Indian J. Genet., 77(1): 145-152 (2017)
DOI: 10.5958/0975-6906.2017.00019.0

Rapid development and characterization of fifteen novel microsatellite loci in vulnerable tree *aquilaria malaccensis* using next generation ion torrent PGM™ sequencing

Priyanka Nanda*, Ashwani Kumar Singh[1], Madan Singh Negi and Shashi Bhushan Tripathi

The Energy and Resources Institute, IHC Complex, Lodhi Road, New Delhi 110 003; [1]Central Pollution Control Board, Ministry of Environment and Forests, Kolkata 700 107, West Bengal

(Received: July 2016; Revised: December 2016; Accepted: January 2017)

Abstract

***Aquilaria malaccensis* is one of the 15 tree species of the genus *Aquilaria* widely distributed in the Indo-Malaysian region. Its resin (Agar or *Gaharu*) is widely known for its use in the production of highly valued incense and traditional medicine throughout Asia. The species has been listed as 'Vulnerable' in the IUCN red list of threatened species due to its over exploitation. In the present study, for the first time, next generation sequencing platform (Ion Torrent PGM™) was used to obtain a large number of sequences containing microsatellites for *A. malaccensis*. A total of 527,425 high quality reads were identified, which contained 4138 SSRs. In these SSRs, the number of dinucleotide repeats were the most frequent one (74.77%). Primers were developed for 93 loci; of these 82 loci were successfully amplified and 47 were polymorphic across a panel of 42 genotypes. Fifteen loci with at least three or more alleles were chosen for detailed characterization. The observed heterozygosity ranged from 0.25 to 0.90, whereas the expected heterozygosity ranged from 0.24 to 0.89, with 5.9 alleles per locus on average. Simple matching dissimilarity coefficients and cluster analysis revealed significant diversity among the genotypes. The study suggested that PGM™ sequencing is a rapid technique for developing microsatellite markers. The microsatellite markers developed in the present study would be useful in comprehensive genetic diversity studies aimed towards the conservation and germplasm management of *Aquilaria*.**

Key words: *Aquilaria malaccensis,* next generation sequencing, microsatellite markers, diversity analysis, conservation, germplasm management

Introduction

Aquilaria malaccensis Lam. belonging to family: Thymelaeaceae (2n = 16), locally known as Agar, is a tropical evergreen tree found in India, Thailand, Myanmar, Sumatra, Peninsular Malaysia, Singapore and the Philippines (Debnath et al. 1995; IUCN 2013). The heartwood of this genus, when wounded and infected by certain fungi, produces an oleoresin locally known as Agar or Gaharu, which is widely used in traditional medicine and in incense and perfume industry. The genus, *Aquilaria,* contains 22 species out of which *A. malaccensis, A. crassna, A. sinensis, A. macrophylla* and *A. khasiana* are commonly known (IUCN 2013). Three species are found in India of which *A. macrophylla* and *A. khasiana,* are restricted to the Nicobar Islands and Khasi Hills of Meghalaya, respectively (Kanjilal et al. 1982). *A. malaccensis* has the widest distribution and occurs in the foothills of the North Eastern states of Assam, Arunachal Pradesh, Nagaland, Meghalaya, Mizoram, Manipur, Tripura and Sikkim (Saikia and Khan 2012). Due to its economic value, the species is currently being commercially planted outside its natural home, namely, West Bengal, Karnataka and Tamil Nadu. Two major variants are loosely recognized, namely, Bhola Sanchi and Jati Sanchi, based on their oleoresin productivity (Saikia and Khan 2012).

In recent decades, aggressive felling and trade activities have resulted in large scale destruction of natural stands of *A. malaccensis* and concerns over the long-term survival potential of *A. malaccensis*. Its population has declined to the point that it is included in CITES Appendix II (CITES 1994) and IUCN Red List of Threatened Species (IUCN 2015) as "Vulnerable" and considered "Critically Endangered in India" and

*Corresponding author's e-mail: priyankasmail@gmail.com
Published by the Indian Society of Genetics & Plant Breeding, F2, First Floor, NASC Complex, PB#11312, IARI, New Delhi 110 012
Online management by indianjournals.com; http://epubs.icar.org.in/journal/index.php/IJGPB

almost “Extinct in wild in Assam” (Anonymous, 2003). Therefore, attention is needed towards conservation and sustainable utilization to prevent the species from extinction. Despite being critically endangered, very little is known about the genetic diversity of *A. malaccensis*. Dominant molecular markers such as ISSR and SRAP (Zou et al. 2012), RAPD and SCAR (Lee et al. 2011) have been used for diversity analysis in *A. sinensis* and other Aquilaria sp. Microsatellite markers are preferred over dominant markers due to their codominant inheritance, ease of use, high reproducibility, high degree of polymorphism and allelic diversity (Weising et al. 2005). However, many times these markers are not readily available, especially in forestry plants such as Aquilaria. So far, very few microsatellite markers on Aquilaria are available in public domain (Tnah et al. 2012; Zhang et al. 2010). In the current study, we used the next generation sequencing platform of Ion Torrent PGM to develop and characterize novel microsatellite markers for *A. malaccensis*, which would be useful for comprehensive genetic diversity studies aimed towards its conservation and germplasm management.

Materials and methods

DNA isolation

To achieve wide representation of various microsatellites, DNA from 16 genetically diverse *A. malaccensis* accessions were pooled in two sets of 8 accessions each (Table 1). These accessions were selected based on AFLP analysis of over 110 accessions from different localities of Assam conducted in our laboratory (unpublished data). DNA was isolated from lyophilized young leaves using a CTAB based method described earlier (Singh et al. 1999). Twenty µl of DNA (100 ng/µl) from each of 16 accessions was pooled as mentioned above and used for library preparation.

Library preparation and SSR enrichment

The library preparation protocol for Ion plus Fragment Library Kit (Part number-4471252, Invitrogen) was used in combination with protocol for SSR enrichment (Glenn and Schable 2005). In brief, 1.2 µg of purified DNA from each pool (Pool 1 and Pool 2) were sheared by adaptive focused acoustics using a Covaris S2 (Covaris Inc.) as per the user manual (Part Number 4467320 Rev. A). The fragments were end repaired using the supplied protocol and again purified using AMPure Magnetic beads (Part number-A63881) and eluted in 50uL of low TE. The two end-repaired libraries resulting from Pool 1 and Pool 2 were ligated to barcode adaptors, BC15 and BC16, respectively and nick repaired. This resulted in two pooled genomic libraries namely, P15 and P16, respectively. The adaptor ligated DNA was purified using AMPure Magnetic beads and eluted in 22 uL of low TE. The DNA fragments were size-selected using 2% size-selection E-gels (Part number G6610-02) and fragments at approx. 330bp region were isolated by excising the gel. The size selected DNA was then amplified for 9 cycles and checked on 2% E-gel and purified using AMPure Magnetic beads and eluted in 26uL of Low TE. The amplified DNA was used for one round of SSR enrichment through hybridization with biotin labeled microsatellite probes (Table 2) and subsequent capture using Streptavidin coated magnetic beads (Dynabeads, Dynal, Oslo, Norway) following the protocol mentioned in Glenn and Schable (2005). The SSR enriched fragments were again amplified for 7 cycles. Template Preparation was done using Ion OneTouch™ 200 Template Kit v2DL (Part number- 4480285) using 26 pM library following manufacturer instructions. All DNA quantifications were carried out using Qubit 2.0 (Life Technologies Inc).

Table 1. List of samples used for library preparation

Accession ID	Original plant ID	Place of collection	Remark	Pool no.
Aq002	AQ018	Dibrugarh University	Status not known	1
Aq005	AQ021	Dibrugarh University	Status not known	1
Aq008	AS/CHM-3	Chamaguri	Infected	1
Aq016	AS/CHM-F	Chamaguri	Not infected	1
Aq027	AS/NMT-2	Namti Chariali	Infected	1
Aq039	AS/GHA-4	Gharphalia	Infected	1
Aq041	AS/GHA-A	Gharphalia	Not infected	1
Aq062	AS/KHA-E	Khanikar	Not infected	1
Aq003	AQ019	Dibrugarh University	Status not known	2
Aq004	AQ020	Dibrugarh University	Status not known	2
Aq022	AS/TC-4A	Shotaichiga	Not infected	2
Aq023	AS/TC-4B	Shotaichiga	Not infected	2
Aq031	AS/NMT-C	Namti Chariali	Not infected	2
Aq035	AS/LM-B	Lalimchiga	Not infected	2
Aq040	AS/GHA-5	Gharphalia	Infected	2
Aq056	AS/KHA-19	Khanikar	Infected	2

Table 2. List of six probe sequences used to segregate the reads for SSR enrichment

S.No.	Repeat motif	Probe Sequence
1	$(GA)_{12}$	GAGAGAGAGAGAGAGAGAGAGA
2	$(CA)_{12}$	CACACACACACACACACACACA
3	$(TAA)_8$	TAATAATAATAATAATAATAA
4	$(GGA)_8$	GGAGGAGGAGGAGGAGGAGGA
5	$(GAT)_8$	GATGATGATGATGATGATGAT
6	$(CTT)_8$	CTTCTTCTTCTTCTTCTTCTT

Ion Torrent PGM™ sequencing

Sequencing was undertaken using 314 chip using Ion Sequencing 200 Kit (Part number-4474004) following the recommended protocol. Before loading the ISPs on the Chip, the ISPs were washed and primed with sequencing primer at 95°C for 2 min and 37°C for 2 minutes. Polymerase was added to these ISPs prior to loading and incubated at RT for 5 minutes. The ISPs were loaded on 314 Chip as per protocol guidelines.

Sequence analysis and primer designing

Torrent Suite 3.6 was used for all analyses. Reads from Ion PGM (obtained in unmapped BAM format) were converted to FASTQ using FASTQCreator plugin. Reads were then extracted in FASTA format using "FASTQ to FASTA Converter" tool of FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) for each of the samples. The unique reads were assembled and screened for the presence of any of the six probe sequences (Table 2) using a Shell script. The filtered reads were processed using Batchprimer 3 (http://probes.pw.usda.gov/batchprimer3/) to identify microsatellite repeats and to design SSR primers. All primers were commercially synthesized (Macrogen, Korea).

Primer evaluation and genotyping

All the 93 primers (Supplementary Table 1) were screened with 42 *Aquilaria* accessions (Table 3) collected from Sibsagar, Jorhat, Golghat and Nagaon districts of Assam and Changlang district of Arunachal Pradesh. Initially, annealing temperature and polymorphism were tested for each primer pair. Based on the results seen on 2.5% agarose gel (out of which 11 primers did not amplify, 36 primers gave monomorphic bands and 47 were polymorphic), twenty primers amplifying at least three or more alleles were selected to be tagged with IRDye (700 or 800) labeled M13 sequences (5'- GTCACGACGTTGTAAAACG-3') at 5' end and were screened with 42 *Aquilaria* accessions. While IRDye 800 was a heptamethine cyanine dye with its absorption and emission maxima at 787nm and 812nm, respectively, IRDye 700 was a pentamethine carbocyanine dye with its absorption and emission maxima at 681 nm and 712 nm, respectively. The M13 primer was added to the PCR product during the first few cycles of amplification. The labeled M13 primer was incorporated in subsequent cycles, thus labeling the PCR product. The amplified SSR loci, incorporated with IRDye labeled M13 sequences, separated according to size on an acrylamide gel. A solid-state diode laser was used to excite the infrared dye on the DNA fragments as they migrated past the detector window. A focusing fluorescence microscope containing a solid-state silicon avalanche photodiode scanned back and forth across the width of the gel and collected data in real time. The raw image data were a series of bands displayed on the computer monitor in a format similar to an autoradiogram. For SSR analysis and genotyping, PCR amplification for each reaction was carried out in a total volume of 10 μL containing 5.5 ng of genomic DNA, 1X PCR buffer, 0.25 mM dNTP mix, 0.04 pM/μL of M13 labeled primer, 0.016 pM/μL of M13 tailed forward primer, 0.3 pM/μL reverse primer and 0.04U/μL Taq polymerase (GENET BIO). The Touchdown PCR conditions for SSRs were optimized as initial cycle of 94°C for 2 min followed by 10 cycles of 94°C for 30 s, 64°C minus 1°C in each cycle of 1 min and 72°C for 1 min followed by 25 cycles of 94°C for 30 s, 54°C for 30 s and 72°C for 1, followed by 8 cycles of 94°C for 30 s, 63°C for 30 s and 72°C for 1 min with final extension at 72°C for 10 min, before cooling to 4°C. Following amplification, 1X formamide dye was added to each sample and the samples were denatured for 5 min at 95°C. Each sample (1 μl) was loaded on a 6.5% polyacrylamide gel containing 7 M urea. Gels were cast using LICOR 25-cm plates with 0.40-mm-thick spacers and comb. Electrophoresis was performed at a constant power of 40W and a constant temperature of 47.5°C for 2 hours. Gels were analyzed on a LI-COR 4300 DNA analyzer (LI-COR Biosciences, Lincoln, NE, USA). Allele scoring was done manually.

Statistical analysis

Observed heterozygosity (H_o), expected heterozygosity (H_e) Hardy-Weinberg equilibrium (HWE), linkage disequilibrium (LD) and polymorphic information content (PIC) were estimated using power

Table 3. Genotype data and details of accessions used for SSR analysis and genotyping in the study

Accession ID	Original ID	Place of collection	Cluster	Group	Infection status
Aq06	AS/CHM-1	Chamaguri (Sibsagar)	I	B	I
Aq07	AS/CHM-2	Chamaguri (Sibsagar)	I	B	I
Aq11	AS/CHM-A	Chamaguri (Sibsagar)	I	B	NI
Aq12	AS/CHM-B	Chamaguri (Sibsagar)	II	B	NI
Aq18	AS/SHC-2	Shotaichiga (Sibsagar)	I	A	I
Aq22	AS/TC-4A	Shotaichiga (Sibsagar)	II	A	NI
Aq24	AS/SHC-4C	Shotaichiga (Sibsagar)	II	A	NI
Aq27	AS/NMT-2	Namti Charialii (Sibsagar)	II	B	I
Aq30	AS/NMT-B	Namti Charialii (Sibsagar)	I	B	NI
Aq34	AS/LM-A	Lalimchiga (Sibsagar)	II	B	NI
Aq37	AS/GHA-1	Gharphalia (Sibsagar)	II	C	I
Aq40	AS/GHA-5	Gharphalia (Sibsagar)	I	A	I
Aq42	AS/GHA-B	Gharphalia (Sibsagar)	II	A	NI
Aq45	AS/JAK-1	Jhakaria gaon (Jorhat)	I	B	I
Aq46	AS/JKS-2	Jhakaria gaon (Jorhat)	II	C	I
Aq48	AS/JKS-4	Jhakaria gaon (Jorhat)	I	B	I
Aq51	AS/JKS-8	Jhakaria gaon (Jorhat)	II	C	I
Aq52	AS/JAK-9	Jhakaria gaon (Jorhat)	I	A	I
Aq59	AS/KHA-B	Khanikar (Jorhat)	II	A	NI
Aq63	AS/AMG/01	Amguri Khat (Jorhat)	II	A	I
Aq66	AS/BAG/02A	Bagmaria (Jorhat)	I	A	NI
Aq68	AS/BAG/04	Bagmaria (Jorhat)	I	B	I
Aq70	AS/Dol/02	Dolerseria (Jorhat)	I	A	I
Aq73	AS/Dol/06	Dolerseria (Jorhat)	I	A	I
Aq76	AS/Goj/01	Goujpuria (Jorhat)	I	A	I
Aq82	AS/KHK/03	Kharkhowa (Jorhat)	II	C	I
Aq83	AS/KHK/04	Kharkhowa (Jorhat)	I	A	I
Aq86	AS/KHT/03	Khatowal (Nagaon)	II	B	I
Aq94	AS/NCH/03	Namchungi (Jorhat)	II	C	I
Aq105	AS/PBC/02	Panbari Chungi (Jorhat)	II	A	I
Aq106	AS/PBC/03	Panbari Chungi (Jorhat)	II	A	I
Aq109	AS/RAJ/01	Rangajan (Golaghat)	II	B	I
Aq119	WA/01/02	Dergaon (Golaghat)	I	A	I
Aq124	WA/03/01	Dergaon (Golaghat)	II	A	I
Aq133	WA/06/01	Dergaon (Golaghat)	II	A	I
Aq138	WA/07/03	Dergaon (Golaghat)	I	A	I
Aq142	WA/09/01	Dergaon (Golaghat)	I	A	I
Aq148	WS/01/01	Jairampur (Changlang)	II	A	I
Aq152	WS/02/02	Jairampur (Changlang)	II	B	I
Aq156	WS/03/03	Jairampur (Changlang)	II	A	I
Aq161	AS/TMG/03	Golaghat	II	A	I
Aq165	AQ/GGR/1.10	Golaghat	II	C	I

marker ver. 3.25 (Liu and Muse 2005). PIC compares the polymorphism levels across markers and is used to determine the usefulness of markers for specific studies (Botstein et al. 1980). For understanding the overall genetic relationships among *Aquilaria* accessions, the diploid allelic data was used to prepare a dendogram based on Neighbor-Joining with simple matching dissimilarity index (1,000 bootstraps) in DARwin 5.0.158 software.

Results and discussion

Ion Torrent PGM™ sequencing

By PGM™ sequencing with a 314 chip, a total of 92.38 Mb bases and 527,425 quality reads were obtained in a single sequencing run from the genomic DNA of Pools 1 (P15) and 2 (P16). The length of the longest read was 355 bp with a mean length of 175 bp. The total number of quality reads in P15 and P16 were 240,406 and 284,134, respectively. The mean read lengths were 178 bp for P15 and 172 bp for P16. The number of reads that had a predicted quality score of Q20 (one error in 100 bases) or better was 74.32 Mbp. In order to segregate the reads into various probe sequences, a small shell script was written resulting in a total of 4138 microsatellites being identified. These were filtered for their length and 2240 sequences which were greater than 150 bp, were selected. Primers were designed for 2240 microsatellites using BatchPrimer 3. Thus, the results reinstate the fact that Ion Torrent PGM™ platform is an efficient and rapid next generation sequencing technique capable of multi-million read level outputs (Whiteley et al. 2012).

Characteristics of microsatellites

In these SSRs, the number of dinucleotide repeats was the most frequent (74.77%) followed by trinucleotide repeats (25.22%). GA repeats (39%) were the most frequently occurring microsatellite, followed by CA (36%), TAA (5%), GGA (3%), GAT (9%) and CTT (9%) repeats (Fig. 1).

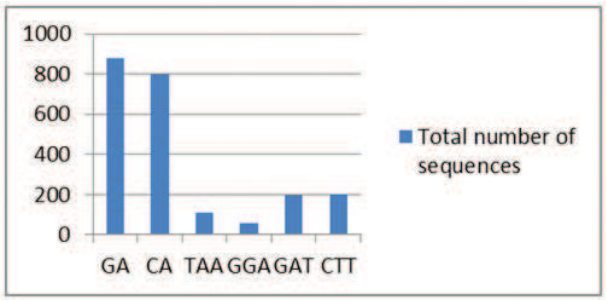

Fig. 1. Number of SSR Repeat Motifs in 2240 sequences

Microsatellite analysis

All the 2240 sequences were filtered for their length, repeat region length and repeat region position using BatchPrimer 3. Henceforth, a total of 174 microsatellite containing sequences were isolated and submitted to the GenBank with accession numbers KR047030, KR052348-KR052440, KR067423-KR067479, KR259825-KR259844 and KR782410-KR782412. Ninety-three of these sequences having repeats with enough flanking nucleotides were used for primer designing. Out of the finally selected twenty

Table 4. Characterization of fifteen polymorphic microsatellite loci and their marker attributes in *Aquilaria malacensis*

S.No.	Locus	Gene bank accession no.	Allele size range (bp)	No. of alleles N_a	Expected heterozygosity H_e	Observed heterozygosity H_o	PIC
1	AQ_SSR_0018	KR047030	160-200	3	0.60	0.54	0.52
2	AQ_SSR_0033	KR259825	145-165	5	0.63	0.48	0.56
3	AQ_SSR_0926	KR052364	176-212	13	0.89	0.29	0.88
4	AQ_SSR_1037	KR052365	158-174	6	0.67	0.36	0.63
5	AQ_SSR_1103	KR052366	177-215	12	0.71	0.71	0.67
6	AQ_SSR_1691	KR052381	142-148	2	0.50	0.48	0.37
7	AQ_SSR_1720	KR067444	145-175	6	0.78	0.74	0.74
8	AQ_SSR_1796	KR052392	185-190	2	0.33	0.28	0.28
9	AQ_SSR_1799	KR052394	141-168	5	0.58	0.25	0.55
10	AQ_SSR_1813	KR052397	109-131	6	0.59	0.28	0.52
11	AQ_SSR_2041	KR259826	150-155	2	0.38	0.40	0.30
12	AQ_SSR_2079	KR052422	129-168	10	0.80	0.55	0.77
13	AQ_SSR_2128	KR052426	155-180	11	0.86	0.90	0.85
14	AQ_SSR_2171	KR052433	165-180	4	0.71	0.38	0.65
15	pGAqu026	KR259830	215-225	2	0.24	0.29	0.21

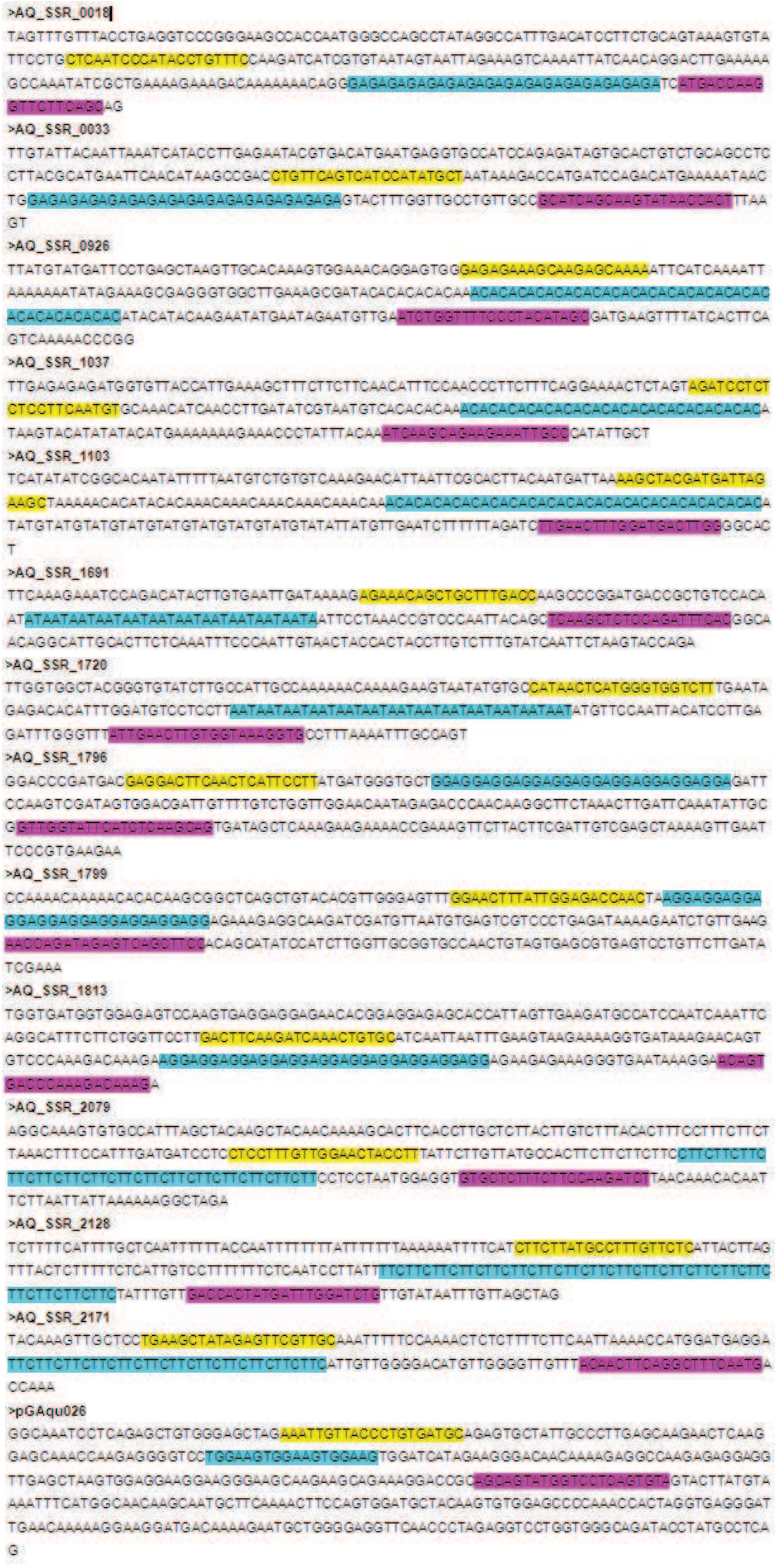

Fig. 2. Fifteen polymorphic microsatellite loci containing sequences of *A. malaccensis* submitted to GenBank, where yellow, blue and magenta shaded sequences refer to forward primer, microsatellite repeat and reverse primer sequences, respectively

microsatellites, fifteen polymorphic microsatellite loci identified in this study showed clear polymorphic bands for each allele, while remaining two markers did not amplify the corresponding fragments and three markers amplified non-specific fragments. Clone sequences of the fifteen polymorphic loci are presented in figure 2 and their characteristics are presented in Table 4. The amplified products were in agreement with the expected size from sequence data for all fifteen microsatellites analyzed. The loci exhibited two to thirteen alleles per locus with an average of 5.9. A total number of 89 alleles were detected with nine single locus microsatellite primer pairs. Major allele frequencies ranged from 0.18 to 0.86 with a mean value of 0.49. Mean expected H_e, H_o and PIC were 0.62, 0.46 and 0.57 respectively. PIC values for microsatellite loci ranged from 0.21 (marker pGAqu026) to 0.88 (marker AQ_SSR_0926). HWE and LD tests were performed and six polymorphic loci *viz*., AQ_SSR_0103, AQ_SSR_0926, AQ_SSR_1037, AQ_SSR_1813, AQ_SSR_2079 and AQ_SSR_2171 were in HWE ($P < 0.005$) while nine of them viz. AQ_SSR_0018, AQ_SSR_0026, AQ_SSR_0040, AQ_SSR_0079, AQ_SSR_0130, AQ_SSR_1103, AQ_SSR_1691, AQ_SSR_1796 and AQ_SSR_2128 were in HWE with no major deviations. A significant LD between pairs of loci was found with negligible deviations.

Diversity analysis

A UPGMA dendrogram was obtained using simple matching dissimilarity coefficients in DARwin 5.0.158 (Fig. 3). Cluster analysis based on coefficient of dissimilarity classified the *Aquilaria* genotypes into two major clusters, which were designated as I and II in this study. Cluster I comprised of 18 genotypes *viz*., 9 from Jorhat, 6 from Sibsagar and 3 genotypes from Golaghat districts of Assam. It was further divided in two groups IA comprising of 11 and IB comprising of 7 genotypes. The average dissimilarity coefficient was 0.59. The lowest similarity

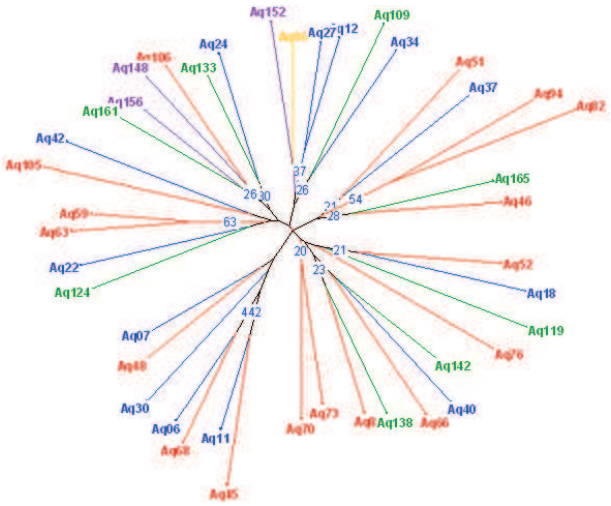

Fig. 3. Dendogram depicting genetic relationship among 42 *Aquilaria malaccensis* genotypes using 15 newly developed SSR markers, where yellow, blue, green, red and magenta colours refer to accessions collected from Nagaon, Sibsagar, Golaghat, Jorhat and Changlang districts, respectively and the numbers at the tree nodes refer to bootstrap values

was between genotypes Aq45 and Aq66 as well as Aq68 and Aq83, all accessions originating from the same district, Jorhat, with the dissimilarity coefficient being 0.83, while the highest similarity was between two genotypes of different districts, Sibsagar and Jorhat, namely Aq06 and Aq68 with the dissimilarity coefficient being 0.33. Cluster II was larger comprising of 24 genotypes *viz*., 8 genotypes from Jorhat, 7 from Sibsagar, 5 from Golghat, 1 from Nagaon districts of Assam and 3 from Changlang district of Arunachal Pradesh, with 0.60 being the average dissimilarity coefficient. It was futher divided into three groups IIA, IIB and IIC comprising of 12, 6 and 6 genotypes respectively. In this cluster Aq42 (Sibsagar) and Aq109 (Golaghat), Aq82 (Jorhat) and Aq105 (Jorhat), Aq82 (Jorhat) and Aq161 (Golaghat) were the most distantly related genotypes with their dissimilarity coefficients being 0.87 whereas Aq12 and Aq27 both from Sibsagar and Aq59 and Aq63 both from Jorhat district were most closely related genotypes with their dissimilarity coefficients being 0.33. The overall average simple matching dissimilarity coefficient in all the 42 accessions was 0.62 with lowest being 0.33 between Aq06 (Sibsagar) and Aq68 (Jorhat) both in group IB, Aq12 and Aq27 both from Sibsagar and in group IIB and Aq59 and Aq63 both from Jorhat and in group IIA, depicting these accessions to be genetically very similar. The overall highest dissimilarity coefficient was 0.90 between Aq94 (Jorhat) in group IIC and Aq119 (Golaghat) in group IA. All the genotypes from different places of origin were proportionally distributed between the two clusters. It was thus evident that there was no clear cut correlation between the place of collection and the degree of similarity or dissimilarity of the genotypes. This further suggested a possible gene flow across genotypes from different geographical origins. The new microsatellite markers developed in this study will serve as a very useful tool for genetic diversity analysis, clonal identification and conservation in *Aquilaria* germplasm.

Authors' contribution

Conceptualization of research (PN, SBT); Designing of the experiments (PN, SBT,MSN); Contribution of experimental materials (SBT); Execution of field/lab experiments and data collection (PN, AKS); Analysis of data and interpretation (PN, SBT); Preparation of manuscript (PN).

Declaration

The authors declare no conflict of interest.

Acknowledements

This work was supported by the Department of Biotechnology, Ministry of Science and Technology, Government of India (Project No. BT/PR11505/NDB/38/223/2011).

References

Anonymous, "Review of significant trade *Aquilaria malaccensis*". 2003. Available from URL: http://www.cites.org/eng/com/PC/14/E-PC14-09-02-02-A2.pdf.

Botstein D., White R. L., Skolnick M. and Davis R. W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet., **32**: 314-331.

CITES, 7-18 November 1994 Resolution of the Conference of the Parties, Ninth meeting of the Conference of the Parties, Fort Lauderdale, USA | CITES. https://www.cites.org/eng/cop/09/doc/index.php. Accessed 10 Aug 2015.

Debnath B., Sil S., Sinha R. K. and Sinha S. 1995. Chromosome Number and Karyotype of Aquilaria agallocha Roxb. (Thymelaeaceae). CYTOLOGIA, **60**: 407-409. doi: 10.1508/cytologia.60.407

Glenn T. C. and Schable N. A. 2005. Isolating microsatellite DNA loci. Meth. Enzymol., **395**: 202-222. doi: 10.1016/S0076-6879(05)95013-1.

IUCN, 2013 Asian Regional Workshop (Conservation & Sustainable Management of Trees, Viet Nam, August 1996) 1998. *Aquilaria malaccensis*. In: IUCN 2013. IUCN Red List of Threatened Species. Version 2013.2.

IUCN Red List of Threatened Species (2015) Version 2.3. Available from URL: http://dx.doi.org/10.2305/IUCN.UK.1998.RLTS.T32056A9677920.en.

Kanjilal U. N., Kanjilal P. C., Dey R. M. and Das A. 1982. Flora of Assam - IV, Government of Assam.

Lee S. Y., Weber J. and Mohamed R. 2011. Genetic Variation and Molecular Authentication of Selected Aquilaria Species from Natural Populations in Malaysia Using RAPD and SCAR Markers. Asian J. Plant Sci., **10**: 202-211. doi: 10.3923/ajps.2011.202.211

Liu K. and Muse S. V. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics, **21**: 2128-2129. doi: 10.1093/bioinformatics/bti282

Saikia P. and Khan M. L. 2012. Aquilaria malaccensis Lam., a Red-listed and highly exploited tree species in the Assamese home garden. Curr. Sci., **102**: 546-547.

Singh A., Negi M. S., Rajagopal J., Bhatia S., Tomar U. K., Srivastava P. S. and Lakshmikumaran M. 1999. Assessment of genetic diversity in Azadirachta indica using AFLP markers. Theor. Appl. Genet., **99:** 272-279. doi: 10.1007/s001220051232

Tnah L. H., Lee C. T., Lee S. L., Ng K. K. S., Ng C. H., Nurul-Farhanah Z., Lau K. H. and Chua L. S. L. 2012. Isolation and characterization of microsatellite markers for an important tropical tree, Aquilaria malaccensis (Thymelaeaceae). Am. J. Bot., **99**: e431-433. doi: 10.3732/ajb.1200165.

Weising K., Nybom H., Wolff K. and Kahl G. 2005. DNA Fingerprinting in Plants: Principles, Methods and Applications, CRC Press (Boca Raton, United States), ISBN 0-8493-1488-7.

Whiteley A. S., Jenkins S., Waite I., Kresoje N., Payne H., Mullan B., Allcock R. and O'Donnell A. 2012. Microbial 16S rRNA Ion Tag and community metagenome sequencing using the Ion Torrent (PGM) Platform. J. Microbiol. Methods, **91**: 80-88. doi: 10.1016/j.mimet.2012.07.008.

Zhang Y-T., Wang Z-F., Cao H-L., Li X-Y., Wu L-F., Zhuo S-B. and Huang X-F. 2009. Isolation and characterization of polymorphic microsatellite loci in Aquilaria sinensis (Lour.) Gilg. Cons. Genet. Resour., **2**: 5-6. doi: 10.1007/s12686-009-9110-z.

Zou M., Xia Z., Lu C., Wang H., Ji J. and Wang W. 2012. Genetic Diversity and Differentiation of (Lour.) Gilg. Revealed by ISSR and SRAP Markers. Crop Sci., **52**: 2304-13. doi: 10.2135/cropsci2011.10.0576.

Supplementary Table 1. List and characteristics of new 93 microsatellite markers submitted to GenBank

S.No.	Primer name	GenBank accession number	T_a (p C)	Repeat Motif	Forward Primer* (5' to 3')	Reverse Primer (5' to 3')	Profile
1	AQ_SSR_0018	KR047030	55	$(GA)_{16}$	CTCAATCCCATACCTGTTTC	GCTGAAGAACCTTGGTCAT	Polymorphic
2	AQ_SSR_0033	KR259825	55	$(GA)_{16}$	CTGTTCAGTCATCCATATGCT	AGTGGTTATACTTGCTGATGC	Polymorphic
3	AQ_SSR_0035	KR052348	55	$(AG)_{17}$	CAAGAGTACGGTTCACAAATC	GGATTTCTGTTTCCAAGTACC	Polymorphic
4	AQ_SSR_0037	KR052349	54	$(GA)_{19}$	ACACGTATTTAAGTGGGGTAG	CATTTTCCTCTTCTCCTTACC	Monomorphic
5	AQ_SSR_0129	KR052350	54	$(AG)_{18}$	CAATGCAGTGTTTACGAATC	GAAACGATTGTAGTGCTCGT	Not Amplified
6	AQ_SSR_0140	KR052351	56	$(GA)_{15}$	AGGGACTTATTTGAATGTGG	TCCTTCCGGTATGATCTCTAT	Monomorphic
7	AQ_SSR_0160	KR067423	54	$(GAT)_{8}$	AGTTCATCTGGGTTAAAGAGC	CACACTGATGAGTGGGTAAGT	Monomorphic
8	AQ_SSR_0178	KR052352	56	$(GA)_{16}$	GACAAACTCGATCTGCAACT	TCCACGTCTTCCTTTCTTC	Polymorphic
9	AQ_SSR_0195	KR052353	56	$(AG)_{15}$	CACAGAAAGCTCAGATGATTG	TGCTTTCTAGTTGTCTTCGTC	Not Amplified
10	AQ_SSR_0234	KR052355	56	$(AG)_{18}$	CCACAGAGTGACAACTCAGAT	ATCTTTACGCTACCTTTGGTC	Not Amplified
11	AQ_SSR_0311	KR052356	56	$(AG)_{17}$	ACATTAGGATCACAATGGAG	ATGTTATTGCCATGAACTCG	Polymorphic
12	AQ_SSR_0357	KR052357	56	$(AG)_{18}$	CACACAAAGGGAAGTATTCTG	CAAAACGTAGCAGAGCTAGTG	Polymorphic
13	AQ_SSR_0424	KR052358	56	$(AG)_{17}$	AGCGTTATCTCTTCTCCATCT	ACCATTAGCAGTTGTCAAGC	Not Amplified
14	AQ_SSR_0527	KR052359	56	$(AG)_{16}$	ACGACTTTTGCATCTACACTC	TCAACCATAAAGGCATTAGC	Polymorphic
15	AQ_SSR_0596	KR052360	55	$(GA)_{16}$	GGTTCAAAGCCTTCCATTA	ACCTCCACTTCCATCTCTTTA	Not Amplified
16	AQ_SSR_0747	KR052361	55	$(AG)_{16}$	CTAGAATTGGGAAACGAGTCT	CCCTGTTTGTCTAGGAAGATT	Monomorphic
17	AQ_SSR_0790	KR052362	55	$(GA)_{17}$	CTTATCCACTGCTAAATGGAG	CATGTTCTTCTTCTCCCTAGAC	Monomorphic
18	AQ_SSR_0908	KR052363	54	$(AC)_{16}$	GTCATGTGGGATAAATTCCT	CATATTTACTTGCACCGTGA	Monomorphic
19	AQ_SSR_0926	KR052364	55	$(AC)_{22}$	GAGAGAAAGCAAGAGCAAAA	GCTATGTAGGGAAAACCAGAT	Polymorphic
20	AQ_SSR_1037	KR052365	55	$(AC)_{16}$	AGATCCTCTCTCCTTCAATGT	GGCAATTTCTTCTGCTTGAT	Polymorphic
21	AQ_SSR_1103	KR052366	55	$(AC)_{20}$	AAGCTACGATGATTAGAAGC	CCAAGTCATCCAAAGTTCAA	Polymorphic
22	AQ_SSR_1170	KR052367	55	$(AC)_{16}$	AATACTCTGGAGCAAATAGGG	AGAAATATACGGCCAGATGAC	Polymorphic
23	AQ_SSR_1207	KR052368	55	$(AC)_{25}$	CATTTATCATCCATCCATCC	CAAGGAAACAAGCAAAGC	Monomorphic
24	AQ_SSR_1215	KR052369	55	$(CA)_{21}$	TTATGTGCAGTGTTCTGTGTC	GCATGAAACTGAGTTGTTAC	Monomorphic
25	AQ_SSR_1349	KR052370	55	$(CA)_{18}$	TATTTCAAGGCTGTTGATGC	AGGATTCGATATGTGTGTGA	Not Amplified
26	AQ_SSR_1385	KR052371	55	$(AC)_{15}$	CGATCAGGATAGGGTGACTAT	GGCACTAACAGAATCCATTTAC	Polymorphic
27	AQ_SSR_1442	KR052372	55	$(AC)_{23}$	AGTGCTCAATGTTTCACAGTC	TAAATCCTGGTTCGCTACTG	Polymorphic
28	AQ_SSR_1495	KR052373	55	$(AC)_{19}$	TTATCAATTCCGGTTCACTC	ATTCAGTCATCCATCCATGT	Polymorphic
29	AQ_SSR_1505	KR052374	55	$(AC)_{23}$	GTCCTAAAGGACTGAAACGTA	TCAGTAAGGTCTTCAACTTGC	Polymorphic
30	AQ_SSR_1573	KR782410	54	$(GTAT)_{7}$	CACCACCTTATTGCTTTGTAG	GTGTGTGTGTGTGTGTGTGT	Monomorphic
31	AQ_SSR_1632	KR052375	55	$(CA)_{17}$	GGAGTATTTCTTAGCGCGTAT	GCAAGTAGAAGGGACTTTGAT	Not Amplified
32	AQ_SSR_1648	KR052376	55	$(CA)_{15}$	CCCAAAATCGAGTTACTGTTC	CTAACCCATAAGTTCAACGAC	Polymorphic
33	AQ_SSR_1679	KR052377	55	$(AAT)_{14}$	GGGCAAATTCTTGATGATTC	CATCCTCATTATCATTGCTG	Polymorphic
34	AQ_SSR_1681	KR052378	55	$(AAT)_{10}$	ATGGTATTCTCCGGTTGTTA	GTGTCGTTTTTAGTTTGTCC	Polymorphic
35	AQ_SSR_1682	KR052379	55	$(ATA)_{13}$	TTGGCTTGTATATCTCGTAGC	ACCCCCTTTGATTATGTTTC	Polymorphic
36	AQ_SSR_1685	KR052380	55	$(AAT)_{13}$	TGAGTGAAGGGTTGAAAACT	CTCTCTTGTTGATGCAATCTC	Monomorphic
37	AQ_SSR_1691	KR052381	55	$(ATA)_{11}$	AGAAACAGCTGCTTTGACC	GTGAAATCTGGAGAGCTTGA	Polymorphic
38	AQ_SSR_1710	KR052382	55	$(TAA)_{8}$	AAGACTGCACGTCTCTAACTG	ACTAACTCGCAGTGAGTCAAA	Monomorphic
39	AQ_SSR_1715	KR052383	52	$(AAT)_{8}$	GTGGTTGACAAGTGGAGTTTA	ACCTAAAGAAGTGTATCAGTCG	Polymorphic
40	AQ_SSR_1720	KR067444	55	$(AAT)_{13}$	CATAACTCATGGGTGGTCTT	CACCTTTACCACAAGTTCAAT	Polymorphic
41	AQ_SSR_1721	KR052384	55	$(ATA)_{9}$	CACCACAGCATACAGTTGATA	ACACTCATACTTGGTGCACTT	Monomorphic
42	AQ_SSR_1730	KR052385	52	$(AAT)_{10}$	AGTTTCTCCAAAAGTGAGGA	CTAAAGTGAGATGGGCAAGT	Monomorphic
43	AQ_SSR_1732	KR052386	55	$(AAT)_{9}$	CTGGTGTACGAGATTGTTCTT	CTCTGAAAGAGTTTGAACTTCC	Monomorphic
44	AQ_SSR_1743	KR052387	55	$(TAA)_{13}$	GGTGATGATGATGATGATGT	TATTGTTAGTGTGAGGCAACC	Monomorphic
45	AQ_SSR_1758	KR052388	55	$(TAA)_{8}$	GAAGGGGAACCAATATACAAG	ATGGAGAAACTGGATTAGCTG	Polymorphic
46	AQ_SSR_1790	KR052389	54	$(GAG)_{9}$	GAAACCCAGAACCCTAATTC	GACTTCCGTGTAAGTTCCAAT	Monomorphic
47	AQ_SSR_1791	KR782411	54	$(GGA)_{8}$	AGAGCAAGAATAAGGAAGTGG	GCCATTTAGTATTTCCTTTCC	Monomorphic
48	AQ_SSR_1792	KR052390	54	$(GAA)_{8}$	GAAAATGGTGAGAAAGAGGAG	CTTCCGTAACATTACCTTCTTC	Monomorphic
49	AQ_SSR_1795	KR052391	55	$(GGA)_{8}$	GCATCCTGAGCAACAATATG	CCACGTGATTCATTAGTGTCT	Monomorphic

50	AQ_SSR_1796	KR052392	55	$(GGA)_{10}$	GAGGACTTCAACTCATTCCTT	CTGCTTGAGATGAATACCAAC	Polymorphic
51	AQ_SSR_1797	KR052393	55	$(GGA)_{8}$	GAGTGGGAGTAGGAACTGAAC	CCGAAATAAAGATTGGAACC	Polymorphic
52	AQ_SSR_1799 (1)	KR052394	55	$(AGG)_{10}$	GGAACTTTATTGGAGACCAAC	GGAAGCTGACTCTATCTGGTT	Polymorphic
53	AQ_SSR_1799 (2)	KR052394	55	$(AGG)_{10}$	GTTGGGAGTTTGGAACTTTAT	GGAAGCTGACTCTATCTGGTT	Polymorphic
54	AQ_SSR_1813	KR052397	55	$(AGG)_{11}$	GACTTCAAGATCAAACTGTGC	CTTTGTCTTTGGGTCACTGT	Polymorphic
55	AQ_SSR_1819	KR052398	55	$(GGA)_{10}$	GATTCAGGGGATGTCATTTAC	GTCCAACTATCGACTTGGAAT	Polymorphic
56	AQ_SSR_1843 (1)	KR052402	55	$(GAT)_{11}$	CAGAAGCTGTACGATTAGATGA	GAACTAAGTGCCTTTGCTTCT	Polymorphic
57	AQ_SSR_1843 (2)	KR052402	54	$(GAT)_{11}$	GATTAGATGAAGTTGGGATCA	GGATCACCTTATCTTTTGCTT	Monomorphic
58	AQ_SSR_1844	KR052403	55	$(GAT)_{8}$	ACGAACAAGTAGAAGATGCAG	ACCATAAGAGCACACACTCAC	Polymorphic
59	AQ_SSR_1850	KR052404	55	$(GAT)_{8}$	AGTTCATCTGGGTTAAAGAGC	CACACTGATGAGTGGGTAAGT	Polymorphic
60	AQ_SSR_1853	KR052405	55	$(TGA)_{10}$	CTGAAACCAGCTGCTACTAAA	ACTAAGCTGGGAAATGGGTAT	Monomorphic
61	AQ_SSR_1864	KR052406	55	$(ATG)_{10}$	ACCTAACTTATGGGGATGATG	TCTATCACCCACTGTTTTGAG	Polymorphic
62	AQ_SSR_1875	KR052407	52	$(GAT)_{8}$	TACTGGAAGAGAACAGAAGCA	CATCATCATCATATCCGAGTC	Polymorphic
63	AQ_SSR_1920	KR052412	52	$(TGA)_{12}$	TCACAGAGAGAGTTCCTCAAA	GGATTAACTTTGCCCATTTC	Not Amplified
64	AQ_SSR_1973	KR052416	55	$(ATG)_{10}$	GTCTTGCTCTTGCTATTCTGA	TGGATTAATCTCATCAGCAG	Monomorphic
65	AQ_SSR_2041	KR259826	55	$(TTC)_{12}$	GTAGGTTGGGTTATTGCTTTT	CATAATTGCAATCAATGACATC	Polymorphic
66	AQ_SSR_2065	KR052420	54	$(TCT)_{16}$	ATTTTCAAGCACATCTGGAC	GCTCTAGGAAAGGGGAAGT	Monomorphic
67	AQ_SSR_2070	KR052421	55	$(TTC)_{14}$	CCTTGTGGCTTCTTCTTCTTA	GTCCACGTTTCTACTACATGC	Polymorphic
68	AQ_SSR_2079	KR052422	55	$(CTT)_{15}$	CTCCTTTGTTGGAACTACCTT	AGATCTTGGAAGAAAGAGCAC	Polymorphic
69	AQ_SSR_2089	KR052423	55	$(TTC)_{9}$	TGCTTGCAGATACACACTCTA	ATCCCAACGCTAAATAAAGG	Polymorphic
70	AQ_SSR_2104	KR052424	55	$(TCT)_{10}$	CGTGTTAGTGTCGTGCTACAT	CGGGTTTCATCAGGTTATACT	Polymorphic
71	AQ_SSR_2128	KR052426	55	$(TTC)_{19}$	CTTCTTATGCCTTTGTTCTC	CAGATCCAAATCATAGTGGTC	Polymorphic
72	AQ_SSR_2140	KR052428	52	$(CTT)_{20}$	GTTTGCATACCAATCTAGGC	GCACACCACACACATCAT	Not Amplified
73	AQ_SSR_2149	KR052429	52	$(CTT)_{22}$	GATCAGTGACCTTCCTTTCC	ACTTAACGCTGAACTCCAAG	Monomorphic
74	AQ_SSR_2171	KR052433	55	$(TTC)_{12}$	TGAAGCTATAGAGTTCGTTGC	CATTGAAAGCCTGAAGTTGT	Polymorphic
75	AQ_SSR_2205	KR782412	54	$(TTC)_{9}$	GACTCCATCTCCTTATCCATC	AAACTTCTGGTTGAGGTTGAT	Polymorphic
76	pGAqu009	KR259827	54	$(GAA)_{4}$	CCGAGGTCAATAAGCTAAAGA	GAGGTCTGGAAATCTTGACA	Monomorphic
77	pGAqu014	KR259828	54	$(TG)_{7}$	GCCTTTAGACAATTCTTCGAC	CAATGACTAAAGGTTAGGGAC	Polymorphic
78	pGAqu025	KR259829	54	$(TC)_{8}$	GAGAAAATTCACCTCTCCAG	GTCAAACCAAGTCAAGTGCT	Monomorphic
79	pGAqu026	KR259830	55	$(TGGAAG)_{3}$	AAATTGTTACCCTGTGATGC	TACACTGAGGACCATACTGCT	Polymorphic
80	pGAqu033	KR259831	54	$(GT)_{6}$	GATAAACTGGGTCTGTTTGTG	GTATGTTTCTTTGCCTCTCTG	Not Amplified
81	pGAqu043	KR259832	54	$(GA)_{18}$	GCATTTAGAGTCACAAGCATC	ACCCTTTTCGCCATACTC	Monomorphic
82	pGAqu070	KR259833	54	$(TG)_{11}$	CTTACGCCAAGGTATATCTCC	GATGAAGTTCAACACTGTAACG	Monomorphic
83	pGAqu072	KR259834	54	$(AC)_{10}$	CACTTCGTTTGGTTGTGTTT	CTGCATGATACTTCCTCCAT	Monomorphic
84	pGAqu074	KR259835	54	$(GACGAG)_{3}$	GTTCAGGAACAGCCATCC	GACATAGCAAACAGTTTCTGG	Monomorphic
85	pGAqu085	KR259836	54	$(AGAAA)_{3}$	GTCATTACACGGTGGAAATAG	TGATCACATTGTCCCTTACTC	Monomorphic
86	pGAqu087	KR259837	54	$(AG)_{15}$	GACAAGCTGGAATGGATATG	GCAAAGGAAGGGACTAATACT	Polymorphic
87	pGAqu099	KR259838	54	$(TG)_{7}$	AAAGGTCCTTCAACCTACAAG	AGCTAGTGCTTGCTAAGAGTG	Monomorphic
88	pGAqu100	KR259839	54	$(CT)_{6}$	TGCTCAGTTCTCTTGTTCTTG	AGCCACTGCTTCTCTACTACA	Monomorphic
89	pGAqu101	KR259840	54	$(CTTG)_{3}$	GTGTTGTTGAGGTAGGATTCA	GCCTAACTTGCTAAGAACTCC	Polymorphic
90	pGAqu107	KR259841	54	$(GATG)_{3}$	GAACAAATCAAGCAACCAAC	TGCTAGGATACGATACAAAGG	Monomorphic
91	pGAqu109	KR259842	54	$(TCAC)_{3}$	AGACTAAAGGGTCGATTCACT	TTCTTCTCCTCCCCTATAGAA	Monomorphic
92	pGAqu117	KR259843	54	$(GT)_{9}$	CACTCGAAAACTTGCAAA	GTGATTCACACACACACACAC	Not Amplified
93	pGAqu119	KR259844	54	(AC)9	GAAATTCTAGCCTTCACAACC	CACTCACTCTCATATGTGCAA	Monomorphic

*All forward primers were tagged with M13 (5'- GTCACGACGTTGTAAAACG -3')-tailed at the 5' end