Breeding Science 68: 545–553 (2018)
doi:10.1270/jsbbs.18057

Research Paper

Assessing hidden parentage and genetic integrity of the "United Fruit Clones" of cacao (*Theobroma cacao*) from Costa Rica using SNP markers

Allan Mata-Quirós[1], Adriana Arciniegas-Leal[1], Wilbert Phillips-Mora[1], Lyndel W. Meinhardt[2], Lambert Motilal[3], Sue Mischke[2] and Dapeng Zhang*[2]

[1] *Programa de Mejoramiento Genético de Cacao*, CATIE 7170, Turrialba 30501, Costa Rica
[2] *USDA-ARS BARC, SPCL*, Rm 223 Bldg 001 BARC-West, 10300 Baltimore Avenue, Beltsville, MD 20705, USA
[3] *Cocoa Research Center, University of the West Indies*, Trinidad & Tobago

The international cacao collection in CATIE, Costa Rica contains nearly 1200 accessions of cacao, mainly from the center of genetic diversity of this species. Among these accessions, the United Fruit clones (UF clones) were developed by the United Fruit Company in Costa Rica, and they represent one of the earliest groups of improved cacao germplasm in the world. Some of these UF clones have been used as key progenitors for breeding resistance/tolerance to Frosty Pod and Black Pod diseases in the Americas. Accurate information on the identity and background of these clones is important for their effective use in breeding. Using Single Nucleotide Polymorphism (SNP) markers, we genotyped 273 cacao germplasm accessions including 44 UF clones and 229 reference accessions. We verified the true-to-type identity of UF clones in the CATIE cacao collection and analyzed their population memberships using maximum-likelihood-based approaches. Three duplicate groups, representing approximately 30% of the UF clones, were identified. Both distance- and model-based clustering methods showed that the UF clones were mainly composed of Trinitario, ancient Nacional and hybrids between ancient Nacional and Amelonado. This result filled the information gap about the UF clones thus will improve their utilization for cacao breeding.

Key Words: CATIE, genebank, genetic diversity, tropical tree, Central America, Nacional, Amelonado.

Introduction

Cacao (*Theobroma cacao* L.), the source of cocoa powder and cocoa butter used for chocolate, is a tropical forest species native to South America that is cultivated extensively in tropical regions. The cacao collection at CATIE (Centro Agronómico Tropical de Investigación y Enseñanza) is one of two international cacao germplasm collections in the world; the other is the "International Cocoa Genebank, Trinidad", which is curated by the University of West Indies, St. Augustine, Trinidad and Tobago (CacaoNet 2012). The CATIE collection was initiated in 1944 to promote the exchange of germplasm of tropical crops. At the time this study was initiated, this collection maintained 1200 clones or accessions from Central America, Mexico, South America, the Caribbean, Asia, and Africa. Some of these accessions, such as the ancient Criollo, Amelonado and accessions from Brazil, are unique in this collection in terms of international distribution (Phillips-Mora *et al.* 2007). The collection has also been an important source for resistance to Frosty Pod and Phytophthora Pod diseases. In 1978, the collection was catalogued by the International Board for Plant Genetic Resources (IBPGR, now Bioversity International) as one of the two "International Cacao Collections". Since 2004, it has been under the auspices of the Food and Agriculture Organization and covered by an international treaty for the protection of plant genetic resources (CacaoNet 2012, Phillips-Mora *et al.* 2006).

Although the CATIE collection has been characterized using microsatellite DNA markers, comprehensive assessment of genetic integrity and genetic diversity of different germplasm groups remains to be accomplished (Zhang *et al.* 2009b). Some information gaps need to be filled in order to improve the accuracy and efficiency of conservation and utilization of this international collection. Among the CATIE cacao germplasm holdings, there are several groups of improved breeding lines which were selected from earlier breeding activities in Costa Rica and other countries in Central America. One of them is the "United Fruit Clones" (UF clones), which were developed by the United Fruit Company in Costa Rica (Engels 1981, Johnson *et al.* 2007). The

Communicated by Norihiko Tomooka
Received April 27, 2018. Accepted August 14, 2018.
First Published Online in J-STAGE on December 1, 2018.
*Corresponding author (e-mail: Dapeng.Zhang@ars.usda.gov)

Table 1. Summary list of 273 analyzed cacao accessions including UF clones, Trinitario clones and the ten reference germplasm groups

Population/group	Origin	Sample size	Provider
United Fruit (UF)	Costa Rica	44	CATIE, Costa Rica
Nacional	Ecuador	20	INIAP, Ecuador
Criollo	Puerto Rico, Honduras, Nicaragua	20	SPCL, TARS, USDA
Amelonado	Puerto Rico, Honduras, Nicaragua	20	SPCL, TARS, USDA
LCT EEN (Curaray)	Ecuador	20	INIAP, Ecuador
IMC (Iquitos)	Peru	20	ICG, T, Trinidad
Nanay	Peru	20	ICG, T, Trinidad
Parinari	Peru	25	ICG, T, Trinidad
Scavina/Ucayali (Contamana)	Peru	20	ICG, T, Trinidad and ICT, Peru
Purus	Brazil	25	SPCL, TARS, USDA
French Guiana	French Guiana	25	CIRAD, France
Trinitario	Trinidad	14	ICG, T, Trinidad

United Fruit Company was formed in 1899 when the Boston Fruit Company merged with the Tropical Trading and Transport Company, then with several other companies that produced, imported and marketed bananas from the Caribbean islands, Central America and Colombia. The principal founder was Minor C. Keith (the banana king of Costa Rica), who developed banana plantations in Costa Rica beginning in the 1870s. The company started cocoa production in the early 20th century, hoping to replace financial losses resulting from problems in the banana industry of this country (Johnson *et al.* 2007, Keithan 1940).

In 1907, cacao seedlings derived from pods introduced from Trinidad and Tobago were planted on the Caribbean coasts of Costa Rica and Panama. The pods from Trinidad were described as a Forastero Amelonado type having much larger pods and beans than the Amelonado cacao grown in Limon at that time. In 1936, after years of random hybridization and selection for pod size, bean size and yield, and after surveying thousands of trees in Limon, Costa Rica, advanced breeding lines were selected. The 12 best trees were clonally propagated and designated with UF (for United Fruit) numbers (**Table 1**). In mid-1940s, the United Fruit Company expanded the planting of cacao (and oil palm) as replacement crops on the west coast of Costa Rica, because of the widespread incidence of Panama disease on banana. Extensive evaluation was carried out in at the Quepos Los Rios farm, using planting materials introduced from Limon and other undocumented sources. An additional group of UF clones were selected after multi-location trials involving Almirante, Limon and Quepos. Today, a total of 44 UF clones are maintained at CATIE, Costa Rica in the "International Cocoa Collection" (IC3). Some of these clones, such as UF-12, UF-273 and UF-712 have been used as key progenitors for breeding new cacao varieties having resistance/tolerance to Frosty Pod disease. This disease, caused by *Moniliophthora roreri*, occurs in most major cacao producing areas in the Western Hemisphere, but will potentially spread to the major cocoa producing regions in Asia and Africa (Evans 2007, Phillips-Mora and Wilkinson 2007).

So far, the utilization of the UF clones as progenitors outside the Americas (e.g. Southeast Asia and West Africa) has been minimal. The lack of information on genetic background of the UF clones is one of the reasons for their limited utilization in breeding. The UF clones have been characterized morphologically and genotyped using SSR markers and a field guide has been developed based on these works (Johnson *et al.* 2007), which significantly improved our understanding about genetic diversity in this germplasm group. However, because at that time the diversity analysis was not conducted in the context of the entire cacao primary gene pool, the genetic background of the UF clones, in terms of their ancestry and relationships with other known germplasm groups, was not analyzed. It is believed that the UF clones were selected from MATINA (an Amelonado cacao variety in Costa Rica) material and Nacional type germplasm from Ecuador (Bartley 2005), but detailed information has not been available. The main objective of the present study was to verify the genetic identity and analyze the ancestry and genetic background of these breeding lines. The resultant information will be highly useful for improving the accuracy and efficiency of cacao genebank management in CATIE and will facilitate the further use of this germplasm in breeding new varieties with enhanced resistance to diseases, productivity and quality attributes.

Materials and Methods

Plant materials

The 44 UF Clones were sampled from the international cacao germplasm collection at CATIE, located in Turrialba in the province of Cartago, Costa Rica (**Table 1**). These trees were morphologically characterized in the 1970s (Engels 1981) based on a comprehensive list of morphological descriptors. Two examples of UF clones, showing different pod shapes were presented in **Fig. 1**. In addition to the 44 UF clones, 229 international clones were included in this experiment as references. These 229 reference clones represent 10 known germplasm groups of cultivated cacao, as classified by Motamayor *et al.* (2008). Genetic identities of these clones were determined through an international initiative of DNA fingerprinting of cacao (Zhang *et al.* 2009a, 2009b) and various studies on cacao germplasm management (Cosme *et al.* 2016, Ji *et al.* 2013, Motilal *et al.* 2010). The majority of the reference clones were maintained in the

Breeding Science BS
Vol. 68 No. 5

Fig. 1. Matured pods (fruits) of two UF clones, showing the pod shape of Nacional hybrid (UF-613) and Trinitario (UF-677).

two international gene banks in Trinidad and Costa Rica (Motilal *et al.* 2010, Zhang *et al.* 2009a, 2009b). These reference trees were sampled from the original collections maintained at Marper Farm and San Juan Estate in Trinidad, and Cabiria Farm at CATIE in Costa Rica. The rest of the reference clones were from the cacao samples collected by the USDA ARS Sustainable Perennial Crops laboratory from various national collections, including the Agricultural Research Institute (INIAP) of Ecuador and the Tropical Crop Institute (ICT) of Peru and Comissão Executiva do Plano da Lavoura Cacaueira and (CEPLAC) of Brazil. The summary list of these ten groups and UF clones were listed in **Table 1**.

Two healthy young leaves were collected from each tree, and the samples were dried in silica gel and sent to the USDA Beltsville agricultural Research Center, Maryland, USA for genotyping. DNA was extracted from dried leaf samples with the DNeasy® Plant Mini Kit (Qiagen, Inc., Valencia, CA), which is based on the use of silica as an affinity matrix. The remainder of the extraction method followed manufacturer's suggestions. DNA was eluted from the silica column with two washes of 50 µL Buffer AE, which were pooled, resulting in 100 µL DNA solution. Using a NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE), DNA concentration was determined by absorbance at 260 nm. DNA purity was estimated by the 260/280 ratio and the 260/230 ratio.

SNP markers and genotyping

Forty-eight SNP markers were selected from 1560 putative candidate SNPs based on cDNA sequences from a wide range of cacao organs (Allegre *et al.* 2012, Argout *et al.* 2008, Boccara, personal communication). The selection of SNPs was based on the level of polymorphism and their distribution across the ten chromosomes in cacao. The chosen 48 markers were used to design and manufacture primers for a SNPtype™ genotyping panel by the Assay Design Group at Fluidigm Corporation (San Francisco, CA). The full list of the 48 SNPs and their flanking sequences are presented in **Table 2**. Genotyping was performed on the high-throughput Fluidigm EP1™ system, using the Fluidigm SNPtype Genotyping Reagent Kit according to the manufacturer's instructions, and nanofluidic 48.48 Dynamic Array™ IFCs (Integrated Fluidic Circuit; Fluidigm Corp.). These chips automatically assemble PCR reactions, enabling simultaneous testing of up to 48 samples with 48 SNP markers. Fluorescent intensity was measured with the EP1™ reader and plotted in two axes. Genotypic calls were made using the Fluidigm SNP Genotyping Analysis program.

Data analysis

Key descriptive statistics for measuring informativeness of the SNP markers were calculated, including observed heterozygosity, expected heterozygosity, and probability of identity (Evett and Weir 1998, Waits *et al.* 2001). The program GenAlEx 6.2 (Peakall and Smouse 2012) was used for computation. For clone or duplicate identification, pairwise multilocus matching was applied among individual varieties and the reference clones, using the same program. Statistical rigor was assessed for match declaration using the probability of identity (PID) that two individuals may share the same multilocus genotype by chance (Waits *et al.* 2001). In computing PID, it was assumed that all individual genotypes were siblings (PID-sib), which was defined as the probability that two sibling individuals drawn at random from a population have the same multilocus genotype (Evett and Weir 1998, Waits *et al.* 2001). The overall PID-sib is 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). This can be computed using the following equation:

$$P_{ID \cdot sib} = 0.25 + (0.5\ \Sigma p_i^2) + [0.5(\Sigma p_i^2)^2] - (0.25\ \Sigma p_i^4)$$

where p_i is the frequency of the ith allele.

The computation was carried out using the program GenAlEx 6.2 (Peakall and Smouse 2012). Accessions with different names that were fully matched at the genotyped SNP loci were declared duplicates or synonymous accessions.

After examining duplicates in the analyzed samples, assignment test was applied to infer population membership and admixed ancestry (hybrids or ancestral forms) of the UF clones, using a model-based clustering method implemented in the software program STRUCTURE (Pritchard *et al.* 2000). The UF clones were analyzed, together with samples in the 10 reference groups that potentially had made ancestral contribution to these farmer selections. The number of clusters (K-value) was set to10, assuming that each of the 10 populations may have contributed to the UF clones. Ten independent runs were assessed for $K = 10$. The run with the highest Ln Pr ($X|K$) value of the 10 runs was chosen and presented in a bar plot. Q-value was used to present the ancestral contribution (membership) from each germplasm group. Accessions possessing ≥25% membership (Q-value)

Table 2. Forty-eight SNPs and their flanking sequences used for genotyping cacao germplasm accessions in CATIE, Costa Rica

SNP code	Chromosome number	SNPs and Flanking sequences
TcSNP25	9	TTTATGTTACTAGTTTTTGTTGGGTGGTTGTGTTTTGTTTGCATTATTTTGTGGAC[C/G]GGACCTTAGTATACTGCTTGTTTGCTGTTTTTTGTAGTAGTGGGTTGTTAGT
TcSNP32	4	TCACTTTTGAATATAGAAGATGGATGATATTGTGCAAAAATAATACCATAGT[A/T]CTAAGAAAATGCACATTTTTGTAAATAGGAAGAATGCCATGGTAATGTTGTGTTGA
TcSNP139	8	AAATAATACTAGTAACATTAGACCGATATTTATGTAGGGAGAAAGACATYTTGA[T/G]TTTGGCTTGGTGGTATGTTGATTTTTGTTGATACCAAAAAAGACCAGATG
TcSNP144	10	ACCGACACAGGATACCGTGAATGGGGAACTATGAAGGCAGCTTGCTATGTGTGGGA[A/C]ATTTGCACTTTGGGTCCAATAAGTTTAAATTACCCGACAGCTTATA
TcSNP150	5	ATACCCTGACGACTACTGCTGACTTGGGAACACCCACCAGCGTTTTGCCGGCAACA[T/G]GGTGACTGGTGATTTTCTAATTTTTATAGAACTTTGAAGCGTTTAAAATA
TcSNP151	8	TGAAGCTGGAGGGATTGTGGAGAGTGTTGGTGAGGGTGTGACTGACAACCAGGTGA[T/C]CATGTCCCATTTACTGGAGAATGCAAGGAGTGGCCACTGCTTGAGAAGAA
TcSNP193	9	GCATAGAAATTACGACAGCCCAATTTCTTGTTAAAAGTTCTGTTTAAT[A/C]GTTTCATCTGAAAATGTGATCAAGCTTTTGGCAGCCGCCGCAGCAA
TcSNP226	9	AAGCCCAAGGCCCAAGAGCTACGCGACGAAGGCAGTGAACCAGGCGCTGAGGGGGG[C/G]CACCGGCAGACTAAAGAAATGATGCTGGTGAACAAGAAAACCGAGGGGCCGG
TcSNP230	10	CAGCAGAGGCCGAAGACAAAGAAGGGGAGAGGATGTGACACCCAAGGTTTTT[A/G]GAAGATACAACGGAACACAAAAAGCTTTTACGCAACATAATTTGCT
TcSNP242	9	AGCAATCCGCCAAACAAGCCAAAAAACCCGTCTGTGCCTT[T/C]CTGTTGCCGAGAGAAACATTTGGGTCTTGCCAAT
TcSNP309	6	ATACGAGTAAGAAAAAAATTTACATTTTAAGTACAAGAAATTCCAAGATGCC[T/C]CTCACAAACAATATAGTAAAGCACACCGTCGTACCACCAGTGTGTT
TcSNP372	4	GGAAATTGAAGATTGTGAAGGAGAACTGATTAGACCTGGAATGATTTAAA[A/T]CTTGATTAAGAGGGGTGGCAATGGCAGGTTTTTGAAGACTGCTGCCTATGGGCATTTT
TcSNP429	2	TGGATGGATTATAAAGGTTGAGATGAATGACGCCGGGGAGTTGAAGAACTTGATGGGC[A/G]GAAGAATACCAAGTTGTGAGGAAGAAGATAAAGCACTGACCAGATGCCCTGGCC
TcSNP469	7	AGCCATGGCCGAGAAAAAGCCAGCCCGGTTAACCCGATTCGACCCGACTCAG[A/G]GAGACGTTACCGCGGCGTGAAAGCGCCCATGGGGCCGTTACGCGGCGGAGACGG
TcSNP529	1	ATAGGCCTTGAATGACTACTGACCTACAAGCAACAAAAGGCT[A/C]AGAAAAATAGGCATGTATGGCTTGCACTTGGGTTGCAGTTGCAAGCCGGCAGTA
TcSNP534	1	TGTAAAGAGTGCCATTGTAAGTGGATGCAGCGCCTTTTAAGCAATGGTACCTAGCA[T/C]TATGGTGGACATTGGGAGGAAGAAGAAAACAGCAGCTAAGAAGGAAGCTACTGAGGAA
TcSNP560	10	GAGGCAGAGAAATACAAGGCCGAGGATGAGGAGCACAAGAAGAAGGCGGAGGCCAAGAAT[T/G]CTTTGGAGAACTACGCCTATAACATGAGGAACACTGAAGGATGAGAAGATTGGCAA
TcSNP577	5	CAATTGCAAACACGCCATGATGAATGATGATGATAATGATGAACAGAAAACAAGAT[C/G]AACGGCCACAAATTTAACACAAATTACAAGGGAAAAAAAGGGGAGAAAATARAAAACA
TcSNP591	1	GGGCGGAGAAGACTGGCTGCTATTAAGGTTGAAACAGAGGGGCTTATGCGGA[A/C]GGAGAAGGCGAAAGATAGCCAGGCCATGCCGGAAAGGGCAAAATGAGGAAGC
TcSNP619	6	GAACAAAAATTGTAAATTAATTATGCATGGAGTAATGACCCCACAGCTTTGCAACACC[T/C]CAAAAATGGTGGTTGCCTTTGTACATAATGATTGGGATGAAATTTGTTTGTA
TcSNP645	5	TGAAGGCACTGGAGCCCTTTGTGAATTGGCTGGAGGAAGCAGAAGAAGAGGAATAAGTGC[A/G]ATAACAATAATATTATGAACTTTGAAAAGAAAAGGGGTGAACCCTGTACTTGTCTT
TcSNP723	10	AAATTGGAACCCCACCAAACCATGCATGGCCATGGATACTAGCAATGATGCTGCTT[T/G]GTTACCTTGCAGGCTGCCTTGGCTGCTTAGTGTTTGACAATGCTAAAAAC
TcSNP750	6	ACCTAGCCTTAAGCTGCAAGCCGCCTTGACAAAACAAACCTCTG[T/C]GTTGGTTGACACAGCATGAGCAGCTGGCTGTGGCAGCCCAACAGAGACCCAACCCA
TcSNP836	2	AAGGGTAACCATGTGGAAAGCTAAGTGCCATCACCATGTACATGGGAC[T/C]CCAACAACCTACATATGAGCCCATACTAAGAGTATGACCATGGAGCCCTG
TcSNP852	3	TGTCTTTACAACCATTGACTGATGGGTGTGAGTAAGACCAAGTGCA[C/G]TACTGTGGGTATAAGGCATTGCTTGGGCCTTTGGAGGCATGATTTGCTTGTA
TcSNP872	4	AGTGATGTTGCACAGGACAAAGCTAACCTTGTTGCCAGAACAAGAGGG[C/G]AGACCAAAGAACAAGGTGGTAGTCCCCAGTTACAAGCTTGTACCTTGC
TcSNP878	3	CCGAGGACGGCCGACACGGGCTGACGCCCGAGTACTAGTGGAACCC[C/G]ACCTGGATGACGGGAATGTAGGCCGGAGGTAAAGCGTACAGGAGGATTTGTTGA
TcSNP886	4	GCGAGTGCCGAGTAGGCAGGAGGAAAGTAGGGAGAAATAAGCGGGCGGTTTGAG[T/C]ACTAGAGACTAGGGGAGCTTTACTAGTATTATGTTTGACTTATGAGGTAAAG
TcSNP891	2	CAGAAAGCACACTTTGCACAAGGTTACACAGTATAAGAAGGGTAAGGATAGTTTGGCTGC[T/C]CAAGGGAAACGCCGTTATGAGCAAACAAAGGTTATGGTGGGCAGACCAAACCAGTG
TcSNP917	10	TATGGAAACTGGGTAAAGGCAAGAATGGGCGGCAATCTGGTGGACAAGCCACTTAC[T/C]TGGGCAACCCTGTCATGGGGTACCCCCTAATTGGAGAATTGCCA

Table 2. (continued)

SNP code	Chromosome number	SNPs and Flanking sequences
TcSNP929	3	TGAAATGGAAATGAAAATACAAAAAACCAAGATGTAATAGCAAGTGATGCTTTT[C/G]TGCTGATTATA ACAGAAGTTGTGTTAATTTTACACTTAGTTGATGTGGTTTTAAAGGG
TcSNP953	4	GGGTAAATGGCACTTGAGGTTGCCAGAGATAAAGCTATATGCCAAGCCAA[A/T]TTTGGGCATCAGGG AAACACCAAAAACAAGTTTTTATGGCAAAATG
TcSNP994	6	AACAAACCGAAATATATTAGGGAAACTTTGCATTTGCAACCCCTATTTGACTTGAT[T/C]TGCAGATGCT TTTGAACAACTTGAATAGAAYATGGGAGGAGGGTAG
TcSNP998	5	AACAAGAACAAGGGAATGGGGACAATATGACCACTAGACATGATGGCTTTCC[A/G]TAGTAGTGAGAA AGGAATAACATAGAATAAGCTGACAGATGCACTGCAACCT
TcSNP1038	5	GGACACCTAGACTAGGAGCAACCTGCTTTTGACAAGAAGCAGTTTGTAACCT[A/G]TGAAGAGGTACA AAGAACTTGACACCCAAATTAGAGCCAGAGAAGCAAGAGTTGTA
TcSNP1060	2	CAGTTTGATTTTGAGATTGAAGCTAACCTTGCACTGAGTACTTGGCACTTGGTAAC[T/C]GTTGGAAGA ATATGTAGCCTGAATAAGACGTGCCATGAACATATATGGATATTAGG
TcSNP1062	3	ATGCAAGAGGGAACTGGAGAGTAGTATGCTAAGCCGCATGAAATGCTGCAAG[A/G]GACCTTGACCCT TGATGAGAAGAACCCACGGGATTTGAGGGTGAGGCTTT
TcSNP1075	1	TTATGGCAGCATGCACTTATAATTTATGATTGCAACCCAACTGATACATAAATG[A/T]GTAGTAGGCCTGT ATAGATGAATTACGAAACAAAGCATGAGAGTGCATGT
TcSNP1144	6	GGTGGACTTTGTGGAGGAGATTTGCAGTATTATATGGATGAGAATTATTGGG[T/C]TGTTTTGCCCAAAC TGGAGAGGTTGTTAGTGAAAGTTACGTAACAAGCAAACAGGC
TcSNP1165	2	GAAATTGCTTAAACATATTTGCATAGCATTATAGATATTTAAAACCATGATGGAGG[T/C]TGCCAAGGTTT CTAGCTTACTTGATGGCCTTAGCCTTGGCCCCCCA
TcSNP1253	9	CACTTGCCACAAGTCACTAAAGCATTGAAACCAGCAAGAAGTGATTTA[T/G]ATTACCAGCACTTAAAA CTTTAAAGGATAGGTGAGTAAAGAAATGAGGCGCT
TcSNP1270	7	ATATTTGAGTTGTTTGATGTTTACTACAAGACCTGCCACTTCCGCAGCTT[T/C]GTACCAAGAGCCCTYC AATTTACCAATTATGTTGATTAGGGATGGCTTTCAGAT
TcSNP1350	1	CAAAATTTTTTAACTATATGATGGACCATATAGCCTAAATAAATATAGCAAAAATG[A/C]ATAACAACAA ATTATATGGCTGGCTTGCAAAAAGACTATAAGGGCTTGTGGCTAGT
TcSNP1414	9	TGTGACTTACGGTTACCCCAAAAGAGCGTGAGGGAGTTGATTTACAAAAGAGGTTA[T/C]GGGAAGTT GAACAAGCAGCGTGTTGCTTTGACTGACAATGAAAATTGAGCAGGCTG
TcSNP1442	9	AAAAGTGATGGAGGAAAGGGAAAGAGATGGGTGGGAGTAGAGATGGCCTTTGGTGTTC[T/C]TATGGA TTTACTACAAAAGCTTTGGTGTGTGATGGAATTTACAGCTACTGTTAT
TcSNP1458	1	TGGATGAGAGCTAAAAGAATTAGAAGGAAAAGCAGATGGCGA[C/G]CAGAGTTTGCAAAGAGAAGCC GACCTCTTACCTTGATTGAGCTG
TcSNP1484	6	AAAACAGAAACGGGGTTGACTTAGCCGCAGCTGTGTAAACACAC[A/G]AGGGGACAGATGGCTGACT GAGAAAGACTGGACCGACGTTGAGTTTAGGG
TcSNP1520	8	ACTGCCTAAATATATATGATGAAGAAAAAGCTTTGGAGAAAACAAGGAAGGTTGAC[T/C]GAGAAGAT TGCAGCTGAACTGCTATTGACGATGTTTGCAGCCGAACTGAA

in a given cluster were considered as receiving a significant ancestry contribution from that cluster (genetic group). Accessions possessing ≥75% membership were considered to be a member of that cluster. Accessions possessing >25% but <75% membership were considered as hybrids of two (or more) clusters.

After assignment test, multivariate analysis was used to provide a complementary assessment of the relationship among the UF clones and their relationships with reference clones from international genebanks. In this analysis, we included only ancestry populations that are relevant to the origin of the UF clones, based on the result of assignment test. Pair-wise genetic distance was computed for every pair of accessions, using the genetic distance procedure in GenAlEx 6.2 (Peakall and Smouse 2006). The same program was then used to perform Principal Coordinates Analysis (PCoA), based on the pairwise distance matrix. Both distance and covariance were standardized.

Results

Of the 48 SNP markers, 44 were successful in genotyping across all 313 samples. The remaining four SNPs (Tc1038, Tc1144, Tc 1165 and Tc226) had a low success rate (< 90%) thus were removed from the data set. A total of 44 polymorphic SNPs were retained for further analysis. Based on the 44 SNP markers, the expected heterozygosity in the UF clones was 0.741, whereas the observed heterozygosity was 0.694. Inbreeding coefficient was negligible in the UF clones. This result revealed that UF clones are mainly composed by hybrids involving different germplasm groups.

Multilocus matching, based on 44 SNP markers, revealed three synonymous groups, involving 16 clones (**Table 3**). SNP profiles of the repeated genotyping on DNA samples that had been independently extracted from the same accessions showed that genotyping results were highly consistent. The probability that two UF clones will have the same genotype at the 44 SNP loci is approximately 1 in

Table 3. Identified synonymous groups (duplicates) in the UF clones maintained in CATIE, Costa Rica based on SNP markers (showing truncated profiles of 48 SNPs)

Synonymous group	Sample name	TcSNP 25	TcSNP 32	TcSNP 139	TcSNP 144	TcSNP 150	TcSNP 151	TcSNP 193	TcSNP 226	TcSNP 230	TcSNP 242	TcSNP 309	TcSNP 372
1	UF 10A	G G	A T	G T	A C	G T	C T	A C	C G	A G	C T	C T	A T
1	UF 601	G G	A T	G T	A C	G T	C T	A C	C G	A G	C T	C T	A T
1	UF 168	G G	A T	G T	A C	G T	C T	A C	C G	A G	C T	C T	A T
1	UF 601	G G	A T	G T	A C	G T	C T	A C	C G	A G	C T	C T	A T
1	UF 650	G G	A T	G T	A C	G T	C T	A C	C G	A G	C T	C T	A T
1	UF 654	G G	A T	G T	A C	G T	C T	A C	C G	A G	C T	C T	A T
1	UF 667	G G	A T	G T	A C	G T	C T	A C	C G	A G	C T	C T	A T
1	UF 668	G G	A T	G T	A C	G T	C T	A C	C G	A G	C T	C T	A T
1	UF676	G G	A T	G T	A C	G T	C T	A C	C G	A G	C T	C T	A T
1	UF 677	G G	A T	G T	A C	G T	C T	A C	C G	A G	C T	C T	A T
2	UF 29	C G	A T	G T	C C	G T	C T	A C	C C	A A	C C	C T	A T
2	UF 242	C G	A T	G T	C C	G T	C T	A C	C C	A A	C C	C T	A T
2	UF 705	C G	A T	G T	C C	G T	C T	A C	C C	A A	C C	C T	A T
3	UF 716	C G	A T	G G	C C	G T	C T	A C	C C	A G	C C	T T	A T
3	UF 717	C G	A T	G G	C C	G T	C T	A C	C C	A G	C C	T T	A T

1,000,000 for the tested UF clones, as computed by the multilocus matching procedure implemented in GenAlEx 6.5. In total, the duplicated accessions accounted for approximately 30% of the UF clones maintained in this collection. The clones in the three synonymous groups were excluded in subsequent analyses of genetic diversity and population structure.

Model-based assignment test showed that out of the 44 UF clones, twelve were identified as classical Trinitario (**Table 4**). These clones (UF 613, 650, 652, 654, 666, 667, 668, 672, 676, 677, 678 and 679) are descendants of Trinidad and Tobago germplasm and were selected in 1936 by the United Fruit Company in Limon, Costa Rica, after years of random hybridization and selection. At that time the germplasm of Upper Amazon Forastero had not been collected and used in cacao breeding. Therefore Criollo and Amelonado were the only available parental lines. The result is comparable with the early selections (e.g. the Imperial College Selections) in Trinidad and Tobago, which are mainly hybrids between Criollo and Amelonado.

In addition to the Trinitario type accessions, two Nacional type clones, UF-20 and UF-712, were revealed in the present study (**Table 4**). Both clones have a Q-value above 95%, thus are assigned to the Nacional group, showing their full Nacional membership. These clones reflected the introduction of native cacao germplasm from Ecuador after the UF Company's purchase of large cacao plantation in Tenguel, Ecuador. After the catastrophic disease attack in the 1910s–1920s, cacao production was replaced by banana in the coast of Ecuador and some of the cacao germplasm was transferred to Costa Rica.

The last group of UF clones was all Nacional hybrids, with Nacional membership ranging from 0.25 to 0.77 (**Table 4**). This group of UF clones probably represented selections in the breeding program of United Fruit Company during a later stage (1944–1950) when the indigenous germplasm from Ecuador was hybridized with Amelonado cacao in Costa Rica.

Result of the distance-based Principal Coordinates Analysis (**Fig. 2**) fully supported the Bayesian clustering outcome. The plane of the first three main PCO axes accounted for 25.3%, 12.9% and 9.0% of total variation, respectively. The relevant reference clones were clustered in six groups, which matched well with their known classification in cacao germplasm groups. The three types of UF clones, including classical Trinitario, ancient Nacional and Nacional hybrids were clearly separated in the PCoA plot (**Fig. 2**).

Discussion

The International Cacao Collection at CATIE (Centro Agronómico Tropical de Investigación y Enseñanza) in Costa Rica and ICGT in Trinidad and Tobago are two universal collections covering all of the known genetic groups (CacaoNet 2012). As with most other cacao germplasm collections, information gaps on passport data remain to be filled. Some 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 (Motilal *et al.* 2013, Wadsworth and Harwood 2000). Significant efforts have been made to solve the problem of mislabeling in some international cacao collections (Motilal *et al.* 2013, Zhang *et al.* 2009b); however, the problem in most of the various national collections has not been fully resolved. In the past few years, microsatellite markers have been widely used in cacao genotyping and individual identification, enabling systematic assessment of genetic identity in national and international cacao genebanks (Motilal *et al.* 2009, 2010, Zhang *et al.* 2009b). Reference SSR profiles of cacao clones have been deposited in the International Cacao Germplasm Database at the University of Reading, UK (http://www.icgd.rdg.ac.uk/index.php). However, comparison of genotyping results from different laboratories has not

Table 4. Assigned population membership representing inferred pedigree of 44 United Fruit clones from the international cacao genebank, CATIE, Costa Rica

		Assigned membership (Q-value)									
Accession	Inferred pedigree	Nacional	Amelonado	Criollo	Parinari	Guiana	IMC	LCT EEN	Nanay	Purus	Scavina
UF 4	Nacional × Amelonado	0.470	0.510	0.000	0.000	0.000	0.010	0.010	0.010	0.000	0.000
UF 10	Trinitario	0.000	0.510	0.470	0.000	0.000	0.000	0.000	0.000	0.010	0.010
UF 11	Trinitario	0.000	0.520	0.460	0.010	0.000	0.010	0.000	0.000	0.000	0.000
UF 12	Trinitario	0.000	0.520	0.460	0.000	0.000	0.000	0.010	0.010	0.010	0.000
UF 168	Trinitario	0.000	0.520	0.460	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UF 221	Trinitario	0.000	0.520	0.460	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UF 601	Trinitario	0.000	0.520	0.460	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UF 650	Trinitario	0.000	0.550	0.430	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UF 654	Trinitario	0.000	0.510	0.470	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UF 667	Trinitario	0.000	0.520	0.460	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UF 668	Trinitario	0.000	0.510	0.480	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UF 676	Trinitario	0.000	0.520	0.460	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UF 677	Trinitario	0.000	0.520	0.460	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UF 242	Nacional × Amelonado	0.470	0.500	0.000	0.010	0.000	0.000	0.000	0.010	0.000	0.000
UF 705	Nacional × Amelonado	0.480	0.500	0.000	0.000	0.000	0.000	0.000	0.010	0.000	0.000
UF 716	Nacional × Amelonado	0.410	0.550	0.010	0.010	0.000	0.010	0.010	0.000	0.000	0.010
UF 717	Nacional × Amelonado	0.410	0.550	0.010	0.010	0.000	0.010	0.010	0.000	0.000	0.010
UF 20	Nacional	0.990	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UF 29	Nacional × Amelonado	0.450	0.510	0.000	0.010	0.010	0.010	0.000	0.010	0.000	0.010
UF 36	Nacional × Amelonado	0.510	0.440	0.010	0.000	0.000	0.010	0.000	0.000	0.010	0.010
UF 38	Amelonado × Scavina	0.020	0.390	0.030	0.010	0.000	0.000	0.010	0.000	0.000	0.540
UF 93	Trinitario	0.000	0.600	0.380	0.010	0.000	0.000	0.000	0.000	0.000	0.000
UF 122	Amelonado	0.000	0.830	0.150	0.000	0.000	0.000	0.010	0.000	0.000	0.000
UF 210	Nacional × Amelonado	0.440	0.530	0.000	0.000	0.000	0.010	0.000	0.000	0.000	0.000
UF 273 T1	Nacional × Amelonado	0.630	0.320	0.000	0.020	0.010	0.010	0.000	0.010	0.010	0.000
UF 273 T2	Nacional × Amelonado	0.630	0.250	0.000	0.090	0.010	0.010	0.000	0.010	0.010	0.000
UF 296	Nacional × Amelonado	0.370	0.590	0.000	0.010	0.010	0.010	0.000	0.010	0.000	0.010
UF 602	Amelonado	0.000	0.950	0.010	0.000	0.010	0.000	0.010	0.010	0.000	0.000
UF 613	Nacional × Amelonado	0.350	0.570	0.040	0.010	0.010	0.010	0.000	0.000	0.000	0.010
UF 666	Trinitario	0.000	0.510	0.470	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UF 672	Trinitario	0.000	0.620	0.370	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UF 700	Trinitario	0.000	0.690	0.280	0.000	0.000	0.010	0.000	0.000	0.000	0.010
UF 701	Trinitario	0.000	0.770	0.210	0.000	0.000	0.000	0.010	0.000	0.000	0.000
UF 703	Nacional × Amelonado	0.440	0.540	0.000	0.000	0.000	0.010	0.000	0.010	0.000	0.000
UF 704	Nacional × Amelonado	0.480	0.490	0.000	0.000	0.000	0.000	0.010	0.010	0.010	0.010
UF 706	Amelonado	0.000	0.790	0.130	0.020	0.020	0.010	0.010	0.000	0.000	0.010
UF 707	Trinitario	0.000	0.730	0.220	0.000	0.000	0.000	0.030	0.000	0.010	0.000
UF 708	Trinitario	0.000	0.650	0.330	0.000	0.000	0.000	0.000	0.000	0.010	0.000
UF 709	Trinitario	0.000	0.730	0.210	0.000	0.010	0.000	0.010	0.010	0.020	0.000
UF 711	Nacional × Amelonado	0.730	0.200	0.000	0.010	0.010	0.010	0.010	0.010	0.020	0.010
UF 712	Nacional	0.990	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
UF 713	Nacional × Amelonado	0.440	0.510	0.000	0.010	0.010	0.010	0.000	0.010	0.010	0.000
UF 714	Nacional × Trinitario	0.250	0.450	0.200	0.020	0.030	0.010	0.000	0.000	0.010	0.020
UF 715	Nacional × Trinitario	0.160	0.520	0.170	0.000	0.000	0.010	0.010	0.000	0.030	0.090

been straightforward. The effectiveness of clone identification via SSR fingerprints depends on the number of loci used for genotyping, as well as the rate of genotyping error. For example, it may require multiple repeated genotyping runs to reach the "consensus genotype". Moreover, data generated from different genotyping platforms can be difficult to compare with each other because the same allele may be binned differently, leading to false conclusions.

The present study demonstrated that SNP-based multilocus fingerprints significantly improved the efficiency of genotype identification. The "UF Clones" in Costa Rica was one of the earliest groups of improved cacao germplasm now maintained in the international germplasm collection in CATIE, Costa Rica. Although morphological characterization has been done on these clones, accurate identification of an individual clone had not been achieved. Our result showed that synonymously mislabeled clones can be accurately identified through the comparison of a small set of SNP markers. Moreover, since SNP genotyping can be done in high-throughput fashion and executed by a centralized service provider, as shown by the recent example in West Africa (Padi *et al.* 2015), verification of large numbers of trees can be achieved rapidly with reasonable cost.

In addition to accurate genotype identification, this small set of SNP markers allowed us to clarify the genetic background of the UF clones, by comparing them with the ten known germplasm groups. Through multivariate clustering analysis and assignment test, we delineated their origin and genetic background. Both distance- and model-based clustering methods showed that the UF clones were composed of Trinitario and Nacional background. This genetic background is highly similar to the Refractario cacao maintained

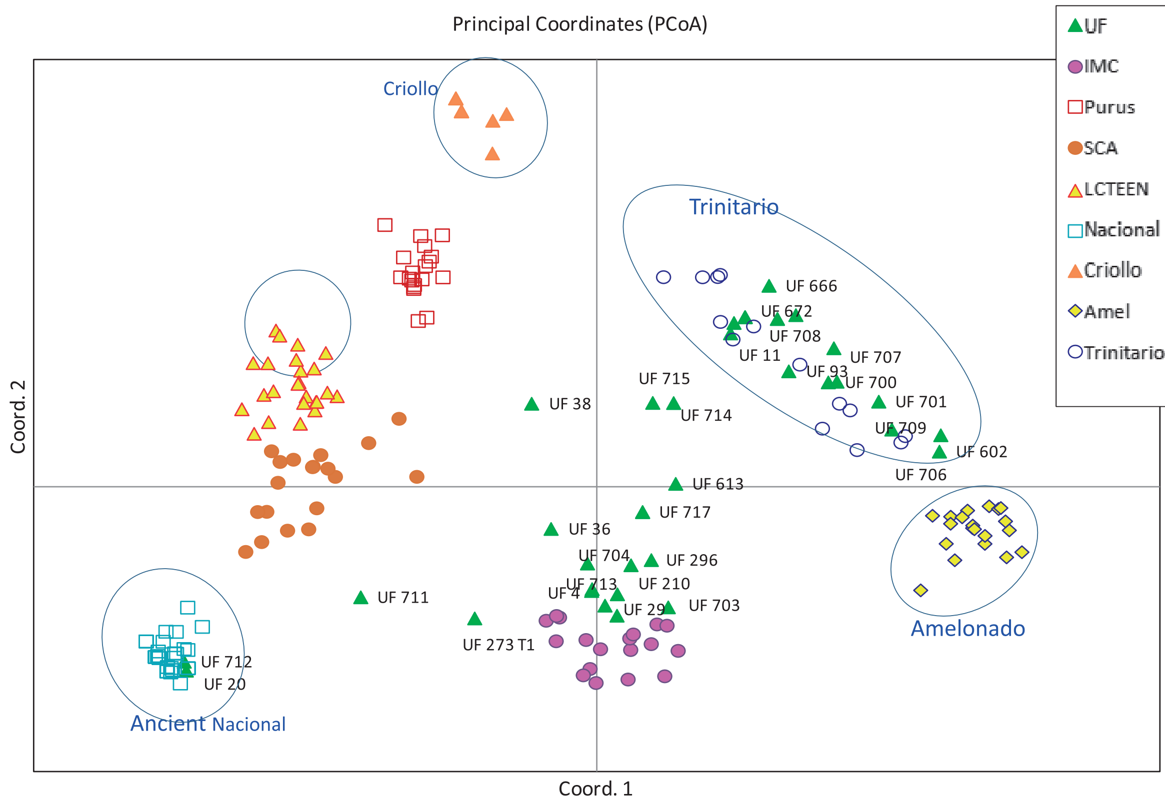

Fig. 2. PCoA plot of UF cacao clones and relevant references from other cacao germplasm collections. The plane of the first three main PCO axes accounted for 47.2% of total variation. First axis 25.3% of total information, the second 12.9% and the third 9.0%.

in the international genebank in Trinidad (Zhang *et al.* 2008). Refractario cacao was collected from the coast of Ecuador by J.F. Pound after the catastrophic disease infection in the 1920s. However the proven resistance to Frosty Pod in some of the UF clones, such as UF-712 and UF-273, make them highly valuable for cacao breeders.

The present study used a germplasm panel of 229 accessions to represent the 10 known germplasm groups. The classification system of 10 germplasm clusters was reported by Motamayor *et al.* (2008) based on 90 SSR markers. In the present study, we were able to differentiate the 10 germplasm clusters based on the 44 SNP markers. This reference germplasm panel allowed us to assess ancestry admixture or infer parentage for the UF clones. In addition to the UF clones, the CATIE cacao collection contains several other groups of improved cacao germplasm including ARF (Área de Recursos Fitogenéticos), CC (Cacao Center Selections) and PMCT (Programa de Mejoramiento de Cultivos Tropicales). For these breeding lines, the reference germplasm panel of 10 clusters is being used to verify population membership and/or recorded pedigrees, which are essential for maintaining correct passport information records and to facilitate a better use of these breeding lines for cacao genetic improvement.

Frosty Pod caused by *Moniliophthora roreri*, is a major concern due to its devastating effects on yields and limited available control measures (Evans 2007, Phillips-Mora and Wilkinson 2007). The breeding program in CATIE, Costa Rica started breeding for improved resistance to FPR in early 1990s. The UF clones have been used as the main source for FPR resistance (Phillips-Mora *et al.* 2005). *Moniliophthora roreri* was confined to northwestern South America until the 1950s; now it is found in 11 countries in tropical America. *M. roreri*, is a very aggressive pathogen that has the capacity to survive under extreme environmental conditions; has a rapid dispersal mechanism and the propensity for human dispersion, and is capable of infecting most commercial cacao genotypes, all of which makes Frosty Pod disease a substantial threat to the worldwide cultivation of cacao (Bailey *et al.* 2013). Preventive breeding in uninfected cacao-producing regions has been proposed but this practice has not been implemented except by Brazil (Gutierrez *et al.* 2016, Phillips-Mora *et al.* 2007). So far, in Southeast Asia and West Africa where 85% of the world's cocoa production are based, sources of resistance to FPR are still needed to be incorporated in the breeding program or in seed gardens. Lack of information on the genetic background of these breeding lines may partially explain breeders' reluctance to use the UF germplasm. The present study fills the information gap regarding their origin and genetic background, and the FPR-resistant UF clones, with a Nacional background that differs from the currently used progenitors in Asia and Africa, should be incorporated in their breeding programs and seed gardens.

Acknowledgments

The authors would like to thank Lin Zhou for assistance in DNA sample preparation and SNP genotyping, and Bernadette LeMasters of the World Cocoa Foundation for logistical support. The authors would also like to acknowledge the USDA-Foreign Agricultural Service and the World Cocoa Foundation for the Borlaug Cocoa Fellowship program.

Literature Cited

Allegre, M., X. Argout, M. Boccara, O. Fouet, Y. Roguet, A. Bérard, J.M. Thévenin, A. Chauveau, R. Rivallan, D. Clement *et al.* (2012) Discovery and mapping of a new expressed sequence tag-single nucleotide polymorphism and simple sequence repeat panel for large-scale genetic studies and breeding of Theobroma cacao L. DNA Res. 19: 23–35.

Argout, X., O. Fouet, P. Wincker, K. Gramacho, T. Legavre, X. Sabau, A.M. Risterucci, C. Da Silva, J. Cascardo, M. Allegre *et al.* (2008) Towards the understanding of the cocoa transcriptome: production and analysis of an exhaustive dataset of ESTs of Theobroma cacao generated from various tissues and under various conditions. BMC Genomics 9: 512.

Bailey, B.A., J. Crozier, R.C. Sicher, M.D. Strem, R.L. Melnick, M.F. Carazzolle, G.G.L. Costa, G.A.G. Pereira, D. Zhang, S. Maximova *et al.* (2013) Dynamic changes in pod and fungal physiology associated with the shift from biotrophy to necrotrophy during the infection of *Theobroma cacao* by *Moniliophthora roreri*. Physiol. Mol. Plant Pathol. 81: 84–96.

Bartley, B.G.D. (2005) The genetic diversity of cacao and its utilization. CAB International, CABI Publishing, Wallingford, Oxfordshire, p. 341.

CACAONET (2012) A global strategy for the conservation and use of cacao genetic resources, as the foundation for a sustainable cocoa economy. Bioversity International, Rome, p. 186.

Cosme, S., H.E. Cuevas, D. Zhang, T.K. Oleksyk and B.M. Irish (2016) Genetic diversity of naturalized cacao (*Theobroma cacao* L.) in Puerto Rico. Tree Genet. Genomes 12: 88.

Engels, J.M.M. (1981) Genetic resources of cacao: a catalogue of the CATIE collection. Technical Series: Technical Bulletin No. 7. CATIE, Turrialba, Costa Rica, p. 196.

Evans, H.C. (2007) Cacao diseases—the trilogy revisited. Phytopathology 97: 1640–1643.

Evett, I.W. and B.S. Weir (1998) Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists. Sinauer, Sunderland, Massachusetts, USA, p. 291.

Gutierrez, O.A., A. Campbell and W. Phillips-Mora (2016) Breeding for disease resistance in cacao. *In*: Bailey, B.A. and L.W. Meinhardt (eds.) Cacao Diseases: A history of old enemies and new encounters, Switzerland, Springer, pp. 567–609.

Ji, K., D. Zhang, L. Motilal, M. Boccara, P. Lachenaud and L.W. Meinhardt (2013) Genetic diversity and parentage in farmer varieties of cacao (*Theobroma cacao* L.) from Honduras and Nicaragua as revealed by single nucleotide polymorphism (SNP) markers. Genet. Resour. Crop Ev. 60: 441–453.

Johnson, E., W. Phillips, F. Bekele, D. Zhang and R.J. Schnell (2007) Field guide to the UF clones of Costa Rica. Proceedings of the International Cocoa Producer's Conference. San Jose, Costa Rica, Oct. 9–14. Vol. I, pp. 641–646.

Keithan, E.F. (1940) Cacao in Costa Rica. Economic Geography 16: 79–86.

Motamayor, J.C., P. Lachenaud, J.W. da Silva e Mota, G. Loor, D.N. Kuhn, J.S. Brown and R.J. Schnell (2008) Geographic and genetic population differentiation of the Amazonian chocolate tree. PLoS ONE 3: e3311.

Motilal, L.A., D. Zhang, P. Umaharan, S. Mischke, M. Boccara and S. Pinney (2009) Increasing accuracy and throughput in large-scale microsatellite fingerprinting of cacao field germplasm collections. Trop. Plant Biol. 2: 23–77.

Motilal, L.A., D. Zhang, P. Umaharan, S. Mischke, V. Mooledhar and L.W. Meinhardt (2010) The relic Criollo cacao in Belize—genetic diversity and relationship with Trinitario and other cacao clones held in the International Cocoa Genebank, Trinidad. Plant Genet. Resour. 8: 106–115.

Motilal, L.A., D. Zhang, S. Mischke, L.W. Meinhardt and P. Umaharan (2013) Microsatellite-aided detection of genetic redundancy improves management of the International Cocoa Genebank, Trinidad. Tree Genet. Genomes 9: 1395–1411.

Padi, F.K., A. Ofori, J. Takrama, E. Djan, S.Y. Opoku, A.M. Dadzie, R. Bhattacharjee, J.C. Motamayor and D. Zhang (2015) The impact of SNP fingerprinting and parentage analysis on the effectiveness of variety recommendations in cacao. Tree Genet. Genomics 11: 44.

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.

Peakall, R. and P.E. Smouse (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28: 2537–2539.

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

Phillips-Mora, W., A. Mora, E.S. Johnson and C. Astorga (2007) Recent efforts to improve the genetic and physical conditions of the international cacao collection at CATIE. Proceedings of the 15th COPAL International Cocoa Research Conference, San Jose, Costa Rica, Oct. 9–14, 2006.

Phillips-Mora, W. and M.J. Wilkinson (2007) Frosty pod of cacao, a disease with a limited geographic range but unlimited potential for damage. Phytopathology 97: 1644–1647.

Pritchard, J.K., M. Stephens and P. Donelly (2000) Inference of population structure from multilocus genotype data. Genetics 155: 945–959.

Wadsworth, R.M. and T. Harwood (2000) International Cocoa Germplasm Database, ICGD 2000 V4.1. London International Financial Futures and Options Exchange and the University of Reading, UK.

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.

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

Zhang, D., M. Boccara, L. Motilal, S. Mischke, E.S. Johnson, D. Butler, B.A. Bailey and L.W. Meinhardt (2009a) Molecular characterization of an earliest cacao (*Theobroma cacao* L.) collection from upper Amazon using microsatellite DNA markers. Tree Genet. Genomes 5: 595–607.

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