Molecular Breeding (2005) 15: 383–393
DOI: 10.1007/s11032-004-7982-x
© Springer 2005

Fingerprinting, embryo type and geographic differentiation in mango (*Mangifera indica* L., Anacardiaceae) with microsatellites

M.A. Viruel[1], P. Escribano[1], M. Barbieri[2], M. Ferri[2] and J.I. Hormaza[1,*]
[1]*Estación Experimental la Mayora – CSIC, 29750 Algarrobo-Costa, Málaga, Spain;* [2]*Facoltá di Agraria, Università degli studi di Modena e Reggio Emilia, Via Kennedy 17, 42100 Reggio Emilia, Italy;* **Author for correspondence (e-mail: ihormaza@eelm.csic.es; phone: 34-952-552-656; fax: 34-952-552-677)*

Received 13 July 2004; accepted in revised form 21 December 2004

Key words: Germplasm resources, Molecular markers, SSRs, Tropical fruits

Abstract

We report the sequence and variability parameters of 16 microsatellite primer pairs obtained from two mango (*Mangifera indica* L.) genomic libraries after digestion of DNA of the cultivar Tommy Atkins with *Hae*III and *Rsa*I and enrichment in CT repeats. Although no significant differences were recorded between the two libraries in the informativeness of the markers obtained, the *Rsa*I library was shown to be more useful than the *Hae*III taking into account the efficiency of the library and the feasibility of clone sequencing. The polymorphism revealed by those microsatellites was evaluated in a collection of 28 mango cultivars of different origins. A total of 88 fragments were detected with the 16 simple sequence repeats (SSRs) with an average of 5.5 bands/SSR. Two primer pairs amplified more than a single locus. The mean expected and observed heterozygosities over the 14 single-locus SSRs averaged 0.65 and 0.69 respectively. The total value for the probability of identity was 2.74×10^{-9}. The SSRs studied allowed the unambiguous identification of all the mango genotypes studied and this discrimination can be carried out with just three selected microsatellites. UPGMA cluster analysis and Principal coordinates analysis group the genotypes according to their origin and their classification as monoembryonic or polyembryonic types reflecting the pedigree of the cultivars and the movement of mango germplasm. The results demonstrate the usefulness of microsatellites for studies on identification, variability, germplasm conservation, domestication and movement of germplasm in mango.

Introduction

Mango (*Mangifera indica* L. $2n = 40$) is a member of the Anacardiaceae, a family of mainly tropical species with a few representatives in temperate regions. The family contains over 600 species classified into 70 genera that include other cultivated species such as pistachio (*Pistacia vera* L.) or cashew (*Annacardium occidentale* L.). The genus *Mangifera* contains about 70 species mostly restricted to tropical Asia and can be divided into two subgenera (Limus and Mangifera) with several sections (Kostermans and Bompard 1993). Mango is diploid ($2n = 40$) with a genome of 8.8×10^8 bp, which is about two and a half times that of *Arabidopsis thaliana* (Bennet and Leitch 2003). Total world mango production has reached over 26 million tons in 2002 making mango one of the five most important fruit species worldwide (together with bananas, oranges, grapes and apples). A few countries (India, China, Thailand, Mexico, Pakistan and Indonesia) account for over 75% of

world production, India being the main producer with over 40% (over 11 millions of tons) (FAOSTAT 2004).

Historical records provide conflicting accounts for mango origin and distribution. Although some authors have considered India as the centre of origin due to the high degree of mango diversity observed in that country (Ravishankar et al. 2000), taxonomic and molecular evidence also supports an evolution of mango within a larger area including northwestern Myanmar, Bangladesh and Northeastern India (Mukherjee 1997). In any case, it is likely that mango cultivation originated in India where over 1000 varieties are recognized, most of them selections from naturally occurring open-pollinated seedlings (Iyer and Degani 1997). Traders spread mango cultivation outside its centre of origin and domestication to other tropical and subtropical regions where selections of the cultivars best adapted to particular conditions was made. Portuguese spread mangoes to eastern and western Africa and Brazil and from Brazil probably to the Caribbean islands (Nakasone and Paull 1998), whereas Spaniards introduced this crop to Mexico from the Philippines (Singh 1978). In the 19th century mangoes were introduced to Florida first from the Caribbean and later from India. The only Indian mango genotype that survived, 'Mulgoba', showed a superior quality that any mango previously grown in Florida and was cultivated until 1912, when one of its seedlings, 'Haden', took its place (Campbell 1992). For 40 years 'Haden' remained the most important commercial mango in Florida and it is present in the pedigree of many of the Florida cultivars. Introduction of new germplasm from Asia, Africa, Central and South America and Caribbean and Pacific Islands has continued in Florida (Campbell 1992) and, as a result, Florida is now considered as a secondary centre of diversity for mango. The adaptability and fruit qualities of this new group of cultivars, called the Florida mangos, has allowed their spread to new growing areas (Mukherjee 1997).

Mango cultivars are commonly divided into two groups based on their mode of reproduction from seeds, monoembryonic and polyembryonic. Monoembryonic mango seeds contain a single zygotic embryo whereas polyembryonic seeds contain at least one embryo and usually one of them is zygotic and the remaining derive from the nucellus. Monoembryonic mangos are mostly subtropical (Indian types) and polyembryonic mangos are mostly tropical (southeast Asian types). There is considerable confusion in mango cultivar nomenclature because many clonally propagated mango cultivars have unique local and regional names and the spelling and name variants have been translated to the Roman alphabet. As in other fruit tree species, mango cultivars are currently identified on the basis of morphological traits based on descriptors (IPGRI 1989). However, cultivar identification based exclusively on phenotypic traits is inaccurate due to the influence of the environment and the limiting number of discriminating traits. Thus, recently, as in other fruit tree species (Wünsch and Hormaza 2002), molecular identification of mango cultivars has been carried out with different molecular systems as isozymes (Degani et al. 1990), minisatellites (Adato et al. 1995), ISSRs (Eiadthong et al. 1999), AFLPs (Eiadthong et al. 2000; Kashkush et al. 2001) and RAPDs (Schnell et al. 1995; Lopez-Valenzuela et al. 1997; Ravishnakar et al. 2000; Hemanth Kumar et al. 2001; Karihaloo et al. 2003). However, microsatellites or simple sequence repeats (SSRs), in addition to their usefulness in mapping and breeding (McCouch et al. 1997), have become the markers of choice for fingerprinting purposes in most plant species (Gupta and Varshney 2000) due to their high polymorphism, codominancy and reproducibility. In species such as mango and most fruit trees, where no or few sequences are available in public sequence databases or in related species, the major drawback for the application of microsatellite analyses is the difficulty and cost of the isolation process that usually involves the construction and screening of genomic libraries to find the sequences flanking the repeat regions. The low yield obtained using conventional library screening methods has been greatly improved by the use of enrichment protocols for specific microsatellite repeats (Zane et al. 2002; Viruel and Hormaza 2004). Thus, in this paper, we report the development of a first set of 16 microsatellites for mango using two genomic libraries enriched in CT repeats. Their polymorphism has been studied in a group of 28 mango genotypes and their value for cultivar identification, variability analysis and mango germplasm movement is discussed.

Materials and methods

Plant material and genomic DNA extraction

Twenty-eight mango genotypes obtained from different geographical areas and maintained in the mango collection of the Estación Experimental de la Mayora in Málaga (Spain) (Table 1) were used for this study. The genotypes include both cultivars and rootstocks as well as monoembryonic and polyembryonic types. DNA extraction was performed on young leaves following the protocol described in Viruel and Hormaza (2004).

Construction and screening of a microsatellite enriched library

Two small-insert libraries enriched with $(CT)_n$ sequences were developed from DNA of the mango cultivar Tommy Atkins digested with *Hae*III and *Rsa*I following the procedures described in Viruel and Hormaza (2004). Positive clones were sequenced to identify the flanking regions that were used to design appropriate primer pairs with the program Primer3 (Whitehead Institute for Biochemical Research, Cambridge, MA, USA).

SSR analysis

The primers obtained were initially studied in a reduced group of 4 mango genotypes by PCR amplification in 15 μl vol containing 16 mM $(NH_4)_2SO_4$, 67 mM Tris–HCl pH 8.8, 0.01% Tween20, 2 mM $MgCl_2$, 0.1 mM each dNTP, 0.4 μM each primer, 25 ng genomic DNA and 0.5 units of BioTaq^{TM} DNA polymerase (Bioline, London, UK). Reactions were carried out on an I-cycler (Bio-Rad Laboratories, Hercules, CA, USA) thermocycler using the following temperature profile: an initial step of 1 min at 94 °C, 35 cycles of 30 s at 94 °C, 30 s at 55 °C and 1 min at 72 °C, and a final step of 5 min at 72 °C. Amplification

Table 1. List of the mango genotypes included in this work.

Genotype	Main use (1)	Geographic origin	Embryo type
Ataulfo	cv	Mexico	Polyembryonic
Carabao	cv	Philippines	Polyembryonic
Dusheri	cv	Northern India	Monoembryonic
Edward	cv	Florida (USA)	Monoembryonic
Glenn	cv	Florida (USA)	Monoembryonic
Gomera-1	rs	Canary Islands (Spain)	Polyembryonic
Gomera-3	rs	Canary Islands (Spain)	Polyembryonic
Gouveia	cv	Hawaii (USA)	Monoembryonic
Haden	cv	Florida (USA)	Monoembryonic
Davis Haden	cv	Florida (USA)	Monoembryonic
Irwin	cv	Florida (USA)	Monoembryonic
Keitt	cv	Florida (USA)	Monoembryonic
Kensington	cv	Australia	Polyembryonic
Kent	cv	Florida (USA)	Monoembryonic
Langra	cv	Northern India	Monoembryonic
Lippens	cv	Florida (USA)	Monoembryonic
Manila	cv	Philippines	Polyembryonic
Manzanillo Nuñez	cv	Mexico	Monoembryonic
Maya	cv	Israel	Monoembryonic
Naomi	cv	Israel	Monoembryonic
Osteen	cv	Florida (USA)	Monoembryonic
Ott	cv	California (USA)	Monoembryonic
Peach	rs	Florida (USA)	Polyembryonic
Sensation	cv	Florida (USA)	Monoembryonic
Tommy Atkins	cv	Florida (USA)	Monoembryonic
Turpentine	rs	Florida (USA)	Polyembryonic
Van Dyke	cv	Florida (USA)	Monoembryonic
Zill	cv	Florida (USA)	Monoembryonic

(1) cv = cultivar, rs = rootstock

products were resolved in 3% high resolution agarose (Metaphor, FMC Bioproducts, Rockland, ME) gel electrophoresis (Hormaza 2002) and the primers that showed clear and scorable amplification patterns were selected for further SSR analysis. The selected SSRs (called LMMA followed by a consecutive number) were analyzed in the 28 mango genotypes using a CEQ^{TM}sans 8000 capillary DNA analysis system (Beckman Coulter, Fullerton, CA, USA). PCR reactions were performed as previously described, except that reverse primers of each primer pair were labeled with WellRED fluorescent dyes D2, D3 and D4 (Proligo, Paris, France). The analyses were repeated at least twice to assure the reproducibility of the results. To check the consistency of the amplification patterns obtained, for three genotypes ('Edward', 'Davis Haden' and 'Langra') the analyses were performed on DNA collected from leaves of 3 different branches in two different trees.

Allelic composition of each accession and the number of total alleles was determined for each SSR locus. Putative alleles were indicated by the estimated size in bp. The genetic information was assessed only for single locus SSRs using the following parameters: number of alleles per locus (A), observed heterozygosity (Ho, direct count), expected heterozygosity ($\mathrm{He} = 1 - \sum pi^2$ where pi is the frequency of the ith allele, Nei 1973), effective number of alleles ($\mathrm{Ne} = 1/1 - \mathrm{He}$), Wright's fixation index ($F = 1 - H\mathrm{o}/\mathrm{He}$) (Wright 1951) and the probability of identity ($\mathrm{PI} = 1 - \sum pi^4 + \sum\sum(2pipj)^2$, where pi and pj are the frequency of the ith and jth alleles respectively) that measures the probability that two randomly drawn diploid genotypes will be identical assuming observed allele frequencies and random assortment (Paetkau et al. 1995). The computations were performed with the programs IDENTITY 1.0 (Wagner and Sefc 1999) and GENEPOP 3.4 (Raymond and Rousset 1995).

A homology search of all the sequenced clones was performed through NCBI Blast to determine whether the sequences had any biological significance.

Genetic diversity

Genetic relationships among the genotypes studied were calculated using UPGMA cluster analysis and Principal coordinates analysis (PCA) of the similarity matrix obtained from the proportion of shared amplification fragments (Nei and Li 1979). The cophenetic coefficient was computed for the dendrogram after the construction of a cophenetic matrix. All those analyses were computed with the program NTSYSpc 2.11 (Exeter Software, Stauket, NY, USA). The robustness of the nodes of the dendrogram was assessed with bootstrap analysis using 2000 iterations with the WinBoot program (Yap and Nelson 1996).

Results

Microsatellite development

A total of 384 clones, 192 of each of the two enriched genomic libraries were hybridized with the probe $(CT)_{15}$ in order to test the presence of the microsatellite. One-hundred and forty clones in the *Hae*III library and 136 clones in the *Rsa*I library produced a positive signal, resulting in enrichments of 73% and 71%, respectively. Of those, 14 clones from each library were sequenced resulting in a total of 25 readable sequences. In all except one clone from each library, the presence of the microsatellite sequence CT/AG was confirmed. Two additional clones, one from each library, were rejected due to the proximity of the microsatellite to the end of the insert. Finally, specific primers for 21 clones, 10 from the *Hae*III and 11 from the *Rsa*I libraries were designed.

The 21 selected SSRs were tested in a group of four mango genotypes in high resolution agarose. All of them produced clear amplification patterns and were further analyzed with the automatic sequencer in all the genotypes. Seven SSRs (LMMA1-LMMA7) from the *Hae*III library and 9 (LMMA8-LMMA16) from the *Rsa*I library produced simple and clear amplification patterns with the automatic sequencer. All the microsatellites were simple. In the *Hae*III library 4 were perfect and 3 imperfect, whereas in the *Rsa*I library 4 were perfect and 5 imperfect. The primer sequences for those loci are given in Table 2 and the microsatellite sequences have been deposited in Genbank (accession numbers from AY628373 to AY628388).

NCBI blast searches showed a highly significant homology of the two microsatellite flanking

Table 2. Locus name, polymorphic microsatellite primer sequences, repeat types, length of the amplified fragments and variability parameters in 28 mango cultivars.

SSR	GenBank accession number	Primer sequence (5′–3′)	Repeat motif	Type[c]	Expected Size (bp)	No. of alleles	Effective No. of alleles (Ne)	Observed heterozygosity (Ho)	Expected heterozygosity (He)	Wright's fixation index (*F*)	Prob. of identity (PI)
LMMA1	AY628373	F:ATGGAGACTAGAATGTACAGAG R:ATTAAATCTCGTCCACAAGT	(GA)13	I	202	9	5.3	0.86	0.81	−0.057	0.11
LMMA2[a]	AY628374	F:AAATAAGATGAAGCAACTAAAG R:TTAGTGATTTTGTATGTTCTTG	(GA)11	P	287						
LMMA3	AY628375	F:AAAAACCTTACATAAGTGAATC R:CAGTTAACCTGTTACCTTTTT	(GA)16	P	207	6	4.2	0.79	0.76	−0.031	0.17
LMMA4	AY628376	F:AGATTTAAAGCTCAAGAAAAA R:AAAGACTAATGTGTTTCCTTC	(GA)13	P	241	3	3.0	0.68	0.66	−0.023	0.33
LMMA5	AY628377	F:AGAATAAGCTGATACTCACAC R:TAACAAATATCTAATTGACAGG	(GA)9	I	283	3	1.3	0.25	0.22	−0.123	0.66
LMMA6[a]	AY628378	F:ATATCTCAGGCTTCGAATGA R:TATTAATTTTCACAGACTATGTTCA	(GA)14	P	118						
LMMA7	AY628379	F:ATTTAACTCTTCAACTTTCAAC R:AGATTTAGTTTTGATTATGGAG	(CT)15	I	212	6	3.7	0.82	0.73	−0.125	0.17
Mean	*Hae*III library					5.4	3.5	0.68	0.64	−0.070	0.29
LMMA8	AY628380	F:CATGGAGTTGTGATACCTAC R:CAGAGTTAGCCATATAGAGTG	(GA)12	P	271	5	2.9	0.75	0.66	−0.135	0.26
LMMA9	AY628381	F:TTGCAACTGATAACAAATATAG R:TTCACATGACAGATATACACTT	(GA)13	I	185	6	4.5	0.86	0.78	−0.100	0.14
LMMA10	AY628382	F:TTCTTTAGACTAAGAGCACATT R:AGTTACAGATCTTCTCCAATT	(GA)10	P	191	4	3.0	0.61	0.67	0.091	0.26
LMMA11[b]	AY628383	F:ATTATTTACCCTACAGAGTGC R:GTATTATCGGTAATGTCTTCAT	(GA)12	I	244	7	4.4	0.64	0.77	0.166	0.14
LMMA12	AY628384	F:AAAGATAGCATTTAATTAAGGA R:GTAAGTATCGCTGTTTGTTATT	(GA)13	I	206	4	2.7	0.68	0.63	−0.085	0.27
LMMA13[b]	AY628385	F:CACAGCTCAATAAACTCTATG R:CATTATCCCTAATCTAATCATC	(GA)17	I	172	7	3.32	0.86	0.70	−0.226	0.21
LMMA14	AY628386	F:ATTATCCCTATAATGCCCTAT R:CTCGGTTAACCTTTGACTAC	(CT)10	P	170	4	1.9	0.39	0.47	0.165	0.52
LMMA15	AY628387	F:AACTACTGTGGCTGACATAT R:CTGATTAACATAATGACCATCT	(CT)11	P	215	6	2.0	0.61	0.49	−0.238	0.40
LMMA16	AY628388	F:ATAGATTCATATCTTCTTGCAT R:TATAAATTATCATCTTCACTGC	(GA)17	I	233	4	3.6	0.86	0.72	−0.191	0.24
Mean	*Rsa*I library					5.2	3.1	0.70	0.65	−0.060	0.27
MEAN	TOTAL					5.3	3.3	0.69	0.65	−0.065	0.28

[a]: Multi-locus SSR.
[b]: Loci with significant *F*.
[c]I: imperfect; P: perfect.

regions of one of the clones sequenced (LMMA9 accession number AY628381) to a mango beta-D-galactosidase mRNA (accession number AF004812). One of the regions showed an alignment score (S) of 204 with a probability value (E) of 1.0E-49 and 99% homology (106 of 107 nt). The other region showed an alignment score (S) of 105 with a probability value (E) of 7.0E-20 and 93% homology (72 of 77 nt). Additional lower significant hits were obtained with beta-galactosidase sequences of other plant species such as *Vigna radiata* (4e-09) or *Asparagus officinalis* (2e-04).

SSRs polymorphism and genetic diversity

The analysis of the 16 SSRs in 28 mango genotypes detected a total of 88 bands, with an average of 5.5 bands/SSR, ranging from 3 to 9 bands/SSR. In 14 of the SSRs, one or two bands were present in each genotype, suggesting the detection of a single locus. In LMMA2, a band of 413 pb was fixed in all the genotypes, and one or two additional bands were obtained in all genotypes except in 'Davis Haden' where a total of 4 bands were amplified. In LMMA6, only 'Langra' showed 3 bands and in the rest of the genotypes 1 or 2 bands were amplified. In order to check the possibility of intraclonal variability, that has been reported for some hypervariable SSRs (Lian et al. 2004), amplification patterns were compared among different branches and trees of 'Davis Haden' and 'Langra' as well as of 'Edward', used as a control. No differences among trees and branches were observed. Consequently, for all the loci except LMMA2 and LMMA6 the genotypes studied were considered homozygous and heterozygous when one or two fragments were present per locus, respectively (Callen et al. 1993).

The variability parameters for the 14 single-locus SSRs are shown in Table 2. The number of alleles detected ranged from 3 to 9 in the *Hae*III library, and from 4 to 7 in the *Rsa*I library, with an average of 5.4 and 5.2 alleles/locus respectively (3.5 and 3.1 for the effective number of alleles, Ne). Allele frequencies ranged from 0.018 to 0.855 (mean 0.209). From the 74 putative alleles detected, 19 (26%) showed frequencies lower than 0.05. The locus LMMA15 accumulated 4 rare alleles from a total of 6 alleles. One allele from the locus LMMA5 was fixed in all the genotypes with a frequency of 0.855. Six genotypes ('Ataulfo', 'Davis Haden', 'Edwards', 'Maya', 'Peach' and 'Van Dyke') showed the presence of one genotype-specific fragment. The Indian genotypes ('Langra' and 'Dusheri') present two genotype-specific fragments each and three additional fragments only present in them. Besides the fragment exclusive of 'Ataulfo', four additional fragments are specific of the Philippine cultivars, one exclusive of 'Manila' and 'Ataulfo', two exclusive of 'Manila' and 'Carabao' and one present in 'Manila', 'Ataulfo' and 'Carabao'.

Observed heterozygosity ranged from 0.25 to 0.86 in the *Hae*III library (mean of 0.68) and from 0.39 to 0.86 in the *Rsa*I library (mean of 0.70) with a mean for all the loci of 0.69. Expected heterozygosity ranged from 0.22 to 0.81 in the *Hae*III library (mean of 0.64) and from 0.47 to 0.78 in the *Rsa*I library (mean of 0.65) with a mean for all the loci of 0.65. Expected and observed heterozygosity values were compared using the fixation index (F), which had an average over all the single locus SSRs of -0.070 and -0.060 for the *Hae*III and *Rsa*I libraries respectively with values between -0.238 (LMMA15) and 0.166 (LMMA11) (Table 2). For 11 loci this parameter was negative (excess of heterozygotes observed) and for 3 (LMMA10, LMMA11 and LMMA14) was positive (excess of homozygotes observed). Allele frequencies of the 14 single-locus loci in the 28 genotypes do not deviate from those expected in a population in Hardy–Weinberg equilibrium. However two of the loci (LMMA13, $F = -0.226$ and LMMA11, $F = 0.166$) differed significantly ($p < 0.05$) from Hardy–Weinberg expectations.

Regarding the probability of identity, the maximum (0.66) was detected in LMMA5 with 3 alleles, one of which is fixed, and the minimum (0.11) in LMMA1 with 9 alleles. The average was 0.29 within the SSRs from the *Hae*III library and 0.27 from the *Rsa*I library (Table 2). Total probability of identity (probability of two cultivars sharing the same genetic profile by chance) was 2.74×10^{-9}.

Molecular fingerprinting and genetic relationships of the mango genotypes

The analysis of the 88 bands detected by the 16 SSRs allowed the unambiguous discrimination of

the 28 mango genotypes included in this study. The same discrimination was obtained with only 3 SSRs: LMMA1, LMMA11 and LMMA12.

Similarity values among different mango cultivars ranged from 0.24 ('Langra'-'Gomera 3') to 0.93 ('Manila'-'Carabao'). The dendrogram obtained after UPGMA analysis with the 28 mango genotypes is shown in Figure 1. A single tree was obtained with a high cophenetic correlation coefficient between the cophenetic matrix and the similarity matrix of 0.87, which reflects a good fit. Three main groups were obtained and principal coordinates analysis (Figure 2) showed a similar distribution of the genotypes. The first three coordinates account for about 37% of the total variance (18%, 10% and 9%, respectively). Bootstrap analysis showed low values for most of the branches (< 50%) except for the node of 'Ataulfo', 'Carabao' and 'Manila' with 60% (in fact 'Carabao' and 'Manila' node is 100%), 'Haden' and 'Manzanillo Nuñez' with 68%, and 'Ott' and 'Gouveia' with 86%.

Discussion

In this work, two mango genomic libraries enriched for CT/AG repeats have been developed simultaneously, after digestion of the DNA from the cultivar Tommy Atkins with the restriction enzymes *Hae*III (methylation-sensitive) and *Rsa*I (methylation-insensitive). The success of the enrichment procedure was similar between the two libraries, 71% in *Hae*III and 73% in *Rsa*I. Taking into account those enrichment percentages and the number of valid SSRs from the sequenced clones (7 and 9 valid SSRs from 14 sequenced clones in the *Hae*III and *Rsa*I libraries, respectively) the total efficiency is higher for *Rsa*I (47%) than for *Hae*III (36%). Moreover, sequencing of the clones was easier in the *Rsa*I library, since the average clone size is about 255 lower than in the *Hae*III library, allowing the complete sequencing of the selected clones in a single reaction. No significant differences in the genetic informativeness obtained were recorded between the two libraries in the 28 mango genotypes studied. Consequently, technically, the *Rsa*I library appears to be more convenient than *Hae*III. However, the location of the markers in a mango genetic map will be needed to determine if the enzyme used for library development has any influence in the isolation of SSRs from specific genomic regions.

Although fragment segregation in the progeny obtained from a cross is the only way to assign alleles to a particular locus, the fragment distribution in the 28 genotypes studied indicates that all SSRs except two (LMMA2 and LMMA6) detect single loci. In LMMA2, one of the bands is fixed and all the genotypes present one or two

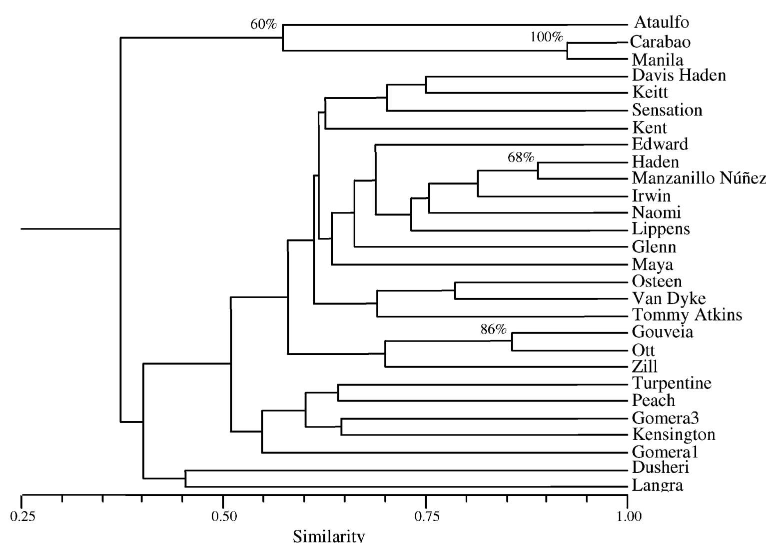

Figure 1. Dendrogram of the 28 mango cultivars studied based on UPGMA analysis using the similarity matrix generated by the Nei and Li coefficient with 88 SSR fragments. Bootstrap values larger than 50% are reported above branches.

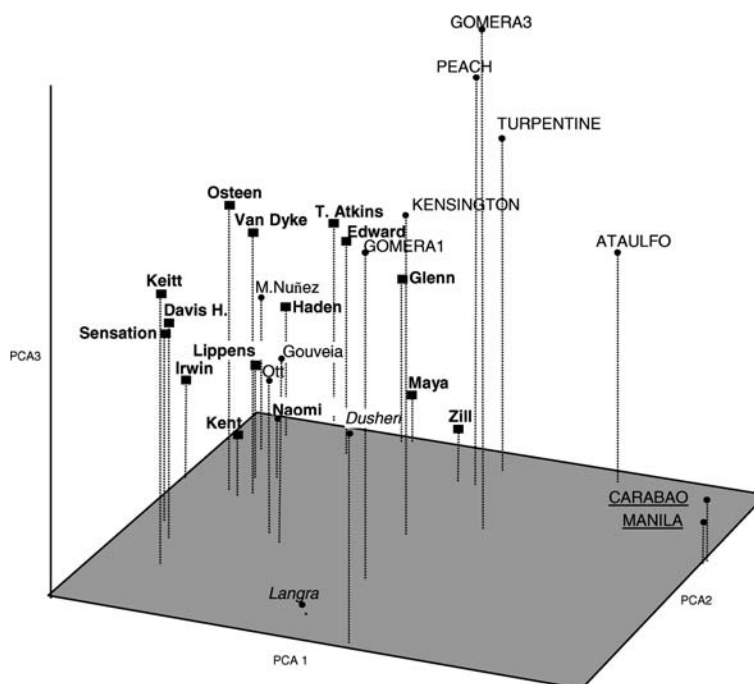

Figure 2. Principal coordinates analysis of the pairwise difference matrix for 28 mango genotypes with 88 SSR fragments. Names of polyembryonic cultivars are shown in capital letters, with Philippines cultivars underlined. In the monoembryonic group, Indian cultivars are shown in *italics*, and Floridian cultivars in bold. ■ indicates cultivars of the Haden family.

additional bands except 'Davis Haden' that presents 3 additional bands. A similar situation is found with LMMA6, where in all the genotypes, except 'Langra' that presents three bands, the amplification pattern seems to indicate the detection of a single locus. Mango has been described as allopolyploid (Mukherjee 1950) and these results suggest a complete diploidization in this species. The presence of additional bands in 'Langra' and 'Davis Haden' could be due to remnants of the ancestral poliploidization and/or to genomic rearrangements accumulated in the course of the long period of cultivation of this species where a wide variability in clones of the same cultivar has been repeatedly reported, particularly, in Indian cultivars (Iyer and Degani 1997).

In spite of the close genetic relationships of many of the cultivars studied, a relatively high average number of alleles per locus (5.3) and expected heterozygosity (0.65) was obtained in this work. Similarly, the low PI (0.28) as well as the low number of fixed alleles (only one in 74 alleles) compared to the number of genotype-specific alleles (6 in 74) indicate that a high variability is detected in the genotypes studied and, consequently, that SSR markers will become a useful tool for genetic variation studies and for genotype identification and similarity analyses in mango. In fact, just 3 selected loci are sufficient to unambiguously identify each of the 28 genotypes analyzed in this work.

The mean F was not different from zero, indicating a global behavior of this group of genotypes similar to a random mating collection. This could be due to the fact that mango cultivars are the result of selection from open-pollinated seedlings, most of them from chance seedlings from natural cross-pollinations. Only a number of cultivars from Florida have resulted from the selection of seedlings from known maternal plants. Although 15 of the 28 genotypes studied are from the Florida breeding program high levels of polymorphism are maintained, reflecting the continuous introduction of plant material from different geographical areas and genetic backgrounds (Campbell 1992).

In addition to their usefulness for fingerprinting mango cultivars, the codominant nature of SSRs allows a better understanding of the pedigree relationships among the cultivars studied. Pedigree

relationships in mango are complex since very few commercial cultivars are the result of controlled breeding crosses. The results obtained in this work confirm most of the pedigree relationships of the cultivars of the Florida group included in this study. Thus, 'Glenn', 'Osteen', 'Lippens', 'Tommy Atkins' and 'Maya' are offspring of 'Haden' (Oppenheimer 1978; Campbell 1992; Adato et al. 1995) and 'Irwin' is offspring of 'Lippens' (Campbell 1992). The data obtained also indicate that 'Edward' seems to be progeny of 'Haden' but not from a cross with 'Carabao' according to Adato et al. (1995) and Degani et al. (1990). In the case of 'Zill', this cultivar is considered as an offspring of 'Haden' (Campbell 1992) but our results show that this is not the case with the 'Zill' genotype studied in this work. 'Manila' has been described sometimes as 'Carabao' (Mathews and Litz 1992). In our work, both genotypes are distinguishable although genetically very close (similarity of 0.93). 'Davis Haden' has been usually described as a sport of 'Haden' (Iyer and Degani 1997) but our results suggest that it is more likely that 'Davis Haden' is a seedling of 'Haden' as suggested by Morton (1987).

UPGMA clustering reflects the history of breeding and selection of the cultivars studied, grouping the cultivars according to geographical origin and type. In this work three main clusters can be observed in the dendrogram (Figure 1). The most distinct cluster includes three polyembryonic cultivars either from Philippines ('Carabao') or from Mexico but related to genotypes from Philippines ('Ataulfo' and 'Manila'; Sauer 1993). Another small differentiated cluster includes two ancient cultivars from the north of the Indian subcontinent, 'Dusheri' or 'Dasherhari', and 'Langra' (Singh 1978). The larger main cluster (23 of the 28 cultivars studied) includes cultivars of the Florida group or somehow related to them. The separation of the Indian and Philippine clusters from the cluster with the Florida material is also reflected in the PCA (Figure 2). This separation between the Florida group and the rest of the groups was already reported with RAPDs (Schnell et al. 1995; Lopez-Valenzuela et al. 1997), but, in contrast to the results of López-Valenzuela et al. (1997), no specific markers of the Floridian group have been found in this work. However, five fragments are present only in the Philippine cultivars, and seven in the Indian cultivars. Both the UPGMA dendrogram and the PCA show that the Philippine cultivars are more separated from the Florida group than the two Indian cultivars studied; this is further stressed by the bootstrap analysis. This result is not surprising since Indian cultivars such as 'Mulgoba' (parent of 'Haden' and 'Keitt') or 'Totapuri' (involved in the 'Kent' pedigree) have been used from the beginning in the Florida breeding program. However, on the other hand, the separation of the two Indian cultivars from the Florida group can be explained by at least two reasons: the large intrapopulation and interregional variability among Indian cultivars (Karihaloo et al. 2003) and the fact that 'Langra' and 'Dusheri' originated in northern India whereas 'Totapuri' and 'Mulgoba' are from the south.

The main Florida cluster can be further divided into two different subclusters. The largest subcluster includes monoembryonic cultivars mainly related to Indian cultivars: 'Kent' (related to 'Totapuri'), 'Keitt' and 'Haden' (seedlings of 'Mulgoba'), and the cultivars directly or indirectly related to 'Haden' ('Davis Haden', 'Edward', 'Glenn', 'Lippens', 'Irwin', 'Osteen', 'Tommy Atkins' and 'Zill'). Florida genotypes of unknown parentage (as 'Sensation' and 'Van Dyke') are also included in this subgroup as well as two genotypes from Israel that are descendant of cultivars from Florida ('Maya', seedling of 'Haden' and 'Naomi', seedling of 'Palmer'), two genotypes of unknown parentage from Hawaii ('Gouveia') and California ('Ott') and another from Mexico ('Manzanillo Nuñez'). The presence of cultivars from other growing areas such as Hawaii, California or Mexico in this subgroup is not unexpected since Florida selections were extended worldwide and in several countries they are grown together with traditional local cultivars. In fact, the Mexican cultivar Manzanillo Nuñez originated as a chance seedling in an area where the main cultivars are from Florida (Nunez-Elisea 1984).

A second subgroup includes only polyembryonic types from different origins ('Turpentine', 'Peach', 'Gomera 3', 'Kensington' and 'Gomera 1'). Initial mango introductions in Florida in the second half of the 19th century involved polyembrionic seeds from the Caribbean such as 'Peach' and 'Turpentine' (Campbell 1992). In addition to the two cultivars from Florida in this group we can find two cultivars from the Canary Islands ('Gomera 1' and 'Gomera 3') probably originating from Cuba

(Galan-Sauco 1979) and 'Kensington', a polyembrionic cultivar discovered in the 1800s in Australia of unknown origin but probably originating from seeds imported from India (Bally 1998).

This work represents the first SSRs developed in mango. These SSRs have been shown to be highly efficient for genotype identification in this species. Moreover, the particular characteristics of microsatellites make such markers very interesting for variability and pedigree studies. The cultivars can be grouped according to their geographical origin and pedigree history and the two main types of mangoes (monoembryonic and polyembrionic) are clearly differentiated. Further studies involving SSRs with a larger number of genotypes will be needed to elucidate the process of mango domestication and movement of plant material among different countries.

Acknowledgements

We gratefully acknowledge J.M. Farré, J.M. Hermoso and E. Guirado for valuable information on mango germplasm. Financial support for this work was provided by MCYT (Project Grant AGL2001-2414-C04-03).

References

Adato A., Sharon D., Lavi U., Hillel J. and Gazit S. 1995. Application of DNA fingerprints for identification and genetic analyses of mango (*Mangifera indica*) genotypes. J. Am. Soc. Hortic. Sci. 120: 259–264.

Bally I.S.E. 1998. Mango Variety Kensington Pride. Queensland Horticultural Institute, QDPI, Brisbane, Australia.

Bennett M.D. and Leitch I.J. 2003. Angiosperm DNA C-values Database (Release 4.0, Jan. 2003), http://www.rbgkew.org.uk /cval/homepage.html (accessed July 2004).

Callen D.F., Thompson A.D., Shen Y., Phillips H.A., Richards R.I., Mulley J.C. and Sutherland G.R. 1993. Incidence and origin of 'null' alleles in the $(AC)_n$ microsatellite markers. Am. J. Human. Genet. 52: 922–927.

Campbell R.J. 1992. A Guide to Mangos in Florida. Fairchild Tropical Gardens, Miami, Florida, USA.

Degani C., El-Batsri R. and Gazit S. 1990. Enzyme polymorphisms in mango. J. Am. Soc. Hortic. Sci. 115: 844–847.

Eiadthong W., Yonemori K., Sugiura A., Utsunomiya N. and Subhadrabandhu S. 1999. Identification of mango cultivars of Thailand and evaluation of their genetic variation using the amplified fragments by simple sequence repeat -(SSR)-anchored primers. Sci. Hort. 82: 57–66.

Eiadthong W., Yonemori K., Kanzaki S. and Sugiura A. 2000. Amplified fragment length polymorphism analysis for studying genetic relationship among *Mangifera* species in Thailand. J. Am. Soc. Hortic. Sci. 125: 160–164.

FAOSTAT. 2004. FAO Statistics Database on the World Wide Web. http://apps.fao.org (accessed April 2004).

Galan-Sauco V. 1979. Actual situation, problems and prospects of mango development in the Canary islands. Acta Horticulturae 102: 7–14.

Gupta P.K. and Varshney R.K. 2000. The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis in bread wheat. Euphytica 113: 163–185.

Hemanth Kumar N.V., Narayanswamy P., Theertha Prasad D., Mukunda G.K. and Sondur S.N. 2001. Estimation of genetic diversity of commercial mango cultivars using RAPD markers. J. Hortic. Sci. Biotech. 76: 529–533.

Hormaza J.I. 2002. Molecular characterization and similarity relationships among apricot (*Prunus armeniaca* L.) genotypes using simple sequence repeats. Theor. Appl. Genet. 104: 321–328.

IPGRI. 1989. Descriptors for Mango. International Board for Plant Genetic Resources, Rome, Italy.

Iyer C.P.A. and Degani C. 1997. Classical breeding and genetics. In: Litz R.E. (ed.), The Mango, Botany, Production and Uses. CAB International, Wallingford, UK.

Karihaloo J.L., Dwivedi Y.K., Archak S. and Gaikwad A.B. 2003. Analysis of genetic diversity of Indian mango cultivars using RAPD markers. J. Hortic. Sci. Biotech. 78: 285–89.

Kashkush K., Jinggui F., Tomer E., Hillel J. and Lavi U. 2001. Cultivar identification and genetic map of mango (*Mangifera indica*). Euphytica 122: 129–136.

Kostermans A.J.G.H. and Bompard J.M. 1993. The Mangoes: Their Botany, Nomenclature, Horticulture and Utilization. IBPGR Academic Press, London, UK.

Lian Ch., Oishi R., Miyashita N. and Hogetsu T. 2004. High somatic instability of a microsatellite locus in a clonal tree *Robinia pseudoacacia*. Theor. Appl. Genet. 108(5): 836–841.

Lopez-Valenzuela J.A., Martinez O. and Paredes-Lopez O. 1997. Geographic differentiation and embryo type identification in *Mangifera indica* L. cultivars using RAPD markers. HortScience 32: 1105–8.

Mathews H. and Litz R.E. 1992. Mango. In: Hammerschlang F.A. and Litz R.E. (eds), Biotechnology of Perennial Fruit Crops Species. CAB Intl, Wallingford, UK.

McCouch S.R., Chen X., Panaud O., Temnykh S., Xu Y., Cho Y.G., Huang N., Ishii T. and Blair M. 1997. Microsatellite marker development, mapping and applications in rice genetics and breeding. Plant. Mol. Biol. 35: 89–99.

Morton J. 1987. Mango. In: Fruits of Warm Climates. Miami, Florida, USA.

Mukherjee S.K. 1950. Mango: its allopolyploid nature. Nature 166: 196–197.

Mukherjee S.K. 1997. Introduction: botany and importance. In: Litz RE (ed.), The Mango, Botany, Production and Uses. CAB International, Wallingford, UK.

Nakasone H.Y. and Paull R.E. 1998. Mango. In: Tropical Fruits. CAB International, Wallingford, UK.

Nei M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA 70: 3321–3323.

Nei M. and Li W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA 76: 5269–5273.

Nunez-Elisea R 1984. 'Manzanillo-Nunez': a new Mexican mango cultivar. Proc. Fl. St. Hortic. Soc. 97: 360–363.

Oppenheimer H. 1978. Subtropical Fruit Trees and Their Cultivation in Israel. Amm-Oved-Tarbth Vehinuch, Tel-Aviv, Israel.

Paetkau D., Calvert W., Stirling I. and Strobeck C. 1995. Microsatellite analysis of population structure in Canadian Polar bears. Mol. Ecol. 4: 347–354.

Ravishankar K.V., Anand L. and Dinesh M.R. 2000. Assessment of genetic relatedness among mango cultivars of India using RAPD markers. J. Hortic. Sci. Biotech. 75: 198–201.

Raymond M. and Rousset F. 1995. GENEPOP (version 1.2): population genetics software for exact test and ecumenicism. J. Hered. 86: 248–249. http://wbiomed.curtin.edu.au/genepop/.

Sauer J.D. 1993. Historical Geography of Crop Plants-a Select Roster. CRC Press, Boca Raton, Florida, USA.

Schnell R.J., Ronning C.M. and Knight R.J. 1995. Identification of cultivars and validation of genetic relationships in *Mangifera indica* L. using RAPD markers. Theor. Appl. Genet. 90: 269–274.

Singh R.N. 1978. Mango. Indian Council of Agricultural Research, New Delhi, India.

Viruel M.A. and Hormaza J.I. 2004. Development, characterization and variability analysis of microsatellites in lychee (*Litchi chinensis* Sonn., Sapindaceae). Theor. Appl. Genet. 108: 896–902.

Wagner H.W. and Sefc K.M. 1999. IDENTITY 1.0. Centre for Applied genetics. University of Agricultural Sciences, Vienna, Austria.

Wright S. 1951. The genetical structure of populations. Ann. Eugen. 15: 323–354.

Wünsch A. and Hormaza J.I. 2002. Cultivar identification and genetic fingerprinting of temperate fruit tree species using DNA markers. Euphytica 125: 59–67.

Yap V. and Nelson R.J. 1996. WinBoot: A Program for Performing Bootstrap Analysis of Binary Data to Determine the Confidence Limits of UPGMA-based Dendrograms. International Rice Research Institute, Manila, Philippines.

Zane L., Bargelloni L. and Patarnello T. 2002. Strategies for microsatellite isolation: a review. Mol. Ecol. 11: 1–16.