IOP Conference Series:
Materials Science and
Engineering

PAPER • **OPEN ACCESS**

Chloroplast DNA sequence of *trn*R-N and *trn*L-F regions in Harumanis mango from different orchards in Perlis, Malaysia

To cite this article: AMA Rahman *et al* 2020 *IOP Conf. Ser.: Mater. Sci. Eng.* **932** 012028

IOP Conf. Series: Materials Science and Engineering **932** (2020) 012028 doi:10.1088/1757-899X/932/1/012028

Chloroplast DNA sequence of *trn*R-N and *trn*L-F regions in Harumanis mango from different orchards in Perlis, Malaysia

AMA Rahman[1], SFM Sabri[1], A Yusuf[1], Z Zakaria[1], SV Kumar[2]

[1]Department of Chemical Engineering Technology, Faculty of Engineering Technology, Universiti Malaysia Perlis (UniMAP) Sungai Chuchuh, 02100 Padang Besar, Perlis

[2]Biotechnology Research Institute, Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia.

Email: mukhlis@unimap.edu.my

Abstract. Harumanis is a premium mango cultivar widely known for its sweet taste, aroma and vibrant flesh colour. To date, the genetic identification of this mango based on multiple conserved DNA region using samples from different orchards has never been reported. The aim of this research is to identify the genetic signature of Harumanis mango at molecular level by analyzing chloroplast DNA sequences of the *trn*L-F and *trn*R-N regions. DNA samples were extracted from a total of 15 Harumanis samples collected from five selected orchards using Cetyl Trimethyl Ammonium Bromide (CTAB) extraction procedure. The extracted DNA and the PCR-amplified products were analyzed through gel electrophoresis and were then subjected to DNA Sequencing and *in silico* analysis. The obtained sequences were compared with the sequences available in the GeneBank. BLAST search for both the *trn*R-N and *trn*L-F regions confirmed that all the 15 samples belong to *Mangifera indica* with a 99% sequence identity. In addition, the trnL-F sequences were 99% identical to a number of specific mango cultivars such as, Tommy and Arunika. However, the trnR-N sequences were less informative as it gave hits to only two mango accessions (e.g. *Mangifera indica* voucher PDBK 2014-0249). It is postulated that the plastid trnR-N may be a potential candidate region for the development of the Harumanis genetic signature. The results may be used to complement other molecular data for the development of a genetic barcode for Harumanis.

1. Introduction

Over the years, the genomic studies have increased in number due to the rapid advances of technology in DNA sequencing [1]. In plants, the chloroplast plays an important role in photosynthesis and carbon fixation, as well as starch, fatty acids, amino acid biosynthesis. In higher plants, the chloroplast genome ranges in size from 120 to 180 Kb and has a quadripartite structure, including a large single copy (LSC) region, a small single copy (SSC) region, and two copies of an inverted repeat (IR) region (IRa and IRb) [2]. The chloroplast genome is considered smaller than the nucleus genome. The differences between the genomes have been compared using comparative genomics. The studies also examined the implications of any similarities or differences of the genomes and provide explanation on evolutionary hypotheses and species differences [3][4]. The average evolution speed of chloroplast DNA is four times slower than nuclear DNA in plants [5]. It is relatively slow and conservative, thus the chloroplast DNA is suitable for comparative genomics study. However, the evolution speed of chloroplast DNA differs from each other and it depends on where it is located in the genome. Commonly, the mutation or evolution speed of small single copy (SSC) and large single copy (LSC) region are much faster than inverted repeat (IR) region [5] [6]. The chloroplast genome has been widely used for phylogenetic analysis and molecular marker development in plant species. These molecular markers are highly useful DNA barcoding tools for the authentication and identification of plant taxanomy. For example, *matK* and *rbcL* genes in chloroplast genomes are used as universal plant DNA barcodes [7].

[1] To whom any correspondence should be addressed.

Mango (*Mangifera Indica* L.) is a dicotyledon broadleaf evergreen fruit that are very popular especially in South East Asia. It is categorized as one of the member under the family of Anacardiaceae commonly known as the flowering plant of genus *Mangifera*. In Perlis, Malaysia there is a popular seasonal fruit that is considered as the "King of Mangoes" called Harumanis mango due to its sweetness, aroma, deliciousness and juiciness [8]. Most of other studies reported on Harumanis are such as on it's market potential [9], improvement of its post-harvest quality [10], flowering stem prediction [11], development of sensor for Harumanis's quality assessment [12] and real-time greenhouse monitoring system [13]. A number of reports are reported on evaluation of genetic and phenotypic diversity of Harumanis in this country with limited number of morphological traits being studied [14][15][16]. Genotype identification to evaluate DNA polymorphism can be conducted by numerous types of DNA based molecular techniques such as polymerase chain reaction (PCR) method, hybridization and sequencing based methods including using chloroplast DNA conserved region amplification and sequencing. In this study, the use of PCR for amplification, DNA sequencing and *in silico* analysis of *trn*L-F and *trn*R-N regions of the Harumanis from different orchards were reported.

2. Materials And Methods

2.1 Biological Samples

The Harumanis mangoes were collected from three districts in Perlis which are Arau, Kangar and Padang Besar. Five different locations have been chosen as the target site for the collection of sample. For genetic characterization study, a few mature leaves from each of the sample were collected. The samples collected were labelled as (1) HM 2, HM 6, HM 9 (from Chelong Balik Bukit), (2) HM 11, HM 12, HM 14 (from Paya Kelubi), (3) HM 21, HM 22, HM 24 (from Alor Ara Timur), (4)HM 32, HM 34, HM 40 (from Ladang Harumanis Santan) and (5) HM 42, HM 44, HM 46 (from Sanglang), with 15 samples in total.

2.2 Genomic DNA Extraction

The DNA from the leaves were extracted according to the protocol of CTAB DNA extraction which was developed by Doyle and Doyle [17]. The purified DNA samples were analyzed before undergo PCR reactions to ensure that the DNA extraction is successful by running on electrophoresis in a 1.5% (w/v) agarose gel and 1X TAE buffer for 60 minutes. 1 kb marker and DNA sample were mixed with loading dye and were loaded into the gels. After electrophoresis has been conducted, the gel were analyzed under the Bio-Rad Gel Imager (Gel Documentation System) to estimate the size of the genomic DNA.

2.3 Polymerase Chain Reaction (PCR) and DNA Sequencing

The primer pairs that were used in the Polymerase Chain Reaction (PCR) are shown in Table 1 . Polymerase Chain Reaction (PCR) for each DNA samples were carried out with two different primer pairs as described by Yang *et al*. [18] and Rahman *et al*. [20] with slight modification. Each of the primer pairs was optimized by conducting a gradient PCR of temperature 60.5 ^{o}C to 65.5 ^{o}C for primer cp104 and 55.4 ^{o}C to 60.4 ^{o}C for primer *trn*L-F at annealing step to find the best annealing temperature. Then the resulted non-purified PCR amplification was analyzed on electrophoresis to examine the quality, sizes of the band and pattern of DNA amplification produced before being directly sent for PCR purification and DNA sequencing. The obtained sequences were then compared to the sequences available in the GeneBank by using BLAST algorithm. The DNA sequences will be deposited in GeneBank of National Centre for Biotechnology Information (NCBI).

Table 1. Primer sequence for amplification of *trn*R-N and *trn*L-F regions

Name/ code	Sequence (5'-3')	Primer Length	Amplification Region/ Location
cp104-F	TCCTCAGTAGCTCAGTGGTAG	21	*trn*R-N
cp104-R	GGCCTGTAGCTCAGAGGATT	20	*trn*R-N
trnL-F3	GGTTCAAGTCCCTCTATCCC	20	*trn*L-F
trnL-F4	ATTTGAACTGGTGACACGAG	20	*trn*L-F

3. Results & Discussion

Many studies have indicated that chloroplast DNA contains a great deal of genetic variation between species and populations [21][22][23]. A representative study was reported by Saski *et al.* [24], who compared variable sequences in chloroplast DNA from *Hordeum vulgare*, *Sorghum bicolor*, and *Agrostis stolonifera* to assess the phylogeny among species. The same protocol was also reported by Yang *et al.* [18] in which the method was used to evaluate the variation in *Oryza sativa, Arabidopsis thaliana, Glycine max, Lotus japonicus, Medicago truncatula* and *Phaseolus vulgaris.*

3.1 Genomic DNA extraction

The extraction was done using conventional method of CTAB DNA extraction under sterile condition. Agarose gel stained with ethidium bromide was used to verify the quality of genomic DNA extracted. The genomic DNA from 15 samples were analyzed using 1.5 % agarose gel electrophoresis and the result obtained as in Figure 1.

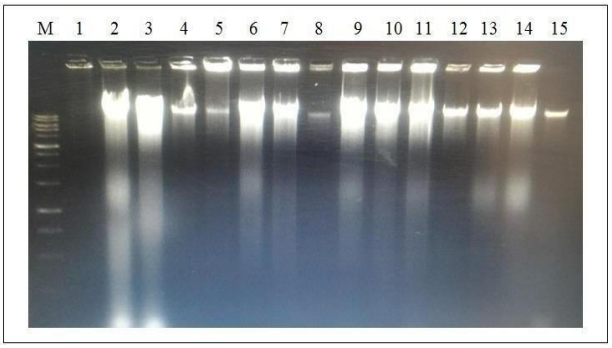

Figure 1 Agarose gel electrophoresis of genomic DNA. Lane M: 1 kb Bio-Rad marker, Lane 1: HM 2, Lane 2: HM 6, Lane 3: HM 9, Lane 4: 11, Lane 5: HM 12, Lane 6: HM 14, Lane 7: HM 21, Lane 8: HM 22, Lane 9: HM24, Lane 10: HM 32, Lane 11: HM 34, Lane 12: HM 40, Lane 13: HM 42, Lane 14: HM 44 and Lane 15: HM 46.

The genomic DNA can be visualized clearly without degradation of DNA by properly treating the DNA with RNase. The size of the band produced are greater than 1 kb ladder, thus it indicates that DNA is successfully isolated. From the result, it can be concluded that there is no RNA contamination during the electrophoresis. The resulted bands formed are very sharp and bright except for the sample HM 2. Thus, the samples which producing a clearer band can be concluded to have high concentration of DNA compared to the other samples. As for the other samples, brighter and sharp bands could be visualized if higher volumes of genomic DNA are loaded.

3.2 PCR Amplification of Harumanis DNA using cp104 and trnL-F Primer Pairs

Figure 2 shows the PCR amplification products of 15 samples by using primer cp104 with optimum annealing temperature of 62.4 ºC. The results showed that there are bands formed from the amplification of 25 cycles using 0.5 µl DNA templates and 0.15 µl primer. The sizes of the bands are almost similar for all the samples which are in the range of 700 bp to 800 bp. The figure clearly showed that all the resulted bands produced are without smears. Based on the Figure 2, all of the samples are producing a bright single band due to the high concentration of DNA loaded. The sizes of the PCR product reported is in the same range as those reported in *Arachis hypogaea*, *Vigna radiata*, *Glycine max* and *Vigna unguiculata* using the same cp104 primer pair [18].

Figure 3 shows the PCR amplification products of 15 samples by using primer *trn*L-F with optimum annealing temperature of 60.4 ºC. The result shows that there are bands formed from the amplification of 25 cycles using 0.5 µl DNA templates and 0.15 µl primer. The sizes of the bands are almost similar for all the samples which are in the range of 400 bp to 500 bp. All of the samples are producing bright and clear bands. The PCR of the *trn*L-F region has been conducted in several species such as *Pergularia daemia*, and *Mangifera indica*, L. var totupura with the size of PCR products of between 350 to 550 bp [25][26].

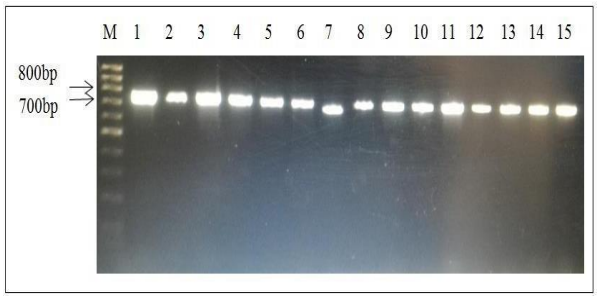

Figure 2. Agarose gel electrophoresis of 15 samples using primer cp 104. Lane M: Lucigen marker. L1: HM 2, L2: HM 6, L3: HM 9, L4: HM 11, L5: HM 12, L6: HM 14, L7: HM 21, L8: HM 22, L9: HM 24, L10: HM 32, L11: HM 34, L12: HM 40, L13: HM 42, L14: HM 44, L15: HM 46.

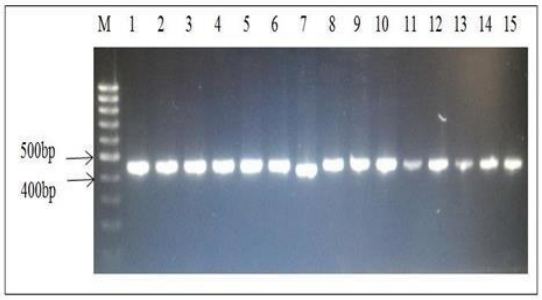

Figure 3. Agarose gel electrophoresis of 15 samples using primer trnL-F. Lane M: 100 bp Lucigen marker. L1: HM 2, L2: HM 6, L3: HM 9, L4: HM 11, L5: HM 12, L6: HM 14, L7: HM 21, L8: HM 22, L9: HM 24, L10: HM 32, L11: HM 34, L12: HM40, L13: HM 42, L14: HM 44, L15: HM 46.

3.3 Sequencing of amplification product of reference DNA target sequence and in silico analysis
The DNA Sequencing results of 30 samples were obtained in a document and all the sequences have been read using software called Chromas Lite and the example of the results are shown in table 2. The DNA sequences obtained from the Chromas Lite software is in the FASTA format. All of 30 samples Harumanis species that have been isolated were identified through online BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) by using the alignment sequences of each samples. The sequences of each isolated were compared with the sequences that are available in GenBank. BLAST search for both the *trn*R-N and *trn*L-F regions are confirm belong to *Mangifera indica* with 99% sequence identity similarity. In addition, the *trn*L-F sequences were 99% identical to a number of specific mango cultivars such as, Tommy and Arunika. However, the *trn*R-N sequences were less informative as it gave hits to only two mango accessions (e.g. *Mangifera indica* voucher PDBK 2014-0249).

Table 2 DNA Sequences of sample HM 12 using primer cp104

Samples	DNA Sequences
HM 12	CTCGACTTCGTTACTGACTGGTCGTAGGTTCGAATCCTACTTGGGGAGATTTG ATTCATTCCGAATTAAAGAATTCAGAATGTCAGAATGAAAGGGTTCGCTT TGACCGTTAAAAGTAGGTAACCCGTTCCCTGTGTCTTTGTTTCTATTGCAT TCTATCTCATCGTATCACATTCTGTTCTGCGATATTTGAGAATCACCGTCA ATACCTCGGTGTAGGTCCGGGATAATCCTTTGTTCCATAGTCCTGGGGCT ATTTACAATTAGCCAATTAAGAATTCTCAGATGTACTAGTACTAGCAAGT GCATCAAAGATGCAGTCATCGATTCTCCCGAGAGGCCACAATTACCGCGA GCAAACATATTAATGACGAGGAACGAATTTTTGCTATGCTACTAATACTT GCTCTGCTATTCTGCCCAAGCCCGGCCGAGGAAGAGTTACGGGGCGTAAAAC AAAAAAAAAATATGCCAATGGGGCCGGGCATACTATAATTAAAAAAACC AAGAATGGTAAATAAAAAAAAAGAAAAAGTAAGGCTACTCCATTTCGAC AAAAGACCCACACCCAAGTTCCATAGCTTTGGGTCCGCTATCCCGATCAT GATTTTCCTACCCCCAGAGGGAAGGGTCCTTCCCTTTTTGGCCGGTTGTGG GCGAGGAGGGATTCGAACCCCCGACACCGTGGTTCGTAGCCACGTGCTCT AATCCTCTAGCTATCAGGCCAAAA

Table 3: BLAST results for the specific target sequence of each 'Harumanis' samples using primer cp 104

Samples	Sequence description	Max score	Total score	Query cover	E value	Max identical	Accession number
HM 2	*Mangifera indica* voucher PDBK 2014-0249 Chloroplast	1295	2591	99%	0.0	99%	KX871231.1
HM 6		1288	2576	96%			
HM 9		1288	2576	99%			
HM 11		1295	2591	97%			
HM 12		1308	2617	98%			
HM 14		1308	2617	96%			
HM 21		1303	2606	98%			
HM 22		1301	2602	97%			
HM 24		1301	2602	96%			
HM 32		1299	2598	98%			
HM 34		1293	2587	96%			
HM 40		1297	2594	97%			
HM 42		1299	2598	97%			
HM 44		1291	2583	96%			
HM 46		1284	2569	95%			

Table 4: BLAST results for the specific target sequence of each 'Harumanis' samples using primer *trn*L-F

Samples	Sequence description	Max score	Total score	Query cover	E value	Max identical	Accession number
HM 2	*Mangifera indica* voucher PDBK 2014-0249 chloroplast	750	750	96%	0.0	99%	KX871231.1
HM 6		741	741	99%			
HM 9		743	743	96%			
HM 11		752	752	98%			
HM 12		749	749	98%			
HM 14		756	756	99%			
HM 21		747	747	99%			
HM 22		743	743	96%			
HM 24		752	752	99%			
HM 32		743	743	97%			
HM 34		754	754	98%			
HM 40		758	758	98%			
HM 42		741	741	98%			
HM 44		750	750	98%			
HM 46		743	743	96%			

4. Conclusion

Variation in the chloroplast DNA sequence is useful for plant phylogenetic studies. The plant genomic DNA extraction which using CTAB DNA extraction method on Harumanis leaves collected from five different locations in Perlis has proven to be effective to isolate the genomic DNA. The best annealing temperature for primer cp104 was found to be at 62.4 ^{o}C while primer

*trn*L- F was found to be at 60.4 ºC. The BLAST results showed that the genetic similarity for all samples extracted are 99% identical to the *Mangifera indica* existed among the Harumanis accessions. We postulate that the plastid *trn*R-N may be a potential candidate region for the development of the Harumanis genetic signature. The results may be used to complement other molecular data for the development of a genetic barcode for Harumanis.

References

[1] Edwards D, Batley J, and Snowdon R J 2013 *Theor. Appl. Genet*. **126** 1-11
[2] Yang M, Zhang X, Liu G, Yin Y, Chen K, Yun Q, Zhao D, Al-Mssallem I S and Yu J 2010 *PLoS ONE*;5:e12762.
[3] Wolfe K H, Li W H, and Sharp P M 1987 *Proc. Nat. Acad. Sci. U.S.A.* **84** 9054-9058
[4] De Las Rivas J, Lozano J J, and Ortiz A R 2002 *Genome Res*. **12** 567-583
[5] Curtis S E and Clegg M T 1984 *Mol. Biol. Evol.* **1** 291-301
[6] Ravi V, Khurana J P, Tyagi A K and Khurana P 2008 *Plant Syst. Evol.* **271** 101-122
[7] Dong W, Liu J, Yu J, Wang L, Zhou S and 2012 *PLoS ONE*. 7:e35071.
[8] Hassan F, Musa R, Yusof J M and Daud N M 2010 *J. of Agribus. Market. Special Ed*, 77–90.
[9] Hassan F H, Musa R, Yusof J M and Yahaya J 2008 *J. of Agribus. Market.* **1** 17-35
[10] Ding P and Syafini N G 2012 *Transac. of The Malay. Soc of Plant Physio*. **20** 42-46
[11] Farook R, Ali H, Harun A, Ndzi, D, Shakaff A Y, Husin Z and Aziz A 2013 *Research Notes in Informatio Sciences (RNIS)* **13** 46-51
[12] Saad M, Kamarudin L M, Kamarudin K and Jaafar M N 2014 *2nd Int. Conf. on Electronic Design* 521-526
[13] Sa'ad F S A, Zakaria A Y M, Shakaff M Z., Abdullah A H, Adom M N, Jaafar M J and Masnan L M 2012 *Proc. Chem.* **6** 165–174
[14] Ariffin Z, Sah M S, Idris S and Hashim 2015 *Int. J. of Biodiv*. 458237
[15] Yusuf A, Rahman A M A, Zakaria Z and Wahab Z 2018a *AIP Conf. Proc.* 2030, 020248.
[16] Yusuf A, Rahman A M A, Zakaria Z, Wahab Z and Kumar S V 2018 *Food Res. J.* **2(6)** 564-571
[17] Doyle J J and Doyle J L 1987 *Phytochem Bull*. **19** 11–15
[18] Yang Y C, Kung T L, Hu C Y and Lin S F 2015 *Genet. Mol. Res.* **14 (4)** 14857-14870
[19] Taberlet P, Gielly L, Pautou G and Bouvet J 1991 *Plant. Mol. Biol.* **17** 1105
[20] Rahman A M A, Zakaria Z, Wahab Z, Kumar S V and Leman N 2017 *AIP Conf.Proc.* 2030 020133.
[21] Wojciechowski M F, Lavin M and Sanderson M J 2004 *Am. J. Bot.* **91** 1846-1862
[22] Byrne M and Hankinson M 2012 *Austr. J. Bot.* **60** 569-574
[23] Yang J B, Tang M, Li H T, Zhang Z R 2013 *BMC Evol. Biol.* **13** 84
[24] Saski C Lee S B, Fjellheim S and Guda C 2007 *Theor. Appl. Genet.* **115** 571-590
[25] Raghavamma S T V, Rao N R, Rao K R S S and Rao G D 2013 *Inter. Jour. of Anal. Pharma. & Biomed. Sci.* **2(2)** 38-44
[26] Sankar J, Rosaiah G, Ravi K K and Rajasekhar P 2011 *Bioinformation* **2(10)** 405-409

Acknowledgements
This work was financially support by the Ministry of Education Malaysia (KPM) through Fundamental Research Grant Scheme (FRGS) FRGS/1/2016/900300574 UniMAP. The authors are thankful to Faculty of Engineering Technology UniMAP for providing lab facilities to perform data analysis.