International Symposium on Cocoa Research (ISCR), Lima, Peru, 13-17 November 2017

UNDERSTANDING THE GENETIC STRUCTURE AND PARENTAGE OF THE CLONAL SERIES OF CACAO UF, CC, PMCT AND ARF PRESERVED IN THE INTERNATIONAL CACAO COLLECTION AT CATIE (IC3)

A. Mata-Quirós[1], A. Arciniegas-Leal[1], W. Phillips-Mora[1], L.W. Meinhardt[2] and D. Zhang[2]

[1]CATIE 7170, Turrialba, Costa Rica.
[2]USDA-ARS BARC PSI SPCL, Rm 223 Bldg 001 BARC-West-10300 Baltimore Avenue Beltsville, MD 20705, USA.

RESUMEN

La Colección Internacional de Cacao del CATIE fue creada en 1944 y conserva bajo el dominio público más de 1200 introducciones. Esta colección custodia 4 series de clones internacionales originadas en Costa Rica. Los clones UF fueron seleccionados por la United Fruit Co., principalmente a partir de materiales amelonados introducidos de Trinidad en 1913 y de tipo Nacional introducidos de Ecuador en 1928. Los clones CC (Centro de Cacao) fueron seleccionados por el Instituto Interamericano de Ciencias Agrícolas (IICA) entre 1957-1966. Son selecciones realizadas a partir de semillas de polinización abierta de clones UF o Matina y cruzamientos dirigidos que incluyen clones internacionales como EET-62, ICS-1, 6 y 39; Pound-12, SCA-6 y 12, etc. Los clones PMCT (Programa de Mejoramiento de Cultivos Tropicales) son selecciones del CATIE realizadas entre 1989 y 1992 a partir de árboles élites de ensayos de campo y fincas de agricultores producto de cruzamientos entre clones internacionales. También incluyen genotipos acriollados colectados en Nicaragua, Belice y Costa Rica. Los clones ARF (Área de Recursos Fitogenéticos) fueron colectados por el CATIE en 1992 en Panamá, Belice, Honduras y Costa Rica. Además incluyen genotipos resistentes a mazorca negra seleccionados en ensayos interclonales del CATIE.
El estudio se realizó en el Sustainable Perennial Crops Laboratory (ARS-USDA) en Beltsville, MD. El objetivo fue analizar el parentesco y composición genética de 266 clones de cacao pertenecientes a las cuatro series mencionadas utilizando 48 marcadores SNP y 228 clones de referencia que representan la diversidad genética conocida de la especie. El grupo genético predominante en los clones fue Amelonado con una presencia que varió entre el 36,7% (Serie ARF) y el 55,6% (Serie CC). Los grupos Nacional y Criollo tienen una presencia del 22,8 y 21,3% en la Serie UF, pero fue menor al 13% en las demás series. El grupo Marañon aportó 17,4% a la serie ARF, mientras el grupo Contamana tuvo una importante presencia (8,2-10,6%) en las Series PMCT, ARF y CC debido al aporte de SCA-6 y SCA-12 en los cruzamientos. La presencia del 8,6% del grupo Iquitos en la serie ARF se deriva de la participación del clon IMC-67 en algunos de los cruces que le dieron origen a estos clones. Los grupos genéticos Guiana, Curaray y Purus tienen una escasa participación (≤ 2%) en la composición de las series analizadas, excepto en el grupo ARF en donde el 7,8% de su composición proviene del grupo Purus.

INTRODUCTION

The future of world cacao production depends on the availability and use of the genetic diversity of the species, which is essential for the generation of new and better varieties. An important part of this diversity is conserved in genebanks, among them the only two international collections, located at the University of the West Indies (CRU / UWI), in Trinidad and Tobago and at the Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) in Turrialba, Costa Rica (CacaoNet 2012).

The International Cacao Collection at CATIE (IC3) was established in 1944; catalogued as International Collection by the IBPGR (now Bioversity) in 1978 and included in the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) in 2004, becoming the first collection of cacao under public domain (Phillips-Mora *et al*. 2006b).

IC3 contains 1235 clones or accessions from Central America, South America, the Caribbean, Asia and Africa, although the wild genotypes from the upper and lower Amazon are predominant. Twenty two percent of the genotypes preserved at CATIE belongs to clonal series developed in Costa Rica by the United Fruit Company (UF Series), IICA (Series CC) and CATIE (Series ARF and PMCT) from 1928 to 1992. CATIE is the global referent and the custodian of these series.

Clones from the series UF (United Fruit) originated from Amelonado genotypes introduced from Trinidad and Tobago in 1913 and Nacional genotypes from Ecuador in 1928 (Bowman, cited by Morera *et al*. 1991;

Johnson *et al.* 2006). It is very probable that the local Amelonado variety (Matina) predominating at that time in Costa Rica had had a very relevant role in the generation of the UF clones. After a hybridization process, the United Fruit Co. selected superior trees based on the size of the pod, bean size and productivity. The best trees were propagated and established in plantations of the company in Limón and Quepos, as well as in Almirante, Panama (Morera *et al*. 1991; Johnson *et al*. 2006).

The CC (Centro de Cacao) clones were selected by IICA between 1957 and 1966 from open pollinated seeds of UF or Matina clones and from crosses among international clones such as EET-62, ICS-1, ICS-6, ICS-39, IMC-67, Pound-12, SCA-6, SCA-12, etc. selected for their high production in experiments established at the Experimental Station of La Hulera in Turrialba and in La Lola Farm in Limón in 1950 and 1959, respectively (Soria and Esquivel 1967; Esquivel and Soria 1967).

The clones from the series PMCT (Programa de Mejoramiento de Cultivos Tropicales) were selected by CATIE during the period 1989-1992 from farms or field trials comprising crosses among high yielding international clones. The selection was part of a strategy to search for new genetic materials with higher production and resistance to diseases (Morera and Mora 1991). This series also includes Criollo-blood genotypes collected in Nicaragua, Belize and Costa Rica, which represent 33% of the clones of this group (CATIE 1990).

ARF clones (Área de Recursos Fitogenéticos) were collected by CATIE in 1992 in Panama, Belize, Honduras and Costa Rica. Also, it includes black pod-resistant genotypes selected in inter-clonal trials in CATIE among the clones CC-42, Catongo, Pound-7, SCA-6, UF-29, UF-613 and UF-676 (Phillips-Mora and Galindo 1991a, Phillips-Mora and Galindo 1991b) and progenies of the cross "Catongo x Pound-12" (Fritz *et al.* 1995; CATIE 1990).

Out of the 266 clones of the four series preserved at CATIE, only the following are present at the International Cocoa Quarantine Centre of the University of Reading: ARF-12, CC-41, CC-137, CC-252, PMCT-93, UF-168, UF-273, UF-613, UF-667, UF-676 and UF-712 (Turnbull and Hadley 2017).

Genotypes belonging to the four series have been widely distributed worldwide, particularly some UF and CC clones which are present in many producing regions and have been used as parents for the generation of new varieties. UF clones have participated as progenitors of clones from the following series: CC and PMCT from Costa Rica; AX from Trinidad and Tobago; EET from Ecuador; IFC from Côte d'Ivoire, and CEPEC from Brazil (Turnbull and Hadley 2017). CC clones also participated in the generation of the CEPEC series of clones.

In Latin America, UF clones have also been widely spread, such as UF-12, UF-221, UF-613, UF-650 (synonym EET-8), etc. They are recommended for commercial plantings in Colombia (Perea *et al*. 2013) and Peru (García 2007). Clones UF-712 and particularly UF-273 have been widely used in crosses in Costa Rica, Mexico and Honduras since their identification as tolerant clones to frosty pod rot in the early 1980s (Phillips-Mora 1986). On the other hand, clones CC-137 and PMCT-58 and three descendants of UF-273 (CATIE-R1, CATIE-R4 and CATIE-R6) are having a significant expansion in Central America as part of the polyclone recommended by CATIE for commercial plantings due to its good production and tolerance to moniliasis (*Moniliophthora roreri)* (Phillips-Mora *et al.* 2012).

One of the most important problems in germplasm collections worldwide is mislabeling, high rate of redundancy and lack of information on the genetic background of the accessions (CacaoNet 2012), which has limited its efficient evaluation and utilization. It is estimated that there are between 15-44% of mislabeling in the genebanks (Motilal and Butler 2003, Motilal 2004, Takrama *et al*. 2005, Sounigo *et al.* 2006). In CATIE's genebank it has been determined that in general, there is a high level of genetic diversity, mainly arisen from the genotypes from South America. However, a close interrelationship between some groups of clones has been also observed, for instance in the series originated in Costa Rica, which apparently have a high level of genetic redundancy (Zhang *et al.* 2009b).

Significant efforts have been made to physically improve the CATIE's Collection and to enrich it genetically, introducing clones primarily wild or resistant to exotic diseases each year (Phillips-Mora *et al*. 2006b). Studies have also been conducted to rationalize the collection by identifying duplicates and misidentification (Johnson *et al.* 2006; Zhang *et al.* 2009b). Studies on the genetic composition of the clones are still pending, especially for those series in which CATIE is the international referent. Knowing the genetic structure of these clones, as well as the contribution of the 10 genetic groups of cacao described by Motamayor *et al*. (2008) in their formation will facilitate the use of these clones in the breeding programs and the interpretation of the results previously obtained.

The objective of the present study was to analyze molecularly the parentage and genetic composition of 266 cacao genotypes belonging the series ARF, CC, PMCT and UF originated in Costa Rica and compare them with a set of reference clones representing the known genetic diversity of *Theobroma cacao*.

MATERIALS Y METHODS

Experimental material

The 266 clones included in this study (44 clones UF, 99 clones CC, 91 clones PMCT and 32 clones ARF) were obtained from the International Cacao Collection at CATIE, located in Turrialba, Costa Rica (604 m.a.s.l., 2645 mm annual precipitation, 22.5°C average temperature). Additionally, 228 international clones that represent the ten genetic groups described by Motamayor *et al.* (2008) were included as genetic referents. These referent clones were obtained from the international genebanks at CRU/UWI (Trinidad) and CATIE (Costa Rica), and from the following national collections: INIAP (Ecuador), ICT (Peru), and CEPLAC (Brazil) (Zhang *et al.* 2009a, Zhang *et al*. 2009b, Motilal *et al*. 2010; Ji *et al.* 2013; Cosme *et al.* 2016).

Two healthy young leaves were collected from each clone using the first tree in the genebank row when it was available. The leaves were cut in squares of 4 x 4 cm and dried in silica gel for 48 hours. Then, they were 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 Tissue Kit (Qiagen, Inc., Valencia, CA) according to manufacturer's instructions.

SNP markers and genotyping

Forty-eight SNP markers were selected for this study based on the level of polymorphism and their distribution across the ten chromosomes in cacao, from 1560 putative candidate SNPs based on cDNA sequences (Allegre *et al.* 2012; Argout *et al*. 2008). Genotyping was performed on the high-throughput Fluidigm EP1™ system, using manufacturer's instructions. 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.

Data analysis

The program GenAlEx 6.2 (Peakall and Smouse 2012) was used for data analysis. 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).

Accessions with different names that were fully matched at the genotyped SNP loci were declared duplicates or synonymous accessions and after that, assignment test was applied to infer population membership and admixed ancestry (hybrids or ancestral forms) of the 266 clones and the 228 samples in the 10 reference groups, using a model-based clustering method implemented in the software program STRUCTURE (Pritchard *et al.* 2000). The number of clusters (K-value) was set to 10, assuming that each of the 10 populations may have contributed to the clones. Ten independent runs were assessed for $K=10$. Q-value was used to present the ancestral contribution (membership) from each germplasm group as follows: accessions possessing ≥25% membership were considered as receiving a significant ancestry contribution from that cluster; accessions possessing ≥75% membership were considered to be a member of that cluster and 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 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 sampled series, based on the result of assignment test. It was performed Principal Coordinates Analysis (PCoA), based on the pairwise distance matrix by using GenAlEx 6.2 (Peakall and Smouse 2012).

RESULTS

Descriptive statistics of genetic diversity

All the 48 SNP markers were used in the data analysis based on the criteria of informativeness about the level of polymorphism, whit an information index mean of 0.617. The inbreeding coefficient was positive for most of the loci, only two loci (TcSNP1038 and TcSNP1484) exhibited negative, the average was 0.295. The observed heterozygosity (H_o) ranged from 0.151 for TcSNP878 to 0.531 for TcSNP1484 with a mean of 0.301; the expected heterozygosity (H_e) varied from 0.221 for TcSNP1075 to 0.500 for TsSPN872 with a mean of 0.428 (Table 1).

Parentage analysis

Amelonado was the predominant group in the four series, with a presence that varied from 55.6% in the CC clones to 36.7% in the ARF. The Criollo and Nacional groups contributed with 16.2 and 14.8% respectively. The other genetic groups had a contribution lower than 7%, which was low for the groups Contamana, Iquitos and Marañon and minimum for the groups Guiana, Curaray, Nanay and Purús (Figure 1).

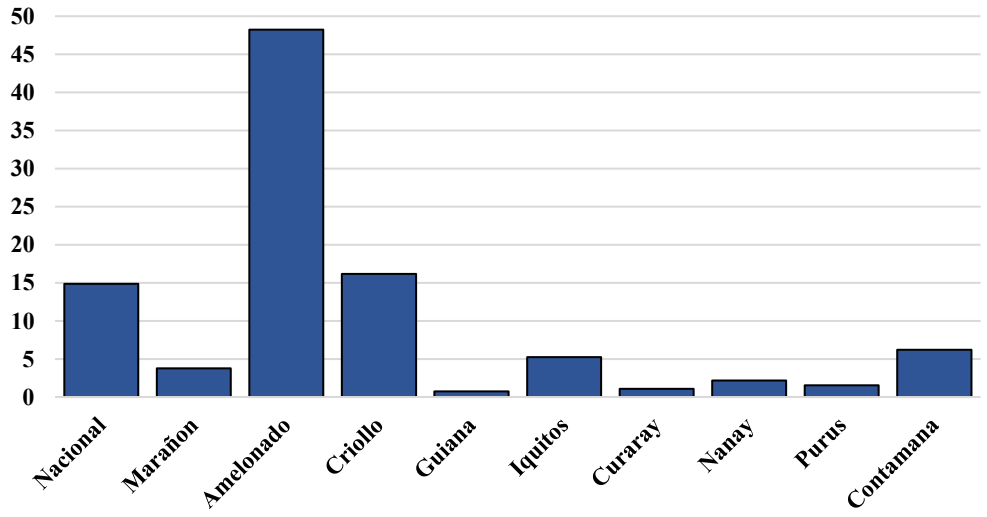


Figure 1. Ancestral contribution of ten genetic groups in the composition of the cacao series from Costa Rica preserved in the International Cacao Collection at CATIE.

The UF clones have a predominant ancestry from the Amelonado (48.7%), Nacional (22.8%) and Criollo (21.3%) groups, while the other groups contributed less than 2.0% in their composition (Table 2). The CC clones have a similar composition, with a predominant presence of the Amelonado group (55.6%), followed by the Criollo (12.4%), Nacional (10.6%) and Contamana (10.5%) groups. The other groups participate in this series with percentages lower than 4%. In the PMCT and ARF series, the Amelonado group also predominated (41.9 and 36.7%, respectively) but there is a higher presence of other genetic groups, particularly Iquitos (14.6%), Criollo (13%), Contamana (8.2%), Nacional (7.2 %) and Nanay (6%) in the series PMCT and Marañon (17.4%), Contamana (10.6%), Iquitos (8.6%), Criollo (8.4%), and Purus (7.8%) in the ARF series. The absence of Nacional in this last group is remarkable. In the PMCT and ARF series the presence of the remaining groups was less than 4.5%. The groups less represented in the four series of clones studied were Guiana and Curaray (Table 2).

Table 1. Summary statistics for the 266 Costa Rican clones and 228 international clones from the International Cacao Collection at CATIE assessed with 48 SNP markers.

SNP Locus	Information Index	Observed Heterozygosity	Expected Heterozygosity	Inbreeding Coefficient
TcSNP25	0.431	0.176	0.262	0.329
TcSNP32	0.686	0.318	0.493	0.355
TcSNP139	0.654	0.372	0.462	0.195
TcSNP144	0.681	0.407	0.488	0.165
TcSNP150	0.553	0.306	0.367	0.165
TcSNP151	0.690	0.330	0.497	0.336
TcSNP193	0.686	0.356	0.493	0.277
TcSNP226	0.676	0.301	0.483	0.376
TcSNP230	0.669	0.374	0.476	0.215
TcSNP242	0.627	0.359	0.435	0.175
TcSNP309	0.554	0.335	0.367	0.088
TcSNP372	0.619	0.245	0.428	0.427
TcSNP429	0.584	0.238	0.395	0.396
TcSNP469	0.685	0.302	0.492	0.386
TcSNP529	0.672	0.331	0.479	0.309
TcSNP534	0.630	0.350	0.439	0.201
TcSNP560	0.606	0.325	0.415	0.218
TcSNP577	0.653	0.241	0.461	0.477
TcSNP591	0.666	0.335	0.473	0.292
TcSNP619	0.689	0.290	0.496	0.416
TcSNP645	0.545	0.241	0.360	0.330
TcSNP723	0.512	0.315	0.330	0.046
TcSNP750	0.471	0.185	0.295	0.373
TcSNP836	0.632	0.200	0.440	0.545
TcSNP852	0.664	0.260	0.472	0.448
TcSNP872	0.693	0.275	0.500	0.449
TcSNP878	0.523	0.151	0.340	0.554
TcSNP886	0.561	0.344	0.374	0.079
TcSNP891	0.680	0.249	0.487	0.489
TcSNP917	0.665	0.298	0.472	0.368
TcSNP929	0.627	0.234	0.436	0.464
TcSNP953	0.502	0.185	0.321	0.423
TcSNP994	0.677	0.349	0.484	0.278
TcSNP998	0.614	0.325	0.423	0.233
TcSNP1038	0.649	0.526	0.456	-0.152
TcSNP1060	0.487	0.257	0.309	0.166
TcSNP1062	0.622	0.256	0.431	0.404
TcSNP1075	0.380	0.164	0.221	0.261
TcSNP1144	0.571	0.265	0.383	0.310
TcSNP1165	0.679	0.321	0.486	0.338
TcSNP1253	0.690	0.363	0.497	0.269
TcSNP1270	0.621	0.251	0.430	0.417
TcSNP1350	0.690	0.313	0.496	0.369
TcSNP1414	0.598	0.280	0.408	0.315
TcSNP1442	0.642	0.396	0.450	0.119
TcSNP1458	0.670	0.281	0.477	0.411
TcSNP1484	0.691	0.531	0.498	-0.066
TcSNP1520	0.551	0.316	0.365	0.133
Mean	0.617	0.301	0.428	0.295
SE	0.011	0.011	0.010	0.022

Table 2. Percentage of contribution of ten genetic group in the composition of the clones ARF, CC, PMCT and UF preserved in the CATIE´s genebank.

	Nacional	Marañon	Amelonado	Criollo	Guiana	Iquitos	Curaray	Nanay	Purus	Contamana
UF	22.8	1.3	48.7	21.3	0.6	1.6	0.4	0.9	0.4	2.0
CC	10.6	3.6	55.6	12.4	0.7	3.0	1.6	0.8	1.2	10.5
PMCT	7.2	4.3	41.9	13.0	1.1	14.6	1.7	6.0	2.0	8.2
ARF	4.4	17.4	36.7	8.4	1.0	8.6	1.7	3.5	7.8	10.6

Assignment of membership from germplasm group

The 42 clones that showed a percentage of contribution of a certain genetic group (≥75%), were assigned to that particular group. Out of the 44 UF studied clones, five were assigned to the Amelonado group (UF-10B, 122, 602, 701, 706); 2 to the Nacional group (UF-20, 712), and one (UF-613) to the Iquitos group. In the UF series, a high proportion of hybrids "Amelonado x Criollo" and "Amelonado x Marañon" were found, as well as other less frequent (Table 3).

Out of the group of 99 CC analyzed clones, 19 were assigned to the Amelonado group, with a composition that varied between 99.3% in CC-267 and 76.1% in CC-35. In fact, the CC-267 clone (aka Matina 1-6) was sequenced by Motamayor *et al*. (2013) due to its largely homozygous condition. In the CC group a high proportion of hybrids was found between the groups "Amelonado x Nacional", "Criollo x Contamana", as well as other combinations (Table 3).

Out of the 32 ARF studied clones, only ARF-4 was assigned to the Marañon group (97.8%) and ARF-30 was assigned to the Amelonado group (77.2%). The other clones correspond to hybrids between different groups, highlighting the combinations between "Amelonado x Marañon", "Criollo x Marañon", "Amelonado x Purus", "Amelonado x Contamana", etc. (Table 3).

Out of the 91 PMCT sampled clones, 10 were assigned to the Amelonado group, with a variation between 99.3% (PMCT-10) and 75.1% (PMCT-67). The clones PMCT-11 and PMCT-29 were assigned to the Criollo group with more than 99% in its composition, while the PMCT-33 contains 84.5% of composition of the Iquitos group. A high proportion of hybrids was found between "Amelonado x Criollo" and "Amelonado x Iquitos" and in less quantity of "Amelonado x Contamana" and "Amelonado x Nacional", among other combinations.

Table 3. Clones of the Costa Rican series assigned as a member to different genetic groups (Ancestral contribution ≥75%).

	Nacional	Marañon	Amelonado	Criollo	Iquitos
UF	20, 712		10B, 122, 602, 701, 706		613
CC			18, 33, 35, 41, 42, 47, 49, 83, 106, 107, 121, 132, 144, 152, 158, 169, 173, 256, 267		
PMCT			10, 14, 18, 22, 25, 26, 27, 28, 51, 67	11, 29	33
ARF		4	30		

Genetic relationship among the Costa Rican clones and reference germplasm groups

The result of PCoA was presented in Figure 2a and Figure 2b. The plane of the first three main PCO axes accounted for 26.45%, 14.27% and 9.76% of total variation, respectively. The Costa Rican clones were found highly diverse and most of the clones were inter-population hybrids, most of which did not fall into any reference group. The four groups of Costa Rican clones differed in terms of their position and range of dispersion. For example, the CC clones showed a broader range of distribution than the other groups in the plane. Also the clones PMCT and ARF are grouped mainly in one side and the UF clones in another side, indicating their difference in parentage (Figure 2a, 2b).

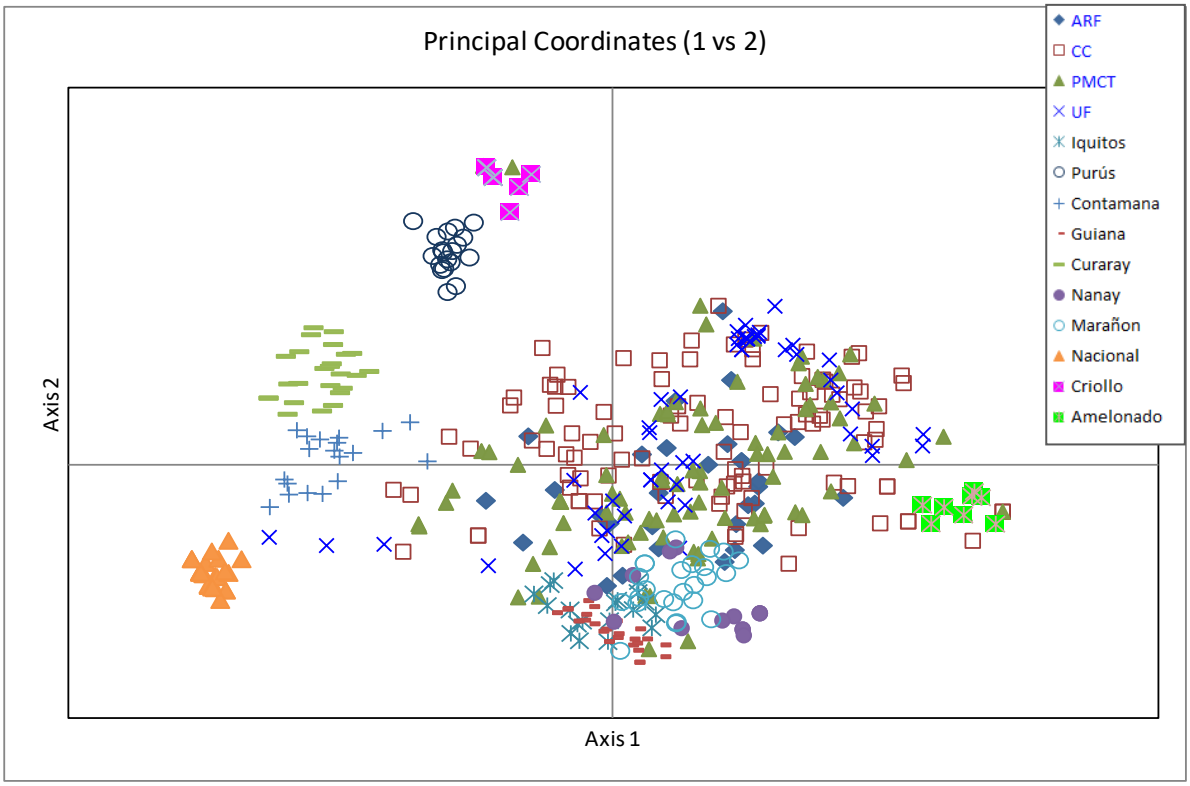
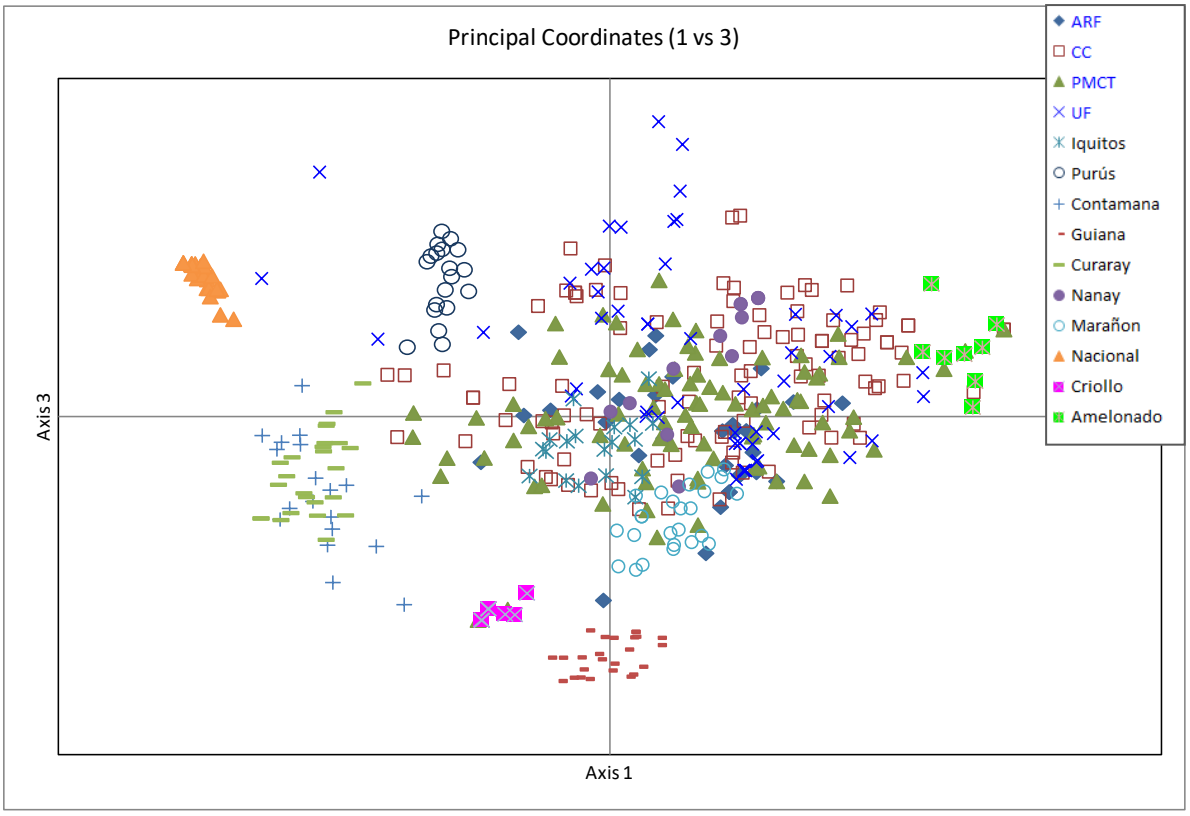


Figure 2a.



Figures 2b. Principal Coordinate Analysis (PCoA) for the 266 cacao accessions from CATIE, Costa Rica and 150 reference trees representing 10 known genetic groups. The plane of the first three main PCO axes accounted for 50.5% of total variation. First axis26.5% of total information, the second 14.3% and the third 9.7%.

DISCUSSION

The three major germplasm suppliers supporting the "Global Strategy for the Conservation and Use of Cacao Genetic Resources" are the two international collections at CRU/UWI (Trinidad and Tobago) and CATIE (Costa Rica), and the International Cocoa Quarantine Centre at the University of Reading (United Kingdom) (CacaoNet 2012). CATIE´s Genebank is the international referent and custodian of four series of clones developed in Costa Rica from 1928 to 1992 (UF, CC, PMCT, and ARF). Clones belonging to these series have an historical and current international relevance as planting materials; parents in breeding programs, and a target clone (CC-267 = Matina 1-6) for the sequencing of the *Theobroma cacao* genome (Motamayor *et al*. 2013).

Important evidence related to the genetic composition and ancestry of the mentioned series was obtained in the present research. A significant level of consanguinity among some clones and a predominant participation of the genetic groups Amelonado, Criollo, and Nacional was found. This information is consistent with the documented origin of the CC, ARF and PMCT clones, and provide novel information for the UF clones for which no available documentation exists indicating their precise genetic origin. Nearly pure clones belonging to the groups Marañon (ARF-4); Nacional (UF-20, UF-712); Criollo (Criollo-11, Criollo-29); Iquitos (UF-613, PMCT-33), and particularly Amelonado (35 genotypes) were identified within the series.

The predominance of the Amelonado group in the Costa Rican series is explained by the putative high participation of Matina in their conformation. This Amelonado variety is not native but introduced to Costa Rica in an uncertain date. It was cultivated in this country since the 17th century, particularly on the Atlantic coast (Mora 1958, MacLeod 1996, Solórzano 2012). The Matina variety predominated in Costa Rica until the arrival of moniliasis in 1978, when most cacao plantations were abandoned due to their high susceptibility to the disease (Phillips-Mora *et al.* 2006a). Its good production, self-compatibility, strong cocoa flavor and acceptance by farmers made the Matina clone a recurrent progenitor in the first series of cacao developed in Costa Rica by the United Fruit Co. and IICA. The variety has remained in its original form over the years and is considered highly homozygous and genetically close to the African Amelonado cacao (Bartley 2005; Motamayor *et al*. 2013). Some clones from the series ARF and particularly PMCT also have a high content of Amelonado ancestry derived probably from the Matina variety.

The second most important group is the Criollo, which was the material introduced and domesticated in Central America a long time ago. Consequently, there is an important presence of these genotypes and their descendants throughout the region (Mora 1958; Bartley 2005). The distinctive traits of the Criollo genotypes such as their good quality, red pod and white seeds color, together with the possibility of creating Trinitarian hybrids through the cross against Matina and other Forastero clones, were some of the possible reasons that justified the utilization of these clones in different crosses. The collection of acriollados genotypes made by CATIE in Central America between 1989 and 1992 is also responsible for the presence of Criollo background in some accessions of the PMCT and ARF series.

Nacional is the third most important genetic background in the Costa Rican series, with a presence that decrease throughout time, being maximum in the UF clones (22.8%) and very low in the ARF clones (4.4%). This presence is due to the early introduction of Nacional genotypes into Costa Rica from Ecuador, and their utilization as parents in the first breeding efforts performed by the United Fruit Co. in 1928 (Bowman, cited by Morera *et al.* 1991).

The Marañon group contributed 17.4% to the ARF series, while Contamana group had a significant presence (8.2-10.6%) in the PMCT, ARF and CC series due to the utilization of SCA-6 and SCA-12 in some crosses. SCA clones were used as parents because of their resistance to witches broom and the hybrid vigor they develop when crossing against genetically distant genotypes. The presence of 8.6% of the Iquitos group in the ARF series is derived from the participation of the IMC-67 clone in some of the crosses. The Guiana, Curaray and Purus genetic groups showed a low participation ($\leq$ 2%) in the composition of the analyzed series, except in the ARF group, where 7.8% of their composition came from the Purus group.

The information generated in this study will be essential to guide the effective conservation, exchange and use of the involved clones, and to understand and interpret the results obtained in previous breeding efforts.

Acknowledgements

The authors would like to thank the USDA-ARS for the support in the laboratory work, Lin Zhou for her help in the preparation of DNA samples and genotyping. To the World Cocoa Foundation, to USDA-FAS

and to the Cocoa Borlaug Fellowship Program for the financial and logistical support during the realization of this study.

REFERENCES

Allegre, M; Argout, X; Boccara, M; Fouet, O; Roguet, Y; Bérard, A; Thévenin, JM; Chauveau, A; Rivallan, R; Clement, D; Courtois, B; Gramacho, K; Boland-Augé, A; Tahi, M; Umaharan, P; Brunel, D; Lanaud, C. 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. doi: 10.1093/dnares/dsr039.

Argout, X; Fouet, O; Wincker, P; Gramacho, K; Legavre, T; Sabau, X; Risterucci, AM; Da Silva, C; Cascardo, J; Allegre, M; Kuhn, D; Verica, J; Courtois, B; Loor, G; Babin, R; Sounigo, O; Ducamp, M; Guiltinan, MJ; Ruiz, M; Alemanno, L; Machado, R; Phillips, W; Schnell, R; Gilmour, M; Rosenquist, E; Butler, D; Maximova, S; Lanaud, C. 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. doi: 10.1186/1471-2164-9-512.

Bartley, BGD. 2005. The genetic diversity of cacao and its utilization. CAB International, CABI Publishing, Wallingford, Oxfordshire.

CacaoNet. 2012. A Global Strategy for the Conservation and Use of Cacao Genetic Resources, as the Foundation for a Sustainable Cocoa Economy (B. Laliberté, compiler). Bioversity International, Montpellier, France.

CATIE, 1990. Libro de introducción de germoplasma de cacao (invernadero) Turrialba, CATIE.

Cosme, S; Cuevas, HE; Zhang, D; Oleksyk, TK; Irish, BM. 2016 Genetic diversity of naturalized cacao (*Theobroma cacao* L.) in Puerto Rico. Tree Genetics and Genomes 12:88 doi:10.1007/s11295-016-1045-4

Esquivel, O; Soria J. 1967. Algunos datos sobre la variabilidad de algunos componentes del rendimiento en poblaciones de híbridos interclonales de cacao. Cacao 12(4):1-8.

Fritz, PJ; Osei, J; Morera, J; Rodríguez, H; Walker, DR; Lainez, J; Lastra, R. 1995. Backcross populations for genetic analysis of *Theobroma cacao*. Catongo X Tree 33 (Catongo X Pound-12). Turrialba 45(1-2):49-56.

García, LF. 2007. Guía de campo. Identificación de cultivares de cacao. Universidad Nacional Agraria de La Selva. Tingo María, Perú. 32 p.

Ji, K; Zhang, D; Motilal, LA; Boccara, M; Lachenaud, P; Meinhardt, LW. 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. Genetic Resour Crop Evol. doi: 10.1007/s10722-012-9847-1

Johnson, ES; Phillips-Mora, W; Leal, AA; Meneses, A; Bekele, FL; Zhang, D; Schnell, RJ. 2006. Field Guide to the UF clones of Costa Rica. *In* Proceedings of the 15th COPAL International Cocoa Research Conference, San Jose, Costa Rica, Oct. 9-14, 2006. Vol. I. p.641-646

MacLeod, P. 1996. Auge y estancamiento de la producción de cacao en Costa Rica 1660-95. Anuario de Estudios Centroamericanos, Universidad de Costa Rica. 22(1):83-107.

Mora, J. 1958. Notas sobre el posible origen y la variabilidad del cacao cultivado en la América Tropical. Turrialba (IICA) 8(1):34-43

Morera, J; Mora, A. 1991. Comparación de 56 cruces interclonales de cacao en Pococí, Costa Rica. Turrialba 41(4):578-582.

Morera, J; Mora, A; Paredes, A. 1991. Caracterización de una Población de Cacao "Nacional" en el CATIE, Costa Rica. Turrialba 41(4):583-588.

Motamayor, JC; Lachenaud, P; da Silva e Mota, JW, Loor, R; Kuhn, DN; Brown, JS; Schnell, RJ. 2008. Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L). PLoS ONE 3(10):e3311. doi:10.1371/journal.pone.0003311.

Motamayor, JC; Mockaitis, K; Schmutz, J; Haiminen, N.; Livingstone D.; Cornejo, O; Findley, SD; Zheng, P; Utro, P; Royaert, S; Saski, C; Jenkins, J; Podicheti, R; Zhao, M; Scheffler, BE; Stack, JC; Feltus, FA; Mustiga, GM; Amores, F; Phillips, W; Marelli, JP; May, GD; Shapiro, H; Ma, J; Bustamante, CD; Schnell, RJ; Main, D; Gilbert, D; Parida, L; Kuhn, DN. 2013. The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. Genome Biology 14:r5. doi:10.1186/gb-2013-14-6-r53

Motilal, L. 2004. The potential of cacao microsatellites amplification across diverse plant taxa. *In* Thangadurai, D; Pullaiah, T; Balatti, PA (eds). Genetic Resources and Biotechnology, Vol 2, Regency Publications, New Delhi, India. pp.24-49.

Motilal, L; Butler, D. 2003. Verification of identities in global cacao germplasm collections. Genet Resour Crop Evol 50:799-807.

Motilal, L; Butler, D. 2003. Verification of identities in global cacao germplasm collections. Genet Resour Crop Evol 50:799-807

Motilal, L; Zhang, D; Umaharan, P; Mischke, S; Mooledhar, V; Meinhardt, LW. 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: Charact Util 8:106-115.

Peakall, R; Smouse, PE. 2006. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6:288-295.

Perea, A; Martínez, N; Aranzazu, F; Cadena, T. 2013. Características de calidad del cacao en Colombia. Catálogo de 26 cultivares. Universidad Industrial de Santander. Bucaramanga, Colombia. 107 p.

Phillips-Mora., W. 1986. Evaluación de la resistencia de cultivares de cacao (*Theobroma cacao* L.) a *Moniliophthora roreri* Cif. Par. Tesis Mag. Sc. CATIE-UCR. Turrialba, CR.

Phillips-Mora, W; Galindo, JJ. 1991a. Effect of inoculation of three isolates of *Phytophthora palmivora* on cacao cultivars. Phytopathology 81(6):698.

Phillips-Mora, W; Galindo, JJ. 1991b. Effect of *Phytophthora palmivora* on three cacao cultivars and their crosses. Phytopathology 81(6):698.

Phillips-Mora, W; Mora, A; Johnson, E; Astorga, C. 2006b. Recent efforts to improve the genetic and physical conditions of the international cacao collection at CATIE. *In* Proceedings of the 15th COPAL International Cocoa Research Conference. 2006 San Jose, Costa Rica Oct. 9-14, 2006. Vol. I p.611-623.

Phillips-Mora, W; Ortiz, CF; Aime, MC. 2006a. Fifty years of frosty pod rot in Central America: Chronology of its spread and impact from Panama to Mexico. *In* Proceedings of the 15th COPAL International Cocoa Research Conference. 2006 San Jose, Costa Rica Oct. 9-14, 2006. Vol. I. p.1039-1047.

Phillips-Mora, W; Arciniegas-Leal, A; Mata-Quirós, A; Motamayor-Arias, JC. 2012. Catálogo de clones de cacao seleccionados por el CATIE para siembras comerciales. Turrialba, CR. Centro Agronómico Tropical de Investigación y Enseñanza. 68 p. (Serie Técnica. Manual Técnico N°105).

Pritchard, JK; Stephens, M; Donelly, P. 2000. Inference of population structure from multilocus genotype data. Genetics 155:945-959.

Solórzano, JC. 2012. La producción de cacao en Matina y la rebelión de los indígenas Urinamas (Costa Rica 1650-1690). Revista Estudios. 25:111-123.

Soria, J; Esquivel, O. 1967. Estudio preliminar sobre el período mínimo y confiable de producción en cacao para su uso en experimentos de evaluación de cultivares. Cacao 12(4):9-14.

Sounigo, O; Risterucci, AM; Clement, D; Fouet, O; Lanaud, C. 2006. Identification of off-types of clones used in the International Clone Trial using DNA analyses. *In* Eskes, AB; Efron, Y (eds). Global approaches to cocoa germplasm utilization and conservation. Final Report of the CFC/ICCO/IPGRI project on Cocoa Germplasm Utilization and Conservation: a Global Approach (1998-2004), CFC, Amsterdam, The Netherlands/ICCO, London, UK/IPGRI, Rome, Italy. pp.82-86

Takrama, JF; Cervantes-Martinez, C; Phillips-Mora, W; Brown, JS; Motamayor, JC; Schnell, RJ. 2005. Determination of off-types in a cocoa breeding programme using microsatellites. INGENIC Newsletter 10:2-8.

Turnbull, CJ; Hadley, P. International Cocoa Germplasm Database (ICGD). [Online Database]. CRA Ltd./ICE Futures Europe/University of Reading, UK. Available: http://www.icgd.reading.ac.uk (5th October, 2017)

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

Zhang, D; Boccara, M; Motilal, L; Mischke, S; Johnson, ES; Butler, DR; Bailey, B; Meinhardt, L. 2009a. Molecular characterization of an earliest cacao (*Theobroma cacao* L.) collection from Upper Amazon using microsatellite DNA markers. Tree Genetics & Genomes 5:595-607. doi: 10.1007/s11295-009-0212-2

Zhang, D; Mischke, S; Johnson, ES; Phillips-Mora, W; Meinhardt, L. 2009b. Molecular characterization of an international cacao collection using microsatellite markers. Tree Genetics & Genomes 5:1-10. doi: 10.1007/s11295-008-0163-z