

Caracterización morfoagronómica y molecular de la colección de cacao de la Federación Nacional de Cacaoteros de Colombia

Nubia Martínez Guerrero¹, Gerardo Gallego², Patricia Zapata², Contanza Quintero², Diannefair Duarte¹, Joe Tohme², Oscar Darío Ramírez¹

1. Departamento de Investigación. Federación Nacional de Cacaoteros. investigacion@fedecacao.com.co
2. Centro Internacional de Agricultura Tropical. g.gallego@cgiar.org

Resumen

La identidad genética y las relaciones evolutivas en colecciones de germoplasma se han estudiado a través de marcadores moleculares, usados en el estudio, dando valor agregado en términos de información genética útil a los bancos de germoplasma. En cacao se han utilizado para establecer la identidad de los genotipos y su relación genética con los grupos establecidos, como base para programas de mejoramiento genético asistido por marcadores moleculares. Con los trabajos previos de selección de genotipos promisorios a través de mejoramiento participativo, la Federación Nacional de Cacaoteros caracterizó a nivel morfoagronómico 42 y molecular 70 genotipos, colectados en diferentes regiones productoras de Colombia, dentro de cultivos tradicionales en fincas de agricultores. La colección, establecida en condiciones de San Vicente de Chucurí, Santander (N 6° 52'57", W 73° 24' 46"), se caracterizó con 56 descriptores distribuidos de la siguiente manera: de planta (2), hoja (12), flor (13), fruto (20) y semilla (9). Para la caracterización molecular se utilizó un chip de 96 SNP's bajo la plataforma Fluidigm®, siguiendo los protocolos previamente establecidos por CIAT. El procesamiento de las muestras y análisis de los datos se realizó entre FEDECACAO y el laboratorio de Biotecnología del Centro Internacional de Agricultura Tropical -CIAT- en Palmira, Colombia. Se utilizó el programa PowerMarker® 3.25 para realizar cálculos de parámetros descriptivos de variación genética y el programa NTSYS-pc para el análisis de agrupamientos y establecer las relaciones entre los individuos bajo estudio, con base en la distancia genética de Nei. Así mismo, se diagramaron los perfiles genéticos (Fingerprinting) de cada genotipo con el programa Flapjack. Basado en los resultados se evidencia la diversidad genética entre los genotipos evaluados y se ratifica la identidad de algunos comerciales para mejorar etiquetados. Se pudo verificar la relación existente entre genotipos seleccionados y parentales utilizados en programas de mejoramiento genético anteriores. La caracterización permitirá establecer futuros programas de mejoramiento genético para atender las necesidades de la demanda creciente de material para nuevas siembras y modernización del cultivo en Colombia.

1. Introducción

Desde el año 2000 Colombia ha promovido el establecimiento de cultivos comerciales de cacao mediante la clonación con el fin de incrementar la producción y productividad, lo que generó el incremento en el uso de genotipos internacionales reconocidos y llevó a procesos de selección de nuevos genotipos regionales como estrategia de aprovechamiento de la diversidad genética.

Se desarrollaron proyectos de evaluación de genotipos internacionales, así como la selección de nuevos en las regiones productoras, permitiendo conformar una colección de materiales promisorios que fueron evaluados y caracterizados.

El objetivo del estudio fue caracterizar la diversidad genética por medio de marcadores tipo SNP a 71 genotipos y evaluar las características morfoagronómicas de 42 genotipos de cacao que hacen parte de la colección de la Federación Nacional de Cacaoteros – FEDECACAO -

2. Metodología

2.1 Caracterización morfoagronómica

La toma de muestras y registro de datos para caracterización, morfoagronómica se realizó en la Granja Villa Mónica, ubicada en San Vicente de Chucurí, Santander, a 900 msnm., con una precipitación promedio de 1.800 mm y una humedad relativa de 80%.

Se utilizaron los descriptores propuestos por Engels *et al.*, (1980), Eskes *et al.*, (2000), Castañeda *et al.*, (2011), García (1997) y Jiménez *et al.*, (2006). Se evaluaron 56 variables: 2 de planta, 11 para hoja, 14 para flor, 19 para fruto y 10 para semilla, de las cuales 31 fueron cualitativas y 25 cuantitativas. El número de muestras utilizadas en la determinación de características cuantitativas y cualitativas fueron: para la hoja 15, para la flor, la planta y el fruto 10 y 50 para el grano.

2.2 Caracterización molecular

Se colectaron y almacenaron en bolsas plásticas con sílica gel hasta su procesamiento, hojas juveniles y adultas de 71 genotipos de cacao, procedentes de plantaciones en granjas ubicadas en San Vicente de Chucurí, Santander y Miranda, Cauca; que hacen parte de la colección de FEDECACAO en la que se encuentran materiales internacionales como ICS, TSH y materiales regionales obtenidos mediante Selección Varietal Participativa, en todas las regiones productoras de Colombia (Tabla 1)

Tabla 1. Lista de genotipos evaluados para caracterización molecular y morfoagronómica por Fedecacao.

Nombre	Procedencia	Nombre	Procedencia
ICS 1 *	Internacional	FEAR 26 *	Regional - Arauquita
ICS 6 *	Internacional	FEC 937	Regional - El Carmen
ICS 8	Internacional	FEC 938	Regional - El Carmen
ICS 39 *	Internacional	FGI 4 *	Regional – Gigante
ICS 60 *	Internacional	FLE 2 *	Regional – Lebrija
ICS 95 *	Internacional	FLE 3 *	Regional – Lebrija
CCN 51 *	Internacional	FLE 4 *	Regional – Lebrija
IMC 67 *	Internacional	FLE 28 *	Regional – Lebrija
EET 8 *	Internacional	FMAC 14 *	Regional – Maceo
EET 62	Internacional	FMA 7 *	Regional – Marsella
EET 96 *	Internacional	FQP 1 *	Regional – Quípama
CHUAO	Internacional	FSA 12 *	Regional – Saravena
SCA 6	Internacional	FSA 13 *	Regional – Saravena
TSA 654	Internacional	FSA 30 *	Regional – Saravena
TSH 565 *	Internacional	FSV 1 *	Regional - San Vicente
TSH 1188	Internacional	FSV 30	Regional - San Vicente
UF6 13	Internacional	FSV 41 *	Regional - San Vicente
UF 667	Internacional	FSV 72 *	Regional - San Vicente
CATONGO	Internacional	FSV 80 *	Regional - San Vicente
TSH 812 *	Internacional	FSV 85 *	Regional - San Vicente
PA 46	Internacional	FSV 89 *	Regional - San Vicente
PA 150	Internacional	FSV 94	Regional - San Vicente
R 30	Internacional	FSV 106	Regional - San Vicente
UF 650	Internacional	FSV 153 *	Regional - San Vicente
SCA 12	Internacional	FSV 155 *	Regional - San Vicente
CAU 39 *	Regional	FTA 2 *	Regional – Tame
CR 38	Regional	FTU 6 *	Regional – Tumaco
CR 39	Regional	FVA 6	Regional – Valledupar
CR 40	Regional	FYC 2 *	Regional – Yacopí
CR 51	Regional	MON 1	Regional – Quimbaya
FBO 1 *	Regional	PENTAGONAL	Regional - San Vicente
FBO 5	Regional	SC 5	Regional – Palmira
FCHI 8	Regional	SC 6	Regional – Palmira
FEC 2 *	Regional	SC 10	Regional – Palmira
FEAR 4 *	Regional	SCC 61*	Regional - San Vicente
FEAR 5 *	Regional	Progenie	
		FSAT 10	Regional - Ataco

* Genotipos con caracterización morfoagronómica.

2.2.1 Extracción de ADN y control de calidad de las muestras.

Se realizó la extracción de ADN a partir de tejido foliar siguiendo el método de Michiels *et al.* (2003), modificado por Zapata (2016), en el laboratorio de Genética Molecular y Cultivo de Tejidos del Centro Internacional de Agricultura Tropical – CIAT -. El ADN aislado fue sometido a control de calidad siguiendo la *Guía Práctica para genotipado de SNPs usando el sistema EPI y SNPtype Assays de Fluidigm* version F_03 (Quintero *et al.*, 2015).

2.2.2 Marcadores moleculares SNPs y Genotipificación

La genotipificación se realizó en la plataforma *Fluidigm EPI System*. Un conjunto de 96 SNPs (Illic *et al.*, 2012; Fang *et al.*, 2013; Ruiz *et al.*, 2004) distribuidos a lo largo de los diez cromosomas del cacao fueron escogidos para establecer los perfiles genéticos.

2.3 Análisis de datos

La relación entre los genotipos evaluados mediante la caracterización morfoagronómica se evaluó con la aplicación del algoritmo UPGMA y la distancia de Gower para construir un dendograma, usando el programa NTSYS-pc.

Se obtuvo el número de *loci* polimórficos, número de *loci* con datos faltantes, número de genotipos observados, frecuencias alélicas, heterocigocidad y contenido de información polimórfica (PIC). Se calcularon las distancias genéticas pareadas usando el estándar de Rogers (1972), a partir de las cuales se construyó un dendograma de distancias con el algoritmo UPGMA. Tanto los estadísticos descriptivos como los cálculos de distancias se realizaron con el programa PowerMarker V3.25 (Liu 2005), mientras que la visualización del dendograma se hizo en FigTree V1.4.2 (Rambaut, 2016). Los análisis de Probabilidad de Identidad (PID, PISib), Genotipos pareados (Matching Multilocus Genotypes), índice de diversidad de Shannon, Análisis de Coordenadas Principales (PCoA) y test de diferenciación entre grupos (FST), fueron realizados en el paquete para Excel GenAlex V6.502 (Peakall y Smouse, 2012). Un análisis para establecer las relaciones de parentesco entre pares de muestras (Kinship), se realizó utilizando el algoritmo de Manichaikul (Manichaikul *et al.*, 2010) incluido en el paquete de VCFtools (Danecek *et al.*, 2011); mientras que la matriz de coancestría resultante fue visualizada por medio de un gráfico heatmap construido con el software para análisis estadísticos R (*R Core Team* 2013).

3. Resultados y discusión

3.1 Caracterización morfoagronómica

En la figura 1 se presenta el dendograma resultado de la caracterización molecular de 42 genotipos a partir de los datos obtenidos de 57 variables cuantitativas y cualitativas. Se puede observar que ICS 39 e ICS 60 presentan menor distancia, teniendo en cuenta la similaridad de sus características morfológicas. De acuerdo con Johnson *et al.*, (2009), los materiales ICS provienen de una sola población genéticamente homogénea que puede ser considerada

como una variedad local con un ancestro común tipo Criollo, de origen centroamericano, con frutos y brotes con escasa antocianina (Bekele *et al.*, 1994). También presentan cercanía los genotipos TSH 565 y FEAR 5 posiblemente por la similitud en características de tipo morfológico y de componentes del rendimiento, como color y forma del fruto, características de las flores, color y medidas de la semilla o grano.

A mayor distancia se forma un grupo con los genotipos ICS 1, ICS 95 y FGI 4, y otro grupo con ICS 6, ICS 39 e ICS 60, por similitud en características morfológicas del fruto. FEAR 26 se ubica aparte de todos los genotipos evaluados.

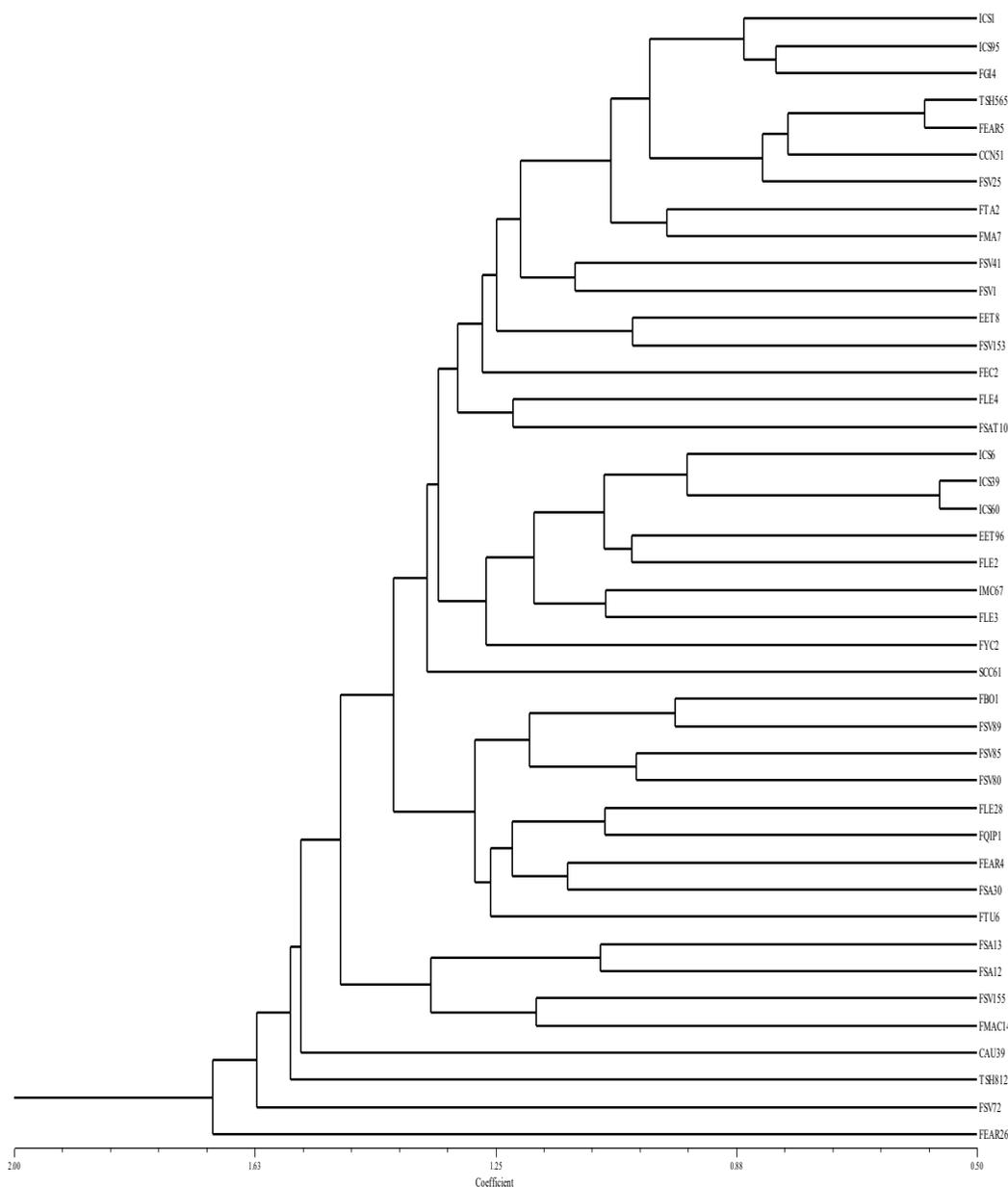


Figura 1. Árbol de distancias para características morfoagronómicas. UPGMA

3.2 Caracterización molecular

Todos los SNPs mostraron ser polimórficos con un promedio de 0.3058; donde TcSNP206 obtuvo el menor contenido de información polimórfica (PIC= 0.0405) y el TcSNP736 el mayor (PIC= 0.3749). La heterocigocidad obtenida para el conjunto de SNPs estuvo entre 0.0422 para el SNP TcSNP206 y 0.6901 para el TcSNP269, con un promedio de 0.4187. Para los marcadores TcSNP344, CL276Contig1, CL1125Contig2, TcSNP152 y TcSNP206 sólo se observó uno de los dos homocigotos posibles. Se encontraron 67 perfiles genéticos diferentes, de los cuales cuatro fueron compartidos para más de un genotipo así:

A = EET 62 y EET 96

B = EET 8 y UF 650

C = MON 1 y SC 5

D = CR40 y CR51

El promedio de heterocigocidad fue de 41.1% con CR40, CR51 y Catongo sin heterocigotos y FSV89 y TSH565 con los mayores valores de 69.5% y 71.6% respectivamente. El valor promedio de heterocigocidad encontrado es cercano al promedio reportado en diversos estudios realizados a partir de microsatélites (Motamayor *et al.* 2002; Sereno *et al.* 2006; Rehem *et al.* 2010; Bertolde *et al.* 2010). El índice de Shannon mostró diversidad moderada (0.5).

El análisis de distancias pareadas calculadas con el estándar de Rogers (1972), y el dendograma obtenido por medio del algoritmo UPGMA muestra que la distancia entre la mayoría de los materiales es reducida, aunque se puede observar la formación de tres grupos: El primero conformado por FBO1, TSH656, SCA12 y SCA6, el segundo por CR40 y CR51, y el tercero que comprende el resto de las muestras (Figura 2). El dendograma evidencia los posibles duplicados mencionados anteriormente, los cuales se encuentran hacia el final de una misma rama. Este análisis permitió observar otros individuos que tienen poca diferenciación genética entre ellos como FSA12 y FSA13; y UF667 con las muestras UF650 y EET8.

El análisis de diferenciación genética entre grupos (F_{st}), arrojó valores entre 0,21 y 0,39 para las tres comparaciones de los grupos observados en el dendograma de distancias (Tabla 2); evidenciando que existe alta diferenciación entre estos grupos de muestras; sin embargo, es importante ampliar el número de materiales evaluados, ya que esto permitirá realizar comparaciones con mayor confianza estadística y mejorará la conformación de grupos genéticos más definidos.

Tabla 2. Valores de F_{st} entre los grupos observados en el dendograma de distancias.

Fst	Grupo 1	Grupo 2	Grupo 3
Grupo 1	0,00		
Grupo 2	0,39	0,00	
Grupo 3	0,21	0,30	0,00

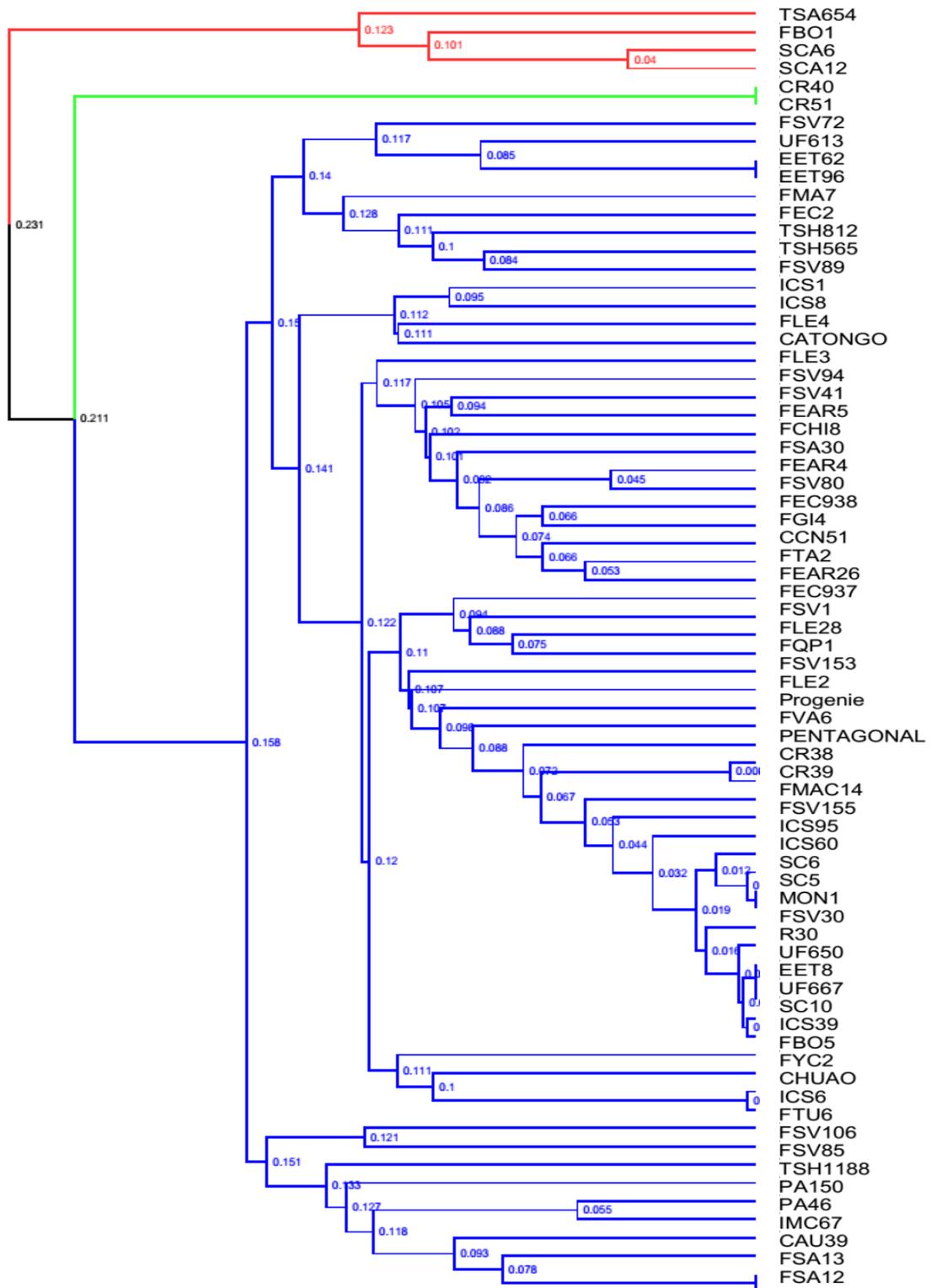


Figura 2. Árbol de distancias pareadas. UPGMA. Caracterización molecular de 71 genotipos

El Análisis de Coordenadas Principales (PCoA) resultante (Figura 3) permitió observar agrupaciones de genotipos que en su mayoría corresponden con los grupos identificados en el dendrograma de distancias (grupo 1 y 2), sin embargo se observa que las muestras FBO1 y TSA654 se encuentran en un punto intermedio entre el grupo 3 y 1.

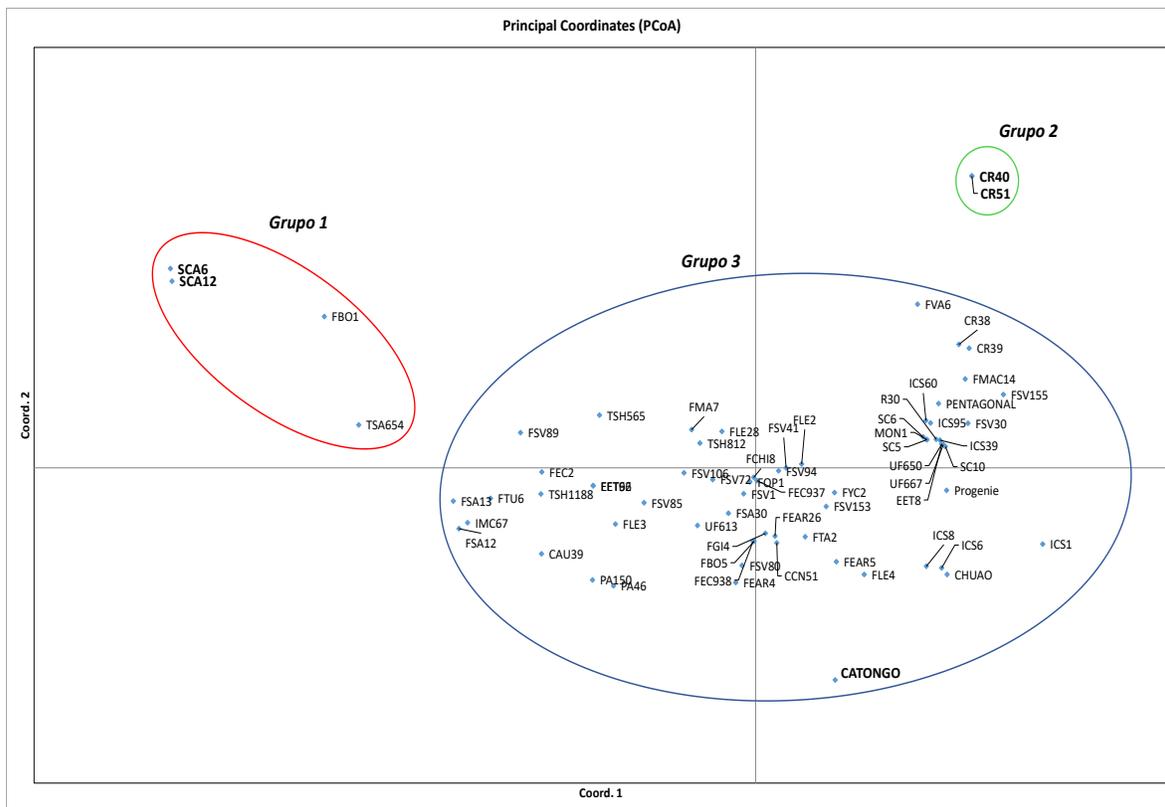


Figura 3. Análisis de Coordenadas Principales (PCoA). Los nombres en negrilla corresponden a las más distantes.

La matriz de coancestría evalúa la similitud genética entre los individuos e incluye la probabilidad de que esta similitud sea idéntica por descendencia, es decir, que un carácter genético sea idéntico en dos individuos porque ha sido heredado de un mismo progenitor (Manichaikul *et al.*, 2010). Se construyó un gráfico tipo *heatmap* con los valores de coancestría calculados y que son ordenados con el algoritmo de UPGMA para identificar grados de parentesco entre los genotipos evaluados. La matriz obtenida clasifica los genotipos evaluados con colores marcando duplicados o clones en rojo, parentesco de segundo grado en amarillo y los que parecen no tener relación cercana con el resto de individuos en azul (Figura 4).

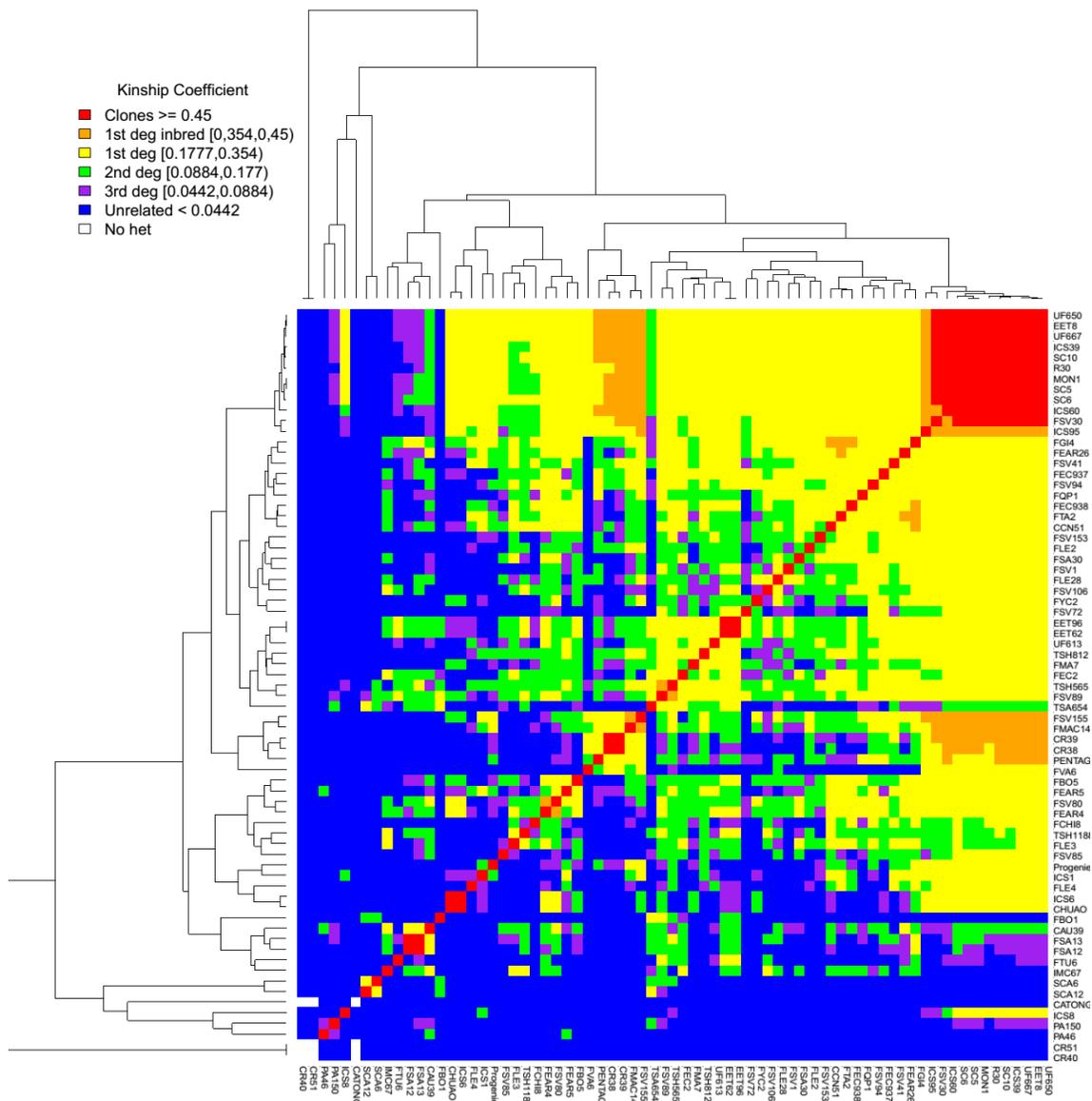


Figura 4. Heatmap a partir de la matriz de coancestría (kinship).

Las muestras UF 650, EET 8, UF 667, ICS 39, SC 10, R 30, MON 1, SC 5, SC 6, ICS 60, y FSV 30; forman un grupo que comparte un alto grado de parentesco. Los individuos que presentan genotipos homocigotos (Catongo, CR40 y CR51) no fueron comparados entre ellos, sin embargo muestran alta diferenciación respecto al resto de individuos (color azul en el heatmap), lo que corresponde con análisis anteriores. Otras muestras que presentan alta diferenciación son SCA6 y SCA12 que fueron identificadas dentro del grupo 1 en el dendograma. Los genotipos FBO 1 y TSA 654 mostraron compartir un mayor grado de parentesco con otros materiales, lo que pudo ser observado igualmente en el gráfico de PCoA.

La alta diferenciación encontrada para genotipos como ICS 8, PA 46 y PA 150 indica que éstas no están cercanamente emparentadas con el resto de materiales evaluados.

4. Conclusiones

Los análisis realizados indican que el conjunto de marcadores SNP usados fueron eficientes en la caracterización de la diversidad genética de cacao, ya que permitieron la obtención 67 perfiles genéticos y la identificación de posibles duplicados.

Es importante aumentar el número de materiales con diferenciación agronómica y procedencias geográficas diversas.

El presente estudio indica un alto grado de relación entre los individuos evaluados y en general, una moderada diferenciación genética

5. Referencias

Argout X., Salse J., Aury J.M., Guiltinan M.J., Droc G., Gouzy J., Allegre M., Chaparro C., Legavre T., Maximova S.N., *et al.* (2011) The genome of *Theobroma cacao*. *Nature Genetics* 43: 101–108.

Beerli P. (2005) Pairwise distance methods. Computational Evolutionary Biology Course. BSC5936-Fall 2005.

Bertolde F.Z., De Almeida A.A., Correa R.X., Gomes F.P., Gaiotto F.A., Baligar V.C., Lopes L. (2010) Molecular, physiological and morphological analysis of waterlogging tolerance in clonal genotypes of *Theobroma cacao* L. *Tree Physiology* 30: 56-67.

Chao A., Jost L., Hsieh T.C., Ma K.H., Sherwin W.B., Rollins L.A. (2015) Expected Shannon Entropy and Shannon Differentiation between Subpopulations for Neutral Genes under the Finite Island Model. *PLoS ONE* 10(6).

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R., Lunter G., Marth G., Sherry S.T., McVean G., Durbin R. y 1000 Genomes Project Analysis Group. (2011) The Variant Call Format and VCFtools. *Bioinformatics* 27 (15): 2156-2158.

Fang W., Meinhardt L.W., Mischke S., Bellato C.M., Motilal L., Zhang D. (2013) Accurate Determination of Genetic Identity for a Single Cacao Bean, Using Molecular Markers with a Nanofluidic System, Ensures Cocoa Authentication. *J Agric Food Chem*

Fluidigm® SNP Genotyping Analysis Software v3. User Guide. © Fluidigm Corporation. All rights reserved. 90p.

Illic K., D. Zhang, X. Wang, R.C. Jones, L. Meinhardt, J. Wang. (2012) Cacao Germplasm Characterization with 48-SNP Genotyping Panel using Fluidigm® SNPtype™ Assays and Dynamic Array™ Integrated Fluidic Circuits. [Poster], P0196. PAG January 14-18. San Diego, California.

Ji K., Zhang D., Motilal L.A., Boccara M., Lachenaud P., Meinhardt L.W. (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 Evol.* 60, 441-453

Kubik C., Honig J., Meyer W., Bonos, S. (2009) Genetic diversity of creeping bentgrass cultivars using SSR markers. *International Turfgrass Society Research Journal* (11).

Lanaud C., Fouet O., Gramacho K., Argout X. *et al.* (2006) A large EST resource for *Theobroma cacao* including cDNAs isolated from various organs and under various biotic and abiotic stresses. In: *Proceedings of the 15th international cocoa research conference*, San Jose, Costa Rica, 185-191.

Liu K., Muse S.V. (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics.* 21(9): 2128-2129.

Manichaikul A., Mychaleckyj J.C., Rich S.S., Daly K., Sale M., Chen W.M. (2010) Robust relationship inference in genome-wide association studies. *Bioinformatics* 26 (22):2867–2873.

Marita J.M., Nienhuis J., Pires J.L., Aitken W.M. (2010) Analysis of Genetic Diversity in *Theobroma cacao* with Emphasis on Witches' Broom Disease Resistance. *Crop Sci.* 41:1305–1316.

Michiels A., Van Den Ende W., Tucker M., Van Riet L., Van Laere A. (2003) Extraction of high-quality genomic DNA from latex-containing plants. *Analytical Biochemistry* 315 (1):85-89.

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

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

Motamayor J.C., Mockaitis K., Schmutz J., Haiminen N., Livingstone D., Cornejo O., Findley S.D., Zheng P., 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* Jun 3; 14:r53

Nei M. (1972) Genetic distance between populations. *American Naturalist.* 106: 283-292.
Peakall R., Smouse P., HUFF D.R. (1995) Evolutionary implications of allozyme and RAPD variation in diploid populations of Buffalograss. *Molecular Ecology* 4 (2): 135-147.

Peakall R., Smouse P.E. (2012) GenAlEx 6.502: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*. Oct 1; 28(19):2537-9.

Quintero C., Zapata Y., Silva E. (2015) Guía Práctica para genotipado de SNPs usando el sistema EP1 y SNPtype Assays de Fluidigm version F_03. Manual interno del Laboratorio de Genética Molecular y Cultivo de Tejidos MGTCL del CIAT.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [URL <http://www.R-project.org/>]

Rambaut A. (2016) FigTree V1.4.2: Tree figure drawing tool. Institute of Evolutionary Biology, University of Edinburgh. [URL: <http://tree.bio.ed.ac.uk/>]

Rehem B.C., Almeida A.F., Corrêa R.X., Gesteira A.S., Yamada M.M., Valle R.R. (2010) Genetic mapping of *Theobroma cacao* (Malvaceae) seedlings of the Parinari series, carriers of the lethal gene Luteus-Pa. *Genetics and Molecular Research* 9 (3): 1775-1784.

Reif J.C., Melchinger A.E., Frisch M. 2005. Genetical and Mathematical Properties of Similarity and Dissimilarity Coefficients Applied in Plant Breeding and Seed Bank Management. *Crop Science* 45:1-7.

Ruiz M., Rouard M., Raboin L.M., Lartaud M., Lagoda P., Courtois B. (2004) TropGENE-DB, a multi-tropical crop information system. *Nucleic Acids Research*, 32: 364-367,

Sereno M.L., Albuquerque P.S.B., Vencovsky R.; Figueira A. (2006) Genetic diversity and natural population structure of cacao (*Theobroma cacao* L.) from the Brazilian Amazon evaluated by microsatellite markers. *Conservation Genetics* 7: 13-24.

Smouse P. E., Peakall R. (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* 82 (5):561-573.

Zapata Y. (2016) Protocolo para extracción de ADN de Cacao V.2. Manual interno del Laboratorio de Genética Molecular y Cultivo de Tejidos MGTCL del CIAT.