

Polymorphism of Microsatellite Markers in Papaya (*Carica papaya* L.)

Eder Jorge de Oliveira · Vanusia Batista Oliveira Amorim ·
Edneide Luciana Santiago Matos · Juliana Leles Costa · Milene da Silva Castellen ·
Juliano Gomes Pádua · Jorge Luiz Loyola Dantas

Published online: 3 February 2010
© Springer-Verlag 2010

Abstract A set of 81 new microsatellite markers for *Carica papaya* L. previously identified by data mining using freely available sequence information from Genbank were tested for polymorphism using 30 germplasm accessions from the Papaya Germplasm Bank (PGM) at Embrapa Mandioca e Fruticultura Tropical (CNPMPF) and 18 landraces. The data were used to estimate pairwise genetic distances between the genotypes. A neighbor-joining based dendrogram was used to define clusters and infer possible genetic structuring of the collection. Most microsatellites were polymorphic (73%), with an observed number of alleles per locus ranging from one to eleven. The levels of observed and expected heterozygosity for 51 polymorphic loci varied from 0.00 to 0.85 and from 0.08 to 0.82, averaging 0.19 and 0.59, respectively. Forty-four percent of microsatellites showed polymorphism information content (PIC) higher than 0.50. The compound microsatellites seem to be more informative than dinucleotide and trinucleotide repeats in average alleles per locus and PIC. Among dinucleotides, AG/TC or GA/CT repeat motifs exhibited more informativeness than TA/AT, GT/CA and TG/AC repeat motifs.

E. J. de Oliveira (✉) · V. B. O. Amorim · J. L. L. Dantas
Embrapa Mandioca e Fruticultura Tropical (CNPMPF),
Cruz das Almas, BA, Brazil
e-mail: eder@cnpmf.embrapa.br

E. L. S. Matos · J. L. Costa
Universidade Federal do Recôncavo da Bahia (UFRB),
Cruz das Almas, BA, Brazil

M. da Silva Castellen
Embrapa Sede,
Brasília, DF, Brazil

J. G. Pádua
Embrapa Recursos Genéticos e Biotecnologia (CENARGEN),
Brasília, DF, Brazil

The neighbor-joining analysis based on shared allele distance could differentiate all the papaya accessions and landraces as well as differences in their genetic structure. This set of markers will be useful for examining parentage, inbreeding and population structure in papaya.

Keywords Germplasm · Microsatellites · Papaya · Polymorphism · Breeding

Introduction

The genus *Carica* is monotypic and only includes papaya (*Carica papaya* L.). Papaya belongs to the small family Caricaceae which consists of six genera and 35 species (Badillo 2000). Most of them are diploid ($2n=18$) and a relatively small genome of 372 Mb was found in *C. papaya* (Arumuganathan and Earle 1991; Parasnis et al. 1999; Kim et al. 2002; Liu et al. 2004). There are three distinct types of *C. papaya* plants: (1) dioecious papayas have male and female flowers on separate trees, (2) gynodioecious papayas have male, female, and hermaphrodite flowers in different plants.

Many landraces and cultivars present hermaphrodite plants, bearing perfect flowers and producing fruits shaped from long-cylindrical to ellipsoidal, which are preferred for commercial production. The economic importance of papaya resides largely in the fruit production, and Brazil is the world's major producer (FAOSTAT 2007).

Papaya exhibits considerable phenotypic variation for many morphological and horticultural traits, including fruit size, fruit shape, flesh color, flavor and sweetness, length of juvenile period, plant stature, stamen carpelody, and carpel abortion (Kim et al. 2002; Ocampo Pérez et al.

2006). Nevertheless, low levels of genetic variation for resistance to major fungal and virus diseases were observed in *Carica* genus (Nishijima 1994). On the other hand, resistance to several diseases which affect papaya are identified in the *Vasconcellea* genepool, including resistance to PRSV-P (*V. cundinamarcensis*, *V. candicans*, *V. stipulata*, *V. cauliflora* e *V. quercifolia*; Horovitz and Jimenez 1967; Alvizo and Rojkind 1987), *Asperisporium Caricae* (*V. cundinamarcensis*; Drew et al. 1998); *Fusarium* and *Meloidogyne* (*V. weberbaueri*; Scheldeman et al. 2003), phytoplasma (*V. parviflora*; Drew et al. 1998), and *Phytophthora* (*V. goudotiana*; Drew et al. 1998).

In recent years, molecular markers have been widely used in practical plant breeding to access the genetic variability available in germplasm banks, manage and develop core collections, target crosses, classify germplasm into interest groups and identify duplicate accessions (Rafalski and Tingey 1993; Manifesto et al. 2001; Martins-Lopes et al. 2007). Additionally, available molecular markers linked to important genes have proved to be useful for early selection of different desirable traits (Deputy et al. 2002; Dillon et al. 2005). These molecular markers have also been used to develop genetic maps, and for analysis of qualitative and quantitative inheritance (Luo et al. 2001; Jansen 2005; Milczarski et al. 2007). Other applications, such as marker-assisted backcrossing can reduce the number of generations needed to obtain a genotype with 98% or 99% genetic similarity to the recurrent parent for a fixed sample size (Liang et al. 2004; Oliveira et al. 2005). The first study of genetic diversity in papaya was made by Tan and Weinheimer (1976) using isozyme. Genetic relationships between papaya and related wild species have been investigated using isozyme (Morshidi 1998), Random Amplification of Polymorphic DNA (RAPD) (Sondur et al. 1996; Stiles et al. 1993; Jobin-Décor et al. 1997), Restriction Fragment Length Polymorphism (RFLP) (Aradhya et al. 1999), Amplified Fragment Length Polymorphism (AFLP) markers (Kim et al. 2002; Van Droogenbroeck et al. 2002), PCR-RFLP (Van Droogenbroeck et al. 2004), SSR (Kyndt et al. 2005; Kyndt et al. 2006) as well as Inter-Simple Sequence Repeats (ISSR) (Carrasco et al. 2009).

A large number of different molecular techniques are at present available and each of them differs in its informational content. Although in principle all types of markers would be suitable for our purpose, microsatellites (Single Simple Repeat (SSR)) are especially useful for diversity studies (Baumung et al. 2004).

Multi-locus approaches, such as RAPD, AFLP, and ISSR, may be convenient but have some technical and analytical drawbacks, such as dominance. Multi-locus data are typically analyzed as pairwise comparison of complex patterns that only have meaning relative to others in the same study, thus results are to a limited extent comparable

among studies. By contrast, single-locus markers, such as SSR, are usually characterized by co-dominance and thus are more flexible and supply more robust and comparable data (Brondani et al. 1998; Rallo et al. 2000; Karp 2002).

Despite the present availability of molecular marker techniques, papaya have just recently received molecular attention, especially by development of a large number of SSR markers with different motifs, by data mining analysis from DNA databases (Oliveira et al. 2008a) and by screening of sequence data from bacterial artificial chromosomes (BAC) ends and complementary DNA (Eustice et al. 2008). The utility of SSR as genetic markers to investigate relationships among plants has been clearly established (Zhou et al. 2003). Microsatellites have also been extensively exploited for fingerprinting, phylogenetic studies, genetic and QTL mapping for a wide range of species. The genetic analysis based on SSR made it possible to investigate the occurrence and variability of simple sequence repeats at the whole genome level in germplasm accessions of papaya. The objective of this investigation was to test the suitability of SSR for genomic analysis in *C. papaya*.

Material and Methods

Plant Material

Thirty papaya accessions and eighteen landraces collected from Muritiba, Bahia-Brazil, were used to screen for SSR polymorphisms (Table 1). These papaya plants were maintained in the Papaya Germplasm Bank (PGM) at Embrapa Mandioca e Fruticultura Tropical (CNPMTF), in Cruz das Almas, BA, Brazil. Among the 30 papaya accessions, two were cultivars, 12 were improved (but not released) breeding lines, and 16 were unimproved germplasm. All landraces were unimproved germplasm (Table 1).

DNA Extraction

Young papaya leaves were harvested and stored at -80°C for long-term storage. Genomic DNA was extracted based on the procedure described by Doyle and Doyle (1990). DNA quantification was carried out in an agarose gel (1.0% w/v) by comparing the fluorescent intensity of the sample stained with ethidium bromide (1.0 mg/mL), relative to a dilution series of Lambda DNA (Invitrogen, Carlsbad, CA) as standard of known concentration.

PCR Amplification

A set of 100 SSR primers developed by Oliveira et al. (2008a) was tested for amplification and polymorphism.

Table 1 Germplasm accessions (code CMF) of *Carica papaya* and landraces (code M) used to analyze levels of microsatellite polymorphism including origin of variety and mating system

| Accession | Germplasm | Origin | Mating system |
|-----------|------------|--------------|---------------|
| CMF008 | Improved | Malaysia | Gynodioecious |
| CMF011 | Unimproved | Costa Rica | Gynodioecious |
| CMF017 | Unimproved | Taiwan | Gynodioecious |
| CMF023 | Unimproved | Malaysia | Gynodioecious |
| CMF024 | Cultivar | Costa Rica | Gynodioecious |
| CMF031 | Unimproved | Costa Rica | Gynodioecious |
| CMF038 | Improved | Brazil | Gynodioecious |
| CMF054 | Unimproved | Hawaii | Gynodioecious |
| CMF058 | Unimproved | Brazil | Gynodioecious |
| CMF068 | Improved | Brazil | Gynodioecious |
| CMF082 | Improved | South Africa | Gynodioecious |
| CMF101 | Unimproved | Brazil | Gynodioecious |
| CMF102 | Unimproved | Brazil | Gynodioecious |
| CMF108 | Improved | South Africa | Gynodioecious |
| CMF115 | Improved | South Africa | Gynodioecious |
| CMF123 | Improved | Thailand | Gynodioecious |
| CMF125 | Improved | Hawaii | Gynodioecious |
| CMF128 | Cultivar | Taiwan | Gynodioecious |
| CMF129 | Unimproved | Brazil | Gynodioecious |
| CMF134 | Unimproved | Brazil | Dioecious |
| CMF135 | Unimproved | Brazil | Gynodioecious |
| CMF138 | Improved | Brazil | Gynodioecious |
| CMF142 | Improved | Brazil | Gynodioecious |
| CMF143 | Improved | Brazil | Gynodioecious |
| CMF147 | Improved | Brazil | Gynodioecious |
| CMF157 | Unimproved | Brazil | Dioecious |
| CMF165 | Unimproved | Brazil | Dioecious |
| CMF188 | Unimproved | Brazil | Dioecious |
| CMF189 | Unimproved | Brazil | Dioecious |
| CMF191 | Unimproved | Brazil | Dioecious |
| M01 | Unimproved | Brazil | Dioecious |
| M02 | Unimproved | Brazil | Dioecious |
| M03 | Unimproved | Brazil | Dioecious |
| M06 | Unimproved | Brazil | Dioecious |
| M09 | Unimproved | Brazil | Dioecious |
| M11 | Unimproved | Brazil | Dioecious |
| M12 | Unimproved | Brazil | Dioecious |
| M13 | Unimproved | Brazil | Dioecious |
| M14 | Unimproved | Brazil | Dioecious |
| M15 | Unimproved | Brazil | Dioecious |
| M16 | Unimproved | Brazil | Dioecious |
| M17 | Unimproved | Brazil | Dioecious |
| M25 | Unimproved | Brazil | Dioecious |
| M27 | Unimproved | Brazil | Dioecious |
| M33 | Unimproved | Brazil | Dioecious |
| M34 | Unimproved | Brazil | Dioecious |
| M38 | Unimproved | Brazil | Dioecious |
| M39 | Unimproved | Brazil | Dioecious |

Each PCR reaction was prepared as follows: 20 ng DNA template, 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 0.3 mM of each primer, 1.5 mM MgCl₂, 0.2 mM dNTPs, and 0.5 U *Taq* DNA Polymerase (Invitrogen, Carlsbad, CA) in a total volume of 20 µL.

PCR cycling consisted of 94°C for 4 min, followed by 35 cycles of 94°C for 40 s, (annealing temperature of 55°C, 60°C, or 62°C according to each SSR primer (Table 2)) for 40 s, and 72°C for 1 min, with a final extension at 72°C for 2 min, on a PTC-100 thermal cycler (MJ Research, Inc., Watertown, MA). After cycling, fragments with size difference shorter than 10 base pairs (bp) were electrophoresed on a 6% (*w/v*) denaturing polyacrylamide gel in a Hoefer SQ3 DNA sequencer gel electrophoresis unit (Pharmacia Biotech Inc., San Francisco, CA) at 70 W for 2.5 h. The gels were stained with silver nitrate, according to Creste et al. (2001). Previous analysis in polyacrylamide gel showed some primers that produces fragments longer than 10 bp in size difference. These loci were electrophoresed on a 3% agarose 1,000 gel (Invitrogen, Carlsbad, CA) at 130 V for 3.5 h. The 50-bp ladder (New England Biolabs, Inc., Beverly, MA) was used as a molecular weight standard to estimate the size of microsatellite alleles.

Data Analysis

Genetic variability was measured as allelic richness determined by the total number of the detected alleles and the number of alleles per locus (N_A), observed heterozygosity (H_O), expected heterozygosity (H_E) and polymorphism information content (PIC). We defined rare alleles as those whose individual frequency is lower than 1% in the investigated materials. Common alleles are those that occur with a frequency between 1% to 20%, while those whose frequency is higher than 20% are classified as most frequent alleles. Genetic distances between individuals were estimated by shared allele distance (SAD). All these analysis were carried out using the software POWER-MARKER version 3.25 (Liu and Muse 2005). The matrix of genetic distance was used to construct the neighbor-joining tree, using the MEGA 4.1 package (Tamura et al. 2007). To assess confidence in the nodes of a tree, bootstrap values were obtained from 1,000 replicates by re-sampling microsatellite loci.

Results

Polymorphism of Microsatellites

All 100 SSR loci were screened for patterns of amplification using two individuals of the PGM-CNPMF. Eighty-one primer pairs were selected according to strength, clarity of

Table 2 Characteristics of 81 microsatellite loci developed for *Carica papaya*

| Marker | Repeat motif | Genbank accession (gi) | Pimer Sequence (5'-3') | | T _a (°C) |
|--------|--|------------------------|------------------------|------------------------|---------------------|
| | | | Forward | Reverse | |
| CP02 | (AGG) ₉ | 33898396 | aggcgaatcggagagag | ctggtaaaacgacgatgacg | 59 |
| CP03 | (AATA) ₇ | 33898427 | gaaggcccgtgtaagtgc | tggtgaaaattgaaaggag | 58 |
| CP04 | (AGA) ₅ +(AG) ₂₁ | 33898550 | aagggagaagagaagcagagt | ctccagttgtccctcaaaag | 57 |
| CP05 | (AT) ₁₇ | 33898578 | gtcctcaatccgaagcat | catcacacctgtggcttct | 57 |
| CP06 | (TGCCA) ₄ | 33898764 | ftgccaccaggcttaat | tgacgttacggttcatctg | 58 |
| CP07 | (GT) ₁₂ gct(GT) ₁₃ | 33898964 | cctagcattgcctfaggtc | gcccactattcacattcacacc | 60 |
| CP09 | (AT) ₁₀ +(CT) ₁₄ (AT) ₁₂ | 33899071 | cccaattcatgtccaaatcc | atgttgacaaaaggaagcaa | 59 |
| CP10 | (TACA) ₄ (TA) ₉ (GA) ₁₀ | 33899084 | aaaaatcacagcagtgatggt | gaaattacaaatgggcaaaaag | 58 |
| CP11 | (AT) ₉ + (TA) ₁₂ + (TA) ₈ | 14211817 | tccgtatgagaaggaattaga | tctctctccaaacattcattt | 58 |
| CP12 | (ATT) ₇ | 33898403 | gggagggtgtgactcttt | ttggattttcccctacctaa | 57 |
| CP13 | (AT) ₇ +(TA) ₆ (TG) ₈ | 33898463 | attgggaaccaaccattcg | tcaccaaccgcaggatataa | 59 |
| CP14 | (AC) ₉ (AT) ₈ | 33898502 | tcaatgttctcgtcagatgct | tggtgatgtgcaaatgggt | 56 |
| CP15 | (TG) ₅ (TA) ₇ tg(TA) ₆ +(AT) ₉ | 33898543 | atgcactcagcgaaggat | tcctgtctgttcaaaagtct | 57 |
| CP16 | (AT) ₁₃ | 33898598 | tcaactattccccgcata | cacctctgtccaaagggt | 60 |
| CP18 | (TA) ₁₂ | 33898620 | ccgtcatgtttcgtcttt | caattctcgttgattctgg | 57 |
| CP19 | (AT) ₇ t(AT) ₈ | 33898700 | taggggtgtgctgctccata | agcaggctaaaaactggta | 58 |
| CP20 | (AT) ₅ (TG) ₈ | 33898744 | tgtgagattgtctgttggtg | gggctcgaatacaaaacat | 58 |
| CP21 | (GT) ₁₂ | 33898842 | atcgaccgaggaaggtagc | tcaaaaaccattgagctcgc | 60 |
| CP22 | (TA) ₁₆ | 33898998 | gttcgctgctctactgtg | tgacacctgataaaggaaga | 59 |
| CP23 | (GT) ₁₁ | 33899013 | aacaataggaagcaagctca | tcattccaaccacaaa | 57 |
| CP24 | (GAA) ₇ | 33899038 | cgagcactctaccattgc | ctgcaaaatccccgaaact | 60 |
| CP25 | (TTC) ₈ | 33899039 | cctcgcattgatctcaaac | aaggacccccacttttgattc | 58 |
| CP26 | (CA) ₁₀ (TA) ₆ | 33899041 | gcccccgcataaaagta | tggtatactcccacgatgctc | 60 |
| CP27 | (AT) ₁₇ | 33899070 | aaaagacggcatctgtgact | tttctatcccaccgacaaa | 58 |
| CP28 | (CA) ₉ (TA) ₁₁ | 33899080 | tgtcagttcacactgcaaat | ggacattactcgtttgacacc | 56 |
| CP29 | (AT) ₂₀ +(AT) ₁₄ | 33899109 | gggtcgcattgtaacgaat | ttcaaaaacctcccacctta | 58 |
| CP30 | (AG) ₁₇ | 37577349 | ggaagaagaggcctgtatgg | gatctggatgctgcatggt | 58 |
| CP31 | (AT) ₆ (GT) ₁₀ | 37577372 | aagggtactcgtgagca | tctgtcgcctttataactctg | 57 |
| CP32 | (AT) ₈ (AG) ₈ | 37577444 | tgctcacaaattatgctctc | gcaaatgcttttggtgga | 58 |
| CP33 | (ATTAA) ₅ +(TAA) ₉ | 37577568 | tgacccccagtttctgatt | tggtgtagcgtccattgtg | 59 |
| CP35 | (TA) ₁₂ | 37577598 | ggacgaagctccacaatca | ggcaatcaaaccaatgagg | 59 |
| CP36 | (AT) ₁₂ +(TA) ₁₀ | 37577610 | gcaagaagaacttggttaa | tcagacaatgactgttactgc | 55 |
| CP37 | (AAT) ₈ | 37577685 | ttccccgcctgcaattatag | tcagagtttgctcatggtt | 59 |
| CP38 | (AAT) ₁₁ | 37577690 | ttcttacagttgctgcat | aacaagttccccgttttca | 57 |
| CP39 | (AG) ₁₂ | 37577761 | tttggttattccagctaccg | ttgcagtgagggctaaggt | 59 |
| CP40 | (AC) ₁₇ | 51243481 | atacgccatctccctgagc | gacaggcagggggttttac | 60 |
| CP42 | (AT) ₁₆ | 90856412 | aacagctctctctatccttg | ttttaggggatgattgaag | 56 |
| CP44 | (AT) ₁₂ | 90856717 | tgacaacgaactacatcccta | cctcatggtttgtactcct | 56 |
| CP46 | (AT) ₅ (AG) ₈ | 90857025 | catgaccggaagctagatg | ccttcaccagaagtggtga | 56 |
| CP47 | (TA) ₁₄ | 90857042 | gagcagattgtcacatgcaga | ccagaatgcaattttgct | 60 |
| CP48 | (TC) ₁₃ | 90857060 | ccatttctgtcacgcatcc | gatgatggccaaattcag | 59 |
| CP49 | (AT) ₁₂ | 90857082 | cctgaaagcaaccatttcta | tcgctggagctgtaagaga | 56 |
| CP51 | (AT) ₁₆ | 90857104 | tgaaaaagacccacacgtaa | gcaatcgaatcttctttacc | 59 |
| CP52 | (AT) ₁₀ (AG) ₁₂ | 90857241 | ggaagatcatagaaacagtg | tgctactgtgtgtctctca | 55 |
| CP53 | (AAC) ₈ | 90857365 | cagacacatgaagatttgg | gtgggtccttctcctttga | 56 |
| CP54 | (AT) ₁₀ +(ATA) ₇ | 90857406 | taagactatattggccaagc | ttaggtcaaggatgaaacca | 56 |
| CP55 | (AT) ₁₂ | 90857407 | tcaaccactctgtctca | caacccttgggctatttg | 60 |

Table 2 (continued)

| Marker | Repeat motif | Genbank accession (gi) | Pimer Sequence (5'-3') | | Ta (°C) |
|--------|---|------------------------|------------------------|-----------------------|---------|
| | | | Forward | Reverse | |
| CP56 | (ATG) ₆ | 90857410 | gaaccgctagcgaagac | cctgcgtaacacttgatga | 59 |
| CP57 | (CA) ₆ (TA) ₇ | 90857692 | ttgagctctggttcaactcc | ttcccactatcttctgttgg | 56 |
| CP58 | (TTG) ₇ | 90857748 | ttccacgagacaaactgtacg | ctcccaccgctgtacttga | 58 |
| CP59 | (TA) ₇ +(AT) ₁₂ | 90857930 | gtgtttgcatcccactgc | ctcgcaccatctgtgt | 60 |
| CP61 | (AT) ₁₂ + (AC) ₈ | 90858177 | ggagtgtatgaagtcctcat | catgagccctactccgaac | 57 |
| CP62 | (AT) ₈ +(AAG) ₁₀ | 90858365 | acctggcccaccactatt | ttgattctgcttggggaga | 59 |
| CP63 | (AT) ₁₄ | 90858411 | aaccagcaggcttccataa | ggcttcaagagagcgtat | 56 |
| CP64 | (TC) ₁₇ | 90858444 | gggaggacaaaagctccaaa | gcatgatccaaggaggag | 60 |
| CP66 | (TG) ₉ cgc(GA) ₁₂ | 90858501 | agtcccatcaggctctcgc | cctttttgtgcgcatatggt | 60 |
| CP67 | (TTC) ₁₀ | 90858512 | tggcaaaagatgtgtgtgt | cttctttgcactttcagg | 59 |
| CP68 | (AG) ₁₅ | 90858681 | agtcctgcatgtgctgct | ccccattctgaaaatctcc | 60 |
| CP69 | (TA) ₁₂ | 90858696 | gctgcatcgacatttacgaa | catgtgattccaacaaggcta | 59 |
| CP70 | (TTC) ₉ | 90858729 | cggagctgacgtgactga | gcctgaggagccattaagc | 60 |
| CP71 | (CT) ₁₄ | 90858774 | tcccaacctcaagaagataa | ttaccaccatcgccatc | 59 |
| CP72 | (ATAC) ₇ (AT) ₆ | 90858906 | cccaaatcaccttttctctc | aacgtgaactgagggtgga | 58 |
| CP73 | (AT) ₅ gt(AT) ₉ | 90858959 | gaaaccgctcatttgcat | ttgattactcccctccatctc | 58 |
| CP74 | (TTA) ₈ | 90859162 | caatgaggggtgataattgg | gctcccccaatgagaatttta | 59 |
| CP75 | (GA) ₁₃ +(AG) ₇ | 90859250 | gcgtattgagagggtgcaa | tttggtagcgccttcgt | 59 |
| CP78 | (TTA) ₇ +(TTTA) ₅ | 90859390 | ggagttggtgtgtatgttacc | attttcaccccttttcc | 57 |
| CP79 | (TA) ₁₇ | 90859529 | gcctcagcatccaacaaa | caaccatgtgaatggagaacc | 60 |
| CP80 | (TC) ₁₂ | 90859560 | gtggcaatgaactcggcta | atcctccctccaccaacag | 59 |
| CP81 | (CT) ₁₅ | 90859561 | gaaaaaccggcaaatcac | ttaccagatttcaccacttg | 58 |
| CP82 | (AAT) ₇ | 90859590 | ccctcccagggaagataca | gaagaggtttggcggatg | 59 |
| CP83 | (TA) ₅ (CT) ₉ | 90859697 | ttgcaccttgaatccact | ggcagcctttaggttctc | 59 |
| CP86 | (AT) ₁₂ | 90859959 | ttcacacggattctacacga | tgggggatagagaagtctcc | 59 |
| CP89 | (AT) ₁₂ + (AT) ₁₁ | 90860069 | agatgattcaccgctccag | ctcaagcaagaaatgcctct | 58 |
| CP90 | (T) ₂₁ (CTT) ₅ | 90860101 | cttgggcttgcacctga | agcaactaatggaggcaaaag | 59 |
| CP92 | (ATT) ₁₁ | 90860241 | ctttttgtctgctgccttc | gaacataataccgtgcgttca | 59 |
| CP94 | (AG) ₁₂ at(AG) ₇ | 90860608 | tcgagagaaagagaagca | tcctctcaccagtattggaa | 58 |
| CP95 | (AG) ₆ tg(AG) ₇ | 90860628 | cagtccttcaggcgatta | ctcaaccttttctcacca | 60 |
| CP97 | (TA) ₁₁ | 90860894 | tttctgtacctctctcgatt | gagatgacaaccatacagcaa | 56 |
| CP98 | (TC) ₁₃ | 90860935 | gattggcgggaaaaactc | agagagagtgtaggacctgaa | 56 |
| CP99 | (AG) ₂₄ | 90861034 | cgaggacaaaacggcacta | tgcacgactctcaaatctc | 60 |
| CP100 | (GA) ₁₁ + (AG) ₂₃ | 90861159 | tgatcgtcttcgcttcaact | gatttactgccacggact | 58 |

Repeat motif is derived from a sequenced clone deposited on Genbank
Ta annealing temperature

banding patterns, and successfully amplified PCR product of high quality, besides the polymorphic amplification product in the expected size. Characteristics of the 81 primer pairs and optimal conditions for their amplification are given in Table 2. Subsequently, all primer pairs that amplified a specific band were used for the genotyping of 30 germplasm accessions and 18 landraces. In total, 59 primer pairs amplified a polymorphic and easily scorable PCR product, while 22 pairs amplified a monomorphic one.

Considering the 59 SSR loci analyzed in the present study and a total of 48 genotyped individuals obtained from partial outcrossing and selfing germplasm, the SSR markers detected a total of 237 alleles. The least and the most variable loci displayed 2 (CP06, CP09, CP11, CP20, CP22, CP23, CP24, CP27, CP28, CP29, CP33, CP36 and CP38) and 11 (CP16) alleles, respectively (Table 3). The average allele number per locus was 4.02 (monomorphic loci excluded). Table 3 summarizes the locus specific descriptive statistics for the 59 SSR markers.

Table 3 Allelic composition, allele size variation (bp), expected and observed heterozygosity, and polymorphic information content (PIC) of the 59 polymorphic SSR loci in 48 individuals of *Carica papaya*

| Marker | Allele | | | | Unique alleles | | Size variation (bp) | H_E | H_O | PIC |
|--------|--------|------|--------|---------------|----------------|------|---------------------|-------|-------|------|
| | Number | Rare | Common | Most frequent | Germ | Land | | | | |
| CP02 | 3 | 0 | 1 | 2 | | | 144–158 | 0.62 | 0.33 | 0.54 |
| CP03 | 4 | 0 | 2 | 2 | 1 | | 238–276 | 0.66 | 0.32 | 0.61 |
| CP05 | 4 | 0 | 1 | 3 | | | 290–320 | 0.69 | 0.04 | 0.63 |
| CP06 | 2 | 0 | 0 | 2 | | | 240–320 | 0.49 | 0.04 | 0.37 |
| CP07 | 4 | 0 | 1 | 3 | 1 | | 175–208 | 0.71 | 0.50 | 0.66 |
| CP09 | 2 | 0 | 1 | 1 | 1 | | 395–411 | 0.30 | 0.00 | 0.26 |
| CP10 | 6 | 0 | 4 | 2 | 3 | | 180–202 | 0.73 | 0.28 | 0.69 |
| CP11 | 2 | 0 | 0 | 2 | | | 568–610 | 0.41 | 0.25 | 0.33 |
| CP14 | 5 | 0 | 3 | 2 | 3 | | 217–232 | 0.52 | 0.00 | 0.46 |
| CP16 | 11 | 1 | 9 | 1 | 1 | 2 | 202–250 | 0.82 | 0.67 | 0.81 |
| CP18 | 4 | 0 | 2 | 2 | 1 | | 238–260 | 0.54 | 0.23 | 0.47 |
| CP19 | 3 | 0 | 1 | 2 | | | 156–166 | 0.51 | 0.21 | 0.43 |
| CP20 | 2 | 0 | 0 | 2 | | | 248–252 | 0.35 | 0.25 | 0.29 |
| CP21 | 4 | 0 | 2 | 2 | | | 142–158 | 0.64 | 0.46 | 0.57 |
| CP22 | 2 | 0 | 1 | 1 | 1 | | 246–254 | 0.08 | 0.00 | 0.08 |
| CP23 | 2 | 0 | 0 | 2 | | | 218–236 | 0.44 | 0.21 | 0.35 |
| CP24 | 2 | 0 | 0 | 2 | | | 236–246 | 0.35 | 0.25 | 0.29 |
| CP26 | 3 | 0 | 1 | 2 | 1 | | 142–152 | 0.48 | 0.04 | 0.41 |
| CP27 | 2 | 0 | 0 | 2 | | | 347–353 | 0.38 | 0.00 | 0.31 |
| CP28 | 2 | 0 | 0 | 2 | | | 186–190 | 0.48 | 0.08 | 0.36 |
| CP29 | 2 | 0 | 0 | 2 | | | 550–556 | 0.41 | 0.13 | 0.33 |
| CP30 | 5 | 1 | 2 | 2 | 2 | | 222–290 | 0.58 | 0.38 | 0.53 |
| CP31 | 4 | 0 | 2 | 2 | | | 160–174 | 0.73 | 0.85 | 0.68 |
| CP33 | 2 | 0 | 1 | 1 | | | 401–411 | 0.29 | 0.23 | 0.25 |
| CP35 | 5 | 0 | 3 | 2 | | | 162–180 | 0.73 | 0.25 | 0.69 |
| CP36 | 2 | 0 | 0 | 2 | | | 452–458 | 0.50 | 0.27 | 0.37 |
| CP38 | 2 | 0 | 0 | 2 | | | 498–510 | 0.39 | 0.35 | 0.31 |
| CP40 | 4 | 0 | 3 | 1 | | 1 | 154–170 | 0.55 | 0.23 | 0.51 |
| CP42 | 5 | 0 | 3 | 2 | 1 | | 260–280 | 0.66 | 0.22 | 0.60 |
| CP44 | 4 | 0 | 1 | 3 | | | 228–236 | 0.75 | 0.56 | 0.70 |
| CP47 | 4 | 0 | 2 | 2 | | | 248–262 | 0.68 | 0.00 | 0.63 |
| CP48 | 3 | 0 | 1 | 2 | | | 210–250 | 0.51 | 0.62 | 0.43 |
| CP49 | 6 | 0 | 4 | 2 | | 2 | 196–216 | 0.73 | 0.63 | 0.68 |
| CP51 | 6 | 0 | 4 | 2 | | 1 | 290–310 | 0.77 | 0.38 | 0.73 |
| CP52 | 3 | 0 | 0 | 3 | | | 186–205 | 0.66 | 0.00 | 0.59 |
| CP53 | 6 | 1 | 2 | 3 | | 1 | 305–320 | 0.70 | 0.08 | 0.65 |
| CP54 | 7 | 0 | 5 | 2 | 3 | | 270–290 | 0.76 | 0.43 | 0.73 |
| CP55 | 5 | 0 | 2 | 3 | 2 | | 250–270 | 0.73 | 0.00 | 0.68 |
| CP57 | 4 | 0 | 1 | 3 | 1 | | 280–290 | 0.70 | 0.00 | 0.63 |
| CP58 | 3 | 0 | 0 | 3 | | | 357–367 | 0.66 | 0.11 | 0.58 |
| CP59 | 5 | 0 | 3 | 2 | 3 | | 367–378 | 0.59 | 0.00 | 0.51 |
| CP61 | 8 | 0 | 7 | 1 | | 2 | 207–220 | 0.81 | 0.21 | 0.78 |
| CP62 | 5 | 0 | 3 | 2 | | 1 | 240–260 | 0.65 | 0.35 | 0.58 |
| CP63 | 3 | 0 | 1 | 3 | 1 | | 235–249 | 0.51 | 0.00 | 0.40 |
| CP66 | 5 | 0 | 3 | 2 | 1 | 1 | 244–270 | 0.59 | 0.18 | 0.50 |
| CP68 | 4 | 0 | 2 | 2 | | 1 | 141–155 | 0.59 | 0.00 | 0.51 |

Table 3 (continued)

| Marker | Allele | | | | Unique alleles | | Size variation (bp) | H_E | H_O | PIC |
|--------|--------|------|--------|---------------|----------------|------|---------------------|-------|-------|------|
| | Number | Rare | Common | Most frequent | Germ | Land | | | | |
| CP69 | 4 | 0 | 2 | 2 | 2 | | 350–377 | 0.52 | 0.00 | 0.47 |
| CP71 | 6 | 0 | 4 | 2 | | 1 | 240–260 | 0.79 | 0.39 | 0.76 |
| CP72 | 4 | 0 | 2 | 2 | 3 | | 190–205 | 0.65 | 0.00 | 0.60 |
| CP73 | 5 | 0 | 3 | 2 | 3 | | 225–242 | 0.73 | 0.00 | 0.68 |
| CP80 | 3 | 0 | 0 | 3 | | | 207–220 | 0.64 | 0.00 | 0.56 |
| CP83 | 4 | 0 | 1 | 3 | | | 240–245 | 0.73 | 0.00 | 0.68 |
| CP86 | 3 | 0 | 2 | 1 | 2 | | 447–451 | 0.27 | 0.08 | 0.24 |
| CP89 | 3 | 0 | 2 | 1 | 1 | | 423–450 | 0.34 | 0.00 | 0.30 |
| CP94 | 6 | 0 | 4 | 2 | 2 | | 170–180 | 0.79 | 0.00 | 0.76 |
| CP95 | 6 | 0 | 5 | 1 | 5 | | 180–192 | 0.73 | 0.00 | 0.70 |
| CP97 | 4 | 0 | 2 | 2 | | 1 | 248–265 | 0.70 | 0.00 | 0.65 |
| CP98 | 4 | 0 | 2 | 2 | | | 194–200 | 0.53 | 0.14 | 0.48 |
| CP100 | 4 | 0 | 2 | 2 | 1 | | 282–305 | 0.67 | 0.00 | 0.61 |

Rare allele, common allele and most frequent allele, present frequency <1%, ≥1% ≥20%, and >20%, respectively; unique alleles present in germplasm accessions (Germ) and landraces (Land)

T_a annealing temperature, H_E expected heterozygosity, H_O observed heterozygosity

Expected heterozygosity was nominally larger than the observed heterozygosity under HWE for all loci, except on CP31 and CP48. Only loci CP21, CP33, and CP38 were found to be in HWE. The observed heterozygosity ranged from 0.00 (CP09, CP14, CP22, CP27, CP47, CP52, CP55, CP57, CP59, CP63, CP68, CP69, CP72, CP73, CP80, CP83, CP89, CP94, CP95, CP97, and CP100) to 0.85 (CP31), and gene diversity (H_E) from 0.08 (CP22) to 0.82 (CP16). The CP16 locus showed the highest polymorphism information content (PIC=0.81) and the CP22 locus the lowest (PIC=0.08), average PIC being 0.53.

When looking at SSR classes and motifs, the compound SSRs showed higher allele numbers (average 4.27 per locus) and PIC values (average 0.55 per marker) followed by dinucleotide (average alleles, 4.08 per locus; PIC value average 0.52 per marker) and trinucleotide SSRs (average alleles, 3.2 per locus; PIC value, average 0.47 per marker). Among dinucleotide SSRs, AG/TC or GA/CT repeat motifs exhibited more informativeness (average alleles, 4.56 per locus and PIC value, average 0.59 per marker) as compared with TA/AT repeat motifs (average alleles, 4.0 per locus and PIC value, average 0.50 per marker) and to GT/CA or TG/AC repeat motifs (average alleles, 3.5 per locus and PIC value, average 0.52 per marker) (Table 3).

Allelic Composition

The allelic composition revealed that rare alleles were represented by < 1.3% of the total number of alleles detected. Of the 237 alleles detected, three were rare, 115

common, and 119 most frequent alleles. Rare alleles were detected from CP16, CP30, and CP53. Common alleles were detected at 46 SSR loci, with an average of 1.95 alleles. In contrast, all SSR loci detected 1 to 3 most frequent alleles in the individuals (Table 3).

This study detected unique alleles within groups (Table 3). The germplasm accessions contained the largest number of unique alleles (46), while 14 unique alleles differentiated landraces from the papaya germplasm. Of the total alleles detected, papaya germplasm and landraces shared 74.7% of alleles (177).

Genetic Diversity

Based on the unique DNA fingerprint profiles of each genotype obtained by the polymorphic markers, a dendrogram was constructed to understand the relationships among the germplasm and landraces surveyed. The neighbor-joining cluster analysis based on shared-allele distance could successfully differentiate all the papaya accessions (group A) and landraces (group B) included in this study (Fig. 1) with the high bootstrap value of 95%. Clearly, four subgroups could be seen in germplasm accessions while landraces presented just two.

Subgroup A1 comprised all gynodioecious accessions, being one cultivar (CM128), eight improved germplasm (CMF068, CMF082, CMF108, CMF115, CMF125, CMF138, CMF142 and CMF143) and five unimproved germplasm (CMF054, CMF058, CMF101, CMF102 and CMF129). Subgroup A2 had two improved (CMF008 and

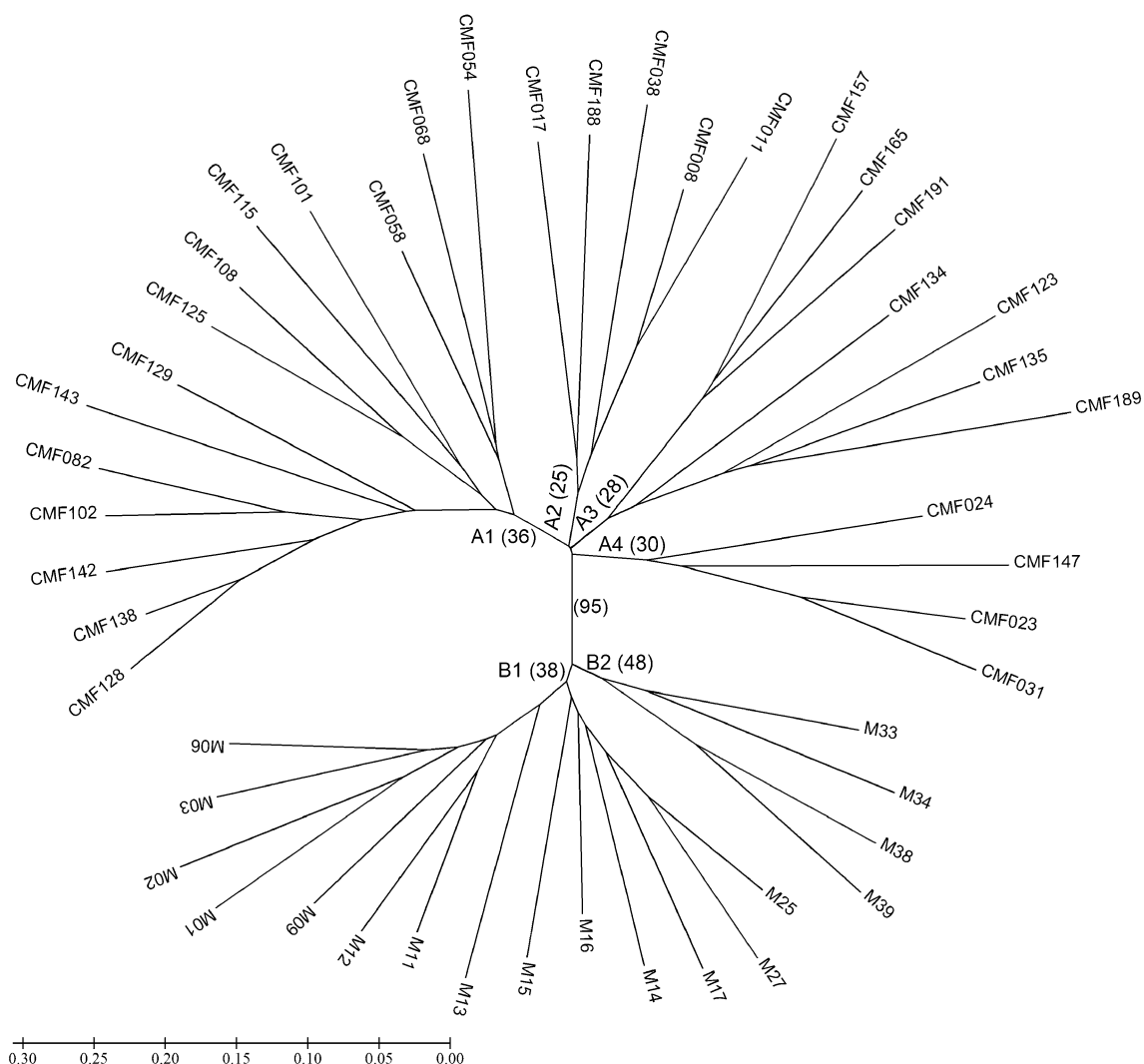


Fig. 1 Neighbor-joining tree based on the shared allele distance of 59 SSR markers using 30 germplasm accessions at PGM-CNPMF (code CMF) and 18 landraces (code M). A1, A2, A3, and A4 are subgroups

of the A group clustered by germplasm accessions; B1 and B2 are subgroups of B group that contained only papaya landraces. Distance bootstrap values are given together with subgroups in parenthesis

CMF038) and three unimproved germplasm (CMF011, CMF017 and CMF188), whereas subgroup A3 was formed by one improved (CMF123) and six unimproved germplasm (CMF134, CMF135, CMF157, CMF165, CMF189 and CMF191). The other cultivars analyzed (CMF024) were clustered with three germplasm accessions in subgroup A4. Subgroups A1, A2, A3 and A4 were clustered with a relatively low bootstrap support value (36%, 25%, 28%, and 30%, respectively), but subgroups B1 and B2, had a moderate bootstrap support value (48% and 38%, respectively).

The genotypes belonging to landraces collected in the State of Bahia (Brazil) were grouped in clusters (B1 and B2), with 14 and four individuals, respectively. No specific grouping was observed for the improved and unimproved cultivars, which dispersed in all subgroups.

Discussion

Microsatellite Informativeness

SSR markers have been extensively used for DNA fingerprinting and elucidating genetic relationships within plant species (Paniego et al. 2002; Stajner et al. 2005). The ability to distinguish between closely related genotypes is a function of the high heterozygosity values of SSR markers. The set of SSR markers characterized in this study proved to be useful for a broad genetic analysis of *C. papaya*.

Previous studies have shown that papaya contains abundant SSRs (Santos et al. 2003; Ocampo Pérez et al. 2006; Eustice et al. 2008; Oliveira et al. 2008a), but the characterization of SSR markers in *C. papaya* germplasm is not yet well studied. Here, we observed relatively high

levels of multiallelism at all 59 SSR loci analyzed. Mean number of alleles per locus (4.02, for 59 loci) and expected heterozygosity (mean of 0.59) were higher than those reported by Ocampo Pérez et al. (2006) who used 26 polymorphic markers, and observed 3.8 alleles per locus and H_O and H_E values of 0.42 and 0.57, respectively, but intermediate for samples collected in Guadeloupe, Venezuela, Colombia, Barbados and Costa Rica, that showed 6.6 alleles per locus for 15 loci, and H_E values between 0.37 and 0.69 (Ocampo Pérez et al. 2007).

The number of alleles per locus reported in our study is most likely a minimum value. The lower number of alleles can be due to the relatively few samples analyzed and to the different mating systems of these genotypes. Half of the accessions were gynodioecious, which is almost exclusively inbreeding and tends to decrease the allele number per locus. The number of alleles should increase when all germplasm accessions of the PGM of CNPMF were sampled.

Except CP11 and CP40, all SSR described here are derived from a genomic sequence and showed considerable polymorphism. This is in agreement with Eustice et al. (2008), who found a relatively high level of polymorphism in the genomic rather than the genic region, using seven papaya accessions (SunUp, Kapoho, 2H94, UH918, Kaek Dum, UH928, and AU9) to screen SSR polymorphisms. As in other crops, the selective pressures from breeding can significantly reduce genetic diversity in the target genes or genic regions, while levels of genetic diversity in the genomic region remain high.

The broad range of observed heterozygosity (0.00 to 0.85) and expected heterozygosity (0.08 to 0.82) result from the broad variation in number of alleles per locus and allele frequency distribution within genotypes. The obtained zero values of H_O at twenty-one loci could be explained by the low number of alleles at these loci (2 to 6 alleles) as well as their combination in homozygous state. Loci with smaller numbers of alleles or with a skewed frequency distribution such as CP06, CP09, CP11, CP20, CP22, CP23, CP24, CP27, CP28, CP29, CP33, CP36, and CP38, tend to have lower heterozygosity values and, consequently, lower probability of paternity exclusion when studying natural population.

Ninety-five percent of the SSR markers surveyed for heterozygosity (56 of 59) deviated significantly from Hardy–Weinberg equilibrium. In all cases, the deviation was in the direction of reduced heterozygosity due to inbreeding, presence of null alleles, natural or artificial selection favoring homozygosity for particular loci and population bottlenecks. Such result has been commonly observed in surveys of other species (Carrasco et al. 2009). A possible reason for high levels of inbreeding in *C. papaya* is the elimination of the male plants in trioic

accessions (male, female, and hermaphrodite plants) and self-compatibility that stimulates crossing among related individuals and increases the degree of selfing.

In relation to polymorphism, microsatellites of papaya showed enrichment for di and trinucleotide SSR repeat motifs (Eustice et al. 2008; Oliveira et al. 2008a). Overall, AT/TA is the predominant dinucleotide motif, and AAT/TTA is the predominant trinucleotide motif. The previous surveys carried out on microsatellite abundance in plant genomes have shown AT as the most frequently occurring dinucleotide repeat motif followed by AG/TC and GT/CA (Condit and Hubbel 1991; Powell et al. 1996; Yonemaru et al. 2009). Compound SSRs, mainly from different dinucleotide repeats, and dinucleotide repeats were more informative (more average alleles per locus and PIC). Among dinucleotides, although AT/TA-rich motif is prominent, AG/TC or GA/CT repeat motifs exhibited more informativeness. The same pattern was observed in other species (Ferguson et al. 2004; Moretzsohn et al. 2005).

Diversity Structuring

Our study based on SSR analysis clearly revealed genetic diversity in *C. papaya* germplasm samples and some landraces cultivated by farmers in the State of Bahia. Neighbor-joining tree broadly separated germplasm accessions from landraces with high bootstrap support (95%). However, clusters from each subgroup were poorly supported by bootstrap values (Fig. 1). A low bootstrap value means that a grouping is sensitive to the combinations of genotypes that are evaluated, implying that more data may alter the grouping.

The main cluster (group A) is composed of germplasm accessions of *C. papaya* subdivided in four smaller clusters. The smaller one (group B) is clustered only by landraces, which have two subgroups. The grouping of genotypes representing landraces in different clusters is in agreement with their origin, and is important as a first indicative of the genetic background of most of the germplasm of PGM at CNPMF.

The genetic variability in landraces grown by small farmers in the region of Recôncavo of Bahia, where the landraces were obtained, was maintained over generations. This variability is quite different from the one in the PGM, because papaya accessions from the Recôncavo of Bahia has not been deposited in the germplasm bank. The additional genetic variability will be used in breeding programmes to enhance the diversity of breeding populations for selection gains in the future.

Ocampo Pérez et al. (2007) analysed genotypes from Costa Rica, Colombia, Venezuela, Guadeloupe and Antillean islands using SSR markers and PCO analysis, and found that, with few exceptions, they were clustered

according to their geographic origin. In our study, although there are few accessions belonging to different countries, no high correlation between the clustering pattern and the geographical location was observed. The Brazilian accessions were clustered in all subgroups. The accessions from South Africa and Hawaii grouped into subgroup A1, Thailand into subgroup A3; whereas accessions from Costa Rica and Malaysia grouped into A2 and A4 subgroup; and Taiwan into A1 and A2 subgroup (Fig. 1).

The 30 germplasm accessions and 18 landraces used for characterization of SSR polymorphism allowed the detection of considerable genetic variability, as showed on six diversity subgroups (Fig. 1). The average similarity based on shared allele distance between accessions was 0.44, which is very similar to the value of 0.48 obtained by Ocampo Pérez et al. (2007), using Dice distance. Previous reports using dominant markers, as RAPD (Stiles et al. 1993) and AFLP (Kim et al. 2002), showed values of 0.78 (Jaccard distance) and 0.88 (Dice distance), respectively. So, the SSRs markers revealed a more important polymorphism than the other nuclear DNA markers used so far in papaya (Ocampo Pérez et al. 2007).

Most of the observed variation occurred among accessions. All genotypes present in A1 and A4 subgroup are gynodioecious, whereas B1 and B2 are dioecious. In addition, the A2 and A3 subgroups are composed of gynodioecious and dioecious genotypes. The gynodioecious tends to show lower genetic variability within accessions than the dioecious, because self-compatibility increases the inbreeding.

Many aspects, such as the breeding system, seed and pollen dispersal, plant longevity and agricultural practices influence the genetic diversity, including the proportion of variation distributed within and between populations (Hamrick and Godt 1996). Kim et al. (2002) analysed the genetic relationships among *C. papaya* cultivars, breeding lines, unimproved germplasm, and related species using AFLP markers and suggested limited genetic variation in papaya, with smaller genetic diversity within the same gene pools, such as gynodioecious and dioecious cultivars. According to Carrasco et al. (2009), the genetic diversity of *Vasconcellea pubescens*, a species of the Caricaceae family, was remarkably low using ISSR (Inter Single Simple Repeat). The major genetic diversity was found within groups (65%) when South and the North samples from Chile were analyzed together. In addition, samples from South and the North analyzed separately showed that 82% and 60% of genetic diversity were within groups, respectively.

Efforts are made to maintain the genetic diversity of papaya breeding through the stocks in germplasm banks to study natural variation. Although artificial by their nature and submitted to several forces, including genetic drift and

man driven artificial selection, germplasm collections play an important role in the way the banks are structured. In this study, we found a great depth of allelic diversity among landraces and germplasm, because this diversity is not distributed randomly among the genotypes, but rather structured into two groups. Thus, many landraces will be incorporated to PGM at CNPMF, to increase the stored variability for future use in conservation and breeding programs.

SSR in Genetic Studies of *C. papaya*

SSRs are powerful molecular biology tools that have been used for a wide variety of applications. Although good estimators of population genetic parameters have been presented for dominant markers (Lynch and Milligan 1994; Zhivotovsky 1999), the codominant nature of SSR allows more precise calculations of heterozygosity, Hardy-Weinberg equilibrium, differentiation and gene flow (Lowe et al. 2004). Polymorphisms based on SSR are more powerful to estimate genetic parameters of populations and understand detailed patterns of gene flow and parentage composition (Dow and Ashley 1996; Collevatti et al. 1999; Ren et al. 2008). Besides, SSRs markers are more useful for genetic mapping (Oliveira et al. 2008b; Wang et al. 2009), and ancestry studies (Guilford et al. 1997; Gianfranceschi et al. 1998).

The papaya microsatellite loci published here provide an abundant set of genetic markers for detailed studies of population genetic structure, hybridization among populations, paternity, migration, phylogeography of the genus *Carica*, creation of linkage and physical maps, and location of genes of interest, particularly those functioning as QTL, encoding important agricultural traits such as disease resistance, yield, fruit type, and fruit size. Specifically, we are now using these markers to understand the genetic variability of 160 papaya germplasm accessions collected from different countries, and that are part of the PGM at CNPMF.

Acknowledgements This work was supported by Embrapa Macroprograma 2, CNPq-Conselho Nacional de Desenvolvimento Científico e Tecnológico (National Counsel of Technological and Scientific Development) and FAPESB-Fundação de Amparo à Pesquisa do Estado da Bahia (Bahia State Research Funding Institution).

References

- Alvizo VHF, Rojkind MC (1987) Resistencia al virus mancha anular del papaya en *Carica cauliflora*. Rev Mex Fitopat 5:61–62
- Aradhya MK, Manshardt RM, Zee F, Morden CW (1999) A phylogenetic analysis of the genus *Carica* L (Caricaceae) based on restriction fragment length variation in a cpDNA intergenic spacer region. Genet Resour Crop Evol 46:579–586

- Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9:208–218
- Badillo VM (2000) *Carica* L vs *Vasconcellea* St Hil (Caricaceae): con la rehabilitación de este úlltimo. *Ernstia* 10:74–79
- Baumung R, Simianer H, Hoffmann I (2004) Genetic diversity studies in farm animals - a survey. *J Anim Breed Genet* 121:361–373
- Brondani RPV, Brondani C, Tarchini R, Grattapaglia D (1998) Development characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theor Appl Genet* 97:816–827
- Carrasco B, Avila P, Perez-Diaz J, Muñoz P, Garcia R, Lavandero B, Zurita-Silva A, Retamales JB, Caligari PDS (2009) Genetic structure of highland papayas (*Vasconcellea pubescens* (Lenné et C Koch) Badillo) cultivated along a geographic gradient in Chile as revealed by Inter Simple Sequence Repeats (ISSR). *Genet Resour Crop Evol* 56:331–337
- Collevatti RG, Brondani RV, Grattapaglia D (1999) Development and characterisation of microsatellite markers for genetic analysis of a Brazilian endangered tree species *Caryocar brasiliense*. *Heredity* 83:748–756
- Condit R, Hubbel SP (1991) Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. *Genome* 34:66–71
- Creste S, Tulmann Neto A, Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Mol Biol Rep* 19:299–306
- Deputy JC, Ming R, Ma H, Liu Z, Fitch MMM, Wang M, Manshardt R, Stiles JI (2002) Molecular markers for sex determination in papaya (*Carica papaya* L.). *Theor Appl Genet* 106:107–111
- Dillon S, Ramage C, Drew R, Ashmore S (2005) Genetic mapping of a PRSV-P resistance gene in “highland papaya” based on inheritance of RAF markers. *Euphytica* 145:11–23
- Dow BD, Ashley MV (1996) Microsatellite analysis of seed dispersal and sapling parentage in bur oak *Quercus macrocarpa*. *Mol Ecol* 5:615–627
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15
- Drew RA, O’Brien CM, Magdalita PM (1998) Development of interspecific *Carica* hybrids. *Acta Hort* 461:285–292
- Eustice M, Yu Q, Lai CW, Hou S, Thimmapuram J, Liu L, Alam M, Moore PH, Presting GG, Ming R (2008) Development and application of microsatellite markers for genomic analysis of papaya. *Tree Genet Genomes* 4:333–341
- Faostat: The United Nations: Food and Agriculture Organization of The United Nations 2007 Available via: <http://faostatfaorg/site/567/default.aspx#ancor> Accessed 04 June 2009
- Ferguson ME, Burow MD, Schultz SR, Bramel PJ, Paterson AH, Kresovich S, Mitchell S (2004) Microsatellite identification and characterization in peanut (*Arachis hypogaea* L.). *Theor Appl Genet* 108:1064–1070
- Gianfranceschi L, Seglias N, Tarchini R, Komjanc M, Gessler C (1998) Simple sequence repeats for the genetic analysis of apple. *Theor Appl Genet* 96:1069–1076
- Guilford P, Prakash S, Zhu JM, Rikkerink E, Gardiner S, Bassett H, Forster R (1997) Microsatellites in *Malus × domestica* (apple): abundance polymorphism and cultivar identification. *Theor Appl Genet* 94:249–254
- Hamrick JL, Godt MJ (1996) Conservation genetics of endemic plant species. In: Avise JC, Hamrick JL (eds) Conservation genetics case histories from nature. Chapman and Hall, New York USA, pp 281–304
- Horovitz S, Jiménez H (1967) Cruzamientos interespecificos e intergenericos en Caricaceas y sus implicaciones fitotecnicas. *Agron Trop* 17:323–344
- Jansen J (2005) Construction of linkage maps in full-sib families of diploid outbreeding species by minimizing the number of recombinations in hidden inheritance vectors. *Genetics* 170:2013–2025
- Jobin-Decor MP, Graham GC, Henry RJ, Drew RA (1997) RAPD and isozyme analysis of genetic relationships between *Carica papaya* and wild relatives. *Genet Resour Crop Evol* 44:471–477
- Karp A (2002) The new genetic Era: will it help us in managing genetic diversity? In: Engels JMM, Ramanatha Rao V, Brown AHD, Jackson MT (eds) Managing plant genetic diversity. IPGRI, Rome, pp 43–56
- Kim M, Moore P, Zee F, Fitch MMM, Steiger D, Manshardt R, Paull R, Drew R, Sekioka T, Ming R (2002) Genetic diversity of *Carica papaya* as revealed by AFLP markers. *Genome* 45:503–512
- Kyndt T, Haegeman A, Van Glabeke S, Maertens I, Van Droogenbroeck B, Roldán-Ruiz I, Gheysen G (2005) Isolation and characterization of microsatellite loci in the highland papaya *Vasconcellea × heilbornii* V. Badillo (Caricaceae). *Mol Ecol Notes* 5:590–592
- Kyndt T, Van Droogenbroeck B, Haegeman A, Roldán-Ruiz I, Gheysen G (2006) Cross-species microsatellite amplification in *Vasconcellea* and related genera and their use in germplasm classification. *Genome* 49:786–798
- Liang F, Deng Q, Wang Y, Xiong Y, Jin D, Li J, Wang B (2004) Molecular marker-assisted selection for yield-enhancing genes in the progeny of “9311 × *O. rufipogon*” using SSR. *Euphytica* 139: 159–165
- Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21:2128–2129
- Liu Z, Moore PH, Ma H, Ackerman CM, Makandar R, Yu Q, Pear HM, Kim MS, Charlton JW, Stiles JI, Zee FT, Paterson AH, Ming R (2004) A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature* 427:348–352
- Lowe A, Harris S, Ashton P (2004) Ecological genetics: design analysis and application. Blackwell Publishing, Oxford
- Luo ZW, Hackett CA, Bradshaw JE, Mcnicol JW, Milbourne D (2001) Construction of a genetic linkage map in tetraploid species using molecular markers. *Genetics* 157:1369–1385
- Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. *Mol Ecol* 3:91–99
- Manifesto M, Schlatter AR, Hopp HE, Suárez EY, Dubcovsky J (2001) Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. *Crop Sci* 41:682–690
- Martins-Lopes P, Lima-Brito J, Gomes S, Meirinhos J, Santos L, Guedes-Pinto H (2007) RAPD and ISSR molecular markers in *Olea europaea* L: genetic variability and molecular cultivar identification. *Genet Resour Crop Evol* 54:117–128
- Milczarski P, Banek-Tabor A, Lebiecka K, Stojawski S, Myskóv B, Masojc P (2007) New genetic map of rye composed of PCR-based molecular markers and its alignment with the reference map of the DS2 × RXL10 intercross. *J Appl Genet* 48: 11–24
- Moretzsohn MC, Leoi L, Proite K, Guimaraes PM, Leal-Bertioli SCM, Gimanes MA, Martin WS, Valls JFM, Grattapaglia D, Bertioli D (2005) A microsatellite-based gene-rich linkage map for the AA genome of *Arachis* (Fabaceae). *Theor Appl Genet* 111:1060–1071
- Morshidi M (1998) Genetic control of isozymes in *Carica papaya* L. *Theor Appl Genet* 103:89–94
- Nishijima W (1994) Papaya. In: Ploetz RC, Zentmyer GA, Nishijima WT, Rohrbach KG, Ohr HD (eds) Compendium of tropical fruit disease. American Phytopath Soc Press, St. Paul, Minnesota, pp 54–70
- Ocampo Pérez J, Dambier D, Ollitrault P, d’Eeckenbrugge GC, Brottier P, Froelicher Y, Risterucci AM (2006) Microsatellite markers in *Carica papaya* L: isolation characterization and transferability to *Vasconcellea* species. *Mol Ecol Notes* 6:212–217

- Ocampo Pérez J, d'Eeckenbrugge GC, Risterucci AM, Dambier D, Ollitrault P (2007) Papaya genetic diversity assessed with microsatellite markers in germplasm from the Caribbean Region. *Acta Hort* 740:93–101
- Oliveira EJ, Marin ALA, Borém A, Fagundes SA, Barros EG, Moreira MA (2005) Molecular-marker assisted selection for development of common bean lines resistant to angular leaf spot. *Plant Breed* 124:572–575
- Oliveira EJ, Dantas JLL, Castellen MS, Machado MD (2008a) Identificação de microssatélites para o mamoeiro por meio da exploração do banco de dados de DNA. *Rev Bras Frutic* 30:841–845
- Oliveira EJ, Vieira MLC, Garcia AAF, Munhoz CF, Margarido GRA, Consoli L, Matta FP, Moraes MC (2008b) An integrated molecular map of yellow passion fruit based on simultaneous maximum-likelihood estimation of linkage and linkage phases. *J Am Soc Hortic Sci* 133:35–41
- Paniego N, Echaide M, Marianne M, Ferrandez L, Tolales S, Faccio P, Fuxan I, Carrera M, Zandomeni R, Suarez E, Hopp HE (2002) Microsatellite isolation and characterization in sunflower (*Helianthus annuus* L). *Genome* 45:34–43
- Parasnis AS, Rama Krishna W, Chowdari KY, Gupta VS, Ranjekar PK (1999) Microsatellite (GATA)₄ reveals sex-specific differences in papaya. *Theor Appl Genet* 99:1047–1052
- Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. *Trends Plant Sci* 1:215–222
- Rafalski JA, Tingey SV (1993) Genetic diagnostics in plant breeding: RAPDs microsatellites and machines. *Trends Genet* 9:275–279
- Rallo P, Dorado G, Martín A (2000) Development of simple sequence repeats (SSRs) in olive tree (*Olea europaea* L). *Theor Appl Genet* 101:984–989
- Ren J, Lu L, Liu X, Tao Z, Zhang C, Wang D, Shen J, Liu W, Tian Y, Zhu Z (2008) Paternity assessment: application on estimation of breeding value in body-weight at first egg trait of egg-laying duck (*Anas platyrhynchos*). *Mol Biol Rep*. doi:10.1007/s11033-008-9432-z
- Santos SC, Ruggiero C, Silva CLSP, Lemos EGM (2003) A microsatellite library for *Carica papaya* L cv Sunrise Solo. *Rev Bras Frutic* 25:263–267
- Scheldeman X, Van Damme P, Urena Alvarez JV, Romero M (2003) Horticultural potential of Andean fruit crops exploring their centre of origin. *Acta Hort* 598:97–102
- Sondur SN, Manshardt RM, Stiles JI (1996) A genetic linkage map of papaya based on randomly amplified polymorphic DNA markers. *Theor Appl Genet* 93:547–553
- Stajner N, Jakse J, Kozjak P, Javornik B (2005) The isolation and characterisation of microsatellites in hop (*Humulus lupulus* L). *Plant Sci* 168:213–221
- Stiles JI, Lemme C, Sondur S, Morshidi MB, Manshardt R (1993) Using randomly amplified polymorphic DNA for evaluating genetic relationships among papaya cultivars. *Theor Appl Genet* 85:697–701
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4. *Mol Biol Evol* 24:1596–1599
- Tan SC, Weinheimer EA (1976) The isoenzyme patterns of developing fruit and leaf of papaya (*Carica papaya* L). *Sains Malays* 5:7–14
- Van Droogenbroeck B, Breyne P, Gotghebeur P, Romeijn-Peeters E, Kyndt T, Gheysen G (2002) AFLP analysis of genetic relationships among papaya and its wild relatives (*Caricaceae*) from Ecuador. *Theor Appl Genet* 105:289–297
- Van Droogenbroeck B, Kyndt T, Maertens I, Romeijn-Peeters E, Scheldeman X, Romero-Motochi J, Van Damme P, Goetghebeur P, Gheysen G (2004) Phylogenetic analysis of the highland papayas (*Vasconcellea*) and allied genera (*Caricaceae*) using PCR-RFLP. *Theor Appl Genet* 108:1473–1486
- Wang X, Wadl PA, Rinehart TA, ScheZer BE, Windham MT, Spiers JM, Johnson DH, Trigiano RN (2009) A linkage map for Xowering dogwood (*Cornus Xorida* L) based on microsatellite markers. *Euphytica* 165:165–175
- Yonemaru J-I, Ando T, Mizubayashi T, Kasuga S, Matsumoto T, Yano M (2009) Development of genome-wide simple sequence repeat markers using whole-genome shotgun sequences of sorghum (*Sorghum bicolor* (L) Moench). *DNA Res* 16:187–193
- Zhivotovsky LA (1999) Estimating population structure in diploids with multilocus dominant DNA markers. *Mol Ecol* 8:907–913
- Zhou H-F, Xie Z-W, Ge S (2003) Microsatellite analysis of genetic diversity and population genetic structure of a wild rice (*Oryza rufipogon* Griff) in China. *Theor Appl Genet* 107:332–339