Genetic Resources and Crop Evolution (2005) 52: 1111–1120
DOI 10.1007/s10722-004-6110-4

Assessing the genetic diversity in the International Cocoa Genebank, Trinidad (ICG,T) using isozyme electrophoresis and RAPD

O. Sounigo[1,*], R. Umaharan[2], Y. Christopher[2], A. Sankar[2] and S. Ramdahin[2]
*[1]CIRAD, Avenue Agropolis, 34398 Montpellier, France; [2]Cocoa Research Unit, University of the West Indies, Saint Augustine, Trinidad and Tobago; *Author for correspondence (e-mail: olivier.sounigo@cirad.fr)*

Received 15 May 2003; accepted in revised form 6 February 2004

Key words: Genetic diversity, Isozyme Electrophoresis, RAPD, *Theobroma cacao* L.

Abstract

Random amplified polymorphic DNA (RAPD) and isozyme electrophoresis (IE) techniques were used to estimate the level of genetic diversity in a sample of cacao germplasm existing at the International Cocoa Genebank, Trinidad. Twenty-six cocoa populations represented by 459 cocoa genotypes were analysed using IE and 22 populations represented by 353 cocoa genotypes were analysed using RAPD. Despite few differences in the classification of the populations, both techniques revealed three major groups: the indigenous trees, the cultivated Trinitario and the cultivated trees from Ecuador. Two-thirds of the partitioned diversity were found within populations and one-third between the populations, with both techniques.

Introduction

Cocoa (*Theobroma cacao* L.) is a diploid (Cope 1984) preferentially allogamous tropical tree species, belonging to the family Malvaceae (Alverson et al. 1999). Its geographical origin is considered to be South America, extending from Amazonia to the Guianas (Cheesman 1944; Motamayor et al. 2002). At present, it is cultivated mostly in the equatorial tropics, with West Africa being the most important producing area.

Most national breeding programmes around the world use locally adapted varieties augmented by genotypes from the Imperial College of Tropical Agriculture (ICTA), Trinidad (Lockwood 1985). In the early 1960s, the ICTA was replaced by the Cocoa Research Unit (CRU). The CRU is, at present, the custodian of the only international cocoa collection in the world, International Cocoa Genebank, Trinidad (ICG,T), which continues the task of distributing germplasm to cocoa breeders worldwide.

The field germplasm collection at ICG,T has been morphologically (Bekele et al. 1994; Bekele and Bekele 1998) and agronomically (Iwaro et al. 2002; Latchman et al. 2000; Umaharan et al. 2001) characterised, and the data is available globally, through the International Cocoa Germplasm Database (Wadsworth et al. 1997).

During the past decade, biochemical (isozyme electrophoresis (IE)) and molecular random amplified polymorphic DNA (RAPD) markers have been used at CRU to study the genetic diversity of the ICG,T (Russell et al. 1993; Warren 1994; Sounigo et al. 1997). The objective of the present study is to assess the genetic diversity of cocoa populations held at ICG,T and its genetic structure, towards establishing a core collection of genotypes that would capture the variability and at the same time contain genotypes with useful traits, such as disease resistance and desirable pod and bean characteristics (Sounigo et al. 2000).

Materials and methods

Materials

The ICG,T (33 ha) is located in the northern part of the island of Trinidad (Republic of Trinidad and Tobago), and contains 90 cocoa populations, represented by over 2000 cocoa genotypes. Each genotype has been vegetatively propagated and is represented by four to sixteen trees per field plot.

The populations were collected from Central and South America as well as from the Caribbean, both from wild and cultivated areas (either for the purpose of capturing genetic diversity or for other useful traits, such as witches' broom resistance, yield or bean characteristics), some of them as early as the 1930s and others just few years ago. The populations are named after the places where they were collected (e.g. country, river, hacienda) or selected (research institutes), or sometimes after the collecting person (e.g. Pound) or the collecting project (e.g. LCT EEN). In the ICG,T, each population contains a number of genotypes ranging from 10 to several hundreds. The degree of relatedness between the genotypes in a population varies with the populations and may include full-sibs, half-sibs or trees without any close relationship.

A total of 459 genotypes, representing 26 populations, were analysed using IE and 353 genotypes, representing 22 populations, were analysed using RAPD. Details of the populations analysed are given in Table 1.

Methods

Isozyme Electrophoresis
Crude enzyme extracts were obtained from flush leaves collected from the 459 trees and subjected to starch gel electrophoresis. The extracts were used in the screening of five isoenzymatic systems, alcohol dehydrogenase (ADH), acid phosphatase (ACP), isocitrate dehydrogenase (IDH), malate dehydrogenase (MDH) and glucose phosphate isomerase (GPI). Gel interpretation was undertaken according to the genetic backgrounds for the five enzyme system as determined by Lanaud (1986). Four out of the five systems are coded by only one locus while MDH is coded by two loci.

Random Amplified Polymorphic DNA
Adult leaves were collected from 353 trees. DNA was extracted according to Russell et al. (1993) and quantified using a TKO-100 DNA fluorometer (Hoefer Scientific). Each DNA sample was amplified separately using a Thermal cycler (Perkin Elmer), using the following protocol: 0.45 ng of template DNA were mixed in a final 25 μL reaction mix containing 15 ng of Operon decameric primer, 0.2 mM of each dNTP, 1 unit of Taq polymerase and 2.5 μL of the Taq buffer (provided by the furnisher). According to the primer used, two versions of Taq polymerase were used: BIOTAQ (Bioline) and Klen Taq I (AB Peptides). Forty cycles of amplifications were performed (5 s at 94 °C, 30 s at 36 °C and 72 °C for 1 min) followed by a final extension step (5 min at 72 °C). The amplification products were separated on a 1.5% agarose gel and photo-documented. Fourteen primers were selected for their ability to produce clear and reproducible polymorphic bands, and 30 of these bands were selected as markers for this study.

Data analyses

Population divergence. The data obtained were used to calculate the mean frequencies of each allele or marker for each population.

- *For IE data*: Nei (1972) distances (D_{AB}) were calculated, according to the following formula:

$$\overline{D_{AB}} = -\ln \overline{I_{AB}}$$

where $\overline{I_{AB}}$ is the arithmetic mean of the Nei similarity indexes between populations A and B calculated for each locus as follows:

$$I_{AB} = \frac{\sum_k (p_{kA} * p_{kB})}{\sum_k (p_{kA}{}^2 * p_{kB}{}^2)^{1/2}},$$

where p_k is the frequency of the allele k in the population

- *For the RAPD data*: Rogers distance modified by Wright (1978) was calculated as follows, at the level of each marker:

$$D_{AB} = (1/2n \sum_k (p_{kA} - p_{kB})^2)^{1/2},$$

where p_k is the frequency of the allele k (0 = absence of the band and 1 = presence of the band) in the population A or B. The distance

Table 1. Details on the studied cocoa populations (populations with names in italics were analysed using IE, those with names underlined were analysed using RAPD and those with names in plain letters were analysed using both techniques).

Accession	Complete name	Country of origin	Purpose of collection	Type of material	References
AM	Hacienda Amalia	Ecuador	Resistance to witches' broom disease	Cultivated	Pound (1943)
AMAZ	Rio Amazonas	Ecuador	Resistance to witches' broom	Indigenous	Chalmers (1973)
B	Hacienda Balao	Ecuador	Resistance to witches' broom disease	Cultivated	Pound (1943)
CAM	Camopi river	French Guiana	Diversity	Cultivated	Lachenaud and Sallée (1993)
CL	Hacienda Clementina	Ecuador	Resistance to witches'broom disease	Cultivated	Pound (1943)
DOM	Dominica	Dominica	Diversity	Cultivated	
ELP	Eulepoussing	French Guiana	Diversity	Indigenous	Lachenaud et al. (1997)
GS	Grenada Selection	Grenada	Yield, fruit and bean characteristics	Cultivated	
ICS	Imperial College Selection	Trinidad	Yield, fruit and bean characteristics	Cultivated	Pound (1932)
IMC	Iquitos Mixed Calabacillo	Peru	Resistance to witches' broom disease	Indigenous	Pound (1938)
JA	Hacienda Javilla	Ecuador	Resistance to witches' broom disease	Cultivated	Pound (1943)
LCT EEN	London Cocoa trade Amazon Project	Ecuador	Diversity	Indigenous	Allen and Lass (1983)
LP	Hacienda La Paz	Ecuador	Resistance to witches' broom disease	Cultivated	Pound (1943)
LX	Large Vuelta X	Ecuador	Resistance to witches' broom disease	Cultivated	Pound (1943)
MAR	Martinique	Martinique	Diversity	Cultivated	
MO	Morona river	Peru	Resistance to witches' broom disease	Indigenous	Pound (1938)
MOQ	Hacienda Moquique	Ecuador	Resistance to witches' broom disease	Cultivated	Pound (1943)
NA	Nanay	Peru	Resistance to witches' broom disease	Indigenous	Pound (1938)
PA	Parinari	Peru	Resistance to witches' broom disease	Indigenous	Pound (1938)
P	Pound	Peru	Resistance to witches' broom disease	Indigenous	Pound (1938)
RIM	Hacienda la Rioja	Mexico	Yield, fruit and bean characteristics	Cultivated	Hunter (1961)
SC	Seleccion Colombiana	Colombia	Yield, fruit and bean characteristics	Cultivated	Hunter (1961)
SCA	Scavina	Peru	Resistance to witches' broom disease	Indigenous	Pound (1938)
SJ	San Juan	Ecuador	Resistance to witches' broom disease	Cultivated	Pound (1943)
SL	Santa Lucia	Ecuador	Resistance to witches' broom disease	Cultivated	Pound (1943)
TRD	Trinidad	Trinidad	Diversity	Cultivated	
UF	United Fruit	Costa Rica	Yield, fruit and bean characteristics	Cultivated	

over all the RAPD markers was obtained by the arithmetic mean of the D_{AB} values.

Phylogenetic analysis. For both IE and RAPD data, cluster analyses were performed using neighbour joining method (Swofford et al. 1996) using the Darwin software, developed by centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD).

Structure of the genetic diversity (Diversity within the populations). Several parameters were calculated using POPGENE 32 software (Yeh et al. 1999). For both IE and RAPD data, the level of diversity was quantified using the Shannon-Weaver diversity index (Magurran 1988), which combines, at every locus, the allelic richness and the level of heterozygosity:

$$H_1 = -\sum_{kl} \ln p(kl),$$

where $p(kl)$ = frequency of the allele k of the locus l (IE) or the frequency of the presence or the absence of the marker l (RAPD); For IE, k corresponds to the alleles present at the locus, while for RAPD k corresponds to the absence and the presence of the marker. The mean value of these indexes is calculated for all the loci or markers. IE data were also used to calculate some other parameters:

- *Proportion of observed heterozygosity*: The proportion of heterozygous genotypes was calculated for each locus and the mean value was calculated for all loci.
- *Mean number of alleles per locus*: Includes all the alleles, whatever their frequencies.
- *Number of polymorphic loci*: Includes all the loci with more than one allele, whatever the frequencies of the alleles.

Partitioning of the genetic diversity

The proportion of the genetic diversity observed between the populations (G_{st}) was calculated according to the following formula:

$$G_{st} = H_{st}/H_T \quad \text{with} \quad H_{st} = H_T - H_m,$$

where H_{st} is the amount of diversity found between populations, H_m is the amount of diversity observed within the populations (mean value of the Shannon and Weaver indexes calculated on each population), H_T is the total genetic diversity (Shannon and Weaver index calculated on all the cocoa genotypes analysed).

Results and discussion

Phylogenetic analysis

Figures 1 and 2 show the dendrograms obtained from IE and RAPD data, respectively. The dendrograms obtained using the IE data (Figure 1) and RAPD data (Figure 2) clearly separate the populations into three major clusters:

- One composed mainly of the indigenous trees (in bold), from Peru, Ecuador and French Guiana;
- One mainly composed of all the populations of cultivated Trinitario (in plain letters);
- One exclusively composed of trees cultivated in Ecuador (in italics).

Originally, the term Trinitario was used to describe cocoa trees from Trinidad, derived from hybridisation between Criollo and Forastero types. More generally the term "Trinitario" designates cultivated cocoa trees derived from recent hybridisation events between Criollo and Forastero, sometimes followed by one or several crossing events with Criollo, Forastero or other Trinitario trees. The term Criollo refers to cocoa anciently cultivated in Central America and in the north of South America, and most probably originating from Venezuela and Colombia (Motamayor et al. 2002). These trees produce white beans, used to

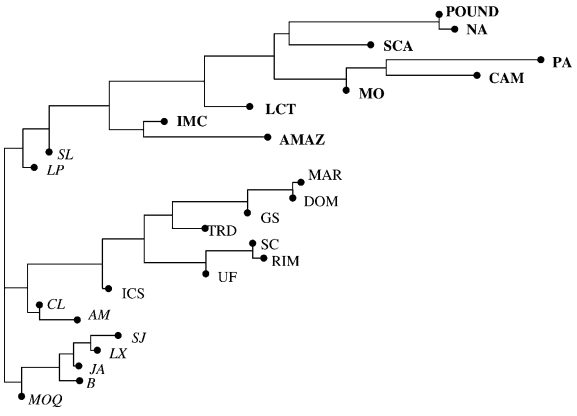

Figure 1. Dendrogram obtained from a Neighbour Joining analysis performed on IE data obtained on 26 *Theobioma cacao* populations (bold – indigenous populations; plain – cultivated Trinitario, italics – cultivated from Ecuador).

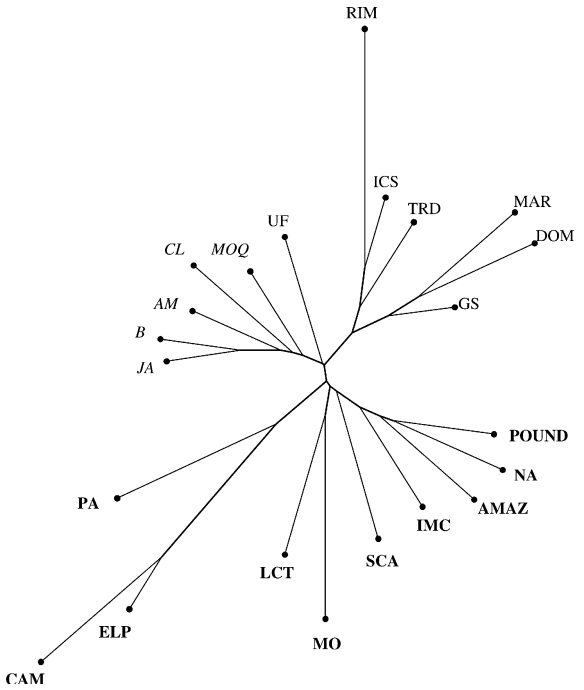

Figure 2. Dendrogram obtained from a Neighbour Joining analysis performed on RAPD data obtained on 22 cocoa populations (bold – indigenous populations; plain – cultivated Trinitario; italics – cultivated from Ecuador).

prepare fine flavour chocolate. The term Forastero is assigned to cocoa trees originating from the Amazonian, Orinocco and the Guyanas, which produce purple beans used in bulk chocolate. Bekele et al. (2002), using morphological descriptors, also indicated that the cocoa populations fell into the same three groups, *viz*. indigenous, Trinitario, cultivated trees from Ecuador, as with molecular data in this study.

Both RAPD and IE data showed further separation of the Trinitario cluster into two sub-clusters, one from the Caribbean (GS, DOM and MAR) region and the other from Central and South America (RIM, UF and SC), but the separation was more distinct with RAPD than with IE.

The populations from Peru (POUND, NA, AMAZ and IMC) fell into a single cluster (with RAPD data), which was expected, because of their geographical proximity (Iquitos region). Furthermore, both RAPD and IE data showed that the Peruvian PA population was more closely related to the two populations from French Guiana (CAM and ELP) than all others. However, the dendrogramme obtained from RAPD data shows a clear separation between these two populations from French Guiana and all the other populations of indigenous trees analysed. In addition, a RAPD marker was found to be specific to the two populations from French Guiana but absent in all the other populations, including the PA population. Lerceteau et al. (1997), using different RAPD markers, also showed that the genotypes from the banks of Camopi river, in French Guiana, were distinct from the others.

The cultivated populations of Ecuador, not surprisingly, clustered distinct from others, because of their complex origin. The cultivated populations of Ecuador originated from trees called "Refractario", which were derived from hybridisation events between cocoa from three main genetic origins and selected for Witches' Broom disease resistance (Pound 1943; Bartley 2001). The genetic origins being (a) one or several Trinitario genotypes from Trinidad, Venezuela or Colombia, introduced into Ecuador at the end of the 19th century; (b) cocoa traditionally cultivated in the western part of Ecuador, referred to as the Nacional and (c) Forastero trees from the Amazonian part of Ecuador or Peru, probably from the banks of the Napo river.

Levels of genetic diversity within the populations

Tables 2 and 3 show the levels of diversity observed using IE and RAPD, respectively, while Table 4 shows the correlation between the different parameters and the sample size for each population.

Significant positive correlation was observed between the number of genotypes per population and population genetic diversity determined as the mean numbers of IE alleles ($R = 0.57$). Hence, between population comparisons for the above parameter were restricted to populations with similar numbers of genotypes. On the other hand, no correlation was observed between the number of genotypes per population and any of the other parameters used to evaluate the genetic diversity, calculated from RAPD or IE data.

The mean number of alleles per locus were very low for the CAM population (1.5), when compared to most of the other populations which had lower or similar sample sizes. On the other hand, LCT EEN showed the highest value (2.67) for this para-

Table 2. Level of genetic diversity observed on 26 cocoa populations using IE.

	Geographic origin	Number of analyzed genotypes	Number of mother trees for the origin of the analyzed genotypes	Shannon index		Observed level of heterozygosity		Mean number of alleles per locus		Number of polymorphic loci
				Mean	SD	Mean	SD	Mean	SD	
AMAZ	Peru	6	≤6	0.54	0.35	0.28	0.25	2	0.63	5
IMC	Peru	34	2	0.43	0.33	0.33	0.28	2.17	0.98	5
MO	Peru	9	1	0.65	0.36	0.26	0.23	2.33	0.82	5
NA	Peru	56	≤14	0.5	0.19	0.2	0.2	2.67	0.82	6
POUND	Peru	8	8	0.47	0.25	0.27	0.33	2	0.63	5
PA	Peru	41		0.41	0.24	0.21	0.14	2.5	1.05	5
SCA	Peru	8	1	0.55	0.33	0.23	0.2	2	0.63	5
CAM	French Guiana	16	16	0.21	0.39	0.11	0.2	1.5	0.84	2
LCT EEN	Ecuador	16	16	0.7	0.33	0.29	0.2	2.67	0.82	6
DOM	Dominica	12		0.24	0.22	0.14	0.13	1.8	0.75	4
GS	Grenada	14		0.43	0.28	0.31	0.21	2	0.63	5
MAR	Martinique	11		0.1	0.16	0.06	0.09	1.33	0.52	2
ICS	Trinidad	34	34	0.66	0.31	0.37	0.22	2.67	0.82	6
TRD	Trinidad	27	27	0.61	0.29	0.26	0.17	2.5	0.55	6
RIM	Mexico	12		0.51	0.3	0.64	0.5	1.83	0.41	5
SC	Ecuador	8		0.6	0.33	0.65	0.46	2.17	0.75	5
UF	Costa Rica	8		0.6	0.34	0.5	0.33	2.17	0.75	5
AM	Ecuador	13	1	0.6	0.41	0.4	0.27	2.17	0.75	5
B	Ecuador	18	11	0.59	0.4	0.34	0.27	2.33	1.03	5
CL	Ecuador	29	5	0.62	0.41	0.35	0.25	2.5	1.04	5
JA	Ecuador	22	5	0.54	0.44	0.45	0.35	2.17	1.17	4
LP	Ecuador	15	4	0.53	0.47	0.39	0.35	2.17	1.17	4
LX	Ecuador	12		0.52	0.43	0.42	0.34	2	1.09	4
MOQ	Ecuador	10	4	0.62	0.41	0.4	0.26	2.33	1.03	5
SJ	Ecuador	7		0.43	0.33	0.45	0.37	1.67	0.52	4
SL	Ecuador	13	2	0.68	0.42	0.29	0.18	2.33	1.03	5

meter, despite a relatively low sample size (16), and can be considered more diverse than other populations with similar or higher sample sizes. Large variation was detected for the level of observed heterozygosity and for Shannon indices, the values being very low for the populations MAR, DOM and CAM (between 0.06 and 0.14 and between 0.1 and 0.24, respectively) but higher for populations such as RIM and SC (observed heterozygosity of 0.65) and for ICS, SL and LCT EEN (Shannon index 0.6–0.7).

The Shannon indices obtained using RAPD data range between 0.09 (RIM) and 0.42 (LCT EEN) but were less contrasting than with the IE data. A highly significant positive correlation was observed between the Shannon indexes observed with IE and RAPD ($R = 0.61$), despite a great discrepancy observed with the RIM population. The discrepancy is because the high value observed with IE only resulted from the high level of observed heterozygosity, since all the genotypes analysed for this population showed the same zymogram. The diversity in the RIM population was not observable with RAPD, because of the dominant nature of these markers.

Surprisingly, both RAPD and IE data failed to show any significant correlation between the level of diversity found in a population and the number of mother trees from which this population is issued (Table 4). Indeed, some populations supposedly derived from a single mother tree, such as MO and SCA show higher genetic diversity values than populations issued from larger numbers of mother trees, such as CAM or B.

The large differences in the level of genetic diversity within populations may reflect partly the

Table 3. Level of genetic diversity observed on 22 cocoa populations using RAPD.

Population	Geographic origin	Sample size	Number of mother trees for the origin of the analyzed genotypes	Shannon index (H)	
				Mean	SD
AMAZ	Peru	10	≤10	0.35	0.27
IMC	Peru	28	2	0.36	0.26
LCT EEN	Ecuador	24	24	0.42	0.26
MO	Peru	14	1	0.36	0.26
NA	Peru	35	≤14	0.31	0.24
POUND	Peru	21	21	0.36	0.22
PA	Peru	37		0.32	0.26
SCA	Peru	15	1	0.36	0.26
CAM	French Guiana	22	22	0.21	0.27
ELP	French Guiana	20	9	0.31	0.26
DOM	Dominica	18		0.23	0.25
GS	Grenada	18		0.26	0.29
MAR	Martinique	13		0.14	0.22
ICS	Trinidad	31	31	0.31	0.24
TRD	Trinidad	14		0.38	0.27
RIM	Mexico	15		0.09	0.17
UF	Costa Rica	15		0.31	0.28
AM	Ecuador	13	2	0.31	0.29
B	Ecuador	18	8	0.24	0.25
CL	Ecuador	16	6	0.38	0.23
JA	Ecuador	22	7	0.24	0.23
MOQ	Ecuador	15	6	0.36	0.28

different strategies adopted for collecting cocoa. For instance, the high genetic diversity observed for LCT EEN can be explained by the very large area (around 100,000 km^2), which was systematically sampled (Allen and Lass 1983). With the CAM and ELP populations, the purpose was also to sample the diversity, but in relatively small areas (around 50 km^2 for each population) (Lachenaud and Sallée 1993; Lachenaud et al. 1997). In the case of the populations from Peru (SCA, MO, NA, PA and POUND), the aim was to collect sources of resistance to Witches' Broom disease (Pound 1938). Thus, even though a large area was explored, only very few mother trees were collected. The Trinitario populations, especially the ICS's, showed a high level of diversity. This population contains trees selected for yield, pod and bean characteristics, from a large number of cocoa farms in Trinidad (Pound 1932). The TRD's were collected from 10 different estates in Trinidad, with a sample size of 10 trees per estate (Warren, pers. commun.).

The low level of diversity in the CAM population was also observed by Lanaud (1986), using IE despite a high level of morphological diversity observed by Lachenaud et al. (1999, 2000) and

Table 4. Correlation between the sample size and the genetic parameters calculated.

	N	Pearson	Spearman	*N*	Pearson	Spearman
IE						
Shannon index	26	0.02	0.01	15	0.05	0.06
Observed level of heterozygosity	26	−0.26	−0.12	15	−0.16	−0.08
Mean number of alleles per locus	26	0.57**	0.59**	15	0.35	0.4
Number of polymorphic loci	26	0.35	0.31	15	0.27	0.32
RAPD						
Shannon index	22	0.12	−0.07	13	−0.09	−0.23

N – number of comparisons; Pearson – Pearson correlation coefficient (calculated on values); Spearman – Spearman correlation coefficient (calculated on ranks); The only significant correlation values are the ones followed by ** indicating significance at 1%.

Lachenaud and Oliver (2001). The contrast between high morphological and low molecular diversity was also observed in cocoa belonging to the Criollo group by Motamayor (2001). The level of genetic diversity observed for the Trinitario populations varies considerably based on their geographical origin. The populations from Dominica (DOM) and from Martinique (MAR) show very low levels of genetic diversity and observed heterozygosity, in contrast to the high diversity seen in populations from Central and South America (SC, UF and RIM). The Trinitario from Trinidad (ICS and TRD) and Grenada (GS) show intermediate values for these two parameters. Our data differ from those of Laurent et al. (1995), who, using RFLP, found a similar level of observed heterozygosity for Trinitario from Trinidad and Grenada (ICS and GS) and those from Central America (UF and RIM). The cultivated cocoa populations of Ecuador show a level of observed heterozygosity intermediate between those for Trinitario and Upper Amazonian Forastero.

Level of genetic differentiation

Table 5 shows the partitioning of the genetic diversity within and among populations, using IE and RAPD data.

The data obtained using both IE and RAPD show G_{st} values of 0.28 (IE) and 0.35 (RAPD), which were lower than that in a previous study (Russell et al. 1993) using RAPD markers on three cocoa populations (0.42), but higher than the mean G_{st} value of 0.08 observed by Hamrick et al. (1992), who studied the partitioning of the genetic diversity of 195 perennial woody species using IE.

Table 5. Partitioning of the genetic diversity observed on 26 cocoa populations using IE and on 23 cocoa populations using RAPD.

	IE	RAPD
H_T	0.71	0.46
H_m	0.51	0.30
H_{st}	0.20	0.16
G_{st}	0.28	0.35

H_T – total diversity; H_m – diversity within populations; H_{st} – diversity among populations; G_{st} – proportion of the diversity found between populations.

Implications for further germplasm collection and for sub-sampling the ICG,T

Collecting material

Our data supports a genetic diversity structured according to geographical origins, indicating the necessity of collecting from diverse areas to capture genetic diversity. Collection efforts are being continued at the ICG,T, with recent additions from French Guiana (Lachenaud et al. 1997) and from Belize (Mooleedhar et al. 1995). Nevertheless, the high level of genetic diversity observed in some populations such as SCA and MO, derived from single mother trees, suggests a strategy of collecting a low number of trees from a large number of populations to capture allelic diversity, at least for alleles represented at a high or fair frequency as suggested also by Warren (1994). The specificity of certain traits to particular populations agrees with the necessity of collecting from a larger number of populations. For example, the "arriba" flavour is attributed to Nacional population from the western part of Ecuador (Fowler 1952). However, in cocoa, some favourable traits seem to result from the expression of alleles present at rather low frequencies. For example, Iwaro et al. (2001) showed that only 5% of 1198 clones tested showed a complete absence of symptoms after pod inoculation using spores of *Phytophthora palmivora* Butl. Phillips-Mora (1999) tested 279 clones for resistance to *Moniliophthora roreri* using an artificial pod inoculation method and found a high level of resistance in only 2.6% of them. These results indicate that if one wants to enhance chances of capturing allelic diversity for favourable traits through a random sampling method, the methodology of collecting should be chosen in such a way as to maximise the chances of capturing alleles present at low frequencies. For this purpose, Dias (2002) proposed a strategy of collecting 400 seeds from 50–80 mother trees, with 5–8 seeds per tree, in each population. This corresponds to a 0.3 ha field surface per collected population, if planting is done at an average density, and if each genotype is represented by one single tree. If such a strategy was adopted, the ICG,T, with its 100 populations, would cover a surface of around 30 ha, which is close to its actual area. In order to prevent possible losses of genotypes, each tree could also be represented in a budwood garden, at high density, to

reduce management costs. A small number of genotypes from each population can be cloned and established in core collections. The representation of each clone by several trees in core collections would allow a better assessment of useful traits.

Sub-sampling the ICG,T

The relatively high G_{st} value suggests that representing a large number of populations in a subsample of the ICG,T should ensure the capturing of a large part of its genetic diversity. In such a sampling strategy, the number of clones selected from a population would be determined by the level of genetic diversity of that population. For most of the populations, the number of representative clones needed is easily decided since a significant positive correlation between the Shannon indices obtained with data from IE and RAPD was observed. For example, populations such as LCT EEN, MO, MOQ, CL, TRD, AMAZ and SCA should be represented by larger numbers of clones while populations such as CAM, DOM and MAR should be represented by lower numbers of clones. Intermediate numbers of clones should represent populations such as PA, POUND, NA, GS and IMC. On the other hand, contradictory values were observed for some populations. The strongest contradiction was observed in the case of RIM, but in this case, the similarity of the IE profiles of all the clones from this population suggests representing it by a single clone. Other strong contradictory values are observed for populations such as B, AM, JA, UF and ICS. In these cases, the other parameters such as the mean number of alleles per locus and the number of polymorphic loci should be considered.

Conclusion

The use of IE and RAPD allowed the determination of the structure and level of genetic diversity existing in the ICG,T. This knowledge can now be used to develop a smaller core collection to represent the genetic diversity, while including all useful traits. Further studies using simple sequence repeat (SSR) (Lanaud et al. 1999), will greatly assist in this process, thanks to the codominance of these markers and to the large numbers of alleles they allow to observe.

Acknowledgements

The authors thank the Biscuit, Cake, Chocolate and Confectionary Alliance (BCCCA) and the French Ministry of Foreign Affairs (MAE) for the funds they provided.

References

Allen J.B. and Lass R.A. 1983. London Cocoa Trade Amazon Project. Final Report Phase 1. Cocoa Growers' Bulletin 34.

Alverson W.S., Whitlock B.A., Nyffeler R., Bayer C. and Baum D.A. 1999. Phylogeny of the core Malvales: evidence from *ndhF* sequence data. Amer. J. Bot. 86: 1474–1486.

Bartley B.G.D. 2001. Refractario - An explanation of the meaning of the term and its relationship to the introductions from Ecuador in 1937. Ingenic Newslett. 6: 10–15.

Bekele F.L., Kennedy A.J., Mc David C., Lauckner B. and Bekele I. 1994. Numerical taxonomic studies on cocoa (*Theobroma cacao* L.) in Trinidad. Euphytica 75: 231–240.

Bekele F. and Bekele I. 1998. A sampling of the phenetic diversity of cacao in the International Cocoa Genebank of Trinidad. Crop Science 36: 57–64.

Bekele F., Bidaisee G. and Mollineau W. 2002. Morphological variation in a sample of germplasm from the International Cocoa Genebank, Trinidad. Annual Report 2001, Cocoa Research Unit, The University of the West Indies, Trinidad and Tobago, pp. 24–33.

Chalmers W.S. 1973. Report on a visit to Ecuador, Peru and Brasil, March 1st–April 15th 1973. Unpublished Report, Cocoa Research Unit, University of the West Indies, Trinidad, 31 pp.

Cheesman E.E. 1944. Notes on the nomenclature, classification possible and relations of cocoa populations. Trop. Agricult. 21: 144–159.

Cope F.W. 1984. Cacao *Theobroma cacao* (Sterculiaceae). In: Longman (ed.), Evolution of Crop Plants, London, pp. 285–289.

Dias L.A.S. and Almeida C.M.V.C. 2002. Cacao germplasm: how and how much to collect? Ingenic Newslett. 7: 8–11.

Fowler R. 1952. Caracteristicas del cacao nacional. Turrialba (Costa Rica) 2(4): 161–165.

Hamrick J.L., Godt M.J.W. and Sherman-Broyles S.L. 1992. Factors influencing levels of genetic diversity in woody plant species. New Forests 6: 95–124.

Hunter J.R. 1961. The status of cacao variety improvement work in the Western hemisphere. Cacao (Costa Rica) 4: 1–16.

Iwaro A.D., Singh V., Barath S. and Jugmohan N. 2001. Germplasm evaluation at the International Cocoa Genebank, Trinidad for resistance to *Phytophthora* pod rot. CRU Annual Report 2000: 34–40.

Iwaro A.D., Bekele F.L. and Butler D.R. 2002. Utilisation of the cacao genetic resources at the International Cocoa Genebank, Trinidad. Ingenic Newslett. 7: 22–24.

Lachenaud P. and Sallée B. 1993. Les cacaoyers spontanés de Guyane. Localisation, écologie et morphologie. Café Cacao Thé 37(2): 101–114.

Lachenaud P., Mooleedhar V. and Couturier C. 1997. Les cacaoyers spontanés de Guyane. Nouvelles prospections. Plantations, recherche, développement 4(1): 25–32.

Lachenaud P., Bonnot F. and Oliver G. 1999. Use of floral descriptors to study variability in wild cocoa trees (*Theobroma cacao L.*) in French Guiana. Genet. Resour. Crop Evol. 46: 491–500.

Lachenaud P., Oliver G. and Letourmy P. 2000. Agronomic assessment of wild cocoa trees (*Theobroma cacao* L.) from the Camopi and Tanpok basins (French Guiana). Plant Genet. Resour. Newslett. 124: 1–6.

Lachenaud P. and Oliver G. 2001. Variability in various agronomic traits of wild cocoa trees (*Theobroma cacao* L.) from the Camopi and Tanpok basins (French Guiana). Plant Genet. Resour. Newslett. 128: 35–40.

Lanaud C. 1986. Genetic studies of *Theobroma cacao* L. with the help of enzymatic markers. I. Genetic control and linkage of nine enzymatic markers. Café Cacao Thé 30(4): 259–270.

Lanaud C., Risterucci A.M., Piereti I., Falque M., Bouet A. and Lagoda P.J.L. 1999. Isolation and characterization of microsatellites in *Theobroma cacao* L. Mol. Ecol. 12(8): 2141–2152.

Latchman B., Umaharan R., Surujdeo-Maharaj S. and Thévenin J.-M. 2000. Field assessment of cacao germplasm for resistance to witches' broom and black pod diseases: year one. Annual Report 1999, Cocoa Research Unit, The University of the West Indies, Trinidad and Tobago, pp. 39–46.

Laurent V., N'Goran J.A.K., Eskes A.B. and Lanaud C. 1995. Estimation of the level of heterozygosity and genetic diversity with RFLP markers. Ingenic Newslett. 1: 5–7.

Lerceteau E., Robert T., Pétiard V. and Crouzillat D. 1997. Evaluation of the extent of genetic variability among *Theobroma cacao* populations using RAPD and RFLP markers. Theoret. Appl. Genet. 95: 10–19.

Lockwood G. 1985. Genetic resources of the cocoa plant. Span 28(1): 14–16.

Magurran A.E. 1988. Ecological Diversity and its Measurement. Croom Helm, London.

Mooleedhar V., Maharaj W. and O'Brien H. 1995. The collection of Criollo cocoa germplasm in Belize. Cocoa Growers' Bull. 49: 26–40.

Motamayor J.C. 2001. Etude de la diversité génétique et de la domestication des cacaoyers du groupe Criollo (*Theobroma cacao* L) à l'aide de marqueurs molèculaires. Thèse. Université Paris XI, pp. 179

Motamayor J.C., Risterucci A.M., Lopez P.A., Ortiz C.F., Moreno A. and Lanaud C. 2002. Cacao domestication I: The origin of the cacao cultivated by the Mayas. Heredity 89: 380–386.

Nei M. 1972. Genetic distance between populations. Amer. Naturalist. 106: 283–292.

Phillips-Maro W. 1999. Studies at CATIE on Moniliasis Resistance (*Moniliophthora roreri* (Cif & Par.) Evans et al.) In: Proceedings of the International Workshop on the contribution of disease resistance to cocoa variety improvement. 24th–26th November 1996. Salvador, Bahia, Brazil. 111–118

Pound F.J. 1932. The genetic constitution of the cacao crop. First Ann. Rep. Cacao Res. 1931, Trinidad, pp. 10–24.

Pound F.J. 1938. Cacao and witches' broom disease of South America with notes on other species of *Theobroma*. Arch. Cocoa Res. 1: 26–71.

Pound F.J. 1943. First report on the selection of cacao trees for resistance to Witches' Broom disease. Unpublished report, Ministry of Agriculture, Trinidad.

Russell J.R., Hosein F., Johnson E., Waugh R. and Powell W. 1993. Genetic differentiation of cocoa (*Theobroma cacao* L.) populations revealed by RAPD analysis. Mol. Ecol. 2: 89–97.

Sounigo O., Umaharan R. and Sankar A. 1997. Assessing genetic diversity in the International Cocoa Genebank of Trinidad (ICG,T) using Isozyme Electrophoresis. Ingenic Newslett. 3: 11–16.

Sounigo O., Mooleedhar V., Iwaro D., Bekele F., Sreenivasan T.N., Thévenin J.M., Khan N. and Butler D. 2000. Strategy to establish a CFC Project Collection. Proceedings of the CFC/ICCO/IPGRI Project Workshop, 1–6 February 1998, Montpellier, France, pp. 29–37.

Swofford D.L., Olsen G.J., Waddell P.J. and Hillis D.M. 1996. Phylogenetic inference. In: Hillis D.M., Moritz C. and Mable B.K. (eds), Molecular Systematics, 2nd edn., Sinauer, Sunderland, pp. 407–514.

Umaharan R., Thévenin J.-M. and Surujdeo-Maharaj S. 2001. Evaluation of cocoa germplasm for resistance to witches'broom disease. Annual Report 2000, Cocoa Research Unit, The University of the West Indies, Trinidad and Tobago, pp. 30–33.

Wadsworth R.M., Ford C.S., End M.J. and Hadley P. 1997. International Cocoa Germplasm Database London, UK. The London International Financial Futures and Options Exchange.

Warren J.M. 1994. Isozyme variation in a number of populations of *Theobroma cacao* L. obtained through various sampling regimes. Euphytica 72: 121–126.

Wright S. 1978. Evolution and the Genetics of Populations. 4. Variability within and Among Natural Populations. University of Chicago Press, Chicago, USA.

Yeh F.C., Yang R.C. and Boyle T. 1999. POPGENE. Microsoft Windows based freeware for population genetic analyses.