Euphytica (2010) 173:279–287
DOI 10.1007/s10681-010-0150-y

Polymorphic microsatellite marker set for *Carica papaya* L. and its use in molecular-assisted selection

Eder Jorge de Oliveira · Aline dos Santos Silva · Fabiana Moraes de Carvalho · Lucas Ferraz dos Santos · Juliana Leles Costa · Vanusia Batista de Oliveira Amorim · Jorge Luiz Loyola Dantas

Received: 8 August 2009 / Accepted: 8 February 2010 / Published online: 25 February 2010
© Springer Science+Business Media B.V. 2010

Abstract Conventional methods for the selection of papaya pure lines are time-consuming. Procedures involving the use of molecular markers and the indirect selection of homozygous plants can reduce this time considerably. The objective of this study was to evaluate the informativeness of a microsatellite marker set when used in marker-assisted selection (MAS) for the development of new papaya lines. Eighty-three lines originating from two segregating F_3 populations and from papaya germplasm were used for the molecular analysis of 27 microsatellite primers. Twenty polymorphic microsatellite primers were identified, allowing the identification of 86 alleles, with an average of 3.18 alleles per primer. The observed heterozygosity values were low for both the markers (0.00–0.29) and the individual lines (0.00–0.35). The inbreeding coefficient (*f*) ranged from 0.634 to 1.00. Eleven lines with $f = 1.00$ and 18 lines with *f* varying from 0.953 to 0.961 were identified. In addition, papaya lines showed high genetic diversity, which will certainly contribute to the development of new varieties. Our results show that the use of microsatellites in MAS is a quick and effective procedure for the development of papaya lines.

Keywords Inbred lines · Microsatellites · Molecular-assisted selection · Papaya

E. J. de Oliveira (✉) · V. B. de Oliveira Amorim · J. L. L. Dantas
Embrapa Mandioca e Fruticultura Tropical, Cruz das Almas, BA 44380-000, Brazil
e-mail: eder@cnpmf.embrapa.br

A. dos Santos Silva · L. F. dos Santos · J. L. Costa
Universidade Federal do Recôncavo da Bahia, Cruz das Almas, BA, Brazil

F. M. de Carvalho
Faculdade Maria Milza, Cruz das Almas, BA, Brazil

Introduction

The annual worldwide production of papaya amounts to 6.9 million tons, of which 27.3% is produced in Brazil (FAOSTAT 2007). Despite these high numbers, however, most of the Brazilian production is based on the culture of only a few varieties, resulting in restricted genetic variability. Most papaya cultivars in Brazil belong to the groups Solo and Formosa. The Solo lines produce small-sized, reddish-fleshed fruits weighing 300–650 g, whereas the Formosa lines bear reddish-fleshed, medium-sized fruits (1,000–1,300 g). Hybrids of these lines are of commercial importance in the internal and external market.

The use of papaya hybrids seems to be a worldwide tendency. According to Chan (1992), the productivity of hybrids will eventually surpass that of the parental lines by 199.6%, thereby representing an excellent alternative for commercial purposes.

Hybrid production of papaya is possible due to the absence of inbreeding depression in the species and to its floral biology, which comprises dioecious and hermaphrodite plants (Manshardt and Drew 1998).

For the development of papaya hybrids, pure lines are needed in order to avoid F_1 segregation. Classical improvement procedures to obtain papaya lines are based on the inbreeding of segregating populations and of germplasm accessions of *Carica papaya* L. This strategy results in the selection of pure lines through progeny testing, which can take up to five to six inbreeding generations (average 12 years), depending on the genetic diversity of the background. The pure lines are then identified among the progenies whose phenotypical segregation is considered to be null. Notwithstanding, this methodology is expensive, laborious, time- and space-consuming, and influenced by negative conditions in the environment.

The use of molecular markers, particularly those used in marker-assisted selection (MAS) has allowed important progress to be made in terms of crop improvement. Among the molecular markers currently available, microsatellites (single simple repeats, SSRs) are of particular importance because they show extensive polymorphism as a consequence of the occurrence of different numbers of repeated units within the SSR structure (Morgante and Olivieri 1993). SSRs have other advantages as well, such as high reproducibility, technical simplicity, low cost, high resolution power and, most important of all, codominance (Brondani et al. 1998; Rallo et al. 2000; Oliveira et al. 2006). Microsatellites were first reported in *C. papaya* L. by Sharon et al. (1992), Parasnis et al. (1999), Santos et al. (2003), Oliveira et al. (2008), and Eustice et al. (2008). Reliable and highly polymorphic SSR markers have been used mainly for genetic map construction (Chen et al. 2007) and sexual differentiation (Parasnis et al. 1999; Santos et al. 2003) and to access genetic diversity (Ocampo Pérez et al. 2007).

The use of codominant markers, such as SSRs, in MAS allows the early identification of plants with high levels of homozygosity in segregating progenies and germplasm accessions. Hence, the objectives of this study were: (1) to access the potential of SSR markers for the detection of polymorphism in papaya lines; (2) to identify papaya pure line using SSR in MAS; (3) to estimate and structure the genetic variability of the lines in order to obtain hybrids, or for other purposes.

Materials and methods

Plant material

Eighty-three papaya lines (46 "Formosa" and 37 "Solo" types), previously selected for their agronomical potential, were evaluated. The fruit of the "Solo" group is generally pear-shaped and ranges in weight from 300 to 700 g; the pulp is orange-yellow, salmon pink, or red when ripe, depending on the cultivar. Fruist of the "Formosa" group are elongate, range from 1,200 to 2,000 g, and are salmon-pink or red when ripe.

Eleven plants were the result of a cross (carried out in field conditions) between the papaya hybrid Calimosa and common papaya, two plants resulted from crossing var. *Golden* and *Sunrise*, and three plants were F_3 inbreds of the Tainung No. 1 hybrid. In addition, 67 plants were selected from among different germplasm accessions of the Active Papaya Germplasm Bank (APGM) at Embrapa Mandioca e Fruticultura Tropical (CNPMF) located in Bahia, Brazil (Table 1). One plant of each line was used in the tests.

DNA extraction

DNA was extracted from young papaya leaves based on the procedure described by Doyle and Doyle (1990). The DNA concentrations of the electrophoresis products on an agarose gel (1.0% w/v) were estimated by comparing the fluorescent signal from the DNA products stained with ethidium bromide (1.0 mg/ml) to a dilution series of commercial Lambda DNA (Invitrogen, Carlsbad, CA) of known concentration.

Selection of SSR marker for MAS

A set of 27 SSR polymorphic primers developed by Oliveira et al. (2008) and tested for polymorphism in *C. papaya* germplasm by Oliveira et al. (2010) was used. All of these SSR primers are independent.

Each PCR reaction mixture consisted of 15 ng template DNA, 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 0.3 mM of each primer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs and 0.5 U *Taq* DNA polymerase (Invitrogen) in a total volume of 15 µl. The reaction mixtures were cycled on a PTC-100 thermal cycler (MJ Research, Watertown, MA) at 94°C for 4 min, followed by 35

Table 1 Papaya (*Carica papaya* L.) lines selected from F_3 segregating populations and from among accessions maintained in the APGB at CNPMF

Lineage	Group[a]	Derived[b]	Origin	Lineage	Group[a]	Derived[b]	Origin
L09	S	F_3 (Cal × Co)	Brazil	L50	F	CMF074	Brazil
L08	S	F_3 (Cal × Co)	Brazil	L51	S	CMF082	South Africa
L10	S	F_3 (Cal × Co)	Brazil	L52	S	CMF092	USA
L11	F	F_3 (Cal × Co)	Brazil	L53	S	CMF094	Brazil
L12	F	F_3 (Cal × Co)	Brazil	L54	S	CMF099	Brazil
L13	F	CMF002	Costa Rica	L55	S	CMF106	Brazil
L14	F	CMF003	Taiwan/China	L56	F	CMF114	South Africa
L15	F	CMF004	USA	L57	F	CMF115	South Africa
L16	S	CMF005	Brazil	L58	S	CMF118	Brazil
L17	S	CMF005	Brazil	L59	S	CMF126	Brazil
L18	S	CMF006	USA	L60	S	CMF128	Taiwan/China
L19	F	CMF010	Malaysia	L61	F	CMF129	Brazil
L20	F	CMF011	Costa Rica	L62	F	CMF129	Brazil
L21	S	CMF012	Malaysia	L63	F	CMF132	USA
L22	S	CMF014	Malaysia	L64	F	CMF135	Brazil
L23	F	CMF018	Taiwan/China	L66	S	CMF138	Brazil
L24	F	CMF018	Taiwan/China	L65	S	CMF138	Brazil
L25	S	CMF020	Brazil	L67	F	CMF146	Brazil
L26	S	CMF021	Brazil	L68	F	CMF147	Brazil
L27	S	CMF021	Brazil	L69	F	CMF176	Brazil
L28	F	CMF022	Malaysia	L70	F	CMF211	Brazil
L29	F	CMF023	Malaysia	L71	S	CMF230	Brazil
L30	F	CMF028	Costa Rica	L72	S	CMF230	Brazil
L31	S	CMF029	Brazil	L73	S	CMF232	Brazil
L32	F	CMF031	Costa Rica	L74	S	CMF232	Brazil
L33	F	CMF032	Costa Rica	L75	F	CMF233	Brazil
L34	F	CMF033	Brazil	L76	F	CMF233	Brazil
L35	F	CMF040	Brazil	L77	S	CMF234	Brazil
L36	F	CMF040	Brazil	L78	S	CMF234	Brazil
L37	F	CMF041	Brazil	L79	S	CMF235	Brazil
L38	F	CMF041	Brazil	L80	S	F_3 (Cal × Co)	Brazil
L39	F	CMF050	Brazil	L81	F	F_3 (Cal × Co)	Brazil
L40	F	CMF052	Brazil	L82	S	F_3 (Cal × Co)	Brazil
L41	F	CMF060	USA	L83	S	F_3 (Cal × Co)	Brazil
L42	F	CMF065	Brazil	L84	S	F_3 (Cal × Co)	Brazil
L43	F	CMF068	Brazil	L85	S	F_3 (Cal × Co)	Brazil
L44	F	CMF069	Brazil	L86	S	F_3 (Gol × Sun)	Brazil
L45	F	CMF070	Brazil	L87	S	F_3 (Gol × Sun)	Brazil
L47	F	CMF070	Brazil	L88	F	F_3 (Tainung)	Brazil
L48	F	CMF070	Brazil	L89	S	F_3 (Tainung)	Brazil

Table 1 continued

Lineage	Group[a]	Derived[b]	Origin	Lineage	Group[a]	Derived[b]	Origin
L46	F	CMF070	Brazil	L90	F	F_3 (Tainung)	Brazil
L49	F	CMF074	Brazil				

[a] S and F, Accessions from the papaya groups Solo and Formosa, respectively

[b] F_3 (Cal × Co), F_3 plants originating from crosses between the papaya hybrid Calimosa × common papaya; F_3 (Gol × Sun), F_3 plants originating from crosses between papaya var. *Golden* and *Sunrise*; Tainung, F_3 inbreds of the Tainung No. 1 hybrid; CMF, code used to identify accessions in the Active Papaya Germplasm Bank (APGB) at Embrapa Mandioca e Fruticultura Tropical (CNPMF)

cycles of 94°C for 40 s, (annealing temperature of 56–60°C depending on the SSR primer; Table 2) for 40 s, and 72°C for 1 min, with a final extension at 72°C for 2 min. Two protocols were used for electrophoresis: (1) fragments with a size difference <10 bp were electrophoresed on a 6% (w/v) denaturing polyacrylamide gel in a Hoefer SQ3 DNA sequencer gel electrophoresis unit (Pharmacia Biotech, San Francisco, CA) at 70 W for 2.5 h and the gel subsequently stained with silver nitrate according to Creste et al. (2001); (2) fragments >10 bp were electrophoresed on a 3% agarose 1000 gel (Invitrogen) at 130 V for 3.5 h. We used the 50-bp ladder (New England Biolabs, Beverly, MA) as a molecular-weight standard to estimate the size of the SSR alleles.

Data analysis

The software program PowerMarker version 3.25 (Liu and Muse 2005) was used to estimate the average number of alleles per locus (N_A), the observed heterozygosity (H_O) and expected heterozygosity (H_E) values, and the inbreeding coefficient (f) in each inbred line across loci and pair-wise shared allele genetic distance. The pair-wise genetic distance matrix between inbred lines was used for constructing the dendrogram using the UPGMA (unweighted pair group method with arithmetic mean) algorithm on Mega ver. 4 (Tamura et al. 2007).

Results

SSR polymorphism analysis

Of the 27 primer pairs analyzed, seven (26%) were monomorphic, while the remaining 20 were polymorphic, showing a distinct variation in allele size (Fig. 1). Table 2 summarizes the marker names, the repeat motifs, the primer sequences, the annealing temperatures (T_a), the number of alleles, H_E, and H_O for single-locus SSRs. In total, there were 86 alleles, and the number of alleles per locus ranged from one (CP17, CP31, CP34, CP36, CP44, CP57, and CP64) to eight (CP35). There was an average of 3.18 alleles per locus when the data set of all individuals was considered. These results show that the H_O ranged from 0.00 to 0.29, with an average of 0.09, and that the H_E ranged from 0.00 to 0.90, with an average of 0.38 (Table 2).

Inbreeding of papaya lines

The results of the heterozygosity analysis of each line are showed in Table 3. The inbreeding coefficient (f) ranged from 0.634 to 1.00 and from 0.83 to 1.00 among the inbred lines derived from papaya germplasm and the F_3 segregant population, respectively. Among the former, nine inbred lines (L36, L40, L42, L44, L45, L54, L58, L68, and L73) had an inbreeding coefficient equal to 1.00.

The average inbreeding coefficient of lines derived from hybrid Tainung No. 1 and those of the two F_3 (Calimosa × Comum) and F_3 (vars. *Golden* × *Sunrise*) lines are 0.849, 0.904 and 1.000, respectively. In this group, only two inbred lines (L86 and L87), with f equal to 1.000, were derived from the last segregant population. The other 18 inbred lines, 15 from germplasm and three from F_3 (Calimosa × Comum), had inbreeding coefficients ranging from 0.953 to 0.961 (Table 3).

Genetic diversity within papaya inbred lines

The H_E in all of the inbred lines was nominally larger that the H_O for all loci and varied from 0.93 (L58,

Table 2 Characteristics of the 27 polymorphic simple sequence repeat loci developed for *C. papaya* L.

Locus	Repeat motif	Primer sequence (5′–3′)		T_a (°C)	N_A	H_E	H_O
		Forward	Reverse				
CP01	$(at)_{15}$	aggggaaggatgtcgttg	acccgcctggaagtaaat	58	5	0.45	0.20
CP02	$(agg)_9$	aggcgaaatcggaagagag	ctggtaaaacgacgatgacg	59	3	0.50	0.24
CP05	$(at)_{17}$	gtcctcaatccgaagcat	catacacccttgtggcttct	58	2	0.30	0.00
CP07	$(gt)_{12}gct(gt)_{13}$	cctagcattgccttgaggtc	gcccactattcacattcacacc	56	3	0.53	0.19
CP10	$(taca)_4(ta)_9(ga)_{10}$	aaaaatcacagcacgtatggtt	gaaattacaaatgggcaaaaag	59	6	0.60	0.14
CP14	$(ac)_9(at)_8$	tcaatgttctcgtcgatagtc	tgggatagtgcaaattggt	59	6	0.71	0.18
CP16	$(at)_{13}$	tcaactatttcccccgcata	cacctccttgtccaaaggtt	58	7	0.70	0.01
CP17	$(taa)_8$	gccatgcagacccaaaaac	caccaccaccaccacctt	60	1	0.00	0.00
CP19	$(at)7t(at)_8$	tagggttgtgcgtccata	agcaggctaaaaactggtca	58	3	0.55	0.16
CP20	$(at)_5(tg)_8$	tgtgagattgtctgttggttg	gggctcgaaaatcaaaacat	60	2	0.47	0.01
CP28	$(ca)_9(ta)_{11}$	tgtcagttcacactgcaaat	ggacattactcgtttgacacc	58	3	0.48	0.07
CP31	$(at)_6(gt)_{10}$	aagggtacgtcatggagca	tctgtcgccttttatactcttg	57	1	0.00	0.00
CP33	$(attaa)_5 + (taa)_9$	tgacccccagttttcgatt	tggtgtagcgtccattgtg	59	2	0.04	0.04
CP34	$(at)_{13}$	tgacaaatgcgatctgttcc	ggcaagcattagcctgcat	59	1	0.00	0.00
CP35	$(ta)_{12}$	ggacgaagctccacaatca	ggcaatcaaaccaaatgagg	56	8	0.69	0.14
CP36	$(at)_{12} + (ta)_{10}$	gcaagaaagcaacttggtaa	tcagacaatgacttgttactgc	56	1	0.00	0.00
CP38	$(aat)_{11}$	ttctttacagttgcctgcat	aacaagtccccgtttttca	58	2	0.21	0.00
CP44	$(at)_{12}$	tgacaacgaactacatcccta	cctcatggtttgtgtactcct	60	1	0.00	0.00
CP47	$(ta)_{14}$	gagcagattgtcacatgcaga	ccagaatgccaatttttgct	57	3	0.55	0.13
CP48	$(tc)_{13}$	ccatttctgtcacgcatcc	gatgatgggccaaattcag	58	2	0.40	0.29
CP49	$(at)_{12}$	cctgaaagcaaccatttcta	tcgctggagctgtaagaga	57	6	0.51	0.05
CP52	$(at)_{10}(ag)_{12}$	ggaaagatcatagaaacagtgg	tgctatcttggttgtctctca	59	6	0.64	0.26
CP57	$(ca)_6(ta)_7$	ttgagtcttggtttcaactcc	ttcccactatcttctgtttgg	60	1	0.00	0.00
CP59	$(ta)_7 + (at)1_2$	gttgtttgcatcccactgc	ctcgccattccatctggt	59	2	0.46	0.00
CP64	$(tc)_{17}$	gggaggacaaagctccaaa	gcatgatccaagggaggag	60	1	0.00	0.00
CP66	$(tg)_9cgc(ga)_{12}$	agtcccatcaggcttctcg	ccttttgtgcgcatatggt	59	4	0.44	0.18
CP72	$(atac)_7(at)_6$	cccaaatcacctttttctctc	aacgtgaactgagggtgga	59	4	0.90	0.15
Average					3.18	0.38	0.09

T_a, Annealing temperature; N_A, number of alleles; H_O, observed heterozygosity; H_E, expected heterozygosity

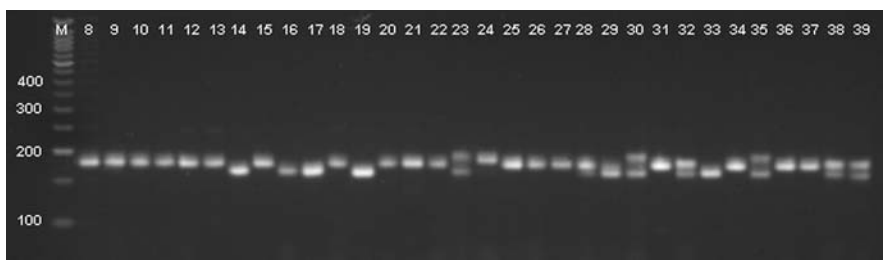

Fig. 1 Simple sequence repeat (SSR) polymorphisms in inbred lines of *Carica papaya* L. from the Active Papaya Germplasm Bank (APGB) at Embrapa Mandioca e Fruticultura Tropical (CNPMF) for locus CP07 visualized on 3% agarose 1000 gels. M 50-bp ladder (New England Biolabs). *Numbers above the lanes* refer to the number of the inbred line, i.e., *lane 8* = L08, *lane 39* = L39

Table 3 Papaya lines from the APGB at CNPMF and corresponding expected heterozygosity (H_E), observed heterozygosity (H_O), and inbreeding coefficient (f)

Genotype	H_E	H_O	f	Genotype	H_E	H_O	f
L36	0.94	0.00	1.000	L41	0.95	0.08	0.912
L40	0.94	0.00	1.000	L70	0.94	0.08	0.911
L42	0.95	0.00	1.000	L15	0.96	0.11	0.884
L44	0.94	0.00	1.000	L25	0.96	0.11	0.884
L45	0.95	0.00	1.000	L31	0.95	0.11	0.884
L54	0.94	0.00	1.000	L35	0.95	0.11	0.884
L58	0.93	0.00	1.000	L48	0.95	0.11	0.883
L68	0.95	0.00	1.000	L67	0.94	0.11	0.882
L73	0.95	0.00	1.000	L19	0.94	0.11	0.881
L86	0.95	0.00	1.000	L65	0.93	0.11	0.881
L87	0.95	0.00	1.000	L21	0.95	0.12	0.879
L33	0.95	0.04	0.961	L30	0.95	0.12	0.879
L43	0.95	0.04	0.961	L32	0.95	0.12	0.879
L78	0.94	0.04	0.961	L53	0.95	0.12	0.879
L14	0.96	0.04	0.960	L79	0.95	0.12	0.879
L23	0.95	0.04	0.960	L16	0.95	0.12	0.878
L37	0.95	0.04	0.960	L63	0.95	0.12	0.878
L50	0.95	0.04	0.959	L89	0.94	0.12	0.878
L52	0.94	0.04	0.959	L57	0.94	0.12	0.877
L66	0.94	0.04	0.959	L85	0.96	0.12	0.875
L10	0.94	0.04	0.959	L61	0.94	0.13	0.861
L38	0.95	0.04	0.958	L81	0.95	0.15	0.845
L72	0.95	0.04	0.958	L76	0.95	0.15	0.844
L80	0.95	0.04	0.958	L46	0.94	0.15	0.843
L83	0.95	0.04	0.958	L17	0.95	0.15	0.839
L20	0.95	0.04	0.956	L18	0.95	0.15	0.839
L39	0.94	0.04	0.954	L28	0.95	0.15	0.839
L60	0.94	0.04	0.954	L34	0.95	0.15	0.839
L56	0.93	0.04	0.953	L47	0.95	0.15	0.839
L77	0.95	0.07	0.922	L12	0.95	0.15	0.839
L11	0.0.95	0.07	0.922	L75	0.95	0.15	0.838
L22	0.96	0.08	0.920	L90	0.95	0.15	0.838
L26	0.95	0.08	0.919	L27	0.95	0.16	0.832
L51	0.95	0.08	0.919	L59	0.95	0.16	0.832
L71	0.95	0.08	0.919	L88	0.95	0.16	0.832
L74	0.95	0.08	0.919	L82	0.95	0.16	0.831
L08	0.95	0.08	0.919	L13	0.95	0.19	0.798
L09	0.94	0.08	0.919	L49	0.94	0.19	0.796
L84	0.94	0.08	0.919	L55	0.94	0.20	0.789
L24	0.94	0.08	0.918	L64	0.93	0.24	0.743
L69	0.95	0.08	0.916	L62	0.94	0.35	0.634
L29	0.94	0.08	0.915				

L56, L19, and L64) to 0.96 (L14, L22, L25, L31, and L85). The H_O of inbred lines ranged from 0.00 (L36, L40, L42, L44, L45, L54, L58, L68, L73, L86, and L87) to 0.35 (L62) (Table 3).

The dendrogram from a UPGMA yielded a tree with two main groups, A and B (Fig. 2). Group A comprises 32 inbred lines and group B, 51. All 16 F_3 plants clustered together with 35 inbred lines derived from germplasm in group A, while group B formed a clearly separated cluster with only germplasm accessions. The H_E of the inbred lines indicates a high genetic variability, as revealed by the 13 major branches in the genetic distance dendrogram, eight on group A and five on group B.

The F_3 (*Golden* × *Sunrise*) inbred lines clustered only in sub-group A8 (L86 and L87), while the F_3 (Calimosa × Comum) was clustered in sub-group A1 (L09 and L11), A2 (L10 and L12), A6 (L08), and A8 (L80, L81, L82, L83, L84, and L85). The F_3 (Tainung) clustered in two sub-groups, A7 (L88) and A8 (L89 and L90). L13, L88, and L59 clustered alone in sub-group A5, A7, and B2, respectively (Fig. 2).

The papaya lines derived from the same germplasm accession were clustered in different subgroups, as observed in L23 and L24 (A1, A3), L35 and L36 (A2, A1), L37 and L38 (A1, A6), L45, L46, L47, and L48 (B1, B5), and L65 and L66 (B1, B4), although all were within the major group. However, the L71 and L72 inbred lines clustered in the B4 and A8 sub-group, respectively.

The 11 inbred lines with a coefficient equal to 1.00 were not grouped into a single cluster, indicating that a considerable genetic divergence exists. These were clustered in three sub-groups: A1 (L36), B1 (L40, L42, L44, L45, L54, L58, and L68) and A8 (L73, L86 and L87). There was was no preferential clustering among genotypes of the same papaya groups, i.e., Solo and Formosa, as shown in Fig. 2.

Discussion

We analyzed a set of 83 hermaphrodite inbred lines that had been adapted to a specific set of environments and optimized for yield and other desirable traits, with the aim of increasing the genetic variability and availability of new F_1 hybrids. A major

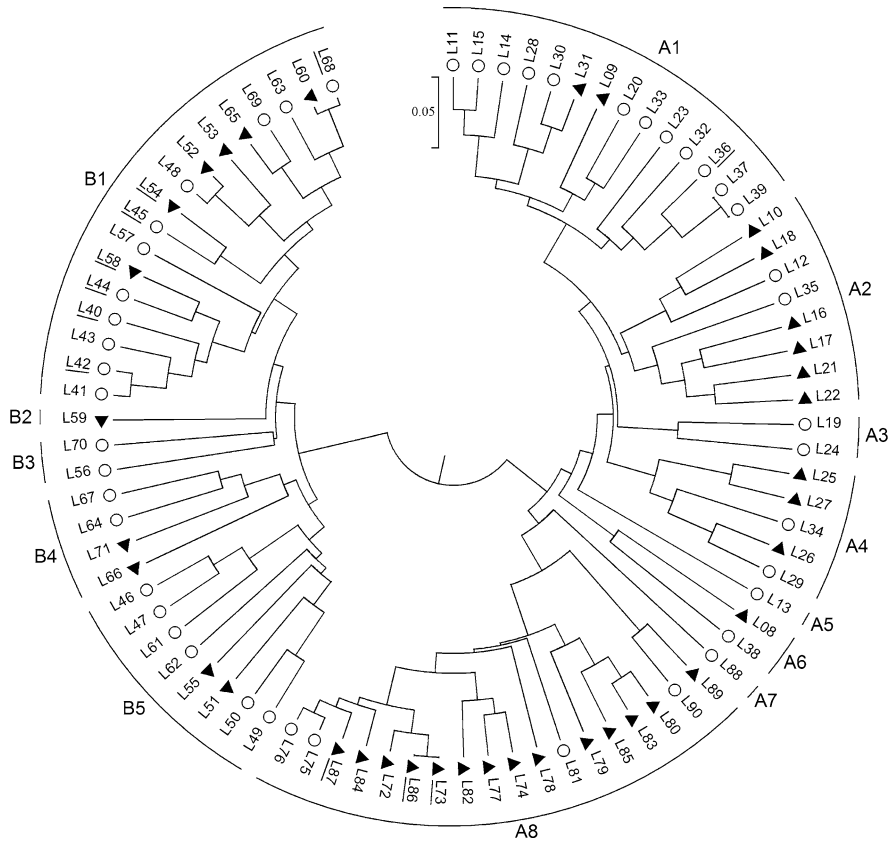

Fig. 2 The unweighted pair group method with arithmetic mean (UPGMA) dendrogram of 83 papaya inbred lines from APGB at CNPMF) constructed from the shared allele genetic distance using SSR markers. *Filled triangle* and *open circle* represent accessions of Solo and Formosa group, respectively. The papaya inbred lines underlined have $f = 1.00$

advantage of pure lines is that one can replicate identical genotypes, replacing individual measurements with group measurements, such as the yield in a plot. Further, individuals can be replicated over environments, such as different locations and years (Walsh and Lynch 2005).

In a number of plants, inbreeding has been shown to have severe negative effects on desirable traits (Cornelius and Dudley 1974; Damgaard and Loeschcke 1994; Huang et al. 1995). However, papaya lines can be obtained and maintained without inbreeding depression (Chan 1992; Dinesh et al. 1992). The finding that increased inbreeding did not disrupt developmental stability is consistent with our own field observation attempts to link developmental stability with homozygosity.

Theoretically, F_3 populations have approximately 75% homozygosity for most loci, whereas in papaya germplasm accessions maintained by outcrossing, this percentage may vary according to the allelic composition of the original population. However, using SSR markers, we were able to identify 11 pure lines with $f = 1.00$ and 18 inbred lines with f ranging from 0.953 to 0.961. The identification of these lines using SSR markers paves the way to advances in papaya breeding programs in terms of time, financial resources, and field space, since the final validation of agronomic performance and crosses aimed at developing of new hybrids can be initiated at this stage. In comparison, conventional breeding procedures would require at least five cycles of selection and self pollination. To the best of our knowledge, this is the first study using SSR in MAS to obtain pure lines of papaya.

Although the genetic map position of these SSR markers is unknown, this high number of markers will enable researchers to select the most informative and well-distributed SSR loci in the papaya genome for use in molecular analysis, since the markers are independent. Another problem with SSR markers in MAS is the presence of a null allele. In the codominant SSR marker system, an individual with a single null allele at a given locus will always appear to be homozygous at that locus. The presence of null alleles decreases the estimated frequency of

heterozygotes and results in incorrect estimates of allele frequencies and overestimates of inbreeding. Even though the identification of exact genotypes is often not possible when sporadic null alleles occur, using a high number of molecular markers may overcame this problem.

Our results show that the use of SSRs allows the detection of high-level polymorphisms among inbred lines. Using 20 pairs of SSR primers, we obtained a total of 86 different alleles among the 83 inbred lines, despite the similarity of these lines in terms of genetic background. We found, on average, a smaller number of alleles per marker (3.18) than Ocampo Pérez et al. (2007), who analyzed 72 papaya accessions from 13 different geographic origins (6.6). The total number of alleles reported in diversity studies is usually proportional to the sample size, and some of the differences seen here may be attributed to sampling differences and plant material. Moreover, the observed heterozygosity values are remarkably low compared to those reported for SSRs in other genotypes, such as germplasm accessions (Ocampo Pérez et al. 2007). The significant deficit in the frequency of heterozygotes observed was due to the origin of the material origin (inbred lines).

We identified a large genetic diversity among the papaya inbred lines. Our cluster analysis based on SSR markers grouped the inbred lines into two main groups. Group A consisted of F_3 plants and 35 lines derived from the germplasm selection, while group B contained only inbreds that belonged to the germplasm selection (Fig. 2). Using isozyme and morphological characteristics, Ocampo et al. (2006) demonstrated high levels of heterozygosity and diversity in Venezuela, Guadeloupe, and Barbados accessions.

This large genetic diversity among the inbred lines will allow the formation of different similarity groups that can then be used to diversify cultivars in commercial fields, ultimately leading to the development of papaya hybrids through crosses between lines of the different similarity groups. On the basis of theoretical and experimental results, Melchinger (1999) demonstrated that the organization of germplasm into genetically divergent heterotic groups is beneficial for a systematic and optimum exploitation of heterosis.

The results of many studies have indicated that MAS may be a more efficient, accurate, and simpler strategy for breeding selection than selection based only on phenotype (Kwon et al. 2001). The former is mainly effective for the selection of early generations and traits that could overcome the difficulty of phenotypic detection. We have demonstrated that the use of SSRs in MAS is an effective procedure for developing papaya inbred lines in that genotyping and selection can be carried out in early generations on those homozygous individuals bearing the desired traits.

Acknowledgments This work was supported by Empresa Brasileira de Pesquisa Agropecuária, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado da Bahia (Fapesb).

References

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

Chan YK (1992) Progress in breeding of F_1 papaya hybrids in Malaysia. Acta Hortic 292:41–50

Chen C, Yu Q, Hou S, Li Y, Eustice M, Skelton RL, Veatch O, Herdes RE, Diebold L, Saw J, Feng Y, Qian W, Bynum L, Wang L, Moore PH, Paull RE, Alam M, Ming R (2007) Construction of a sequence-tagged high-density genetic map of papaya for comparative structural and evolutionary genomics in Brassicales. Genetics 177:2481–2491

Cornelius PL, Dudley JW (1974) Effects of inbreeding by selfing and full-sib mating in a maize population. Crop Sci 14:815–819

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

Damgaard C, Loeschcke V (1994) Inbreeding depression and dominance-suppression competition after inbreeding in rapeseed (*Brassica napus*). Theor Appl Genet 88:321–323

Dinesh MR, Iyer CPA, Subramanyam MD (1992) Genetical study in papaya (*Carica papaya* L.). Acta Hortic 321: 152–163

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

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 (2007) Food and Agriculture Organization of the United Nations statistical database. Available at: http://faostat.fao.org/site/567/default.aspx#ancor. Accessed 04 June 2009

Huang H, Harding J, Byrne T, Famula T (1995) Effects of inbreeding on cut-flower yield in gerbera. Euphytica 81:157–161

Kwon YS, Eun MY, Sohn JK (2001) Marker-assisted selection for identification of plant regeneration ability of seed-derived calli in rice (*Oryza sativa* L.). Mol Cells 1:103–106

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129

Manshardt RM, Drew RA (1998) Biotechnology of papaya. Acta Hortic 461:65–73

Melchinger AE (1999) Genetic diversity and heterosis. In: Coors JG, Pandey S (eds) The genetics and exploitation of heterosis in crops. American Society of Agronomy, Crop Science Society of America (ASA–CSSA), Madison, pp 99–118

Morgante M, Olivieri AM (1993) PCR amplified microsatellites in plant genetics. Plant J 3:175–182

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 Hortic 740:93–101

Ocampo J, d'Eeckenbruggeb GC, Bruyère S, de Lapeyre de Bellaire L, Ollitrault P (2006) Organization of morphological and genetic diversity of Caribbean and Venezuelan papaya germplasm. Fruits 61:25–37

Oliveira EJ, Padua JG, Zucchi MI, Vencovsky R, Vieira MLC (2006) Origin, evolution and genome distribution of microsatellites. Genet Mol Biol 29:294–307

Oliveira EJ, Dantas JLL, Castellen MS, Machado MD (2008) Identificação de microssatélites para o mamoeiro por meio da exploração do banco de dados de DNA. Rev Bras Frutic 30:841–845

Oliveira EJ, Amorim VBO, Costa JL, Matos ELS, Castellen MS, Dantas JLL, Pádua JG (2010) Polymorphism of microsatellite markers in papaya (*Carica papaya* L.). Plant Mol Biol Report. doi: 10.1007/s11105-010-0180-6

Parasnis AS, Krishna WR, Chowdari KY, Gupta VS, Ranjekar PK (1999) Microsatellite $(GATA)_4$ reveals sex-specific differences in papaya. Theor Appl Genet 99:1047–1052

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

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

Sharon D, Hillel A, Vainstein A (1992) Applications of DNA fingerprints for identification and genetic analysis of *Carica papaya* and other *Carica* species. Euphytica 62:119–125

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Walsh B, Lynch M (2005) Selection and development of pure lines. In: Walsh B, Lynch M (eds) Evolution and selection of quantitative traits. Available at: http://nitro.biosci.arizona.edu/zbook/volume_2/vol2.html. Accessed 15 April 2009