Reexamining the Classification of *Theobroma cacao* L. Using Molecular Markers

Antonio Figueira¹, Jules Janick², Morris Levy³, and Peter Goldsbrough

Department of Horticulture, West Lafayette, IN 47907-1165

Additional index words. RFLP, RAPD, *Herrania*, cocoa, taxonomy

Abstract. Genetic similarities among eight *Theobroma* and two *Herrania* species, including 29 genotypes of *T. cacao*, were estimated by rDNA polymorphism. A phenogram based on these genetic similarities significantly separated two clusters: one cluster included all *Herrania* and *Theobroma* species, except *T. cacao*, while the second contained 28 of 29 *T. cacao* genotypes. There was no clear distinction between *Herrania* and *Theobroma* species. Separation of 29 *T. cacao* genotypes, representing all races and various origins, had no congruency with the conventional classification into three horticultural races: Criollo, Forastero, and Trinitario. Genetic similarities in *T. cacao*, estimated with RAPD markers, indicated continuous variation among the generally similar but heterogeneous genotypes. The wild genotypes formed an outgroup distinct from the cultivated genotypes, a distinction supported by the rDNA data. The phenograms constructed from RAPD and rDNA data were not similar within the wild and cultivated cacao subsets.

The genus *Theobroma* consists of small understory trees of the lowland neotropical rainforest from the Amazon basin through southern Mexico (Purseglove, 1968). Cuatrecasas (1964) recognized 22 species and divided the genus into six sections, based on mode of germination, branching pattern, and flower morphology. Interspecific crosses can produce viable hybrids and certain species are graft compatible (Addison and Tavares, 1951). Most species have a restricted geographical distribution, with the Andes establishing a major separation between species (Baker et al., 1953).

The genus *Herrania* has a similar geographic distribution, with only one species occurring in Central America and the other 16 limited to South America (Schultes, 1958). *Herrania* species are morphologically similar to *Theobroma*, being distinguished by the unbranched orthotropic growth, palmately compound leaves, very long petal laminae, and nonconvoluted cotyledons (Baker et al., 1953). Until the revision of Schultes (1958), *Herrania* was considered a section of *Theobroma* (Cuatrecasas, 1964). The chromosome number for both genera is $2n = 20$ (Muñoz Ortega, 1948).

Theobroma cacao (cacao) is the only species of major economic importance, because its oil-rich seed is the unique source of cocoa solids and cocoa butter. Cacao was domesticated in pre-Colombian times by Meso-American natives (Bergmann, 1969). Cuatrecasas (1964) postulated that a natural cacao population once extended from the Guyana-Amazon region through southern Mexico and ultimately developed into two forms separated by the Panama isthmus. Cheesman (1944) proposed the headwaters of the Amazon basin region as the center of diversity for cacao. Isozyme studies by Lanaud (1987) have confirmed that populations found in this region exhibit the greatest genetic diversity, and contain the majority of known isozyme alleles.

Various attempts have been made to categorize cacao genotypes into three horticultural races (i.e., Criollo, Forastero, and Trinitario) based on morphological descriptors and geographic origin (Engels, 1981; Enriquez and Soria, 1967). Criollo and Forastero races have distinct historical and commercial features,

and have been considered as subspecies (Cuatrecasas, 1964). Domesticated in Meso-America, the Criollo race is characterized by having pods that are red or yellow when ripe, deeply furrowed, warty, and with a pointed end and a thin husk. Criollo seeds are plump with white or pale violet cotyledons, and have superior flavor. Historically, all nonCriollo cacaos were considered to be Forastero, but the latter name is currently used to describe cacao types that have nonpigmented pods, with thick, hard pod husks, and flat, dark purple seeds. Forastero can be subdivided into Upper Amazonian [wild or semi-wild cacaos as described by Pound (1938)] and Lower Amazonian, characterized by a rather uniform pod type called Amelonado (nonpigmented, smooth, melon-shaped pods with a blunt end). Lower Amazonian cacaos now constitute the most prevalent cultivated type worldwide, grown notably in West Africa and Brazil. These types grow wild in the Guyanas and the eastern Brazilian Amazon, and may have been partially domesticated by pre-Colombian Amazonian natives for their aromatic pulp (Barrau, 1979). Cacaos with red pods have not been found in the Amazon region, and pod color remains the unique characteristic that discriminates Amazonian and other populations (B.G.D. Bartley, unpublished).

Trinitario is considered to be either an intermediate type between Criollo and Forastero (Lockwood and Gyamfi, 1979) or a group of hybrids that display characteristics that include the total range of variation (Cheesman, 1944). Trinitario types have not been found under wild conditions. Because Criollos and Forastero types are so heterogeneous, the resulting hybrids might not be distinct from the parental populations, making Trinitarios impossible to define except by geographical origin (B.G.D. Bartley, unpublished). Consequently, the race classification system based on pod and seed characters is, at best, imprecise.

Isozyme markers have been used in an attempt to validate the conventional classification, but limited polymorphism does not allow discrimination between cacao types (Lanaud, 1987). The development of molecular genetic markers such as restriction fragment length polymorphisms (RFLPs) and random amplified polymorphic DNAs (RAPDs) has made it possible to evaluate phylogenetic relationships more efficiently (Figueira et al., 1992; Lerceteau et al., 1992; Wilde et al., 1992).

The objectives of this study were 1) to use molecular markers to examine phylogenetic relationships among various *Theobroma* and *Herrania* species based on RFLPs of ribosomal RNA genes (rDNAs) and 2) to determine if the organization of genetic diver-

Accepted for publication 3 Mar. 1994. Received for publication 12 Nov. 1993. Journal paper no. 13,791 of the Purdue Univ. Agricultural Experiment Station. The cost of publishing this paper was defrayed in part by the payment of page charges. Under postal regulations, this paper therefore must be hereby marked *advertisement* solely to indicate this fact.

¹Current address: Alimirante Cacau, Caixa Postal 55, Itajuipá, Bahia, Brazil.

²To whom reprint requests should be addressed.

³Dept. of Biological Sciences.

sity among a broad geographic sample of *Theobroma cacao* genotypes examined with rDNA polymorphisms and RAPDs was congruent with the traditional separation of *T. cacao* into the horticultural races Criollo, Forastero, and Trinitario.

Materials and Methods

Plant material. Thirty-eight accessions of *Theobroma* and *Herrania*, including 30 genotypes of *T. cacao* representing various populations and origins, were collected at the International Cocoa Germplasm repository, Centro Agronomico Tropical de Investigacion y Ensenanza, Turrialba, Costa Rica. The origin and description of genotypes are given in Tables 1 and 2 and Fig. 1.

DNA extraction method. Young leaves from each accession were brought into the laboratory, washed in distilled water, blotted dry, and frozen in liquid nitrogen. The frozen leaves were kept in dry ice, transported to the United States, and stored at -70C for up to 6 months. DNA was extracted from frozen leaves using the method described by Figueira et al. (1992). Final DNA concentrations were measured spectrophotometrically.

DNA hybridization experiments. DNA samples (0.5 µg) were digested for 4 h with 20 units of *DraI*, *EcoRV*, *EcoRI*, (Promega Corp., Madison, Wis.), or *BamHI* (New England Biolabs, Beverly, Mass.). The digested DNAs were separated on a 0.7% agarose gel using a voltage gradient of 2 V/cm in TAE buffer (40 mM Tris-base, 5 mM sodium acetate, 1 mM EDTA, pH 7.9). DNA was transferred to a nitrocellulose membrane (Schleicher & Schuell, Keene, N.H.) according to Southern (1975). The hybridization probe was pBG35, a plasmid containing a flax rDNA repeat (Goldsbrough and Cullis,

1981), and was labelled with ³²P-dCTP by random priming using a Decaprime kit (Ambion, Austin, Texas). The hybridization conditions were as described by Goldsbrough et al. (1990).

Random amplified polymorphic DNAs. Amplification reactions were performed as described by Williams et al. (1990) with minor modifications. The total reaction volume was 25 µliters, containing 10 mM Tris-HCl (pH 8.8), 50 mM KCl, 1.5 mM MgCl₂, 0.1% Triton X-100, 100 µM dATP, dCTP, dGTP, TTP, 0.2 µM primers, 25 ng genomic DNA, and 1.5 units *Taq* DNA polymerase (Promega Corp., Madison, Wis.). The primers used in this study were from Operon (Alameda, Calif.), Kits B and C, with the following sequences: B01, GTTTCGCTCC; B03, CATCCCCCTG; B04, GGAAGTGGAGT; B05, TGCGCCCTTC; B07, GGTGACGCAG; B08, GTCCACACGG; B10, CTGCTGGGAC; B11, GTAGACCCGT; B12, CCTTGACGCA; B13, TTCCCCCGCT; B15, GGAGGGTGTT; B16, TTTGCCCGGA; B17, AGGGAACGAG; B18, CCACAGCAGT; C02, GTGAGGCGTC; C04, CCGCATCTAC; C05, GATGACCGCC; C06, GAACGGACTC; C08, TGGACCGGTG; C09, CTCACCGTCC; and C13, AAGCCTCGTC. Amplification was conducted in a Cetus DNA thermalcycler (Perkin-Elmer, Norwalk, Conn.) for 45 cycles of 1 min at 94C, 1 min at 37C, and 1 min at 72C, followed by a final incubation at 72C for 7 min, using the most rapid transition time between temperatures. Products were analyzed on 1.4% agarose gels run at 6.4 V/cm in TAE buffer.

Statistical analysis. Ribosomal DNA restriction fragments and RAPD products were scored visually for presence or absence. Variation among genotypes was evaluated from pairwise comparisons of the proportion of shared fragments among samples, i.e., two times the number of shared fragments divided by the total number of fragments. This method was equivalent to calculating

Table 1. Accessions of *Theobroma* and *Herrania* used in this study.

Genus	Section	Species	Races	Genotypes
<i>Herrania</i>		<i>H. albiflora</i> Goudot		
		<i>H. sp.</i> (Colombia) ²		
<i>Theobroma</i>	Andropetalum	<i>T. mammosum</i> Cuatr. & León		
	Glossopetalum	<i>T. simiarum</i> Dann. Smith		
		<i>T. angustifolium</i> Mocino & Sessé		
		<i>T. grandiflorum</i> (Willd. & Spreng.) Schum.		
		<i>T. subincanum</i> Mart		
	Oreanthes	---		
	Rhytidocarpus	<i>T. bicolor</i> Humb. & Bonpl.		
	Telmatocarpus	<i>T. microcarpum</i> Mart.		
	Theobroma	<i>T. cacao</i> L.		
			Criollo ^y	
			Central American	BS 2; MEX 2; R 48; R 52; SGU 3
			South American	SC 5
			Forastero	
			Amazonian	
			Upper	APA 4; CSUL 7; IMC 67; PA 121; P 7; RB 29; SCA 6
			Lower	BE 4; MA 13; SIAL 70; SIAL 325; SIC 6
			Catongo ^y	
			Unknown	EET 59; EET 62
			Trinitario	CC 71; DR 1; ICS 29; ICS 39; ICS 135; SNK 12; TSH 792; UF 296; UF 613

²Species unknown, based on identification label of CATIE, Turrialba, Costa Rica.

^yHorticultural classification based on Cheesman (1944) and Toxopeus (1987). Used in RAPD analysis only.

Table 2. Brief description of cacao clones (End, 1991; Engels, 1981; Enriquez and Soria, 1967; Lockwood and Gyamfi, 1979; Pound, 1938, 1943; Soria, 1970).

Genotype	Place of origin	Pedigree	Color	
			Pod	Cotyledon
APA 4	Guaviare river, Colombia	Origin uncertain	Green	Purple
BE 4	Belem, Amazon, Brazil	Budwood, semi-wild plantings, unknown origin	Green	Purple
BS 2	Buenos Aires, Honduras	Origin uncertain	Green	White
Catongo	Urucuca, Bahia, Brazil	Seedling from albino mutant from Bahia, Brazil	Green	White
CC 71	Turrialba, Costa Rica	Selection from Criollo x Amazonian Matina population	Green	Purple
CSUL 7	Jurua River, Acre, Brazil	Seedling from wild trees	Green	Purple
DR 1	Djati Roengo, Indonesia	Selection from Java Criollo with Forastero	Red	White
EET 59	Los Rios, Ecuador	“Nacional” selection from farm at San Javier	Green	Purple
EET 62	Los Rios, Ecuador	Selection from Trinitario x “Nacional” at Porvenir	Green	Purple
ICS 29	River State, Trinidad	Selection from Trinitarios with Nicaraguan Criollo	Red	NA ²
ICS 39	River State, Trinidad	Selection from Trinitarios with Nicaraguan Criollo	Green	Purple
ICS 135	Trinidad	Selection from progeny of ICS 6 (Trinitario) x SCA 6	Green	Purple
IMC 67	Iquitos, Peru	Seedling from wild tree	Green	Purple
MA 13	Manaus, Amazon, Brazil	Budwood, semi-wild variable planting, unknown origin	Green	Purple
MEX 2	West Coast of Mexico	Unknown, but probably Forastero collected at Izapas	NA	NA
PA 121	Parinari, Peru	Seedling collected at Mara�on river, above Parinari	Green	Purple
P 7	Iquitos, Peru	Budwood from tree probably from Nanay river	Green	Purple
R 48	Chiapas, Mexico	Selection on Criollo population	Green/red	White
R 52	Chiapas, Mexico	Selection on Criollo population	Green	NA
RB 29	Rio Branco, Acre, Brazil	Budwood from wild trees	Green	Purple
SC 5	Palmira, Colombia	Origin uncertain	Red	White
SCA 6	Contamana, Peru	Seedling from wild tree at Ucayali river	Green	Purple
SGU 3	Brillantes, Guatemala	Selection from Matina (Amazon) x Criollo	Green	Purple
SIAL 70	Jussari, Bahia, Brazil	Selection on cultivated cacao	Green	Purple
SIAL 325	Buerarema, Bahia, Brazil	Selection on cultivated cacao	Green	Purple
SIC 6	Urucuca, Bahia, Brazil	Selection on cultivated cacao	Green	Purple
SNK 12	N’Koemvobe, Cameroon	Selection from African Trinitarios population	Red	Purple
TSH 792	Trinidad	Selection OP progeny of TSA 644 (SCA 6 x IMC 67)	Red	Purple
UF 296	Limon, Costa Rica	Selection from Trinitarios and “Nacional” population	Red	Purple
UF 613	Limon, Costa Rica	Selection from Trinitarios and “Nacional” population	Red	Purple

²Information not available to authors.

Nei and Li’s (1979) index of genetic similarity (S_{xy}) for RFLP comparisons. Relationships among samples were evaluated with a phenetic cluster analysis using the unweighted pair group method for arithmetic averages (UPGMA), and plotted in a phenogram. Statistical confidence intervals for the branching arrangement of the phenogram were determined using a bootstrap procedure (Efron and Gong, 1983) of 1000 repetitive samplings of the rDNA and RAPD data and computing the 95% range of the branch point values in the resultant UPGMA phenograms (modified after Nei et al., 1985). Confidence intervals were used to delimit clustering structure in the following fashion. All samples involved in branch points whose confidence interval included 100% similarity were judged to be not significantly different. Branch points whose confidence intervals did not include 100% similarity joined genotypes that were significantly different. Where confidence intervals overlapped, the branches involved joined genotypes whose average similarities were not significantly different. All computations and statistical analyses were performed using SAS programs.

Results

rDNA RFLP analysis. DNA from each of 29 *Theobroma cacao* genotypes, seven related *Theobroma* species, and two *Herrania* species was digested with four restriction enzymes (*EcoRV*, *EcoRI*, *DraI*, and *BamHI*). Hybridization with a flax rDNA probe

produced a total of 46 fragments. In *EcoRV* digests, the length of the hybridized fragment ranged from 10.6 to 12.3 Kb for *T. cacao* and 11.5 to 13.5 Kb in the related species (Table 3). These fragments probably represented the full length rDNA repeat (Jorgensen and Cluster, 1988). A *DraI* fragment of 5.0 Kb hybridized in all cacao genotypes (Fig. 2). In some genotypes, this was the only fragment that hybridized, and therefore *DraI* must digest the rDNA repeat of ≈ 10 Kb to produce two fragments of similar size. Smaller *DraI* fragments of between 3.5 and 4.2 Kb were detected in genotypes that also contained smaller rDNA repeats. *BamHI* digests revealed additional small variations in the size of the rDNA repeat, not detected with *DraI* and *EcoRV* (Table 3).

Overall, five size classes of rDNA repeat (*EcoRV* fragments of 12.3 Kb = A; 11.6 Kb = B; 11.0 Kb = C; 10.6 Kb = D; 10.5 Kb = E) were identified in *T. cacao*. The estimated rDNA repeat length should be the same for all restriction enzymes used for any genotype. The discrepancy in sizes resulted from either failure to detect small fragments or error in estimating the size of the restriction fragments. The largest rDNA repeat (A) was the predominant class size in *T. cacao*, and was present in all but one genotype (BE 4). There were variants of this major repeat, which might represent a small insertion (RB 29) or deletion (IMC 67, BS 2, TSH 792), and were only detected in digests with *BamHI* or *EcoRI*. The second largest rDNA repeat (B) was detected most frequently in association with A repeat, but also appeared alone

Fig. 1. Geographic origin of *Theobroma cacao* genotypes analyzed (after Cheesman, 1944).

(BE 4), or in the case of ICS 29, with a smaller size variant. The three smallest rDNA repeats each occurred in two genotypes. Only a single size of rDNA repeat was detected in the other *Theobroma* and *Herrania* species, except for *T. simiarum* in which two different sizes of rDNA repeats were found.

Genetic similarity of *Theobroma* and *Herrania* based on rDNA RFLPs. The average similarity among all *Theobroma* and *Herrania* species was 68%, ranging from 25% to 100%. Considering only *T. cacao*, similarity averaged 88%, ranging from 61% to 100%. The average similarity between *Herrania* and *T. cacao* was 39%, less than the 60% between *Herrania* and all the other *Theobroma* species (Table 4). The average similarity between *T. cacao* and the other *Theobroma* species was 46%, ranging from 34% (*T. bicolor*) to 53% (*T. mammosum*).

Bootstrap analysis divided the phenogram into two significantly different clusters. One cluster included all *Herrania* and all *Theobroma* species except *T. cacao*, and a second cluster included all *T. cacao* genotypes, except RB 29 (Fig. 3). In the *Theobroma*-*Herrania* cluster, *T. mammosum* and *T. subincanum* were not different from each other; all other species were significantly different. The extensive 95% confidence interval overlap in this cluster indicated that further statistical separation among this heterogeneous group of species was not possible. It is noteworthy that no clear distinction was observed between *Theobroma* and *Herrania*. There is a superficial subdivision in the *Theobroma*-*Herrania* cluster (Fig. 3). Subset 1 includes *T. bicolor* (section Rhytidocarpus), *T. microcarpum* (section Telmatocarpus), and *H. colombia*. The second subset contains *T. mammosum* (section Andropetalum), *T. subincanum*, *T. grandiflorum*, *T. angustifolium*, and *T. simiarum* (section Glossopetalum), and the *T. cacao* genotype RB 29 and *H. albiflora*.

The second cluster, which included only *T. cacao* genotypes, is a generally homogeneous group with several genotypes with no statistical differences. Extensive confidence interval overlap in this cluster indicated no meaningful further subdivision.

RAPD analysis. To further analyze the organization of genetic diversity in *Theobroma cacao*, DNA polymorphisms were identified with RAPDs. From the original 33 primers tested (primers B1 through B20 and C1 through C13 from Operon), 23 produced reproducible amplified DNA fragments and were used in this study. A total of 128 fragments were scored for presence or absence. With the 30 genotypes the primers produced an average of 5.7 ± 0.5 fragments per genotype ranging from 1 (primer B05) to 11 (primer B07) fragments. Of the 128 fragments scored, $\approx 80\%$ were polymorphic.

Genetic similarity in *Theobroma cacao* based on RAPD. The average similarity among all the cacao samples is 81%, ranging from 61% to 96%. All genotypes analyzed in the present study were significantly different from each other (Fig. 4). The bootstrap analysis indicated essentially continuous variation among the generally similar but heterogeneous genotypes (Fig. 4). In the continuum, a few genotypes could be statistically separated in terms of degrees of similarity. For example, SNK 12 and DR 1 were significantly more similar to one another than to any other genotype, except for R 52. The most meaningful separation revealed by bootstrap analysis was between a subset of the Upper Amazonian genotypes (CSUL 7, SCA 6, RB 29, P 7, and IMC 67), which differ either marginally ($P < 0.10$) or significantly ($P < 0.05$) from all others examined, apart from PA 121, SGU 3, and BE 4. The variability found in this subset, containing genotypes collected in the wild, was significantly higher than among most other

Table 3. Sizes of restriction enzyme fragments that hybridized to the flax rDNA probe.

Genotypes	Estimated rDNA repeat length (Kb)			
	Restriction enzymes			
	<i>EcoRV</i>	<i>DraI</i>	<i>BamHI</i>	<i>EcoRI</i>
<i>Theobroma cacao</i>				
RB 29	12.3	10.0	11.8	10.1
SIC 6, IMC 67, SIAL 70, CSUL 7, BS 2, UF 613, SNK 12	12.3	10.0	11.7	13.1
UF 296, SGU 3, EET 59, R 48, SIAL 325, SC 5, ICS 135, DR 1, EET 62 TSH 792, R 52, ICS 39, CC 71	12.3/11.6	10.0/9.2	11.7/11.0	13.1/12.6
BE 4	11.6	9.2	11.0	12.6
ICS 29	12.3/11.6/10.6	10.0/9.2/ 8.8	11.7/11.0/10.3	13.1/12.6/?
PA 121, MA 13	12.3/11.0	10.0/9.0	11.7/10.7	13.1/12.4
P 7, APA 4	12.3/10.6	10.0/8.8	11.7/10.3	13.1/12.2
SCA 6, MEX 2	12.3/10.5	10.0/8.5	---	---
<i>Herrania albiflora</i>	12.3	10.5	11.9	10.3
<i>H. sp. (Colombia)</i>	12.3	9.8	11.4	12.9
<i>T. bicolor</i>	11.5	9.8	11.4	13.1
<i>T. microcarpum</i>	12.3	10.0	11.9	10.3
<i>T. simiarum</i>	13.5/12.3	11.2/9.7	12.1	14.3/10.3
<i>T. angustifolium</i>	12.3	10.5	11.9	(12.6)
<i>T. grandiflorum</i>	12.3	10.1	11.6	13.2
<i>T. subincanum</i>	12.3	9.9	11.6	13.2
<i>T. mammosum</i>	12.3	9.9/9.5	11.5	nd

genotypes, which were derived from cultivated populations.

The phenogram constructed from RAPD data was arbitrarily subdivided into six subsets. Each subset groups together some genotypes consistent with the conventional classification of cacao (Fig. 4). Subset A includes all but one of the Upper Amazonian genotypes examined. PA 121, a genotype normally considered Upper Amazonian, is not present in subset A. Conversely, BE 4 and SGU 3, genotypes not typically considered to be of Upper Amazonian origin, are included. The great genetic diversity of subset A supports the hypothesis that the center of origin of cacao is located in the headwaters of the Amazon basin. Subset B also includes genotypes of diverse origin, primarily those believed to be of Ecuadorian and/or Trinitario background, e.g., UF 296 and EET 59. MEX 2 has an unknown origin and BS 2 may have some Criollo background. Subset C includes most of the genotypes considered to be Trinitarios, except for UF 296 and the hybrid TSH 792. Subset D is comprised mainly of genotypes considered to be Criollo (R 52, DR 1, ICS 29, and R 48) and/or red podded genotypes (SNK 12, DR 1, ICS 29, and R 48). However, another Criollo (BS 2) and other genotypes with red pods (UF 296, TSH 792, UF 613, and SC 5) are placed in other subsets. Subset E contains the Lower Amazonian Forastero selections from naturalized populations occurring in Bahia, Brazil. The origin of APA 4 is uncertain. Subset F contains genotypes with no obvious relationships.

Discussion

Considerable DNA sequence diversity could be detected in the ribosomal RNA genes of *Theobroma cacao* and related species. At least five classes of rDNA repeat length variants were detected in

cacao. The two largest rDNA repeat variants (Table 3) were present in 96% and 52% of the cacao genotypes tested. Other rDNA length variants were detected in cacao genotypes that originated in the Upper Amazon region (PA 121, P 7, and SCA 6). Laurent et al. (1993) evaluated 192 cacao genotypes and observed 15 rDNA repeat variants, but only two rDNA repeats accounted for 80% of the genotypes analyzed.

The analysis of only one genotype for each of the cacao-related species limited the phylogenetic assessments of species in genera. However, the genetic similarity observed between *Herrania* and *T. cacao* (39%) was only slightly less than between *T. cacao* and the other *Theobroma* species (46%) (Table 4). Additionally, the genetic similarity between *Herrania* species and *Theobroma* species (60%) was higher than between the two *Herrania* species (54%). These results suggest that *Herrania* and *Theobroma* may not be distinct genera. This conclusion is supported by the compatibility and seed viability found in crosses between *Herrania* species and *Theobroma*, including *T. cacao*, (Addison and Tavares, 1951). However, further molecular investigations using more genotypes from *Herrania* and *Theobroma* species are required before proposing a formal systematic reorganization.

The subdivision of *Theobroma* into sections based on morphological characters was loosely supported by rDNA polymorphisms. In the present study the *Theobroma-Herrania* cluster was subdivided into two arbitrary subsets. Subset 2 (Fig. 3) contained all genotypes from the section Glossopetalum and Andropetalum, considered to be the most primitive sections in *Theobroma* (Cuatrecasas, 1964). The section Andropetalum (*T. mammosum*) was initially considered to be in Glossopetalum by Bernoulli in 1869, but was separated by Cuatrecasas (1964). *Theobroma an-*

Fig. 2. Detection of rDNA sequences in *Theobroma* and *Herrania* DNA digested with *Dra*I and hybridized with flax rDNA repeat. The positions of DNA size standards are shown on the left.

gustifolium, *T. simiarum*, and *T. mammosum* are native to Central America, while *T. subincanum* and *T. grandiflorum* are native to eastern Amazonia. *Theobroma subincanum* has been considered the most ancient species of the genus, with a broad range of distribution (Cuatrecasas, 1964), and, based on rDNA polymorphism data, had the highest average genetic similarity with other *Theobroma* species. *Theobroma mammosum* was not significantly different from *T. subincanum* in this study (Fig. 3), and may represent a recent derivative.

The section *Theobroma*, containing only *Theobroma cacao*, was considered by Cuatrecasas (1964) to be more evolved than the

others. This section separated into a distinct rDNA cluster in the present study. There was no clear relationship between the separation of genotypes of *T. cacao* based on rDNA polymorphism and the conventional classification of the species into three major "horticultural races" (Criollo, Forastero, and Trinitario). Laurent et al. (1993) identified a number of rDNA polymorphisms in *T. cacao*. While some rDNA units were associated predominantly with either Criollo or Forastero types, there was significant overlap and rDNA genotype could not be used as a sole criterion for classification. The level of genetic similarity in *T. cacao* based on rDNA polymorphism (88%, ranging from 61% to 100%) was similar to levels obtained using RAPD markers (81%, ranging from 61% to 96%). However, the phenograms constructed from the rDNA and RAPD data are not similar. This discrepancy may reflect the nature of the DNA sequences examined. The rRNA genes are present in tandem repeats localized at a small number of sites in the genome (Jorgensen and Cluster, 1988; Polans et al., 1986). In contrast, the DNA sequences that give rise to RAPD markers are considered to be distributed throughout the genome (Williams et al., 1993), and may provide a more representative view of differences between genomes.

Classification of cacao types based on 39 morphological descriptors (Engels, 1986) also failed to separate genotypes into the conventional horticultural races. Isozyme polymorphisms were similarly unable to distinguish between various cacao populations (Lanaud, 1987). In the isozyme analysis, Upper Amazonian Forastero genotypes expressed the entire range of variation, and there was overlap between Lower Amazonian Forasteros, Criollos, and Trinitarios. The RAPD data presented here indicate that the conventional classification has a superficial rationale. However, statistical analysis indicates that there is no rigorous distinction among the so-called Criollo, Forastero, and Trinitario genotypes. Lerceteau et al. (1992) also concluded that no well-defined associations of cacao groups could be obtained using RAPD markers. Recently, Wilde et al. (1992) concluded that *Theobroma cacao* clones (actually open-pollinated seedlings) could be classified reliably into the traditional *Theobroma* groups (races) using RAPD markers. However, no statistical analysis of genetic distance data was presented in support of their conclusion. A phenogram constructed from their data does not confirm their conclusions, because no clear separation of cacao groups is apparent (Fig. 5). Further analysis of a collection of Forastero types collected from three distinct regions indicates that RAPD polymorphisms can be useful in distinguishing between cacao genotypes (Russell et al., 1993).

Table 4. Average genetic similarity among *Theobroma* and *Herrania* species.

Spp.	Average genetic similarities (%)								
	<i>H. albiflora</i>	<i>H. sp. (Colombia)</i>	<i>T. bicolor</i>	<i>T. microcarpum</i>	<i>T. simiarum</i>	<i>T. angustifolium</i>	<i>T. grandiflorum</i>	<i>T. subincanum</i>	<i>T. mammosum</i>
<i>Herrania</i> sp. (Colombia)	54	100							
<i>Theobroma</i>									
<i>bicolor</i>	42	67	100						
<i>microcarpum</i>	71	63	77	100					
<i>simiarum</i>	56	48	36	50	100				
<i>angustifolium</i>	70	60	44	67	52	100			
<i>grandiflorum</i>	61	61	50	71	64	76	100		
<i>subincanum</i>	64	64	42	59	75	70	87	100	
<i>mammosum</i>	64	54	38	63	72	67	78	95	100
<i>cacao</i>	34	46	34	45	44	46	49	51	53

Fig. 3. Phenogram generated using UPGMA clustering, demonstrating relationships among *Theobroma* and *Herrania* genotypes based on rDNA polymorphisms. Error bars denote 95% confidence interval of branch point based on bootstrap analysis.

Lack of distinction between the conventional groupings of cacao based on molecular markers indicates that the classification proposed by Cheesman (1944) is inappropriate. The existence of a real Criollo group is arguable. Cheesman (1944) thought that Criollos were already rare, and Mora (1958) recognized that Criollos, once cultivated in Costa Rica, Nicaragua, and Mexico, had been either extinguished or hybridized with Amazonian introductions. Cuatrecasas (1964) considered Criollo and Forastero to be subspecies; Criollo was designated *Theobroma cacao* ssp. *cacao*, and subdivided into four botanic forms, while Forastero was designated *T. cacao* ssp. *sphaerocarpum*. However, traditional taxonomy tends to overclassify cultivated species (Harlan, 1992). Large morphological alterations may occur in the organs most valued by humans without any substantial change in the genetic background. Harlan and de Wit (1971) have proposed an informal classification for cultivated crops, where the principal separation is between cultivated and wild or spontaneous groups. This is the only separation supported by the rDNA and RAPD analyses for *T. cacao* (Table 5). Two arbitrary subsets were identified in the *T. cacao* cluster by rDNA polymorphism. Subset 1 included 13 of the 15 cultivated cacaos, which contained the B rDNA repeat unit (AB, ABC, and B), while subset 2 included 7 of 13 wild cacaos, which lacked the B repeat (A, AC, AD, and AE).

Similarly, subset A identified by RAPD polymorphism contained only noncultivated cacaos.

On the basis of our data, we conclude that the traditional horticultural races cannot be defined by molecular markers. However, on the basis of geographic origin and morphology, cultivated cacaos can be conveniently subdivided into Amazonian and Circumcaribbean "races" (Table 5) as proposed by B.G.D. Bartley (unpublished). The cultivated Amazonian cacaos are exclusively green-podded, typically with purple cotyledons, while the Circumcaribbean cacaos are often red-podded, with white to pale violet cotyledons. Further subdivisions in each race depends on geographical and commercial history. For example, the cultivated Amazonian "races" can be conveniently subdivided into "Cacao Comum da Bahia" from Brazil, "Cacao Nacional" from Ecuador, and "amelonado" from West Africa. Similarly, in the Circumcaribbean cultivated cacaos, there are "Trinitarios" from Trinidad, Central America, Cameroon, Java, and the Philippines. The subgroup of wild or spontaneous cacaos from the Upper Amazon can be also subdivided into various populations, according to location of collection. Unfortunately, a complete and rational analysis of wild populations of *T. cacao* may be preempted by the destruction of natural flora in the Amazon, caused by the encroachment of colonization.

Fig. 4. Cacao phenogram generated using UPGMA clustering demonstrating the relationships among genotypes based on RAPD analysis. Error bars denote 95% confidence interval of branch point based on bootstrap analysis.

Literature Cited

- Addison, G.O. and R.M. Tavares. 1951. Observações sobre as espécies do gênero *Theobroma* que ocorrem na Amazônia. Bol. Téc. Inst. Agronomico Norte 25:3–20.
- Baker, R.E.D., F.W. Cope, P.C. Holliday, B.G. Bartley, and J. Taylor. 1953. The Anglo-Colombian cacao collecting expedition. In: H. Toxopeus (ed.). Reprint archives cocoa research. vol. 1. Amer. Cocoa Res. Inst.
- Barrau, J. 1979. Sur l'origine du cacaoyer, *Theobroma cacao* Linne, Sterculiacées. J. d'Agr. Trad. Bot. Applied 26(3–4):171–180.
- Bergmann, J.F. 1969. The distribution of cacao cultivation in pre-columbian America. Ann. Assn. Amer. Geographer 59:85–96.
- Cheesman, E.E. 1944. Notes on nomenclature, classification and possible relationships of cacao populations. Trop. Agr. 21(8):144–159.
- Cuatrecasas, J. 1964. Cacao and its allies. A taxonomic revision on the genus *Theobroma*. Cont. Natl. Herb (USA) 35(6):379–614.
- Efron, B. and G. Gong. 1983. A leisurely look at the bootstrap, the jackknife, and crossvalidation. Amer. Stat. 37:36–48.
- End, M. 1991. The international cocoa germplasm database. Software. Dept. Botany and Horticulture, Univ. Reading, U.K.
- Engels, J.M.M. 1981. Genetic resources of cacao: A catalogue of the CATIE collection. Tech. Bul. 7. Turrialba, Costa Rica.
- Engels, J.M.M. 1986. The systematic description of cacao clones and its significance for taxonomy and plant breeding. PhD diss., Agr. Univ., Wageningen, The Netherlands.
- Enriquez, G.A. and J.V. Soria. 1967. Cacao cultivars register. IICA, Turrialba, Costa Rica.
- Figueira, A., J. Janick, and P.B. Goldsbrough. 1992. Genome size and DNA polymorphism in *Theobroma cacao*. J. Amer. Soc. Hort. Sci. 117:673–677.
- Goldsbrough, P.B. and C.A. Cullis. 1981. Characterization of the genes for ribosomal RNA in flax. Nucleic Acids Res. 9:1301–1309.
- Goldsbrough, P.B., E.M. Hatch, B. Huang, W.G. Kosinski, W.E. Dyer, K.M. Herrmann, and S.C. Weller. 1990. Gene amplification in glyphosate tolerant tobacco cells. Plant Sci. 72:53–62.
- Harlan, J.R. 1992. Crops and man. Amer. Soc. Agron. and Crop Sci. Soc. Amer., Madison, Wis.
- Harlan, J.R. and J.M.J. de Wit. 1971. Toward a rational classification of cultivated plants. Taxon 20:509–517.
- Jorgensen, R.A. and P.D. Cluster. 1988. Modes and tempos in the evolution of nuclear ribosomal DNA: new characters for evolutionary studies and new markers for genetic and population studies. Ann. J. AMER. SOC. HORT. SCI. 119(5):1073–1082. 1994.

Fig. 5. Phenogram generated based on Nei's estimate of similarity data presented by Wilde et al. (1992). F = Forastero; C = Criollo; T = Trinitario.

Table 5. Proposed classification of *Theobroma cacao*.

Group	Genotypes evaluated
Cultivated cacaos	
Amazonian	
Brazilian (comum da Bahia, amelonado)	SIAL 70; SIAL 325; SIC 6; Catongo
Costa Rican (Matina)	
Ecuadorian (Nacional)	EET 59; EET 62
Suriname amelonado	
West African amelonado	
Unknown	MA 13; APA 4
Circumcaribbean	
Trinidad (Trinitario)	ICS 29; ICS 39; ICS 135; TSH 792
Costa Rican (Trinitario)	CC 71; UF 296; UF 613
Mexican (Trinitario)	BS 2; R 48; R 52; MEX 2
Cameroon (African Trinitario)	SNK 12
Indonesian (Java Criollo)	DR 1
Wild cacaos	
Colombian	
Ecuadorian	
Peruvian	
Iquitos	IMC 67
Nanays	P 7
Ucayali	SCA 6
Parinari	PA 121
Brazilian	
Acre	CSUL 7; RB 29

Missouri Bot. Gard. 75:1238–1247.
Lanaud, C. 1987. Nouvelles donnés sur la biologie du cacaoyer (*Theobroma cacao* L.): diversité des populations, système d'incompatibilité, haploïdes spontanés. Leurs conséquence pour l'amélioration génétique de cette espèce. PhD diss., Université de Paris Sud, Centre d'Orsay, France.
Laurent, V., A.M. Risterucci, and C. Lanaud. 1993. Variability for nuclear ribosomal genes within *Theobroma cacao*. Heredity 71:96–103.
Lerceteau, E., D. Crouzillat, and V. Petiard. 1992. Use of random amplified polymorphic DNA (RAPD) and restriction length polymor-

phism (RFLP) to evaluate genetic variability within the *Theobroma* genus. Intl. Cocoa Genetic: Resource Workshop, Port-of-Spain, Trinidad Tobago, 13–17 Sept. 1992.
Lockwood, G. and M.M.O. Gyamfi. 1979. The CRIG germplasm collection with notes on codes used in the breeding programme at Tafo and elsewhere. Tech. Bul. 10, Cocoa Research Institute, Tafo, Ghana.
Muñoz Ortega, J.M. 1948. Estudios cromosómicos en el género *Theobroma* L. M. Agr. Diss., IICA, Turrialba, Costa Rica.
Nei, M. and W.H. Li. 1979. Mathematical model for studying genetic

- variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. (USA)* 76:5269–5273.
- Nei, M., J. Stephens, and N. Saitou. 1985. Methods for computing the standard errors of branching points in an evolutionary tree and their application to molecular data from humans and apes. *Mol. Biol. Evolution* 2:66–85.
- Polans, N.O., N.F. Weeden, and W.F. Thompson. 1986. Distribution, inheritance and linkage relationships of ribosomal DNA spacer length variants in pea. *Theoretical Applied Genet.* 72:289–295.
- Pound, F.J. 1938. Cacao and witchbroom disease in South America. In: H. Toxopeus (ed.). *Reprint archives cocoa research*. vol. 1. Amer. Cocoa Res. Inst.
- Pound, F.J. 1943. Cacao and witches' broom disease (*Marasmius perniciosus*). In: H. Toxopeus (ed.). *Reprint archives cocoa research*. vol. 1. Amer. Cocoa Res. Inst.
- Purseglove, J.W. 1968. *Tropical crops, Dicotyledons 2*. Wiley, New York.
- Russell, J.R., F. Hosein, E. Johnson, R. Waugh, and W. Powell. 1993. Genetic differentiation of cacao (*Theobroma cacao* L.) populations revealed by RAPD analysis. *Mol. Ecol.* 2:89–97.
- Schultes, R.E. 1958. A synopsis of the genus *Herrania*. *J. Arnold Arboretum* 39:216–278.
- Soria, J.V. 1970. The latest cocoa expeditions to the Amazon basin. *Cacao* 15:5–15.
- Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503–517.
- Toxopeus, H. 1987. Botany, types and populations, p. 11–37. In: G.A.R. Wood and R.A. Lass (eds.). *Cocoa*. Longman, London.
- Wilde, J., R. Waugh, and W. Powell. 1992. Genetic fingerprinting of *Theobroma* clones using randomly amplified polymorphic DNA markers. *Theoretical Applied Genet.* 83:871–877.
- Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski, and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18:6531–6535.
- Williams, J.G.K., J.A. Rafalski, and S.V. Tingey. 1992. Genetic analysis using RAPD markers. *Meth. Enzymol.* 218:704–740.