GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Genetic identification of *Theobroma cacao* L. trees with high Criollo ancestry in Soconusco, Chiapas, Mexico

J.A. Vázquez-Ovando[1,4], F. Molina-Freaner[2], J. Nuñez-Farfán[3], I. Ovando-Medina[4] and M. Salvador-Figueroa[4]

[1]Graduate Program in Biological Sciences,
National Autonomous University of Mexico, Mexico City, Mexico
[2]Institute of Ecology, Ecology of Biodiversity Department,
National Autonomous University of Mexico, Hermosillo, Sonora, Mexico
[3]Institute of Ecology, Evolutionary Ecology Department,
National Autonomous University of Mexico, Mexico City, Mexico
[4]Center of Biosciences, Autonomous University of Chiapas, Tapachula,
Chiapas, Mexico

Corresponding author: M. Salvador-Figueroa
E-mail: miguel.salvador@unach.mx

Genet. Mol. Res. 13 (4): 10404-10414 (2014)
Received October 11, 2013
Accepted July 7, 2014
Published December 12, 2014
DOI http://dx.doi.org/10.4238/2014.December.12.2

ABSTRACT. Criollo-type cacao trees are an important pool of genes with potential to be used in cacao breeding and selection programs. For that reason, we assessed the diversity and population structure of Criollo-type trees (108 cultivars with Criollo phenotypic characteristics and 10 Criollo references) using 12 simple sequence repeat (SSR) markers. Cultivars were selected from 7 demes in the Soconusco region of southern Mexico. SSRs amplified 74 alleles with an average of 3.6 alleles per population. The overall populations showed an average observed heterozygosity of 0.28, indicating heterozygote deficiency (average fixation index F = 0.50). However, moderate allelic diversity was found within populations

(Shannon index for all populations I = 0.97). Bayesian method analysis determined 2 genetic clusters (K = 2) within individuals. In concordance, an assignment test grouped 37 multilocus genotypes (including 10 references) into a first cluster (Criollo), 54 into a second (presumably Amelonado), and 27 admixed individuals unassigned at the 90% threshold likely corresponding to the Trinitario genotype. This classification was supported by the principal coordinate analysis and analysis of molecular variance, which showed 12% of variation among populations (F_{ST} = 0.123, P < 0.0001). Sampled demes sites (1-7) in the Soconusco region did not show any evidence of clustering by geographic location, and this was supported by the Mantel test (R_{xy} = 0.54, P = 0.120). Individuals with high Criollo lineage planted in Soconusco farms could be an important reservoir of genes for future breeding programs searching for fine, taste, flavor, and aroma cocoa.

Key words: Population structure; Genetic diversity; Genetic resources; Fine flavor cocoa; Simple sequence repeat markers

INTRODUCTION

The cacao tree (*Theobroma cacao* L.) is the most economically important species of the genus *Theobroma* (Cuatrecasas, 1964). Cocoa butter, cocoa powder, and chocolate can be obtained from its beans. In addition, it is a source of products for the food, pharmaceutical, and cosmetic industries (Wood and Lass, 1985). This plant, which is almost exclusively allogamous, appears to be native to South America (Cuatrecasas, 1964; Motamayor et al., 2008), although it was domesticated in Mesoamerica (Gómez-Pompa et al., 1990; De la Cruz et al., 1995; Coe and Coe, 1996). In this region of the American continent, the history of cacao dates from ancient times (Henderson et al., 2007). The earliest evidence of cacao use dates from 1500 to 1900 BC and was reported by Powis et al. (2007, 2008) in Paso de Amada, in the southern Mexican state of Chiapas. Cacao was also cultivated and used in the pre-Columbian era in countries such as Guatemala, Honduras, and Belize (Hurst et al., 2002; Powis et al., 2007). The first group classification accepted of *T. cacao* is based on morpho-geographical traits of trees and fruits. From this classification, Forastero, Trinitario (admixture of Forastero and Criollo), and Criollo, likely domesticated by Mesoamerican natives (Cheesman, 1944), are commonly known. More recently, a study by Motamayor et al. (2008) using simple sequence repeat (SSR) markers reported 10 genetic groups in cacao: Amelonado, Contamana, Criollo, Curaray, Guiana, Iquitos, Marañón, Nacional, Nanay, and Purús.

The Criollo variety produces fruits with high-quality sensory characteristics, such as sweet pulp, less bitter beans, and attractive aroma, flavor, and taste (Smith, 1999). Although the Criollo cacao spread in Mesoamerica and was socio-economically attractive, the presence of unknown diseases and other socio-cultural events caused the loss of cacao crops in Southern Mexico and Central America (Zhang et al., 2011). However, Whitkus et al. (1998) and Motamayor et al. (2002) reported that, in that region, there are still naturally grown Criollo cacao trees. Such individuals of the Criollo-type could be a useful resource for cacao breed-

ing and selection programs. Especially, gourmet markets (niche for the Criollo-type) have increased in recent years (Vázquez-Ovando et al., 2012) up to 6.8% of the current cacao bean demand in the world (ICCO, 2012).

In 2011, Mexico produced 21,400 tons of cacao beans (FAO, 2012). The Soconusco region of southern Mexico produces 30% of the national production of cacao beans (SIAP, 2013). During the second half of the twentieth century, Mexican government programs promoted the renovation of old, sick Criollo orchards with Trinitario and Forastero varieties (López-Mendoza, 1987; Ogata, 2003). Nevertheless, experienced cacao producers are still able to identify and select native Criollo trees based on phenotypic characteristics from traditional plantations.

The finding of cacao trees of near-pure Criollo lineage in the area of Soconusco suggests that seedling propagation practices by cacao farmers have disseminated this material in the area. The knowledge of the population genetic diversity and structure of that germplasm or farmer selections compared with wild Criollo individuals can provide guidelines for future selection and improvement programs of fine cacao. The objectives for this study were to identify trees with high Criollo ancestry and to compare the molecular genetic diversity of trees selected based on Criollo fruit traits in the Soconusco area in Mexico.

MATERIAL AND METHODS

Plant material and sample collection

A total of 118 leaf samples comprising 108 cultivars from farms in Soconusco, Mexico, and 10 Criollo references from wild collections were analyzed. Cultivars were selected from 19 farms in 10 municipalities in Soconusco, Chiapas, Mexico. They were *a priori* grouped in 7 demes according to their geographical proximity and similar environmental conditions (Table 1; Figure 1). The individuals were selected based on traits of fruits (pod) and seeds, which resembled those of the Criollo-type. Pods were elongated, deeply grooved, pointed at the pod end, had a lumpy surface with a warty appearance outside, white or slightly pigmented seeds, and sweet mucilage as described by Engels (1983). Samples were collected from 1 to 16 cacao trees per farm (Table 1). The average age of cacao trees was 30 years, and particular care was taken to include only sexually propagated plants (farmer's selections). Reference accessions with Criollo lineage referenced as Carmelo, Lacandón 06, Lacandón 28, Xocen, Yaxcabá, Loxicha, and Pentagona from the germplasm bank of the Instituto Nacional de Investigaciones Forestales Agrícolas y Pecuarias (National Research Institute of Forestry, Agriculture, and Livestock; INIFAP), Campo Experimental Rosario Izapa, Chiapas, Mexico, were kindly provided by Hugo Avendaño-Arrazate. The other reference accessions (SL1, SL2, and SL3) were collected in the Lacandon rainforest (Selva Lacandona, SL), where expeditions to collect wild Criollo cacao individuals have previously been reported (Whitkus et al., 1998; Motamayor et al., 2008). Leaves were sampled and placed in plastic bags, taken to the laboratory (4°C), and processed on the same day of sampling. Because of the long distance to the SL site, 2 leaves were collected and kept at ambient temperature in 20 mL aqueous solution [35% NaCl, 1.5% cetyltrimethylammonium bromide (CTAB), and 0.02% sodium azide] to maintain DNA integrity (Bhattacharjee et al., 2004).

Table 1. Location and number of *Theobroma cacao* L. trees considered in this study.

Soconusco location	Geographic location	Code	Sample size (N)	*a priori* population assignment (deme)
Tapachula	14°52'55"N, 92°21'42"W	TASG	16	1
Cacahoatán	14°59'53"N, 92°10'44"W	CAAM	14	2
Tuxtla Chico	14°56'41"N, 92°09'59"W	TCHR	8	2
Villa Comaltitlán	15°10'31"N, 92°38'06"W	VCHL	4	3
Villa Comaltitlán	15°11'17"N, 92°36'55"W	VCLB	3	3
Villa Comaltitlán	15°15'06"N, 92°36'17"W	VCAV	5	3
Mapastepec	15°28'07"N, 92°48'42"W	MAJH	3	3
Frontera Hidalgo	14°47'31"N, 92°11'11"W	FHSA	6	4
Suchiate	14°38'27"N, 92°13'47"W	SUED	7	4
Mazatán	14°38'58"N, 92°29'06"W	MAMG	13	5
Huehuetán	14°59'27"N, 92°26'34"W	HUTG	4	6
Huehuetán	14°59'28"N, 92°26'44"W	HUJF	6	6
Huehuetán	14°59'09"N, 92°26'41"W	HUCM	2	6
Huehuetán	15°00'46"N, 92°26'32"W	HUCG	3	6
Huehuetán	15°03'22"N, 92°21'54"W	HUIL	3	7
Tuzantán	15°08'19"N, 92°24'51"W	TUCA	2	7
Tuzantán	15°08'32"N, 92°25'11"W	TURA	1	7
Tuzantán	15°09'37"N, 92°23'39"W	TURH	2	7
Tuzantán	15°04'59"N, 92°24'01"W	TUMM	6	7
Several[a]	Several	INIFAP	7	8
Benemérito de las Américas[b]	16°64'37"N, 90°56'26"W	SL	3	8

[a]Trees identified as Criollo wild cacaos (sampled from the germplasm bank of the National Research Institute of Forestry, Agriculture, and Livestock, INIFAP). [b]Samples collected in the Lacandon Forest (wild Criollo).

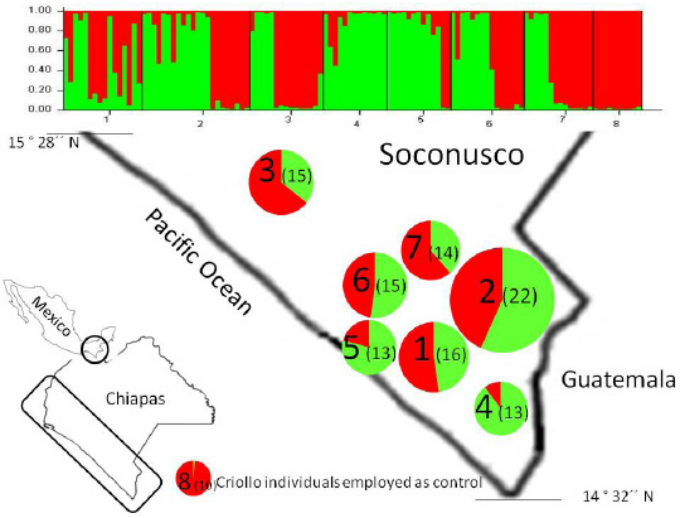

Figure 1. Genetic structure of the 7 demes of Criollo cacao trees sampled in Soconusco, Chiapas, Mexico, and one reference group with high Criollo ancestry. Results were obtained using the Structure 2.3.2 software (Pritchard et al., 2000) with 200,000 burn-in iterations, 400,000 iterations after burn-in, and 20 repetitions of each genetic population (K1-K5), admixture ancestry, and allele frequency-correlated model. The numbers in the pie charts denote the geographic deme and the number of individuals in parentheses. Red and green represent Criollo and presumably Amelonado types, respectively.

DNA extraction and SSR analysis

Total DNA extraction was performed by modifying the method described by Doyle and Doyle (1990). Leaves were washed with sterile water and 70% ethyl alcohol. Approximately

200 mg cacao leaves were ground with liquid nitrogen with 60 mg polyvinyl pyrrolidone and 1 mL CTAB buffer [2% CTAB (w/v), 20 mM ethylenediaminetetraacetic acid (EDTA), 1.4 M NaCl, 100 mM Trizma® base, pH adjusted to 8 with HCl, and 1% 2-mercaptoethanol (v/v)]. DNA extractions were performed with chloroform-isoamyl alcohol and precipitation with isopropanol. The extracted DNA was purified with a mixture of phenol:chloroform:isoamyl alcohol (25:24:1). The DNA integrity (dissolved in 60 µL Milli-Q water) was checked by 0.8% agarose electrophoresis and quantified by spectrophotometry at 260 nm (GBC Cintra 10e™ spectrophotometer, Dandenong, VIC Australia), and the purity was inferred by the 260/280 and 260/230 absorbance ratios.

Twelve highly polymorphic SSR markers were used in this test (Lanaud et al., 1999; Riju et al., 2009; Table 2). After evaluating reagent concentration and temperature conditions of several protocols, polymerase chain reaction (PCR) was performed according to Schnell et al. (2005) with modifications. These loci markers were chosen because they are distributed throughout the *T. cacao* genome (TropGene DB, Hamelin et al., 2013). They are part of the international set of SSR markers for cacao tree characterization reported by Saunders et al. (2004) with good discriminatory power (Motilal et al., 2009). Only those that showed reproducible and polymorphic bands in at least 3 trials were selected.

The PCR mixture contained 30 ng genomic DNA, 0.4 µM of each primer (Integrated DNA Technologies®, Coralville, IO USA), 0.2 mM dNTP Mix (Promega®, Madison, WI, USA), 1X PCR assay buffer ViBuffer A (Vivantis Oceanside, CA, USA), 4 mM $MgCl_2$, and 1.25 U Taq DNA polymerase (Vivantis, USA) in a reaction volume of 25 µL. DNA template was initially denatured at 94°C for 4 min, followed by 32 cycles of PCR amplification in a thermal cycler TC3000 (Techne, Cambridge, UK) with the following conditions: 30 s denaturation at 94°C, 1 min at appropriate annealing temperature for each primer (Table 2), and 1 min primer extension at 72°C, followed by a final extension at 72°C for 5 min. PCR products (13 samples per gel) were separated on 12% polyacrylamide gels (size 13 x 12 cm) using 0.5X Tris-borate-EDTA buffer, stained with ethidium bromide (0.6 ng/µL) for 30 min, and visualized under ultraviolet light and photographed with a Gel Doc™ EZ Imager gel documentation system (Bio-Rad, Hercules, CA, USA). Fragment sizes were estimated using the Image Lab software utility v. 4.0.1 (Bio-Rad Laboratories) option of the gel documentation system and by integrating GeneRuler™ Low Range DNA Ladder (Fermentas, Carlsbad, CA, USA). Weak bands (ghost bands) and smeared products were ignored. To confirm the reproducibility of results, DNA was extracted twice, amplified, and quantified (4 amplifications per individual).

Data analysis

Using the GenAlEx v6.5b3 software (Peakall and Smouse, 2006), genetic diversity analysis was performed by calculating global and population parameters of polymorphism, number of alleles, effective number of alleles, Shannon index, observed heterozygosity, expected heterozygosity, and fixation index. Chi-square tests were performed to test the balance-imbalance of Hardy-Weinberg equilibrium by locus in all populations. Private alleles were identified by individual and by population. Genetic and geographic linear distances were calculated for all individuals except for reference trees. Mantel correlation test (10,000 permutations) was performed using the GenAlEx v6.5b3 software. The presence of null alleles and scoring for errors were analyzed using the Micro-Checker software (Van Oosterhout et al., 2004).

The population structure and genetic ancestry of individuals were inferred using the Bayesian statistical methods from Structure v 2.3.2 (Pritchard et al., 2000). The following parameters were set up: admixture model of ancestry, 200,000 iterations during the burn-in period, 400,000 iterations after the burn-in period, and 20 repetitions for each genetic subpopulation (*K*1-*K*5) using the allele frequencies correlated model. The number of *K* was estimated following the procedure described by Evanno et al. (2005) using the Structure Harvester v0.6.93 (Earl and von Holdt, 2012). To determine the degree of differentiation within and among populations, analysis of molecular variance (AMOVA) was performed with 10,000 iterations using the GenAlEx v6.5b3.

On the basis of the coefficient of membership to 1 of 2 distinct genetic clusters generated by structure analysis, all individuals were separated into genotypic classes (assignment percentage threshold of 90%). The first class included individuals with higher than 90% membership to cluster 1 (primarily red vertical bars in Figure 1). A second class included trees with higher than 90% membership to cluster 2 (primarily green vertical bars in Figure 1). Unassigned individuals (less than 90% membership to any cluster) were considered trees with admixed ancestry. We performed an additional AMOVA for these classes. The grouping was further verified by the assignment method of Paetkau et al. (2004) using the GenAlEx v6.5b3 software. The assignment was considered correct when 100% of individuals converged with their respective group. A principal coordinate analysis (PCoA) was conducted using a pairwise distance matrix with distance and covariance standardized. The plot was built with the 2 axes that best explain the total variation (43%) of subpopulations.

RESULTS

Twelve SSR loci detected 12 homozygous individuals (range 0-3 individuals per population) and distinguished 74 alleles among the 118 individuals analyzed, with a range of 3 to 13 alleles per locus (average of 6.2 alleles per locus). Fragment sizes were as expected, except for the mTcCIR3 locus, which was shorter than expected [previously reported by Lanaud et al. (1999); Table 2]. The Micro-Checker analysis confirmed no scoring error due to stuttering and no evidence for large allele dropout for mTcCIR3. However, the analysis suggests that null alleles may be present at the mTcCIR3 locus. Seven private alleles were detected in 5 of the 12 microsatellites. Individuals HUTG01, MAMG07, and TASG18 have a unique private allele at locus mTcCIR 6, mTcCIR 7, and mTcCIR 28, respectively. The Tapachula subpopulation had the highest proportion of individuals with private alleles detected at the mTcCIR28 locus (Table 3).

Table 2. Microsatellite primers used in molecular characterization, melting temperatures (Tm), and fragment sizes obtained.

Microsatellite locus	Chromosome[a,b]	Physical position[a,b] (cM)	Tm (°C)	Size of expected fragment[c] (bp)	Size range in study (bp)
mTcCIR1	8	0.0	64.7	143	133-155
mTcCIR3	2	19.5	57.1	249	197-227
mTcCIR6	6	10.3	54.2	231	163-265
mTcCIR7	7	27.3	56.2	160	142-194
mTcCIR8	9	52.6	51.5	301	265-319
mTcCIR11	2	92.2	49.7	298	278-370
mTcCIR12	4	45.7	54.4	188	168-274
mTcCIR15	1	18.0	53.4	254	228-296
mTcCIR19	2	10.3	56.7	376	214-310
mTcCIR21	3	25.8	61.0	157	142-170
mTcCIR25	6	45.0	54.1	153	128-175
mTcCIR28	6	55.3	57.3	336	322-362

[a]Risterucci et al. (2000); [b]TropGene database (Hamelin et al., 2013); [c]Lanaud et al. (1999).

Table 3. Private alleles by population and individual of *Theobroma cacao* L. trees sampled in Soconusco, Chiapas, Mexico.

Population	Locus	Private allele	Frequency	Individuals
Huehuetán (Pop 6)	mTcCIR 6	233	0.077	HUTG01
Huehuetán	mTcCIR 8	307	0.200	HUTG02, HUJF05
Mazatán (Pop 5)	mTcCIR 7	194	0.083	MAMG07
Suchiate (Pop 4)	mTcCIR 19	352	0.111	SUED02, SUED05
Tapachula (Pop 1)	mTcCIR28	328	0.036	TASG18
Tapachula	mTcCIR28	330	0.357	TASG03, TASG09, TASG10, TASG13, TASG14, TASG15, TASG16, TASG17
Tapachula	mTcCIR28	338	0.250	TASG14, TASG15, TASG16, TASG17, TASG18

Additionally, the microsatellite loci revealed polymorphism rates for subpopulations with values over 91.7%, effective number of alleles between 2.35 and 2.63, and values for the Shannon diversity index between 0.87 and 1.02 (Table 4). Looking at the expected heterozygosity (H_E = 0.54), observed heterozygosity (H_O = 0.28), and positive value of fixation index (F), all of the *a priori* determined subpopulations presented heterozygote deficiency ($H_E > H_O$). Except for mTcCIR 1 (which revealed a value of 4.29), the other SSR loci had low numbers of migrants for all subpopulations (0.284-1.784), with an average value of 1.39 for all SSR markers.

Table 4. Average values as indicators of population diversity in *Theobroma cacao* L. trees sampled from Soconusco, Chiapas, Mexico. Values were obtained using the GenAlEx v6.5b3 software (Peakall and Smouse, 2006).

Population	N	N_A	N_E	I	H_O	H_E	F
Pop 1	13.8	3.75	2.53	1.02	0.34	0.56	0.41
Pop 2	15.2	4.00	2.59	1.03	0.24	0.56	0.54
Pop 3	12.8	3.67	2.60	1.02	0.23	0.56	0.64
Pop 4	9.0	3.17	2.39	0.89	0.25	0.52	0.53
Pop 5	8.6	3.67	2.63	1.02	0.20	0.56	0.66
Pop 6	10.8	3.50	2.60	0.95	0.35	0.52	0.32
Pop 7	10.8	3.42	2.51	0.95	0.22	0.55	0.63
Pop 8	8.0	3.17	2.35	0.87	0.38	0.50	0.19
Mean	11.1	3.54	2.52	0.97	0.28	0.54	0.49
SE	0.4	0.16	0.11	0.04	0.03	0.02	0.05

N = number of individuals; N_A = number of different alleles; N_E = effective number of alleles; I = Shannon diversity index, H_O = observed heterozygosity; H_E = expected heterozygosity; F = fixation index; SE = standard error.

The effective number of alleles (N_E = 2.52) for all subpopulations was lower than the number of alleles (N_A = 3.54). The Hardy-Weinberg test revealed disequilibrium in five loci (mTcCIR 3, mTcCIR 6, mTcCIR 8, mTcCIR 25, and mTcCIR 28) in all subpopulations ($P < 0.05$). The subpopulations Tapachula (1) and Villa Comaltitlán (3) exhibited this disequilibrium for 11 of 12 tested SSR markers.

AMOVA was conducted to detect inter- and intra-population variation. The analysis revealed that the largest variation (88%) was explained by the variation within populations, while the remaining (12%) was among populations, with an F_{ST} value of 0.123 ($P < 0.0001$). Bayesian statistical analysis using an admixture model and calculating the best value of K by Evanno et al. methods (Evanno et al., 2005) indicated that the 118 individuals grouped in 2 genetic clusters ($K = 2$, $\Delta K = 294.21$). Structure analysis without *a priori* deme designations indicated that the 118 individuals were effectively grouped into 2 genetic clusters ($\Delta K = 278.89$).

Geographic demes Villa Comaltitlán (3) and Tuzantán (7) shared a high proportion of alleles (red color key in Figure 1) with Criollo references (8). The other group had individuals from demes Cacahoatán (2), Suchiate (4), and Mazatán (5), with more than 50% of non-Criollo alleles. Tapachula (1) and Huehuetán (6) shared an equal proportion of alleles from both genetic clusters. Sampled demes sites (1-7) in the Soconusco region did not show any evidence of clustering by geographic location (Figure 1), and this was supported by a non-significant Mantel test (R_{xy} = 0.54, P = 0.120).

Assignment analysis performed with the GenAlEx v6.5b3 software defining 3 clusters *a priori* grouped 37 individuals in cluster 1 (red color key in Figure 2) that included the Criollo references, 54 individuals in cluster 2 (green color key in Figure 2), and 27 individuals in cluster 3 (blue color key in Figure 2), which was called "hybrids", with 100% probability. We found moderate differentiation among that clusters (AMOVA F_{ST} = 0.111). The PCoA plot showed more clearly the differences among clusters. Coordinates 1, 2, and 3 explain 22.7, 20.01, and 17.15% of the total variation, respectively. Figure 2 shows the separation between the Criollo-type trees, non-Criollo, and "hybrids" in the overlapping region where allele frequencies are shared.

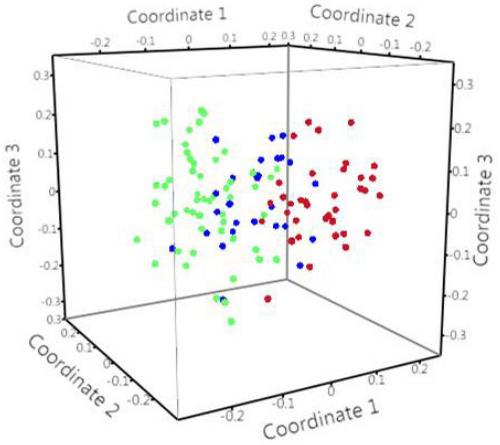

Figure 2. Principal coordinate analysis of 118 cacao trees comprising 108 cultivars sampled from 7 demes in Soconusco, Chiapas, Mexico, and 10 Criollo references belonging to 2 genetic clusters $K = 2$ based on the Structure software. Criollo cluster (red) and genetic cluster 2 (green) are shown. Blue indicates hybrid individuals that likely belong to the Trinitario cultivar. Analysis was conducted with data from 12 microsatellites and 100% assignment by GenAlEx v6.5b3 (Peakall and Smouse, 2006). Coordinates 1, 2, and 3 explain 22.7, 20.01, and 17.15% of the total variation, respectively.

DISCUSSION

In this study, we investigated the genetic structure and relationships among 7 demes of Criollo-type cacao trees cultivated in Soconusco, Chiapas, Mexico, using SSR markers. To our knowledge, this is the first report on the genetic diversity of cacao grown in the Soconusco area where theobromine residues have tested positive on archaeological pottery specifically at the Mokaya site of Paso de Amada (Powis et al., 2007, 2008). There are studies on genetic diversity of the germplasm collected in southeast Mexico; however, these have focused on wild accessions. For example, Motamayor et al. (2008) analyzed, among others, wild cacaos from

the Lacandon Forest with SSRs markers. In the same way, Whitkus et al. (1998) employed random amplified polymorphic DNA (RAPD) markers to examine the genetic diversity of cultivars from the Tabasco area, which is located 300 km north of our study, and wild trees collected from the Lacandon Rainforest.

The fixation rates found in our study for all *a priori* subpopulations represents high levels of inbreeding or possible crossover between siblings as supported by a deficiency of heterozygotes ($H_E > H_O$). Such levels of homozygosity may suggest intensive cultivation and selection of Criollo genotypes for a long period of time. Selection pressure increases levels of inbreeding (Loor Solorzano et al., 2012) by 2 mechanisms: self-compatibility, as reported in ancient Criollo (Cheesman, 1944; Haddon, 1961), and selection (unintended) of homozygous trees by local cacao farmers. It is not surprising that populations had a deficiency of heterozygotes because of the effect of deliberate sampling of individuals with Criollo characteristics. Another hypothesis to explain the heterozygote deficiency is the presence of a founder effect, i.e., the populations analyzed descended from a narrow number of parents and were continuously cultivated in the Soconusco region by local cacao farmers. The values obtained in the Hardy-Weinberg test support this hypothesis.

The sampling effort in this study was influenced by the intention toward a specific group of individuals (Criollo-type trees), reflected by the $N_E < N_A$ values, setting limits for the population study. It would be desirable to assess the diversity of all individuals in a greater sampling effort in future studies. It is important to note that, despite the high homozygosity, the Shannon diversity index shows significant allelic diversity within populations, which is consistent with the total number of alleles revealed by microsatellites (total = 74; mean = 6.2/SSR) for all individuals. Johnson et al. (2009), using microsatellite markers (260 alleles, using 35 SSRs; mean 7.4/SSR), reported a genetic diversity that was similar to that of our study. This proximity in mean number of alleles could be an indication that more than 50% of the individuals analyzed in our study (particularly demes 2, 4, 5, and 6) corresponded to Trinitario types (cluster 2, green color key in Figures 1 and 2). This may have 2 explanations. First, the phenotypic characteristics utilized for collecting individuals in the farmer's field were not discriminatory enough to select only Criollo trees. Second, when selection was based on sensory characteristics, trees were also the Trinitario type.

By analyzing the genetic structure of populations and finding $K = 2$, a genetic differentiation can be observed between the 2 clusters (Figures 1 and 2). Almost 75% of the total individuals were assigned to the groups with high levels of correspondence; however, the remaining individuals (27) are located in the transition region between the pools. The genetic structure analysis clustered wild individuals that were collected in the Mayan area (INIFAP and SL Table 1; mean coefficient of membership to cluster 1 = 0.984) together with 27 cultivated individuals (mean coefficient of membership to cluster 1 = 0.974). This finding suggests that a group of Criollo-type cacao was found in the Mokaya region. In contrast, Whitkus et al. (1998), using RAPD marker analysis, did not find an association between wild Criollo (from the Lacandon Forest and Yucatan State, a zone of influence of the Mayan culture in Mexico) and accessions cultivated from Tabasco, Mexico, and South America. Individuals in the overlapping area (blue color key in Figure 2), which were identified as "hybrids," are admixed individuals that share alleles with Criollos and could be considered candidates for Trinitario. The parental contribution of Mesoamerican Criollos to the appearance of Trinitario cacao has been documented in the past (Motamayor et al., 2003).

Cluster 2 grouped 54 individuals (green in Figures 1 and 2) with a clear non-Criollo

lineage, but these individuals share alleles with "hybrids." Possibly, these individuals belong to a particular genetic group lacking reference accessions in this study; therefore, they are now unassigned. A strong contender for this group, based on Trinitario history, is the Amelonado. More studies, including parental controls of Trinitario, are needed to address this question.

Criollo-type individuals represent a gene pool with potential to be included in breeding programs searching for fine flavor cocoa; however, is necessary to evaluate their agronomic potential, disease tolerance, and sensory qualities. Some of these materials will also serve as the basis for future selection studies to propagate and establish plantations with enhanced cocoa qualities. On the other hand, it is important to increase the heterozygosity of populations without neglecting the important features of the local cacaos. Private bands could be isolated, sequenced, and converted to sequenced-characterized amplified region or other markers, which would assist the identification of improved genotypes, because they are specific tags.

ACKNOWLEDGMENTS

We thank Dr. Hugo Avendaño-Arrazate, who kindly provided the leaf material of wild cacao accessions belonging to the germplasm bank of the National Research Institute of Forestry, Agriculture, and Livestock (INIFAP, Mexico); the Graduate Program in Biological Sciences of the National Autonomous University of Mexico (UNAM) for the academic instruction in doctoral studies for the first author; and the National Council of Sciences and Technology (CONACyT, Mexico) for financial grant awarded to the first author. Special thanks to Samuel Guillén for contact with farmers and to Carlos Amores, Julio Magaña, Dafne Castillo, Nayelly Hernández, Nidia Gutiérrez, and Julio Coutiño for technical support.

REFERENCES

Bhattacharjee R, Kolesnikova-Allen M, Aikpokpodion P, Taiwo S, et al. (2004). An improved semiautomated rapid method of extracting genomic DNA for molecular marker analysis in cocoa, *Theobroma cacao* L. *Plant Mol. Biol. Rep.* 22: 435-436.

Cheesman E (1944). Notes on the nomenclature, classification and possible relationships of cocoa populations. *Trop. Agr.* 21: 144-159.

Coe SD and Coe MD (1996). The True History of Chocolate. 1st edn. Thames and Hudson Ltd., New York.

Cuatrecasas J (1964). Cacao and its allies: A taxonomic revision of the genus *Theobroma*. *Contr. U. S. Natl. Herb.* 35: 379-614.

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

Doyle JJ and Doyle JL (1990). Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15.

Earl DA and von Holdt BM (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4: 359-361.

Engels JMM (1983). A systematic description of cacao clones. I. The discriminative value of quantitative characteristics. *Euphytica* 32: 377-385.

Evanno G, Regnaut S and Goudet J (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620.

FAO (Food and Agriculture Organization) (2012). FAOSTAT Online Statistical Service. Available at [http://faostat.fao.org/]. Accessed September 12, 2013.

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

Haddon AV (1961). Variety trials of seedling cacao in Malaya. *Malayan Agric. J.* 43: 169-232.

Hamelin C, Sempere G, Jouffe V and Ruiz M (2013). TropGeneDB, the multi-tropical crop information system updated and extended. *Nucleic Acids Res.* 41: D1172-D1175.

Henderson JS, Joyce RA, Hall GR, Hurst WJ, et al. (2007). Chemical and archaeological evidence for the earliest cacao beverages. *Proc. Natl. Acad. Sci. U. S. A.* 104: 18937-18940.

Hurst WJ, Tarka SM Jr, Powis TG, Valdez F Jr, et al. (2002). Cacao usage by the earliest Maya civilization. *Nature* 418: 289-290.

ICCO (International Cocoa Organization) (2012). Quarterly Bulletin of Cocoa Statistics. Vol. XXXVIII-No.2. Cocoa Year 2010/11 London (May 2012).

Johnson ES, Bekele FL, Brown SJ, Song Q, et al. (2009). Population structure and genetic diversity of the Trinitario cacao (*Theobroma cacao L.*) from Trinidad and Tobago. *Crop Sci.* 49: 564-572.

Lanaud C, Risterucci AM, Pieretti I, Falque M, et al. (1999). Isolation and characterization of microsatellites in *Theobroma cacao* L. *Mol. Ecol.* 8: 2141-2143.

Loor Solorzano RG, Fouet O, Lemainque A, Pavek S, et al. (2012). Insight into the wild origin, migration and domestication history of the fine flavour Nacional *Theobroma cacao* L. variety from Ecuador. *PLoS One* 7: e48438.

López-Mendoza R (1987). El cacao en Tabasco. *Colección Cuadernos Universitarios. Serie Agronómica* no. 13. Universidad Autónoma de Chapingo, Texcoco.

Motamayor JC, Risterucci AM, Lopez PA, Ortiz CF, et al. (2002). Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity* 89: 380-386.

Motamayor JC, Risterucci AM, Heath M and Lanaud C (2003). Cacao domestication II: progenitor germplasm of the *Trinitario cacao* cultivar. *Heredity* 91: 322-330.

Motamayor JC, Lachenaud P, da Silva e Mota J, Loor R, et al. (2008). Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L.). *PLoS One* 3: e3311.

Motilal LA, Zhang D, Umaharan P, Mischke S, et al. (2009). Increasing accuracy and throughput in large-scale microsatellite fingerprinting of cacao field germplasm collections. *Trop. Plant Biol.* 2: 23-37.

Ogata N (2003). Domestication and Distribution of the Chocolate Tree (*Theobroma cacao* L.) in Mesoamerica. In: The Lowland Maya Area. Three Millennia at the Human-Wildland Interface (Gómez-Pompa A, Allen MF, Fedick SL and Jiménez-Osorio JJ, eds.). The Haworth Press Inc., New York, 415-438.

Paetkau D, Slade R, Burden M and Estoup A (2004). Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Mol. Ecol.* 13: 55-65.

Peakall R and Smouse PE (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6: 288-295.

Powis T, Hurst J, Rodríguez MC, Ortíz P, et al. (2007). Oldest chocolate in the New World. *Antiquity* 81: 314.

Powis T, Hurst J, Rodríguez MC, Ortíz P, et al. (2008). The origins of cacao use in Mesoamerica. *Mexicon* 30: 35-38.

Pritchard JK, Stephens M and Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Riju A, Rajesh MK, Sherin PT, Chandrasekar A, et al. (2009). Mining of expressed sequence tag libraries of cacao for microsatellite markers using five computational tools. *J. Genet.* 88: 217-225.

Risterucci AM, Grivet L, N'Goran JAK, Pieretti I, et al. (2000). A high-density linkage map of *Theobroma cacao* L. *Theor. Appl. Genet.* 101: 948-955.

Saunders JA, Mischke S, Leamy EA and Hemeida AA (2004). Selection of international molecular standards for DNA fingerprinting of *Theobroma cacao*. *Theor. Appl. Genet.* 110: 41-47.

SIAP (Servicio de Información Agroalimentaria y Pesquera, SAGARPA México) (2013). Produccion Agrícola por Cultivo, Reporte 2012. Available at [http://www.siap.gob.mx/index.php?option=com_wrapper&view=wrapper&Itemid=350.] Accessed July 30, 2013.

Schnell RJ, Olano CT, Brown JS, Meerow AW, et al. (2005). Retrospective determination of the parental population of superior cacao (*Theobroma cacao* L.) seedlings and association of microsatellite alleles with productivity. *J. Am. Soc. Hortic. Sci.* 130: 181-190.

Smith N (1999). The Amazon River Forest: a Natural History of Plants, Animals, and People. Oxford University Press, New York.

Van Oosterhout C, Hutchinson WF, Wills DPM and Shipley P (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* 4: 535-538.

Vázquez-Ovando A, Molina-Freaner F, Nuñez-Farfán J and Salvador-Figueroa M (2012). Potencial de los marcadores moleculares para el rescate de individuos de *Theobroma cacao* L. de alta calidad. *BioTecnología* 16: 36-56.

Whitkus R, de la Cruz M, Mota-Bravo L and Gómez-Pompa A (1998). Genetic diversity and relationships of cacao (*Theobroma cacao* L.) in southern Mexico. *Theor. Appl. Genet.* 96: 621-627.

Zhang D, Gardini EA, Motilal LA, Baligar V, et al. (2011). Dissecting genetic structure in farmer selections of *Theobroma cacao* in the Peruvian Amazon: implications for on farm conservation and rehabilitation. *Trop. Plant Biol.* 4: 106-116.

Wood GAR and Lass RA (1985). Cocoa. Tropical Agriculture Series. 4th edn. Blackwell Science Logman Group Ltd., New York.