Indian J. Genet. Plant Breed., (2022) 82(2): 236-239

doi: 10.31742/IJGPB.82.2.15

ISSN: 0975-6906

www.isgpg.org

SHORT RESEARCH ARTICLE

Genome-wide mining of simple sequence repeats and development of polymorphic SSR markers in *Aquilaria malaccensis*

Satyam Bordoloi*, N. Ravi[1], M. K. Modi[2] and R. S. C. Jayaraj

Abstract

Aquilaria malaccensis is one of the economically important tree species distributed in North-east India which produces oleoresinous heartwood known as Agarwood - the most expensive wood in the world. There are limited genomic resources in this species. Therefore, the publicly available whole-genome sequence of *Aquilaria agallocha*, which is a synonym of *A. malaccensis*, was utilized for genome-wide mining of Simple Sequence Repeats (SSR) and to develop polymorphic SSR markers. In total 2,53,155 SSRs were identified from the draft genome. Primers were designed for di-nucleotide and tri-nucleotide repeats and screening of 90 such primers have resulted in the identification of 10 polymorphic markers. The present report is the first one for exhaustive identification of SSRs from the draft genome of the species. The SSR markers identified in the present study may be efficient in genomic analyses and useful in the genetic assessment and conservation of *A. malaccensis*.

Keywords: Agarwood, *Aquilaria malaccensis*, genome-wide, microsatellite, SSR

Aquilaria malaccensis Lam. is a tree species of extremely high economic importance owing to its fragrant resinous heartwood, commonly known as agarwood. Aloeswood, eaglewood, *agaru*, and *gaharu* are all synonyms of agarwood (Naziz et al. 2019). Besides several uses, agarwood has also been useful in medicine and widely used in pharmaceutical industries (Wang et al. 2021). Agarwood is considered to be the most expensive wood in the world. This evergreen tree species belongs to the family Thymelaeaceae and is well distributed in countries of South-east Asia namely, Bangladesh, Bhutan, the North-eastern States of India, Sumatra and Kalimantan of Indonesia, Malaysia, Myanmar, Philippines, Singapore and Thailand (Oldfield et al. 1998). There are 21 accepted species of genus *Aquilaria* of which 13 produce agarwood (Lee and Mohamed 2016; Farah et al. 2020). *A. malaccensis* Lam. and *Aquilaria khasiana* Hallier are found in India (Gardenfors et al. 2001). The chromosome number of *A. agallocha* (synonym *A. malaccensis*) has been determined as $x = 8$ ($2n = 16$) and the karyotype formula as $2n = 2x = 16 = 10m + 6sm$ (Debnath et al. 1995). Variations in genome size have been reported in *Aquilaria* species. For example, the genome size estimation for *A. agallocha* of Myanmar origin is $2C = 1.51$ pg (Chen et al. 2014) while the estimation for *A. malaccensis* of Peninsular Malaysia is $2C = 1.86 \pm 0.02$ pg (Siti Suhaila et al. 2013), while Ding et al. (2020) reported the genome architecture of *Aquilaria sinensis* by using an integrated Nanopore, Illumina and Hi-C sequencing, i.e., about 726.5 Mb in size, which reached a high level of contig N50 of 1.1 <Mb.

Unfortunately, due to extensive felling for its exquisitely fragrant heartwood, the population of the species has declined by over 80 per cent and therefore, the International Union for Conservation of Nature (IUCN) has declared the species as critically endangered (Harvey-Brown

Rain Forest Research Institute, Jorhat 785 010, Assam, India; [1]Institute of Wood Science and Technology, Bengaluru 560 003, Karnataka, India; [2]Assam Agricultural University, Jorhat 785 013, Assam, India

***Corresponding Author:** Satyam Bordoloi, Rain Forest Research Institute, Jorhat 785 010, Assam, India, E-Mail: bordolois@icfre.org; satyamrfri@gmail.comChen

How to cite this article: Bordoloi S., Ravi N., Modi M. K., Jayaraj R. S. C. 2022. Genome-wide mining of simple sequence repeats and development of polymorphic SSR markers in *Aquilaria malaccensis*. Indian J. Genet. Plant Breed., **82**(2): 236-239.

Source of support: Indian Council of Forestry Research and Education, Dehradun

Conflict of interest: None.

Received: Jan. 2022 **Revised:** April 2022 **Accepted:** May 2022

2018). Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) has also enlisted *A. malaccensis* in its Appendix II to regulate international trade in a sustainable way (CITES 2010). Though in North-East India the cultivation of the species is gaining popularity in recent years, the species has already become rare in its natural habitat. It is listed in the Red Data Book of Indian Plants (Nayar and Sastry 1990) and as species facing genetic erosion (Kumar1997).

For genetic diversity analysis different molecular markers are available. However, among the different marker types Single Nucleotide Polymorphisms (SNPs) and Simple Sequence Repeat (SSR) markers are most commonly used nowadays (Davey 2011). The high level of polymorphism in microsatellites makes these markers powerful tools for assessing genetic similarity between individuals or closely related taxa (Guichoux et al. 2011; Kalia et al. 2011). Though microsatellites have been discovered in *A. malaccensis* (Tnah et al. 2012; Singh et al. 2014) the number is very less which limits its applicability. With the advent of Next Generation Sequencing (NGS) techniques, genomes of many species have easily been sequenced which has opened up possibility of genome-wide identification of microsatellite markers in large numbers. The draft genome sequence of *A. agallocha*, which is a synonym of *A. malaccensis*, is available in the public domain (Chen et al. 2014; http://www.ncbi.nlm.nih.gov/assembly/GCA_000696445.1). This genome sequence was utilized for genome-wide mining of SSRs and developing polymorphic SSR markers in this study. In this study, the genome sequence of *A. agallocha* was mined and SSR markers were developed for future use in genetic and genomic analyses of *A. malaccensis*.

Tender leaf samples of 14 genotypes of *A. malaccensis* stored in -80°C, each collected from different populations located at different geographical areas of Northeast India were used for DNA isolation, primary evaluation and screening of polymorphic microsatellites. Isolation of DNA was done with Nucleospin plant II (MACHEREY-NAGEL, Germany) and the samples of DNA (30 ng/µL) was used for PCR reactions. The draft genome sequence of *A. agallocha* was mined with microsatellite identification tool (Thiel *et al*. 2003) for genome-wide identification of microsatellites. The minimum repeat unit was defined for identification as: ten for mono-nucleotide repeat motifs, six for di-nucleotide repeat motifs and five for all other motifs, including tri-nucleotide repeats, tetra-nucleotide repeats, penta-nucleotides repeats and hexa-nucleotide repeats.

The forward and reverse primers of only di- and tri-nucleotide SSRs were designed based on the flanking sequences of SSR repeat motifs using Primer3 software (Untergasser et al. 2012) considering the following criteria:

1. primer length ranging from 20–30 bp, 2. PCR product size ranging from 100-300 bp, 3. annealing temperature ranging from 50^0-65^0 C and 4. GC content ranging from 40 to 60%.

A set of 90 primers was selected for lab validation and screening of polymorphic markers. The top 45 repeats in terms of length in both the repeat classes were chosen. Each set of primers was subjected to PCR reaction in 14 different genotypes. The primers showing variable size bands among different genotypes were selected as polymorphic and the primers showing uniform bands across the genotypes were considered monomorphic and were discarded. PCR reactions were performed in a thermocycler (Biorad T100) in a total reaction mixture of 10 µL, containing: 30 ng of genomic DNA, 5 µL PCR Master Mix (EmeraldAmp® GT), 0.25 µL each of forward and reverse primers and sterile water to reach the final volume. PCR products were electrophoresed on 3% agarose gel (SeaKem® LE) with a 100 bp ladder.

Genome-wide assessment

The size of the haploid *A. agallocha* genome was estimated to be approximately 736 Mb (Chen et al. 2014). In the present study, a total of 1,10,477 sequences containing 689.34 Mb (approx. 94% of the genome) were downloaded and 2,53,155 SSRs were identified. The number of SSR containing sequences was found to be 47,726. The number of sequences containing more than 1 SSR was 32,332. Out of total microsatellites, 9913 (3.91%) SSRs were found to be in compound formation. In total, 193 different types of repeat motifs were detected during the search. For all SSR types, SSR frequency decreased with the increase of the repeat units number (Fig. 1). Mono-nucleotides were the most abundant, accounting for 68.69% (1,73,899) of all SSR types. Of the remaining SSR types *viz.,* di-nucleotide, tri-nucleotide, tetra-nucleotide, penta-nucleotide and hexa-nucleotide repeats accounted for 19.18% (48,553), 9.90% (25,051), 1.87% (4,736), 0.27% (687) and 0.09% (229), respectively.

The frequency of each SSR motif type in the draft genome was also analyzed. In the case of mono-nucleotide repeats, A/T repeats were most abundant (1,71,290) while C/G repeats were comparatively less (2,609). Of the di-nucleotide motifs, AT/AT was the most common (8.82%,

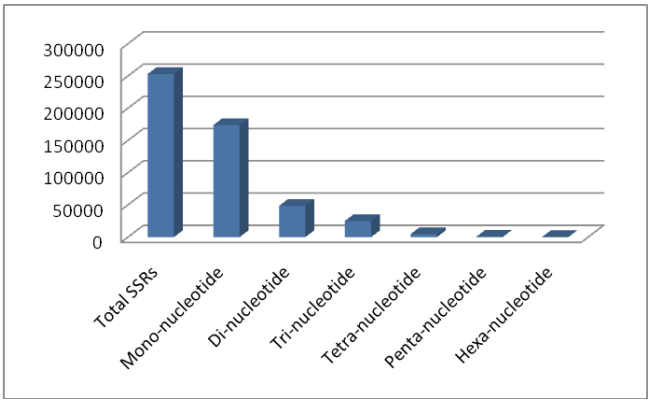

Fig. 1. Distribution of different repeat type classes

Table 1. Polymorphic SSR markers in *A. malaccensis*

Locus Name	SSR	FORWARD PRIMER (5′-3′)	REVERSE PRIMER (5′-3′)	T_a (°C)	Product size (bp)
JMHV01019863.1	(TC) 31	TGAAAAATGTTTGTGGGTGC	GGTTGCTTTATGGGCTTGAG	51.0	275
JMHV01018494.1	(AT) 29	ATATGGCAGGTTTCGATTGC	GGAAGGGAAGCCATCATACA	57.3	269
JMHV01024811.1	(TA) 29	ATTTCGGAATGAATTGTCGG	GTAAATCAGAGCGATGGGGA	57.5	248
JMHV01034634.1	(CT) 29	GCATCCCAACCTTTTTACCA	CGACATTTGCTTGCATGTTT	55.2	266
JMHV01000757.1	(AT) 27	CAACTCAGTCTGGTCGGTGA	ATTTGTGGCCAATCGCTAAC	59.0	218
JMHV01001275.1	(ATA) 24	AACGAACCGAGCTTTACGAA	TACCCGATTTCACGGTTTGT	56.3	250
JMHV01053491.1	(ATA) 24	GTGGCACCCAATTAGGCTAT	GCTTCCCACATTTCATCCAT	57.3	179
JMHV01008521.1	(TAT) 24	TGGCTCGACTTGAACATCTG	GGCGGGTGAGTGAGATTAAC	51.0	279
JMHV01056685.1	(ATT) 23	AATCCATACCGAATCCCTCC	TGTTGCCAAAATGCCATAAA	55.2	265
JMHV01002392.1	(AAT) 23	CTCCAAACCACTCTTGCTCA	CATGTTTAAGGGGTCTGCCT	59.0	207

22,319) followed by AC/GT (5.81%, 14,706) and AG/CT (4.52%, 11,431) while CG/GC motif repeats were very rare (0.04%, 97). The tri-nucleotide repeat motif consists of 10 types and the predominant motif of tri-nucleotide was AAT/ATT which accounted for 5.88% (14,895) followed by AAG/CTT 2.01% (5,087). In the case of tetra-nucleotide repeats, 27 motif types were observed among which- AAAT/ATTT was most abundant (1,717) followed by ACAT/ATGT (1,606). Among 60 different penta-nucleotide repeats, AAAAT/ATTTT were most abundant (247) followed by AAAAG/CTTTT (151). In the case of hexa-nucleotide repeats, AAAAAT/ATTTTT was observed to be in the highest number (37) among 90 different motif types.

Using the primer 3 software tool, a total of 54,087 primers were designed for di- and tri-nucleotide microsatellites. Out of these, a set of 90 primers (45 each for both di- and tri-nucleotide SSRs) were selected for polymorphic study. Ten sets of primers produced clear polymorphic bands (Table 1); five of the polymorphic SSRs were di-nucleotide repeats and the other 5 were tri-nucleotide repeat markers. In *A. malaccensis,* very few reports exist on microsatellite marker development. Tnah et al. (2012) developed 17 polymorphic microsatellite markers while Singh et al. (2015) identified 18 novel highly polymorphic microsatellite markers. They also observed that GA/TC (85.94%) and AGG/ CCT (70.59%) motifs were prominent in microsatellites containing di- and tri repeats, respectively. However, in the present study AT/AT and AAT/ATT were the most abundant di-nucleotide repeat and tri-nucleotide repeat in the genome which are in agreement with the results of Tan et al. (2014). The disagreement with the previous report on *A. malaccensis* is perhaps due to the difference in approach to the identification of SSRs. The present report is the first one for exhaustive identification of SSRs from the draft genome of the species. These markers may help in efficient genetic and genomic analyses and conservation of *A. malaccensis*.

Author's contribution

Conceptualization of research (NR, MKM, RSCJ); Designing of the experiments (NR, MKM, SB); Contribution of experimental materials (SB, MKM); Execution of field/ lab experiments and data collection (SB); Analysis of data and interpretation (SB, MKM, NR); Preparation of the manuscript (SB, RSCJ, MKM).

References

Chen C. H., Kuo T. C. Y., Yang M. H., Chien T. Y., Chu M. J., Huang L. C., Chen C. Y., Hsiao F. L., Shih T. J. and Chen L. F. O. 2014. Identification of cucurbitacins and assembly of a draft genome for *Aquilaria agallocha*. BMC Genomics, **15**(1): 1-11. doi: 10.1186/1471-2164-15-578.

CITES. Appendix II of Convention on International Trade in Endangered Species of Wild Fauna and Flora. 2010. (https://cites.org/eng/app/appendices.php.)

Davey J. W., Hohenlohe P. A., Etter P. D., Boone J. Q., Catchen J. M. and Blaxter M. L. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet., **12**(7): 499-510.

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

Ding X., Mei W., Lin Q., Wang H., Wang J., Peng S., Li H., Zhu J., Li W., Wang P. et al. 2020. Genome sequence of the agarwood tree *Aquilaria sinensis* (Lour.) Spreng: the first chromosome-level draft genome in the Thymelaeceae family. Giga Science, **9**(3): giaa013, https://doi.org/10.1093/gigascience/giaa013.

Farah A.H., Lee S.Y., Gao Z., Yao T.L., Madon M. and Mohamed R. 2020. Genome Size, Molecular Phylogeny, and Evolutionary History of the Tribe Aquilarieae (Thymelaeaceae), the Natural Source of Agarwood. Front. Plant Sci., **29**: https://doi.org/10.3389/fpls.2018.00712

Gärdenfors U., Hilton T. C., Mace G.M. and Rodriguez J. 2001. The application of IUCN red list criteria at regional levels. Conserv. Biol., **15**: 1206–1212

Guichoux E., Lagache L., Wagner S., Chaumeil P., Léger P., Lepais O. and Petit R.J. 2011. Current trends in microsatellite genotyping. Mol. Ecol. Resour., **11**(4): 591-611.

Harvey-Brown Y. 2018. Aquilaria malaccensis. The IUCN Red List of Threatened Species. e.T32056A2810130. https://dx.doi.org/10.2305/IUCN.UK.2018-1.RLTS.T32056A2810130.en.

Kalia R. K., Rai M. K., Kalia S., Singh R. and Dhawan A. K. 2011. Microsatellite markers: an overview of the recent progress in plants. Euphytica, **177**(3): 309-334.

Kumar S., Singh J., Shah N. C. and Ranjan V. 1997. Indian Medicinal and Aromatic Plants Facing Genetic Erosion. Central Institute of Medicinal and Aromatic Plants. Lucknow.

Lee S. Y., Ng W. L., Mahat M. N., Nazre M. and Mohamed R. 2016. DNA Barcoding of the endangered *Aquilaria* (Thymelaeaceae) and its application in species authentication of agarwood products traded in the market. PLoS One, **11**: e0154631. doi: 10.1371/journal.pone.0154631.

Mason A. S. 2015. Challenges of genotyping polyploid species. Plant Genotyping, 161-168.

Nayar M. P. and Sastry A. R. K. 1990. Red Data Book of Indian Plants, Vol. 3. Botanical Survey of India, Calcutta.

Naziz P. S., Das R. and Sen S. 2019. The Scent of Stress: Evidence From the Unique Fragrance of Agarwood. Front. Plant Sci., **10**: 840. DOI: 10.3389/fpls.2019.00840.

Oldfield S., Lusty C. and MacKinven A. 1998. In: The Word List of Threatened Trees. World Conservation Press, Cambridge, UK: 650.

Singh P., Sharma H., Nag A., Bhau B. and Sharma R. 2015. Development and characterization of polymorphic microsatellites markers in endangered *Aquilaria malaccensis*. Conservation Genetics Resources, **7**: 61-63.

Suhaila A. S., Norihan M. S., Norwati M., Azah M. N., Mahani M. C., Parameswari N., Kodi I.K., Mailina J., Azrina A., Hasnida H.N. and Haliza I., Nazirah A. and Fuad Y. M. 2015. Aquilaria malaccensis polyploids as improved planting materials. J. Trop. Forest Sci., 376-387.

Tan M., Wu K., Wang L., Yan M., Zhao Z., Xu J., et al. 2014. Developing and characterizing Ricinus communis SSR markers by data mining of whole-genome sequences. Mol. Breed., **34**: 893-904.

Thiel T., Michalek W., Varshney R. K. and 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): 411-22.

Tnah L. H., Lee C. T., Lee S. L., Ng K. K., Ng C. H., Nurul-Farhanah Z., Lau K. H. and Chua L. S. 2012. Isolation and characterization of microsatellite markers for an important tropical tree, *Aquilaria malaccensis* (Thymelaeaceae). Am. J. Bot., **99**(11): e431-3.

Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B.C., Remm M. et al. 2012. Primer3–new capabilities and interfaces. Nucleic Acids Res., **40**: e115.

Wang Y., Hussain M., Jiang Z., Wang Z., Gao J. , Fengxian Ye , Mao R. and Li H. 2021. *Aquilaria* species (Thymelaeaceae) Distribution, Volatile and Non-Volatile Phytochemicals, Pharmacological Uses, Agarwood Grading System, and Induction Methods. Molecules, **26**: 7708. https://doi.org/10.3390/molecules26247708. https://www.mdpi.com/journal/molecules.