J. Agr. Sci. Tech. (2016) Vol. 18: 437-452

De Novo Characterization of the Root Transcriptome and Development of EST-SSR Markers in *Paris polyphylla* Smith var. *yunnanensis,* an Endangered Medical Plant

L. Wang[1], Y. Yang[2], Y. Zhao[2], S. Yang[1], S. Udikeri[3], and T. Liu[1*]

ABSTRACT

***Paris polyphylla* Smith var. *yunnanensis* (Liliaceae) is an important traditional medicinal plant of the Yunnan Province in China. However, the genomic information regarding this plant is limited. To further understand its molecular background, we conducted Illumina HiSeq 2000 second-generation sequencing of this plant species. Approximately 30,198,679 reads with an average length of 202 bases were obtained from its root cells. These reads were assembled into 56,095 unique sequences and approximately 49.7% of the unique sequences were annotated by Basic Local Alignment Search Tool (BLAST) similarity searches against public sequence databases. Most of these unigenes were mapped to carbohydrate metabolism, energy metabolism, and secondary metabolite biosynthetic pathways. Additionally, 3,853 EST-SSRs were identified as potential molecular markers in our unigenes. Of these, 9 nuclear SSR markers were employed to assess genetic diversity and structure of 11 geographically disjunct populations. The present study revealed a moderate genetic diversity (He= 0.527) and low genetic differentiation (Fst= 0.103), which may be ascribed to an earlier period of more pronounced gene flow when the species had a more continuous distribution. The 11 studied populations were divided into two clusters based on the UPGMA dendrogram, which were not congruent with their geographical distributions. Overall, the root transcriptome sequences generated in this study reveal novel gene expression profiles and offer important clues for further study of the molecular mechanism of *Paris*' root secondary metabolite synthesis and population genetics. The EST-SSR markers identified will also facilitate marker-assisted selection in *Paris* breeding.**

Keywords: HiSeq second-generation sequence, Saponin, Simple sequence repeat.

INTRODUCTION

P. polyphylla is a highly valued medicinal plant in Asiatic countries especially in China, India, and Nepal. The plant is used much more in China than other countries as traditional medicine as well as contemporary therapy. It is mainly distributed in southwestern China, particularly in the Yunnan and Sichuan provinces (Zhang, 2007). In India, it is grown in Manipur, Uttarakhand, Himachal Pradesh, and in Lushai and Aka Hills (Tiwari *et al.,* 2010). The rhizome of *P. polyphylla* has been used in traditional Chinese medicine for the treatment of various inflammations and injuries (Zhao *et al*., 2010) and is an important ingredient of particular Chinese patent medicines such as ''Biyan Qingdu Keli'', which is widely used in southern China for the treatment of chronic rhinitis

[1] Yunnan Research Center on Good Agricultural Practice for Dominant Chinese Medicinal Materials, Yunnan Agricultural University, Kunming, 650201, People's Republic of China.
*Corresponding author; e-mail: yantao618@126.com
[2] China Tobacco Yunnan Industrial Co., Ltd, Kunming 650201. People's Republic of China.
[3] University of Agricultural Sciences, Dharwad, Pin-580005, Karnataka, India.

[Downloaded from jast.modares.ac.ir on 2024-04-18]

[DOR: 20.1001.1.16807073.2016.18.2.7.7]

and nasopharyngeal cancer (Guo *et al*., 2006). The main active ingredients of the plant are steroidal saponins (Zhang, 2007), with at least 30 steroidal saponins isolated through phytochemical methods (Liu *et al*., 2006; Xu *et al*., 2007). Therefore, understanding the processes regulating *Paris* root secondary metabolite production is of particular importance.

Steroidal saponins are synthesized via the Mevalonic Acid (MVA) pathway in cytoplasm (Haralampidis *et al*., 2002), or through non-Mevalonate Pathway (MEP) in plastids (Rohmer, 2003). Cyclization of 2, 3-oxidosqualene leads to the formation of cycloartenol, which is catalyzed by OxidoSqualene Cyclase (OSC). Then, some specific CYP450s and UDP-GlycosylTransferases (UGTs) may catalyze the conversion of cycloartenol to various steroidal saponins (Kumar *et al*., 2012). Until now, several OSC genes have been cloned from various plant systems (Corey *et al*., 1993; Herrera *et al*., 1998), however, little is known about the molecular mechanism of the biosynthetic pathway downstream of cyclization. Despite its pharmacological importance, the very limited information on the transcriptome and genome of *P. polyphylla* greatly hinders investigations on the mechanism of steroidal saponin biosynthesis.

Simple Sequence Repeat (SSR) has been widely used in the study of genetic identification and fingerprint mapping with the characteristics of high polymorphic information content, simple technology, and good reproducibility (Abbasi *et al.*, 2014; Zhang *et al*., 2014; Ghaedrahmati *et al.*, 2014). EST collections can also contribute to the development of molecular markers for further genetic research on *Paris* species. This Paris species are distributed in the areas of tropical and temperate Eurasia (Ji *et al*., 2006; Li, 1998). Molecular genetic studies have been few in number and no simple sequence repeats (SSRs) have been reported. To optimize the conservation and utilization of *P. polyphylla,* the development of Expressed Sequence Tag (EST–SSR) markers is very useful for germplasm identification and research into the genetic diversity of this species.

Next-Generation Sequencing (NGS) technologies such as pyrosequencing circumvent lengthy and relatively low-throughput steps associated with Sanger sequencing and provide rapid and economical technologies for transcriptomics (Margulies *et al*., 2005; Chi, 2008; Mardis, 2008; Morozova and Marra, 2008; Wang *et al*., 2010; Wang *et al*., 2009). During the last decade, a large number of transcriptomic sequences have been generated and collected in model and non-model organisms, which have greatly accelerated the understanding of the complexity of gene expression, regulation and networks in higher plants. Furthermore, the large number of Expressed Sequence Tags (ESTs) generated from transcriptome sequencing have provided valuable genetic resources for functional genomics and molecular marker development. However, to date, there is only one paper about the study on embryo transcriptome sequencing of *P. polyphylla'* seeds, and no reports on its root transcriptome sequencing is reported. This is far from enough for genomic study and functional gene identification in it.

In the present study, we utilized Illumina HiSeq 2000 second-generation sequencing technology to characterize the root transcriptome of *Paris* and to develop EST-SSR markers. We developed and characterized 8 novel polymorphic EST-SSR markers for this species. To the best of our knowledge, this study is the first to profile the root transcriptome of *Paris* through the analysis of large-scale transcript sequences and generate a few EST-SSR marker resources for further study. These EST-SSR markers provide an important tool for the study of genetic diversity in *P. polyphylla*. Also, our study could make it possible to construct high density microarrays for further characterization of gene expression profiles during secondary metabolite production in the future.

[Downloaded from jast.modares.ac.ir on 2024-04-18]

[DOR: 20.1001.1.16807073.2016.18.2.7.7]

MATERIALS AND METHODS

Plant Material

Nine-year-old *P. polyphylla* plants cultivated on farms are routinely harvested for medical purposes. The *P. polyphylla* plants were collected from the fields of Kunming City, Yunnan Province, China. After cleaning, the roots were cut into small pieces, immediately frozen in liquid nitrogen, and stored at -80°C until further processing. Total RNA was isolated from each sample using RNAiso Plus (TaKaRa). RNA quality was initially characterized on an agarose gel and NanoDrop ND1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and then further assessed by RNA Integrity Number (RIN) value (> 8.0) using an Agilent 2100 Bioanalyzer (Santa Clara, CA, USA). In addition, based on an extensive field survey, we collected 11 indigenous accessions of *P. polyphylla* from different geographical populations (Table 1) for the amplification validation of designed EST-SSR primers. Five samples from each accession were used to screen polymorphic EST-SSR markers and investigate the genetic relatedness of *P. polyphylla* among the accessions. These germplasms represented a relatively broad genetic diversity. Genomic DNA of each individual was isolated and purified from silica gel-dried using the method of Doyle and Doyle (1987).

Library Preparation and Sequencing

The cDNA libraries were prepared according to the manufacturer's instructions (Illumina, San Diego, CA). Poly-A mRNA molecules were purified using Sera-mag Magnetic Oligo (dT) Beads (Illumina) from 20 µg of total RNA from each sample and eluted with 10 mM of Tris–HCl. To avoid priming bias during cDNA synthesis, the purified mRNA was first fragmented into small pieces using RNA fragmentation reagents (Ambion, Austin, TX, USA) before cDNA synthesis. Double-stranded cDNA was generated from cleaved mRNA fragments using random hexamer primers (Illumina) and a SuperScript Double-Stranded cDNA Synthesis kit (Invitrogen, Camarillo, CA). The resulting cDNAs were purified using a QIAQuick PCR Purification Kit (Qiagen, Valencia, CA) and then subjected to end-repair and phosphorylation using T4 DNA polymerase, Klenow DNA polymerase, and T4 PNK (NEB, Ipswich, MA, USA). Repaired cDNA fragments were 3′ adenylated using Klenow Exo- (Illumina), producing cDNA fragments with a single 'A' base overhang at their 3′ ends for subsequent adapter ligation. Illumina paired-end adapters were ligated to the ends of these 3′ adenylated cDNA fragments. To select a size range of templates for downstream enrichment, the products of the

Table 1. List of *P. polyphylla* accessions used to determine genetic diversity.

No	Accession name	Origin	Longitude/Latitude	Altitude (m)
1	YNND0001	Qiaojia, Yunnan	103.06/27.65	2400
2	YNND0002	Xuanwei, Yunnan	104.47/26.29	2153
3	YNND0003	Luquan, Yunnan	102.54/26.13	2721
4	YNND0004	Weixi, Yunnan	99.38/27.13	2517
5	YNND0005	Longli, Yunnan	108.71/27.89	1178
6	YNND0006	Yongsheng, Yunnan	100.82/26.70	2595
7	YNND0007	Funing, Yunnan	105.47/23.52	1456
8	YNND0008	Zhenxiong, Yunnan	110.67/26.46	2482
9	YNND0009	Xichang, Sichuan	102.31/27.72	2476
10	YNND0010	Diqing, Yunnan	99.23/28.22	3052
11	YNND0011	Yongping, Yunnan	99.69/26.09	2285

[Downloaded from jast.modares.ac.ir on 2024-04-18]

[DOR: 20.1001.1.16807073.2016.18.2.7.7]

ligation reaction were purified on a 2% TAE-agarose gel (Certified Low-Range Ultra Agarose, BioRad, Hercules, CA). A specific range of cDNA fragments (200±25 bp) was excised from the gel and extracted using QIAquick Gel Extraction Kit (Qiagen). Fifteen rounds of PCR amplification were performed to enrich the purified cDNA template using primers complementary to the ends of the adapters [PCR Primer PE 1.0 and PCR Primer PE 2.0 (Illumina) with Phusion DNA Polymerase]. Finally, after validating on an Agilent Technologies 2100 Bioanalyzer using the Agilent DNA 1000 chip kit, the cDNA library products were sequenced on a paired-end flow cell using an Illumina Genome Analyzer II at Beijing Genomics Institute (BGI), China.

Data Processing and *De Novo* Assembly

Because the algorithms used in the *de novo* transcriptome construction of the short reads provided by the Illumina platform might be severely inhibited by sequencing errors, a stringent cDNA sequence filtering process was employed to select clean reads. First, Illumina's Failed-Chastity filter software was used to remove raw reads that fell into the relation "Failed-chastity≤ 1", using a chastity threshold of 0.6 on the first 25 cycles. Second, all raw reads showing signs of adaptor contamination or ambiguous trace peaks (denoted by an "N" in the sequence trace) were removed. Finally, raw reads showing more than 10% of bases with a Phred-scaled probability (Q) of < 20 were discarded.

The resulting clean short reads that showed sufficient overlap with other reads were joined using the SOAPdenovo software to generate longer, contiguous sequences (i.e., contigs). Contigs were rejected unless their K-mers were conjoined along an unambiguous path. The identities of the contigs generated from a transcript and their distances from each other were established by mapping clean reads back to the corresponding contigs based on their paired-end information. Joining of these contigs and filling of the unknown interspaces (i.e., gaps) using "Ns" (i.e., ambiguous base calls) resulted in the generation of scaffolds. Finally, the gaps of scaffolds were filled using the paired-end clean reads according to their sequence complementarity to scaffolds, resulting in sequences with the fewest "Ns" that also could not be further extended on either end, or unigenes. To obtain distinct sequences, the unigenes from the two different phases were clustered using the TGI Clustering tool.

Unigenes were then aligned to a series of protein databases using BLASTx (E-value< 10^{-5}). Databases included the NCBI Non-redundant Protein (Nr), Swiss-Prot, the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway (Kanehisa and Goto, 2000), and the Cluster of Orthologous Groups of proteins (COG) (http://www.ncbi.nlm.nih.gov/COG/) databases. Sequence directionality was assigned according to the best alignments. When these databases generated different results, the following priority structure was used in selecting one unigene: NCBI Nr, Swiss-Prot, KEGG, and COG. When a unigene failed to align to any of the four databases, ESTScan (Iseli *et al*., 1999) was used to predict its coding regions and ascertain its sequence direction.

Primer Designing, PCR Amplification, and Visualization of SSR Loci

The MISA Perl program (Thiel *et al*., 2003) was used to identify EST-SSRs in the unigenes. In this study, EST-SSRs were defined as regions with two- to six-nucleotide motifs with at least five repetitions and a minimum length of 10 bp. BatchPrimer3 (Frank *et al*., 2008) was used to design PCR primers in the flanking regions of the SSRs. The parameters were as follows: product length, 100–500 bp; primer size, 18–24 bp (optimum, 20 bp); and melting temperature between 40 and 60°C.

[Downloaded from jast.modares.ac.ir on 2024-04-18]

[DOR: 20.1001.1.16807073.2016.18.2.7.7]

Polymerase Chain Reaction (PCR) was performed in a 10-μL reaction volume containing 1 μL genomic DNA (10 ng μL^{-1}), 1X PCR master mix (TransGen Biotech, Beijing, China), and 0.3 μL of both forward and reverse primers (10 pmol μL^{-1}) using an Eppendorf master cycler. The PCR conditions used were as follows: initial denaturation for 5 minutes at 95°C, followed by 30 cycles of denaturation for 30 seconds at 94°C, annealing for 45 seconds at 57–63°C (primer specific), and extension for 90 seconds at 72°C. After PCR amplification was confirmed on 1.5% agarose gel, PCR products were electrophoresed and separated on 6% polyacrylamide gels (acrylamide/bis-acrylamide, 29:1). The sizes of PCR products on polyacrylamide gels were visualized by silver staining.

Data Acquisition and Analysis

The polymorphic SSR loci were analyzed with POPGENE version 1.32 software (Yeh *et al*., 1999) for the Number of alleles per locus (A), expected Heterozygosity (He), and Fixation index (F_{IS}). The Polymorphism Information Content (PIC) was calculated using Power Marker (Liu and Muse, 2005). In addition, in order to test for a correlation between Nei's genetic distance and geographical distances (in kilometers) between populations, a Mantel test was performed using tools for population genetic analysis (Miller, 1997) (999 permutations were calculated).

RESULTS

Transcriptome Sequencing and Sequence Assembly

Using the latest HiSeq 2000 platform, the cDNA library of 9-year old *P. polyphylla* roots produced 30,198,679 reads representing a total of 6,100,133,158 (6.10 Gb) nucleotides, with an average length of 202 bp. The average read size, CycleQ20 percentage, and GC percentage were 201 bp, 100, and 50.46%, respectively. The short reads were assembled into 2,510,576 contigs with a mean length of 58 bp and a contig N50 of 62 bp. From these contigs, 114,941 scaffolds were built using SOAPdenovo, with a mean length of 833 bp and an N50 of 1,321 bp. Because all the annotations and bioinformatic analyses in this study were based on unigenes, the N50 sizes of the contigs and scaffolds did not significantly influence the analysis. The results of an assembly are influenced by the assembly software used, as well as by the sequencing depth utilized in the analysis and, thus, the more the sequencing data available, the longer the assembled contigs. A total of 56,095 unigenes with a mean length of 573 bp and an N50 size of 823 bp were generated from the analysis (Additional file 1). Most unigenes ranged from 200 to 2,000 bp in length, with 23,011 (41.02%) sequences containing 200–300 bp, 15,746 (28.07%) containing 300–500 bp, 9,296 (16.57%) containing 500–1000 bp, and 5,908 (10.53%) containing 1,000–2,000 bp. However, there were also 2,134 (3.80%) unigene sequences longer than 2,000 bp. An overview of the sequencing and assembly of *P. polyphylla* is summarized in Table 2.

GO Analysis and KEGG Assignment

The GO annotation describes gene products according to their associated molecular functions, cellular components, and biological processes, illustrating the broad overview of the groups of genes catalogued in the transcriptome (Berardini *et al*., 2004). In this study, plant-specific GO slim terms associated with 20,766 (37.2%) of the 56,095 assembled EST unigenes were available based on sequence similarity to proteins in the TAIR database. A total of 20,766 unigenes were assigned to 45 functional groups using GO assignments (Figure 1). Within each of the three main categories of the GO classification scheme (biological process, cellular component, and

[Downloaded from jast.modares.ac.ir on 2024-04-18]

[DOR: 20.1001.1.16807073.2016.18.2.7.7]

Table 2. Statistical summary of cDNA sequences of *P. polyphylla* generated by the HiSeq 2000 platform.

Length range	Total number (Percentage)	
	T1 transcripts	All unigenes
200-300	29985 (26.09%)	23011 (41.02%)
300-500	25649 (22.31%)	15746 (28.07)
500-1000	26173 (22.77%)	9296 (16.57)
1000-2000	23463 (20.41%)	5908 (10.53%)
2000+	9671 (8.41%)	2134 (3.80%)
Total number	114941	56095
Total length	95850462	32172750
N50 length	1321	823
Mean length	8339101104	573.5404225

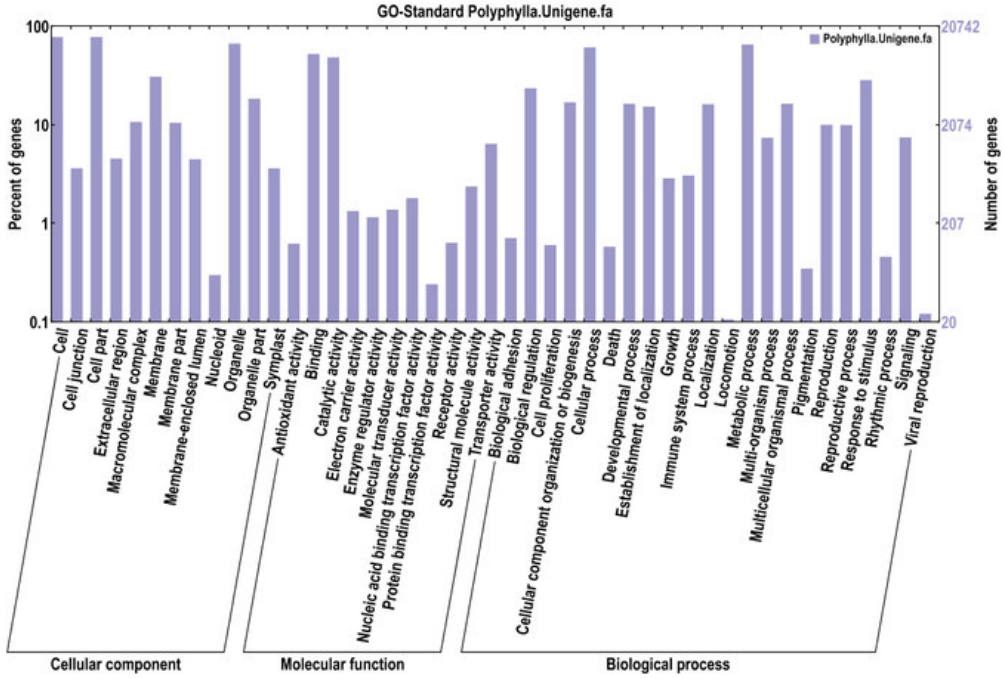

Figure 1. GO analysis of *P. polyphylla* unique sequences based on cellular component, molecular function, and biological process.

molecular function), the dominant subcategories were "metabolic process", "cell part", and "binding", respectively. "Cellular process", "catalytic activity", "organelle", and "cell part" were also well represented. However, only a few genes were assigned to the category "viral reproduction", and almost no genes were found in the "locomotion" cluster. The GO categories represented in the *P. polyphylla* root transcriptome did not show any significant biases and showed similar distribution patterns reported in other plant species (Luo *et al*., 2011).

Out of 56,095 hits in the public databases, 7,309 sequences were classified into 25 Clusters of Orthologous Groups (COG) categories (Figure 2), among which "general function prediction only" represented the largest group (1,582; 21.6%), followed by "replication, recombination, and repair" (1,088; 14.9%), "transcription" (926; 12.7%), "translation, ribosomal structure, and biogenesis" (798; 10.9%), and "posttranslational modification, protein

[Downloaded from jast.modares.ac.ir on 2024-04-18]

[DOR: 20.1001.1.16807073.2016.18.2.7.7]

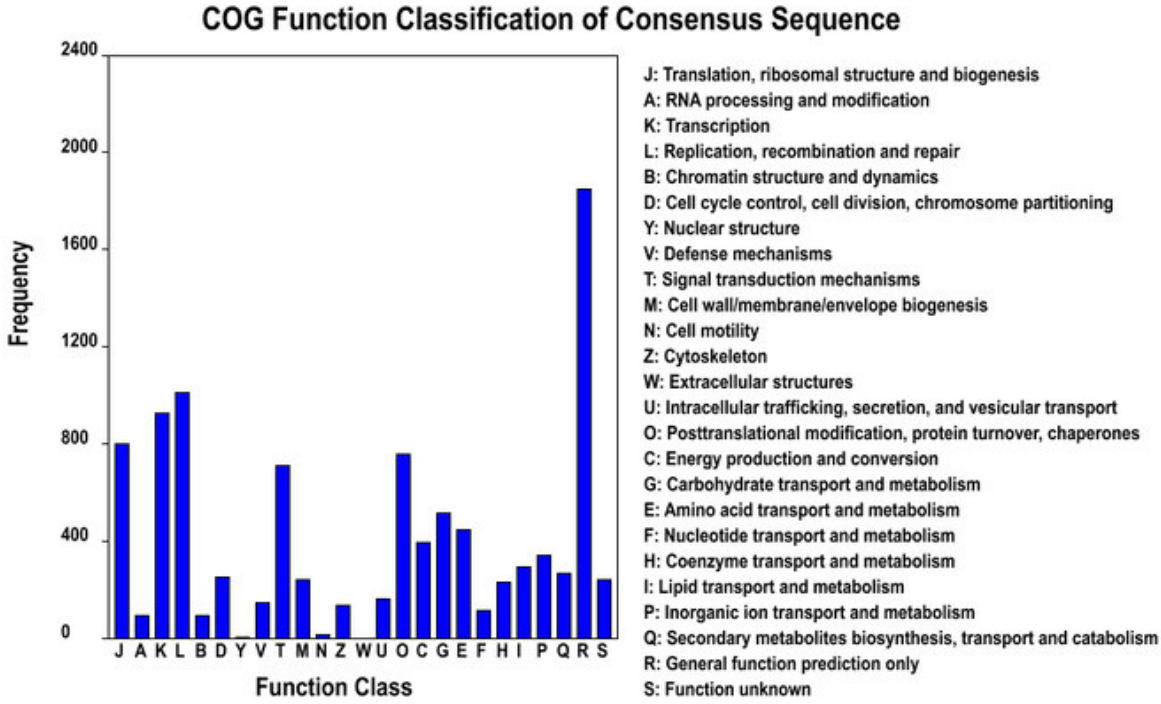

Figure 2. Histogram of Clusters of Orthologous Groups (COG) classification.

turnover, chaperones" (756; 10.3%). "nuclear structures" (2; 0.00%) and "extracellular structures" (0; 0.00%) were the smallest groups. In addition, 268 (3.6%) unigenes were classified as having secondary metabolites biosynthesis, transport, and catabolism functions (Figure 2).

The KEGG assignments provide an alternative functional annotation of genes associated with biochemical pathways with their corresponding Enzyme Commission (EC) numbers (Kanehisa and Goto, 2000). To identify the biological pathways that were active in *P. polyphylla*, we mapped the 56,095 annotated sequences to the reference canonical pathways in KEGG. A total of 5,831 unigene sequences were assigned to 287 KEGG pathways, which included categories such as carbohydrate metabolism, energy metabolism, amino acid metabolism, the biosynthesis of secondary metabolites, and lipid metabolism. Many active metabolic processes such as purine metabolism and pyrimidine metabolism also occurred in root, covering a large number of unigenes (161, and 128, respectively).

The results of our sequence similarity searches against public databases such as SwissProt, KEGG, COG, GO, TrEMBL, NCBI Non-redundant Protein (Nr), and NCBI Non-redundant Nucleotide (Nt) databases using BLAST, and the annotated unique sequences are summarized as an additional file (available upon request). As expected, high percentages were found in the Nr, Nt, SwissProt, TrEMBL, GO, and COG databases; 26,655 (47.52%), 17,072 (30.43%), 18,931 (33.75%), 26,839 (47.85%) and 20,766 (37.02%) unigenes showed significant similarity to known proteins in the respective databases, indicating that the sequencing method in this study recovered a substantial fraction of *Paris* root genes. In contrast, only 5,831 (10.39%) and 7,309 (13.03%) unigenes had BLAST hits in the KEGG and COG, respectively, showed a low ratio to known proteins in the respective databases. Altogether, sequence similarity searches of the above eight public databases found that

[Downloaded from jast.modares.ac.ir on 2024-04-18]

[DOR: 20.1001.1.16807073.2016.18.2.7.7]

27,904 unigenes could be annotated with gene descriptions or conserved protein domains, accounting for 49.74% of all unique sequences. In addition, our result showed that 86.25% of unigenes over 600 bp in length had BLAST matches, compared to 44.64% of unigenes ranging from 200 to 400 bp and 13.21% of unigenes shorter than 200 bp, which is just in agreement with the opinion that longer contigs were more likely than shorter ones to have BLAST matches in the protein databases (Zhou *et al*., 2009).

SSR Detection

SSRs are the most effective genetic markers for plant breeding and genetic applications (Sharma *et al*., 2009). The assembled *Paris* unigenes with annotations were used for identifying SSRs. This analysis using the 8,042 assembled unigenes with annotations identified 2,849 unigenes that contained SSRs between 2–6 nucleotides in length using the MISA program, in which a total of 3,853 putative SSR motifs were identified. Among the SSR-containing unique sequences, the majority (2,108; 73.99%) had a single SSR motif in every sequence, whereas the rest contained more than 1 SSR. The frequency of SSRs is shown in Table 3. These motifs included di-, tri-, tetra-, penta-, and hexa-nucleotides, with lengths ranging from 2 to 6 bp, which was similar to the EST-SSRs reported in other dicotyledonous species (Luo *et al*., 2011). The dinucleotide repeats were the most abundant (42.4%), which is consistent with other findings, including that of *Panax notoginseng* (Luo *et al*., 2011). Mononucleotide repeats (1,612) were the second most common SSRs, followed by trinucleotide (594), tetranucleotide (8), hexanucleotide (3), and pentanucleotide (2) repeats. Among the dinucleotide repeats, 6–11 repeat units were the most common, while 5–6 repeat units were the most common for trinucleotide repeats. SSRs are the most feasible genetic markers for plant breeding and genetic applications (Sharma *et al*., 2009). The unique sequence-derived markers generated in this study represent a valuable genetic resource for future studies of this species, as well as related *Paris* species (Additional file 3).

DISCUSSION

Unique Sequence Annotation and Highly Expressed Transcript Analyses

Previous studies have shown that approximately 87% of *Arabidopsis* 454-derived ESTs could be aligned to predicted genes (Weber *et al*., 2007), whereas 72% could be identified based on homology in cucumber (Guo *et al*., 2010) and 70.2% in *Panax notoginseng* root using the RefSeq database of highly curated genes (Luo *et al*., 2011). Although the annotation rate of *P. polyphylla* unique sequences is lower than that of *P. notoginseng* root (70.2%) and *P. ginseng* root (63.6%) (Chen *et al*., 2011) transcriptomes, this study succeeded in identifying putative transcripts of *P. polyphylla* root, which earlier had limited genomic information. In fact, "non-BLASTable" sequences have been reported in all studied plant transcriptomes, with the proportion varying from 13% to 80%, depending on the species, the sequencing depth, and the parameters of the BLAST search (Wang *et al*., 2010; Parchman *et al*., 2010). Except for the technical issues associated with sequencing, biological factors may be responsible for the large population of non-BLASTable sequences, including rapidly evolving genes (orthologs that are so highly divergent that efficient recognition is precluded), species-specific genes (present in the studied species but absent from the databases), and the persistence of non-coding fractions mainly from untranslated regions of the sampled transcripts.

Through the method of sequence similarity searches against public databases, including SwissProt, KEGG, the Arabidopsis Information Resource (TAIR), NCBI Non-

[Downloaded from jast.modares.ac.ir on 2024-04-18]

[DOR: 20.1001.1.16807073.2016.18.2.7.7]

redundant Protein (Nr), and NCBI Non-redundant Nucleotide (Nt) database, and the annotated unique sequences were summarized in Additional file 2. From the file, we can see that 27,904 (49.7%) *P. polyphylla* unique sequences were annotated and the remaining (50.3%) unique sequences had no match to any sequence in the public databases. Previous studies have shown that approximately 87% of Arabidopsis 454-derived ESTs could be aligned to predicted genes (Weber *et al*., 2007) and 70.2% in *Panax notoginseng* root using the RefSeq database of well-annotated genes (Luo *et al*., 2011). Though the annotation rate for *P. polyphylla* unique sequences is lower than *P. notoginseng* root (70.2%) and *P. ginseng* root (63.6%) (Chen *et al*., 2011) transcriptomes, this study succeeded in assigning putative identification to a significant proportion of the discovered *P. polyphylla* root transcripts given the lack of genomic information for this species. In fact, "non-BLASTable" sequences have been reported in all studied plant transcriptomes, with the proportion varying from 13 to 80%, depending on the species, the sequencing depth and the parameters of the BLAST search (Wang *et al*., 2010; Parchman *et al*., 2010). Excepting the technical issues derived from sequencing, biological factors may be responsible for the large population of non-BLASTable sequences, including rapidly evolved genes (having orthologs in other species, but so highly divergent that efficient recognition of orthologs is precluded), species-specific genes (present in the studied species but absent from the databases) and the persistence of non-coding fractions mainly from untranslated regions of the sampled transcripts.

Saponin is considered to be derived from metabolites of phytosterol anabolism, which is the current assumption of saponin biosynthesis in plants (Lee *et al*., 2004; Qin *et al*., 2010). In plants, all terpenoids derive from condensation of five-carbon building blocks designated IPP (3-Isopentenyl PyroPhosphate, C5) and DMAPP (DiMethylallyl PyroPhosphate, C5), which mainly derive from condensation of acetyl-CoA in the cytosolic mevalonate pathway, although they may sometimes be from pyruvate and phosphoglyceraldehyde in the plastidial MEP (also: DXP) pathway. IPP and DMAPP undergo condensation to GPP (Geranyl PyroPhosphate, C10), which taking with a second IPP unit leads to FPP (Farnesyl PyroPhosphate, C15), FPP is the common precursor of the vast array of sesquiterpenes produced by plants. Linkage of two FPP units leads to formation of squalene (C30), which subsequently is epoxygenated to 2,3-oxidosqualene (C30). 2,3-Oxidosqualene is considered the last common precursor of triterpenoid saponins, of phytosterols and steroidal saponins (Kalinowska *et al*., 2005; Phillips *et al*., 2006; Vincken *et al*., 2007). However, although oxidosqualene has been suggested as a precursor of steroidal saponins, the steps at which steroidal saponin and phytosterol biosynthesis diverge have not been elucidated (Kalinowska *et al*., 2005; Vincken *et al*., 2007). In plants, oxidosqualene, as a precursor in the biosynthesis of saponins, was cyclized by OSCs (e.g. DS or AS). The step is rate-limited step for steroidal saponin biosynthesis. Next, this enzyme of CYP450s and UGTs, in turn, will play important roles in the hydroxylation and glycosidation of the product of cyclization, which is important in the production of various steroidal saponins. In this study, .most of all the known enzymes involved in MVA pathway for steroidal saponin biosynthesis were discovered in *P. polyphylla* EST dataset, which indicating the root can be a place synthesizing steroidal saponins for Paris, although further study must be undertaken including the leaf. By the way, it is noteworthy that one singleton sequences matched to *β*-AS of *P. polyphylla* was found in paris root library. The function of *β*-AS is used for the production of oleanane-type ginsenosides, which just seemingly in according with the recent paper about the founding of oleanane-type ginsenosides found in *P. polyphylla* (Wu *et al*., 2013). The existence of the transcripts for *β*-AS in

[Downloaded from jast.modares.ac.ir on 2024-04-18]

[DOR: 20.1001.1.16807073.2016.18.2.7.7]

P. polyphylla furtherly supports the existence of oleanane-type ginsenosides in paris root.

Development and Validation of Genic-SSR Markers

In total, 31,653 EST-SSR loci were identified in *P. polyphylla* transcriptome sequence data, which were then analyzed for potential SSRs using Simple Sequence Repeat Identification Tool (SSRIT) software (Temnykh *et al*., 2001). Primer pairs were designed with the following criteria: primer lengths of 18–24 bp, GC content of 40–65%, Annealing Temperature (Ta) ranging from 40 to 60°C, and a predicted PCR product size ranging from 100 to 500 bp. Eithy EST-SSR primer pairs were designed and synthesized (Shanghai Sangon Co. Ltd., Beijing, China). Thirty nine primer pairs were identified that yielded stable, clear, and repeatable amplicons in *P. polyphylla*. While for 41 primer pairs (51.2%), PCR completely failed, amplified too weakly, or amplified multiple bands, and then the 41 primers were excluded from further analyses. The possible reason might be because of the introns existing between the two primers (Varshney *et al*., 2005). The genotyping data of 9 primers were polymorphic within 55 samples, whose proportion of polymorphic primers was 21.9%. In total, 21 alleles were identified, ranging from one to three at each locus, with an average of 2.57 alleles (Table 2). The *H*o and *H*e values were 0.3656–0.5914 and 0.4086– 0.6452 with averages of 0.4731 and 0.5269, respectively. The coefficient hierarchical Fst, estimated according to Wright, ranged from –0.2402 to 0.6618.

PIC value is usually grouped into high (PIC> 0.5), moderate (0.5> PIC> 0.25), and low (PIC< 0.25) categories. It is often used to assess the informativeness level of markers developed (Botstein *et al*., 1980). The number of alleles and frequency distribution within the population can both affect SSR locus' *PIC* value. In our study, 9 EST-SSR markers represented power and marker index of most SSR loci with moderate *PIC* values. *PIC* values ranged from 0.467 to 0.589 (Mean= 0.535), suggesting that the EST-SSR markers developed had a moderate level of polymorphism in *P. polyphylla*.

The nine SSR markers were used to test the genotypes of 55 *P. polyphylla* samples representing 11 accessions. The samples used in our study were primarily for the development of EST-SSR markers and investigation of genetic relatedness among *P. polyphylla* accessions collected, not for revealing the extent of genetic diversity at the population level. The genetic variations among *P. polyphylla* populations (Mean value of He= 0.527) revealed in this study were similar to the general trend of high average microsatellite heterozygosity found in the species with narrow distribution (Mean He= 0.56) (Nybom 2004). It is known that breeding system of a species usually greatly impacts its genetic diversity and structure. Out-crossing species generally tend to be more genetically diverse (Hamrick and Godt, 1996; Nybom, 2000). Although *P. polyphylla,* is a self-compatible species and can reproduce selfed seeds, it is an insect pollinated outcrossing species (Li, 1998), which may account for its high variations among populations. In addition, the low inbreeding coefficient (Fis values= 0.0093) obtained for the populations in most of the populations also suggest that outcrossing in *P. polyphylla* populations was predominant. And the mean value of *Fst* (0.103) also suggests that this species exhibits most of the overall genetic variation between, rather than within, its populations.

The overall gene flow (Nm) provides an estimate of the average number of migrants between two populations. In this study, the inferred gene flow values between populations based on the *FST* values were shown to be high (Nm values ranged from 1.258 to 4.2889). But we would not propose efficient ongoing gene flow between extant populations of *P. polyphylla*, considering significant fragmentation of its modern

[Downloaded from jast.modares.ac.ir on 2024-04-18]

[DOR: 20.1001.1.16807073.2016.18.2.7.7]

habitats; instead, we suggest that the considerably high gene flow might be indicative of an earlier period of more pronounced gene flow when the species had a more continuous distribution.

Within all the accessions, the generated unrooted tree constructed using neighbor-joining criteria suggested that the 11 populations were separated into two main groups (Figure 3). In order to identify any geographical correlations between the recovered gene pools or genetic groups and sampled populations, we performed a Mantel test with 1,000 permutations. The results suggested that the genetic divergence of populations (Nei's genetic distance) was significantly correlated with geographic distance (P= 0. 003).

In summary, the present-day *P. polyphylla* populations maintain high degree of intra-population genetic diversity and exhibit low levels of inter-population differentiation by SSR microsatellite analysis. However, He *et al.* (2007) showed different conclusions of high genetic variation and differentiation in *P. polyphylla* using ISSR markers. This situation should be attributed to difference in detection methods and the collected populations. Genetic parameters detected by SSR and ISSR markers are obviously different, which can't be compared directly (Nybom, 2004). SSR markers are more informative and versatile in assessing genetic diversity and structure (Selkoe and Roonen, 2006; Zalapa *et al*., 2012). Therefore, moderate genetic diversity and low genetic differentiation observed by SSR probably reflects more accurately the situation of this species. As an endangered plant, *P. polyphylla* has been listed as the first class of protected plants. Appropriate conservation strategies are required for the long-time survival of this species. Protecting more habitats should be considered as a priority, since damage to the habitats is the prevalent causative factor for the decline in populations and number of individuals, as genetic diversity plays important role in species survival and evolution and maintenance of genetic variation is a major objective of conservation for endangered plants (Hamrick and Godt, 1996).

The ESTs derived EST-SSRs in the present study, were also characterized with BLAST annotation, in which 88% of ESTs had a putative function. Thiel *et al*. (2003) also reported earlier studies with a significant portion of the ESTs with putative functions. Most of the present EST-SSRs with known biological processes related to cell development and so on (Table 4). This might reveal the functional identity of a

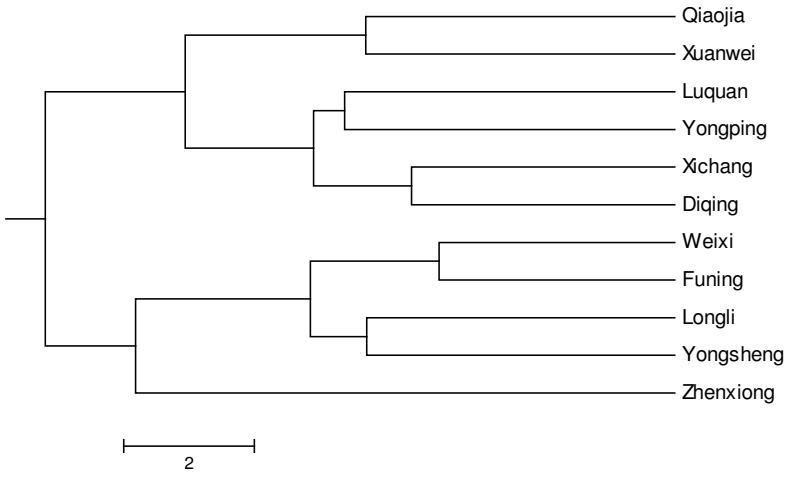

Figure 3. UPGMA dendgram of 11 collected population based on Nei's (1978) genetic distances using SSR data.

[Downloaded from jast.modares.ac.ir on 2024-04-18]

[DOR: 20.1001.1.16807073.2016.18.2.7.7]

Table 3. Summary of frequency of SSR nucleotide repeats in *P. polyphylla.*

Searching item	Numbers
Total number of sequences examined	8042
Total size of examined sequences (bp)	14185066
Total number of identified SSRs	3853
Number of SSR containing sequences	2849
Number of sequences containing more than 1 SSR	741
Number of SSRs present in compound formation	327
Mono-nucleotide	1612
Di-nucleotide	1634
Tri-nucleotide	594
Tetra-nucleotide	8
Penta-nucleotide	2
Hexa-nucleotide	3

Table 4. Characteristics of 9 ploymorphic *P. polyphylla* SSR markers.

Primer sequence (5'-3')	SSR motif	Expected size (bp)	*Tm*	*He*	*PIC*	Best matched protein
Forward: *AGATACTGGCCGGAAGGAGT* Reverse: *GCTTCAGCATTCCACTCCAG*	(GGT)5	143	57	0.5269	0.4665	Serine/arginine repetitive matrix protein
Forward: *GCACCCAATTCTACCACACC* Reverse: *ACTGGAACGTCCAGCTCGT*	(ATC)5	150	59	0.6344	0.584	Monoglyceride lipase
Forward: *AAAGTTCGCCTCCCTTTCTC* Reverse: *CCATTACCTGAGGCCTGAAA*	(TGC)5	149	56	0.5699	0.572	Cytoskeleton-associated protein
Forward: *GCTGCGATGCAAAACCTTAT* Reverse: *GGCAACCACCACCCTACTAA*	(TA)8	151	56	0.6452	0.568	E3 ubiquitin-protein ligase
Forward: *CCTTTGTAGCATGGGTGGTT* Reverse: *GACAATTGCTCCGACTCAAAA*	(AT)7	169	55	0.4516	0.486	rRNA-processing protein
Forward: *GTATCGACGGTCGCGATTAT* Reverse: *AGCAGGAGATTGAACCCTCA*	(CT)8	178	57	0.4409	0.544	Vesicle transport
Forward: *CATTAGCCGAGAAAGGCTTG* Reverse: *ACTGGAGCCTCGATCACAAT*	(AG)8	141	55	0.4086	0.501	ATP-dependent RNA helicase
Forward: *GGAGGAAGACGATGATCGAA* Reverse: *GCCATGTGCAGTCTCTCAAA*	(GAG)6	169	56	0.5269	0.589	No hit
Forward: *CCTCCATCACCACCTAAACC* Reverse: *AACTGAAGGTGGGGTCAGTG*	(CAC)5	168	57	0.5376	0.509	Proline-rich receptor-like protein kinase

[Downloaded from jast.modares.ac.ir on 2024-04-18]

[DOR: 20.1001.1.16807073.2016.18.2.7.7]

particular marker locus in the future. Therefore, working with these EST-SSR markers may provide a shortcut to candidate genes and gene based functional markers.

Hitherto, little work has been done on the development and application of SSR markers in *P. polyphylla* genetic and breeding studies. Our study is the first report on the development of EST-SSRs in *P. polyphylla*. In this study, a large-scale EST investigation involving the root of the medicinal plant *P. polyphylla* was performed using a HiSeq 2000 platform. This dataset contributes essential transcriptome information for gene discovery. The description of the expressed genes and distribution of gene functions was illustrated according to the results of GO analysis and KEGG assignments. Several transcription factors and EST-SSR markers were identified as well. These data will provide comprehensive information on gene discovery, transcriptome profiling, transcriptional regulation, and molecular markers for *P. polyphylla*. The findings of this study may facilitate marker-assisted breeding or genetic engineering schemes involving this species, as well as other medicinal plants of the Liliaceae family.

ACKNOWLEDGEMENT

This study was supported by Natural Science Foundation of Yunnan Province (31260075), Science and Technology Department of Yunnan Province (2012FB146), and The Major Project of Yunnan Provincial Development and Reform Commission (20121956).

REFERENCES

1. Abbasi, Z., Arzani, A. and Majidi, M. M. 2014. Evaluation of Genetic Diversity of Sugar Beet (*Beta vulgaris* L.) Crossing Parents Using Agro-morphological Traits and Molecular Markers. *J. Agr. Sci. Tech.,* **16(6):** 1397-1411
2. Berardini, T. Z., Mundodi, S., Reiser, L., Huala, E., Garcia-Hernandez, M., Zhang, P., Mueller, L. A., Yoon, J., Doyle, A. and Lander, G. 2004. Functional Annotation of the Arabidopsis Genome Using Controlled Vocabularies. *Plant. Physiol*., **135:** 745-755.
3. Botstein, D., White, R. L., Skolnick, M. and Davis, R. W. 1980. Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *Am. J. Hum. Genet.* **32:** 314–331
4. Chen, S., Luo, H., Li, Y., Sun, Y., Wu, Q., Niu, Y., Song, J., Lv, A., Zhu, Y., Sun, C., Steinmetz, A. and Qian, Z. 2011. 454 EST Analyses Detects Genes Putatively Involved in Ginsenoside Biosynthesis in *Panax ginseng*. *Plant. Cell. Report.*, 30: 1593-1601.
5. Chi, K. R. 2008. The Year of Sequencing. *Nat. Method*., **5:** 11–14.
6. Corey, E. J., Matsuda, S. P. T. and Bartel, B. 1993. Isolation of an *Arabidopsis thaliana* Gene Encoding Cycloartenol Synthase by Functional Expression in a Yeast Mutant Lacking Lanosterol Synthase by the Use of a Chromatographic Screen. *Proc. Natl. Acad. Sci. USA*, **90:** 11628–11632.
7. Doyle, J. J. and Doyle, J. L. 1987. A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue. *Phytochem. Bull.,* **19:** 11-15.
8. Frank, M. Y.**,** Naxin, H.**,** Yong, Q. G.**,** Ming, C. L.**,** Ma, Y. Q.**,** Hane, D.**,** Lazo, G. R.**,** Dvorak, J. and Anderson, O. D. 2008. BatchPrimer3: A High Throughput Web Application for PCR and Sequencing Primer Design. ***BMC Bioinforma.***, 9: 253.
9. Ghaedrahmati, M., Mardi, M., Naghavi, M.R., Majidi Heravan, E., Nakhoda, B., Azadi, A. and Kazemi, M. 2014. Mapping QTLs Associated with Salt Tolerance Related Traits in Wheat (*Triticum aestivum* L.). *J. Agr. Sci. Tech.,* **16(6):** 1413-1428.
10. Guo, C. K., Zhang, S., Kong, W. J. and Li, Q. T. 2006. Clinical Study of Bi Yan Qing Du Granule and Bi Yan Shu Oral Liquid in Treatment Patients with Nasopharyngeal Carcinoma after Radiotherapy. *Chin. Cancer,* **15:** 113–115.
11. Guo, S. Y., Zheng, J., Joung, G., Liu, S., Zhang, Z., Crasta, O. R., Sobral, B. W., Xu, Y., Huang, S. and Fei, Z. 2010. Transcriptome Sequencing and Comparative Analysis of Cucumber Flowers with Different Sex Types. *BMC Genom.,* **11:** 384

[Downloaded from jast.modares.ac.ir on 2024-04-18]

[DOR: 20.1001.1.16807073.2016.18.2.7.7]

12. Hamrick, J. L. and Godt, M. J. W. 1996. Effects of Life History Traits on Genetic Diversity in Plant Species. *Phil. Trans. R. Soc. B,* 351: 1291–1298.
13. Haralampidis, K., Trojanowska, M., Osbourn, A. E. 2002. Biosynthesis of Triterpenoid Saponins in Plants. *Adv. Biochem. Eng. Biotechnol*. **75:** 31-49.
14. He, J., Yang, B. Y., Chen, S. F., Gao, L. M. and Wang, H. 2007. Assessment of genetic diversity of *Paris ploiphylla* (Trilliaceae) by ISSR markers. *Acta Botanica Yunnanica*, **29:** 388-392.
15. Herrera, J. B., Bartel, B., Wilson, W. K. and Matsuda, S. P. 1998. Cloning and Characterization of the *Arabidopsis thaliana Lupeol Synthase* Gene. *Phytochem.* ,**49:** 1905–11.
16. Iseli, C., Jongeneel, C. V. and Bucher, P. 1999. ESTScan: A Program for Detecting, Evaluating, and Reconstructing Potential Coding Regions. Proc. *Int. Conf. Intell. Syst. Mol. Biol.,* **4**: 138–148.
17. Ji, Y. H., PETER, W. F., Li, H., Xiao, T. J. and Zhou, Z. K. 2006. Phylogeny and Classification of Paris (Melanthiaceae) Inferred from DNA Sequence Data. *Ann. Bot*., **98:** 245–256.
18. Kalinowska, M., Zimowski, J., Paczkowski, C. and Wojciechowski, Z. A. 2005. The Formation of Sugar Chains in Triterpenoid Saponins and Glycoalkaloids. Phytochem. Rev., **4:** 237–257.
19. Kanehisa, M. and Goto, S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic. Acids. Res.,* **28:** 27-30.
20. Kumar, S., Kalra, S., Kumar, S., Kaur, J. and Singh, K. 2012. Differentially Expressed Transcripts from Leaf and Root Tissue of *Chlorophytum borivilianum:* A Plant with High Medicinal Value. *Gene.,* **511:** 79–87.
21. Lee, M. H., Jeong, J. H., Seo, J. W., Shin, C. G.., Kim, Y. S., In, J. G., Yang, D. C., Yi, J. S. and Choi, Y. E. 2004. Enhanced Triterpene and Phytosterol Biosynthesis in *Panax ginseng* Overexpressing *Squalene Synthase* Gene. *Plant Cell. Physiol.,* **45:** 976–984.
22. Lee, R. K. Y., Ong, R. C. Y., Cheung, J. Y. N., Li, Y. C., Chan, J. Y. W. and Lee, M. M. S. 2009. Polyphyllin D: A Potential Anti-cancer Agent to Kill Hepatocarcinoma Cells with Multi-drug Resistance. *Curr. Chem. Biol*., **3:** 89–99.
23. Li, H. 1998. The Phylogeny of the Genus *Paris* L. In: "The Genus *Paris* (Trilliaceae)", (Ed.): Li, H.. Science Press, Beijing, PP. 8–65.
24. Liu, H., Zhang, T., Chen, X. Q., Huang, Y. and Wang, Q. 2006. Steroidal Saponins of *Paris polyphylla Smith* var. *yunnanensis*. *Chin. J. Nat. Med.,* **4:** 265–7.
25. Liu, K. and Muse, S. V. 2005. PowerMarker: Integrated Analysis Environment for Genetic Marker Data. *Bioinforma.,* **21:** 2128-2129.
26. Luo, H. M., Sun, C. and Sun, Y. Z. 2011. Analysis of the Transcriptome of *Panax notoginseng* Root Uncovers Putative *Triterpene Saponin-biosynthetic* Genes and Genetic Markers. *BMC Genom.,* **12(Suppl. 5) S5**.
27. Mardis, E. R. 2008. The Impact of Next-generation Sequencing Technology on Genetics. *TIG,* **24:** 133–141.
28. Margulies, M., Egholm, M., Altman, W. E., Attiya, S. and Bader, J. S. 2005. Genome Sequencing in Microfabricated High-density Picolitre Reactors. *Nature,* **37:** 376–380.
29. Matsuda, H., Pongpiriyadacha, Y., Morikawa, T., Kishi, A., Kataoka, S. and Yoshikawa, M. 2003. Protective Effects of Steroid Saponins from *Paris polyphylla* var. *yunnanensis* on Ethanol- or Indomethacin-induced Gastric Mucosal Lesions in Rats: Structural Requirement for Activity and Mode of Action. *Bioorg. Med. Chem. Lett*., **13:** 1101–1106.
30. Miller, M. P. 1997. *Tools for Population Genetic Analysis, Version 1.3.* Department of Biological Sciences, Northern Arizona University, Flagstaff AZ.
31. Morozova, O. and Marra, M. 2008. Applications of Next-generation Sequencing Technologies in Functional Genomics. *Genom.,* **2**: 255–264.
32. Nybom, H. 2000. Effects of Life History Traits and Sampling Strategies on Genetic Diversity Estimates Obtained with RAPD Markers in Plants. *Perspect. Plant. Ecol*., **3:** 93–114.
33. Nybom, H. 2004. Comparison of Different Nuclear DNA Markers for Estimating Intraspecific Genetic Diversity in Plants. *Mol. Ecol.,* 13: 1143–1155.
34. Parchman, T. L., Geist. K. S., Grahnen, J. A., Benkman, C. W. and Buerkle, C. A. 2010. Transcriptome Sequencing in an Ecologically Important Tree Species: Assembly, Annotation, and Marker Discovery. *BMC Genom*., **11:** 180.

[Downloaded from jast.modares.ac.ir on 2024-04-18]

[DOR: 20.1001.1.16807073.2016.18.2.7.7]

35. Phillips, D. R., Rasbery, J. M., Bartel, B. and Matsuda, S. P. T. 2006. Biosynthetic Diversity in Plant Triterpene Cyclization. *Curr. Opin. Plant Biol.,* **9:** 305–314.
36. Qin, B., Eagles, J., Mellon, F. A., Mylona, P., Pena-Rodriguez, L. and Osbourn, A. E. 2010. High throughput Screening of Mutants of Oat that Are Defective in Triterpene Synthesis. *Phytochem.,* **71:** 1245–1252.
37. Rohmer, M. 2003. Mevalonate-independent Methylerythritol Phosphate Pathway for Isoprenoid Biosynthesis. Elucidation and Distribution. *Pure. Appl. Chem*., **75:** 375–387.
38. Selkoe, K. A. and Roonen, R. J. 2006. Microsatellites for Ecologists: A Practical Guide to Using and Evaluating Microsatellite Markers. *Ecol. Lett.,* 9: 615–629.
39. Sharma, R. K., Bhardwaj, P., Negi, R., Mohapatra, T. and Ahuja, P. S. 2009. Identification, Characterization and Utilization of Unigene Derived Microsatellite Markers in Tea (*Camellia sinensis* L.). *BMC Plant. Biol*., **9:** 53.
40. Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S. and McCouch, S. 2001. Computational and Experimental Qnalysis of Microsatellites in Rice (*Oryza sativa* L.): Frequency, Length Variation, Transposon Associations, and Genetic Marker Potential. *Genom. Res*., **11:** 1441-1452.
41. Thiel, T., Michalek, W., Varshney, R. K. and Graner, A. 2003. Exploiting EST Databases for the Development and Characterization of Gene-derived SSR-markers in Barley (*Hordeum vulgare* L.). *Theor. Appl. Genet*., **106:** 411–422.
42. Tiwari, J. K., Ballabha, R. and Tiwari, P. 2010. Ethnopaediatrics in Garhwal Himalaya Uttarkhand India (Psychomedicine and Medicine). *New York Sci. J.*, **3:** 123-126.
43. Varshney, R. K., Graner, A., Sorrells, M. E. 2005. Genic Microsatellite Markers in Plants: Features and Applications. *Trend. Biotechnol*., **23**: 48–55.
44. Vincken, J. P., Heng, L., deGroot, A. and Gruppen, H. 2007. Saponins: Classification and Occurence in the Plant Kingdom. *Phytochem.,* **68**: 275–297.
45. Wang, L., Li, P. and Brutnell, T. P. 2010. Exploring Plant Transcriptomes Using Ultra High-throughput Sequencing. *Brief Funct. Genom.,* **9:** 118–128.
46. Wang, W., Wang, Y., Zhang, Q., Qi, Y. and Guo, D. 2009. Global Characterization of *Artemisia annua* Glandular Trichome Transcriptome Using 454 Pyrosequencing. *BMC Genom.,* **10:** 465.
47. Weber, A. P., Weber, K. L., Carr, K., Wilkerson, C. and Ohlrogge, J. B. 2007. Sampling the *Arabidopsis* Transcriptome with Massively Parallel Pyrosequencing. *Plant. Physiol*., **144:** 32–42.
48. Wu, X., Wang, L., Wang, G.. C., Wang, H., Dai, Y., Yang, X. X., Ye, W. C. and Li, Y. L. 2013. Triterpenoid Saponins from Rhizomes of *Paris polyphylla* var. *yunnanensis. Carbohydr. Res.,* **4:** 3681–3687.
49. Xu, T. H., Ma, X. X., Xu, Y. J., Xie, S. X., Zhao, H. F., Si, Y. S., Han, D., Li, Y., Niu, J. Z. and Xu, D. M. 2007. New Steroidal Saponin from *Paris polyphylla Sm*. var. *yunnanensis* (France). Hand-Mazz. *Chem. J. Chin*. *U*, **28:** 2303–6.
50. Yeh, F. C., Yang, R. C. and Boyle, T. 1999. *POPGENE 32, Version 1.31.* Population Genetics Software, Edmonton University of Alberta.
51. Zalapa, J. E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E. and McCown, B. 2012. Using Nest-generation Sequencing Approaches to Isolate Simple Sequence Repeat (SSR) Loci in the Plant Sciences. *Am. J. Bot.,* **99:** 193-208.
52. Zhang, C. B., Peng, B., Zhang, W. L., Wang, S. M., Sun, H., Dong, Y. S. and Zhao, L. M. 2014. Application of SSR Markers for Purity Testing of Commercial Hybrid Soybean (*Glycine max* L.). *J. Agr. Sci. Tech.,* **16(6):** 1389-1396.
53. Zhang, S. P. 2007. Research Progress on Chemical Constituents and Pharmacological Effect of Genus *Paris*. *Strait. Pharm. J.,* **19:** 4–7.
54. Zhao, J. L., Mou, Y., Shan, T. J., Li, Y., Zhou, L. G., Wang, M. G.. and Wang, J. G. 2010 .Antimicrobial Metabolites from the Endophytic Fungus *Pichia guilliermondii* Isolated from *Paris polyphylla* var. *yunnanensis. Mol.,* **15:** 7961–7970.
55. Zhou, X., Su, Z., Sammons, R. D., Peng, Y. H., Tranel, P. J., Stewart, C. N. and Yuan, J. S. 2009. Novel Software Package for Cross-Platform Transcriptome Analysis (CPTRA). *BMC Bioinforma.,* **10:** S16.

[Downloaded from jast.modares.ac.ir on 2024-04-18]

[DOR: 20.1001.1.16807073.2016.18.2.7.7]

توصیف جدید ترانسکریپتوم ریشه و توسعه مارکر هایEST-SSR در گیاه دارویی *Paris polyphylla* Smith var. *yunnanensis* در معرض انقراض به نام

ل. وانگ، ی. یانگ، ی. ژائو، س. یانگ، س. اودیکری، و ت. لیو

چکیده

گیاه دارویی *Paris polyphylla* Smith var. *yunnanensis* (Liliaceae) گیاه دارویی مهمی در استان یونان چین است. با این همه، اطلاعات ژنومیکی این گیاه محدود است. برای درک بیشتر از زمینه ملکولی این گونه گیاه، آزمون Illumina HiSeq 2000 توالی یابی نسل دوم انجام شد و تقریبا 30,198,679 قرائت از یاخته های ریشه آن به دست آمد. این قرائت ها در56,095 توالی تک نسخه ای (unique sequence) گردهم شد و تقریبا ۴۹/۷٪ این توالی ها با جستجوی مشابه یابی در بانک عمومی توالی ها با کاربرد Basic Local Alignment Search Tool (BLAST) تفسیر و یادداشت نویسی شد. بیشتر این تک ژن ها(unigenes) در متابولیسم کربوهیدارت ها، متابولیسم انرژی و در مسیر زیست ساخت متابولیت های ثانویه مکان یابی (map) شد. افزون بر این، 3,853 عدد EST-SSRs به عنوان ملکولی های مارکر مستعد در تک ژن ها شناسایی شد.. از این ها، ۹ مارکر SSRهسته ای برای ارزیابی تنوع ژنتیکی و ساختار ۱۱ جمعیت با پراکندگی جغرافیایی (adjunct) مورد استفاده قرار گرفت. در ادامه بررسی ها، پژوهش حاضر تنوع ژنتیکی متوسط (He= ۰/۵۲۷) و تفاوت ژنتیکی کم (Fst= ۰/۱۰۳) را آشکار ساخت که این نتایج را می توان به یک جریان ژنی(gene flow) چشمگیرتر در زمانی که گونه ها از توزیع (جغرافیایی) پیوسته ای برخوردار بودند نسبت داد. ۱۱ جمعیت مورد مطالعه بر مبنای دندروگرام UPGMA به خوشه دسته تقسیم شد که با توزیع جغرافیایی آن ها هماهنگ نبود. به طور کلی، توالی مجموعه آر.ان.ا. (transcriptome) ریشه که در این پژوهش به دست آمد اطلاعات جدیدی در مورد مشخصات بیان ژن را اشکار ساخت و راهنمایی های مهمی برای مطالعه بیشتر در زمینه سازوکار مولکولی زیست ساخت متابولیت های ثانویه ریشه گیاه Paris و ژنتیک جمعیتی آن ارایه می کند. همچنین، مارکر هایEST-SSRs شناسایی شده این گیاه، بهگزینی به کمک مارکر را در گیاه مزبور تسهیل خواهد کرد.

[Downloaded from jast.modares.ac.ir on 2024-04-18]

[DOR: 20.1001.1.16807073.2016.18.2.7.7]