



Research article

Development of novel microsatellite markers using RAD sequencing technology for diversity assessment of rambutan (*Nephelium lappaceum* L.) germplasm

Shahril Ab Razak^{a,*}, Salehudin Mad Radzuan^b, Norkhairi Mohamed^c,
Nor Helwa Ezzah Nor Azman^a, Alny Marlynni Abd Majid^a, Siti Norhayati Ismail^a,
Muhammad Fairuz Mohd Yusof^a, Johari Sarip^d, Khairun Hisam Nasir^a

^a Biotechnology & Nanotechnology Research Centre, MARDI Headquarters, 43400, Serdang, Selangor, Malaysia

^b Horticulture Research Centre, MARDI Sintok, 06050, Bukit Kayu Hitam, Kedah, Malaysia

^c Department of Agriculture, Hulu Paka Agricultural Centre, 23300, Dungun, Terengganu, Malaysia

^d Horticulture Research Centre, MARDI Headquarters, 43400, Serdang, Selangor, Malaysia

ARTICLE INFO

Keywords:

Agricultural sciences
Crop breeding
Horticulture
Plant biology
Evolutionary biology
Biotechnology
Genetics
Molecular biology
Nephelium lappaceum
RAD sequencing
SSR marker

ABSTRACT

The trend of microsatellite marker discovery and development revolved as a result of the advancement of next generation sequencing (NGS) technology as it has developed numerous microsatellites within a short period of time at a low cost. This study generated microsatellite markers using RAD sequencing technologies for the understudied *Nephelium lappaceum*. A total of 1403 microsatellite markers were successfully designed, which consisted of 853 di-, 525 tri-, 17 tetra-, 5 penta-, and 3 hexanucleotide microsatellite markers. Subsequently, selection of 39 microsatellites was made for the evaluation of genetic diversity of the selected 22 rambutan varieties. Twelve microsatellites, which exhibited high call rates across the samples, were used to assess the diversity of the aforementioned rambutan varieties. The analysis of 12 microsatellites revealed the presence of 72 alleles and six alleles per locus in average. Furthermore, the polymorphic information content (PIC) value ranged from 0.326 (NlaSSR20) to 0.832 (NlaSSR32), which included an average of 0.629 per locus, while the generated Neighbour Joining dendrogram showed two major clusters. The pairwise genetic distance of shared alleles exhibited a range of values from 0.046 (R134↔R170) to 0.818 (R5↔R170), which suggested highest dissimilarity detected between R5 and R170. Notably, these research findings would be useful for varietal identification, proper management and conservation of the genetic resources, and exploitation and utilization in future breeding programs.

1. Introduction

Generally known as 'hairy litchi', rambutan (*Nephelium lappaceum* L.) falls under the Sapindaceae species, which consists of other fruit crops, such as pulasan (*Nephelium mutabile* Blume), longan (*Dimocarpus longan* Lour), and litchi (*Litchi chinensis* Sonn.). Although rambutan was believed to be a native to Malaysia and Indonesia, but now, it has proliferated throughout South East Asia, with Thailand being the most renowned country for producing rambutan. Nevertheless, Philippines, China, Australia, Malaysia, and Indonesia are also the centres of rambutan commercial production.

As a cross-pollinated crop, rambutan is dependent on insects for fruit set and pollination (Free, 1993). Furthermore, the scent of rambutan flowers strongly attracts various insects, including flies, butterflies, and bees. While naturally-grown rambutan trees have significant size and a height of 20 m, while the propagative clonal cultivars are less in size, with 4–7 m of height (Muhamed et al., 2019). Although some researches were conducted on the morphological and phonological characteristics of rambutan (e.g; Leaf and fruit shape and color), the molecular-based characterisation of rambutan remained limited (Muhamed and Kurien, 2018).

Also known as simple sequence repeats (SSR), microsatellites are the recurring motifs of 1–6 nucleotides, which spread around the genomes of

* Corresponding author.

E-mail address: shahrilf@mardi.gov.my (S. Ab Razak).

<https://doi.org/10.1016/j.heliyon.2020.e05077>

Received 14 May 2020; Received in revised form 18 July 2020; Accepted 23 September 2020

2405-8440/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

eukaryotes (Selkoe and Toonen, 2006). In comparison to other PCR-based methods, including ISSR, RAPD, SSRs, and AFLP, microsatellites represent a powerful marker due to their codominant inheritance, prevalent in the genome, multiallelic in nature which lead to high polymorphic. Microsatellite has been numerously used in phylogeography studies, population and parentage analysis, and genome mapping (Guichoux et al., 2011; Hodel et al., 2016; Qin et al., 2017). However, in Rambutan, the molecular based genetic study is extremely limited. Only a few studies were reported on diversity rambutan using ISSR and AFLP marker. Currently, provided that the species-specific microsatellites for the rambutan are still limited. To our knowledge, only 10 microsatellites for rambutan is available as developed by Xing'e et al., 2019. Hence, there is a needed to develop species-specific microsatellites for rambutan.

Conventionally, the establishment of SSR for non-model or understudied species requires the development of a genomic library and the identification of clones containing SSR through the Sanger sequencing, which may be pricy and taxing (Zalapa et al., 2012). However, the molecular marker development for microsatellites and single nucleotide polymorphism (SNP) marker evolve through the advancement in next

generation sequencing (NGS) technology. Furthermore, NGS enables development of numerous microsatellite loci with a least cost and time compared to conventional approach (Guichoux et al., 2011; Zalapa et al., 2012). Moreover, the restriction site-associated DNA sequencing (RAD-seq) is among the NGS-based approaches used for the large-scale identification of microsatellite and SNP markers (Baird et al., 2008; Bonatelli et al., 2015).

The theory of RAD sequencing follows the acquirement of the sequence parallel to particular restriction enzyme recognition sites and then sequence by NGS sequencer. Notably, the implementation of RAD sequencing-based Illumina platform contributes to the generation of comparatively long paired-end sequencing reads (100–150 bp). This platform is also cost-efficient and adequate for the development of microsatellites (Castoe et al., 2012; Wang et al., 2017). However, a reference genome is not required in this method and it enables the assembly of de-novo genome (Willing et al., 2011; Wang et al., 2017). Despite the high implementation of technologies based on NGS for SNP identification and genotyping, they are not the replacement for SSRs in population genetic research, especially the research with large sample

Table 1. List of 39 microsatellite markers for diversity assessment study.

SSR ID	Repeat motif	Forward sequence	Reverse sequence	Expected size (bp)	Accession no*
NlaSSR1	(AG) ²⁰	TGCAAGTGCATTTCATAAGAGC	TGTGTCAAGATCAGTTGTAGGG	223	MT385963
NlaSSR2	(CT) ¹⁹	ACACTGGCGTTCGGTGTTC	TGTTCCGGGAGCACATCGG	183	MT385964
NlaSSR3	(CT) ¹⁹	AGAAGTTGTGAGTCAACTGAGAC	AGATTTGGTACATACAGAGCAGG	240	MT385965
NlaSSR4	(CT) ¹⁹	AGAGATCAGTCCGGTGCC	AATTCGTGGGAAAGCGACC	346	MT385966
NlaSSR5	(AG) ¹⁷	AGTCTACAATTCGCCACAAAG	GTAACCTCAAGCAACCGCC	195	MT385967
NlaSSR6	(AG) ¹⁷	GCTTGGCAGAGAGGTTTG	GACGCTGGTACGGCGAG	186	MT385968
NlaSSR7	(AG) ¹⁷	ATTCAACAGGGCTCTGCC	CAGTGTCTCGTGTCTTACC	189	MT385969
NlaSSR8	(AG) ¹⁷	GCAAGAAATCGACACTTCAATCC	AAACTTGGCGCCGGTTCATC	162	MT385970
NlaSSR9	(GT) ¹⁷	TGATTGAGTTTGAACGTATCCAC	GCTAAACCTCAACGCCTGC	265	MT385971
NlaSSR10	(CTT) ¹⁶	ACGGTAGGTCTGGTTTGC	GAGACAGTCTGTAACGC	151	MT385972
NlaSSR11	(CT) ¹⁶	GAGCCTTGCTGACTTGCAG	GGTGGCTGTGGAACAACCTC	163	MT385973
NlaSSR12	(AG) ¹⁶	TCCAGTCTAAGAGCAGCAAATC	TGGTGGTGGTGCATCTG	213	MT385974
NlaSSR13	(AAT) ¹⁶	CTTCGAAATCTTATGCACCGC	CACCGAATCTGTTCGCC	158	MT385975
NlaSSR14	(AG) ¹⁶	GATGGTGTGTGTCAAGCTC	GGTGTGTGATGGCAGAAGCC	171	MT385976
NlaSSR15	(CT) ¹⁶	ACCTCTAAAGATCCCAAGAAGCTC	TGCAAGCAGAGTATGAACCC	378	MT385977
NlaSSR16	(AT) ¹⁵	AGCATGTGTAGAAGTTTGGCG	TGGATCCCAACAATGTGACAAG	152	MT385978
NlaSSR17	(CTT) ¹⁵	CAAGTCTGGTCAACGAGGTC	AACAGATGACAGCCTCGCC	187	MT385979
NlaSSR18	(AG) ¹⁵	TGTCCACATCATCTCCGGC	CTGGTGTGGAGACGACCC	192	MT385980
NlaSSR19	(GT) ¹⁵	GACTAACCTAGACACCGCG	CGTAGCACAAAGTTTAGGAAG	198	MT385981
NlaSSR20	(CT) ¹⁵	CCAGTTTGGATCCTGGACG	CTCCAACAAGGCCATGACAG	182	MT385982
NlaSSR21	(AG) ¹⁵	CCGACTGTAGTTTACTGGG	ACCAGGCCATCTTTATAGTCCC	187	MT385983
NlaSSR22	(AG) ¹⁵	CITTCATGGCGCTTCACCG	GAGACATAATCGTGAAGGCAG	204	MT385984
NlaSSR23	(AG) ¹⁵	GTTGGTGCAGTCAAGGCTC	GCTGACGACCCGACG	227	MT385985
NlaSSR24	(AT) ¹⁵	GCATCCAAC TAGAAGTCCCG	TGAGGAGACTGGAGCTTGC	398	MT385986
NlaSSR25	(CTT) ¹⁵	GGCGATGGTCGATCAAAGCTG	AAAGCTAAGGGACGAGACC	334	MT385987
NlaSSR26	(AG) ¹⁴	AGAAAATTGTGATGGTGGC	CCACATTCAAAATCCATCTGACG	157	MT385988
NlaSSR27	(CT) ¹⁴	TCTGGTATTTCCCAATCAGCTAC	GGTTGGGTAGTGGAAACCG	149	MT385989
NlaSSR28	(CT) ¹⁴	GGTACCAACAATGGAAGGAG	CTGCAGTGGGATAACTTTCTG	190	MT385990
NlaSSR29	(AG) ¹⁴	TGGTGGCTGGTCTCTCAAG	CATCGAAGCCGATTTGGAGC	170	MT385991
NlaSSR30	(AC) ¹⁴	TGTGACGTTCACTTGTGGC	ACACAGATCATTCTGCAAGCTG	155	MT385992
NlaSSR31	(AT) ¹⁴	TGTTCTGCATGTTTATGGC	TCTGTACGTTGGCAGCTTG	187	MT385993
NlaSSR32	(AT) ¹⁴	TCGGTGACTCTAGCAAAGAGG	GGTTCAAACCTTCGTGGAC	192	MT385994
NlaSSR33	(AT) ¹⁴	TGACTCATATAATCCCTACAAGGC	GCTGAGATCATGTAAGGTAAGAAC	211	MT385995
NlaSSR34	(CT) ¹⁴	GAGCAGGTGTAAGCCCAAG	AGAAAGAAGATTCGGGTTCC	153	MT385996
NlaSSR35	(AG) ¹⁴	GCTGGTCTGGGTTTGTATCG	TGGCATTCGATGGATTGGG	198	MT385997
NlaSSR36	(CT) ¹⁴	ACCGGGAAATCAACAGAGG	AGTAGAAATGGTATCACATTGAGG	167	MT385998
NlaSSR37	(AG) ¹⁴	GATGGTGAAGGCCAGAACAG	AGAGAGAGAGTTTCTTCGGTG	180	MT385999
NlaSSR38	(CT) ¹⁴	ATTTCCCTTCTTAGGCCCG	AAGGAGGTCGGCAATGGAG	150	MT386000
NlaSSR39	(AG) ¹⁴	GAGAGCATGGCGATCAAGTC	CTCACCATCGTTCCAACCC	194	MT386001

* Accessions number of DDBJ/NCBI database.

Table 2. Statistical summary of RAD sequencing data for microsatellites.

Clean bases	1,733,815,414
Clean reads	5,920,729
Total contigs	172,166
Contigs with repeats	6,563
Total of repeats found	7,535
Total of successfully design Primer	1,403
Number of selected SSR for varieties identification	39

Table 3. The number of repeat motifs found in this study.

Type of repeat	Number of repeats
Dinucleotide	853
Trinucleotide	525
Tetranucleotide	17
Pentanucleotide	5
Hexanucleotide	3

sizes and high associated price. Besides, the genetic structure developed by SSR was reflective of the finding obtained through the SNPs analysis (Elbers et al., 2017; Jeffries et al., 2016). For this reason, microsatellites are essential in population genetic research even though in the NGS era (Hodel et al., 2016; Qin et al., 2017).

2. Material and methods

2.1. RAD sequencing

Fresh leaf tissues of R134 were collected prior to DNA extraction. The extraction of the overall genomic DNA was performed using QIAquick DNeasy kit (Qiagen, Hilden, Germany) based on the manufacturer's guide. The high quality DNA (O.D > 1.8) was transferred to MyTACG Bioscience Enterprise (Petaling Jaya, Malaysia) for RAD sequencing according to the standard manufacturer's protocol, in which the total genomic DNA was digested with *MseI* before being sequenced Illumina HiSeq 2000 sequencer. Then, the raw data generated using Stacks (Catchen et al., 2013) were processed, which were used for de novo assembly using Velvet (Namiki et al., 2012).

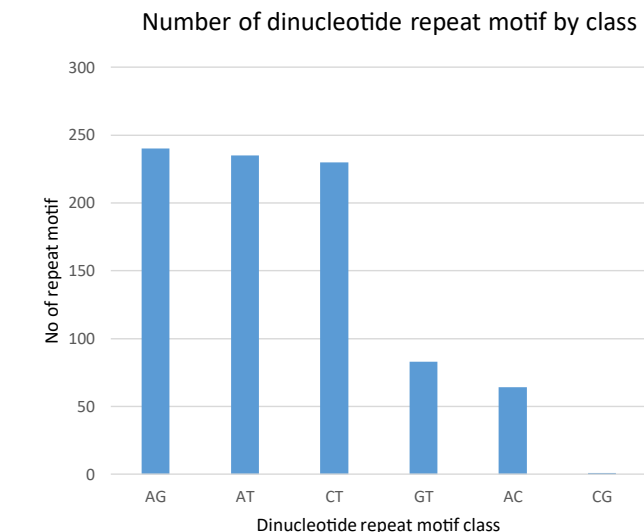


Figure 1. Distribution of dinucleotide repeat motif.

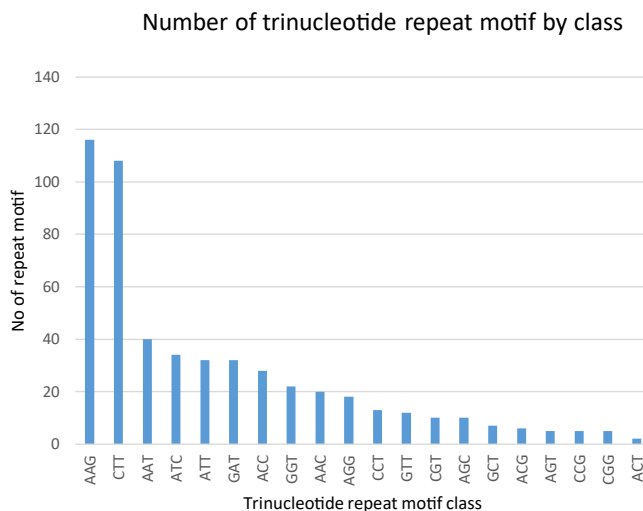


Figure 2. Distribution of trinucleotide repeat motif.

2.2. Microsatellite analysis and design

Microsatellites were mined using MISA, a Micro-SATellite identification tool for the screening of the assembled contigs for the repeat motifs of microsatellite (Beier et al., 2017). The search criteria consisted of the following parameters 6, 5, 4, and 3, with 3 repeated for the di-, tri-, tetra-, penta-, and hexa-repeat motifs, respectively. However, the mononucleotide repeats were omitted from the search criteria. The primer pairs targeting the repeat motif region were designed using the Primer3Plus software programme with the criterion of the amplicon size ranging from 100 and 250 bp. A total of 39 microsatellite markers were randomly selected for optimisation to evaluate the genetic diversity of rambutan clones (Table 1).

2.3. Diversity assessment using Novel Microsatellite Markers

Further implementation was made on the successfully optimised microsatellites to assess the genetic diversity of 22 rambutan clones (Table S1). The PCR was performed based on Schuelke (2000) by ligating the primers (either forward or reverse) and fluorescent dye (FAM, PET, NED or VIC) with M13 sequence (TGTAACAACGACGGCCAGT). The PCR cocktails of 10 µL volume consist of 1x buffer (Invitrogen, United States), 10 µM of every reverse and forward primer, 5 µM fluorescence-labelled

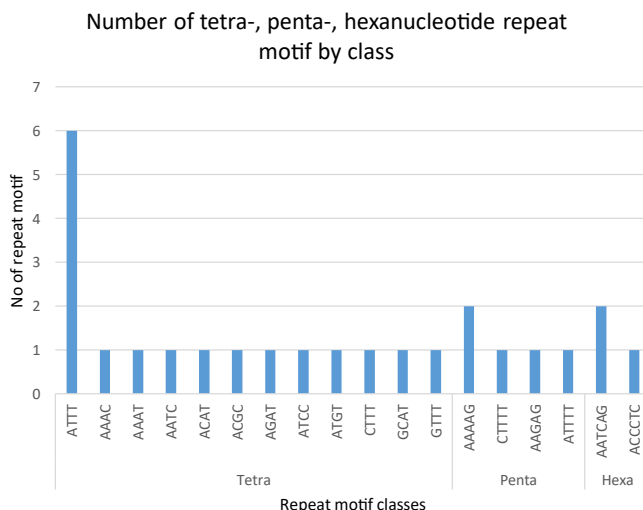


Figure 3. Distribution of tetra-, penta-, and hexanucleotide repeat motif.

Table 4. Characterisation of novel polymorphic microsatellite markers.

Marker	Ta	Major allele frequency	Allele no.	Gene diversity	Heterozygosity	PIC
NlaSSR1		0.294	9.000	0.811	0.809	0.786
NlaSSR3		0.331	6.000	0.770	0.593	0.734
NlaSSR5		0.382	5.000	0.709	0.647	0.660
NlaSSR7		0.492	7.000	0.647	0.541	0.591
NlaSSR12		0.517	5.000	0.668	0.633	0.630
NlaSSR20		0.750	3.000	0.385	0.206	0.326
NlaSSR21		0.728	5.000	0.442	0.368	0.411
NlaSSR23		0.618	7.000	0.568	0.485	0.529
NlaSSR27		0.331	5.000	0.766	0.794	0.728
NlaSSR31		0.259	6.000	0.804	0.839	0.775
NlaSSR32		0.257	10.000	0.849	0.824	0.832
NlaSSR36		0.566	4.000	0.601	0.838	0.548
Mean		0.460	6.000	0.668	0.631	0.629

M13 adaptor, 2 μM of each dNTP (Invitrogen, United States), 0.1 μL of bovine serum albumin (BSA) as PCR enhancer, and 1 U of Taq polymerase (Invitrogen, United States). A GeneAmp® PCR System 9700 (Applied Biosystems, United States) was used for amplification process. Pre-denaturation was conducted at 94 °C for two minutes on the PCR profile. Following that was 35 cycles at 94 °C for 30 s, 41–65 °C for 45 s, and 72 °C for 45 s, and a post-extension at 72 °C for five minutes.

After the amplification process, the PCR products were multiplexed up to four primers with a different combination of fluorescent dyes. This was followed by mixing the products with GeneScan 500 LIZ (standard molecular weight ladder) and Hi-Di formamide prior resolving the

products using an ABI 3130xL Genetic Analyser (Applied Biosystems, United States). GeneMapper Version 5 (Thermo Fisher Scientific, United States) was used to score the size of the allele, while the generated electropherograms were scored and analysed according to the description by Arif et al. (2010).

PowerMarker was used to calculate the number of alleles, the Major allele frequency, Gene Diversity, heterozygosity, and Polymorphism Information Content (PIC) of each microsatellite marker. The pairwise genetic distance based on shared allele also was generated using the same software. (Liu and Muse, 2005). MEGA7 was used to visualize the Neighbour Joining dendrogram, constructed by applying the generated

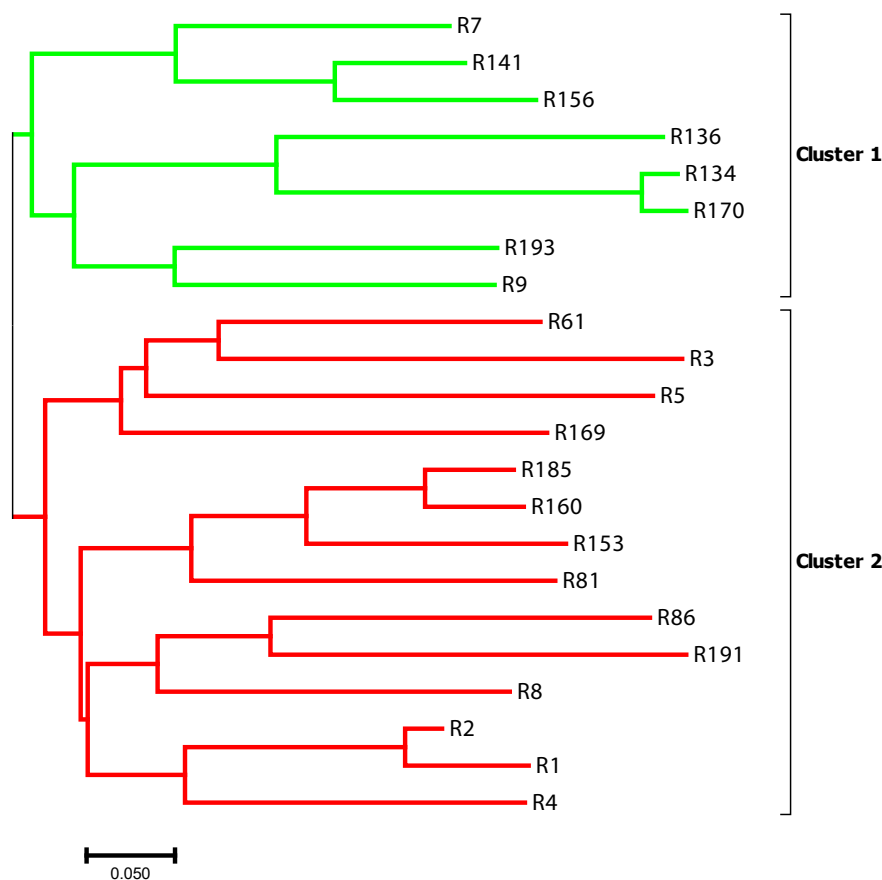


Figure 4. Neighbour Joining dendrogram of 22 rambutan varieties generated through the shared allele genetic distance using 12 polymorphic microsatellite markers.

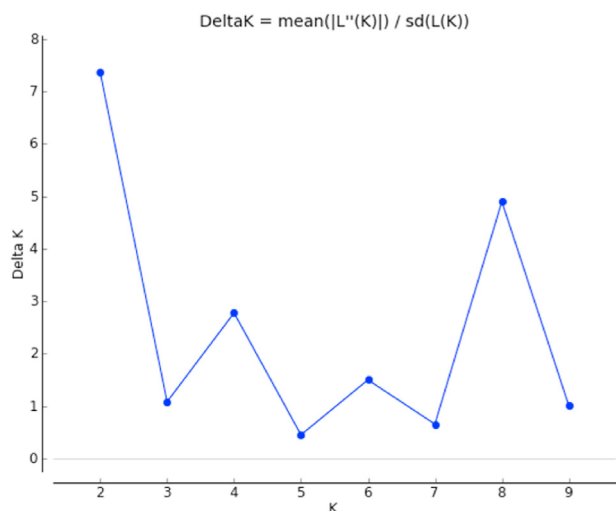


Figure 5. Estimation of the amount of K (population structure), where the maximum point of ΔK, ($\Delta K = m(|L''(K)|)/sd[L(K)]$), K = 2 was indicated by the amount of K.

genetic distance matrices of the shared alleles (Kumar et al., 2016). STRUCTURE software was used to determine the structure group of studied varieties. It involved the calculation of the K values by varying the K values from 1 to 10 with 10 independent runs per K value, with a 100,000 Markov chain Monte Carlo (MCMC) repetitions and 100,000 burn-in period (Pritchard et al., 2000). Then, the optimal value of K was determined using web-based STRUCTURE HARVESTER (Earl and von-Holdt, 2012) using the formula as described by Evanno et al. (2005) $\Delta K = \text{mean}(|L''(K)|)/sd[L(K)]$. The PCoA and AMOVA analysis was performed using GenAlEx 6 (Peakall and Smouse, 2006).

3. Results

3.1. RAD-seq sequencing and microsatellite analysis

In this study, 5.9 million of clean reads were obtained. These reads were then assembled into 172,166 contigs using Velvet (Table 2). Overall, 6,563 contigs were found to have a total of 7,535 repeats. Out of these, only 1403 repeats were successfully designed their flanking primer in which 853 of the repeat motifs (60.8%) were dinucleotide, 525

(37.4%) were trinucleotide, 17 (1.2%) were tetranucleotide, five (0.4%) were pentanucleotide, and three (0.2%) were hexanucleotide (refer to Table 3). Notably, the highest number of dinucleotide repeat motifs class was for AG, which consisted of 240 repeats, while the lowest number of dinucleotide repeat motifs was for CG with only one repeat found. Furthermore, a total of six types of dinucleotide repeat motifs (AG > AT > CT > GT > AC > CG) were found in this study (Figure 1). Specifically, a total of 20 types of trinucleotide repeat motifs were found, where the AAG repeat motif contained the highest number of trinucleotide repeat motif with 116 repeats recorded (22.1%), while the lowest number of repeat motif for trinucleotide was ACT with only two repeats identified (Figure 2). The distribution of the tetra-, penta- and hexanucleotide repeat motifs were described in Figure 3.

3.2. Characterisation of novel microsatellite markers.

Based on the genotyping of 39 microsatellites on 22 different rambutan clones, only twelve microsatellites were successfully optimized, polymorphic and call rates above 90% across the total samples. The remaining microsatellites were either monomorphic, exhibited low call rates, produced multiple bands, or appeared with a stutter band across the local rambutan clones. Furthermore, an analysis of 12 microsatellites revealed a total of 72 alleles, which ranged from three (NlaSSR20) to 10 (NlaSSR32) alleles per locus and a mean of six alleles per locus. The range of the polymorphic information content (PIC) value was from 0.326 (NlaSSR20) and 0.832 (NlaSSR32) with an average of 0.629, while the value of gene diversity and heterozygosity ranged from 0.385 (NlaSSR20) to 0.206 (NlaSSR20) and 0.849 (NlaSSR32) to 0.839 (NlaSSR31), respectively. The characterisation of the novel polymorphic microsatellite markers is summarised in Table 4.

3.3. Genetic diversity of rambutan clones using Novel Microsatellite Markers

Two clusters were revealed through the neighbour joining dendrogram tree, namely Cluster 1 and Cluster 2 (Figure 4). Specifically, Cluster 1 consisted of eight clones, namely R7, R9, R134, R136, R141, R156,

Table 5. Percentage of variation explained by the first three axes.

Axis	1	2	3
%	14.81	13.27	10.70
Cumulative %	14.81	28.08	38.78

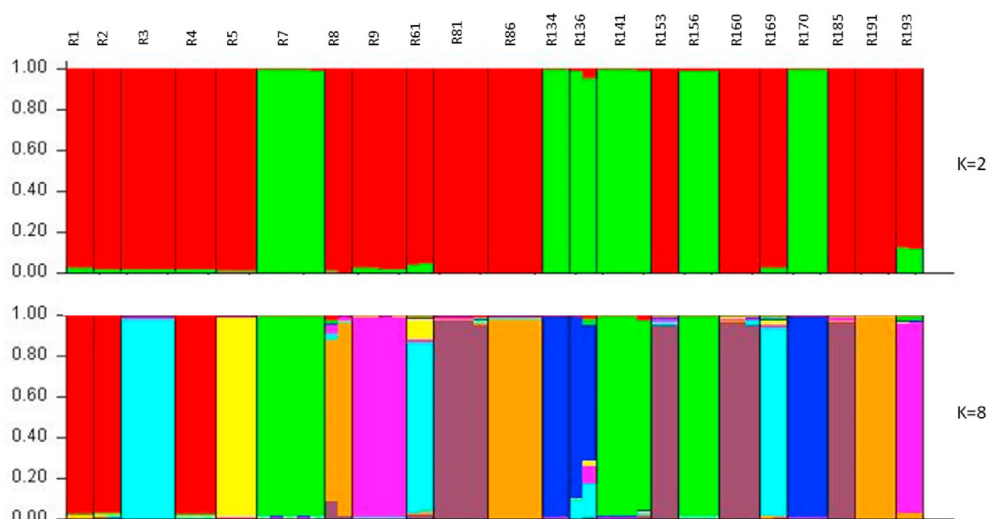


Figure 6. The structure diagram generated by the STRUCTURE programme according to K = 2 (above) and when K = 8 (below).

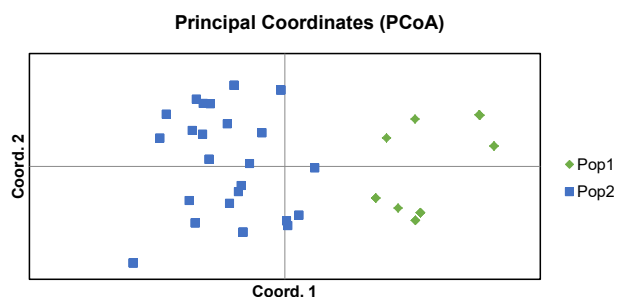


Figure 7. The Principle Component Analysis (PCoA) of the rambutan clones based on 12 polymorphic microsatellite markers through a model-based approach. Pop 1 represents a green structure group, while Pop 2 represents a red structure group.

R170 and R193, while the remaining clones were grouped in Cluster 2. It was suggested from the dendrogram analysis that all rambutan clones were differentiated from each other using twelve novel polymorphic microsatellites. Furthermore, it was found, from the calculated pairwise genetic distance values based on shared alleles showed that the value ranged from 0.0455 (R134↔R170) to 0.8182 (R5↔R170). The details of the pairwise genetic distance were described in Table S2. This analysis was reinforced by the results from the STRUCTURE analysis, which recorded the highest ΔK value at $K = 2$ according to Evanno et al. (2005) and the second highest ΔK value at $K = 8$ (Figure 5). Unlike dendrogram analysis, when $K = 2$, only 6 clones namely R7, R134, R136, R141, R156 and R170 were grouped in Green Group while R9 and R193 were clustered together with the remaining varieties in Red Group. Notably, when $K = 8$, six subgroups were derived from Red Group while the remaining two subgroups were derived from Green Group (Figure 6). No clear morphological or phenotypical similarity or difference was found between the clones in the same cluster. However, we found that all the yellow skin varieties (R2, R5, R61 and R153) were clustered together in Cluster 2 except R156. Besides, variety of R134 and R170 which scored the highest pairwise similarity with the genetic distance of 0.0455 showed several similarities such as red skin colour and globose in seed shape. Besides, the varieties with least similarities also differed in term of fruit skin colour and seed shape as the R5 is yellow skin colour with oblong seed shape meanwhile R170 is red skin colour with globose seed shape.

A model-based method acquired from the STRUCTURE analysis was implemented to perform the PCoA and AMOVA analysis using Genalex. As a result, highly diverse genetic was found among the investigated rambutan varieties. Table 5 presents the first three axes, which represented 38.78% of the cumulative variation. Figure 7 indicates the clear classification of the green and red groups in the PCoA, which were analysed by AMOVA through a model-based approach. As a result, it could be seen in Table 6 that 14% variance was observed from these two groups, while 6% variance was observed among the individuals, and 80% variance was observed from the individual. Overall, these values indicated a significant genetic diversity between individuals, leading to highly diverse genetics between the varieties in these two groups.

4. Discussion

This study aims towards a cost-effective and rapid development of microsatellite marker for non-model plant species, *Nephelium lappaceum*,

through the RAD sequencing based NGS technology. This approach allowed the generation of an adequate length of assembled sequence to design microsatellite markers. In contrast to the conventional selective hybridisation approach for microsatellite development, this approach has a higher effectiveness in producing a high amount of microsatellite markers for non-model or understudied species. This feature contributes to a responsive and cost-effective development of a hundred to thousands of microsatellites (Davey et al., 2011; Tian et al., 2016). Furthermore, the RAD sequencing approach is more efficient in terms of time and cost compared to the conventional selective hybridisation technique for microsatellite development.

The selective hybridisation conventional approach requires approximately two to four weeks to perform a tedious laboratory work for the development of a low or limited amount of microsatellite markers. Meanwhile, the current approach only requires one to two weeks to obtain a high number of microsatellite markers (Zane et al., 2002). Notably, the microsatellite development using NGS was investigated in numerous works of research (Bonatelli et al., 2015; Yang et al., 2016). For instance, Yang et al. (2016) found that the loci of 650 microsatellites was identified through 4.5 million raw RAD reads, with a successful development taken place on 285 (43.84%) primer pairs. In this research, 7,535 repeats were found in 172,166 contigs, and 1,403 primer pairs were designed successfully. Badri et al. (2014) acquired 350 white colonies through the conventional selective hybridisation approach, with only 52 of the colonies shown to have unique SSR regions for sesame. Similarly, Paliwal et al. (2016) applied the conventional approach to develop microsatellite markers and obtain 356 white colonies and 114 (32.02%) colonies with SSR repeats. Overall, this research indicated that the RAD-seq approach for microsatellite development exhibits higher efficiency in terms of time, cost, and labour.

4.1. Genetic diversity of rambutan varieties

The assessment of crop diversity provides an insight to plant breeders regarding the development of cultivars. An accurate information is required in order to exploit the genetic diversity for developing new cultivars. Following that, the intended features or characteristics could be acquired through breeding programmes while retaining the genetic variability at the same time. Essentially, rambutan is a cross-pollinated crop, which pollination and fruit set rely on insects, particularly honeybees (Free, 1993). Provided that the cross-pollination of rambutan leads to high genetic variability, the propagation of seedlings constantly performs through grafting compared to direct seed cultivation as this process would result in uncharacterised or unforeseen progenies.

Conventionally, the characterisation of rambutan is based on the morphological characterisation of its leaves, flowers, fruit, and seeds among others. Although the use of the morphological characterisation for the cultivar identification remains significant, the use of morphological characteristics often results in conflicting outcomes due to environmental factors and agronomic practices. Besides, the approach also often limited by the number of different characteristics. As a result, molecular markers offer a more reliable option compared to morphological characterisation approach. Additionally, as one of molecular marker types, microsatellite is highly useful for various genetic purposes, including diversity assessment, cultivar identification through DNA fingerprinting, linkage and QTL analysis (Kalia et al., 2011).

Table 6. Summarised details of the AMOVA analysis.

Source	df	SS	MS	Est. Var.	%
Among groups	1	39.548	39.548	0.627	14%
Among individuals	66	273.158	4.139	0.264	6%
Within individual	68	245.500	3.610	3.610	80%
Total	135	558.206		4.502	100%

The conservation of rambutan requires identification, characterisation, and evaluation of the genotypes for efficient and practical breeding procedures and suitable conservation management programs. Specific methodologies and techniques have been applied worldwide, such as biochemical, morphological, and molecular-based identification of different plant species. Notably, The molecular based identification prove to be the best and ideal method as the application offered an accurate and consistent approach in identifying and characterizing plant varieties. Last but not least, the data acquired in this research contributes to a significant insight for breeders to properly design a breeding programme to obtain genetic heterosis.

Declarations

Author contribution statement

Shahril Ab Razak:Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Salehudin Mad Radzuan:Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Norkhairi Mohamed:Contributed reagents, materials, analysis tools or data.

Nor Helwa Ezzah Nor Azman, Alny Marlynni Abd Majid, Siti Norhayati Ismail: Performed the experiments.

Muhammad Fairuz Mohd Yusof: Analyzed and interpreted the data.

Johari Sarip, Khairun Hisam Nasir:Conceived and designed the experiments.

Funding statement

This work was supported by Malaysian government under NKEA EPP14 project (KRBNA1).

Competing interest statement

The authors declare no conflict of interest.

Additional information

Supplementary content related to this article has been published online at <https://doi.org/10.1016/j.heliyon.2020.e05077>.

Acknowledgements

The author would like to thank all parties who are indirectly involved in this project.

References

- Arif, I.A., Khan, H.A., Shobrak, M., Al Homaidan, A.A., Al Sadoon, M., Al Farhan, A.H., Bahkali, A.H., 2010. Interpretation of electrophoretograms of seven microsatellite loci to determine the genetic diversity of the Arabian Oryx. *Genet. Mol. Res.* 9 (1), 259–265.
- Badri, J., Yepuri, V., Ghanta, A., Siva, S., Siddiq, E.A., 2014. Development of microsatellite markers in sesame (*Sesamum indicum* L.). *Turk. J. Agric. For.* 38 (5), 603–614.
- Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One* 3 (10), 1–7.
- Beier, S., Thiel, T., Münch, T., Scholz, U., Mascher, M., 2017. MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33 (16), 2583–2585.
- Bonattelli, I.A.S., Carstens, B.C., Moraes, E.M., 2015. Using next generation RAD sequencing to isolate multispecies microsatellites for *Pilosocereus* (Cactaceae). *PLoS One* 10 (11), 1–14.

- Castoe, T.A., Poole, A.W., de Koning, A.P.J., Jones, K.L., Tomback, D.F., Oyler-McCance, S.J., Fike, J.A., Lance, S.L., Streicher, J.W., Smith, E.N., 2012. Rapid microsatellite identification from Illumina paired-end genomic sequencing in two birds and a snake. *PLoS One* 7 (2), 1–10.
- Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22 (11), 3124–3140.
- Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L., 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12 (7), 499–510.
- Earl, D.A., vonHoldt, B.M., 2012. Structure HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Res.* 4 (2), 359–361.
- Elbers, J.P., Clostio, R.W., Taylor, S.S., 2017. Population genetic inferences using immune gene SNPs mirror patterns inferred by microsatellites. *Mol. Ecol. Res.* 17 (3), 481–491.
- Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14 (8), 2611–2620.
- Free, J.B., 1993. *Insect Pollination of Crops*. Academic press.
- Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Léger, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F., 2011. Current trends in microsatellite genotyping. *Mol. Ecol. Res.* 11 (4), 591–611.
- Hodel, R.G.J., Segovia-Salcedo, M.C., Landis, J.B., Crowl, A.A., Sun, M., Liu, X., Gitzendanner, M.A., Douglas, N.A., Germain-Aubrey, C.C., Chen, S., 2016. The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century. *Appl. Plant Sci.* 4 (6), 1600025.
- Jeffries, D.L., Copp, G.H., Lawson Handley, L., Olsén, K.H., Sayer, C.D., Hänfling, B., 2016. Comparing RAD seq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, *Carassius carassius*, L. *Mol. Ecol.* 25 (13), 2997–3018.
- Kalia, R.K., Rai, M.K., Kalia, S., Singh, R., Dhawan, A.K., 2011. Microsatellite markers: an overview of the recent progress in plants. *Euphytica* 177 (3), 309–334.
- Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33 (7), 1870–1874.
- Liu, K., Muse, S.V., 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21 (9), 2128–2129.
- Muhamed, S., Kurien, S., 2018. Phenophases of rambutan (*Nephelium lappaceum* L.) based on extended BBCH-scale for Kerala, India. *Curr. Plant Biol.* 13, 37–44.
- Muhamed, S., Kurien, S., Iyer, K.S., Remzeena, A., Thomas, S., 2019. Natural diversity of rambutan (*Nephelium lappaceum* L.) in Kerala, India. *Genet. Resour. Crop Evol.* 66 (5), 1073–1090.
- Namiki, T., Hachiya, T., Tanaka, H., Sakakibara, Y., 2012. MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res.* 40 (20), e155–e155.
- Paliwal, R., Kumar, R., Choudhury, D.R., Singh, A.K., Kumar, S., Kumar, A., Bhatt, K.C., Singh, R., Mahato, A.K., Singh, N.K., 2016. Development of genomic simple sequence repeats (g-SSR) markers in *Tinospora cordifolia* and their application in diversity analyses. *Plant Genet.* 5, 118–125.
- Peakall, R., Smouse, P.E., 2006. Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6 (1), 288–295.
- Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155 (2), 945–959. <https://www.genetics.org/content/genetics/155/2/945.full.pdf>.
- Qin, H., Yang, G., Provan, J., Liu, J., Gao, L., 2017. Using MiddRAD-seq data to develop polymorphic microsatellite markers for an endangered yew species. *Plant Divers.* 39 (5), 294–299.
- Schuelke, M., 2000. An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* 18 (2), 233–234.
- Selkoe, K.A., Toonen, R.J., 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecol. Lett.* 9 (5), 615–629.
- Tian, Z., Zhang, F., Liu, H., Gao, Q., Chen, S., 2016. Development of SSR Markers for a Tibetan medicinal plant, *Lancea tibetica* (Phrymaceae), based on RAD sequencing. *Appl. Plant Sci.* 4 (11), 1600076.
- Wang, B., Xie, X., Liu, S., Wang, X., Pang, H., Liu, Y., 2017. Development and characterization of novel microsatellite markers for the Common Pheasant (*Phasianus colchicus*) using RAD-seq. *Avian Res.* 8 (1), 4.
- Willing, E.M., Hoffmann, M., Klein, J.D., Weigel, D., Dreyer, C., 2011. Paired-end RAD-seq for de novo assembly and marker design without available reference. *Bioinformatics* 27 (16), 2187–2193.
- Xing'e, L., Junhai, N., Ying, C., Jianhong, M., Hongmao, G., Yu, G., Zhaoxi, Z., 2019. Construction of DNA fingerprinting with SSR markers for 68 rambutan (*Nephelium lappaceum* L.) accessions. *Chin. J. Top. Crops* 40 (4), 708–714.
- Yang, G.-Q., Chen, Y.M., Wang, J.P., Guo, C., Zhao, L., Wang, X.Y., Guo, Y., Li, L., Li, D.Z., Guo, Z.H., 2016. Development of a universal and simplified ddRAD library preparation approach for SNP discovery and genotyping in angiosperm plants. *Plant Methods* 12 (1), 39.
- Zalapa, J.E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., McCown, B., Harbut, R., Simon, P., 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *Am. J. Bot.* 99 (2), 193–208.
- Zane, L., Bargelloni, L., Patarnello, T., 2002. Strategies for microsatellite isolation: a review. *Mol. Ecol.* 11 (1), 1–16.