

Cacao domestication I: the origin of the cacao cultivated by the Mayas

JC Motamayor^{1,5}, AM Risterucci¹, PA Lopez², CF Ortiz³, A Moreno⁴ and C Lanaud¹

¹CIRAD, Centre de Coopération Internationale en Recherche Agronomique pour le Développement, TA 40/03, Av. Agropolis, 34398 Montpellier cedex 5, France; ²INIFAP, C.P. 86400, A.P. 17, Huimanguillo, Tabasco, Mexico; ³Colegio de Postgraduados, Carr. Cardenas-Huimanguillo Km. 3.5 A. P. 24, Huimanguillo, Tabasco, Mexico; ⁴FONAIAP, Estación Experimental Miranda, Calle El Placer, Cauagua, Edo. Miranda, Venezuela

Criollo cacao (*Theobroma cacao* ssp. *cacao*) was cultivated by the Mayas over 1500 years ago. It has been suggested that Criollo cacao originated in Central America and that it evolved independently from the cacao populations in the Amazon basin. Cacao populations from the Amazon basin are included in the second morphogeographic group: Forastero, and assigned to *T. cacao* ssp. *sphaerocarpum*. To gain further insight into the origin and genetic basis of Criollo cacao from Central America, RFLP and microsatellite analyses were performed on a sample that avoided mixing pure Criollo individuals with individuals classified as Criollo but which might have been introgressed with Forastero genes. We distinguished these two types of individuals as Ancient and Modern Criollo. In contrast to previous studies, Ancient Criollo individuals formerly classified as 'wild', were found to

form a closely related group together with Ancient Criollo individuals from South America. The Ancient Criollo trees were also closer to Colombian-Ecuadorian Forastero individuals than these Colombian-Ecuadorian trees were to other South American Forastero individuals. RFLP and microsatellite analyses revealed a high level of homozygosity and significantly low genetic diversity within the Ancient Criollo group. The results suggest that the Ancient Criollo individuals represent the original Criollo group. The results also implies that this group does not represent a separate subspecies and that it probably originated from a few individuals in South America that may have been spread by man within Central America.

Heredity (2002) 89, 380–386. doi:10.1038/sj.hdy.6800156

Keywords: *Theobroma cacao* L.; Criollo; genetic diversity; domestication; RFLP; microsatellites

Introduction

Theobroma cacao ($2n = 2x = 20$) is a tree native to humid tropical regions of the northern part of South America and, according to some reports, of Central America (Miranda, 1962). Indeed, there is still some controversy about the origin and domestication of cacao.

Some authors believe that cacao was introduced into Central America. Although the first centre of domestication and culture has been identified as Central America, Van Hall (1914) stated that the most probable origin of cacao is the region of the Orinoco and Amazon basins, in the valleys of their tributaries. Cheesman (1944) considered that the centre of origin of cacao was the Upper Amazon near the Colombian-Ecuadorian border, on the eastern flanks of the Andes. Cheesman (1944) argued that although cacao has been cultivated in Mexico and Central America for over 2000 years, no truly wild populations were present in this region, suggesting that cacao was introduced into Central America and Mexico. Schultes (1984) hypothesized that once cacao had spread throughout the Amazon Valley, it could have dispersed

along two routes: one leading north and the other west. In this way, the domestication of cacao occurred in South America and then spread to Central America and Southern Mexico, carried by migrating Indians (Schultes, 1984).

Cuatrecasas (1964) suggested separate simultaneous origin in South and Central America, a hypothesis that has been supported by most subsequent authors (Cope, 1976; Wood and Lass, 1985; Gómez-Pompa *et al.*, 1990; Laurent *et al.*, 1994; De la Cruz *et al.*, 1995; Whitkus *et al.*, 1998). Based on the great morphological diversity observed in Central as well as in South America, Cuatrecasas (1964) proposed that North and South American cacao populations developed into two forms, separated by the Panama Isthmus. Both populations evolved independently and are recognised as separate subspecies (*T. cacao* ssp. *cacao* and *T. cacao* ssp. *sphaerocarpum*; Cuatrecasas, 1964). Furthermore, Cuatrecasas (1964) hypothesized that wild plants from the Lacandona rainforest from Mexico were possible ancestors of domesticated cacao. Separate simultaneous origin and domestication in both Central and South America has been suggested for the common bean, *Phaseolus vulgaris* (Velasquez and Gepts, 1994).

The subspecies proposed by Cuatrecasas correspond to the two morphogeographic groups proposed by Cheesman (1944); *T. cacao* ssp. *cacao* and *T. cacao* ssp. *sphaerocarpum* are synonyms for the Criollo and Forastero groups,

JC Motamayor

⁵Present address: USDA-ARS, 13601 Old Cutler Road, Miami, FL 33158, USA. E-mail: motamayor@hotmail.com

Received 26 March 2001; accepted 14 May 2002

respectively. A third hybrid group originating from crosses between Criollo and Forastero was called Trinitario. The Forastero group is composed of very diverse populations with different geographic origins: Upper Amazon, Lower Amazon, Orinoco and the Guianas. Forastero trees are also identified according to pod morphology, eg the Amelonado type characterised by a melon shape.

Criollo cacao was cultivated in Latin America during the pre-Columbian and colonial period, and had a higher quality than Forastero types, but with a low vigour and yield (Cheesman, 1944). Since 1825, it has been steadily replaced by more disease resistant and productive Trinitario clones in countries such as Venezuela (Pittier, 1935).

This paper describes our work on the genetic structure of Criollo, and its genetic relationships with other cacao populations, based on molecular analyses of a large sample of Criollo from several Latin American countries.

Materials and methods

Plant material

Samples were classified as Ancient Criollo, Modern Criollo, Trinitario, Lower Amazon Forastero, Orinoco Forastero, French Guiana Forastero, Upper Amazon Forastero and hybrids with at least one Upper Amazon Forastero parent (Table 1). A complete list of individuals used in the study is available upon request.

Ancient Criollo individuals consisted of trees showing the morphological traits described by Cheesman (1944) for the Criollo group, sampled from places where gene flow between Criollo and Trinitario or Forastero trees

was absent or limited due to the improbability of Trinitario or Forastero introductions. Most samples came from trees on old or abandoned farms and in private gardens in difficult to access towns. In Mexico, samples of Ancient Criollo were also collected in the Lacandona rainforest where wild trees had previously been reported and studied (Miranda, 1962; Cuatrecasas 1964; De la Cruz *et al*, 1995; Whitkus *et al*, 1998), and in places where Mayan peoples cultivated cacao; in the Yucatan sinkholes (Gómez-Pompa *et al*, 1990; De la Cruz *et al*, 1995; Whitkus *et al*, 1998) and on the Pacific coast of Mexico (Lopez-Mendoza, 1987). Samples from the Belizian rainforest associated with Mayan ruins (Mooledhar *et al*, 1995) were also included in the analysis.

Modern Criollo individuals were defined as those showing the morphological traits described by Cheesman (1944) for the Criollo group, but sampled on modern farms or on farms where significant introductions of Trinitario or Forastero were suspected. This class also included material from germplasm collections in Costa Rica, Côte d'Ivoire, Mexico, Venezuela and France. 'Modern Criollo' represents the genotypes studied as Criollo in previous biochemical and molecular studies (Lanaud, 1987; Bekele and Bekele, 1998; Ronning and Schnell, 1994; Laurent *et al*, 1995; N'Goran *et al*, 1994; De la Cruz *et al*, 1995; Lerceteau *et al*, 1997; Whitkus *et al*, 1998).

Trinitario, Forastero clones from Orinoco, French Guiana, the Lower and Upper Amazon (LAF and UAF) and hybrids with at least one Upper Amazon Forastero parent, were studied to compare the structure of their genetic diversity to that of Ancient and Modern Criollo.

Table 1 Number of individuals analysed for each morphogeographic group, geographic origin and molecular technique (RFLP/microsatellites)

	Ancient Criollo (92/41)	Modern Criollo (68/13)	Trinitario (67/14)	Lower Amazon Amelonado type (9/7)	French Guiana (4/4)	Orinoco Amelonado (3/3)	Upper Amazon Forastero (29/20)	Hybrids* (11/-)
Belize	2/3	-	-	-	-	-	-	-
Brazil	-	-	-	2/2	-	-	2/2	-
Cameroon	-	-	5/1	-	-	-	-	-
Colombia	8/2	1/-	4/-	-	-	-	3/3	-
Costa Rica	-	2/1	5/2	1/1	-	-	-	-
Ecuador	-	-	1/-	-	-	-	2/2	-
Côte d'Ivoire	-	-	8/3	5/1	-	-	-	6/-
French Guiana	-	-	-	-	4/4	-	-	-
Ghana	-	1/-	1/-	-	-	-	-	-
Grenada	-	-	1/-	-	-	-	-	-
Honduras	-	-	-	1/-	-	-	-	-
Indonesia	-	-	3/1	-	-	-	-	-
Mexico								
Lacandona	31/13	-	-	-	-	-	-	-
Yucatan	7/7	-	-	-	-	-	-	-
Michoacan	3/3	-	-	-	-	-	-	-
Tabasco	-	13/4	1/1	-	-	-	-	-
Malaysia	-	-	1/-	-	-	-	-	-
Nicaragua	3/3	2/2	-	-/3	-	-	-	-
Panama	-	1/1	1/1	-	-	-	-	-
Peru	-	-	-	-	-	-	22/13	-
Samoa	-	-	1/-	-	-	-	-	-
São Tomé	-	-	1/-	-	-	-	-	-
Trinidad	-	6/3	9/5	-	-	-	-	5/-
Venezuela	38/10	42/2	25/1	-	-	3/3	-	-

*Hybrids with at least one Upper Amazon Forastero parent.

DNA extraction

For RFLP analyses, DNA was isolated using ultracentrifugation in a caesium chloride-ethidium bromide gradient as described in Lanaud *et al* (1995). For microsatellite analysis, DNA isolation was performed as described in Risterucci *et al* (2000); after total resuspension, a QIAGEN genomic-tip was used to purify the DNA according to the manufacturer's instructions.

RFLP analysis

RFLP procedures: All RFLP procedures were conducted as described previously (Lanaud *et al*, 1995). *EcoRI* and *HindIII* were used for genomic DNA restriction.

DNA probes: Seventeen cDNA and eight genomic DNA probes, chosen for their coverage of the genetic map of *T. cacao* (Lanaud *et al*, 1995), were used to study 283 individuals.

Microsatellite analysis

Microsatellites: Sixteen microsatellites (Lanaud *et al*, 1999) were used to study the genetic diversity of 102 individuals.

Primers were end-labelled with γ -³³P ATP and PCR amplification was performed in a MJ Research PTC 100 thermal cycler, in 20 μ l of reaction mix containing 10 ng cacao DNA, 0.2 mM dNTP mix, 2 mM MgCl₂, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 0.2 μ M primer (5' end labelled with γ -³³P ATP) and 1 unit Taq polymerase (Eurobio). The samples were denatured at 94°C for 4 min, and subjected to 32 repeats of the following cycle: 94°C for 30 sec, 46°C or 51°C for 1 min and 72°C for 1 min. After adding 20 μ l of loading buffer (98% formamide, 10 mM EDTA, bromophenol blue, xylene cyanol), the mix was denatured at 92°C for 3 min and 3 μ l of each sample were loaded in a 5% polyacrylamide gel with 7.5 M urea and electrophoresed in 0.5% TBE buffer at 55 W for 1 h 40 min. The gel was dried for 30 min at 80°C and exposed overnight to X-ray film (Fuji RX).

Data analysis

RFLP alleles observed after hybridization with 25 probes were scored. Microsatellite allele sizes were scored by comparing PCR product lengths to the sequence of the genomic clone from which primers were designed. Repeat number variance was calculated as the sum of the squares of the difference between the mean number of repeats at the locus and the repeat number for each allele multiplied by its frequency in the population.

The genetic diversity statistic (Nei, 1978), the mean number of alleles per locus, the percentage of polymorphic loci at the 95% significance level, and the observed heterozygosity were calculated using Genetix 4.0 software. Hybrid clones were not included in the calculation of these statistics for the Forastero group.

For RFLP and microsatellite data the shared allele distance (D_{AS}) (Chakraborty and Jin, 1993) was calculated. This distance is equal to 1 minus the proportion of shared alleles: $D_{AS} = 1 - (a/2n)$, where a is the number of common alleles to individuals i and j and n the number of loci studied. This distance was computed by averaging the values over all available loci between two individuals.

Multidimensional scaling (MDS) was performed on a D_{AS} matrix calculated from RFLP data using the MDS procedure of SAS (version 6.12) software.

To determine the relatedness between Criollo and Forastero individuals a neighbour-joining (N-J) tree (Saitou and Nei, 1987) was constructed from the shared allele distance between individuals obtained from microsatellite data. The D_{AS} estimation, N-J tree construction and bootstrapping procedures were conducted using a computer program kindly provided by Jean-Marie Cornuet and Sylvain Piry (Laboratoire de Modelisation et Biologie Evolutive, INRA, Montpellier, France). Hybrid genotypes were excluded from the analysis.

Microsatellite and RFLP analyses were compared using a Mantel test on D_{AS} matrices of 92 individuals. The Mantel test was performed after 1000 permutations on the order of individuals in one of the matrices using the software Genetix 4.0.

Results

RFLP analysis

After hybridization with 25 probes, 66 alleles were detected. The unbiased gene diversity (Nei, 1978) was higher for Forastero than for Ancient Criollo (Table 2). The average number of alleles per locus was highest for the Forastero group, as were the percentage of polymorphic loci and the observed heterozygosity. This latter parameter was very low (0.002) for the Ancient Criollo group, indicating almost complete homozygosity as a genetic characteristic of this group (Table 2).

All Ancient Criollo individuals ($n = 92$) mapped in the right half of the MDS plane, with a cluster of homogeneous trees and several near-identical individuals in the third quadrant (Figure 1). Only eight multilocus genotypes were observed among the Ancient Criollo individuals (three of them are represented superimposed on Figure 1), whereas each Forastero clone had a unique RFLP genotype. Some Ancient Criollo individuals that shared identical RFLP genotypes were members of different morphotypes or from diverse geographical areas (Venezuela, Colombia, Nicaragua, Belize and Mexico). For instance, some cacao trees from the Lacandona rainforest had identical RFLP profiles to genotypes putatively cultivated by the Mayas (found in the sinkholes of Yuca-

Table 2 RFLP diversity within Ancient Criollo, Forastero, Modern Criollo and Trinitario

	No.	H_{nb}	H_{obs}	$P_{0.95}$	A
Ancient Criollo (ssp. cacao)	92	0.002 (0.008)	0.002 (0.008)	0.00	1.08
Forastero (ssp. sphaerocarpum)	37	0.39 (0.17)	0.18 (0.11)	0.92	2.36
Modern Criollo	68	0.39 (0.18)	0.47 (0.23)	0.84	2.04
Trinitario	67	0.42 (0.18)	0.43 (0.19)	0.84	2.12

H_{nb} = unbiased gene diversity (Nei, 1978); No. = sample size; H_{obs} = observed heterozygosity; $P_{0.95}$ = proportion of polymorphic loci when most frequent allele does not exceed 95%; A = mean number of alleles per locus. Standard deviations are indicated in parentheses.

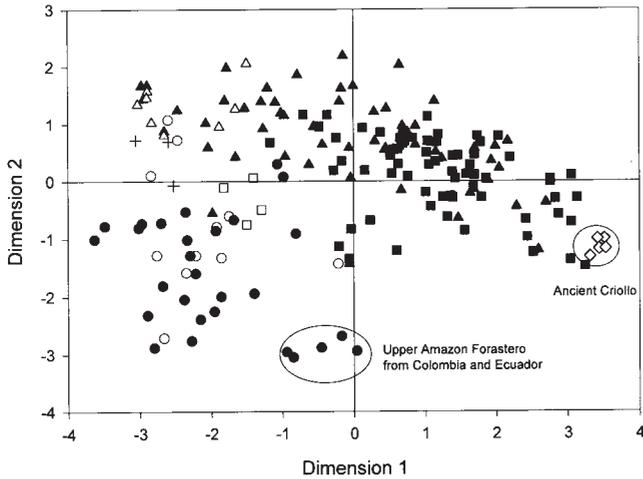


Figure 1 Multidimensional scaling plot of 283 genotypes based on RFLP D_{AS} matrix. \diamond = Ancient Criollo; \blacksquare = Modern Criollo; \blacktriangle = Trinitario; \triangle = Lower Amazon Amelonado type; \square = French Guyanan individuals; $+$ = Orinoco Amelonado individuals; \bullet = Upper Amazon Forastero individuals; \circ = hybrids with at least one Upper Amazon Forastero parent. R^2 of the regression of the genetic distances versus the graphical distances was 0.55.

tan, on the Pacific Coast of Mexico and in Belize) as well as to individuals cultivated today in South America.

Modern Criollo individuals are superimposed onto Trinitario (Figure 1). Furthermore they are distributed continuously from Ancient Criollo in the third quadrant to Amelonado Forastero in the first. Only three Modern Criollo and one Trinitario are found in the fourth quadrant of the plane of the MDS. These individuals show introgression of alleles specific to Upper Amazon Forastero trees from Peru (see Table 3). Most of the hybrids with at least one Upper Amazon Forastero parent were represented in the fourth quadrant.

The statistics for the genetic diversity of Modern Criollo were similar to those of Trinitario (Table 2). The unbiased genetic diversity and the observed heterozygosity of Modern Criollo were higher than those of Ancient Criollo. Similarities between Modern Criollo and Trinitario are to be expected given that the distinction, based on morphological traits, between the two types is subjective.

Microsatellite analysis

The 16 microsatellite loci detected 150 alleles. Despite the increased number of alleles, the genetic diversity of the Ancient Criollo group (0.04) observed from microsatellite

Table 3 Number of RFLP and microsatellite alleles that are specific to groups of individuals from different geographic regions, number of individuals studied in brackets

	Colombia-Ecuador	Peru	French Guiana	Ancient Criollo South and Central America	Orinoco	Lower Amazon (Brazil)
RFLP	0 (5)	6 (22)	1 (4)	3 (92)	0 (3)	0 (9)
Microsatellites	28 (5)	21 (13)	5 (4)	5 (41)	5 (3)	1 (8)

data was still very low compared to that observed for the Forastero group (0.78; Table 4). Within geographic regions, the genetic diversities of 13 individuals from Peru and five individuals from Colombia-Ecuador were similarly high (0.70). The observed heterozygosity was 0.00 for the Ancient Criollo and 0.34 for Forastero. The average number of alleles per locus was the highest for the Forastero group (8.69 and 1.19 for Forastero and Criollo, respectively), as was the percentage of polymorphic loci (1.00 vs 0.06). Allele size variance was much higher for Forastero than for Ancient Criollo (14.02 vs 0.08).

Within the Ancient Criollo individuals ($n = 41$), only six genotypes were found among different morphotypes or across diverse geographic zones, as was observed for the RFLP analysis (Figure 2, group A). In this tree, Ancient Criollo individuals were more related to Colombian-Ecuadorian Forastero individuals (EBC5, EBC 6, EBC 10, Lcteen 37 and Lcteen 355) than the latter are to some Peruvian, French Guiana or Lower Amazon Forastero individuals. The clustering pattern reflects the geographic origin of individuals. Medium to high bootstrap values support this last result (Figure 2).

In general, for the microsatellite analysis, Modern Criollo had similar values for genetic diversity statistics to those obtained for the Trinitario group (Table 4).

As observed in the RFLP analysis, Modern Criollo and Trinitario microsatellite alleles were also present in Ancient Criollo individuals and Amelonado clones. This finding, combined with the evidence from genetic diversity statistic (similar values obtained for both Modern Criollo and Trinitario), supports the hybrid character of Modern Criollo.

Comparison of RFLP and microsatellite analyses

The 25 RFLP probes detected 66 alleles, whilst 16 microsatellites loci detected 150 alleles. D_{AS} values were slightly lower for RFLP than for microsatellite data; mean distance between individuals was 0.41 (s.d. 0.21) and 0.64 (s.d. 0.30), respectively. The relationship between individuals was very similar for both techniques. For example, Upper Amazon individuals from Colombia and Ecuador were more related to Ancient Criollo for D_{AS} values than to other Forastero individuals using both RFLP and microsatellite markers (Figures 1 and 2). For the mantel test of D_{AS} RFLP and microsatellite matrices, a Pearson

Table 4 Microsatellite diversity within Ancient Criollo, Forastero, Modern Criollo and Trinitario. Genetic diversity within Forastero from Peru and Ecuador is shown

	No.	H_{nb}	H_{obs}	$P_{0.95}$	A	Av
Ancient Criollo (ssp. <i>cacao</i>)	41	0.04 (0.13)	0.00 (0.01)	0.06	1.19	0.08
Forastero (ssp. <i>sphaerocarpum</i>)	28	0.78 (0.13)	0.34 (0.13)	1.00	8.69	14.02
Peru	13	0.70 (0.18)	0.48 (0.25)	1.00	5.50	–
Colombia-Ecuador	5	0.70 (0.13)	0.39 (0.19)	1.00	3.94	–
Modern Criollo	13	0.52 (0.03)	0.59 (0.14)	1.00	2.37	–
Trinitario	14	0.50 (0.04)	0.68 (0.14)	1.00	2.19	–

No. = sample size; H_{nb} = unbiased gene diversity (Nei, 1978); H_{obs} = observed heterozygosity; $P_{0.95}$ = proportion of polymorphic loci when most frequent allele does not exceed 95%; A = mean number of alleles per locus; Av = allele size variance. Standard deviations are indicated in parentheses.

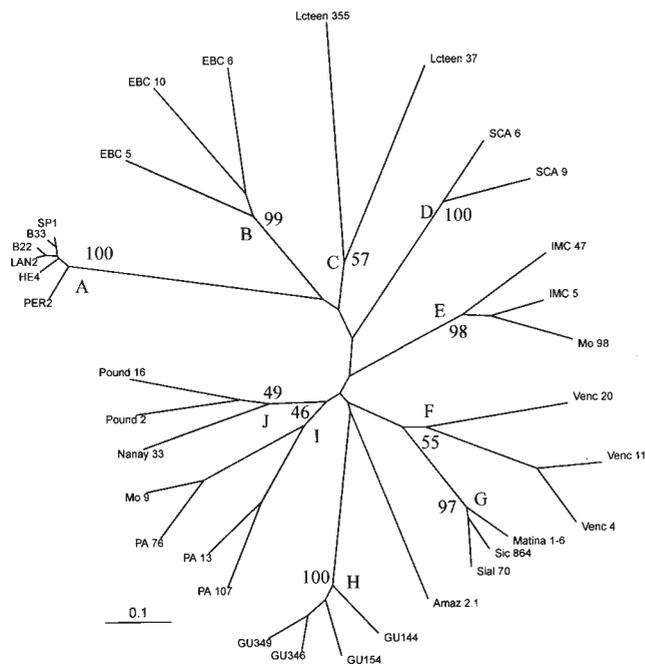


Figure 2 Neighbour-joining tree of Forastero ($n = 28$) and Ancient Criollo ($n = 41$) genotypes based on the shared allele distance calculated from microsatellite data. A: All the 41 Ancient Criollo individuals analysed cluster under this node; B: Forastero from Colombia; C: Forastero from Ecuador; D: Forastero from Peru; E: Forastero from Peru (Iquitos); F: Forastero from Venezuela (Orinoco river); G: Lower Amazon Amelonado type; H: Forastero from Guyana; I: Forastero from Peru (Parinari river, except MO 9); J: Forastero from Peru (Nanay river). Bootstrap values have been computed over 2000 replications by resampling loci and noted as percentages. Names of clones included in the analysis are given in the graphic.

correlation coefficient of 0.9 ($r = 0.9$) was obtained with a probability of one for dependence between the two matrices.

Discussion

The analyses of RFLP and microsatellite markers presented here shed new light on the patterns of genetic diversity and genetic relationships amongst *T. cacao* populations. Both techniques yielded equivalent results, despite the number of alleles detected per locus being significantly higher for microsatellites than for RFLPs. The correlation between RFLP and microsatellite D_{AS} matrices was highly significant. There is no apparent effect of the higher mutation rate for microsatellites (Dallas, 1992; Dietrich *et al*, 1992) on the determination of relatedness between cacao individuals. Congruency between RFLP and microsatellite diversity patterns has also been observed for other species (Pejic *et al*, 1998; Desplanque *et al*, 1999).

Both our RFLP and microsatellite analyses clearly distinguished Ancient Criollo individuals from Modern Criollo (Ancient Criollo individuals introgressed with Forastero genes). It is important to note that previous studies using isozymes (Lanaud, 1987; Ronning and Schnell, 1994), RFLP (Laurent *et al*, 1994; Lerceteau *et al*, 1997), and RAPD markers (N'Goran *et al*, 1994; Lerceteau *et al*, 1997) have analysed what are defined here as Modern Criollo (usual representatives of the Criollo group). The

present study is the first to show that the two types (Modern and Ancient) can be distinguished, and this was made possible by using a sample that avoided mixing pure Criollo individuals with individuals classified as Criollo but likely to have been introgressed with Forastero genes. Thus, individuals classified as Ancient Criollo constitute the true Criollo group comprised of cacao genotypes cultivated before the introduction of Forastero individuals to cacao plantations. Natural hybridisation between these two groups later gave rise to the appearance of Modern Criollo or Trinitario.

De la Cruz *et al* (1995) and Whitkus *et al* (1998) found that cacao trees from the Lacandona rainforest and 'Criollo' from germplasm collections could be clearly distinguished. These studies used dominant markers (RAPD), and the relatedness between what was called Criollo and what was called 'wild' (individuals from the Lacandona rainforest) was not clearly established, in contrast to the present study. They too found that samples from the Lacandona rainforest and those from the sinkholes of the Yucatan were clearly different.

In the present study, we analysed seven individuals from three sinkholes from Yucatan (near the towns of Yaxcaba, Tixcaltuyub and Chechmil). Using 25 RFLP probes and 16 microsatellites we found that these seven individuals shared an identical genotype. Contrary to the findings of De la Cruz *et al* (1995) and Whitkus *et al* (1998) little differentiation was observed between individuals from Yucatan and the Lacandona rainforest: The genotype found in the seven individuals from Yaxcaba, Tixcaltuyub and Chechmil was also found in nine out of 13 individuals from the Lacandona rainforest.

The origin of the cacao cultivated by the Mayas

Very low diversity (Figure 1, Tables 2 and 3) was found within the Ancient Criollo group comprising individuals from the Lacandona rainforest, even though some of them were obtained from distant sites. It has been suggested that the Criollo group originated in the Lacandona rainforest where such trees are apparently present in the wild state (Miranda, 1962; Cuatrecasas, 1964; Gómez-Pompa *et al*, 1990 and De la Cruz *et al*, 1995). This hypothesis does not agree with the results present here. Indeed, a wild population should exhibit levels of genetic diversity similar to that observed within geographic areas (for example, in Peru or Colombia-Ecuador, Table 4). This was not the case; very low diversity associated with high homozygosity was observed in Central America (including the Lacandona rainforest). Moreover, cacao from the Lacandona rainforest was found to be identical at a molecular level to individuals putatively cultivated by the Mayas (those found in sinkholes, the Pacific Coast of Mexico and Belize) and to individuals from the regions of southwestern Venezuela and northeastern Colombia. Therefore, the population consisting of trees found at the Lacandona rainforest should neither be considered wild nor as originating from this region. Another element that must be considered is the absence of palynological evidence of the presence of *Theobroma* in the forests of Chiapas before the human colonization. Pollen of genera belonging to the modern vegetation of Chiapas has been observed from Tertiary deposits, but not *Theobroma* or related genera such as *Herrania* (Graham, 1999). In addition, in the Lacandona rainforest, where material was sampled, vestiges of the Mayan civilization were fre-

quently found. Thus, the presence of Criollo cacao trees in the Lacandona rainforest may be a remnant of cacao cultivation by the Mayas.

Our results contradict Cuatrecasas' (1964) hypothesis that Criollo is a separate subspecies that evolved independently to South American populations in Central America, and suggest rather that the Criollo group had a South American origin. If Cuatrecasas' hypothesis is true, all wild Forastero individuals should be clustered independently of Criollo in the analysis of genetic relatedness between individuals (Figure 2). In contrast, Ancient Criollo individuals are more related to Forastero from Colombia and Ecuador (EBC 5, EBC 6, EBC 10, Lcteen 37 and Lcteen 355) than the latter are to other Forastero individuals from French Guiana, the Orinoco, the Lower Amazon or some from Peru (ie, GU154, Matina 1-6, Venc 4, PA 107). Consequently, the Criollo group does not form a separate subspecies (ssp. *cacao*) from the one comprising individuals from South America (ssp. *sphaerocarpum*). Moreover, since microsatellite mutations tend to change allelic size by small amounts (Schlötterer and Tautz, 1992), the low allele size variance found for Ancient Criollo (0.08) compared to Forastero (14.02) also indicates a recent origin for this group.

Classification within the species

Since genetic distances between some Forastero individuals are equivalent to that observed between some Forastero and Ancient Criollo, a classification of cacao based on two main populations (Criollo and Forastero) has no genetic base. Indeed, the classification based on Criollo and Forastero mentioned by Cheesman (1944) and first proposed by Morris (1882) was simply based on the terms used by the Venezuelan cacao producers of the central coastal zone. At the time of Morris, the terms Criollo and Forastero were employed to distinguish the local cultivated trees (with a specific pod morphology) from the introduced foreign material.

Evolution and domestication

Allopatric divergence of cacao populations is suggested by the clustering pattern of individuals (Figure 2). Clear divergence of cacao from specific origins such as French Guiana and Ecuador has been reported (Lanaud, 1987; Laurent *et al*, 1994; Sounigo *et al*, 1996; Lerceteau *et al*, 1997). Although a reduced number of Forastero genotypes for each South American region were studied (Table 3), RFLP and microsatellite alleles, specific (allelic frequency higher than 0.05) to groups of individuals from different geographic areas were identified. Furthermore, mitotypes as well as rDNA alleles specific to different geographic origins have been observed (Laurent *et al*, 1993a, b). Collection expeditions in Amazonia (Allen and Lass, 1983; Young, 1994) revealed striking differences in morphology among populations from different river tributaries or other topographic features. Patterns of genetic diversity in other Amazonian plant species, eg *Hevea brasiliensis* (Seguin *et al*, 1999) and *Elaeis oleifera* (Barcelos, 1998), have also been reported as being associated with streams and explained according to 'refuge' theory (Simpson and Haffer, 1978; Haffer, 1982). The geological changes on which the 'refuge' theory is based could also explain divergence among *T. cacao* populations (including the population at the origin of Criollo individuals). Isolated cacao populations in constricted

forest 'refuges', possibly in contracted gallery forest along scattered tributaries, could have survived during the adverse climatic conditions that occurred during the Quaternary period. These populations could then have evolved independently into different variants prior to a subsequent phase of forest expansion (Lanaud, 1987; Young, 1994). Stepwise founder events over repeated cycles of forest contraction and expansion could have then led to the loss of much natural genetic diversity in Criollo prior to domestication.

Subsequently, bottlenecks could have occurred during domestication. The intervention of man through selection during cultivation could also have reduced the effective number of individuals of the original Criollo population. People would then have been able to fix and maintain the extreme phenotypes that could have appeared due to mutations in a few genes, and spread the crop into Central America. Some cacao types may have been of special interest to people and therefore selected through collection, maintenance and use. Indeed, such extremely different phenotypes as Porcelana and Pentagona (one very smooth and the other very rough) contrast strongly with other pod types. The Pentagona or Lagarto type, for example, has the finest pod cortex, increasing the bean weight to pod weight ratio and facilitating the extraction of the beans from the fruit. Characteristic traits of Criollo trees such as the sweet pulp of its beans and the fact that it needs less fermentation, could be seen as targets of selection by man during more than 1500 years of cultivation.

The results of the present study, combined with the evidence presented above, uphold the theories of Van Hall (1914), Cheesman (1944) and Schultes (1984), that cacao originated in South America, and was later introduced by man in Central America.

Acknowledgements

This study was partly financed by the International Plant Genetic Resources Institute (IPGRI) through the 1998 Vavilov-Frankel Fellowship, by the Consejo Nacional de Ciencia y Tecnología of Venezuela (CONICIT) by the Fundación para el Desarrollo de la Ciencia y la Tecnología del Estado Aragua (FUNDACITE-ARAGUA, Venezuela) and by the Centre de Coopération Internationale en Recherche Agronomique pour le Développement (CIRAD, France).

References

- Allen JB, Lass RA (1983). London cocoa trade Amazon project: final report, phase 1. *Cocoa Grower's Bull* 34: 1-71.
- Barcelos E (1998). Etude de la diversité génétique du palmier à huile américain (*Elaeis oleifera*) par marqueurs moléculaires (RFLP et AFLP). PhD Dissertation, Ecole Nationale Supérieure Agronomique de Montpellier.
- Bekele F, Bekele I (1998). A sampling of the phenetic diversity of cacao in the international cocoa gene bank of Trinidad. *Crop Sci* 36: 57-64.
- Chakraborty R, Jin L (1993). A unified approach to study hyper-variable polymorphisms: statistical considerations of determining relatedness and populations distances. In: Pena SDJ, Chakraborty R, Epplen JT, Jeffreys AJ (eds) *DNA fingerprinting: State of the Science*, Birkhäuser Verlag: Basel. pp 153-175.
- Cheesman EE (1944). Notes on the nomenclature, classification

- and possible relationships of cocoa populations. *Trop Agricult* **21**: 144–159.
- Cope FW (1976). Cacao. *Theobroma cacao* L. (Sterculiaceae). In: Simmonds NW (ed) *Evolution of Crop Plants*, Longman: London. pp 207–213.
- Cuatrecasas J (1964). Cacao and its allies: a taxonomic revision of the genus *Theobroma*. *Contrib US Herbarium* **35**: 379–614.
- Dallas JF (1992). Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. *Mammal Genome* **5**: 32–38.
- De la Cruz M, Whitkus R, Gomez-Pompa A, Mota-Bravo L (1995). Origins of cacao cultivation. *Nature* **375**: 542–543.
- Desplanque B, Boudry P, Broomberg K, Saumitou-Laprade P, Cuguen J, Van Djkj H (1999). Genetic diversity and gene flow between wild, cultivated and weedy forms of *Beta vulgaris* L. (Chenopodiaceae), assessed by RFLP and microsatellite markers. *Theor Appl Genet* **98**: 1194–1201.
- Dietrich W, Katz H, Lincoln SE, Shin HS, Friedman J, Dracopoli N *et al* (1992). A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* **131**: 423–447.
- Genetix Version 4.0. *Logiciel sous Windows TM pour la génétique des populations*. Laboratoire Génome et Populations, CNRS UPR 9060, Université de Montpellier II, Montpellier, France.
- Gómez-Pompa A, Flores JS, Fernandez MA (1990). The sacred cacao groves of the Maya. *Latin Am Antiquity* **1**: 247–257.
- Graham A (1999). Studies in Neotropical paleobotany. XIII. An oligo-miocene palynoflora from Simojovel (Chiapas, Mexico). *American. Am J Bot* **86**: 17–31.
- Haffer J (1982). General aspects of the refuge theory. In: Prance GT (ed) *Biological Diversification in the Tropics*, Colombia University Press: New York. pp 6–24.
- Lanaud C (1987). Nouvelles données sur la biologie du cacaoyer (*Theobroma cacao* L.): diversités de populations, systèmes d'incompatibilité, haploïdes spontanés; leurs conséquences pour l'amélioration génétique de cette espèce. PhD Dissertation, Université Paris XI.
- Lanaud C, Risterucci AM, N'Goran AKJ, Clément D, Flament MH, Laurent V *et al* (1995). A genetic linkage map of *Theobroma cacao* L. *Theor Appl Genet* **91**: 987–993.
- Lanaud C, Risterucci AM, Pieretti I, Falque M, Bouet A, Lagoda PJJ (1999). Isolation and characterization of microsatellites in *Theobroma cacao* L. *Mol Ecol* **8**: 2141–2152.
- Laurent V, Risterucci AM, Lanaud C (1993a). Chloroplast and mitochondrial DNA diversity in *Theobroma cacao*. *Theor Appl Genet* **87**: 81–88.
- Laurent V, Risterucci AM, Lanaud C (1993b). Variability for nuclear ribosomal genes within *Theobroma cacao* L. *Heredity* **71**: 96–103.
- Laurent V, Risterucci AM, Lanaud C (1994). Genetic diversity in cocoa revealed by cDNA probes. *Theor Appl Genet* **88**: 193–198.
- Lerceteau E, Robert T, Pétiard V, Crouzillat D (1997). Evaluation of the extent of genetic variability among *Theobroma cacao* L. accessions using RAPD and RFLP markers. *Theor Appl Genet* **95**: 10–19.
- Lopez Mendoza R (1987). *El cacao en Tabasco*. Colección Cuadernos Universitarios. Serie Agronomía No. 13. Universidad Autónoma de Chapingo, Mexico.
- Miranda F (1962). *Wild Cacao in the Lacandona Forest*, Chiapas, Mexico. *Cacao (Turrialba)*, **7**: 7. CATIE: Costa Rica.
- Mooledhar V, Maharaj WW, O'Brien H (1995). The collection of Criollo cocoa germplasm in Belize. *Cocoa Grower's Bull* **49**: 26–40.
- Morris D (1882). *Cocoa: how to grow and how to cure it*. Jamaica.
- Nei M (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**: 583–590.
- Pejic I, Ajmone-Marsan P, Morgante M, Kozumplick V, Castiglioni P, Taramino G *et al* (1998). Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs. *Theor Appl Genet* **97**: 1248–1255.
- Pittier H (1935). Degeneration of cacao through natural hybridization. *Heredity* **36**: 385–390.
- Risterucci AM, Grivet L, N'Goran JAK, Pieretti I, Flament MH, Lanaud C (2000). A high-density linkage map of *Theobroma cacao* L. *Theor Appl Genet* **101**: 948–955.
- Ronning CM, Schnell RJ (1994). Allozyme diversity in a germplasm collection of *Theobroma cacao* L. *J Hered* **85**: 291–295.
- Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 685–691.
- Schlötterer C, Tautz D (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acids Res* **20**: 211–215.
- Schultes RE (1984). Amazonian cultigens and their northward and westward migrations in pre-Columbian times. In: Stone D (ed) *Pre-Columbian plant migration, Papers of the Peabody Museum of Archaeology and Ethnology*. Vol **76**: Mass.: Harvard University Press: Cambridge. pp 69–83.
- Seguin M, Flori A, Legnaté H, Clément-Demange A (1999). L'hévea. In: CIRAD (ed) *Diversité génétique des plantes tropicales cultivées*, Collection Repères, Montpellier: France. pp 241–269.
- Simpson BB, Haffer J (1978). Speciation patterns in the Amazonian forest biota. *Ann Rev Ecol Systemat* **9**: 497–518.
- Sounigo O, Christopher Y, Umaharan R (1996). Genetic diversity assessment of *Theobroma cacao* L. using isoenzyme and RAPD analyses. In: *CRU, Annual Report 1996* Port of Spain, Trinidad and Tobago. pp 35–51.
- Van Hall CJJ (1914). *Cocoa*. Macmillan: London.
- Velasquez VLB, Gepts P (1994). RFLP diversity of common bean (*Phaseolus vulgaris*) in its centres of origin. *Genome* **37**: 256–263.
- Whitkus R, De la Cruz M, Mota-Bravo L (1998). Genetic diversity and relationships of cocoa (*Theobroma cacao* L.) in southern Mexico. *Theor Appl Genet* **96**: 621–627.
- Wood GAR, Lass RA (1985). *Cocoa*. Longman: London.
- Young AM (1994). *The Chocolate Tree: a natural history of cacao*. Smithsonian Institution Press: Washington.
- N'Goran JAK, Laurent V, Risterucci AM, Lanaud C (1994). Comparative genetic diversity of *Theobroma cacao* L. using RFLP and RAPD markers. *Heredity* **73**: 589–597.