7, 1–14.

Zhang, S., Wu, Y., Jin, J., Hu, B., Zeng, W., Zhu, W., et al. (2015b). De novo characterization of *Panax japonicus* C. A. Mey transcriptome and genes related to triterpenoid saponin biosynthesis. *Biochemical and Biophysical Research Communications*, 466(3), 450–455.

Zhang, G. Q., Xu, Q., Bian, C., Tsai, W. C., Yeh, C. M., Liu, K. W., et al. (2016b). The *Dendrobium catenatum* Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. *Scientific Reports*, 6, 1–10.

Zhao, S., Tuan, P. A., Li, X., Kim, Y. B., Kim, H., Park, C. G., et al. (2013). Identification of phenylpropanoid biosynthetic genes and phenylpropanoid accumulation by transcriptome analysis of *Lycium chinense*. *BMC Genomics*, 14(1), 802–812.

Zheng, X., Pan, C., Diao, Y., You, Y., Yang, C., & Hu, Z. (2013). Development of microsatellite markers by transcriptome sequencing in two species of *Amorphophallus* (Araceae). *BMC Genomics*, 14(1), 1–11.