Theor Appl Genet (1998) 96: 621–627

R. Whitkus · M. de la Cruz · L. Mota-Bravo
A. Gómez-Pompa

Genetic diversity and relationships of cacao (*Theobroma cacao* L.) in southern Mexico

Received: 21 May 1997/ Accepted: 9 October 1997

Abstract Neotropical tree crops are affected by a combination of biological and human factors that complicate the study of genetic diversity and crop evolution. Genetic diversity and relationships among southern Mexican populations and horticultural collections of *Theobroma cacao* (chocolate, cocoa, cacao) are examined in light of the agricultural practices of the Maya. Collections of cacao were obtained from the extremes of its geographic range including archeological sites in southern Mexico where cacao was first domesticated. Genetic diversity was assayed by 57 informative random amplified polymorphic DNA (RAPD) marker loci. A unique sample of the total diversity found in this study exists in the southern Mexican populations. These populations are significantly different from all other cacao with regards to their profile of RAPD bands, including the 'criollo' variety, their morphological and geographical group. A population of cacao found in a sinkhole (cenote) in northern Yucatan with genetic affinities to populations in Chiapas suggests the Maya maintained plants far away from their native habitat. This finding concurs with known agroforestry practices of the Maya. Modern efforts to increase germplasm of tropical tree crops such as cacao should carefully examine archeological sites where genetic diversity, either deliberately or by chance, was collected and maintained by ancient cultures.

Key words Cacao · *Theobroma cacao* · Genetic diversity · Crop evolution · RAPD

Communicated by G. E. Hart

R. Whitkus (✉) · A. Gómez-Pompa
Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, USA
Fax: (909) 787–4437
E-mail: whitkus@moe.ucr.edu

M. de la Cruz
Instituto de Ecologia, UNAM, Apdo. Post. 70-275, C.P. 04510, Mexico

L. Mota-Bravo
ITESM Campus Morelos, Av. Refoma 182-A, Cuernavaca, Morelos 62589, Mexico

Introduction

The genetic diversity of Neotropical tree crops is influenced by a combination of biological attributes and human-precipitated changes. Some intrinsic characteristics of the plants that directly affect genetic variability are extensive geographic ranges of numerous species, fairly small populations sizes, and outcrossing breeding systems. Historical evidence indicates a long history of management of species by native cultures (Gómez-Pompa and Kaus 1990), while the greatest modern human factor is extensive deforestation of the region. These factors make it difficult to identify wild progenitors, locate centers of diversity, decipher the history of the domestication process, and conserve genetic diversity of important tropical crops.

One of the more important Neotropical tree crops is *Theobroma cacao* L. (chocolate, cacao), ranked 29th in planted hectares of world crops (N. Ellstrand, personal communication). Cacao is an understory tree in tropical rainforests extending from the Amazonian basin of South America to southern Mexico. Plants from Mesoamerica (Mexico and Central America) are classified as *T. cacao* subsp. *cacao* (Cuatrecasas 1964). Cultivated forms of the subspecies represent the horticultural variety 'criollo', considered to have been domesticated by the Maya (or their ancestors) more than 2000 years ago. South American plants are placed in *T. cacao* subsp. *sphaerocarpum* (Cuatrecasas 1964), and its cultivars represent the horticultural variety 'forastero'. This variety is the basis of most chocolate production because of its higher yield and greater disease

resistance. A third horticultural form, 'trinitario', represents a hybrid derivative between 'criollo' and 'forastero' (Cheesman 1944; Kennedy 1995).

As in other crops, a small portion of the genetic variability in cacao has been used in breeding programs, but efforts are underway to increase germplasm diversity (Lockwood 1985; Kennedy 1995). The search for additional genetic diversity is concentrated on the upper Amazonian basin because the region is considered the center of origin for the species (Cheesman 1944) and has the highest known level of diversity (Laurent et al. 1993a, 1994; Figueira et al. 1994; N'Goran et al. 1994). Mesoamerica is not being intensely searched since wild forms of cacao are considered rare or non-existent in the region (Cheesman 1944; Purseglove 1974). The discovery of *T. cacao* subsp. *cacao* occurring naturally in the Lacandon forest of Chiapas and sinkholes (cenotes) of northern Yucatan, Mexico (Gómez-Pompa et al. 1990) suggests that novel genetic variation may exist within these populations which could be distinct from South American plants (Cheesman 1944; Lockwood 1985).

A preliminary analysis of cacao from Chiapas and Yucatan found that the plants are genetically distinctive from all other cacao, including the 'criollo' and 'forastero' varieties (de la Cruz et al. 1995a). The lack of population samples from Mexico prevented a comparison of the organization of diversity in these collections with that of the cultivated varieties or South American native plants. The study presented here compared genetic diversity in natural populations of cacao from southern Mexico with cacao from South America and the horticultural varieties, and re-examined the relationships of the subspecies and horticultural varieties. Findings from this study indicate future efforts in collecting natural diversity of cacao should include Mesoamerica since the region contains a unique source of genetic variation.

Materials and methods

Plant material was obtained from trees growing in southern Mexico or in germplasm banks. Collections in southern Mexico were made in the states of Yucatan, Chiapas, and Tabasco. The Yucatan collection consists of 5 individuals from a cenote near the village of Yaxcaba in central Yucatan. Although the surrounding area is a seasonal tropical forest with alternating wet and dry seasons, cenotes in this region contain vegetation typical of mesic tropical forests and a number of plants that were likely cultivated by the Maya (Gómez-Pompa et al. 1990). Two naturally occurring populations were collected in Chiapas, one of 6 individuals near the Maya ruins at Bonampak and a second of 26 individuals along the Lacantun River in eastern Chiapas, approximately 120 km NE of Comitan. The three main cultivars ('criollo', 'forastero', and 'trinitario') were included in this study, with a concentration on criollo since this variety is considered to be derived from Mesoamerican plants (Cuatrecasas 1964). Criollos were collected from plantations at Chajul in the state of Chiapas (27 individuals) and near Villahermosa (9 individuals) and Comacalco in Tabasco (2 individuals). Additional criollos (4 clones) as well as forasteros (6 clones), South American wilds (6 clones), and trinitarios (4 clones) were obtained from the USDA germplasm collection in Mayaguez, Puerto Rico (clone identification list available upon request).

Table 1 Polymorphism of cacao collections based on 57 polymorphic RAPD loci (*ID* groups · *n* sample size · *P* percentage polymorphic bands · *U* total unique bands)

Collection	ID	*n*	P	U
Chiapas	1	28	12.3	5
Yucatan	2	5	8.8	5
Cultivars	3	48	73.4	11
South America	4	5	33.3	0
Southern Mexico	1&2	33	49.1	12
Cultivars and South America	3&4	53	77.2	29

Collections were organized into a hierarchy (Table 1) that reflected geographic origin, cultivation status, and relationships suggested in a previous analysis (de la Cruz et al. 1995a). In the first hierarchical level, four groups represented the southern Mexican populations of Chiapas (ID1) and Yucatan (ID2), the cultivars (ID3), and South American native plants (ID4). The second level in the hierarchy combined the southern Mexican populations (ID1 and 2) and the cultivars and South American natives (ID3 and 4).

Leaves of individual trees were placed into plastic bags and stored on dry ice in the field for transportation and at $-80\,^{\circ}C$ in the laboratory. Total genomic DNA extraction and random amplified polymorphic DNA (RAPD) band amplification followed the procedure of de la Cruz et al. (1995b). Our earlier study of cacao (de la Cruz et al. 1995a) showed 13 decamer primers (Operon Technologies, Almeda, Calif.) which give strong, reliable banding patterns.

Data were recorded from pictures of ethidium-stained, 2% agarose gels run in TRIS-acetate buffer (Sambrook et al. 1989) in which amplification products were separated. Pictures were examined for strong, clearly defined bands. Bands were identified by their molecular sizes relative to a 123-bp marker ladder on each gel. Each band was scored across all individuals as either present or absent. Individuals that did not give clear, easily scored bands for 90% or more of the total number of bands were not included in the analysis. Bands that were not clear in an individual were scored as missing data.

An estimate of the percentage of the nuclear genome covered by the markers was calculated (Beckmann and Soller 1983) with adjustments for chromosome ends (Lange and Boehnke 1982). Each marker was assumed to represent an independent locus, and the cacao genome length was taken as 1200 centiMorgans (cM), following Crouzillat et al. (1996). A maximum distance of 20 cM between a marker and any genomic region was used in the calculation.

Pair-wise correlations between different RAPD bands were determined by Spearman rank correlation. Bands that formed groups of perfect correlation (1 or -1) were removed, save for one to represent the data of the group. This procedure insures that a single locus was sampled.

RAPD bands are typically inherited as dominant Mendelian factors (Williams et al. 1990; Ronning et al. 1995), and several methods are available for calculating genetic diversity. Assuming that RAPD loci are in Hardy-Weinberg equilibrium, population frequencies of the two alleles (presence vs. absence) are estimated. Standard population genetic statistics can be calculated, although RAPDs tend to produce biased results (Lynch and Milligan 1994), or an analysis of variance of band frequencies provides an assessment of diversity apportionment (Stewart and Excoffier 1996). An alternative that does not rely on Hardy-Weinberg equilibrium calculates diversity on the basis of band phenotypes and assumes that the resulting estimate of phenotypic diversity approximates genetic diversity (Lewontin 1972; King and Schaal 1989). Diversity estimates obtained with this approach are downwardly biased with increased

outcrossing. Since cacao is primarily an outcrossing species (Knight and Rogers 1955) the diversity estimates based on band phenotypes will not accurately reflect true genetic diversity. However, we aim to compare diversity levels between groups; an approach facilitated by assuming that the bias is randomly distributed over populations.

In the present study, RAPD band diversity was calculated on polymorphic loci with the Shannon-Weaver information statistic using the Brillouin formula to eliminate the bias of finite sample sizes (Peet 1974). For each band: $H = (1/N)(N!/\prod_{i=1,2} n_i!)$, where N is the total number of individuals scored for the band, and n_i the number of individuals in the alternative band categories ($n_{present} + n_{absent} = N$). Calculation of diversity is facilitated by conversion of the Brillouin formula to logarithms, $H = (1/N)(\log N! - \Sigma \log n_i!)$ and the use of a table of logarithms of factorials (Lloyd et al. 1968).

Diversity was estimated for the primary (H_p) and secondary (H_s) hierarchical groups of collections, and over all collections (H_t). Diversity within a group of collections is the average H over all RAPD bands: $H = (1/M)\Sigma_{i=1}^{M} H_i$, where H_i is the diversity at each band and M is the total number of bands. Apportionment of diversity within and between groups follows Lewontin (1972). Average diversity within the lowest two hierarchical levels is calculated first ($\bar{H}_p$ and $\bar{H}_s$), and the proportion of diversity within and between levels is compared. The proportion of diversity within the first level is $\bar{H}_p/H_s$, and in the second level is $\bar{H}_s/H_t$. The proportion of diversity between groups in the first level is $(H_s - \bar{H}_p)/H_s$, and in the second level is $(H_t - \bar{H}_s)/H_t$. The relative contribution of each group to the total diversity is their ratio: H_p/H_t and H_s/H_t.

A significance test of between-group diversity in each hierarchical level was performed with Fisher's exact test (Fisher 1935). The comparison consisted of a 2×2 contingency table for each RAPD band with the number of presences and absences in paired cells for each group of collections. The probability that the observed distribution was not significantly greater than expected if the two groups of collections were samples of a single population was computed. The significance values obtained for individual band phenotypes were combined using Fisher's combined probability test (Fisher 1932) to give a one-tailed Chi-square value.

Relationships among collections were summarized in a neighbor-joining tree (Saitou and Nei 1987). Band phenotypes for each individual were converted to a dinucleotide sequence of **a** for band present and **t** for band absent. This conversion allowed us to import the data into the Mega program (Kumar et al. 1993) and calculate distances among individuals as the proportion of shared **a**'s and **t**'s subtracted from 1. Pair-wise distances were used to construct the neighbor-joining tree. One thousand bootstrap resamplings over band phenotypes in the original data provided support values for branches in the tree.

Results

We obtained 90 bands of different molecular weight with 13 primers for 95 individuals (primer, number of bands: A01, 7; A09, 3; F03, 11; F04, 6; F05, 15; F06, 3; G02, 3; G03, 5; G04, 3; G05, 4; G06, 6; G16, 15; G17, 9). Nine individuals were excluded from further analysis because of failure to obtain information for 90% or more of the 90 bands. In the remainder of the dataset missing data were recorded for 36 of 7740 data points (0.5% of total). The maximum number of missing data points was 7 for an individual and 5 for a band.

Spearman rank correlation analysis reduced the dataset to 57 polymorphic markers that lacked perfect correlation. A total of ten correlation groups was obtained. Seven groups were comprised of correlations among bands within a primer, and three groups comprised of correlation among bands between 2 different primers. Correlations among bands could be a result of internal priming of repetitive sequences, amplification of the same locus by different primers, codominant alleles, or close linkage among marker loci. Southern hybridization experiments have indicated internal priming in one case, linkage in two cases, and codominance in a final experiment (de la Cruz and Whitkus, unpublished data). Without complete information, a conservative approach of deleting perfectly correlated bands reduces the chance of resampling the same locus and thereby downwardly biasing diversity estimates. If we assume that all remaining RAPD bands represent random, independent loci, the reduced data set of 57 bands potentially covers 90% of the genome with a 20-cM distance between any marker and region of the genome.

Table 2 Diversity analysis of cacao collections based on 57 polymorphic RAPD loci. First hierarchical level = A, second hierarchical level = B. Group designations (ID) follow Table 1

A

ID	H_p	$\bar{H}_p$	H_s	$\bar{H}_p/H_s$	$(H_s - \bar{H}_p)/H_s$	H_p/H_t
1	0.01					0.05
2	0.01	0.01	0.07	0.14	0.86*	0.05
3	0.12					0.63
4	0.06	0.11	0.13	0.85	0.15*	0.32

B

ID	H_s	$\bar{H}_s$	H_t	$\bar{H}_s/H_t$	$(H_t - \bar{H}_s)/H_t$	H_s/H_t
1&2	0.07					0.37
3&4	0.13	0.11	0.19	0.58	0.42*	0.68

* $P \leqslant 0.0005$

RAPD band polymorphism levels are lowest in the collection from Yucatan and highest in the cultivars (Table 1). As a group, the southern Mexican collections are less polymorphic than the cultivars and the South American collections. A similar pattern is revealed by the number of unique bands.

Average within-group diversity in the first hierarchical level ($\bar{H}_p$) is lower in the southern Mexican collections (0.01) than the cultivars and South American natives (0.11; Table 2A). The apportionment of diversity is also quite different. In the southern Mexican collections, the proportion of diversity found between Chiapas and Yucatan is larger than that within each region (0.86 vs. 0.14, respectively). The pattern is reversed in the cultivars and the South American plants which carry a larger component of their diversity within their respective groups (0.85) rather than between (0.15). Despite this difference in the patterns of diver-

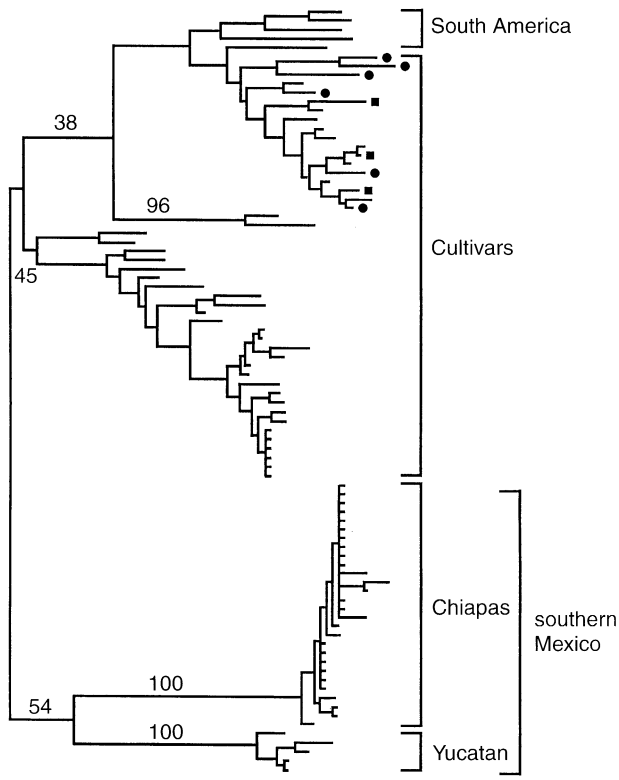

Fig. 1 Unrooted neighbor-joining tree of 86 cacao collections based on 57 RAPD marker loci. Terminal branch labels are given for 'forastero' (●) and 'trinitario' (■); all remaining cultivars are 'criollo'. All terminal branches belong to groups designated on *right side* of figure. Bootstrap support values of $> 90\%$ are given on their respective branches. Additional bootstrap support values are given for major branches as discussed in text

sity, significant divergence is seen for each group ($P < 0.0005$, Table 2A).

In the second hierarchical level (Table 2B), a larger amount of diversity is seen within groups than between groups (0.58 vs. 0.42, respectively). The significant value obtained for the between-group proportion indicates substantial differentiation.

The proportion of the total diversity in groups of collections reflects the level of polymorphism seen in Table 1. In the first hierarchical level, the highest diversity occurs in the cultivars and the lowest in the two southern Mexican regions. Surprisingly, the South American plants were represented by only 5 individuals yet contain over 30% of the total diversity found in this study. This finding concurs with previous molecular analyses that have shown a high level of diversity in collections from South America, especially the Upper Amazon basis (Laurent et al. 1993a, 1994; Figueira et al. 1994; N'Goran et al. 1994). The cultivars contain the greatest proportion of the total diversity observed in this study (0.63; Table 2A), possibly as a consequence of the larger sample size. In the second hierarchical level, the cultivars and South American plants carry 68% and the southern Mexican populations carry 37% of the total diversity found in this study (Table 2B).

Relationships among individuals, based on pair-wise distance (Fig. 1), agree with the earlier study of de la Cruz et al. (1995a). The South American plants group with the cultivars, whereas the southern Mexican populations are distinct and cluster within their respective geographic region. Within the cultivars, two major groups are found. The first consists of a mixture of all three cultivars that shares a branch with the South American plants. A cluster of two criollos is a sister group to this cluster. The second major group consists of criollos from Chiapas and Tabasco. As the majority of the collections used for the cultivars are representatives of the 'criollo' type, this clustering suggests a fair amount of diversity within the variety (Laurent 1993b; Lerceteau et al. 1997). The two major clusters of cultivars and South American plants are not well supported by the data, however, with bootstrap values of 38% and 45% (Fig. 1). The cluster of the southern Mexican populations is supported by a bootstrap value of 54%, and the branches leading to the states of Yucatan and Chiapas have 100% support values (Fig. 1). A pair of criollos from Comacalco, Tabasco, have the remaining highest support value in the neighbor-joining tree (96%).

Discussion

Genetic diversity in cacao has been examined with isozymes (Ronning and Schnell 1994; Warren 1994), mitochondrial and chloroplast DNA (Laurent et al. 1993b), nuclear restriction fragment length polymorphisms (Laurent et al. 1993a, 1994; Figueira et al. 1994; N'Goran et al. 1994; Lerceteau et al. 1997) and RAPDs (Russell et al. 1993; Figueira et al. 1994; N'Goran et al. 1994; Lerceteau et al. 1997). The pattern that emerges from many of these analyses is a distinction between the 'criollo' and 'forastero' varieties (Laurent et al. 1993a,b; Russell et al. 1993; N'Goran et al. 1994; Ronning and Schnell 1994; but compare Figueira et al. 1994 and Lerceteau et al. 1997). A second outcome of previous studies is that the greatest diversity in cacao is found in South American plants from the upper Amazonian basin (Laurent et al. 1993a, 1994; Russell et al. 1993; N'Goran et al. 1994). The higher genetic diversity of plants from the upper Amazonian region, along with their greater morphological diversity and disease resistance, has supported the idea that this region is a center of diversity and possibly the center of origin for the species (Cheesman 1944). Previously noted distinctions between 'criollo' and 'forastero' varieties, however, argues for an early differentiation of the two forms, supporting the hypothesis of Cuatrecasas (1964) of natural differentiation between South American and Mesoamerican cacao (Laurent et al. 1993a,b, 1994).

The previous cited studies compared cultivated forms of cacao ('criollo' and 'forastero') or cultivated forms and South American collections from the wild. The present study is the first to examine genetic diversity in cacao that has included samples of southern Mexican plants collected from natural populations. Our findings indicate these populations represent a unique segment of genetic diversity in cacao. The southern Mexican populations carry 12 unique alleles (Table 1), cluster as a separate group (Fig. 1), and are significantly differentiated from all other cacao (Table 2B). These findings add a new dimension to questions on the natural distribution of cacao and the origin of the domesticated crop.

Two hypotheses exist on the origin and distribution of domesticated cacao. One is that wild populations of cacao (*T. cacao*) existed exclusively in South America. The 'criollo' predecessor was transported by humans from the upper Amazonian basin to Mesoamerica where it was eventually domesticated by the Maya (or their ancestors) more than 2000 years ago (Cheesman 1944; Schultes 1984; Young 1994; Coe and Coe 1996). The diversity of cacao in the upper Amazon (Cheesman 1944; Laurent et al. 1993a, 1994; Figueira et al. 1994; N'Goran et al. 1994), the long history of cacao cultivation in Mesoamerica, and traditional doubts about native cacao in the region support this scenario (Purseglove 1974).

An alternative hypothesis (Cuatrecasas 1964) states that cacao had a natural geographic distribution from the Amazonian region to southern Mexico, as is common for many Neotropical tree genera. Differentiation in the wild produced the two recognized subspecies, *T. cacao* subsp. *cacao* in Mesoamerica and *T. cacao* subsp. *sphaerocarpum* in South America. Independent domestication from these subspecies produced the two main cultivated varieties: 'criollo' in Mesoamerica and 'forastero' in South America (Cuatrecasas 1964). Evidence for this view is based on the discovery of presumptively wild *T. cacao* subsp. *cacao* in the Lacandon forest of Chiapas (Cuatrecasas 1964), some genetic marker-based studies indicating that criollos and forasteros can be distinguished (Laurent et al. 1993a; N'Goran et al. 1994; Ronning and Schnell 1994), and a preliminary study of southern Mexican collections of *T. cacao* subsp. *cacao* showing that these plants are highly differentiated from cultivated and South American plants (de la Cruz et al. 1995a).

The results from our study vary with one expectation from previous hypotheses of cacao relationships. According to both Cuatrecasas (1964) and Cheesman (1944), all cacao in Mesoamerica is related to the 'criollo' cultivar. Yet despite a wide sampling of criollos in this study, the 'criollo' horticultural variety exhibits less affinity to the plants from southern Mexico than to the 'forastero' horticultural variety and the South American collections (Fig. 1). The distinction seen in the neighbor-joining analysis between the southern Mexican populations and 'criollo' is maintained, with the same bootstrap values for the southern Mexican populations, when the analysis is repeated for just these two groups (results not shown). Although distinct in the genetic analyses, the southern Mexican populations are morphologically similar to the 'criollo' variety in lacking the formation of the jorquett, having red, elongated pods with ten ridges, and white to pale-pink seeds (Cheesman, 1944; Cuatrecasas, 1964; Toxopeus, 1985). Finally, the plants from Chiapas occur in the native rainforest without evidence of cultivation or management. This combination of features for the southern Mexican plants suggests that the Chiapas populations may represent wild cacao and support the contention of Cuatrecasas (1964) that cacao occurs naturally in Mesoamerica.

Significant genetic diversity and differentiation is seen between the Chiapas and Yucatan collections. Differentiation of the Yucatan collections may have arisen by founder effect in an isolated population, where the cenote represents a refugial patch of a previously extensive mesic tropical forest. The current vegetation of the central and northern Yucatan is seasonal tropical forest, unsuited for the growth of a rain forest species such as cacao, and the closest available rainforest is approximately 300 km to the south. Paleoclimatic and microfossil evidence indicate that northern and central Yucatan did not have extensive moist, tropical forests during the Holocene and that the lowland mesic tropical forests of the southern Yucatan peninsula were established only approximately 9000 years before the present time (Leyden et al. 1994, 1996; Islebe et al. 1996). Without evidence of mesic rainforests in northern and central Yucatan, an alternative explanation for the observed differentiation between Chiapas and Yucatan plants would be that the cenote population represents a human-mediated introduction into northern Yucatan.

Cacao was cultivated by the Maya civilization in a sophisticated agroforestry system (Gómez-Pompa and Kaus 1990) from southern Mexico to El Salvador and Honduras (Bergmann 1969). Originally grown in wet tropical rain forest areas (the habitat of wild cacao), cacao was eventually introduced to regions which required specialized knowledge for its cultivation and irrigation during dry periods (Bergmann 1969). Evidence of the diversity of the plants grown at that time is seen in an early description of four varieties of cacao in Mexico at the end of the 16th century (Hernández 1514–1587). One intriguing description is of a variety with reddish seeds (Sahagún 1590), which are typical of the 'forastero' variety, whereas the 'criollo' variety is characterized by white seeds (Cheesman 1944). The most interesting sites known of ancient cacao cultivation were in cenotes in the Yucatan called 'sacred groves' (Gómez-Pompa et al. 1990). Plants of the sacred groves, including cacao, would have been transported from other regions and maintained as introduction

gardens, similar to those described for Central Mexico (Maldonado-Koerdell 1941). If the plants growing in the cenotes today are descendants of those transported and managed by the Maya, then the plants may represent the closest living material of the primitive cultivars of cacao. In addition, the divergence of the Yucatan collection from populations in Chiapas indicates the Maya, either deliberately or by chance, sampled and maintained a unique portion of diversity of cacao in their territory.

The maintenance of plant diversity appears to have been an agricultural strategy employed by ancient civilizations to enhance production and control pests and diseases. Brush (1986) stated that the preservation of different land races and cultivars was a widespread practice of traditional agriculturists, and the ancient Mesoamericans may have attempted to do the same thing with cacao and other tropical crops. Our results illustrate that novel diversity in widespread Neotropical trees may be revealed by sampling near archeological sites (Peters 1983). Genetic diversity conservation in cacao is currently focusing on South America for new germplasm (Lockwood 1985). In light of the new genetic diversity found in this study, and archeological and historical evidence indicating the maintenance of diversity in cacao in Mesoamerica, we recommend that future efforts should include Mexico and Central America.

Acknowledgements N. C. Ellstrand, B. B. Simpson, J. F. Wendel, and G. A. Zentmyer are gratefully acknowledged for their comments on an early version of the manuscript. This work was supported by a UC Mexus grant to R.W. and A.G.-P., by the University of California Agricultural Experiment Station funds to R.W. and A.G.-P. and a CONACYT grant to M.C.

References

Beckmann JS, Soller M (1983) Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. Theor Appl Genet 67 : 35–43

Bergmann JF (1969) The distribution of cacao cultivation in pre-Columbian America. Ann Assoc Am Geogr 59 : 85–96

Brush SB (1986) Genetic diversity and conservation in traditional farming systems. J Ethnobiol 6 : 151–167

Cheesman EE (1944) Notes on the nomenclature, classification and possible relationships of cacao populations. Trop Agric 21 : 144–159

Coe SD, Coe MD (1996) The true history of chocolate. Thames and Hudson, London

Crouzillat D, Lerceteau E, Petiard V, Morera J, Rodriguez H, Walker D, Phillips W, Ronning C, Schnell R, Osei J, Fritz P (1996) *Theobroma cacao* L.: a genetic linkage map and quantitative trait loci analysis. Theor Appl Genet 93 : 205–214

Cuatrecasas J (1964) Cacao and its allies: a taxonomic revision of the genus *Theobroma*. Contrib US Natl Herb 35 : 379–614

de la Cruz M, Whitkus R, Gómez-Pompa A, Mota-Bravo L (1995a) Origins of cacao cultivation. Nature 375 : 542–543

de la Cruz M, Whitkus R, Mota-Bravo L (1995b) Tropical tree DNA isolation and amplification. Mol Ecol 4 : 787–789

Figueira A, Janick J, Levy M, Goldsbrough P (1994) Reexamining the classification of *Theobroma cacao* L. using molecular markers. J Am Soc Hortic Sci 119 : 1073–1082

Fisher RA (1932) Statistical methods for research workers, 4th edn. Oliver and Boyd, Edinburgh

Fisher RA (1935) The logic of inductive inference. J R Stat Soc 98 : 39–54

Gómez-Pompa A, Kaus A (1990) Traditional management of tropical forests in Mexico. In: Anderson AB (ed) Alternatives to deforestation: steps toward sustainable use of the Amazon rain forest. Columbia University Press, New York, pp 45–64

Gómez-Pompa A, Flores JS, Fernández MA (1990) The sacred cacao groves of the Maya. Lat Am Antiq 1 : 247–257

Hernández F (1514–1587) Historia de las plantas de Nueva España, Tomo III. Imprenta Universitaria, National Autonomous University of Mexico, 1946

Islebe GA, Hooghiemstra H, Brenner M, Curtis JH, Hodell DA (1996) A Holocene vegetation history from lowland Guatemala. Holocene 6 : 265–271

Kennedy AJ (1995) Cacao. In: Smartt J, Simmonds NW (eds) Evolution of crop plants, 2nd edn. Longman Scientific & Technical, London, pp 472–475

King LM, Schaal BA (1989) Ribosomal-DNA variation and distribution in *Rudbeckia missouriensis*. Evolution 43 : 1117–1119

Knight R, Rogers HH (1955) Incompatibility in *Theobroma cacao*. Heredity 9 : 69–77

Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetics analysis, version 1.0. The Pennsylvania State University, University Park, Pa.

Lange K, Boehnke M (1982) How many polymorphic genes will it take to span the human genome. Am J Hum Genet. 34 : 842–845

Laurent V, Risterucci AM, Lanaud C (1993a) Variability for nuclear ribosomal genes within *Theobroma cacao*. Heredity 71 : 96–103

Laurent V, Risterucci AM, Lanaud C (1993b) Chloroplast and mitochondrial DNA diversity in *Theobroma cacao*. Theor Appl Genet 87 : 81–88

Laurent V, Risterucci AM, Lanaud C (1994) Genetic diversity in cacao 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* accessions using RAPD and RFLP markers. Theor Appl Genet 95 : 10–19

Lewontin RC (1972) The apportionment of human diversity. Evol Biol 6 : 381–398

Leyden BW, Brenner M, Hodell DA, Curtis JH (1994) Orbital and internal forcing of climate on the Yucatan Peninsula for the past ca. 32 ka. Palaeogeogr Palaeoclimatol Palaeoecol 109 : 193–210

Leyden BW, Brenner M, Whitmore T, Curtis JH, Piperno DR, Dahlin BH (1996) A record of long- and short-term climatic variation from northwest Yucatan: cenote San Jose Chulchaca. In: Fedick SL (ed) The managed mosaic: ancient Maya agriculture and resource use. University of Utah Press, Salt Lake City, pp 30–50

Lloyd M, Zar JH, Karr JR (1968) On the calculation of information-theoretical measures of diversity. Am Midl Nat 79 : 257–272

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

Lynch M, Milligan BG (1994) Analysis of population genetic structure with RAPD markers. Mol Ecol 3 : 91–99

Maldonado-Koerdell M (1941) Los jardines botánicos de los antiguos Mexicanos. Rev Soc Mex Hist Nat 2 : 79–84

N'Goran JAK, Laurent V, Risterucci AM, Lanaud C (1994) Comparative genetic diversity studies of *Theobroma cacao* L. using RFLP and RAPD markers. Heredity 73 : 589–597

Peet RK (1974) The measurement of species diversity. Annu Rev Ecol Syst 5 : 285–307

Peters CE (1983) Observations on Maya subsistence and the ecology of a tropical tree. Am Antiq 48 : 609–615

Purseglove JW (1974) Tropical crops, dicotyledons. John Wiley & Sons, New York

Ronning CM, Schnell RJ (1994) Allozyme diversity in a germplasm collection of *Theobroma cacao* L. J Hered 85:291–295

Ronning CM, Schnell RJ, Kuhn DN (1995) Inheritance of random amplified polymorphic DNA (RAPD) markers in *Theobroma cacao* L. J Am Soc Hortic Sci 120:681–686

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

Sahagún B de (1590) Historia general de las cosas de Nueva España. Tomo II. (An annotated Spanish edition published in 1955). Alfa, Mexico, DF

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York

Schultes RE (1984) Amazonian cultigens and their northward and westward migrations in pre-Columbian times. In: Stone D (ed) Pre-Columbian plant migration. Harvard University Press, Cambridge, pp 32–33

Stewart CN, Excoffier L (1996) Assessing population genetic structure and variability with RAPD data: application to *Vaccinium macrocarpon* (American Cranberry). J Evol Biol 8:153–171

Toxopeus H (1985) Botany, types and populations. In: Wood GAR, and Lass RA (eds) Cocao, 4th edn. Longman, London, pp 11–37

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

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535

Young AM (1994) The chocolate tree: a natural history of cacao. Smithsonian Institution Press, Washington D.C.