Pak. J. Bot., 54(4): 1327-1333, 2022. DOI: http://dx.doi.org/10.30848/PJB2022-4(23)

ANALYSIS OF SSR LOCI INFORMATION IN TRANSCRIPTOME OF *SINOPODOPHYLLUM HEXANDRUM* (ROYLE) YING BY RNA-SEQ

ZHENG ZHANG[1], ZHENZHEN SUI[2], MINGGE CHEN[2], HUANXING CHANG[2], LUYAO YU[1], TONG ZHANG[1], DONGXUE YIN[3] AND WEI LIU[1*]

[1]College of Agriculture, Henan University of Science and Technology, Luoyang 471000, P.R. China
[2]Institute of Agricultural Sciences of Sui County in Henan Province, Sui County 476900, China
[3]College of Agricultural Equipment Engineering, Henan University of Science and Technology, Luoyang 471003, P.R. China
**Corresponding author's email: 15729111052@163.com*

Abstract

Sinopodophyllum hexandrum (Royle) Ying is a rare and endangered medicinal plant in China, and the podophyllotoxin contained in its rhizome is a precursor substance of synthetic anti-cancer drugs, which has special effects on antineoplastic activities. In this study, 75245 Unigenes were assembled by splicing clean reads with Trinity software based on the transcriptome sequencing data of *S. hexandrum*. A total of 19934 SSR loci were identified in the total Unigenes by using MISA software, and the number of SSR containing sequences was 14809, accounting for 19.68% of the total Unigenes. The dominant repeat motif was mononucleotide, accounting for 55.51% of the total SSR loci, followed by dinucleotide and trinucleotide, accounting for 22.87% and 19.02%, respectively. A/T, AG/CT and AAG/CTT were repeat units with high frequency in mononucleotide, dinucleotide and trinucleotide, respectively. Furthermore, ten was the most frequent number of repeats, up to 5063 SSR loci, accounting for 25.40% of the total SSR loci, followed by 5 repetitions, the number of loci was 2584, accounting for 12.96% of the total. This study utilized transcriptome high-throughput sequencing technique to analyze SSR loci information, which can provide a new basis for genetic diversity research, systematic differentiation analysis, effective molecular marker development and molecular directional breeding of *S. hexandrum*.

Key words: *Sinopodophyllum hexandrum*, Podophyllotoxin, Transcriptome, SSR loci information, RNA-Seq.

Introduction

Sinopodophyllum hexandrum (Royle) Ying, also known as *Podophyllum*, is a perennial medicinal plant belonging to the *Sinopodophyllum* of the *Berberaceae* family. It has the functions of preventing cancer and anti-cancer, clearing heat-toxin, dissipating phlegm and resolving masses, removing blood stasis and detumescence (Stähelin & Von, 1991; Giri & Narasu, 2000; Canel *et al.*, 2001; Rickard-Bon & Thompson, 2003; Mortensen & Larsen, 2010), mainly distributed in Yunnan, Sichuan, Tibet, Gansu, Qinghai, Shaanxi and other places (Anon., 2011). Breast cancer has become the most common cancer in the world, and podophyllotoxin is a precursor substance for the synthesis of etoposide (VP-16), teniposide (VM-26), GP-7 and NK-611, which have excellent effects to treat breast cancer, cervical cancer and skin cancer (Ayres & Kee, 1976; Kamil & Dewick, 1986; Holthuis, 1988; Stähelin & Von, 1991; Imbert, 1998; Giri & Narasu, 2000; Mortensen & Larsen, 2010). *S. hexandrum* is the preferred medicinal plant for extracting podophyllotoxin, as its rhizome contains a large amount of podophyllotoxin, and its content is even more than three times that of American Mayapple (*Podophyllum peltatum*) (Moraes *et al.*, 2002; Yousefzadi *et al.*, 2010). Due to excessive logging, the plant resources of *S. hexandrum* are becoming more and more scarce and have been included in the "List of Rare and Endangered Plants in China" (Fu, 1991). Therefore, it is necessary to carry out research on the genetic diversity to protect the plant resources of *S. hexandrum*, and the transcriptome SSR locus analysis can provide a technical guarantee for the genetic diversity research, innovation and development of *S. hexandrum*.

Simple Sequence Repeat (SSR), or microsatellite DNA, is short tandem repeats consisting of 1-6 nucleotides, and the length is generally less than 200bp (Li *et al.*, 2019). Transcriptome sequencing is based on high-throughput sequencing technology for cDNA sequencing, which can comprehensively and quickly obtain almost all transcripts of a specific organ or tissue of a species in a certain state (Zhou *et al*., 2012). Transcriptome sequencing has opened up a new way for the study of SSR molecular markers. For example, Li *et al*., (2019) used the transcriptome sequencing data of *Eleutherococcus senticosus* to screen the genes sequences related to saponins, fatty acids, sugar and acid, vitamins and SOD, and it was found that 56 pairs of primers could amplify the clear bands among 107 pairs of target gene SSR primers. Sa *et al*., (2019) screened 37 pairs of SSR primers and identified 182 alleles at 37 SSR loci based on the transcriptome data of *Perilla frutescens*. Jia *et al*., (2019) obtained transcripts of leaves, stems and flowers of *Notopterygium incisum* by RNA-Seq and found that 19 pairs of SSR primers were polymorphic through SSR analysis. Wang *et al*., (2017) screened 99 pairs of SSR primers to analyze the genetic diversity of centipedegrass and divided the materials into two germplasm resources from different regions according to 79 polymorphic markers. Wang *et al*., (2015) found that 14 SSR markers were polymorphic by SSR analysis of transcriptome sequencing data of *Veratrilla baillonii* Franch (Gentianaceae).

At present, the related researches on *S. hexandrum* were mainly focused on chemical composition (Shang *et al*., 2000), pharmacological effect (Stähelin & Von, 1991; Giri & Narasu, 2000; Canel *et al.*, 2001; Rickard-Bon & Thompson, 2003; Mortensen & Larsen, 2010), biological characteristics (Zhang, 2009), the production technology of podophyllotoxin (Nandagopal *et al.*, 2017) and the genes of podophyllotoxin

synthesis (Lau & Sattely, 2015), etc. Only a few scholars have reported the genetic diversity of *S. hexandrum* based on AFLP (Xiao *et al.*, 2006), ISSR (Xiao *et al.*, 2006; Shen *et al.*, 2019), SCoT (Chen *et al*., 2013), and RAPD (Xiao *et al.*, 2015). However, there are few reports on SSR molecular markers of *S. hexandrum*. Therefore, this study selected *S. hexandrum* samples from Gansu, Shaanxi and Tibet as test materials, and then used RNA-Seq to analyze SSR loci information in transcriptome based on the homogenous garden experiment. The result of study can provide a new tool for genetic diversity research, systematic differentiation analysis, effective molecular marker development and directional breeding of *S. hexandrum* at molecular level, which is of great significance for the innovation, development and reasonable protection of *S. hexandrum* resources.

Material and Methods

Plant samples: The samples were taken from the *S. hexandrum* Germplasm Resource Nursery (E107°35'46.75", N33°49°10.44") in Wanhuashan Medicinal Botanical Garden, Huangbaiyuan Town, Taibai County, Shaanxi Province. In *S. hexandrum* resource nursery, five-year rhizomes of *S. hexandrum* provenances from Gansu, Shaanxi and Tibet were taken in May, 2020, and three plants were randomly selected from each provenance. Approximately 20g rhizomes from each single *S. hexandrum* plant were stored in RNAhold preservation solution [Shenggong Bioengineering (Shanghai) Co., Ltd.] at -80°C for further analysis.

Extraction and detection of total RNA: The rhizomes of *S. hexandrum* were treated with liquid nitrogen and extracted total RNA with TRIzol kit (Invitrogen, USA). The specific operation steps were performed according to the instructions of the TRIZOL kit. The purity and integrity of the RNA was used agarose gel electrophoresis, NanoPhotometer® spectrophotometer (IMPLEN, CA, USA), Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA) to accurately detect (Yan *et al.*, 2018).

Library construction, sequencing and assembly: A total amount of 1.5 μg RNA per sample was used as input material for the RNA sample preparations, and sequencing libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA). The library was preliminarily quantified by Qubit2.0 Fluorometer, then used Agilent 2100 bioanalyzer to detect the insert size of the library. Finally, the effective concentration of the library was accurately quantified by qRT-PCR technology (the effective concentration of the library was higher than that of 2nM) to ensure the quality of the library (Ou *et al*., 2019). Paired-end sequencing was performed using Illumina Novaseq 6000 for the 9 cDNA libraries of *S. hexandrum* rhizomes after the library quality inspection was qualified. The clean reads were obtained after the sequenced fragments passed the quality control, and then used Trinity (v2.4.0) to splice and assemble the clean reads (Grabherr *et al.*, 2011).

Analysis and screening of SSR loci: In this study, MISA software (version 1.0) was used to perform the analysis of the *S. hexandrum* unigenes, in order to search and identify Simple Sequence Repeat (SSR). The minimum number of repeats for each corresponding unit size was set as follows: when using mononucleotide as repeat unit, the minimum number of repeats was at least 10; when dinucleotide was used as repeat unit, the number of repeats was at least 6; when using trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide as repeat units, the minimum number of repeats was 5 (Wang *et al.*, 2018). At the same time, if the distance of two SSR sequences was less than 100 bp, they will be considered as one SSR locus (Wang *et al*., 2018; Zhang *et al*., 2019).

Results

Quality control analysis of transcriptome sequencing data: Table 1 showed that the transcriptome sequencing results of *S. hexandrum*. After 9 samples were sequenced by the transcriptome, 195681371 raw reads were obtained. The sequencing data of each sample was filtered to remove reads containing adapter, reads containing ploy-N and low quality reads from raw reads, and a total of 191065788 clean reads were obtained. The clean base was above 6 Gb, the average GC-content was 43.59%, Q20 (The percentage of bases with a Phred value greater than 20 in the total bases) and Q30 (The percentage of bases with a Phred value greater than 30 in the total bases) were more than 97% and 93%, respectively. These data indicated that the quality of the clean data was relatively high, which could be used for the subsequent analysis of this study.

Table 1. Quality control analysis of transcriptome sequencing data of S*inopodophyllum hexandrum.*

Sample	Raw reads	Clean reads	Clean bases	Error rate	Q20	Q30	GC%
Gansu 1	21327501	20718294	6.22G	0.02	98.02	94.22	43.31
Gansu 2	21328231	20798416	6.24G	0.02	98.05	94.34	43.74
Gansu 3	22235107	21799613	6.54G	0.03	97.77	93.67	43.71
Shaanxi 1	23673883	23196443	6.96G	0.03	97.84	93.74	42.87
Shaanxi 2	21158826	20667859	6.20G	0.03	98.01	94.13	43.87
Shaanxi 3	20666977	20196666	6.06G	0.02	98.21	94.53	43.81
Tibet 1	23150071	22663283	6.80G	0.02	98.05	94.19	43.75
Tibet 2	20875840	20332098	6.10G	0.02	98.14	94.38	43.92
Tibet 3	21264935	20693116	6.21G	0.02	98.10	94.28	43.34
Total	195681371	191065788	—	—	—	—	—
Average	—	—	—	—	—	—	43.59

Q20 and Q30 means the percentage of bases with a Phred value greater than 20 and 30 in the total bases, respectively [Phred=-10Log10(e)]

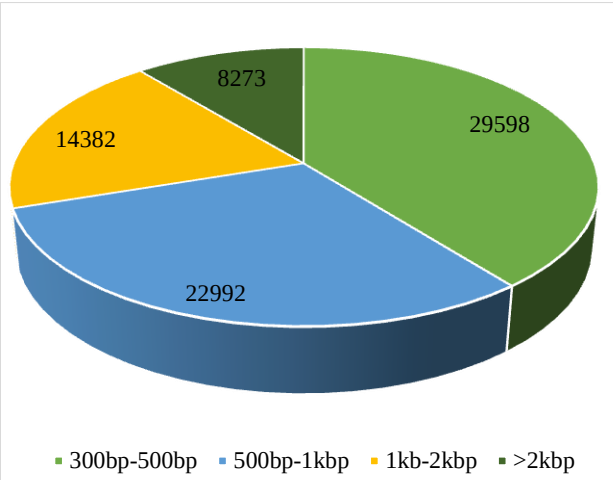

Fig. 1. Unigenes length distribution map.

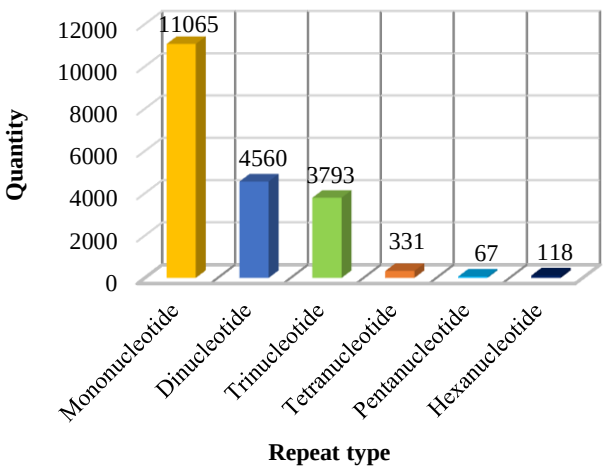

Fig. 2. SSR locus repeat type.

Sequence assembly and cluster analysis: Clean reads was assembled and spliced by de novo using Trinity (v2.4.0), and then used Corset software to carry out cluster analysis (Fig. 1). The results showed that a total of 75245 Unigenes were obtained, ranging in length from 301 to 16637 bp, with an average length of 969 bp and a total length of 72926778 bp. Fig. 1 showed that there were 29598 Unigenes in the range of 300 to 500 bp in length, 22992 Unigenes with lengths ranging from 500 bp to 1 Kbp, 14382 Unigenes ranging in length from 1 Kbp to 2 Kbp, and 8273 Unigenes ranging in length over 2 Kbp. These data indicated that there were more Unigenes distributed between 300 bp-500 bp and 500 bp-1 Kbp. In addition, the N50 (the length of the spliced transcript ≥ 50% of the total length) was 1397nt and N90 (the length of the spliced transcript ≥ 90% of the total length) was 418nt.

The number and distribution of SSR loci in *S. hexandrum* transcriptome: Table 2 showed the number and distribution of SSR loci in the transcriptome of *S. hexandrum*. It found that a total of 19934 SSR loci were identified from 75245 Unigenes using MISA software. The number of SSR containing sequences was 14809, accounting for 19.68% of the total Unigenes, and the frequency of occurrence of SSR was 26.49%. Furthermore, there were 3671 of these Unigenes containing more than one SSR locus, and the number of SSRs present in compound formation was 1818.

Table 2. SSR loci information of the transcriptome in *S. hexandrum*.

Type	Number of SSRs
Total number of sequences examined	75 245
Total size of examined sequences (bp)	72 926 778
Total number of identified SSRs	19 934
Number of SSR containing sequences	14 809
Number of sequences containing more than 1 SSR	3 671
Number of SSRs present in compound formation	1 818

Figure 2 showed that six nucleotide repeat types were detected based on the transcriptome of *S. hexandrum*, and the proportion of different nucleotide repeat types were quite different. Similarly, the distribution map of SSR motifs also illustrated this point very well (Fig. 3). In detail, mononucleotide, dinucleotide and trinucleotide were the dominant repeat motifs, the most abundant motif was mononucleotide (11065, accounting for 55.51% of the total SSR loci), followed by dinucleotide and trinucleotide, with the number of 4560 and 3793, accounting for 22.87% and 19.02% of the total SSR loci, respectively. The number of repeats of tetranucleotide, pentanucleotide and hexanucleotide were relatively less, which were 331, 67 and 118, respectively, accounting for 1.66%, 0.34% and 0.60% of the total SSR loci, respectively.

Table 3. Distribution of repeat times of SSR locus in *S. hexandrum* transcriptome.

Motif type	Number of repeat									
	5	6	7	8	9	10	11	12	13	≥14
Mononucleotide	0	0	0	0	0	4676	1891	1026	722	2750
Dinucleotide	0	1235	741	575	443	303	225	178	131	729
Trinucleotide	2203	766	389	178	117	79	11	16	13	21
Tetranucleotide	239	61	19	5	1	4	1	0	0	1
Pentanucleotide	47	19	0	0	1	0	0	0	0	0
Hexanucleotide	95	7	5	4	2	1	0	0	3	1
Total	2584	2088	1154	762	564	5063	2128	1220	869	3502
Proportion（%）	12.96	10.47	5.79	3.82	2.83	25.40	10.68	6.12	4.36	17.57

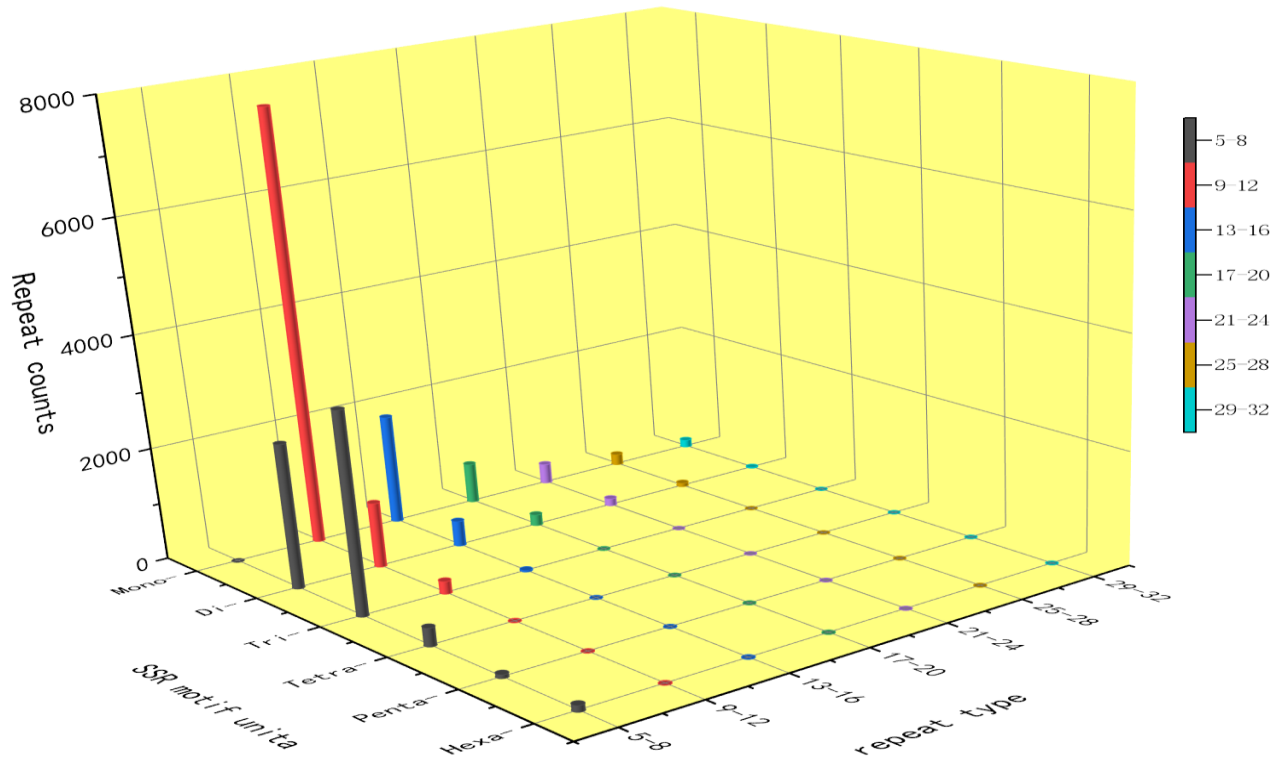

Fig. 3. Distribution of SSR motifs. The X coordinate (SSR motif unit) is the motif type, Y coordinate (repeat type) corresponds to the legend according to the color for the specific number of repeats, and Z coordinate (repeat counts) is the number of SSR.

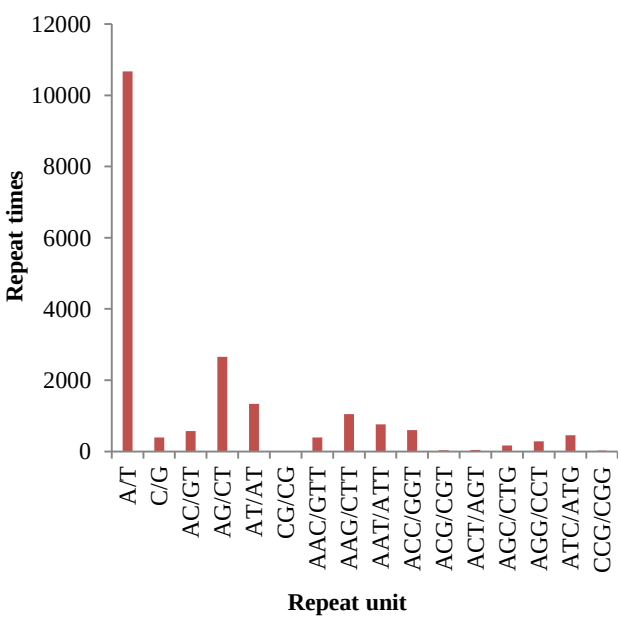

Fig. 4. Distribution map of the number of repeat units in *S. hexandrum* transcriptome (only displaying mononucleotide, dinucleotide and trinucleotide).

Repeat unit types and frequencies of SSR loci: Fig. 4 showed that the distribution map of repeat units of SSR loci in *S. hexandrum* transcriptome. A total of 120 SSR loci with different types of repeat unit were detected in the transcriptome of *S. hexandrum*, and the quantity of the repeat units from mononucleotide to hexanucleotide were 2, 4, 10, 22, 15 and 67, respectively. There were two types of repeat units in mononucleotide, namely A/T and C/G, which appeared 10674 times and 391 times respectively, accounting for 53.55% and 1.96% of the total SSR loci. A/T was not only the most repeat unit in mononucleotide, but also the most repeat unit of all repeat units. Of the dinucleotide repeat motifs, AG/CT and AT/AT were the most abundant repeat units that appeared 2654 times and 1337 times respectively, with frequencies of 13.31% and 6.71% of the total SSR loci, respectively. The most abundant units of trinucleotide repeat motifs were AAG/CTT, which appeared 1043 times, accounting for 5.23% of the total, followed by AAT/ATT and ACC/GGT, which appeared 756 times (3.79%) and 602 times (3.02%), respectively.

Table 3 showed the results of statistical analysis of SSR loci repetitions in the transcriptome of *S. hexandrum*. The results exhibited that the repetitions of SSR loci were mostly concentrated in 5-13, with a total of 16432 loci, accounting for 82.43% of the total SSR loci. With the increase of repeat times, the number of loci gradually decreased. The highest times of repeats was 10, reaching 5063 loci, accounting for 25.40% of the total SSR loci, followed by 5 times, with 2584 SSR loci, accounting for 12.96% of the total SSR loci.

Discussion

Molecular marker is a kind of genetic markers based on the polymorphism of biological macromolecules. In the broad sense, molecular genetic markers include DNA markers and protein markers, while in the narrow sense, molecular markers only refer to DNA markers (Zhao *et al.*, 2000). At present, some molecular markers such as RFLP, RAPD, SSR,

ISSR, AFLP, SNP were commonly used in study. SSR markers have the advantages of wide distribution, rich polymorphism, good repeatability, convenient detection and stable detection results. The traditional SSR marker technology is to study the expression and function of genes at the genome level, and the cost of research is expensive. With the development of high-throughput sequencing technology, using RNA-Seq to obtain transcripts of certain tissues or organs of a species under a certain growth state not only reduces the cost of research, but also provides a platform for the development and utilization of molecular markers for species that lack genomic information (Yao *et al.*, 2019). At present, there are more and more studies on SSR analysis based on transcriptome data, which has been widely used in traditional Chinese medicine, horticultural crops, economic animals and other fields. In a word, RNA-Seq provides a new research method for SSR analysis.

This study was based on second-generation transcriptome sequencing technology to analyze SSR loci of different provenances of *S. hexandrum*. Among the 75245 Unigenes, 19934 SSR loci were detected on 14809 Unigenes, and the occurrence rate of SSR (the ratio of SSR loci to total Unigenes) was 26.49%, which was higher than that of some important medicinal plants, such as *Forsythia suspense* (25.15%) (Wang *et al.*, 2015), *Sambucus nigra* (24.87%) (Yao *et al*., 2019), *Gleditsia sinensis* (20.72%) (Lin *et al*., 2017), *Camellia fascicularis* (19.63%) (Ye *et al*., 2019), *Sophora flavescens* (16.23%) (Zhang *et al*., 2019), *Eucommia ulmoides* (11.85%) (Feng *et al*., 2016), *Psammosilene tunicoides* (8.54%) (Ye *et al*., 2019), *Houttuynia cordata* Thunb (7.51%) (Lin *et al*., 2014), Blackberry (5.33%) (Li *et al*., 2016). However, it was lower than that of other medicinal plants such as *Xanthoceras sorbifolia* (27.81%) (Dong *et al*., 2019), *Cordyceps sinensis* (34.99%) (Zhang *et al*., 2019), *Magnolia officinalis* (52.75%) (Dai *et al*., 2017). Transcriptome sequencing is divided into first-generation sequencing, second-generation sequencing and third-generation sequencing. There are also many kinds of sequencing instruments. The difference in the occurrence rate of SSR loci may be due to the differences in transcriptome sequencing technologies and platforms, SSR determination parameters and software types, or it may be related to the differences and specificity of genetic characteristics among species.

By consulting a large number of recent literatures, it was found that the dominant repeat motifs of most medicinal plants were mononucleotide, dinucleotide and trinucleotide. In this study, the dominant repeat motif was mononucleotide, followed by dinucleotide and trinucleotide, accounting for 55.51%, 22.87% and 19.02% of the total SSR loci, respectively, which was consistent with the research results of *Eleutherococcus senticosus* (Li *et al.*, 2019), *Sambucus nigra* (Yao *et al.*, 2019), *Gleditsia sinensis* (Lin *et al*., 2017) and *Eucommia ulmoides* (Feng *et al.*, 2016), but the proportion of mononucleotide, dinucleotide and trinucleotide was different. However, although the dominant motifs of *Prunella vulgaris* were mononucleotide, dinucleotide and trinucleotide, dinucleotide accounted for the largest proportion of repeat motifs, followed by mononucleotide and trinucleotide (Wang *et al.*, 2018). The dominant motifs of *Camellia fascicularis* (Ye *et al*., 2019), *Psammosilene tunicoides* (Ye *et al*., 2019), *Platycodon grandiflorum* (Kim *et al*., 2015) and *Coriandrum sativum* L. (Tulsani *et al*., 2019) were dinucleotide and trinucleotide. Blackberry took mononucleotide and trinucleotide as the dominant motifs (Li *et al*., 2016). In addition, among the SSR loci in the transcriptome of *S. hexandrum*, the number of repeats of A/T was the most abundant, accounting for 53.55% of the total number of SSR loci, which was the dominant repeat unit, followed by AG/CT, AT/AT and AAG/CTT, accounting for 13.31%, 6.71% and 5.23%, respectively. The dominant repeat units of *Eleutherococcus senticosus* (Li *et al.*, 2019), *Prunella vulgaris* (Wang *et al*., 2018), *Sophora flavescens* (Zhang *et al*., 2019), *Sambucus nigra* (Yao *et al*., 2019), *Gleditsia sinensis* (Lin *et al*., 2017), *Eucommia ulmoides* (Feng *et al*., 2016), *Houttuynia cordata* Thunb (Lin *et al*., 2014), BlackBerry (Li *et al*., 2016) and *Magnolia officinalis* (Dai *et al*., 2017) were very similar to *S. hexandrum*, However, the dominant repeat units of *Camellia fascicularis* (Ye *et al*., 2019), *Psammosilene tunicoides* (Ye *et al.*, 2019) and *Cordyceps sinensis* (Zhang *et al.*, 2019) were quite different from *S. hexandrum*. The causes of these phenomena may be related to the minimum number of repeats set in the SSR analysis software, as well as the growth environment of the species and the inherent genetic factors.

The repetitions of SSR loci of *S. hexandrum* were mostly concentrated in 5-13, accounting for 82.43% of the total SSR loci, with 10 times being the most frequent, followed by 5 times, accounting for 25.40% and 12.96% of the total SSR loci, respectively. Similarly, the repetitions of SSR loci in *Eucommia ulmoides* (Feng *et al*., 2016), *Magnolia officinalis* (Dai *et al*., 2017) and BlackBerry (Li *et al*., 2016) were also 10 times as the maximum, and the repetitions of SSR loci in most medicinal plants were within 5-13 times, but the maximum repeat times were different. For example, the maximum number of repeats of *Psammosilene tunicoides* (Ye *et al*., 2019) was 5, followed by 6, accounting for 28.95% and 26.68% of the total SSR loci, respectively. The maximum number of repeats of *Prunella vulgaris* (Wang *et al*., 2018), *Camellia fascicularis* (Ye *et al*., 2019) were 6 times. Although there were some differences in the concentration range of repetitions and the maximum number of repeats, the number of SSR loci in these medicinal plants decreased with the increase of repetitions. The reason why the number of SSR repetitions has changed is that the SSR sequence may be affected by selection pressure in the process of evolution, which makes SSR sequence slip in the process of replication and cause the original sequence length to increase (Sia *et al.*, 1997; Tóth *et al*., 2000). Changes in the number of repetitions may lead to gene mutations, which is a noteworthy point in the genetic evolution of species (Meng *et al*., 2019).

Conclusions

The study used high-throughput sequencing technology to analysis the transcriptome sequence of *S. hexandrum* germplasm resources from Gansu, Shaanxi and Tibet. SSR analysis was carried out based on the sequencing data, which obtained a batch of medium-level SSR loci. In the follow-up research work, excellent genes can be excavated, SSR primers can be designed and the SSR molecular markers can be exploited based on the data of the present research. Furthermore, the genetic map can be constructed by inductive statistics of the distribution of SSR loci in this study. The results can facilitate the genetic diversity study, systematic differentiation analysis, development of effective molecular markers, and directional breeding, possessing a great significance for the innovative, development and rational protection of *S. hexandrum* resources.

Acknowledgments

The authors acknowledge the Projects for National Natural Science Foundation of China (Nos. 81803659 and 31902129) and the Programs of Doctoral Scientific Research Start-up Foundation of Henan University of Science and Technology (Nos. 13480075 and 13480019). The authors also thank other colleagues in the same laboratory for helpful comments on the manuscript.

References

Anonymous. 2011. Editorial Committee of Flora of China. *Flora of China*. Vol: 19. Science Press, Beijing; Missouri Botanical Garden Press, St. Louis.

Ayres, D.C. and L.C. Kee. 1976. Lignans and related phenols. Part XV. Remote substituent effects on the rates and products of some reactions of aryltetrahydronaphthalenes. *J. Chem. Soc., Perkin Trans.*, 1, 8: 832-836.

Canel, C., F.E. Dayan, M. Ganzera, A. Rimando, C. Burandt, I. Khan and R.M. Moraes. 2001. High yield of podophyllotoxin from leaves of *Podophyllum peltatum* by in situ conversion of podophylotoxin 4-O-β-Dglucopyranoside. *Planta Med*., 67: 97-99.

Chen, D.X., J.F. Zhao, X. Liu, C.H. Wang and G.Y. Zhong. 2013. Genetic diversity and genetic structure of endangered wild *Sinopodophyllum emodi* by start codon targeted polymorphism. *China J. Chin. Mater. Med*., 38: 278-283.

Dai, J., X.D. Shi, Y.X. Gu, Y.Z. Sheng and G.Q. Zhuang. 2017. Development and function analysis of SSR molecular markers in *Magnolia officinalis* transcriptome. *Chin. Tradit. Herb. Drugs.*, 48: 2726-2732.

Dong, X., C.J. Ruan, L.R. Liu, R. Yan, P. Han and Y. Pang. 2019. Development of SSR markers for target genes in *Xanthoceras sorbifolia* based on high throughput RNA-seq technology. *Mol. Plant Breed.*, 17: 2943-2949.

Feng, Y.Z., F.D. Li, Q.Q. Wei, W.J. Mo, L. Wang, D.G. Huang and J.M. Fu. 2016. Microsatellites characteristics of transcriptomic sequences from *Eucommia ulmoides* Oliv. based on RNA-Seq. *J. China Agric. Univ.*, 21: 68-79.

Fu, L.G. 1991. *Red Book of Chinese*. Science press, Beijing.

Giri, A. and M.L. Narasu. 2000. Production of podophyllotoxin from *Podophyllum hexandrum*: A potential natural product for clinically useful anticancer drugs. *Cytotechnology*, 34: 17-26.

Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury and Q. Zeng. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol*., 29: 644-652.

Holthuis, J.J.M. 1988. Etoposide and teniposide. Bioanalysis, metabolism and clinical pharmacokinetics. *Pharm. Weekbl. Sci.*, 10: 101-116.

Imbert, T.F. 1998. Discovery of podophyllotoxins. *Biochimie*, 80: 207-222.

Jia, Y., J.Q. Bai, M.L. Liu, Z.F. Jiang, Y. Wu, M.F. Fang and Z.H. Li. 2019. Transcriptome analysis of the endangered *Notopterygium incisum*: Cold-tolerance gene discovery and identification of EST-SSR and SNP markers. *Plant Diver.*, 41: 1-6.

Kamil, W.M. and P.M. Dewick. 1986. Biosynthetic relationship of aryltetralin lactone lignans to dibenzylbutyrolactone lignans. *Phytochemistry*, 25: 2093-2102.

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

Lau, W. and E.S. Sattely. 2015. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science*, 349: 1224-1228.

Li, W.L., C.H. Zhang, H.Y. Yang, W.L. Wu and L.F. Lyu. 2016. SSR marker information mining in blackberry from transcriptome sequences. *Acta Hort.*, 1133: 97-102.

Li, X.R., J. Ding, R. Chengjiang, Y. Zhou and S. Yang. 2019. Development of SSR Markers of Target-Genes in *Eleutherococcus senticosus* by RNA-seq Technology. *Mol. Plant Breed.*, 17: 4722-4728.

Lin, F.R., J.L. Xing, Y.Q. Meng, P. Huang and Y.Q. Zheng. 2017. Development and assessment of EST-SSR for *Gleditsia sinensis* Lam. based on transcriptome sequences. *J. Plant Genet. Res.*, 18: 148-154.

Lin, W., S. Li, S. Liu, A. He, W. Dan, W. Jie, Y. Tang and X. Wu. 2014. Transcriptome analysis of *Houttuynia cordata* Thunb. by illumina paired-end rna sequencing and SSR marker discovery. *Plos One*, 9: e84105.

Meng, W., Y.L. Jiang, L. Zhang, T.Y. Yang and J.G. Zhou. 2019. SSR loci information analysis in the transcriptome of Burbot (*Lota lota*) based on RNA-seq. *Freshw. Fish.*, 49: 10-14.

Moraes, R.M., H. Lata, E. Bedir, M. Maqbool, K. Cushman, J. Janick and A. Whipkey. 2002. The American Mayapple and its potential for podophyllotoxin production. In: (Eds.): Janick, J. and A. Whipkey. *Trends in New Crops and New uses*. ASHS Press, Alexandria, VA, pp. 527-532.

Mortensen, G.L. and H.K. Larsen. 2010. The quality of life of patients with genital warts: a qualitative study. *BMC Public Health*, 10: 113-140.

Nandagopal, K., M. Halder, B. Dash, S. Nayak and S. Jha. 2017. Biotechnological approaches for production of anti-cancerous compounds resveratrol, podophyllotoxin and zerumbone. *Curr. Med. Chem.*, 24: 4693-4717.

Ou, D.D., G. Zhao, Y. He, P. Wang, Z. Xu, H. Kong, J. Zhang and Y. Xu. 2019. Genetic analysis and gene mapping for melon rind color. *Chin. Agric. Sci. Bull.*, 13: 70-75.

Rickard-Bon, S.E. and L.U. Thompson. 2003. The role of flaxseed lignans in hormone-dependent and independent cancer. In: (Eds.): Muir, A.D. and N.D. Westcott. *Flax, the genus Linum*. Taylor and Francis Inc, London, pp. 181-203.

Sa, K.J., E.L. Su, I.Y. Choi, K.C. Park and K.L. Ju. 2019. Development and characterization of new microsatellite markers for *Perilla frutescens* (L.) Britton. *Amer. J. Plant Sci.*, 10: 1623-1630.

Shang, M., J. Li, S. Cai, L. Ping, L. Xu and G. Xu. 2000. Studies on the chemical constituents of the fruit of *Sinopodophyllum emodi*. *Chin. Trad. Herb. Drugs*, 31: 569-571.

Shen, Q., D. Zhang, C.X. Zang, W.J. Zhu, Q.I. Yue, M.Y. Ren, X. Guan, Y.D. Zhang and B. Center. 2019. Genetic diversity analysis of *Sinopodophyllum hexandrum* (Royle) T.S. Ying germplasm accessions using ISSR markers. *J. Plant Genet. Res.*, 20: 129-136.

Sia, E.A., R.J. Kokoska, M. Dominska, P. Greenwell and T.D. Petes. 1997. Microsatellite instability in yeast: Dependence on repeat unit size and DNA mismatch repair genes. *Cell. Mol. Biol.*, 17: 2851-2858.

Stähelin, H.F. and W.A. Von. 1991. The chemical and biological route from podophyllotoxin glucoside to etoposide: ninth cain memorial award lecture. *Cancer Res.*, 51: 5-15.

Tóth, G., Z. Gáspári and J. Jurka. 2000. Microsatellites in different eukaryotic genomes: Survey and analysis. *Genom. Res.*, 10: 967-981.

Tulsani, N.J., R. Hamid, F. Jacob, N.G. Umretiya, A.K. Nandhae, R.S. Tomar and B.A. Golakiya. 2019. Transcriptome landscaping for gene mining and SSR marker development in Coriander (*Coriandrum sativum* L.). *Genomics*, 112: 1-9.

Wang, H.M., S.U. Xiu-Hong, C.M. Dong, Y.H. Zhu and S.O. Pharmacy. 2018. Data mining of simple sequence repeats in *Prunella vulgaris* L. transcriptome. *Mod. Chin. Med.*, 20: 29-33.

Wang, L., Z.K. Wang, J.B. Chen, C.Y. Liu, W.L. Zhu, L.Y. Wang and L.H. Meng. 2015. De Novo Transcriptome Assembly and Development of Novel Microsatellite Markers for the Traditional Chinese Medicinal Herb, *Veratrilla baillonii* Franch (Gentianaceae). *Evol. Bioinform. Online.*, 11: 39-45.

Wang, P., L. Yang, E. Zhang, Z. Qin, H. Wang, Y. Liao, X. Wang and L. Gao. 2017. Characterization and development of EST-SSR markers from a cold-stressed transcriptome of centipedegrass by illumina paired-end sequencing. *Plant Mol. Biol. Rep.*, 35: 215-223.

Wang, X.C., H.L. Tan, C. Zhao, L.Z. Meng and S.C. Fan. 2015. Assembly and characterization of the transcriptome and development of SSR markers in *Forsythia suspensa* based on RNA-Seq technology. Sci*entia Sinica Vitae*, 45: 301-310.

Xiao, M., L.I. Qun, L. Guo, L. Tang, L. Wang and F. Chen. 2015. RAPD analysis of genetic diversity of an endangered species *Sinopodophyllum hexandrum* (Royle) Ying from Western Sichuan Province, China. *Acta Ecol. Sin.*, 35: 1488-1495.

Xiao, M., Q. Li, G. Liang, L. Tong, W.X. Duan, W.X. He, L. Wang, and C. Fang. 2006. AFLP analysis of genetic diversity of the endangered species *Sinopodophyllum hexandrum* in the Tibetan region of Sichuan Province, China. *Biochem. Genet.*, 44: 44-57.

Xiao, M., Q. Li, L. Wang, L. Guo, L. Jing, T. Lin and C. Fang. 2006. ISSR analysis of the genetic diversity of the endangered species *Sinopodophyllum hexandrum* (Royle) Ying from western Sichuan Province, China. *J. Integr. Plant Biol.*, 48: 1140-1146.

Yan, J., T. Xiao, Q. Yuan, P. Xie, Z. Peng, Y. Mao and M.A. Yuhua. 2018. SSR and SNP polymorphism analysis based on *Hylocereus polyrhizus* transcriptome. *Chin. J. Trop. Crops*, 39: 1338-1343.

Yao, X., S. Li, F. Ren, Q. Li, X. Liu, D. Wu and L. Yan. 2019. Study on SSR fluorescent markers based on transcriptome of *Sambucus nigra*. *J. Cen. South. Univ. T.*, 39: 123-129.

Ye, P., X. Li, C. Jie, X. Qu, Z. Dong, H. Lei, C. Li, P. Xin and S. Yang. 2019. The distribution and characteristics of SSR in transcriptome of *Psammosilene tunicoides*. *J. Yunnan Agric. Univ.*, 34: 808-814.

Ye, P., X. Li, J. Tang, B. Li, G. Zhang, C. Liu, H. Lei and P. Xin. 2019. Distribution and characteristics of SSR in transcriptome of *Camellia fascicularis*. *J. Cent. South Univ.*, 39: 86-91.

Yousefzadi M., M. Sharifi M. Behmanesh E. Moyano, M. Bonfill, R.M. Cusido and J. Palazon. 2010. Podophyllotoxin: Current approaches to its biotechnological production and future challenges. *Eng. Life Sci.*, 10: 281-292.

Zhang, F.C. 2009. Biological characteristics and cultivation techniques of medicinal plant *Sinopodophyllum emodi*. *Pract. For. Tech.*, 9: 45-46.

Zhang, H., F. Wang, J. Bai, Q. Dong and J.L. Guo. 2019. SSR loci information analysis in transcriptome of *Cordyceps sinensis*. *China J. Chin. Mater. Med.*, 44: 4605-4611.

Zhang, N., M. Yin, Q. Tan, Y. Wen, Y. Wang and J. Wang. 2019. Analysis of SSR loci and gene function annotation in transcriptome of *Sophora flavescens*. *Jiangsu Agric. Sci.*, 47: 41-44.

Zhao, Z., H. Zheng and C. Zhang. 2000. Development of molecular markers and their application in botany research. *Life Sci. Res.*, 4(2): 68-74.

Zhou, H., X. Zhang, T.Y. Liu and F.X. Yu. 2012. Data processing and gene discovery of high-throughput transcriptome sequencing. *Jiangxi Sci.*, 30: 607-611.

(Received for publication 5 April 2021)