Genet Resour Crop Evol (2013) 60:605–619
DOI 10.1007/s10722-012-9860-4

RESEARCH ARTICLE

Genetic diversity, conservation, and utilization of *Theobroma cacao* L.: genetic resources in the Dominican Republic

Edward J. Boza · Brian M. Irish · Alan W. Meerow · Cecile L. Tondo · Orlando A. Rodríguez · Marisol Ventura-López · Jaime A. Gómez · J. Michael Moore · Dapeng Zhang · Juan Carlos Motamayor · Raymond J. Schnell

Received: 2 December 2011 / Accepted: 14 May 2012 / Published online: 7 July 2012
© Springer Science+Business Media B.V. (outside the USA) 2012

Abstract Cacao (*Theobroma cacao* L.) is a significant agricultural commodity in the Dominican Republic, which ranks 11th in the world for cacao exports. To estimate genetic diversity, determine genetic identity, and identify any labeling errors, 14 SSR markers were employed to fingerprint 955 trees among cacao germplasm accessions and local farmer selections (LFS). Comparisons of homonymous plants across plots revealed a significant misidentification rate estimated to be 40.9 % for germplasm accessions and 17.4 % for LFS. The 14 SSRs amplified a total of 117 alleles with a mean allelic richness of 8.36 alleles per locus and average polymorphism information content (PIC) value of 0.67 for the germplasm collection. Similar levels of variation were detected among the LFS where a total of 113 alleles were amplified with a mean of 8.07 alleles per locus and PIC of 0.57. The observed heterozygosity (H_{obs}) was 0.67 for the germplasm collection and 0.60 for LFS. Based on population structure analysis 43.9 % of the germplasm accessions and 72.1 % of the LFS are predominantly of the Amelonado ancestry. Among these Amelonado, 51.7 % for the germplasm collection and 50.6 % for LFS corresponded to Trinitario hybrid lineage. Criollo ancestry was found in 7.6 and 9.5 % of the germplasm accessions and LFS, respectively. The

Electronic supplementary material The online version of this article (doi:10.1007/s10722-012-9860-4) contains supplementary material, which is available to authorized users.

E. J. Boza · A. W. Meerow · C. L. Tondo ·
J. M. Moore · J. C. Motamayor · R. J. Schnell
C/O USDA-ARS Subtropical Horticulture Research
Station, 13601 Old Cutler Road, Miami, FL 33158, USA

B. M. Irish
USDA-ARS Tropical Agriculture Research Station,
2200 P. A. Campos Ave., Suite 201, Mayagüez,
PR 00680, USA

O. A. Rodríguez · M. Ventura-López
Instituto Dominicano de Investigaciones Agropecuarias y
Forestales (IDIAF), Mata Larga, San Francisco de
Macorís, Dominican Republic

J. A. Gómez
Confederación Nacional de Cacaocultores Dominicanos,
Inc. (CONACADO), Calle Altagracia Saviñon No. 11,
Los Prados, Santo Domingo, Dominican Republic

D. Zhang
USDA-ARS Sustainable Perennial Crops Laboratory,
10300 Baltimore Avenue, Bldg 001 BARC-West,
Beltsville, MD 20705, USA

J. C. Motamayor
MARS, Inc., Hackettstown, NJ, USA

R. J. Schnell (✉)
MARS, Inc., Elizabethtown, PA, USA
e-mail: Ray.Schnell@effem.com

Contamana, Nacional, and Iquitos backgrounds were also observed in both populations, but the Curaray background was only detected in the germplasm accessions. No Purús or Guiana ancestry was found in either of the populations. Overall, significant genetic diversity, which could be exploited in the Dominican Republic breeding and selection programs, was identified among the germplasm accessions and LFS.

Keywords *Theobroma cacao* · Cacao improvement · Gene diversity · Genetic groups · Germplasm mislabeling

Introduction

Cacao (*Theobroma cacao* L.), a member of the Malvaceae sensu lato, is an allogamous diploid species with a small genome (~430 Mb), largely cultivated by small farmers in the humid tropics as a cash crop (Alverson et al. 1999; Cope 1984; Lanaud et al. 1992; Rice and Greenberg 2000; Steinberg 2002; Argout et al. 2010; Dantas and Guerra 2010). It also contributes significantly to the economy of many regions of the world, including countries in West Africa, Asia, South and Central America, and the Caribbean. Its seeds or 'beans' are the source of a multibillion dollar industry for the production of cocoa butter, cocoa powder, chocolates, and confectionary. It is also important for the cosmetic industry and has recently been reported as a source of antioxidants, a promoter of cardiovascular health, and as having antitumor properties (Rusconi and Conti 2010). It is native to South America where the Upper Amazon region is the center of its genetic diversity (Cheesman 1944; Cuatrecasas 1964; Coe and Coe 1996; Motamayor et al. 2008). It was domesticated, likely from wild ancestors, through centuries of cultivation in Central America and Mexico where the Criollo type was most extensively found (Cuatrecasas 1964; Hunter 1990; Motamayor et al. 2002). It has been reported that the Mayas (300–900 AD) and Olmecs (400–1200 BC) consumed Criollo cacao (Henderson et al. 2007) which is a distinct group compared to other major cultivated varieties such as Amelonado, Nacional, or Trinitario (Pound 1945; Motamayor et al. 2008).

T. cacao is an important agricultural commodity in the Dominican Republic which ranks 11th in the world for total production at ~42,000 tons/year (FAOSTAT 2008) and number one in organic cacao exports, ca. 23,500 tons/year (Estadísticas del Departamento de Cacao en República Dominicana 2010). The overall area of cultivated cacao in the country is ~152,000 ha and yields an average of ~450–500 kg/ha. It is grown at elevations of 50–1,300 m above sea level with rainfall range between 1,500 and 1,800 mm/year (Batista 2009; Plan Operativo Annual, Instituto Dominicano de Investigaciones Agropecuarias y Forestales [IDIAF] 2007). In 1962–1966 the first hybrid crosses, improved cacao that had originated at Trinidad and Tobago and the Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), Turrialba, Costa Rica (Sofreco and Ecocaribe 2001), were introduced to the Dominican Republic. In 1972 the Dominican Republic Institute for Agriculture, Animal Husbandry and Agroforestry Research (IDIAF) established a cacao germplasm collection at the Mata Larga Experiment Station in San Francisco de Macorís.

Some discrepancy in the literature exists as to where clones were imported from for the plantings at Mata Larga (Sofreco and Ecocaribe 2001; Informe Anual del Departamento de Cacao 2006). Clonal stocks from CATIE, the USDA-Agricultural Research Service (ARS) Tropical Agriculture Research Station (TARS) in Puerto Rico, and from the USDA-ARS Subtropical Horticulture Research Station (SHRS) quarantine facility in Miami, FL (Sofreco and Ecocaribe 2001), with additional contributions from Brazil, Ecuador, Peru, Venezuela, Central America, Mexico, and Cameroon (Informe Anual del Departamento de Cacao 2006), were all introduced starting in 1970. Since then, the Dominican Republic's cacao genetic resources have been primarily maintained, propagated, and distributed out of the IDIAF's Mata Larga research station. The collection includes clonally propagated germplasm accessions with broad genetic variability and of distinct geographical origin, as well as a number of local farmer selections (LFS) chosen for their disease resistance, and productivity. Superior cacao qualities, including fine flavors and aroma, along with a high fat content and large bean size, have historically commanded premium prices in the market (Eskes and Lanaud 1997; Rusconi and Conti 2010).

Most cacao accessions do not breed true-to-type and high amounts of heterogeneity are often found when using molecular markers (Toxopeus 1985; Schnell et al. 2005, 2007; Zhang et al. 2006a, b; Motamayor et al. 2008; Irish et al. 2010). In cacao self-incompatibility (SI) plays a role in heterogeneity and contributes to inconsistent and poor bean yields (Royaert et al. 2011). Knight and Rogers (1953, 1955) reported that SI in cacao is genetically under sporophytic control suggesting a single S-locus with five alleles. Genes controlling SI in cacao have not been reported; however, a region associated with SI is positioned on linkage group 4 (Crouzillat et al. 1996; Royaert et al. 2011). Cacao germplasm collections, like most horticultural tree crop collections, are maintained as clonally propagated living trees (Zhang et al. 2006b; Schnell et al. 2005, 2007). Unfortunately, as in many clonally propagated collections, mistakes associated with propagation tend to accumulate over time (Motilal and Butler 2003).

In the past, characterization of germplasm collections was based on morphology and agronomic characteristics of individuals (Engels 1983; Bekele and Butler 2000; Iwaro et al. 2003). More recently however, efforts have been underway to assess, describe, and characterize cacao collections based on population structure, genetic diversity, and evolutionary relationships using molecular markers (Zhang et al. 2006a, b; Lerceteau et al. 1997; Johnson et al. 2009; Aikpokpodion et al. 2009; Ventura-López et al. 2006; Motilal et al. 2010; Loor et al. 2009; Motamayor et al. 2002, 2003, 2008; Borrone et al. 2007; Schnell et al. 2005; Irish et al. 2010). Microsatellites or Simple Sequence Repeats (SSRs) and Single Nucleotide Polymorphisms (SNPs) are among the molecular markers being used to characterize cacao germplasm collections (Motilal et al. 2010; Zhang et al. 2006b, 2009; Schnell et al. 2005; Motamayor et al. 2008; Lanaud et al. 1999). The objectives for this study were (a) to characterize the genetic diversity in the cacao germplasm collection at IDIAF Mata Larga; (b) to determine the genetic identity and relationship of trees among the LFS; and (c) to assess mislabeling errors in both collections. The correct identification of genetic resources utilized in the cacao breeding and propagation programs is a critical first step for future increases in the productivity of cacao farms in the country. The characterized genetic diversity of a germplasm collection is a useful resource for allele mining, marker trait associations, genetic mapping, and cloning of genes of interest.

Materials and methods

Germplasm collection and LFS sampling

Cacao leaves were collected from germplasm accessions (803) and LFS trees (55) for a total of 858 trees maintained at the IDIAF Mata Larga research station located at San Francisco de Macorís, Dominican Republic. Leaves from an additional 97 clonally propagated LFS trees were collected on Confederación Nacional de Cacaocultores Dominicanos (CONACADO) producer farms for a total of 955 tree samples from these two sources. The germplasm samples included 59 named cultivars (accessions), each represented by from 2 to 25 replicate trees (Supplemental Material S1). The LFS samples included 110 selections. The number of replicate trees varied from one to three, and replicates were presumably clones (Supplemental Material S2). New, fully-expanded leaves were harvested from each tree and shipped to the USDA-ARS SHRS in Miami, FL for analysis. Leaf material was maintained at 4.0 °C until DNA extractions were conducted.

DNA isolation, SSR markers, and polymerase chain reaction (PCR) amplifications

Genomic DNA was extracted from 200 mg of fresh leaf sample using the Fast-DNA SPIN kit (MP Biomedicals, Irvine, CA) as previously described by Schnell et al. (2005). The study used 14 SSR loci identified as an international standard set for cacao germplasm characterization as reported by Lanaud et al. (1999) and Saunders et al. (2004) (Table 1), These SSR markers were chosen on the basis of their high allelic polymorphism, ease of amplification, and reproducibility (Lanaud et al. 1999; Saunders et al. 2004). These loci are distributed throughout the cacao genome and no linkage disequilibrium has been reported among them (Lanaud et al. 1999; Saunders et al. 2004; Brown et al. 2008). PCR amplifications were carried out in a DNA Engine Tetrad 2, Peltier Thermal Cycler (MJ Research, Watertown, MA) as previously described (Schnell et al. 2005).

Table 1 Description and summary statistics for 14 SSR loci genotyped in the cacao (*Theobroma cacao* L.) germplasm collection and local farmer selections (LFS) at the Dominican Republic's IDIAF Mata Larga station

Locus name	Linkage group	Distance (cM)	T_m	Size range (bp)	Germplasm Collection				LFS			
					No of alleles	H_{obs}	H_{exp}	PIC value	No of alleles	H_{obs}	H_{exp}	PIC value
mTcCir37	10	4.0	46	133–185	12	0.65	0.81	0.79	11	0.66	0.67	0.62
mTcCir33	4	62.9	51	264–346	11	0.73	0.83	0.80	11	0.70	0.71	0.69
mTcCir12	4	45.4	59	188–251	11	0.73	0.80	0.77	9	0.67	0.64	0.61
mTcCir15	1	–	46	232–256	10	0.80	0.83	0.80	10	0.72	0.78	0.75
mTcCir26	8	36.3	46	282–307	9	0.73	0.71	0.66	9	0.54	0.63	0.59
mTcCir60	2	53.6	51	187–223	9	0.71	0.73	0.68	7	0.63	0.62	0.55
mTcCir11	2	95.3	46	288–317	9	0.62	0.69	0.66	9	0.59	0.61	0.57
mTcCir18	4	25.8	51	331–355	8	0.76	0.78	0.74	9	0.79	0.69	0.64
mTcCir6	6	–	46	222–247	8	0.71	0.73	0.69	11	0.64	0.59	0.55
mTcCir40	3	12.6	51	259–284	7	0.67	0.80	0.77	6	0.64	0.73	0.69
mTcCir8	9	54.6	46	288–304	7	0.65	0.69	0.64	5	0.46	0.49	0.46
mTcCir22	1	95.7	46	279–290	6	0.59	0.60	0.55	7	0.51	0.57	0.53
mTcCir24	9	29.2	46	185–203	6	0.58	0.53	0.48	6	0.44	0.44	0.39
mTcCir1	8	1.7	51	127–144	4	0.42	0.43	0.35	3	0.38	0.48	0.37
Mean					8.36	0.67	0.71	0.67	8.10	0.60	0.62	0.57

T_m melting temperature for microsatellite primers, H_{obs} observed heterozygosity, H_{exp} expected heterozygosity, *PIC* polymorphic content values

Capillary electrophoresis and genotypic characterization

The electrophoresis of all PCR products was conducted on an ABI Prism 3730 Genetic Analyzer (Applied Biosytems, Foster City, CA) using Performance Optimized Polymer 7 [POP 7] (Applied Biosystems, Foster City, CA) as previously described by Schnell et al. (2005). Briefly, after PCR reactions, each sample was prepared by combining 1.0 µl of PCR product with 20.0 µl of dH_2O and 0.1 µl of GeneScan ROX 400HD size standard (Applied Biosystems, Foster City, CA), denatured at 95 °C for 5 min and placed immediately on ice. Electrophoresis was carried out using the default run module for fragment analysis with a 36 cm array. SSR alleles were analyzed in terms of fragment size, allelic designations, and internal standard using GeneMapper^{TM} software version 4.0 (Applied Biosystems, Foster City, CA) as previously described (Schnell et al. 2005). A dataset in which each tree sample had a corresponding multi-locus genotype (MLG) was generated as a result.

Exclusion of duplicates and clone identification

Nine hundred and fifty-five trees were genotyped for 14 loci and after eliminating duplicates and mislabeled accessions, 66 germplasm accessions and 115 LFS individuals possessing unique MLGs were analyzed to assess genetic diversity, population structure and differentiation. Based on pairwise comparison, when trees had the same accession name and the same MLG, replicates of clones (intra-accession duplications) were eliminated. Next, inter-accession duplicates or synonymous groups were eliminated (i.e. SYN GROUPs identification), as previously described (Irish et al. 2010) (Tables 2, 3). A probability of identity or identity check, calculated using Cervus v3.0.3 (Marshall et al. 1998; Kalinowski et al. 2007), was conducted as previously described (Waits et al. 2001) to evaluate the discriminatory power of each locus in the study. This test, among simulated siblings from a given sample (PID-sib), provides an estimate of the probability that two individual accessions selected at random would be differentiated by their allelic profiles; it can also be used to detect identical MLGs

Table 2 Accessions from the Dominican Republic's IDIAF Mata Larga cacao (*Theobroma cacao* L.) germplasm collection with unique multi-locus genotypes (MLG) ordered based on synonymous group number (SYN GROUP no)

SYN GROUP no	Accession	SYN GROUP no	Accession	SYN GROUP no	Accession
1	§EET 250 a	2	SIC 1	8	§EET 250 b
1	§EET 95 c	2	SIC 2	8	EET 390
1	§ICS 6 b	3	EET 103	9	§EET 95 b
1	§PA 150 a	3	EET 95	9	SIAL 93
1	§PA 81 a	3	§EET 397 b	10	§IMC 67 c
1	R 117 [MEX]	4	§§EET 228 a	10	§SCA 12 b
1	R 15 [MEX]	4	§IMC 67 d	11	§NA 34 a
1	R 2 [MEX]	4	§POUND 12 b	11	§SPA 9 b
1	R 52 [MEX]	5	§ICS 1 a	12	POUND 7
1	R 75 [MEX]	5	§SPA 9 c	12	§SIC 2 b
1	§TSA 644 a	5	§UF 242 b	13	§§S 1 a
1	§UF 221 c	6	ICS 39	13	§§SGU 26 b
1	UF 676	6	§SCA 12 a	14	UF 168
1	UF 677	6	§TSH 565 b	14	UF 668
2	CATONGO	7	§§EET 228 c	15	§UF 168 b
2	SIAL 98	7	§§Silecia 5 [ECU] a	15	UF 29

Genotype name and matching fingerprint profile label are given for each accession

Lowercase letters (a, b, c, or d) indicate an accession had more than one MLG; §misidentified genotypes within SYN GROUPs based on ICGD, SPCL, TARS, or SHRS references; §§ no reference profile to compare to from ICGD, SPCL, TARS, or SHRS; accession names with no symbol are true-to-type

Table 3 Accessions from local farmer selections (LFS) of cacao (*Theobroma cacao* L.) in the Dominican Republic with unique multi-locus genotypes (MLG) ordered based on synonymous group number (SYN GROUP no)

SYN GROUP no	Accession	SYN GROUP no	Accession	SYN GROUP no	Accession
1	AH 7	4	IML 16	9	ML 106 a
1	ER 1	4	Criollo 16 b	9	ML 102 b
1	ER 5	5	IML 18	10	ML 4 [MEDIOPESO]a
1	ML 22 b	5	Criollo 18	10	ML 3 [PEPINO]
1	TCS 39	6	IML 22	11	ML 64
2	AM 2	6	Criollo 20 a	11	ML 66 b
2	JK 2	7	PL 2	12	ML 71
2	JK 3	7	JR 7 b	12	ML 73
3	IML 9	8	ML 102 a	13	SMR 1
3	Criollo 9	8	ML 106 b	13	SMR 2

Genotype name and matching fingerprint profile label are given for each accession

Lowercase letters (a or b) indicate an accession had more than one MLG

(Evett and Weir 1998; Waits et al. 2001). The probability can be obtained for both the sample under study and for an infinitely large theoretical population with the same genotype frequencies as found in the sample population (Kloosterman et al. 1993).

Comparisons to reference genotypes

Accessions from publicly available databases were used as reference genotypes; their microsatellite profiles were compared against those generated from the

Dominican Republic germplasm. The reference accessions used in this process were obtained from the International Cacao Germplasm Database (ICGD; Turnbull and Hadley 2011), USDA-ARS Sustainable Perennial Crops Laboratory (SPCL; Zhang et al. 2006a, 2008, 2009), USDA-ARS TARS (Irish et al. 2010), and the USDA-ARS SHRS (Motamayor et al. 2008; Schnell et al. 2005, 2007). Accessions from SPCL originated from collections located at CATIE and the International Cacao Collections at the Cocoa Research Unit (CRU) in Trinidad and Tobago. A MLG for each of the reference samples was extracted from each database for the 14 SSR loci used in this study. The MLG of each sample from the IDIAF germplasm collection included in the dataset with inter-accession replications removed, was compared to the MLG of the reference genotype with a corresponding accession name. Each sample with a MLG that matched the MLG of its reference genotype was considered true-to-type. Each sample with an MLG that did not match the MLG of its reference genotype was considered misidentified. Some accessions did not have a reference genotype within any of the databases with which to do a comparison and are so noted in Table 2.

Analysis of genetic diversity

Allele frequencies were estimated for each locus of the entire population (both the germplasm collection and the LFS) using Cervus v3.0.3 (Marshall et al. 1998; Kalinowski et al. 2007). Statistical analyses were carried out on the reduced dataset (66 germplasm and 115 LFS MLGs) and included alleles per locus as well as observed (H_{obs}) and expected (H_{exp}) heterozygosity values. In addition, the polymorphic information content (PIC), equal to $1 - \Sigma P_i^2$, where P_i is equal to the frequency of the ith allele at the locus, was then calculated (Powell et al. 1996). PIC values express the level of polymorphism associated with each of the loci used in the study. High PIC value loci effectively discriminate better than less informative loci.

Genetic relationships of accessions

Population structure analysis was carried out separately for the germplasm collection and the LFS using the model-based Bayesian cluster analysis software Structure v2.3.3 (Pritchard et al. 2000, 2010). Structure uses the probability Pr(X|*K*) given the data (X) and the log Pr(X|*K*) to determine the most likely number of clusters (Pritchard et al. 2000) based on all unique MLGs. All MLGs in the reduced dataset were used to infer the genetic structure of the germplasm accessions (66 profiles) and LFS individuals (115 profiles). An admixture model was used with 200,000 iterations after a burn-in period of 100,000. The two sets of samples were analyzed independently in two steps: (1) the number of clusters (*K*) tested was from 1 to 12 for the germplasm collection and from 1 to 10 for the LFS using the default option to infer the best number of clusters, and (2) both populations were then re-analyzed using the prior population information option and 20 reference samples from each of the ten previously described cacao genetic groups (Motamayor et al. 2008). The number of clusters was reset to $K = 10$ in order to assign a respective genetic group of correspondence to each of the genotypes being evaluated. In this study, an individual was considered to belong to the cluster in which it had the highest coefficient of membership (Q value). Additionally, for step one, Pritchard's ad hoc (Pritchard et al. 2000) and Evanno's (Evanno et al. 2005) methods were used to calculate the most optimal value of (*K*) in the two populations studied. Results of ten iterations of each replicated run in Structure were matched by permutation using Clumpp v1.1.2 ([Cluster Matching and Permutation Program]; Jakobsson and Rosenberg 2007); 50,000 independent runs were applied to generate the best optimum alignment over multiple runs at optimal (*K*). The Clumpp software facilitates the interpretation of population genetic clustering results based on membership coefficients of individual MLG so that the multi-modality can be detected and quantified (Jakobsson and Rosenberg 2007). Principal Component Analysis (PCA) was performed using SAS v9.2 (SAS 2010). The germplasm collection and the LFS datasets (inter- and intra-accession duplications removed) were used for PCA along with MLG data for the 200 reference samples that represent the 10 genetic groups of cacao (20 individuals per genetic group).

Results

Germplasm collection and LFS identity

Genotype data was used to identify intra-accession and inter-accession errors in the collections and to assess the rate of those errors. Based on the comparisons of MLGs, 15 SYN GROUPs were identified in the germplasm collection and 48 accessions were placed in these groups (Table 2). Although there were fewer individuals (30) assigned to SYN GROUPs in the LFS, a similar number of SYN GROUPs (13) were identified among them (Table 3). Another source of putative mislabeling is indicated by the difference between the genotype expected, based on the reference genotype, and the genotype observed for a given accession. Out of the 66 MLGs in the germplasm collection, 27 (40.9 %) were different than expected based on comparisons against reference accessions available at ICGD, SPCL, TARS, and SHRS (Supplementary Material S1). No reference profiles currently exist for the LFS. However, two or more MLG were identified for trees with the same accession name in 18.2 % of the same-labeled accessions analyzed (20 of out 110 accessions; Supplementary Material S2).

Analysis of genetic diversity

Multiple alleles were identified at each of the 14 loci. Null-alleles were not observed in the study and homozygous genotypes were called when only one allele was detected. The set of SSR markers utilized in the present research has been widely used for DNA fingerprinting, assessing diversity and for identity analysis in cacao populations (Sereno et al. 2006; Opoku et al. 2007; Zhang et al. 2006a, b, 2008, 2009; Efombagn et al. 2008; Johnson et al. 2009; Aikpokpodion et al. 2009; Motilal et al. 2010, 2011; Irish et al. 2010; Trognitz et al. 2011). These microsatellite markers have been mapped in several cacao populations and no apparent null alleles have been reported (Brown et al. 2008). The total number of alleles detected among the 14 loci was 117 in the germplasm collection and 113 in LFS, with an average of 8.36 and 8.10 alleles per locus, respectively. The number of alleles per locus ranged from four associated with mTcCIR1–12 associated with mTcCIR37 in the germplasm collection and from three associated with mTcCIR1–11 associated with markers mTcCIR6, mTcCIR33, and mTcCIR37 in the LFS (Table 1). Observed heterozygosity (H_{obs}) in the germplasm collection ranged from 0.42 for mTcCIR1 to 0.80 for mTcCIR15 and in the LFS from 0.38 for mTcCIR1 to 0.79 for mTcCIR18 with an average over all 14 loci of 0.67 and 0.60 in the germplasm collection and the LFS, respectively. Overall, gene diversity expresses the probability in which two randomly chosen alleles from the population are different. Additionally, expected heterozygosity (H_{exp}) in the germplasm collection ranged from 0.43 for mTcCIR1 to 0.83 for mTcCIR33 and mTcCIR15 and averaged 0.71 over all 14 loci; and in the LFS H_{exp} ranged from 0.44 for mTcCIR24 to 0.78 for mTcCIR15 and averaged 0.62 (Table 1). The PIC ranged from 0.35 for mTcCIR1 to 0.80 for mTcCIR33 and mTcCIR15 and had a mean of 0.67 (Table 1). In the LFS, PIC ranged from 0.37 for mTcCIR1 to 0.75 for mTcCIR15 and had a mean of 0.57 (Table 1). PIC was greater than 0.60 for 11 loci in the germplasm collection and six loci in the LFS. Highly polymorphic markers, as determined by PIC values, are useful in identifying genetically diverse germplasm. This genetic diversity is often the best resources for local cacao selection and breeding programs.

Population structure

When germplasm accessions were evaluated in terms of their relationship to the 10 representative cacao genetic groups as defined by Motamayor et al. (2008), almost half of them (43.9 %, 29 out of 66) had a predominantly Amelonado lineage (Fig. 1). Among these Amelonados, 51.7 % out of 29 were typical Trinitario hybrids. Contamana, Nacional and Iquitos genetic groups were also found in the Dominican Republic cacao germplasm, although the proportion was small (10.6 % each). These lineages comprised misidentified and true-to-type MLGs from the germplasm accessions (Fig. 1). Two other genetic groups, Criollo and Marañon, each accounted for 7.6 % of the germplasm genotypes. They consisted of true-to-type, misidentified MLGs and accessions which had no reference genotypes. Another two genetic groups, Nanay and Curaray had 6.0 and 3.0 % of the germplasm genotypes, respectively. These two lineages were comprised by true-to-type, misidentified MLGs, and accessions with no reference genotypes. Based on the highest of coefficient of membership, the majority of the LFS (72.1 %, 83 out of 115 genotypes) grouped with the Amelonado genetic group followed by Criollo 9.5 % (Fig. 2). From the 83 LFS that grouped into the Amelonado lineage, approximately 50.6 % were Trinitario hybrids. There were also 7.8 % for Iquitos, 6.0 % for Contamana, and 2.6 % for

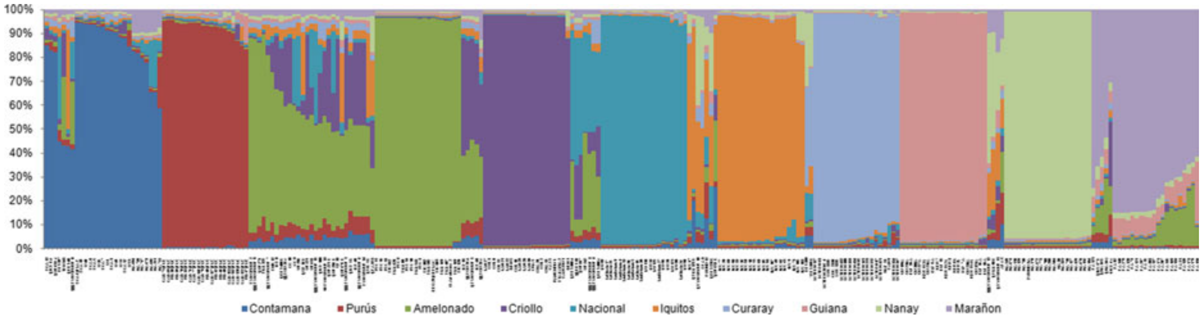

Fig. 1 Population structure of the Dominican Republic's IDIAF Mata Larga cacao (*Theobroma cacao* L.) germplasm collection and representative genotypes of the ten cacao genetic groups described by Motamayor et al. (2008) produced using Structure v2.3.3. Each individual *vertical line* represents a genotype. Admixed individuals are denoted with *multiple colors* representing the 10 representative genetic groups (see *color* key)

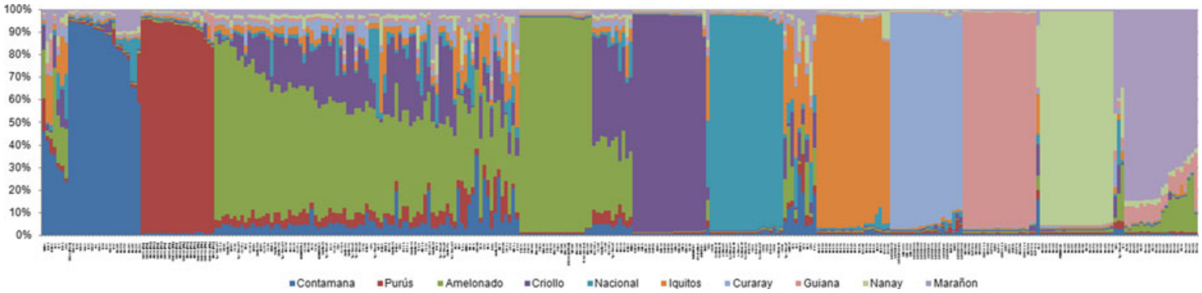

Fig. 2 Population structure of the Dominican Republic cacao (*Theobroma cacao* L.) local farmer selections (LFS) and representative genotypes of the ten cacao genetic groups described by Motamayor et al. (2008) produced using Structure v2.3.3. Each individual *vertical line* represents a genotype. Admixed individuals are denoted with *multiple colors* representing the ten representative genetic groups (see *color* key)

Marañon. Genetic groups Nacional and Nanay were least represented among the LFS with 0.87 % (Fig. 2). When the Structure analysis was carried out without inclusion of the representative genotypes of the 10 cacao genetic groups as defined by Motamayor et al. (2008), nine lineages ($K = 9$) for the germplasm collection (Supplementary Material S3) and seven ($K = 7$) for LFS (Supplementary Material S4) were identified among the genotypes comprising each of these two collections.

Discussion

Theobroma cacao was introduced to the Dominican Republic in the mid-seventeenth century. Since that time, its genetic diversity has been broadened as a result of important introduction events. The allelic diversity observed in this study of Dominican Republic populations is undoubtedly related to the introduction of cacao from several South American countries and Mesoamerica. Initial efforts were directed at bringing improved material from Venezuela and Central America. In the 19th century introductions were also made from neighboring islands in the Caribbean, especially Trinidad. According to Sea 1983, cited by Sofreco and Ecocaribe (2001), these early introductions constituted the genetic foundation to what is known as 'native' cacao in the Dominican Republic. The original admixtures were thought to be a makeup of Amelonado from Brazil, Criollo originating from Venezuela and Central America, Trinitario derived from Trinidad, and cacao Nacional from Ecuador. In 1962, six cacao hybrids were introduced from Trinidad with admixtures of Iquitos and Calabacillo (IMC), Scavina (SCA), Parinari (PA) from Upper Amazon, Trinidad Selected Amazon (TSA), and Trinidad Selected Hybrid (TSH) series. These hybrids were utilized as mother trees for crossing to a 'native' parent tree Genoveva 5 selected from a famers' field (Sofreco and Ecocaribe 2001). Several other hybrids with admixtures of Iquitos and Calabacillo (IMC), United Fruits (UF) selections, Imperial College Selections (ICS) with Criollo ancestry cultivated in Nicaragua, Scavina (SCA), and Parinari (PA) were also introduced from CATIE in 1964. Large quantities of hybrid seed (Barranca hybrid selections or SHB series) were distributed to cacao farmers as result of local breeding efforts with the aforementioned introductions (Batista 1984). The findings further support the hypothesis that the original genetic foundation of cacao resources in Dominican Republic had admixed ancestries of Amelonado, Criollo, Nacional and Trinitrios hybrids (Fig. 1). A bottle neck may have been associated when selecting for specific agronomic and disease resistant traits on more recent introductions and this may have influenced the current genetic structure of the LFS. Germplasm collections with significant amounts of genetic variation have resulted from large number of diverse accessions being introduced, such is the case for CATIE, West Africa, and Cameroon populations (Zhang et al. 2009; Aikpokpodion et al. 2009; Efombagn et al. 2008).

The allelic diversity present in the Dominican Republic germplasm collection (8.36 alleles per locus; 116 total alleles) is comparable to that identified by Irish et al. (2010) in the cacao collection maintained in Puerto Rico by USDA-ARS TARS (8.80 alleles per locus; 132 total alleles) and the allelic diversity recently identified in a young cacao production area in Nicaragua (7.73 alleles per locus; 116 total alleles; Trognitz et al. 2011). The allelic diversity observed is also similar to that reported for 'Cacao Nacional Boliviano' from the Amazonia regions of La Paz and Beni, Bolivia (7.30 alleles per locus; 110 total alleles; Zhang et al. 2012) and Ghanaian cacao collections (7.50 alleles per locus; 127 total alleles) which contain accessions collected locally and of international origin (Opoku et al. 2007). In contrast, allelic diversity is greater in the CATIE cacao collection (14.2 alleles per locus; 231 total alleles) which contains a larger number of accessions from a much broader geographic area (Zhang et al. 2009). Collections in West Africa (12.5 alleles per locus; 144 total alleles; Aikpokpodion et al. 2009) and Cameroon (9.41 alleles per locus; 125 total alleles; Efombagn et al. 2008) are also highly diverse. On the other hand, Loor et al. (2009) reported a low amount of genetic diversity in accessions collected along the Pacific coast of Ecuador (4.22 alleles per locus; 169 total alleles). This report is similar in the number of alleles per locus for a 'Refractario' population in Ecuador but different for the total number of alleles (4.20 alleles per locus; 63 total alleles; Zhang et al. 2008). Low allelic diversity was also reported for Peruvian collections from the Huallaga (3.68 alleles per locus) and Ucayali (5.7 alleles per locus) valleys with 161 total alleles (Zhang et al. 2006a). Sereno et al. (2006) described similar low genetic diversity values (4.45 alleles per locus; 49 total alleles) in accessions collected from 19 Amazon River basins, a result likely related to physical isolation and fewer admixtures among wild populations.

High levels of heterozygosity ($H_{exp} = 0.71$) in the Dominican Republic germplasm collection and in the LFS ($H_{exp} = 0.62$) were detected. The high gene diversity indicates substantial levels of admixture present in the gene pool for the germplasm collection. The observed gene diversity (H_{exp}) in the germplasm collection is only lower than that reported for a population in Ucayali Valley in Peru ($H_{exp} = 0.74$; Zhang et al. 2006a) and from Ghana ($H_{exp} = 0.74$; Opoku et al. 2007). Conversely, it is higher than that described for the germplasm collection in Puerto Rico ($H_{exp} = 0.66$; Irish et al. 2010), and in Nicaragua ($H_{exp} = 0.476$; Trognitz et al. 2011). Similarly, it is also higher than that found in a population in the Huallaga valley in Peru ($H_{exp} = 0.61$; Zhang et al. 2006a), for the Ecuadorian Refractario ($H_{exp} = 0.561$; Zhang et al. 2008), and Nacional populations ($H_{exp} = 0.496$; Loor et al. 2009), the population in Bolivia ($H_{exp} = 0.56$; Zhang et al. 2012), Cameroon ($H_{exp} = 0.55$; Efombagn et al. 2008), CATIE ($H_{exp} = 0.51$; Zhang et al. 2009), and Brazil ($H_{exp} = 0.497$; Sereno et al. 2006). A higher number of private alleles were detected in the germplasm collection (25) compared to the LFS (15). This significant difference is due to the fact that diverse clonal introductions were made from countries regarded as being within the center of origin for cacao, where the greatest genetic diversity is found (Cheesman 1944; Cuatrecasas 1964; Pound 1945; Coe and Coe 1996; Motamayor et al. 2008). Selection pressure on the LFS may also explain this difference, creating a bottleneck for alleles present in the germplasm collection and not detected in the LFS. For years, Dominican Republic cacao farmers have been selecting elite trees with high quality beans of superior organoleptic characteristics from their own fields (i.e., LFS). To our knowledge, this is the first effort to characterize the Dominican Republic cacao genetic resources using microsatellite markers.

In addition to estimating its genetic diversity, the current study provides a comprehensive examination of the population structure and mislabeling errors in the cacao germplasm collection at the Dominican Republic's IDIAF Mata Larga research station. The error rate among the germplasm accessions was 40.9 %. This percentage is consistent with other studies conducted to characterize and define clonal *T. cacao* germplasm collections in other countries (Motilal and Butler 2003; Turnbull et al. 2004) although it differs, perhaps related to sampling size or methodologies employed, from others (Zhang et al. 2009; Irish et al. 2010; Trognitz et al. 2011). Several mislabeled genotypes tagged as 'EET', 'UF', 'NA', 'ICS', 'TSA', and 'PA' clonal series belonging to the Nacional, Amelonado and Criollo backgrounds, were identified. A number of these clones were identified as mislabeled in other collections as well (Irish et al. 2010) suggesting that these errors could have originated at the time of introduction into the Dominican Republic. MLGs for those accessions in the Dominican Republic's germplasm collection with no reference genotype can be now incorporated into the ICGD to serve as references in future studies. This does not implicitly mean that the genotype for a particular accession is correct, only that a MLG is available for comparison. This study also documents the first time that many of the LFS have been fingerprinted and the genotypes generated for those accessions (deposited into the ICGD) can serve as the reference profiles from this point forward. Comparison to reference profiles has become an important tool in delineating and helping to correct the mislabeling and duplicating errors that inevitably occur during maintenance of germplasm collections (Irish et al. 2010) and has allowed us to correctly identify the clonal germplasm trees at the IDIAF Mata Larga research station. However, caution must be advised when basing clone/accession identification solely on a MLG from a limited number of markers. Phenotypic characterization needs to accompany the MLG data to confidently identify an accession. Efforts are underway at Mata Larga to mark true-to-type clones in old germplasm plots and new clonal plantings, from correctly identified clonal plots, are being established at multiple sites. The 17.4 % of the LFS that were identified as mislabeled trees would significantly affect a breeding program by reducing possible genetic gains in the progeny when misidentified parents are used in breeding.

The software Structure (Pritchard et al. 2000) has been widely used for Bayesian clustering analysis, which assigns groups of individuals based on coefficients of population membership and to determine the degree of admixture or allelic contributions within a population (Kaeuffer et al. 2007; Rosenberg et al. 2002; Kalinowski 2010). When the 66 MLGs in the germplasm collection were analyzed together with representative genotypes of the ten genetic groups previously reported by Motamayor et al. (2008), each individual had a highest coefficient of membership to one of the ten genetic groups. Eight of the 10 genetic groups were represented within the germplasm collection and a majority of accessions corresponded to the Amelonado, Contamana, Nacional and Criollo lineages. Seven of the 10 genetic groups were represented within the LFS as no MLG with highest coefficient of membership for the Guiana, Purús, or Curaray genetic groups was identified among the LFS (Fig. 2). The highest coefficient of membership for the majority of the LFS corresponded to the Amelonado, Criollo, Iquitos, and Contamana lineages (Fig. 2). These four lineages are important among the Dominican Republic's cacao genetic resources since they have been historically associated with high quality chocolate processing traits valued by the global market. The Dominican Republic has both a niche production for organic cacao and a good reputation for producing conventionally grown cacao associated with flavor, aroma and fat content that have high demand worldwide (http://www.conacado.com.do/site/index.php?lang=en). Modern Criollo cacao accessions have traditionally been associated with the production of chocolate with milder, fine and nutty flavors (Elwers et al. 2009; Ed Seguine, Mars, Inc., Hackettstown, NJ, personal communication 2011) while the Amelonado types are characterized by more intense cacao flavors (Ed Seguine, Mars, Inc., Hackettstown, NJ, personal communication 2011). The Nacional (Ecuadorian) genetic identities have been linked to chocolate with good organoleptic qualities and fine floral aroma characteristics (Deheuvels et al. 2004; Loor et al. 2009). These organoleptic quality traits, together with the organic production component, are current triggers of competitive international markets (Eskes and Lanaud 1997; Rusconi and Conti 2010).

Originally, Structure analysis was carried out without inclusion of representative genotypes corresponding to the ten genetic groups. Significant levels of admixture were detected in the cacao germplasm collection, however only nine clusters were identified (Supplementary Material S3). The 66 MLGs retained in the study after removal of duplicates and mislabels is a small number compared to the 952 individuals that were analyzed by Motamayor et al. (2008). The germplasm analysis was also limited to an accession collection from only one location and by the small number of SSR markers that were used, especially as compared to the large geographic area covered and the greater number of SSR markers (106) utilized by Motamayor et al. (2008). When the germplasm collection and the ten genetic groups were analyzed together (Fig. 1; Supplementary Material S3), none of the MLGs in the cacao germplasm collection had a highest coefficient of membership for the Guiana or Purús genetic groups. Although, in two cases, ‘POUND 12 [POU]’ and ‘APA 5’, a low admixture of the Guiana and Purús backgrounds was identified, those data comprised the only evidence indicating that these genetic lineages had been imported into the country (Supplementary Material S3). High levels of admixture, like those observed in most of the individuals in the germplasm accessions and LFS in the current study (Figs. 1, 2; Supplementary Materials S3–S4), are common in *T. cacao* (Cheesman 1944; Cuatrecasas, 1964; Motamayor et al. 2008; Schnell et al. 2005; Zhang et al. 2006a, 2008). The additional ninth cluster detected in the cacao germplasm collection corresponds to the Trinitario group. Trinitario is a hybrid group that is comprised primarily of the Amelonado and Criollo lineages (Motamayor et al. 2003) and it was excluded from the Motamayor et al. (2008) study. The substantial levels of admixture between the Amelonado and Criollo groups must be the result of hybridization and recombination events that had taken place in cacao breeding nurseries and farms prior to their introduction in the Dominican Republic. The International Cocoa Genebank in Trinidad has been a repository of clonal and hybrid material for distribution worldwide (Johnson et al. 2009; Motilal et al. 2010). The Dominican Republic has relied heavily upon this source when importing cacao germplasm to the country. Imperial College Selections (ICS) have Trinitario ancestry presumably introgressed from either Guyana, Venezuela or the Lower Amazon (Cheesman 1944; Bartley 2005; Motamayor et al. 2002) and Criollo admixture from Nicaragua (Sofreco and Ecocaribe 2001). Furthermore, ICS Trinitario appears to have introgression from Upper Amazon as well (Motilal et al. 2010). More recent Trinidad Selected Hybrids (TSH), which are present in the Dominican Republic germplasm collection, were derived from selected Peruvian Upper Amazon cacao which are resistant to *Moniliophthora perniciosa* (Stahel) Aime & Phillips-Mora, the causal agent of witches’ broom disease (Aime and Phillips-Mora 2005) and contain additional resistance to *Moniliophthora roreri* H.C. Evans, Stalpers, Samson & Benny, the causal agent of frosty pod rot or moniliasis disease (Phillips-Mora et al. 2005). This substantial genetic base observed in introgressed Trinitario in the Dominican Republic explains the predominant diversity and admixture detected in the cacao germplasm collection and LFS. In addition to the typical Trinitario lineage in the germplasm collection, admixtures of Amelonado of Lower Amazon origin and Nacional of Ecuador; Amelonado and Iquitos Mezcla Calabacillo (IMC) from Peru; Amelonado, Iquitos and Nanay from Peru were found (Fig. 1). For LFS, the Trinitario hybrid was also detected (Fig. 2). However, it appears that a number of LFS have multiple admixtures other than Trinitario lineage. Such is the case for mixture of Amelonado, Nacional and Criollo (Lacadona and Santa Marta from Mexico and Venezuela, respectively); and/or admixtures including one or two mentioned before either with Scavina (Contamana), Parinari (Marañon), Acre R (Purús), or Camopi R (Guiana) from Peru, Brazil, and French Guyana, respectively.

A PCA was conducted to distinguish and separate subgroups within the germplasm accessions, LFS and the ten genetic groups, combined (Fig. 3). Although a small percentage of the total variation for the PCA was captured in this study, a similarity in grouping patterns was observed between the PCA and the population structure analysis conducted using Structure. A number of the germplasm accessions and the LFS clustered together on the right quadrant (Fig. 3). This group of accessions corresponded to Trinitario group in both collection. Also, a continuous variation can also be observed with the germplasm collection, LFS and the

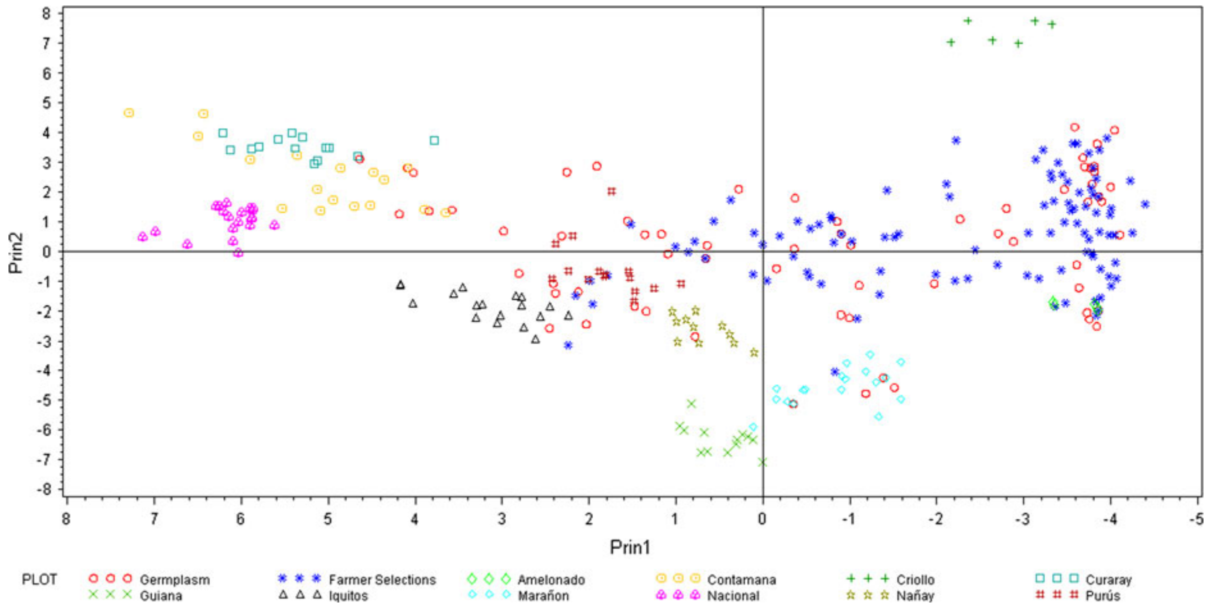

Fig. 3 Principal Component Analysis for 66 cacao (*Theobroma cacao* L.) germplasm accessions belonging to the Dominican Republic's IDIAF Mata Larga station collection, 115 local farmer selections (LFS), and genotypes representing the ten cacao genetic groups described by Motamayor et al. (2008). The first and second principal component axis accounted for 7.0 and 5.8 % of the total variation respectively. See legend for key to the groups included in this analysis

genetic groups utilized from the Motamayor et al. (2008). The degree of genetic diversity and thus the admixture level observed as a result of both analyses is consistent with the diverse introduction of material that has taken place in the country. Even though the number of accessions is small, these results confirmed a high level of error associated with clonal propagation and relatively high genetic diversity and levels of admixture in both the cacao germplasm collection and the LFS. The knowledge and understanding of the genetic diversity, population structure, and degree of admixture in this collection will greatly enhance the ability to select for elite cacao trees. Accessions in the Dominican Republic germplasm collection have been reported to be resistant to black pod (*Phytophthora* spp.), witches' broom, frosty pod rot, or ceratocystis wilt (*Ceratocystis fimbriata*/*C. cacaofunesta*) in other countries (Turnbull and Hadley 2011). Among them 'SCA 12', 'POUND 7 [POU]', 'POUND 12 [POU]', 'EET 95 [ECU]', 'PA 121', 'SPA 9', 'TSH 565', and 'SNK 12' may constitute a group of accessions that, along with elite LFS, would be useful in a national breeding and selection program. The Dominican Republic is free of some of the more important diseases that affect cacao globally (http://www.conacado.com.do/site/index.php?lang=en) and the country should continue to strictly enforce quarantine regulations on cacao imports from other areas of the world. But, it would be prudent to utilize the disease-resistant accessions that already exist, as well as add accessions with potential for introgressing desirable disease-related traits to the existing germplasm, in a national breeding program. Cacao breeding for disease resistance, an environmentally and economically sound long term strategy, along with agronomic characteristics and desirable organoleptic properties will improve local cacao production and the productivity of several thousand cacao farms in the country.

Acknowledgments The authors would like to thank Land O'Lakes, Inc., International Development, for their collaborative efforts and partial financial support through the USAID-funded Cooperative Development Program (CDP). We also would like to thank MARS, Inc., for funding Trust Agreement #58-6631-6-123: Genetic Improvement of *Theobroma cacao*, and the USDA-ARS SHRS cacao program for their partial financial support. We would like to extend our appreciation to Drs. Belinda Martineau and Tomás Ayala-Silva for their constructive comments and suggestions to the manuscript.

References

Aikpokpodion PO, Motamayor JC, Adetimirin VO, Adu-Ampomah Y, Ingelbrecht I, Eskes AB, Schnell RJ, Kolesnikova-Allen M (2009) Genetic assessment of sub-samples of cacao, *Theobroma cacao* L. collections in West Africa using simple sequence repeats marker. Tree Genet Genomes 5:699–711

Aime MC, Phillips-Mora W (2005) The causal agents of witches' broom and frosty pod rot of cacao (chocolate, *Theobroma cacao*) form a new lineage of Marasmiaceae. Mycologia 97:1012–1022

Alverson WS, Whitlock BA, Nyffeler R, Bayer C, Baum DA (1999) Phylogeny of the core Malvales: evidence from ndhF sequence data. Am J Bot 86:1474–1486

Argout X, Salse J, Aury JM et al (2010) The genome of *Theobroma cacao*. Nat Genet 43:101–108

Bartley BGD (2005) The genetic diversity of cacao and its utilization. CABI Publishing, Wallingford

Batista L (1984) Progreso en 10 años de investigacion en el mejoramiento genetic del cacao en República Dominicana. In: Proceedings international cocoa research conference, Lomé, Togo

Batista L (2009) Guía técnica del cultivo de cacao en República Dominicana. Ministerio de Agricultura, Santo Domingo

Bekele F, Butler DR (2000) Proposed list of cocoa descriptors for characterization. In: Eskes AB, Engels JMM, Lass RA (eds) Working procedures for cocoa germplasm evaluation and selection. Proceedings of the CFC/ICCO/IPGRI project workshop, Montpellier, France, 1–6 Feb 1998. IPGRI, Montpellier, pp 41–48

Borrone JW, Meerow AW, Kuhn DN, Whitlock BA, Schnell RJ (2007) The potential of the WRKY gene family for phylogenetic reconstruction: an example from the Malvaceae. Mol Phylogenet Evol 44:1141–1154

Brown JS, Sautter RT, Olano CT, Borrone JW, Kuhn DN, Motamayor JC, Schnell RJ (2008) A composite linkage map from three crosses between commercial clones of cacao, *Theobroma cacao* L. Trop Plant Biol 1:120–130

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

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

Cope FW (1984) Cacao *Theobroma cacao* (Sterculiaceae). In: Simmonds NW (ed) Evolution of crops plants. Longman, London, pp 285–289

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. Contributions from the United States Nacional Herbarium, vol 35. Smithsonian Institution Press, Washington, pp 375–614

Dantas LG, Guerra M (2010) Chromatin differentiation between *Theobroma cacao* L. and *T. grandiflorum* Schum. Genet Mol Biol 33:94–98

Deheuvels O, Decazy B, Perez R, Roche G, Amores F (2004) The first Ecuadorean 'Nacional' cocoa collection based on organoleptic characteristics. Trop Sci 44:23–27

Efombagn IBM, Motamayor JC, Sounigo O, Eskes AB, Nyassé S, Cilas C, Schnell R, Manzanares-Dauleux MJ, Kolesnikova-Allen M (2008) Genetic diversity and structure of farm and GenBank accessions of cacao (*Theobroma cacao* L.) in Cameroon revealed by microsatellite markers. Tree Genet Genomes 4:821–831

Elwers S, Zambrano A, Rohsius C, Lieberei R (2009) Differences between the content of phenolic compounds in Criollo, Forastero and Trinitario cocoa seed (*Theobroma cacao* L.). Eur Food Technol 229:937–948

Engels JMM (1983) A systematic description of cacao clones. III. Relationships between clones, between characteristics and some consequences for the cacao breeding. Euphytica 32:719–733

Eskes A, Lanaud C (1997) Cocoa. In: Charrier A (ed) Tropical plant breeding. Montpellier, France, pp 78–105

Estadísticas del Departamento de Cacao en República Dominicana (2010) Ministerio de Agricultura. Santo Domingo, República Dominicana

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

Evett IW, Weir BS (1998) Interpreting DNA evidence: statistical genetics for forensic scientists. Sinauer, Sunderland

FAOSTAT (2008) http://faostat.fao.org/site/339/default.aspx. Verified 10 Jan 2011. FAO, Rome

Henderson JS, Joyce RA, Hall GR, Hurst WJ, McGovern PE (2007) Chemical and archeological evidence for the earliest cacao beverages. Proc Natl Acad Sci USA 104:18937–18940

Hunter RJ (1990) The status of cocoa (*Theobroma cacao*, Sterculiaceae) in the western hemisphere. Econ Bot 44:425–439

Informe Anual del Departamento de Cacao (2006) Secretaría de Estado de Agricultura. Santo Domingo, República Dominicana

Irish BM, Goenaga R, Zhang D, Schnell RJ, Brown JS (2010) Microsatellite fingerprint of the USDA-ARS Tropical Agriculture Research Station cacao (*Theobroma cacao* L.) germplasm collection. Crop Sci 50:656–667

Iwaro AD, Bekele FL, Butler DR (2003) Evaluation and utilization of cacao (*Theobroma cacao* L.) germplasm at the International Cocoa Genebank, Trinidad. Euphytica 130: 207–221

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806

Johnson ES, Bekele FL, Brown SJ, Song Q, Zhang D, Meinhardt LW, Schnell RJ (2009) Population structure and genetic diversity of the Trinitario cacao (*Theobroma cacao* L.) from Trinidad and Tobago. Crop Sci 49:564–572

Kaeuffer R, Reale D, Coltman DW, Pontier D (2007) Detecting population structure using STRUCTURE software: effect of background linkage disequilibrium. Heredity 99:374–380

Kalinowski ST (2010) The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. Heredity 2010:1–8

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106

Kloosterman AD, Budowle B, Daselaar P (1993) PCR-amplifications and detection of the human D1S80 locus. Int J Leg Med 105:257–264

Knight R, Rogers HH (1953) Sterility in *Theobroma cacao* L. Nature 172:164

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

Lanaud C, Hammon CP, Duperray C (1992) Estimation of the nuclear DNA content of *Theobroma cacao* by flow cytometry. Cafe Cacao 36:3–8

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

Lerceteau E, Quiroz J, Soria J, Flipo S, Pétiard V, Crouzilat D (1997) Genetic differentiations among Ecuadorian *Theobroma cacao* L. accessions using DNA and morphological analyses. Euphitica 95:77–87

Loor RG, Risterucci AM, Courtois B, Fouet O, Jeanneau M, Rosenquist E, Amores F, Vasco A, Medina M, Lanaud C (2009) Tracing the native ancestors of the modern *Theobroma cacao* L. population in Ecuador. Tree Genet Genome 5:421–433

Marshall TC, Slate J, Kruuk L, Pemberton JJ (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639–655

Motamayor JC, Risterucci AM, Lopez PA, Ortiz CF, Moreno A, Lanaud C (2002) Cacao domestication I: the origin of the cacao cultivated by the Mayas. Heredity 89:380–386

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

Motamayor JC, Lachenaud P, e Mota JW, Loor R, Kuhn DN, Brown JS, Schnell RJ (2008) Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L). PLoS One 3:10

Motilal LA, Butler D (2003) Verification of identies in global cacao germplasm collections. Genet Resour Crop Evol 50:799–807

Motilal LA, Zhang D, Umaharan P, Mischke S, Mooleedhar V, Meinhardt LW (2010) The relic Criollo cacao in Belize—genetic diversity and relationship with Trinitario and other cacao clones held in the International Cacao Genebank, Trinidad. Plant Genet Resour Charact Util 8:106–115

Motilal LA, Zhang D, Umaharan P, Mischke S, Pinney S, Meinhardt LW (2011) Microsatellite fingerprinting in the International Cocoa Genebank, Trinidad: accession and plot homogeneity information for germplasm management. Plant Genet Resour Charact Util 9:430–438

Opoku SY, Bhattacharjee R, Kolesnikova-Allen M, Motamayor JC, Schnell R, Ingelbrecht I, Enu-Kwesi L, Adu-Ampomah Y (2007) Genetic diversity in Cocoa (*Theobroma cacao* L.) germplasm collection from Ghana. J Crop Improv 20:73–87

Phillips-Mora W, Castillo J, Krauss U, Rodriguez E, Wilkinson MJ (2005) Evaluation of cacao (*Theobroma cacao*) clones against seven Colombian isolates of *Moniliophthora roreri* from four pathogen genetic groups. Plant Pathol 54:483–490

Plan Operativo Anual (2007) Instituto Dominicano de Investigaciones Agrícolas y Forestales (IDIAF). Programa Nacional de Cacao, Santo Domingo

Pound FJ (1945) A note on the cacao population of South America. In: Report and proceedings of the cacao research conference held at colonial office. The Colonial Office, His Majesty's Stationery Office, London, May–June 1945, pp 131–133

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breeding 2:225–238

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

Pritchard JK, Wen X, Falush D (2010) Documentation for STRUCTURE software: version 2.3. http://pritch.bsd.uchicago.edu/structure_software/release_versions/v2.3.3/structure_doc.pdf. Verified 15 Feb 2011

Rice RA, Greenberg R (2000) Cacao cultivation and the conservation of biological diversity. Ambio 29:3

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. Science 298:2381–2385

Royaert S, Phillips-Mora W, Arciniegas Leal AM, Carriaga K, Brown JS, Kuhn DN, Schnell RJ, Motamayor JC (2011) Identification of marker-trait associations for self-compatibility in a segregating mapping population of *Theobroma cacao* L. Tree Genet Genomes 7:1159–1168

Rusconi M, Conti A (2010) *Theobroma cacao* L., the food of the Gods: a scientific approach beyond myths and claims. Pharm Res 61:5–13

SAS Institute Inc. (2010) Version 9.2. SAS Institute Inc, Cary

Saunders JA, Mischke S, Leamy EA, Hemeida AA (2004) Selection of international molecular standards for DNA fingerprinting of *Theobroma cacao* L. Theor Appl Genet 110:41–44

Schnell RJ, Olano CT, Brown JS, Meerow AW, Cervantes-Martinez C, Nagami C, Motamayor JC (2005) Retrospective determination of the parental population of superior cacao (*Theobroma cacao* L.) seedlings and association of microsatellites alleles with productivity. J Am Soc Hortic Sci 130:181–190

Schnell RJ, Kuhn DN, Brown JS, Olano CT, Phillips-Mora W, Amores FM, Motamayor JC (2007) Development of a marker assisted selection program for cacao. Phytopathology 97:1664–1669

Sereno ML, Albuquerque PSB, Vencovsky R, Figueira A (2006) Genetic diversity and natural population structure of cacao (*Theobroma cacao* L.) from a the Brazilian Amazon evaluated by microsatellite markers. Conserv Genet 7:13–24

Sofreco, Ecocaribe (2001) Proyecto piloto de mejoramiento de la producción y comercialización del cacao en República Dominicana. Secretariado Técnico de la Presidencia, República Dominicana

Steinberg MK (2002) The globalization of a ceremonial tree: the case of cacao (*Theobroma cacao*) among the Mopan Maya. Ecol Bot 56:58–65

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

Trognitz B, Scheldeman X, Hansel-Hohl K, Kuant A, Grebe H, Hermann M (2011) Genetic population structure of cacao plantings within a young production area in Nicaragua. PLoS One 6:1

Turnbull CJ, Hadley P (2011) International Cocoa Germplasm Database (ICGD). [Online Database]. NYSE Liffe/CRA Ltd./University of Reading, UK. Available: http://www.icgd.reading.ac.uk. Verified 21 Apr 2011

Turnbull CJ, Butler DR, Cryer NC, Zhang D, Lanaud C, Daymond AJ, Ford CS, Wilkinson MJ, Hadley P (2004) Tackling mislabeling in cocoa germplasm collections. INGENIC Newsl 9:8–11

Ventura-López M, González A, Batista L (2006) Selección de arboles de cacao (*Theobroma cacao* L.) nativo y híbrido de buen rendimiento y con indicadores de calidad. In: Eskes AB, Efron Y, End MJ, Bekele F (eds) Proceeding of the international workshop on cocoa breeding for farmer's needs. INGENIC, San Jose, Costa Rica, 15–17 Oct 2006

Waits LP, Luikart G, Taberlet P (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol Ecol 10:249–256

Zhang D, Arevalo-Gardini E, Mischke S, Zúñiga-Cernades L, Barreto-Chavez A, Adriazola del Aguila J (2006a) Genetic diversity and structure of managed and semi-natural populations of cocao (*Theobroma cacao*) in the Huallaga and Ucayali valleys of Peru. Ann Bot 98:647–655

Zhang D, Mischke S, Goenaga R, Hemeida AA, Saunders JA (2006b) Accuracy and reliability of high-throughput microsatellite genotyping for cacao clone identification. Crop Sci 46:2084–2092

Zhang D, Boccara M, Lambert M, Butler DR, Umaharan P, Mischke S, Meinhardt L (2008) Microsatellite variation and population structure in the "Refractario" cacao of Ecuador. Conserv Genet 9:327–337

Zhang D, Mischke S, Johnson ES, Phillips-Mora W, Meinhardt L (2009) Molecular characterization of an international cacao collection using microsatellite markers. Tree Genet Genomes 5:1–10

Zhang D, Martinez WJ, Johnson WS, Somarriba E, Phillips-Mora W, Astorga C, Mischke S, Meinhardt LW (2012) Genetic diversity and spatial structure in a new distinct *Theobroma cacao* L. population in Bolivia. Genet Resour Crop Evol 59:239–252