Journal of Forest Research

ISSN: 1341-6979 (Print) 1610-7403 (Online) Journal homepage: https://www.tandfonline.com/loi/tjfr20

Development of 15 novel microsatellite markers for a *Haloxylon ammodendron* (Amaranthaceae) using next-generation sequencing

Nyam-Osor Batkhuu, Sang-Chul Kim, Jei-Wan Lee & Kyung-Nak Hong

To cite this article: Nyam-Osor Batkhuu, Sang-Chul Kim, Jei-Wan Lee & Kyung-Nak Hong (2019): Development of 15 novel microsatellite markers for a *Haloxylon ammodendron* (Amaranthaceae) using next-generation sequencing, Journal of Forest Research, DOI: 10.1080/13416979.2019.1675253

To link to this article: https://doi.org/10.1080/13416979.2019.1675253

JOURNAL OF FOREST RESEARCH
https://doi.org/10.1080/13416979.2019.1675253

SHORT COMMUNICATION

Development of 15 novel microsatellite markers for a *Haloxylon ammodendron* (Amaranthaceae) using next-generation sequencing

Nyam-Osor Batkhuu[a], Sang-Chul Kim[b], Jei-Wan Lee[b] and Kyung-Nak Hong[b]

[a]Department of Environment and Forest Engineering, School of Engineering and Applied Sciences, National University of Mongolia, Ulaanbaatar, Mongolia; [b]Department of Forest Bio Resources, National Institute of Forest Science, Suwon-si, Gyeonggi-do, Republic of Korea

ABSTRACT

Microsatellite primers were developed in *Haloxylon ammodendron* (Amaranthaceae). This species is an ecologically important component of the desert ecosystem and is one of the main tree species used for restoration. These markers will facilitate population genetic studies within *Haloxylon ammodendron*. Fifteen polymorphic microsatellite markers were developed from Ion torrent data.The four to twenty-six alleles were detected for each locus, the levels of observed and expected heterozygosity ranged from 0.375 to 0.933 and form 0.413 to 0.931, respectively. In 6 populations of plants, the levels of observed and expected heterozygosity ranged from 0.613 to 0.730 and from 0.633 to 0.676, respectively. These new microsatellite markers will be useful in future studies of the phylogeography, reproductive genetics, and population genetics of *H. ammodendron*.

ARTICLE HISTORY
Received 20 June 2019
Accepted 27 September 2019

KEYWORDS
Amaranthaceae; *Haloxylon ammodendron*; microsatellite; Psammophyte; next-generation sequencing (NGS)

Introduction

Haloxylon Bunge *ex* Fenzl is a genus comprises 10 species of shrubs or subcanopy trees, belonging to the Amaranthaceae. Among them, *Haloxylon ammodendron* (C.A.Mey.) Bunge *ex* Fenzl is an ecologically important component of the desert ecosystem and is one of the main tree species used for restoration (Sheng et al. 2005). This tree species is widely distributed in Middle and Central Asia, through the Iran, West-Afghanistan, Turcomania, from Aralo-Caspian to Amu Darya, the lowland areas of Central Asia, Mongolia and China. As a dominant desert plant, *H. ammodendron* is known to play an important role in promoting the settlement and growth of other desert plants by maintaining the structure and function of the whole ecosystem, reducing wind speed, and improving the forest micro-climate (Shamsutdinov and Ubaidullaev 1988). Owing to human activities and climate change, natural populations of *H. ammodendron* have been destroyed or become extinct. Research on genetic differentiation and population structure is needed to protect the genetic resources of this species. Recently, expressed sequence tag simple sequence repeats (EST-SSR) markers for this species (Long et al. 2014) and *Haloxylon persicum* identifiable marker have been developed (Suo et al. 2012). But there was no information on the nuclear simple sequence repeats (nSSR) markers. nSSR marker is a codominant marker, which is of good repeatability, easy operation, and wide coverage and also shows higher polymorphism compared with other markers (Russell et al. 1997). Conservation and utilization of the ecological restoration of this important desert species is currently hampered by the paucity of genetic resources available for this species.

Microsatellites are the most frequently used genetic markers in a wide range of applications in population genetics, conservation biology and evolutionary biology. Their codominant nature, high levels of polymorphism, and reproducibility are particularly appropriate for estimating population structure and genetic diversity (Mariette et al. 2002). In recent years, the emergence of next-generation sequencing has enabled fast and cost-effective identification of microsatellite loci. The random sequence-based approach also has the advantage of identifying thousands of useful microsatellite sites in previously unexplored species. In the present study, we isolated microsatellite (simple sequence repeat, SSR) markers in *H. ammodendron* to investigate its genetic diversity.

Materials and methods

Dried-plant materials were obtained from 6 populations of *H. ammodendron* samples from the conservation area of the Southern Gobi Desert in Mongolia (44° 12′ N, 110° 01′ E; 20 individuals per group.) and its genomic DNA was isolated from dried leaves using a Plasmid SV mini kit (GeneAll Biotechnology, Seoul, Korea) following the manufacturer's instructions. The voucher specimens of this samples were deposited in the herbarium of National University of Mongolia (UBU; Table S1 in Supplementary Material). DNA was quantified using a spectrophotometer (ND-1000, NanoDrop Technologies, Wilmington, Delaware, USA). The fragment DNA library was constructed using the Ion Xpress ™ Plus Fragment Library Kit (Life Technologies, Carlsbad, California, USA) using one sample of good DNA purity (UBU20180071). Libraries were amplified with the Ion PI™ Hi-Q™ Sequencing 200 Kit (Life Technologies) and subsequently sequenced with the Ion PGM Sequencing 200 kit (Life Technologies) and the Ion PI™ Chip v3 kit (Life Technologies).

De novo assembly and analysis of raw files

In total, 71,890,513 paired reads were produced. The generated Raw file was assembled into the published chloroplast genome of *H. ammodendron* (NC027668) and mitochondrial

CONTACT tlsb -0.01w?>CONTACT Jei-Wan Lee leejeiwan@korea.kr Department of Forest Bio Resources, National Institute of Forest Science, 39 Ojeong-ro, Gwonseon-gu, Suwon-si, Gyeonggi-do, 16631, Republic of Korea
Supplemental data for this article can be accessed here.

genome of Related Genus (*Beta macrocarpa*, NC015994) using the Geneious (Kearse et al. 2012) to isolate only nuclear DNA reads. The isolated nuclear reads were assembled using SOAPdenovo2 ver. 2.04 (Luo et al. 2012). They were first combined to form longer fragments (i.e. contigs) and then mapped back to the contigs.

SSR mining and primer design

MISA (Thiel et al. 2003) was employed for microsatellite mining. nSSRs were considered to contain motifs with two to six nucleotides and a minimum of four contiguous repeat units. Based on MISA results, sequences longer than 500 bp were searched for randomly set repeat motifs. The criteria of no fewer than six repeat units for di-, five for tri-, and four for tetra- to penta-nucleotide repeats were adopted. Primer pairs were designed for 440 loci with Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/; Untergasser et al. 2012).

PCR amplification and validation of selected SSRs

The M13 (–19) sequence (5′-CACGACGTTGTAAAACGAC-3′; Imtiaz and Naz 2012) was affixed to the 5′ end of all forward primers to facilitate fluorescence labeling. To investigate the amplification of 440 loci, we conducted a preliminary PCR analysis with eight individuals. We selected 30 loci that amplified successfully using a GeneAmp PCR System 2700 Thermal Cycler (Applied Biosystems, Foster City, California, USA). All the amplifications were carried out in a 12 µL volume containing 1.2 µL 10× buffer (with Mg^{2+}), 1mM dNTP, 1µM M13 primer, 0.2µM primer pairs, 1U A-Star *Taq* DNA polymerase (BioFACT, Daejeon, Korea), and 20ng genomic DNA. The reaction cycle was 5 min at 95°C, followed by 30 cycles of 45 s at 95°C, 45 s at 56°C, 1 min at 72°C, a final extension at 72°C for 10 min, and storage at 4°C. Genotyping of the microsatellite fragments was conducted on an ABI PRISM 3730 Analyzer (Applied Biosystems). The fragment size was determined based on the GeneScan-500 ROX size standard (Applied Biosystems) using GeneMapper 5 software (Applied Biosystems) and manually checked for consistency and accuracy.

Data analysis

Allelic frequencies were used to calculate species and population measures of genetic diversity including the mean number of different alleles per locus (*A*), effective alleles (*Ae*), observed (*Ho*) and expected (*He*) heterozygosities using GenAlEx 6.5 (Peakall and Smouse 2006). Deviations from Hardy–Weinberg equilibrium (HWE) for each locus, and linkage disequilibrium (LD) between all loci were tested using the online version of GENEPOP (Rousset 2008). Null allele frequency (F_{Null}) was calculated using Cervus 3.0 (Marshall et al. 1998) for each locus. The presence of null alleles was checked by MICROCHECKER version 2.2.3 software (Van Oosterhout et al. 2004).

Results and discussions

Sequencing by ion proton™

By Ion Proton™ sequencing with an Ion PI™ chip, a total of 12.6 Gb data and 71,890,513 quality reads were obtained in a single sequencing run from the genomic DNA of one *H. ammodendron* individual. The length of the reads was average of 175 bp. Using Genious, 855,202 reads (1.19%) and 1,161,332 reads (1.64%) were removed with reference to the chloroplast of *H. ammodendron* (NC027668) and the mitochondria of *Beta macrocarpa* (NC015994), respectively. The result of SOAPdenovo2, all reads were assembled into 243,379 contigs with mean length of 778 bp, and the sum of the total lengths was 189,453,116 (Table S2). There was a total of 20,076 confirmed SSRs, 7,201 dinucleotides, 5,269 trinucleotides, 4,523 tetranucleotides, 2,656 pentanucleotides and 427 hexanucleotides (Table S3). Of these, a total of 440 nSSR marker candidates were identified as a result of searching for contigs capable of producing a pair of primers. We developed 30 nSSR markers in the presence of polymorphic and fifteen loci were discarded because of the presence of null alleles ($F_{Null} > 0.05$). Finally, we developed primer pairs for 15 loci.

Characteristics of microsatellites and nSSR polymorphism

Fifteen of the pairs of primers tested amplified high-quality PCR products that exhibited polymorphisms. Most of the microsatellite primers contained from dinucleotide to tetranucleotide repeat units. The *A* varied between 4 and 26 (mean, 10.4). The *Ho* varied between 0.375 and 0.933 (mean, 0.660) and the *He* between 0.413 and 0.931 (mean, 0.676; Table 1). Linkage disequilibrium was detected among four pairs of loci ($P < 0.01$; HAS006 and HAS275, HAS124 and HAS289, HAS006 and HAS327, HAS111 and HAS420). Null alleles are likely to occur in all loci, but do not occur at the same frequency across populations. The Null allele frequencies ranged from −0.0205 to 0.049 (Table 1).

Genetic diversity values for 6 populations of *H. ammodendron* were as follows (Table 2): The overall diversity in all populations analyzed was moderate and nSSR loci revealed moderate genetic diversity in the analyzed populations; the average number of *Ae* was 3.952 ± 0.289 (mean ± sd), the *Ho* was 0.661 ± 0.02, whereas *He* was 0.653 ± 0.018. The highest average number of *Ae* was detected in population BZ (4.216 ± 0.930), whereas the lowest number of *Ae* was found in population TZ (3.669 ± 0.598). The highest average number of *Ho* was detected in population KHE (0.730 ± 0.046), whereas the lowest number of *Ho* was found in population TZ (0.613 ± 0.054). The highest average number of *He* was detected in population DG1 (0.676 ± 0.036), whereas the lowest number of *He* was found in population TZ (0.633 ± 0.048). The highest average mean value of *F* was detected in population DG1 (0.037 ± 0.047), whereas the lowest number of *F* was found in population KHE (−0.104 ± 0.028).

In this study, we examined the genetic variations of the fifteen nSSR loci and identified a high level of polymorphism within each population (Table 2). The genetic diversity tended to be similar compared to other tree species with similar vitality (long-lived perennial: *He* = 0.680; widespread: *He* = 0.620; Seed dispersal (wind): *He* = 0.610, Nybom 2004). It is well known that species with large geographical distribution often have high genetic diversity (Hamrick et al. 1979). The sampled populations were separated by long distances. Our genetic analyses indicate that the *H. ammodendron* populations are characterized by a high level of genetic diversity (Table 2). Due to the frequently strong winds and the low coverage of vegetation

Table 1. Characteristics of 15 polymorphic microsatellite loci isolated from Haloxylon ammodendron.

Locus	Primer sequence (5′–′)	Repeat motif	Allele size range (bp)	*A*	*Ho*	*He*	F_{Null}		GenBank accession no.
HAS_006	F: AAAAGAGAAGACAACCGATGGA R: AGCGGACTGCTATGTGAAGG	$(CAT)^7$	220–229	4	0.375	0.428	0.0374		MH918744
HAS_016	F: TCTGAAAACAAGGGCGATTT R: CAACATGAAAAAGACTGTCACCTT	$(TATT)^4$	214–250	4	0.592	0.674	0.049		MH918745
HAS_111	F: GGTGCACATTGATTGTTCCA R: CCGAGCAAGAACTGCTAAGG	$(TTA)^7$	204–222	7	0.692	0.703	0.0066		MH918746
HAS_124	F: CATCGCTTAGACTTCAACCGTA R: CGGGGCGATATTAGGAAATC	$(TAA)^6$	193–217	7	0.425	0.413	−0.0087		MH918747
HAS_128	F: CAGCTGCGATTACAAAGGAG R: TCCACTTAACTCCCCTCACC	$(GA)^7$	219–241	8	0.608	0.655	0.0282		MH918748
HAS_164	F: GGTGGATGGAAGCACAGTTT R: GGGACACTCCATACCCTCCT	$(GA)^{10}$	243–255	7	0.508	0.546	0.0243		MH918749
HAS_275	F: GGTACGTACGCTCCTTTTGC R: TGTAATGAATCAGACTGAACACAA	$(GA)^8$	221–263	8	0.742	0.724	−0.0101		MH918750
HAS_281	F: AGCCTTGAATCCTTTGACCA R: TTAGGGATGGCGATTGGTAG	$(AC)^9$	189–229	17	0.883	0.898	0.0077	***	MH918751
HAS_312	F: TGATTTTAATGTGCTGGCTGA R: TCATCCATATTTCCGAAACAGA	$(GT)^8$	224–258	5	0.658	0.685	0.0157		MH918752
HAS_321	F: AAGACGGAGAAGGGTTGAGG R: AGCACATGTAAAGGGCAGGA	$(TC)^6$	190–222	10	0.433	0.457	0.0165	***	MH918753
HAS_374	F: ATAAGGGCGCTTTCATTGTC R: GCCGTACGACTGTACTCAGAC	$(TTAT)^{10}$	175–203	7	0.717	0.702	−0.0087	***	MH918754
HAS_388	F: TTGTTTCATTGGGCTTTTCC R: CCAACCCTAACAAAATATTCGAT	$(TCAT)^9$	210–234	7	0.650	0.639	−0.0069		MH918755
HAS_415	F: TTCGAAGTCCCATGTTAAGGA R: CTGGATTGGTTGCGGTATTT	$(AAG)^{13}$	149–221	24	0.867	0.904	0.0197		MH918756
HAS_420	F: CCCATAATCCCATGAGTCAA R: TTTTTCACTCCAGTGCTTCAAA	$(ACA)^{16}$	194–251	15	0.825	0.788	−0.0205		MH918757
HAS_429	F: GCACACACACTCCCTTGATG R: AATTTTCCCTTGCATTGCTG	$(AAT)^{15}$	166–247	26	0.933	0.931	−0.001	***	MH918758

Note: Annealing temperature was 56°C for all loci, *A*: number of alleles, *Ho*: Observed Heterozygosity, *He*: Expected Heterozygosity, F_{Null}: Null allele frequency, ***: Significant evidence ($p < 0.001$) against Hardy-Weinberg equilibrium model

Table 2. Number of individuals and average (standard deviation) of genetic diversity parameters for the six populations of Haloxylon ammodendron.

Population	*N*	*A*		*Ae*		*I*		*Ho*		*He*		*F*	
DG1	20	7.200	(1.139)	4.063	(0.707)	1.457	(0.143)	0.653	(0.048)	0.676	(0.036)	0.037	(0.047)
DG2	20	6.733	(1.040)	3.795	(0.634)	1.398	(0.139)	0.673	(0.045)	0.658	(0.038)	−0.024	(0.034)
BZ	20	7.333	(1.149)	4.216	(0.930)	1.433	(0.158)	0.667	(0.055)	0.649	(0.050)	−0.024	(0.035)
TZ	20	5.933	(0.753)	3.669	(0.598)	1.308	(0.138)	0.613	(0.054)	0.633	(0.048)	0.025	(0.042)
MG	20	6.267	(0.928)	3.815	(0.706)	1.321	(0.152)	0.627	(0.055)	0.636	(0.047)	0.019	(0.042)
KHE	20	6.867	(1.146)	4.154	(0.736)	1.427	(0.158)	0.730	(0.046)	0.665	(0.043)	−0.104	(0.028)
mean	20	6.722	(0.414)	3.952	(0.289)	1.391	(0.059)	0.661	(0.020)	0.653	(0.018)	−0.012	(0.016)

N: sample size, *A*: number of alleles, *Ae*: effective number of alleles, *I*: Shannon's Information Index, *Ho*: Observed Heterozygosity, *He*: Expected Heterozygosity, *F*: Fixation Index

in the desert, long-distance dispersal of pollen by wind is possible and this is likely to promote gene flow between isolated populations.

H. ammodendron is a very important plant in the desert, and the new microsatellite markers developed in this study will be very useful for conservation genetic studies of this species. In the process, these markers will be effectively applied in terms of identifying selected breeding lines and individuals.

Conclusions

A set of polymorphic microsatellite markers for the *Haloxylon ammodendron* is reported. The 15 nSSR markers developed in the present study could serve as powerful tools for assessing genetic diversity, genetic connectivity, and genetic structure in *H. ammodendron* populations, which will facilitate our understanding and conservation of *H. ammodendron* in Mongolia.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was supported by the Korea Foundation of Advanced Studies "International Scholar Exchange Fellowship for the academic year of 2017-2018" and the National Institute of Forest Sciences, Grant number: FG0801-2013-01, Republic of Korea;National Institute of Forest Sciences [FG0801-2013-01].

References

Hamrick JL, Linhart YB, Mitton JB. 1979. Relationships between life-history characteristics and electrophoretically detectable genetic variation in plants. Annu Rev Ecol Syst. 10:173–200.

Imtiaz A, Naz S. 2012. A rapid and cost-effective protocol for screening known genes for autosomal recessive deafness. Pak J Zool. 44 (3):641–647.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–1649.

Long Y, Zhang J, Tian X, Wu S, Zhang Q, Zhang J, Pei XW. 2014. De novo assembly of the desert tree *Haloxylon ammodendron* (C. A. Mey.) based on RNA-Seq data provides insight into drought response, gene discovery and marker identification. BMC Genomics. 15(1):1111.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. GigaScience. 1(1):18.

Mariette S, Le Corre V, Austerlitz F, Kremer A. 2002. Sampling within the genome for measuring within-population diversity: trade-offs between markers. Mol Ecol. 11(7):1145–1156.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM. 1998. Statistical confidence for likelihood- based paternity inference in natural populations. Mol Ecol. 7:639–655.

Nybom H. 2004. Comparison of different nuclear DNA markers for estimating ntraspecific genetic diversity in plants. Mol Ecol. 13 (5):1143–1155.

Peakall ROD, Smouse PE. 2006. genalex6: genetic analysis in excel. Population genetic software for teaching and research. Mol Ecol Notes. 6(1):288–295.

Rousset F. 2008. genepop'007: a complete re-implementation of the genepop software for windows and linux. Mol Ecol Resour. 8(1):103–106.

Russell JR, Fuller JD, Macaulay M, Hatz BG, Jahoor A, Powell W, Waugh R. 1997. Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theor Appl Genet. 95(4):714–722.

Shamsutdinov ZS, Ubaidullaev SR. 1988. Distribution of Poa bulbosa L. and Carex pachystylis within the phytogenous field of black saxaul. Prob Desert Dev. 1:38–43.

Sheng Y, Zheng W, Pei K, Ma K. 2005. Genetic variation within and among populations of a dominant desert tree *Haloxylon ammodendron* (Amaranthaceae) in China. Ann Bot. 96(2):245–252.

Suo Z, Jia Z, Lu Q, Pan B, Jin X, Xu G, Tao Y. 2012. Distinguishing *Haloxylon persicum* and *H. ammodendron* (*Haloxylon* Bunge, Amaranthaceae) using DNA marker. AASRI Procedia. 1:305–310.

Thiel T, Michalek W, Varshney R, 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 (3):411–422.

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3—new capabilities and interfaces. Nucleic Acids Res. 40(15):e115.

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes. 4 (3):535–538.