

Development and characterization of microsatellite loci from mango (*Mangifera indica* L.)

K V Ravishankar^{1*}, B Padmakar¹, B Lavanya¹, B H Mani¹ and M R Dinesh²

¹Division of Biotechnology and ²Division of Fruit Crops, ICAR-Indian Institute of Horticultural Research
Hessaraghatta Lake Post, Bangalore 560 089, India

Received 27 October 2015 ; revised 30 May 2016; accepted 14 June 2016

In the present study, a microsatellite enriched partial genomic library was constructed for mango. Twenty sequence tagged microsatellite site loci were characterized by employing M13 tailed PCR technique. All the 20 microsatellite loci were found to be efficient in discriminating and identifying the 20 diverse cultivars of mango used in this study. The genetic analysis for 20 microsatellite loci showed that expected heterozygosity ranged 0.659 to 0.965, with a mean of 0.892, and the observed heterozygosity ranged 0.350 to 0.850, with a mean of 0.505. The polymorphic information content values ranged 0.624 to 0.938, with a mean of 0.860. The probability of identity (PI) values ranged 0.012 to 0.182, with a mean of 0.050. The total PI was 1.06×10^{-28} . These novel SSR markers could be further exploited for genetic diversity studies, cultivar identification, linkage map development and association studies.

Keywords: Diversity, mango, microsatellite markers, PIC

Introduction

The genus *Mangifera* belongs to the family Anacardiaceae and *M. indica* L. (mango) is the most important species in this genus for commercial fruit production. Mango (2n=40), which is indigenous to India, has been under cultivation for more than 4000 years¹. India has been considered as the world's richest germplasm centre for mango. Commonly known as 'the king of fruits', mango is one of the most popular fruits of Asia. The outbreeding nature and a wide range of agroclimatic conditions prevailing in this country continue to contribute to the diversity in this crop. Commercially grown cultivars have arisen through seedling selections made for different fruit characters like color, taste, flavor, size etc.

Simple sequence repeats (SSRs) have become the markers of choice for DNA fingerprinting purposes in most plant species due to their high polymorphism, codominancy and reproducibility. In the present study, 20 novel SSR markers have been developed, which would basically help in the characterization of mango cultivars. Further, they could be useful in distinguishing cultivars from wild relatives and in managing mango germplasm collections. In spite of their importance, till date only a few SSR markers are available for mango²⁻⁴ and, considering its genome size,

their number is very low. Therefore, there is an urgent need to develop more number of SSR markers for mango.

Material and Methods

Microsatellites were isolated and optimized for amplification as described by Ravishankar *et al*⁴. Total genomic DNA was isolated⁵ from fresh leaf samples of mango cultivar 'Alphonso'. DNA was restriction digested with *Rsa*I, and a microsatellite-enriched library was constructed following Glenn and Schable⁶. Using linker-specific primers, these fragments were further enriched by PCR. The enriched fragments were cloned using a T/A cloning kit (MBI Fermentas, Burlington, Canada) and sequenced. Sequences containing microsatellites were selected for primer design. Primers were designed using Primer3 software⁷.

Twenty primer pairs were designed for novel microsatellite loci. The microsatellite sequences were deposited in NCBI GenBank with accession numbers as mentioned in Table 1. High throughput genotyping, by exploiting M13 tailed PCR technique⁸ was employed to characterize 20 diverse mango cultivars, namely, Jalal, Totapuri, Alphonso, Lemon, Malgash, Neelum, Shidadaka Appe, Chettali, Banganapalli, Muffarai, Arya Samaj, Himsagar, Papayaraju Goa, Hamlet, Janardhan Pasand, Vanraj, Bomaby Green, Bombay No. 1, Kesar and Dashehari. The plant material was collected from the mango germplasm collection of the ICAR-Indian Institute of Horticultural

*Author for correspondence:

Tel: +91-80-28466420 extn 406; Fax: +91-80-28466291
kv_ravishankar@yahoo.co.in

Research (IIHR), Bangalore, India. The twenty SSR primers are referred to as MiIHR38-MiIHR72 (Table 1). All the forward primers were modified by adding a tail of M13 sequence, *i.e.*, 'GTAAACGACGGCCAGT' (17 mer) at their 5' ends, while a pig tail of 'CTTGT' sequence was added at the 5' end of all the reverse primers. Additionally, four M13 sequences (17 mers) were further labeled with four different dyes, namely, FAM, PET, NED and VIC at their 5' ends and used as probes.

The total genomic DNA was isolated from young leaves following modified CTAB method⁵. PCR amplification was carried out in 20 μ L reaction mixture

containing: 16 mM (NH₄)₂SO₄, 67 mM Tris-HCl pH 8.8, 0.01% Tween-20, 1.5 mM MgCl₂, 0.1 mM each dNTP, 0.25 μ M each primer, 0.25 μ M M13 probe, 40 ng genomic DNA, and 0.5 unit of Taq DNA polymerase (Bioron GmbH, Ludwigshafen, Germany). PCR was carried out on a Master Cycler Gradient (Eppendorf AG, Hamburg, Germany) thermal cycler with the following temperature profile: initial denaturation at 94°C for 1 min, 35 cycles of 30 sec at 94°C, 30 sec at 55°C, and 1 min at 72°C, and a final extension at 72°C for 5 min. Amplification products were initially screened on 3% agarose gel and then analyzed using an automated DNA sequencer ABI

Table 1 — Characteristics of microsatellite markers isolated from *M. indica*

Locus	GenBank acc. no.	**Primers (5'-3')	Repeat motif	No. of alleles	Range of alleles size (bp)	<i>H_e</i>	<i>H_o</i>	PIC	PI
MiIHR38	EU421160	F: AGTGGTCATGATGGGGAAGA R: TGCAGCAATCAAAGAAGGTG	(TG) ₉	11	132-203	0.845	0.350	0.802	0.096
MiIHR39	EU421161	F: CGTGCAAAATTTTAGCCATT R: GAGAGAGCCCAAAGGGAGTT	(CAA) ₈	22	224-290	0.949	0.850	0.921	0.017
MiIHR40	EU421162	F: ACAGTCAGGCTTCCTCCAAG R: AGGGAACCTGGGGTAAGAAA	(GA) ₁₀	17	261-287	0.859	0.500	0.826	0.059
MiIHR41	EU421163	F: ACGCGTCATCATTTGTGTGT R: CCCACTCTACAGTATTGCTCTTAAA	(GA) ₁₀	14	158-212	0.899	0.500	0.866	0.047
MiIHR43	EU421165	F: AATCAAGGGGTTTCGCTTTT R: CCCACCCACAGCAGTAACCTA	(CT) ₁₀	16	196-268	0.846	0.500	0.814	0.062
MiIHR44	EU421166	F: GGGACGAAAATCGAACTCTTC R: GAACTTAAGCCAGCGCAAAG	(CT) ₄ (TC) ₂ TGAC (CT) ₅	12	117-192	0.864	0.350	0.827	0.071
MiIHR48	EU421170	F: CATGGAATGAAATGCTTTGG R: CGGCACCTGGATAATGAAAT	(TG) ₅ (GT) ₁₀	15	149-216	0.908	0.300	0.875	0.042
MiIHR50	EU421172	F: TCCACAATGGTGAAAACACG R: TGTGGGTCTTTGTGGGATT	(GT) ₅ (AT) ₄	13	132-185	0.835	0.650	0.793	0.097
MiIHR51	EU421173	F: CCGTCGTATGAATTGCTTGA R: ACCAAATTATCGCGATCTCC	(AAG) ₄ (AGGA) ₂	17	93-121	0.897	0.700	0.864	0.050
MiIHR52	EU421174	F: CCTCGCTTTGAATAATACCG R: TCAACACCGGACACATTGAT	(GA) ₃ GG (GA) ₇ (AG) ₃ GG (AG) ₇	21	112-206	0.953	0.600	0.925	0.017
MiIHR55	EU421177	F: GGAGAGCTGAGAGGCCATAA R: CCACGAGACGACAGTCAAAA	(TGG) ₂ (AG) ₇	21	225-301	0.956	0.450	0.929	0.016
MiIHR56	EU421178	F: ACAGGCTGCACAAAACACAA R: TTTTGCTCATGCAATTGAG	(AG) ₃ AA (AG) ₆ AA (AG) ₃	16	191-204	0.899	0.400	0.867	0.043
MiIHR57	EU421179	F: GATTATGGCCACGAAGATGG R: GGAAGATTGGAGCCTTCGT	(GAA) ₄ (GAG) ₂	14	155-209	0.909	0.200	0.877	0.041
MiIHR58	EU421180	F: TAGTGGTCGGCCTTCCTACA R: GGGTCTTTCCTTCGCCTATC	(CCT) ₂ TG (GTTCT) ₂	18	162-216	0.933	0.450	0.904	0.026
MiIHR59	EU421181	F: ACTGCTGTGGTGCTCCTTT R: ACTGCAAAGAGCACCGTTTT	(GTG) ₃ (TTG) ₉	13	138-175	0.909	0.500	0.876	0.044
MiIHR60	EU421182	F: TGCATCACCCCTTGGTTATT R: CATCATCTGCTCTCGCAAAA	(GTT) ₆ (GCT) ₄ (GAT) ₃ (GAA) ₅	20	244-301	0.871	0.400	0.841	0.046
MiIHR64	EU421186	F: GCAAGCAAGGGACAAAGAAG R: TGCACTCTGCTTTGTTCAGC	(CCT) ₃ (GCT) ₂ (GCA) ₂	17	143-209	0.946	0.350	0.917	0.021
MiIHR65	EU421187	F: TGATATGGGCGTAGCTGTTG R: TCTTCTGCGAGCTTCTGACA	(AGG) ₂ (AAG) ₃ GAG (GAT) ₃	11	192-228	0.659	0.450	0.624	0.182
MiIHR68	EU421190	F: GGAAAGGGCTTGGTGTGTGA R: CACTATCAACGCCTTCAGCA	(GA) ₆ (TGG) ₂ (AAG) ₂ (AG) ₂	21	161-222	0.954	0.750	0.926	0.017
MiIHR72	EU421194	F: CCGGGGGTTAGAATTACGAT R: AAACCTTGGCCTTGGCTATT	(AAG) ₂ (GAA) ₃ (TAG) ₂	23	121-169	0.965	0.850	0.938	0.012

H_e: Expected heterozygosity, *H_o*: Observed heterozygosity, PIC: Polymorphic information content (calculated using CERVUS), PI: Probability of identity (calculated using IDENTITY).

** Primer sequences without M13 (17 mer) tail sequence and pigtail

3730. Finally, the SSR alleles were scored from the raw data using peak scanner software (Applied Biosystems, USA).

The allelic composition of each accession and the number of total alleles were determined for each SSR locus. Putative alleles were indicated by the estimated size in bp. The genetic information was assessed by analyzing the following parameters: number of alleles per locus (A), observed heterozygosity (Ho, direct count), expected heterozygosity ($H_e = 1 - \sum p_i^2$ where p_i is the frequency of the i^{th} allele⁹) and Polymorphic Information Content (PIC)¹⁰.

$$\widehat{PIC}_i = 1 - \sum_{u=1}^k \bar{p}_{iu}^2 - \sum_{u=1}^{k-1} \sum_{v=u+1}^k 2\bar{p}_{iu}\bar{p}_{iv}$$

Where p_{lu} is the frequency of the u^{th} allele, and p_{lv} is the frequency of the v^{th} allele.

Additionally, the probability of identity (PI) that measures the probability of two randomly drawn diploid genotypes will be identical, assuming observed allele frequencies and random assortment¹¹, was analyzed using following computation

$$PI = 1 - \sum p_i^4 + \sum \sum (2p_i p_j)^2$$

Where p_i and p_j are the frequency of the i^{th} and j^{th} alleles, respectively.

The computations were performed with the CERVUS 3.0.3 software¹² and IDENTITY 4.0 software¹³. Genetic relationships among the genotypes were calculated by computing the dissimilarities through simple matching coefficient. A dendrogram was constructed through Ward's minimum variance method using DARWin 5 software¹⁴.

Results and Discussion

The genetic parameters for the 20 microsatellite markers are shown in Table 1. The analysis of 20 SSR primers in 20 cultivars detected a total of 332 alleles, with an average of 16.60 alleles per locus, ranging 11 to 23 alleles per locus. The expected heterozygosity ranged 0.659 to 0.965, with a mean of 0.892, and the observed heterozygosity ranged 0.350 to 0.850, with a mean of 0.505. The calculated PIC values ranged 0.624 to 0.938, with a mean of 0.860, showing that all the 20 STMS markers as informative markers. The maximum PI (0.182) was detected in MiIHR65 with 11 alleles and the minimum (0.012) in MiIHR60 with 23 alleles, with an average of 0.050. The total PI (probability of two cultivars sharing the same genetic profile by chance) was 1.06×10^{-28} .

The dendrogram obtained through Ward's minimum variance method resulted in the formation of two major

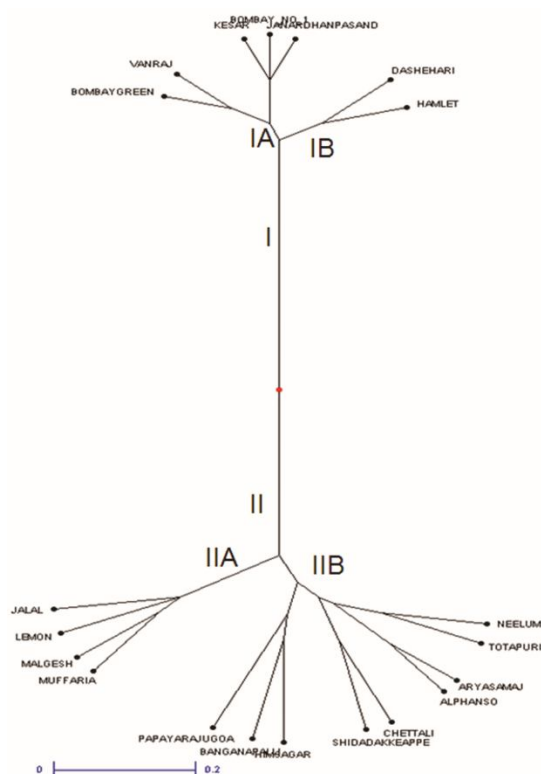


Fig. 1 — Cluster analysis of mango cultivars using twenty two microsatellite marker data.

clusters (Fig. 1) based on genetic dissimilarity. Cluster I included Bombay Green, Vanraj, Banganapalli, Bombay No. 1, Janardhan Pasand, Dashehari and Hamlet that have been further sub-grouped into two more clusters as IA and IB. Cluster II had Jalal, Lemon, Malgesh, Muffarai, Papayaraju Goa, Kesar, Himsagar, Shidadakke Appe, Chettalli, Alphonso, Arya Samaj, Totapuri and Neelum that have been further sub-grouped into two sub-clusters IIA and IIB.

The developed SSR markers exhibited relatively higher levels of heterozygosity and PIC values compared to earlier studies^{2,3}. The low PI values for these markers indicate their utilization in fingerprinting mango cultivars. These SSR markers would be further exploited for characterizing large mango germplasm collection, linkage map development, analysis of pedigree, population structure and marker-trait associations.

Acknowledgement

We thank the Department of Biotechnology (DBT), Government of India, New Delhi for funding the present study.

References

- 1 Mukherjee S K, Origin of mango, *Econ Bot*, 26 (1972) 260-264.
- 2 Duval M F, Bunel J, Sitbon C & Risterucci, A M, Development of microsatellite markers of mango, *Mol Ecol Notes*, 5 (2005) 824-826.
- 3 Honsho C, Nishiyama K, Eiadthong W & Yonemori, K, Isolation and characterization of new microsatellites markers in mango, *Mol Ecol Notes*, 5 (2005) 152-154.
- 4 Ravishankar K V, Mani B H, Anand L & Dinesh M R, Development of new microsatellite markers from Mango (*Mangifera indica*) and cross-species amplification, *Am J Bot*, 98 (2011) e96-e99.
- 5 Ravishankar K V, Anand L & Dinesh M R, Assessment of genetic relatedness among mango cultivars of India using RAPD markers, *J Hortic Sci Biotechnol*, 75 (2000) 198-201.
- 6 Glenn T C, & Schable T C, Isolating microsatellite DNA loci, *Methods Enzymol*, 395 (2005) 202-222.
- 7 Rozen S & Skaletsky H J, Primer3 on the www for general users and for biologist programmers, in *Bioinformatics methods and protocols: Methods in molecular biology*, edited by S Krawetz & S Misener (Humana Press, Totowa, NJ, USA) 2003, 365-386.
- 8 Schuelke M, An economic method for the fluorescent labeling of PCR fragment, *Nat Biotechnol*, 18 (2000) 233-234.
- 9 Nei M, Analysis of gene diversity in subdivided populations, *Proc Natl Acad Sci USA*, 70 (1973) 3321-3323.
- 10 Botstein D, White R L, Skolnick M & Davis R W, Construction of a genetic linkage map in man using restriction fragment length polymorphism, *Am J Hum Genet*, 32 (1980) 314-331.
- 11 Paetkau D, Calvert W, Stirling I & Strobeck C, Microsatellite analysis of population structure in Canadian polar bears, *Mol Ecol*, 4 (1995) 347-354.
- 12 Kalinowski S T, Taper M L & Marshall T C, Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment, *Mol Ecol*, 16 (2007) 1099-1106.
- 13 Wagner H W, & Sefc K M, IDENTITY 1.0 (Centre for Applied Genetics, University of Agricultural Sciences, Vienna, Austria) 1999. [<http://www.boku.ac.at/zag/forsch/identity.htm>]
- 14 Perrier X, Flori A & Bonnot F, Data analysis methods, in *Genetic diversity of cultivated tropical plants*, edited by P Hamon, M Seguin, X Perrier & J C Glaszmann (Science Publishers and CIRAD, Enfield, NH, USA) 2003, 43-76.