Transcriptome Analysis and EST-SSR Molecular Marker Development for Anemone shikokiana in Heterogeneous Habitats (Postprint)

Authors: Shan Xiaohan, well-being, Zhou Chunxia, Zhang Xin, Pang Yujuan, Lixia Li, Bian Fuhua

Date: 2023-12-26T00:00:00+00:00

Abstract

To investigate the ecological adaptation mechanisms of *Anemone shikokiana* in two distinct habitats—fully illuminated mountaintop shrubland and shaded coniferous-broadleaf mixed forest—and to develop EST-SSR molecular markers for this species, this study employed Illumina high-throughput sequencing technology to perform transcriptome sequencing on leaves of *A. shikokiana* at the flowering stage, obtaining functional annotation and differentially expressed genes. The results showed that: (1) Transcriptome sequencing yielded a total of 53,536 Unigene sequences, of which 27,448 were successfully annotated. (2) There were 5,635 differentially expressed genes, with 1,600 up-regulated and the remaining 4,035 down-regulated in *A. shikokiana* from mountaintop shrubland. A total of 2,460 differentially expressed genes were annotated to 2,533 tertiary terms in the GO database, and 1,051 differentially expressed genes were annotated to 113 metabolic pathways in the KEGG database. (3) The metabolic pathways involved in the adaptation of *A. shikokiana* to heterogeneous habitats primarily included the photosynthesis-antenna protein pathway and the flavonoid biosynthesis pathway. In the photosynthesis-antenna protein pathway, the *lhca5* gene was up-regulated while *lhca1-3* genes were down-regulated; in the flavonoid biosynthesis pathway, the *chs*, *c4h*, *f3'h*, *f3h*, *fls*, *ans*, *chi*, *ccoaomt*, and *hct* genes were all up-regulated. (4) A total of 7,146 SSR loci were identified from the transcriptome data of *A. shikokiana*, distributed across 6,006 Unigene sequences, comprising 106 types of repeat motifs, with mononucleotide repeats being the dominant repeat motif. Among 100 pairs of EST-SSR primers designed and synthesized, 68 pairs were validated as effective, of which 11 pairs were polymorphic, amplifying a total of 24 polymorphic fragments. These findings contribute to a deeper understanding of the adaptive regulatory mechanisms of *A. shikokiana* in different habitats, and represent the first development of EST-SSR molecular markers for this species, filling a gap in this area

and providing important molecular marker resources for the conservation and utilization of this organism.

Full Text

Transcriptome Analysis and Development of EST-SSR Molecular Markers in *Anemone shikokiana* Under Heterogeneous Habitats

Authors: SHAN Xiaohan, AN Kang, ZHOU Chunxia, ZHANG Xin, PANG Yujuan, LI Lixia, BIAN Fuhua
Affiliation: College of Life Sciences, Yantai University, Yantai 264005, Shandong, China

Abstract

This study investigated the ecological adaptation mechanisms of *Anemone shikokiana* in two distinct habitats—full-light hilltop scrub and shady mixed broadleaved-coniferous forest—using Illumina high-throughput sequencing of leaves collected during the flowering stage. Additionally, EST-SSR molecular markers were developed based on SSR locus distribution characteristics. The results were as follows: (1) A total of 53,536 Unigene sequences were obtained, of which 27,448 were successfully annotated. (2) After filtering low-abundance genes, 5,635 differentially expressed genes (DEGs) were identified, with 1,600 up-regulated and 4,035 down-regulated genes when comparing *A. shikokiana* from full-light hilltop scrub versus shady mixed coniferous forest. GO classification showed that 2,460 DEGs were annotated to 2,533 functional categories. Furthermore, 1,051 DEGs were involved in 113 KEGG pathways. (3) Comprehensive analysis of photosynthesis-antenna protein pathway-related genes revealed that *lhca5* expression was significantly higher, while *lhca1*, *lhca2*, and *lhca3* expression was significantly lower. Meanwhile, analysis of flavonoid biosynthesis pathway-related genes showed that *chs*, *c4h*, *f3'h*, *f3h*, *fls*, *ans*, *chi*, *ccoaomt*, and *hct* expression was significantly higher. (4) Using MISA software, 7,146 SSRs were identified in 6,006 Unigenes from the transcriptome data. Among these SSRs, single-nucleotide repeats were the dominant motif type across 106 repetitive motif categories. Of the 100 designed EST-SSR primer pairs, 68 were effective and 11 showed polymorphism, amplifying 24 polymorphic fragments. Overall, this study analyzed the adaptation mechanisms of *A. shikokiana* in heterogeneous habitats at the molecular level and, for the first time, developed EST-SSR molecular markers for this species, providing important molecular resources for its conservation and utilization.

Keywords: heterogeneous habitat; *Anemone shikokiana*; transcriptome; EST-SSR; ecological adaptation mechanisms

Introduction

Anemone shikokiana (Makino) Makino is a perennial herbaceous plant endemic to China. It is classified as Vulnerable (VU) on the IUCN Red List and is a key protected wild plant in Shandong Province. The species is primarily distributed at altitudes around 600 m, where it inhabits two distinct habitat types: full-light hilltop scrub and shady mixed broadleaved-coniferous forest understory. Previous studies have demonstrated that *A. shikokiana* exhibits significant phenotypic plasticity in response to heterogeneous habitats and altitude changes, with variations observed in leaf morphology, photosynthetic characteristics, and root structure. However, the molecular mechanisms underlying these adaptations remain unclear.

Transcriptome sequencing technology has become a powerful tool for studying plant responses to environmental heterogeneity. This approach enables large-scale discovery of functional genes and molecular markers. For non-model organisms without reference genomes, de novo transcriptome assembly can generate comprehensive sequence information for gene expression analysis and marker development. In recent years, transcriptome analysis has been successfully applied to study ecological adaptation in various plant species, including *Coreopsis tinctoria*, *Alisma orientale*, and *Medicago ruthenica*.

Expressed sequence tag-simple sequence repeat (EST-SSR) markers are developed from transcriptome data and offer several advantages, including high polymorphism, good reproducibility, and direct association with functional genes. These markers have been widely used in genetic diversity analysis, fingerprint construction, and molecular breeding of various species such as hemp (*Cannabis sativa*), Chinese sweetgum (*Liquidambar formosana*), and moso bamboo (*Phyllostachys vivax*). However, no EST-SSR markers have been reported for *A. shikokiana* to date.

To address this knowledge gap, this study employed Illumina high-throughput sequencing to analyze the transcriptome of *A. shikokiana* leaves from two contrasting habitats. The objectives were to: (1) characterize the transcriptome profile and identify differentially expressed genes associated with habitat adaptation; (2) develop and validate EST-SSR markers for future genetic studies; and (3) provide molecular insights into the ecological adaptation mechanisms of this rare and endangered species.

Materials and Methods

1.1 Plant Material and Sampling Fresh leaves of *A. shikokiana* were collected during the flowering stage from two distinct habitats in Kunyu Mountain, Yantai, Shandong Province: (1) full-light hilltop scrub (coded as D) and (2) shady mixed broadleaved-coniferous forest (coded as L). For transcriptome sequencing, three biological replicates were collected from each habitat, totaling

six samples. For EST-SSR validation, additional leaf samples were collected from five populations across different locations. Detailed sampling information is provided in and .

1.2 RNA Extraction and Transcriptome Sequencing Total RNA was extracted using the modified CTAB method. RNA quality was assessed using 1% agarose gel electrophoresis and a NanoDrop spectrophotometer. High-quality RNA samples (concentration $ $20ng \cdot \mu L\hat{}\{-1\}$) were sent for library construction and sequencing on the Illumina HiSeq platform. Paired-end sequencing generated raw reads that were subsequently filtered to obtain clean data.

1.3 De Novo Assembly and Functional Annotation Clean reads were assembled de novo using Trinity software. The resulting transcripts were clustered using CD-HIT to generate non-redundant Unigenes. Functional annotation was performed by searching against multiple databases including NR, KOG, GO, Swiss-Prot, eggNOG, and KEGG using diamond and HMMER tools. Pfam domain annotation was also conducted.

1.4 Differential Expression Analysis Gene expression levels were quantified as read counts using the DESeq package. Differentially expressed genes (DEGs) were identified based on the criteria of $|\log_2\text{FoldChange}| > 1$ and $p < 0.05$. GO and KEGG pathway enrichment analyses were performed on the DEGs.

1.5 SSR Marker Development and Validation SSRs were detected in the assembled Unigenes using MISA software with the following parameters: mono-nucleotide repeats $ $10, di-nucleotide repeats $ $6, tri-nucleotide to hexa-nucleotide repeats $ 5, $and minimum distance between two SSRs of 100 bp. Primer pairs were designed using Prim
$L\hat{}\{-1\}$), and template DNA. Amplification products were separated on 8% polyacrylamide gels and visualized by silver staining.

Results

2.1 Transcriptome Assembly and Annotation A total of 42.34 Gb of clean data was obtained, with Q30 values ranging from 93.68% to 94.20%. De novo assembly yielded 53,536 Unigenes with a total length of 57,909,453 bp, an average length of 1,081.69 bp, and an N50 of 1,502 bp. The assembly quality metrics D and L were 0.8689 and 0.8318, respectively, indicating high assembly completeness.

Functional annotation revealed that 27,448 Unigenes (51.27%) matched known sequences in public databases. The annotation rates for individual databases were: NR (50.52%), eggNOG (44.71%), Swiss-Prot (36.68%), Pfam (33.36%),

GO (32.76%), KOG (29.19%), and KEGG (10.24%). The Venn diagram of database annotations is shown in [Figure 1: see original paper].

2.2 Differentially Expressed Genes Comparative analysis between the two habitats identified 5,635 DEGs, including 1,600 up-regulated and 4,035 down-regulated genes in the shady forest habitat compared to the hilltop scrub. GO enrichment analysis annotated 2,460 DEGs to 2,533 GO terms, while KEGG analysis mapped 1,051 DEGs to 113 pathways.

The top enriched GO terms included biological processes such as abscisic acid metabolic process, menthol biosynthetic process, and sulfolipid biosynthetic process; cellular components such as cytosolic ribosomal subunits; and molecular functions such as $NADP^+$-dependent dehydrogenase activities. The volcano plot of DEGs is presented in [Figure 2: see original paper].

KEGG pathway enrichment revealed significant representation in photosynthesis-antenna proteins, flavonoid biosynthesis, ribosome, plant-pathogen interaction, MAPK signaling, and glutathione metabolism pathways. The top 30 enriched KEGG pathways are shown in [Figure 4: see original paper].

2.3 Key Pathway Analysis Photosynthesis-Antenna Proteins: Fifteen DEGs were annotated to the photosynthesis-antenna protein pathway. Notably, *lhca5* (light-harvesting complex I chlorophyll a/b binding protein 5) was significantly up-regulated, while *lhca1*, *lhca2*, and *lhca3* were down-regulated. Additionally, *lhcb1*, *lhcb2*, and *lhcb3* (light-harvesting complex II proteins) showed decreased expression in the shady habitat.

Flavonoid Biosynthesis: Eleven DEGs were identified in the flavonoid biosynthesis pathway, with all key enzyme-coding genes showing up-regulation in the shady habitat, including *chs* (chalcone synthase), *c4h* (cinnamate 4-hydroxylase), *f3'h* (flavonoid 3'-hydroxylase), *f3h* (flavanone 3-hydroxylase), *fls* (flavonol synthase), *ans* (anthocyanidin synthase), *chi* (chalcone isomerase), *ccoaomt* (caffeoyl-CoA O-methyltransferase), and *hct* (hydroxycinnamoyl transferase).

2.4 SSR Marker Development From the 53,536 Unigenes, 7,146 SSRs were identified in 6,006 Unigenes, with a frequency of 13.35% and an average density of one SSR per 8.1 kb. The SSR distribution showed that mono-nucleotide repeats were most abundant (3,248 SSRs, 45.59%), followed by tri-nucleotide (2,041, 28.56%), di-nucleotide (1,645, 23.02%), hexa-nucleotide (101, 1.41%), tetra-nucleotide (81, 1.13%), and penta-nucleotide (20, 0.28%) repeats. A total of 106 motif types were identified, with AG/CT being the most frequent di-nucleotide repeat and AAG/CTT the most common tri-nucleotide repeat.

One hundred EST-SSR primer pairs were synthesized and tested across four populations. Sixty-eight primers successfully amplified products, yielding 11 polymorphic loci that generated 24 polymorphic fragments. The polymorphic

markers showed clear population differentiation. Details of the polymorphic SSR markers are provided in .

Discussion

3.1 Ecological Adaptation Mechanisms in Heterogeneous Habitats The differential expression of photosynthesis-related genes reflects *A. shikokiana*'s adaptation to contrasting light environments. The up-regulation of *lhca5* and down-regulation of *lhca1-3* and *lhcb1-3* in shady habitats suggests a remodeling of the light-harvesting complex to optimize photon capture under low-light conditions. Lhca5 is known to play a specialized role in photoprotection and energy dissipation, while the reduced expression of major LHC proteins may represent an energy-saving strategy under light limitation.

The concerted up-regulation of flavonoid biosynthesis genes indicates enhanced production of secondary metabolites in response to shade stress. Flavonoids serve multiple protective functions, including UV screening, antioxidant activity, and signaling. The increased expression of *chs*, *chi*, *f3h*, *f3'h*, *fls*, *ans*, *ccoaomt*, and *hct* suggests elevated synthesis of flavonols, anthocyanins, and other phenolic compounds that protect against oxidative damage and facilitate acclimation to low-light conditions.

3.2 EST-SSR Marker Development and Application This study represents the first development of EST-SSR markers for *A. shikokiana*. The high amplification success rate (68%) and polymorphism level (11 polymorphic loci) demonstrate the effectiveness of our marker development pipeline. These markers provide valuable tools for future research on genetic diversity, population structure, and conservation genetics of this vulnerable species. The functional annotation of SSR-containing Unigenes links these markers to candidate genes involved in stress response and adaptation, enhancing their utility for marker-assisted selection in conservation programs.

References

BIAN J, WANG XN, CAO K, et al., 2023. Genetic structure analysis and fingerprint construction of hemp by EST-SSR analysis[J]. *J Plant Genet Resour*: 1–16[2023-10-20]. https://doi.org/10.13430/j.cnki.jpgr.20230531001.

CHEN LC, GU DZ, LIU Y, et al., 2018. Establishment of tissue culture regeneration system of *Anemone shikokiana*[J]. *Jiangsu Agric Sci*, 46(24): 47–50.

DAI MJ, 2022. Molecular cloning and functional analysis of CHIL and F3H in flavonoid biosynthesis pathway in mulberry[D]. Zhenjiang: Jiangsu Univ of Sci and Technology.

GUO LA, MO RY, TAN J, et al., 2021. Analysis on differential expressed genes in the transcriptome of *Atractylodes macrocephala* from different producing areas[J]. *Chin Med Mat*, 44(12): 2787–2792.

HOU YT, LIU B, 2010. *Anemone shikokiana* (Makino) Makino[J]. *Encyclopedic Knowl*, (12): 40.

JIANG Q, 2019. Genomic analysis and response factors under abiotic stress in *Oenanthe javanica*[D]. Nanjing: Nanjing Agricultural University.

LI H, FENG YH, TANG SS, et al., 2023. Development and validity evaluation of *Liquidambar formosana* EST-SSR primers based on transcriptome sequencing[J]. *Guihaia*, 43(2): 327–335.

LIU Q, 2014. Research on genetic diversity of *Anemone shikokiana*[D]. Jinan: Shandong Normal University.

LOU YF, GAN R, ZHU CL, et al., 2022. Development and application of EST-SSR markers based on transcriptome of *Phyllostachys vivax* McClure f. *aureocaulis* N. X. Ma[J]. *Plant Sci J*, 40(3): 355–364.

LU YH, YU FB, GUO JH, et al., 2021. Analysis of transcriptomic differences between winged and wingless bird cherry-oat aphid *Rhopalosiphum padi* on wheat[J]. *Acta Phytophy Sin*, 48(6): 1303–1309.

PANG YJ, SUN W, WANG XS, et al., 2022. Influence of soil heterogeneity on root structure of *Anemone shikokiana*[J]. *J Yantai Univ*, 35(1): 28–34.

SUN HN, LIU DY, LI MY, et al., 2022. Transcriptome data analysis of *Coreopsis tinctoria* based on high-throughput sequencing[J]. *Plant Breed*: 1–18[2023-10-20]. http://kns.cnki.net/kcms/detail/46.1068.S.20221017.1205.010.html.

TIAN HZ, GUO H, TIAN H, et al., 2022. Transcriptome characteristics analysis and EST-SSR marker development of *Capsicum annuum* under waterlogging stress[J/OL]. *Guihaia*: 1–14[2023-10-20]. http://kns.cnki.net/kcms/detail/45.1134.q.20221227.1548.004.html.

WANG YF, ZHU RJ, CHENG XP, 2023. Research on phenotypic plasticity of different diameter class *Bombax ceiba* in heterogeneous habitats[J]. *Ecol Sci*, 42(3): 61–66.

WANG YH, 2020. Identification and expression profiling of Lhc genes in Ramie[D]. Wuhan: Huazhong Agricultural University.

WANG Z, PANG YJ, LIU CL, et al., 2014. Distribution pattern and analysis of influencing factors of rare plant *Anemone shikokiana* (Makino) Makino[J]. *Bot Stud*, 34(4): 440–445.

WEI CB, GU W, TIAN R, et al., 2022. Transcriptomic data analysis of *Alisma orientale* tubers grew in different ecological environments by high-throughput sequencing[J]. *Mod Chin Med*, 24(12): 2342-2349.

WU RN, XU B, SHI FL, 2022. Analysis of SSR characteristics for *Medicago ruthenica* ( ) Zhilixing ( ) based on transcriptome data[J]. *J Northwest A & F*

Univ (Nat Sci Ed), 50(5): 1–8.

XIAO YZ, HAN SM, QIN Z, et al., 2020. Analysis of transcriptome sequencing and related genes of flavonoids biosynthesis from *Polygonatum kingianum*[J]. *J Henan Agric Univ*, 54(6): 931-940.

YAN MM, WEI GC, PAN XH, et al., 2008. A method suitable for extracting genomic DNA from animal and plant—modified CTAB method[J]. *J Anhui Agric Sci*, (20): 8488+8558.

YANG X, YANG N, YUAN QH, et al., 2021. Development and application of EST-SSR molecular markers in *Pinus bungeana*[J]. *J Beijing For Univ*, 43(7): 111.

YU WY, GAO Y, PANG YJ, et al., 2019. Response of leaf morphology and structure of *Anemone shikokiana* to heterogeneous habitats and altitude changes[J]. *Acta Ecol Sin*, 39(12): 4413–4420.

YUE XY, YUE W, YAN K, 2023. Adaptation of functional traits of *Phragmites australis* to different water habitats in the Yellow River Delta[J]. *J NE For Univ*, 51(3): 18–21.

YUE YH, YAN LJ, HUANG X, et al., 2022. Development EST-SSR markers based on transcriptome sequences of *Tilia miqueliana* Maxim[J]. *Plant Breed*: 1–18[2023-10-20]. http://kns.cnki.net/kcms/detail/46.1068.S.20220715.1525.008.html.

ZHANG CY, LONG Y, FENG J, et al., 2007. Transcriptional regulation of plant genes and its significance in biology[J]. *Hereditas*, (7): 793–799.

ZHENG Y, JIN GL, WU WR, 2012. Relationship between sequence completeness and polymorphism of microsatellites[J]. *Genom Appl Biol*, 31(6): 587–591.

ZHOU CM, JIANG LP, DONG GQ, et al., 2022. Analysis of differential expression genes of *Chrysanthemum indicum* var. *aromaticum* associated with different altitude habitats using RNA-seq technology[J]. *World Sci Technol-Mod Chin Med*, 24(4): 1356–1368.

ZOU FL, SHOU SY, YE ZZ, et al., 2004. Advances in the research on flavonoid biosynthesis and plant stress response[J]. *Chin J Cell Biol*, (1): 39–44.

BATSCHAUER A, 1993. A plant gene for photolyase: an enzyme catalyzing the repair of UV-light-induced DNA damage[J]. *Plant J*, 4(4): 705–709.

BUCHFINK B, CHAO X, DANIEL HH, 2015. Fast and sensitive protein alignment using DIAMOND[J]. *Nat Methods*, 12: 59–60.

CHEN YY, LU HQ, JIANG KX, et al., 2023. The flavonoid biosynthesis and regulation in *Brassica napus*: A review[J]. *Int J Mol Sci*, 24, 357.

FANG L, MARTRE P, JIN K, et al., 2022. Neglecting acclimation of photosynthesis under drought can cause significant errors in predicting leaf photosynthesis in wheat[J]. *Glob Change Biol*, 29, 505–521.

GANETEG U, KIMMEK F, JANSSON S, 2004. Lhca5—an LHC-type protein associated with photosystem I[J]. *Plant Mol Biol*, 54(5): 641–51.

GHANBARI MSR, TOHIDFAR M, AHMADIKHAH A, 2022. Gene profile analysis and molecular-physiological evaluation of tomato genotypes under drought stress[J]. *J Appl Biotechnol Rep*, 9(3): 747–62.

GUO LA, TAN J, DENG XS, et al., 2023. Integrated analysis of metabolome and transcriptome reveals key candidate genes involved in flavonoid biosynthesis in *Pinellia ternata* under heat stress[J]. *J Plant Res*, 136(3): 359–369.

HELENA S, MAKIO Y, ATSUSHI T, et al., 2023. Reversible down-regulation of photosystems I and II leads to fast photosynthesis recovery after long-term drought in *Jatropha curcas*[J]. *J Exp Bot*, 336–351.

LUCINSK R, JACKOWSKI G, 2006. The structure, functions and degradation of pigment-binding proteins of Photosystem II[J]. *Acta Biochim Pol*, 53(4): 693–708.

MANFRED G, BRIAN J, MORAN Y, et al., 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data[J]. *Nat Biotechnol*, 29(7): 644.

MOZZO M, MANTELLI M, PASSARINI F, et al., 2010. Functional analysis of Photosystem I light-harvesting complexes (Lhca) gene products of *Chlamydomonas reinhardtii*[J]. *Biochim Biophys Acta*, 1797: 212–221.

PANG YJ, LI LX, BIAN FH, 2020. Photosynthetic and ultrastructural adaptability of *Anemone shikokiana* leaves to heterogeneous habitats[J]. *Braz J Bot*, 43, 979–988.

ROZEN S, SKALETSKY H, 2000. Primer3 on the WWW for general users and for biologist programmers[J]. *Methods Mol Biol*, 132: 365–86.

SEBASTIAN B, THOMAS T, THOMAS M, et al., 2017. MISA-web: a web server for microsatellite prediction[J]. *Bioinform*, 33(16): 2583–2585.

SHEN N, WANG TF, GAN Q, et al., 2022. Plant flavonoids: Classification, distribution, biosynthesis, and antioxidant activity[J]. *Food Che*, 132531. https://pubmed.ncbi.nlm.nih.gov/35413752/.

SIMON A, WOLFGANG H, 2012. Differential expression of RNA-Seq data at the gene level EMBL. https://www.bioconductor.org/help/course-materials/2012/BioC2012/DESeq.pdf.

STORF S, STAUBER EJ, HIPPLER M, et al., 2004. Proteomic analysis of the photosystem I light-harvesting antenna in tomato (*Lycopersicon esculentum*)[J]. *Biochemistry-USA*, 43(28): 9214–9224.

TEMNYKH S, DECLERCK G, LUKASHOVA A, 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency,

length variation, transposon associations, and genetic marker potential[J]. *Genom Res*, 11(8): 1441–1452.

XIANG NY, LU BJ, YUAN T, et al., 2023. De Novo transcriptome assembly and EST-SSR marker development and application in *Chrysosplenium macrophyllum*[J]. *Genesis*, 14: 279.

XIONG S, TIAN N, LONG JH, et al., 2016. Molecular cloning and characterization of a flavanone 3-hydroxylase gene from *Artemisia annua* L.[J]. *Plant Physiol Biochem*, 105(5): 29–36.

ZHUANG WB, LI YH, SHU XC, et al., 2023. The classification, molecular structure and biological biosynthesis of flavonoids, and their roles in biotic and abiotic stresses[J]. *Molecules*, 28(8): 3599.

Note: Figure translations are in progress. See original paper for figures.