Research Report

Open Access

SSR Site Information Analysis of The Transcriptome of *Glehnia littoralis*Li Lingli¹, Zhou Yifeng¹, Xu Ming¹, Li Mimi^{1,2} ✉, Xu Zenglai¹ ✉¹ Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing, 210014, China² The Jiangsu Provincial Platform for Conservation and Utilization of Agricultural Germplasm, Nanjing, 210014, China✉ Corresponding author email: limm001@163.com; xuzl5194@sohu.comMedicinal Plant Research, 2021, Vol.11, No.3 doi: [10.5376/mpr.2021.11.0003](https://doi.org/10.5376/mpr.2021.11.0003)

Received: 09 Dec., 2021

Accepted: 17 Dec., 2021

Published: 24 Dec., 2021

Copyright © 2021 Li et al., This article was first published in Molecular Plant Breeding in Chinese, and here was authorized to translate and publish the paper in English under the terms of Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:

Li L.L., Zhou Y.F., Xu M., Li M.M., and Xu Z.L., 2021, SSR site information analysis of the transcriptome of *Glehnia littoralis*, Medicinal Plant Research, 11(3): 1-6 (doi: [10.5376/mpr.2021.11.0003](https://doi.org/10.5376/mpr.2021.11.0003))

Abstract *Glehnia littoralis*, also known as Beishashen, is a medicine and food homologous plant and is one of the national second-class protected plants. The natural populations of *G. littoralis* are scarce. In this study, MISA (MicroSatellite) was used to search and analysis the SSRs from the transcriptome data of *G. littoralis*. As a result, 14 421 SSRs were identified in 78 828 unigenes, accounting for 15.43%. The numbers of SSR loci ranged from 1 to 6. The mononucleotide and dinucleotide are the most frequent types, accounting for 39.98% and 40.98%, respectively. The repeat motifs of SSR ranged from 5 to 24, of which 6 and 10 were the most, with 3 102 and 2 672 respectively, accounting for 21.51% and 18.53%. A total of 2 894 pairs of primers were designed by Primer 6.0 software, of which, 20 pairs of primers were randomly selected for amplification and validation, and 12 pairs of primers successfully amplified. The above results showed that the SSR loci in the transcriptome of *G. littoralis* provided technical basis for the construction of genetic map, analysis of genetic diversity, species identification and breeding of *G. littoralis*.

Keywords *Glehnia littoralis*; Transcriptome; SSR site

Glehnia littoralis, which is a perennial herb, belongs to the family Apiaceae. Its rhizome and roots are used as medicine, named Beishashen. As for the historical records of Beishashen, it was first found in the Shennong's Classic of Material Medica. Tao Hongjing, a famous medical expert, included *G. littoralis*, *Panax Ginseng*, *Scrophularia ningpoensis*, *Salvia miltiorrhiza* and *Sophora flavescens* as 'Wu Shen' (Wang et al., 2015). Beishashen tastes sweet and has obvious effect of nourishing yin. It can also ventilate lung, dissipate phlegm and promote the production of body fluid. Recent studies had found that it has the effect of anti-aging (National Pharmacopoeia Committee, 2010, China Medical Science Press, Beijing, pp. 699). Beishashen is not only a famous Chinese medicinal material, but also one of the medicine and food homologous plants. It has the effect of improving immune, so it is very popular in China and Japan, even in Southeast Asia and the United States (Bi et al., 2006, China Agricultural Technology Extension, (1): 40-41). *Glehnia* is a monotypic genus. The natural populations of *G. littoralis* are mainly distributed in the coastal beaches along the Pacific rim of East Asia and North America. In China, they are mainly distributed along the coasts of Liaoning, Shandong, Jiangsu, Zhejiang, Taiwan, Fujian, Guangdong and Hainan (Shan and She, 1992, Flora of China, Beijing: Science Press, 55(3): 77).

Since the reform and opening up, with the rise of China's economy, the booming of tourism industry, the construction of ports and the development of seaside tourism, coastal beaches have been severely destroyed by humans and the living environment of *Glehnia littoralis* has become narrow. The natural population of *G. littoralis* is endangered due to over excavation. In 1992, *G. littoralis* was listed in the Red Book of Chinese Plants, which is one of the national second-class protected plants (Fu, 1992) and is known as the 'giant panda' in the plant world (Liu et al., 2017). With the development of cultivation technique of *G. littoralis*, artificial cultivation has become the main source of Beishashen. However, the potential high-quality provenances and genes of natural populations of *G. littoralis* are important for the variety improvement of artificial cultivation. At present, the research on *G. littoralis* was limited to phytochemistry and cultivation, and the research on its genetic diversity and genetic differentiation is not comprehensive.

At present, a variety of molecular marker techniques have been used to analyze the genetic diversity of endangered plants, such as randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), simple sequence repeats (SSR) and so on. SSR has many advantages, such as co-dominant, extensive coverage, high levels of allelic diversity and easy detection. Therefore, SSR is used extensively in research on conservation biology of endangered plants (Hou et al., 2018; Zhao et al., 2018).

1 Results and Analysis

1.1 Statistical results and distribution of SSR loci in transcriptome of *Glehnia littoralis*

After transcriptome of *G. littoralis* assembly, a total of 78 828 unigenes were obtained. The SSR loci were searched by MISA software, and 14 424 SSR sites were retrieved from 12 166 unigenes; The occurrence and frequency of SSR were 15.43% and 18.29% respectively; There were 10 267 unigenes containing only a single SSR locus, there were 1 899 unigenes containing more than one SSR locus, and 1 079 composite SSR loci were retrieved (Table 1).

Table 1 Statistical analysis of SSR loci in transcriptome of *Glehnia littoralis*

Type	Statistic
Number of sequence	78 828
Total length of sequence (bp)	97 984 098
Number of SSR loci	14 421
Number of sequence containing SSR loci	12 166
Number of sequence containing more than 1 SSR locus	1 899
Number of sequence containing composite SSR loci	1 079
Frequency of SSR loci (%)	15.43

The mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide repeats were found from the transcriptome data of *G. littoralis*. The statistics of the number of SSR sites of different repeat types showed that the number of SSR sites of different repeat types varied. Among them, the number of mononucleotide and dinucleotide repeats were the largest, accounting for 39.98% and 40.98%, respectively, followed by trinucleotide repeats, accounting for 17.54%, and tetranucleotide, pentanucleotide and hexanucleotide repeats were very few, accounting for 1.50% together of the total SSRs. From the distribution of SSR sites in the transcriptome data of *G. littoralis*, it can be seen that the average distance of the distribution of SSR sites of different repeat types was also very different. The average distance of the distribution of mononucleotide and dinucleotide repeats was 16.99 kb/SSR and 16.58 kb/SSR respectively, the average distance of the distribution of trinucleotide repeats was 38.74 kb/SSR, and the number of tetranucleotide, pentanucleotide and hexanucleotide repeats was very small, therefore, the average distance of their distribution was 675.75 kb, 2 332.95 kb and 3 378.76 kb respectively. The average distance of the distribution of all SSR sites in transcriptome of *G. littoralis* was 6 794 bp/SSR (Table 2).

Table 2 Statistical analysis of SSR repeat types in transcriptome sequences of *Glehnia littoralis*

Repeat type	Number	Proportion (%)	Frequency (%)	Average distance (kb)	Total length (bp)	Average length (bp)
Mononucleotide	5766	39.98	7.31	16.99	70 858	12
Dinucleotide	5910	40.98	7.50	16.58	86 300	15
Trinucleotide	2529	17.54	3.21	38.74	43 128	17
Tetranucleotide	145	1.01	0.18	675.75	3 024	21
Pentanucleotide	42	0.29	0.05	2 332.95	1 125	27
Hexanucleotide	29	0.20	0.04	3 378.76	1 032	36

1.2 SSR repeat types in the transcriptome sequence of *Glehnia littoralis*

The total length of all SSR sites was 205 467 bp, and the average length of each SSR site was about 14 bp. The average length and number of SSR sites of different repeat types were counted, and the results showed that the

average length of SSR sites of mononucleotide, dinucleotide, trinucleotide, tetranucleotide, pentanucleotide and hexanucleotide repeats were 12, 15, 17, 21, 27 and 36 bp respectively, and the number of SSR sites were 5 677, 5 910, 2 529, 145, 42 and 29 times, accounting for 39.98%, 40.98%, 17.54%, 1.01%, 0.29% and 0.20%, respectively. The number of SSR repeats of mononucleotide and dinucleotide repeat types was the largest, which was 10 and 6 respectively; The most SSR sites of trinucleotide, tetranucleotide and pentanucleotide repeat types were 5 repeats; The SSR sites of hexanucleotide repeat types were mostly 5 and 6 repeats. The number of SSR sites decreased with the increase of the number of SSR repeats of different repeat types. When the number of repeats was larger than 10, the number of mononucleotide repeat types accounted for the largest proportion. Among the mononucleotide repeat types, A and T were the dominant repeat units, accounting for 46.41% and 47.69% respectively; Among the dinucleotide repeat types, GA and TC were the dominant repeat units, accounting for 20.52% and 18.14% respectively. The minimum times of SSR repeats was 5, the maximum times of SSR repeats was 24, and the largest number of SSR sites repeated 6 and 10 times, with 3 102 and 2 672 sites respectively, accounting for 21.51% and 18.53% of the total number of SSRs, respectively (Figure 1).

Figure 1 The number of SSR repeat types in the transcriptome of *Glehnia littoralis*

1.3 Design and validation of SSR primers

In order to test the polymorphic of primers, the unigenes with above SSR sites were screened with a length longer than 20 bp that were not at the beginning and end of the sequence. For the selected SSR sites, Primer 6.0 (Clarke and Gorley, 2006) software was used to design primers for unigenes containing SSR sites. A total of 2 894 pairs of primers were designed, accounting for 20.06%. 20 pairs of primers were randomly selected to carry out in 15 individuals of 3 populations of *G. littoralis*, including 5 individuals in each population. The amplified products were detected by polyacrylamide gel electrophoresis. According to results, 12 pairs of primers that can successfully amplify. Among them, 11 pairs had the same fragment size as the expected (Figure 2), and the amplification success rate was 55%.

2 Discussion

Because the transcriptome was derived from the gene data directly involved in expression, its practicability had very significant advantages, which also significantly increased the candidate sequence of SSR. These advantages were of great significance for species with relatively less genomic information (Wang et al., 2016). Existing research data showed that using transcriptome data to screen SSR site had extensive coverage, high polymorphism and effective amplification, which was suitable for the development and utilization of SSR sites (Jia et al., 2019).

Figure 2 Polyacrylamide gel electrophoresis results of three primers

The development and application of SSR sites had certain universality in different species, and were of great significance in plant conservation (Zhang et al., 2019). For example, after GO and KEGG functional annotation in the transcriptome sequence of *Ophiocordyceps sinensis*, it was found that unigenes containing SSR locus was mainly related to genetic and environmental functions, which laid a foundation for resource protection and gene function research of *O. sinensis* (Zhang et al., 2019). The development and application of SSR sites was conducive to the construction of genetic linkage map, which was benefit for the genetics, functional genomics and genetic breeding of many species. It was also conducive to the study of genetic relationship and genetic diversity among species. Moreover, it was helpful to understand the evolutionary mechanism of biodiversity and to analyze the distribution of gene resources, which was of great significance for the study of species adaptability and provided a theoretical basis for the rational allocation, utilization and protection of plant germplasm resources. Qin (2007) used SSR marker to study the genetic diversity and genetic structure of *Abies yuanbaoshanensis*, so as to provide basis for formulating scientific protection measures of *A. yuanbaoshanensis*. It was conducive to variety identification and fingerprint construction. Because SSR site analysis started from the gene level, it had higher accuracy than species morphological analysis. In the SSR experiment of *Reaumuria trigyna*, it was found that three pairs of primers had obvious identification effect on *R. songarica* (Qi, 2015).

In this study, the distribution frequency, distribution distance, average length, proportion and nucleotide distribution of SSR sites in the transcriptome of *Glehnia littoralis* were analyzed. At present, SSR markers were used in a variety of plants, such as in the transcriptome sequence of *Zingiber officinale*, the frequency of SSR was 9.39% (Zou et al., 2016); In the transcriptome sequence of *Polygonatum cyrtonema*, the frequency of SSR was 9.73% (Chen et al., 2020); In the transcriptome sequence of *Bergenia purpurascens*, the frequency of SSR was 9.50% (Diao et al., 2019). Compared with them, the number of SSR sites in the transcriptome sequence of *G. littoralis* was significantly larger. The frequency of SSR in the transcriptome sequences of *Ophiocordyceps sinensis*, *Tussilago farfara* and *Sambucus nigra* were 34.99%, 24.75% and 24.87% respectively (He et al., 2019; Yao et al., 2019; Zhang et al., 2019). In comparison, the frequency of SSR in these transcriptome sequences was slightly higher than that in *G. littoralis*. There were many reasons, such as species differences, SSR search criteria and so on. The main repeat types of SSR sites in the transcriptome of *G. littoralis* were mononucleotide and dinucleotide repeats. In mononucleotide repeat types, A and T were dominant repeat units, and in dinucleotide repeat types, GA and TC were dominant repeat units, which was similar to the transcriptome of *Carya cathayensis* (Jia et al., 2019), *Zingiber officinale* (Zou et al., 2016), *Polygonatum cyrtonema* (Chen et al., 2020), *Tussilago farfara* (He et al., 2019), *Bergenia purpurascens* (Diao et al., 2019) and other plants, but was different from the transcriptome of *Ophiocordyceps sinensis* with a large number of C/G, CG/CG, CCG/CGG SSR motifs (Zhang et al., 2019). The high frequency, density and types of SSR sites in the transcriptome of *G. littoralis* laid a foundation for the research of genetic diversity and the development of molecular marker assisted breeding of *G. littoralis*.

3 Materials and Methods

The transcriptome data of *Glehnia littoralis* were derived from the data (SRX547159) that published by NCBI (<http://www.ncbi.nlm.nih.gov/>). Taking the roots, stems and leaves of *Glehnia littoralis* as materials, the transcriptome of *G. littoralis* was sequenced by using the sequencing platform Illumina HiSeq 2000. The data was submitted by the Institute of Botany, Jiangsu Province and Chinese Academy of Sciences. The Sickle software (<https://github.com/najoshi/sickle>) and SeqPrep software (<https://github.com/jstjohn/SeqPrep>) were used to decontaminate the original transcriptome data of *G. littoralis* and obtain high-quality sequencing results, and the Trinity software (<http://trinityrnaseq.sourceforge.net/>) was used to *de novo* assemble of high-quality sequences. The MISA software (MicroSatellite identification tool, <http://pgrc.ipk-gatersleben.de/misa/>) was used to search the SSR sites in transcriptome of *G. littoralis*. The search criteria were conducted based on mono-, di-, tri-, tetra-, penta- and hexa-nucleotide motifs minimum number of 10, 6, 5, 5, 5 and 5 respectively. The results were calculated by Microsoft Excel, including SSR type, repeat numbers and distribution frequency.

The Primer 6.0 software (Clarke K.R., Gorley R.N., 2006, PRIMER v6: user manual/tutorial, PRIMER-E, Plymouth) was used to design primers for the detected SSR sites. The default parameters were as follows: primer length was 18~23 bp; annealing temperature was 55°C~65°C; the difference between upstream and downstream annealing temperature was ≤5°C; product length was 100~300 bp; GC content was 40%~70%; Avoid primer dimers, hairpin structures and mismatches.

The PCR amplification was carried out in 15 individuals of 3 populations of *G. littoralis* using 20 pairs of random selected primers. The total PCR volume was 20 µL: 10 µL 2×PCR Mix (provided by Dongsheng Biological Co., Ltd.); 1 µL forward primer; 1 µL reverse primer; 1 µL DNA (20 ng). The PCR amplification conditions were as follows: 94°C incubation for 4 min; 94°C denaturation for 30 s, 56°C 30 s, 72°C 30 s, for 35 cycles; and then 72°C extension for 10 min. The amplified products were detected by polyacrylamide gel electrophoresis, and the primers were initially screened according to the electrophoresis results.

Authors' contributions

LLL was the executor of the experimental design and research of this study; LLL completed data analysis and wrote the first draft of the manuscript; ZYF, XM and XZL participated in the experimental design and analyzed the experimental results; LMM was the conceiver and person in charge of the project, guiding the experimental design, data analysis, manuscript writing and revision. All authors read and approved the final manuscript.

Acknowledgments

This study was funded by National Natural Science Foundation of China (31600169).

References

- Chen Y.W., Liao R.J., Ye B.H., Song Q.Y., Yang Y., Li N., and Li H.B., 2020, Analysis on SSR loci in transcriptome and development of molecular markers in *Polygonatum cyrtoneura*, *Zhongcaoyao* (Chinese Traditional and Herbal Drugs), 51(1): 182-189
- Diao Y.C., Li P.X., Guo F.G., Wang S.Y., and Li J., 2019, SSR Loci characteristics and primer development of *Bergenia purpurascens* based on transcriptome sequences, *Fenzi Zhiwu Yuzhong* (Molecular Plant Breeding), 17(22): 7428-7432
- Fu L.G., eds, 1992, Chinese plant red book - Rare and endangered plants, Science Press, Beijing, China, pp.699
- He R.L., Han Y.L., Duan J.J., Liu J.Q., and Wang L.H., Development and polymorphism analysis of SSR molecular markers in *Tussilago farfara* transcriptome, *Zhongcaoyao* (Chinese Traditional and Herbal Drugs), 50(20): 5026-5032
- Hou L., Cui Y.H., Li X., Chen W., Zhang Z.Y., Pang X.M., and Li Y.Y., 2018, Genetic evaluation of natural populations of the endangered conifer *Thuja koraiensis* using microsatellite markers by restriction-associated DNA sequencing, *Genes* (Basel), 9(4): 218
<https://doi.org/10.3390/genes9040218>
PMid:29673217 PMCID:PMC5924560
- Jia Z.H., Xuan J.P., Zhang J.Y., Wang G., Jia X.D., and Guo Z.R., 2019, Analysis of SSR information in *Carya illinoensis* transcriptome, *Fenzi Zhiwu Yuzhong* (Molecular Plant Breeding), 17(10): 3305-3311
- Liu X.F., Zhang M.X., Chen Y., Cao S.Y., Wu J.B., Huang Y.H., and Chen S.T., 2017, Sample plot investigation and community characteristics of *Glehnia littoralis* in Changjiang'ao bay, Fujian province, *Zhongguo Yesheng Zhiwu Ziyuan* (Chinese Wild Plant Resources), 36(6): 57-64
- Qi Q., 2015, Genetic diversity and population structure of the endangered and endemic relict plant *Reaumuria trigyna* based on SSRs developed from transcription, Thesis for M.S., Life Science Institute, Inner Mongolia University, Supervisor: Wang Y.C, pp.1

- Qin Y.X., 2007, Study on conservation genetics of the endangered plant *Abies yuanbaoshanensis*, Thesis for M.S., Guangxi Normal University, Supervisor: Tang S.Q., pp.1
- Wang A.L., Li W.W., and Liu X., 2015, Genetic diversity analysis of *Glehnia littoralis* populations, an endangered medicinal plant, based on RAPD molecular markers, *Anhui Nongye Daxue Xuebao* (Journal of Anhui Agricultural University), 147(5): 148-152
- Wang J.Y., Liang Y.L., Hai M.R., Chen J.W., Gao Z.J., Hu Q.Q., Zhang G.H., and Yang S.C., 2016, Genome-Wide transcriptional excavation of *Dipsacus asperoides* unmasked both cryptic asperosaponin biosynthetic genes and SSR markers, *Front Plant Sci.*, 29(7): 339
<https://doi.org/10.3389/fpls.2016.00339>
- Yao J.X., Li S.W., Ren F., Li Q.H., Liu X.L., Wu D.J., and Yan L.P., 2019, Study on SSR fluorescent markers based on transcriptome of *Sambucus nigra*, *Zhongguo Linye Keji Daxue Xuebao* (Journal of Central South University of Forestry and Technology), 39(12): 123-129
- Zhai S.H., Yin G.S., and Yang X.H., 2018, Population genetics of the endangered and wild edible plant *Ottelia acuminata* in southwestern China using novel SSR markers, *Biochem. Genet.*, 56(3): 235-254
<https://doi.org/10.1007/s10528-018-9840-2>
PMid:29350309
- Zhang H., Wang F., Bai J., Dong Q., Tong X.X., Yuan Y., and Guo J.L., 2019, SSR loci information analysis in transcriptome of *Cordyceps sinensis*, *Zhongguo Zhongyao Zazhi* (China Journal of Chinese Materia Medica), 44(21): 4605-4611
- Zou Y., Huang K., Jiang Y.S., and Liu Y.Q., 2016, SSR Information in transcriptome of white ginger, *Zuowu Zazhi* (*Zingiber officinale* Roscoe), (3): 171-174