UNIVERSIDAD NACIONAL DE COLOMBIA

Estructura genética de poblaciones naturales de palma de aceite (*Elaeis guineensis* Jacq.) procedentes de Angola

Sandra Nereida Barrera Galvis

Universidad Nacional de Colombia

Facultad de Agronomía

Bogotá, Colombia

2011

Estructura genética de poblaciones naturales de palma de aceite (*Elaeis guineensis* Jacq.) procedentes de Angola

Sandra Nereida Barrera Galvis

Tesis o trabajo de investigación presentada(o) como requisito parcial para optar al título de:

Magister en ciencias agrarias con énfasis en genética y fitomejoramiento

Director (a):

Ph.D., Maria Isabel Chacón Sánchez

Codirector (a):

M.Sc., Diana Marcela Arias Moreno

Línea de Investigación:

Genética y fitomejoramiento

Universidad Nacional de Colombia

Facultad de Agronomía

Bogotá, Colombia

2011

Dedico este trabajo a Dios, por ser mi guía, fortaleza y por haberme dado la oportunidad de continuar con mi realización profesional y permitir que esta etapa de mi vida culmine con éxito.

A mi esposo Edison quien, incansablemente y sin dudar, me ha apoyado en cada uno de los retos que en vida he decidió emprender, por su paciencia, compañía, consejo, amor y en fin por poner su tiempo y su vida a disposición de mis sueños.

A mi familia, principalmente a mis padres, Evelia y Héctor, por ser los artífices de la mujer que ahora soy, y por su constante apoyo y compañía.

El hombre encuentra ha Dios detrás de cada puerta que logra abrir.
Albert Einstein

Agradecimientos

A la ***Universidad Nacional de Colombia,*** a todos sus profesores y administrativos quienes con su trabajo aportaron un granito de arena para poner a mi disposición su sabiduría y los recursos para mi superación académica.

Al ***Centro de investigación en palma de aceite (CENIPALMA)*** y especialmente a Diana Arias coordinadora del laboratorio de marcadores moleculares, por el apoyo económico y logístico para realizar este trabajo, por su confianza y al grupo de apoyo técnico de laboratorio Duzley, Diana, Cristian, a las señoras Gladicita y Elena.

A la profesora ***María Isabel Chacón Sánchez***, por orientación en la elaboración de este documento, por brindarme su apoyo, confianza, tiempo, experiencia y sus enseñanzas al decir acompañarme en el desarrollo de este proyecto mi tesis y en general en el transcurso de la maestría.

A los profesores **Gustavo Ligarreto, Gerardo Cayón, Manuel Ruiz** y al doctor **Leonardo Rey**, por sus aportes, sugerencias y comentarios para la obtención de un buen documento resultado de este proyecto.

A María Antonia y Pilar, por su amistad incondicional, durante toda esta etapa de mi vida. A mis compañeros y amigos de universidad Alejandro, Fernan, Isaac, Cesar, Juanito, Deissy, José, Sonia y Diana, por su amistad y confianza, porque en algún momento estuvieron allí compartiendo aquellos momentos de alegría y de angustia, haciendo de este tiempo más grato para mí. A mis hermanas Angela, Patricia y Marcela, por hacer parte importante de mi vida, y ser inspiración en busca de una mejor calidad de vida. A la señora Gloria, a Edwin y Cesar, y a quienes de una u otra manera siempre me han acompañado y apoyado.

Resumen

El objetivo del presente estudio fue evaluar la diversidad y estructura genética de una colección de 455 entradas de palma africana (*Elaeis guineensis* Jacq*)* colectadas en 5 zonas geográficas de la República de Angola y 9 híbridos interespecíficos, con 19 loci microsatélite. Un total de 79 alelos fueron detectados con un promedio de 5.61, trece loci fueron polimórficos con un PIC (índice de información polimórfica) promedio de 0,533. La heterocigosidad esperada promedio (0,561) mostro una variabilidad genética alta respecto a los materiales comerciales. La diferenciación genética entre zonas geográficas fue baja (F_{ST}= 0,028, P<0,01) y entre familias fue moderada (F_{ST}= 0,137, P<0,01), revelando mayor variación entre familias que entre zonas geográficas. La población evaluada no mostro estructura genética debida a la ubicación geográfica. Las topologías permitieron vislumbrar relaciones genéticas de disimilaridad, sin embargo, la mayoría no mostraron tener soporte estadístico.

Palabras clave: palma de aceite, microsatélite, estructura poblacional, diversidad genética

Abstract

The aim of this study was to assess the diversity and genetic structure of a collection of 455 accessions of african oil palm (*Elaeis guineensis* Jacq.), collected in 5 geographic areas of the Republic of Angola and 9 hybrids, with 19 SSR loci. A total of 79 alleles were detected with an average of 5.61, thirteen markers were polymorphic with average polymorphic index content (PIC) of 0.533. The average expected heterozygosity (0.561) showed a high genetic variability compared to commercial materials. Genetic differentiation between geographic areas was low (F_{ST}= 0.028, P<0.01) and between families was moderate (F_{ST}= 0.137, P<0.01) indicating greater genetic variation among families within geographic areas than among geographic areas. The study population showed no genetic structure due to geographic location. The cluster analysis made possible to see genetic relationships of dissimilarity, although most of the groups did not show statistical support.

Keywords: oil palm, microsatellite, population structure, genetic diversity.

Contenido

Pág.

Resumen VII
Lista de figuras XI
Lista de tablas XII
Introducción 1
Objetivos 3
General 3
Específicos 3

1. Capítulo 1. El cultivo de la palma de aceite y las colecciones de germoplasma . 5
1.1 El cultivo de la palma de aceite (*Elaeis guineensis* Jacq.) 5
1.1.1 Generalidades 5
1.1.2 Tipos de palma 6
1.1.3 Usos y productos 7
1.1.4 Importancia Económica 7
1.2 Colecciones genéticas ex situ de palma de aceite (CENIPALMA) 8

2. Capítulo 2. Diversidad y estructura genética de *E. guineensis* de la república de Angola 11
2.1 Marcadores moleculares y estudios de diversidad genética en palma de aceite. Antecedentes 11
2.2 Materiales y métodos 13
2.2.1 Lugar del estudio 13
2.2.2 Material vegetal y zona de estudio 13
2.2.3 Análisis molecular 16
2.2.4 Análisis de datos 18
2.3 Resultados 20
2.3.1 Extracción de ADN genómico 20
2.3.2 Datos faltantes 20
2.3.3 Diversidad alélica 21
2.3.4 Diversidad genética en las zonas geográficas 22
2.3.5 Diversidad genética dentro de familias 23
2.3.6 Estructura genética 25
2.3.7 Relaciones genéticas 29

2.4 Discusión 35
2.4.1 Diversidad alélica 36
2.4.2 Estructura poblacional 38
2.4.3 Relaciones genéticas 40

3. Conclusiones y recomendaciones 42
3.1 Conclusiones 42
3.2 Recomendaciones 43

A. Anexo: Información del grosor de cuesco en las familias y zonas geográficas en este estudio 43

B. Anexo: Perfil de los loci PR01, PR06, PR07 y PR16 de las muestras 127 a 136 ... 44

C. Anexo: Test exacto de desequilibrio de ligamiento. Pares de loci con desequilibrio de ligamiento significativo 45

D. Anexo: F_{ST} y valor-p entre pares de zonas geográficas 46

E. Anexo: F_{ST} y valor-p entre pares de familias 47

F. Anexo: Distancias genéticas entre pares de zonas geográficas 50

G. Anexo: Distancia genética de Nei (1972) entre pares de familias 51

4. Bibliografía 55

Lista de figuras

Pág.

Figura 2- 1: Distribución geográfica de las cinco zonas de colecta de *E. guineensis* en la República de Angola. 15

Figura 2- 2: Geles de agarosa (0,8%). ADN de palma *E. guineensis* extraído con el kit de QIAGEN DNeasy® Plant Min 20

Figura 2- 3: Gráficas para la exploración de la estructura poblacional de las entradas analizadas utilizando el programa Structure. (A) Valores promedio del Ln P(D) (Log likelihood) vs K, (B) ΔK vs K, de acuerdo a Evanno *et al*. (2005). (I) simulación sin información *a priori* de la zona geográfica de las entradas (II) simulación con información *a priori* de la zona geográfica de las entradas evaluadas. 28

Figura 2- 4: Gráfico del K óptimo (ΔK) de las simulaciones del programa Structure incluyendo el grupo de híbridos. A. Sin información *a priori* del lugar de colecta. B. Con información geográfica *a priori* del lugar de colecta. 29

Figura 2- 5: Análisis de coordenadas principales (primeras dos dimensiones) basado en la matriz de distancias genéticas dm^2 para las 5 zonas geográficas incluyendo los híbridos…… 31

Figura 2- 6: Topología basada en el método de agrupamiento Neighbor-joining y la distancia genética de Nei (1972), para las 5 zonas geográficas y el grupo de híbridos usado como grupo externo. 32

Figura 2- 7: Topología basada en el método de agrupamiento Neighbor-joining y la distancia genética de Nei (1972), incluye 44 familias correspondientes a las 5 zonas geográficas colectadas de *E. guineensis* en la República de Angola y un grupo de híbridos……. 33

Figura 2- 8: Topología basada en el método de agrupamiento Neighbor-joining y el índice de disimilaridad, incluye 464 entradas de *E. guineensis* distribuidas en 5 zonas geográficas y un grupo de híbridos usado como grupo externo. Valores de bootstrap superiores a 50. 34

Lista de tablas

Pág.

Tabla 2-1: Información de la estructura de la colecta y número de entradas evaluada……16

Tabla 2- 2: Información y condiciones de amplificación de los loci SSR evaluados en el presente estudio. 17

Tabla 2- 3: Parámetros descriptivos para los loci microsatélites evaluados 22

Tabla 2- 4: Parámetros de variabilidad y diversidad genética para las zonas geográficas evaluadas..... 23

Tabla 2- 5: Variabilidad y diversidad genética dentro de las familias de cada zona geográfica 24

Tabla 2- 6: Coeficientes de diversidad y diferenciación genética de Nei (1987) entre grupos (familias y zonas geográficas)..... 26

Tabla 2- 7: Análisis de varianza molecular (AMOVA F)..... 26

Tabla 2- 8: Estadísticos F (Wright, 1965) de diferenciación genética entre zonas geográficas y familias. 27

Tabla 2- 9. Estimadores de flujo génico con base en el valor promedio del número de migrantes 27

Tabla 2- 0: Distancias genéticas. Valores máximos, mínimos y promedios entre pares de zonas geográficas con inclusión y exclusión del grupo de híbridos..... 30

Tabla 2- 1. Análisis de Coordenadas principales. Porcentaje de variación explicado por las 3 primeras coordenadas principales basado en diferentes distancias genéticas. 31

Introducción

La palma de aceite africana es una monocotiledónea diploide que pertenece a la familia Palmae, género *Elaeis* y especie *E. guineensis*. Es una especie perenne, monoica, alógama (Corley y Tinker, 2003), su genoma es diploide y comprende 16 pares de cromosomas (Singh *et al.,* 2008). La distribución geográfica de la palma de aceite se encuentra en la franja ecuatorial y es ampliamente cultivada por su fruto como fuente de aceite. Se cree que su origen está en las tierras bajas tropicales de África occidental y central extendiéndose de los 16°N en Senegal a los 15°S en Angola. Los procesos de dispersión han sido principalmente realizados por el hombre y la polinización es esencialmente entomófila y anemófila (Corley y Tinker, 2003).

La palma de aceite fue introducida a Colombia como cultivo comercial en 1945 y en la actualidad Colombia se ha posicionado como el primer productor de aceite de palma en Suramérica y el quinto en el mundo (Fedepalma, 2010). Debido a la creciente demanda de aceite de palma y el previsible aumento en los costos del cultivo, es necesario poner a disposición de los cultivadores de palma de aceite un material vegetal con mayor potencial genético, que permita una mayor producción y reducción de costos de manejo del cultivo. Las plantaciones comerciales alrededor de mundo y en Colombia se caracterizan por poseer una base genética muy reducida, circunscrita a la limitada diversidad encontrada en las líneas paternas Deli dura y La Me, Yagambi y AVROS, y ante esta situación, el programa de mejoramiento de Cenipalma ha conformado unas colecciones de germoplasma de *E. guineensis* con el fin de ampliar el acervo genético, para introducir nuevas variantes alélicas que pudieran tener repercusiones favorables en el desempeño agronómico del cultivar en los materiales comerciales de palma de aceite.

Es así como Rey y colaboradores (2004) realizaron colectas de materiales espontáneos en regiones que se postulan como centro de origen. Sin embargo, el conocimiento de la diversidad genética y la estructura poblacional del material colectado permitirá el uso más eficiente del recurso genético existente en el programa de mejoramiento de Cenipalma. La naturaleza perenne del cultivo, implica largos ciclos de selección que involucran amplias áreas experimentales en las cuales se tendría que esperar por lo menos siete años para definir el comportamiento agronómico y seleccionar parentales para el diseño de cruzamientos. Frente a esta situación, el uso de marcadores moleculares podría acelerar los procesos de selección a través de la generación de información que permita determinar la diversidad genética, estructura poblacional y relaciones de parentesco, similitud y origen entre los distintos materiales en diferentes estados de crecimiento y sin el efecto del medio ambiente, permitiendo entender la complementariedad de dichos materiales al momento de planear esquemas de mejoramiento enfocados a dirigir cruces para generar nuevos materiales genéticos comerciales y determinando cómo se integran las colecciones dentro de los programas de mejoramiento y cómo se adaptan las

estrategias de selección. Esta información, junto con las evaluaciones morfoagronómicas, permitirá un uso racional y sistemático de los recursos genéticos del germoplasma (colección núcleo) de Cenipalma y una selección de parentales Dura y Pisífera que logre combinaciones deseables con el objetivo de obtener materiales comerciales altamente productivos, y de esta manera, incrementar la producción de aceite de palma en el país.

Con el propósito de comprender mejor la diversidad y la estructura genética de la población de estudio, y evaluar la hipótesis de diferenciación genética entre el germoplasma colectados en las diferentes zonas geográficas, es necesario tener en cuenta algunos fenómenos que posiblemente pudieron haber afectado la estructura. En este contexto, es importante tener en cuenta las generalidades de los factores de flujo génico como la polinización, la cual es principalmente entomófila atribuida a los gorgojos de la familia Derelominae (*Elaeidobius*) y anemófila que desempeña una parte relativamente pequeña en las polinizaciones. Otro aspecto importante es el efecto antrópico en los procesos de dispersión de semillas, sin embargo, algunos autores describen una importante influencia debida que el fruto de la palma es comestible en fresco en las regiones africanas, proceso que ha influido en la dispersión de la especie. Para las zonas de colecta no se tiene ninguna información de factores que pudieran explicar el flujo génico en la región

El objetivo del presente estudio fue evaluar la diversidad y estructura genética de una colección de 455 entradas de *E. guineensis* de las colecciones de germoplasma de Cenipalma procedentes de cinco zonas contrastantes y representativas del cultivo de la palma en la República de Angola, con 19 marcadores moleculares tipo microsatélite. La información será utilizada en el programa de mejoramiento genético de Cenipalma, el cual será fortalecido por el conocimiento generado en el presente estudio, permitiendo acelerar el proceso de selección de materiales genéticamente distantes para planear cruces adecuados que permitan lograr materiales comerciales altamente productivos y que a su vez, favorezcan la diversidad fenotípica del cultivo.

Objetivos

General

Evaluar la diversidad genética y determinar la estructura genética de una colección de 455 entradas de palma de aceite (*E. guineensis* Jacq.) procedentes de la República de Angola por medio de marcadores microsatélites, con el fin de aportar información a la colección de germoplasma de Cenipalma y su programa de mejoramiento genético.

Específicos

- Cuantificar y describir la diversidad genética de una muestra representativa de *E. guineensis* de la colección de germoplasma procedente de la República de Angola con marcadores tipo microsatélite.

- Determinar la estructura genética de las entradas de *E. guineensis* evaluadas en este estudio.

- Evaluar las relaciones genéticas entre las entradas de *E. guineensis* incluidas en este estudio.

1. Capítulo 1. El cultivo de la palma de aceite y las colecciones de germoplasma

1.1 El cultivo de la palma de aceite (*Elaeis guineensis* Jacq.*)*

1.1.1 Generalidades

La palma de aceite africana es una monocotiledónea diploide perteneciente a la familia Arecaceae, sub-familia Cocoidae, género Elaeis y especie E. guineensis. Es una especie perenne, monoica, alógama (Corley y Tinker, 2003), su genoma es diploide y comprende 16 pares de cromosomas, su contenido de ADN nuclear ha sido estimado en 2C= 3.76 pg por citometría de flujo (Rival *et al.,* 1997) con un tamaño de genoma haploide de 1700 Mpb (Singh *et al.,* 2008).

El origen de la palma de aceite se ubica en las costas del golfo de Guinea en el África occidental y central, distribuida entre los 11°N y 15°S, desde ahí se expandió en forma natural y con intervención humana. La palma de aceite entró a Asia por Java, en 1848, con cuatro palmas plantadas en el Jardín Botánico Bogor en Indonesia y en Brasil en el siglo XVI con la llegada de los esclavos africanos. El crecimiento del la palma de aceite empezó hacia 1910 en el periodo colonial, en África y en el sureste de Asia. Posteriormente, el desarrollo de la palma se dio junto con el establecimiento de organizaciones de investigación y el establecimiento de colecciones de germoplasma con materiales de las regiones vecinas, el cual se convierte en la base genética para los programas de mejoramiento del mundo. En el 2006, la palma de aceite se convierte en el primer cultivo de oleaginosas en producción superando a la soya (Cochard *et al.,* 2009).

Uno de los factores que incrementó significativamente la expansión de la palma fue el descubrimiento de un mejor potencial de producción entre los cruces de origen Africano x Deli (desarrollado en el sureste de Asia). Posteriormente, se implementaron los programas de selección usando solamente recursos genéticos locales, logrando así una ganancia genética rápida. Desde entonces, todos los programas de mejoramiento genético de palma de aceite se han basado en los esquemas de selección recurrente

reciproca (SRR) o selección de familias individuales (SFI), en los cuales se han realizado cruces Deli x orígenes africanos (principalmente La Mé o Congo). Muy pocas palmas fundadoras forman la base genética de estos cruces: cuatro Deli, tres para La Mé y alrededor de 10 para Congo. De esta manera, la base genética eficiente es muy estrecha (Cochard *et al.,* 2009). El material comercial de muchas plantaciones en el mundo proviene de estos pocos parentales, situación que en la actualidad genera inconvenientes para el productor.

Florentino Clases introdujo la palma africana de aceite a Colombia en 1932. Estas primeras palmas fueron sembradas con fines ornamentales en lugares públicos de algunos pueblos de la región amazónica y en la Estación Agrícola de Palmira, en el Valle del Cauca. Sin embargo, el cultivo comercial de esta planta oleaginosa sólo comenzó años después, en 1945 cuando la United Fruit Company estableció una plantación en la zona bananera del departamento del Magdalena, con plantas procedentes de Honduras. La expansión del cultivo ha variado enormemente en cada uno de los países en donde se ha sembrado la palma (Bernal, 2001).

1.1.2 Tipos de palma

Los tipos de palma africana más relevantes se establecen de acuerdo al grosor del cuesco o endocarpio del fruto, característica relacionada con la producción de aceite. También existe otra clasificación de acuerdo al color del fruto.

Según el grosor del cuesco:

- Pisífera (PxP, TxP, TxT). Son palmas cuyos frutos no tienen cuesco, sino un cartílago blando. Los árboles se caracterizan por tener un gran porte y una alta producción de flores femeninas que generalmente no logran culminar la formación de frutos.

- Dura (DxD, TxD, TxT). Este tipo de plantas se cultivó comercialmente en el mundo entero hasta finales de la década de los sesenta. Su principal característica es la presencia de un gran cuesco de 2 a 8 mm de espesor en los frutos.

- Ténera (DxP, TxD, TxT). Por ser un híbrido proveniente del cruzamiento de Dura por Pisífera, el cuesco del fruto es delgado y la producción de pulpa bastante mayor. Por ende, el contenido de aceite es significativamente más abundante. En un corte transversal de un fruto de Ténera, se observa un anillo de fibras oscuras adyacentes al cuesco que le es característico (Bernal, 2001).

Según el color del fruto:

- Frutos nigrescens: son los más comunes, se caracterizan por el color violeta oscuro a negro antes de la maduración y rojo ladrillo en estado de madurez.

- Frutos virescens: son de color verde oliva, que cambia a un claro anaranjado-rojizo cuando maduran.

- Frutos albescens: son frutos poco frecuentes, éstos almacenan muy poco o nada de caroteno en el mesocarpio por esto se observan de color blanco o amarillo pálido (Corley y Tinker, 2003).

1.1.3 Usos y productos

La palma de aceite juega un papel muy importante en la economía del país debido a su cultivo para la explotación de aceite de su fruto.

Los productos primarios del cultivo de la palma de aceite son de tres tipos: el aceite que se extrae de la pulpa de la fruta, el que se obtiene de las almendras y la torta que queda del proceso de extracción del aceite de las almendras. El aceite producido en mayor cantidad y más utilizado es el extraído de la pulpa del fruto.

El aceite de palma crudo es utilizado para la fabricación de concentrados animales y para la obtención de ácidos grasos mediante desdoblamiento por hidrólisis. El aceite de palma se fracciona en oleína (liquida) y estearina (sólida), de las que se produce el 56% de los aceites y grasas que se consumen en Colombia (Fedepalma, 2007).

El aceite comestible es materia prima en la industria de alimentos y concentrados para animales. El aceite no comestible es usado en la industria de combustibles, velas, cosméticos y tintas. Otros compuestos del aceite de palma son los oleoquímicos (Fedepalma, 2007) y constituyentes menores como el caroteno vitamina E y esteroles (Corley y Tinker, 2003).

1.1.4 Importancia Económica

La palma de aceite tiene gran importancia económica a nivel mundial debido al alto contenido de aceites en sus frutos, contenido en el mesocarpio (aceite de palma) y en la almendra (aceite de almendra).

En el mundo, el área total en producción de palma de aceite se incrementó pasando de 9.389 millones de hectáreas (mha) en el 2005 a 12.117 mha en el 2009. Indonesia es el país con mayor área en producción registrando 5.350 mha en el 2009 y Malasia posee el mejor rendimiento de aceite de palma crudo obteniendo 4,39 toneladas por hectárea, seguido de Costa Rica con 4,15 ton/hect. (Fedepalma, 2010).

Colombia ocupa el quinto lugar a nivel mundial en área en producción registrando 236 mha para el 2009 y el primero en sur América (Fedepalma, 2010). La producción se encuentra en algunas zonas del país, identificadas por tener las condiciones agroecológicas adecuadas para el cultivo, las cuales se distribuyen en los departamentos Cesar, Atlántico, Guajira, Córdoba, Meta, Casanare, Caquetá, Magdalena, Nariño,

Cauca, Santander, Norte de Santander y Bolívar, conformando las cuatro zonas palmeras de Colombia, las cuales han cambiado su participación en área sembrada debido a la incidencia de la enfermedad de Pudrición del Cogollo (PC) (Fedepalma, 2010).

En los últimos años ha aumentado considerablemente el número de hectáreas sembradas en el país, pasando de 161.210 ha en 2001 a 360.537 ha en el 2009. El área en producción aumentó 7,7% en el 2009, la producción de aceite de palma crudo acumulada en el tercer trimestre (enero/septiembre) del 2009 fue de 620.950 toneladas, cayendo 1,8% por debajo de la producción del mismo periodo en el 2008, lo cual significa un cambio en la tendencia de crecimiento que se venía presentado desde el año 2002. En el rendimiento promedio de aceite de palma por hectárea desde el 2004 cuando se alcanzó 4,1 toneladas por hectárea (t/ha), se ha observado una tendencia al descenso pasando de 3,5 t/ha en el 2008 a 3,4 t/ha en el 2009. Así mismo, la producción nacional de aceite de palma crudo redujo su crecimiento promedio anual de 6% en 2008 a 3,2% en 2009 (Fedepalma, 2010). Esta tendencia afecta de manera importante la economía del país ya que el producto interno bruto (PIB) en el tercer trimestre de 2010 presentó una reducción de la producción respecto al 2009 de -4,0% para palma y soya (Agronet, 2010). Estos descensos son atribuidos a algunos factores como la propagación de enfermedades especialmente la alta incidencia de PC y problemas fitosanitarios en algunas zonas, el cambio en la estructura de edades del área productiva, el ingreso de nuevos palmicultores que no han afianzado el manejo eficiente de sus cultivos, y las condiciones climáticas adversas (Fedepalma, 2010; Agronet, 2010).

El consumo aparente de aceite de palma en el mundo aumentó 27,4% y en el país aumentó 32,2% en 2009, principalmente debido a la demanda para fabricación de biodiesel, que absorbió 153.496 toneladas (Fedepalma, 2010). De acuerdo a los estimativos de crecimiento poblacional para el año 2020, se espera que el consumo mundial de aceite de palma sea de 37,9 millones de toneladas, representando una tasa de crecimiento promedio anual de 3,3%. Dada esta dinámica de crecimiento del aceite de palma, la metas nacionales en área sembrada para el 2020 son de 996.296 ha con una producción de 3.383.892 toneladas, metas que solo se podrán cumplir con materiales comerciales de alta producción y rendimiento.

1.2 Colecciones genéticas ex situ de palma de aceite (CENIPALMA)

Los bancos de germoplasma en el mundo son un recurso creado para ampliar la base genética de los cultivos y conservar buena parte de la variabilidad genética existente, caracterizando agronómica y morfológicamente materiales con uso potencial en fitomejoramiento, debido a que las líneas paternas de las plantaciones comerciales, no poseen algunos caracteres genéticos de interés agronómico y tienen una limitada variabilidad (Rey *et al.*, 2004).

Cenipalma en el año 2002 inició la conformación de una colección de germoplasma de palma de aceite en el centro experimental “La Vizcaína” ubicado en zonas rurales de Barrancabermeja. La colección genética de Cenipalma cuenta con dos especies de interés económico: palma africana (*Elaeis guineensis* Jacq.) y palma americana o noli (*Elaeis oleifera* H.B.K.).

La palma americana está representada en la colección por entradas (una entrada está definida como una palma que conforma la colección de germoplasma) colectadas en el trapecio amazónico entre Leticia y San Juan de Atacuarí constituyendo 13 poblaciones, 92 familias y 137 entradas, y por 10 poblaciones y 250 entradas provenientes de la Costa Norte, del Valle del Río Sinú, la Amazonía y el Magdalena medio (Rey *et al.,* 2003, 2004). La palma americana es considerada en la colección de germoplasma como una fuente valiosa de variabilidad genética en los programas de fitomejoramiento para solucionar algunos de los problemas fitosanitarios y de calidad de aceite que presenta la palma de aceite comercial, y que se están solucionando con el uso de híbridos interespecíficos OxG (Meunier, 1991 citado por Rey *et al.,* 2004). Rey y colaboradores en 2004 publicaron la caracterización realizada para componentes de producción y calidad de aceite, vitamina E y otros metabólitos, de una muestra representativa de la colecta donde se evidenció una variabilidad interesante.

La palma de aceite *E. guineensis* es de origen africano y es en África donde se encuentra la mayor variabilidad genética de esta especie, motivo por el cual Cenipalma, en convenio con el Instituto Nacional del Café (Inca) de la República de Angola, hicieron colectas en zonas contrastantes y representativas en las cuales la palma de aceite se encuentra en forma espontánea. Esto, con el fin de obtener buena parte de la variabilidad genética existente en este país para mejorar las bases genéticas de las poblaciones actuales, y de esta manera obtener nuevos materiales comerciales con bases genéticas amplias que le confieran mayor estabilidad ante los factores ambientales adversos. Esta especie está representada en la colección por material genético de Angola distribuido en 20 poblaciones, 38 familias y 4876 entradas, y de Camerún distribuido en 18 poblaciones, 72 familias y 3500 entradas.

La colección genética palma cuenta con material genético mejorado Dura, Pisífera, híbridos interespecíficos comerciales y materiales genéticos utilizados como parentales en los diferentes cruces. Para caracterizar el material genético existente en la colección genética de palma de aceite, Cenipalma ha implementado a través de diferentes líneas de investigación estudios genéticos, bioquímicos, moleculares, fisiológicos y de reacción a plagas y enfermedades.

El germoplasma de la República de Angola fue establecido en campo bajo una estructura poblacional que ha permitido identificar después de tres años de evaluaciones morfoagronómicas, variaciones muy importantes entre las características de interés agroindustrial, como las altas producciones de racimos de fruta fresca (RFF) (40 ton/ha/año) y contenido de aceite en pulpa (CAP) (7,5 tonpalma/año), dado por altos porcentajes de pulpa en fruto (75% en familias duras y 95% en familias teneras), bajas tasas de crecimiento (30cm/palma año), estípite largo (40cm) para facilitar la cosecha, altos contenidos de carotenos, tocoferoles y tocotrienoles. Estas características han permitido establecer líneas de mejoramiento por características y combinaciones de ellas, es así que se tienen familias de alta producción de aceite y lento crecimiento, familias con alto contenido de ácidos grasos insaturados similares al híbrido OxG es decir, alto acido oleico, la asistencia en la selección dada por los marcadores

moleculares busca dentro de estas familias evidenciar la variabilidad genética existente con el fin de garantizar un avance genético significativo y determinar los ciclos de selección recurrente necesarios para agotar la varianza aditiva (Comunicación personal Rey, 2011).

.

2. Capítulo 2. Diversidad y estructura genética de *E. guineensis* de la república de Angola

2.1 Marcadores moleculares y estudios de diversidad genética en palma de aceite. Antecedentes

Los marcadores moleculares son etiquetas de ADN utilizadas para identificar y evaluar los genes de interés en una planta. Son neutros con respecto al medio ambiente y pueden detectarse en cada etapa del crecimiento de la planta, particularmente útiles en especies perennes con tiempo generacional largo, ya que permiten pruebas de selección temprana en material mejorado (Billote *et al.,* 2007).

Los estudios de diversidad genética han dado un entendimiento general de la estructura genética, a través de uso de marcadores isoenzimáticos (Hayati *et al.,* 2004) y marcadores moleculares.

Inicialmente los marcadores moleculares fueron utilizados en Malasia y Reino Unido a principio de la década de los noventa en estudios de diversidad genética utilizando RFLP (*Restriction Fragment length polymorphism*) (Cheah *et al.,* 1991; Barcelos *et al.,* 2002; Maizura *et al.,* 2006), RAPD (*Random Amplification of polymorphic DNA)* (Shah *et al.,* 1994) y AFLP (*Amplified Fragment length polymorphism*) (Purba *et al.,* 2000; Singh *et al.,*1998; Barcelos *et al.,* 2002), en pruebas de genotipificación de entradas de *E. guineensis* y *E. oleifera* (Jack *et al.,* 1995; Mayes *et al*., 1996), en la búsqueda de marcadores de genes (ADNc) comprometidos en el desarrollo floral (Singh y Cheah, 2000), en el control genético de enzimas de la biosíntesis de ácidos grasos (Rashid y Shan, 1996; Shah y Cha, 2000), en la búsqueda de variantes somaclonales (Paranjothy *et al.,* 1993; Chowdhury, 1993; Rival *et al.*, 1998) y en la detección de tolerancia a la pudrición del cogollo (Ochoa *et al.,* 1997). Se han realizado trabajos de cartografía genética con marcadores moleculares inicialmente con RFLP (Mayes *et al.,* 1997), RAPD (Moretzsohn *et al.,* 2000) y posteriormente con SSR (*Sequence Simple Repeat*) (Billote *et al.,* 2005).

Estos estudios han demostrado la utilidad de los marcadores moleculares en la

evaluación de la diversidad genética, para el manejo de cultivares y establecimiento de colecciones núcleo para la conservación *ex situ* del acervo genético de la palma de aceite (Singh *et al.,* 2008).

Los marcadores microsatélites (*Simple Sequense Repeats*: SSR) son arreglos de secuencias de nucleótidos repetidos consecutivamente en el genoma, se presentan en motivos di, tri o tetranucleotidos y están presentes en todos los organismos (Ferreira y Grattapaglia, 1998). Estos marcadores han sido ampliamente utilizados en muchas especies por ser considerados ideales en trabajos de identificación varietal, análisis de pedigrí, cartografía de genomas y detección de QTL para el mejoramiento asistido por marcadores, mapeo genético, medidas de diversidad y estructura poblacional, debido a que son altamente variables y permiten distinguir entre cultivares estrechamente relacionados (Mohammadi y Prasanna, 2003; Billote *et al.,* 2007). En palma de aceite, de todos los marcadores moleculares existentes para la evaluación de germoplasma, los SSR parecen ser los más promisorios para el entendimiento de la estructura genética poblacional y el flujo génico debido a que presentan herencia mendeliana (Billote *et al.,* 2007), alta reproducibilidad, co-dominancia, naturaleza multialélica, son abundantes y ampliamente distribuidos en el genoma, son de detección simple por medio de PCR y son fácilmente transferibles entre laboratorios (Singh *et al.,* 2008).

El uso de SSR para evaluar germoplasma de *E. guineensis* ha sido descrito por Billote y colaboradores quienes en el 2001 reportaron 20 secuencias de marcadores SSR y posteriormente en el 2005 publicaron el uso de marcadores SSR en la construcción de un mapa de ligamiento de alta densidad basado en microsatélites. En este estudio se utilizaron 390 marcadores SSR y AFLP permitiendo la construcción de un mapa de ligamiento constituido por 288 SSR, 688 AFLP y el locus del gen *Sh.* Los marcadores se distribuyeron en 16 grupos de ligamiento con una longitud de 1,743 cM. Este trabajo constituye el primer mapa con los grupos de ligamiento igual al número de cromosomas y un avance importante para el análisis de QTL (*Quantitative trait loci*).

En el 2005 Montoya y colaboradores estimaron la diversidad genética e infirieron la estructura genética de 48 entradas procedentes de Angola pertenecientes a 6 poblaciones con 20 combinaciones de marcadores SSR. En este estudio se identificaron 69 alelos, una diversidad genética estimada en términos de heterocigosidad esperada (H_e) de 0,45 y un índice de endogamia global F_{IS} de -0,0604, presentando un leve exceso de heterocigotos. La diferenciación genética entre poblaciones fue baja (F_{ST}= 0,0836) y el flujo génico estimado entre poblaciones fue alto (Nm= 2,73).

Bakoumé y colaboradores (2007) determinaron la diversidad alélica dentro de 4 poblaciones naturales que representan 10 países africanos, 3 materiales mejorados y 1 material semi-silvestre mediante 16 loci SSR. En este estudio se detectaron 209 alelos. Se encontraron alelos poco comunes en poblaciones de áreas con clima seco, sin tener en cuenta el país. Los alelos excepcionales DeliMPOB, eran poco comunes en poblaciones naturales de palma de aceite, presentaban una reducción debido a los muchos años de selección. El número medio de alelos osciló entre 1,1 y 1,67. El número efectivo de alelos (A_e) osciló entre 1,1 y 4,7.

Singh y colaboradores (2008), usaron etiquetas de secuencias expresadas (ESTs) de palma de aceite para buscar marcadores SSR. Diez marcadores ESTs-SSR fueron

desarrollados y usados para evaluar 76 muestras de palma de aceite de 7 países de África y una población estándar de Deli dura. El número promedio de alelos observados y A_e fue de 2,56 y 1,84, respectivamente. Los marcadores ESTs-SSR fueron polimórficos presentando un PIC (contenido de información polimórfica) de 0,53. La diferenciación genética (F_{ST}= 0,249) entre poblaciones indicó un alto nivel de divergencia genética. El análisis UPGMA (*unweighted pair-group method with arithmetic mean*) reveló una fuerte asociación entre distancia genética y localización geográfica de la poblaciones estudiadas. El germoplasma evaluado presentó alta diversidad comparado con Deli dura. Las poblaciones de Nigeria, Congo y Camerún mostraron alta diversidad. Los resultados del estudio mostraron el potencial de los ESTs-SSR, en estudios de diversidad, estructura poblacional y análisis de filogenia.

Cochard y colaboradores (2009) analizaron 318 individuos de 26 orígenes y 8 países con 14 marcadores microsatélites. Con el propósito de describir la estructura de los recursos genéticos involucrados en la historia del mejoramiento de la palma en relación a los orígenes africanos subespontáneos. Los datos revelaron dos grupos originales, el grupo I formado por África occidental y “Benin-Nigeria-Camerum-Congo-Angola” formando el grupo II. El tercer grupo Deli es derivado del grupo II, resultado de la selección artificial. Con los resultados obtenidos se propuso una estrategia de selección basada en la complementariedad de los grupos genéticos africanos, donde fueron sugeridos los cruces (Deli x grupo II) x Grupo I.

2.2 Materiales y métodos

2.2.1 Lugar del estudio

El análisis de las muestras se realizó en el laboratorio de biología molecular (LBM) de Cenipalma (Bogotá).

2.2.2 Material vegetal y zona de estudio

Las 455 entradas (una entrada está definida como una palma que conforma la colección de germoplasma) utilizadas en este estudio representan una muestra significativa (10%) del germoplasma colectado en Angola y establecido en el Centro Experimental La Vizcaína ubicado en el departamento de Santander.

Este recuso genético de *E. guineensis* fue colectado por Rey *et al.,* (2004) en la República de Angola bajo el marco de un convenio entre el Instituto Nacional del Café (Inca) de la República de Angola y Cenipalma, con el fin de obtener buena parte de la variabilidad genética existente en este país para mejorar la base genética de las poblaciones comerciales actuales. La colecta se realizó en cinco regiones naturales consideradas zonas contrastantes y representativas del cultivo de palma de aceite *E. guineensis*, donde se encontró como poblaciones espontanéas. Al momento de la

colecta se realizó una caracterización *in situ*, de 17 características morfoagronómicas entre las cuales están el contenido de la pulpa en el fruto, peso medio del racimo, forma del racimo, presencia de espinas, color de frutos, altura de la palma, entre otras características. Se presentaron diferencias significativas para el contenido de mesocarpio en fruto, entre las zonas de colecta con una variación entre 31 y 67% (tipo Dura) y entre 60 y 77% (tipo Ténera), este criterio fue tenido en cuenta al realizar la colecta (Rey *et al,* 2004). De las zonas analizadas en este estudio se desconoce información acerca del aprovechamiento y actividades culturales que involucren el transporte de semillas de una región a otra.

Según Rey *et al.* (2004), las zonas se encuentran localizadas entre los 5,5° y 12° de latitud sur, entre la zona desde el litoral hasta la montaña, con regímenes climáticos contrastantes, así (Figura 2-1):

- Caixito (1). Presenta alturas entre los 250 y 400 msnm, con pluviometría bimodal entre 600 y 900 mm al año y temperaturas entre 23 y 26°C, clima subhúmedo seco a semiárido de litoral bajo. La vegetación es de sabana arbóreo-arbustiva y suelos ferralíticos.

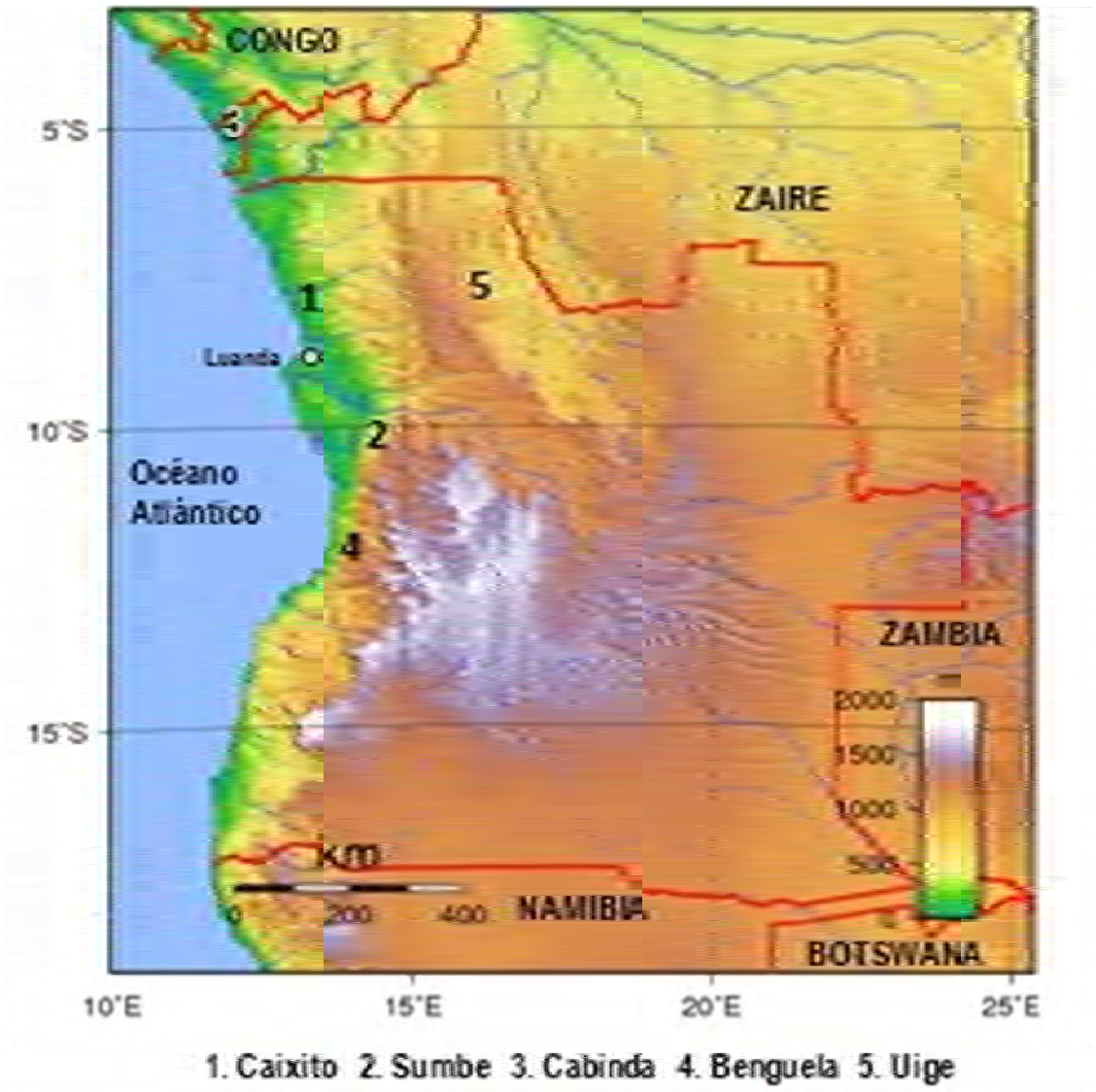
- Sumbe (2). Comprendida entre los 0 y 250 msnm, precipitación monomodal entre 300 y 600 mm, temperaturas entre 24 y 26°C, clima árido a semiárido, suelos de vega sedimentarios y arídicos tropicales, pardo anaranjados, con vegetación de sabana arbórea.

- Cabinda (3). Zona de sublitoral, muy accidentada con alturas entre los 400 y 1.000 msnm, clima húmedo, con precipitaciones anuales entre 900 y 1.400 mm y temperaturas entre 22 y 25°C, vegetación densa y siempre verde, intercalada con pequeñas sabanas arbustivas – arbóreas, con suelos ferralíticos. Zona de colecta en el extremo norte de Angola.

- Benguela (4). Comprendida entre los 0 y 250 msnm, precipitación monomodal entre 300 y 600 mm, temperaturas entre 24 y 26°C, clima árido a semiárido, suelos de vega sedimentarios y arídicos tropicales, pardo anaranjados, con vegetación de sabana arbórea. Zona de colecta en el extremo sur de Angola.

- Uige (5). Presenta alturas sobre el nivel del mar entre 600 y 900 metros, con clima húmedo megatérmico, con dos estaciones de lluvias bien definidas, precipitaciones pluviales entre 1.300 y 1.400 mm al año y temperaturas entre 23 y 24°C, la vegetación es de sabana arbórea, con galerías de selva densa y húmeda y suelos ferralíticos.

Figura 2-1: Distribución geográfica de las cinco zonas de colecta de *E. guineensis* en la República de Angola.



Rey *et al.* (2004) realizaron la colecta bajo un diseño genético estructural el cual posee tres niveles jerárquicos. El primer nivel conformado por cinco zonas geográficas, el segundo por las familias colectadas cada zona y el tercero los individuos (frutos) de cada racimo. En la palma de aceite, cada racimo es una familia polihíbrida de medios hermanos formados a través de polinización abierta, donde se conoce solamente la identidad del progenitor femenino, ya que el polen proviene de cualquier palma circundante (Tabla 2-1). Las 455 entradas analizadas incluyen material genético tipo Dura y Ténera (Anexo A). Se incluyó en el estudio un grupo de 9 híbridos interespecíficos (HB) (*Elaeis oleifera* x *Elaeis guineensis*) para fines de comparación en los análisis de distancias genéticas.

Tabla 2-1: Información de la estructura de la colecta y número de entradas evaluadas.

ZONA GEOGRÁFICA	FAMILIAS	Muestras Vegetales (cód. Laboratorio)	TOTAL INDIVIDUOS
1. Caixito	9	127 al 214	88
2. Sumbe	9	215 al 315	100
3. Cabinda	11	317 al 435	119
4. Benguela	6	436 al 494	58
5. Uige	10	496 al 585	90
6. HB	-	1 al 9	9
TOTAL	**45**	**-**	**464**

Fuente: Programa de Biología y Mejoramiento Genético de Cenipalma

2.2.3 Análisis molecular

- **Extracción de ADN**

La extracción de ADN genómico total se realizó a partir de tejido foliar joven de palma de aceite (*E. guineensis* Jacq.) utilizando el kit de QIAGEN DNeasy® Plant Mini y siguiendo las recomendaciones del fabricante.
La cuantificación del ADN extraído se realizó mediante un espectrofotómetro y la calidad del ADN se verificó utilizando el método comparativo de electroforesis en gel de agarosa.

- **Amplificación de microsatélites (SSR)**

Para desarrollar la técnica de marcadores moleculares microsatélites se seleccionaron 19 loci y se amplificaron bajo las condiciones reportadas por Billote *et al.* (2001) y Singh *et al.* (2008) (Tabla 2-2).

Tabla 2- 2: Información y condiciones de amplificación de los loci SSR evaluados en el presente estudio.

CÓD.	LOCUS	A	TA	MOTIVO	Secuencia del Cebador (5'-3')	
					Forward	Reverse
PR01	mEgCIR0008	6	195–220	(GA)18	CGGAAAGAGGGAAGATG	ACCTTGATGATTGATGTGA
PR02	mEgCIR0009	4	162–204	(GA)20	CAGTCTTTAAGTACGGCTATGAT	GAATTTTTAGTTCAACCAGGTAGA
PR03	mEgCIR0018	11	158–177	(GA)18	CCTTATTTTCTTTGCTTACC	TTCTATTTTATTTTCTTCCT
PR04	mEgCIR0046	5	198–262	(GA)19	AGCCTTAGTATTTTGTTGAT	CCTCTGATTTGTCCTTTTGG
PR05	mEgCIR0067	8	135–187	(GA)17	TACACAACCCATGCACAT	AAAAACATCCAGAAATAAAA
PR06	mEgCIR0219	7	205–233	(GA)17	TTTGCTCGGCGGATACAT	CTCACTGGCCTCTTTCTT
PR07	mEgCIR0221	7	195–213	(GA)11	TGCCATGTTCCAGAGAGC	TTCAGATTTTTCCGACTTC
PR10	mEgCIR0304	1	275–281	(GT)4(GC)2(GT)2	CCACAAACAATCCAAGCAAGT	TGGCATACACGAAAGCATAA
PR11	mEgCIR0326	1	131–147	(GT)9	GCTAACCACAGGCAAAAACA	AAGCCGCACTAACATACACATC
PR12	mEgCIR0332	4	269–281	(GT)8	ATTTCGTAAGGTGGGTGT	CCTCCAAACTCCTCTGT
PR13	mEgCIR0350	4	80–102	(GT)8	AAATCCTAAATCCTAAACTC	TCTACCTGTACTGGTGACAA
PR16	mEgCIR1772	12	293–297	(GT)22	ACCTTGTATTAGTTTGTCCA	CTTCCATTGTCTCATTATTCTCTTA
PR23	sEg00066	10	192-215	$(AT)^9$	TTGCTCCAACTGACTGATGC	ACATTCCAGATCCCAGCAAG
PR24	sEg00067	4	150- 170	(TGTA)6	GTCAGCCCGTAGAAGATTGC	CTTTCGGATAGCCAAAACGA
PR27	sEg00090	8	200–210	(AT)9	TATGCGGGTGATCAAGTGAA	CCACCATGGTTCTCAGGAAA
PR28	sEg00125	4	152- 170	(GCG)6	TACCCTTTTCCCTCCCTCCATA	CATCATCTCCGTTGCCAGTATT
PR29	sEg00126	2	212–215	(CGC)7	CCGTCTCAAAAGCCCTAAAC	TTGTTGTCCCACTCCCTCTT
PR31	sEg00140	5	210–226	(GA)10	AAGTGAGACGGTGGATTTGG	GTTCCAGTTGTCCTCGCATT
PR37	mEgCIR3282	12	232-272	(GA)20	GTAACAGCATCCACACTAAC	GCAGGACAGGAGTAATGAGT

A: alelos detectados por marcador, TA: Rango del tamaño de los alelos en pares de bases. Fuente: Billote *et al.* (2001) y Singh *et al.* (2008)

Para la amplificación de las muestras se optimizaron las condiciones de amplificación reportadas por los autores para algunos de los loci. En general, las reacciones de PCR se llevaron a cabo en un volumen final de 15 µl que contenían 20 a 30 ng de ADN, 1X de buffer (100 mM Tris-HCl pH 8,8, 500 mM KCl) para PCR, 2 ó 1,5 mM $MgCl_2$, 0,2 mM dNTPs, 0,15 uM de cada marcador ("forward" y "reverse") y 0,5; 1 ó 1,5 µl de *Taq* polimerasa dependiendo del marcador. Las reacciones de PCR se llevaron a cabo en un termociclador iCycler 1000 de BioRad. Las temperaturas utilizadas para los loci fueron 2 minutos a 94ºC, seguidas de 30 ciclos a 94ºC por 30 s, 1 minutos a 52ºC y 30 s a 72ºC y 10 minutos a 72 ºC de extensión.

El producto de amplificación fue detectado en geles denaturantes de poliacrilamida 6%, urea 5M, en cámaras de secuenciación BioRad y visualizado mediante tinción en nitrato de plata y revelado con carbonato de sodio (Anexo B). El marcador de peso molecular utilizado fue de 10 pares de bases (pb) (invitrogen, cat. No. 10821-015), el cual determinó fragmentos entre 10 pb y 330 pb, rango en el cual se encuentran todos los fragmentos amplificados.

Para la electroforesis, se llevó a cabo una fase de precorrido de 20 minutos, posteriormente el corrido electroforético se realizó durante 1 hora y 30 minutos a 80W. El tiempo de corrida difirió según el tamaño del fragmento. Se generó registro fotográfico para cada gel y la lectura de los alelos fue realizada visualmente y registrada en matrices digitales.

2.2.4 Análisis de datos

Las matrices fueron codificadas para ser analizadas en los diferentes programas de computador.

Los loci monomórficos no fueron incluidos en la descripción y cuantificación de la diversidad genética, teniendo en cuenta los siguientes parámetros: número de individuos evaluados por locus (N), número de alelos por locus (A), número de alelos efectivos por locus (A_e), heterocigosidad observada (H_o), heterocigosidad esperada (H_e) de Nei 1978, alelos privados por zona geográfica (APP), porcentaje de loci polimórficos (%P), riqueza alélica (R_a) Índice de fijación (F) y número total de alelos, estimados con los algoritmos incluidos en el programa GenALEX 6.1 (Peakall y Smouse, 2006). Se estimó el contenido de información polimórfica promedio (PIC), el cual mide la diversidad alélica de un locus, estimando y comparando el poder de discriminación de los marcadores moleculares (Botstein *et al.,* 1980). El test exacto de equilibrio Hardy-Weinberg (EHW) (Guo y Thompson, 1992) fue realizado tomando como unidades de análisis las zonas geográficas basado en 1000 permutaciones. Se evaluó el desequilibrio de ligamiento entre pares de loci basado en 1000 permutaciones utilizando el programa Arlequín versión 3.1.1 (Excoffier, 2006).

La estructura y diferenciación genética entre las zonas geográficas se calculó mediante los índices F (F_{ST}, F_{IS} y F_{IT}) y sus diversas maneras de estimación (Nei, 1987; Wright 1946, 1951 y 1965). El test estadístico de significancia para estos índices se basó en 30000 cadenas de Markov de genotipos multilocus entre pares de zonas geográficas y familias utilizando el programa Arlequín versión 3.1.1 (Excoffier, 2006). El parámetro Nm (número de migrantes) se estimó a partir del estadístico F_{ST} (Nei, 1987). Para evaluar la estructura genética de la muestra se realizó además, en un contexto jerárquico, un Análisis Molecular de la Varianza (AMOVA) a partir del cual se calcularon los estadísticos F_{ST}, cuya significancia estadística se obtuvo a partir de 999 permutaciones, incluyendo como fuentes de variación los individuos, las familias y las zonas geográficas, mediante el uso del programa GenALEX 6.1 (Peakall y Smouse, 2006) y Arlequín versión 3.1.1 (Excoffier, 2006).

Incluso en ausencia de discontinuidades geográficas extremas en el flujo génico, la conectividad limitada entre subpoblaciones distribuidas continuamente podrían producir patrones de aislamiento por distancia (IBD), donde la divergencia subpoblacional aumenta con la distancia geográfica (Wright, 1943). Para evaluar la posible influencia de procesos de IBD, se realizó un análisis usando el coeficiente de Mantel para probar la hipótesis nula de no asociación entre divergencia genética y distancia geográfica. Para este análisis, se usaron matrices de F_{ST} y F_{ST} linearizado (Slatkin, 1995) como medida de

divergencia genética entre pares de zonas geográficas estimadas con el programa Arlequín versión 3.1.1 (Excoffier, 2006), y una matriz de distancias geográficas en kilómetros entre los lugares de colecta estimada usando distancias euclidianas. La significancia del test se realizó a partir de 10000 permutaciones con el programa IBMWS versión 3.16 (Jensen *et al*., 2005)

Para explorar la estructura poblacional sin asumir poblaciones *a priori* se utilizó el programa Structure 2.3 (Pritchard *et al.,* 2010), el cual implementa un modelo bayesiano basado en un algoritmo de agrupamiento. Sus aplicaciones incluyen la deducción de la presencia de poblaciones distintas, para asignar individuos a las poblaciones. En este análisis se intenta estimar el número más probable de grupos (K) en la muestra minimizando las desviaciones al equilibrio Hardy-Weinberg y al equilibrio de ligamiento dentro de cada grupo y asignando los individuos a un número K de poblaciones. El método emplea un algoritmo Monte Carlo de cadenas de Markov (MCCM) para estimar las frecuencias alélicas en cada uno de los K, y para cada individuo, la proporción de su genoma derivado de cada población (q_K). El análisis se realizó inicialmente sin asumir poblaciones *a priori* y posteriormente incluyendo información del origen de colecta. Los parámetros de corrida fueron 100000 de burnin y 2000000 de iteraciones de MCCM, bajo un modelo de mezcla y de frecuencias alélicas correlacionadas. También se calculó el ΔK según Evanno *et al*. (2005) utilizando 5 iteraciones y explorando de 2 a 20 K. Finalmente se realizó un análisis de coordenadas principales (ACoP) a partir de 1000 permutaciones utilizando como unidades de análisis los individuos con el programa DARwin versión 5.0. 158 (Perrier y Jacquemoud-Collet, 2009) y posteriormente utilizando las zonas geográficas como unidad de análisis con el programa GenAIEX 6.1 (Peakall y Smouse, 2006).

Para evaluar las relaciones genéticas entre individuos y zonas geográficas se utilizaron dos tipos de análisis descriptivos a partir de la estimación de distancias genéticas. Las distancias genéticas son una medida de disimilaridad entre los individuos o grupos evaluados de la misma o diferentes especies. Se estimaron diferentes distancias entre las que están la distancia genética estándar de Nei (DSG) (1972), distancia cuadrática media (ASD) (Goldstein, 1995b y Slatkin, 1995) y la distancia $d\mu^2$ (Goldstein 1995a). Las últimas dos distancias mencionadas asumen modelos mutacionales para loci microsatélites. También se estimó la distancia geométrica de Cavalli-Sforza y Edwards (1967) usando el programa MSA 4.05 (Dieringer y Schlötterer, 2003). El análisis coordenadas principales (ACoP) se realizó utilizando como unidades de análisis los individuos para estimar una matriz de disimilaridad calculada con el índice de concordancia simple basado en 1000 permutaciones con el programa DARwin versión 5.0.158 (Perrier y Jacquemoud-Collet, 2009) y posteriormente se realizó nuevamente utilizando las zonas geográficas como unidades de análisis con el programa GenAIEX 6.1 (Peakall y Smouse, 2006). El segundo fue la reconstrucción de topologías utilizando el método de agrupamiento Neighbour-joining usando un bootstrap de 1000 réplicas con el programa Phylip versión 3.69 (Felsenstein, 2005). La significancia de los nodos en las topologías solo fue tenida en cuenta para valores superiores al 50%.

2.3 Resultados

2.3.1 Extracción de ADN genómico

Se extrajo ADN de 464 entradas de *E. guineensis* con el kit de QIAGEN DNeasy® Plant Mini, con el cual se obtuvo ADN de calidad adecuada para el uso de la técnica de microsatélites. La calidad y concentración de ADN visualizada en el gel arrojó concentraciones entre 30 y 100 ng/µl (Figura 2-2):

Figura 2- 2: Geles de agarosa (0,8%). ADN de palma *E. guineensis* extraído con el kit de QIAGEN DNeasy® Plant Min

2.3.2 Datos faltantes

Los datos faltantes de genotipos en los loci no superaron el 1% para cada marcador en toda la muestra de estudio, con excepción de los loci PR03, PR04 y PR05, los cuales fueron descartados del análisis. Las entradas colectadas fueron codificadas con los números consecutivos desde el 127 al 585, sin embargo, en este estudio las entradas 314, 316, 466 y 495 no fueron incluidas por falta del material vegetal al momento de la colecta. La información de pasaporte de cada entrada fue registrada y almacenada en la base de datos del laboratorio.

2.3.3 Diversidad alélica

Se evaluaron 19 loci, solo para 16 visualizaron perfiles genéticos claramente diferenciables (Anexo B). Los loci PR03, PR04 y PR05 presentaron porcentajes de amplificación muy bajos, por esta razón, no se incluyeron en los análisis. De los 16 loci amplificados, 13 mostraron ser polimórficos. Tres fueron monomórficos (PR10, PR11 y PR12) con los cuales se detectó 1 alelo por locus, estos loci no fueron incluidos en el análisis de datos.

Los loci microsatélites son conocidos por presentar un gran número de alelos y los parámetros que describen la diversidad detectada por los loci se presentan en la Tabla 2-**¡Error! No se encuentra el origen de la referencia.**3. Un total de 79 alelos fueron detectados con los 13 loci microsatélites evaluados en las 455 entradas, con un promedio de 5,61 alelos por locus. El porcentaje de amplificación para la mayoría de los loci fue del 100% exceptuando el locus PR23 (95,6%). El número máximo de alelos detectados fue de 9 alelos para los loci PR37 y PR23, el mínimo fue para los loci PR10, PR11 y PR12, en los cuales se detectó un solo alelo. El nivel de polimorfismo entre las 455 entradas de Angola fue evaluado con el índice de información polimórfica (PIC) para cada locus. El PIC mostró un valor máximo de 0,843 para el locus PR16 y un valor mínimo de 0,240 para el marcador PR28, con un promedio de 0,533. Según los valores reportados en la Tabla 2-3, se puede decir que los loci más informativos en esta población son el PR16, PR23 y PR37 y los menos informativos fueron los loci monomórficos (PR10, PR11 y PR12), seguidos de PR28, PR31 y PR29. Los loci con mayor número de alelos efectivos (A_e) se correlacionan con los mayores PIC y viceversa.

El máximo número de alelos efectivos detectado fue de 7.107 en el locus PR16 de un total de 8 alelos identificados y el mínimo fue de 1,3 para el locus PR28. La máxima heterocigosidad observada fue detectada en el locus PR16 (0,781) y el mínimo valor en el locus PR28 (0,209). La riqueza alélica (R_a), basada en la estimación de las variantes alélicas por locus, osciló entre 1,995 y 6,093, con un promedio de 3,636.

El análisis del equilibrio de Hardy–Weinberg (EHW) en una población nos permite inferir si las frecuencias alélicas evaluadas cumplen el modelo bajo asunciones como que el organismo sea diploide, reproducción sexual, generaciones no sobrelapadas, frecuencias alélicas iguales en hembras y machos, apareamiento al azar, tamaño poblacional infinito, ausencia de migración, mutación y que los loci evaluados no estén bajo selección (Hartl y Clark, 1997). El conocimiento de la biología de la especie nos permitirá inferir acerca de los posibles procesos evolutivos y biológicos que dentro de la población pudieron llevarla al desequilibrio. El test exacto de EHW fue significativo a un nivel del 0,05 para la mayoría de los loci, rechazando la hipótesis nula. Los loci PR01, PR02, PR07, PR013, PR024, PR27 y PR31, mostraron valores de probabilidad que oscilaron entre 0,136 y 0,999, interpretados como no significativos para rechazar la hipótesis nula de equilibrio HW. El desequilibrio presentado en los otros loci es debido a un leve déficit de heterocigotos detectados en la muestra, sugiriendo la ausencia de apareamiento aleatorio. El test exacto de desequilibrio de ligamiento (DL) global (Anexo C) mostró un desequilibrio significativo para el 56,4% de los pares de loci, lo cual indica no independencia entre los loci evaluados. Todas las zonas geográficas presentaron un DL significativo en algunos pares de loci, principalmente Sumbe (36) seguida de Caixito, Cabinda, Uige y Benguela (29, 24, 23 y 8), respectivamente.

El índice de Fijación F_{IS} muestra un exceso significativo de homocigotos detectados para los loci PR06, PR16, PR23, PR28, PR29 y PR37 y exceso no significativo de heterocigotos para los loci PR02, PR24, PR27 y PR31.

Tabla 2- 3: Parámetros descriptivos para los loci microsatélites evaluados

Locus	N	A	PIC	A_e	H_o	H_e	R_a	F_{IS}	P<0,05
PR01	442	7	0,644	3,110	0,649	0,678	4,317	0,029	0,136
PR02	444	5	0,534	2,427	0,568	0,588	3,473	0,021	0,214
PR06	455	7	0,645	3,284	0,505	0,696	4,059	0,251	**0,001**
PR07	455	5	0,457	2,075	0,501	0,518	2,869	0,026	0,246
PR13	455	4	0,441	1,971	0,545	0,493	2,878	-0,174	0,999
PR16	453	8	0,843	7,107	0,781	0,859	6,093	0,059	**0,002**
PR23	435	9	0,836	6,818	0,779	0,853	6,014	0,070	**0,001**
PR24	455	3	0,442	2,161	0,580	0,537	2,506	-0,075	0,979
PR27	455	4	0,460	1,983	0,484	0,496	3,138	0,017	0,350
PR28	455	5	0,240	1,332	0,209	0,249	2,293	0,138	**0,003**
PR29	455	2	0,352	1,838	0,277	0,456	1,995	0,388	**0,001**
PR31	455	5	0,298	1,458	0,332	0,314	2,528	-0,088	0,993
PR37	452	9	0,741	4,327	0,695	0,769	5,110	0,082	**0,001**
PROMEDIO	**451,2**	**5,61**	**0,533**	**3,069**	**0,531**	**0,577**	**3,636**	**0,057**	**0,001**

N: número de individuos evaluados por locus, A: número de alelos detectados por locus, A_e: número efectivo de alelos, PIC: indice de información polimórfica, H_o: hetetocigosidad observada, H_e: heterocigosidad esperada, R_a: riqueza alélica, P<0,05: probabilidad del equilibrio Hardy-Weinberg estadísticamente significativa menor de 0,05, F_{is}: índice de Fijación

2.3.4 Diversidad genética en las zonas geográficas

La diversidad genética es un atributo que mide la frecuencia esperada de heterocigotos en las poblaciones, independiente de su sistema de apareamiento. La diversidad genética dentro de las cinco zonas geográficas es presentada en la Tabla 2-4. Los 13 loci evaluados fueron polimórficos en todas las zonas geográficas, el porcentaje de loci polimórficos (%P) para todas las zonas geográficas fue de 100%. El número promedio de alelos por locus (A) dentro de cada zona geográfica varió entre 5,308 y 4,846, mientras que el número de alelos efectivos por zona geográfica (A_e) osciló entre 2,998 y 2,780, con un promedio de 2,846. La máxima heterocigosidad observada fue detectada en Uige (0,612) seguida por Benguela (0,586), Cabinda (0,517), Caixito (0,496) y finalmente Sumbe (0,474) que presentó la heterocigosidad observada más baja. El promedio de la heterocigosidad esperada (H_e) a través de las zonas geográficas fue de 0,561. Estos valores muestran que la magnitud de la diversidad genética a lo largo de las cinco zonas geográficas fue relativamente alta y varió en un rango muy estrecho. Los valores del índice de Fijación F_{IS} (F) para Sumbe, Caixito y Cabinda fueron positivos y significativos, indicando un déficit de heterocigotos. Por el contrario, en Benguela y Uige se observa un

exceso no significativo de heterocigotos, tendencia que se esperaría encontrar en la mayoría de las zonas debido al sistema de reproducción de la palma. Solo en Caixito se detectaron dos alelos privados en los loci PR02 (184pb) y PR06 (221pb), con una frecuencia de 0,47 y 0,068, respectivamente.

Tabla 2- 4: Parámetros de variabilidad y diversidad genética para las zonas geográficas evaluadas.

ZG	N	%P	A	A_e	H_o	H_e	R_a	F	P<0,05	AP
CX	88	100	5,077	2,780	0,403	0,559	4,993	0,100	**0,000**	2
SU	100	100	5,154	2,883	0,385	0,537	4,886	0,106	**0,000**	0
CB	117	100	5,308	2,779	0,420	0,561	5,047	0,076	**0,000**	0
BG	57	100	4,846	2,792	0,476	0,551	4,787	-0,063	0,599	0
UG	89	100	5,154	2,998	0,497	0,598	4,975	-0,019	0,951	0
PROMEDIO	**90**	**100**	**5,108**	**2,846**	**0,537**	**0,561**	**4,938**	**0,040**	**0,000**	**0,4**
DE		0,00	0,258	0,187	0,025	0,023		0,023	0,000	

ZG: zona geográfica, CX: Caixito, SU: Sumbe, CB: Cabinda, BG: Benguela, UG: Uige, P: promedio, DE: desviación estándar, N: entradas evaluadas por zona geográfica, %P: porcentaje de loci polimórficos, A: número promedio de alelos por zona geográfica, A_e: número efectivo de alelos, H_o: hetetocigocidad observada, H_e: heterocigosidad esperada, R_a: riqueza alélica, F: índice de fijación F_{IS}, P<0,05: probabilidad del test de EHW estadísticamente significativa menor de 0,05, AP: alelos privados

2.3.5 Diversidad genética dentro de familias

La estimación de la diversidad genética dentro de familias es importante ya que ésta puede ser usada para detectar familias con diversidad genética relativamente más alta y que a su vez, puedan marcar diferencias genéticas entre zonas geográficas. De esta manera, se pueden diseñar estrategias de mejoramiento y mantenimiento que las involucren.

En la Tabla 2-5 se observan los parámetros que describen la diversidad genética dentro de familias, donde el porcentaje de loci polimórficos (%P) fue en promedio de 94,58%. Sin embargo, la familia 5 de Sumbe (SU5) y la familia CX8 de Caixito presentaron los valores más bajos de %P (61,54 y 76,92, respectivamente). El valor máximo para el promedio de alelos detectados por familia (A) correspondió a las familias SU2 y CB3 (4,077), seguido por la familia BG5 (3,769) y contrario a la familia CX8 (2,154). Igualmente, para el número promedio de alelos efectivos (A_e) las mismas familias registraron los valores extremos incluyendo a la familia UG4 (2,718) que obtuvo uno de los mayores A_e. La mayoría de las familias de Uige superaron el promedio (2,273) de alelos efectivos.

Los valores de heterocigosidad esperada (H_e) oscilaron en un rango entre 0,564 (UG9) y 0,265 (CX8), con un promedio de 0,478 (Tabla 2-5). El índice de Fijación F_{IS} (F) describe

para la mayoría de las familias un exceso de heterocigotos reflejado en los valores negativos de F, los cuales oscilaron entre -0,348 (*p*=0,000) (UG9) y 0,080 (*p*=0,046) (SU1), con un valor promedio de -0,1111 (*p*=0,002). Las familias CX6, SU1, SU5, SU6, BG2, BG4, BG6, UG2, UG7, UG9 y UG10, se desvían significativamente del EHW por exceso de heterocigotos, por el contrario, las familias SU1 y SU9 presentaron desviación significativa del EHW por deficiencia de heterocigotos. Benguela y Uige presentaron mayor proporción de familias con desviación significativa del EHW por exceso de heterocigotos (3:6, y 4:9, respectivamente).

No se identificaron alelos privados en ninguna familia, y a nivel de zonas geográficas se identificaron dos alelos privados en Caixito, con lo cual se puede suponer que a pesar de su aislamiento geográfico, las zonas geográficas comparten los mismos alelos por un presumible alto flujo génico o divergencia reciente entre estas zonas.

Tabla 2- 5: Variabilidad y diversidad genética dentro de las familias de cada zona geográfica

Pop	N	%P	A	Ae	Ho	He	F	P<0,05
CX1	10	100,00%	3,154	2,018	0,531	0,455	-0,123	0,432
CX2	10	92,31%	2,923	2,065	0,485	0,435	-0,105	0,165
CX3	9	100,00%	3,692	2,322	0,470	0,472	-0,015	0,956
CX4	10	92,31%	3,769	2,539	0,538	0,522	-0,002	0,879
CX5	10	92,31%	3,077	2,111	0,513	0,451	-0,123	0,117
CX6	10	92,31%	3,692	2,351	0,588	0,486	-0,205	0,005
CX7	10	92,31%	3,615	2,373	0,500	0,495	-0,026	0,848
CX8	9	76,92%	2,154	1,475	0,274	0,265	0,024	0,224
CX9	10	100,00%	3,385	2,280	0,546	0,500	-0,108	0,265
SU1	10	100,00%	3,462	2,527	0,477	0,522	0,080	0,046
SU2	15	100,00%	4,077	2,131	0,508	0,470	-0,070	0,354
SU3	17	100,00%	3,615	2,098	0,480	0,482	-0,009	0,917
SU4	10	92,31%	3,077	2,217	0,545	0,460	-0,184	0,088
SU5	10	61,54%	2,308	1,953	0,385	0,333	-0,149	0,042
SU6	10	84,62%	2,462	1,921	0,477	0,393	-0,211	0,044
SU7	10	84,62%	3,385	2,364	0,469	0,433	-0,083	0,790
SU8	10	84,62%	3,462	2,329	0,485	0,449	-0,071	0,197
SU9	8	92,31%	3,385	2,181	0,413	0,438	0,014	0,037
CB1	10	92,31%	3,538	2,215	0,474	0,447	-0,034	0,349
CB2	10	100,00%	3,231	2,116	0,510	0,475	-0,056	0,244
CB3	10	100,00%	4,077	2,563	0,652	0,557	-0,180	0,200
CB4	10	92,31%	2,846	2,153	0,514	0,447	-0,135	0,443
CB5	10	92,31%	3,231	1,761	0,377	0,365	-0,029	0,499
CB6	12	100,00%	3,615	2,205	0,490	0,495	0,003	0,113
CB7	17	100,00%	3,692	2,007	0,409	0,399	-0,034	0,591
CB8	10	92,31%	3,538	2,416	0,592	0,509	-0,159	0,092
CB9	10	100,00%	3,615	2,259	0,536	0,496	-0,073	0,492

Tabla 2- 5: (Continuación)

Pop	N	%P	A	Ae	Ho	He	F	P<0,05
CB10	10	100,00%	3,385	2,347	0,615	0,532	-0,182	0,381
CB11	10	100,00%	3,692	2,508	0,605	0,549	-0,110	0,237
BG1	10	100,00%	3,538	2,661	0,521	0,537	0,025	0,080
BG2	10	100,00%	3,462	2,468	0,638	0,522	-0,192	0,006
BG3	10	100,00%	3,154	2,107	0,517	0,478	-0,064	0,481
BG4	9	84,62%	3,000	2,343	0,606	0,457	-0,328	0,000
BG5	10	100,00%	3,769	2,536	0,587	0,507	-0,172	0,086
BG6	9	100,00%	3,385	2,462	0,650	0,531	-0,244	0,031
UG1	10	100,00%	3,308	2,322	0,568	0,513	-0,114	0,267
UG2	10	92,31%	3,538	2,184	0,603	0,485	-0,225	0,065
UG3	10	100,00%	3,538	2,520	0,607	0,513	-0,168	0,214
UG4	10	92,31%	3,615	2,718	0,532	0,530	-0,021	0,752
UG6	10	100,00%	3,538	2,503	0,541	0,549	0,008	0,113
UG7	10	100,00%	3,308	2,345	0,689	0,530	-0,301	0,001
UG8	10	92,31%	3,308	2,126	0,531	0,466	-0,151	0,169
UG9	10	100,00%	3,692	2,556	0,768	0,564	-0,348	0,000
UG10	10	92,31%	3,538	2,371	0,669	0,536	-0,257	0,003
PROMEDIO	10	94,58	3,383	2,273	0,534	0,478	-0,111	0,002
DE	0,074	1,16	0,064	0,041	0,012	0,009	0,013	0,0001

ZG: zona geográfica, CX: Caixito, SU: Sumbe, CB: Cabinda, BG: Benguela, UG: Uige, N: entradas evaluadas por familia, %P: porcentaje de loci polimórficos, A: número promedio de alelos por familia, A_e: número efectivo de alelos, H_o: hetetocigosidad observada, H_e: heterocigosidad esperada, F: índice de fijación F_{IS}, P<0,05: probabilidad del test de EHW estadísticamente significativa menor de 0,05, DE: desviación estándar.

2.3.6 Estructura genética

La Tabla 2-6 muestra los parámetros de diferenciación genética entre familias y entre zonas geográficas según Nei (1987), los cuales pueden ser aplicados sin considerar el número de alelos por locus y las fuerzas evolutivas (Laurentin, 2009). La Tabla 2-6 muestra valores promedio de heterocigosidad observada (H_o) y diversidad genética dentro de grupos (H_s). Se observa que la diversidad genética dentro de familias es un poco menor que dentro de zonas geográficas, también se muestra una diversidad genética total relativamente alta (0,581). El promedio de diversidad genética entre subgrupos (D_{ST}) muestra un valor de tan solo 0,079 entre familias y de 0,016 entre zonas geográficas, reflejando menor variabilidad entre las zonas geográficas, con lo cual se concluye que la mayor parte de la diversidad genética se encuentra distribuida dentro de grupos (familias y zonas geográficas) y no entre ellos. El valor calculado para el coeficiente de diferenciación genética (G_{ST}) fue de 0,136, que aunque significativo (P<0,01) determina una diferenciación genética más baja entre zonas geográficas que entre familias. El promedio del G_{is} entre zonas geográficas fue de 0,049 y entre familias

de -0,063, mostrando leves desviaciones del EHW, entre familias un exceso significativo de heterocigotos (P<0,01) y entre zonas geográficas un déficit altamente significativo (P<0,01) de heterocigotos.

Tabla 2- 6: Coeficientes de diversidad y diferenciación genética de Nei (1987) entre grupos (familias y zonas geográficas)

Agrupamiento	H_O	H_S	D_{ST}	H_T	G_{ST}	G_{IS}	P<0,05
Por familias	0,534	0,502	0,079	0,581	0,136	-0,063	**0,002**
Por zonas geográficas	0,537	0,565	0,016	0,581	0,028	0,049	**0,000**

H_o: promedio de diversidad genética observada, H_s: promedio de diversidad genética dentro de grupos, H_t: diversidad genética total, D_{ST}: promedio de diversidad genética entre subgrupos, G_{ST}: coeficiente de diferenciación genética, G_{IS}: coeficiente de endogamia, P<0,05: probabilidad del test de EHW estadísticamente significativa menor de 0,05.

El análisis molecular de varianza (AMOVA) se realizó teniendo en cuenta como fuentes de variación de nivel jerárquico las zonas geográficas, las familias y los individuos, con el fin de identificar la variación molecular dentro y entre estos niveles. El AMOVA (Tabla 2-**¡Error! No se encuentra el origen de la referencia.**7), muestra que la diferenciación genética entre zonas geográficas es solo del 1% y entre individuos dentro de familias no hay diferenciación genética, así que la varianza se encuentra distribuida principalmente entre familias y dentro de individuos. La varianza entre familias dentro de zonas geográficas representa el 12% de la varianza total, con valores de diferenciación genética moderada (F_{ST}= 0,137) pero significativos (Tabla 2-8). Es claro que la mayor parte de la varianza genética está dentro de los individuos (87%) con una baja pero significativa consanguinidad individual (F_{IT}= 0,029) en la población total. Estos valores son concordantes con los valores reportados en la Tabla 2-6, concluyendo que la mayor parte de la varianza se encuentra entre familias dentro de las zonas geográficas y dentro de los individuos en la población total. La diferenciación genética entre todos los pares de zonas geográficas y familias aunque fue baja, en todos los casos fue significativa (Anexo D y E).

Tabla 2- 7: Análisis de varianza molecular (AMOVA F)

Fuente de Variación	gl	Suma de cuadrados	Componentes δ^2	% δ^2
Entre zonas geográficas	4	92,711	0,057	1,4
Entre familias dentro de zonas geográficas	39	487,009	0,465	11,8
Entre Individuos dentro de familias	411	1295,048	0,000	0
Dentro de Individuos	455	1554,00	3,415	86,8
Total	**909**	**3438,768**	**3,937**	**100**

gl: grados de libertad, % δ^2 : porcentaje de variación

Tabla 2- 8: Estadísticos F (Wright, 1965) de diferenciación genética entre zonas geográficas y familias.

	ZONAS GEOGRÁFICAS		FAMILIAS	
	F	**P<0,05**	**F**	**P<0,05**
F_{ST}	0,028	**0,001**	0,137	**0,001**
F_{IT}	0,077	**0,001**	0,102	**0,001**
F_{IS}	0,103	**0,001**	-0,040	1,000

El test de Mantel mostró que entre las 5 zonas geográficas la correlación entre la divergencia genética (valorada como F_{ST}) y la distancia geográfica no fue estadísticamente significativa (Z = 122,6220, r = -0,3917, R^2= 0,154, valor-p= 0,8250), Por lo tanto en el presente estudio no se evidenció un proceso de aislamiento por distancia entre las zonas geográficas.

- **Flujo génico**

El flujo génico es una fuerza homogenizadora que actúa en contra de la divergencia genética entre las zonas geográficas. Para la estimación del número de migrantes no se consideró el método de alelos privados puesto que su resultado podría subestimar el estadístico Nm, ya que solo se registraron dos en las zonas geográficas evaluadas. Por lo tanto, la estimación de Nm se basó en el estadístico $F_{ST.}$ La Tabla 2-9 muestra que los valores promedio de flujo génico son moderadamente altos entre zonas geográficas (Nm= 8,664) y entre familias (Nm= 1,572) basado en un numero efectivo de 50 individuos para zonas geográficas y 10 para familias. Estas estimaciones del flujo génico entre familias eran esperadas debido al tipo de polinización alógama característico de la palma, sin embargo, entre zonas geográficas no se esperaba un flujo génico moderadamente alto, pero estos resultados corroboran los resultados de la AMOVA y el G_{ST}.

Tabla 2- 9. Estimadores de flujo génico con base en el valor promedio del número de migrantes

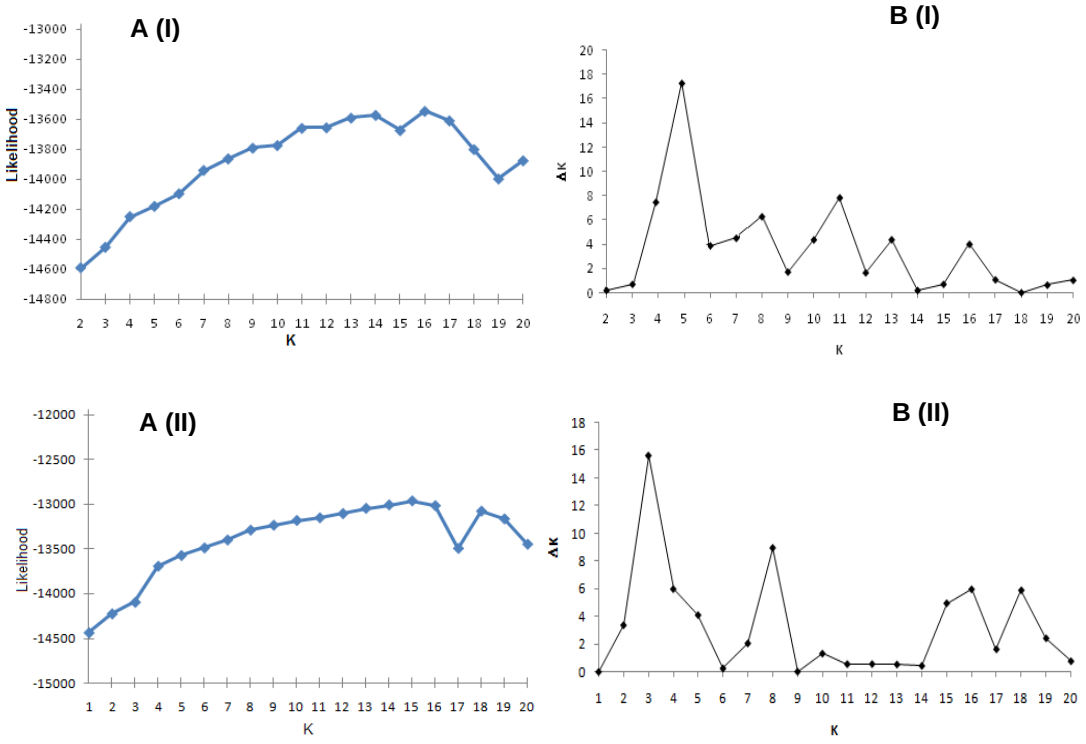
	N_e	**Nm**
Zonas geográficas	50	8.664
Familias	10	1,572

Nm: número de migrantes, Ne: Número efectivo

- **Estructura poblacional y aproximación bayesiana**

Para el análisis basado en métodos de estadística bayesiana, se calculó el K optimó (ΔK) según la metodología de Evanno *et al.* (2005), realizando varias simulaciones y explorando diferentes parámetros. El análisis fue realizado sin información *a priori* y con información *a priori* de la zona geográfica donde se realizó la colecta, y los resultados de las dos simulaciones son mostrados en la Figura 2-3. Las Figuras 2- 3 A muestran que a medida que aumenta el K, el valor de la likelihood tiende a estabilizarse, sin embargo no se logra estabilidad completa ya que a partir de K=14 (Figura 2- 3 A (I)) y K=16 (Figura 2-3 A (II)) el valor de la likelihood vuelve a descender. La Figura 2-3 B (I) presenta un ΔK máximo en un K=5 poblaciones, mientras que la simulación con información *a priori* calculó un ΔK máximo en un K=3 poblaciones (Figura 2-3 B (II)).

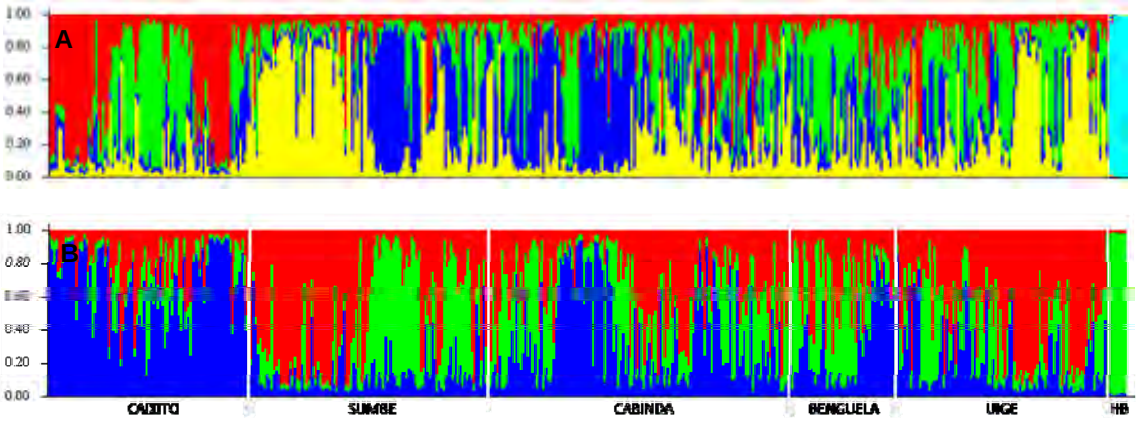
Figura 2- 3: Gráficas para la exploración de la estructura poblacional de las entradas analizadas utilizando el programa Structure. (A) Valores promedio del Ln P(D) (Log likelihood) vs K, (B) ΔK vs K, de acuerdo a Evanno *et al*. (2005). (I) simulación sin información *a priori* de la zona geográfica de las entradas (II) simulación con información *a priori* de la zona geográfica de las entradas evaluadas.



Las simulaciones sin información *a priori* y con información del lugar de colecta mostraron diferentes resultados de agrupamiento, como se muestra en las Figuras 2-3 y 2-4, respectivamente. En la Figura 2-4 se observa que cada zona geográfica está compuesta por un conjunto de individuos heterogéneos evidenciado en el patrón de asignación de los individuos a cada una de las K=5 (Figura 2-4 A) y K=3 (Figura 2-4 B) poblaciones estadísticas sugeridas por el análisis. Por lo cual, el análisis sugiere desde el punto de vista genético, que las zonas geográficas no muestran una clara diferenciación entre ellas, lo que está en corcondancia con los resultados anteriores de AMOVA y de índices de fijación.

Después de inspeccionar la asignación de los individuos a cada una de las K=3 poblaciones, aún no es claro cuál podría ser el significado de cada una de estas tres poblaciones estadísticas ni desde el punto de vista biológico ni geográfico.

Figura 2- 4: Gráfico del K óptimo (ΔK) de las simulaciones del programa Structure incluyendo el grupo de híbridos. A. Sin información *a priori* del lugar de colecta. B. Con información geográfica *a priori* del lugar de colecta.



2.3.7 Relaciones genéticas

- **Distancias genéticas entre las zonas geográficas**

La distancia genética refleja la diferenciación a nivel de frecuencias génicas expresada en el número de sustituciones alélicas por locus que se han dado en la evolución separada de dos subpoblaciones, grupos o especies. Por eso es de suponer que los individuos más distantes en las topologías construidas con disimilaridades comparten menor número de variantes alélicas. Los promedios y valores máximos de las diferentes distancias genéticas entre todos los pares de zonas geográficas (Tabla 2-10) permiten inferir que a nivel general la diferenciación genética es baja entre las zonas geográficas donde se encuentra *E. guineensis* consideradas en este trabajo. Por otra parte, se

observa que el promedio de las distancias aumenta significativamente cuando se incluye la población de híbridos para todas las clases de distancias genéticas. La distancia dm^2, la cual asume un modelo mutacional *"Stepwise"* o paso a paso para loci microsatélites y tiene una varianza menor (Goldstein y Pollock, 1997), muestra diferencias mayores entre los promedios de las comparaciones. Finalmente, se concluye que la distancia genética entre *E. guineensis* de las 5 zonas geográficas de estudio es mínima comparada con la distancia genética entre *E. guineensis* y el híbrido.

Tabla 2-100: Distancias genéticas. Valores máximos, mínimos y promedios entre pares de zonas geográficas con inclusión y exclusión del grupo de híbridos.

COMPARACIONES	DISTANCIA	MÁXIMO	MÍNIMO	PROMEDIO
Entre *E. guineensis*	CAS	0,2003	0,1413	0,1650
***E. guineensis* vs HB**	CAS	0,6078	0,1413	0,3062
Entre *E. guineensis*	DSG	0,0557	0,0202	0,0375
***E. guineensis* vs HB**	DSG	0,4675	0,0202	0,1941
Entre *E. guineensis*	dm^2	0,4223	0,0589	0,2042
***E. guineensis* vs HB**	dm^2	906,704	0,0589	276,594
Entre *E. guineensis*	ASD	930860	777857	858896
***E. guineensis* vs HB**	ASD	2071106	777857	1226346,2

CAS: distancia de Cavalli Sforza y Edwards (1967) (Valor máximo 2), DSG: distancia genética estándar de Nei (1972) (Valor máximo ∞), $dm^2 = \delta\mu^2$ (Goldstein, 1995a) (Valor máximo ∞), ASD: distancia cuadrática media (Goldstein, 1995b y Slatkin 1995) (Valor máximo ∞).

- **Análisis de Coordenadas principales (ACoP)**

El análisis de coordenadas principales (ACoP) fue usado como una herramienta adicional para explorar patrones de agrupamiento entre individuos y corroborar los datos de la estructura poblacional. Al realizar el análisis de coordenadas principales con una distancia de disimilaridad estimada con un índice de concordancia simple y aplicando un test de bootstrap de 1000 repeticiones, el resultado de las dos primeras coordenadas principales representa en conjunto tan solo el 8,36 % de la varianza total de la población (CoP1= 4,32 y CoP2= 4,04) no siendo significativo.

Es claro notar que las coordenadas encontradas para el análisis anterior explican un porcentaje muy bajo de la variación, por esta razón, se exploraron cuatro distancias genéticas (Nei, 1972; Cavalli Sforza y Edwards, 1967; dm^2 (Goldstein, 1995a) y cuadrática media (ASD) (Goldstein, 1995b y Slatkin, 1995)) (Anexo F), incluyendo distancias como dm^2 y ASD que asumen modelos mutacionales para marcadores

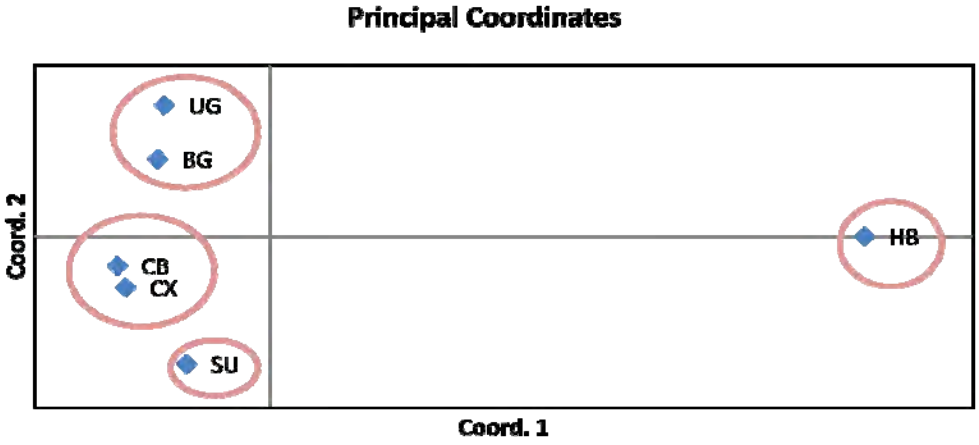
microsatélites (Tabla 2-11).

Tabla 2-11. Análisis de Coordenadas principales. Porcentaje de variación explicado por las 3 primeras coordenadas principales basado en diferentes distancias genéticas.

Distancia	CoP1	CoP2	CoP3	% Acumulado
dm^2	100	0	0	100
Nei	85,20	6,24	4,18	95,62
Cavalli Sforza	56,94	13,63	10,52	81,09
ASD	44,15	15,30	14,43	73,88

Como se observa, la distancia genética dm^2 explicó mejor la variación de los datos en su primera coordenada principal acumulando un 100%. En este ACoP se utilizaron las zonas geográficas como unidades de análisis incluyendo los híbridos, aumentando significativamente el porcentaje de explicación de la variación en los tres primeros ejes. La Figura 2-5 muestra que las zonas geográficas de Uige y Benguela están más cercanas entre ellas y de forma similar las zonas geográficas de Cabinda y Caixito. Sumbe se ubica más distante respecto a las otras. Por otra parte, se puede observar que el grupo de híbridos se encuentra muy alejado de las cinco zonas geográficas de estudio como se esperaba.

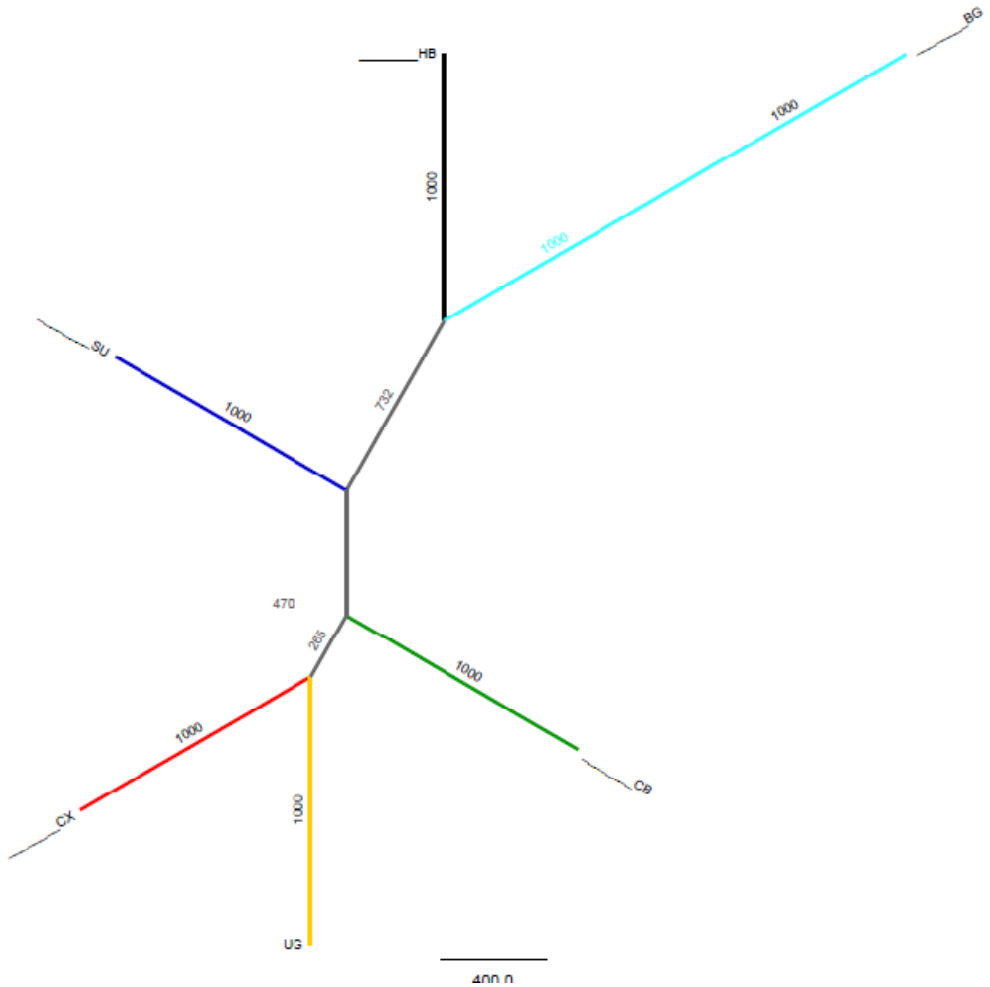
Figura 2- 5: Análisis de coordenadas principales (primeras dos dimensiones) basado en la matriz de distancias genéticas dm^2 para las 5 zonas geográficas incluyendo los híbridos.



- **Análisis de agrupamiento realizado para zonas geográficas, familias e individuos**

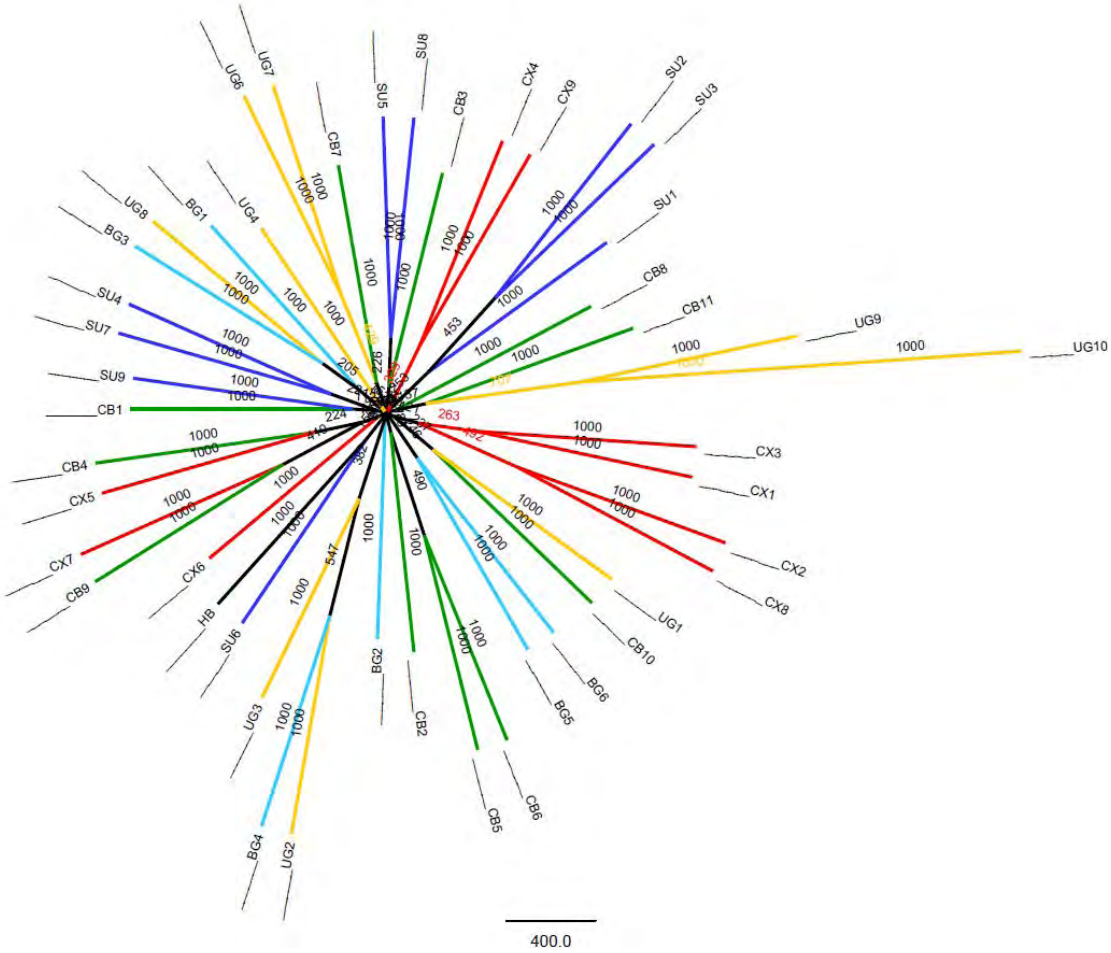
El análisis de agrupamiento obtenido usando el algoritmo NJ y la distancia genética de Nei (1972), en las 5 zonas geográficas y el grupo de híbridos de palma de aceite es mostrado en la Figura 2-6. Se observan las zonas geográficas de Caixito, Uige y Cabinda como las más cercanas, sin embargo, los valores de bootstrap no son significativos para la mayoría se los agrupamientos ya que no superan el 50% (470 y 265). La zona de Benguela y los híbridos se observan más distantes respecto a las otras.

Figura 2- 6: Topología basada en el método de agrupamiento Neighbor-joining y la distancia genética de Nei (1972), para las 5 zonas geográficas y el grupo de híbridos usado como grupo externo.



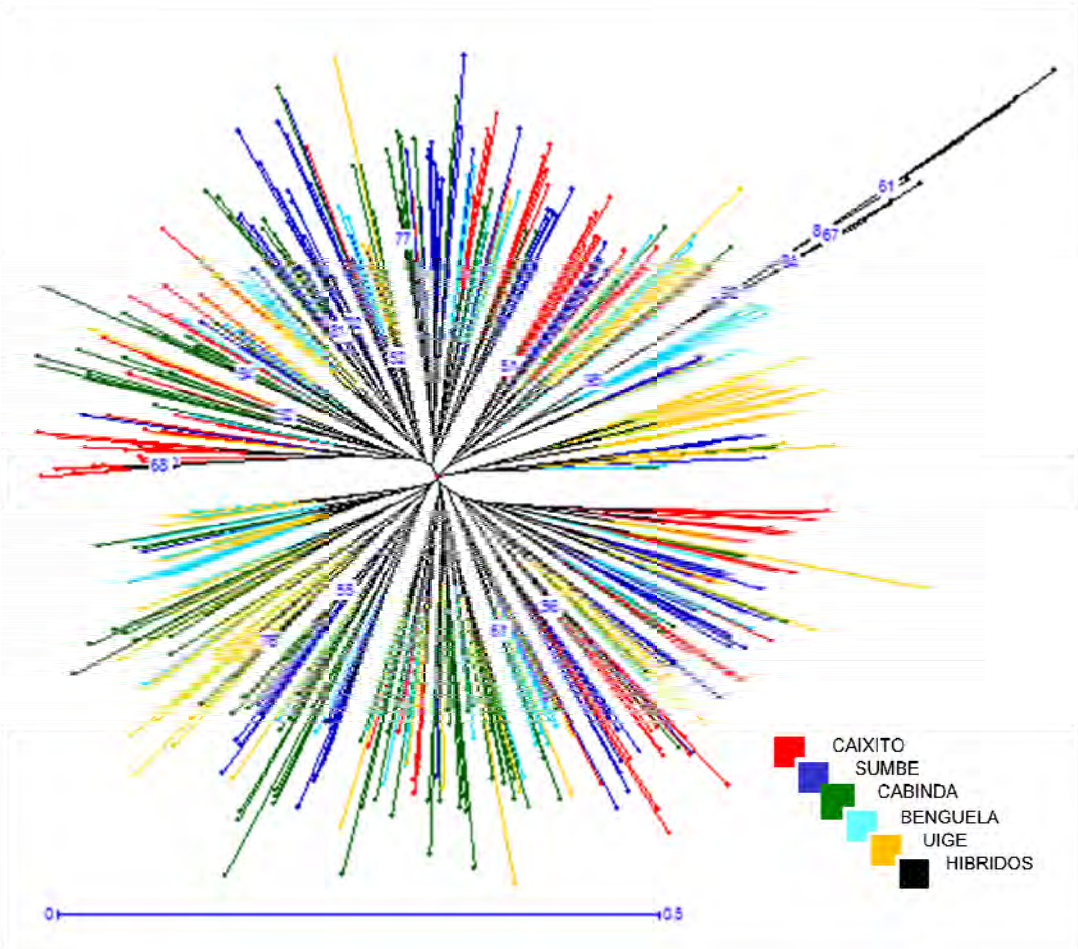
La topología de agrupamiento NJ tomando como unidad de análisis las cuarenta y cuatro familias de palma de aceite, es mostrada en la Figura 2-7, donde se puede identificar la familia UG10 como la más distante del grupo seguida por las familias CX2, CX7, CX8, CB5, CB6, CB9, BG4, SU2, SU3, SU6, SU5, UG2, UG6, UG7 y UG9, respectivamente. La topología no muestra agrupamientos definidos que presenten un patrón claro. Sin embargo, muestra cuales familias son cercanas o distantes, vislumbrando relaciones entre ellas definidas por las distancias genéticas entre pares de familias (Anexo G), lo cual en términos prácticos es importante para la selección de germoplasma. La mayoría de los nodos de la base de la topología no presentan soporte estadístico, puesto que no superan el 50% de las repeticiones (1000).

Figura 2- 7: Topología basada en el método de agrupamiento Neighbor-joining y la distancia genética de Nei (1972), incluye 44 familias correspondientes a las 5 zonas geográficas colectadas de *E. guineensis* en la República de Angola y un grupo de híbridos.



La topología de agrupamiento obtenida al aplicar el algoritmo de NJ (Figura 2-8) y tomando como unidad de análisis los individuos no permitió identificar grupos grandes con soporte estadístico según los valores de bootstrap. Al examinar la Figura 2-8 se observan agrupamientos muy heterogéneos conformados por individuos provenientes de orígenes geográficos diferentes. En la gráfica se muestran los valores bootstrap mayores a 50. Los valores más altos corresponden a los híbridos los cuales están separados significativamente del resto de individuos.

Figura 2- 8: Topología basada en el método de agrupamiento Neighbor-joining y el índice de disimilaridad, incluye 464 entradas de *E. guineensis* distribuidas en 5 zonas geográficas y un grupo de híbridos usado como grupo externo. Valores de bootstrap superiores a 50.



Finalmente, se puede concluir que no se detectó ninguna estructura en la cual las zonas geográficas, familias o individuos pudieran ser agrupados separadamente unas de otras, por criterios geográficos o por la característica grosor de cuesco de la cual se tiene información y para la cual se evaluó la posible presencia de un patrón de agrupamiento. La topología obtenida nos brinda una aproximación descriptiva de las relaciones de disimilitud entre individuos, pero no tiene soporte estadístico significativo para la mayoría de las agrupaciones.

2.4 Discusión

2.4.1 Diversidad alélica

Los loci microsatélites utilizados en este estudio permitieron evaluar y cuantificar los patrones de variación genética existente en *E. guineensis* de las cinco zonas geográficas evaluadas para 13 loci.

El número promedio de alelos por locus (5,61) es relativamente menor respecto al número de alelos reportados por Billote *et al*. (2001) y Singh *et al*. (2008), en los loci evaluados, sin embargo, se detectaron tres alelos nuevos en tres loci (PR01, PR02 y PR028), lo cual podría explicarse debido a la naturaleza del material analizado correspondiente a diferentes orígenes africanos y americanos. Por otra parte, Bakoumé e*t al.* (2007), reporta un número total de 209 alelos detectados con 16 loci microsatélites en 494 individuos procedentes de 10 países africanos, número que atribuyen al alto número de repeticiones de las secuencias microsatélites. En el presente estudio, con zonas geográficas de un solo país, no podemos corroborar estos resultados ya que a pesar del alto número de repeticiones en algunos loci, el número de alelos detectados no fue similar a los reportados por Bakoumé *et al.* (2007), para el mismo locus y para otros con igual o mayor número de repeticiones. De lo anterior, podemos concluir que independientemente del número de repeticiones de cada locus, el número de los alelos detectados depende de la naturaleza del germoplasma y por ende de su variabilidad.

La mayoría de loci utilizados en este estudio para la amplificación de regiones microsatélites en *E. guineensis* mostraron ser útiles en la identificación de variabilidad genética. Los promedios del número de alelos (A= 5,61) y el número de alelos efectivos en la población de Angola (A_e= 3,069), presentan una clara diferencia lo cual podría indicar la presencia de alelos raros en todas las zonas geográficas principalmente en Caixito y Cabinda, donde la diferencia es mayor, teniendo en cuenta que diferencias grandes entre estos parámetros indican bajas frecuencias de algunos alelos presentes solo en pocos individuos, diferencia usada para identificar alelos raros (Laurentin, 2009). La presencia de alelos raros en las zonas geográficas de condiciones climáticas diferentes podrían ser atribuidas a la presencia de nuevas variantes alélicas como consecuencia de las adaptaciones especificas a cada región, resultados similares fueron reportados por Bakoumé *et al*. (2007), quien atribuye su presencia no solo a procesos de deriva genética sino también de selección. Por ende, es de suponer que las poblaciones mejoradas tras muchos ciclos de selección han perdido muchos de estos alelos raros.

El polimorfismo encontrado en este estudio permitió hacer estimaciones de diversidad genética (H_e= 0,577) que corroboran los resultados obtenidos por Montoya *et al.* (2005), en la exploración preliminar del germoplasma de Angola de Cenipalma (H_e= 0,455). En otros estudios realizados con isoenzimas (Hayati *et al.*, 2004) y RFLP (Maizura *et al.*, 2006), incluyendo germoplasma de varios países africanos, reportan valores de heterocigosidad esperada menores (0,184 y 0,199, respectivamente) y en donde las poblaciones de Angola mostraron valores superiores al promedio total (0,194 y 0,211, respectivamente). Por el contrario, con loci microsatélites Cochard *et al.* (2009) encontraron que la población de Angola evaluada mostró la riqueza alélica más baja (4,877), así mismo, fue la menor heterocigosidad observada (0,632). Es de notar que, lo anterior no puede ser una comparación debido a los loci y tipo de germoplasma utilizado

en cada estudio, pero si podría ser un referente para concluir que los loci microsatélite debido al poder de detección y en general a sus características particulares (Goldstein y Pollock, 1997), respecto a los otros loci, pueden hacer mejores estimaciones de la diversidad de las poblaciones, también que hay otras poblaciones africanas que podrían tener mayor y menor diversidad genética.

En general, las poblaciones espontáneas de palma africana colectadas en las regiones a las cuales se les atribuye su origen, poseen altos valores de diversidad genética. Los coeficientes de diversidad genética (Tablas 2-4, 2-5 y 2-6) entre zonas geográficas no variaron mucho (H_e varía entre 0,537 y 0,598), pero a nivel de familias se puede observar una variación en un rango más amplio (H_e varía entre 0,265 y 0,564). En términos de mejoramiento genético en palma, estos niveles de variabilidad genética detectados son muy importantes, ya que superan el promedio de H_e de las poblaciones mejoradas de Deli y Dura reportadas en diferentes estudios (Montoya *et al.*, 2005; Maizura *et al.*, 2006; Singh *et al.*, 2008; Tingh *et al.,* 2010), constituyendo a la colección de Angola del centro de investigación de Cenipalma en un recurso genético útil para el propósito de ampliar la base genética de las poblaciones mejoradas y establecer en los acervos genéticos seleccionados la presencia de características de interés agroindustrial y de esta manera, guiar los cruzamientos que generan segregantes transgresivos, además de identificar las regiones de alta diversidad con el fin de realizar nuevas colectas en estos sitios.

El progreso de los procesos de mejoramiento genético depende de la utilización eficiente y sistemática del recurso genético permitiendo la obtención de materiales mejorados agronómicamente, a través de la inclusión de germoplasma de poblaciones naturales, silvestres, nativas y especies relacionadas con valor potencial que contengan buena parte de la diversidad genética resultante de procesos de selección y mutación en respuesta a la adaptación a factores bióticos y abióticos (Rey, 2007), en este contexto, la variabilidad genética encontrada en el germoplasma de Angola puede ser muy útil e interesante para el programa de mejoramiento genético de Cenipalma, información que permitirá identificar familias, las cuales presentaron mayor diversidad genética que podría aportar variantes alélicas de interés en la planeación de nuevos cruces y ganar eficiencia en el proceso del fitomejoramiento, teniendo en cuenta la correlación con los estudios bioquímicos, morfoagronómicos, de plagas y enfermedades.

Por otra parte, el entendimiento de los patrones de desequilibrio de ligamiento entre los loci evaluados en el germoplasma disponible podría ayudar a vislumbrar fenómenos poblaciones que han afectado a la población. El desequilibrio de ligamiento puede ser causado principalmente por la selección, la deriva genética y la mezcla de poblaciones de diferentes frecuencias alélicas. Sin embargo, hay otros factores que juegan un papel importante en los niveles del desequilibrio de ligamiento en una población como son la distancia física entre los loci, las tasas de recombinación, el apareamiento al azar, los patrones de reproducción (sexual, asexual, alógama o autógama), los procesos de mutación, el tamaño poblacional, el efecto fundador y los cuellos de botella (Flint-García *et al.,* 2003; Hartl y Clark, 1997). Estos últimos pueden incrementar dramáticamente los niveles de DL (Charlesworth *et al.,* 2003) reduciendo la variabilidad.

Estas fuerzas posiblemente han actuado en diferentes niveles a lo largo de la historia de la población de Angola, principalmente en las zonas de Sumbe, Caixito, Cabinda y Uige, las cuales presentaron los mayores niveles de DL. Por el contrario, Benguela presentó la menor proporción de pares de loci en DL (8:78), por lo cual podría inferirse que esta zona ha tenido menor efecto de algunas fuerzas evolutivas o el apareamiento al azar y la recombinación han sido más efectivos. Sin embargo, ninguna zona supera el 50% lo cual puede ser posiblemente un efecto principal del sistema de reproducción de la palma (sexual, alógama) y puede decrecer según la dinámica poblacional.

En los estudios de estructura y diversidad genética de palma realizados hasta el momento no se han encontrado reportes de los niveles de DL en las poblaciones analizadas, siendo éste un atributo importante de las poblaciones ya que la estructura incrementa el DL en el genoma (Aranzana *et al.,* 2010). Adicionalmente, el DL afecta los análisis de datos produciendo distorsiones, patrones inestables y poco confiables (Mohammadi y Prasanna, 2003; Kaeuffer *et al.,* 2007), además tiene implicaciones importantes en los procesos de mejoramiento, ya que reduce la variabilidad genética y puede ser importante en futuros estudios de asociación que involucren los loci o germoplasma evaluados.

El índice de endogamia (F_{IS}), mostró ser levemente positivo (0,103, P<0,001) entre zonas geográficas, reflejando un efecto menor de la endogamia dentro de las zonas geográficas, resultado similar al encontrado por Cochard *et al.* (2009) (F_{IS}= 0,125), lo cual podría ser una de las causas que explicaría el desequilibrio de Hardy-Weinberg detectado en las zonas geográficas evaluadas, debido posiblemente al sistema de reproducción de la palma de aceite, donde rara vez, en algunas inflorescencias femeninas se desarrollan flores masculinas que causarían esos bajos porcentajes de endogamia. Sin embargo, se podría decir que en general parece predominar un patrón de apareamiento y de distribución de la diversidad genética que no se aleja significativamente de lo esperado bajo un modelo de apareamiento al azar. En términos prácticos, la homocigocis observada aunque es baja, puede representar una desventaja en términos de mejoramiento si se involucran familias como CX8 y SU1, sin conocimiento de esta condición en algún programa de mejoramiento o conservación de germoplasma, adicionalmente es probable que aumente positivamente el F en estas familias, si las siguientes generaciones no tienen cruzamiento al azar y flujo génico efectivo.

2.4.2 Estructura poblacional

La mayoría de la diferenciación genética ocurre dentro de individuos, sin embargo, entre familias dentro de las zonas geográficas, se detectaron niveles de diferenciación genética bajos pero significativos. Resultados similares han sido reportados en poblaciones no domesticadas y espontáneas de chontaduro (*Bactris gasipaes* Kunth*)* colectadas en el Amazonas ($F_{ST=}$ 0,19 y G_{ST}= 0,16) (Adin *et al*., 2004; Rodrígues *et al.,* 2004). Por el contrario, otros estudios en palma de aceite de origen africano en los cuales se estima la diferenciación genética entre poblaciones (Hayati *et al.,* 2004; Singh *et al*., 2008;

Chochard *et al.,* 2009) han encontrado grados superiores de diferenciación, debido posiblemente a que fueron incluidos dentro de la colecta de germoplasma de diferentes orígenes en varios países Africanos donde se encuentra la palma de aceite.

Los resultados de este estudio mostraron una diferenciación genética menor entre zonas geográficas y un poco mayor entre familias. A nivel de familias, el proceso de flujo génico es dado por la dispersión natural de polen por efecto entomófilo y anemófilo, el cual ocurre a cortas distancias, posiblemente haciéndolo más restringido a algunas familias dentro de cada zona, lo cual, explicaría la diferenciación entre familias. El movimiento dinámico de las semillas de palma entre poblaciones a pesar de las amplias distancias geográficas es principalmente atribuido al efecto antrópico dado que el fruto de la palma es comestible en fresco y comercializado en los mercados locales de toda la República de Angola, actividad cultural que ha podido durante miles de años tener un efecto homogenizador. Otros dispersores de semillas de palma son los animales y el agua, este último con un impacto importante en el proceso de migración de semillas al que se le atribuye parte del proceso de dispersión de la palma de aceite a través de los ríos y manantiales africanos (Hayati *et al.,* 2004; Corley y Tinker, 2003). La baja diferenciación genética es consistente con la ausencia de correlación significativa de aislamiento por distancia entre las zonas geográficas evaluadas en este trabajo. Igualmente, Elsibli y Korpelainen, (2008) y El-Assar *et al.* (2005), evaluando la estructura genética para poblaciones de palma de dátil (*Phoenix dactylifera L.*) de Sudan y Egipto no encontraron estructura debida a la localización geográfica de las poblaciones. Los resultados del análisis bayesiano no mostraron ninguna estructura genética aparente debida a la localización geográfica en las entradas evaluadas. Sin embargo, es fundamental tener en cuenta que se desconoce el estatus de aprovechamiento de la población evaluada por lo cual, no se puede afirmar concretamente cual ha sido el efecto de este fenómeno en los proceso de flujo génico entre las zonas geográficas.

El ΔK calculado en el análisis podría ser causado por el desequilibrio de ligamiento entre los loci usados como lo menciona Kaeuffer *et al*. (2007), quien evaluó el efecto del desequilibrio de ligamiento en el análisis de estructura poblacional con el programa Structure, sugiriendo que el análisis puede ser sensible incluso a un par de loci con DL fuerte en algunos casos, donde la tasa de recombinación, la dinámica poblacional y el efecto fundador pueden incrementar o decrecer los niveles de DL alterando las estimaciones de probabilidades de agrupamientos. Los pares de loci evaluados en este estudio, en su mayoría mostraron desequilibrio de ligamiento significativo (Apéndice C), lo cual podría explicar en parte el ΔK obtenido, ya que Kaeuffer *et al.* (2007) cita que desequilibrios de ligamiento “fuertes” o desviaciones de EHW podrían generar una sobrestimación del numero de agrupamientos detectados, ambas condiciones son presentadas en los resultados de este estudio. Por otra parte, el perfil de agrupamiento mostrado en la Figura 2- 5 infiere una población heterogénea (poblaciones compuestas por plantas con genotipos diferentes) e individuos mayoritariamente heterocigotos, características que tienden a tener las poblaciones naturales de especies alógamas (Vallejo y Estrada, 2002).

Lo cual podría ser posiblemente explicado por el efecto mayor de algunas fuerzas evolutivas como la selección que estarían actuando más que la recombinación efectiva

(factores que podrían afectar como el sistema de apareamiento y el tamaño poblacional) favoreciendo algunas asociaciones de loci, que podrían pertenecer a genotipos que tienen altos valores de fitness.

Al analizar la falta de estructura genética y no encontrar un patrón de agrupamiento por ubicación geográfica, es importante tener en cuenta la naturaleza del muestreo, el cual fue basado en el criterio de contenido de la pulpa en el fruto y no de ubicación geográfica, concluyendo que el muestreo puede estar sesgado por este criterio de colecta, posiblemente excluyendo otros materiales genéticos presentes en las zonas geográficas. El número de entradas y los loci (464 y 13, respectivamente) empleados en este análisis proveen un poder estadístico significativo para rechazar la hipótesis nula de estructura geográfica dentro de la muestra evaluada.

2.4.3 Relaciones genéticas

La exploración de las relaciones genéticas a través de la estimación de las distancias genéticas y métodos de agrupamiento, permitió distinguir cuales individuos, familias y zonas geográficas son más cercanos o distantes respecto a los otros. Los valores máximos y los promedios de las diferentes distancias genéticas entre zonas geográficas (Tabla 2-10) incluyendo y excluyendo el grupo de híbridos son una evidencia más de la baja pero significativa diferenciación genética en *E. guineensis*, por lo tanto, al incluir los híbridos se observa un aumento en los valores de las distancias.

Las diferentes distancias muestran una mayor similitud entre las zonas de Cabinda, Benguela y Uige (Anexo F). La mayor divergencia se presenta entre Sumbe y Caixito (Anexo D). Por la ubicación geográfica y las condiciones climáticas que caracterizan cada zona de colecta se esperaba la mayor similitud entre Benguela y Sumbe ya que están ubicadas en zonas con altitud, condiciones de pluviosidad y temperatura similares, sin embargo, en los resultados de los análisis se observaron relativamente distantes, y por el contrario Benguela se observa más cercana de Uige. Este resultado muestra que las relaciones genéticas entre las zonas geográficas para los loci evaluados podrían ser explicadas por procesos de flujo génico entre ellas y no por ubicación geográfica o adaptaciones climáticas. Sin embargo, esto podría ser una evidencia que refleja que el muestreo fue sesgado a la característica contenido de pulpa en fruto, y no en base a la zona geográfica de colecta, por cual los resultados obtenidos expresan la ausencia de correlación significativa entre la ubicación geográfica y la diversidad colectada en cada zona.

En la topología mostrada en la Figura 2-8 se observa una población altamente heterogénea y heterocigota donde no se observan agrupaciones correlacionadas bajo un criterio aparente. Los agrupamientos con valores de bootstrap superiores al 50 por ciento, aunque son pocos pueden ser tenidos en cuenta a la hora de seleccionar germoplasma.

3. Conclusiones y recomendaciones

3.1 Conclusiones

Los 13 loci microsatélite y la metodología usada permitieron la determinación de la variabilidad, estructura y relaciones genéticas de una muestra representativa de la colección procedente de la República de Angola establecida y mantenida en el centro experimental La Vizcaína.

Los índices de variabilidad genética mostraron ser superiores a los descritos para las poblaciones Deli mejoradas reportadas en diferentes estudios. Las zonas geográficas evaluadas presentaron índices de diversidad relativamente homogéneos. Sin embargo, a nivel de familias se detectaron algunas con índices de diversidad relativamente mayores con respecto al promedio, lo cual indica que este material genético podría ser un recurso importante de nuevos genes para su introgresión en materiales mejorados y ampliación de su base genética.

Todos los análisis realizados mostraron resultados que permiten rechazar la hipótesis de investigación, la cual evalúa la estructura genética del germoplasma colectado en Angola basada en la ubicación geográfica, donde no mostraron agrupamientos bajo ningún criterio aparente, definiendo una población heterogénea y formada de individuos heterocigotos, la cual ha mantenido un alto flujo génico a través de los años determinando un efecto homogenizador entre zonas geográficas.

Este trabajo generó gran cantidad de información que, junto a las evaluaciones morfoagronómicas y su análisis a partir de la estructura genética establecida en la colección de germoplasma (componentes de las varianzas) puede efectivamente asistir el proceso de selección de parentales que generen segregaciones transgresivas para las características de interés y junto a las evaluaciones morfoagronómicos, permitirá un uso racional y sistemático de las colecciones de germoplasma (colección núcleo) y una selección de parentales que logren, combinaciones deseables entre Dura y Pisífera, con el objetivo de obtener materiales comerciales altamente productivos, y de esta manera, fortalecer el programa de mejoramiento genético de Cenipalma.

3.2 Recomendaciones

Este estudio se constituye como uno los primeros en evaluar una muestra representativa de la estructura poblacional y caracterización de la diversidad genética del germoplasma de poblaciones naturales de Cenipalma, por cual es importante, seguir desarrollando investigaciones que contemplen otros marcadores, poblaciones y estudios de asociación.

Se recomienda para futuros estudios diseñar estrategias de muestreo y realizarlo sistemáticamente definiendo el criterio principal de colecta para evitar sesgos. También es importante a la hora de colectar el material vegetal tomar datos acerca de los factores culturales que podrían estar jugando un papel importante en la diversidad y estructura genética de la especie en la región, como por ejemplo agentes dispersores, polinizadores y aprovechamiento de frutos.

Este estudio aporta la información de diversidad y estructura genética de la colección de Angola al programa de mejoramiento genético de Cenipalma, la cual se recomienda analizar en conjunto con la información obtenida de los estudios morfoagronómicos, bioquímicos, fisiológicos, de plagas y enfermedades permitiendo definir de manera más clara y contundente las relaciones genéticas y estructura poblacional logrando la selección correcta de parentales para planear nuevos cruces, y así involucrar eficientemente este germoplasma para ampliar la base genética y obtener materiales comerciales con mejores características agronómicas.

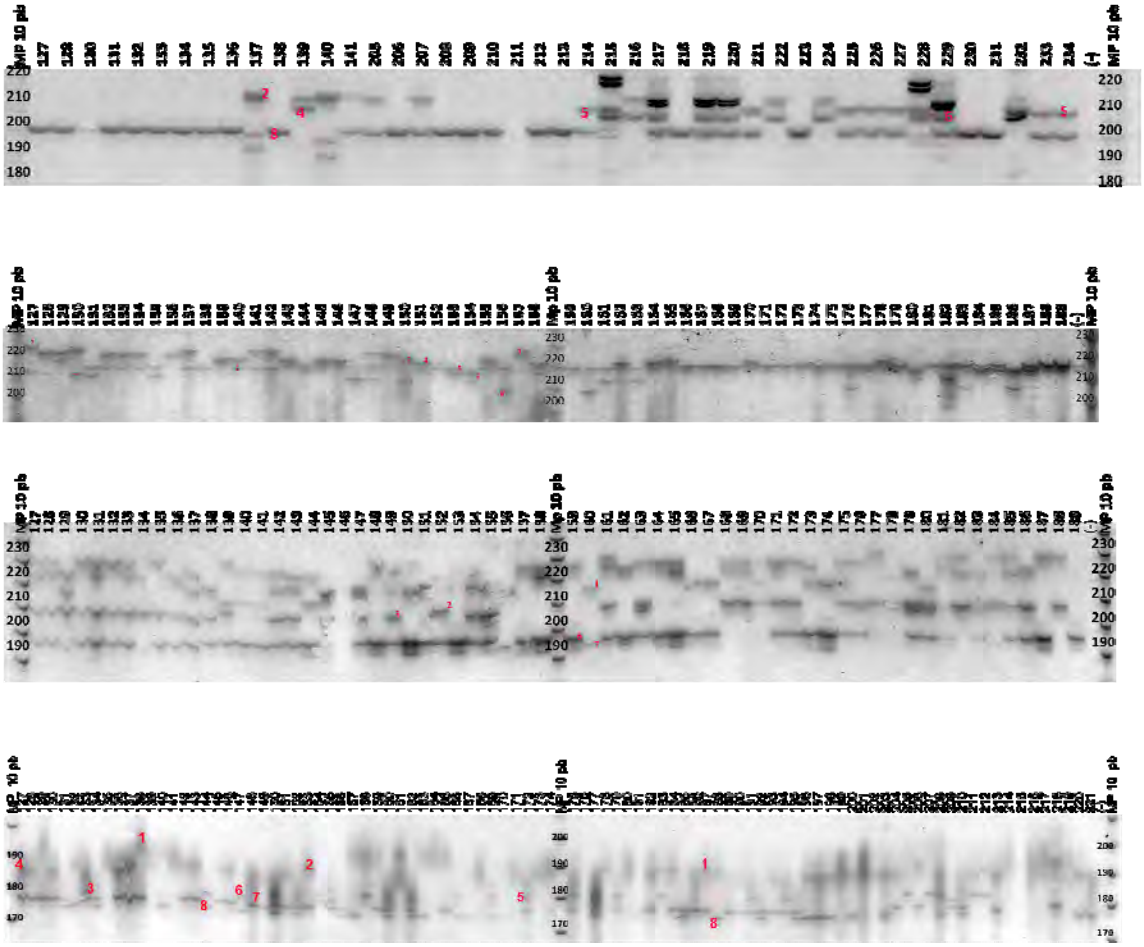
La mayor parte de la varianza se encuentra distribuida dentro de individuos (87%) y entre familias (12%) haciendo de estos grupos un recurso valioso de mejoramiento donde la selección debe hacerse cuidadosamente ya que soy muy variables entre ellos y esto puede repercutir de manera importante.

Así mismo, la información obtenida en este estudio y complementada con los resultados de los otros programas de Cenipalma podría en el futuro servir de lineamiento para formar una colección núcleo del germoplasma de Angola. Teniendo en cuenta únicamente los resultados de este trabajo se recomienda para propósitos de colectar la mayor diversidad genética en pocas entradas, realizar un exhaustivo y sistemático muestreo de las familias con mayor diversidad genética pero sin considerar las zonas geográficas.

A. Anexo: Información del grosor de cuesco en las familias y zonas geográficas analizadas en este estudio

FAMILIA	TIPO DE FRUTO	FAMILIA	TIPO DE FRUTO
CX1	DURA	CB5	DURA
CX2	DURA	CB6	TENERA
CX3	DURA	CB7	TENERA
CX4	DURA	CB8	DURA
CX5	DURAS Y TENERAS	CB9	DURA
CX6	DURAS Y TENERAS	CB10	DURA
CX7	DURA	CB11	DURAS Y TENERAS
CX8	DURA	BG1	DURA
CX9	DURA	BG2	DURA
SU1	TENERA	BG3	DURAS Y TENERAS
SU2	DURA	BG4	DURAS Y TENERAS
SU3	TENERA	BG5	DURA
SU4	DURA	BG6	DURA
SU5	DURAS Y TENERAS	UG1	TENERA
SU6	DURA	UG2	DURA
SU7	DURA	UG3	DURAS Y TENERAS
SU8	DURAS Y TENERAS	UG4	DURA
SU 9	ANG 4	UG6	DURAS Y TENERAS
CB1	DURA	UG7	DURA
CB2	DURA	UG8	DURA
CB3	DURA	UG9	DURAS Y TENERAS
CB4	DURA	UG10	DURAS Y TENERAS

B. Anexo: Perfil de los loci PR01, PR06, PR07 y PR16 en las muestras 127 a 139



C. Anexo: Test exacto de desequilibrio de ligamiento. Pares de loci con desequilibrio de ligamiento significativo.

LOCUS1	LOCUS1	P<0.05		LOCUS1	LOCUS1	P<0.05
PR01	PR02	0,03846		PR07	PR27	0,02244
PR01	PR06	0,04487		PR07	PR31	0,00705
PR01	PR07	0,03782		PR07	PR37	0,00064
PR01	PR16	0,01859		PR13	PR23	0,01410
PR01	PR23	0,00321		PR13	PR29	0,00064
PR02	PR06	0,01474		PR13	PR31	0,01154
PR02	PR07	0,00513		PR16	PR23	0,00064
PR02	PR16	0,00256		PR16	PR24	0,00449
PR02	PR23	0,01538		PR16	PR27	0,04744
PR02	PR24	0,00192		PR16	PR31	0,03077
PR02	PR27	0,00962		PR16	PR37	0,01346
PR02	PR28	0,04487		PR23	PR27	0,00064
PR02	PR37	0,00064		PR23	PR29	0,03013
PR06	PR07	0,00064		PR23	PR31	0,00385
PR06	PR16	0,00064		PR23	PR37	0,00385
PR06	PR23	0,00064		PR24	PR28	0,02756
PR06	PR24	0,03205		PR24	PR37	0,00128
PR06	PR27	0,01987		PR27	PR29	0,02115
PR06	PR28	0,02051		PR27	PR31	0,01987
PR06	PR31	0,04744		PR28	PR29	0,03590
PR06	PR37	0,01731		PR29	PR37	0,02628
PR07	PR16	0,04615		PR31	PR37	0,00192
		GLOBAL		**0.000641*** **		

Probabilidad de significancia P<0,05, P<0,01***

D. Anexo: F_{ST} y valor-p entre pares de zonas geográficas.

P<0.01 / F_{ST}	CX	SU	CB	BG	UG
CX	0	0,00000	0,00000	0,00000	0,00000
SU	0,04352	0	0,00000	0,00000	0,00000
CB	0,02644	0,02547	0	0,00000	0,00000
BG	0,0368	0,02156	0,02074	0	0,00000
UG	0,03761	0,03278	0,01988	0,01439	0

E. Anexo: F_{ST} y valor-p entre pares de familias

CX1	CX2	CX3	CX4	CX5	CX6	CX7	CX8	CX9	SU1	SU2	SU3	SU4	SU5	SU6
0.00000	0.00000	0.00020	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.17471	0.00000	0.00119	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.07933	0.10189	0.00000	0.01139	0.00020	0.00000	0.00673	0.00020	0.00000	0.00307	0.00000	0.00000	0.00000	0.00000	0.00000
0.13695	0.15654	0.04809	0.00000	0.00010	0.00000	0.00000	0.00000	0.00139	0.00059	0.00000	0.00000	0.00000	0.00010	0.00000
0.22263	0.16946	0.12089	0.12522	0.00000	0.00000	0.00000	0.00000	0.00020	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000
0.16457	0.17003	0.11930	0.14027	0.08958	0.00000	0.00059	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.16372	0.10504	0.05607	0.11403	0.08422	0.07018	0.00000	0.00000	0.00000	0.00020	0.00000	0.00000	0.00000	0.00000	0.00000
0.29316	0.23216	0.25609	0.28647	0.33379	0.32746	0.27721	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00010	0.00000
0.17938	0.17586	0.09579	0.07520	0.14904	0.14465	0.13121	0.34452	0.00000	0.00099	0.00000	0.00000	0.00000	0.00000	0.00000
0.17751	0.12696	0.07003	0.07159	0.08242	0.09428	0.07816	0.28405	0.07669	0.00000	0.00000	0.00020	0.00000	0.00000	0.00000
0.22301	0.20965	0.16754	0.15701	0.12836	0.11841	0.11046	0.31034	0.17508	0.07557	0.00000	0.00000	0.00000	0.00000	0.00000
0.18266	0.15097	0.08656	0.11921	0.14548	0.12751	0.10973	0.33646	0.10298	0.06616	0.08655	0.00000	0.00000	0.00000	0.00000
0.20759	0.22132	0.17502	0.16970	0.18188	0.10309	0.14469	0.28640	0.18151	0.12522	0.13775	0.15639	0.00000	0.00040	0.00000
0.26483	0.26807	0.21772	0.25852	0.21654	0.12035	0.19879	0.35071	0.26591	0.21748	0.25993	0.27219	0.12772	0.00000	0.00000
0.27242	0.24386	0.20201	0.22614	0.13517	0.07627	0.12978	0.40397	0.20976	0.14467	0.18906	0.18170	0.12849	0.14851	0.00000
0.25640	0.21218	0.18830	0.18406	0.20048	0.17464	0.15000	0.26031	0.20454	0.09717	0.11454	0.16199	0.08422	0.22864	0.18834
0.16584	0.21649	0.12594	0.16687	0.20854	0.15078	0.15334	0.34178	0.15219	0.13169	0.19751	0.16874	0.11532	0.15914	0.16820
0.14986	0.16208	0.10738	0.13949	0.13883	0.11000	0.11572	0.26760	0.14315	0.07096	0.12276	0.12586	0.03319	0.14421	0.12242
0.23310	0.22775	0.13216	0.17227	0.16100	0.14904	0.10506	0.33528	0.17479	0.07437	0.13711	0.16148	0.13161	0.22348	0.11765
0.14890	0.11332	0.08391	0.11450	0.14132	0.10824	0.07931	0.28244	0.11913	0.09251	0.15130	0.12785	0.09542	0.18054	0.13363
0.12515	0.14528	0.10390	0.11540	0.12212	0.07053	0.09187	0.26262	0.09784	0.07663	0.10300	0.11082	0.06477	0.12063	0.11517
0.16733	0.17164	0.11568	0.12922	0.09141	0.13943	0.12152	0.25869	0.14677	0.09074	0.12827	0.17280	0.18430	0.24429	0.19380
0.21403	0.24972	0.13916	0.16182	0.23862	0.18450	0.16713	0.33759	0.17456	0.20354	0.22545	0.20812	0.22106	0.23487	0.23869
0.18417	0.18184	0.10814	0.14252	0.16225	0.12514	0.09278	0.32347	0.13631	0.13191	0.18460	0.16590	0.14291	0.15864	0.12592
0.16191	0.19739	0.13110	0.15949	0.20830	0.14544	0.15660	0.31251	0.18539	0.13727	0.17852	0.13546	0.10109	0.17694	0.19504
0.16830	0.15255	0.10035	0.07004	0.11828	0.12794	0.08879	0.30991	0.11411	0.04808	0.12213	0.09922	0.10661	0.22009	0.16499
0.18561	0.16160	0.09471	0.12680	0.10406	0.11528	0.03954	0.33418	0.10679	0.10785	0.13939	0.15309	0.18285	0.25836	0.18426
0.09753	0.11597	0.09214	0.07970	0.15323	0.11634	0.12141	0.29959	0.12642	0.09460	0.18331	0.12107	0.13628	0.18374	0.15061
0.15063	0.13792	0.13734	0.10625	0.15729	0.11884	0.11447	0.28954	0.16744	0.10345	0.15149	0.15889	0.11584	0.19835	0.15949
0.11781	0.14843	0.08450	0.09562	0.07914	0.04594	0.08274	0.31841	0.07412	0.04718	0.10271	0.06706	0.05540	0.17477	0.07020
0.15903	0.11081	0.06961	0.09885	0.10453	0.07709	0.07545	0.24998	0.08772	0.05667	0.13765	0.09175	0.08005	0.12434	0.08686
0.18445	0.17546	0.10011	0.12181	0.09129	0.08309	0.07450	0.32229	0.09679	0.08620	0.14606	0.12211	0.07241	0.14103	0.09667
0.21942	0.22030	0.14996	0.13886	0.16807	0.13476	0.13853	0.36105	0.14808	0.09134	0.17067	0.16119	0.11996	0.21578	0.11456
0.12293	0.13270	0.09890	0.11747	0.11308	0.06147	0.07185	0.27314	0.09920	0.05418	0.14439	0.11824	0.07611	0.14136	0.08361
0.10006	0.14072	0.05490	0.10289	0.09233	0.06070	0.03638	0.22559	0.09703	0.05747	0.10911	0.09094	0.08507	0.16224	0.10913
0.09223	0.11968	0.04740	0.05923	0.13390	0.12367	0.07806	0.26759	0.08669	0.06452	0.13931	0.09432	0.13550	0.23017	0.17894
0.20523	0.21794	0.14347	0.12213	0.16440	0.11465	0.12615	0.34791	0.13277	0.11008	0.17903	0.16381	0.10320	0.19329	0.11010
0.17893	0.18255	0.07302	0.07756	0.14702	0.12867	0.10298	0.27438	0.09908	0.06554	0.13884	0.11758	0.12110	0.22039	0.16171
0.14689	0.15567	0.12549	0.11575	0.12946	0.07365	0.07868	0.25836	0.16209	0.08468	0.15993	0.16763	0.07464	0.14804	0.12820
0.10504	0.13072	0.12218	0.11460	0.14842	0.09124	0.13328	0.26813	0.12985	0.10127	0.16357	0.13605	0.09369	0.15722	0.17556
0.18543	0.19920	0.16932	0.13752	0.20790	0.16712	0.15847	0.28369	0.17550	0.11641	0.15880	0.15200	0.13266	0.24170	0.21969
0.15564	0.15240	0.10213	0.11892	0.08219	0.08211	0.10379	0.32111	0.08000	0.05744	0.13263	0.06864	0.09221	0.17560	0.11389
0.16559	0.22411	0.13508	0.10629	0.18426	0.15106	0.13471	0.38321	0.12690	0.10948	0.14196	0.11969	0.15042	0.28119	0.21586
0.11693	0.21596	0.10566	0.09289	0.18860	0.13478	0.12286	0.30588	0.11198	0.08045	0.16034	0.15935	0.11627	0.22477	0.19520

Anexo E: (Continuación)

SU7	SU8	SU9	CB1	CB2	CB3	CB4	CB5	CB6	CB7	CB8	CB9	CB10	CB11	BG1
0.00000	0.00000	0.00010	0.00010	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010
0.00000	0.00000	0.00010	0.00010	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000
0.00000	0.00010	0.00010	0.00000	0.00059	0.00020	0.00020	0.00000	0.00000	0.00000	0.00030	0.00030	0.00000	0.00000	0.00139
0.00010	0.00000	0.00000	0.00000	0.00000	0.00010	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000
0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00010	0.00000	0.00000	0.00020	0.00020	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00079
0.00000	0.00000	0.00020	0.00000	0.00010	0.00040	0.00040	0.00010	0.00000	0.00000	0.00010	0.01198	0.00000	0.00000	0.00010
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000
0.00000	0.00000	0.00010	0.00000	0.00000	0.00020	0.00020	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00069
0.00020	0.00010	0.00614	0.00198	0.00010	0.00050	0.00050	0.00000	0.00000	0.00000	0.00822	0.00000	0.00010	0.00000	0.02158
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00050
0.00000	0.00040	0.02930	0.00000	0.00000	0.00030	0.00030	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00059
0.00000	0.00000	0.00010	0.00000	0.00000	0.00010	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00050
0.00000	0.00000	0.00020	0.00000	0.00000	0.00010	0.00010	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000
0.21061	0.00000	0.00188	0.00000	0.00000	0.00010	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00010	0.00000
0.07972	0.10487	0.00000	0.00149	0.00010	0.00129	0.00129	0.00020	0.00079	0.00347	0.00050	0.00000	0.00000	0.00030	0.01257
0.09632	0.17426	0.05963	0.00000	0.00020	0.00040	0.00040	0.00000	0.00069	0.00000	0.00000	0.00000	0.00000	0.00000	0.00050
0.13782	0.15485	0.06079	0.11108	0.00000	0.00099	0.00099	0.00000	0.00050	0.00000	0.00000	0.00059	0.00050	0.00059	0.04485
0.10351	0.08603	0.06043	0.10618	0.06118	0.00000	0.00089	0.00000	0.00000	0.00000	0.00000	0.00010	0.00010	0.00129	0.01238
0.14474	0.22674	0.12539	0.13319	0.11972	0.08377	0.00000	0.00000	0.00000	0.00000	0.00000	0.00089	0.00000	0.00010	0.00010
0.24254	0.22480	0.17252	0.18912	0.14296	0.14590	0.21263	0.00000	0.00020	0.00000	0.00000	0.00000	0.00010	0.00000	0.00010
0.17211	0.16217	0.09456	0.08390	0.06900	0.07729	0.17115	0.09662	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010
0.15701	0.14731	0.06385	0.15492	0.09244	0.09883	0.20080	0.18007	0.11168	0.00000	0.00000	0.00000	0.00000	0.00000	0.00050
0.11238	0.14131	0.07150	0.12700	0.08468	0.08393	0.13128	0.22940	0.13335	0.13125	0.00000	0.00020	0.00000	0.00000	0.00000
0.18565	0.22163	0.15353	0.12943	0.09358	0.10079	0.09606	0.20731	0.12086	0.23174	0.12773	0.00000	0.00000	0.00000	0.00000
0.19702	0.11908	0.08816	0.17338	0.07001	0.08435	0.16274	0.20334	0.11535	0.09947	0.07852	0.16423	0.00000	0.00248	0.00337
0.11805	0.14662	0.10552	0.13110	0.06143	0.06947	0.13816	0.21791	0.12598	0.13397	0.07142	0.14711	0.04849	0.00000	0.00109
0.12745	0.10208	0.05202	0.09622	0.02931	0.03716	0.10835	0.18369	0.09736	0.09247	0.07465	0.08821	0.05137	0.06608	0.00000
0.11085	0.10456	0.04300	0.06273	0.03597	0.04790	0.10804	0.12332	0.05635	0.08932	0.07127	0.12089	0.06832	0.09277	0.04592
0.12975	0.11029	0.06098	0.10113	0.05621	0.06172	0.12078	0.16551	0.09504	0.10437	0.10010	0.10074	0.10589	0.11776	0.02617
0.15335	0.14990	0.08488	0.11318	0.11383	0.10711	0.20191	0.23488	0.12544	0.15973	0.11049	0.17803	0.11741	0.14106	0.06923
0.12403	0.09381	0.02449	0.09414	0.05301	0.06010	0.13411	0.17240	0.09568	0.08716	0.07404	0.13048	0.04230	0.07761	0.02834
0.12323	0.12099	0.05221	0.10096	0.05816	0.05070	0.07928	0.16500	0.10824	0.09375	0.07293	0.06887	0.06381	0.09283	0.02871
0.14010	0.14674	0.06950	0.13872	0.06065	0.09363	0.11025	0.20008	0.13659	0.11230	0.04396	0.09729	0.02752	0.07087	0.04419
0.15526	0.16723	0.09837	0.12010	0.06110	0.07882	0.18245	0.18562	0.09216	0.15600	0.12459	0.14894	0.09425	0.09740	0.04790
0.13879	0.14798	0.08453	0.09664	0.04454	0.08276	0.12744	0.15234	0.08784	0.14056	0.11391	0.12422	0.11258	0.11057	0.05794
0.10545	0.14116	0.07238	0.12287	0.07142	0.07857	0.13473	0.22459	0.13167	0.13243	0.09018	0.12887	0.06866	0.06194	0.04200
0.16125	0.10677	0.08534	0.15393	0.08947	0.07441	0.16341	0.22174	0.13418	0.10492	0.07173	0.17051	0.05993	0.06916	0.05826
0.13998	0.20392	0.13180	0.14810	0.12082	0.11139	0.17505	0.22056	0.14338	0.13115	0.12704	0.19979	0.13370	0.13237	0.10770
0.11953	0.12250	0.05742	0.12652	0.07786	0.07364	0.11377	0.20148	0.13404	0.08729	0.07676	0.14106	0.07697	0.11899	0.00781
0.20047	0.14644	0.15801	0.15145	0.12757	0.08766	0.20239	0.24437	0.14259	0.16544	0.08618	0.13527	0.11023	0.09415	0.06662
0.16433	0.08369	0.08919	0.09906	0.09646	0.06675	0.15173	0.20951	0.11837	0.13643	0.06944	0.14163	0.08839	0.07290	0.06105

Anexo E: (Continuación)

BG2	BG3	BG4	BG5	BG6	UG1	UG2	UG3	UG4	UG6	UG7	UG8	UG9	UG10
0.00000	0.00000	0.00000	0.00000	0.00020	0.00020	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00089	0.00188	0.00000	0.00000	0.00485	0.00762	0.00000	0.00010	0.00000	0.00050	0.00000	0.00000	0.00000	0.00010
0.00020	0.00010	0.00000	0.00000	0.00000	0.00020	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00020	0.00000
0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010
0.00000	0.00010	0.00000	0.00020	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00030	0.00000	0.00000	0.00475	0.00000	0.00000	0.00000	0.00020	0.00000	0.00000	0.00010	0.00000	0.00000
0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00010
0.00020	0.00050	0.00000	0.00010	0.00010	0.00000	0.00010	0.00010	0.00000	0.00000	0.00000	0.00020	0.00000	0.00000
0.00347	0.00089	0.00030	0.00079	0.00149	0.00050	0.00000	0.00030	0.00000	0.00000	0.00000	0.00673	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000
0.00010	0.00020	0.00000	0.00010	0.00010	0.00000	0.00000	0.00010	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010
0.00000	0.00010	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00010	0.00040	0.00010	0.00010	0.00010	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00020
0.00228	0.00525	0.00010	0.04544	0.00079	0.00178	0.00000	0.00020	0.00129	0.00277	0.00010	0.00248	0.00000	0.00000
0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000
0.00624	0.00634	0.00000	0.00000	0.00000	0.00020	0.00000	0.00505	0.00059	0.00020	0.00000	0.00040	0.00000	0.00010
0.00040	0.00149	0.00000	0.00010	0.00119	0.00010	0.00000	0.00010	0.00000	0.00000	0.00000	0.00030	0.00000	0.00000
0.00040	0.00020	0.00000	0.00000	0.00079	0.00010	0.00000	0.00000	0.00000	0.00010	0.00000	0.00010	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000
0.00059	0.00000	0.00000	0.00000	0.00000	0.00000	0.00020	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000
0.00020	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00010	0.00000	0.00000
0.00010	0.00040	0.00000	0.00000	0.00010	0.00109	0.00010	0.00000	0.00000	0.00059	0.00000	0.00000	0.00000	0.00000
0.00000	0.00000	0.00000	0.00000	0.00020	0.00000	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.00000	0.00010	0.00000	0.00228	0.00000	0.02327	0.00010	0.00000	0.00010	0.00099	0.00000	0.00010	0.00000	0.00000
0.00000	0.00000	0.00000	0.00020	0.00010	0.00030	0.00010	0.00000	0.00208	0.00366	0.00000	0.00010	0.00000	0.00010
0.00257	0.07880	0.00010	0.02930	0.02237	0.00772	0.00317	0.00010	0.00921	0.00386	0.00000	0.28235	0.00020	0.00030
0.00000	0.15226	0.00000	0.00317	0.00000	0.00000	0.00000	0.00040	0.00030	0.00386	0.00000	0.00030	0.00010	0.00000
0.00753	0.00000	0.00010	0.00208	0.00475	0.00030	0.00020	0.00020	0.00119	0.00010	0.00000	0.04594	0.00010	0.00000
0.06103	0.07438	0.00000	0.00089	0.00000	0.00000	0.00129	0.00000	0.00010	0.00000	0.00000	0.00000	0.00000	0.00010
0.03010	0.05401	0.04715	0.00000	0.03138	0.00168	0.00000	0.00010	0.00535	0.00208	0.00000	0.01673	0.00000	0.00050
0.05621	0.04522	0.10845	0.01680	0.00000	0.01733	0.00020	0.00000	0.00099	0.00000	0.00000	0.00030	0.00000	0.00000
0.07566	0.09494	0.11569	0.04897	0.03091	0.00000	0.00000	0.00000	0.00040	0.00020	0.00000	0.00010	0.00000	0.00000
0.07521	0.07456	0.03594	0.05995	0.11446	0.10848	0.00000	0.00099	0.00020	0.00000	0.00000	0.00000	0.00000	0.00000
0.04487	0.07881	0.05399	0.08069	0.07363	0.08079	0.04480	0.00000	0.00010	0.00000	0.00010	0.00000	0.00000	0.00000
0.05895	0.05660	0.08361	0.03999	0.05099	0.07482	0.07626	0.08479	0.00000	0.00455	0.00000	0.00149	0.00000	0.00010
0.04617	0.09718	0.12289	0.05574	0.08567	0.07360	0.13127	0.11600	0.05171	0.00000	0.00000	0.00149	0.00010	0.00129
0.08903	0.12445	0.15440	0.12558	0.11633	0.12439	0.15334	0.10469	0.09721	0.07481	0.00000	0.00000	0.00010	0.00000
0.04821	0.02594	0.11611	0.02931	0.05008	0.06753	0.11085	0.11015	0.06317	0.07150	0.12490	0.00000	0.00000	0.00000
0.13372	0.14472	0.13453	0.11720	0.11792	0.10799	0.10811	0.12615	0.15499	0.11123	0.14495	0.14793	0.00000	0.00267
0.09250	0.11733	0.10206	0.05703	0.06742	0.07109	0.10184	0.06873	0.08575	0.06937	0.11293	0.12140	0.04146	0.00000

Probabilidad de significancia P<0,01

F. Anexo: Distancias genéticas entre pares de zonas geográficas

DSG	CX	0,00000					
	SU	0,05572	0,00000				
	CB	0,03561	0,03328	0,00000			
	BG	0,05102	0,02914	0,02683	0,00000		
	UG	0,05334	0,04243	0,0273	0,02015	0,00000	
	HB	0,43981	0,43097	0,42698	0,40433	0,46746	0,0000
dm^2	CX	0,00000					
	SU	0,42232	0,00000				
	CB	0,11285	0,36741	0,00000			
	BG	0,13440	0,28900	0,08457	0,00000		
	UG	0,16408	0,33243	0,0759	0,05893	0,0000	
	HB	887,631	744,883	906,704	812,451	797,244	0,0000
DAN	CX	0,00000					
	SU	0,06001	0,00000				
	CB	0,04344	0,03922	0,00000			
	BG	0,05569	0,04094	0,03361	0,00000		
	UG	0,05161	0,03941	0,03252	0,03354	0,0000	
	HB	0,47990	0,46072	0,44692	0,44007	0,4859	0,0000
ASD	CX	0,00000					
	SU	904,886	0,00000				
	CB	913,132	863,119	0,00000			
	BG	837,865	777,857	796,607	0,00000		
	UG	930,860	872,227	885,767	806,649	0,00000	
	HB	2,071,106	1,865,400	2,046,415	1,872,235	1,952,068	0,0000

DSG: distancia genética estándar de Nei (1972), DN: distancia corta de Nei (1983), dm^2: $d\delta^2$ (Goldstein 1995b), ASD: distancia cuadrática media (Goldstein 1995b y Slatkin 1995)

G. Anexo G: Distancia genética de Nei (1972) entre pares de familias

	CX1	CX2	CX3	CX4	CX5	CX6	CX7	CX8	CX9	SU1	SU2	SU3	SU4	SU5
CX1														
CX2	0,187													
CX3	0,082	0,103												
CX4	0,166	0,185	0,055											
CX5	0,259	0,175	0,128	0,147										
CX6	0,190	0,186	0,135	0,186	0,098									
CX7	0,195	0,111	0,062	0,148	0,089	0,080								
CX8	0,249	0,172	0,207	0,272	0,293	0,303	0,246							
CX9	0,220	0,205	0,112	0,094	0,171	0,185	0,166	0,348						
SU1	0,227	0,144	0,082	0,094	0,093	0,122	0,097	0,268	0,096					
SU2	0,268	0,239	0,195	0,195	0,139	0,138	0,124	0,293	0,214	0,084				
SU3	0,214	0,164	0,094	0,144	0,161	0,150	0,126	0,345	0,118	0,074	0,092			
SU4	0,242	0,250	0,203	0,216	0,201	0,115	0,169	0,239	0,223	0,151	0,151	0,178		
SU5	0,251	0,245	0,196	0,273	0,186	0,090	0,182	0,250	0,273	0,214	0,258	0,284	0,097	
SU6	0,301	0,249	0,209	0,265	0,124	0,069	0,127	0,362	0,231	0,150	0,193	0,189	0,118	0,107
SU7	0,303	0,227	0,211	0,224	0,216	0,198	0,167	0,201	0,246	0,105	0,116	0,178	0,081	0,197
SU8	0,181	0,241	0,136	0,207	0,236	0,172	0,178	0,311	0,178	0,156	0,228	0,192	0,120	0,128
SU9	0,159	0,166	0,113	0,167	0,142	0,118	0,128	0,205	0,165	0,078	0,129	0,136	0,031	0,109
CB1	0,272	0,255	0,143	0,216	0,171	0,168	0,117	0,302	0,206	0,082	0,148	0,182	0,141	0,198
CB2	0,166	0,118	0,093	0,146	0,157	0,125	0,093	0,247	0,140	0,114	0,174	0,146	0,106	0,159
CB3	0,155	0,175	0,132	0,170	0,150	0,090	0,122	0,247	0,134	0,110	0,125	0,137	0,072	0,095
CB4	0,177	0,177	0,120	0,152	0,090	0,156	0,134	0,197	0,171	0,104	0,140	0,200	0,205	0,220
CB5	0,204	0,241	0,122	0,159	0,229	0,173	0,159	0,254	0,170	0,214	0,228	0,212	0,212	0,178
CB6	0,221	0,206	0,123	0,188	0,186	0,149	0,108	0,312	0,172	0,172	0,228	0,203	0,165	0,132
CB7	0,150	0,185	0,118	0,160	0,200	0,136	0,152	0,253	0,188	0,131	0,176	0,128	0,086	0,135
CB8	0,198	0,168	0,116	0,092	0,131	0,162	0,108	0,285	0,143	0,063	0,144	0,116	0,121	0,206
CB9	0,225	0,180	0,107	0,166	0,113	0,139	0,044	0,324	0,130	0,139	0,162	0,184	0,221	0,256
CB10	0,108	0,129	0,111	0,116	0,191	0,157	0,164	0,274	0,175	0,139	0,251	0,159	0,174	0,168
CB11	0,198	0,166	0,188	0,157	0,205	0,165	0,158	0,290	0,251	0,153	0,198	0,214	0,146	0,198
BG1	0,146	0,177	0,107	0,139	0,093	0,066	0,110	0,323	0,102	0,069	0,129	0,081	0,064	0,161
BG2	0,197	0,122	0,081	0,134	0,117	0,093	0,093	0,218	0,112	0,074	0,167	0,106	0,089	0,099
BG3	0,216	0,199	0,119	0,166	0,102	0,101	0,098	0,302	0,115	0,113	0,174	0,142	0,087	0,126
BG4	0,261	0,252	0,173	0,175	0,186	0,157	0,166	0,342	0,177	0,108	0,197	0,188	0,132	0,192
BG5	0,137	0,145	0,114	0,157	0,125	0,072	0,087	0,241	0,120	0,068	0,172	0,138	0,085	0,118
BG6	0,114	0,163	0,062	0,143	0,105	0,075	0,044	0,182	0,128	0,077	0,130	0,107	0,098	0,142
UG1	0,102	0,133	0,053	0,077	0,158	0,160	0,097	0,236	0,112	0,086	0,171	0,112	0,163	0,229
UG2	0,251	0,258	0,169	0,156	0,186	0,135	0,153	0,339	0,163	0,139	0,215	0,196	0,111	0,170
UG3	0,222	0,218	0,083	0,100	0,174	0,163	0,129	0,248	0,123	0,083	0,164	0,138	0,142	0,214
UG4	0,175	0,179	0,153	0,159	0,151	0,089	0,096	0,223	0,224	0,114	0,199	0,214	0,083	0,121
UG6	0,123	0,150	0,156	0,166	0,183	0,117	0,183	0,247	0,183	0,146	0,215	0,174	0,112	0,135
UG7	0,244	0,258	0,232	0,199	0,279	0,230	0,226	0,269	0,253	0,169	0,206	0,198	0,165	0,249
UG8	0,174	0,162	0,111	0,145	0,081	0,091	0,118	0,292	0,089	0,064	0,146	0,072	0,095	0,148
UG9	0,221	0,313	0,185	0,156	0,251	0,219	0,194	0,466	0,181	0,164	0,182	0,152	0,199	0,326
UG10	0,141	0,284	0,133	0,128	0,248	0,184	0,165	0,308	0,150	0,110	0,204	0,207	0,144	0,229

Anexo G: (Continuación)

	SU6	SU7	SU8	SU9	CB1	CB2	CB3	CB4	CB5	CB6	CB7	CB8	CB9	CB10	CB11
SU7	0,178														
SU8	0,184	0,231													
SU9	0,108	0,074	0,104												
CB1	0,104	0,096	0,190	0,056											
CB2	0,129	0,151	0,175	0,062	0,116										
CB3	0,116	0,117	0,099	0,066	0,126	0,075									
CB4	0,190	0,147	0,261	0,126	0,136	0,129	0,097								
CB5	0,208	0,231	0,215	0,150	0,170	0,127	0,141	0,199							
CB6	0,120	0,195	0,188	0,101	0,088	0,077	0,097	0,198	0,082						
CB7	0,169	0,137	0,132	0,051	0,139	0,081	0,087	0,192	0,148	0,099					
CB8	0,164	0,123	0,162	0,075	0,140	0,095	0,114	0,141	0,238	0,164	0,121				
CB9	0,193	0,216	0,279	0,175	0,142	0,106	0,134	0,099	0,207	0,146	0,246	0,155			
CB10	0,158	0,251	0,142	0,100	0,212	0,086	0,127	0,191	0,210	0,145	0,090	0,099	0,223		
CB11	0,178	0,137	0,188	0,130	0,166	0,077	0,108	0,173	0,248	0,170	0,130	0,094	0,208	0,075	
BG1	0,063	0,144	0,121	0,060	0,110	0,032	0,058	0,126	0,194	0,123	0,076	0,089	0,115	0,069	0,091
BG2	0,078	0,121	0,119	0,045	0,067	0,042	0,064	0,122	0,113	0,064	0,075	0,085	0,155	0,093	0,132
BG3	0,089	0,146	0,126	0,068	0,103	0,061	0,081	0,128	0,154	0,113	0,090	0,112	0,118	0,128	0,159
BG4	0,102	0,164	0,164	0,086	0,116	0,128	0,134	0,229	0,231	0,145	0,149	0,124	0,217	0,149	0,189
BG5	0,074	0,137	0,103	0,023	0,098	0,057	0,078	0,149	0,165	0,112	0,073	0,084	0,160	0,050	0,106
BG6	0,105	0,138	0,142	0,054	0,115	0,072	0,072	0,083	0,163	0,136	0,079	0,090	0,084	0,085	0,138
UG1	0,194	0,160	0,175	0,076	0,162	0,072	0,134	0,121	0,209	0,176	0,105	0,050	0,121	0,035	0,101
UG2	0,103	0,171	0,193	0,103	0,131	0,069	0,098	0,212	0,177	0,106	0,147	0,151	0,183	0,120	0,129
UG3	0,168	0,158	0,174	0,091	0,106	0,050	0,115	0,146	0,142	0,107	0,134	0,145	0,157	0,158	0,161
UG4	0,126	0,116	0,166	0,077	0,141	0,084	0,110	0,155	0,237	0,167	0,121	0,116	0,167	0,095	0,089
UG6	0,193	0,196	0,125	0,097	0,192	0,111	0,110	0,205	0,246	0,175	0,090	0,092	0,244	0,086	0,103
UG7	0,262	0,166	0,272	0,160	0,187	0,160	0,168	0,227	0,241	0,190	0,123	0,178	0,294	0,207	0,208
UG8	0,104	0,120	0,130	0,056	0,133	0,083	0,087	0,121	0,193	0,156	0,071	0,088	0,166	0,095	0,155
UG9	0,266	0,267	0,189	0,206	0,195	0,172	0,138	0,281	0,290	0,200	0,170	0,117	0,193	0,177	0,147
UG10	0,224	0,200	0,094	0,102	0,114	0,121	0,099	0,187	0,228	0,156	0,133	0,089	0,195	0,130	0,108

	BG1	BG2	BG3	BG4	BG5	BG6	UG1	UG2	UG3	UG4	UG6	UG7	UG8	UG9	UG10
BG1	0,058														
BG2	0,025	0,013													
BG3	0,085	0,069	0,079												
BG4	0,032	0,035	0,060	0,052											
BG5	0,039	0,072	0,058	0,132	0,019										
BG6	0,060	0,098	0,119	0,141	0,057	0,037									
UG1	0,062	0,090	0,087	0,039	0,069	0,147	0,136								
UG2	0,078	0,056	0,098	0,061	0,099	0,097	0,104	0,052							
UG3	0,059	0,074	0,070	0,095	0,048	0,066	0,098	0,091	0,108						
UG4	0,081	0,057	0,126	0,157	0,069	0,119	0,099	0,176	0,163	0,068					
UG6	0,162	0,122	0,170	0,204	0,176	0,171	0,182	0,209	0,143	0,135	0,106				
UG7	0,007	0,052	0,028	0,126	0,029	0,055	0,079	0,123	0,128	0,070	0,082	0,159			
UG8	0,099	0,202	0,201	0,179	0,166	0,182	0,160	0,146	0,184	0,243	0,176	0,227	0,200		
UG9	0,089	0,128	0,156	0,125	0,072	0,093	0,095	0,133	0,089	0,118	0,101	0,164	0,152	0,058	
UG10	0,067	0,082	0,104	0,119	0,080	0,107	0,119	0,133	0,136	0,127	0,116	0,183	0,176	0,058	0,000

4. Bibliografía

Adin, A., J.C. Weber, C. Sotelo-Montes, H. Vidaurre, B. Vosman y M.J.M. Smulders. 2004. Genetic differentiation and trade among populations of peach palm (*Bactris gasipaes* Kunth) in the Peruvian Amazon—implications for genetic resource management. Theoretical and Applied Genetics 108, 1564–1573

Aranzana, M. J., E. Abbassi, W. Howad y P. Arús. 2010. Genetic variation, population structure and linkage disequilibrium in peach commercial varieties. BMC Genetics 69, 1-11. En: http://www.biomedcentral.com/1471-2156/11/69

Bakoumé C., R. Wickneswari, N. Rajanaidu, A. Kushairi, P. Amblard y N. Billote. 2007. Allelic diversity of natural oil palm (*Elaeis guineensis* Jacq.). Populations detected by microsatellites markers. Implications in conservation. Palmas 28 (1), 149- 158.

Barcelos, E., P. Amblard, J. Berthaud y M. Seguin. 2002. Genetic diversity and relationship in American and African oil palm as revealed by RFLP and AFLP molecular markers. Pesquisa agropecuária brasileira 37 (8), 1105-1114.

Bernal, NF. 2001. El cultivo de la palma de aceite y su beneficio. Fedepalma.186 p.

Billote, N., A. M. Risterucci, E. Barcelos, J. L. Noyer, P. Amblard y F. C. Baurens, 2001. Development, characterization, and across-taxa utility of oil palm (*Elaeis guineensis* Jacq.) microsatellite markers. Genome 44, 413-425.

Billote, N., N. Marseillac, A-M. Risterucci, B. Adon, P. Brottier, F-C. Baurens, R. Singh, A. Herrán, H. Asmady, C. Billlot, P. Amblard, T. Durand-Gasselin, B. Courtois, D. Asmono, S.C. Cheah, W. Rohde, E. Ritter y A. Charrier. 2005. Microsatellite-based high density linkage map in oil palm (*E. guineensis* Jacq). Theoretical and Applied Genetics 110, 754-765.

Billote N., P. Amblard, T. Durand-Gasselin, A. Flori, B. Nouy, F. Potier, F. Richaud, A. Rival, T. Beule, F. Morcillo, B. Adon, MF. Jourjon, F. Aberlenc –Bertossi, H. Adam, Y. Duval, S. Jounnic, J. Tregear y Z. Hayun. 2007. Oil palm biotechnology at CIRAD: Breeding y seed production. Palmas 28 (I), 123-143.

Botstein, D., R. L. White, M. Skolnick y R. W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics 32, 314-331

Cavalli-Sforza, L. L. y W. F. Edwards. 1967. Phylogenetic analysis: Models and estimation procedures. American Journal of Human Genetics 19, 233-257.

Charlesworth B., D. Charlesworth y N.H. Barton. 2003. The effects of genetic and geographic structure on neutral variation. The Annual Review of Ecology, Evolution and Systematic 34, 99–125.

Cheah, SC., SN. Akmar Abdullah, LC-L. Ooi, R. Abdul Rahman y M. Madon. 1991. Detection of DNA variability in the oil palm using RFLP probes. Proc PORIM International palm oil Conference, Malaysia, 144-150.

Chowdhury, MKU. 1993. Detection of variation in tissue culture-derived normal and abnormal clones of oil palm using RAPD method. Proc 1993 ISOPB Symposium Malaysia 190-194.

Cochard, B., B. Adon, , S. Rekima, N. Billotte, R. Desmier, A. Koutou, B. Nouy, A. Omoré, A. Purba, J. Glazsmann y J. Noyer. 2009. Geographic and genetic structure of african oil palm diversity suggests new approaches to breeding. Tree Genetics and Genomes 5, 493–504

Corley, R. y P. Tinker. 2003. La palma de aceite. Cuarta edición (versión en español traducida por Fedepalma). Editorial Blackwell Publishing Ltd. Oxford.

Dieringer, D. y C. Schlštterer. 2003. Microsatellite analyzer (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes 3 (1), 167-169. En: http://i122server.vu-wien.ac.at/MSA/info.html/MSA_info.html; consultada: octubre 2010.

El-Assar, A.M., R. Krueger, P. Devanand y C-C. Chao. 2005. Genetic analysis of Egyptian date (*Phoenix dactylifera* L.) accessions using AFLP markers Genetic. Resources and Crop Evolution 52, 601–607

Elsibli S. y H. Korpelainen. 2008. Microsatellite markers reveal high genetic diversity in date palm *(Phoenix dactylifera* L.) germplasm from Sudan. Genética 134, 251–260

Evanno, G., S. Regnaut y J. Goudet. 2005. Detecting the number of cluster of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14, 2611-2620

Excoffier, L., G. Laval y S. Schneider. 2007. Arlequin ver. 3.1.1: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1, 47-50. En: http://cmpg.unibe.ch/software/arlequin3/; consulta: octubre 2010.

FEDEPALMA. 2007. Aceite de palma en la salud humana. Editorial Kimpres Ltda. Bogotá. Colombia. 122 p.

FEDEPALMA. 2010. Anuario estadístico 2010. La agroindustria de la palma de aceite en Colombia y en el mundo 2005-2010. Bogotá. Colombia. 161 pp.

Felsenstein, J. 2005. PHYLIP (Phylogeny Inference Package) version 3.69. Department of genome sciences. University of Washington, Seattle, USA. En: http://evolution.genetics.washington.edu/phylip.html; consulta: noviembre 2010.

Ferreira, M. y D. Grattapaglia. 1998. Introducción al uso de loci moleculares en el análisis genético. EMBRAPA-CENARGEN, Brasil. 221p.

Flint-García, S., M. J. Thornsberry y E. S. Buckler IV. 2003. Structure of linkage disequilibrium in plants. Annual Review of Plant Biology 54, 357–74.

Goldstein, D.B., A. Ruiz-Linares, L. L. Cavalli-Sforza y M. W. Feldman. 1995a. Genetic absolute dating based on microsatellites and the origin of modern humans. Proceedings of the National Academy of Sciences of USA 92, 6723-6727.

Goldstein, D.B., A. Ruiz-Linares, L. L. Cavalli-Sforza y M. W. Feldman. 1995b. An evaluation of genetic distances for use with microsatellite loci. Genetics 139, 463-471.

Goldstein, D. B. y D. D. Pollock. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. Journal of Heredity 88**,** 335-342

Guo, S. W. y E. A. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48, 361-372.

Hartl, D.L. y A. G. Clark. 1997. Principles of population genetics. Tercera edición. Editorial Sinauer Associates, Sunderland Massachusetts. 542 pp

Hayati, A., R. Wickneswari, I. Maizura y N. Rajanaidu. 2004. Genetic diversity of oil palm (*Elaeis guineensis* Jacq.) germplasm collections from Africa: implications for improvement and conservation of genetic resources. Theoretical and Applied Genetics 108, 1274-1284.

Jack, PL., TAF. Dimitrijevic y S. Mayes. 1995. Assessment of nuclear, mitochondrial and chloroplast RFLP markers in oil palm (*E. guineensis* Jacq.). Theoretical and Applied Genetics 90, 643-649.

Jensen, J.L., A.J. Bohonak y S.T. Kelley. 2005. Isolation by distance, web service. En: http://ibdws.sdsu.edu/; consulta: octubre 2010.

Kaeuffer, R., D. Réale, DW Coltman y D. Pontier. 2007. Detecting population structure using STRUCTURE software: effect of background linkage disequilibrium. Heredity 99, 374–380.

Laurentin, H. 2009. Data analysis for molecular characterization of plant genetic resources 56 (2), 277-292.

Maizura I., N. Rajanaidu, A.H. Zakri y S.C. Cheah. 2006. Assessment of genetic diversity in oil palm (*Elaeis guineensis* Jacq.) using Restriction Fragment Length Polymorphism (RFLP). Genetic Resources and Crop Evolution 53, 187–195.

Mayes, S., CM. James, SF. Horner, PL. Jack y RHV Corley. 1996. The application of restriction fragment length polymorphism for the genetic fingerprinting of oil palm (*E. guineensis* Jacq). Molecular Breeding 2, 175-180.

Mayes, S., PL. Jack, DF. Marshall y RHV Corley. 1997. Construction of a RFLP genetic linkage map for oil palm. (*Elaeis guineensis* Jacq.). Genome 40 (1), 116-122.

Ministerio de agricultura y protección social. 2010. Producto interno bruto agropecuario. En: Agronet-Indicadores económicos, http://www.agronet.gov.co/agronetweb/Boletines/tabid/75/Default.aspx. 4 p.; consulta: noviembre 2010.

Mohammadi, S.A. y B M. Prasanna. 2003. Analysis of genetic diversity in crop plants-Salient statistical tools and considerations. Crop Science 43(4), 1235-1248.

Montoya, C., D. Arias, L. Rey y P. Rocha. 2005. Caracterización molecular de materiales *Elaeis guineensis* Jacq. procedentes de Angola. Fitotecnia Colombiana 5 (2), 1-10.

Moretzsohn, M.C., C.D.M. Nunes, M.E. Ferreira y D. Grattapaglia. 2000. RAPD linkage mapping of the shell thickness locus in oil palm (*Elaeis guineensis* Jacq.). Theoretical and Applied Genetics 100, 63–70.

Nei, M. 1972. Genetic distance between populations. American Naturalist 106, 283-292.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences of USA 70, 3321- 3323.

Nei, M. y RK. Chesser. 1983. Estimation of fixation indexes and gene diversities. Annals of Human Genetics 47, 253-259.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press. New York, USA. 164 p.

Ochoa, I, V. Villegas y S. Beebe. 1997. Identification of RAPD molecular markers associated with resistance to the bud rot complex in oil palm (*Elaeis guineensis* Jacq.). Palmas 18(1), 33-38.

Paranjothy, K., LM. Ong y S. Sharifah. 1993. DNA and protein changes in relation to clonal abnormalities. Proc. 1993 ISOPB Symposium, Malaysia. pp. 86-97.

Peakall, R. y P. Smouse. 2006. Program note GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6(1), 288-295.

Perrier, X. y J.P. Jacquemoud-Collet. 2009. DARwin (Dissimilarity Analysis and Representation for Windows) software, version 5.0. 158. En: http://darwin.cirad.fr/darwin; consulta: noviembre 2010.

Pritchard, J., X. Wen y D. Falush. 2010. Structure software: Version 2.3. University of Chicago and University of Oxford. En: http://pritch.bsd.uchicago.edu/software; consulta: octubre 2010.

Purba, A. R., J. L. Noyer, L. Bandouin, X. Perrier, S. Hamon y PJL. Lagoda. 2000. A new aspect of genetic diversity of Indonesian oil palm (*Elaeis guineensis* Jacq.) revealed by isoenzyme and AFLP markers and its consequences for breeding. Theoretical and Applied Genetics 101, 956-961.

QIAGEN Sciences, 2006. DNeasy® Plant Mini Kit (250), Maryland 20874, USA.

Singh, R., C. Suan-Choo y A. Rahimah. 1998. Generation of molecular markers in oil palm (*Elaeis guineensis*) using AFLP^{TM} analysis. Plant Biotechnology 20, 26-27.

Singh, R., M. Noorhariza, T. Ngoot-Chin, R. Rozana, T. Soon-Guan, L. Eng-Ti, I. Maizura y C. Suan-Choo. 2008. Exploiting an oil palm EST database for the development of gene-derived SSR markers and their exploitation for assessment of genetic diversity. Biology 63 (2), 227-235.

Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139, 457-462.

Rashid, O. y FH. Shah. 1996. The isolation, characterization and sequencing of cDNA clones coding for stearyl-ACP desaturase gene from oil palm (*Elaeis guineensis*). Proc 1996 PORIM Int Palm Oil Congress Malaysia 575-582

Rey, L., I. Ayala, W. Delgado y P. Rocha. 2003. Colecta de material genético de Palma Americana Nolí *Elaeis oleifera* (H.B.K.) Cortez en el Trapecio Amazónico. Ceniavances 101. 4 p.

Rey, L., P.L. Gómez, I. Ayala, W. Delgado y P. Rocha. 2004. Cenipalma: Características de importancia en el sector palmicultor. Palmas 25 (2), 39-48.

Rey, L. 2007. La variabilidad del germoplasma y su relación con el éxito de un programa de mejoramiento en palma de aceite. Palmas 28 (1) 166-175.

Rival, A., T. Beule, P. Barre, S. Hamon, Y. Duval y M. Noirot. 1997. Comparative flow cytometric estimation of nuclear DNA content in oil palm (*Elaeis guineensis* Jacq.) tissue cultures and seed-derived plants. Plant Cell Reports 16, 884–887.

Rival, A., L. Bertrand, T. Beule, M.C. Combes, P. Trouslot y P. Lashermes. 1998. Suitability of RAPD analysis for the detection of somaclonal variants in oil palm (*Elaeis guineensis* Jacq.). Plant Breeding 117, 73–76.

Rodrígues, D.P., S.A.A. Filho y C.R. Clement. 2004. Molecular marker-mediated validation of morphologically defined landraces of Pejibaye (*Bactris gasipaes*) and their phylogenetic relationships. Genetic Resources and Crop Evolution 51 (8), 871-882.

Shah, F.H., O. Rashid, A.J. Simons y A. Dundson. 1994. The utility of RAPD markers for determination of genetic variation in oil palm *(Elaeis guineensis).* Theoretical and Applied Genetics 89 (6), 713-718.

Shah, FH. y TS. Cha. 2000. A mesocarp and species-specific cDNA clone from oil palm encodes for sesquiterpene synthase. Plant Science 154, 153-160.

Singh, R. y SC. Cheah. 2000. Differential gene expression during flowering in the oil palm (*Elaeis guinnensis* Jacq.). Plan Cell Reports 19, 804-809.

Singh, R., M. Noorhariza, T. Ngoot-Chin, R. Rozana, T. Soon-Guan, L. Eng-Ti, I. Maizura y C. Suan-Choo. 2008. Exploiting an oil palm EST database for the development of gene-derived SSR markers and their exploitation for assessment of genetic diversity. Biology 63 (2), 227-235.

Vallejo, F. y Estrada, E. 2002. Mejoramiento genético de plantas. Universidad Nacional de Colombia- sede Palmira. ISBN 958-8095-11-5.

Weir, BS. y C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38 (6), 1358-1370.

Wright, S. 1943. Isolation by distance. Genetics 28, 114-138.

Wright, S. 1946. Isolation by distance under diverse systems of mating. Genetics 31, 39-59.

Wright, S. 1951. The genetical structure of populations. Annals of Eugenics 15, 323-354.

Wright, S. 1965. The interpretation of population structure by F-statistics with especial regard to systems of mating. Evolution 19, 395-420.