IDENTIFICATION OF SIMPLE SEQUENCE REPEAT MARKERS USING RESTRICTION SITE-ASSOCIATED DNA SEQUENCING (RADSeq) OF *GARCINIA PRAINIANA*

Khairun Hisam Nasir
Centre for Marker Discovery and Validation,
Biotechnology and Nanotechnology Research Centre,
MARDI Headquarters, Persiaran MARDI-UPM,
Serdang, Selangor,
e-mail:hairin@mardi.gov.my

Muhammad Fairuz Mohd Yusof,
Centre for Marker Discovery and Validation,
Biotechnology and Nanotechnology Research Centre,
MARDI Headquarters, Persiaran MARDI-UPM,
Serdang, Selangor,

Siti Norsaidah Ibrahim
Centre for Marker Discovery and Validation,
Biotechnology and Nanotechnology Research Centre,
MARDI Headquarters, Persiaran MARDI-UPM,
Serdang, Selangor,

Siti Nadrah Abd Hisham
Centre for Marker Discovery and Validation,
Biotechnology and Nanotechnology Research Centre,
MARDI Headquarters, Persiaran MARDI-UPM,
Serdang, Selangor,

Azuan Amron
Agrobiodiversity and Environments Research Centre,
MARDI Headquarters, Persiaran MARDI-UPM
Serdang, Selangor

ABSTRACT

Garcinia prainiana (cerapu) is an underutilized fruit and native to Malaysia. A restriction site-associated DNA sequencing (RADSeq) of G. prainiana DNA using restriction enzyme EcoR1combined with Illumina paired-end sequencing generated a clean data of 910.3 Mbp reads. Data assemblies produced 245,721 scaffolds with 38.05% guanine-cytosine (GC-content). These sequences were mined for simple sequence repeats (SSRs) and a total of 5,086 SSRs were identified which consisted of mono- (33.21%), di- (19.94%), tri- (13.43%), tetra- (1.91%), penta- (0.59%) and hexanucleotide (0.11%) repeats. From the 100 SSRs selected for fragment analysis of cerapu accessions, 11 polymorphic SSRs were identified, and a phylogenetic tree was generated, which consisted of two main groups.

Keywords: *Garcinia prainiana*, Restriction site-associated DNA sequencing (RADSeq), Simple sequence repeats (SSRs), fragment analysis, phylogenetic tree

INTRODUCTION

Garcinia prainiana (cerapu) belongs to the family Clusiaceae and grows naturally in Peninsular Malaysia and Southern Thailand. Cerapu is closely related to mangosteen and not less attractive compared to other known *Garcinia* species. In Peninsular Malaysia, it is mainly found in Pahang, Perak and Negeri Sembilan. The Cerapu bark is greyish-brown and produces a white sap when slashed. The leaves are oppositely arranged, large, ovate, 15-23 cm long, and 5-10 cm broad; the emerging young leaves are reddish, turning glossy dark-green when mature (Zawiah & Othaman, 2012).

Cerapu is a dioecious tree with male and female flowers on separate trees. The red or orange fragrant flowers emerge like jewels from the green branch tips (Aman, 2006). Both male and female flowers are identical in size, shape, and color. While the female flowers do not produce pollen grains, the male flowers produce largely visible pollen grains that are spherically arranged in its center. The fruits are round, 3-5 cm in diameter, shiny and green when raw, and yellowish-orange when ripe. These cerapu fruits are edible and have a pleasant sweet-sour taste (Adnan et al., 2018). In a traditional Malay recipe, the raw cerapu fruits are cooked together with dried fish (Zawiah & Othaman, 2012). The fruits are borne once a year, from July to September. The cerapu trees are slow-growing, long-lived, and can produce fruits even when only a few feet tall. Cerapu is an excellent plant for landscape gardens in parks and considered as Least Concern (LC) in the IUCN Red List version 3.1, 2011 (Nparks Flora Fauna website, 2020) because it is widely grown and does not face any threat of extinction.

Basic genomic resources of DNA markers have not yet been identified for *G. prainiana* for its genetic conservation. Some studies have reported the application of plastid coding region. Maturase K gene (matK) was used to establish the genetic relationship among selected *Garcinia* spp. (Zulhairil et al., 2020). A genetic diversity study on apomictic fruit species *Garcinia mangostana* was conducted using Randomly Amplified DNA Fingerprinting (RAF) (Carl et al., 2004). Some molecular markers to study the *Garcinia* spp. diversity using Inter Simple Sequence Repeat (ISSR) (Utpala et al., 2016) have been reported. To date, there are no published reports that have explored the simple sequence repeats (SSRs) of *G. prainiana*.

Molecular markers offer several advantages over the morphological and biochemical markers for the identification of species, phylogenetic studies, and generating the genetic linkage maps (Suh et al. 2011). Molecular marker such as the SSRs has earlier been used to identify and characterize fruit species *Baccaurea motleyana* (Khairun et al., 2020) and *Averrhoa carambola* (Khairun et al., 2019).

The use of cost-effective and practical strategies such as restriction site-associated DNA sequencing (RADseq), an extension of Next-generation sequencing (NGS) technologies to develop novel SSR markers, can be considered. In this method, sequencing of a reduced representation of genomic library is carried out using restriction site-associated DNA, which targets genomic regions adjacent to restriction sites. The fragments thus generated are sequenced via a suitable NGS platform and assembled to produce contigs that can be screened for SSR markers. This methodology has been effectively used in diverse applications such as marker identification (Miller et al., 2007), genome scans (Hohenlohe et al., 2010), population genetic differentiation (Emerson et al., 2010), and phylogeography (McCormack et al., 2012 and Zellmer et al., 2012).

Malaysian Agricultural Research and Development Institute (MARDI) is making efforts to sustain cerapu at MARDI Headquarters, Serdang, Selangor to MARDI Station Jelebu, Negeri Sembilan due to land acquisition for Mass Rapid Transit (MRT) depot. To assist in the mission to conserve cerapu, the Centre of Marker Discovery and Validation (CMDV) has embarked on RAD Sequencing, SSR mining, and DNA fragment analysis of the cerapu accessions. The cerapu SSR's data generated will be used for diversity study and redundancy checking of accessions in the future.

MATERIALS AND METHODS

Preparation of plant material

Leaf samples from cerapu plants were collected from a botanical garden in Wilayah Persekutuan Putrajaya. The plant material was frozen and stored in liquid nitrogen until further use.

Extraction of DNA

Total genomic DNA was extracted from frozen cerapu leaves with the Cetyl trimethylammonium bromide (CTAB) method (Turaki et al., 2017). CTAB is a cationic detergent that facilitates the separation of polysaccharide during purification while polyvinylpyrrolidone aid in removing polyphenols. Approximately 5 g of leaves were disintegrated in a tissue lyser (Qiagen, GmbH, Germany) with stainless steel beads. About 2000 µL of CTAB extraction buffer (2% (w/v) CTAB, 2% (w/v) polyvinylpyrrolidone (PVP), 4 mM diethyldithiocarbamate (DDC), 5 mM ascorbic acid, 1.4M NaCl, 100 mM Tris-HCl pH 8.0 and 20 mM EDTA) was added to the ground sample and incubated at 65 °C for 1 h with intermittent mixing. An equal volume of isopropanol was added to the above solution, mixed by inversion, and centrifuged at 5,500 rpm for 15 min; the supernatant was collected and washed twice with isopropanol. The DNA pellet was air-dried and resuspended in 100 µL tris-EDTA buffer.

For extraction of DNA from cerapu accessions, 11 accessions grown in a nursery at Agrobiodiversity and Environment Research Centre (MARDI Headquarters Serdang, Selangor) were collected, and the DNA was extracted following the same procedure.

Purification of DNA and determination of quality

To the 100 µL volume of DNA extract, 10 mg/mL RNAase I was added and incubated at 37 °C for 30 min. Following the incubation, the DNA was resuspended in 200 µL isopropanol, 125 µL 2.5 M LiCl, incubated at –20 °C for 2 h and was centrifuged at 15,000 rpm at 4 °C for 20 min. The DNA pellet was washed with absolute ethanol and again with 70% (v/v) ethanol, air-dried, and resuspended in 80 µL distilled water. The DNA yield was measured using Thermo Labsystems Fluoroskan Ascent™ (Thermo Fisher Scientific, USA). The DNA integrity was tested in 0.8% agarose gel electrophoresis.

Construction of RADSeq library

To construct a RADSeq library, 1 µg of genomic DNA was digested using *EcoR1* (New England Biolabs, USA) for 60 min at 37 °C, followed by a heat-inactivation step for 20 min at 65 °C. Modified Illumina adapter primers (barcoded EcoR1-P1) at 100 nM concentration (top:5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTxxxxx-3'and bottom: 5'phos-AATTxxxxxAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT-3', where xxxxx =barcode) were added to each sample along with 0.6 µL rATP (100 mM, Promega Corporation, USA), 1 µL NEB Buffer (10X), 0.5 µL T4 DNA Ligase (1000U) (New England Biolabs, USA), remaining volume made up with PCR-grade H_2O and incubated at room temperature for 30 min. The ligated samples with unique barcodes were heat-inactivated for 20 min at 65 °C and sheared with Covaris (Covaris, USA) with a DNA fragment peak of 500 bp. The sheared DNA was separated by gel electrophoresis in 1.5% agarose (0.5X TBE buffer), and a selected size (350–650 bp) of sheared DNA was purified with MinElute Gel Extraction Kit (Qiagen, Germany).

The recovered DNA was blunt end-repaired using a Quick blunting kit (New England Biolabs, USA) and the 3'-adenine overhangs were added to 10U of the 3'-5' exo-Klenow Fragment (New England Biolabs, USA) by incubating the purified sample at 37 °C for 30 min. The sequencing paired-end-P_2 Adapter (PE-P_2; 10 µM), which is a divergent modified Illumina adapter (top oligo: 5'-Phos-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAGACCGATCAGAACAA-3' and bottom oligo: 5'-CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT-3') were then ligated to the DNA fragments at room temperature, the library was column purified and gel size selected (350–650 bp). Finally, the prepared library was sequenced using the Illumina NovaSeq6000 sequencing platform (Illumina, USA) on PE150 mode.

SSR mining and SSR primers design

Raw sequence reads with RAD barcode sequence GGCTAC were filtered by customized Perl scripts and converted into FASTQ format files. The clean reads were then assembled, and the optimal assembly results were reached after multiple iterations.

Using MIcroSAtellite identification tool (MISA) (Sebastian et al., 2017) the SSRs were mined based on their minimum motif length (12 bp) and mono-, di-, tri-, tetra-, penta- and hexanucleotide repeat motifs with the compound as 0bp to be the maximum permitted size of interruption between two different SSRs. The SSR markers were identified by examining the distribution of repeats using a customized Perl script and R programming software.

SSR primers were designed using the Primer3 software (Andreas et al., 2012) with the following parameters: primer length of 18-23bp, with an optimum length of 21 bp, a PCR product size of 100-300 bp, an annealing temperature of 58–64 °C, with an optimum setting of 60 °C and a GC content of 45-75%.

Validation of SSR markers and PCR amplification

To validate the markers, one hundred SSRs were selected and PCR amplified. Amplified PCRs were used for fragment analysis. For the amplification, three specific oligomers were used: a forward primer (with an M13-tail sequence), an unlabeled reverse primer, and an M13 infrared fluorescent dye-labeled primer (Arif et al., 2010). The PCR reactions were performed in 10 µL reaction volumes with a mixture of 40 ng of template DNA (1·µL), 1 µL 10 X PCR buffer, 1.5–3.0 mM $MgCl_2$, 2 mM mix dNTP, 10 µM of each forward and reverse primer, 5 µmol of fluorescent reporter dye (6-FAM/VIC/NED/PET), and 1 U of *Taq* DNA polymerase (Invitrogen, USA). PCR amplifications were performed with the Peltier Thermal Cycler, DNA Engine Tetrad 2 (BioRad, USA). The PCR profile consisted of an initial denaturation for 5 min at 94 °C, followed by 34 cycles of 30s denaturation at 94 °C, 45s annealing (Table 5), and 45s extension at 72 °C, and a final extension step of 7 min at 72 °C. The obtained DNA fragments were further analyzed using the 3730xl DNA Analyzer (Thermo Fisher Scientific, USA).

Scoring and Data Analyses

The DNA fragment analysis was done using GeneMapper® 5.0 software (Applied Biosystems, USA) using the size standard GS500LIZ. The fluorescent data/peaks were scored as alleles in electropherograms (Glaubitz et al., 2004) and analyzed using MicroChecker v2.2.3 (Van Oosterho et al., 2004) to determine the presence of null alleles, large allelic dropouts, and stutter peak errors. The scores were then analyzed by the Power Marker software version 3.0 to evaluate the number of alleles per locus, allele frequencies of each variety, and their genetic diversities. The phylogenetic tree was constructed using the MEGA-X software (Sudir et al., 2018) by the unweighted pair group method with arithmetic mean (UPGMA) approach to generate the genetic distance matrix by the PowerMarker software.

RESULTS AND DISCUSSION

Sequencing and assembly

The cerapu sequence clean reads obtained using the Illumina sequencing were 910.3 Mbp compared to the raw 973.1 Mbp reads. The GC content and data Q20 of the clean reads were 38.63% and 97.58%, respectively (Table 1). A total of 245,721 scaffolds could be assembled from the cerapu clean reads from the base count of 42,787,211. The longest scaffold of these was of length 1,053 with 166 N50. The assembled sequence had 38.05% GC content which is slightly higher to dicods, typically range between 33 and 36% (e.g., *Carica papaya*, 34%; *Arabidopsis thaliana*, 36%) and lower to monocotyledonous (e.g., *Sorghum bicolor*, 43.9%; *Oryza sativa* ,43.6%) (Singh et al., 2016) (Table 2).

Table 1. Illumina statistic on cerapu sequence data

Sample ID	Raw data (Mb)	Clean data (Mb)	Clean data GC(%)	Clean data Q20(%)
Cerapu	973.1	910.3	38.63	97.58

Table 2. Genome assembly statistics

No. of all scaffolds	245,721
Bases in all scaffolds	42,787,211
Largest length	1,053
Scaffold N50	166
G+C content	38.05%
N rate	00.06%

Note: No. of all scaffolds refer to the total number of assembled sequences; bases in all scaffolds refer to the number of base pairs in the total assembled sequence; largest length refers to the longest scaffold; scaffold N50 refers to the length of shortest contigs among the large contigs that account for 50% of the assembly, G + C content refers to GC content of the assembled sequence; N rate refers to the N ratio in the assembled sequence.

Identification of candidate SSRs in cerapu

A total of 5,405 putative SSRs were identified by screening the 245,721 cerapu scaffold sequences. Of these, more than one SSRs were identified in 5,086 sequences and with 286–293 SSRs present in the compound forms (Table 3). A majority of SSR motifs obtained were dinucleotide (50.7%), follow by mono- (33.2%), tri- (13.4%), tetra- (1.49%), penta- (0.6%) and hexanucleotide (0.11%). High dinucleotide motif in cerapu was also found in many species (48-67%) (Wang et al., 1994; Schug et al., 1998) but in primates mononucleotides [mainly, poly (A/T)] are the most copious classes of SSRs (Toth et al. 2000; Wren et al. 2000). The most frequent motifs of cerapu SSRs were A/T for mono- AG/CT for di-, AAT/ATT for tri-, AAAT/ATTT for tetra-, AAAAG/CTTTT for penta-and AAAAAT/ATTTTT for hexanucleotide sequence (Table 4).

Table 3. Statistic of cerapu SSRs mining

Total number of sequences examined:	**245,721**
Total size of examined sequences (bp):	**42,787,211**
Total number of identified SSRs:	**5,405**
Number of SSR containing sequences:	**5,086**
Number of sequences containing more than 1 SSR:	**286**
Number of SSRs present in compound formation:	**293**

Table 4. Types of cerapu SSRs and their frequency

SSR type	Number	Most frequent repeat motif
Mononucleotide	1,795 (33.2%)	A/T
Dinucleotide	2,743 (50.7%)	AG/CT
Trinucleotide	726 (13.4%)	AAT/ATT
Tetranucleotide	103 (1.9%)	AAAT/ATTT
Pentanucleotide	32 (0.6%)	AAAAG/CTTTT
Hexanucleotide	6 (0.11%)	AAAAAT/ATTTTT
Total	5,405	

Construction of phylogenetic tree from scored SSRs

Of the hundred SSRs that were fragment analyzed, only 11 polymorphic SSRs were identified whereas the remaining SSRs were monomorphic and no/low allele calls. The scored SSRs produced allele numbers per locus ranging from 2 to 17(mean: 5.1765), allele frequency ranging from 0.3636 to 0.9167(mean: 0.7654), availability ranging from 0.5091 to 1 (mean: 0.9412), gene diversity ranging from 0.1553 to 0.7155 (mean: 0.3167), heterozygosity from 0.1429 to 1 (mean: 0.3586), polymorphic information content (PIC) ranging from 0.1480 to 0.6637 (mean: 0.2891) (Table 6).

Previous records of cerapu accession, a total of 40 accessions were conserved at MARDI Headquarters Serdang. In this study we managed to conserve only 11 accessions with genetic distance range from 0.0000-0.3203 (Table 7). This result showed low cerapu diversity due to some accessions was not able to be rescued during land acquisition by MRT. With this low diversity, lost and new cerapu accession has to be discovered in Malaysia forest in the future collection mission.

Further the genetic distance data were used for phylogenetic tree. The phylogenetic tree construct consisted of 2 groups of cerapu varieties: Group I contained ACC10, ACC8, ACC11, ACC5, ACC7, Kg Kelibang, and Kg Kuala Lhng and group II contained Kg Beris, Kg Pwngnong, ACC6, and ACC9 (Figure 1).

Table 5. List of 100 cerapu SSRs

No	Seq Id	Forward	Reverse	Annealing temperature (°C)	SSR repeat	Product size (base pair)
1	scaf5	5'-TGTAAAACGGCCAGTTTCCCTCTTCAACTCAGCGC-3'	5'-CCGGAGTCCGCCTAAAGAAG-3'	60	(TA)7	238
2	scaf99	5'-TGTAAAACGGCCAGTCCTGATTTCGGCGTCGGATA-3'	5'-TGTTCCAAGCTTTCCATTGCC-3'	60	(AG)7	168
3	scaf200	5'-TGTAAAACGGCCAGTTCCTCGTGTCCTGAGGGTAG-3'	5'-ACGATCCGAACCATCAGTCA-3'	59	(CT)6	168
4	scaf225	5'-TGTAAAACGGCCAGTTGCACGAGACAAATGATAGAAGGA-3'	5'-AGAATCATTTCTGTATCTGCAATTCA-3'	57	(TA)6	153
5	scaf228	5'-TGTAAAACGGCCAGTAGGGTACGAGAAAATGTGCCT-3'	5'-GCTTTTGAGCAGCACACGAA-3'	60	(AT)7	254
6	scaf268	5'-TGTAAAACGGCCAGTGACCTCCATCTTGGGCTAGC-3'	5'-ACCTAACCATCTGCACCTGT-3'	59	(AT)6	126
7	scaf416	5'-TGTAAAACGGCCAGTACCCAAAAGGCCCAGAGATG-3'	5'-CCCCTAAGGTAGTAGTTCAACTGG-3'	60	(TA)8	153
8	scaf429	5'-TGTAAAACGGCCAGTGCCCAGGAGTGTAAAGTCCC-3'	5'-ACCCCTAGAAACAACCCACT-3'	60	(AG)7	113
9	scaf464	5'-TGTAAAACGGCCAGTAGCCTTTTCTCCACACACTCT-3'	5'-GGCGCAACCTGTAAAATCCC-3'	60	(CT)9	230
10	scaf514	5'-TGTAAAACGGCCAGTAGAAAACCCAGTGAGTGGCA-3'	5'-GGCCGCTGTTGGTATATGGT-3'	60	(CA)10	168
11	scaf599	5'-TGTAAAACGGCCAGTTGATCTCCGCCTTCCCAGTA-3'	5'-CTCGAGTGGACATTACCGGG-3'	60	(AC)21	200
12	scaf621	5'-TGTAAAACGGCCAGTCCCAAACTCACCCCAAACCA-3'	5'-TGGAGTGCTTACCTTCCTCT-3'	60	(AG)6	174
13	scaf664	5'-TGTAAAACGGCCAGTGCCAGAATCGCGGGAAAATC-3'	5'-CGGCTAGTTTCCTCGGGTTT-3'	60	(A)14	170
14	scaf672	5'-TGTAAAACGGCCAGTTGGTGAGTTCGGTGTGTCTG-3'	5'-CGAAGACACGAGCGATGTGA-3'	60	(AT)6	104
15	scaf686	5'-TGTAAAACGGCCAGTCCGATGATCACTCTCTGCCC-3'	5'-TCACAACTCATTAACTTATTCCACGT-3'	60	(TG)7	256
16	scaf701	5'-TGTAAAACGGCCAGTACGCCGTTCTGTCACTTTCT-3'	5'-CGGTCCGCGACATTTTGAAA-3-	60	(CT)6	143
17	scaf719	5'-TGTAAAACGGCCAGTTTCCCATCACTTTCTCTCTCT-3'	5'-AAAACTGCAAGAAACGGCCC-3'	60	(TC)8	227
18	scaf795	5'-TGTAAAACGGCCAGTTCACAGTGAAGTGCCCGAAG-3'	5'-TTTTTGCGTTCACCGGAA-3'	60	(AT)6	279
19	scaf867	5'-TGTAAAACGGCCAGTCGGTTCTCACCCAACTCTCA-3'	5'-GGAAGGTGCTTTGGTGAGGT-3'	60	(AC)6	215
20	scaf871	5'-TGTAAAACGGCCAGTTGTGAGACAGTGTGACGTGG-3'	5'-GGCACGTCTAGACAAGGGAG-3'	60	(GT)6	203
21	scaf875	5'-TGTAAAACGGCCAGTAGTTGCGATTGGAGCTGGAA-3'	5'-ACGTTTCTAGCTTTCCATTGCC-3'	60	(AG)8	135
22	scaf926	5'-TGTAAAACGGCCAGTACTTGTGACTTGTGAATGTGT-3'	5'-CCTCTTGGGTTGCCCTTTTT-3'	59	(AT)6	236
23	scaf1032	5'-TGTAAAACGGCCAGTACATACACAACCTTCCACACA-3'	5'-AGGGTTGATCTGAGGCCTAGA-3'	60	(CA)7	245
24	scaf1096	5'-TGTAAAACGGCCAGTTGTTCCAAGCTCTCTATTGCCT-3'	5'-GGATACAATCCGGCGCTGTA-3'	60	(TC)6	194
25	scaf1134	5'-TGTAAAACGGCCAGTATCACTGTTCGGGGCCATTT-3'	5'-CTCGTGTTGGTGACAGTCGA-3'	60	(GA)6	142

26	scaf1372	5'-TGTAAAACGGCCAGTTGTTGCTAGACCTTTGGAGCT-3'	5'-ACCTGTCACATTGCCACACA-3'	60	(AG)8	192
27	scaf1524	5'-TGTAAAACGGCCAGTAGAGGGTGAGCACAACAACA-3'	5'-CACTGGGGATGCCTGAGATC-3'	60	(GT)6	264
28	scaf1633	5'-TGTAAAACGGCCAGTAGTCTAAGTGAGGTCATATCATGTCA-3'	5'-ATGAGCCAGGCATACCAGTG-3'	60	(CA)11	251
29	scaf1643	5'-TGTAAAACGGCCAGTCCATGTATCTAGCTTCCATTTGCC-3'	5'-TGGTGCTGTATCCGGACCTA-3'	60	(CT)10	191
30	scaf1772	5'-TGTAAAACGGCCAGTAACTTTAGGATCAAAACATCTTACTT-3'	5'-TGGGTGGATATGGGTTTCCT-3'	58	(AT)6	152
31	scaf1818	5'-TGTAAAACGGCCAGTTGCATCTTGAGAGAACAAGTGT-3'	5'-GGTTTCACATCATCACACATGCA-3'	60	(TG)7	161
32	scaf1907	5'-TGTAAAACGGCCAGTTCATGAACTTTCTCTCTCTAACC-3'	5'-AACTGCAGGAAACGAGCCTT-3'	60	(CT)8	216
33	scaf2053	5'-TGTAAAACGGCCAGTAAGTAGTTGCTGTCCGTGGG-3'	5'-TGACTTTCTCTCTCTAACCCTCT-3'	60	(GA)6	192
34	scaf2094	5'-TGTAAAACGGCCAGTACCCATAATCTCCATGTTTCTAGCT-3'	5'-ACGAGTCCCTTCCAGTGACT-3'	60	(TC)8	265
35	scaf2108	5'-TGTAAAACGGCCAGTGCCCAGCCCATCAAAATGTG-3'	5'-TGGTTGGCTGCTGAATTTGC-3'	60	(AT)6	131
36	scaf2111	5'-TGTAAAACGGCCAGTACGCCGAATACACCTACAGC-3'	5'-TCATTCTCCTTGTTGAATGGGA-3'	60	(AG)9	213
37	scaf2158	5'-TGTAAAACGGCCAGTACGTACTCTATGAGGGGGTAAGT-3'	5'-GCAAGAGGAAAGGCATTGCA-3'	60	(AT)8	239
38	scaf2241	5'-TGTAAAACGGCCAGTAACATGAGGGGTGTTGGTCC-3'	5'-AAGCCGGGTACGCGATAAAA-3'	60	(AT)8	133
39	scaf2323	5'-TGTAAAACGGCCAGTTGGATATGAGTTGAAATGATGCGA-3'	5'-CCAACTTGCACTCACATGCC-3'	60	(TG)9	176
40	scaf2370	5'-TGTAAAACGGCCAGTTGGTTCACACCCTCAAGTCAG-3'	5'-TGCTTTTATTTGGAAATGGATGCA-3'	60	(AC)8	180
41	scaf2445	5'-TGTAAAACGGCCAGTCGGGCTAAGGCTCATCAGAG-3'	5'-TGCATGCTCTCTCACATGGT-3'	60	(AC)7	245
42	scaf2536	5'-TGTAAAACGGCCAGTATTCCACCCTCCAAGCTTGT-3'	5'-AAGGGACCTCATTTCTCACA-3'	59	(CT)6	114
43	scaf2577	5'-TGTAAAACGGCCAGTGGAGCCCACGTACATCTCAC-3'	5'-GGTAGTAAACTACTCCATATTATCTCA-3'	60	(TC)7	157
44	scaf2628	5'-TGTAAAACGGCCAGTCTCCTTGCCTCCACCTCAAG-3'	5'-TCCCTAACCTACACATGCTCA-3'	60	(AT)7	149
45	scaf2630	5'-TGTAAAACGGCCAGTCGCCATCGGTGCAATATGTG-3'	5'-TAAGGAGGTGGGCGGTTTTC-3'	60	(AG)7	138
46	scaf2823	5'-TGTAAAACGGCCAGTGAGTACTTAACCATGAAGTGAGGGT-3'	5'-TGCCCTTCTTGTTAAGTTCCCA-3'	60	(GA)6	100
47	scaf2862	5'-TGTAAAACGGCCAGTTCCATGTATCTAGCTTTCATTTGCC-3'	5'-AGTAAGTATCCGGCCCTGGT-3'	60	(CT)10	240
48	scaf2959	5'-TGTAAAACGGCCAGTAGCATGACAATTTCATTTGCGT-3'	5'-TGTGACTCGCGAATGTGGAT-3'	59	(TA)7	214
49	scaf2979	5'-TGTAAAACGGCCAGTTCCACTTTCTCTCTCTAGCCAC-3'	5'-CCTAAAGGGGGTTGCGACAT-3'	60	(TC)9	133
50	scaf3032	5'-TGTAAAACGGCCAGTTGCCTATGTATTCCTCATATGTTACGT-3'	5'-GGACATTTCCTCCCTGGCAA-3'	60	(TA)7	101
51	scaf111	5'-TGTAAAACGGCCAGTGCAAGAAACCAGGGCACATC-3'	5'-CGGCTTCATCCTCACTCCAA-3'	60	(GAA)5	108
52	scaf343	5'-TGTAAAACGGCCAGTGGAGGTCCTTGGACCCTGTA-3'	5'-TCAAACGGACAGAACCTCCC-3'	60	(TGA)5	250
53	scaf752	5'-TGTAAAACGGCCAGTTACAACAACAGGCACAGGCA-3'	5'-TAGTAGATGGTCCACCGGGG-3'	60	(CCA)5	237

54	scaf1051	5'-TGTAAAACGGCCAGTACACCCCTCACCCCATAAGA-3'	5'-AACCCCATGCGTTGTCTGAA-3'	60	(ATT)5	140
55	scaf1653	5'-TGTAAAACGGCCAGTCCGTGTGACAGGATGTGTTT-3'	5'-CAACCCACCTCCCATCGATC-3'	60	(ATA)5	248
56	scaf1708	5'-TGTAAAACGGCCAGTTGGAAGTGACCAATTTGAAGATGA-3'	5'-AGTCTAAGTCAAAACACACACACA-3'	59	(ATG)7	260
57	scaf1887	5'-TGTAAAACGGCCAGTGCAACAAGAGAAAGAGGGGC-3'	5'-TGCTCGATCTTTGCCTTCATG-3'	60	(AAG)11	104
58	scaf2110	5'-TGTAAAACGGCCAGTAGGCTACACCCAATGCAGAC-3'	5'-TGAAGCACCTGATCTGCCTG-3'	60	(CCA)5	255
59	scaf2186	5'-TGTAAAACGGCCAGTCCATCCTGCAGCTTCTCCAA-3'	5'-AGGGATAGGAGCCTGCTGAA-3'	60	(CAG)5	137
60	scaf2443	5'-TGTAAAACGGCCAGTACGACGCGCATCATATACCA-3'	5'-CTGCCTTCTAACTAGGGTGCA-3'	60	(GAA)5	254
61	scaf2683	5'-TGTAAAACGGCCAGTGTCTGTGTCCTCTCTGGCAC-3'	5'-ATTAGTGAAGGCGTGGGTGG-3'	60	(ACA)9	104
62	scaf2795	5'-TGTAAAACGGCCAGTTGGAGGAAGATAGGGTCCCG-3'	5'-GGGGCTTCTCATTCGGATCC-3'	60	(GAG)8	189
63	scaf3144	5'-TGTAAAACGGCCAGTCCCACATTGCATGAAGAGGG-3'	5'-TTGTTGTGGTGGTGGAGGTC-3'	60	(AAC)5	234
64	scaf3357	5'-TGTAAAACGGCCAGTGAGTGGGGTGGATCATGAGC-3'	5'-TGGCATGGGTTGAGCTACAG-3'	60	(ACA)5	193
65	scaf4245	5'-TGTAAAACGGCCAGTATCACCTCGAAGCCTGATCG-3'	5'-AGTGTCTGATTGTGGCCTGA-3'	59	(TAA)10	162
66	scaf4378	5'-TGTAAAACGGCCAGTTCCTTCTCACCTTACAATGAAAAGC-3'	5'-TCTCTCTAGACTTTCCTCCATCT-3'	59	(CCA)5	132
67	scaf4463	5'-TGTAAAACGGCCAGTACCACCTCAACAACAACCACT-3'	5'-ACCATTCTCCCACTTGTTTGC-3'	60	(ATC)5	252
68	scaf4808	5'-TGTAAAACGGCCAGTATCTTTTGGCGCTTCAACTG-3'	5'-GAGCTTGTTGACGCTCCTCT-3'	60	(GCA)5	123
69	scaf5102	5'-TGTAAAACGGCCAGTTTGTTGTGGTGGTGGAGGTC-3'	5'-TGGAGTAAACAAATGCTATTGGAGG-3'	60	(GTT)6	206
70	scaf5328	5'-TGTAAAACGGCCAGTAAGGTACCACGAGAAGCAGC-3'	5'-AATGACTCCACAGTACAAGTG-3'	60	(AT)7	163
71	scaf5883	5'-TGTAAAACGGCCAGTTCTCCAACAGCAGTAGCAGC-3'	5'-CAGCTGATCTGCCTGCTAGG-3'	60	(CCA)5	211
72	scaf6132	5'-TGTAAAACGGCCAGTAGAGCACACGGTGTACAAGA-3'	5'-CCCCAACAAGGCCCCAATAA-3'	60	(TTG)6	135
73	scaf6251	5'-TGTAAAACGGCCAGTGTGACCTAGCCATATCCAGCA-3'	5'-CGAGGATGGAAGGCTTTGCT-3'	60	(TTA)9	227
74	scaf6306	5'-TGTAAAACGGCCAGTGGAGATGCTGACACAGGCAT-3'	5'-TCTCAGAATCACCAGATCCAGA-3'	60	(TGA)5	238
75	scaf6771	5'-TGTAAAACGGCCAGTATTCCGACAAGAGGGCGAAA-3'	5'-TGCTGAGAATGGCGAGCTAG-3'	60	(AGG)5	150
76	scaf6886	5'-TGTAAAACGGCCAGTGCTATTGCCGTTGCTGCTAC-3'	5'-TGAAGCCCTTATCAGGTGCG-3'	60	(TGT)6	107
77	scaf6968	5'-TGTAAAACGGCCAGTCTCATTGAAGGCCTCTAGGTGT-3'	5'-AGTCTGTGAGTACTTGGTCCA-3'	60	(AAG)12	112
78	scaf7625	5'-TGTAAAACGGCCAGTTGGCCTTATGATTATTTTTGGGA-3'	5'-ACGGATAGAGTCTCCCCTGT-3'	59	(TGA)6	183
79	scaf7792	5'-TGTAAAACGGCCAGTACGTCGAGTGATCAATAGACCT-3'	5'-GCACTAGCAAACAAATTGTAGCC-3'	59	(TAA)5	122
80	scaf8110	5'-TGTAAAACGGCCAGTTGGGGTGGGGATTGATGTTG-3'	5'-TGGATGGAAGCAACACCCAA-3'	60	(TTG)5	102
81	scaf2421	5'-TGTAAAACGGCCAGTTGCCTTATTTAAGTGAAGGGCG-3'	5'-ACTCTTTCAAATGACATTCTTGAGTCA-3'	60	(AAAT)5	133

82	scaf4506	5'-TGTAAAACGGCCAGTGAGCCGACCGGTTTTTGTTC-3'	5'-CCCCTTATGCTCTTCGCCAA-3'	60	(ATAA)6	106
83	scaf7949	5'-TGTAAAACGGCCAGTCACACCTGGCTACTGCTGAT-3'	5'-TGTCTGTGGATGGAGGTTTCC-3'	60	(TTAA)5	199
84	scaf9130	5'-TGTAAAACGGCCAGTTACCAGGCATCCCCTAACCT-3'	5'-CTTGGTCATCGCTCCCTTGT-3'	60	(AAAG)5	108
85	scaf19231	5'-TGTAAAACGGCCAGTAAACAAATCGGCCAACCAGC-3'	5'-CCAGGAGCTGTTGATTGGGT-3'	60	(AACC)6	192
86	scaf22436	5'-TGTAAAACGGCCAGTAGTAACACCCACTTTTACACAA-3'	5'-AGAGAGAGGAGAAGCGGAG-3'	60	(AAAC)5	159
87	scaf27210	5'-TGTAAAACGGCCAGTCTCTAGCTCCACTGCCAAGC-3'	5'-CGGGCCCTTGTACTCTTTG-3'	60	(AATT)5	166
88	scaf29549	5'-TGTAAAACGGCCAGTAACACGAAGAGAAGTCGGCT-3'	5'-TGGAGGAGATGGATGTAACAGA-3'	59	(AAAT)5	160
89	scaf31908	5'-TGTAAAACGGCCAGTTGGTCCTCAGGATAGGAGAGT-3'	5'-AGTGCTTATCTTGGGATGCA-3'	59	(AAAT)5	155
90	scaf34068	5'-TGTAAAACGGCCAGTAGGCTGCGAGTTATTGTCGT-3'	5'-TGCGAAGGTAAAATCCCTGCT-3'	60	(TTTA)5	138
91	scaf5044	5'-TGTAAAACGGCCAGTGACCCCGAGATCTTTTGGGT-3'	5'-GGGACGCCTAGAGGTCAGAT-3'	60	(ATCTC)5	157
92	scaf11939	5'-TGTAAAACGGCCAGTTTGCGCTATATGGGTCGGAC-3'	5'-ACCTCATCTCATCACTCACCA-3'	60	(TTTTA)5	168
93	scaf14754	5'-TGTAAAACGGCCAGTTCTGGTCCTTTTGGGTCACC-3'	5'-GGTATGACGGCCGATCAGAG-3'	60	(ATCTC)6	156
94	scaf26339	5'-TGTAAAACGGCCAGTTGGTTGAGGGACCCAAAAGG-3'	5'-CTGGCTTGTTGGGCTAGTCA-3'	60	(GAGAT)5	114
95	scaf28882	5'-TGTAAAACGGCCAGTGACGTGGTTGAAGGACCCAA-3'	5'-TCTTCACTGGCTTGTCGAACT-3'	60	(TGAGA)6	124
96	scaf31591	5'-TGTAAAACGGCCAGTTGTGCATTTTCTTCGTTAATGCT-3'	5'-ACATGAGTAGTCCATCTGCGT-3'	59	(AAAAG)5	132
97	scaf32973	5'-TGTAAAACGGCCAGTTCCCCTCCTCTCAAACAAACC-3'	5'-TGAAGCAATACCCTTCCCCG-3'	60	(ATTTT)5	171
98	scaf38880	5'-TGTAAAACGGCCAGTTGAGGGGCCTATGAGATCCA-3'	5'-CCTATTGGCTCTTTTCACCCA-3'	60	(AAAGA)9	100
99	scaf14697	5'-TGTAAAACGGCCAGTACAAAGGAAAAGGAAAAGGAAAA-3'	5'-CGAGGTGATCCGACCGATC-3'	60	(AAAGGA)5	195
100	scaf17449	5'-TGTAAAACGGCCAGTAGAACTCCTTCATGTGTCGT-3'	5'-GCTTGTTAGGAGGCAACCCA-3'	60	(AATAAA)5	207

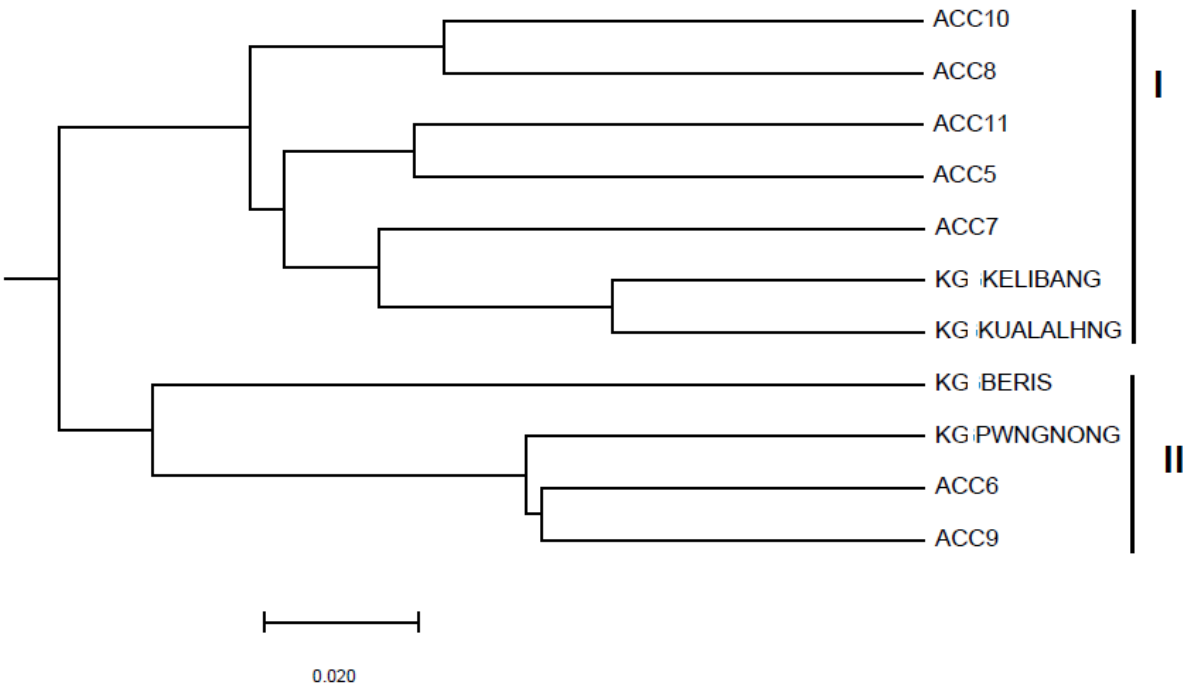
Table 6. Characterization of polymorphic SSRs in cerapu

No	Marker	Allele Frequency	Allele No	Availability	Gene Diversity	Heterozygosity	PIC
1	Scaf228	0.4907	4.0000	0.9818	0.6325	1.0000	0.5639
2	Scaf621	0.8214	8.0000	0.5091	0.3202	0.1429	0.3132
3	Scaf672	0.6727	5.0000	1.0000	0.5020	0.2909	0.4604
4	Scaf795	0.4727	17.0000	1.0000	0.6545	0.8727	0.5994
5	Scaf2094	0.5849	14.0000	0.9636	0.6072	0.3585	0.5718
6	Scaf1708	0.5000	4.0000	1.0000	0.5494	1.0000	0.4483
7	Scaf2110	0.3636	4.0000	1.0000	0.7155	0.9818	0.6637
8	Scaf2443	0.5455	7.0000	1.0000	0.5977	0.7455	0.5335
9	Scaf4378	0.8727	8.0000	1.0000	0.2336	0.2000	0.2262
10	Scaf29549	0.9167	4.0000	0.9818	0.1553	0.1667	0.1480
11	Scaf31908	0.8679	2.0000	0.9636	0.2293	0.2264	0.2030
	Mean	0.7654	5.1765	0.9412	0.3167	0.3586	0.2891

Table 7. Genetic distance of selected cerapu accessions

OTU	ACC10	ACC11	ACC5	ACC6	ACC7	ACC8	ACC9	KG BERIS	KG KELIBANG	KG KUALA LHNG	KG PWNG NONG
ACC10	0.0000	0.1297	0.1396	0.1875	0.2203	0.1250	0.2000	0.2500	0.2000	0.1688	0.2125
ACC11	0.1297	0.0000	0.1328	0.1529	0.1969	0.1912	0.1294	0.2265	0.1574	0.1234	0.1407
ACC5	0.1396	0.1328	0.0000	0.1990	0.2036	0.2108	0.1931	0.2755	0.1990	0.1208	0.2049
ACC6	0.1875	0.1529	0.1990	0.0000	0.2484	0.2221	0.1000	0.1588	0.2324	0.2063	0.1039
ACC7	0.2203	0.1969	0.2036	0.2484	0.0000	0.1734	0.2359	0.3203	0.1609	0.1234	0.2578
ACC8	0.1250	0.1912	0.2108	0.2221	0.1734	0.0000	0.2397	0.3162	0.1868	0.1375	0.2456
ACC9	0.2000	0.1294	0.1931	0.1000	0.2359	0.2397	0.0000	0.2118	0.2147	0.1938	0.1039
KG BERIS	0.2500	0.2265	0.2755	0.1588	0.3203	0.3162	0.2118	0.0000	0.2912	0.2875	0.2333
KG KELIBANG	0.2000	0.1574	0.1990	0.2324	0.1609	0.1868	0.2147	0.2912	0.0000	0.0813	0.2294
KG KUALA LHNG	0.1688	0.1234	0.1208	0.2063	0.1234	0.1375	0.1938	0.2875	0.0813	0.0000	0.2063
KG PWNG NONG	0.2125	0.1407	0.2049	0.1039	0.2578	0.2456	0.1039	0.2333	0.2294	0.2063	0.0000

Figure 1. Phylogenetic tree of the cerapu SSRs generated via MEGA-X software



CONCLUSION AND RECOMENDATION

Conclusion

Cerapu SSRs were identified using the novel RADSeq technique. By SSR mining, 245,721 scaffold sequences in which 5,405 SSRs with dinucleotide repeats were identified, also with the majority of nucleotide repeats. Of the hundred cerapu SSRs selected for fragment analysis, 11 SSRs were characterized for polymorphism. A phylogenetic tree was constructed from which two cerapu groups were identified. Low genetic diversity of cerapu accessions was detected indicated that discovery of new accession is pivotal to increase MARDI cerapu germplasm. The cerapu genetic information generated from this study will be utilized for cerapu conservation and new accession of cerapu at MARDI germplasm in the future.

Recommendation

1. Conservation of available cerapu at MARDI Headquaters Serdang, Selangor to MARDI station Jelepu, Negeri Sembilan is important for short term and also for long term management.
2. MARDI station, Jelebu,Negeri Sembilan should be nominated as a Center of excellent for cerapu and other underutilized crops.
3. Others potential MARDI stations should be identified for cerapu conservation as a backup material for unforeseen circumstances.
4. New cerapu accessions to MARDI germplasm by activate discovery mission into Malaysian forest.

REFERENCES

Adnan, M., Zainuddin, A.F., Hamzah, M.A., Moorthy, M. and Mohamad Zaki, M.I. (2018). Koleksi Pokok Taman Botani Kepong. Institut Penyelidikan Perhutanan Malaysia (FRIM), Malaysia, pp. 1–234. (https://www.mybis.gov.my/pb/2290).

Aman, R. (2006). Buah-buahan nadir Semenanjung Malaysia. Kuala Lumpur: Dewan Bahasa dan Pustaka, Kementerian Pendidikan Malaysia, pp. 178.

Andreas, U., Ioana, C., Triinu, K., Jian Y., Brant, C.F., Maido, R. and Steven G. R. (2017). Primer3-new capabilities and interfaces. Nucleic Acids Research, Vol 40(15).

Arif, I. A., Khan, H. A., Shobrak, M., Al Homaidan, A. A., Al Sadoon, M., Al Farhan, A. H. and Bahkali, A. H. (2010). Interpretation of Electrophoretograms of Seven Microsatellite Loci to Determine the Genetic Diversity of the Arabian Oryx. Genetics and Molecular Research, 9 (1): 259–65. doi: 10.4238/vol9–1 gmr714.

Carl, M. R., Lillian, S., Cameron, P. P., Bernard, J. Carroll. and Roderick, A. D. (2004). Genetic diversity revealed in the apomictic fruit species *Garcinia mangostana* L. (mangosteen). Euphytica, 136: 1–10.

Emerson, K., Merz, C., Catchen, J., Hohenlohe, P., Cresko, W. and Bradshaw, W. (2010). Resolving postglacial phylogeography using high-throughput sequencing. Proc Natl Acad Sci USA, 107:16196–16200. doi: 10.1073/pnas.1006538107 PMID: 20798348.

Glaubitz, J. C. (2004). Convert: A User-Friendly Program to Reformat Diploid Genotypic Data for Commonly Used Population Genetic Software Packages. Molecular Ecology Notes, 4(2): 309–10.

Hohenlohe, P., Bassham, S., Etter, P., Stiffler, N., Johnson, E., Cresko, W. (2010). Population genomics of parallel adaptation in three spine stickleback using sequenced RAD tags. PLoS Genet, 6: e1000862. doi: 10.1371/journal.pgen.1000862 PMID: 20195501.

Khairun, H. N., Muhammad Fairuz, M. Y., Mohd Shahril, F., Siti Zainab, J., Mira Farzana, M. M., Siti Norsaidah, I., Noor Baiti, A. A. and Joanna Cho L. Y. (2019). Development of Starfruit Simple Sequence Repeat (SSR) Using Next Generation Sequencing. Journal of Food Science and Engineering, 9: 95–121 doi: 10.17265/2159–5828/2019.03.003.

Khairun, H. N., Muhammad Fairuz, M. Y., Mohd Shahril, F. A. R., Siti Norsaidah, I., Mira Farzana, M. M., Khadijah, A. and Mohd Shukri, M. Ali. (2020). Discovery of Simple Sequence Repeat Markers through Transcriptome Analysis of *Baccaurea motleyana*. Journal of Food Science and Engineering, 1 doi: 10.17265/2159–5828/2020.

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T. (2012). Applications of next-generation sequencing to phylogeography and phylogenetics. Mol Phylogenet Evol, 62: 397–406.

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., Johnson, E.A. (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res, 17: 240–248. PMID: 17189378.

Nparks Flora Fauna website. Garcinia prainiana (2020). National Parks Board. Retrieved from https://florafaunaweb.nparks.gov.sg.

Sebastian, B., Thomas, T., Thomas, M., Uwe, S. and Martin, M. (2017). MISA-web: a web server for microsatellite prediction. Bioinformatics, 33(16): 2583–2585.

Singh, R., Ming, R. and Yu, Q.Y.(2016). Comparative analysis of GC content variations in plant genomes. Trop Plant Biol, 9(3):136–49.

Sudhir K., Glen, S., Michael, Li., Christina, K. and Koichiro, T. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol, 35(6):1547–1549.

Schug, M.D., Wetterstrand, K.A. and Gaudette, M.S. (1998). The distribution and frequency of microsatellite loci in *Drosophila melanogaster*. Molecular Ecology, **7**: 57-70.

Suh, J.K., Hetterscheid, W., Lee, A.K., Hong, J. and Roh, M.S. (2011). Identification and evaluation of Forsythia germplasm using molecular markers. Genet Resour Crop Evol, 58: 1225–1235.

Toth, G., Gaspari, Z., Jurka, J. (2000). Microsatellites in different eukaryotic genomes: survey and analysis. Genome Research, 10:967-981.

Turaki, A.A., Ahmad, B., Magaji, U.F., Abdulrazak, U.K., Yusof, B.A. and Hamza, A.B. (2017). Optimised cetyltrimethylammonium bromide (CTAB) DNA extraction method of plant leaf with high polysaccharide and polyphenolic compounds for downstream reliable molecular analyses. African Journal of Biotechnology Vol, 16(24):1354-1365.

Utpala, P., Nandakishore, O.P., Rosana, O.B., Nirmal Babu, K., Sentril Kumar, R. and Parthasarathy, V.A. (2016). Identification of molecular marker to study the *Garcinia spp*. Diversity. Indian Journal of Experimental Biology Vol, 54:400–405.

Wren, J.D., Forgacs, E. and Fondon, J.W. (2000). Repeat polymorphisms within gene regions: Phenotypic and evolutionary implications. American Journal of Human Genetics, 67:345-356.

Van Oosterho, C., Hutchinson, W. F. and Wills, D. P. (2004). "MICRO-CHECKER: Software for Identifying and Correcting Genotyping Errors in Microsatellite Data." Molecular Ecology Notes, 4 (3): 535–8. https://doi.org/10.1111/j.1471–8286.2004.00684.x.

Wang, Z., Weber, J, L., Zhong, G., Tanksley, S.D. (1994). Survey of plant short tandem DNA repeats. Theoretical and Applied Genetics, 88:1-6.

Zellmer, A.J, Hanes, M.M., Hird, S.M. and Carstens, B.C. (2012). Deep phylogeographic structure and environmental differentiation in the carnivorous plant Sarracenia alata. Syst Biol, 61:763–777. doi: 10.1093/sysbio/sys048 PMID: 22556200 14.

Zawiah, N. and Othaman, H. (2012). 99 Spesies Buah di FRIM. Institut Penyelidikan Perhutanan Malaysia, Kementerian Sumber Asli dan Alam Sekitar, Malaysia, pp. 22–23, 256.

Zulhairil, A., Azuan, A. and Nuradni, H. (2020). Phylogenetic relationship on *Garcinia* species using matK. Buletin Teknologi MARDI Bil. 21, Khas Agrobiodiversiti 201–210.