TÜBİTAK

Turkish Journal of Botany

http://journals.tubitak.gov.tr/botany/

Research Article

Turk J Bot
(2022) 46: 26-36

doi:10.3906/bot-2108-61

Mining and analysis of chloroplast simple sequence repeats (SSRs) from eight species of *Aquilaria*

Muhammad Syahmi HISHAMUDDIN[1,2,*], Samsuddin Ahmad SYAZWAN[1,2], Rozi MOHAMED[1]

[1]Forest Biotechnology Laboratory, Department of Forestry Science and Biodiversity, Faculty of Forestry and Environment, Putra Malaysia University, Selangor, Malaysia

[2]Mycology and Pathology Branch, Forest Biodiversity Division, Forest Research Institute Malaysia (FRIM), Selangor, Malaysia

Received: 27.08.2021 • **Accepted/Published Online:** 22.12.2021 • **Final Version:** 31.01.2022

Abstract: *Aquilaria* is a tropical forest tree, producer of the famed and expensive agarwood. Aggressive collection of agarwood put strain on the natural stands of *Aquilaria* species, sparking efforts to domesticate the tree and cultivate agarwood in plantations. However, tree domestication progress is hampered by the scarcity of genomic resources that is crucial for breeding programs. In this study, the complete chloroplast (cp) genome sequences from eight *Aquilaria* species were analyzed in silico. For identification of the simple sequence repeats (SSRs), MISA PERL script which had a repeat length of 12 for mononucleotides (mono-), 6 for dinucleotides (di-), 4 for trinucleotides (tri-), 3 for tetranucleotides (tetra-), pentanucleotides (penta-), and hexanucleotides (hexa-), respectively, along with frequency were utilized. From a total of 312 SSRs that were discovered, merely 50 (16%) were found localized within the coding region while the majority (84%) were within the intergenic regions, with an average of one SSR per 4.5 kb. The mean length of the SSRs were 11.63 bp. Mono- repeats were the predominant motifs (29.2%), followed by tetra- (28.8%), di- (20.5%), tri- (19.9%), and penta- (1.6%). Whereas the most recurring motifs were A/T (97.8%) for mono-, AT/AT (87.5%) for di-, AAT/ATT (48.4%) for tri-, and AAAT/ATTT (45.6%) for tetra-. GO analysis using the REVIGO software identified four molecular functions, six biological processes and three cellular components. In conclusion, findings of this study offer a scientific foundation for future phylogenetics, evolutionary genetics, diversity studies and breeding programs on *Aquilaria* species.

Key words: Gene ontology analysis, in silico PCR, transferability, molecular marker

1. Introduction

Aquilaria spp. are endemic to the Indomalesian region, with a total of 22 species reported (The Plant List, 2013). *Aquilaria* thrives on marginal terrain and under a wide variety of environmental conditions. They are rapidly growing trees that can reach a diameter of 10 cm diameter at breast height (DBH) in four to six years in locations with optimum moisture (Blanchette et al., 2015). They are best known for producing the fragrant resin agarwood, which is widely used as a raw material in perfumes, incenses, and herbal medicines (Mohamed and Lee, 2016). Agarwood chips of high grade are frequently offered for around $4000 USD per kilo, while far higher prices of $10,000 USD or more per kilo are frequently encountered on the market (Blanchette et al., 2015). The high commercial value of agarwood fuels indiscriminative cutting of *Aquilaria* Lam., trees, reducing their natural populations (Barden et al., 2000). To minimize unsustainable harvesting of *Aquilaria* genetic resources, the species has been listed on the International Union for Conservation of Nature's (IUCN) Red List of Threatened Species in 1998 (IUCN, 2021), and have also been listed in Appendix II of the Convention on International Trade in Endangered Species (CITES) in 2005 (CITES, 2004). A study suggested an option for mitigating the impact of *Aquilaria* tree overexploitation by establishing large-scale ex situ plantations coupled with techniques to increase agarwood production in young plants (Faridah-Hanum et al., 2009). Except, necessary tools for identifying seeds and seedlings are required for the production and breeding of *Aquilaria* in the nursery.

A range of different molecular markers have been created and is widely used in genetic and breeding studies, where it is well understood that breeding studies mainly determined by the extent of genetic variability (Smith et al., 1991). Among them is simple sequence repeats (SSRs markers) (Tautz, 1989), also known as microsatellites. SSRs markers is a family of molecular markers comprised of tandem repeats of short (1-6 nucleotide) DNA sequences that are abundant in both the coding and noncoding regions of the cp genome (Vieira et al., 2016).

* Correspondence: muhdsyahmihisham@gmail.com

SSRs are classified as mono-, di-, tri-, and tetranucleotide repeats, and one of the most important factors influencing mutation rate is the length of the microsatellite. Andru et al. (2011) developed and applied SSR markers for sugarcane breeding in 2011. In their study, they utilized SSRs to develop a correlation map of LCP 85-384, which allowed them to analyze the segregation pattern in the mapping population and chromosome pairing behavior during meiosis. LCP 85-384 contains good agronomic properties as well as resistance to biotic and abiotic stresses, and the S1 progeny formed by selfing LCP 85-384 segregate in response to these attributes. The framework map in this study will provide crucial information that will be useful for clone crossing and selection in the breeding program.

The SSRs markers system was created for both nuclear and chloroplast genomes. However, due to the rapid evolution of sequences, SSRs markers derived from the nuclear genome are insufficient for phylogenetic analysis of distinct species or genera. This hinders comparisons of sequences and allele sizes across the species level (Powell et al., 1996). On the other hand, chloroplast SSRs (cpSSRs) obtained from the chloroplast genome sequence, deems a suitable complimentary molecular tool as a nuclear genetic marker. This is due to the fact that SSRs loci in the chloroplast genome are frequently dispersed throughout noncoding regions of the chloroplast genome and exhibit a higher degree of sequence variation than coding regions, despite the fact that chloroplast DNA has a low evolutionary rate devoid of recombination rate (McCauley, 1995; Provan et al., 2001). Therefore, cpSSRs (microsatellite) markers can be utilized to study population genetics and biogeography, as well as to shed light on the genetic relationships between closely related species.

The length of the microsatellite is one of the most significant parameters influencing mutation rate (Ellegren, 2004). Development of SSRs markers was a laborious approach because it consumed a lot of time, costly method of genomic library generation and sequenced a vast number of clones to identify SSRs-containing DNA regions (Kale et al., 2012). To facilitate this challenge, conventional techniques for generating SSRs markers from genomic libraries have been rapidly replaced by in silico mining of SSRs from DNA sequences deposited in biological databases (Shanker, 2014). In the early 1990s, SSRs markers were first demonstrated to be beneficial in plant molecular genetics in the *Glycine* subgenus *soja*, a subgroup of the seasonal cultivated soybean *Glycine max* (L.) Merr. and its suspected wild progenitor *Glycine sugenus soja* Siebold & Zucc. (Akkaya et al., 1992; Morgante and Olivieri, 1993). For *Aquilaria* species, studies on SSRs began in 2010 with Eurlings et al. (2010), who successfully assessed the polymorphisms of five SSRs isolated from an *Aquilaria crassna* Pierre ex Lecomte genomic library and, based on the results of the cross specificity using identified SSRs markers, determined two related *Aquilaria* species bred in China and Vietnam as well as one SSRs locus fruitfully amplified from wood and incense samples. In the same year, Zhang et al. (2010) reported their findings from assessing polymorphisms of eight SSRs loci (isolated from *Aquilaria sinensis* Merr. genomic libraries) on 31 individuals in the wild. In the following years of 2012 and 2015, Tnah et al. (2012) and Singh et al. (2015) reported that 17 and 18 SSRs derived from enriched genomic libraries for *Aquilaria malaccensis* Lam. were characterized using 24 and 45 samples from a natural population, respectively. Thereafter, in 2016, Chua et al. (2016) attempted to analyze 963 samples of *A. malaccensis* trees from 35 populations using a total of 17 SSRs reported by Tnah et al. (2012). Other than nuclear SSRs, cpSSRs are widely used to assess the genetic structure and diversity of other species such as the red pine (Echt et al., 1998), *Pyrus calleryana* Decne. (Kato et al., 2013), *Salix caprea* L. (Perdereau et al., 2014) and rubber tree (Phumichai et al., 2015) within a population. A distinguishing aspect of cpSSRs is their nonrecombinant, uniparentally inherited nature, which is characterized by the presence of eight to fifteen repeating mononucleotide repeats (Ozyigit et al., 2015).

The complete *Aquilaria* cp genome has been sequenced: it is a circular double-stranded DNA molecule with a size of 174,693 bp and two inverted repeat regions of 42,090 bp each (Hishamuddin et al., 2020). The accessibility of complete cp genome of eight *Aquilaria* species from our previous study (*Aquilaria beccariana* Tiegh. MN125347, *Aquilaria crassna Pierre ex Lecomte*MN125348, *Aquilaria hirta* Ridl. MN125349, *Aquilaria malaccensis Lam.* MH286934, *Aquilaria macrocarpa* Baill. MN125350, *Aquilaria rostrata* Ridl. MN125351 *Aquilaria sinensis (Lour.) Spreng* MN147870 and *Aquilaria subintegra* Ding Hou MN147871) at the National Centre of Biotechnology Information (NCBI) database allows for the mining and identification of SSRs. As a result, the current effort was designed to discover SSRs in the *Aquilaria* cp genomes computationally (in silico method), estimate their frequency and distribution, and also determine SSRs crosstransferability within the genus *Aquilaria*.

2. Materials and methods

2.1. Retrieval of chloroplast genome sequences

The complete cp genome sequences of eight species of *Aquilaria* were retrieved in FASTA and GenBank formats from NCBI (http://www.ncbi.nlm.nih.gov/genomes) (Table 1).

2.2. Mining of simple sequence repeats (SSRs) and primer designing

Classification of cpSSRs was performed using the MISA PERL script, which is available online at http://pgrc.ipk-gatersleben.de/misa/. MISA is capable of determining

Table 1. The information of chloroplast genomes used in SSR mining.

No	Organism	Accession number	Genome size (kb)
1	*Aquilaria beccariana*	MN125347	174.831
2	*Aquilaria crassna*	MN125348	174.830
3	*Aquilaria hirta*	MN125349	174.761
4	*Aquilaria malaccensis*	MH286934	174.832
5	*Aquilaria microcarpa*	MN125350	174.819
6	*Aquilaria rostrata*	MN125351	174.693
7	*Aquilaria sinensis*	MN147870	174.907
8	*Aquilaria subintegra*	MN147871	174.828

the number and distribution of perfect SSRs (where each repeat comes after the previous one without a separation) as well as compound SSRs (repeats that are close to each other). The minimum repetition size was defined as >=12 mononucleotides, >=6 dinucleotides, >=4 trinucleotides, and >=3 tetranucleotides, pentanucleotides, and hexanucleotides. The maximum difference between two SSRs was 0. The Primer3 tool (http://bioinfo.ut.ee/primer3-0.4.0/primer3/input.htm) was used to create primer pairs for each SSR's flanking sequences with default function parameters of GC content, melting temperature (Tm), primer, and size of the PCR product that was further crosschecked with PCR Primer Stats which is available in the Sequence Manipulation Suite (Stothard, 2000).

2.3. Gene ontology analysis

Blast2GO was used to functionally annotate the coding sequences of *Aquilaria* species (Conesa et al., 2005). A FASTA file including all coding SSRs was matched to sequences annotated with the gene ontology (GO) through the Blast2GO program based on Fisher's exact test and a false discovery rate (FDR) of 0.05, whereby GO enrichment analysis have been used to classify the functional categories that were abundant in the plant UniProt database. (Conesa et al., 2005). The enrichment analysis results were displayed and summarized using REVIGO software (Reduce and Visualize Gene Ontology; http://revigo.irb.hr/) by eliminating redundant GO terms.

2.4. In silico polymerase chain reaction (PCR)

To evaluate product amplification and study transferability, SnapGene software (from Insightful Science; available at snapgene.com) with default function parameters was used to perform in silico PCR as well as gel simulation on primer pairs optimized for cpSSRs of *Aquilaria* species. As templates, the complete cp genome sequences of *Aquilaria* species were utilized, and virtual amplicons were generated for each primer pair. On the basis of different repeats (mono, di, tri and tetra), a total of 25 candidate primer pairs were identified for transferability validation, and primer selection will be based on a 100% transferability result.

3. Results

3.1. SSRs frequency and distribution in *Aquilaria* chloroplast genomes

A total of eight cp genomes of *Aquilaria* species were examined for the presence of cpSSRs, and 312 perfect cpSSRs were identified, of which 50 (16%) were within the coding regions while 262 (84%) were within the intergenic regions. Supplementary file 1 contains information on the cpSSRs that were identified. The density of cpSSRs discovered in the *Aquilaria* cp genome is as follows: *A. beccariana* (1 SSR/4.3 kb), *A. crassna* (1 SSR/4.5 kb), *A. hirta* (1 SSR/4.7 kb), *A. malaccensis* (1 SSR/4.3 kb), *A. macrocarpa* (1 SSR/4.6 kb), *A. rostrata* (1 SSR/4.4 kb), *A. sinensis* (1 SSR/4.6 kb) and *A. subintegra* (1 SSR/4.6 kb). According to data from *Aquilaria* cp genomes, the average length of cpSSRs is 11 bp (Table 2). Among the *Aquilaria* species, the largest number of cpSSRs (41) were discovered in *A. beccariana* and *A. malaccensis*, while the lowest number of cpSSRs (37) was discovered in *A. hirta*. There were no penta- SSRs identified in *A. crassna*, *A. sinensis* and *A. subintegra* (refer Table 2). Mononucleotides were found to have the most repeats, followed by tetra- (Figure 1), with A being the most frequent repeat followed by T. The most of SSRs in the *Aquilaria* cp genome are mononucleotide repeats (A/T). The A/T motifs (97.8%), AT/AT motifs (87.5%), AAT/ATT motifs (48.4%) and AAAT/ATTT motifs (45.6%) had the highest frequencies for mono-, di-, tri- and tetra- in all *Aquilaria* cp genomes, respectively (Tables 3 and S1).

Based on gene ontology analysis, the molecular functions of cpSSRs identified were categorized according to its binding, catalytic activity, signaling receptor activator activity and transporter activity; whereas for the biological process, it was categorized according to its cellular component organization or biogenesis, cellular process, cell

Table 2. Information of SSRs identified in chloroplast genomes of *Aquilaria*.

Parameters	Organisms							
	Aquilaria beccariana	*Aquilaria crassna*	*Aquilaria hirta*	*Aquilaria malaccensis*	*Aquilaria microcarpa*	*Aquilaria rostrata*	*Aquilaria sinensis*	*Aquilaria subintegra*
Chloroplast genome size (bp)	174,831	174,830	174,761	174,832	174,819	174,693	174,907	174,828
Total SSRs identified	41	39	37	41	38	40	38	38
Density of SSR (kb)	1 SSR/4.3	1 SSR/4.5	1 SSR/4.7	1 SSR/4.3	1 SSR/4.6	1 SSR/4.4	1 SSR/4.6	1 SSR/4.6
Coding region	5 (12.2%)	7 (18%)	5 (13.5%)	7 (17%)	6 (15.8%)	6 (15%)	7 (18.4%)	7 (18.4%)
Average length of SSR (bp)	11.85	11.38	11.76	11.85	11.58	11.75	11.47	11.39
Repeat type								
Mononucleotides	13 (31.7%)	12 (30.8%)	11 (29.7%)	13 (31.7%)	9 (23.7%)	11 (27.5%)	11 (28.9%)	11 (28.9%)
Dinucleotides	8 (19.5%)	8 (20.5%)	8 (21.6%)	8 (19.5%)	8 (21.1%)	8 (20%)	8 (21.1%)	8 (21.1%)
Trinucleotides	8 (19.5%)	8 (20.5%)	6 (16.2%)	8 (19.5%)	8 (21.1%)	8 (20%)	8 (21.2%)	8 (21.2%)
Tetranucleotides	11 (26.8%)	11 (28.2%)	11 (29.7%)	11 (26.8%)	12 (31.5%)	12 (30%)	11 (28.9%)	11 (28.9%)
Pentanucleotides	1 (2.5%)	-	1 (2.8%)	1 (2.5%)	1 (2.6%)	1 (2.5%)	-	-

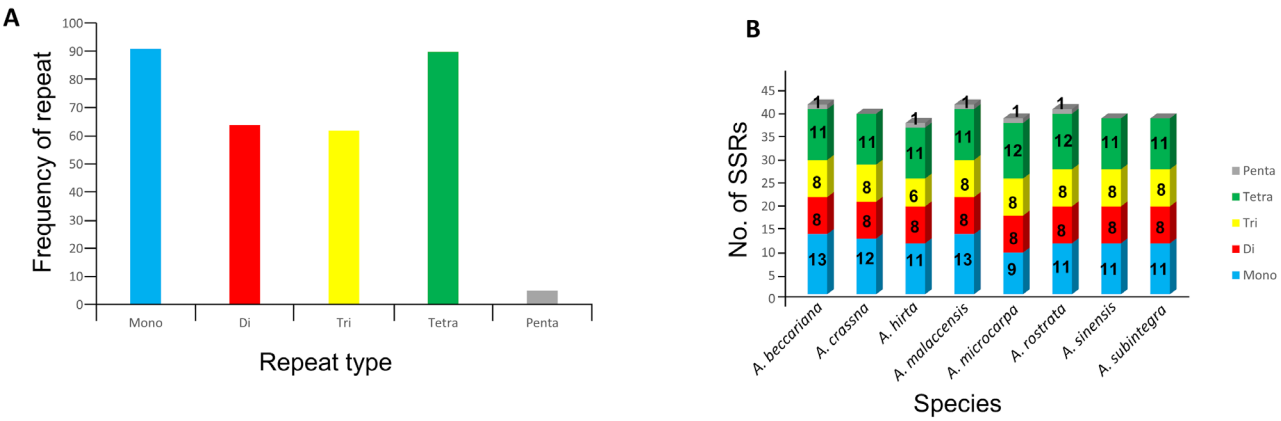

Figure 1. (A) Distribution of cpSSRs in genus *Aquilaria*, (B) types of SSRs detected in eight chloroplast genomes of genus *Aquilaria*.

cycle, growth, signaling, and mitotic cell cycle process. The cellular component was classified as membrane, (obsolete) membrane part, integral component of membrane and intrinsic component of membrane (Figures 2A–2C). In this study, the most enriched molecular function, biological process and cellular component are displayed with larger components/boxes. Though the Treemap depicts only the molecular function, biological process, and cellular component, an interactive graph is constructed to display the relationship and functional network associated with each GO word (Figures 3A–3C).

To minimize functional category redundancy among enriched GO terms, a semantic similarity technique was used. The gene ontology interactive graph-based network in the molecular function comprised of five nodes and eight edges, nine nodes and three edges in the biological process, whereas five nodes and two edges in the cellular component. The network interaction in the molecular function showed that all five nodes (GO term related to binding (GO:0005488), catalytic activity (GO:0003824), molecular function (GO:0003674), transporter activity (GO:0005215) and signaling receptor activator activity (GO:0030546)) correlated with each other. Notwithstanding, in the biological process, only three out of nine nodes correlated with each other that was nuclear division (GO:0000280), organelle fission (GO:0048285) and organelle organization (GO:0006996). However, in the cellular component display, intrinsic component of membrane (GO:0031224) correlated with two nodes which were an integral component of membrane (GO:0016021) and membrane (GO:0044425).

Table 3. The most frequent motif types of all *Aquilaria* species for mono-, di-, and tri- and tetra- repeat types.

Repeat types	*Aquilaria beccariana*	*Aquilaria crassna*	*Aquilaria hirta*	*Aquilaria malaccensis*	*Aquilaria microcarpa*	*Aquilaria rostrata*	*Aquilaria sinensis*	*Aquilaria subintegra*
A/T	13	12	11	13	9	10	10	11
AT/AT	7	7	7	7	7	7	7	7
AAT/ATT	4	4	2	4	4	4	4	4
AAAT/ATTT	5	5	5	5	6	5	5	5

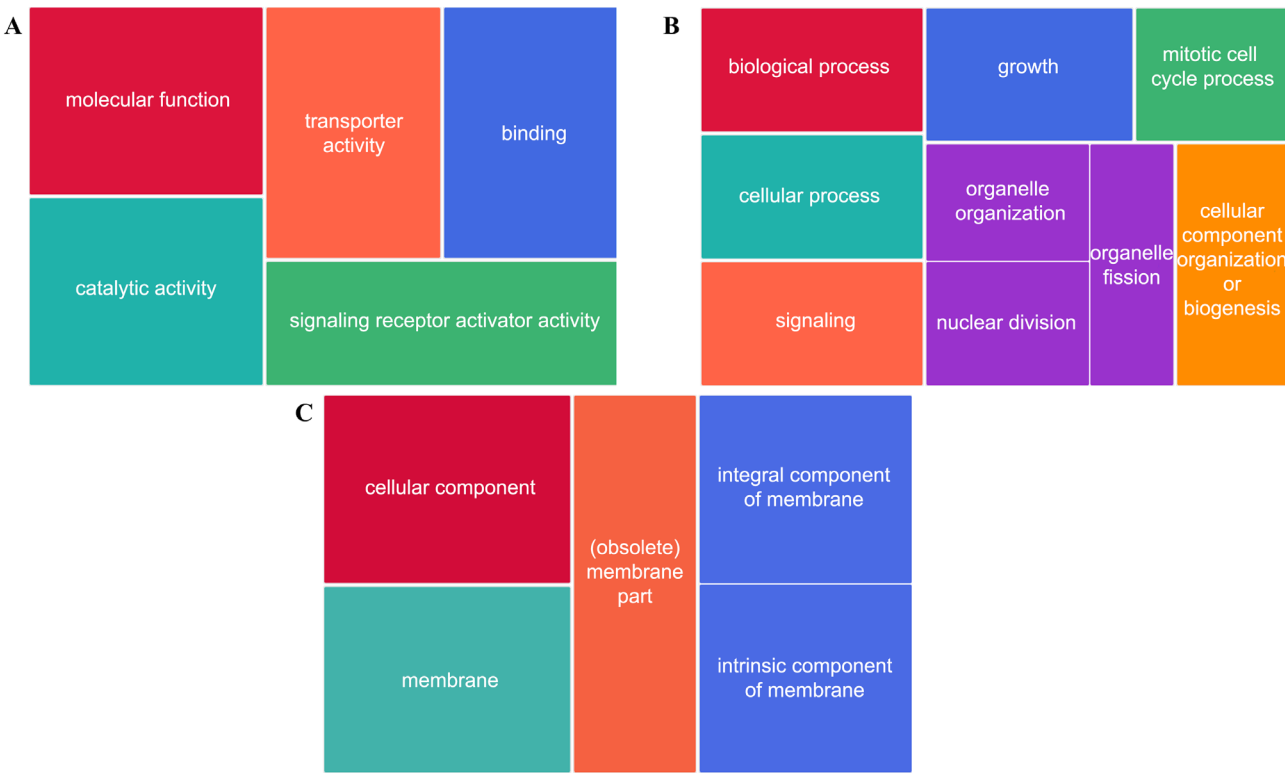

Figure 2. Treemap of gene ontology distribution into molecular functions, biological processes and cellular components of cpSSRs for *Aquilaria* species. (A) Molecular function category, (B) biological process category, (C) cellular component category. The most enriched biological process is shown as larger components within the map. Treemap was generated using REVIGO (Supek et al., 2011).

3.2. Crosstransferability of SSRs

Ten out of 25 cpSSRs ((A)14, (TA)6, (TA)7, (TC)7, (ACA)4, (TGA)4, (TTA)4, (ATAA)3, (ATTC)3 and (TAAT)3) showed complete transferability (100%) within the genus using in silico PCR and a gel simulator provided by SnapGene software for the *Aquilaria* cp genomes (*A. beccariana*, *A. crassna*, *A. hirta*, *A. malaccensis*, *A. microcarpa*, *A. rostrata*, *A. sinensis* and *A. subintegra*) (Figure 4, Table 4).

4. Discussion

Simple sequence repeats are dispersed through the sequence of plant genomes even though the frequency as well as patterns of distribution are different (Sonah et al., 2011). SSRs can be classified into three groups according to their genomic location: (1) nuclear SSRs (nSSRs), (2) chloroplast SSRs (cpSSRs), and (3) mitochondrial SSRs (mtSSRs) (Kalia et al., 2011). In accordance with previous studies, there was a vast definition of dominant repetition in each plant. For instance, the dominant repetitions that was found in mung beans was defined as mono- (Chen et al., 2015), tri- within citrus (Liu et al., 2013), tetra- within cucumber (Cavagnaro et al., 2010), and penta- within cotton (Zou et al., 2012). Moreover, the number of repeats was not same at the genomic and transcriptional levels (Moe et al., 2011). Simple sequence repeats were found in lower numbers in coding regions than in noncoding regions (Ueno et al., 2012). Our results corroborated the identification of nearly all SSRs found in noncoding DNA.

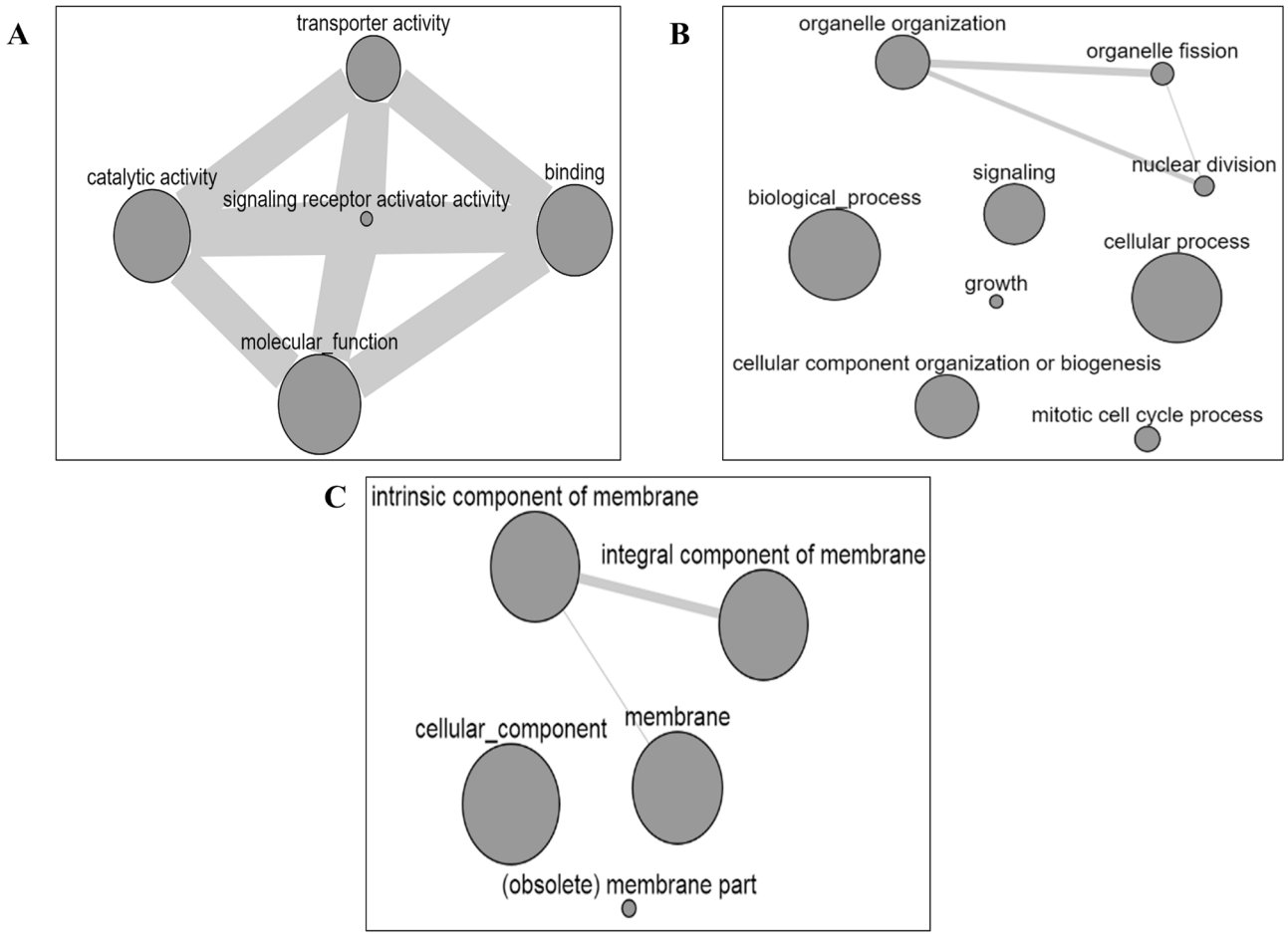

Figure 3. Interactive graph-based enriched gene ontology view from REVIGO. (A) Molecular function category, (B) biological process category, (C) cellular component category. The bubble size indicates the frequency of the GO term. Highly similar GO terms are linked by its edges in the graph, where the width of the line indicates the degree of similarity.

The present study used bioinformatics software (MISA PERLscript) to screen for cpSSRs in eight *Aquilaria* cp genomes available in GenBank. The majority of SSRs discovered in the *Aquilaria* cp genomes are mononucleotide repeats (T/A) (Table 1 and Figure 1), and were consistent with the findings in monocots and dicots (Sonah et al., 2011). In addition, AT/TA repeats instead of AG/CT were the most common dinucleotide found in *Aquilaria* species. Although, the most frequent AG/CT dinucleotide was identified in plant peppers (Cheng et al., 2016). CG/CG repeats were extremely rare among dinucleotide repeats in the organellar genomes coding and noncoding regions (Shukla et al., 2018), which intriguingly did not occur in this study.

A sum of 312 perfect cpSSRs were discovered within the cp genomes of *Aquilaria*, with an average frequency of 1 SSR/4.5 kb. The average cpSSRs density in *Aquilaria* species was found to be less than that of the Nightshade (1 SSR/1.26 kb) and Hornwort (1 SSR/2.4 kb) (Tambarussi et al., 2009; Shanker, 2013). *Aquilaria* species were found to have a higher density than *Oryza* (1 SSR/6.5 kb) and EST-SSRs in barley, maize, rye, sorghum, and wheat (Varshney et al., 2002; Rajendrakumar et al., 2007) (Table 2). In general, the number of intergenic SSRs in *Aquilaria* species is greater than the number of coding SSRs (Table 2), which is likely due to the lower polymorphism of coding regions compared to noncoding ones.

The frequency of SSRs repeats within a diverse region of the cp genome varies greatly, suggesting that disparate repeats play different roles in gene ontology studies (Qian et al., 2013). Besides that, tetranucleotide repeats were discovered to be more common than trinucleotide repeats in *Aquilaria* cp genomes, which is not consistent with previous reports where tetranucleotide repeats were typically low (Liu et al., 2018) (Table 3). Gene annotation of coding cpSSRs in various *Aquilaria* species (*A. beccariana*, *A. crassna*, *A. hirta*, *A. malaccensis*, *A. microcarpa*, *A. rostrata*, *A. sinensis* and *A. subintegra*) were categorized according to many functions. The distribution

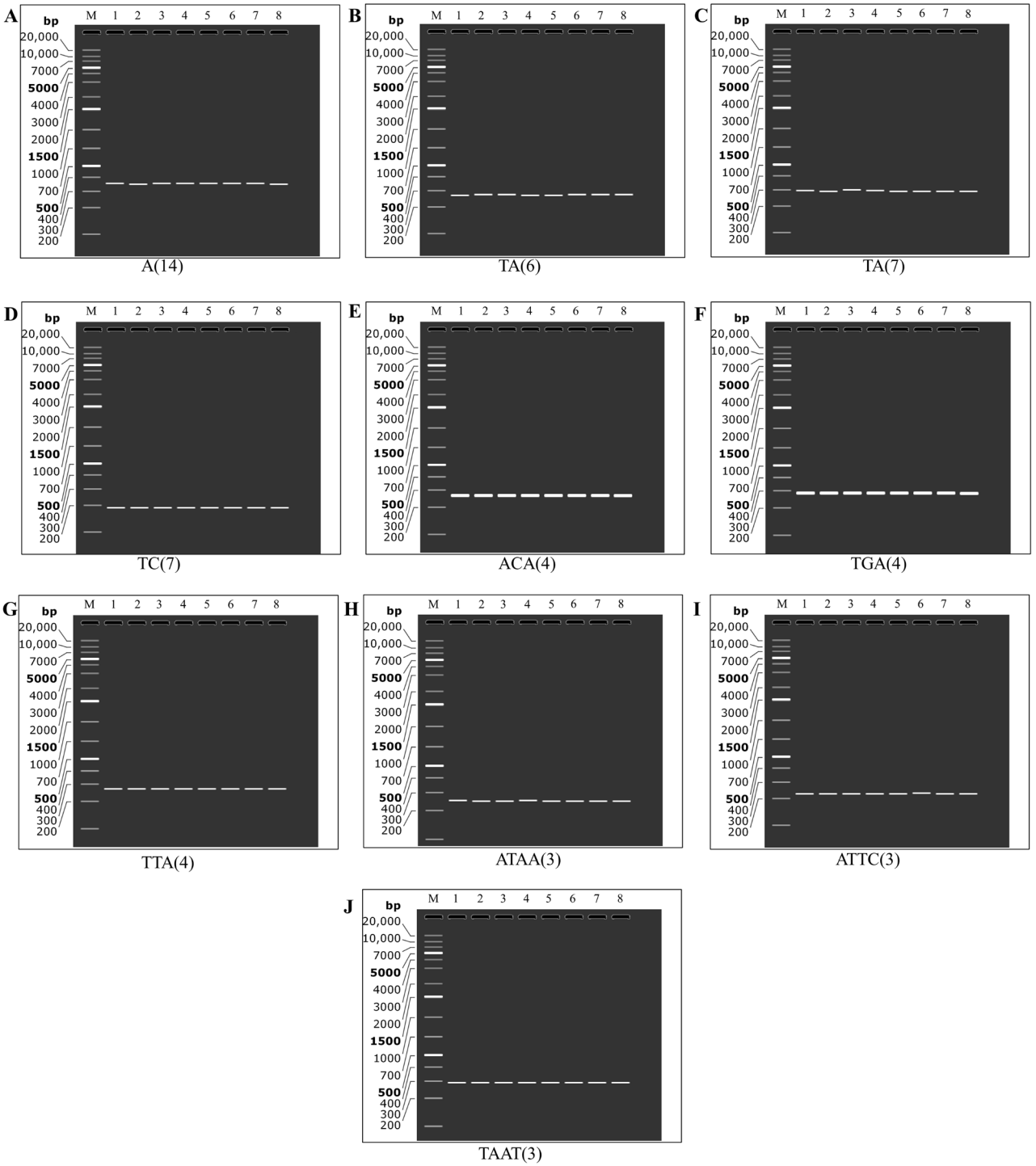

Figure 4. In silico transferability of cpSSRs for *Aquilaria* species. Gel lanes are labeled across the top as follows: M: DNA ladder, Figures 4A–4J (primer for repeat motif A(14), TA(6), TA(7), TC(7), ACA(4), TGA(4), TTA(4), ATAA(3), ATTC(3) and TAAT(3) lane 1-8: *A. beccariana*, *A. crassna*, *A. hirta*, *A. malaccensis*, *A. microcarpa*, *A. rostrata*, *A. sinensis,* and *A. subintegra*.

of sequences including SSRs shared by *A. beccariana*, *A. crassna*, *A. hirta*, *A. malaccensis*, *A. microcarpa*, *A. rostrata*, *A. sinensis,* and *A. subintegra* was observed to have descended from a common ancestor. Therefore, the data of SSRs in coding region generated from this study would improve the utilization of coding sequences in the development of genetic relationships and evolutionary studies among *Aquilaria* species.

Table 4. Primers designed for SSRs identified in *Aquilaria* chloroplast genomes.

No	Motif	Start	End	Forward (F)/Reverse Primer (R)	Primer Length (unit)	Tm °C	GC%
1	(A)14	7694	7707	F: GACACGTCTAGATATAGAATTCAACCG R: AGCATAGGCTTCGGGCAATTTGGC	27 24	53 60	40.7 54.2
2	(TA)6	62972	62983	F: GTCGAGTGGATTCAGTCAC R: GCAATGTGTGGATTGTTC	19 18	52.7 49.3	52.6 44.4
3	(TA)7	33570	33583	F: TCGGATCTCTGTCAAATTGC R: GGGCCATTTGATTCAATTAC	20 20	52.7 49.6	45 40
4	(TC)7	17672	17685	F: ATGCCCAAAATGAACTCCTG R: CGACCAATCTCCGGTAGAAG	20 20	53.5 54.7	45 55
5	(ACA)4	112770	112781	F: GAGATGTCGTTTCTAGTCTATC R: GACCCCAACGACCGAATTGC	22 21	49.8 59.1	40.9 60
6	(TGA)4	94579	94590	F: CCCGTCTCATGGAAAACCC R: CAAGGTCACCGTCACTAGCA	19 20	55.9 56.9	57.9 55
7	(TTA)4	39706	39717	F: TCACTCAAGGACGGAACC R: CGGGCAAATTTGGTTAATTC	18 20	53.9 50.4	55.6 40
8	(ATAA)3	6686	6697	F: TTGAATCCGCCTTACATTGTC R: GTTCAGCCTACCTGATAGATACGG	21 24	53.2 54.2	42.9 50
9	(ATTC)3	25133	25144	F: GGGTTTTGTCACCTCTCC R: CGATCTTTGGCTTTGGAAC	18 19	52.9 51.8	55.6 47.4
10	(TAAT)3	30483	30494	F: CGTAGTATGGGGAAGAAGTGG R: TCGGAATACGAAAATGGG	21 18	54.6 49.5	52.4 44.4

Based on the GO visualization results generated by the REVIGO software, the most enriched molecular function, biological process and cellular component is depicted with larger boxes and was found exhibiting different types of processes. In the molecular function category, there are four processes/GO termed: (1) GO:0005488 (binding - for ligands that bind to receptors for signal transduction); (2) GO:0003824 (catalytic activity - involve in enzyme activity); (3) GO:0005215 (allows for the controlled movement of substances (such as macromolecules, tiny compounds, and ions); and (4) GO:0030546 (activity of a signaling receptor activator - the capability of interacting with receptors (directly or indirectly) in order to raise the number of receptors in the active state).

Whereas in the biological process category there are several important GO term such as: GO:0009987 (cellular process - involve in cell communication), GO:0071840 (biogenesis - the process through which constituent macromolecules are biosynthesized), GO:0040007 (growth - part of an organism or a cell) and GO:0023052 (signaling - engage in the process of information transmission inside a biological system), while in the cellular component categories there are also several important GO term such as: GO:0016020 (membrane: a lipid bilayer that contains and is linked to all the proteins and protein complexes) and GO:0044425 (obsolete membrane part - a component of the membrane) (Figures 2A–2C) (Binns et al., 2009).

Ten cpSSRs obtained from *Aquilaria* species were tested for transferability using in silico PCR and gel simulator methods (Table 4). This study determined that the SSRs had significant transferability properties between *Aquilaria* species and this discovery is verified to a recent research (Pern et al., 2020). The transferability of microsatellite markers is determined by the preservation of primer binding sites (Fagundes et al., 2016). The relatively high success rate of crosstransferability markers tested in this study indicated that *Aquilaria* species have comparable genetic relationships (Figure 4). Due to the high accuracy and rigor with which microsatellite data can be explored to generate DNA profiles of individuals for comparison to a reference sample database, it may benefit in identifying forest timber resources, particularly endangered plant species (Dormontt et al., 2015; Hung et al., 2017).

5. Conclusion

In conclusion, this study shows the number and distribution of cpSSRs identified in the complete cp genomes of *Aquilaria* species acquired from GenBank

database. Since there is a dearth of information of the Aquilarieae tribe from genetic studies, findings from this study may be applied in breeding programs as well as serve as a valuable reference for future studies pertinent to population and conservation genetics.

Supporting information

Supplementary 1: The most frequent motif types for mono-, di-, tri- and tetra- repeat types in the cp genome of eight *Aquilaria* species.

Acknowledgment

The authors would like to acknowledge University of Putra Malaysia for awarding Muhammad Syahmi Hishamuddin with the Graduate Research Fellowship.

Contribution of authors

Muhammad Syahmi Hishamuddin designed the study, conducted the experiments, drafted, and revised the manuscript as well as analyzing the data with the guidance from Rozi Mohamed and Samsuddin Ahmad Syazwan. Samsuddin Ahmad Syazwan assisted in the screening of SSRs and figures formatting. Rozi Mohamed supervised and acquired funding for the project. All authors reviewed and approved the final manuscript.

Funding

This research was funded by the Research University Grant provided by University of Putra Malaysia (Grant no. UPM/800-3/3/1/9630100).

References

Akkaya MS, Bhagwat AA, Cregan PB (1992). Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132 (4): 1131–1139.

Andru S, Pan YB, Thongthawee S, Burner DM, Kimbeng CA (2011). Genetic analysis of the sugarcane (*Saccharum* spp.) cultivar 'LCP 85-384'. I. Linkage mapping using AFLP, SSR, and TRAP markers. Theoretical and Applied Genetics 123 (1): 77–93. doi: 10.1007%2Fs00122-011-1568-x

Barden A, Anak NA, Mulliken T, Song M (2000). Heart of the matter: agarwood use and trade and CITES implementation for *Aquilaria malaccensis*. Cambridge, UK: TRAFFIC International.

Binns D, Dimmer E, Huntley R, Barrell D, O'donovan C et al (2009). QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics 25 (22): 3045–3046. doi: 10.1093/bioinformatics/btp536

Blanchette RA, Jurgens JA, Beek HHV (2015). Growing Aquilaria and production of Agarwood in hill agro-ecosystems. In: Eckman K, Ralte L (editors). Integrated Land Use Management in the Eastern Himalayas, India: Akansha Publishing House Delhi, pp. 66–82.

Cavagnaro PF, Senalik DA, Yang L, Simon PW, Harkins TT et al. (2010). Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.). BMC Genomics 11 (1): 1–18. doi: 10.1186/1471-2164-11-569

Chen H, Wang L, Wang S, Liu C, Blair MW et al. (2015). Transcriptome sequencing of mung bean (*Vigna radiate* L.) genes and the identification of EST-SSR markers. PloS One 10 (4): e0120273. doi: 10.1371/journal.pone.0120273

Cheng J, Zhao Z, Li B, Qin C, Wu Z et al. (2016). A comprehensive characterization of simple sequence repeats in pepper genomes provides valuable resources for marker development in *Capsicum*. Scientific Reports 6 (1): 1–12. doi: 10.1038/srep18919

Chua LSL, Lee SL, Lau KH, Zakaria NF, Tnah LH et al. (2016). Conservation action plan for the threatened agarwood species *Aquilaria malaccensis* (Thymelaeaceae) in Peninsular Malaysia. Kepong, Malaysia: Forest Research Institute Malaysia.

CITES (2004). Convention on international trade in endangered species of Wild Fauna and Flora. CoP13 Prop. 49. Consideration of proposals for amendment of Appendices I and II-*Aquilaria* spp. and *Gyrinops* spp. Thirteenth meeting of the conference of the parties, 2–14 Oct 2004, Bangkok. International Environment House, Geneva, pp 1–9

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M et al. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21 (18): 3674–3676. doi: 10.1093/bioinformatics/bti610

Dormontt EE, Boner M, Braun B, Breulmann G, Degen B et al. (2015). Forensic timber identification: It's time to integrate disciplines to combat illegal logging. Biological Conservation 191: 790–798. doi: /10.1016/j.biocon.2015.06.038

Echt CS, DeVerno LL, Anzidei M, Vendramin GG (1998). Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. Molecular Ecology 7: 307–316.

Ellegren H (2004). Microsatellites: simple sequences with complex evolution. Nature Reviews Genetics 5 (6): 435–445. doi: 10.1038/nrg1348

Eurlings MC, van Beek HH, Gravendeel B (2010). Polymorphic microsatellites for forensic identification of agarwood (*Aquilaria crassna*). Forensic Science International 197 (1-3), 30–34. doi: 10.1016/j.forsciint.2009.12.017

Fagundes BS, da Silva LF, Giacomin RM, Secco D, Díaz-Cruz JA et al. (2016). Transferability of microsatellite markers among Myrtaceae species and their use to obtain population genetics data to help the conservation of the Brazilian Atlantic Forest. Tropical Conservation Science 9 (1): 408–422. doi: 10.1177/194008291600900122

Faridah-Hanum I, Mustapa MZ, Lepun P, Tuan Marina TI, Nazre M et al. (2009). Notes on the distribution and ecology of *Aquilaria* Lam. (Thymelaeaceae) in Malaysia. Malaysian Forester 72 (2): 247–259.

Hishamuddin MS, Lee SY, Ng WL, Ramlee SI, Lamasudin Du et al. (2020). Comparison of eight complete chloroplast genomes of the endangered *Aquilaria* tree species (Thymelaeaceae) and their phylogenetic relationships. Scientific Report 10 (1): 1–13. doi: 10.1038/s41598-020-70030-0

Hung KH, Lin CH, Ju LP (2017). Tracking the geographical origin of timber by DNA fingerprinting: a study of the endangered species *Cinnamomum kanehirae* in Taiwan. Holzforschung 71 (11): 853–862. doi: 10.1515/hf-2017-0026

IUCN (2021). The IUCN Red List of Threatened Species. Version 2021-2 [online]. Website https://www.iucnredlist.org/search?query=Aquilaria&searchType=species [Accessed 22 April 2021].

Kale SM, Pardeshi VC, Kadoo NY, Ghorpade PB, Jana MM et al. (2012). Development of genomic simple sequence repeat markers for linseed using next-generation sequencing technology. Molecular Breeding 30 (1): 597–606. doi: 10.1007/s11032-011-9648-9

Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011). Microsatellite markers: an overview of the recent progress in plants. Euphytica 177 (3): 309–334. doi: 10.1007/s10681-010-0286-9

Kato S, Imai A, Rie N, Mukai Y (2013). Population genetic structure in a threatened tree, Pyrus calleryana var. dimorphophylla revealed by chloroplast DNA and nuclear SSR locus polymorphisms. Conservation Genetics 14 (5): 983–996. doi: 10.1007%252Fs10592-013-0489-4

Liu G, Xie Y, Zhang D, Chen H (2018). Analysis of SSR loci and development of SSR primers in *Eucalyptus*. Journal of Forestry Research 29 (2): 273–282. doi: 10.1007/s11676-017-0434-3

Liu SR, Li WY, Long D, Hu CG, Zhang JZ (2013). Development and characterization of genomic and expressed SSRs in citrus by genome-wide analysis. PloS One 8 (10): e75149. doi: 10.1371/journal.pone.0075149

McCauley DE (1995). The use of chloroplast DNA polymorphism in studies of gene flow in plants. Trends in Ecology & Evolution 10(5): 198–202. doi: 10.1016/S0169-5347(00)89052-7

Moe KT, Chung JW, Cho YI, Moon JK, Ku JH et al. (2011). Sequence information on simple sequence repeats and single nucleotide polymorphisms through transcriptome analysis of mungbean. Journal of Integrative Plant Biology 53 (1): 63–73. doi: 10.1111/j.1744-7909.2010.01012.x

Mohamed R, Lee SY (2016). Keeping up appearances: Agarwood grades and quality. In: Mohamed R (editor) Agarwood. Tropical Forestry. Singapore: Springer, pp. 149–167. doi: 10.1007/978-981-10-0833-7_10

Morgante M, Olivieri AM (1993). PCR-amplified microsatellites as markers in plant genetics. The Plant Journal 3 (1): 175–182. doi: 10.1046/j.1365-313X.1993.t01-9-00999.x

Ozyigit II, Dogan I, Filiz E (2015). *In silico* analysis of simple sequence repeats (SSRs) in chloroplast genomes of '*Glycine*' species. Plant Omics 8 (1): 24-29. doi: 10.3316/INFORMIT.023705269856161

Perdereau AC, Kelleher CT, Douglas GC, Hodkinson TR (2014). High levels of gene flow and genetic diversity in Irish populations of *Salix caprea* L. inferred from chloroplast and nuclear SSR markers. *BMC Plant Biology 14*(1), 1-12.

Pern YC, Lee SY, Ng WL, Mohamed R (2020). Cross-amplification of microsatellite markers across agarwood-producing species of the Aquilarieae tribe (Thymelaeaceae). 3 Biotech 10 (3): 1–9. doi: 10.1007/s13205-020-2072-2

Phumichai C, Phumichai T, Wongkaew A (2015). Novel chloroplast microsatellite (cpSSR) markers for genetic diversity assessment of cultivated and wild *Hevea* rubber. Plant Molecular Biology Reporter 33 (5): 1486–1498. doi: 10.1007/s11105-014-0850-x

Powell W, Morgante M, Andre C, Hanafey M, Vogel J et al. (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Molecular Breeding 2 (3): 225–238. doi: 10.1007/BF00564200

Provan J, Powell W, Hollingsworth PM (2001). Chloroplast microsatellites: new tools for studies in plant ecology and evolution. Trends in Ecology & Evolution 16 (3): 142–147. doi: 10.1016/S0169-5347(00)02097-8

Qian J, Song J, Gao H, Zhu Y, Xu J et al.(2013). The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. PloS One 8 (2): e57607. doi: 10.1371/journal.pone.0057607

Rajendrakumar P, Biswal AK, Balachandran SM, Srinivasarao K, Sundaram RM (2007). Simple sequence repeats in organellar genomes of rice: frequency and distribution in genic and intergenic regions. Bioinformatics 23 (1): 1–4. doi: 10.1093/bioinformatics/btl547

Shanker A (2013). Identification of microsatellites in chloroplast genome of *Anthoceros formosae*. Archive For Bryology 191: 1–6.

Shanker A (2014). Computationally mined microsatellites in chloroplast genome of *Pellia endiviifolia*. Archive For Bryology 199: 1–5.

Shukla N, Kuntal H, Shanker A, Sharma SN (2018). Mining and analysis of simple sequence repeats in the chloroplast genomes of genus *Vigna*. Biotechnology Research and Innovation 2 (1): 9–18. doi: 10.1016/j.biori.2018.08.001

Singh P, Sharma H, Nag A, Bhau BS, Sharma RK (2015). Development and characterization of polymorphic microsatellites markers in endangered *Aquilaria malaccensis*. Conservation Genetics Resources 7 (1): 61–63. doi: 10.1007/s12686-014-0287-4

Smith SE, Al-Doss A, Warburton M (1991). Morphological and agronomic variation in North African and Arabian alfalfas. Crop Science 31 (5): 1159–1163. doi: 10.2135/cropsci1991.0011183X003100050016x

Sonah H, Deshmukh RK, Sharma A, Singh VP, Gupta DK et al. (2011). Genome-wide distribution and organization of microsatellites in plants: an insight into marker development in *Brachypodium*. Plos One 6 (6): e21298. doi: 10.1371/journal.pone.0021298

Stothard P (2000). The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28 (6): 1102–1104. doi: 10.2144/00286ir01

Tambarussi EV, Melotto-Passarin DM, Gonzalez SG, Brigati JB, de Jesus FA et al. (2009). *In silico* analysis of Simple Sequence Repeats from chloroplast genomes of Solanaceae species. Crop Breeding and Applied Biotechnology 9 (4): 344–352. doi: 10.12702/1984-7033.v09n04a09

Tautz D (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research 17 (16): 6463–6471. doi: 10.1093/nar/17.16.6463

The Plant List (2013) onward (continuously updated). A working list of all plant species. Version 1.1 [online] Website http://www.theplantlist.org/tpl1.1/search?q=Aquilaria [accessed 28 June 2021]

Tnah LH, Lee CT, Lee SL, Ng KKS, Ng CH et al. (2012). Isolation and characterization of microsatellite markers for an important tropical tree, *Aquilaria malaccensis* (Thymelaeaceae). American Journal of Botany 99 (11): e431–e433. doi: 10.3732/ajb.1200165

Ueno S, Moriguchi Y, Uchiyama K, Ujino-Ihara T, Futamura N et al. (2012). A second generation framework for the analysis of microsatellites in expressed sequence tags and the development of EST-SSR markers for a conifer, *Cryptomeria japonica*. BMC Genomics 13 (1): 1–16. doi: 10.1186/1471-2164-13-136

Varshney RK, Thiel T, Stein N, Langridge P, Graner A (2002). *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. Cellular & Molecular Biology Letters 7 (2A): 537–546. doi:

Vieira MLC, Santini L, Diniz AL, Munhoz CDF (2016). Microsatellite markers: what they mean and why they are so useful. Genetics and Molecular Biology 39 (3): 312–328. doi: 10.1590/1678-4685-GMB-2016-0027

Zhang YT, Wang ZF, Cao HL, Li XY, Wu LF et al. (2010). Isolation and characterization of polymorphic microsatellite loci in *Aquilaria sinensis* (Lour.) Gilg. Conservation Genetics Resources 2 (1): 5–6. doi: 10.1007/s12686-009-9110-z

Zou C, Lu C, Zhang Y, Song G (2012). Distribution and characterization of simple sequence repeats in *Gossypium raimondii* genome. Bioinformation 8 (17): 801–806. doi: 10.6026/97320630008801