RESEARCH

Microsatellite Fingerprinting of the USDA-ARS Tropical Agriculture Research Station Cacao (*Theobroma cacao* L.) Germplasm Collection

Brian M. Irish,* Ricardo Goenaga, Dapeng Zhang, Raymond Schnell, J. Steve Brown, and Juan Carlos Motamayor

ABSTRACT

Cacao (*Theobroma cacao* L.) is an important cash crop in many tropical countries. Cacao accessions must be propagated vegetatively to conserve genetic integrity due to its allogamous nature and its seed recalcitrance (lack of dormancy). Therefore, cacao germplasm is usually maintained as living trees in field collections and has resulted in varying rates of misidentification and duplication. Using a high throughput genotyping system with 15 microsatellite loci, all 924 trees in the USDA-ARS Mayaguez cacao collection were fingerprinted. Nineteen accessions (12.3%) were found to have intraplant errors while 14 (9.1%) synonymous sets were identified that included replicates of 49 accessions. The average number of alleles (8.8; SE = 0.56) and gene diversity (H_{Obs} = 0.65; SE = 0.026) indicate a high allelic diversity in this collection. A distance-based cluster analysis and a Bayesian assignment test showed that the cacao accessions can be classified into four distinct clusters, with their geographical origins covering most of the cacao growing regions in the Americas. Assessment of the representative diversity of the collection led to the identification of several genetic gaps, including underrepresented genetic populations and particular traits of economic and agronomic value. The improved understanding of identities and structure in the USDA-ARS cacao collection will contribute to more efficient use of cacao in conservation and breeding.

B.M. Irish and R. Goenaga, USDA-ARS, Tropical Agriculture Research Station, 2200 P. A. Campos Ave., Suite 201, Mayaguez, PR 00680; D. Zhang, USDA-ARS, Sustainable Perennial Crops Lab., 1300 Baltimore Ave., Bldg. 50 BARC-W, Beltsville, MD 20705; R. Schnell and J.S. Brown, USDA-ARS, Subtropical Horticulture Research Station, 13601 Old Cutler Rd., Miami, FL 33158; J.C. Motamayor, Mars, Inc., c/o USDA-ARS Subtropical Horticulture Research Station, 13601 Old Cutler Rd., Miami, FL 33158. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. Received 12 June 2009. *Corresponding author (Brian.Irish@ars.usda.gov).

Abbreviations: AFLP, amplified fragment length polymorphism; CATIE, Centro Agronómico Tropical de Investigación y Enseñanza; CIRAD, Centre de Coopération Internationale en Recherche Agronomique pour le Dévelopement; CRU, Cocoa Research Unit; PID, probability of identity; RAPD, random amplified polymorphic DNA; RFLP, restriction fragment length polymorphism; TARS, Tropical Agriculture Research Station; UPGMA, unweighted pair group method with arithmetic mean.

In the Malvaceae family, cultivated cacao (*Theobroma cacao* L.) is one of the most important cash crops grown in tropical regions, mostly in developing nations. Production estimates indicate that more than 4.0 million metric tons of commercial cacao beans were produced in 2007 (FAOSTAT, 2007). The bulk of the crop is produced in Western Africa, with Republic of Côte d'Ivoire and Ghana producing 1,300,000 and 690,000 MT in 2007, respectively, and ranking first and second in worldwide production. Other important cacao producing countries include Indonesia

Published in Crop Sci. 50:656–667 (2010).
doi: 10.2135/cropsci2009.06.0299

(620,000 MT), Nigeria (500,000 MT), Brazil (221,699 MT), and Cameroon (179,239 MT) (FAOSTAT, 2007).

Genetic erosion of cultivated tropical and subtropical fruit crop species has become a paramount problem worldwide. Natural disasters, environmental changes, disease and insect pests, changing intellectual property rights and genetic resources legislation, political unrest, and lack of financial support for collection, research, and maintenance of germplasm collections have all led to a decline in accessibility to valuable plant germplasm (Gepts, 2006). Currently, commercially cultivated cacao is composed of a narrow genetic base and many cultivars are susceptible to numerous damaging insects and diseases of commercial importance (Motamayor et al., 2002, 2003; Bennett, 2003). Some of the most economically important diseases and insect pests include black pod (*Phytophthora* spp.), *Cacao swollen-shoot virus* (Willson, 1999) vectored by sap sucking capsids/mirids (several genera and species) and mealybugs (*Pseudococcidae* spp.), witches' broom [*Moniliophthora perniciosa* (Stahel) Aime and Phillips-Mora], frosty pod [*M. roreri* (Cif.) H.C. Evans et al.], and the cocoa pod borer [*Conopomorpha cramerella* (Snelling)]. Witches' broom and frosty pod diseases are only found in the Americas (Bowers et al., 2001; Schnell et al., 2007), whereas *Phytophthora megakarya* Brasier and M.J. Griffin, an aggressive species causing black pod (Ducamp et al., 2004), and *Cacao swollen-shoot virus* are confined to the African continent. If these aforementioned cacao pests were to spread to currently noninfested continents, the negative impact on cacao production and availability would be significant (Bowers et al., 2001; Schnell et al., 2007).

Pest management techniques that have focused on cultural practices and pesticide use have had marginal results, suggesting that the best method for pest management is the incorporation of resistance. Breeding for pest and disease resistance in cacao has had only moderate success due to the lack of well-developed screening procedures and the lack of readily available resistant germplasm (Ploetz, 2007). This has led to an increased interest in the evaluation of existing germplasm collections and the acquisition of cacao genotypes in their centers of origin or "wild" germplasm in the hope of identifying new sources of resistance (Giron et al., 2004).

In general, germplasm collections are difficult to manage and maintain due to the large numbers of individual accessions. Mislabeling of cacao accessions has been found to be one of the principal problems in clonal germplasm collections with some estimates of mislabeling reaching 40% (Saunders et al., 2001; Sounigo et al., 2001; Motilal and Butler, 2003; Turnbull et al., 2004). Cacao may be propagated from seed, but due to the seed's recalcitrant (lack of dormancy) nature (Vanitha et al., 2005) and because the seed lacks the ability to produce plants that are true-to-type, cacao must be propagated via grafting. Traditionally, the identification of accessions relied on a few phenotypic traits that could assist in distinguishing accessions (Engels et al., 1980; Bekele and Butler, 2000; Bartley, 2005; Bekele et al., 2006). However, accurate genotype identification based on morphological traits has proven difficult, even for trained individuals.

DNA fingerprinting techniques (restriction fragment length polymorphisms [RFLPs], random amplified polymorphic DNA [RAPD], amplified fragment length polymorphisms [AFLPs], microsatellites, single nucleotide polymorphisms, sequencing, etc.) allow rapid and accurate identification of accessions in germplasm collections. Several of these molecular biology techniques have been applied successfully to distinguish cacao genotypes, including RAPDs (Leal et al., 2008) and AFLPs (Perry et al., 1998). More recently, efforts have focused on the use of microsatellite markers, also known as simple sequence repeats, for germplasm characterization (Fregene et al., 2003; Volk et al., 2006; Kameswara et al., 2007) because of their reproducibility, codominant nature, versatility, and amenability to high throughput. In cacao germplasm characterization, an internationally accepted group of 15 microsatellite primers has been advocated for fingerprinting germplasm worldwide (Swanson et al., 2003; Saunders et al., 2004; Cryer et al., 2006; Zhang et al., 2006a, 2006b, 2008, 2009). Microsatellite primers were chosen based on the relatively high number of allelic polymorphisms generated at each locus and their distribution across chromosomes. While 15 microsatellite markers are usually sufficient to differentiate cacao accessions, Cervantes-Martinez et al. (2006) showed that a higher number of markers per linkage group (approximately 10) is required to enable reliable inferences of genetic variance on the entire genome.

The USDA-ARS, Tropical Agriculture Research Station (TARS) in Mayaguez, PR, is part of the National Plant Germplasm System and is the primary site for maintenance and evaluation of the USDA cacao germplasm collection. As such, our objectives were to utilize microsatellite markers to fingerprint all accessions in the current cacao collection with the goal of using the fingerprint profiles to (i) verify the genetic identity of the cacao accessions, (ii) determine the degree of mislabeling within accessions, (iii) estimate the genetic diversity in the USDA-ARS collection, and (iv) identify potential diversity gaps.

MATERIALS AND METHODS

Plant Material and DNA Extraction

The current USDA-ARS cacao germplasm collection consists of 154 clones located on the TARS grounds in Mayaguez, PR. The trees were planted in a randomized complete block design with three blocks and two trees per block for a total of 924 trees. Five leaves from each tree were collected and frozen at −20°C. DNA was extracted using a Fast DNA SPIN Kit (MP Biomedicals, Irvine, CA) as described by Schnell et al. (2005).

Polymerase Chain Reaction and Microsatellites

Fifteen microsatellite primer combinations were used in this study. All primers were originally designed and produced at the Centre de Coopération Internationale en Recherche Agronomique pour le Dévelopement (CIRAD), Montpellier, France and have been chosen as the international set for fingerprinting cacao germplasm (Saunders et al., 2004). Forward primers were labeled with one of three fluorescent dyes (Applied Biosystems, Foster City, CA) on the 5′ end. The 15 primer pairs were used to genotype all individuals in 10 μL polymerase chain reaction amplifications as described by Schnell et al. (2005).

Electrophoresis

Capillary electrophoresis was performed on an ABI Prism 3730 Genetic Analyzer (Applied Biosystems) as described by Schnell et al. (2005). Electrophoresis results were analyzed with GeneMapper 3.0 software (Applied Biosystems) to determine alleles as well as for internal standard and fragment size determination.

Identification of Duplicates and Mislabeling

For the purpose of this study, three types of mislabeling in this collection were defined. The first mislabeling error type was homonymous or "intraplant error," meaning that trees had the same name in this collection but different multilocus microsatellite genotypes. The second was synonymous mislabeling or "duplicate error," meaning that accessions had different names but shared the same microsatellite genotype. The third type of error refers to accessions that had a unique microsatellite genotype in this collection, but did not match with the established reference genotype (having the same name) in the original genebank or "nonmatching error."

For the identification of intraplant error, pairwise matching of multilocus microsatellite profiles were performed among the six individual trees representing each accession in the collection. If the microsatellite fingerprint profiles for all six trees of a given accession were identical (matching alleles at all loci), then there was no intraplant error in the accession and their profiles were condensed to one. If one or more of the six fingerprint profiles did not match, these were considered intraplant errors and were treated as separate accessions in the following analysis.

For our subsequent analyses purposes, accessions with different names that were fully matched at 15 microsatellite loci were declared synonymously mislabeled accessions or duplicates. Rigor was assessed for match declaration using the probability of identity (PID)—that the two individuals may share the same multilocus genotype by chance (Waits et al., 2001). Probability of identity was computed assuming all individual genotypes were siblings (PID_sib), which was defined as the probability that two sibling individuals drawn at random from a population had the same multilocus genotype (Evett and Weir, 1998; Waits et al., 2001). The overall PID_sib was the upper limit of the possible ranges of PID in a population, thus providing the most conservative number of loci required to resolve all individuals, including relatives (Waits et al., 2001).

Using accessions with an established reference genotype in the International Cacao Collections at the Cocoa Research Unit (CRU) in Trinidad and Tobago and the Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) in Turrialba, Costa Rica, microsatellite profiles generated from the accessions in the USDA-ARS Mayaguez cacao collection were compared with the microsatellite profiles from the reference genotypes, which were established after the process of identity verification, including duplicate identification, population assignment test, and pedigree reconstruction (Boccara and Zhang, 2006; Johnson et al., 2007; Motamayor et al., 2008; Zhang et al., 2008, 2009). If an accession completely matched at all loci with the reference genotypes in the international cacao collections, this accession was considered true to type; otherwise, they were considered as mislabeled.

Analysis of Genetic Diversity

After the exclusion of duplicates, summary descriptive statistics were computed for this collection. The descriptive statistics included the number of loci, allele frequencies, and observed and expected heterozygosity values. All statistics were calculated using POPGENE version 1.32 (Francis C. Yeh, University of Alberta, Edmonton, AB, Canada). Polymorphic information content (PIC) values were calculated using the following formula: $PIC = 1 - \Sigma p_i^2$, where p_i is the frequency of the allele. The genetic relationship among the cacao accessions was assessed using a cluster analysis. A similarity matrix, using the simple matching coefficient, was calculated between all possible pairs of accessions using the SimQual function in the program of NTSYS pc v 2.2e (Exeter Software, Setauket, NY). The corresponding matrices were used to build a dendrogram using the unweighted pair group method with the arithmetic mean (UPGMA) mathematical averaging function implemented in the same program.

The genetic structure of the USDA collection was examined using a Bayesian cluster analysis (Pritchard et al., 2000). The program Structure v2.1 (Pritchard et al., 2000) was used for computation. An admixture model with 200,000 iterations after a burn-in period of 100,000 was used. The number of clusters (K-value) was set from 2 to 10. Ten independent runs were assessed for each fixed number of clusters (K). The ΔK value was computed to detect the most probable number of clusters (Evanno et al., 2005). The run with the highest Ln Pr $(X|K)$ value of the 10 was chosen and presented as bar plots per genotype.

The level of genetic diversity in the USDA collection was also assessed by comparing the allele richness in this collection with that in the international cacao collection maintained in CATIE (Zhang et al., 2009). The data of the CATIE collection was based on the 548 unique accessions available at the Cabiria farm in 1999, which reflects part of the genetic diversity in the current collection (W. Phillips-Mora, personal communication, 2008). The two collections were compared for total number of alleles, number of major alleles (number of alleles that have frequency >5%), and total molecular variance. The allelic richness and molecular variance were computed for each collection independently using the frequency and analysis of molecular variance procedures in the program GenAlEx 6.0 (Peakall and Smouse, 2006).

After the exclusion of duplicates, the genetic redundancy caused by closely related accessions in the collection was assessed by simulating the relationship between genetic diversity and

different sizes of sampled individual accessions, following the sampling method of random sampling and maximization strategy (Schoen and Brown, 1993). The maximization procedure was originally designed for the development of germplasm core collections implemented in the MSTRAT computer program (Gouesnard et al., 2001). For each simulated sampling, Shannon's diversity index was used to represent the sampled diversity. For each sample size, an average value of Shannon's diversity index based on 10 replicated runs was presented.

RESULTS

Identification of Mislabeling and Duplicates

Fingerprint profiles for all 924 trees were generated with all 15 microsatellite loci. Reproducibility of the identical amplification profiles was evident when all six trees of a given accession were compared. Matching fingerprint profiles were condensed into one consensus profile, generating 174 unique fingerprint profiles (data not shown) that were used in further analyses. There were 19 cases (in one of the 19 cases there were three genotypes) of homonymous mislabeling (intraplant error) out of the 154 accessions (12.3%) (Table 1).

Pairwise comparisons among the 174 genotypes that passed the test of intraplant error led to the identification of 14 synonymous sets, involving 49 accessions (9.1%) (Table 2). The size of the synonymously mislabeled sets ranged from 2 to 19. From each synonymous set, only one individual accession from each duplicate group was selected for the subsequent diversity analysis and the rest were eliminated from the data set, which led to a total of 139 unique fingerprint profiles in this collection. A total of 64 accessions that have established reference genotypes in the two international cacao collections (CATIE and CRU) were used for pairwise comparisons (data not shown) and the results of the comparisons are presented in Table 1.

Descriptive Statistics and Genetic Diversity

After the elimination of duplicates, the 139 accessions with unique individual genotypes were included in the data set and used for diversity analysis. The results of descriptive statistics showed that the 15 loci had an average of 8.8 alleles per locus with mTcCIR1 having five alleles and both mTcCIR37 and mTcCIR60 having 12 alleles (Table 3) at their respective loci. The observed heterozygosity values ranging from 0.47 to 0.82, with a mean of 0.65 and expected heterozygosity (Levene, 1949) values ranged from 0.45 to 0.81. Polymorphic information content ranged from 0.45 to 0.99 with a mean of 0.78 (Table 3).

Cluster analysis showed that accessions generally grouped together according to their geographical origin and traditional genetic background (Fig. 1). At the similarity level of 0.81 to 0.82, the dendrogram split into three tightly grouped clusters (at the upper part of the dendrogram) and numerous small clusters (at the lower part of the dendrogram). The first cluster on the top consisted mostly of accessions from Mexico, Central America, and the Caribbean region, represented mainly by the Trinitario type varieties and breeding lines. The second cluster consisted mostly of accessions that originated from Brazil, including Amelonado, SIAL, and SIC accessions, and it was called "lower Amazon Forastero" for practical purposes. The third cluster included mostly the domesticated Ecuadorian varieties, including the EET and UF accessions from the coastal plains of Ecuador, which have various degrees of ancestry from the "Nacional" cacao. At the lower bottom of the dendrogram were mostly accessions from the upper Amazon, including APA SPEC and SPA accessions from Colombia and IMC from Peru, and breeding lines (e.g., APA and HY) also from the upper Amazon. They were generally referred as "upper Amazon Forastero". Two accessions with distinctive genotypes grouped as outliers and share some exclusive morphological features, including small, rounded leaves (personal observation).

The result of Bayesian clustering analysis largely agreed with the distance-based cluster analysis. Based on the value of ΔK (Evanno et al., 2005), the 139 accessions could be grouped into four most probable clusters representing the four main clusters mentioned above, Trinitario (51 accessions), "Upper Amazon" (44 accessions), "Lower Amazon and Parinari" (17 accessions), and "Nacional hybrids" (27 accessions) (Fig. 2). The four clusters, on average, had a coefficient of membership (Q value) of 0.874. A Q value of 0 corresponds to an individual of purely exogenous origin, whereas a value of 1 is a purely native individual. Accessions with a Q value <0.75 were considered a "failed match" to their home cluster membership (based on their recorded passport information) thus were categorized as putative mislabeled (Table 1; Fig. 2).

The amount of genetic diversity as measured by the number of alleles in the USDA-ARS collection was proportional to its size when compared to the CATIE collection (Fig. 3). A total of 132 alleles from 139 accessions were found in the USDA-ARS collection. In contrast, data collected from the cacao collection at CATIE in 1999 showed the collection having 231 alleles in 548 unique accessions (Zhang et al., 2009). The difference was negligible when comparing the number of major alleles (allele frequency >5%) between the two collections (Fig. 3). However, approximately 43% of the alleles at CATIE are not represented in the USDA-ARS collection, demonstrating that there are still various diversity gaps that remains to be filled (Fig. 3).

The simulation between sample size and diversity representation showed that 90% of the genetic diversity, as measured by Shannon's index, can be captured at a sample size of 37 if a random sampling approach is taken (Fig. 4). The curvilinear relationship between sample size and genetic diversity (Fig. 4) suggests that the accessions in

Table 1. Name, source and results of identification verification (RIV) of cacao accessions maintained at the USDA-ARS Tropical Agriculture Research Station (TARS) in Mayaguez, PR.

	Clone name†	Source‡	RIV		Clone name	Source	RIV
1	AC T 1/1 [TTO]	Trinidad		38	EET 397 [ECU]	Ecuador	
2	AC T 2/8 [TTO]	Trinidad		39	EET 400 [ECU]	Ecuador	
3	AC T 2/11 [TTO]	Trinidad		40	EET 401 [ECU]	Ecuador	
4	AC T 2/18 [TTO]	Trinidad		41	EET 407 [ECU]	Ecuador	
5	AMELONADO	Ghana		42	GA 57	Haiti	
6	APA 4	Costa Rica		43	GC 7 [SUR]	Costa Rica	
7	APA 5	Colombia		44	GS 7	Grenada	
8	BE 10	Brazil	1§,2¶	45	GS 29	Grenada	
9	C 87/56	Trinidad		46	GS 46	Grenada	
10	CAS 1	Costa Rica		47	HY 27 1418	Puerto Rico	
11	CC 10 A	Costa Rica	1	48	HY 27 1419	Puerto Rico	
11	CC 10 B	Costa Rica	1	49	HY 27 1420	Puerto Rico	
12	CC 11	Costa Rica		50	ICS 1	Trinidad	
13	CC 34	Costa Rica		51	ICS 6	Trinidad	
14	CC 37 A	Costa Rica	1	52	ICS 16	Trinidad	
14	CC 37 B	Costa Rica	1	53	ICS 22	Trinidad	
15	CC 38 A	Costa Rica	1	54	ICS 29	Trinidad	
15	CC 38 B	Costa Rica		55	ICS 39	Trinidad	
16	CC 39	Costa Rica	1	56	ICS 40	Guatemala	
17	CC 40	Costa Rica		57	ICS 41	Trinidad	
18	CC 41	Costa Rica		58	ICS 45	Trinidad	
19	CC 49	Costa Rica	1	59	ICS 48	Trinidad	1,2
20	CC 54	Costa Rica	1	60	ICS 55	Trinidad	
21	CC 57	Costa Rica		61	ICS 60	Trinidad	
22	CC 60	Costa Rica		62	ICS 61	Trinidad	
23	CC 71	Costa Rica		63	ICS 88	Trinidad	
24	CC 80	Costa Rica		64	ICS 95	Trinidad	
25	EET 40 [ECU] A	Ecuador		65	ICS 129	Trinidad	
25	EET 40 [ECU] B	Ecuador		66	IMC 20	–#	
26	EET 54 [ECU]	Ecuador		67	IMC 47	Trinidad	
27	EET 64 [ECU]	Ecuador		68	IMC 67 A	Guatemala	
28	EET 67 [ECU]	Ecuador		68	IMC 67 B	Guatemala	1
29	EET 75 [ECU]	Ecuador		69	LAFI 7	Guatemala	
30	EET 94 [ECU]	Ecuador		70	MO 20	Trinidad	
31	EET 103 [ECU]	Ecuador		71	MOCORONGO	Brazil	
32	EET 164 [ECU]	Ecuador		72	MX 75/3 A	–	
33	EET 236 [ECU]	Ecuador		72	MX 75/3 B	–	
34	EET 283 [ECU]	Ecuador		73	P 10 [MEX] A	Mexico	
35	EET 353 [ECU] A	Ecuador		73	P 10 [MEX] B	Mexico	2
35	EET 353 [ECU] B	Ecuador		74	P 22 [MEX]	Mexico	
36	EET 381 [ECU]	Ecuador		75	P 43 [MEX]	Costa Rica	
37	EET 390 [ECU]	Ecuador		76	PA 4 [PER]	Trinidad	
77	PA 13 [PER]	Haiti	1,2	116	SIC 1	Brazil	
78	PA 16 [PER] A	England		117	SIC 2	Costa Rica	
78	PA 16 [PER] B	England	1,2	118	SIC 5	Brazil	
79	PA 39 [PER]	Trinidad		119	SIC 7	Brazil	
80	PA 44 [PER]	Peru		120	SIC 72 A	Brazil	1,2
81	PA 51 [PER]	Trinidad	1	120	SIC 72 B	Brazil	
82	PA 121 [PER]	Puerto Rico		121	SNK 12 A	Cameroon	
83	PA 185 [PER] A	Trinidad	1,2	121	SNK 12 B	Cameroon	
83	PA 185 [PER] B	Trinidad	1,2	122	SPA 4	Colombia	
84	PA 303 [PER]	Ghana		123	SPA 7	Colombia	
85	POUND 7 [POU] A	Haiti	2	124	SPA 9	Colombia	
85	POUND 7 [POU] B	Haiti	2	125	SPA 10	Colombia	

(cont'd)

Table 1. Continued.

	Clone name†	Source‡	RIV		Clone name	Source	RIV
86	POUND 16 [POU]	Trinidad		126	SPEC 194/16	Trinidad	
87	POUND 19 [POU]	Costa Rica		127	STAHEL	Surinam	
88	POUND 25 [POU]	USPIS		128	TARS #1	Puerto Rico	
89	POUND 25/A [POU] A	–	2	129	TARS #9	Puerto Rico	
89	POUND 25/A [POU] B	–		130	TARS #14	Puerto Rico	
90	POUND 32 [POU]	Trinidad		131	TARS #15 A	Puerto Rico	
91	RIM 2 [MEX]	–		131	TARS #15 B	Puerto Rico	
92	RIM 6 [MEX]	Guatemala		131	TARS #15 C	Puerto Rico	
93	RIM 10 [MEX]	Guatemala		132	TARS #23	Puerto Rico	
94	RIM 13 [MEX] A	Guatemala		133	TARS #27	Puerto Rico	
94	RIM 13 [MEX] B	Guatemala	2	134	TARS #30	Puerto Rico	
95	RIM 15 [MEX]	Guatemala		135	TARS #31	Puerto Rico	
96	RIM 30 [MEX]	Mexico		136	TARS #34	Puerto Rico	
97	RIM 34 [MEX]	Mexico		137	TSAN 812	Trinidad	
98	RIM 41 [MEX]	Mexico		138	TSH 1112	Trinidad	
99	RIM 48 [MEX]	Mexico		139	UF 10	Costa Rica	
100	RIM 52 [MEX]	Mexico		140	UF 29	Costa Rica	
101	RIM 75 [MEX]	Guatemala		141	UF 36	Costa Rica	
102	RIM 78 [MEX]	Mexico		142	UF 122	Costa Rica	1
103	RIM 105 [MEX]	Guatemala		143	UF 221	Guatemala	
104	SC 49	Colombia		144	UF 601	Costa Rica	
105	SCA 6	Ecuador	1	145	UF 613	Costa Rica	1
106	SCA 9 A	England	1,2	146	UF 652 A	Costa Rica	
106	SCA 9 B	England	1,2	146	UF 652 B	Costa Rica	
107	SCA 12	Ecuador	1	147	UF 666	Costa Rica	
108	SCR 2	Costa Rica		148	UF 667	Costa Rica	
109	SCR 4	Costa Rica		149	UF 668	Costa Rica	
110	SGU 3	Guatemala		150	UF 703 A	Costa Rica	
111	SGU 69	Guatemala		150	UF 703 B	Costa Rica	1
112	SIAL 42	Brazil		151	UF 705	Costa Rica	
113	SIAL 44	Brazil		152	UF 710	Costa Rica	
114	SIAL 56	Brazil		153	UF 715	Costa Rica	
115	SIAL 98	Brazil		154	UF 717	Costa Rica	1

†The International Cocoa Germplasm Database preferred name for each clone is used.

‡Based on passport data maintained at USDA-ARS TARS. Source in some cases is synonymous with the origin of an accession. USPIS, U.S. Plant Introduction Station.

§Mislabeling determined by comparing fingerprint profiles generated in this study to those generated for matching clones at Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) and International Cacao Collections at the Cocoa Research Unit (CRU) in Trinidad and Tobago.

¶Mislabeling determined using the assignment test, which determined the population of origin of a given single individual using the Bayesian clustering method (Pritchard et al., 2000).

#Passport for information on source missing.

this collection overlapped their contribution to the overall genetic diversity. Redundancy was caused by closely related breeding lines of the various Trinitario hybrids as revealed in the UPGMA tree (Fig. 1). These redundant Trinitario hybrids could be replaced by accessions that bring complementary allelic contribution to this collection.

DISCUSSION

Molecular markers have been widely used to assess duplicates and mislabeling in the national and international cacao gene banks. In contrast to identification methods that use dominant markers, identification methods using multilocus microsatellite profiles are significantly more accurate because identical genotypes can have a full match in the multilocus microsatellite profiles. The present study obtained reliable identification of genotypes using this method. Microsatellite fingerprinting is both a practical and cost-effective method for assessing the genetic identity of a large number of cacao germplasm accessions. However, there are exceptional cases in which closely related clones are indistinguishable based on 15 loci, such as point mutations that may cause phenotypic change (e.g., the change of pod or seed color is often associated with few mutations). Other cases include genetic groups with low genetic diversity such as Criollo, Amelonado, Trinitario, Nacional, and Nanay (Lercetau et al., 1997; Motamayor et al., 2003, 2008) in cacao. Low genetic diversity may have been the reason why the use of 15 markers showed no differences among some of the accessions (Table 2). Therefore, phenotypic examination, which is currently being conducted on the

collection, remains an important tool that can play a complementary role in the identification of duplicates in cacao germplasm. Another approach would be to use additional markers, known to be polymorphic in those low genetic diversity groups. Screening of polymorphic markers for specific groups and their utilization could be cost effective.

All cacao accessions in the USDA-ARS Mayaguez repository were introduced from various collections in Central and South America. As with most other cacao germplasm collections, passport records documenting introductions of some genotypes into the collection are incomplete. It is noteworthy that several of the primary and secondary contributors of germplasm were unable to guarantee the authenticity of the material supplied. This is considered a common cause of the introduction of mislabeled accessions into cacao collections (Turnbull et al., 2004). Recent studies on the genetic identity of cacao germplasm in the international collections held in Costa Rica and Trinidad showed that in many instances, mislabeling occurred before the materials were introduced into ex situ collections. Therefore, verification and correct mislabeling in the USDA-ARS collection using "reference profiles" of the original trees in the source collections must be conducted. In the present study, 64 reference genotypes from the two international cacao collections (Costa Rica and Trinidad) were used to verify the genetic identity of the corresponding accessions held in the USDA-ARS Mayaguez collection. However, reference genotypes originating from other countries, such as Ecuador and Colombia, are still in development as the source trees in the original collections are in the process of being genotyped. Moreover, some genotypes, such as the breeding lines of Trinitario hybrids, do not have original references for comparison. For this reason, only a fraction of the mislabeled accessions in the USDA-ARS collection can be confirmed in this study. In Motamayor et al. (2008) an exhaustive list of genotypes from reference clones (from the most important germplasm collections) is provided (indicating which genotypes are correctly labeled and which not). In the future such a list, with the corresponding publicly available microsatellite genotypes, should be increased with additional accessions to be used as the database source of reference genotypes.

In addition to the use of multilocus matching, a model-based assignment test was also employed, which determined the population of origin of any given single individual using the Bayesian clustering method (Pritchard et al., 2000). This method needs a relatively small number of loci to identify population structure and assign individuals appropriately (Pritchard et al., 2000). It is thus highly suitable for resolving mislabeling problems in this cacao germplasm collection by identifying if a given cacao genotype belongs to a specific "home population." This method allowed us to detect mislabeling based on their posterior assignment probability (Fig. 2), because many accessions in the international cacao germplasm collections have a clear population identity label. The combination of assignment test with multilocus matching offered a powerful tool to detect mislabeling in the cacao germplasm collection. However, it is noteworthy to point out that the resolution of the assignment test may be improved with the addition of more maker loci. With 15 loci, the present study grouped the 139 distinctive accessions into four main clusters. Some clusters (e.g., the Upper Amazon cluster) may actually include more than one population corresponding to the 10 populations defined by Motamayor et al. (2008). The amount of genetic diversity in the USDA-ARS cacao collection at Mayaguez, PR (as measured by allele richness and gene diversity) is approximately proportional to its size in comparison to the international cacao germplasm collection maintained in CATIE. The UPGMA dendrogram and the Bayesian cluster analysis both show that the accessions can be primarily grouped into four clusters that correspond to the traditional cacao germplasm groups. The geographical origin of accessions in the Mayaguez collection covers the majority of the major cacao producing countries in the Americas. However, several known genetic groups are absent in this collection. Motamayor et al. (2008) suggested that the structure of the cacao germplasm diversity goes beyond the traditional classification of Criollo and lower and upper Amazon "Forasteros" and a new classification

Table 2. Fourteen synonymous groups (including 49 accessions) within the USDA-ARS Mayaguez cacao collection identified by microsatellite DNA analysis. Accessions in the same synonymous set shared identical multilocus microsatellite profiles.

Set	Accessions	Set	Accessions	Set	Accessions
1	CC 10 A	3	GS 46	10	CC 38 A
1	EET 353 [ECU] B	3	UF 668	10	RIM 13[MEX] B†
1	EET 381 [ECU]				
1	P 10 [MEX] A	4	GS 7	11	CC 39
1	P 22 [MEX]	4	ICS 29	11	CC 49
1	P 43 [MEX]			11	EET 40 [ECU] A
1	RIM 10 [MEX]	5	EET 236 [ECU]		
1	RIM 13 [MEX] A	5	TSAN 812	12	CC 10 B
1	RIM 15 [MEX]			12	CC 11
1	RIM 105 [MEX]	6	ICS 60		
1	RIM 2 [MEX]	6	ICS 61	13	UF 666
1	RIM 34 [MEX]			13	UF 705
1	RIM 41 [MEX]	7	CC 57		
1	RIM 48 [MEX]	7	GA 57 [MAY]	14	EET 397 [ECU]
1	RIM 52 [MEX]			14	UF 717
1	RIM 6 [MEX]	8	SIAL 98		
1	RIM 75 [MEX]	8	SIC 1		
1	RIM 78 [MEX]	8	SIC 2		
1	SGU 69 [MEX]	8	SIC 72 B		
2	ICS 39	9	POUND 7 [POU] B†		
2	POUND 7 [POU] A†	9	UF652 A		
2	SIC 72 A†				

†Means accession did not match population of origin using the model-based assignment test (Fig. 2).

Table 3. Characteristics and summary statistics for the 15 international set of microsatellite primers utilized for fingerprinting the USDA-ARS Tropical Agriculture Research Station cacao (*Theobroma cacao*) collection.

Primer name	Forward and reverse sequences (5′–3′)	Chromosome	Tm	Repeat motif	Allele range	Alleles/ locus†	H_{Obs}‡	H_{Exp}‡	PIC
mTcCIR1§	F: GCAGGGCAGGTCCAGTGAAGCA R: TGGGCAACCAGAAAACGAT	8	51	$(CT)_{14}$	127–144	5	0.47	0.45	0.44
mTcCIR6	F: TTCCCTCTAAACTACCCTAAAT R: TAAAGCAAAGCAATCTAACATA	6	46	$(TG)_7(GA)_{13}$	222–247	9	0.64	0.64	0.96
mTcCIR7	F: ATGCGAATGACAACTGGT R: GCTTTCAGTCCTTTGCTT	7	51	$(GA)_{11}$	148–163	6	0.61	0.65	0.65
mTcCIR8	F: CTACTTTCCCATTTACCA R: TCCTCAGCATTTTCTTTC	9	46	$(TC)_5$ TT$(TC)_{17}$ TTT$(CT)_4$	288–304	6	0.56	0.62	0.92
mTcCIR11	F: TTTCCTCATTATTAGCAG R: GATTCGATTTGATGTGAG	2	46	$(TC)_{13}$	288–317	11	0.61	0.66	0.74
mTcCIR12	F: TCTGACCCCAAACCTGTA R: ATTCCAGTTAAAGCACAT	4	46	$(CATA)_4$ N_{18} $(TG)_6$	188–251	10	0.73	0.74	0.80
mTcCIR15	F: CAGCCGCCTCTTGTTAG R: TATTTGGGATTCTTGATG	1	46	$(TC)_{19}$	232–256	11	0.82	0.81	0.87
mTcCIR18	F: GATAGCTAAGGGGATTGAGGA R: GGTAATTCAATCATTTGAGGATA	4	51	$(GA)_{12}$	331–355	9	0.66	0.67	0.72
mTcCIR22	F: ATTCTCGCAAAAACTTAG R: CATCCAAGGAGTGTAAATAG	1	46	$(TC)_{12}$ N_{146} $(CT)_{10}$	279–290	6	0.60	0.58	0.59
mTcCIR24	F: TTTGGGGTGATTTCTTCTGA R: TCTGTCTCGTCTTTTGGTGA	9	46	$(AG)_{13}$	185–203	7	0.57	0.50	0.95
mTcCIR26	F: GCATTCATCAATACATTC R: GCACTCAAAGTTCATACTAC	8	46	$(TC)_9C(CT)_4TT(CT)_{11}$	282–307	9	0.71	0.67	0.69
mTcCIR33	F: TGGGTTGAAGATTTGGT R: CAACAATGAAAATAGGCA	4	51	$(TG)_{11}$	264–346	10	0.71	0.72	0.73
mTcCIR37	F: CTGGGTGCTGATAGATAA R: AATACCCTCCACACAAAT	10	46	$(GT)_{15}$	133–185	12	0.67	0.70	0.72
mTcCIR40	F: AATCCGACAGTCTTTAATC R: CCTAGGCCAGAGAATTGA	3	51	$(AC)_{15}$	259–284	9	0.70	0.79	0.84
mTcCIR60	F: CGCTACTAACAAACATCAAA R: AGAGCAACCATCACTAATCA	2	51	$(CT)_7(CA)_{20}$	187–223	12	0.64	0.73	0.86
Mean						8.8	0.65	0.66	0.78

†Summary statistics for alleles/locus and observed and expected heterozygosity generated with POPGENE 1.32.

‡Observed (H_{Obs}) and expected (H_{Exp}) heterozygosity computed using Levene (1949) algorithm and polymorphic information content (PIC) calculated by PIC = $1 - \Sigma p_i^2$ where p_i is the frequency of the allele.

§mTcCir, microsatellite *Theobroma cacao* CIRAD (Centre de Coopération Internationale en Recherche Agronomique pour le Développement).

into 10 different populations or genetic groups, which reflects more accurately the large genetic diversity of the species, should be implemented. Using these 10 populations as a point of reference, then the USDA-ARS collection still has several diversity gaps that need to be filled. For example, the "Criollo" group from Mexico and Central America, the "Guiana" group from Guiana, and the "Nanay" population from Peru, among others, were absent. The difference in the total number of alleles found between the USDA-ARS and the CATIE collections also indicated that the genetic diversity of cacao in this international collection is not fully represented, although all of the common alleles have been well sampled (Fig. 3). Moreover, simulation of the relationship between sample size and Shannon's diversity index also suggests that the amount of allelic diversity in the USDA-ARS repository can be captured with a much smaller sample size if the maximization strategy (Schoen and Brown, 1993; Gouesnard et al., 2001) is used to sample the subset. The present result thus suggests the potential to rationalize this collection by replacing the redundant accessions with those that can make a complementary contribution to genetic diversity. However, it needs to be pointed out that the estimation of genetic diversity and simulation of genetic redundancy were based on microsatellite marker-defined diversity parameters and index alone, without taking into consideration economic and agronomic traits. These estimations should be considered as indicators for cacao genebank management. There are many accessions that may not have an outstanding contribution in terms of the microsatellite allele richness, but they may possess variation in valuable agronomic and economic traits (e.g., fine flavors, as shown in the landraces from Mesoamerica). It is well known that diversity quantified by morphological and agronomic traits do not necessarily correspond to marker-defined genetic diversity. For this reason, a further exercise of diversity estimation would be to include major agronomic traits (presently being conducted on the germplasm collection), together with the neutral microsatellite

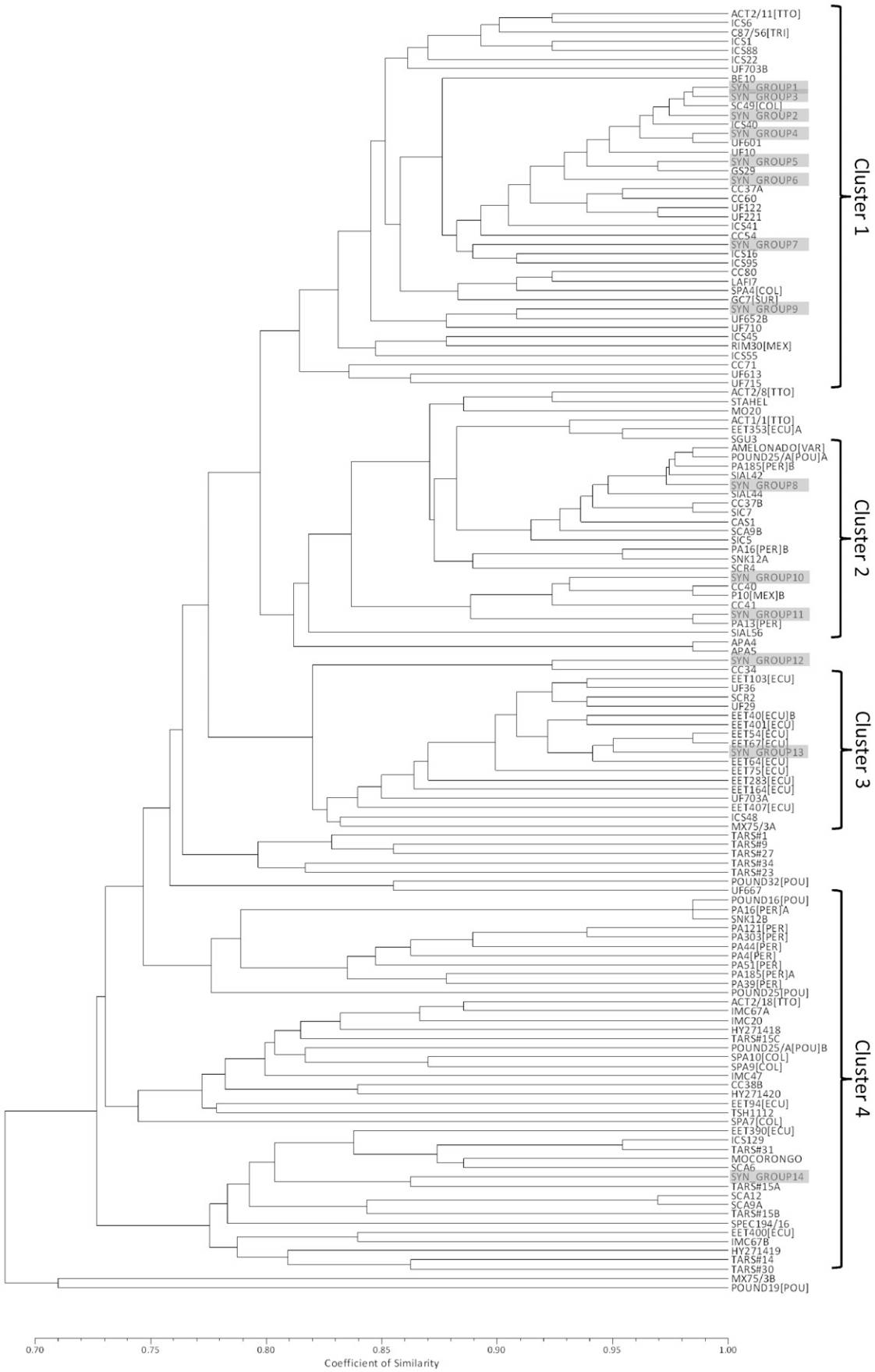

Figure 1. Dendrogram of 139 cacao genotypes maintained in the USDA-ARS Mayaguez collection. Dendrogram includes 19 cases in which at least more than one fingerprint profile was identified for a given accession (identified with the letter A and B). The cluster analysis was based on simple matching coefficient with the unweighted pair group method with the arithmetic mean clustering method. Synonymous groups are numbered and shaded and correspond to those listed in Table 2.

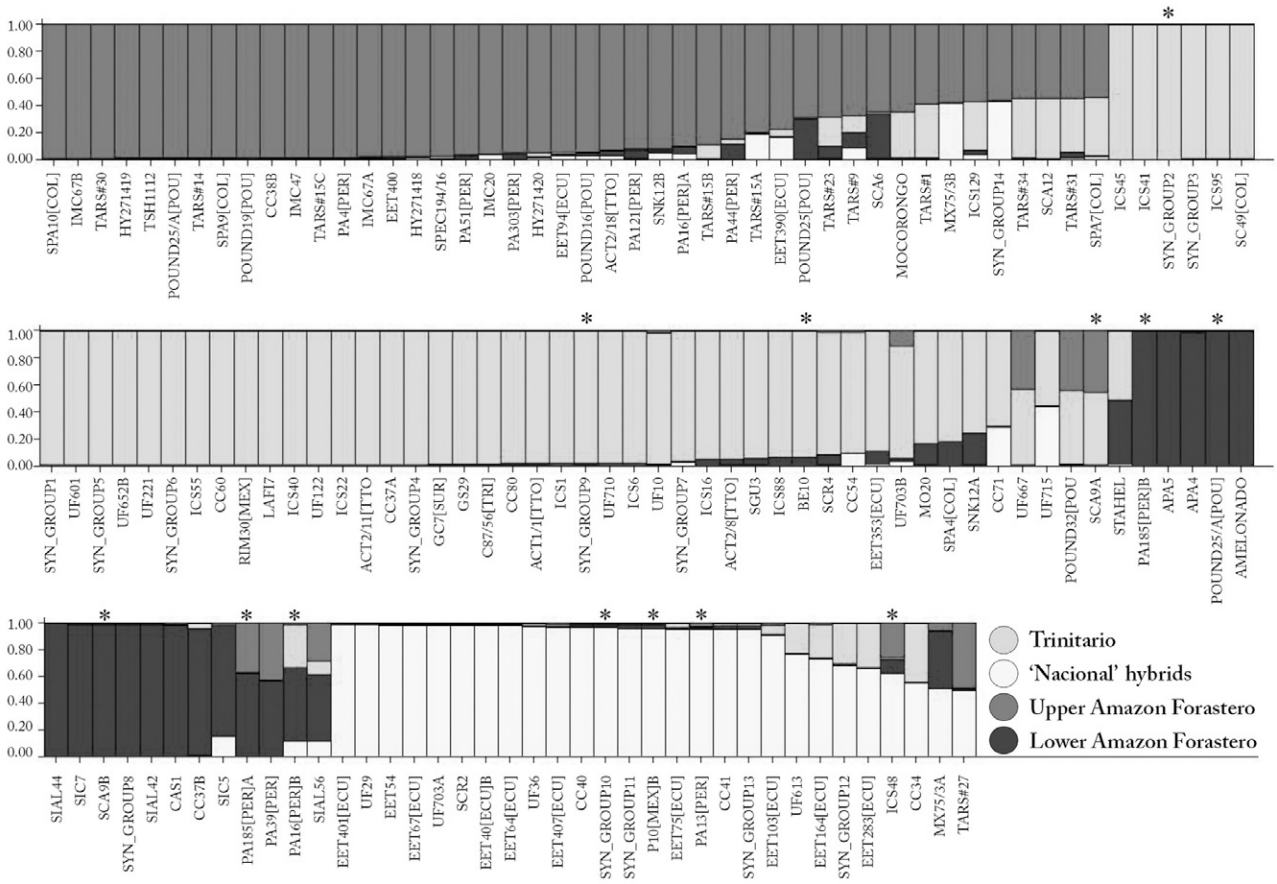

Figure 2. Inferred clusters in the USDA-ARS Mayaguez collection using STRUCTURE. The most probable clusters were obtained at K = 4. Each vertical line represents one individual genotype. Individuals with multiple colors have admixed genotypes from multiple clusters. Mislabeled clones (with an assignment probability <0.75) are marked with an asterisk.

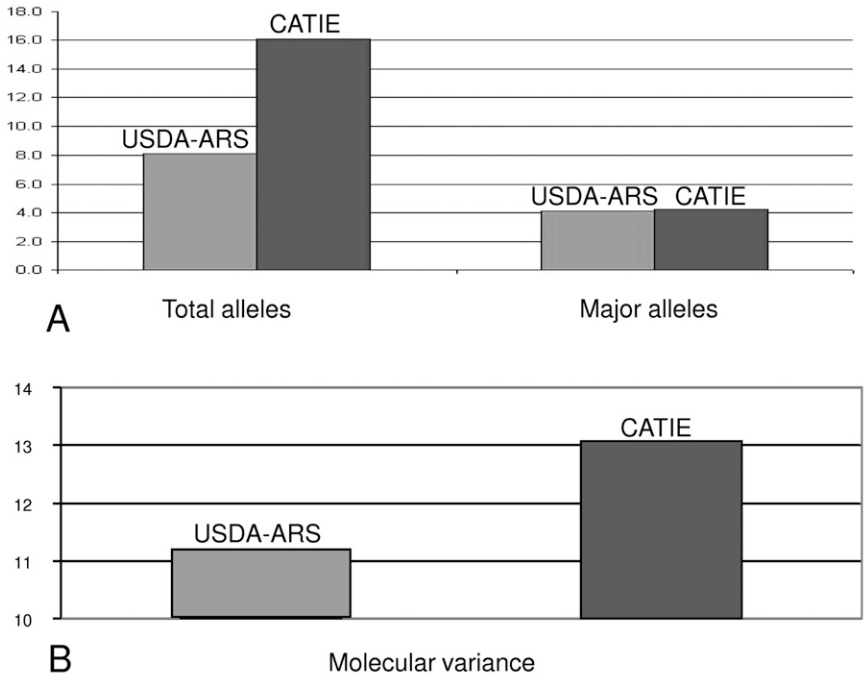

Figure 3. Comparison of genetic diversity between the USDA-ARS Mayaguez collection and the Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) international cacao collection. (A) Total number of alleles and major alleles (frequency >5%). (B) Mean square of molecular variance calculated using the program of GenAlex 6.0 (Peakall and Smouse, 2006)

markers. Currently, a core collection of cacao germplasm representing the genetic diversity in the international cacao collections in Trinidad and Costa Rica is being developed (D. Zhang, personal communication, 2008). The development of this core set is based on the diversity defined by molecular markers, agronomic traits, and geographical representation. This core set will serve as the base for introducing new germplasm into the USDA-ARS collection in the next few years.

In conclusion, the availability of multilocus microsatellite profiles for every tree allowed the unambiguous identification of intraplant errors as well as putative duplicates in the 924 cacao trees in the USDA-ARS collection. Comparisons with reference genotypes and assignment tests also allowed the detection of mislabeling in this collection. In addition, the assessment of the representative diversity in the USDA-ARS collection was conducted through the comparison of genetic diversity between the local collection and an international collection and through comparisons with other diversity studies. This study also identified several diversity gaps and proposed a potential approach, through appropriate quarantines, to fill these gaps. To our knowledge, this study is the first to genotype and analyze the DNA fingerprints of every tree in a cacao collection. The results of this study will be very useful in improving the genetic accuracy and efficiency in cacao germplasm conservation at the USDA-ARS Mayaguez repository. Fingerprint profiles for cacao accessions will be made available through the USDA National Plant Germplasm System Germplasm Resource Information Network database (http://www.ars-grin.gov/).

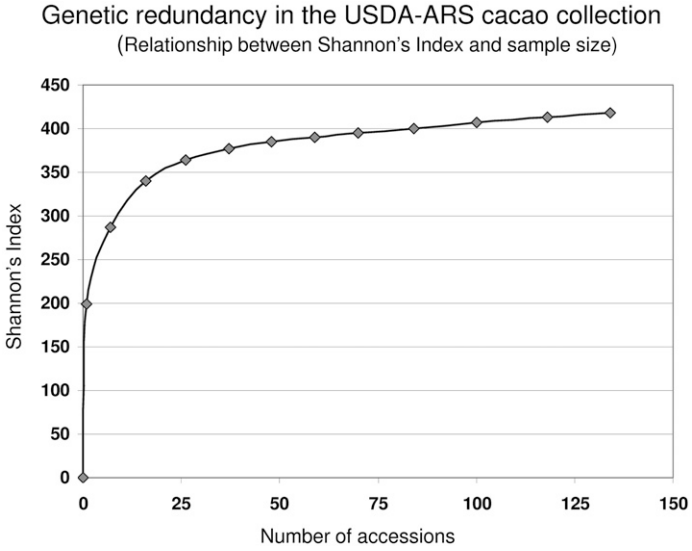

Figure 4. Simulated relationship between sample size and genetic diversity (measured by Shannon's index) in the USDA-ARS Mayaguez collection.

Acknowledgments

The authors would like to thank Mr. Wilber Quintanilla for his help generating data, Mr. Carlos Rios for his technical help with the figures, and Drs. Dimuth Siritunga, Wilberth Phillips-Mora, Timothy Porch, and Mark Guiltinan for their critical review of the manuscript.

REFERENCES

Bartley, B.D.G. 2005. The genetic diversity of cacao and its utilization. CABI Publishing, CAB International, Wallingford, Oxfordshire, UK.

Bekele, F., I. Bekele, D.R. Butler, and G.G. Bidaisee. 2006. Patterns of morphological variation in a sample of cacao (*Theobroma cacao* L.) germplasm form the International Cocoa genebank, Trinidad. Genet. Resour. Crop Evol. 53:933–948.

Bekele, F., and D.R. Butler. 2000. Proposed short list of descriptors for characterization. *In* A.B. Eskes et al. (ed.) Working Procedures for Cocoa Germplasm Evaluation and Selection: Proceedings of the CFC/ICCO/IPGRI Project Workshop, Montpellier, France. 1–6 Feb. 1998. IPGRI, Rome.

Bennett, A.B. 2003. Out of the Amazon: *Theobroma cacao* enters the genomic era. Trends Plant Sci. 8:561–563.

Boccara, M., and D. Zhang. 2006. Progress in resolving identity issues among the Parinari accessions held in Trinidad: The contribution of the collaborative USDA/CRU project. *In* Annual report for 2005. Univ. of the West Indies, St. Augustine, Trinidad and Tobago.

Bowers, J.H., B.A. Bailey, P.K. Hebbar, S. Sanogo, and R.D. Lumsden. 2001. The impact of plant diseases on world chocolate production. Available at www.plantmanagementnetwork.org/php/. Plant Health Prog. DOI: 10.1094/PHP-2001-0709-01-RV.

Cervantes-Martinez, C., J.S. Brown, R. Schnell, J.C. Motamayor, A.W. Meerow, and D. Zhang. 2006. A computer simulation study on the number of loci and trees required to estimate genetic variability in cacao (*Theobroma cacao* L.). Tree Genet. Genomes 2:152–164.

Cryer, N.C., M.G.E. Fenn, C.J. Turnbull, and M.J. Wilkinson. 2006. Allelic size standards and reference genotypes to unify internal cocoa (*Theobroma cacao* L.) microsatellite data. Genet. Resour. Crop Evol. 53:1643–1652.

Ducamp, M., S. Nyassee, L. Grivet, J.M. Thevenin, G. Blaha, D. Despreaux, and C. Cilas. 2004. Genetic diversity of cocoa tree *Phytophthora* pathogens. p. 45–75. *In* C. Cilas and D. Despreauz (ed.) Improvement of cocoa tree resistance to *Phytophthora* diseases. CIRAD, Montpellier, France.

Engels, J.M.M., B.G.D. Bartley, and G.A. Enriquez. 1980. Cacao descriptors, their states and modus operandi. Turrialba (Costa Rica) 30:209–218.

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

Evett, I.W., and B.S. Weir. 1998. Interpreting DNA evidence: Statistical genetics for forensic scientists. Sinauer, Sunderland, MA.

FAOSTAT. 2007. http://faostat.fao.org/site/291/default.aspx (verified 11 Dec. 2009). FAO, Rome.

Fregene, M., M. Suárez, J. Mkumbira, H. Kulembeka, E. Ndedya, A. Kulaya, S. Mitchel, U. Gullberg, H. Rosling, A. Dixon, R. Dean, and S. Kresovich. 2003. Simple sequence repeat marker diversity in cassava landraces: Genetic diversity and differentiation in an asexually propagated crop. Theor. Appl. Genet. 107:1083–1093.

Gepts, P. 2006. Plant genetic resources conservation and utilization.

Crop Sci. 46:2278–2292.
Giron, C., J. Toledo, and P. Sanchez. 2004. *Theobroma cacao* L. (Sterculiaceae) en la región nororiental de la Isla de Margarita, Estado Nueva Esparta, Venezuela. Plant Genet. Resour. Newsl. 2004:1–4.
Gouesnard, B., T.M. Bataillon, G. Decoux, C. Rozale, D.J. Schoen, and J.L. David. 2001. MSTRAT: An algorithm for building germplasm core collections by maximizing allelic or phenotypic richness. J. Hered. 92:93–94.
Johnson, S.E., A. Mora, and R.J. Schnell. 2007. Field guide efficacy in the identification of reallocated clonally propagated accessions of cacao. Genet. Resour. Crop Evol. 54:1301–1313.
Kameswara, R.K., M.D. Burow, G. Burow, J. Burke, and P. Naveen. 2007. Molecular characterization of the U.S. peanut mini core collection using microsatellite markers. Crop Sci. 47:1718–1727.
Leal, J.B., L.M. Santos, C.A.P. dos Santos, J.L. Pires, D. Ahnert, and R.X. Correa. 2008. Genetic diversity among farm and germplasm bank accessions of cacao in Bahia, Brazil. Pesqui. Agropecu. Brasil. 43:851–858.
Lercetau, E., T. Robert, V. Petiard, and D. Crouzillat. 1997. Evaluation of the extent of genetic variability among *Theobroma cacao* accessions using RAPD and RFLP makers. Theor. Appl. Genet. 95:10–19.
Levene, H. 1949. On a matching problem in genetics. Ann. Math. Stat. 20:91–94.
Motamayor, J.C., P. Lachenaud, J.W. da Silva e Mota, R. Loor, D.N. Khun, S. Brown, and R.J. Schnell. 2008. Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L.). PLoS ONE 3(10):E3311, DOI: 10.1371/journal.pone.0003311.
Motamayor, J.C., A.M. Risterucci, M. Heath, and C. Lanaud. 2003. Cacao domestication II: Progenitor germplasm of the Trinitario cacao cultivar. Heredity 91:322–330.
Motamayor, J.C., A.M. Risterucci, P.A. Lopez, C.F. Ortiz, and A. Moreno. 2002. Cacao domestication I: The origin of the cacao cultivated by the Mayas. Heredity 89:380–386.
Motilal, L., and D. Butler. 2003. Verification of identities in global cacao germplasm collections. Genet. Resour. Crop Evol. 50:799–807.
Peakall, R., and P.E. Smouse. 2006. GenAlEx 6: Genetic analysis in Excel: Population genetic software for teaching and research. Mol. Ecol. Notes 6:288–295.
Perry, M.D., M.R. Davey, J.B. Power, K.C. Lowe, H.F.J. Blight, P.S. Roach, and C. Jones. 1998. DNA isolation and AFLP™ genetic fingerprinting of *Theobroma cacao*. Plant Mol. Biol. Rep. 16:49–59.
Ploetz, R.C. 2007. Cacao diseases: Important threats to chocolate production worldwide. Phytopathology 97:1634–1639.
Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.
Saunders, J.A., A.A. Hemeida, and S. Mischke. 2001. USDA DNA fingerprinting programme for identification of *Theobroma cacao* accessions. p. 108–114. *In* Proc. of the Int. Workshop on New Technologies and Cocoa Breeding, Kota Kinabalu, Sabah, Malaysia. 16–17 Oct. 2000. Int. Network for the Improvement of Cacao.
Saunders, J.A., S. Mischke, E.A. Leamy, and A.A. Hemeida. 2004. Selection of international molecular standards for DNA fingerprinting of *Theobroma cacao*. Theor. Appl. Genet. 110:41–47.
Schnell, R.J., J.S. Brown, D.N. Khun, C. Cervantes-Martinez, J.W. Borrone, C.T. Olano, W. Phillips, E. Johnson, E.J. Monteverde-Penso, J.C. Motamayor, F. Amores, and U. Lopes. 2007. Current challenges of tropical tree crop improvement: Integrating genomics into an applied cacao breeding program. Acta Hortic. 738:129–144.
Schnell, R.J., C.T. Olano, J.S. Brown, A.W. Meerow, C. Cervantes-Martinez, C. Nagani, and J.C. Motamayor. 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.
Schoen, D.J., and A.H.D. Brown. 1993. Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. Proc. Natl. Acad. Sci. USA 22:10623–10627.
Sounigo, O., Y. Christopher, F. Bekele, M. Vishnarayan, and F. Hosein. 2001. The detection of mislabeled trees in the International Cocoa Genebank, Trinidad (ICG,T) and options for a global strategy for identification of accessions. p. 34–39. *In* Proc. of the Int. Workshop on New Technologies and Cocoa Breeding, Kota Kinabalu, Sabah, Malaysia. 16–17 Oct. 2000. Int. Network for the Improvement of Cacao.
Swanson, J.D., A.C. Lee, and M.J. Guiltinan. 2003. USDA cacao DNA fingerprinting ring test: Results from Penn State University. Ingenic Newsl. 8:22–24.
Turnbull, C.J., D.R. Butler, N.C. Cryer, D. Zhang, C. Lanaud, A.J. Daymond, C.S. Ford, M.J. Wilkinson, and P. Hadley. 2004. Tackling mislabeling in cocoa germplasm collections. Ingenic Newsl. 9:8–11.
Vanitha, C., K. Ramamoorthy, A. Vijayakumar, and K. Sivasubramaniam. 2005. Moist and conditioning to minimize loss of viability in cocoa (*Theobroma cacao* Linn.) seed. Natural Product Radiance 4:487–491.
Volk, G.M., C.M. Richards, A.D. Henk, A.A. Reilley, N.V. Bassil, and J.D. Postman. 2006. Diversity of wild *Pyrus communis* based on microsatellite analyses. J. Am. Soc. Hortic. Sci. 131:408–417.
Waits, L.P., G. Luikart, and P. Taberlet. 2001. Estimating the probability of identity among genotypes in natural populations: Cautions and guidelines. Mol. Ecol. 10:249–256.
Willson, K. (ed.). 1999. Coffee, cocoa, and tea. CABI Publishing, Univ. Press, Cambridge.
Zhang, D., E. Arevalo-Gardini, S. Mischke, L. Zuñiga-Cernades, A. Barreto-Chavez, and J. Adrizola del Aguila. 2006a. Genetic diversity and structure of managed and semi-natural populations of cacao (*Theobroma cacao*) in the Huallaga and Ucayali valleys of Peru. Ann. Bot. (Lond.) 98:647–655.
Zhang, D., M. Boccara, M. Lambert, D.R. Butler, P. Umaharan, S. Mischke, and L. Meinhardt. 2008. Microsatellite variation and population structure in the "Refractario" cacao of Ecuador. Conserv. Genet. 9:327–337.
Zhang, D., S. Mischke, R. Goenaga, A.A. Hemeida, and J.A. Saunders. 2006b. Accuracy and reliability of high-throughput microsatellite genotyping for cacao clone identification. Crop Sci. 46:2084–2092.
Zhang, D., S. Mischke, E. Johnson, W. Phillips-Mora, and L. Meinhardt. 2009. Molecular characterization of an international cacao collection using microsatellite markers. Tree Genet. Genomes 5:1–10.