Crop Breeding and Applied Biotechnology 7:36-42, 2007
Brazilian Society of Plant Breeding. Printed in Brazil

Monitoring of the genetic variability in papaya parent 'Formosa' of 'UENF/CALIMAN 01' hybrid via RAPD

Francisco Filho da Silva[1], Messias Gonzaga Pereira[1*], Pedro Corrêa Damasceno Junior[1], Rogério Figueiredo Daher[1], Telma Nair Santana Pereira[1], Gonçalo Apolinário de Souza Filho[1], Alexandre Pio Viana[1], and Geraldo Antônio Ferreguetti[2]

Received 09 December 2005

Accepted 25 April 2006

ABSTRACT – *The genetic improvement of papaya in Brazil, in hybridizations and studies of combining ability, can broaden the genetic base of the species and brush aside the constraints in choosing cultivars for planting. The objective of this study was to monitor the advance of selfing generations in the parent Formosa of 'UENF/CALIMAN 01', the first Brazilian papaya hybrid, using RAPD markers. For this purpose 27, 30 and 31 plants of the C_1, C_2 and C_3 cycles, respectively, were genotyped. The coefficients of mean genetic dissimilarity were 0.1766, 0.0555 and 0.0445 in the C_1, C_2 and C_3 cycles, respectively. The reduced mean distance in the C_3 cycle indicates a greater genetic homogeneity of these plants, fulfilling a basic requirement for the creation of uniform hybrids. The establishment of the C_4 cycle would therefore achieve no significant gains in reducing the genetic distance but cause a delay in the release of hybrid seeds.*

Key words: *Carica papaya*, molecular markers, Jaccard index, UPGMA, Tocher clustering method.

INTRODUCTION

One of the main problems concerning the cultivation of papaya (*Carica papaya* L.) in Brazil are the limited alternatives for the choice of varieties and/or commercial hybrids for planting that would feed the national as well as the international market. Moreover, the priceyness of hybrid papaya seed of the group 'Formosa', usually imported from Taiwan at 4,200 US dollars per kilogram (Silva et al. 2004), has caused many fruit producers to plant the F_2, F_3 and F_4 generations of the hybrid 'Tainung 01' in succession. This has triggered a number of problems, in the first place, the loss of vigor and segregation of fruit shape.

A measure for solving this problem is to broaden the genetic base of papaya, by means of intra and interspecific hybridizations. The use of this latter possibility is however limited by interspecific barriers that must be overcome through appropriate procedures such as the use of the cultivation technique of in vitro embryos. Interespecific hybridization alongside a backcross program effectively enables the introgression of new genes of interest into cultivated varieties, releasing broad genetic variability in the subsequent generations (Siqueira et al. 1988).

The study of combining ability in diallel crosses can contribute substantially to the success of programs of plant genetic improvement, making the identification of parents for hybridization and of hybrids with superior

[1] Laboratório de Melhoramento Genético Vegetal, Centro de Ciências e Tecnologias Agropecuárias, Universidade Estadual do Norte Fluminense Darcy Ribeiro. Av. Alberto Lamego, 2000, 28.013.600, Campos dos Goytacazes, RJ, Brasil. *E-mail: messias@uenf.br
[2] Caliman Agrícola S.A., C. P. 52 – 29.900-970, Linhares, ES, Brasil

qualities possible as well as helping in the identification of the most adequate selection method, since the genetic parameters are known beforehand (Cruz and Regazzi 1997).

Due to the high allogamy rate in the ‘Formosa’ genotypes the parents available have variable degrees of loci in heterozygosis. The evolution of the fixation degree of the parents must therefore be monitored in order to produce vigorous and uniform hybrids.

On this background, along with the classic improvement procedures, the use of the DNA marker techniques has opened a completely new outlook on plant genetic improvement and is being successfully used for several crops, including papaya. Some of the diverse applications of DNA markers are in studies on the genetic diversity in germplasm accessions (Lanza et al. 1997, Wadt et al. 2004), in the construction of genetic linkage maps (Ferreira and Grattapaglia 1998), QTL mapping (quantitative trait loci) (Kato 2004, Pereira and Lee 1995), aside from representing an important tool in selection studies.

The objective of this study was to estimate the mean genetic dissimilarity coefficients in the 1st, 2nd and 3rd cycle (C_1, C_2 and C_3) of successive selfings of the ‘Formosa′ parent of ‘UENF/CALIMAN 01’, first Brazilian papaya hybrid (Solo x Formosa), released in 2003, using RAPD markers.

MATERIAL AND METHODS

Genetic material

Three generations (C_1, C_2 and C_3) and the reference papaya population ‘Formosa’ in the S_0 generation (UENF-1594), designated C_1, were used. The C_2 and C_3 cycles were originated through selfing of one tree of generation C_1 and C_2, respectively.

This ‘Formosa’ parent was chosen owing to its morpho-agronomic traits and, above all, based on previous studies on the general and specific combing abilities (GCA and SCA) of this genotype crossed with genotypes from the ‘Solo’ group, to which the second parent of the hybrid ‘UENF/CALIMAN 01’ belongs. For this purpose, young leaves were randomly collected from hermaphrodite and female plants, from 27, 30 and 31 trees of the C_1, C_2 and C_3 cycles, respectively, originated from the papaya germplasm of the UENF, located at the CALIMAN AGRÍCOLA S.A company in Linhares, state of Espirito Santo. The variation in the sampled tree numbers (27, 30 and 31) was brought about by discarding the trees whose amplification standards were not reliable enough. The leaf samples were sent to the laboratory of plant genetic improvement of the Centro de Ciências e Tecnologias Agropecuárias of the Universidade Estadual do Norte Fluminense Darcy Ribeiro (LMGV/CCTA/UENF), in Campos dos Goytacazes, state of Rio de Janeiro, where they were grinded in liquid nitrogen and deep-frozen (-86 ºC), in closed 15 mL tubes (Falcon).

Genomic DNA extraction

From each ground leaf sample, approximately 200 mg tissue was transferred to 1.5 mL Eppendorf tubes placed in liquid nitrogen. The genomic DNA was extracted by the CTAB method (Doyle and Doyle 1990), with some modifications described by Daher et al. (2002). After the extraction, the DNA concentrations in the samples were estimated using papaya DNA at a known concentration as standard. The total genomic DNA was evaluated through electrophoresis in 0.8 % agarose gel as indicator of the integrity and purity of the extracted DNA. After quantification and gauging of the DNA purity and integrity, the samples were standardized at a concentration of 10 ng.µL-1.

RAPD markers

Thirty RAPD primers were used in polymerase chain reactions (PCR), as listed below: OPA 12. OPD 04. OPD 15. OPD 20. OPE 06. OPF 12. OPG 10. OPN 09. OPN 13. OPO 10. OPO 15. OPR 15. OPV 14. OPW 02. OPAA 12. OPAA 17. OPAB 01. OPAB 08. OPAC 01. OPAC 14. OPAE 01. OPAE 02. OPAE 11. OPAF 07. OPAG 11. OPAH 04. OPAH 14. OPAH 18. OPAI 03 and OPAJ 20.

The polymerase chain reactions were performed according to Williams et al. (1990) with modifications in a final volume of 20 µL, containing Tris-HCl 10 mM (pH 8.3), KCl 50 mM, $MgCl_2$ 2.4 mM, 100 µM of each one of the deoxyribonucleotides (dATP, dCTP, dGTP, and dTTP), 0.3 µM of each initiator, 20 ng genomic DNA, and one unit of the enzyme Taq DNA polymerase. The amplifications were run in the thermocycler (Perkin Elmer GeneAmp PCR System 9700) set at at 95 ºC for 1 min followed by 45 cycles of 1 min at 94 ºC, 1 min at 36 ºC and 2 min at 72 ºC plus a final extension step of 7 min at 72 ºC. After the amplifications, the sample temperature was reduced to 4 ºC. The amplification products were electrophoretically analyzed in 1.0 % agarose gel and visualized after staining with ethidium bromide under UV light, by the Eagle Eye II imaging system.

Molecular data analysis

The RAPD profiles of each genotype were obtained by the presence (1) or absence (0) of bands considering the joint analysis of the three cycles. The same analysis was carried out in the cycles 1, 2 and 3 separately, and the complement of the Jaccard index (Alfenas et al. 1991) was calculated as 1- a/(a+b+c), where **a** corresponds to the number of bands found simultaneously in both genotypes, **b** to the number of bands in genotype A only and **c** to the number of bands in genotype B only.

Statistical analysis

The genetic distance matrix, considering the three cycles jointly, was used to perform the clustering analysis by the hierarchic method (UPGMA - Unweighted Pair Group Method with Arithmetic Mean), as well as by Tocher´s optimization method, cited by Rao (1952). A second clustering analysis used the genetic distance matrix of each cycle separately, by means of the hierarchic method (UPGMA). For the analysis of the polymorphic markers, the Chi-square test was applied at a probability level of 5 %.

To run the statistical analyses we used the software packages GENES, version 0.1.0 and STATISTICA version 95 (StatSoft Inc, Tulsa, Oklahoma, EUA).

RESULTS AND DISCUSSION

A descriptive analysis of the monomorphic and polymorphic markers for each selfing cycle, and all of them jointly, considering the three cycles simultaneously, is presented in Table 1. Of all the 134 markers, 63 were polymorphic and 71 were monomorphic, that is, each primer generated a mean of 2.10 polymorphic and 2.37 monomorphic bands. The total number of bands per primer varied from 2 to 7; 47.01 % of the bands presented polymorphisms.

The chi-square test indicated that the reduction in the number of polymorphic markers from C_1 (60) to C_2 cycle (23) and from C_2 (23) to C_3 cycle (17), is near the expected, which would be from 60 in C_1 to 30 in C_2 cycle and from 30 in C_2 to 15 in C_3 cycle (Table 1). Based on this testing of the hypothesis, there is no reason at all to disagree that the reduction of the polymorphism throughout the selfing cycles was 50% in each generation. One must bear in mind that the polymorphic markers in a particular cycle reflect the proportion of non-fixed loci, and are therefore related to loci in heterozygosis. Considering moreover that the subsequent cycle was derived from the selfing of a single tree, the polymorphic loci of the following cycle express the proportion of loci in heterozygosis of the selfed plant. As expected, the proportion of loci in heterozygosis is reduced by half in each selfing generation, so the polymorphism is reduced as well, according to the observed results. Moreover, as expected, an increase in the number of monomorphic markers was observed from C_1 to C_2 cycle and from C_2 to C_3 cycle and a decrease of the total number of markers considering C_1 for C_2 cycle and C_2 for C_3 cycle (Table 1).

The extreme and mean values of the genetic dissimilarity coefficient estimated for each cycle with the respective amplitudes and standard deviations are shown in Table 2. The amplitude of these values was widest in the C_1 cycle, indicating considerable divergence in this cycle. In comparison these amplitudes were reduced by 49.46 and 54.76 % in the C_2 and C_3 cycles, respectively. However in a comparison of the C_2 and C_3 cycles, the magnitude of reduction of the amplitude of genetic divergence was only 10.48 %.

In a comparative analysis of the means and their respective standard deviations, the reduction of both parameters was fairly expressive from the C_1 to C_2 cycle and in spite of a smaller magnitude between the C_2 and C_3 cycles, the means were significantly different by the t test at 5 % probability. Despite the significant difference between the means of the C_2 and C_3 cycles, the magnitude of the difference between them was small (0.011) compared with the magnitude of the difference between the C_1 and C_2 cycles (0.1211). This indicates sufficient genetic uniformity in the C_3 cycle available for procedures of hybridization and establishment of uniform hybrids,

Table 1. Descriptive analysis of the monomorphic and polymorphic markers in each generation separately and together

Markers											
Monomorphic				Polymorphic				Total			
C1	C2	C3	General	C1	C2	C3	General	C1	C2	C3	General
73	93	97	71	60	23	17	63	133	116	114	134

Table 2. Maximum, minimum and mean values and the total amplitude of the genetic dissimilarity coefficient estimates of each cycle, obtained with the arithmetic complement of the Jaccard index, with the respective mean standard deviations

Cycle	Genetic dissimilarity				Standard deviation
	Minimum	Maximum	Mean	Amplitude	
C_1	0.0700	0.2833	0.1766 a	0.2133	0.0429
C_2	0.0093	0.1171	0.0555 b	0.1078	0.0205
C_3	0.0000	0.0965	0.0445 c	0.0965	0.0179

Means followed by the same letter do not differ from each other by the t test at 5 % probability

including genetically identical trees (genetic distance = 0.000), considering the experimental conditions of this study.

The clustering analysis based on the genetic distance took the three cycles into consideration, using the hierarchical clustering method UPGMA (Figure 1). Based on this clustering analysis one large group with 87 trees (98.86 %) was formed, considering the three cycles, at a genetic distance of 21.0 %. One single tree of group C_1 was left outside this clustering, expressing considerable divergence. When a distance of 6.5 % was taken as reference, the C_2 and C_3 cycles were separated in two groups with small distortions: one C_2 genotype (3.33 %) was inserted into group C_3, one C_3 genotype (3.22 %) into group C_2 and one C_1 genotype (1A) (3.70 %) into group C_2, while four C_2 genotypes (13.33 %) were excluded from this clustering. The analysis of the dendrogram as a whole shows a strong reduction in the genetic distance between the genotypes of the C_2 and C_3 cycles in comparison with the C_1 cycle, as well as a greater dispersion in distances among the genotypes of the C_1 and C_2 cycles, corroborating the results presented in Table 1.

Plant 1A had been selected to generate the C_2 cycle and its presence in this cycle had therefore been expected. On the other hand, the identity of the C_2 plant selfed to generate the C_3 cycle was not recorded and

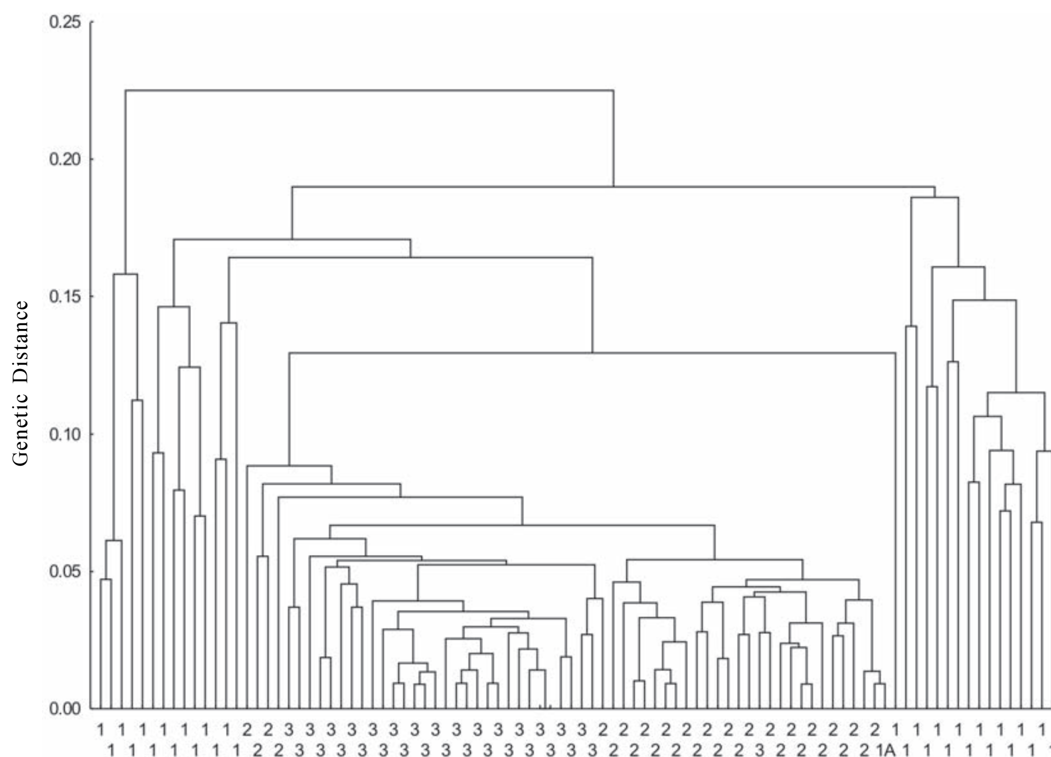

Figure 1. Genetic dissimilarity among and between the C_1, C_2 and C_3 papaya selfing cycles, assessed by the UPGMA, based on the genetic distance matrix obtained through RAPD markers. 1A corresponds to the selfed C_1 plant to generate the C_2 cycle

one can therefore only infer that the C_2 tree that grouped together with the C_3 plants probably represented it. The allocation of trees from previous in subsequent cycles is expected since their alleles are passed on to the following generation or generations.

For a better illustration of the evolution in the reduction of the mean genetic distance in the C_1, C_2 and C_3 cycles a separate clustering was performed for each cycle, based on the genetic distance matrix and using the UPGMA method. The evolution in the reduction of the genetic distance between the trees of the C_1, C_2 and C_3 cycles is well characterized (Figure 2) by the marked dispersion in the C_1 cycle, with a range of genetic distance varying from 0.04 to 0.24. At the level of 6.0 % diversity all trees were differentiated and at a genetic distance of 8.0 %, six genotypes (22.22 %) grouped into the C_2 cycle. In comparison with the C_1 cycle, dispersion was somewhat reduced within a range of 0.00 to 0.09, so that eight trees (26.67 %) were grouped at a genetic distance of 1.0 %. In the C_3 cycle though less expressive, considering the comparison of the C_1 and C_2 cycles, a reduction in the dispersion of the distances was observed in a range from 0.00 to 0.07, where eight trees (25.81 %) were grouped at 1.0 % genetic distance.

The results of the clustering of the 88 papaya trees of the C_1 (27), C_2 (30) and C_3 (31) cycles by means of Tocher´s optimization method are displayed in Table 3. In a first attempt of clustering, five groups were formed which contained the C_2 and C_3 cycles and nine trees of the C_1 cycle in the first group, and the others (18 C_1) in the groups II, III, IV and V. In a next step, the trees of the first group were re-grouped in six sub-groups (Ia, Ib, Ic, Id, Ie, and If) (Table 3). With this re-clustering, barring C_1a (selfed C_1 plant to generate cycle C_2), all C_1 trees were moved to different sub-groups (Ib, Ic, Id, Ie, and If) while the C_2 and C_3 cycles remained in a single sub-group (Ia).

With the re-clustering of sub-group Ia, thirteen new sub-groups were formed (Table 3) (Iaa, Iab, Iac, Iad, Iae, Iaf, Iag, Iah, Iai, Iaj, Ial, Iam and Ian), which distinguished the C_2 and C_3 more clearly. In this last analysis, a greater uniformity of the C_3 cycle with the formation of five distinct sub-groups (Iaa, Iac, Iae, Iaj and Iam) was observed, of which the first contains 77.42 % of the trees. On the other hand, the C_2 trees were regrouped in eight sub-groups (Iab, Iad, Iaf, Iag, Iah, Iai, Ial and Ian), of which Iab contained 70 %, along with one tree of each one of the C_1 and C_3 cycles.

Stiles et al. (1993), who used RAPD markers in the evaluation of the genetic divergence in ten papaya tree cultivars, verified the formation of three distinct groups by a dendrogram, discriminating within and among the study cultivars, in agreement with the origin and the pedigree as it is known.

In the present study, the techniques of clustering (Tocher´s optimization and UPGMA hierarchic clustering method) indicate that the RAPD marker data agreed with the genetic variability that would be expected between the different selfing cycles, evidencing a significantly greater genetic dispersion in the C_1 cycle and a reduced one in the C_3 cycle. This greater genetic dispersion in the C_1 cycle is justified

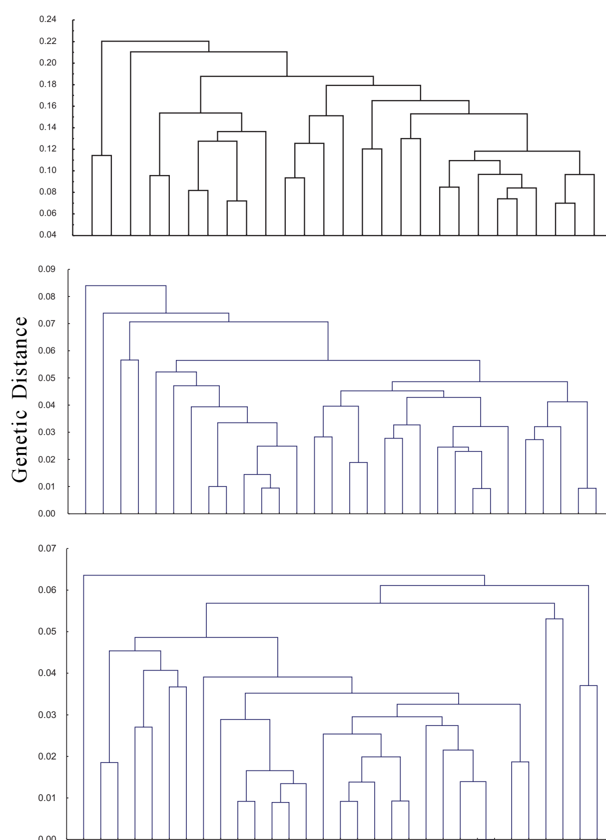

Figure 2. Genetic dissimilarity in 27, 30 and 31 papaya plants of the C_1 (A), C_2 (B) and C_3 (C) cycles, respectively, assessed by the UPGMA, based on the genetic distance matrix obtained through RAPD markers.

Table 3. Clustering analysis through Tocher´s optimization method, based on the arithmetic complement of the Jaccard index of 88 papaya plants of the C_1, C_2 and C_3 cycles, using RAPD markers

Groups/Sub-groups	Plants
I a	C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C2, C3, C3, C3, C2, C2, C2, C3, C2, C2, C2, C2, C2, C2, C2, C2, C2, C2, C2, C2, C1A, C2, C2, C2, C2, C3, C2, C3, C2, C2, C2, C2, C2, C2, C2, C2, C2
I aa	C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3, C3
I ab	C2, C2, C2, C2, C2, C2, C2, C1a, C2, C2, C2, C2, C2, C2, C2, C2, C2, C2, C2, C2, C3, C2, C2
I ac	C3, C3
I ad	C2, C2
I ae	C3, C3
I af	C2, C2
I ag	C2
I ah	C2
I ai	C2
I aj	C3
I al	C2
I am	C3
I an	C2
I b	C1, C1
I c	C1, C1
I d	C1, C1
I e	C1
I f	C1
II	C1, C1, C1, C1, C1, C1, C1, C1, C1, C1, C1, C1
III	C1, C1, C1
IV	C1, C1
V	C1

C_1A, corresponds to the C_1 plant selfed to generate the C_2 cycle

since the cultivars of papaya tree ‘Formosa’ tend to have a marked allogamy rate, promoting genetic recombinations and therefore preserve a certain genetic variability.

Even in cultivated autogamous plants such as common bean, which has a low rate of cross fecundation, molecular genetic variability within cultivars has been observed (Menezes et al. 2004). These authors identified broad molecular genetic variability in common bean lines sampled within the cultivars Carioca and Pérola. They inferred that the cultivar Carioca used by farmers is completely different from the original cultivar. This variability is ascribed to the mixture of cultivars and mutations generated in the vast areas planted with common bean which is most welcome for broadening the genetic base of these genotypes for improvement.

ACKNOWLEDGEMENTS

We thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for granting one of the authors a doctoral scholarship and the Financiadora de Estudos e Projetos (FINEP) and the Caliman Agrícola S/A (CALIMAN) company for financial and logistic support.

Monitoramento da variabilidade genética em progenitor de mamoeiro 'Formosa' do híbrido 'UENF/ CALIMAN 01' via RAPD

RESUMO – *O melhoramento genético do mamoeiro no Brasil, por meio de hibridações e estudos de capacidade combinatória, pode possibilitar a ampliação da base genética da espécie e superar as limitações quando da escolha de cultivares para o plantio. O objetivo deste trabalho foi monitorar o avanço de gerações de autofecundação no progenitor Formosa do 'Uenf/ Caliman 01', primeiro híbrido de mamão brasileiro, utilizando-se marcadores RAPD. Para tanto, 27, 30 e 31 plantas dos cycles C_1, C_2 e C_3, respectivamente, foram genotipadas. Os coeficientes de dissimilaridade genética médios foram de 0,1766, 0,0555 e 0,0445, nos cycles C_1, C_2 e C_3, respectivamente. No cycle C_3, a reduzida distância média indica uma maior homogeneidade genética deste material, adequando-se a um pré-requisito básico para produção de híbridos uniformes. Portanto, a obtenção do cycle C_4 não resultaria em ganhos significativos na redução da distância genética e implicaria atraso na liberação das sementes híbridas.*

Palavras-chave: *Carica papaya*, marcadores moleculares, índice de Jaccard, UPGMA, método de agrupamento de Tocher.

REFERENCES

Alfenas AC, Peters I, Brune W and Passador GC (1991) **Eletroforese de proteínas e isoenzimas de fungos e essências florestais**. Editora UFV, Viçosa, 242p.

Cruz CD and Regazzi AJ (1997) **Modelos biométricos aplicados ao melhoramento genético**. Editora UFV, Viçosa, 390p.

Daher RF, Pereira MG, Tupinambá EA, Amaral Júnior AT, Aragão WM, Ribeiro FE, Oliveira LO and Sakiyama NS (2002) Assessment of coconut tree genetic divergence by compound sample RAPD marker analysis. **Crop Breeding and Applied Biotechnology 2**: 431-438.

Doyle JJ and Doyle JL (1990) Isolation of plant DNA from fresh tissue. **Focus 12**: 13-15.

Ferreira ME and Grattapaglia D (1998) **Introdução ao uso de marcadores moleculares em análise genética**. Embrapa – Cenargen, Brasília, 220p.

Kato T (2004) Quantitative trait loci controlling the number of spikelets and components traits in rice: the main effects and interaction with years. **Breeding Science 54**: 125-132.

Lanza LLB, Souza Junior CL, Ottoboni LMM, Vieira MLC and Souza AP (1997) Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. **Theoretical and Applied Genetics 94**: 1023-1030.

Menezes MVCO, Santos JB and Menezes M (2004) Genetic diversity in Carioca and Pérola cultivars of common bean based on RAPD markers. **Crop Breeding and Applied Biotechnology 4**: 178-182.

Pereira MG and Lee M (1995) Identification of genomic regions affecting plant height in sorghum and maize. **Theoretical and Applied Genetics 90**: 380-388.

Rao RC (1952) **Advanced statistical methods in biometric research**. Jonh Wiley and Sons, New York, 390p.

Silva MCA, Tarsitano MAN and Corrêa LS (2004) Análise do custo de produção e lucratividade do mamão Formosa Cultivado no município de Santa Fé do Sul (SP). **Revista Brasileira de Fruticultura 26**: 40-43.

Siqueira WJ, Fonseca MIS and Sondhal MR (1988) Regeneração de plantas híbridas entre *Lycopersicon esculentum* e *L. peruvianum* a partir de calos com dois anos de cultura in vitro. **Bragantia 47**: 1-8.

Stiles JI, Lemme C, Sondur S, Morshidi MB and Manshardt R (1993) Using randomly amplified polymorphic DNA for evaluating genetic relationships among papaya cultivars. **Theoretical and Applied Genetics 85**: 697-701.

Wadt LHO, Ehringhaus C and Kageyama PY (2004) Genetic diversity of "Pimenta Longa" genotypes (*Piper spp*., Piperaceae) of the Embrapa Acre germplasm collection. **Genetics and Molecular Biology 27**: 74-82.

Williams JG, Kubelik AR, Livak KJ, Rafalsky LA and Tingey SV (1990) DNA polymorphism amplified by arbitrary primers are useful as genetic markers. **Nucleic Acids Research 18**: 6531-6535.