Bioagro 32(2): 75-86. 2020

GENETIC DIVERSITY OF ECUADORIAN COCOA FROM THE GERMPLASM BANK OF TENGUEL-GUAYAS ECUADOR BASED IN SNP'S

Mercedes Susana Carranza[1], Yeny Patricia Zapata[2], Gerardo Gallego[2], José Nieto Rodríguez[1], Jaime Morante Carriel[1], Nicolás Cruz Rosero[1], Sila Medina Jara[1] and Jaime Eduardo Muñoz[3]

ABSTRACT

The genetic diversity and structure of 80 cocoa introductions of *Theobroma cacao* L. Nacional variety existing in the farm "La Buseta", Tenguel-Guayas, Ecuador, was studied. The characterization included 19 controls and the use of 96 markers of simple nucleotide polymorphisms (SNPs), in the Fluidigm EP1 system platform. The SNPs showed to be informative, with a mean content of polymorphic information of 0.289, and an average observed heterozygosity of 0.479 and expected heterozygosity of 0.378. The genetic profiles of the introductions were obtained from which 16 groups of identical introductions were identified. The identity probability analysis, including siblings (PIsib), concluded that the SNPs were enough to differentiate the introductions. The average of heterozygosity of the samples was 0.432 being higher in the group F (0.646) and lower in the group P (0.253). The Criollo and Contamana genotyping controls showed greater genetic distance than the Trinitario, Forastero and Ecuadorian Criollo controls. The maximum genetic distance between the introductions was 0.167. The results are consistent with the history of cultivation of the Nacional cacao, which over time received genetic contributions from other varieties. The present study constitutes a significant advance in the knowledge of the genetic diversity of Ecuador's Nacional cacao.
Additional keywords: Genetic diversity, germplasm, molecular markers, SNPs

RESUMEN

Diversidad genética del cacao Nacional ecuatoriano del banco de germoplasma de Tenguel-Guayas, Ecuador, con base en SNP'S

Se estudió la diversidad genética y estructura de 80 introducciones de *Theobroma cacao* L. 'Nacional' existente en la finca "La Buseta", Tenguel-Guayas, Ecuador. La caracterización incluyó 19 controles y 96 marcadores de polimorfismo de nucleótidos simples (SNP), en la plataforma del sistema Fluidigm EP1. Los SNP's mostraron información polimórfica de 0,289 y 0,395; con heterocigosidad observada de 0,479 y esperada de 0,378. Se obtuvieron los perfiles genéticos de las introducciones, a partir de los cuales se identificaron 16 grupos de introducciones idénticas. El análisis de probabilidad de identidad, incluyendo hermanos (PIsib), concluyó que los SNP's fueron suficientes para diferenciar las introducciones. El promedio de heterocigosidad de las muestras fue de 0,432 siendo mayor en el grupo F (0,646) y menor en el grupo P (0,253). Los controles de genotipeo Criollo y Contamana, mostraron mayor distancia genética que los controles Trinitario, Forastero y Criollo ecuatoriano. La distancia genética máxima entre las introducciones fue de 0,167. Los resultados son consistentes con la historia del cultivo del cacao Nacional, el cual a lo largo del tiempo recibió contribuciones genéticas de otras variedades. El presente estudio constituye un avance significativo en el conocimiento de la diversidad genética del cacao Nacional de Ecuador.
Palabras clave adicionales: Diversidad genética, germoplasma, marcadores moleculares, SNPs

INTRODUCTION

Theobroma cacao L. is a diploid species (2n 2x 20), native in the rainforests in the Amazon region of South America (Motamayor et al., 2002). This region has been considered the geographic origin center of the species because it has the greatest genetic diversity of cocoa (Thomas et al., 2012). Although traditionally three morpho-geographical groups called Criollo, Forastero and Trinitario

Received: June 24, 2019 Accepted: March 9, 2020
[1] Quevedo State Technical Univ. (UTEQ). Ecuador. e-mail mcarranza@uteq.edu.ec (autor de correspondencia)
[2] International Center for Tropical Agriculture (CIAT). Cali-Palmira, Colombia. e-mail g.gallego@cgiar.org
[3] National University of Colombia (UNAL), Faculty of Agricultural Sciences, Molecular Biology Laboratory. Palmira, Valle del Cauca, Colombia. e-mail jemunozf@unal.edu.co

have been recognized, molecular studies by Motamayor et al. (2008), have differentiated ten major groups, which indicate the genetic diversity of cocoa in South America. The cacao presents a high genetic variability and Criollo group is clearly differentiated from the rest of the genetic (Cornejo et al., 2018).

The Nacional type variety according to CacaoNet (2012) is one of the oldest commercially cultivated populations found in the coastal regions of Ecuador, west of the Andes. Due to its morphology and allelic diversity, the upper Amazon basin is the most accepted origin of this variety (Pound, 1945; Zhang et al., 2012).

The Nacional Fine Aroma cocoa is recognized worldwide for its fruity and floral fragrances. These properties give an added value which makes it a cocoa recognized by the international market especially in the confectionery industry (Asociación Nacional de Exportadores de Cacao e Industrializados del Ecuador, 2010).

The Cocoa Center of Aroma Tenguel (CCAT), is a Nacional germplasm bank of cocoa diversity considered the second most important in the country of which there are records of agronomic behavior of eight years. This makes it possible to select those cocoa introductions that are outstanding in terms of productivity and disease tolerance (Amores et al., 2009; Sánchez et al., 2015).

Improvement of cocoa by conventional genetics may take several years and errors at the time of selection when it involves the analysis of phenotypic traits influenced by environmental conditions, dominance of some genes, multigenic and quantitative inheritance (Wilde et al., 1992; Arvelo et al., 2017).

However, molecular markers like microsatellites and single sequence markers (SNPs) make it possible to evaluate genetic diversity, being a useful tool, especially if they are associated with traits of agronomic or economic interest (Danial and Rojas, 2007). The obtaining of (SNPs) has been facilitated by sequencing of complete Criollo B97-61/B2 and Matina 1-6 cocoa genome, as well as recent publications of large collections of ESTs (Expressed Sequence Tag) (Argout et al., 2008; Argout et al. 2011; Lanaud et al., 2006; Motamayor et al., 2013). SNP markers are changes in a single nucleotide and are the most common form of polymorphism in plant genomes. Their use allows for automated and standardized testing through laboratories and they have diverse applications such as: identification of introductions within collections, follow-up of improvement programs through analysis, characterization of parents and derived lines, among others (Ji et al., 2012).

Different studies carried out with cocoa from Honduras and Nicaragua (Ji et al., 2012), Brazil (DuVal et al., 2017), Colombia (Osorio et al., 2017) and Costa Rica (Mata et al., 2018), have used SNP markers, considering even a small number of them, that have shown high efficiency in the characterization of genotypes. Based on the above, the main aim of this study was to determine genetic diversity and genotypically characterize 80 introductions and 19 controls of introductions of Nacional cocoa of Ecuador, as a complement to morphoagronomic studies of this variety, for selection purposes for genetic improvement programs.

MATERIALS AND METHODS

Plant source. The selection and genotyping of the 80 introductions from the Nacional cocoa of Ecuador was made from 2222 introductions of the CCAT, farm "La Buseta", located in the parish Tenguel, canton Guayaquil, Guayas province, at 3º00' S and 79º 47' W. The introductions were established in that germplasm bank during the decade of the 1940s, coming from the coastal region of Ecuador, considering characters of productivity and health conditions (Table 1). In addition, 19 controls of the farm "La Represa", property of the Universidad Técnica Estatal de Quevedo, (01º 03' S and 79º 25' W), Fayta precinct, province Los Ríos (Table 2), for a total of 99 individuals. Three leaves were collected at an intermediate stage of maturity and stored in hermetically sealed bags containing 30 g of silica gel. Such samples were transferred to the Molecular Genetics and Tissue Culture laboratory of the International Center for Tropical Agriculture (CIAT). In addition, four cocoa controls available at CIAT (C-Criollo, C-Trinitario 1, C-Trinitario 2, C-Contamana) were included in the evaluation.

Table 1. Nacional genotypes of cocoa from Ecuador evaluated at CCAT

N°	Genotype	N°	Genotype	N°	Genotype	N°	Genotype
1	L10H28	21	L20H50	41	L29H7	61	L33H49
2	L11H25	22	L20H53	42	L29H47	62	L33H56
3	L12H8	23	L21H6	43	L29H48	63	L33H65
4	L12H27	24	L21H38	44	L30H1	64	L34H67
5	L12H28	25	L21H56	45	L30H3	65	L40H49
6	L12H29	26	L21H57	46	L30H25	66	L42H80
7	L12H30	27	L23H36	47	L30H45	67	L44H88
8	L13H11	28	L24H12	48	L30H46	68	L45H11
9	L13H37	29	L24H14	49	L30H9	69	L46H66
10	L15H31	30	L24H19	50	L32H54	70	L46H68
11	L15H32	31	L25H45	51	L31H66	71	L46H70
1	L15H34	32	L25H59	52	L32H48	72	L46H75
13	L16H48	33	L25H60	53	L32H65	73	L48H23
14	L17H25	34	L25H64	54	L32H68	74	L48H92
15	L17H30	35	L26H64	55	L32H72	75	L49H4
16	L17H36	36	L27H19	56	L33H8	76	L49H86
17	L17H38	37	L27H48	57	L33H25	77	L49H98
18	L18H36	38	L28H22	58	L33H26	78	L51H59
19	L18H58	39	L28H48	59	L33H45	79	L52H98
20	L20H49	40	L29H4	60	L33H47	80	L53H4

Table 2. Introductions of cocoa considered as controls or wild type from the farm "La Represa" UTEQ- Ecuador

N°	Introduction /variety	Code	N°	Introduction /variety	Code
1	Trinitario(L48H89)	(1-06)	11	Forastero IMC-67	(1-1)
2	Trinitario (L18H53)	(14-09)	12	Forastero (103)	(1-4)
3	Trinitario (ISS 95-IMC67-canelo)	(24-2)	13	Forastero (109)	(1-15)
4	Trinitario	(28-4)	14	Forastero (L11H19)	(1-21)
5	Trinitario	(29-4)	15	Onzole-Criollo	8PL4
6	CCN51	(2-2)	16	Onzole-Criollo	88PL1
7	CCN51	(2-3)	17	Onzole-Criollo	78PL1
8	CCN51	(2-4)	18	Onzole-Criollo	79PL2
9	CCN51	(3-4)	19	Onzole-Criollo	5PL1
10	CCN51	(3-5)			

DNA extraction technique. DNA extraction was carried out from cocoa foliar tissue following the method of Michiels et al. (2003) modified by Zapata (2016) in the laboratory of CIAT. The isolated DNA was subjected to quality control following the practical guide for genotyping of SNPs using Fluidigm's EP1 and SNPtype Assays F_03 (Corporation Fluidigm, 2018). To this end, an electrophoresis in agarose gel (1 % agarose, 0.5 % TBE, stained with SYBR Safe) and quantification by spectrophotometry (Synergy-H1m) were carried out while all the 99 samples were diluted to 60 $ng \cdot \mu L^{-1}$ for subsequent processing.

SNPs markers and genotyping. Genotyping was carried out on the Fluidigm EP1 System platform and samples were processed in accordance with the aforementioned technical manual using the allele-specific detection methodology, SNPtype developed by Fluidigm. A PCR pre-amplification,

called PCR-mold, of the genomic regions containing the SNPs was performed. Both the PCR-molds and the initiators for each SNP marker were transferred to the solid support called the Integrated Fluidics System (IFC). Once loaded in the central part of the IFC, an allele-specific PCR was performed, the result of which was the determination of fluorescent signals captured as images, which were subsequently analyzed using the Fluidigm SNP Genotyping Analysis Software 4.1.3 (Corporation Fluidigm, 2018).

Cocoa introductions were genotyped with a total of 96 SNPs (Table 3). A total of 48 were developed by researchers from the United States Department of Agriculture (USDA) and Fluidigm Corp. (San Francisco, California, USA) (Illic et al., 2012; Fang et al., 2014), and the other 48 were obtained from the TropGene database of the CIRAD Research Center (Ruiz et al., 2015).

Data analysis. Analysis of the fluorescence's obtained was carried out with the Fluidigm SNP Genotyping Analysis Software version 4.1.3., to group as homozygotes or heterozygotes. In order to evaluate the efficacy of the SNP markers used, descriptive statistics of observed Heterozygosity (Ho), expected heterozygosity (He) and polymorphic information content (PIC) were obtained using Power Marker V3.25 25 (Liu and Muse, 2005).

GenAlex V6.502 (Peakall and Smouse, 2012), was used to calculate the probability of identity (IP), which refers to the probability that two randomly chosen individuals have the same genotype and the probability of identity by including relatives (PIsib), which is the probability that two randomly chosen related individuals have the same genotype. These probabilities assess the discriminatory ability from a specific combination of markers while their values are expected to be significant ($P \leq 0.0001$).

According to the above probability, genetic copies were identified with the Matching Multilocus Genotypes option of GenAlex, which compares all samples to identify those with the same genetic profile.

To observe the distribution of samples, the genetic distances paired in PowerMarker were calculated, comparing the genetic profiles in pairs and transforming the differences found in distance values (Beerli, 2005). These distances were obtained following different standards according to the nature of the data, as is the case of the standard of Rogers (1972) that does not assume evolutionary forces *a priori*, so it has been considered adequate for the evaluation of improving collections (Reif et al., 2005). This standard distributes samples in a range of 0 to 1, where 0 means that there are no genetic differences and 1 means that there is total differentiation between individuals. The numerical matrix resulting from this analysis was grouped in a dendrogram by means of the UPGMA algorithm, where the lengths of each branch indicate the genetic distance between the individuals. The resulting trees were edited in FigTree V1.4.2 (Rambaut, 2016).

The structure of the introductions was inferred with the grouping algorithm with a Bayesian model in the software Structure v 2.3.4 (Pritchard et al. 2000) to estimate the probable number of groups of introductions (K) with hierarchical information. The probabilities were calculated for K values from 1 to 5. The consistency of the results was evaluated using six repetitions from 100,000 interactions for the burn-in stage followed by 500,000 Monte Carlo Markov Chain (MCMC) simulation interactions for each K. The ideal value of K was calculated using the online software Structure Harvester v 6.94 (Earl and VonHoldt, 2012) observing the logarithm of the probability of the L(K) data and the Delta K estimates.

The analysis of molecular variance (AMOVA) and Wright's statistical parameters F (F IS, F IT and F ST) was performed using GenAlex V6.502 (Peakall and Smouse, 2012).

RESULTS

SNPs markers. Ninety five SNPs were polymorphic, having an average of polymorphic information content (PIC) = 0.289, while marker TcSNP353 was excluded due to its high number of missing data. The highest value for this indicator was PIC 0.375, shared by 20 markers, while the lowest 0.010 was observed in four markers. The (Ho) and (He) averages were 0.468 and 0.378, respectively (Table 4). The genotyping analysis with the Fluidigm platform made it possible to obtain the genetic profiles of 99 introductions with 95 SNPs for a total of 9405 data points (profiles not shown).

Table 3. Composition of SNPs arrangements for genotyping Nacional variety cocoa of Ecuador

No.	SNP ID (USDA)	Chromosome	Allele 1	Allele 2	No.	SNP ID (TropGene)	Chromosome	Allele 1	Allele 2
1	CL317Contig1	1	A	C	49	TcSNP865	1	T	C
2	CL527Contig1	1	T	C	50	TcSNP1058	1	A	C
3	CL77Contig2	1	A	C	51	TcSNP510	1	C	G
4	CL276Contig1	1	A	T	52	TcSNP418	1	A	G
5	CL3336Contig1	1	A	C	53	TcSNP316	2	A	T
6	CL1454Contig1	1	C	G	54	TcSNP1159	2	A	C
7	CL1125Contig2	2	A	G	55	TcSNP206	2	A	G
8	CL646Contig2	2	T	C	56	TcSNP798	2	A	T
9	CL1Contig69	2	T	C	57	TcSNP152	2	T	C
10	CL1002Contig1	2	T	C	58	TcSNP353	2	A	G
11	CL1Contig277	2	T	C	59	TcSNP437	2	A	T
12	CL4Contig14	3	C	G	60	TcSNP689	3	A	G
13	CL1312Contig1	3	C	G	61	TcSNP595	3	A	G
14	CL209Contig1	3	C	G	62	TcSNP928	3	T	C
15	CL132Contig1	3	A	G	63	TcSNP413	3	T	C
16	CL3696Contig1	4	A	T	64	TcSNP395	4	T	C
17	CL552Contig2	4	A	T	65	TcSNP344	4	T	C
18	CL359Contig1	4	C	G	66	TcSNP382	4	T	C
19	CL588Contig1	4	T	C	67	TcSNP1209	4	T	G
20	CL2987Contig1	4	A	T	68	TcSNP174	4	A	G
21	CL695Contig1	5	T	G	69	TcSNP475	5	C	G
22	CL1Contig128	5	C	G	70	TcSNP736	5	T	C
23	CL318Contig1	5	A	G	71	TcSNP1111	5	A	G
24	CL1086Contig1	5	A	G	72	TcSNP28	5	T	G
25	CL218Contig1	5	A	G	73	TcSNP1453	5	T	G
26	CL581Contig1	6	T	C	74	TcSNP602	6	T	C
27	CL456Contig1	6	T	C	75	TcSNP1212	6	T	C
28	CL745Contig1	6	T	C	76	TcSNP1390	6	T	C
29	CL171Contig2	6	T	C	77	TcSNP1383	7	A	G
30	CL192Contig2	6	T	C	78	TcSNP1063	7	T	G
31	CL423Contig1	6	A	G	79	TcSNP791	7	A	T
32	CL532Contig1	7	A	G	80	TcSNP1201	7	T	G
33	CL2205Contig1	7	T	C	81	TcSNP1194	7	T	G
34	CL858Contig1	8	T	G	82	TcSNP606	7	A	G
35	CL235Contig1	8	T	C	83	TcSNP189	8	A	G
36	CL1Contig129	8	T	C	84	TcSNP23	8	T	C
37	CL8Contig4	9	C	G	85	TcSNP1309	8	T	C
38	CL1Contig135	9	A	C	86	TcSNP269	8	A	G
39	CL1957Contig1	9	C	G	87	TcSNP899	8	A	G
40	CL918Contig1	9	T	C	88	TcSNP1064	9	A	G
41	CL1600Contig1	9	T	G	89	TcSNP264	9	A	T
42	CL139Contig1	9	T	C	90	TcSNP184	9	T	G
43	CL1030Contig1	9	T	C	91	TcSNP1305	9	A	G
44	CL639Contig1	10	A	C	92	TcSNP1517	9	T	C
45	CL88Contig2	10	A	G	93	TcSNP731	10	A	G
46	CL1Contig113	10	T	G	94	TcSNP1041	10	C	G
47	CL282Contig2	10	T	G	95	TcSNP653	10	A	G
48	CL702Contig1	10	T	C	96	TcSNP1392	10	T	C

A: adenina, C: citosina, G: guanina, T: timina

Analysis of conglomerate. These profiles were compared using the Matching Multilocus Genotypes function in GenAlex and we identified 16 groups of samples sharing similar profile. Such groups were identified with letters A through P for subsequent analysis, adding the number of

introductions belonging to each group (Table 5).

The largest group was designated as a C-13, with 13 samples sharing similar genetic profile. Some Forastero controls (Forastero1_4, Forastero1_21) were grouped with other introductions of Nacional cacao, while the CCN51 (CCN) controls shared a single profile (group M-5). According to the IP and PIsib identity probability analysis, it was established that a set of 21 SNPs are sufficient to distinguish the evaluated introductions, even when relatives were included (PIsib 0.00008) and that with the total of 95 SNPs evaluated an IP 2.4E-31 and a PIsib 1.1E-16 were obtained, showing the capacity of distinction of these molecular markers.

The heterozygosity from cocoa introductions averaged 43.2 %, the samples from group F-3 (L46H70, L17H36, L15H31) had 64.6 % of heterozygous; while the lowest value was shown by group P-3 (L12H30, L12H29, L12H28) with 25.3 % (Figure 1).

The dendrogram (Figure 2) shows the 16 groups plus the 43 remaining introductions (on the right) (Table 5) and their genetic similarities according to the Rogers' standard. In both cases, a single branch is observed with separations of greater distance between the C-Criollo-Cent and C-Cont-CIAT controls (Contamana) from which the other Nacional cocoa introductions and controls of Ecuador are located, so that consistency can be inferred in the grouping according to the evaluated standards.

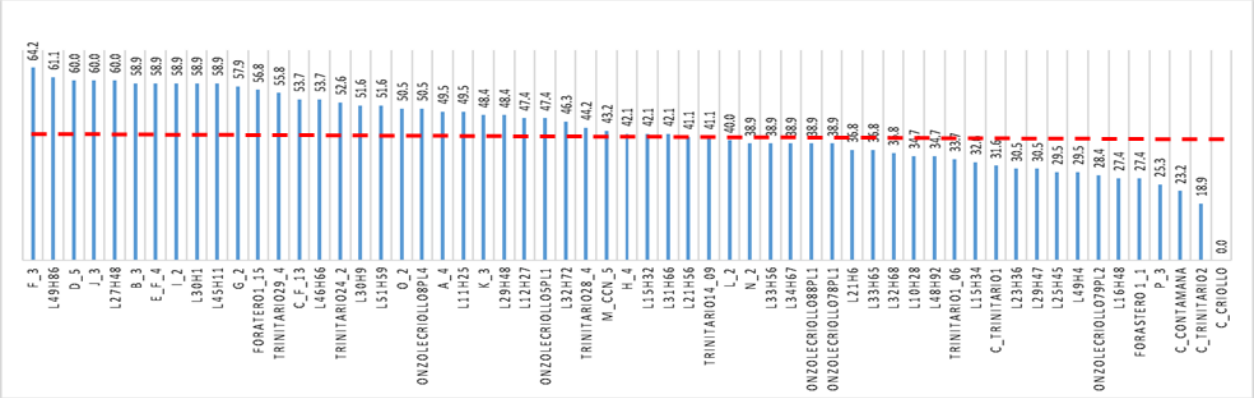

Figure 1. Average percentage of heterozygosity of analysed cocoa introductions. The dotted red line shows the average value (43.2 %)

The probability of group assignment by Delta K Bayesian cluster analysis suggests a value of K=2. The average assignment probability of the first group in red was 0.167 and the second in green was 0.834 (Figure 3). Group two introductions can be inferred to belong to Nacional cocoa; however, L10H28, L11H25, L12H27, L16H48, L32H72, L46H66. L52H98 and L53H4 have shared proportions with the first group in ranges of 0.231 to 0.549, which allows inferring that they contain information not only Nacional cocoa.

The introductions used as controls were placed in group one with the exception of Forastero 1-4 and Forastero 1-21 which showed 0.994 and 99.15 of allocation ratio with group two respectively, and could be considered as Nacional (Figure 3); they were also consistent with the UPGMA grouping.

According to the AMOVA the Fst indicates that there is little differentiation between groups, but this is significant. The percentage of differentiation between the two populations was 7 %, within individuals 93 %, and between individuals was zero (Table 6).

DISCUSSION

According to HE and HO there is a high genetic diversity and a high proportion of heterozygous individuals observed for a locus in the evaluated introductions. The high heterozygosity values in this study exceed those obtained by Ji et al (2013) who reported 0.206 heterozygosity in Honduran and Nicaraguan varieties.

Low heterozygosity values have been associated with the Central American Criollo

morphological group, while high values have been frequently found in individuals from the Forastero group of the upper Amazon region (Livingstone et al., 2017). The levels of heterozygosity obtained in this study are partly explained by the self-incompatibility characteristics of some cocoa materials, which avoid high frequencies of self-pollination or consanguinity (Scheltema, 1989; Ruiz et al., 2015).

Table 4. Descriptive statistics for the 95 SNPs assessed in 99 cocoa introductions. Single nucleotide sequence (SNPs); major allele frequency (FAM); observed genotypes (GO); heterozygosity observed (H_O) expected heterozygosity (H_E), polymorphic information content (PIC)

SNP	FAM	GO	HO	HE	PIC	SNP	FAM	GO	HO	HE	PIC
CL1002Contig1	0.859	3	0.242	0.276	0.213	TcSNP1041	0.960	2	0.081	0.084	0.075
CL1030Contig1	0.505	3	0.768	0.500	0.375	TcSNP1058	0.505	3	0.747	0.500	0.375
CL1086Contig1	0.692	3	0.556	0.428	0.335	TcSNP1063	0.732	3	0.455	0.405	0.315
CL1125Contig2	0.722	3	0.414	0.396	0.321	TcSNP1064	0.919	3	0.141	0.175	0.138
CL1312Contig1	0.556	3	0.384	0.495	0.372	TcSNP1111	0.606	3	0.667	0.477	0.363
CL132Contig1	0.510	3	0.677	0.499	0.375	TcSNP1159	0.646	3	0.606	0.460	0.353
CL139Contig1	0.591	3	0.717	0.481	0.367	TcSNP1194	0.884	3	0.212	0.249	0.184
CL1454Contig1	0.505	3	0.788	0.500	0.375	TcSNP1201	0.995	2	0.010	0.029	0.010
CL1600Contig1	0.657	3	0.646	0.455	0.349	TcSNP1209	0.626	3	0.667	0.463	0.359
CL171Contig2	0.904	3	0.172	0.206	0.158	TcSNP1212	0.652	3	0.475	0.455	0.351
CL192Contig2	0.545	3	0.768	0.496	0.373	TcSNP1305	0.611	3	0.737	0.471	0.362
CL1957Contig1	0.783	3	0.273	0.347	0.282	TcSNP1309	0.576	3	0.364	0.492	0.369
CL1Contig113	0.520	3	0.657	0.499	0.375	TcSNP1383	0.995	2	0.010	0.029	0.010
CL1Contig128	0.561	3	0.677	0.489	0.371	TcSNP1390	0.571	3	0.616	0.491	0.370
CL1Contig129	0.505	3	0.768	0.500	0.375	TcSNP1392	0.990	2	0.020	0.038	0.020
CL1Contig135	0.581	3	0.596	0.486	0.368	TcSNP1453	0.758	3	0.465	0.377	0.300
CL1Contig277	0.520	3	0.596	0.500	0.375	TcSNP1517	0.985	2	0.030	0.047	0.029
CL1Contig69	0.540	3	0.697	0.495	0.373	TcSNP152	0.990	2	0.020	0.029	0.020
CL209Contig1	0.990	2	0.020	0.038	0.020	TcSNP174	0.909	3	0.162	0.183	0.152
CL218Contig1	0.515	3	0.687	0.499	0.375	TcSNP184	0.525	3	0.788	0.499	0.374
CL2205Contig1	0.535	3	0.525	0.495	0.374	TcSNP189	0.545	3	0.788	0.495	0.373
CL235Contig1	0.571	3	0.535	0.492	0.370	TcSNP206	0.990	2	0.020	0.029	0.020
CL276Contig1	0.631	3	0.677	0.460	0.357	TcSNP23	0.813	3	0.333	0.313	0.258
CL282Contig2	0.732	3	0.475	0.396	0.315	TcSNP264	0.566	3	0.788	0.489	0.371
CL2987Contig1	0.520	3	0.758	0.498	0.375	TcSNP269	0.505	3	0.606	0.500	0.375
CL317Contig1	0.540	3	0.859	0.495	0.373	TcSNP28	0.995	2	0.010	0.029	0.010
CL318Contig1	0.687	3	0.485	0.446	0.338	TcSNP316	0.525	3	0.667	0.498	0.374
CL3336Contig1	0.515	3	0.808	0.499	0.375	TcSNP344	0.520	3	0.737	0.498	0.375
CL359Contig1	0.874	3	0.111	0.249	0.196	TcSNP382	0.854	3	0.253	0.269	0.219
CL3696Contig1	0.535	3	0.727	0.497	0.374	TcSNP395	0.929	2	0.141	0.151	0.123
CL423Contig1	0.515	3	0.747	0.500	0.375	TcSNP413	0.626	3	0.667	0.468	0.359
CL456Contig1	0.737	3	0.263	0.387	0.312	TcSNP418	0.677	3	0.545	0.439	0.342
CL4Contig14	0.540	3	0.758	0.495	0.373	TcSNP437	0.990	2	0.020	0.038	0.020
CL527Contig1	0.515	3	0.828	0.500	0.375	TcSNP475	0.596	3	0.646	0.479	0.366
CL532Contig1	0.576	3	0.626	0.488	0.369	TcSNP510	0.505	3	0.747	0.500	0.375
CL552Contig2	0.535	3	0.707	0.497	0.374	TcSNP595	0.924	3	0.131	0.159	0.130
CL581Contig1	0.616	3	0.626	0.475	0.361	TcSNP602	0.571	3	0.596	0.489	0.370
CL588Contig1	0.586	3	0.768	0.486	0.368	TcSNP606	0.788	3	0.202	0.363	0.278
CL639Contig1	0.904	3	0.172	0.213	0.158	TcSNP653	0.505	3	0.586	0.500	0.375
CL646Contig2	0.520	3	0.657	0.500	0.375	TcSNP689	0.869	3	0.222	0.242	0.202
CL695Contig1	0.500	3	0.677	0.500	0.375	TcSNP731	0.652	3	0.596	0.458	0.351
CL702Contig1	0.566	3	0.727	0.489	0.371	TcSNP736	0.662	3	0.475	0.452	0.348
CL745Contig1	0.616	3	0.566	0.468	0.361	TcSNP791	0.601	3	0.535	0.475	0.365
CL77Contig2	0.970	2	0.061	0.066	0.057	TcSNP798	0.949	2	0.101	0.110	0.091
CL858Contig1	0.505	3	0.545	0.500	0.375	TcSNP865	0.995	2	0.010	0.029	0.010
CL88Contig2	0.848	3	0.283	0.276	0.224	TcSNP899	0.505	3	0.582	0.500	0.375
CL8Contig4	0.611	3	0.657	0.473	0.362	TcSNP928	0.833	3	0.212	0.288	0.239
CL918Contig1	0.813	3	0.313	0.319	0.258	Mean	0.685	-	0.479	0.378	0.289

Table 5. Groups of introductions sharing similar genetic profiles according to the "Matching Multilocus Genotypes" function of GeneAlex

Group	Number	Code/group	Code/Introduction
A	4	A-4	L21H57, L20H53, L20H50, L20H49
B	3	B-3	L40H49, L18H36, L17H30
C	13	C-13	Forstero1_21, L44H88, L33H49, L33H47, L33H26, L33H25, L32H65,L32H54, L32H48, L26H64, L25H64, L25H60, L25H59
D	5	D-5	L33H45, L30H3, L29H7, L29H4, L28H48
E	4	E-4	Forastero 1_4, L46H75, L33H8, L30H25
F	3	F-3	L46H70, L17H36, L15H31
G	2	G-2	L30H46, L30H45
H	4	H-4	L46H68, L24H19, L24H14, L24H12
I	2	I-2	L48H23, L21H38
J	3	J-3	L18H58, L17H25, L13H37
K	3	K-3	L17H38, L13H11, L12H8
L	2	L-2	L28H22, L27H19
M	5	M-5	CCN513_5, CCN513_4, CCN512_4, CCN512_3, CCN512_2
N	2	N-2	L49H98, L42H80
O	2	O-2	L53H4, L52H98
P	3	P-3	L12H30, L12H29, L12H28

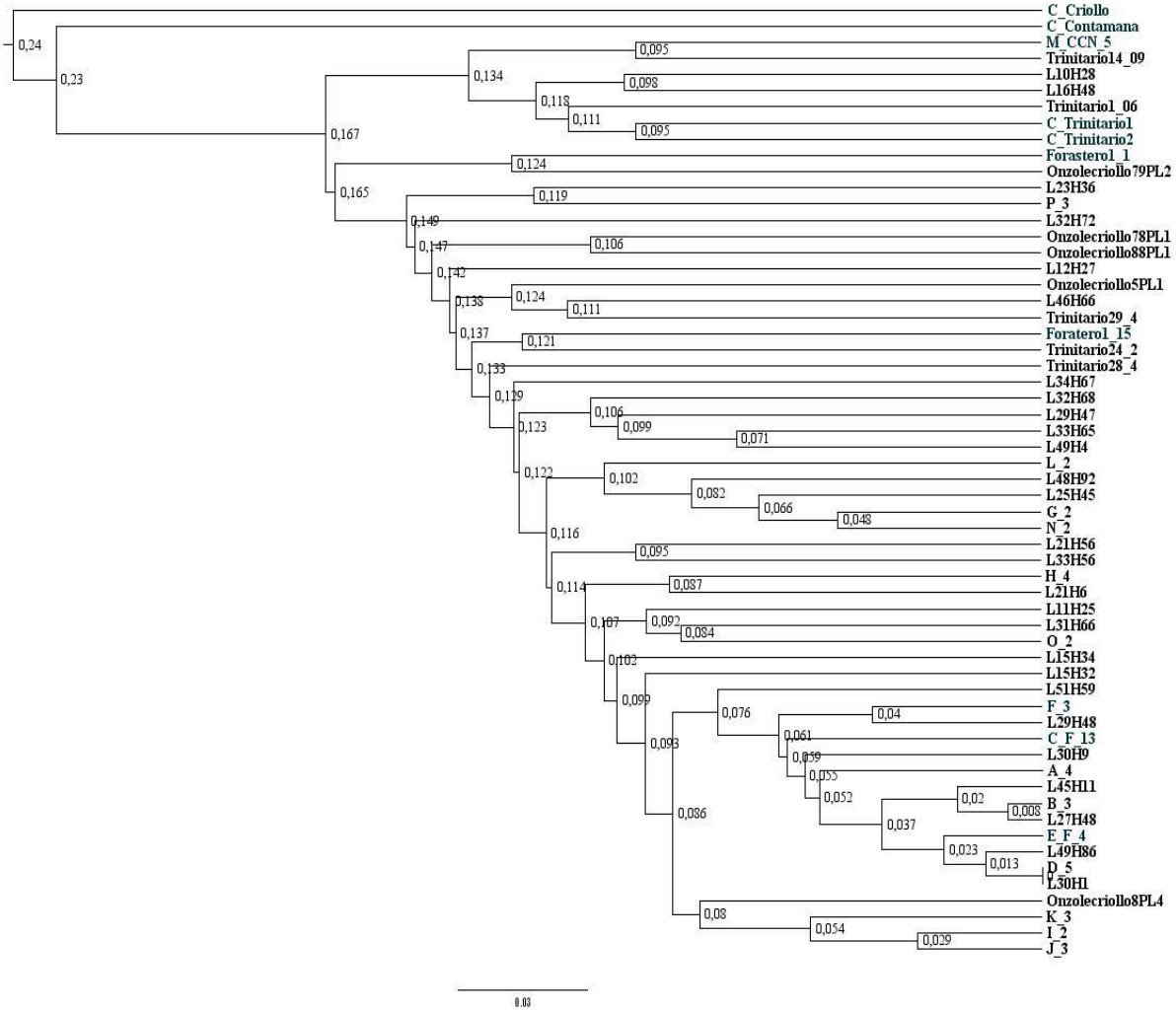

Figure 2. Dendrogram of genetic distances elaborated with the Rogers' standard and UPGMA algorithm based in 90 snips markers

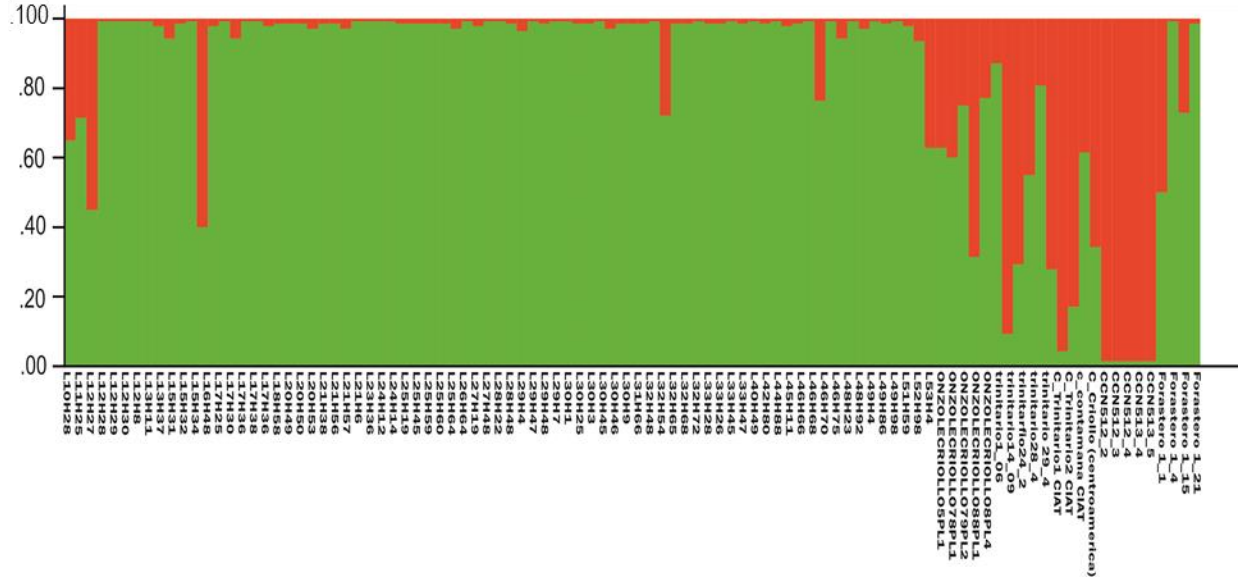

Figure 3. Determination of the number of genetic groups (ΔK) using the Structure Harvester program. Each vertical line represents an introduction, and each color represents the most likely groups (Red: first group, averaging 0.167; Green: second group, averaging 0.834). Lines with both colors represent a mixture of the introductions

Table 6. Summary of statistics of analysis of molecular variance (AMOVA) for *Theobroma cacao* L.

Source	Df	SS	MS	SD	Difference	F-statistics
Among pops	1	128.025	128.025	1.613	7 %	Fst=0.084
Among individuals	101	1288.014	12.753	0.000	0 %	Fis=-0,271
Within individuals	103	2288.500	22.218	22.218	93 %	Fit=-0.135
Total	205	3704.539		23.832	100 %	

The genetic variation detected by the content of polymorphic information was informative for the evaluation of the diversity of the introductions in this study, considering also that the selected SNPs are represented along the 10 cocoa chromosomes. Therefore, these results are comparable with those obtained by Osorio et al. (2017), who obtained a PIC in the range of 0.113 for TcSNP1383 to 0.460 for TcSNP709, and with Cosme et al. (2016), who obtained a PIC above 0.4 in a collection of Colombian cocoa.

According to the analysis of identity probability and identity probability including relatives, the possibility is inferred that introductions located in the same group may be synonymous or duplicated. The number of SNPs considered in this study are similar to those reported by Ji et al. (2013) who conducted a study on 84 fine-tasting cocoa varieties from farms in Honduras and Nicaragua; as well as 31 clones from international cocoa collections as a reference; who demonstrated that 26 SNPs with 99.99 % confidence are sufficient to identify individual cocoa varieties.

According to the genetic distances of the dendrogram, the separation of the C-Criollo control sample, which corresponds to the Central American Criollo, seems distant from the C-Contamana control, as well as from the other introductions. This corresponds to what Berdugo et al. (2017) described about the Criollo genotype as the most genetically differentiated group. The results of this research are equally consistent with those reported by Loor et al. (2009) and Ruiz et al (2015) for the Nacional cocoa, reporting that within cocoa individuals there are low levels of genetic differentiation, but high levels of heterozygosity. This consolidates what was reported by Lerceteau et al (1997) when they argued that the high genetic diversity of the native

populations of Nacional cocoa was due to the successive introduction of cocoa germplasm since the late 1980s, including the foreign group. Likewise, Romero et al (2010) affirm that there is a genetic relationship between the Nacional cocoa of Ecuador and the Criollos and Forasteros groups.

Motamayor et al. (2008) defined cocoa groups, where some introductions of the Nacional variety are located in groups close to Criollo, Trinitario and Forastero, which had been corroborated by Lerceteau et al. (1997) when mentioning that the Nacional cocoa and from other origins are related to each other. For the Trinitarian controls the assignment is shared between group one and two according to the analysis of the structure; therefore, its nature of mixture of the other genetic groups is inferred, being consistent these results with what was reported by Lindo et al. (2018) who express that the Trinitarian cocoa is a mixture of Criollo and Amelonado. In addition, the new classification of cocoa genetics by Motamayor et al (2008) excluded Trinitario as a mixture of Criollo and Forastero.

The genetic characterization with SNPs markers of the Nacional cocoa variety of the CCAT of Ecuador constitutes an important contribution to develop and implement genetic improvement programs. This is especially the case of the selection of parents that allows to widen the genetic base, besides considering morpho-agronomic characteristics that allow to obtain introductions with a high yield and tolerance to diseases in the cultivation of this variety of world-wide renown and importance in the markets of cocoa of fine aroma.

CONCLUSIONS

The set of SNP markers used in this study were efficient in the characterization of cocoa introductions, since they allowed the obtaining of 99 genetic profiles including the profiles of 80 introductions of Nacional cocoa cultivated in Ecuador. Genetically identical introductions possibly from the same clone were identified. The genetic distribution of the samples suggests that the Nacional cocoa evaluated corresponds to introductions that share genetic information with Trinitarios, possibly due to the history of early hybridization of this group. The comparison of the genetic data obtained with the morpho-agronomic, geographic and other information of characteristics of agronomic interest will allow the selection of outstanding cocoa introductions in terms of productivity and disease tolerance. This will serve as a basis for genetic improvement programs, in addition to facilitating the comparison of molecular profiles with other varieties of fine aroma cocoa from other cocoa germplasm banks in Ecuador and in Latin America and the Caribbean.

LITERATURE CITED

1. Amores, F., J. Agama, F. Mite, J. Jiménez, G. Loor and J. Quiroz. 2009. EET 544 y EET 558: Nuevos clones de cacao Nacional para la producción bajo riego en la Península de Santa Elena. Instituto Nacional de Investigaciones Agropecuarias (INIAP). Estación. Boletín Técnico 134. Quevedo. 54 p.
2. Asociación Nacional de Exportadores de Cacao e Industrializados del Ecuador. 2010. Cacao Nacional, Un producto emblemático del Ecuador. http://www.anecacao.com/es/quienes-somos/cacao-nacional (retrieved Mar 19, 2019).
3. Argout, X., O. Fouet, P. Wincker, K. Gramacho, T. Legavre, X. Sabau et al. 2008. Towards the understanding of the cocoa transcriptome: Production and analysis of an exhaustive dataset of ESTs of *Theobroma cacao* L. generated from various tissues and under various conditions. BMC Genomics 9(1): 512.
4. Argout, X., J. Salse, J. Aury, M. Guiltinan, G. Droc, J. Gouzy, M. Allegre et al. 2011. The genome of *Theobroma cacao*. Nature genetics 43(2): 101-8.
5. Arvelo, M., D. Gonzáles, S. Maroto, T. Delgado and P. Montoya. 2017. Manual Técnico del Cultivo de Cacao Prácticas Latinoamericanas. San José Costa Rica: IICA.
6. Beerli, P. 2005. Pairwise distance methods. Computacional Evolucionary Biology, N° BSC5936-Fall: 1-7.
7. Cornejo, O., M. Yee, V. Dominguez, M. Andrews, A. Sockell, E. Strandberg and J. Motamayor. 2018. Population genomic analyses of the chocolate tree, *Theobroma*

cacao L., provide insights into its domestication process. Communications Biology 1(1): 1-12.

8. Corporation Fluidigm. 2018. Biomark/EP1 Software v4. fluidigm.com/software.
9. Cosme, S., H. Cuevas, D. Zhang, T. Oleksyk and B. Irish. 2016. Genetic diversity of naturalized cacao (*Theobroma cacao* L.) in Puerto Rico. Tree Genetics and Genomes 12(5): 88.
10. Danial, D. and J. Rojas-Beltran. 2007. Uso de marcadores moleculares en el mejoramiento genético de plantas. Quito-Ecuador: Instituto Nacional de Investigaciones Agropecuarias [INIAP]. 199 p.
11. DuVal, A., S. Gezan, G. Mustiga, D. Livingstone, J. Chaparro, J. Marelli et al. 2017. Genetic Parameters and the Impact of Off-Types for *Theobroma cacao* L. in a Breeding Program in Brazil. Frontiers in Plant Science 8: 1-12.
12. Earl, D.A. and B.M. VonHoldt. 2012. Structure Harvester: A website and program for visualizing Structure output and implementing the Evanno method. Conservation Genetics Resources 4(2): 359-361.
13. Fang, W., L. Meinhardt, S. Mischke, C. Bellato, L. Motilal and D. Zhang. 2014. Accurate determination of genetic identity for a single cacao bean, using molecular markers with a nanofluidic system, ensures cocoa authentication. Journal of Agricultural and Food Chemistry 62(2): 481-87.
14. Illic, K., D. Zhang, X. Wang, R. Jones, L. Meinhardt and L. Wang. 2012. Cacao Germplasm Characterization with 48-SNP Genotyping Panel using Fluidigm SNPtypeTM Assays and Dynamic ArrayTM Integrated Fluidic Circuits. San Diego California (https://pag.confex.com/pag).
15. Ji, K., M. Boccara, L. Motilal, D. Zhang, P. Lachenaud and L. Meinhardt. 2012. 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 Resources and Crop Evolution 60(2): 441-453.
16. Laliberté, B. 2012. A Global Strategy for the Conservation and use of Cacao Genetic Resources, as the Foundation for a Sustainable Cocoa Economy. Bioversity Internacional 66: 186.
17. Lanaud C., O. Fouet, K. Gramacho and X. Argout. 2006. A large EST resource for *Theobroma cacao* including cDNAs isolated from various organs and under various biotic and abiotic stresses. Proc. 15th International Cocoa Research Conference. pp. 185-191.
18. Lerceteau, E., S. Flipo, J. Quiroz, J. Soria, V. Pétiard and D. Crouzilat. 1997. Genetic differentiation among Ecuadorian *Theobroma cacao* L. accessions using DNA and morphological analyses. Euphytica 95(1): 77-87.
19. Liu, K. and S. Muse. 2005. PowerMaker: An integrated analysis environment for genetic maker analysis. Bioinformatics 21(9): 2128-2129.
20. Livingstone, D., C. Stack, G. Mustiga, D. Rodezno, C. Suarez, F. Amores et al. 2017. A larger chocolate chip-development of a 15K *Theobroma cacao* L. SNP array to create high-density linkage maps. Frontiers in Plant Science 8: 1-18.
21. Lindo, A.A., D. Robinson, P. Tennant, L. Meinhardt and D. Zhang. 2018. Molecular characterization of cacao (*Theobroma cacao*) germplasm from Jamaica using single nucleotide polymorphism (SNP) markers. Tropical Plant Biology 11(3-4): 93-106.
22. Loor, R., A. Risterucci, B. Courtois, O. Fouet, M. Jeanneau, E. Rosenquist et al. 2009. Tracing the native ancestors of the modern *Theobroma cacao* L. population in Ecuador. Tree Genetics and Genomes 5(3): 421-33.
23. Mata-Quiroz, A., A. Arciniegas-Leal, W. Phillips-Mora, S. Mischke, A. Mata-Quirós, L. Motilal et al. 2018. Assessing hidden parentage and genetic integrity of the "united fruit clones" of cacao (*Theobroma cacao*) from Costa Rica using SNP markers. Breeding Science 68(5): 545-53.
24. Michiels, A., W. Ende, M. Tucker and L. Riet. 2003. Extraction of high-quality genomic DNA from latex-containing plants. Analytical Biochemistry 315(1): 85-89.
25. Motamayor, J., P. Lachenaud, J. da Silva e Mota, R. Loor, D. Kuhn, J. Brown and R. Schnell. 2008. Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L.). PLoS

One 3(10): 8.

26. Motamayor, J., A. Risterucci, P. Lopez, C. Ortiz, A. Moreno and C. Lanaud. 2002. Cacao domestication I: the origin of the cacao cultivated by the Mayas. Heredity 89(5): 380-86.

27. Motamayor, J., R. Schnell, D. Kuhn, W. Phillips, N. Haiminen, D. Livingstone et al. 2013. The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. Genome Biology 14(6): 2-24.

28. Osorio-Guarín, J., R. Yockteng, C. Quintero, Y. Zapata, R. Coronado, G. Gallego-Sánchez and J. Berdugo-Cely. 2017. Colombia a source of cacao genetic diversity as revealed by the population structure analysis of the Germplasm Bank of *Theobroma cacao* L. Frontiers in Plant Science 8(11): 1-13.

29. Peakall, R. and P. Smouse. 2012. GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28(19): 2537-2539.

30. Pritchard, J.K., M. Stephens and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

31. Pound, F. 1945. A note on the cocoa population of South America. Reprinted in 1982 in Arch. Cacao Res. 1: 96-97.

32. Rambaut, A. 2016. FigTree V1.4.2: Tree figure drawing tool. Institute of Evolutionary Biology, University of Edinburgh. http://tree.bio.ed.ac.uk.

33. Reif, J., A. Melchinger and M. Frisch. 2005. Assessing the genetic diversity in crops with molecular markers: theory and experimental results with CIMMYT wheat and maize elite germplasm and genetic resources. Crop Science 45: 1-7.

34. Rogers, J. 1972. Measures of genetic similarity and genetic distance. Studies in genetics. University of Texas Pub. 7213: 145-153.

35. Romero, C., J. Bonilla., E. Santos., E. Peralta and X. Zhong. 2010. Identificación varietal de 41 plantas seleccionadas de cacao (*Theobroma cacao* L.) provenientes de cuatro cultivares distintos de la región amazónica ecuatoriana, mediante el uso de marcadores microsatélites. Revista Tecnológica 23(1): 121-28.

36. Ruiz, X., M. Almanza, Y. Morillo, A. Morillo, A. Gonzalez, Ä. Caicedo and J. Muñoz. 2015. Comparación genética de tres fuentes del cacao *Theobroma cacao* L. mediante el uso de marcadores microsatélites. Biotecnología en el Sector Agropecuario 13(1): 10-18.

37. Sánchez-Mora, F., M. Medina-Jara, G Díaz-Coronel, R. Ramos-Remache, J. Vera-Chang, V. Vásquez-Morán et al. 2015. Potencial sanitario y productivo de 12 clones de cacao en Ecuador. Revista Fitotecnia Mexicana 38(3): 265-74.

38. Scheltema, T. 1989. La autoincompatibilidad en los híbridos de cacao del CATIE. San José (Costa Rica): CATIE 43: 1-90.

39. Thomas, E., M. van Zonneveld, J. Loo, T. Hodgkin, G. Galluzzi and J. van Etten. 2012. Present spatial diversity patterns of *Theobroma cacao* L. in the neotropics reflect genetic differentiation in pleistocene refugia followed by human-influenced dispersal. PLoS One 7(10): 1-17.

40. Wilde, J., R. Waugh and W. Powell. 1992. Genetic fingerprinting of *Theobroma* clones using randomly amplified polymorphic DNA markers. Theoretical and Applied Genetics 83(6): 871-77.

41. Zapata, Y. 2016. Protocolo para extracción de ADN de cacao. Manual interno del Laboratorio de Genética Molecular y Cultivo de Tejidos. CIAT. Cali, Colombia.

42. Zarrillo, S., N. Gaikwad, C. Lanaud, T. Powis, C. Viot, I. Lesur et al. 2018. The use and domestication of *Theobroma cacao* during the mid-Holocene in the upper Amazon. Nature Ecology and Evolution 2(12): 1879-1888.

43. Zhang, D., W. Martínez, E. Johnson, E. Somarriba, W. Phillips-Mora, C. Astorga et al. 2012. Genetic diversity and spatial structure in a new distinct *Theobroma cacao* L. population in Bolivia. Genetic Resources and Crop Evolution 59(2): 239-52.