THE UNIVERSITY OF THE WEST INDIES

COCOA RESEARCH UNIT

REPORT FOR 1995

THE UNIVERSITY OF THE WEST INDIES
ST. AUGUSTINE, TRINIDAD & TOBAGO

COCOA RESEARCH UNIT

REPORT FOR 1995

The Use of RAPD for Characterization and Genetic Assesment of Cacao

Y. Christopher and O. Sounigo

This study reports on the use of RAPD analysis for the realisation of two objectives:

1. Estimation of the genetic diversity of the populations, held at the ICG,T.
2. The identification of the clones of the ICG,T.

The first part of the study consisted in the evaluation of the impact of some potential sources of variation, such as the model of the thermal cycler and the age of the DNA extracts, on the reproducibility of the RAPD patterns. Over 150 10-mer oligonucleotide primers were screened for their ability to reveal stable polymorphisms in Theobroma cacao. This study reports on the use of five of these primers:

- in the DNA fingerprinting analysis of a subset of 47 clones, for which complete data were were available.
- in the estimation of the genetic variability of the 47 fingerprinted clones and of 88 clones representing 11 populations (between 5 and 14 clones per population). The diversity existing within the populations has been evaluated as well as the relationship among these populations.

Material and methods

Plant material

DNA was extracted from mature green leaves of accessions located on campus fields, the ICG,T and Marper Estate. Details on the populations studied are given in Table 1.

DNA Extraction

DNA was extracted according to a modified version of Edwards (Russell *et al.*, 1991). The DNA, suspended in sterile distilled water was then quantitated using a TKO-100 DNA Fluorometer (Hoefer Scientific).

DNA amplification and Electrophoresis

PCR was performed in a 0.5 µL microfuge tube. The final reaction volume was 25 µL, consisting of 0.45ng of DNA, 15ng primer (Operon), reaction buffer, PC2 - 50mM Tris HCL pH 9.1, 16mM ammonium sulphate, 2.5 mM $MgCl_2$ and 150µg/mL BSA). 2 mM each of dATP, dCTP, dGTP, dTTP and 1.25 unit of Klentaq I polymerase.

Table 1: Informations on the studied populations .

Population	group	Country of Origin	Region	Reasons for collection or selection	Sample size in the study
B	Forastero	ECUADOR	Hacienda Bolivar	Collection for Resistance to Witche's broom	8
PA	Forastero	PERU	Parinari	Collection for Resistance to Witches' broom	7
P	Forastero	PERU	Iquitos	Collection for Resistance to Witches' broom	6
AMAZ	Forastero	ECUADOR	Rio Amazonas	Collection for Germplasm Conservation	5
RIM	Trinitario	MEXICO	Izapa station	Selection for Agronomical Traits	5
ICS	Trinitario	TRINIDAD	River Estate	Selection for Yield and Vigour	14
IMC	Forastero	PERU	Iquitos	Collection for Resistance to Witches' broom	10
NA	Forastero	PERU	Rio Nanay	Collection for Resistance to Witches' broom	8
SCA	Forastero	PERU	Hacienda Scavina	Collection for Resistance to Witches' broom	11
SPEC	Criollo Forastero	COLOMBIA	large area	Collection for Germplasm Conservation	7
UF	Forastero Trinitario	COSTA-RICA	Limon, Almirante	Selection for Agronomical Traits	7

PCR reactions were performed in a Perkin Elmer Cetus 480 or Techne PHC-3 automated Thermal Cycler. Thermal cycling parameters were as follows:

94°C for 5 seconds
36°C for 30 seconds
72°C for 1 minute

This was performed for 40 cycles, followed by a final extension step at 72°C for 5 minutes to ensure that all reactions proceeded to completion.

Amplification products were separated on a 1.5% Agarose gel, visualised by staining with ethidium bromide and viewed under a UV light.

Primer Screening

Over 150 primers from Operon Technologies Inc. were screened for their usefulness in revealing reproducible polymorphic amplification products under the defined reaction and thermal cycling conditions. Eighteen interesting primers were identified, and the five primers used in this analysis and their sequences are listed in the Table 2.

Table 2: Primers that revealed polymorphisms among the cacao genotypes.

PRIMER	SEQUENCE
AO8	GTGACGTAGG
AO9	GGGTAACGCC
M17	TCAGTCCGGG
O15	TGGCGTCCTT
P14	CCAGCCGAAC

Reproducibility test of RAPD technique

a) Influence of the thermal-cycler

PCR amplification was performed using either a Perkin Elmer Cetus or Techne PHC-3 thermal cycler. The experiment was performed where the only source of potential variation was the different machines used for thermal cycling. All other variables, that is, DNA sample, reaction mixes and thermal cycling programmes were identical.

b) Influence of the age of the DNA extracts.

PCR amplification was performed with samples that had been stored at 4°C for different periods of time. This was necessary to determine the time limit within which extracts can be used before DNA degradation resulted in changes in RAPD patterns.

Data Analysis

A binary code was used to score the presence or absence of stable polymorphic bands where 1 represented the presence of a particular band and 0 represented its absence.

A simple correspondence analysis was performed on the 47 fingerprinted accessions, after removing the bands present or absent at a frequency lower than 5%. A hierarchical clustering analysis was performed on the 11 populations, using UPGMA. The index used was Nei index, and this was calculated from the frequency of each marker in each population. Only 12 of the markers have been used in this study.

In each population a mean pairwise similarity index (Sokal and Milchener) has been calculated in order to evaluate the within diversity .

Results

Primer Screening

The five primers used revealed a total of 17 stable polymorphic bands ranging in size for 492-2027 base pairs. Fifteen of these are indicated in Figure 1. The number of stable polymorphic loci varied from 1 in primers A09 and M17 to 7 in primer P14. Both monomorphic and polymorphic bands were observed for all primers. Primers differed in their ability to reveal polymorphic variation. P14 was the most discriminatory primer with 7 polymorphic loci.

Reproducibility

a) Thermal Cyclers

No difference has been observed, when using the two different models of thermal cyclers. Identical results were obtained for all extracts and all primers tested. In Figure 2, even minor faint polymorphic bands (600 bp) are clearly reproducible between cyclers. Even in cases where a poor extract seemed to have affected the quality of amplification - lanes (2,3), (8,9) and (14,15) - both cyclers showed identical results.

Once identical programmes and ramp rates are chosen, identical molecular profiles should be obtained. It appears though that a difference in ramp rates (time taken to go from one temperature to the next), drastically affects RAPD results. This was discovered when, due to a faulty block, the 3.5 hour thermal cycling programme used at CRU took approximatedly 5 hours to go to completion on the PHC-3 Techne cycler. This was at first accompanied by fuzzy amplification products, however, the results were still reproducible with those of the other cycler. As the ramp rate of this cycler decreased and the time taken for the PCR to go to completion increased, major and minor bands, not previously observed, began to appear. Re-fitting of the cycler with a new block resulted in a reappearance of previous molecular profiles that were consistent with those obtained using the Perkin Elmer cycler. Hence in transferring this technique important details such as ramp rate of the cyclers used, whether the same model of cycler or not, must be considered.

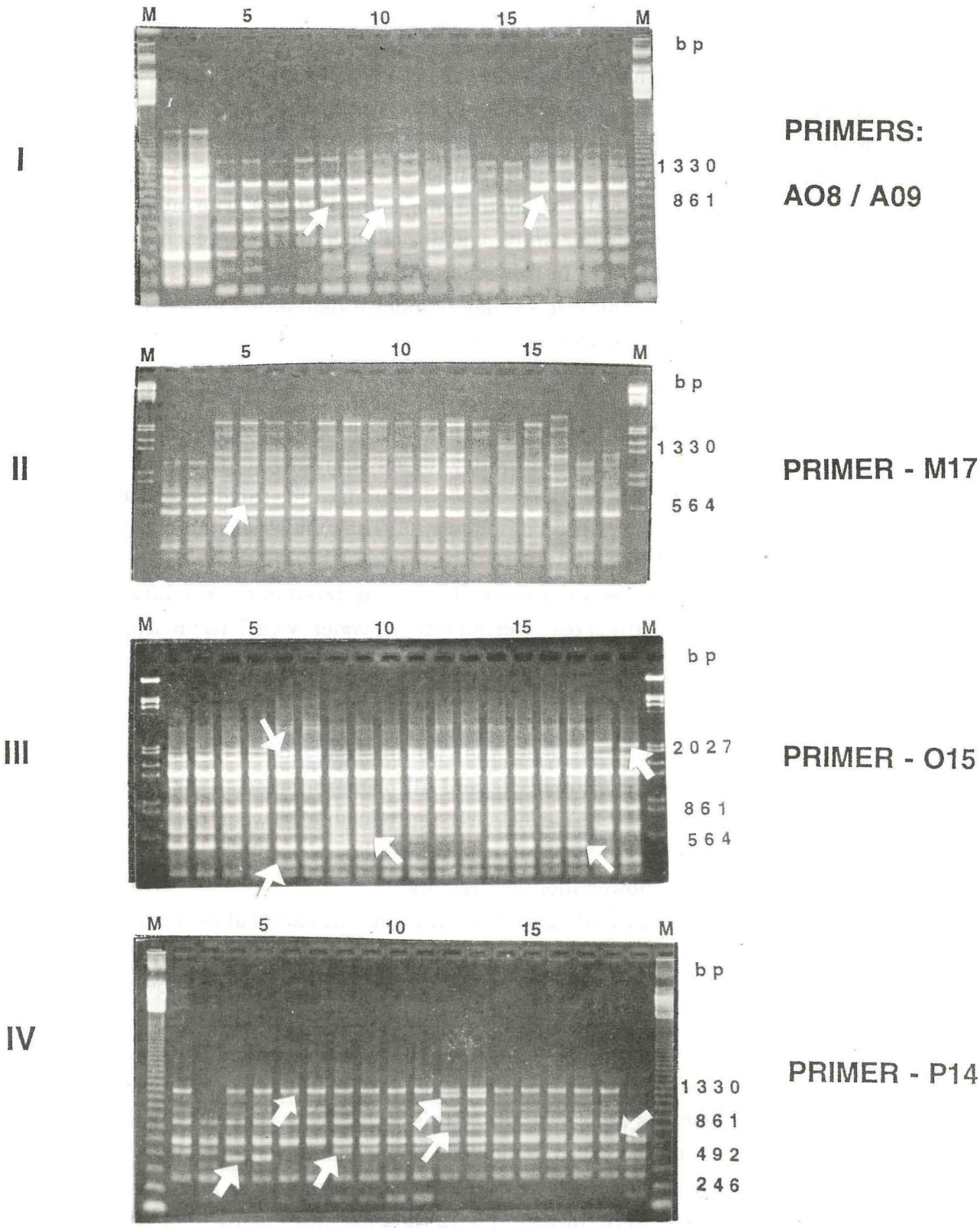


Figure 1. RAPD patterns obtained using five primers I: lanes 4-11:-AO8; lanes 12-19:- AO9, II:M17,III:O15 and IV:P14. The stable polymorphic bands are indicated. The band of molecular weight 2027 base pairs (III:lanes 17-18) is a very rare band. Also the pattern exhibited in (IV: lanes 12-13) is very rare.

Reproducibility Test - Use of Different Models of Thermal Cyclers

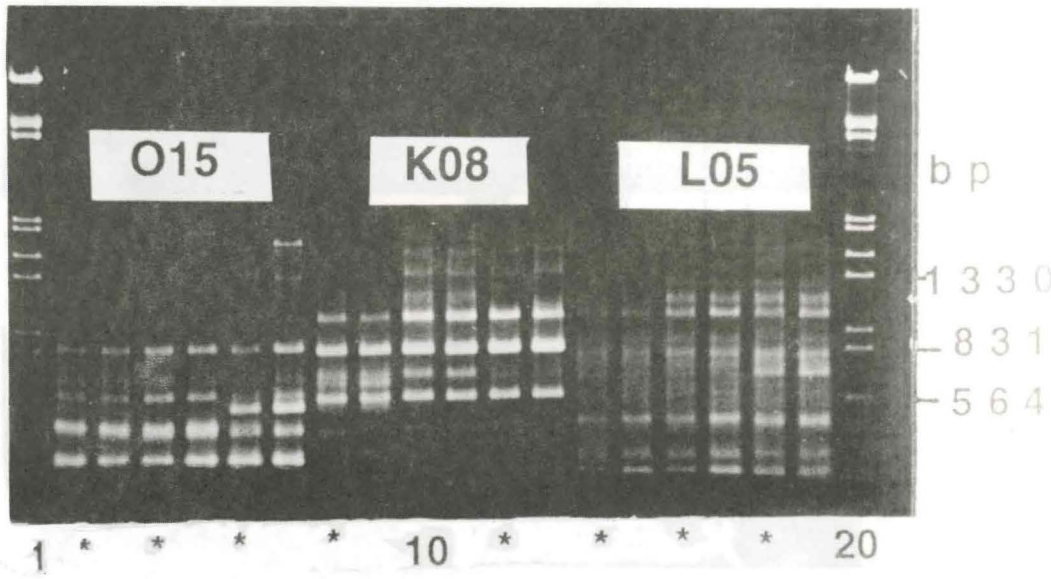


Figure 2. **Comparison of RAPD patterns obtained using two different Thermal Cyclers. Lanes 2, 4, 6, 8, 10, 12, 14, 16 and 18 labelled '*' represent results obtained using the Techne PHC-3 Cycler. Lanes 3, 5, 7, 9, 13, 15 and 17 represent the corresponding results obtained using the Perkin Elmer 480 Cycler. Note that with all three primers used, O15, KO8 and LO5, identical RAPD patterns were obtained.**

REPRODUCIBILITY TEST - Integrity of DNA

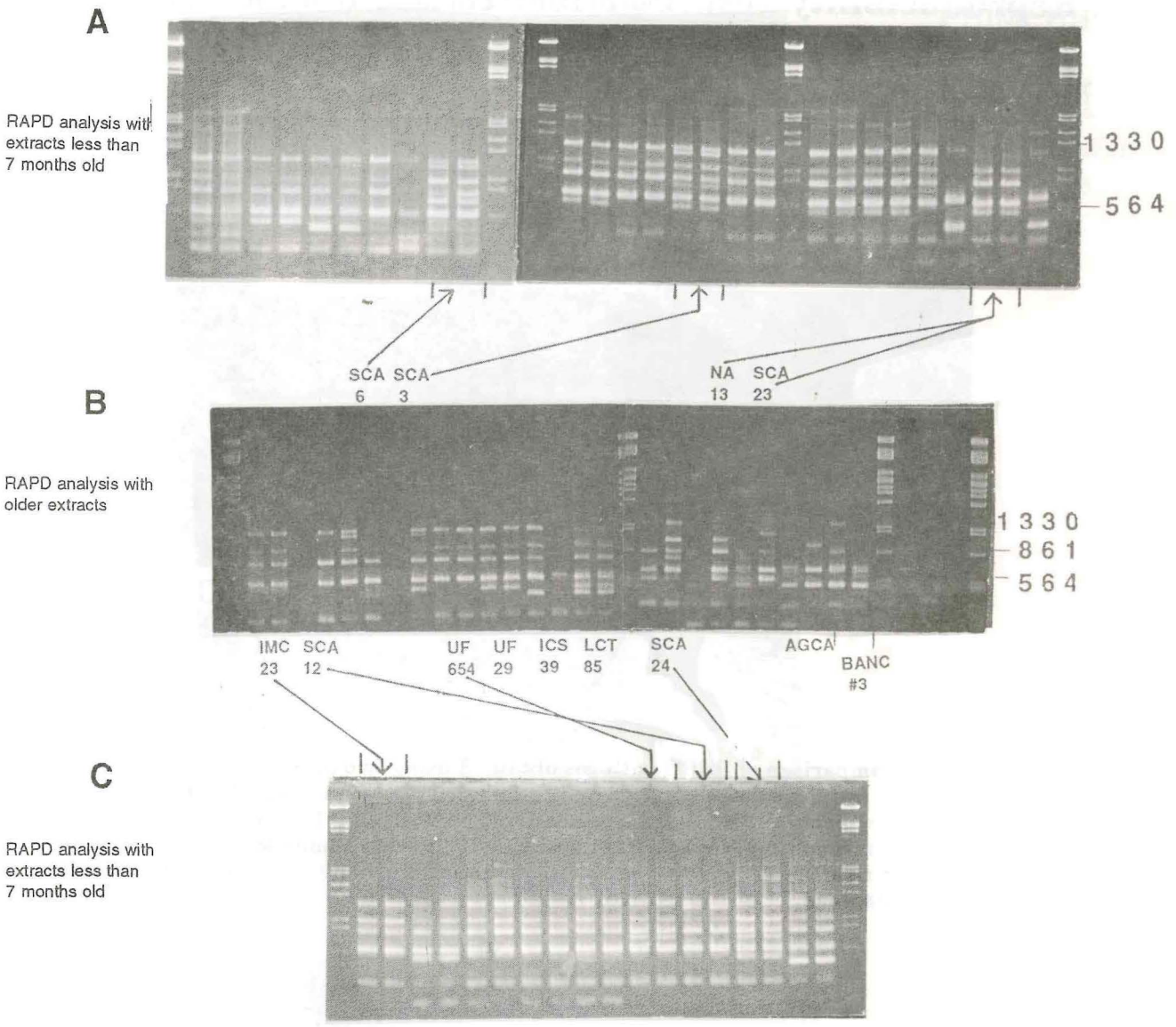
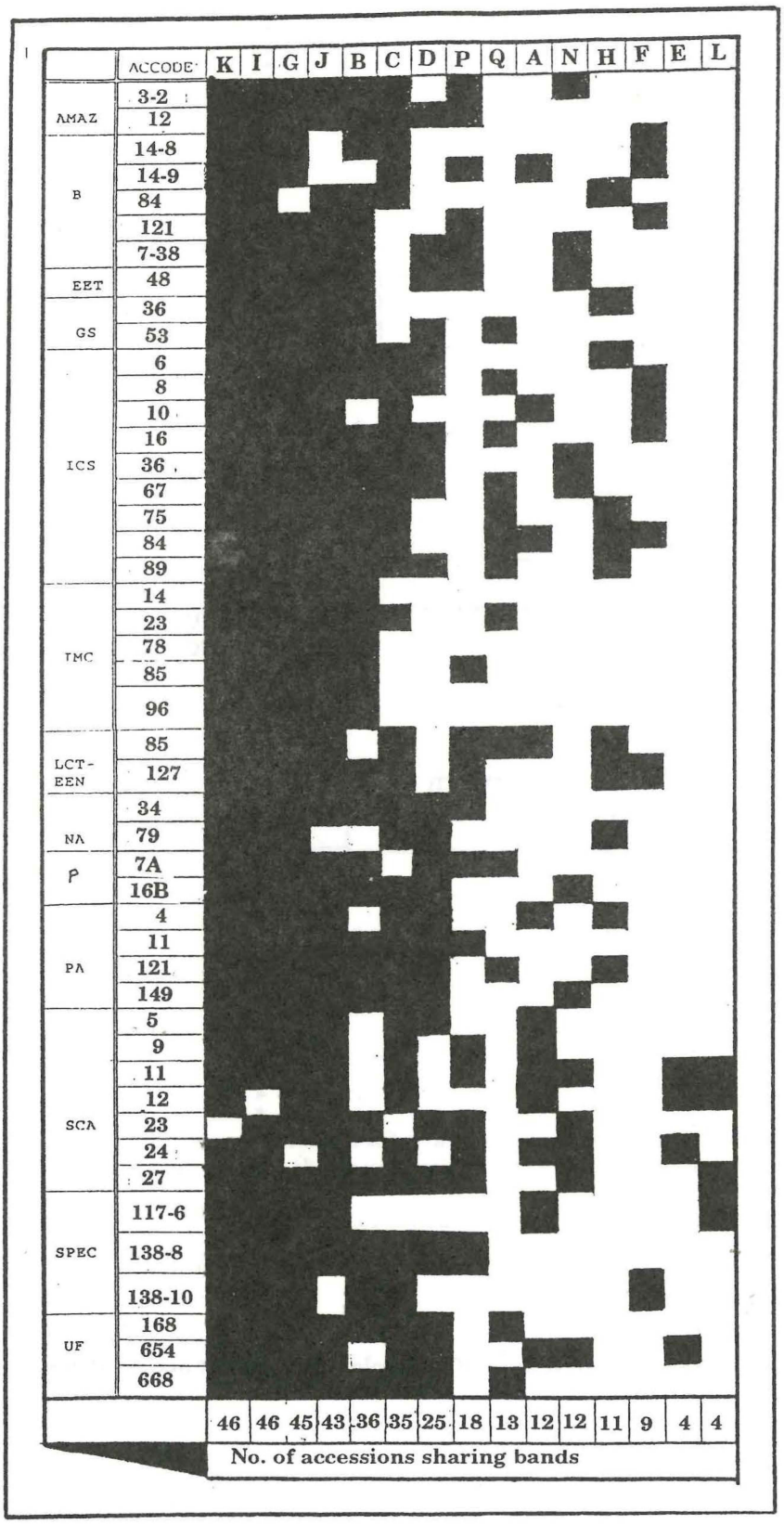


Figure 3. **Comparison of RAPD results obtained with extracts of different ages. A and C show the high degree of reproducibility obtained with relatively new extracts (less than 7 months old), B shows results obtained with older extracts. Note that the patterns observed for both or one sample of the following accessions changed when older extracts were used -SCA 3, 6, 12, 23, 24, UF 654, AGCA, BANC#3. The bands lost were at least one of the following, 1107bp, 984bp, 768.75bp.**



Code for RAPD loci

A - $A08_{1000}$
B - $AO8_{950}$
C - $A09_{1107}$
D - $M17_{620}$
E - $O15_{2027}$
F - $O15_{1968}$
G - $O15_{615}$
H - $O15_{553}$
I - $O15_{461}$
J - $O15_{369}$
K - $P14_{1168}$
L - $P14_{1107}$
N - $P14_{738}$
P - $P14_{523}$
Q - $P14_{430}$

primer fragment size (bp)

Figure 4. Diagram showing the molecular fingerprints of the 47 accessions analysed.

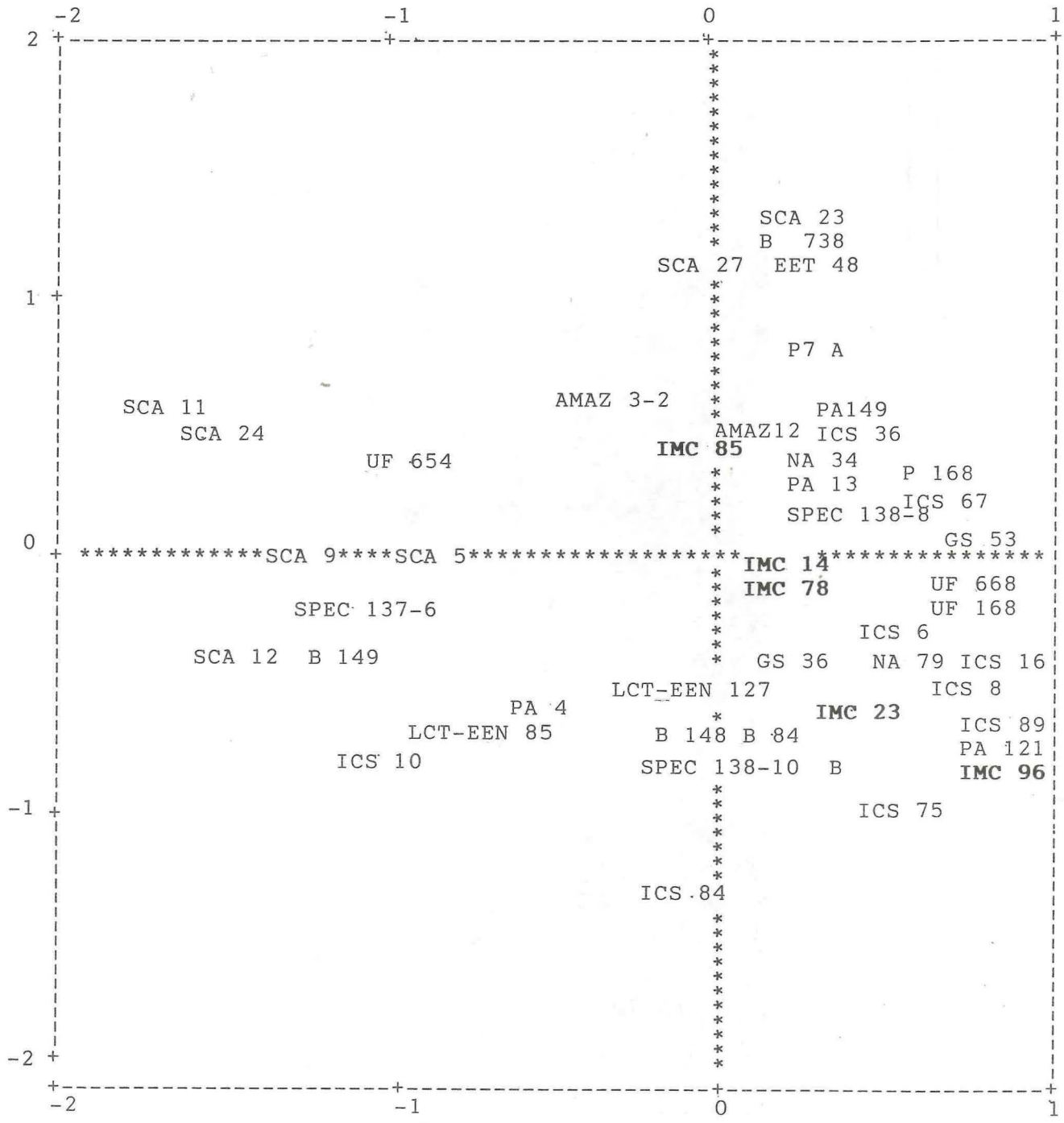


Figure 5. Plan defined by the two first axes of a simple Correspondence Analysis.

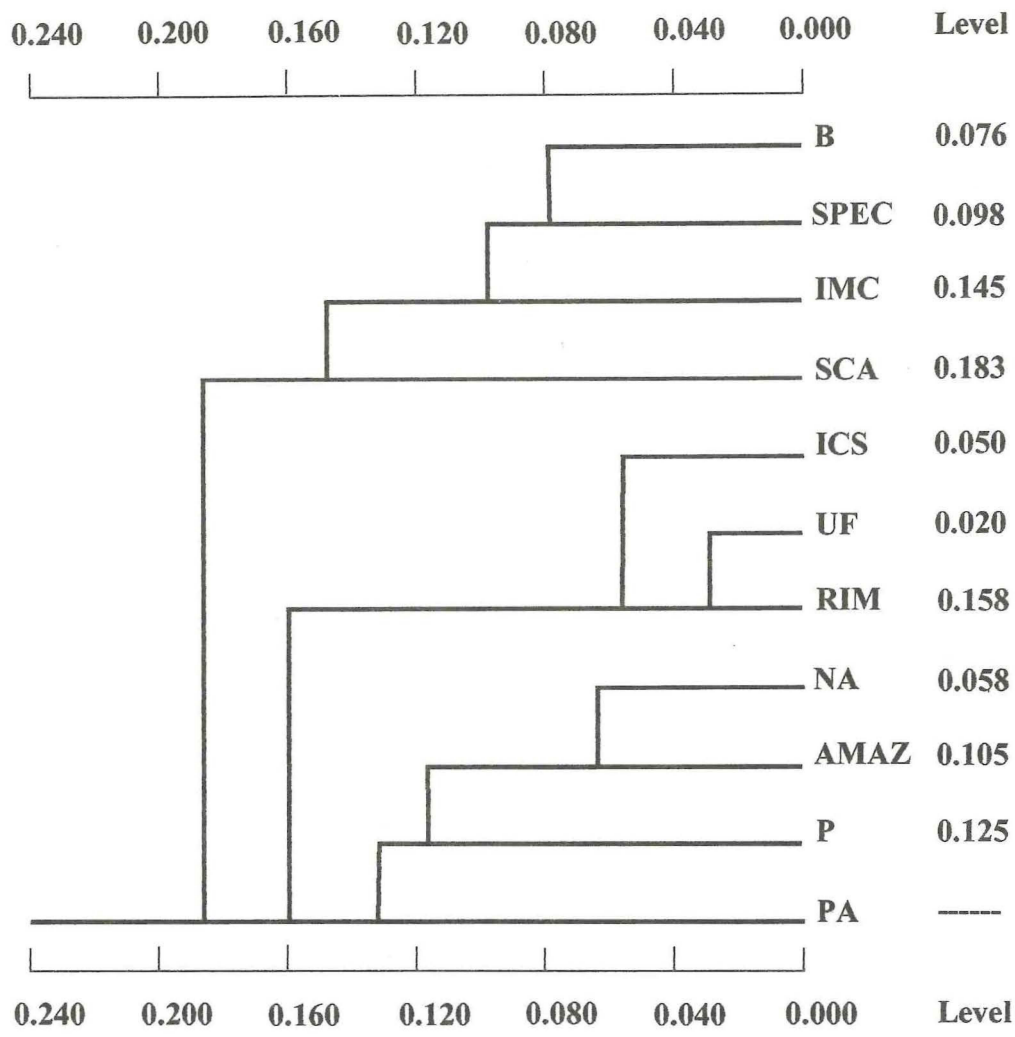


Figure 6. Dendrogram obtained after a clustering analysis based on Nei and Li coefficients calculated from the frequencies of the different markers in each population.

b) Age of the extracts

Extraction of DNA from cocoa using complicated methods such as that perfected by Laurent (1993) allows the obtention of very pure DNA samples, which may be conserved for at least one year (RISTERUCCI, pers. comm.). In our extraction procedure, the integrity and the purity of DNA were compromised, with the result that extracts older than seven months gave less readable and in some cases completely modified patterns. These included a complete loss of some higher or lower molecular weight bands, as shown in Figure 3.

DNA fingerprinting analysis

Analysis of the molecular profiles of 47 clones generated by the 18 polymorphic loci of the five primers revealed 35 different fingerprints, 27 clones were uniquely fingerprinted. The maximal number of clones sharing the same fingerprint was 4.

The molecular profiles of the 47 clones are represented in Figure 4 where rare patterns were sometimes due to the absence of a generally common band as in SCA 23. (This pattern has been observed for only one other accession, NA13, of the total clones analysed by this primer, P14). Less frequently, rare patterns arose due to the presence of a rare band as in Scavinas 11, 12 and 24. The unique molecular profiles were generally due to the combinations of banding patterns of all five primers rather than by the absence of common bands or the presence of rare bands.

Genetic Diversity

Figure 5 shows the plan obtained with the two first axes of the simple component analysis performed on the 47 clones. On this, most of the Trinitario (ICS, GS, UF) are fairly closely gathered, with the exception of three: ICS84, ICS10 and UF654. The SCAs are separated into two groups, while the IMCs are fairly gathered. Some of the Bs are gathered, whereas others scattered.

Figure 6 shows the dendrogram obtained after a hierarchical ascending cluster analysis, based on Nei distances, calculated from the frequencies of the different bands in each population. Once a maximal distance of 0.13 is decided between two populations for them to be considered to belong to the same cluster, four main groups are observed:

- one composed of PA, P, AMAZ and NA, the two latter populations being clustered fairly tightly.
- another composed of RIM, UF and ICS, all three being clustered fairly tightly.
- the third one represented by SCA alone.
- the last one composed of B, SPEC and IMC, these populations were clustered rather loosely.

Table 3 shows the values of mean pairwise Sokal and Milchener similarity index for each population. From these results, it can be seen that the populations showing the highest diversity are:B, SPEC, SCA and PA while the most homogenous ones are: UF, AMAZ and P.

Table: Mean pairwise similarity index (Sokal and Milchener) in each population.

Population	Mean Similarity Index	Standard Deviation
B	**0.57**	0.19
PA	**0.66**	0.2
P	0.8	0.17
AMAZ	0.84	0.09
RIM	0.7	0.32
ICS	0.75	0.12
IMC	0.77	0.16
NA	0.71	0.13
SCA	**0.64**	0.16
SPEC	**0.61**	0.16
UF	0.88	0.15

Discussion

The rather low percentage of fingerprinting (57%) obtained with 18 markers emphasises the need to use a larger number of primers and hence generation of a greater number of markers, in order to improve the accuracy of fingerprinting. This implies an increase in labour and cost of RAPD as a fingerprinting technique. However, this can be reduced if primers selected reveal polymorphism at the level of several primers, as P14 does. Unfortunately, such primers are rather uncommon, most of them allowing the obtention of one or two clear, reproducible and polymorphic bands. Nevertheless, in some case, clones can be distinguished by the very uncommon absence of a particular band, as it is the case for SCA 12, 23, 24 and B 84. This kind of marker is very helpful to detect very quickly the presence of "off-types".

When genetic diversity among the populations is assessed, the 11 populations studied here can be separated into four major clusters. The most homogenous major cluster is composed of the three Trinitario populations (ICS, RIM and UF). The populations collected in Peru by Pound are found in three of the four major clusters, NA, P and PA being in the same one, while SCA and IMC are in the two others. The two populations obtained from collection in Ecuador (AMAZ and B) are found in two of the major clusters. If heterosis can be obtained through crosses between populations from the same or from different geographical groups, then it will be more probably obtained through crosses between populations belonging to different major clusters. In this case, crosses between AMAZ, NA, P and PA should be avoided, as well as the ones between B, IMC and SPEC. SCA should be used in crosses with any other population, with slightly better chances for crosses with AMAZ, NA, P and PA.

However, one has to be cautious with these indications since differences are found between the populations, for the level of diversity found inside them. Indeed, a rather large level of diversity is found within some populations, such as B, PA, SCA and SPEC. This is not surprising in the case of SPEC, which is a collection from a wide area, with the objective of germplasm conservation, but more unexpected in the case of the three other populations since they were obtained from collections in small areas, with the objective of finding sources of resistance to witches' broom disease. In the latter case, all budwood had been collected from a large number of trees of the same area,once these trees had been found free of witches' broom symptoms. This strategy should have resulted in a rather high degree of consanguinity among the trees collected in a same population, resulting in a rather limited level of diversity, but this is not what appears from these results. As expected, a lower level of diversity is found within AMAZ, P and UF populations, these populations being obtained from collections or selections in small areas. The estimation of genetic diversity within each population can be helpful in the establishment of core collections, by helping to choose the minimal sample size to reduce the loss of diversity inside each population. From this study, it appears that populations such as B, NA, SCA and SPEC would require a larger sample size than populations such as AMAZ, NA, P and PA. In addition, genetic diversity analyses of the clones are helpful to determine which particular clones should be selected from the populations, in order to maintain the maximum diversity after the sampling. For example, the results obtained from the simple correspondence analysis show that in case of sampling the SCA population, SCA 23 or SCA 27 should be included in this sample, as well as SCA 5 or SCA 9 or SCA 12 and also SCA 11 or SCA 24. A subsample of the clones of the B population represented in our study should include B14-9 as well as B7-38 as well as B-14-8 or B 84 or B121. However such a sampling should be performed also based on agronomical traits, this information being the most important for the breeders.

However, the results obtained here are far from definitive since the number of clones per population and the number of markers used are too low. Using a bootstrap method, Christopher *et al.* (1992) found a minimal sample size of 10 clones per population from results obtained by the use of isozyme electrophoresis, in order to ensure reliable estimations of diversity. The same type of analysis should also be performed in order to evaluate the minimal number of markers to use for the same objective.

References

Christopher, Y., Hosein, F., Johnson, E., and Warren, J. (1992) Genetic diversity and the conservation of germplasm in *Theobroma cacao* L. Cocoa Research Unit Report for 1992, 25 - 37.

Johnson, E., Russel, J.R. Hosein, Powell, W., and Waugh, R. (1992) A laboratory manual for RAPD analyses in cocoa. pp.17.

Laurent, V. (1993) Etude de la diversité génétique du cacaoyer (*Theobroma cacao* L.) basée sur le polymorphisme de la longueur des fragments de restriction (RFLP). Thèse de l'Université, Paris XI