TRANSCRIPTOME SEQUENCING OF *LEPISANTHES FRUTICOSA* TO DISCOVER SSR MARKERS

Zulkifli Ahmad Seman [1]✉, Azrin Ahmad [2]✉, Rabiatul Adawiah Zainal Abidin [1], Siti Zainab Jantan [1]✉, Mohd Hanif Azhari Noor [1]✉, Yun Shin Sew [1]✉, Mohd Norfaizal Ghazalli [3]✉, Khairun Hisam Nasir [1]✉, Sanimah Simoh [4]✉, Mohd Shukri Mat Ali [5]✉

[1] Research Officer, Biotechnology and Nanotechnology Research Centre, MARDI Headquarters, Persiaran MARDI-UPM 43400 Serdang, Selangor Malaysia.
[2] Scientist, Malaysia Genome and Vaccine Institute (MGVI), National Institute of Biotechnology Malaysia (NIBM), Jalan Bangi,43000 Kajang, Selangor, Malaysia.
[3] Deputy Director, Bio-Agrodiversity and Environment Research Centre, MARDI Headquarters, Persiaran MARDI-UPM 43400 Serdang, Selangor Malaysia.
[4] Director, Strategic Planning and Innovation Management Centre, MARDI Headquarters, Persiaran MARDI-UPM 43400 Serdang, Selangor Malaysia
[5] Director, Bio-Agrodiversity and Environment Research Centre, MARDI Headquarters, Persiaran MARDI-UPM 43400 Serdang, Selangor Malaysia
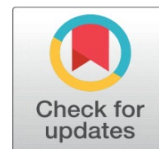
ABSTRACT

Lepisanthes fruticosa (ceri Terengganu) is one of the important underutilized fruit plants with high value of bioactive compounds and pharmacological properties. Current studies have focused mainly on the bioactive compounds which are essential for functional food and pharmaceutical applications. However, studies on the diversity and conservation of L. fruticosa are still scarce since genomic and genetic resources for this plant species are still lacking. In this study, RNA sequencing of L. fruticosa leaf was carried out using Illumina HiSeq to identify potential unigenes and simple sequence repeats (SSRs). A total of 52,657 unigenes were identified from about 91,043,356 million raw sequence reads. Mining of SSRs from these unigenes have predicted a total of 23,958 SSRs which was approximately 45.58% of total unigenes obtained. Dinucleotide repeats motif was the highest (21.48%) and the next were trinucleotide repeats motif (14.65%). A total of 4,620 SSRs ranging from 12 to 116 bp were selected for experimental validation. Bioinformatic analysis via GO and KEGG platforms showed that a total of 1,861 (40.28%) SSR-containing unigenes matched to Gene Ontology (GO) terminology and 48 biochemical pathways. The SSR-containing unigenes of L. fruticosa were involved in various cell functions and a majority of their functions were associated with purine and thiamine metabolism. In addition. A majority of SSR-containing unigenes were involved in organic and heterocylic compounds bindings, indicating an active event of biosynthesis process of secondary metabolites in L. fruticosa. SSR markers obtained from this study provides new genetic information that can be utilized to facilitate future characterization of L. fruticosa accessions at molecular levels.

Keywords: EST-SSRs, Lepisanthes Fruticosa, Transcriptome Sequencing

1. INTRODUCTION

There is multiplicity of underutilized fruits which are natively grown at the regions of Peninsular Malaysia, Sabah and Sarawak. The plants bear less attractive fruits compared to commercial plant species, however many of them

have high nutritional value and medicinal properties Rizvi et al. (2015). Ibrahim et al. (2010) has shown the important of native fruits to be traditionally used as a medicine to treat several common diseases. However, narrow down of research has been focused on different part of plants including fruit to further scientifically studied of their medicinal benefits. Research has shown that the bioactive compounds, carotenoid, and other terpenoids are the primary contributors to compounds identified from the plants extracts which include phenolic their medicinal properties.

Lepisanthes fruticosa or locally known as ceri Terengganu is one of the valuable underutilized fruits in Malaysia with the potential to be exploited for commercial production. *L. fruticosa* is a non-seasonal woody plant of which the fruits are available throughout the year. Typically, the fruits are arranged closely and attractively in a big bunch or cluster (20 fruits/bunch). The flesh is soft and fairly sweet taste with 1-3 seeds/fruit. The tree is small but can reach medium height with spreading out canopy. The purplish colour of the young leaves adds to the attractiveness of the tree Mirfat et al. (2017). Studies have shown that *L. fruticosa* ripe fruits contain the highest free radical scavenging and total phenolic contents compared to numbers of underutilized and commercial fruits Mirfat and Salma (2015). This suggests the fruit of *L. fruticose* is a good candidate for alternative medicine and health benefiting food supplement. Umikalsum and Mirfat (2014), Dayang et al. (2012), Ibrahim et al. (2010), Ikram et al. (2009). Nevertheless, notwithstanding the rich genetic diversity of *L. fruticosa,* there are limited reports on germplasm diversity and molecular markers data. For instance, the nucleotide sequences of *L. fruticosa* deposited in NCBI database (https://www.ncbi.nlm.nih.gov/gquery/?term=Lepisanthes+alata+) were found to be scarce (as little as nucleotide sequences as of March 2018).

Large-scale sequencing data possibly be generated from both genome and transcriptome via recent advance in sequencing technology. Likewise, this next-generation sequencing technologies and bioinformatics analysis has led to large-scale identification of EST-SSRs from various crops Wang et al. (2010), Garg et al. (2011), Zeng et al. (2010), Zhou et al. (2016)). Simple sequence repeats (SSRs) represent arrays of short motifs which are characterized based on their hypervariability, abundance, reproducibility, Mendelian inheritance, codominant nature Scott et al. (2000), Gupta et al. (1996) and convenient to be applied as compared to the molecular markers Zhou et al. (2016). SSRs can be either predicted from genomic or transcriptome which are known as genomics SSRs and EST-SSRs, respectively Song et al. (2012). While EST-SSRs are derived from expressed sequence tags and these types of SSRs are more evolutionary conserved compared to genomic SSRs derived from noncoding sequences with relatively high transferability Wei et al. (2011b). SSRs is powerful tool that have been extensively used in population study to determine genetic diversity and also to analyze genetic structure Yoichi et al. (2017). The present study aimed to generate and identify EST-SSRs from leaf of *L. fruticosa* using Illumina paired end sequencing technology. Genic SSRs markers (in genic sequences) will then be characterized based on their frequency and distribution followed by analysis the functional properties of those SSRs. This data and results obtained from the present study will be a valuable genomic and genetic resources for future studies of *L. fruticosa*.

Zulkifli Ahmad Seman, Azrin Ahmad, Rabiatul Adawiah Zainal Abidin, Siti Zainab Jantan, Mohd Hanif Azhari Noor, Sew Yun Shin, Mohd Norfaizal Ghazalli, Khairun Hisam Nasir, Sanimah Simoh, and Mohd Shukri Mat Ali

2. MATERIALS AND METHODS
2.1. PLANT MATERIAL

The plants of *L. fruticosa* were grown at MARDI's germplasm located at MARDI headquarter, Serdang, Malaysia. Harvested young leaves of *L. fruticosa* were snap-frozen in liquid nitrogen and kept in -80°C freezer for further use.

2.2. RNA EXTRACTION FOR ILLUMINA SEQUENCING

A total of 100 mg of young leaves tissue were powdered in liquid nitrogen prior to RNA extraction using TRI Reagent (SIGMA-Aldrich, St. Louis, USA) according to provide instruction. Extracted RNA was treated with RNase-free DNase I Recombinant (QIAGEN, USA) to prevent sample from genomic DNA contamination. The quantification of extracted total RNA was performed using Nanodrop ND-100 spectrophotometer (Thermo Scientific, Wilmington, USA). The integrity of RNA bands was checked with 1% TAE agarose gel electrophoresis. The quality and RNA integrity number (RIN) of total RNA was assessed using a 2100 Bioanalyzer (Agilent Technology, Santa Clara, CA, USA). Illumina sequencing was performed at Novogen Co., Ltd. Beijing, China, the cDNA libraries were sequenced using Illumina HiSeq™ 2500 system under effective concentration.

2.3. SEQUENCE PREPROCESSING AND *DE NOVO* ASSEMBLY

The raw reads of sequencing data were filtered to generate high quality data. This includes removing adaptor contaminants, removing reads with more than 10% of uncertainty nucleotides and low-quality base reads with a cut-off value of Phred score, q <= 20. The cleaned reads were then assembled using Trinity version r20140413p1, with parameter minimum kmer coverage =2 and other parameters were by default. *De novo* transcripts assembly Trinity was then clustered using Corset version 1.05 to obtain unigene sequences.

2.4. SSR MARKERS DISCOVERY

SSR markers were discovered using MISA (MIcroSAtellite) (http://pgrc.ipk-gatersleben.de/misa/). Parameters chosen for detection of a SSR motif with minimum length of 12 base pairs (bp) and repeat length of mono-10, di-6, tri-5, tetra-5, penta-5 and hexa-5. The maximum size of interruption allowed between two different SSRs in a compound SSR was 100 bp. The SSRs filtering was performed using a custom Perl script and R programming. Criteria used in the SSRs filtering were choosing the markers which represent more than two alleles and in a single contig. Approximately 200 bp of each side of repeat motif region were extracted using bedtools version 2.21.0 (Quinlan & Hall 2010). SSR markers were annotated using BLAST program (blastx) against non-redundant (nr) and SwissProt protein databases with E-value of $1x10^{-5}$. Gene ontology (GO) enrichment and KEGG pathway analyses were conducted using BLAST2GO software.

3. RESULTS
3.1. DE NOVO ASSEMBLY OF TRANSCRIPTOMIC SEQUENCING DATA

A total of 91,043,356 paired end raw reads were generated from transcriptome sequencing of *L. fruticosa* leaf. Approximately, the total reads comprise of approximately 13.66 Gigabase (Gb) with an average read length of 150 bp were generated using Illumina HiSeq™ 2500 sequencer. After reads quality assessment, approximately 89,441,736 (98.24%) of high-quality data were recovered (Table 1). This cleaned pair-end reads were then assembled into 52,657 transcripts. Clustering of transcripts resulted in 52,569 unigenes (Table 2).

Table 1 Statistic of data quality control

Type of Data	Total Reads	Total Bases (Base Pair)	Average Read Length (Base Pair)
Raw data[a]	91,043,356	13,656,503400	150
Clean data[b]	89,441,736	13,416,260,400	150
Percentage of data quality (%) [c]	98.24	98.24	100

[a] original raw reads from sequencer Illumina HiSeq™ 2500

[b] Reads after quality filtering.

[c] Figure obtained based on,

$$\frac{\text{Total reads of clean data}}{\text{Total reads of raw data}} \times 100\%$$

Table 2 Statistic of de novo L. fructicosa transcriptome assembly

	Transcripts[a]	Unigenes[b]
Number of transcripts/unigenes	52,657	52,569
Total bases transcripts/unigenes (base pair)	77,451,424	77,428,927
Range of length transcripts/unigenes (base pair)	201-17,572	201-17,572
Average of length transcripts/unigenes (base pair)	1,471	1,473
Length of N50 transcripts/unigenes (base pair)	2,136	2,136

A total of 23,958 of SSRs were identified which was accounted for 45.58% of the total unigenes. All the markers were classified into seven types, consisting of type c, type p1 (mono), p2 (di-), p3 (tri-), p4 (tetra-), p5 (penta-) and p6 (hexa-) nucleotide repeats (Table 3).

Table 3 Occurrence of SSRs in L. fructicosa transcriptome

SSRs repeat type	Total number
SR markers type c	1,941
SSR markers type p1	13,179
SSR markers type p2	5,146
SSR markers type p3	3,510
SSR markers type p4	113
SSR markers type p5	11
SSR markers type p6	12
Total	23,958

Mononucleotide repeats were the most abundant (55.01%) in the *L. fruticosa* transcriptome, followed by dinucleotide repeats (21.48%) and trinucleotide repeats (14.65%). Among them, the most frequent repeat motifs were AT/AG (9%), followed by TC/GA/CT (7%), TA (5%), TG/AC (2%), CA and GAAAA (1%) (Table 4). A total of 4,620 SSRs with the motif size ranging from 12 to 116 bp were selected as potential SSRs for genotyping validation. The analysis was not included mononucleotide repeats and compound formation.

Table 4 Summary of putative SSRs in L. fruticose transcriptome

SSRs repeat type	Total number	Total length (base pair)	Most frequent repeat motif
SSR markers type p2	2,305	33,746	AT/AG
SSR markers type p3	2,251	36,768	GAA/TTC
SSR markers type p4	51	1,108	CTCA
SSR markers type p5	10	250	GAAAA
SSR markers type p6	3	90	AAAAGA/GCTGGT/GGGCAA
Total	4,620	71,962	

3.3. FUNCTIONAL ANNOTATION OF SSRS UNIGENE

Following BlastX homology search analysis, only 954 (20.65%) of 4,620 unigenes of *L. fruticosa* returned significant hits to the NCBI non-redundant (nr) database with 10^{-5} E-value (Table 5). Based on the similarity search analysis, sequences were found to be homologous to *Citrus sinensis* (197, 34%), *Citrus clementina* (113, 20%), *Theobroma cacao* (63, 11%), *Hevea brasiliensis* (49, 9%), *Cephalotus follicularis* (32, 6%) and *Corchorus capsularis* (32, 6%) (Figure 1). These results indicated that *L. fruticosa* has high similarity with other perennial woody tree plants.

Table 5 Homology search of L. fruticosa transcriptome against NCBI nr database

	Number of SSRs
With BLAST hits[a]	954
Without BLAST hits	3,666
Total of sequences	4,620

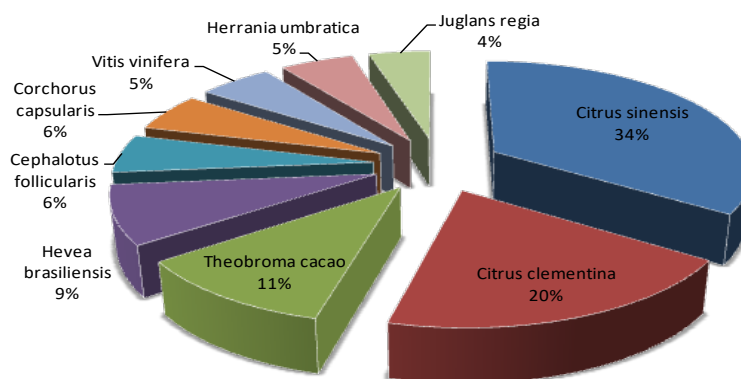[a] based on an E-value cutoff of 1x10-5

Figure 1 Top hit species classification upon homology searches of SSR against nr database

Gene ontology analysis on SSRs associated unigenes was performed in order to gain an understanding on the biological mechanisms involved in leaf of *L. fruticosa*. Results showed that a total of 52,569 unigenes were annotated, of which 1,861 (40.28%) unigenes represent identified SSRs. GO annotation of SSR- containing unigenes were classified into three categories: biological process (613), cellular component (747) and molecular function (501), respectively. For biological process, the SSR-containing unigenes were assigned into top 5 categories were organic substance metabolic process (18%), primary metabolic process (18%), cellular metabolic process (17%), nitrogen compound metabolic process (11%) and biosynthetic process (9%). For cellular component, the intracellular (23%), intracellular part (20%), intracellular organelle (16%), intrinsic component of membrane (16%) and membrane-bounded organelle (14%) were those with higher abundancy. While most of the SSR-containing unigenes were involved in the heterocyclic compound binding (21%), organic cyclic compound binding (21%), ion binding (17%), small molecule binding (11%) and carbohydrate derivative binding (9%) (Figure 2).
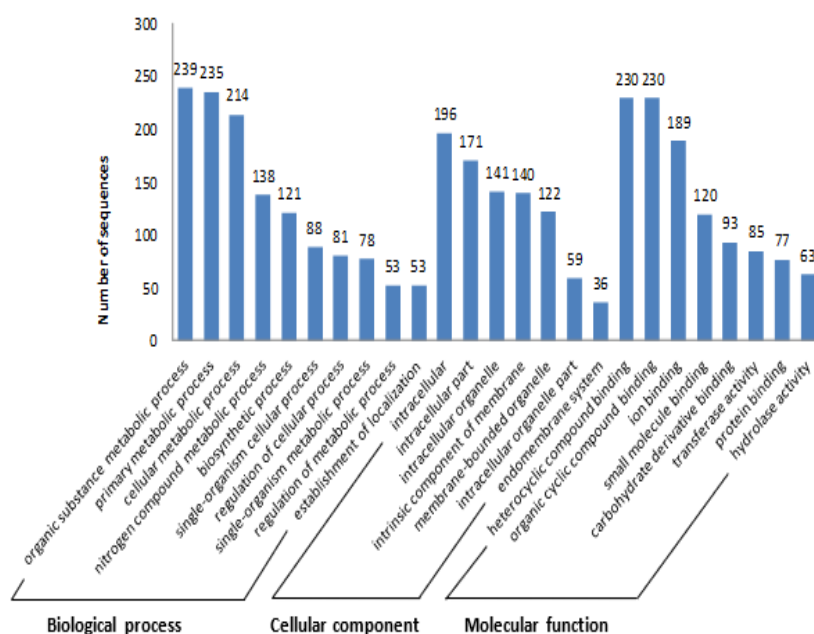

Figure 2 Gene Ontology analyses of 1,861 SSR-containing unigenes of L. fructicosa transcriptome underg0 classifications of biological process, cellular component and molecular function

Pathway prediction for SSR-containing unigenes was performed to gain an insight into the biosynthesis of bioactive compounds in leaf of *L. fruticosa*. SSR-containing unigenes were assigned into 48 KEGG pathways. The most represented pathways were metabolism, secondary metabolite and signaling pathway (Table 6). Purine and thiamine metabolism were the two most highly represented metabolisms among all, with 23 and 21 predicted SSRs respectively. These was followed by 13 predicted SSRs that involved in aminoacyl-tRNA biosynthesis. There was a total of 12 SSR-containing unigenes were assigned to pathways associated with secondary metabolite biosynthesis which include isoflavonoid biosynthesis (3), zeatin biosynthesis (3), diterpenoid biosynthesis (3), flavonoid biosynthesis (2), phenylpropanoid biosynthesis (1) and ubiquinone and other terpenoid-quinone biosynthesis (1). In addition, there were SSR-containing unigenes with functions

Zulkifli Ahmad Seman, Azrin Ahmad, Rabiatul Adawiah Zainal Abidin, Siti Zainab Jantan, Mohd Hanif Azhari Noor, Sew Yun Shin, Mohd Norfaizal Ghazalli, Khairun Hisam Nasir, Sanimah Simoh, and Mohd Shukri Mat Ali

related to cell signalings such as T cell receptor signaling pathway (4), mTOR signaling pathway (2) and phosphatidylinositol signaling system (2), respectively.

Table 6 Pathway predicted for SSR-containing unigenes via KEGG pathway analysis

Pathway Categories	Pathway	Number of SSR
Metabolism	Purine metabolism	23
	Thiamine metabolism	21
	Glycerophospholipid metabolism	5
	Glycerolipid metabolism	4
	Starch and sucrose metabolism	4
	Pyrimidine metabolism	4
	Drug metabolism-other enzymes	3
	Methane metabolism	2
	Cysteine and methionine metabolism	2
	Glycine, serine and threonine metabolism	2
	Glyoxylate and dicarboxylate metabolism	1
	Glutathione metabolism	1
	Cyanoamino acid metabolism	1
	alpha-Linolenic acid metabolism	1
	Galactose metabolism	1
	Drug metabolism-cytochrome P450	1
	Retinol metabolism	1
	Ether lipid metabolism	1
	Phenylalanine metabolism	1
	Metabolism of xenobiotics by cytochrome P450	1
	Tyrosine metabolism	1
	Phosphonate and phosphinate metabolism	1
	Taurine and hypotaurine metabolism	1
	Inositol phosphate metabolism	1
	Ascorbate and aldarate metabolism	1
	Alanine, aspartate and glutamate metabolism	1
	Fructose and mannose metabolism	1
	Arginine and proline metabolism	1
Biosynthesis	Aminoacyl-tRNA biosynthesis	13
	Biosynthesis of antibiotics	5
	Isoflavonoid biosynthesis	3
	Zeatin biosynthesis	3
	Diterpenoid biosynthesis	2
	Flavonoid biosynthesis	2
	Phenylpropanoid biosynthesis	1
	Ubiquinone and other terpenoid-quinone biosynthesis	1
Others	Aminobenzoate degradation	4
	Th1 and Th2 cell differentiation	4
	T cell receptor signaling pathway	4
	Glycolysis/Gluconeogenesis	2
	mTOR signaling pathway	2
	Naphthalene degradation	1
	Oxidative phosphorylation	1

Chloroalkane and chloroalkene degradation	1
Phosphatidylinositol signaling system	1
Fatty acid degradation	1
Pentose phosphate pathway	1
Other glycan degradation	1

4. DISCUSSION

Understanding of genetic variation in the germplasm, e.g determination of plant genetic diversity by utilizing DNA molecular markers is the prerequisite for crop improvement. Previous studies have identified the application of Random Amplification of Polymorphic DNA (RAPD) markers from *Annona* species Anuragi et al. (2016) and loquat Badenes et al. (2004) of which the markers were utilized for genetic diversity study. Nevertheless, massive data generated from transcriptome sequencing has been reported to be comprehensive and useful genomic and genetic resources that facilitate gene discovery and SSRs development in future study (Zheng et al. (2013), Huang et al. (2014), Chen et al. (2015). Compared to coffee (1/2.16 kb) Zheng et al. (2013), marker density of *L. fruticosa* is lesser but appeared at a much higher frequency than *Arabidopsis* (1/14 kb), chickpea (1/8.66 kb), Jhanwar et al. (2012). Cardle et al. (2000) and cereal plants such as wheat (1/15.6kb) and barley (1/6.3 kb) Kantety et al. (2002). Different distribution frequency of markers among plant species most probably due to composition and size of the genome and also the criteria chosen to screen the marker.

This study showed the major common repeat units were mononucleotide and followed by di-nucleotide repeats which comprised 55.0% and 21.5% of the total SSRs respectively. This class of SSR repeats (Mononucleotide) were found also abundant in other plant species Jin et al. (2016) but for certain plant species di-nucleotide repeat was the most Izzah et al. (2014), Silva et al. (2013), Zhang et al. (2012). Dominant motif of di-nucleotide repeats detected in this study was (AT/AG) n (48.9%) whereas (GAA/TTC) n (48.7%) was the dominant motif of tri-nucleotide repeats. However, different dominant motif detected in *Dipteronia* Oliver (*Aceraceae*) and rubber tree which AG/CT and AAG/CTT as their dominant di-nucleotide trinucleotide repeat motif respectively (Li et al. 2012; Triwitayakorn et al. (2011), Zhou et al. (2016)Zhou et al. 2016). Interestingly, GGGCAA motif was considered as a rare motif for *L. fruticosa* with a lower frequency of 3 in this present study.

Sequence annotation showed that about 20.6% of SSR markers had significantly hits to the NCBI database. However, hits percentage against the existing data in the database are relatively low (20.65%). This unlikely due to the size of unigenes as the average size indicated in the analysis was 1473 bp. Low homology percentage could be probably contributed by certain genes that do not hit against database upon blast or possibly matched to unknown proteins. Assumption made on the basis that a scanty genetic information available for *L. fruticosa* and its close relatives in the current public database. Similarity search analysis carried out on SSR-containing unigenes clearly indicated that *L. fruticosa* has a closer genetic distance with woody plant species such as *Citrus sinensis* (34%), *Citrus clementina* (20%) and *Theobroma cacao* (11%). These data suggested that some of genetic information of *L. fruticosa* such as unigenes and SSR-containing unigenes obtained from this current study could be useful and applicable to other woody plant species such as Citrus and date palm.

The results of Gene Ontology (GO) analysis suggested that SSR-containing unigenes of *L. fruticosa*'s leaf have diversified biochemical functions. More

interestingly, those a number of detected SSR-containing unigenes have functions related to compounds binding such as heterocyclic and organic cyclic compound binding. Many of the natural plant secondary metabolites have cyclic compound structure that carry one or more atoms that connecting each other to form a ring. Organic cyclic compounds may contain carbon atom while heterocyclic compounds can be formed by a combination of carbon and non-carbon atoms Smith and March (2006). For instance, the compounds of flavan-3-ols, proanthocyanins and flavanones contain heterocyclic C-ring Crozier et al. (2007), a highly abundant SSR-containing unigenes with binding functions to those compounds observed in *L. fruticosa* suggests that a majority of detected SSRs are actively involved and being an essential component in modulating the biosynthesis process of secondary metabolites.

KEGG pathway analysis revealed the SSR-containing unigenes are highly associated with purine and thiamine metabolism. This finding is correlated with Study by Suzuki and Waller (1985) found that purine alkaloids were abundantly available in the fruits of *Camellia sinensis* L. and *Coffea arabica* L. during fruit development stages. It is therefore suggested that high purine content is likely to be observed among fruit bearing tree plants including *L. fruticosa* which is essential for purine alkaloid synthesis during fruit formation. Thiamine diphosphate (vitamin B (1)) is known to be an enzymatic cofactor for various metabolic pathways in central metabolism such as glycolysis, pentose phosphate and the tricarboxylic acid cycle Goyer (2010). In addition, previous studies have shown that thiamine was a cofactor involved to elicit the genes expression in phenylpropanoid pathways of grapevine and those genes were associated with accumulation of phenolics, flavonoids, lignin and stilbenes Boubakri et al. (2013).

The above-mentioned findings are in agreement with results obtained from KEGG pathway analysis, where a number of SSR were detected in unigenes that involved in biosynthesis pathways of secondary metabolites. Likewise, zeatin biosynthesis, diterpenoid biosynthesis, terpenoid-quinone biosynthesis, flavonoid and isoflavonoid biosynthesis were among those pathways represented in the secondary metabolite biosynthesis pathways, well explaining that *L. fruticosa* is a rich source of phytochemicals. This is supported by previous studies which showed that leaf of *Lepisanthes* species contains various phytochemicals such as proanthocyanidins Zhang et al. (2016), alkaloids, terpenoids and flavonoids Kuspradini et al. (2012).

5. CONCLUSION

Study carried out was the first transcriptome sequencing of RNA sample extracted from leaf of *L. fruticosa* using NGS technology. Both EST-SSRs and NGS-based SSRs are functional markers identified from expressed transcripts of an organism, however application of NGS technology is comparatively a more reliable and higher throughput approach for novel SSR markers discovery in rapid, convenient and cost-effective manners than the traditional EST-SSRs identification processes. The present study has detected and characterized 23,958 microsatellite loci from 52,569 non-redundant unigenes. These data have substantially increased the existing genomic information of under explored plant species of *L. fruticosa*. Also, our data suggested that a high similarity of SSR-containing unigenes between *L. fruticosa* and date palm and citrus plants. These newly obtained genetic information will be useful for screening and profiling *L. fruticosa* accessions in particular for their secondary metabolites content. This will support future breeding

program for *L. fruticosa* from under ultilized fruit to local food market as an enhanced nutritional food product.

REFERENCES

Anuragi, H., H.L Dhaduk, S. Kumar, J.J. Dhruve, M.J. Parekh, and A.A. Sakure. (2016). Molecular diversity of Annona species and proximate fruit composition of selected genotypes. 3 Biotech. 6 :1-10. Retrieved from https://doi.org/10.1007/s13205-016-0520-9

Badenes, M.L., T. Canyamas, C. Romero, J. Martinez-Calvo, E. Giordani and G. Llacer. (2004). Characterization of under-utilized fruits by molecular markers. A case study of loquat. Genetic Resources and Crop Evolution 51 :335-341. Retrieved from https://doi.org/10.1023/B:GRES.0000024017.57973.6f

Boubakri, H., A. Poutaraud, M.A.Wahab, C. Clayeux, R. Baltenweck-Guyot, D. Steyer, C. Marcic, A. Mliki, and I. Soustre-Gacougnolle. (2013). Thiamine modulates metabolism of the phenylpropanoid pathway leading to enhanced resistance to Plasmopara viticola in grapevine. BMC Plant Biology, 13 : Retrieved from https://doi.org/10.1186/1471-2229-13-13

Cardle, L., L. Ramsay, D. Milbourne, M. Macaulay, D. Marshall and R. Waugh. (2000). Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156 : 847-854. Retrieved from https://doi.org/10.1093/genetics/156.2.847

Chen, H., L. Wang, S. Wang, C. Liu, M.W. Blair and X. Cheng. (2015). Transcriptome sequencing of mung bean (Vigna radiate L.) genes and the identification of EST-SSR markers. Plos ONE 10 : e0120273 : 10.1371/journal.pone.0120273. Retrieved from https://doi.org/10.1371/journal.pone.0120273

Crozier, A., M.N. Clifford. And H. Ashihara. (2007). In : H. Ashihara (ed), Plant secondary metabolites: Occurrence, structure and role in the human diet. John Wiley and Sons, New York. Retrieved from https://doi.org/10.1002/9780470988558

Dayang, F.B., F. Ahmad, M.H. Ruslan, and M.A. Alghoul. (2012). Drying kinetics of Malaysian Canarium odontophyllum (Dabai) fruit. WSEAS Transaction on Biology and Biomedicine 9 :77-82.

Garg, R., R.K. Patel, A.K. Tyagi. And M. Jain. (2011). De novo assembly of chickpea transcriptome using short reads for gene discovery and marker identification. DNA Res. 18 : 53-63. Retrieved from https://doi.org/10.1093/dnares/dsq028

Goyer, A. (2010). Thiamine in plants : aspects of its metabolism and functions. Phytochemistry 71 :1615-1624. Retrieved from https://doi.org/10.1016/j.phytochem.2010.06.022

Gupta, P.K., I.S. Balyan, P.C. Sharma and B. Ramesh. (1996). Microsatellites in plants : A new class of molecular markers. Curr. Sci. 70 : 45-54. Retrieved from https://www.jstor.org/stable/24097472

Huang, D., Y. Zhang, M. Jin, H. Li, Z. Song, Y. Wang and J. Chen. (2014). Characterization and high cross-species transferability of microsatellite markers from the floral transcriptome of Aspidistra Saxicola (Asparagacea). Mol. Ecol. Resource 14 : 569-577. Retrieved from https://doi.org/10.1111/1755-0998.12197

Ibrahim, M., K.N. Prasad, I. Amin, A. Azrina and A.H. Azizah. (2010). Physiochemical composition and antioxidant activities of underutilized Mangifera pajang fruit. African Journal of Biotechnology 9 :4392-4397.

Ikram, E.H.K., K.H. Eng, A.M.M. Jalil, A. Ismail, S. Idris, A. Azlan, H.S.M Nazri, N.A.M Diton and R.A.M. Mokhtar. (2009). Antioxidant capacity and total phenolic content of Malaysian underutilised fruits. J. Food Comp. Anal. 22 :388-393. Retrieved from https://doi.org/10.1016/j.jfca.2009.04.001

Izzah, N. K., J. Lee, M. Jayakodi, S. Perumal, M. Jin, B.S. Park, K. Ahn and T.J. Yang. (2014). Transcriptome sequencing of two parental lines of cabbage (Brassica oleracea L. var. capitata L.) and construction of an EST-based genetic map. BMC Genom. 15 : DOI : 10.1186/1471-2164-15-149. Retrieved from https://doi.org/10.1186/1471-2164-15-149

Jhanwar, S., P. Priya, R. Garg, S.K. Parida, A.K. Tyagi and M. Jain. (2012). Transcriptome sequencing of wild chickpea as rich resources for marker development. Plant Biotechnology J. 10 : 690-702. Retrieved from https://doi.org/10.1111/j.1467-7652.2012.00712.x

Jin, M., M.Y. Guo, L. Han, J.L. Li, S.Y. Yang and Y.H. Su. (2016). Transcriptome analysis of potential simple sequence repeat markers in Ammopiptanthus mongolicus. Genetic and Molecular Research 15 : DOI : 10.4238/gmr.15038581. Retrieved from https://doi.org/10.4238/gmr.15038581

Kantety, R.V., M. La Rota, D.E. Matthews and M.E. Sorrells. (2002). Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Mol. Biol. 48 : 501-510. Retrieved from https://doi.org/10.1023/A:1014875206165

Kuspradini, H., D. Susanto and T. Mitsunaga. (2012). Phytochemical and comparative study of antimicrobial activity of Lepisanthes amoena leaves extract. Journal of biology, agricultural and healthcare 2 :80-86.

Li, D.J., Z. Deng, B, Qin, X.H. Liu and Z.H. Men. (2012). De novo assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (Hevea brasiliensis Muell. Arg.). BMC Genome 13 : DOI : 10.1186/1471-2164-13-192. Retrieved from https://doi.org/10.1186/1471-2164-13-192

Mirfat, A.H.S. and I. Salma. (2015). Ceri Terengganu : The future antioxidant superstar. MARDI Scientia 6 : 6.

Mirfat, A.H.S., O. Zaulia, C.L.Y. Joanna, S.M.N. Erny and I. Salma. (2017). Antioxidant activity and phytochemical content of fresh and freeze-dried Lepisanthes fruticosa fruits at different maturity stages. Journal of Agricultural Science 9 :1916-9760. Retrieved from https://doi.org/10.5539/jas.v9n2p147

Rizvi, A., A. Mishra, A. Mahdi, M. Ahmad and A. Basit, (2015). Natural and herbal stress remedies: à review. International Journal of Pharmacognosy 2 :155-160.

Scott, K.D., P. Eggler, G. Seaton, M. Rossetto, E.M. Ablett, L.S. Lee and R.J. Henry. (2000). Analysis of SSRs derived from grape ESTs. Theor. Appl. Genet. 100 : 723-726. Retrieved from https://doi.org/10.1007/s001220051344

Silva, P. I., A.M. Martins, E.G. Gouvea, M. Pessoa-Filho and M.E. Ferreira. (2013). Development and validation of microsatellite markers for Brachiaria ruziziensis obtained by partial genome assembly of Illumina single-end

reads. BMC Genome 14 : DOI.org/10.1186/1471-2164-14-17. Retrieved from https://doi.org/10.1186/1471-2164-14-17

Smith, M.B and J. March. (2006). March's advanced organic chemistry: Reactions, mechanisms and structure. 6th Ed. John Wiley & Sons., New York.

Song, Y.P., X.B. Jiang, M. Zhang, Z.L. Wang, W.H. Bo, X.M. An, D.Q. Zhang and Z.Y. Zhang. (2012). Differences of EST-SSR and genomic-SSR markers in assessing genetic diversity in poplar. Forestry studies in China 14 : 1-7. Retrieved from https://doi.org/10.1007/s11632-012-0106-5

Suzuki, T. and G.R. Waller. (1985). Purine alkaloids of the fruits of Camellia sinensis L. and Coffea arabiaca L.during fruit development. Annals of Botany 56 :537-542. Retrieved from https://doi.org/10.1093/oxfordjournals.aob.a087038

Triwitayakorn, K., P. Chatkulkawin, S. Kanjanawattanawong, S. Sraphet, T. Yoocha, D. Sangsrakru, J. Chanprasert, C. Ngamphiw, N. Jomchai and K. Therawattanasuk. (2011). Transcriptome sequencing of Hevea brasiliensis for development of microsatellite markers and construction of a genetic linkage map. DNA Res. 18 : 471-482. Retrieved from https://doi.org/10.1093/dnares/dsr034

Umikalsum, H.Z. and A.H. Mirfat, (2014). Proximate composition of Malaysia underutilized fruits. J. Trop. Agric. & Food. Sc. 42 : 63-72.

Wang, Z., B. Fang, J. Chen, X. Zhang, Z. Luo, L. Huang, X. Chen and Y. Li. (2010). De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of SSR markers in sweet potato (Ipomoea batatas). BMC Genomics 11 :726-739. Retrieved from https://doi.org/10.1186/1471-2164-11-726

Wei, W., X. Qi, L. Wang, Y. Zhang, W. Hua, D. Li, H. Lv. And X. Zhang. (2011b). Characterization of the sesame (Sesamum indicum L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers. BMC Genomics 12 :451-463. Retrieved from https://doi.org/10.1186/1471-2164-12-451

Yoichi, W., S. Sakaguchi, S. Ueno, N. Tomaru and K. Uehara. (2017). Development and characterization of EST-SSR markers for the genus Rhododendron section Brachycalyx (Ericaceae). Plant Spec. Biol. 32 :455-459. Retrieved from https://doi.org/10.1111/1442-1984.12155

Zeng, S., G. Xiao, J. Guo, Z. Fei, Y. Xu, B.A Roe and Y. Wang. (2010). Development of a ET data set and characterization of EST-SSRs in a traditional Chinese medicinal plant, Epimedium sagittatum (Sieb. Et Zucc.) Maxim. BMC Genomics 11 : 94-104. Retrieved from https://doi.org/10.1186/1471-2164-11-94

Zhang, J.N., S. Liang, J.L. Duan, J. Wang, S.L. Chen, Z.S. Cheng, Q. Zhang, X.Q. Liang and Y.R. Li. (2012). De novo assembly and characterization of the transcriptome during seed development, and generation of genic-SSR markers in Peanut (Arachis hypogaea L.). BMC Genom. 13 : DOI : 10.1186/1471-2164-13-90. Retrieved from https://doi.org/10.1186/1471-2164-13-90

Zhang, Y., A.I.C. Wong, J. Wu, N.B.A. Karim and D. Huang. (2016). Lepisanthes alata (Malay cherry) leaves are potent inhibitors of starch hydrolases due to proanthocyanidins with high degree of polymerization. Journal of Functional Foods 25 : 568-578. Retrieved from https://doi.org/10.1016/j.jff.2016.06.035

Zheng, X., C. Pan, Y. Diao, Y. You, C. Yang and Z. Hu. (2013). Development of microsatellite markers by transcriptome sequencing in two species of

Zulkifli Ahmad Seman, Azrin Ahmad, Rabiatul Adawiah Zainal Abidin, Siti Zainab Jantan, Mohd Hanif Azhari Noor, Sew Yun Shin, Mohd Norfaizal Ghazalli, Khairun Hisam Nasir, Sanimah Simoh, and Mohd Shukri Mat Ali

Amorphophallus (Araceae). BMC Genomic 14 : 490. doi 10.1186/1471-2164-14-490. Retrieved from https://doi.org/10.1186/1471-2164-14-490

Zhou, Q., D. Luo, L. Ma, W. Xie, Y. Wang, Y. Wang and Z. Liu. (2016). Development and cross-species transferability of EST-SSR markers in Siberian wildrye (Elymus sibiricus L.) using Illumina sequencing. Sci. Rep. 6 : 266-278 Retrieved from https://doi.org/10.1038/srep20549