Journal of Tropical Forest Science 30(3): 279–291 (2018)
https://doi.org/10.26525/jtfs2018.30.3.279291

ISOLATION AND CHARACTERISATION OF SSR MARKERS IN TONGKAT ALI (*EURYCOMA LONGIFOLIA*) USING NEXT-GENERATION SEQUENCING APPROACH

Lee CT*, Norlia B, Tnah LH, Lee SL, Ng CH, Ng KKS, Nor-Hasnida H, Nurul-Farhanah Z, Suryani CS & Nur-Nabilah A

Forest Research Institute Malaysia, 52109 Kepong, Selangor, Malaysia

**leechait@frim.gov.my*

Submitted January 2018; accepted June 2018

Eurycoma longifolia is an important medicinal plant in Malaysia. Due to the high demand for *E. longifolia* roots, the natural populations are under tremendous harvesting pressure. Hence, attention should be given to conserve the genetic resources as well as to produce elite planting materials for cultivation. As simple sequence repeats (SSRs) are the most versatile markers to be used in genetic diversity assessment and molecular breeding, the study was aimed to develop SSR markers based on transcriptome sequences of *E. longifolia* root. Root transcriptome of a 10-year-old *E. longifolia* was sequenced via Illumina paired-end sequencing. From the 39,100,208 trimmed sequencing reads, a total of 51,551 contigs were generated with an average length of 672 bp. The assembled contigs were annotated and functionally classified. Based on these transcriptome sequences, we have successfully developed and validated 46 polymorphic SSR markers applicable across 102 *E. longifolia* samples from five natural forest reserves in Peninsular Malaysia. A total of 232 alleles were observed among the five populations, with an average of 5.0 alelles per locus. The SSR markers developed are valuable genetic tools for a wide spectrum of genetic analyses for *E. longifolia*, be it for conservation or breeding-related applications.

Keywords: Tongkat ali, Simaroubaceae, tropical medicinal plant, next-generation sequencing, simple sequence repeats

INTRODUCTION

Eurycoma longifolia, locally known as tongkat ali, is one of the most popular medicinal plants in Malaysia. It belongs to the family Simaroubaceae and is indigenous to South-East Asian countries such as Malaysia, Indonesia and Vietnam. The plant extract, particularly from the root, has been used in traditional medicines for treatment of various illnesses (Burkill 1966). It was reported to have anti-malarial (Ang et al. 1995), anti-ulcer (Tada et al. 1991) and cytotoxic (Morita et al. 1990) activities. However, its root is more well-known for aphrodisiac properties (Low et al. 2013) and various chemical compounds have been isolated and characterised (Bhat & Karim 2010). It was also claimed to have the potential in prevention and treatment of male osteoporosis (Effendy et al. 2012). A recent review on *E. longifolia* (Rehman et al. 2016) detailed its myriad of traditional uses, chemical constituents as well as evidence-based pharmacology and toxicology.

With the advent of next-generation sequencing (NGS) technologies, research on *de novo* transcriptome assemblies has been increasing exponentially (Han et al. 2015). Transcriptome sequencing has become the tool of choice for gene discovery, transcript quantification and marker discovery, especially for non-model plants (as reviewed by Bräutigam & Gowik 2010). Discovery of molecular markers in numerous organisms has also been accelerated tremendously by NGS, in particular simple sequence repeats (SSRs) (Zhou et al. 2016, Lu et al. 2017, Du et al. 2018). Conventional methods of developing SSR markers from libraries involve several steps and at each stage SSR loci may be lost, resulting in high costs and work load for a modest number of SSRs (Squirrell et al. 2003). In contrast, due to the massive amount of sequence data and much lower cost compared with traditional Sanger methods (Metzker 2010), the NGS approach is much faster and

more cost-effective (Zalapa et al. 2012, Hodel et al. 2016).

In this study, we report the isolation and characterisation of SSR markers using NGS approach. The purpose of marker development is to facilitate future conservation and breeding programmes in view of the diminishing genetic resources due to overexploitation and the rising needs for conservation and breeding research.

MATERIALS AND METHODS

Root harvesting and RNA extraction

The root of a 10-year-old *E. longifolia* was collected from Bukit Hari, Forest Research Institute Malaysia. The harvested root was immediately immersed in liquid nitrogen and subsequently kept at -80 °C before RNA extraction using RNeasy Plant Mini Kit. The RNA quality was determined on an Agilent 2100 Bioanalyzer and an RNA Integrity Number (RIN) was attributed to the samples to standardize the quality of the extracts.

Illumina sequencing, sequence data processing, assembly and functional annotation

The root transcriptome sequencing was outsourced to Beijing Genomics Institute. The RNA sample with absorbance 260/280 nm ratio of ~2.0 and RIN number more than 8.0 was used to generate a cDNA library of 100 bp-paired-end sequencing reads. The cDNA libraries were generated using Sera-mag Magnetic Oligo (dT) Beads (Illumina) whereby beads with Oligo(dT) were used to isolate poly(A) mRNA. Short fragments of mRNA produced by adding fragmentation buffer were used as templates, with random hexamer-primer in the synthesis of the first-strand cDNA. The second-strand cDNA was synthesised using buffer, dNTPs, RNaseH and DNA polymerase I. Short fragments were purified using PCR extraction kit and resolved with EB buffer for end reparation with the addition of poly(A), prior to ligation with sequencing adapters. Subsequently, after agarose gel electrophoresis, suitable fragments were selected for the PCR amplification. Finally, the library was sequenced using Illumina HiSeq 2000 platform.

The raw sequence data quality was checked with FastQC v0.11.3 (Andrews 2010), and adapter and low quality bases were trimmed using Trimmomatic v0.32 (Bolger et al. 2014). Trimmed sequence raw data was deposited in the NCBI Sequence Read Archive under the accession number SRP064289. Trinity (version trinity/2013-02-25) (http://trinityrnaseq.sourceforge.net) was used to assemble the sequence reads into contigs. All the contigs were then blasted against NCBI non-redundant and SwissProt databases (E-value < 10^{-6}), functionally annotated by gene ontology (GO) analysis and mapped against KEGG pathways using Blast2GO software (version 2.4.4; http://www.blast2go.org), a universal data mining tool suited for transcriptome analysis in non-model species (Conesa et al. 2005).

Isolation and characterisation of SSRs

SSR-containing sequences were identified from the contigs using MIcroSAtellite identification tool (Thiel et al. 2003), with minimum criteria of five repeats for mono-, di-, tri-, tetra-, penta-, hexa- and hepta- nucleotide motifs; and the maximal number of 50 bases interrupting two SSRs. OLIGO 6.67 software was used to design primer pairs from randomly selected contigs containing SSR motifs with adequate flanking regions. However, contigs containing dinucleotide repeats were excluded from primer design because all the 18 SSR markers previously developed for *E. longifolia* via conventional enrichment approach comprised dinucleotide motifs (Tnah et al. 2011). Besides, dinucleotide microsatellites are more prone to have stutter peaks (Ellegren 2004). A total of 72 primer pairs were designed.

Initial screening of the primer pairs was carried out on four *E. longifolia* individuals following the PCR protocol reported in Tnah et al (2011), except that the annealing temperature was set at 50 °C. A total of 62 primer pairs yielded specific amplification products of the expected size and were selected for labelling with either 6-FAM or HEX at the 5'-end of the forward primers.

In order to characterise and validate the SSR loci, leaf samples from a total of 102 *E. longifolia* trees from five natural forest reserves (FR) located in Peninsular Malaysia were collected (32 samples from Semangkok FR, Selangor,

Pondok Tanjong FR (13) and Bukit Larut FR (8), Perak, Gunung Raya FR (26) and Pulau Singa Besar FR (23), Langkawi). Approximately 5 g of leaf sample was used for DNA extraction using the modified CTAB method (Murray & Thompson 1980) and subsequently purified using the high pure PCR template preparation kit. For initial characterisation, the 62 primer pairs were applied in PCR reactions with 32 samples from Semangkok FR under the same reaction conditions as described above. The PCR products were then subjected to fragment analysis using an ABI 3130xl genetic analyser. Allele sizes were assigned according to the internal size standard and individuals were genotyped using Genemarker v2.6.4.

Genotypic data were analysed using Microsatellite toolkit (Park 2001). The observed heterozygosity (H_O), expected heterozygosity (H_E), and polymorphic information content (PIC) were calculated based on the formula as described in Chung et al. (2013). The exact tests for Hardy–Weinberg expectation and genotypic disequilibrium were conducted using Genetic Data Analysis version 1.0 (Lewis & Zaykin 2002). Microchecker version 2.2.1 (van Oosterhout et al. 2004) was used to detect presence of null alleles.

A total of 46 of the 62 SSR primer pairs which yielded specific and consistent genotypes were further validated using all the 102 samples collected, including the initial 32 samples from Semangkok FR as multiplex PCRs were conducted. There were in total seven multiplex sets (Appendix 1). The multiplex PCRs were performed in 8 µL reaction mixture, with about 10 ng of template DNA, 0.032 µM of each forward and reverse primer, and 1× master mix of Type-it® Multiplex PCR Kit. The PCR programme included 5 min at 95 °C, 35 cycles of 95 °C for 30 s, 57 °C for 90 s, and 72 °C for 30 s, followed by a 30 min final extension at 60 °C. The procedures for genotyping and subsequent data analyses were the same as with the products from single PCR.

RESULTS AND DISCUSSION

Sequencing and *de novo* assembly

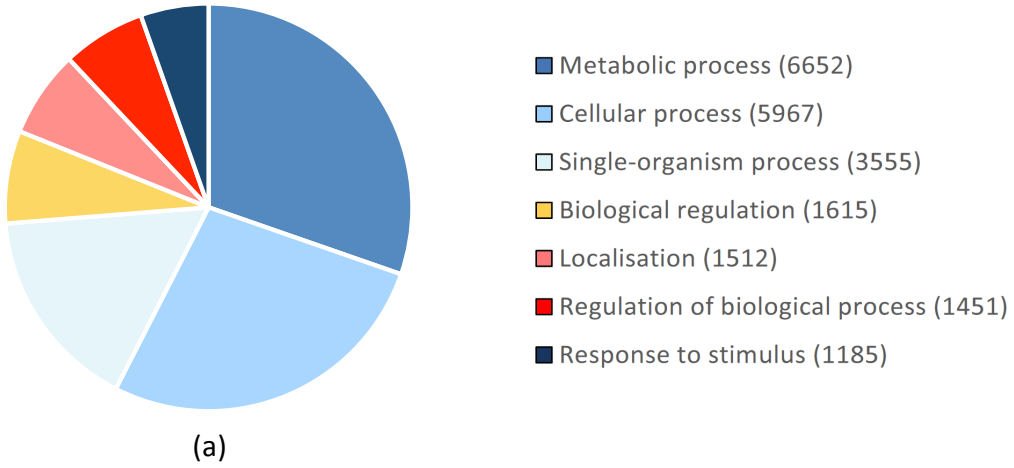
The root transcriptome of a 10-year-old *E. longifolia* tree was obtained by Illumina sequencing. A total of 41,357,870 sequencing reads were generated. After filtering adapter sequences and discarding low quality reads, a total of 39,100,208 paired-end reads with a mean length of 100 bp were obtained. *De novo* assembly conducted using Trinity program (version trinity/2013-02-25) (http://trinityrnaseq.sourceforge.net) generated 69,754 contigs with an average length of 724 bp and N50 of 1006 bp. Further analysis was carried out to reduce redundancy which finally resulted in only 51,551 contigs representing unique transcripts with an averange length of 672 bp and N50 of 934 bp were served as the mRNA transcript of *E. longifolia* root. The length of assembled contigs ranged from 201 to 8064 bp.

Functional annotation and gene ontology classification

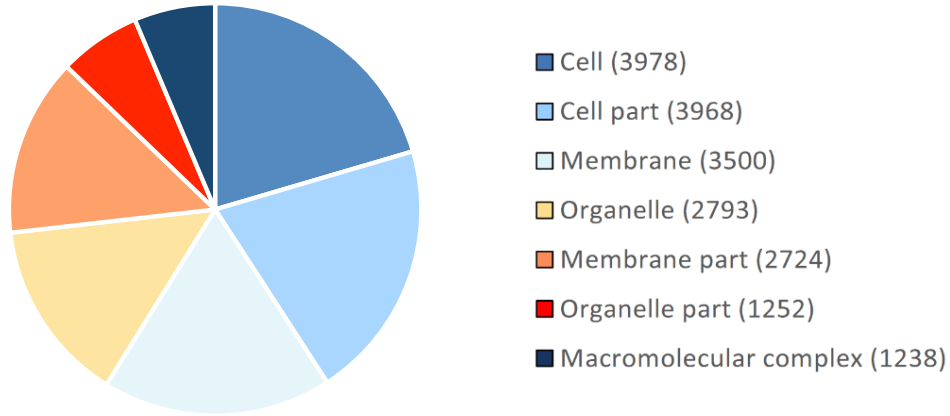
BLASTX sequence similarity analysis of the assembled contigs against the GenBank non-redundant protein database resulted in the identification of 36,882 (71.54%) sequences with at least one significant alignment to an existing gene; 14,669 sequences (28.45%) did not have any hit in the GenBank. The BLASTX-hit transcripts exhibited the highest homology with those from *Citrus sinensis* (44.57%), followed by *C. clementina* (20.88%), *C. unshiu* (16.76%) and *Theobroma cacao* (1.92%).

Blast2GO analysis showed that in *E. longifolia* root, 24,259 (47.05%) contigs were assigned to at least one GO term. The contigs were assigned to three main categories: biological process (8466, 34.8%), cellular component (6549, 26.9%) and molecular function (9244, 38.1%) (Figure 1). Within the biological process category, the majority of the GO terms was grouped into either metabolic (6652) or cellular (5967) processes. For the cellular component, cell (3978) and cell part (3968) were the majority, while for the molecular function, the sequence were related to binding (7759) and catalytic (6334) activities. Similiar pattern of GO terms was also found in other medicinal plants as reported in *Silvia miltiorrhiza* (Wenping et al. 2011), *Camellia chekiangoleosa* (Wang et al. 2014) and *Youngia japonica* (Peng et al. 2014).

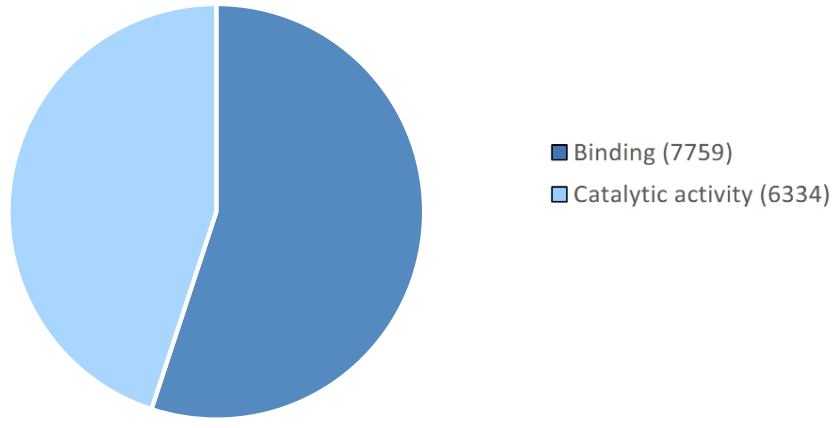
From the metabolic pathway analysis, a total of 1039 contigs were mapped to 139 known biological pathways, with some unigenes assigned to more than one KEGG pathways. The highest number of contigs (121) were



(a)



(b)



(c)

Figure 1 Functional classification of the assembled contigs (a) biological process, (b) cellular component and (c) molecular function; the number of transcripts in each class is shown in brackets

detected in biosynthesis of antibiotics pathways, followed by purine metabolism (34), cystein and methionine metabolism (29), starch and sucrose metabolism (29) and amino sugar and nucleotide sugar metabolism (24). KEGG data analysis, together with GO analysis, provides important information to annotate potential genes and their function at a transcription level for future research into gene functions.

Isolation and characterisation of SSR markers

A total of 46 SSR primer pairs yielded consistent and scorable genotypes based on the fragment analysis results. Information on the GenBank accession numbers, repeat motifs, primer sequences and allele size range of these SSR markers are given in Table 1. Of these SSR loci, there were 28 with trinucleotide motifs, 13 tetra- and 5 penta-. Based on the initial characterisation using 32 individuals from Semangkok FR, all the 46 SSR loci exhibited polymorphism, with number of alleles per locus ranging from two to eight (Table 2). Locus *EloT*040 showed the lowest observed and expected heterozygosities (both 0.031), while *EloT*050 showed the highest, 0.844 and 0.811 respectively. From the 32 samples, there were in total 196 alleles observed. Significant deviation from Hardy–Weinberg equilibrium was only detected in one locus (*EloT*028; $p < 0.001$), after Bonferroni correction. This particular locus also showed evidence of null alleles (across all five populations), besides *EloT*024 (only in Semangkok and Gunung Raya FRs).

Multiplex PCR and subsequent SSR analysis demonstrated that all the markers are applicable across the 102 samples from five different locations. The majority of the SSR loci (41 out of 46) were polymorphic within each of the five populations surveyed, while the remaining five loci (*EloT*018, *EloT*021, *EloT*038, *EloT*040 and *EloT*044) were monomorphic for one or more populations (Table 2). A total of 232 alleles were observed among the five populations, with an average of 5.0 per locus. Despite limited number of samples from Pondok Tanjong (n = 13) and Bukit Larut (n = 8), Hardy Weinberg disequilibrium deviation was only detected within population in loci *EloT*024 (Gunung Raya) and *EloT*028 (all populations except Pondok Tanjong), after Bonferroni correction. Excluding locus *EloT*028, linkage disequilibrium was detected in *EloT*008 × *EloT*024, *EloT*047 × *EloT*050, *EloT*050 × *EloT*056, *EloT*050 × *EloT*058, *EloT*050 × *EloT*061 in Gunung Raya; *EloT*050 × *EloT*051 in Pulau Singa Besar and *EloT*024 × *EloT*025 in Semangkok (at $p < 5.05 \times 10^{-5}$ with Bonferroni correction), among the 990 possible combinations. Therefore, in general, most of the markers were robust.

SSRs derived from transcriptome sequences or expressed sequence-tags (ESTs) tend to be less polymorphic compared with genomic SSRs because they are located within more conserved transcribed regions (Varshney et al. 2005, Tabbasam et al. 2014). Likewise, the transcriptomic *E. longifolia* SSRs developed in this study were also relatively less polymorphic in comparison with the genomic SSRs previously developed via enrichment approach (Tnah et al. 2011), albeit the repeat motifs for the latter were all dinucleotide, which might also contribute to the higher level of polymorphism to some extent (Zhao et al. 2013). The mean number of alleles for the genomic SSRs was 8.7 (Tnah et al. 2011, sample size, n = 28), about twice higher than that of the transcriptome-derived SSRs [4.3 when n = 32; 5.0 when n = 102]. However, the NGS approach of SSR marker discovery is much more efficient and cost effective. Lower polymorphism of these transcriptomic SSR markers can be compensated by increasing the number of loci in a genetic assay, which is amenable with multiplex PCR. Another feature of the transcriptomic SSRs is the advantage of being able to reveal functional diversity of the samples investigated (Varshney et al. 2005).

The main source of *E. longifolia* supply is from the natural forests. Therefore, the destructive harvesting of *E. longifolia* roots will eventually lead to the decline of its natural populations, moreover, there is high demand for *E. longifolia* products in the health-food market (Bhat & Karim 2010). Hence, in order to safeguard the existing genetic resources and prevent irreversible genetic erosion, effective conservation measures should be taken apart from promoting plantation (Ismail et al. 2004).

One way to assist the forest managers in making informed decision is through conservation genetic studies. However, the baseline genetic information of *E. longifolia* is still lacking. Asiah et al. (2003) assessed its genetic diversity using single nucleotide polymorphisms with limited

Table 1 Details of the 46 transcriptomic SSRs of *Eurycoma longifolia* developed from the root transcriptome sequences, their GenBank accession numbers, repeat motifs, primer sequences and allele size range based on 32 individuals from Semangkok Forest Reserve

Locus	GenBank accession no.	Repeat motif	Primer sequence (5'–3')	Allele size range (bp)
*EloT*002	KR084329	$(CAT)_7$	F: CTAGGATTAATGGCATTCGGAAGC	391–400
			R: ATGCGCCAATTTAGGTCAA	
*EloT*003	KR084330	$(AGA)_7$	F: GGCGTCTGCTATGTTGGAGAGTTG	116–119
			R: TCATCCGTCGCCATCTTTCAT	
*EloT*005	KR084331	$(AAC)_8$	F: CTAATAATGGTGCCAATACGAATA	226–229
			R: CTGTGACTGCGATTGTGATTTCT	
*EloT*006	KR084332	$(AAT)_{12}$	F: TATCATGGCGGAAGCAGTAAC	136–157
			R: CATCAGCAGTAGCAGTAGCCGACA	
*EloT*007	KR084333	$(AGC)_7$	F: ACACGCCAGGTACTTGCAGAAGA	234–246
			R: CGGCTTTATGCTTTGATGGTACAG	
*EloT*008	KR084334	$(TGC)_7$	F: TGGTTGGACAAGTGGCATCAGAGC	225–237
			R: CCATGCCCAGCAGTCACATCAACT	
*EloT*010	KR084335	$(GAAA)_7$	F: ACAGATGCGTAGGCTACTCT	121–133
			R: AAAAATCCACAAGAACCAAACTAT	
*EloT*011	KR084336	$(TGAT)_7$	F: GGCATCAAGAAGAGGAGACGAA	301–317
			R: CAGAACGCAGGAAGGAGATATGGT	
*EloT*012	KR084337	$(AAAT)_7$	F: CGCTTTCTGTTCTGTTTCGTA	140–144
			R: AACTTAACATCTCTGGTCGATTCT	
*EloT*013	KR084338	$(AAAG)_6$	F: CACATCTCATCTTTTATTGGGTTG	270–283
			R: ACTCTGGTTGCGTCGTTCATC	
*EloT*014	KR084339	$(CTTT)_8$	F: GGCAGTCCTCAAGAACCTATAAT	277–289
			R: CACGCATACACATACACTCATAGA	
*EloT*015	KR084340	$(AAAG)_6$	F: ATGGTCCGGCAGTGGTAGTGTC	384–397
			R: ATAAATCAGCAATCACGCGATATT	
*EloT*016	KR084341	$(TTTA)_7$	F: GACACAGGCCTTTGAGCATCATAA	222–230
			R: AGACTGAGCGCAAGACGATAA	
*EloT*018	KR084342	$(TCTGA)_5$	F: ACAGTACATGTGAGGGTCTGATCT	338–348
			R: CCCAGTTTGAATGACGTTGAA	
*EloT*019	KR084343	$(GGGAA)_6$	F: AAGTGGCATCTTGATATGATATAA	150–171
			R: GGCCACCAGTAATTCTATA	
*EloT*020	KR084344	$(TCTTC)_6$	F: CGAAGAAGACCAAAGAGTATCTG	206–222
			R: ATATCATCACCATTGACCGTACTA	
*EloT*021	KR084345	$(ATTGT)_5$	F: TGATCGTGAGTGCTTTCACAGTAG	297–323
			R: GAATCAAGCAATATCTGATGTGAG	
*EloT*024	KR084346	$(TTCTC)_6$	F: AGCGACAGACTTACAGTTGAGCTT	360–365
			R: TTATTCTTCTCGTCGTCGTTT	
*EloT*025	KR084347	$(AAAG)_7$	F: CGCTACGGTCTCCTCAAGGTAAGT	170–186
			R: TTTGGGTTTTGCAGAGATTAAAGT	
*EloT*028	KR084348	$(TAAA)_6$	F: GGCAACAAAAGGATATATAATCA	116–128
			R: ATAAGGCATGACTATCTTCACAAC	
*EloT*031	KR084349	$(CATA)_6$	F: CCTATTGTAAGCCCTATGTCAC	298–330
			R: ATTTGGGAGTCCTTACGATAATGT	
*EloT*032	KR084350	$(AAAC)_6$	F: AACAGGACCAGAAGCATAAACA	260–268
			R: TGGCAGAGAGTGAAGTACTACATC	

continued

Table 1 (continued)

Locus	GenBank accession no.	Repeat motif	Primer sequence (5'–3')	Allele size range (bp)
*EloT*034	KR084351	$(TCAA)_7$	F: GCTCCCTGCATTCCGATCAATTCC	202–254
			R: TTTCGCCTTCGCTTCCTCTACAAT	
*EloT*036	KR084352	$(TAGA)_6$	F: TTGAAGAAACTTTGCAGATGCG	172–184
			R: TTCAGCTGCTCTATTTGCCTATCC	
*EloT*037	KR084353	$(ATC)_8$	F: TCACCGTCCGATCCTCATAACTAT	105–120
			R: CCTATCGTCCGATGAAGATGATT	
*EloT*038	KR084354	$(CAG)_7$	F: TCGGCTTCAGTAACAGTGGAATTA	104–113
			R: TGTGATGGGCTTCCTATTCAATG	
*EloT*039	KR084355	$(CAA)_{13}$	F: CAGCACTTATACTTGGAAACTC	128–149
			R: CTCTGTAGTCTAATTGGTGAACAT	
*EloT*040	KR084356	$(TAA)_7$	F: CGCGGACAGTGGATGGATTG	190–193
			R: GCCGTTTCACCGTCCATCATTGTT	
*EloT*041	KR084357	$(GAA)_9$	F: CGCCAAAAGACGATGTCTCAG	154–174
			R: AATTTAACCGCTCCTGTTCTTCAC	
*EloT*042	KR084358	$(ATG)_{10}$	F: CAGATCGGAGCTCTTGGAGAAC	180–192
			R: GATCCCAACGGTCACATTCTTTAC	
*EloT*044	KR084359	$(GAA)_8$	F: TCTTTCTGGGTCCCGTTTCTG	240–249
			R: ATCAATGGCCACTCTCTCAACTCC	
*EloT*045	KR084360	$(GAA)_9$	F: TAGGTTTCTGTTTTTCGCAAATGA	233–250
			R: CGCCAGTCACCTCTGTAGCTGTG	
*EloT*046	KR084361	$(CAT)_8$	F: ACCCTGAAACATTGAAAGTGACTA	269–284
			R: GCTTCTGAAACTCTTTGCTG	
*EloT*047	KR084362	$(AAC)_8$	F: TATGGCACAGGCAATGAGTTTTAC	244–253
			R: CTTTAAGCAGGTCCTACTCCAAG	
*EloT*048	KR084363	$(TAA)_7$	F: GCCCTAATTCTTCTGTGATGG	269–275
			R: AGTCAGTGAACTCGCTGTTGTGAT	
*EloT*049	KR084364	$(CTG)_7$	F: GGCGATTCACCTTCATCTGCTG	291–309
			R: AGAATCCTCCTGCCGAGTGTAAAC	
*EloT*050	KR084365	$(GAG)_{11}$	F: TTTGCTCTTGCCCAGATGCCTAC	293–320
			R: GAGAAAGACGCACCGCTCTATCAC	
*EloT*051	KR084366	$(CAA)_9$	F: CCCAACAACTCGTCCTCGGATAAT	254–286
			R: TTTGCTTTCTCCTATTGCGTTCA	
*EloT*056	KR084367	$(AGA)_7$	F: CAAAGAAAGCACCAAAAGTGAAA	115–130
			R: CATCAATCCCAACAACAACTATCA	
*EloT*058	KR084368	$(GGT)_{10}$	F: GATGGGGCTGCTCTTGTGATT	89–128
			R: TTTCTCTATCTCTCCGCTCACCAG	
*EloT*061	KR084369	$(TCT)_9$	F: GCACGGGAGACGAGGAGAG	211–239
			R: CGCAGGTGGATAAGAGAGGGAAGA	
*EloT*063	KR084370	$(TGA)_{12}$	F: GATTTGCCTTTGCTTTCTTGA	96–125
			R: CTTCATCCCATTTCTTCTTCCTCT	
*EloT*065	KR084371	$(AAC)_8$	F: TGGTTTGCCGAGTATTCAGAAAGA	240–255
			R: GCCGATGCTCAAGGTATTTGC	
*EloT*066	KR084372	$(CAG)_8$	F: CCGCATTTACCTCCACAACAAGCA	312–321
			R: CGGCTCATTGTAGGACCCACAGCA	
*EloT*068	KR084373	$(TTC)_{10}$	F: TAAACGAGGAAGCAGCCCACA	335–347
			R: CCAGTTAGTGAAGCGGTCAGGAAC	
*EloT*069	KR084374	$(AGA)_{12}$	F: GCCCTCGTCTCATTCACAAG	284–301
			R: AGTAAGCCGAGTGTCTATTTGTTT	

Table 2 Genetic diversity measures estimated for the 46 SSR loci in *Eurycoma longifolia* across five forest reserves in Peninsular Malaysia

Location	Semangkok FR				Pondok Tanjong FR				Bukit Larut FR				Gunung Raya FR				Pulau Singa Besar FR			
Coordinate	3° 37' N, 101° 44' E				5° 5' N, 100° 47' E				4° 52' N, 100° 46' E				6° 22' N, 99° 48' E				6° 13' N, 99° 44' E			
Locus	*A*	H_O	H_E	PIC	*A*	H_O	H_E	PIC	*A*	H_O	H_E	PIC	*A*	H_O	H_E	PIC	*A*	H_O	H_E	PIC
*EloT*002	4	0.281	0.302	0.274	2	0.077	0.077	0.071	2	0.625	0.525	0.371	2	0.346	0.340	0.278	2	0.435	0.487	0.363
*EloT*003	2	0.219	0.198	0.176	2	0.385	0.471	0.350	2	0.250	0.233	0.195	2	0.385	0.462	0.350	2	0.478	0.476	0.357
*EloT*005	2	0.531	0.468	0.354	2	0.500	0.391	0.305	2	0.250	0.533	0.375	2	0.577	0.449	0.343	2	0.522	0.464	0.351
*EloT*006	6	0.625	0.584	0.548	5	0.417	0.377	0.346	3	0.500	0.592	0.456	6	0.577	0.599	0.502	2	0.478	0.449	0.343
*EloT*007	3	0.531	0.493	0.426	3	0.750	0.670	0.566	3	0.500	0.567	0.468	3	0.385	0.358	0.318	3	0.348	0.305	0.273
*EloT*008	3	0.563	0.520	0.396	3	0.500	0.554	0.428	3	0.625	0.492	0.398	2	0.423	0.503	0.372	2	0.565	0.511	0.375
*EloT*010	2	0.250	0.222	0.195	2	0.167	0.159	0.141	2	0.500	0.400	0.305	2	0.346	0.340	0.278	2	0.304	0.322	0.265
*EloT*011	4	0.594	0.544	0.456	3	0.833	0.540	0.420	2	0.875	0.525	0.371	4	0.500	0.439	0.366	3	0.739	0.585	0.480
*EloT*012	2	0.156	0.146	0.134	2	0.636	0.455	0.340	3	0.750	0.625	0.520	2	0.577	0.491	0.366	2	0.478	0.476	0.357
*EloT*013	4	0.625	0.567	0.460	3	0.583	0.540	0.444	3	0.625	0.492	0.398	2	0.385	0.317	0.262	4	0.478	0.591	0.511
*EloT*014	4	0.313	0.307	0.286	3	0.417	0.489	0.391	3	0.375	0.433	0.371	4	0.769	0.673	0.598	4	0.739	0.703	0.625
*EloT*015	6	0.281	0.519	0.476	2	0.538	0.471	0.350	3	0.500	0.508	0.427	3	0.423	0.405	0.352	3	0.348	0.474	0.390
*EloT*016	3	0.344	0.392	0.354	2	0.417	0.344	0.275	2	0.750	0.500	0.359	2	0.500	0.449	0.343	2	0.348	0.394	0.311
*EloT*018	2	0.125	0.119	0.110	2	0.154	0.148	0.132	1	0.000	0.000	0.000	2	0.308	0.265	0.226	2	0.435	0.394	0.311
*EloT*019	5	0.281	0.332	0.309	4	0.538	0.443	0.383	3	0.500	0.425	0.354	4	0.538	0.498	0.458	4	0.609	0.573	0.507
*EloT*020	5	0.219	0.206	0.197	2	0.077	0.077	0.071	2	0.125	0.125	0.110	3	0.308	0.278	0.255	3	0.261	0.237	0.213
*EloT*021	3	0.161	0.154	0.146	1	0.000	0.000	0.000	1	0.000	0.000	0.000	3	0.115	0.112	0.107	2	0.043	0.043	0.042
*EloT*024	3	0.094	0.466	0.411	3	0.154	0.283	0.255	3	0.375	0.342	0.294	3	0.154	0.382	0.335	3	0.217	0.305	0.273
*EloT*025	4	0.406	0.513	0.465	4	1.000	0.714	0.620	3	0.750	0.708	0.590	4	0.654	0.672	0.609	4	0.739	0.669	0.594
*EloT*028*	4	0.250	0.690	0.618	4	0.273	0.524	0.467	6	0.125	0.842	0.759	5	0.333	0.723	0.653	3	0.143	0.547	0.434
*EloT*031	7	0.625	0.646	0.569	4	0.615	0.495	0.433	3	0.500	0.592	0.456	5	0.808	0.730	0.668	6	0.870	0.808	0.758
*EloT*032	4	0.594	0.607	0.517	3	0.692	0.600	0.513	4	0.625	0.675	0.570	4	0.538	0.511	0.458	4	0.609	0.673	0.596
*EloT*034	5	0.688	0.639	0.577	2	0.615	0.492	0.361	2	0.250	0.233	0.195	2	0.462	0.434	0.335	2	0.696	0.487	0.363
*EloT*036	4	0.250	0.258	0.244	4	0.636	0.502	0.427	2	0.250	0.233	0.195	4	0.231	0.215	0.200	4	0.304	0.310	0.284

continued

Table 2 (continued)

Location	Semangkok FR				Pondok Tanjong FR				Bukit Larut FR				Gunung Raya FR				Pulau Singa Besar FR			
Coordinate	3° 37' N, 101° 44' E				5° 5' N, 100° 47' E				4° 52' N, 100° 46' E				6° 22' N, 99° 48' E				6° 13' N, 99° 44' E			
Locus	*A*	H_O	H_E	PIC	*A*	H_O	H_E	PIC	*A*	H_O	H_E	PIC	*A*	H_O	H_E	PIC	*A*	H_O	H_E	PIC
*EloT*037	6	0.688	0.596	0.532	4	0.308	0.403	0.363	4	0.750	0.742	0.636	4	0.615	0.531	0.475	4	0.522	0.509	0.454
*EloT*038	2	0.125	0.119	0.110	2	0.083	0.083	0.077	1	0.000	0.000	0.000	2	0.538	0.401	0.316	2	0.522	0.464	0.351
*EloT*039	6	0.375	0.377	0.348	4	0.308	0.403	0.363	4	0.375	0.350	0.313	4	0.654	0.717	0.646	4	0.826	0.730	0.662
*EloT*040	2	0.031	0.031	0.030	1	0.000	0.000	0.000	2	0.250	0.233	0.195	1	0.000	0.000	0.000	1	0.000	0.000	0.000
*EloT*041	5	0.406	0.452	0.405	3	0.231	0.335	0.290	2	0.125	0.125	0.110	4	0.423	0.428	0.369	2	0.261	0.294	0.246
*EloT*042	5	0.844	0.746	0.686	3	0.692	0.628	0.536	3	1.000	0.708	0.590	4	0.423	0.533	0.448	3	0.609	0.478	0.377
*EloT*044	4	0.188	0.179	0.171	3	0.231	0.218	0.198	2	0.375	0.325	0.258	1	0.000	0.000	0.000	1	0.000	0.000	0.000
*EloT*045	5	0.750	0.767	0.714	4	0.538	0.603	0.520	7	1.000	0.817	0.735	4	0.577	0.689	0.616	5	0.783	0.714	0.640
*EloT*046	4	0.531	0.445	0.391	3	0.750	0.620	0.523	3	0.375	0.567	0.468	3	0.538	0.510	0.408	3	0.565	0.594	0.505
*EloT*047	4	0.563	0.648	0.569	3	0.154	0.335	0.290	3	0.750	0.542	0.428	4	0.615	0.565	0.477	3	0.652	0.553	0.434
*EloT*048	3	0.313	0.319	0.281	2	0.500	0.507	0.368	3	0.625	0.508	0.427	2	0.500	0.382	0.305	2	0.391	0.372	0.298
*EloT*049	6	0.563	0.562	0.523	4	0.417	0.370	0.330	4	0.375	0.350	0.313	4	0.538	0.512	0.461	5	0.478	0.435	0.394
*EloT*050	8	0.844	0.811	0.770	4	0.500	0.525	0.456	3	0.625	0.608	0.496	6	0.385	0.698	0.631	6	0.478	0.716	0.665
*EloT*051	7	0.813	0.754	0.699	6	0.667	0.641	0.583	6	0.500	0.750	0.655	5	0.538	0.623	0.535	6	0.565	0.649	0.592
*EloT*056	4	0.281	0.282	0.262	4	0.385	0.345	0.310	3	0.250	0.242	0.215	4	0.423	0.447	0.409	4	0.348	0.344	0.315
*EloT*058	7	0.781	0.625	0.563	4	0.923	0.726	0.641	7	0.875	0.825	0.741	6	0.577	0.665	0.617	8	0.652	0.655	0.615
*EloT*061	5	0.656	0.699	0.626	3	0.385	0.342	0.303	3	0.500	0.575	0.447	3	0.500	0.520	0.398	5	0.826	0.698	0.624
*EloT*063	7	0.781	0.737	0.686	4	0.692	0.591	0.476	4	0.750	0.642	0.547	6	0.769	0.667	0.594	4	0.391	0.375	0.341
*EloT*065	4	0.219	0.232	0.218	2	0.462	0.369	0.292	2	0.375	0.325	0.258	2	0.423	0.340	0.278	2	0.261	0.232	0.201
*EloT*066	3	0.125	0.149	0.142	2	0.231	0.212	0.183	3	0.250	0.242	0.215	2	0.115	0.111	0.103	3	0.174	0.165	0.154
*EloT*068	3	0.563	0.670	0.585	3	0.462	0.480	0.404	3	0.125	0.508	0.427	3	0.538	0.539	0.472	3	0.478	0.581	0.504
*EloT*069	5	0.781	0.767	0.715	6	0.636	0.792	0.720	5	0.875	0.750	0.657	5	0.692	0.695	0.637	6	0.739	0.653	0.598
Mean (x/46)	4.3	0.440	0.453	0.407	3.1	0.446	0.421	0.355	3.0	0.473	0.464	0.384	3.4	0.457	0.461	0.396	3.3	0.472	0.467	0.399

A = number of alleles, H_O = observed heterozygosity, H_E = expected heterozygosity, PIC = polymorphism information content; *Significant deviation from Hardy–Weinberg equilibrium was only detected in *EloT*028 for populations in Semangkok, Bukit Larut, Gunung Raya and Pulau Singa Besar FRs ($p < 0.001$, after Bonferroni correction)

number of populations, while Razi et al. (2013) reported on the relationship between Malaysian cultivars of *E. longifolia* inferred from RAPD markers, which was found to be associated with their geographical distribution. In order to identify priority areas for in-situ conservation and germplasm collection, comprehensive population genetic study of *E. longifolia* from the natural populations is necessary. The transcriptomic SSR markers developed will be useful for this purpose. Besides, these markers will facilitate DNA profiling in breeding-related programmes, for instance, cultivar/variety/clonal identification, plant variety protection and marker assisted selection. Having said that, loci with trinucleotide repeats should be tested for neutrality as they are more prone to selection (Hodel et al. 2016).

In Malaysia, apart from *E. longifolia*, there is another lesser-known medicinal plant species of the same genus, namely *E. apiculata*. Its leaves are used by indigenous people as an ingredient in post-partum bath (Ong et al. 2012). The majority of the transcriptomic SSRs developed in this study might likely be applicable to *E. apiculata* because transcriptomic SSRs have higher transferability across related species due to more conserved priming sites derived from coding regions of the genome. Guo et al. (2014) reported nearly 100% interspecific transferability of EST-SSRs from *Dysosma versipellis* (a traditional Chinese medicinal plant) to six other *Dysosma* spp. Of 14 loci tested, 13 loci were 100% transferable, while the remaining one locus had 83.3% success rate. In the case of *Helianthus annuus*, EST-SSRs were found to have more than three times higher transferability rate across species compared with genomic SSRs (73% vs. 21%, respectively) (Pashley et al. 2006). Ng et al. (2009) encountered a few *Shorea leprosula* EST-SSR loci that conferred 100% transferability across all the 36 dipterocarp species of several genera tested while Yan et al. (2017) reported 61.4% transferability of 114 *Melilotus albus* EST-SSR markers to 18 species of the same genus.

CONCLUSION

In this study, we utilised the root transcriptome data of a 10-year-old *E. longifolia* as a resource for SSR marker discovery. A total of 46 novel polymorphic transcriptomic SSR markers have been isolated and characterised for this important medicinal plant. These markers will complement the existing suite of genomic SSRs as valuable genetic tools for a wide spectrum of genetic analyses, not only for *E. longifolia* but also potentially for *E. apiculata*. Conservation genetics, cultivar and clonal identification, germplasm evaluation as well as marker assisted selection are among the manifold potential applications.

ACKNOWLEDGMENTS

We thank Saw JH, Lee PS, Yeow SK, Ramli P, Ghazali J, Yahya M, Yasri B, Sharifah T and Nurul Hudaini M for their excellent assistance in the laboratory and field. This research was supported by the Ministry of Agriculture and Agro-Based Industry (e-Science Project No. 05-03-10-SF1009) and the Ministry of Natural Resources and Environment, Malaysia, under the 10th Malaysia Plan.

REFERENCES

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Ang HH, Chan KL & Mak JW. 1995. *In vitro* antimalaria activity of quassinoids from *Eurycoma longifolia* against Malaysian chloroquine-resistant *Plasmodium falciparum* isolates. *Planta Medica* 61: 177–178. https://doi.org/10.1055/s-2006-958042.

Asiah O, Barbara J, Lessard PA. et al. 2003. Genetic diversity of *Eurycoma longifolia* inferred from single nucleotide polymorphisms. *Plant Physiology* 131: 1249–1301. https://doi.org/10.1104/pp.012492.

Bhat R & Karim AA. 2010. Tongkat ali (*Eurycoma longifolia* Jack): a review on its ethnobotany and pharmacological importance. *Fitoterapia* 81: 669–679. https://doi.org/10.1016/j.fitote.2010.04.006.

Bolger AM, Lohse M & Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina Sequence Data. *Bioinformatics* 30: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Bräutigam A & Gowik U. 2010. What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. *Plant Biology* 12: 831–841. https://doi.org/10.1111/j.1438-8677.2010.00373.x.

Burkill IH. 1966. *A Dictionary of the Economic Products of the Malay Peninsula*. Volumes 1 & 2. Ministry of Agriculture, Malaysia.

Chung JW, Kim TS, Suresh S, Lee SY & Cho GT. 2013. Development of 65 novel polymorphic cDNA-SSR markers in common vetch (*Vicia sativa* subsp. *sativa*) using next generation sequencing. *Molecules* 18: 8376. https://doi.org/10.3390/molecules18078376.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M & Robles M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676. https://doi.org/10.1093/bioinformatics/bti610.

Du M, Li N, Niu B et al. 2018. *De novo* transcriptome analysis of *Bagarius yarrelli* (Siluriformes: Sisoridae) and the search for potential SSR markers using RNA-Seq. *PLoS ONE* 13: e0190343. https://doi.org/10.1371/journal.pone.0190343.

Effendy NM, Mohamed N, Muhammad N, Mohamad IN & Shuid AN. 2012 *Eurycoma longifolia*: medicinal plant in the prevention and treatment of male osteoporosis due to androgen deficiency. *Evidence-Based Complementary and Alternative Medicine*: Article ID 125761. http://dx.doi.org/10.1155/2012/125761.

Ellegren H. 2004. Microsatellites: simple sequences with complex evolution. *Nature Review Genetics* 5: 435–445. https://doi:10.1038/nrg1348.

Guo R, Mao YR, Cai JR, Wang JY, Wu J & Qiu YX. 2014. Characterization and cross-species transferability of EST-SSR markers developed from the transcriptome of *Dysosma versipellis* (Berberidaceae) and their application to population genetic studies. *Molecular Breeding* 34: 1733-1746. https://doi: 10.1007/s11032-014-0134-z.

Han S, Wu Z, Jin Y, Yang W & Shi H. 2015. RNA-Seq analysis for transcriptome assembly, gene identification, and SSR mining in ginkgo (*Ginkgo biloba* L.). *Tree Genetics and Genomes* 11: 37. https://doi: 10.1007/s11295-015-0868-8.

Hodel RGJ, Gitzendanner MA, Germain-Aubrey CC et al. 2016. A new resource for the development of SSR markers: millions of loci from a thousand plant transcriptomes. *Applications in Plant Sciences* 4:1600024. https://doi.org/10.3732/apps.1600024.

Ismail SM, Abd Rahim AR & Ibrahim Z. 2004. The prospect of plantation and conservation of tongkat ali (*Eurycoma longifolia*) in Peninsular Malaysia. In: Chang YS, Mohtar M, Yunos NM (eds) *Proceedings of the Seminar on Medicinal Plants* 20-21 August 2002.

Lewis PO & Zaykin D. 2002. Genetic Data Analysis: computer program for the analysis of allelic data. Version 1.1. http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php.

Low B-S, Kumar Das P & Chan K-L. 2013. Standardized quassinoid-rich *Eurycoma longifolia* extract improved spermatogenesis and fertility in male rats via the hypothalamic-pituitary-gonadal axis. *Journal of Ethnopharmacology* 145: 706–714. https://doi.org/10.1016/j.jep.2012.11.013.

Lu Y-P, Su X, Luo W-C et al. 2017. Development of SSR markers from transcriptomes for *Orinus* (Poaceae), an endemic of the Qinghai–Tibetan plateau. *Applications in Plant Sciences* 5: 1700029. https://doi.org/10.3732/apps.1700029.

Metzker ML. 2010. Sequencing technologies—the next generation. *Nature Review Genetics* 11: 31–46. https://doi:10.1038/nrg2626.

Morita H, Kishi E, Takeya K, Itokawa H & Tanaka O. 1990. New quassinoids from the roots of *Eurycoma longifolia*. *Chemistry Letters* 19: 749–752. https://doi.org/10.1246/cl.1990.749.

Murray M & Thompson WF. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* 8: 4321–4325. https://doi.org/10.1093/nar/8.19.4321.

Ng KKS, Lee SL, Tsumura Y, Ueno S, Ng CH & Lee CT. 2009. EST-SSRs isolated from *Shorea leprosula* and their transferability to 36 species within the Dipterocarpaceae. *Molecular Ecology Resources* 9: 393–398. https://doi.org/10.1111/j.1755-0998.2008.02238.x.

Ong HC, Lina E & Milow P. 2012. Traditional knowledge and usage of medicinal plants among the Semai Orang Asli at Kampung Batu 16, Tapah, Perak, Malaysia. *Ethno-Medicine* 6: 207–211. https://doi.org/10.1080/09735070.2012.11886440.

Park SDE. 2001. Trypano tolerance in West African cattle and the population genetic effects of selection. PhD thesis, University of Dublin, Dublin.

Pashley CH, Ellis JR, McCouley DE & Burke JM. 2006. EST database as a source for molecular markers: lessons from *Helianthus*. *Journal of Heredity* 9: 381–188. https://doi.org/10.1093/jhered/esl013.

Peng Y, Gao X, Li R & Cao G. 2014. Transcriptome sequencing and *de novo* analysis of *Youngia japonica* using the illumine platform. *PLoS ONE* 9: e90636. https://doi.org/10.1371/journal.pone.0090636.

Razi ARM, Abdul-Aziz A, Alwee SSBS & Aziz R. 2013. Relationships between Malaysian cultivars of tongkat ali (*Eurycoma longifolia* Jack) obtained through RAPD analysis. *International Journal of Biotechnology for Wellness Industries* 2: 45–50.

Rehman SU, Choe K & Yoo HH. 2016. Review on a traditional herbal medicine, *Eurycoma longifolia* Jack (tongkat ali): its traditional uses, chemistry, evidence-based Pharmacology and Toxicology. *Molecules* 21: 331. https://doi:10.3390/molecules21030331.

Squirrell J, Hollingsworth M, Woodhead M et al. 2003. How much effort is required to isolate nuclear microsatellites from plants? *Molecular Ecology* 12: 1339–1348. https://doi.org/10.1046/j.1365-294X.2003.01825.x.

Tabbasam N, Zafar Z & Mehboob-ur-Rahman. 2014. Pros and cons of using genomic SSRs an EST-SSRs for resolving phylogeny of the genus *Gossypium*. *Plant Systematics and Evolution* 300: 559–575. https://doi.org/10.1007/s00606-013-0891-x.

Tada H, Yasuda F, Otano K, Doteuchi M, Ishihara Y & Shiro M. 1991. New antiulcer quassinoids from *Eurycoma longifolia*. *European Journal of Medicinal Chemistry* 26: 345-349. https://doi.org/10.1016/0223-5234(91)90069-Y.

Thiel T, Michalek W, Varshney RK & Graner A. 2003. Exploiting EST databases for the development and characterisation of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 106: 411–422. https://doi.org/10.1007/s00122-002-1031-0.

Tnah LH, Lee CT, Lee SL, Ng KKS, Ng CH & Hwang SS. 2011. Mircosatellite markers of an important medicinal plant, *Eurycoma longifolia* (Simaroubaceae), for DNA profiling. *American Journal of Botany* e130–e132. https://doi.org/10.3732/ajb.1000469.

van Oosterhout C, Hutchinson WF & Wills DM. 2004. MICRO-CHECKER: A software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4: 535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x.

Varshney RK, Graner A & Sorrells ME. 2005. Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* 23: 48–55. https://doi.org/10.1016/j.tibtech.2004.11.005.

Wang Z, Jiang C, Wen Q, Wang N, Tao Y & Xu L. 2014. Deep sequencing of the *Camellia chekiangoleosa* transcriptome revealed candidate genes for anthocyanin biosynthesis. *Gene* 538: 1–7. https://doi.org/10.1016/j.gene.2014.01.035.

Wenping H, Yuan Z, Jie S, Lijun Z & Zhezhi W. 2011. *De novo* transcriptome sequencing in *Silvia miltiorrhiza* to identify genes involved in the biosynthesis of active ingredients. *Genomics* 98: 272–279. https://doi.org/10.1016/j.ygeno.2011.03.012.

Yan Z, Wu F, Luo K et al. 2017. Cross-species transferability of EST-SSR markers developed from the transcriptome of *Melilotus* and their application to population genetics research. *Scientific Reports* 7: 17959. https://doi.org/10.1038/s41598-017-18049-8.

Zalapa JE, Cuevas H, Zhu H et al. 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *American Journal of Botany* 99: 193–208. https://doi.org/10.3732/ajb.1100394.

Zhao Y, Williams R, Prakash CS & He G. 2013. Identification and characterization of gene-based SSR markers in date palm (*Phoenix dactylifera* L.). *BMC Plant Biology* 12: 237. https://doi.org/10.1186/1471-2229-12-237.

Zhou T, Li ZH, Bai GQ et al. 2016. Transcriptome sequencing and development of genic SSR markers of an endangered Chinese endemic genus *Dipteronia* Oliver (*Aceraceae*). *Molecules* 21: 166. https://doi:10.3390/molecules21030166.

Appendix 1 Marker combination of the seven multiplex sets used for validation

Multiplex set	SSR loci	5' fluorescent labelling at the forward primer	Number of alleles	Allele size range
E1	*EloT*058	Hex	11	79–128
	*EloT*063	6-Fam	7	96–125
	*EloT*019	Hex	7	150–171
	*EloT*025	6-Fam	4	170–186
	*EloT*061	6-Fam	6	211–239
	*EloT*050	Hex	8	293–320
	*EloT*031	6-Fam	8	298–330
E2	*EloT*039	Hex	7	126–149
	*EloT*006	6-Fam	10	127–160
	*EloT*034	6-Fam	5	202–254
	*EloT*044	Hex	5	234–249
	*EloT*049	Hex	7	288–309
	*EloT*024	Hex	3	350–365
	*EloT*015	6-Fam	5	384–397
E3	*EloT*037	Hex	7	98–120
	*EloT*028	6-Fam	6	116–137
	*EloT*041	6-Fam	5	154–174
	*EloT*020	Hex	4	206–222
	*EloT*045	6-Fam	7	233–253
	*EloT*032	Hex	6	253–276
	*EloT*069	6-Fam	7	284–304
	*EloT*068	Hex	3	335–347
	*EloT*002	Hex	4	391–400
E4	*EloT*047	6-Fam	5	244–256
	*EloT*056	Hex	5	115–136
	*EloT*036	Hex	5	160–184
	*EloT*016	Hex	3	222–230
	*EloT*014	Hex	5	277–293
E5	*EloT*038	6-Fam	3	104–119
	*EloT*012	6-Fam	3	132–144
	*EloT*042	6-Fam	6	180–205
	*EloT*051	6-Fam	9	254–286
	*EloT*066	6-Fam	3	312–321
	*EloT*007	Hex	3	234–246
E6	*EloT*010	6-Fam	2	121–133
	*EloT*005	6-Fam	2	226–229
	*EloT*048	6-Fam	4	263–275
	*EloT*011	6-Fam	5	301–317
	*EloT*018	6-Fam	2	338–348
	*EloT*046	Hex	4	269–284
E7	*EloT*065	6-Fam	4	240–255
	*EloT*003	Hex	2	116–119
	*EloT*013	6-Fam	5	270–283
	*EloT*040	Hex	3	189–193
	*EloT*008	Hex	3	225–237
	*EloT*021	Hex	4	297–328

The number of alleles and allele size range for each locus were calculated based on 102 *Eurycoma longifolia* samples collected from Peninsular Malaysia