# Genetic similarity of *Theobroma cacao* L. accessions maintained in duplicate at the cacao research center germplasm collection based on RAPD markers

**Fábio Gelape Faleiro\*[1, 2]; Milton Macoto Yamada[2]; Uilson Vanderlei Lopes[2]; Alessandra Silva Gelape Faleiro[2]; Rita de Cássia Siqueira Bahia[2]; Luana Mahé Costa Gomes[2]; Ronaldo Carvalho Santos[2] and Reinaldo Figueiredo dos Santos[2]**

[1] Embrapa Cerrados, CP 08223, 73301-970, Planaltina, DF, Brazil; [2] Laboratório de Biotecnologia, Seção de Genética, Centro de Pesquisas do Cacau, Caixa Postal 07, CEP 45600-000, Itabuna, BA, Brazil. (\* Corresponding Author. E-mail: ffaleiro@cpac.embrapa.br)

## ABSTRACT

The objective of this work was to use RAPD markers to analyze the genetic similarity of 18 accessions maintained in 37 rows in the germplasm collection of Cacao Research Center (CEPEC/CEPLAC), Itabuna, Brazil. Within each row a plant was chosen at random, from which genomic DNA was extracted. DNA sample of each row was amplified by the RAPD. Fifty one RAPD markers were generated, and then transformed to a matrix of binary data. These binary data were used to estimate genetic similarities and to perform cluster analyses. Genetic similarities of 0.70, 0.94 and 0.96 were observed between plants of the accessions RB 29, Ca 1 and C. Sul 4, respectively. The mistakes in the identification of accessions in different germplasm collections seems to be more common than expected. Also, considering the great importance of the germplasm collections, an international effort would be necessary for the characterization of all important cacao germplasm collections around the world.

KEY WORDS: Cocoa breeding, DNA analysis, molecular markers.

## INTRODUCTION

The germplasm collection is the place where the *ex situ* conservation of accessions is processed, that is to say, conservation out of its habitats, in the form of permanent collections of pollen, seeds, tissue cultures or collections of plants maintained in the field. In the case of cacao, the difficulty of working with its seeds demands that genetic resources are maintained in field collections of living plants, for the preservation of species (Almeida and Dias, 2001). Usually, germplasm collections of *Theobroma cacao* (a perennial plant, with a high genetic diversity) are very big and occupy very extensive areas, making them burdensome and difficult to manage appropriately, demanding much work with various cultivation practices.

The germplasm collection of CEPEC/CEPLAC, in Ilhéus, BA, Brazil, possesses around 1500 accessions of *T. cacao* occupying an area of approximately 20 ha. The accessions in this collection are from several countries around the world, representing practically all the variability existing in the species (Pires et al., 2000). To facilitate the process of statistical data analysis, some accessions are maintained in two or more rows. Such duplication, together with synonymy problems, homonymous and identification mistakes, make the transfer of results and recommendations among different breeding programs quite difficult and increase the level of time and resources needed for evaluating the genetic potential of the preserved accessions. Besides that, confusion of genetic identity between rows reduces the accuracy of the phenotypic evaluations. In that sense, the verification of the genetic similarities of those accessions maintained in two or more rows would be of great importance.

Several genetic markers can be used in the verification of genetic identity, such as the morphological, agronomic and molecular (Figueira, 1998; Christopher et al., 1999; Figueira and Cascardo, 2001). However, the molecular markers, especially those based on PCR (Polymerase Chain Reaction), like RAPD (Random Amplified Polymorphic DNA), present advantages in relation to the other genetic markers, once they are able to yield a practically unlimited number of markers unaffected by the environment. According to Gilmour (1995), RAPD analyses are very precise for the identification of duplicated genotypes, when evaluated in the same laboratory, although reproducibility problems of some bands can happen among laboratories.

The objectives of this study were to use RAPD markers to analyze the genetic identity of 18 accessions maintained in duplicate within the germplasm collection of the Cacao Research Center (CEPEC/CEPLAC).

## MATERIAL AND METHODS

### Genetic material

Eighteen accessions of *Theobroma cacao* L., maintained in 37 rows in the germplasm collection of the Cacao Research Center (CEPEC/CEPLAC) (Table 1) were used in this study. Each accession is maintained in two rows, but the Ca-6 is maintained in three rows. To represent each row, randomly chosen leaves from one plant were collected for DNA extraction.

### DNA extraction

Genomic DNA of one plant in each row was extracted using the CTAB method (Doyle and Doyle, 1990) with some modifications (Araújo et al., 2000). After the extraction, the concentration of the DNA was estimated by spectrophotometer at 260 nm (Sambrook

et al., 1989). Bands of total genomic DNA separated by electrophoresis in 0.8% agarose gel were used as an indication of the integrity and of the purity of the DNA extracted. After the quantification, the DNA samples were diluted to the concentration of 10 ng/mL.

### RAPD markers obtention

DNA samples of each row were amplified by RAPD technique (Random Amplified Polymorphic DNA). The amplification reactions were made in a total volume of 25 μL, containing Tris-HCl 10 mM (pH 8,3), KCl 50 mM, MgCl2 2 mM, 100 μM of each one of the deoxyribonucleotides (dATP, dTTP, dGTP and dCTP), 0.4 μM of a "primer" (Operon Technologies Inc., Boulevard, CA, USA), one unit of the enzyme Taq polimerase and, approximately, 30 ng of DNA. The following decamer primers were used: OPB19, OPC05, OPC19, OPD16, OPD19, OPD20, OPE19 and OPE20. The amplifications were made in a thermocycler, programmed for 40 cycles, each one with the following sequence: 15 seconds at 94 ºC, 30 seconds at 35 ºC and 90 seconds at 72 ºC. After 40 cycles, a final extension stage of 7 minutes at 72 ºC and finally, the temperature was reduced to

**Table 1.** Genetic similarity of accessions of *Theobroma cacao* L. in duplicates in the germplasm collection of the Cacao Research Center(CEPEC/CEPLAC).

| Accessions [1] | Rows number in the field ([2]) | Genetic similarity | *p value* |
|---|---|---|---|
| 52 R [1 and 2] | 402 (2); 666 (1) | 0.97 | 0.15515 |
| Ca 6 [3, 4 and 5] | 359 (1); 507 (5); 508 (10) | 1.00 | - |
| Be 5 [6 and 7] | 264 (2); 514 (1) | 0.97 | 0.15469 |
| Ca 1 [8 and 9] | 512 (10); 257 (1) | 0.94 | 0.07007 |
| C. Sul 4 [10 and 11] | 261 (4); 353 (9) | 0.96 | 0.10430 |
| C. Sul 5 [12 and 13] | 354 (7); 630 (1) | 1.00 | - |
| C. Sul 9 [14 and 15] | 356 (2); 631 (5) | 0.98 | 0.23975 |
| IQ 1 [16 and 17] | 357 (3); 461 (9) | 0.99 | 0.23975 |
| EEG 64 [18 and 19] | 115 (1); 540 (5) | 0.99 | 0.23975 |
| H 17 [20 and 21] | 1006B (2); 1129A (1) | 1.00 | - |
| H 39 [22 and 23] | 1121A (1); 1137 (A) | 0.99 | 0.23975 |
| P 7 [24 and 25] | 738 (1); 447 (7) | 1.00 | - |
| RB 29 [26 and 27] | 252 (6); 711 (10) | 0.70 | 0.0002 |
| RIM 15 [28 and 29] | 163 (3); 406 (1) | 0.97 | 0.15504 |
| RIM 76 [30 and 31] | 164 (1); 416 (1) | 0.99 | 0.23975 |
| SIC 23 [32 and 33] | 103 (10); 330 (1) | 0.97 | 0.15525 |
| SIC 24 [34 and 35] | 104 (4); 331 (2) | 0.99 | 0.23975 |
| SPA 5 [36 and 37] | 349 (9); 456 (5) | 1.00 | - |

[1] Numbers used to identify the accession in the dispersion graph (Figure 1); [2] Number of the sampled plant in the row; *p value* - smaller significance level for which the hypothesis of genetic identity between plants of each accession is rejected.

4 ºC. After the amplification, 3 μL of a mixture of bromophenol blue (0.25%), glicerol (60%) and water (39,75%) were added to each sample. The samples were then applied in agarose gel (1.2%), and submerged in TBE buffer (Tris-borate 90 mM, EDTA 1 mM). The electrophoretic separation was approximately four hours, at 90 volts. At the end of the run, the gels were stained with etidium bromide and photographed under ultraviolet light.

## Statistical analysis

The RAPD markers generated were transformed into a matrix of binary data. From that matrix, genetic similarities were estimated between plants supposedly belonging to the same accession, using the Nei and Li (1979) coefficient of similarity of the Genes program (Cruz, 1997). The hypothesis that similarity coefficients are 1.0 (genetic identity) was tested based

on test $z = \frac{Sij}{\sqrt{\frac{Sij(1-Sij)}{n}}}$ , where, Sij is the genetic similarity

between the rows i and j of the accession in study and n is the number of molecular markers used in the computation of similarity. The value of probability (p value) of the test was calculated using the SAS. Based on the complement of the similarity coefficients, the accessions were plotted in a bi-dimensional dispersion graph based on the minimization of the differences between the original genetic distances and the graphic distances (Cruz and Viana, 1994).

## RESULTS AND DISCUSSION

The primers used in the amplifications generated a total of 51 DNA bands, with an average of 6.4 bands per primer. Out of the 51 bands, 35 (68.6%) were polimorphic and 16 (31.4%) monomorphic. The genetic similarities between plants of the same accession, maintained in different rows, varied between 0.70 and 1.00. For 15 (83.3%) out of the 18 accessions, the similarity was equal or superior to 0.97 (Table 1).

The minimum level of genetic similarity for two accessions to be considered identical depends on the number of markers used in the analyses, as well as the precision of the technique used, and is debated among scientists. Figueira et al. (1997), using RAPD markers to analyze the genetic similarity among accessions maintained in the collections of CEPEC/CEPLAC (Brazil) and BAL Plantations (Malaysia), considered as non-identical those plants with genetic

similarity lower than 0.95. According to the authors, three (27.3%) of the 11 accessions did not show the same genetic identity in the two collections. Later, Figueira (1998) confirmed the results obtained by RAPD analysis based on pod characteristics. Several factors can cause non-identity in collections, such as problems of administration, mistakes in the collections, labelling mistakes, mixture of propagation material in the introduction or in the establishment of the accession in the collection, growth of branches starting from the rootstock, occurrence of homonyms and synonyms, etc. Figueira (1998) also considered that small genetic differences could happen among derived accessions of populations of open pollination collected by different research institutes.

The consideration of genetic identity among two plants supposedly belonging to the same accession depends on the similarity coefficient among them and also the number of markers used. Here, we used the p-value, the smallest significant level for which the hypothesis of genetic identity between plants of the same accession is rejected. If a significance level α=0.15 is adopted (in our work, such value corresponds the genetic similarity inferior to 0.97), three out of the 18 accessions analyzed would have had some problem in the establishment of the germplasm collection (Table 1). Genetic similarities of 0.70, 0.94 and 0.96 were observed between plants of the accessions RB 29, Ca 1 and C. Sul 4, respectively. Phenotypic observations of pods and plant architecture evidenced genetic differences among the two rows of the accession RB 29 and the occurrence of genetic variations inside of the two rows of the accession Ca 1. In the case of the accession C. Sul 4, phenotypic differences were not verified between plants of the two rows evaluated. If a level of significance α=0.10 is adopted, plants of the accession C. Sul 4 would be considered identical.

The analysis of dispersion of the 37 cacao accessions based on the complements of the coefficients of genetic similarity is presented in Figure 1. Such analysis confirmed the differences observed between plants of the accessions Ca 1 and RB 29. The plants of the accession C. Sul 4, although having a similarity of 0.96, were grouped together in the dispersion graph. The dispersion of the accessions followed the same tendency observed by Pires et al. (2000), i.e., a high similarity among accessions of the common variety (SIC, EEG) and the grouping of some series as C. Sul and Criollo accessions (52R, RIM 15 and RIM 76). The accessions RB 29 and SPA 5 and IQ 1 were the most divergent among those analyzed.

Mistakes in the identification of accessions in different germplasm collections seems to be more common than expected. In this work the mistake frequency was 16.7% of the accessions analyzed. Frequency of mistakes of 27.3%, 29.9% and 35% were observed by Figueira et al. (1997), Christopher et al. (1999) and Risterucci et al. (2000), respectively.

Considering the great importance of germplasm collections for the preservation of the variability of the species and agronomic evaluation for use in breeding programs, an international effort would be necessary for the characterization of all the important collections of cacao germplasm of the world. For that, the standardization of the characterization process is necessary using stable, precise molecular markers, with high reproducibility and the high variability degree associated with the different accessions of *T. cacao*. In that sense, Saunders et al. (2000), based on the analysis of the characteristics of different types of molecular markers of the DNA, proposed the establishment of a protocol of molecular characterization of cacao accessions using 15 microsatellite loci. The 15 primer combinations for microsatellites used as pattern were selected considering the level of identified polymorphism per locus, distribution of the loci in different chromosomes and occurrence of different allelic frequencies.

## ACKNOWLEDGEMENTS

## RESUMO

**Identidade genética de acessos de *Theobroma cacao* L. mantidos em duplicatas no banco de germoplasma do Centro de Pesquisas do Cacau com base em marcadores RAPD**

Neste trabalho, objetivou-se utilizar marcadores RAPD para analisar a identidade genética de 18 acessos mantidos em 37 filas no banco de germoplasma do Centro de Pesquisas do Cacau (CEPEC/CEPLAC), Itabuna, Brasil. Para representar cada fila, foi amostrada aleatoriamente uma planta, a partir da qual foi extraído o DNA genômico. Amostras de DNA de cada fila foram amplificadas pela técnica de RA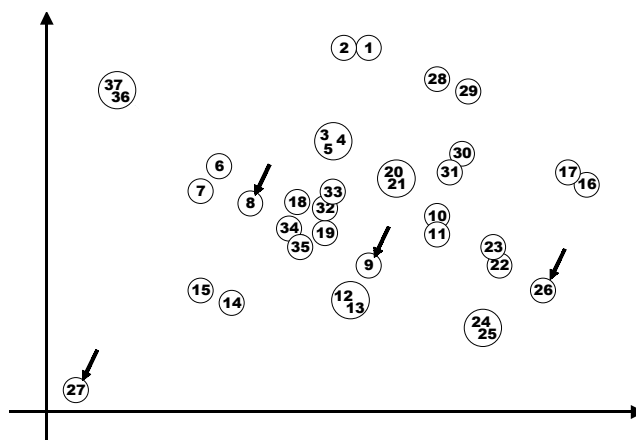PD. Foram gerados 51 marcadores RAPD, os quais foram convertidos em uma matriz de dados binários, a partir da qual foram calculadas similaridades genéticas e realizadas análises de agrupamento. Foram observadas similaridades genéticas de 0,70, 0,94 e 0,96 entre plantas dos acessos RB 29, Ca 1 e C. Sul 4, respectivamente. A presença de erros na identificação de acessos em diferentes bancos de germoplasma parece ser mais comum do que se espera e, nesse sentido, visto a grande importância dos bancos de germoplasma, um esforço internacional seria necessário para a caracterização de todas as importantes coleções de germoplasma de cacaueiro do mundo.



**Figure 1.** Dispersion analysis of 37 accessions of *Theobroma cacao* L (Table 1) based on the genetic distances calculated using RAPD markers. The accessions pointed by the arrows presented genetic similarity lower than 0.95 between the plants.

## REFERENCES

Almeida, C.M.V.C. and Dias, L.A.S. 2001. Recursos genéticos. p.163-216. In: Dias, L.A.S. (Ed.). Melhoramento genético do cacaueiro. Editora Folha de Viçosa, Viçosa.

Araújo, I.S.; Bahia, R.C.S.; Santos, R.F.; Faleiro, F.G. and Ahnert, D. 2000. Otimização da extração e amplificação de DNA de *Theobroma cacao* L. visando a obtenção de marcadores moleculares RAPD. Genetics and Molecular Biology. 23(Supplement):219-220.

Christopher, Y.; Mooleedhar, V.; Bekele, F. and Hosein, F. 1999. Verification of accessions in the ICG, T using botanical descriptors and RAPD analysis. p.15-18. CRU Report for 1998.

Cruz, C.D. 1997. Programa Genes: aplicativo computacional em genética e estatística. Ed. UFV, Viçosa.

Cruz, C.D. and Viana, J.M.S. 1994. A methodology of genetic divergence analysis based on sample unit projection on two-dimensional space. Revista Brasileira de Genética. 17:69-73.

Doyle, J.J. and Doyle, J.L. 1990. Isolation of plant DNA from fresh tissue. Focus. 12:13-15.

Figueira, A. 1998. Homonymous genotypes and misidentification in germplasm collections of Brazil and Malaysia. Ingenic Newsletter. 4:4-8.

Figueira, A. and Cascardo, J.C.M. 2001. Marcadores moleculares no melhoramento. p.385-437. In: Dias, L.A.S. (Ed.). Melhoramento genético do cacaueiro. Editora Folha de Viçosa, Viçosa.

Figueira, A.; Lambert, S.V.; Carpenter, D.; Pires, J.L.; Cascardo, J.C.M. and Romanczyk, L. 1997. The similarity of cocoa flavour of fermented seeds from fingerprinted genotypes of *Theobroma cacao* from Brazil and Malaysia. Tropical Agriculture. 74:132-139.

Gilmour, M. 1995. The BCCCA ringtest on the RAPD analysis of cocoa. p.135-138. In: Proceedings of the International Workshop on Cocoa Breeding Strategies, Malaysian. Cocoa Board, Kuala Lumpur.

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

Pires, J.L.; Marita, J.M.; Lopes, U.V.; Yamada, M.M.; Aitken, W.M.; Melo, G. P.; Monteiro, W.R. and Ahnert, D. 2000. Diversity for phenotypic traits and molecular markers in CEPEC's germplasm collection in Bahia, Brazil. p.72-78. In: Proceedings of the International Workshop on New Technologies and Cocoa Breeding, Ingenic, Malaysia.

Risterucci, A.M.; Eskes, B.; Fargeas, D.; Motamayor, J.C. and Lanaud, C. 2000. Use of microsatellite markers for germplasm identity analysis in cocoa. p.25-33. In: Proceedings of the International Workshop on New Technologies and Cocoa Breeding, Ingenic, Malaysia.

Sambrook, J.; Fritsch, E.F. and Maniats, T. 1989. Molecular cloning: a laboratory manual. 2ed. Cold Spring Harbor, Cold Spring Harbor Laboratory, New York.

Saunders, J.A.; Hemeida, A.A. and Mischke, S. 2000. USDA DNA fingerprinting program for identification of *Theobroma cacao* accessions. p.108-114. In: Proceedings of the International Workshop on New Technologies and Cocoa Breeding, Ingenic, Malaysia.