

UNIVERSIDAD DE COSTA RICA
FACULTAD DE CIENCIAS BÁSICAS
ESCUELA DE BIOLOGÍA

Tesis para optar por el grado de Licenciatura en Biología con énfasis en
Biología Molecular y Biotecnología

**Evaluación de microsatélites para la identificación del germoplasma
de especies del género *Coffea* en Costa Rica**

Elodia María Sánchez Barrantes

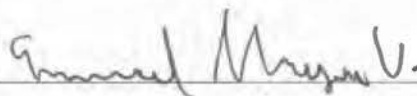
Ciudad Universitaria Rodrigo Facio

2017

MIEMBROS DEL COMITÉ



Dr. Andrés Gatica Arias
Director de tesis



M.Sc. Emanuel Araya Valverde
Miembro Comité Asesor



Dr. Federico Albertazzi Castro
Miembro Comité Asesor



Dra. Marta Valdez Melara
Miembro del Tribunal Examinador



Dr. Germán Vidaurre Fallas
Presidente del Tribunal Examinador



Elodia María Sánchez Barrantes
Postulante

DEDICATORIA

A mis padres, hermanos y hermanas.

PREFACIO

“Son muchas las manos y los corazones que contribuyen al éxito de una persona”

-Walt Disney-

AGRADECIMIENTOS

Primero agradezco a Dios por todas las bendiciones, por la sabiduría y por permitirme disfrutar de todo lo que he logrado.

A mis padres por su amor, su comprensión y su apoyo incondicional. A mis hermanos, a mis hermanas y a toda mi familia por su presencia y su cariño.

Al programa de becas del Centro Nacional de Alta Tecnología (CENAT) quien financió y facilitó sus instalaciones para que el desarrollo del proyecto fuera posible.

A Andrés Gatica, por ser mi tutor, profesor y consejero, por todos los aportes, las enseñanzas, la confianza y el apoyo incondicional

A Emanuel Araya por su gran ayuda, por sus enseñanzas en el laboratorio, por buscar siempre lo mejor para el trabajo, por ser mi guía en todo momento y sobretodo, por convertirse en un gran amigo.

A Max Chavarría por confiar en mí desde el primer momento, por abrirme las puertas del laboratorio y por toda la colaboración brindada.

A Federico Albertazzi por sus revisiones, correcciones y sugerencias durante todo el desarrollo de este proyecto.

A Roberto Avendaño por sus consejos, por su ayuda y su amistad.

Al Instituto del Café de Costa Rica (ICAFFE) y al Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) por brindarme material de sus bancos de germoplasma para desarrollar el proyecto.

A Eric Fuchs por la accesibilidad, la amabilidad y la orientación que me brindó ante mis dudas en la parte estadística

Al personal del Centro Nacional de Innovaciones Biotecnológicas (CENIBiot) por toda la colaboración brindada, por recibirme y hacerme sentir parte de esta gran familia.

A todos los asistentes del laboratorio 150 de biotecnología vegetal por su apoyo y sus consejos.

A Luis Barboza, por la confianza, por todo el apoyo, y principalmente, gracias por darme ese empujón final para concluir este trabajo.

A Daniela Masis, Kathia Alfaro, Gabriel Ríos y Nelson Chaves, por estar siempre a mi lado, por sus consejos, enseñanzas y su gran amistad.

A todos mis amigos y compañeros de la Escuela de Biología por acompañarme durante este largo viaje.

Muchas gracias!!!

ÍNDICE GENERAL

MIEMBROS DEL COMITÉ.....	ii
DEDICATORIA.....	iii
PREFACIO	iv
AGRADECIMIENTOS	v
ÍNDICE DE FIGURAS	ix
ÍNDICE DE CUADROS	x
LISTA DE ABREVIATURAS, SIGLAS Y UNIDADES	xi
RESUMEN	xiii
1. INTRODUCCIÓN.....	1
2. JUSTIFICACIÓN	5
3. OBJETIVOS.....	8
3.1 Objetivo general	8
3.2 Objetivos específicos.....	8
4. HIPÓTESIS.....	9
5. MARCO TEÓRICO.....	9
5.1 Centros de diversidad y botánica.....	9
5.2 Fenología	11
5.3 Genética	11
5.4 Origen de <i>Coffea arabica</i> y su llegada a Costa Rica	12
5.5 Identificación molecular de <i>Coffea</i> en Costa Rica	13
6. METODOLOGÍA	15
6.1 Localización del estudio	15
6.2 Material vegetal para la valoración de los microsatélites	15
6.3 Extracción de ADN total.....	18
6.4 Selección y amplificación de los microsatélites.....	20
6.5 Programa de amplificación de PCR.....	22
6.6 Electroforesis en gel de agarosa	23
6.7 Evaluación de polimorfismos.....	23

6.8 Amplificación de microsatélites fluorocromados.....	26
6.9 Electroforesis capilar.....	29
6.10 Análisis estadístico.....	29
7. RESULTADOS Y DISCUSIÓN.....	32
7.1 Análisis de diversidad genética y polimorfismo de los marcadores microsatélites.....	32
7.2 Análisis de diferenciación genética.....	38
7.3 Perfil molecular.....	43
8. CONCLUSIONES.....	47
9. RECOMENDACIONES.....	48
10. REFERENCIAS.....	49
11. ANEXOS.....	57
Anexo I Protocolo de extracción de ADN genómico.....	57
Anexo II Fragmentos amplificados de los 3 pooles genéticos con los marcadores M32, CaM03, CM5, SSRCa87, M793, M837 y DCM01 en gel de agarosa al 3% (m/v). (MM: marcador molecular (GeneRuler™ 100 pb Thermo Scientific™)), (-): control negativo (H ₂ O).	59
Anexo III. Fragmentos amplificados de los 3 pooles genéticos con los marcadores M32, M24, CaM03, SSRCa87, SSR04 y M774 en gel desnaturalizante de poliacrilamida al 10% (MM: marcador molecular GeneRuler™ 100 pb (Thermo Scientific™)). Las flechas indican algunas bandas polimórficas.....	60
Anexo IV Frecuencia de alelos nulos en las muestras de café arábico y café no arábico.	61
Anexo V Número y frecuencia de alelos obtenidos en las muestras de arábicas y no arábicas con 14 marcadores microsatélites	62
Anexo VI Errores estándar de AU (p-valores) para cada bootstrap generado en el dendrograma de distancias genéticas. El 19 pertenece al único bootstrap que obtuvo un error estándar mayor a 0,01.....	63

ÍNDICE DE FIGURAS

- Figura 1** ADN extraído de 14 muestras de *Coffea* obtenidas del banco de germoplasma del ICAFE en gel de agarosa al 1% (m/v)..... 19
- Figura 2** PCoA utilizando las distancias de Bruvo, los círculos negros representan las muestras de café arábico y los círculos rojos las muestras de café no arábico. El porcentaje de varianza explicado por cada eje se proporciona dentro de la figura. 41
- Figura 3** Dendrograma con valores (%) AU/BP obtenidos a partir de las distancias de Bruvo; los valores rojos corresponden a AU (Approximately Unbiased) y los valores verdes son BP (Bootstrap Probability). **A.** Agrupación de muestras con padres de procedencia etíope. **B.** Agrupación de muestras que presentaron el mismo perfil molecular 42
- Figura 4** Electroferograma mostrando los alelos del SSR locus M774 para 3 muestras distintas de *Coffea* 44
- Figura 5** Perfil molecular obtenido a partir de los 14 marcadores para 3 muestras distintas de café arábico y 3 muestras distintas de café no arábico. Se señalan los microsatélites que no presentaron polimorfismos en las muestras de café arábico..... 46

ÍNDICE DE CUADROS

Cuadro 1. Genotipos de <i>Coffea</i> provenientes del banco de germoplasma del ICAFE, Barva de Heredia, Costa Rica	17
Cuadro 3. Genotipos de <i>Coffea</i> provenientes del banco de germoplasma del CATIE, Turrialba, Costa Rica.....	18
Cuadro 3. Imprimadores seleccionados para amplificar las diferentes especies de <i>Coffea</i>	21
Cuadro 4. Pooles genéticos conformados a partir de 14 genotipos de <i>Coffea</i> procedentes del banco de germoplasma del ICAFE	22
Cuadro 5. Imprimadores que presentan posibles polimorfismos seleccionados para marcar con fluorocromos	25
Cuadro 6. Asignación del fluorocromo a los microsatélites seleccionados de acuerdo al tamaño esperado de alelos.	26
Cuadro 7. Concentración final de imprimador necesario para cada reacción de PCR	27
Cuadro 8. Mezclas asignadas de los productos de PCR para la electroforesis capilar	28
Cuadro 9. Parámetros de diversidad genética obtenida por 14 marcadores microsatélites en muestras de café arábico y no arábico	37
Cuadro 10. Parámetros de diferenciación genética de cada marcador entre muestras de café arábico y no arábico.	39

LISTA DE ABREVIATURAS, SIGLAS Y UNIDADES

° C: Celcius

μL: Microlitros

μM: Micromolar

ABI: Applied Biosystems

ADN: Ácido desoxirribonucleico

AFLP: “Amplified fragment length polymorphism” (Polimorfismos de longitud de fragmentos amplificados)

APS: Persulfato de amonio

AU: “Approximately Unbiased” (Aproximación imparcial)

BP: “Bootstrap Probability” (Probabilidad de bootstrap)

CATIE: Centro Agronómico Tropical de Investigación y Enseñanza

CENIBiot: Centro Nacional de Innovaciones Biotecnológicas

CIBCM: Centro de Investigación en Biología Celular y Molecular

Cn: Concentración

CTAB: Cetil-trimetil bromuro de amonio

EDTA: Ácido etilendiaminotetraacético

EST: “Expressed Sequence Tag” (Marcas de secuencias expresadas)

F_{ST}: Índice de fijación

G_{ST}: Coeficiente de diferenciación de genes

H': Índice de diversidad de Shannon

He: Heterocigosidad esperada

He(c): Heterocigosidad esperada corregida

H₂O: Agua

ICAFFE: Instituto del Café de Costa Rica

ICO: Organización Internacional del Café

kg: Kilogramos

MAG: Ministerio de Agricultura y Ganadería

mg: Miligramos

min: Minutos

mL: Mililitros

MM: Marcador molecular

msnm: Metros sobre el nivel del mar

N_{Alelos}: Número de alelos

ng: Nanogramos

pb: Pares de bases

PCoA: Análisis de coordenadas principales

PCR: “Polymerase chain reaction” (Reacción en cadena de la polimerasa)

PIC: Contenido de información del polimorfismo

QTL: “Quantitative trait loci” (Loci de caracteres cuantitativos)

RAPD: “Random amplified polymorphic DNA” (ADN polimórfico amplificados al azar)

RFLP: “Restriction fragment length polymorphic” (Polimorfismos de longitud de fragmentos de restricción)

R_{ST}: Coeficiente de diferenciación de genes asumiendo la mutación paso por paso

s: Segundos

SNP: “Single nucleotide polymorphism” (Polimorfismo de un solo nucleótido)

SSR: “Simple sequence repeats” (Secuencias simples repetidas)

TAE: Tris-acetato-EDTA

var.: Variedad

TEMED: N,N,N',N'-tetrametiletlenodiamina

RESUMEN

El café es uno de los productos agrícolas más importantes en el comercio mundial. En Costa Rica el 100% del café cultivado comercialmente es de la especie *Coffea arabica*, ya que esta produce un grano de mayor calidad y una taza con mejores características organolépticas que la especie robusta. Sin embargo *C. arabica* presenta baja diversidad genética debido a su origen y proceso de domesticación, lo cual ha provocado grandes dificultades en la identificación de individuos. Además, esta estrecha base genética ha hecho que la especie sea susceptible a plagas y enfermedades. La identificación y caracterización de cultivares del género *Coffea* basadas en el estudio de caracteres morfológicos, botánicos y agronómicos tienen limitaciones, ya que muchas de las características fenotípicas utilizadas pueden ser moduladas por determinantes ambientales. Mientras tanto la caracterización a nivel molecular cada vez demuestra ser más efectiva y necesaria, ya que las técnicas moleculares están evolucionando, se realizan independientemente de factores ambientales y funcionan como complemento de los descriptores morfológicos, los cuales son insuficientes para identificar a un individuo.

El objetivo de este trabajo fue evaluar marcadores moleculares microsatélites en la identificación del germoplasma de especies del género *Coffea* en Costa Rica. Para ello se analizaron 55 muestras de *Coffea* con 14 marcadores (10 SSR's y 4 EST-SSR). Las muestras se dividieron en 2 grupos: las arábicas (en las cuales se incluyen variedades e híbridos) (n=52) y las no arábicas (n=3). Se encontraron 90 alelos diferentes en total. En las muestras de café arábico se encontraron 64 alelos diferentes, obteniendo un promedio de 4,57 alelos por marcador. Mientras que en las otras especies no arábicas se encontraron 49 alelos, con un promedio de 3,50 alelos por marcador.

Las muestras de café no arábico presentaron mayor variación genética y mayor polimorfismo (PIC=0,5792, He=0,6220, He_(c)=0,6786, H'=1,267) que las muestras de café arábico (PIC=0,5183, He= 0,5885, He_(c)=0,5913, H'=1,0422), esto debido a que *Coffea*

arabica posee una base genética estrecha por su origen, reproducción, y su proceso de domesticación. En cuanto a la diferenciación genética entre ambos grupos se obtuvieron valores promedio de $G_{ST} = 0,1457$, y $R_{ST} = 0,3853$, que indican que la mayoría de los marcadores muestran heterogeneidad genética entre moderada y alta entre muestras de arábicas y no arábicas. Tanto con el estadístico R_{ST} como con G_{ST} todos los marcadores revelaron diferencias estadísticamente significativas entre ambos grupos.

Además se realizó un PCoA y un dendrograma a partir de las distancias genéticas de Bruvo, los resultados muestran que el grupo de 14 marcadores utilizados en este estudio permiten diferenciar claramente entre especies. Sin embargo, entre las muestras de arábicas las distancias genéticas continúan siendo bajas, lo cual no permite una separación clara de las muestras. Para las muestras de especies arábicas, este conjunto de 14 marcadores no fueron efectivos para diferenciar entre las accesiones de Borbón, Oeiras, Villa Sarchí y Venecia. Sin embargo los marcadores CaM03, CaM16, M24, M32, M753, SSR03, SSRCa18 y SSRCa87 son polimórficos en muchas de las accesiones de café arábico, y se observó que utilizando en conjunto este grupo de 8 microsatélites podemos diferenciar entre todas las accesiones estudiadas, con excepción de las mencionadas anteriormente (Borbón, Oeiras, Villa Sarchí y Venecia) que mostraron el mismo perfil molecular.

Podemos concluir que este conjunto de marcadores son altamente polimórficos y son efectivos, ya que permitieron diferenciar claramente entre las muestras de arábicas y no arábicas. Además, permitió diferenciar entre las diferentes accesiones de arábicas. Este grupo de marcadores facilitaría la caracterización molecular del germoplasma de café en Costa Rica y puede convertirse en una gran herramienta para realizar estudios de paternidad, diversidad, estructura poblacional e incluso para el registro o certificación de café con alta calidad de taza.

Palabras claves: *Coffea arabica*, microsatélites, polimorfismo, marcadores moleculares, imprimadores, electroforesis capilar, diversidad genética, diferenciación genética

1. INTRODUCCIÓN

El café es la bebida no alcohólica más popular del mundo y el segundo producto más importante en el comercio mundial después de los productos derivados del petróleo (Dessalegn *et al.*, 2009). Según el reporte de la Organización Internacional del Café (ICO) para el año 2016, la producción promedio mundial fue de 153,869 millones de sacos de 60 kg para este período. Brasil fue el principal país productor aportando el 35,7% de la producción mundial, mientras que Costa Rica ocupó el décimo cuarto lugar entre los países productores, aportando el 0,97% (<http://www.ico.org/>).

El género *Coffea* presenta gran importancia económica en la familia Rubiaceae (Montoya *et al.*, 2006), consta de más de 100 especies conocidas (Missio *et al.*, 2009) de las cuales solo *Coffea arabica* (conocido como el café arábica) y *C. canephora* (conocido como el café robusta) se cultivan comercialmente (Tornincasa *et al.*, 2010).

Actualmente, *C. arabica* es la especie más importante en el comercio ya que produce café de alta calidad en comparación con robusta, y contribuye con alrededor del 70% de la producción total mundial de café (Anthony *et al.*, 2002, Tornincasa *et al.*, 2010). En nuestro país el 100% del café cultivado comercialmente es de la especie arábica, ya que desde el año 1989 se prohíbe por ley la siembra de café robusta por su inferior calidad de taza (<http://www.icafe.go.cr/>). Arábica, a diferencia de las otras especies, posee condiciones como buen rendimiento del fruto, tamaño del grano y principalmente, es conocido por su excelente calidad de la taza (Pérez, 1977). Sin embargo, tiene un gran inconveniente, ya que sufre de una baja diversidad genética debido a su origen y proceso de domesticación, esta baja diversidad se refleja en una mayor susceptibilidad a enfermedades y plagas (Pereira *et al.*, 2015, Prakash *et al.*, 2002). Por lo tanto, el café requiere adaptarse a los diferentes ambientes en los que se cultiva para mantener los requerimientos del mercado. Para lograr esto se requiere del mejoramiento genético del café, el cual principalmente se lleva a cabo mediante el uso de métodos convencionales, los cuales son por lo general lentos y laboriosos (Hendre

& Aggarwal, 2007). Esta situación exige el desarrollo de herramientas que aceleren y brinden fiabilidad en la caracterización del acervo genético, que conllevará a una utilización más eficiente del germoplasma disponible en los programas de mejora genética de café (Hendre & Aggarwal, 2014). En este contexto, los marcadores genéticos basados en polimorfismo de ADN (ácido desoxirribonucleico) se convierten en una importante herramienta que ha demostrado ser de gran valor en la caracterización y mejora genética de los recursos genéticos vegetales.

Entre los principales marcadores moleculares utilizados en cultivares de importancia comercial están las aloenzimas, RFLP (Restriction fragment length polymorphic), RAPD (Random amplified polymorphic DNA), AFLP (Amplified fragment length polymorphism), SNP (Single nucleotide polymorphism) y microsatélites SSR (Simple sequence repeats) (Azofeifa, 2006).

Las aloenzimas fueron los primeros marcadores moleculares usados en genética de plantas. Son marcadores codominantes y han sido de gran valor en estudios de mejoramiento tanto en poblaciones naturales como en plantaciones, sin embargo es una técnica muy laboriosa, se debe conocer muy bien las enzimas y es poco reproducible entre laboratorios (Azofeifa, 2006).

Los RFLP, también son marcadores codominantes que se pueden utilizar tanto con ADN nuclear como con ADN de organelas. Para realizar esta técnica se requiere de grandes cantidades de ADN de buena calidad para la detección de *loci* de copias únicas; se necesita de sondas específicas, requiere de muchas manipulaciones y solamente se detecta una fracción de la variabilidad de secuencias existentes en el genoma, es decir, su información es limitada (Alcántara, 2007).

Los RAPDs amplifican regiones tanto codificantes del ADN como las no codificantes y revelan niveles de variación más altos que los RFLPs e isoenzimas (Parker *et al.*, 1998). Es una técnica que no necesita conocimiento previo de la secuencia de ADN y no requiere de sondas específicas para cada especie (Azofeifa, 2006), sin embargo presenta baja

reproducibilidad y no se puede diferenciar entre individuos homocigotos y heterocigotos ya que son marcadores dominantes (Zhivotovsky, 1999). Los AFLP son altamente polimórficos, no requieren ninguna información previa de la secuencia para su análisis y pueden ser marcadores dominantes o codominantes. Esta técnica es útil para generar huellas genéticas y mapeo; también se ha utilizado para la caracterización de germoplasma, estudios filogenéticos en plantas, bacterias, hongos y en estudios de genética de poblaciones (Alcántara, 2007). Sin embargo, requiere de una alta cantidad de ADN y la técnica es más complicada de ejecutar que la de los RAPD (Azofeifa, 2006). Los SNP son marcadores que se han utilizado ampliamente en estudios de diversidad genética, sin embargo poseen un contenido de información bastante bajo, por lo que hay que usar muchos para llegar a un buen nivel de información (Rischkowsky & Pilling, 2010).

Finalmente, están las secuencias simples repetidas (SSR) o marcadores microsatélites que son secuencias de ADN compuestas de motivos cortos repetidos en tándem (Cristacho & Gaitán, 2008, Poncet *et al.*, 2004). Estos motivos tienen un amplio polimorfismo de longitud específica del sitio debido a las diferentes cantidades de unidades de repetición; también son robustos, transferibles, codominantes, cromosómicamente se encuentran en un solo lugar y se pueden desarrollar como marcadores basados en la técnica de reacción en cadena de la polimerasa (PCR), con la amplificación de las regiones repetitivas complementarias del genoma. Las regiones de acompañamiento de secuencias microsatélites son conservadas, esto permite una alta reproducibilidad de la técnica debido a que los imprimadores están diseñados para estas regiones deseadas (Motta *et al.*, 2015). Además los microsatélites son multialélicos y presentan numerosas ventajas en comparación con otros tipos de marcadores, porque son muy informativos, necesitan poco material genético de muestra y tienen posibilidad de automatización (Rovelli *et al.*, 2000).

En un inicio, la identificación de marcadores microsatélites es costosa, ya que se necesita ser aislados *de novo* en la mayoría de las especies que se analizan por primera vez y el desarrollo de los imprimadores, requiere un proceso muy largo y laborioso (Zane *et al.*, 2002). Actualmente, existen diferentes estrategias para el desarrollo de estos imprimadores,

sin embargo el método más utilizado es el de biblioteca genómica enriquecida con hibridación selectiva (Zane *et al.* 2002).

Los microsatélites se encuentran en regiones codificantes y no codificantes del genoma (Missio *et al.*, 2011), por lo que la estrategia de diseñar iniciadores universales no funciona muy bien para este tipo de marcadores. Sin embargo, se ha reportado que algunos microsatélites tienen regiones flanqueantes muy conservadas, lo que permite amplificar estos microsatélites en especies divergentes (Zane *et al.*, 2002).

En variedades con una base genética estrecha como las arábicas, el uso de marcadores moleculares ha sido de gran utilidad para diversos estudios. Berthou y Trouslot (1977) realizaron un estudio con isoenzimas para determinar el polimorfismo enzimático de *Coffea*, y posteriormente Lashermes *et al.* 1999 realizaron un análisis con RFLP para realizar la caracterización y el origen de *Coffea arabica*, sin embargo ambos estudios encontraron que el polimorfismo es extremadamente bajo entre las accesiones de *Coffea* por lo que estos marcadores no resultaron ser lo suficientemente informativos.

Otro marcador molecular que ha sido importante en los estudios de *Coffea* son los RAPD, los cuales se ha usado para construir dendrogramas que resultaron ser consistentes con la historia y evolución conocida de *Coffea arabica* (Lashermes *et al.*, 1996, Orozco *et al.*, 1994). Además Orozco *et al.*, 1994 reporta que los marcadores RAPD también han sido buenos marcadores para diferenciar entre *C. arabica* var. *typica* y *C. arabica* var. *bourbon*, así como para detectar introgresión natural e interespecífica entre diploides *C. canephora* y *C. arabica* (accesión Rume Sudan RS-510).

Unos años después, Steiger *et al.* 2002, utilizó marcadores AFLP con el objetivo de determinar la diversidad entre diferentes cultivares de *Coffea arabica*. Ellos mencionan que la técnica AFLP utilizada en este estudio es tan fiable como RFLP y SSR a un costo menor, y es más confiable que los marcadores RAPD. Ellos encontraron polimorfismos entre algunos cultivares que permitieron la diferenciación de los mismos, sin embargo la variación genética entre los cultivares de arábica fue similar a la variación dentro de los cultivares, y

no se detectó ningún marcador de ADN específico del cultivar. La principal desventaja que mencionan con el uso de la técnica es que pueden surgir datos erróneos de los marcadores debido a la digestión parcial de ADN genómico.

Por otro lado los marcadores SSR o microsatélites son potencialmente útiles, sobre todo para explorar regiones altamente variables del genoma entre los individuos o poblaciones de la misma especie (Hurtado & Herrera, 2013, Missio *et al.*, 2009). Estos marcadores son ideales para el estudio de la diversidad genética, estructura poblacional, relaciones filogenéticas, construcción de mapas de ligamiento, asignación de intervalo de QTL, entre otros (Hendre & Aggarwal, 2014).

En estudios anteriores de *Coffea arabica* el grado de diversidad alélica revelado en los SSR loci en la mayoría de estos organismos indica que los SSR son ideales para el análisis de ligamiento, selección de características agronómicas, evaluación de germoplasma y la identificación de variedades (Powell *et al.*, 1996). Sin embargo, a pesar de todas las ventajas que proveen estos marcadores, la disponibilidad de marcadores de SSR polimórficos sigue siendo una restricción en *C. arabica* (Ferrao *et al.*, 2015, Missio *et al.*, 2011), esto debido a la poca abundancia de SSR que presenta en su genoma, y que ha provocado una gran dificultad en el desarrollo de marcadores específicos (Hendre & Aggarwal, 2007).

2. JUSTIFICACIÓN

Costa Rica es un país reconocido alrededor del mundo por su producción cafetalera. La producción y exportación de café, ha tenido un aumento sostenido en los últimos años, principalmente en lo que respecta a cultivares de alta calidad (<http://www.icafe.go.cr>), los cuales han aumentado desde hace más de dos décadas, se vende a precios elevados y ha representado una alternativa prometedora para superar la crisis de los precios internacionales

(Castro *et al.*, 2004). Esta condición motiva a un mayor cuidado en la calidad del producto final, siendo aspectos significativos en este sentido la certificación y su identidad genética.

La identificación y diferenciación de los cultivares del género *Coffea* han representado un problema de difícil solución para las técnicas basadas en el estudio de caracteres morfológicos, botánicos y agronómicos (forma, tamaño y color de planta, hojas, frutos, resistencia a enfermedades, etc.) (Vieira *et al.*, 2010). Además, estas características pueden ser moduladas por determinantes ambientales, tales como el efecto de patógenos, la etapa del crecimiento o el ambiente agroclimático en que se encuentre la planta. La caracterización morfológica depende del criterio y experiencia individual de cada especialista. Esto ha llevado a que en numerosas ocasiones existan genotipos mal clasificados (Vieira *et al.*, 2010), especialmente cuando se trata de cultivares que presentan fenotipos muy similares como el caso de las arábicas.

Los cultivares del café arábica son considerados como cafés especiales con características organolépticas peculiares y estas poseen un alto valor comercial, es por ello que existen razones económicas para aparentar las garantías de la autenticidad de las especies y variedades de café (Tornincasa *et al.*, 2010). En las últimas décadas a nivel comercial se han detectado adulteración de café de origen arábica con cafés de origen robusta, y entre las variedades de arábica. En la actualidad, para distinguir arábica de robusta en mezclas de café se utilizan análisis químicos de compuestos tales como esteroides, ácido clorogénico y cafeína, ácidos grasos, tocoferol, entre otros (Tornincasa *et al.*, 2010) aunque estos no siempre dan resultados confiables. Lo que nos lleva a la conclusión de que aún no existe ningún método fiable para garantizar totalmente la autenticidad del café, por lo que el uso y desarrollo de herramientas analíticas basadas en el análisis directo del genoma (ADN) como los microsatélites supera varios de estos inconvenientes debido a que el análisis se hace independiente de factores ambientales (Morell *et al.*, 1995). De esta manera, se obtienen marcadores moleculares que definen un "perfil molecular" ("fingerprinting") propio para cada cultivar, herramienta que sería de gran ayuda para productores y exportadores, ya que

permite diferenciar entre las diferentes variedades producidas en el país y les permite garantizar la autenticidad del producto que están ofreciendo.

Además de la identificación y la autenticidad, los SSR pueden utilizarse con el fin de analizar con mayor profundidad la estructura del genoma del café, y de esta manera hacer más eficiente la utilización de los recursos genéticos en el mejoramiento vegetal. Una evaluación más rápida y precisa del germoplasma, puede agilizar de gran forma la selección de características agronómicas de interés para estudios de investigación, tal como la resistencia a la enfermedad de roya, el cual es un hongo que debilita las plantas y provoca que el fruto del café caiga antes de su maduración (www.mag.go.cr). Esta enfermedad afecta principalmente al café de origen arábico y ha provocado grandes pérdidas a los agricultores cafetaleros del país desde 1983. Según Bigirimana *et al.* (2013) el desarrollo y cultivo de variedades resistentes es la opción más viable y eficaz para hacerle frente a esta enfermedad, ya que el control químico, además de presentar un constante incremento en sus costos (insumos y mano de obra), puede ocasionar un desequilibrio biológico, atribuido a las constantes aplicaciones.

El panorama actual del número de SSR desarrollados para el género *Coffea* muestra que es necesario aumentar los esfuerzos para el desarrollo y validación de nuevos marcadores, en la literatura se enumeran cerca de 400 SSR genómicos para las especies *C. canephora* y *C. arabica* (Aggarwal *et al.*, 2007, Cristacho & Gaitán, 2008, Dessalegn *et al.*, 2009, Geleta *et al.*, 2012, Hendre *et al.*, 2008, Hendre & Aggarwal, 2014, Missio *et al.*, 2009, Motta *et al.*, 2015, Poncet *et al.*, 2004, Vieira *et al.*, 2010) lo cual representa un número modesto cuando se compara con otras especies cultivadas de importancia comercial como el maíz.

A pesar de las grandes ventajas que nos puede ofrecer el uso de los SSR en los cultivos de café en Costa Rica, en la actualidad no se han reportado estudios desarrollados en el país con estos marcadores moleculares, es por ello que se consideramos de gran importancia realizar una evaluación preliminar de estos marcadores moleculares en algunas accesiones

de *Coffea* representativas del país, esto con el fin de aportar un poco al estudio, conocimiento y entendimiento de esta especie. Al mismo tiempo se pretende desarrollar y brindar al país una herramienta que contribuya a la identificación y al mejoramiento del cultivo de café en Costa Rica.

3. OBJETIVOS

3.1 Objetivo general

- Evaluar marcadores microsatélites (SSR) diseñados previamente para café en la identificación de accesiones del género *Coffea* en Costa Rica.

3.2 Objetivos específicos

- Establecer una metodología óptima para la amplificación por la técnica PCR de regiones microsatélites presentes en el ADN genómico de varias especies del género *Coffea*.
- Estimar mediante la técnica de electroforesis capilar el número, tamaño de alelos y frecuencia alélica que cada microsatélite presenta en los individuos analizados para identificar las diferentes especies del género *Coffea*.
- Determinar el nivel de polimorfismo para diferentes microsatélites presentes en el ADN genómico con el fin de identificar y diferenciar las distintas especies del género *Coffea*.

4. HIPÓTESIS

Los marcadores moleculares SSR son efectivos para la caracterización genética del género *Coffea*, lo que permite identificar y diferenciar las especies del género en Costa Rica mediante la información del contenido polimórfico obtenido.

5. MARCO TEÓRICO

5.1 Centros de diversidad y botánica

El género *Coffea* pertenece a la familia Rubiaceae y posee 103 especies, de las cuales hay 41 especies en África, 59 en Madagascar y 3 en Mascarenes (Islas Mascareñas). No se encuentran especies de *Coffea* naturales fuera de estas tres áreas (Davis *et al.*, 2006).

Los tres principales centros de diversidad de especies son Madagascar, principalmente en los bosques siempre verdes y húmedos, Camerún (14 especies) y Tanzania (16 especies, principalmente en las montañas del arco oriental (Davis & Mvungi, 2004). Madagascar tiene una gran variedad de tipos de bosque, y esto puede explicar de alguna manera la alta diversidad de especies de *Coffea* en esta isla. Las especies de *Coffea* en África habitan generalmente en bosques húmedos y perennes. No hay especies de origen natural compartidas entre África, Madagascar y Mascarenes, cada área tiene una endemividad del 100% para sus especies de *Coffea* (Davis *et al.*, 2006).

En cuanto a su botánica, el género *Coffea* presenta una combinación de caracteres que permite su reconocimiento: Todas son leñosas, de madera dura y densa, comprenden desde

arbustos hasta árboles que por lo general presentan ramificación horizontal o casi horizontal (ramificación plagiotrópicas) (Davis *et al.*, 2006).

Sus hojas varían de forma de elípticas a lanceoladas, la lámina de la hoja puede medir entre 12 a 24 cm de largo y de 5 a 12 cm de ancho, el tamaño va a variar dependiendo no solo de la especie sino también de condiciones ambientales a las que se encuentre sometida la planta. Además, el color de las hojas puede variar entre diferentes tonalidades de verdes, bronce o con matices purpúreos (Rojo, 2014).

Presentan peciolo cortos y pequeñas estípulas, y en el envés pueden aparecer unas pequeñas cavidades de forma irregular llamados domocios, cuya función no se conoce aún, sin embargo se ha observado que generalmente albergan pequeños artrópodos como ácaros (Rojo, 2014). Sus inflorescencias son pareadas, axilares y se forman sólo en el crecimiento corriente del año, por lo tanto la floración depende estrechamente del crecimiento de las ramas laterales, en particular de las primarias. Las yemas seriadas dan origen a yemas florales o a ramas laterales secundarias. La yema superior de cada serie axilar de las ramas laterales origina la mayor inflorescencia, mientras que las otras son cada vez menores (De Oliveira *et al.*, 2012).

Las flores son hermafroditas, con corolas blancas o rosado claro y presentan un cáliz reducido que rara vez supera el disco, este es un carácter muy importante, ya que puede usarse para distinguir a *Coffea* de los otros géneros. Los frutos son carnosos, de tipo drupa y contiene dos (raramente una) semillas, cada semilla con una ranura profunda (invaginación) en el lado plano ventral de la semilla (Davis *et al.*, 2006).

5.2 Fenología

La naturaleza fisiológica de la bienalidad de la producción del café puede explicarse por la competencia entre las funciones vegetativas y reproductivas. En los años de gran producción, el crecimiento de los frutos absorbe la mayor parte de la actividad metabólica de la planta, reduciendo el desarrollo vegetativo (De Oliveira *et al.*, 2012). En *Coffea arabica*, el fruto se desarrolla en las partes nuevas de las ramas del año anterior, habiendo, consecuentemente, menor producción. El crecimiento de las ramas nuevas depende de la cantidad de frutos en desarrollo, y el volumen de producción es proporcional al vigor vegetativo, al número de nudos y gemas florales formadas en la estación vegetativa anterior (Camargo & Camargo 2001).

Las dos especies más importantes desde el punto de vista económico son *Coffea arabica* L. (café arábica) y *Coffea canephora* Pierre ex Froehner (café robusta). Otra especie también conocida pero cultivada en menor escala es *Coffea liberica* (café liberica) *Coffea excelsa* y *Coffea dewevrei* (Rojo, 2014).

5.3 Genética

Las especies del género *Coffea* se divide en dos grupos según el número cromosómico, el grupo grande de las especies diploides ($2n=22$ cromosomas) (Aggarwal *et al.*, 2007, Davis *et al.*, 2006, Geleta *et al.*, 2012) conformado por *C. canephora*, *C. liberica* *C. stenophylla*, *C. racemosa* y otros, y el grupo de los tetraploides ($2n=4x=44$ cromosomas) conformado únicamente por *C. arabica* la cual es una especie anfídiploide que se originó a partir de una hibridación entre *C. canephora* y *C. eugenioides* en las mesetas de Etiopía (Lashermes *et al.*, 1999). Esta especie se caracteriza por ser auto-fértil en aproximadamente el 90%, mientras que otras especies de café son generalmente auto-incompatibles (Anthony *et al.*, 2001), a

pesar de ser la única especie tetraploide del género, presenta un comportamiento meiótico regulado y de tipo disómico, lo que significa que prevalece el apareamiento preferencial entre cromosomas homólogos (Romero *et al.*, 2010).

5.4 Origen de *Coffea arabica* y su llegada a Costa Rica

El café arábico (*Coffea arabica* L.) tiene su origen en las tierras altas (de más 1000 msnm.) de Etiopía, país donde se inició su cultivo (Anthony *et al.*, 1999) y en el que actualmente crece en forma silvestre sobre los 1500 msnm (León, 2000).

La introducción del café a Arabia y a Yemen lo realizaron los persas y los árabes durante los siglos XIV y XV, mientras que los nativos africanos lo introdujeron a Mozambique y Madagascar, esto gracias a que extendieron sus cultivos por toda esta zona (Alvarado & Rojas, 1998).

De Oliveira *et al.* (2012) menciona que durante este tiempo el café era cultivado para el consumo de la pulpa del fruto, fue hasta el siglo XVI, en Persia donde los primeros granos de café fueron tostados para transformarse en la bebida que hoy se conoce. El café se hizo de gran importancia para los árabes, que tenían completo control sobre el cultivo y la preparación de la bebida. En la época, el café era un producto resguardado por los árabes y se prohibía que los extranjeros se acercaran a las plantaciones. Durante este tiempo y hasta el siglo XVII, el café era producido exclusivamente por los árabes. Sin embargo los alemanes, franceses e italianos buscaban por todos los medios una técnica de desarrollar la siembra en sus colonias. Pero fueron los holandeses quienes lograron obtener las primeras plantas y las cultivaron en los invernaderos del jardín botánico de Ámsterdam. A partir de estas plantas los holandeses iniciaron en 1699, plantíos experimentales en Java. Esta iniciativa de éxito traía beneficios, alentando a otros países a intentar el mismo.

Las primeras introducciones al continente americano se dan a inicios del siglo XVIII cuando se realizaron envíos de plantas de Ámsterdam hasta Surinam y de París a la isla de Martinica en las Antillas de donde su cultivo se expandió hacia el norte de Brasil y hacia el resto de América tropical (León, 2000).

Los cafés de Yemen dieron origen a dos variedades: *C. arabica* var. *arabica (typica)*, que se conoce comúnmente como Arábigo o Típica y que constituyó la base genética de las primeras variedades cultivadas en América y Asia (Anthony *et al.* 1999), y *C. arabica* var. *bourbon* que fue difundida desde las islas Borbón (Calle, 2012).

Costa Rica fue el primer país Centroamericano en iniciar la industria del café, este se introdujo hasta finales del siglo XVIII y se cree que las primeras semillas fueron de la especie *Coffea arábica*, variedad *typica* procedentes de isla Martinica (<http://www.icafe.go.cr/>). Hasta mediados del siglo XX todas las variedades cultivadas en América Latina compartieron la base genética del café introducido de Ámsterdam (Anthony *et al.*, 1999).

5.5 Identificación molecular de *Coffea* en Costa Rica

Los marcadores moleculares se han utilizado en gran cantidad de estudios de café con el fin de estudiar su origen, transferibilidad, contenido de información polimórfica, identificación y caracterización de variedades, diversidad y estructura genética, mapas de ligamiento, diversidad de especies resistentes a enfermedades, huellas genéticas, entre otros (Aggarwal *et al.*, 2007, Alekcevetch *et al.*, 2013, Baruah *et al.*, 2003, Bigirimana *et al.*, 2013, Ferrao *et al.*, 2015, Hendre *et al.*, 2008, Misio *et al.*, 2009, Lashermes *et al.*, 1999, López *et al.*, 2009, López & Moncada 2005). Sin embargo en Costa Rica existen pocos estudios con marcadores moleculares en café, a pesar de ser un cultivo de gran importancia comercial y que ha llevado al país a grandes reconocimientos a nivel mundial debido a su calidad de taza (López, 2004).

Uno de los estudios realizados en el país es el presentado por Astorga (1999), para este estudio utilizó RAPD con el fin de caracterizar las variedades *Typica* y *Bourbon*, estimó las distancias genéticas entre grupos, y generó la clasificación de los grupos genéticos. Logró identificar marcadores diferentes entre cada grupo y marcadores comunes entre grupos. Sin embargo obtuvo un número de marcadores polimórficos bajo. Además la distancia genética entre el grupo *Typica* y *Bourbon* no fue significativa, lo que demostró la estrecha base genética que caracteriza estos grupos, sin embargo los análisis permitieron observar una separación clara entre las dos bases genéticas.

Posteriormente Anthony *et al.* (2001), también realizaron un estudio utilizando marcadores RAPD con el fin de estudiar la diversidad genética de accesiones silvestres y cultivadas de *C. arabica* procedentes del banco de germoplasma del CATIE. De acuerdo a los resultados y a los análisis de agrupamiento encontró que el material etíope formó cuatro grupos genéticos que se separan claramente de las accesiones derivadas de *Typica* y *Bourbon*. Luego compararon las diferencias entre los cafés silvestres y cultivados, y encontraron bajas distancias genéticas entre ambos, principalmente con las accesiones silvestres.

Por lo tanto, los marcadores RAPD fueron eficientes para distinguir las variedades del *Typica* y del *Bourbon*. Sin embargo, no detectaron diferencias dentro de cada variedad, y tampoco dentro de las accesiones silvestres, por lo que concluyeron que utilizar estos marcadores para proteger variedades de *C. arabica* resulta difícil y costoso.

Posteriormente Anthony *et al.* (2002) realizaron otro estudio más para determinar el polimorfismo entre cafés cultivados (variedades provenientes de *Typica* y *Bourbon*) y cafés silvestres etíopes, pero esta vez utilizando los marcadores moleculares AFLP y SSR. El estudio mostró que los marcadores AFLP demostraron ser eficaces para caracterizar las dos bases genéticas que se difundieron en el siglo XVIII, identificando sus híbridos y distinguiendo las plantas de café derivadas de especies silvestres etíopes. Los cultivares se separaron claramente de las accesiones etíopes derivadas, como se demostró en el estudio previo basados en los marcadores RAPD (Anthony *et al.*, 2001). Los marcadores SSR

confirmaron el bajo polimorfismo detectado en las accesiones cultivadas por los marcadores AFLP y el alto polimorfismo dentro de las accesiones silvestres.

Estos resultados permitieron un análisis de las reducciones de la diversidad genética que ocurrieron sucesivamente durante la diseminación de *C. arabica* desde su centro principal de diversidad.

6. METODOLOGÍA

6.1 Localización del estudio

El estudio se realizó en el Laboratorio de Genómica del Centro Nacional de Innovaciones Biotecnológicas (CENIBiot) ubicado en Pavas, San José, Costa Rica.

6.2 Material vegetal para la valoración de los microsatélites

Se seleccionaron 55 genotipos del género *Coffea* con el fin de abarcar la mayor diversidad alélica posible para la valoración de los microsatélites. El material utilizado es de origen genético conocido, 33 genotipos provienen de la colección de germoplasma del Instituto del Café de Costa Rica (ICAFFE) ubicado en Barva de Heredia (10°01'51"N, 84°08'23"O, 1.170 msnm.) (Cuadro 1), y 22 genotipos provienen de la colección de germoplasma del Centro Agronómico Tropical de Investigación y Enseñanza (CATIE) ubicado en Turrialba (9°53'44" N, 83°38'7"O, 602 msnm) (Cuadro 2). Se tomaron muestras de hojas (tanto jóvenes como adultas) de cada genotipo, sin importar la posición en la planta,

pero se procuró coleccionar hojas que tuvieran color verde intenso y que no mostrarán ningún tipo de daño. Posteriormente, las hojas se transportaron en tubos plásticos falcon de 50 mL con gel sílica desecante de Merck®, esto con el fin de reducir el contenido de agua del tejido vegetal. Estas muestras permanecieron con la sílica durante una semana a temperatura ambiente, asegurándonos de esta forma de que el material quede totalmente seco, y así evitar la degradación de la muestra.

Cuadro 1. Genotipos de *Coffea* provenientes del banco de germoplasma del ICAFE, Barva de Heredia, Costa Rica.

Germoplasma	Identificación de ICAFE
1. Borbón	<i>C. arabica</i>
2. Catigua MC2	Híbrido <i>C. arabica</i> X <i>C. canephora</i>
3. Catimor CR95	Híbrido <i>C. arabica</i> X <i>C. canephora</i>
4. Catimor S175	Híbrido <i>C. arabica</i> X <i>C. canephora</i>
5. Catuaí amarillo	<i>Catuaí (C. arabica)</i>
6. Catuaí CR	Catuaí (Caturra X Mundo Novo)
7. Catuaí-SH3	Cavimor (Catuaí X Híbrido Timor)
8. CaturraIC	<i>C. arabica</i>
9. Cavimor-A	Cavimor (Catuaí X Híbrido Timor)
10. Coffea	<i>C. arabica</i> silvestre
11. Excelsa	<i>C. excelsa</i>
12. I-59	Híbrido <i>C. arabica</i> X <i>C. canephora</i>
13. L02A30	Híbrido <i>C. arabica</i>
14. L04A20	Híbrido (Catimor X Rume Sudán)
15. L04A42	Híbrido (Catuaí X Etiopo)
16. L04A5IC	Híbrido
17. L09A22	Híbrido (Caturra X Anfilo)
18. L10A25	Híbrido (Catuaí X Rume Sudán)
19. L11A26	Híbrido (Caturra X Etiopo)
20. L12A28IC	Híbrido
21. L13A12	Híbrido (Sarchimor X Etiopo)
22. L13A22IC	Híbrido
23. L13A44IC	Híbrido
24. Liberica	<i>C. liberica</i>
25. Oeiras	Híbrido <i>C. arabica</i> X <i>C. canephora</i>
26. Robusta 35	<i>C. canephora</i>
27. S-288	Híbrido <i>C. arabica</i> X <i>C. liberica</i>
28. S288-6	Híbrido <i>C. arabica</i> X <i>C. liberica</i>
29. S296	Híbrido <i>C. arabica</i> X <i>C. canephora</i>
30. T.5296IC	Híbrido Sarchimor
31. TUPI	Híbrido <i>C. arabica</i> X <i>C. canephora</i>
32. Venecia	<i>C. arabica</i>
33. Villa Sarchí	<i>C. arabica</i>

Cuadro 2. Genotipos de *Coffea* provenientes del banco de germoplasma del CATIE, Turrialba, Costa Rica.

Genotipo	Identificación de CATIE
1. L04A5	Híbrido (ET25 X T5296)
2. L12A28	Híbrido (Rume Sudan X T5296)
3. L13A22	Híbrido (E531 X Caturra)
4. L13A44	Híbrido (Rume Sudan X T5296)
5. L4A34	Híbrido (ET41 X Caturra 7)
6. L5A26	Híbrido (E531 X Caturra 7)
7. T. 02694	BA-10
8. T. 03099	Arussii S-11
9. T. 04492	E-044
10. T. 04551	E-478
11. T. 04873	E-506
12. T. 16691	ET-03
13. T. 16728	E-044
14. T. 21236	PDRY-07
15. T. 21239	PDRY-14
16. T. 21242	PDRY-22
17. T. 2724	Rume Sudan
18. T. 2739	SL-28
19. T. 2744	Rume Sudan
20. T.04900 (E5314900)	E-531
21. T.16712	ET-25
22. T.16725	ET-41

6.3 Extracción de ADN total

Del material seco se utilizaron 40 mg de tejido y se realizó la extracción de ácidos nucleicos totales utilizando protocolo de extracción el basado en CTAB (cetil-trimetil bromuro de amonio) con limpieza de cloformo:octanol de Araya, 2010 (Anexo I).

La idoneidad del ADN extraído se valoró mediante su cantidad y calidad. Para cuantificar el ADN se utilizó un espectrofotómetro (NanoDrop®) utilizando 2 μ L de cada muestra, y para observar la integridad, se realizó una electroforesis en un gel de agarosa al 1% (m/v) el cual se corrió por 35 minutos en buffer TAE (Tris-acetato-EDTA) 1X a 100 voltios constantes. Se colocaron 5 μ L del ADN y 1 μ L de buffer de carga con GelRed de Fermentas ® en cada pozo, luego el gel fue expuesto al transiluminador y fotografiado (Figura 1). Todos los ADN se diluyeron a una concentración de 50 ng/ μ L.

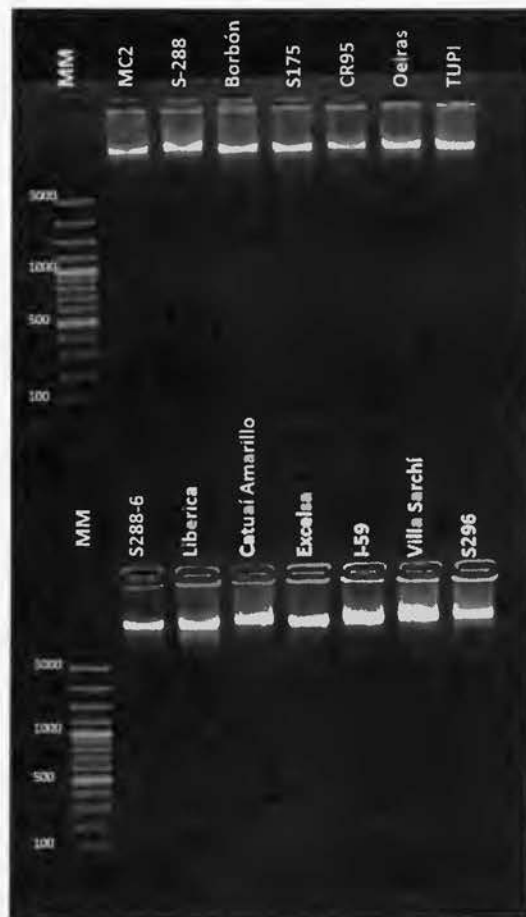


Figura 1. ADN extraído de 14 muestras del género *Coffea* obtenidas del banco de germoplasma del ICAFE en gel de agarosa al 1% (m/v) (MM: marcador molecular GeneRuler™ 100 pb (Thermo Scientific™))

6.4 Selección y amplificación de los microsatélites

Los imprimadores para la detección de SSR se seleccionaron con base en la literatura existente relacionada con el uso de este tipo de marcadores moleculares en café (Cuadro 3).

Los criterios utilizados para seleccionar los imprimadores fueron los siguientes:

- Que hayan amplificado al menos 2 veces en estudios anteriores con especies del género *Coffea*, y que estas amplificaciones sean en diferentes regiones del mundo.
- Que un mismo imprimador amplifique en varias especies del género *Coffea*.
- Que la literatura reporte un PIC (Contenido de información del polimorfismo) diferente a cero para las muestras de Arábica.

En total se seleccionaron 27 microsatélites que cumplieron con al menos 2 de los 3 criterios de selección. Una vez realizada la selección se procedió a realizar la amplificación de los microsatélites mediante la técnica de PCR en 14 genotipos de *Coffea* procedentes del ICAFE. Con estos genotipos se realizaron 3 “pooles” con las muestras (Cuadro 4), esto con el fin de agilizar el proceso de selección de los imprimadores. Cada pool tiene un volumen de 100 μL y una concentración aproximada de 50 $\text{ng}/\mu\text{L}$.

Cada reacción tuvo en concentración final los siguientes componentes: 100 $\text{ng}/\mu\text{L}$ de ADN genómico aproximadamente, 1X de Master Mix (Maxima Hot Start PCR de Thermo ScientificTM), 2 mM de cloruro de magnesio, 0.2 μM de cada imprimador (forward y reverse). El volumen final de cada reacción fue de 25 μL .

Cuadro 3. Imprimadores seleccionados para amplificar las diferentes especies de *Coffea*

Marcador	Forward Primer	Reverse Primer	Repetición	Referencia
CaM03	CGCGCTTGCTCCCTCTGTCTCT	TGGGGGAGGGGCGGTGTT	(AC) ₁₁	8, 9
CaM16	AAGGCAGCTGAAGCGGGACAAA	TGGGGAGAGCTGCAGTTGGAGG	(TC) ₁₁	8, 9
CaM36	TGGTTTTAGTTTGTATTTTATTTGATGTGAT	CGAGCCCTCCCTTGCA	(TTA) ₇	8
CaM49	CCGGTTAATACATTGGTCTTT	ATGACATTGTTGACTTTGCTATAA	(A) ₃₃	8
CM2	TGTGATGCCATTAGCCTAGC	TCCAACATGTGCTGGTGATT	(AC) ₁₀ (AT) ₉	2
CM5	TGGAGGTAACGGAAGCTCTG	GCTAAGTTCAATTGCCCTGT	(CCT) ₇	2, 9
DCM01*	TTTTTGGGAAATGAAGGTGC	TGCACTTCAAGATCCCCTTT	(AG) ₁₅	1
DCM04*	TCAAAAACACATGGAAGCCA	CAACAGATCAGCTTTTGCCA	(ATA) ₆	1
M24	GGCTCGAGATATCTGTTTAG	TTTAATGGGCATAGGGTCC	(CA) ₁₅ (CG) ₄ CA	14, 15
M32	AACTCTCCATCCCCGCATTC	CTGGGTTTTCTGTGTTCTCG	(CA) ₃ /(CA) ₃ /(CA) ₁₈	10, 11
M47	TGATGGACAGGAGTTGATGG	TGCCAATCTACCTACCCCTT	(CT) ₉ (CA) ₈ /(CT) ₄ /(CA) ₅	6, 14
M753	GGAGACGCAGGTGGTAGAAG	TCGAGAAGTCTTGGGGTGTT	(AC) ₁₀	7, 14
M764	CTGGCATTAGAAAGCACCTTG	GCTTGGCTCACTGTAGGACTG	(CA) ₁₀	7, 14
M774	GCCACAAGTTTCGTGCTTTT	GGGTGTCGGTGTAGGTGTATG	(CA) ₇	7, 14
M779	TCCCCCATCTTTTCTTTCC	GGGAGTGTTTTGTGTTGCTT	(GT) ₁₇	7, 14
M793	CTGAGCGCATGGAAGGAGTA	GGAGACGCAGGTGGTAGAAG	(GT) ₁₀	7, 14
M837	CTCGCTTTCACGCTCTCTCT	CGGTATGTTCTCGTTCCTC	(GT) ₁₆ (GA) ₁₁	7, 14
SSR03*	GGACAAAACACCGCCAAAATA	AGCGAGACAGAGGAAGGGAATATT	(TC) ₁₃	1, 9
SSR04*	GGTCGCTCACTCATATCTCCAG	GCCTGGAAAGCAAACGTCTCA	(AC) ₁₀	1, 3
SSR09*	TTGGCTTTTGTCCCTCCTTCTCTG	AGCCCATTTCCCTCTCATATTCAAG	(GT) ₁₀	1, 9
SSR073*	GAGGTCTTCCCACCACAACA	GGATACGAGAGTCCCTTCCA	(CGC) ₂ (GGC) ₅	12, 13
SSRCa 018	GTCTCGTTTCACGCTCTCTC	ATTTTTGGCACGGTATGTTC	(GT) ₁₈ (GA) ₁₀	4, 5, 12, 13
SSRCa 023	GACCCTTGCCTTTTGTG	GCCATTCATCCATTCATTC	(AATG) ₃	12
SSRCa 082	GCTTGTTTCCATCGCTAAA	TTACACGTCAACCCACAAAC	(CT) ₁₇ CG(CT) ₆	12, 13
SSRCa 087	TCACTCTCGCAGACACTAC	GCAGAGATGATCACAAGTCC	(TC) ₂₂	4, 5, 12, 13
SSRCa 088	TACCTCTCCTCCTCCTCCT	ATTTCTATGGACCGGCAAC	(TTTTCT) ₃	4, 5, 12, 13

Referencias: 1. Aggarwal *et al.*, 2007 2. Baruah *et al.*, 2003 3. Bhat *et al.*, 2005 4. Brandao *et al.*, 2014 5. Brandao *et al.*, 2015 6. Combes *et al.*, 2000 7. Cubry *et al.*, 2007 8. Hendre *et al.*, 2008 9. Geleta *et al.*, 2012 10. Maluf *et al.*, 2005 11. Dessalegn *et al.*, 2009 12. Missio *et al.*, 2009 13. Missio *et al.*, 2011 14. Poncet *et al.*, 2004 15. Vieira *et al.*, 2010. *Marcadores son EST-SSR

Cuadro 4. Pooles genéticos conformados a partir de 14 genotipos de *Coffea* procedentes del banco de germoplasma del ICAFE, Barva de Heredia, Costa Rica.

Pool 1	Pool 2	Pool 3
Híbrido S296	Híbrido TUPI	Híbrido S-288
<i>C. arabica</i> variedad Villa Sarchí	Híbrido Oeiras	<i>C. liberica</i>
Híbrido I-59	Híbrido Catimor CR95	Híbrido S288-6
<i>C. excelsa</i>	Híbrido Catimor S175	Híbrido Catigua MC2
<i>C. arabica</i> var. Catuaí Amarillo	<i>C. arabica</i> var. Borbón	

6.5 Programa de amplificación de PCR

Las reacciones se llevaron a cabo en un termociclador Veriti® de Applied Biosystems, y se evaluaron los siguientes programas de amplificación:

- A. 1 ciclo de desnaturalización inicial a 95 °C durante 5 min; 30 ciclos de 95 °C durante 30 s, 52 °C durante 30 s y 72 °C durante 1 min, con una extensión final de 72 °C durante 8 min.
- B. 1 ciclo de desnaturalización inicial a 95 °C durante 5 min; 30 ciclos de 95 °C durante 30 s, 56 °C durante 30 s y 72 °C durante 1 min, con una extensión final de 72 °C durante 8 min.
- C. 1 ciclo de desnaturalización inicial a 95 °C durante 5 min; 35 ciclos de 95 °C durante 30 s, 55 °C durante 30 s y 72 °C durante 1 min, con una extensión final de 72 °C durante 8 min.

Tomando en cuenta la cantidad de SSR que amplificaron con cada programa y la calidad (intensidad) de la amplificación que mostraba el gel de agarosa, se seleccionó el programa de amplificación C. El programa tiene una duración de una hora y 44 minutos.

6.6 Electroforesis en gel de agarosa

Una vez realizada la PCR, se evaluó la presencia o ausencia de fragmentos amplificados para cada marcador mediante la visualización de una electroforesis en un gel de agarosa al 3 % (m/v), el cual se corrió por 45 minutos en buffer TAE 1X a 100 voltios constantes, colocando 5 μ L del producto de PCR y 1 μ L de buffer de carga con GelRed de Fermentas® en cada pozo (Anexo II).

6.7 Evaluación de polimorfismos

Para visualizar la amplificación de los posibles polimorfismos en los productos de PCR, se realizaron geles desnaturizantes de poliacrilamida al 10% utilizando la cámara DCode Universal Mutation Detection System (Bio-Rad®). Para preparar la solución desnaturizante se utilizaron: 25mL de una solución madre de Acrilamida/Bisacrilamida al 10% de Sigma®, 2 mL del buffer TAE 50X, 42 g de urea y se aforó a 100 mL con dH₂O. La polimerización de cada gel se realizó con 25 mL de la solución desnaturizante de poliacrilamida, 450 μ L de APS (persulfato de amonio) al 10% y 45 μ L de TEMED (N,N,N',N'-tetrametiletlenodiamina).

Una vez polimerizado se colocaron 10 μ L de los productos de PCR y 5 μ L de buffer de carga con formamida en cada pozo. Cada gel se corrió por 5 horas y media con buffer

TAE 1X a 200 voltios constantes. Los geles se colocaron en una bandeja con una solución de 100 mL de TAE (1X) y 10 μ L GelRed, finalmente se dejaron reposando por 45 min en la oscuridad para la tinción.

Los geles teñidos se visualizaron en el transiluminador, se fotografiaron y se analizaron (Anexo III), se seleccionaron los marcadores tomando en cuenta los siguientes criterios:

- Que se visualizara más de una banda por marcador
- Que se visualizará algún tipo de diferencia en el patrón de bandas entre los 3 pools.

Una vez analizados, se seleccionaron 14 microsatélites (Cuadro 5) que aparentaban presentar polimorfismos, y tomando en cuenta el tamaño aproximado en pares de bases de los alelos encontrados con electroforesis en geles de acrilamida, se marcó el imprimador “forward” con alguno de los fluorocromos 6FAMTM (azul), VIC® (verde), NEDTM (amarillo) o PET® (rojo) (Cuadro 6) para proceder a la evaluación mediante electroforesis capilar de los microsatélites seleccionados como polimórficos.

Cuadro 5. Imprimadores que presentan posibles polimorfismos seleccionados para marcar con fluorocromos

Marcador	Forward Primer	Reverse Primer	Repetición	Referencia
CaM03	CGCGCTTGCTCCCTCTGTCTCT	TGGGGGAGGGGCGGTGTT	(AC) ₁₁	6, 7
CaM16	AAGGCAGCTGAAGCGGGACAAA	TGGGGAGAGCTGCAGTTGGAGG	(TC) ₁₁	6, 7
M24	GGCTCGAGATATCTGTTTAG	TTTAATGGGCATAGGGTCC	(CA) ₁₅ (CG) ₄ CA	12, 13
M32	AACTCTCCATTCCCGCATTC	CTGGGTTTTCTGTGTTCTCG	(CA) ₃ / (CA) ₃ / (CA) ₁₈	8, 9
M753	GGAGACGCAGGTGGTAGAAG	TCGAGAAGTCTTGGGGTGT	(AC) ₁₀	5, 12
M764	CTGGCATTAGAAAGCACCTTG	GCTTGGCTCACTGTAGGACTG	(CA) ₁₀	5, 12
M774	GCCACAAGTTTCGTGCTTTT	GGGTGTCCGGTGTAGGTGTATG	(CA) ₇	5, 12
SSR03*	GGACAAAACACCGCCCAAAATA	AGCGAGACAGAGGAAGGGAATATT	(TC) ₁₃	1, 9
SSR04*	GGTCGCTCACTCATATCTTCCAG	GCCTGGAAAGCAAACGTCTCA	(AC) ₁₀	1, 2
SSR09*	TTGGCTTTTGTCCCTCCTTCCTCTG	AGCCCATTTCCCTCTCATCATTTCAAG	(GT) ₁₀	1, 9
SSR073*	GAGGTCTTCCCACCACAACA	GGATACGAGAGTCCCTTCCA	(CGC) ₂ (GGC) ₅	10, 11
SSRCa 018	GTCTCGTTTCACGCTCTCTC	ATTTTTGGCACGGTATGTTC	(GT) ₁₈ (GA) ₁₀	3, 4, 10, 11
SSRCa 087	TCACTCTCGCAGACACACTAC	GCAGAGATGATCACAAGTCC	(TC) ₂₂	3, 4, 10, 11
SSRCa 088	TACCTCTCCTCCTCCTTCT	ATTTCTATGGACCGGAAC	(TTTTCT) ₃	3, 4, 10, 11

Referencias: 1. Aggarwal *et al.*, 2007 2. Bhat *et al.*, 2005 3. Brandao *et al.*, 2014 4. Brandao *et al.*, 2015 5. Cubry *et al.*, 2007 6. Hendre *et al.*, 2008 7. Geleta *et al.*, 2012 8. Maluf *et al.*, 2005 9. Dessalegn *et al.*, 2009 10. Missio *et al.*, 2009 11. Missio *et al.*, 2011 12. Poncet *et al.*, 2004 13. Vieira *et al.*, 2010. *Marcadores son EST-SSR

Cuadro 6. Asignación del fluorocromo a los microsatélites seleccionados de acuerdo al tamaño esperado de alelos.

Marcador	Mínimo	Máximo	Fluorocromo
SSR04	110	150	NED
SSRCa87	115	140	6-FAM
CaM03	140	200	NED
M753	280	300	VIC
M774	175	230	PET
SSRCa18	100	130	6-FAM
SSRCa88	170	200	6-FAM
CaM16	160	210	6-FAM
M24	145	165	VIC
M764	130	160	NED
SSR09	90	130	6-FAM
M32	85	130	NED
SSR03	130	160	VIC
SSR073	140	160	PET

6.8 Amplificación de microsatélites fluorocromados

Una vez seleccionados y marcados los microsatélites se procedió a realizar la amplificación mediante la técnica de PCR utilizando el programa: 1 ciclo de desnaturalización inicial a 95 °C durante 5 min; 35 ciclos de 95 °C durante 30 s, 55 °C durante 30 s y 72 °C durante 1 min, con una extensión final de 72 °C durante 8 min, y se realizaron pruebas de amplificación con los 14 genotipos procedentes del ICAFE (Figura 1) mediante electroforesis capilar en el analizador genético ABI 3130xl de Applied Biosystems. Esto con el fin de determinar la cantidad de alelos y la forma de mezclar los productos de PCR para obtener una buena señal en la amplificación de los alelos.

Cada reacción de PCR tuvo en concentración final los siguientes componentes: 100 ng/ μ L de ADN genómico aproximadamente, 1X de Master Mix (Maxima Hot Start PCR de

Thermo Scientific™), 2 mM de cloruro de magnesio, la cantidad de cada imprimador (forward fluorocromado y reverse) varía dependiendo de la señal de amplificación (Cuadro 7).

Las reacciones de PCR se realizaron de forma individual en los 56 genotipos de *Coffea* (Cuadro 1 y 2).

Cuadro 7. Concentración final de imprimador necesario para cada reacción de PCR

Imprimador	Foward y Reverse c/u (μM)
M779	0,14
CaM16	0,14
SSR09	0,14
SSRCa87	0,15
CaM03	0,15
SSRCa88	0,15
M764	0,15
SSR04	0,2
SSRCa18	0,2
M753	0,2
M774	0,2
SSR03	0,2
M32	0,2
SSR073	0,2

En total se realizaron 4 mezclas con los productos de PCR marcados (Cuadro 8). En las mezclas donde se repiten los mismos colorantes fluorescentes, se tomó en cuenta el tamaño de los productos de PCR y se pusieron juntos los que tienen al menos 30 pb de diferencia para evitar la superposición de los alelos.

Cuadro 8. Mezclas asignadas de los productos de PCR para la electroforesis capilar

Mezclas	Imprimador	Producto de PCR (μL)
Mezcla 1	SSRCa87	4
	M779	3
	SSR04	3
Mezcla 2	CaM03	4
	SSRCa18	4
	SSRCa88	3
	M753	5
	M774	4
Mezcla 3	CaM16	2
	M24	3
	M764	4
	SSR09	3
Mezcla 4	SSR03	4
	M32	4
	SSR073	4

6.9 Electroforesis capilar

Los productos de la amplificación se separaron mediante electroforesis capilar en un analizador genético ABI 3130xl (Applied Biosystems). Cada muestra analizada se preparó utilizando 1,5 µL de la mezcla de productos de PCR (Cuadro 8), 0,4 µL de marcador de peso molecular GeneScan™-600 LIZ® de Thermo Scientific (para determinar el tamaño de los fragmentos) y 8.5 µL de formamida altamente desionizada.

Los datos de salida del ABI 3130xl fueron analizados en el programa GeneMapper® V4.0 el cual permitió etiquetar los alelos para cada imprimador según el ámbito esperado en pares de bases. Este etiquetado se trasladó a Excel (MS Office) para realizar el análisis estadístico.

6.10 Análisis estadístico

Se realizó la estimación del grado de polimorfismo de cada microsatélite mediante el cálculo del índice de contenido polimórfico (PIC), el cual expresa el grado de polimorfismo que es capaz de detectar un determinado marcador molecular dentro de una población (Powell *et al.*, 1996). El valor varía dependiendo del número de alelos detectables y la distribución de su frecuencia, por lo tanto, proporciona una estimación del poder de discriminación del marcador (Nagy *et al.*, 2012). Este parámetro fue medido de acuerdo a la fórmula de Botstein *et al.* (1980):

$$PIC = 1 - \left(\sum_{i=1}^n p_i^2 \right) - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i^2 p_j^2$$

Donde p_i y p_j son las frecuencias alélicas en los alelos i y j , respectivamente, y n es el número de alelos. Para el cálculo de este índice se utilizó el paquete Polysat 1.7 del programa R (Clark & Jasieniuk, 2011).

Para los análisis descriptivos de la diversidad genética se determinó el número de alelos observados, el número de alelos raros, se realizó el cálculo de la frecuencia alélica, de heterocigosidad esperada (HE) (Nei, 1987) y de heterocigosidad esperada corregida por tamaño de muestra ($HE_{(c)}$) (Van Puyvelde *et al.*, 2010):

$$HE = (1 - \sum_{i=1}^s p_i^2)$$

$$HE_{(c)} = \frac{4n}{4n - 1} (1 - \sum_{i=1}^s p_i^2)$$

Además se calculó el índice de diversidad de Shannon- Wiener (H') (Van Puyvelde *et al.*, 2010) mediante la fórmula:

$$H' = \sum_{i=1}^s (p_i \log p_i)$$

Donde p_i es la frecuencia del alelo i en una población en un locus dado, n es el tamaño de la muestra y s es el número de genotipos.

La diferenciación genética entre poblaciones se calculó mediante el coeficiente de diferenciación de genes (GST) el cual es un índice similar a FST pero derivado especialmente para tratar con locus multialélicos (Nei, 1973). Este sistema está diseñado teniendo en cuenta la migración, selección y mutación. Este índice se calculó mediante la siguiente fórmula:

$$G_{ST} = \frac{D_{ST}}{H_T}$$

Donde D_{ST} es la diversidad genética entre subpoblaciones y H_T es la diversidad genética entre la población total.

También se calculó el índice de diferenciación R_{ST} , el cual se diseñó bajo el modelo de mutación por pasos (SMM) y refleja con mayor precisión el patrón de mutación de los microsatélites (Slatkin, 1995). Para su cálculo se utilizó la fórmula:

$$R_{ST} = \frac{D_m}{H_s}$$

Donde D_m es la diversidad genética interpoblacional y H_s la diversidad genética dentro de poblaciones.

Tanto la diversidad genética como la diferenciación genética fueron estimadas con el software ATETRA, el cuál es un programa diseñado especialmente para el análisis de datos obtenidos por microsatélites a partir de especies tetraploides (Van Puyvelde *et al.*, 2010). Para realizar estos análisis se utilizaron 10 000 permutaciones con un intervalo de confianza del 95%.

Además se calcularon las distancias genéticas de Bruvo, esta es una medida de distancia genética similar a los índices de distribución de bandas utilizados con los datos dominantes, pero teniendo en cuenta las distancias mutacionales entre alelos, y permite la comparación de individuos con diferentes niveles de ploidía (Bruvo *et al.*, 2004).

A partir de la matriz obtenida por las distancias genéticas se realizó un análisis de coordenadas principales (PCoA) mediante el paquete InfoStat (Di Rienzo *et al.*, 2017). Posteriormente utilizando la misma matriz se utilizó el paquete pvclust del programa R, el

cual permite calcular los valores de probabilidad (p-valores) para cada grupo utilizando técnicas de remuestreo “bootstrap” y realiza el análisis de agrupamiento, el cual tiene como objetivo clasificar varios objetos en grupos (clusters) de acuerdo a las similitudes entre ellos (Suzuki & Shimodaira, 2006).

7. RESULTADOS Y DISCUSIÓN

7.1 Análisis de diversidad genética y polimorfismo de los marcadores microsatélites

Se analizaron en total 55 muestras del género *Coffea* con 14 marcadores (10 SSR's y 4 EST-SSR) (Cuadro 5) y se encontraron 90 alelos diferentes en total. Las muestras se dividieron en 2 grupos: las arábicas (en las cuales se incluyen variedades e híbridos) y las no arábicas. En las muestras de café arábico (n=52) se encontraron 64 alelos diferentes, obteniendo un promedio de 4,57 alelos por marcador. Mientras que en las otras especies no arábicas (n=3) se encontraron 49 alelos, con un promedio de 3,50 alelos por marcador (Cuadro 9). Además se contabilizaron 23 alelos en común, es decir que se encontraron tanto en las muestras de arábica como en las no arábicas.

El número total de alelos y el promedio de alelos por marcador obtenido en este estudio en las muestras de arábicas, es superior al reportado por otros autores. En el caso de Aggarwal *et al.*, (2007) utilizaron 16 SSR's (n=15), ellos encontraron 62 alelos y obtuvieron un promedio de 3,4 alelos por marcador, mientras que Combes *et al.*, 2000 utilizaron 11 SSR's (n=32), encontraron 40 alelos y obtuvieron un promedio de 3,6 alelos por marcador. Por otro lado, Missio *et al.*, (2009) utilizaron una mayor cantidad de marcadores (26 SSR's) y encontraron 47 alelos en 19 muestras, sin embargo el promedio de alelos por marcador fue

mucho menor (1,8) que el reportado por los autores anteriormente mencionados. Hendre *et al.*, 2008, si obtuvieron un número mayor de alelos (177) utilizando 44 SSR's en tan solo 8 muestras, pero el número promedio de alelos por marcador sigue siendo menor (2,66) que el encontrado en este estudio.

Es importante recalcar que a pesar de que los autores utilizaron muestras de café arábico, todos estos resultados se obtuvieron a partir de híbridos y variedades distintos a las utilizadas en este estudio, y también por distintos SSR's, por lo que la diversidad cambia. Sin embargo podemos ver que el rango de promedios acá mostrados es bajo, esto debido a que las muestras de café arábico mantienen una base genética estrecha debido a su reproducción, evolución del genoma y su proceso de domesticación (Hendre *et al.*, 2008). Además de estos factores es importante tomar en cuenta la técnica utilizada para la caracterización molecular, ya que tanto Missio *et al.*, 2009 como Combes *et al.*, 2000 encontraron los alelos por medio de geles desnaturizante de poliacrilamida, lo cual es una técnica de bajo costo pero tiene el limitante de que presentan baja reproducibilidad, no siempre se puede calcular con precisión los tamaños de los alelos y ni se obtiene el registro de los datos en formato digital, por lo que el análisis de los datos puede ser problemático (Wang *et al.*, 2009), mientras que Aggarwal *et al.*, 2007 utilizó geles de poliacrilamida pero bajo un sistema automatizado que analiza los fragmento amplificados de ADN (marcados con fluorescencia) por medio de electroforesis en gel que mejora la reproducibilidad. Finalmente Hendre *et al.*, 2008, amplificó el ADN usando imprimadores universales M13 para ambas cadenas, para ello se utilizó un analizador automatizado de ADN usando la química BigDye™ (Applied Biosystems), esta técnica aunque es más costosa que las anteriores tiene mayor sensibilidad y reproducibilidad, además presenta un principio similar al que se utilizó en este estudio, el cual fue bajo electroforesis capilar mediante un analizador genético automatizado (ABI 3130xl), esta técnica en combinación con el uso de imprimadores marcados con fluorescencia proporciona una alta sensibilidad de detección de fragmentos amplificados (Wang *et al.*, 2009, Huang *et al.*, 2006) y permite detectar polimorfismos de un único nucleótido (Wang *et al.*, 2009) además presenta un software que permite visualizar el tamaño de los fragmentos mediante electroferogramas,

y almacena los datos en forma digital, los cuales pueden ser exportarlos a diferentes programas estadísticos para su análisis.

Por otro lado, pudimos observar que en general la presencia de alelos nulos fue menor en las muestras de arábicas (Anexo IV), y que no hubo presencia de alelos raros (alelos con una frecuencia menor a 0,01) en las muestras de café no arábico. Sin embargo para las muestras de café arábico con los marcadores CaM03, M24, M32, SSR03, SSR09 y SSRCa87 si se observó presencia de alelos raros (Anexo V).

Según Jain *et al.*, 2004 los alelos raros son altamente informativos para la genotipificación de variedades, ya que pueden ser indicadores de la presencia de variantes genéticas únicas. Sin embargo estos resultados podrían deberse a la metodología que se empleó para el cálculo de estas mediante el paquete estadístico Polysat, ya que este programa permite estimar las frecuencias alélicas de loci con ambigüedad en el dosaje alélico, asumiendo que todos los alelos tienen la misma probabilidad de estar presentes en múltiples copias, y bajo esta estimación hay una tendencia a subestimar la frecuencia de los alelos comunes y a sobreestimar la frecuencia de los alelos raros (Clark & Jasieniuk, 2011).

En cuanto al Índice de Contenido Polimórfico (PIC) en el café arábico los valores se encontraron en un rango entre 0,4095 (M764, M774 y SSRCa88) a 0,7171 (CaM13) con un promedio de 0,5183. Mientras que en las especies no arábicas los valores de PIC se encontraron en un rango entre 0,3654 (M774) a 0,7973 (SSR03) con un promedio de 0,5792. Podemos observar que este grupo de marcadores son bastante informativos tanto para arábicas como para no arábicas. Sin embargo el valor de PIC para las especies no arábicas sigue siendo mayor, tal y como lo describen Ferrao *et al.*, 2015, Missio *et al.*, 2009 y Hendre *et al.*, 2008, en estudios anteriores. Estos autores atribuyen este mayor contenido polimórfico en las especies no arábicas a su proceso de reproducción por cruzamiento, mientras que las arábicas son autocompatibles, además mencionan que las arábicas siempre van a presentar menor polimorfismo que el resto de las especies de *Coffea* esto debido a la evolución del genoma y su proceso de domesticación.

El marcador CaM03 fue el que presentó mayor cantidad de alelos (10 alelos) en la población de café arábica ($n=52$), estos con un tamaño dentro de un rango de 161-188 pares de bases. Hendre *et al.*, 2008 encontró 6 alelos entre 164-184 pares de bases entre con este mismo marcador, lo cual coincide con el rango encontrado en este estudio.

En la población de café no arábicas SSR03 fue el marcador con mayor cantidad de alelos (6 alelos) en 3 muestras, fue el marcador con más variación genética y uno de los marcadores más polimórficos tanto en arábicas ($PIC=0,6712$, $He=0,7248$, $He_{(c)}=0,7283$) como en no arábicas ($PIC=0,7973$, $He=0,7731$, $He_{(c)}=0,8434$), esto a pesar de que SSR03 es uno de los marcadores EST-SRR que provienen de regiones genómicas altamente conservadas, y que generalmente ocasiona un menor grado de polimorfismo en comparación con los microsátélites originarios a partir de bibliotecas genómicas (Motta *et al.*, 2015), es por ello que Missio *et al.*, (2011) recomienda el uso combinado de los marcadores SSR y EST-SSR debido a pueden proporcionar información complementaria importante.

En general pudimos observar que este grupo de marcadores utilizados presentan una gran variación genética (He) y un alto grado de polimorfismo (PIC) en ambos grupos de muestras (Cuadro 9), como se esperaba, ya que estos imprimadores se seleccionaron por ser polimórficos en otros estudios con estas especies (Aggarwal *et al.*, 2007, Dessalegn *et al.*, 2009, Geleta *et al.*, 2012, Hendre *et al.*, 2008, Hendre & Aggarwal, 2014, Missio *et al.*, 2009, Motta *et al.*, 2015, Poncet *et al.*, 2004, Vieira *et al.*, 2010).

En cuanto a la diversidad genética de ambos grupos, las especies no arábicas mostraron en promedio un índice de Shannon mayor ($H'=1,1268$) que las muestras de arábicas ($H'=1,0422$), estos resultados están de acuerdo con Anthony *et al.*, 2001, Lashermes *et al.*, 1996, Maluf *et al.*, 2005, Moncada & McCouch 2004, quienes ya han definido la estrecha base genética de las plantas cultivadas de *C. arabica* debido a su origen. Hechos históricos indican que las plantas en el continente americano tienen un origen común, ya que se obtuvieron a partir de unas pocas plantas introducidas a principios del siglo XVIII y estas

plantas se originaron de las primeras plantas cultivadas en Yemen (Anthony *et al.* 1999, León, 2000, Silvestrini *et al.*, 2007).

Cuadro 9. Parámetros de diversidad genética obtenida por 14 marcadores microsatélites en muestras de café arábico y no arábico

N_{Alelos} : número de alelos, He : Heterocigosidad esperada, $He_{(C)}$: Heterocigosidad esperada corregida, H' : índice de diversidad de Shannon, PIC : Índice de contenido polimórfico, DS : Desviación estándar

<u>Arábicas (n=52)</u>							<u>No arábicas (n=3)</u>				
Marcador	Rango de alelos pb	N_{Alelos}	He	$He_{(C)}$	H'	PIC	N_{Alelos}	He	$He_{(C)}$	H'	PIC
CaM03	161-188	10	0,7137 [DS= 0,0112]	0,7171 [DS= 0,0112]	1,5287 [DS= 0,0270]	0,7171	4	0,6475 [DS= 0,0468]	0,7064 [DS= 0,0511]	1,1930 [DS= 0,0822]	0,6678
CaM16	184-200	4	0,5834 [DS= 0,0102]	0,5862 [DS= 0,0102]	1,0008 [DS= 0,0242]	0,4959	4	0,7036 [DS= 0,0349]	0,7676 [DS= 0,0381]	1,2860 [DS= 0,0733]	0,6620
M24	149-169	8	0,4573 [DS= 0,0174]	0,4595 [DS= 0,0175]	1,0359 [DS= 0,0405]	0,5221	4	0,6482 [DS= 0,0461]	0,7071 [DS= 0,0503]	1,1943 [DS= 0,0811]	0,6678
M32	99-131	8	0,6966 [DS= 0,0138]	0,6999 [DS= 0,0139]	1,4845 [DS= 0,0366]	0,6566	2	0,4907 [DS= 0,0065]	0,5353 [DS= 0,0071]	0,8384 [DS= 0,0205]	0,4095
M753	283-296	3	0,5513 [DS= 0,0060]	0,5540 [DS= 0,0060]	0,8720 [DS= 0,0145]	0,4635	5	0,7500 [DS= 0,0351]	0,8181 [DS= 0,0383]	1,4881 [DS= 0,0752]	0,7485
M764	148-157	2	0,4984 [DS= 0,0022]	0,5008 [DS= 0,0022]	0,6916 [DS= 0,0022]	0,4095	4	0,6480 [DS= 0,0465]	0,7070 [DS= 0,0508]	1,1941 [DS= 0,0817]	0,6678
M774	193-224	2	0,4984 [DS= 0,0023]	0,5008 [DS= 0,0022]	0,6915 [DS= 0,0023]	0,4095	2	0,2692 [DS= 0,0908]	0,2937 [DS= 0,0991]	0,4341 [DS= 0,1130]	0,3654
SSR03	134-156	8	0,7248 [DS= 0,0030]	0,7283 [DS= 0,0030]	1,4914 [DS= 0,0129]	0,6712	6	0,7731 [DS= 0,0344]	0,8434 [DS= 0,0375]	1,6327 [DS= 0,0722]	0,7973
SSR04	116-148	2	0,6593 [DS= 0,0017]	0,6625 [DS= 0,0017]	1,0874 [DS= 0,0023]	0,4095	2	0,6017 [DS= 0,0461]	0,6564 [DS= 0,0503]	0,9918 [DS= 0,0792]	0,3935
SSR073	147-155	4	0,5532 [DS= 0,0024]	0,5558 [DS= 0,0024]	0,9337 [DS= 0,0032]	0,5174	2	0,6667 [DS= 0,0000]	0,7273 [DS= 0,0000]	1,0986 [DS= 0,0000]	0,4611
SSR09	118-126	3	0,5438 [DS= 0,0043]	0,5464 [DS= 0,0044]	0,8743 [DS= 0,0158]	0,4198	4	0,639 [DS= 0,0800]	0,6979 [DS= 0,0872]	1,1861 [DS= 0,1336]	0,5436
SSRCa18	110-126	4	0,6305 [DS= 0,0098]	0,6336 [DS= 0,0099]	1,0900 [DS= 0,0228]	0,5758	4	0,7129 [DS= 0,0065]	0,7778 [DS= 0,0071]	1,3006 [DS= 0,0205]	0,5730
SSRCa87	116-135	4	0,6294 [DS= 0,0117]	0,6324 [DS= 0,0117]	1,1174 [DS= 0,0199]	0,5789	3	0,6667 [DS= 0,0000]	0,7273 [DS= 0,0000]	1,0986 [DS= 0,0000]	0,5789
SSRCa88	171-193	2	0,4984 [DS= 0,0022]	0,5008 [DS= 0,0022]	0,6916 [DS= 0,0022]	0,4095	3	0,4908 [DS= 0,0066]	0,5354 [DS= 0,0072]	0,8387 [DS= 0,0206]	0,5730
Promedio		4.57	0,5913	0,5885	1,0422	0,5183	3.50	0,6220	0,6786	1,1268	0,5792

7.2 Análisis de diferenciación genética

Los índices de diferenciación de genes para alelos múltiples (G_{ST}) variaron entre 0,0412 y 0,4472 con un valor promedio de 0,1457. De acuerdo a las guías de Wright (1978), cuando los marcadores obtienen valores que van desde 0-0,05 indican poca diferenciación genética, 0,05-0,15 indican que es moderado, de 0,15-0,25 indican que la diferenciación genética es grande y los valores por encima de 0,25 indican diferenciación muy grande.

En el cuadro 10 podemos observar que la mayoría de los marcadores poseen valores de G_{ST} entre 0,05 y 0,25, y de acuerdo a las guías se consideran como marcadores que indican diferenciación genética moderada y alta entre las muestras de café arábico y las muestras de café no arábico. Sin embargo, hay 2 marcadores que se salen de este rango de valores, el marcador CaM16 que posee un valor de 0,0412, indicando poca diferenciación genética y el marcador M774 es cual indica una diferenciación genética muy grande, ya que posee un valor de G_{ST} mayor a 0,25.

Sin embargo Whitlock (2011) ha notado que cuando la diversidad genética es alta, G_{ST} está limitado a un valor máximo menor que 1 y en el caso de microsatélites con alta heterocigosidad el G_{ST} máximo se encuentra generalmente entre 0,1 y 0,2, como sucedió en este estudio con la mayoría de los marcadores. Además G_{ST} es sensible a la variación de la tasa de mutación entre loci, principalmente cuando la tasa de mutación se aproxima o supera la tasa de migración, lo que conduce a que ciertos loci principalmente microsatélites posean valores de G_{ST} muy bajos con respecto a otro marcadores. Por lo tanto es en estos casos cuando hay que tener cierto cuidado a la hora de interpretar los datos, ya que las guías de Wright mencionadas anteriormente pueden ser engañosas o no muy informativas. Es por ello que se decidió complementar esta información con el índice de diferenciación R_{ST} , propuesto por Slatkin, (1995) que refleja de mejor forma el patrón de mutación de los microsatélites.

Los valores obtenidos por el estadístico de diferenciación poblacional R_{ST} calculado entre ambos grupos (arábicas y no arábicas) para los 14 microsatélites analizados, oscilaron entre 0,0866 y 1,6457, con un valor promedio de 0,3853 (Cuadro 10). Este valor es mayor que el obtenido mediante el estadístico G_{ST} (0,1457). Podemos observar que para los valores de R_{ST} , el marcador CaM16 ahora muestra diferenciación moderada, y el resto de marcadores

poseen valores que indican una heterogeneidad genética grande y muy grande entre las muestras de café arábica y no arábicas. Tanto con el estadístico R_{ST} como con G_{ST} todos los marcadores revelaron diferencias estadísticamente significativas entre ambos grupos (Cuadro 10).

Cuadro 10. Parámetros de diferenciación genética de cada marcador entre muestras de café arábico y no arábico.

Marcador	G_{ST}	R_{ST}	95% Conf
CaM03	0,1873	0,4621	0,0003
CaM16	0,0412	0,0866	0,0003
M24	0,2079	0,5295	0,0007
M32	0,0982	0,2177	0,0001
M753	0,1226	0,2828	0,0007
M764	0,0748	0,1635	0,0005
M774	0,4472	1,6457	0,0009
SSR03	0,0888	0,1952	0,0002
SSR04	0,1615	0,3862	0,0003
SSR073	0,1118	0,2519	0,0001
SSR09	0,0726	0,1580	0,0005
SSRCa18	0,1578	0,3758	0,0003
SSRCa87	0,1866	0,4588	0,0001
SSRCa88	0,0819	0,1809	0,0006
Promedio	0,1457	0,3853	

El análisis de coordenadas principales (PCoA) mostró un agrupamiento de todas las muestras de cafés arábicos, mientras que las otras especies estaban aisladas, lo cual demuestra que este grupo de marcadores permite diferenciar claramente entre las muestras arábicas y

las no arábicas (Figura 2). El primer eje que separa los grupos explica el 35,5% de la variación total en el conjunto de datos, mientras que el segundo eje que explica la variación entre las muestras individuales dentro de los grupos explica sólo el 11,9% de la variación total.

La forma de agrupamiento observada en el PCoA muestra concordancia con el dendrograma (Figura 3), que a partir de las distancias genéticas también muestra una clara separación entre las muestras de arábicas y las otras especies, casi todos los agrupamientos (clusters) tienen errores estándar inferiores a 0,01 (Anexo VI).

Además en este dendrograma podemos distinguir algunas pequeñas agrupaciones entre híbridos derivados de padres etíopes (Figura 3 A.), también vemos una agrupación entre las accesiones que no pudieron ser separadas con este grupo de marcadores (Figura 3 B.) sin embargo las agrupaciones entre las especies de arábicas no es clara y seguimos observando una distancia genética muy pequeña entre las muestras, esto producto de la base genética estrecha de la especie (Anthony *et al* 2002). Otro factor que influye en la alta similitud genética entre las muestras es que muchas de las accesiones utilizadas en este estudio son híbridos cuyos padres se derivan principalmente de híbridos Timor que generalmente son utilizados como fuente de genes resistentes a la roya, y se sabe que los investigadores utilizan las mismas fuentes de genes resistentes a la roya en todo el mundo (Sousa *et al.*, 2017)

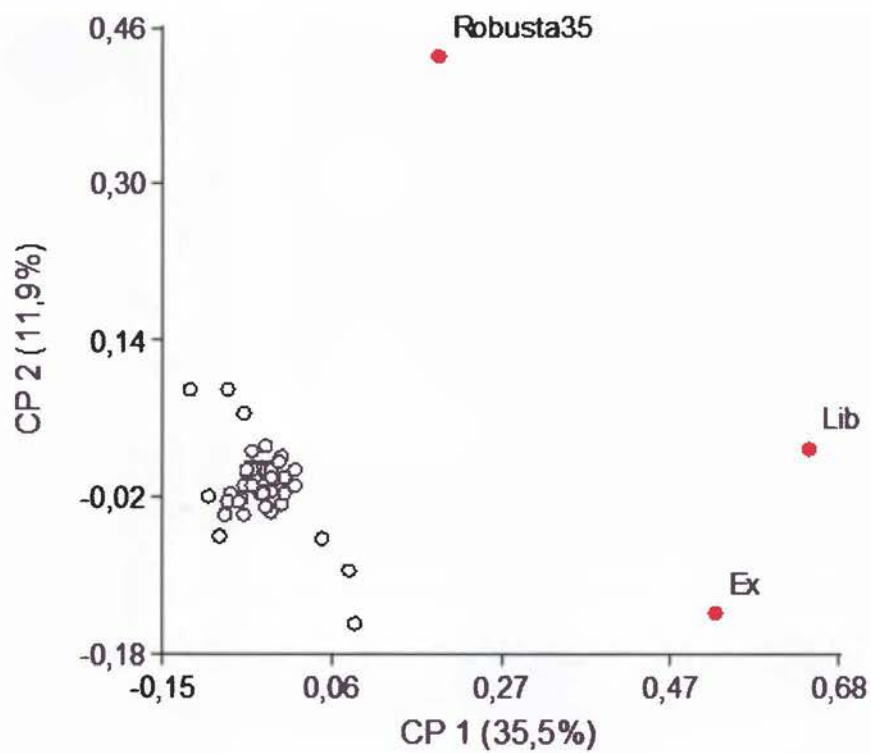


Figura 2. PCoA utilizando las distancias de Bruvo, los círculos blancos representan las muestras de café arábico y los círculos rojos las muestras de café no arábico. El porcentaje de varianza explicado por cada eje se proporciona dentro de la figura.

7.3 Perfil molecular

En los perfiles moleculares obtenidos, pudimos ver claramente como los 14 marcadores permiten diferenciar las especies arábicas de las no arábicas. Las 3 accesiones que presentaron mayor cantidad de marcadores polimórficos para diferenciar entre este grupo de muestras fueron liberica, excelsa y robusta35. En caso de la accesión de café liberica “Lib” fue la única muestra en la que pudimos observar polimorfismos en los 14 marcadores, por lo que tanto todos ellos permiten diferenciar esta accesión del resto de las muestras, en el caso de la accesión de excelsa “Ex” se presentaron polimorfismos en 10 de los marcadores y en la accesión de robusta35, 9 marcadores presentaron polimorfismos.

En cuanto a las especies arábicas, este conjunto de 14 marcadores no fueron efectivos para diferenciar entre las accesiones de Borbón, Oeiras, Villa Sarchí y Venecia. Además se obtuvo el mismo perfil molecular para las accesiones S288 y S288-6 las cuales pertenecen a la misma especie. También hubo homología entre las accesiones T.21239 y T.21242, sin embargo en el caso de estas 2 muestras no contamos con la información de su origen para saber si se trata del mismo híbrido.

En la figura 4 se muestra un ejemplo de los electroferogramas obtenidos para las muestras de arábica CR95 y MC2, las cuales no pueden ser diferenciadas por el marcador M774 ya que no hay polimorfismo entre estas, sin embargo podemos ver como se diferencia claramente de la muestra Liberica la cual corresponde a una especie distinta (*Coffea liberica*).

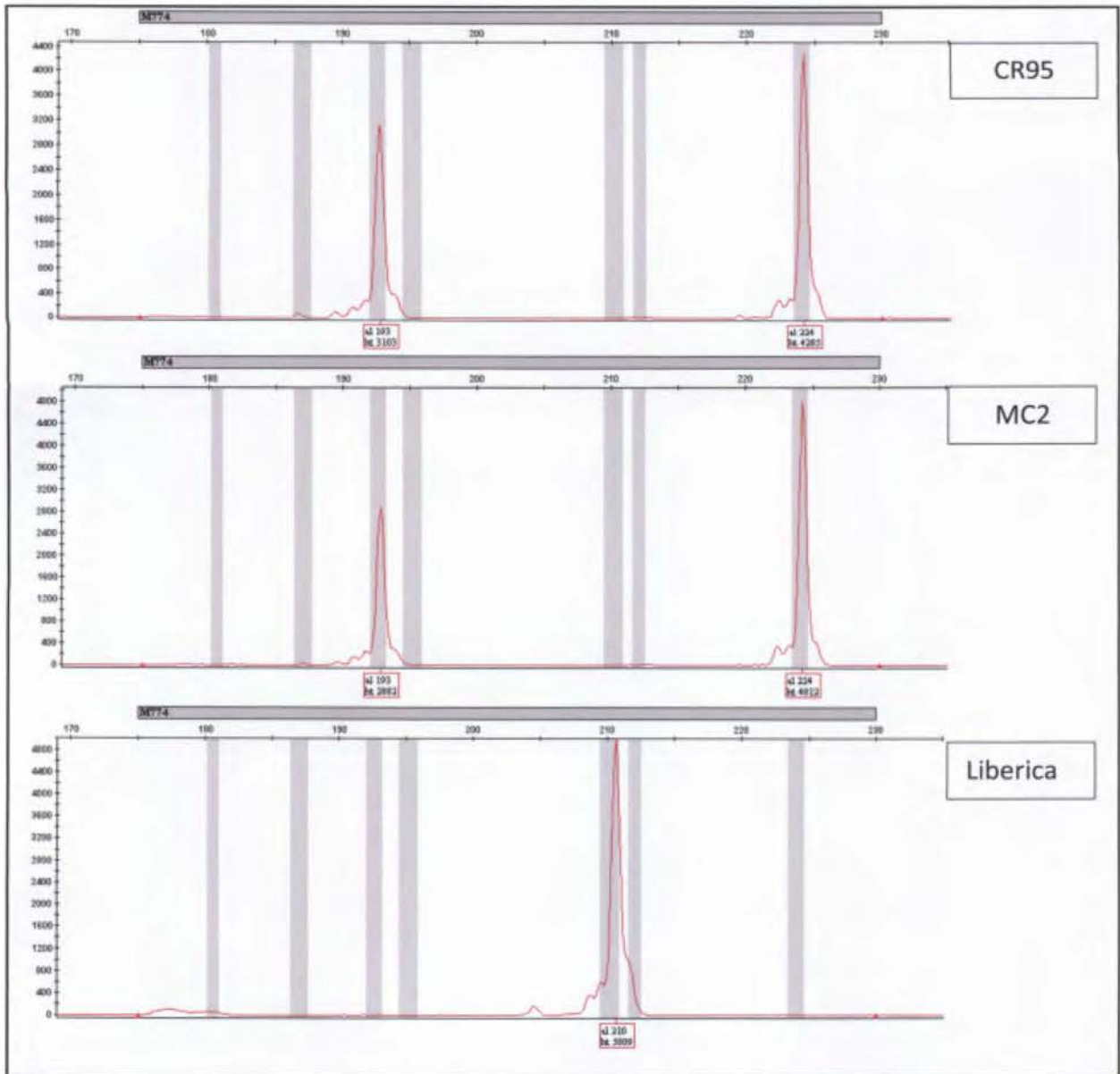


Figura 4. Electroferograma mostrando los alelos del SSR locus M774 para 3 muestras distintas de *Coffea*

Los marcadores SSR04, SSRCa88, M764 y M774 solo funcionaron para diferenciar entre las arábicas y las no arábicas, ya que dentro de las arábicas estos 4 marcadores no muestran polimorfismos, y para todas las muestras se observan solo 2 alelos (Figura 5), lo cual puede sugerir que sea el resultado de un loci fijo duplicado en el genoma de arábica, ya

que hay que recordar que esta especie es un alotetraploide producto de la hibridación de dos genomas (*C. eugenoides* y *C. canephora*) (Hendre *et al.*, 2008).

El marcador SSR04 además de no ser polimórfico entre arábicas, presentó el inconveniente de que no amplificó en 14 accesiones, lo que corresponde al 27% de las muestras estudiadas. Otros marcadores que presentaron solo 2 alelos en arábicas son el SSR09 y el SSR073, sin embargo el primero presentó un polimorfismo en la accesión S175 y no amplificó en las accesiones T.3099 y L02A30, mientras que el segundo presentó polimorfismos en las accesiones L12A28 y L13A22IC, y este no amplificó en la accesión L10A25.

Los pares de muestra L04A5 y L04A5IC, L12A28 y L12A28, L13A22 y L13A22IC, L13A44 y L13A44IC proceden de diferente banco de germoplasma (terminación IC proceden del ICAFE) y poseen un código de identificación igual pero no presentaron un perfil molecular idéntico, sin embargo no fue posible conocer el origen de las muestras para saber si se trata del mismo híbrido. Los marcadores CaM03, CaM16, M24, M32, M753, SSR03, SSRCa18 y SSRCa87 son polimórficos en muchas de las accesiones de café arábico y pudimos observar que utilizando en conjunto este grupo de 8 microsatélites podemos diferenciar entre todas las accesiones estudiadas, con excepción de las 4 accesiones que mostraron el mismo perfil molecular explicado anteriormente.

Por lo tanto, podemos decir que este conjunto de marcadores son altamente polimórficos y son bastante efectivos, ya que nos permiten diferenciar claramente entre las muestras de arábicas y no arábicas. Además nos permite diferenciar entre las diferentes accesiones de arábicas la cuales incluyen híbridos y variedades.

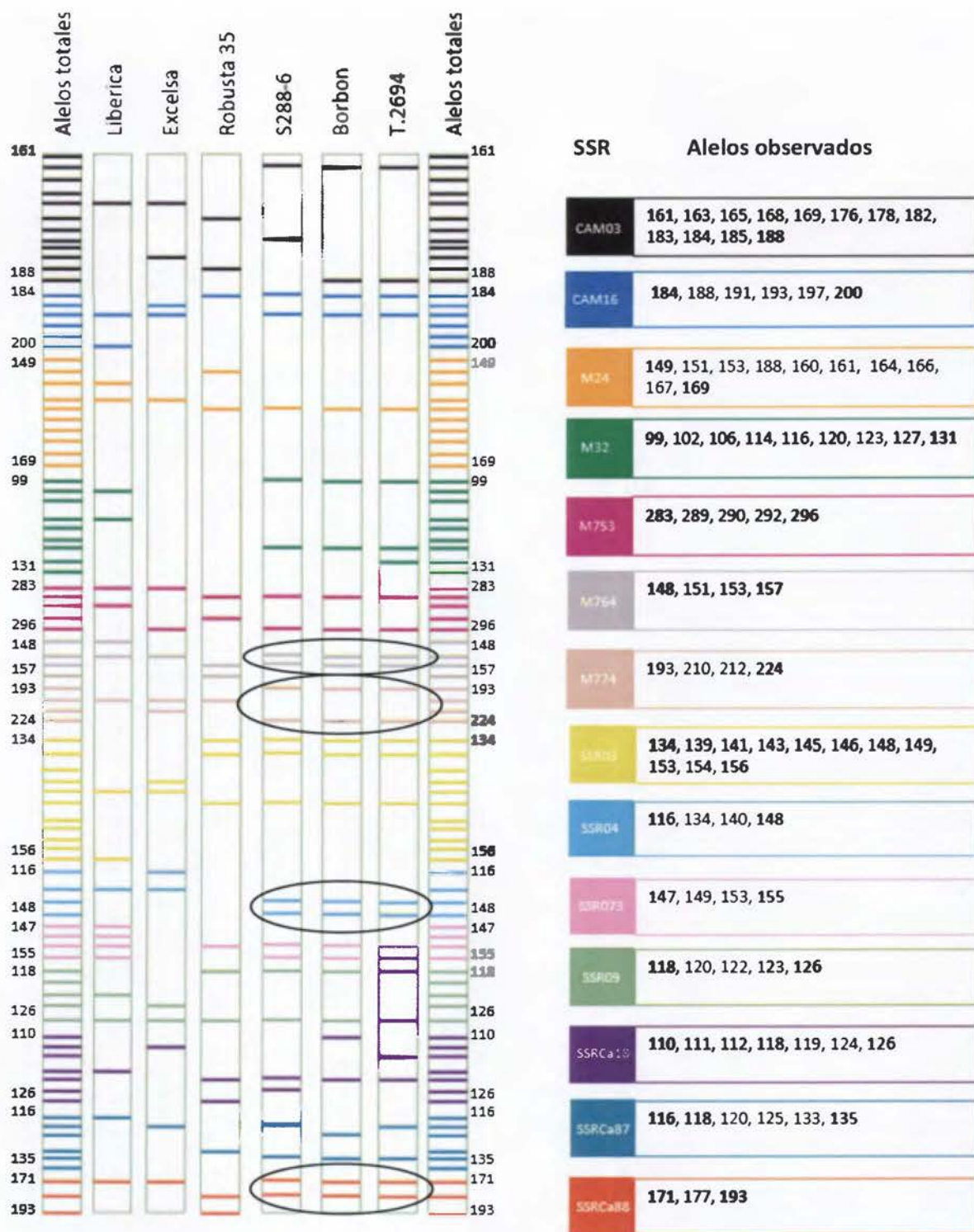


Figura 5. Perfil molecular obtenido a partir de los 14 marcadores para 3 muestras distintas de café arábico y 3 muestras distintas de café no arábico. Se señalan los microsatélites que no presentaron polimorfismos en las muestras de café arábico.

8. CONCLUSIONES

- El grupo de 14 marcadores microsatélites utilizados en este estudio permitieron diferenciar eficientemente entre las especies de café arábico y las de café no arábico.

En las muestras de arábicas (híbridos y variedades) el material de estudio resultó ser bastante homogéneo, lo que no permitió que se diferenciara el 100% de las muestras analizadas y dificultó un buen análisis de agrupamiento, lo cual respalda la teoría de que en la especie existe una base genética estrecha debido a su origen y proceso de domesticación.

- El programa de amplificación: 1 ciclo de desnaturalización inicial a 95 °C durante 5 min; 35 ciclos de 95 °C durante 30 s, 55 °C durante 30 s y 72 °C durante 1 min, con una extensión final de 72 °C durante 8 min, resultó ser el óptimo para este grupo de marcadores, permitió una buena amplificación tanto en las muestras de ADN individuales como en los pools genéticos.
- Una vez optimizado el protocolo, el análisis por medio de electroforesis capilar resultó ser bastante rápido y eficiente, nos permitió evaluar los marcadores, y obtener información tanto de diversidad como de diferenciación genética entre las muestras.
- Los marcadores microsatélites analizados en este estudio resultaron ser altamente polimórficos, nos permitieron realizar un perfil molecular con el cual podemos diferenciar entre las diferentes especies de *Coffea*, y nos permitió identificar la mayoría de las muestras de café arábico. Este grupo de marcadores facilitaría la caracterización molecular del germoplasma de café en Costa Rica y puede convertirse en una gran herramienta para realizar estudios de paternidad, diversidad, selección, estructura poblacional e incluso para el registro o certificación de café con alta calidad de taza.

9. RECOMENDACIONES

- Optimizar la metodología para implementar PCR multiplex, esto con el fin de reducir el costo y el tiempo de la genotipificación
- En el caso de trabajar con híbridos se recomienda conocer los padres de todos los cruces que se hacen para obtener el mismo, esto con el fin de poder establecer relaciones y analizar mejor los agrupamientos producidos por distancias las distancias genéticas.
- Debido a la estrecha base genética del café arábico, para el análisis de los híbridos y variedades de esta especie se recomienda valorar un grupo mayor de marcadores.
- Este es un estudio preliminar acerca del potencial que tienen los marcadores microsatélites para la identificación de especies, sin embargo es recomendable realizar más estudios incrementando la cantidad de muestras.

10. REFERENCIAS

- Aggarwal, R. K., Hendre, P. S., Varshney, R. K., Bhat, P. R., Krishnakumar, V., & L. Singh. 2007. Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *Theoretical and Applied Genetics*, 114(2): 359-372.
- Alcántara, M. R. 2007. Breve revisión de los marcadores moleculares. *Ecología molecular*, 541-566.
- Alekcevetch, J. C., De Araújo, F., Da Silva, É. C., Guerra, A. F., Ferreira, G., Gava, M. A., & A. Carvalho. 2013. Estudo da diversidade genética de uma população de *Coffea canephora* var. Conilon por meio de marcadores moleculares do tipo SSR. VIII Simpósio de Pesquisa dos Cafés do Brasil
- Alvarado, M & G. Rojas. 1998. El Cultivo y Beneficiado del Café. San José, C.R. EUNED. 184 p.
- Anthony, F., Bertrand, B., Quirós, O., Wilches, A., Lashermes, P., Berthaud, J., & A. Charrier. 2001. Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. *Euphytica*, 118(1):53-65.
- Anthony, F., Astorga, C. & J. Berthaud, 1999. Los recursos genéticos: las bases de una solución genética a los problemas de la caficultura latinoamericana. In: B. Bertrand & B. Rapidel (Eds.), Desafíos de la caficultura centroamericana San José, Costa Rica IICA. 369–406.
- Anthony, F., Combes, M. C., Astorga, C., Bertrand, B., Graziosi, G., & P. Lashermes. 2002. The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *Theoretical and Applied Genetics*, 104(5): 894-900.
- Araya, E. 2010. Relaciones genéticas entre fuentes de germoplasma de palma aceitera americana (*Elaeis oleifera* Cortés) del programa de mejoramiento genético de ASD Costa Rica. Tesis MSc. Universidad de Costa Rica.

- Astorga, C. G. 1999. Caracterización de variedades cultivadas de café (*Coffea arabica* L.) conservadas en el banco de germoplasma del CATIE (Doctoral dissertation, Tesis Mag. Sc. Turrialba, CR, CATIE). 112 p.
- Azofeifa, A. 2006. Uso de marcadores moleculares en plantas; aplicaciones en frutales del trópico. *Agronomía mesoamericana*, 17(2):221-242.
- Baruah A, Naik V, Hendre PS, Rajkumar R, Rajendrakumar P & RK. Aggarwal. 2003. Isolation and characterization of nine microsatellite markers from *Coffea arabica* L., showing wide cross-species amplifications. *Molecular Ecology Notes* 3: 647–650.
- Berthou F & P. Trouslot 1977. L'analyse du polymorphisme enzymatique dans le genre *Coffea*: adaptation d'une méthode d'electrophorèse en série. Eighth Conference of ASIC, Abidjan (Ivory Coast), 373-383.
- Bigirimana, J., Njoroge, K., Muthomi, J. W., Gahakwa, D., Phiri, N., Gichuru, E. K., & D. J. Walyaro. 2013. Genetic diversity among disease resistant coffee varieties and cultivars in Rwanda based on RAPD and SSR markers. *Journal of Renewable Agriculture*. 1(6): 106-112
- Botstein D, White RL, Skolnick M & RW. Davis 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32:314–331.
- Bruvo R, Michiels NK, D'Souza TG & H. Schulenburg. 2004. A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Molecular Ecology*. 13:2101–2106
- Calle, F. 2012. Características de calidad e incidencia de enfermedades en 24 genotipos de *Coffea arabica* bajo tres condiciones agroecológicas de Panamá. Tesis Magister: Scientiae. Agricultura ecológica: Centro Agronómico Tropical de Investigación y Enseñanza. CATIE. 86p.
- Camargo, A. D., & M. D. Camargo. 2001. Definição e esquematização das fases fenológicas do cafeeiro arábica nas condições tropicais do Brasil. *Bragantia*, 60(1):65-68.

- Castro, P., Contreras, Y., Laca, D., & K. Nakamatsu. 2004. Café de especialidad: alternativa para el sector cafetalero peruano. The bi-annual academic publication of Universidad ESAN. 9(17):61-84.
- Clark. L. V. & M. Jasieniuk. 2011. POLYSAT: an R package for polyploid microsatellite analysis, *Molecular Ecology Resources*. (11): 562–566.
- Combes MC, Andrzejewski S, Anthony F, Bertrand B, Rovelli P, Graziosi G & P. Lashermes. 2000 Characterization of microsatellite loci in *Coffea arabica* and related coffee species. *Molecular Ecology* 9: 1178-1180.
- Cristancho, M. A., & Á. L. Gaitán. 2008. Isolation, characterization and amplification of simple sequence repeat loci in coffee. *Crop Breed Appl Biotechnol*, 8:321-329.
- Davis AP & EF. Mvungi. 2004. Two new and endangered species of *Coffea* (Rubiaceae) from the Eastern Arc Mountains (Tanzania) and notes on associated conservation issues. *Botanical Journal of the Linnean Society* 146: 237–245.
- Davis, P., Govaerts, R., Bridson, D.M. & P. Stoffelen. 2006. An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae), *Botanical Journal of the Linnean Society*, 152(4): 465–512.
- De Oliveira, I P., Oliveira, L C., & C S. De Mo. 2012. Cultura de café: histórico, classificação botânica e fases de crescimento. *Revista Eletrônica Faculdade Montes Belos*, 5(4):17-32
- Dessalegn, Y., Herselman, L., & M. Labuschagne. 2009. Comparison of SSR and AFLP analysis for genetic diversity assessment of Ethiopian arabica coffee genotypes. *South African Journal of Plant and Soil*, 26(2):119-125.
- Di Rienzo, J. A., Casanoves, F., Balzarini, M. G., González, L., Tablada, M., & Robledo, Y. C. 2017. InfoStat versión 2017e. Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina. URL <http://www.infostat.com.ar>.
- Ferrao, L. F., Caixeta, E. T., Pena, G., Zambolim, E. M., Cruz, C. D., Zambolim, L. & N. S. Sakiyama. 2015. New EST–SSR markers of *Coffea arabica*: transferability and

- application to studies of molecular characterization and genetic mapping. *Molecular Breeding*, 35(1): 1-5.
- Geleta, M., Herrera, I., Monzón, A., & T. Bryngelsson, 2012. Genetic diversity of Arabica coffee (*Coffea arabica* L.) in Nicaragua as estimated by simple sequence repeat markers. *The Scientific World Journal*. 11 pp.
- Hendre P. S., & R. K. Aggarwal. 2007. DNA markers: development and application for genetic improvement of coffee, In: Varshney RK, Tuberosa R (eds.), *Genomic Assisted Crop Improvement: Genomics Applications in Crops*. 2: 399–434.
- Hendre, P. S., & R. K. Aggarwal. 2014. Development of genic and genomic SSR markers of robusta coffee (*Coffea canephora* Pierre Ex A. Froehner). *PloS one*, 9(12), 34 pp.
- Hendre, P. S., Phanindranath, R., Annapurna, V., Lalremruata, A., & R. K. Aggarwal. 2008. Development of new genomic microsatellite markers from robusta coffee (*Coffea canephora* Pierre ex A. Froehner) showing broad cross-species transferability and utility in genetic studies. *BMC plant biology*, 8(1): 51.
- Huang Y F., Huang C C., Hu, C C. & H T., Chang. 2006. Capillary electrophoresis-based separation techniques for the analysis of proteins. *Electrophoresis* 27: 3503-3522.
- Hurtado, I., & J. Herrera- 2013. Estudio de la diversidad morfológica y molecular del germoplasma de *Coffea liberica* bull ex hiern en Colombia. *Cenicafé*, 64(2):31-47
- Instituto del café de Costa Rica (ICAFFE). Historia del café de Costa Rica en: <http://www.icafe.go.cr/> Recuperado el 7 de diciembre del 2015.
- Jain, S., Jain, R. K., & S. R. McCouch. 2004. Genetic analysis of Indian aromatic and quality rice (*Oryza sativa* L.) germplasm using panels of fluorescently-labeled microsatellite markers. *Theoretical and Applied Genetics*, 109(5):965-977.
- Lashermes, P., Combes, M. C., Robert, J., Trouslot, P., D'hont, A., Anthony, F., & A. Charrier. 1999. Molecular characterisation and origin of the *Coffea arabica* L. genome. *Molecular and General Genetics*, 261(2): 259-266.

- Lashermes, P., Trouslot, P., Anthony, F., Combes, M. C., & A. Charrier. 1996. Genetic diversity for RAPD markers between cultivated and wild accessions of *Coffea arabica*. *Euphytica*, 87(1):59-64.
- León, J. 2000. Botánica de los cultivos tropicales. San José, C. R. IICA. (3): 350-364
- López G., Cortina, H., McCouch, S. R., & M. D. Moncada. 2009. Analysis of genetic structure in a sample of coffee (*Coffea arabica* L.) using fluorescent SSR markers. *Tree genetics & genomes*, 5(3):435-446.
- López, G., & M. D. Moncada, 2005. Construcción de un mapa de ligamiento genético preliminar en *Coffea liberica* x *C eugenoides*. *Cenicafé*. 56(4):319-338.
- López, K. 2004. El mercado mundial del café tostado. Dirección de Inteligencia Comercial. Promotora del Comercio Exterior de Costa Rica (PROCOMER). URL <http://infocafes.com/portal/wp-content/uploads/2016/08/El-mercado-mundial-del-café-tostado.pdf>
- Maluf, M. P., Silvestrini, M., Ruggiero, L.M.D.C., Guerreiro Filho, O., & C. A. Colombo. 2005. Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. *Scientia Agricola*, 62(4): 366-373.
- Ministerio de Agricultura y Ganadería (MAG). Programa Nacional de Café: Información sobre la roya del cafeto en: <http://www.mag.go.cr/> Recuperado el 22 de abril del 2016.
- Missio, R. F., Caixeta, E. T., Zambolim, E. M., Pena, G. F., Zambolim, L., Dias, L. A. S., & N. S. Sakiyama. 2011. Genetic characterization of an elite coffee germplasm assessed by gSSR and EST-SSR markers. *Genetic and Molecular Research*, 10(4):2366-2381.
- Missio, R. F., Caixeta, E. T., Zambolim, E. M., Zambolim, L., Cruz, C. D., & N. S. Sakiyama. 2009. Polymorphic information content of SSR markers for *Coffea spp.* *Crop breeding and applied biotechnology*, 10:89-94.
- Moncada P & S. McCouch. 2004. Simple sequence repeat diversity in diploid and tetraploide *Coffea* species. *Genome* 47: 501–509.
- Montoya, G., Cristancho, M. A., & M. Moncada. 2006. Análisis de secuencias de genes de *Coffea arabica* var Caturra. *Cenicafé*, 57(2):79-87

- Morell, M.K., Peakall, R., Appels, R., Preston, L.R. & H.L. Lloyd. 1995. DNA profiling techniques for plant variety identification. *Australian Journal of Experimental Agriculture*. 35:807-819.
- Motta, L. B., Alvez, D. S., Sheu, Y., Demolinari, F., Passos, L., Gava, M. A., Gava, R., Almeida, A. F. & T. C. Bastos. 2015. Validacao de marcadores microssatélites em genótipos de café (*Coffea arabica* e *Coffea canephora*) .XV Encontro Latino Americano de Iniciação Científica e XI Encontro Latino Americano de Pós-Graduação – Universidade do Vale do Paraíba
- Nagy, S., Poczai, P., Cernák, I., Gorji, A. M., Hegedüs, G., & J. Taller. 2012. PICcalc: an online program to calculate polymorphic information content for molecular genetic studies. *Biochemical genetics*, 50: 670-672.
- Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences of the USA, 70:3321-3323.
- Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press. 512 p.
- Organización internacional del café (ICO). Estadísticas del comercio mundial en: <http://www.ico.org/> Recuperado el 7 de diciembre del 2015.
- Orozco, C. C., Chalmers, K. J., Waugh, R., & W. Powell. 1994. Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. *Theoretical and Applied Genetics*, 87(8):934-940.
- Parker P.G., Snow, A.A., Schug, M.D., Booton, G.C & P.A. Fuerst. 1998. What molecules can tell us about population: choosing and using a molecular marker. *Ecology*, 79:361-382.
- Pereira, T. B., Setotaw, T. A., Santos, D. N., Salgado, S. M. D. L., Rezende, R. M., Mendes, A. N. G., & G. R. Carvalho. 2015. Diversidade genética em progênies de cafeeiro (*Coffea arabica* L.) avaliada por marcadores microssatélites. IX Simpósio de Pesquisa dos Cafés do Brasil, Curitiba.
- Pérez, V. 1977. Venticinco años de investigación sistemática del cultivo del café en Costa Rica: 1950–1975. *Agronomía Costarricense*, 1(2): 169-185.

- Poncet, V., Hamon, P., Minier, J., Carasco, C., Hamon, S., & M. Noiro. 2004. SSR cross-amplification and variation within coffee trees (*Coffea spp.*). *Genome*, 47(6): 1071-1081.
- Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S. & A. Rafalski. 1996. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*. 2: 225–238
- Prakash, N.S., Combes, M.C., Somanna, N. & P. Lashermes. 2002. AFLP analysis of introgression in coffee cultivars (*Coffea arabica* L.) derived from a natural interspecific hybrid. *Euphytica* 124:265-271.
- Rischkowsky, B., & D. Pilling. 2010. La situación de los recursos zoogenéticos mundiales para la alimentación y la agricultura. FAO. 596 pp.
- Rojo, E. 2014. Café I (G. *Coffea*). Reduca (Biología). Serie Botánica. 7 (2): 113-132
- Romero, J.V.; Camayo, G.C.; González, L.F.; Cortina, H.A. & J.C. Herrera. 2010. Caracterización citogenética y morfológica de híbridos interespecíficos entre *C. arabica* y las especies diploides *C. liberica* y *C. eugenioides* Cenicafé 61 (3):206-221.
- Rovelli, P., Mettullo, R., Anthony, F., Anzueto, F., Lashermes, P., & G. Graziosi. 2000. Microsatellites in *Coffea Arabica* L. In Coffee biotechnology and quality. *Springer Netherlands*. pp. 123-133.
- Silvestrini, M., Junqueira, M. G., Favarin, A. C., Guerreiro-Filho, O., Maluf, M. P., Silvarolla, M. B., & C. A. Colombo. 2007. Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers. *Genetic Resources and Crop Evolution*, 54(6):1367-1379.
- Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*. 139:457-462
- Sousa, T. V., Caixeta, E. T., Alkimim, E. R., De Oliveira, A. C. B., Pereira, A. A., Zambolim, L., & N. S. Sakiyama. 2017. Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. *Euphytica*, 213(3):75.

- Steiger, D., Nagai, C., Moore, P., Morden, C., Osgood, R., & R. Ming. 2002. AFLP analysis of genetic diversity within and among *Coffea arabica* cultivars. *Theoretical and Applied Genetics*, 105(2):209-215.
- Suzuki, R., & H. Shimodaira. 2006. PvcLust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*. 22 (12):1540-1542.
- Tornincasa, P., Furlan, M., Pallavicini, A., & G. Graziosi. 2010. Coffee species and varietal identification. *Tools for Identifying Biodiversity: Progress and Problems*. pp. 307-313.
- Van Puyvelde, K., Van Geert, A., & L. Triest. 2010. ATETRA, a new software program to analyse tetraploid microsatellite data: comparison with TETRA and TETRASAT. *Molecular Ecology Resources*, 10(2):331-334.
- Vieira, E. S., Von Pinho, É. V., Carvalho, M. G., Esselink, D. G., & B. Vosman. 2010. Development of microsatellite markers for identifying Brazilian *Coffea arabica* varieties. *Genetics and molecular biology*, 33(3):507-514.
- Wang, X., Rinehart, T. A., Wadl, P. A., Spiers, J. M., Hadziabdic, D., Windham, M. T., & R. N. Trigiano. 2009. A new electrophoresis technique to separate microsatellite alleles. *African Journal of Biotechnology*, 8:11.
- Whitlock, M. C. 2011. GST and D do not replace FST. *Molecular Ecology*, 20(6):1083-1091.
- Wright, S. 1978. *Evolution and the Genetics of Population, Variability Within and Among Natural Populations*. The University of Chicago Press, Chicago.
- Zane, L., Bargelloni, L., & T. Patarnello. 2002. Strategies for microsatellite isolation: a review. *Molecular ecology*, 11(1):1-16.
- Zhivotovsky L.A. 1999. Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology*, 8:907-913.

11. ANEXOS

Anexo I Protocolo de extracción de ADN genómico

Materiales:

- ✓ Muestras a analizar frescas o liofilizadas
- ✓ Buffer de extracción, para 100 mL (1 M Tris/HCl (8,0); 5 M NaCl; 0,5 M EDTA (8,0))
- ✓ TE 1X (10 mM Tris/HCl (8,0) y 1 mM 0,5 M EDTA (8,0))
- ✓ Cloroformo:octanol (24:1)
- ✓ ARNasa (conc. 10 mg/mL)
- ✓ Isopropanol frío
- ✓ Etanol 70% v/v frío
- ✓ Tubos Eppendorf 1.5 y 2 mL
- ✓ Centrifuga
- ✓ Vortex
- ✓ Termobloque

Buffer de extracción para 100 mL:

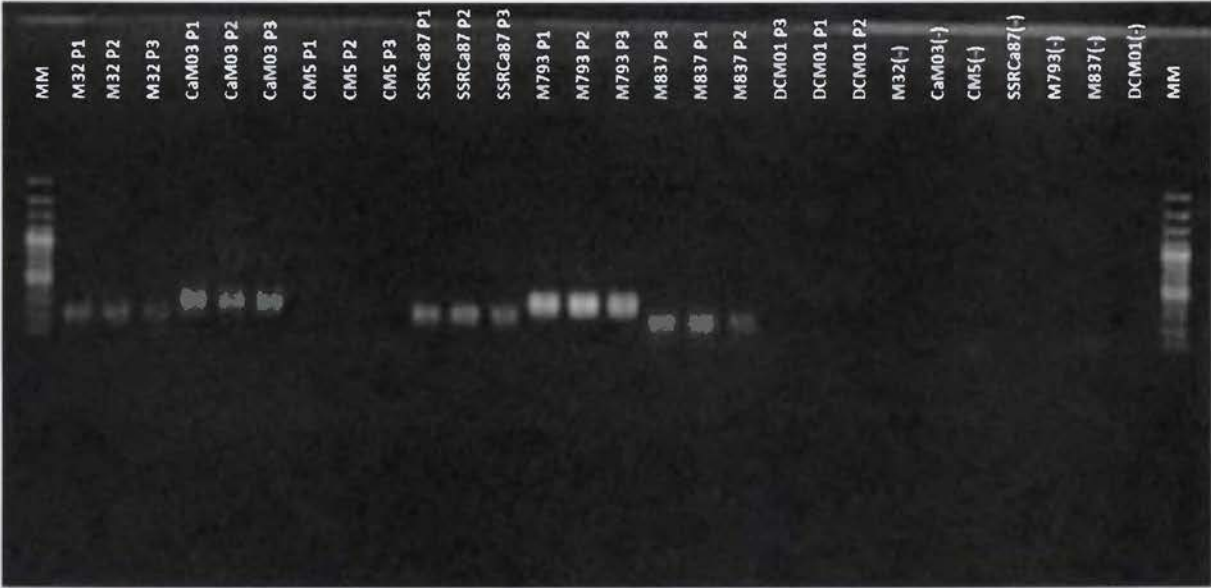
1 M Tris/HCl (8,0)	10,0 ml (= 100 mM)
5 M NaCl:	28,0 ml (= 1,4 M)
0,5 M EDTA (8,0):	4,0 ml (= 20 mM)
CTAB 2%	2 g
PVP 2%	2 g
H ₂ O:	58,0 ml

Agregar fresco por cada 20 mL de buffer de extracción: 40 uL mercaptoetanol (conc. final 0.2 %).

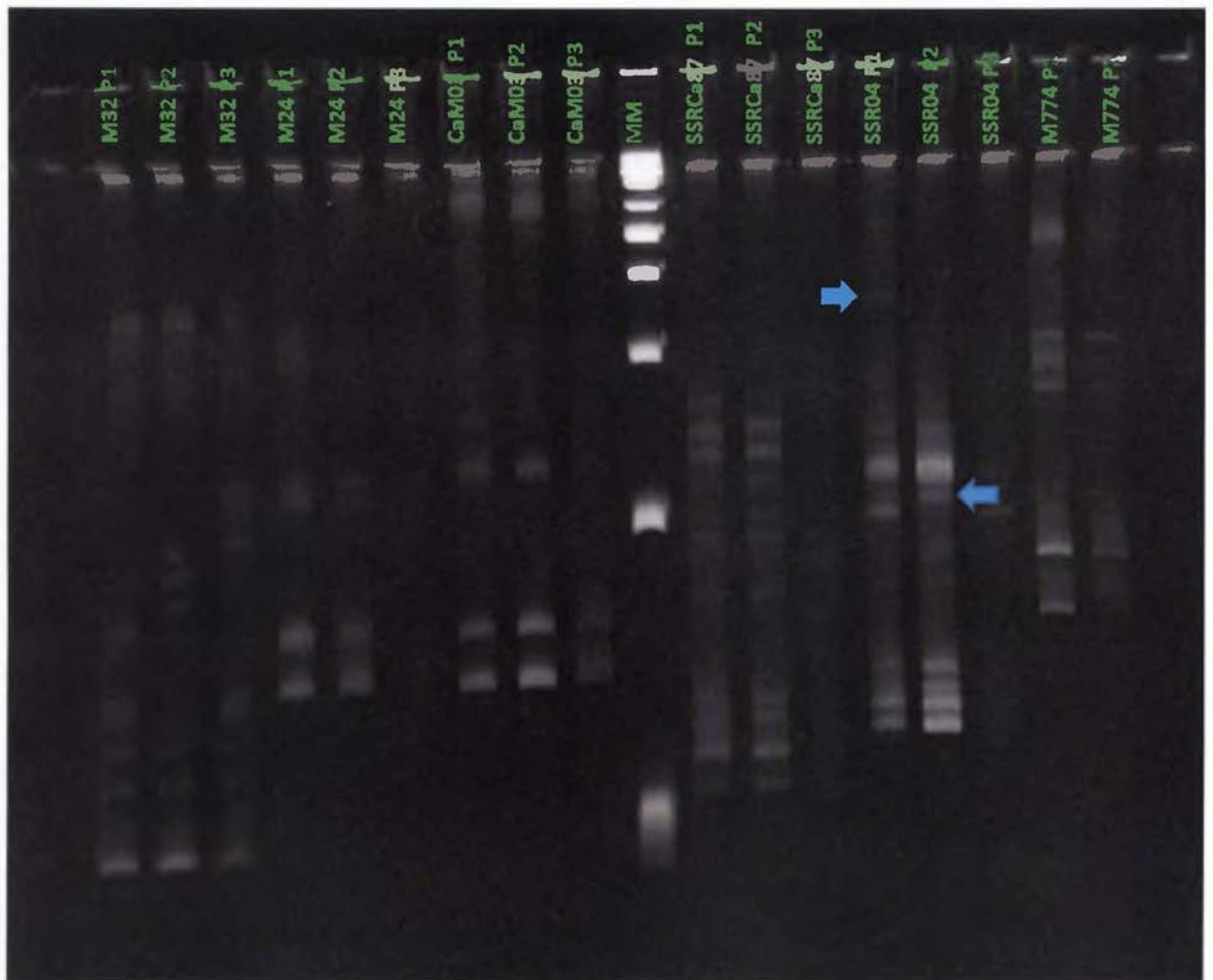
Procedimiento:

1. Colocar 40 mg del material vegetal en un tubo de 2 mL y macerar con un micropistilo
2. Agregar 750 μ L de buffer de extracción precalentado a 65 °C y mezclar con el vórtex.
3. Incubar la muestra a 65 °C por 20 min y agitar por inversión los tubos cada 10 minutos.
4. Agregar 750 μ L de cloroformo:octanol (24:1) y agitar 20 veces la muestra por inversión (suavemente). No utilizar vórtex.
5. Centrifugar la muestra por 5 min a 13000 rpm, a temperatura ambiente.
6. Transferir 300 μ L del sobrenadante a un nuevo tubo Eppendorf de 1,5 mL.
7. Agregar 300 μ L de isopropanol, frío a -20 °C y agitar 10 veces por inversión.
8. Incubar 5 min a temperatura ambiente y centrifugar la muestra por 5 min a 13000 rpm
9. Descartar con cuidado el sobrenadante y agregar 500 μ l de etanol 70% frío
10. Centrifugar la muestra por 2 min a 13000 rpm y descartar el etanol. Tener cuidado de NO perder el pellet.
11. Secar el pellet a 42 °C por 20-30 min
12. Resuspender el pellet en 50 μ l de TE 1X y agregar 1 μ L de ARNasa (conc. solución stock 10 mg/mL). Incubar por 30 min a 37 °C
13. Determinar la concentración del ADN

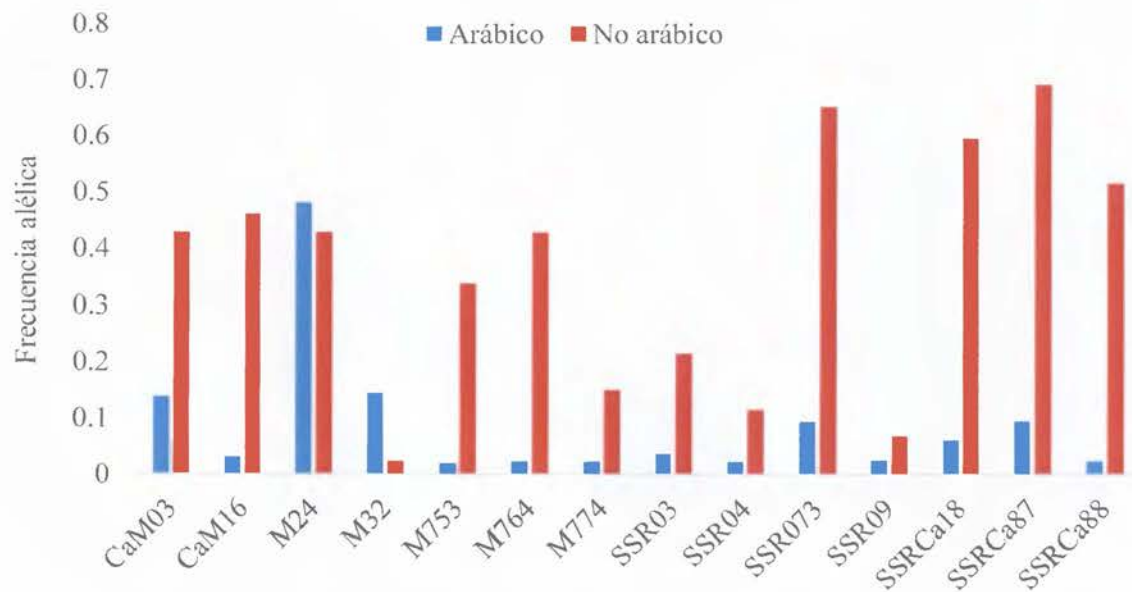
Anexo II Fragmentos amplificados de los 3 pooles genéticos con los marcadores M32, CaM03, CM5, SSRCa87, M793, M837 y DCM01 en gel de agarosa al 3% (m/v). (MM: marcador molecular (GeneRuler™ 100 pb Thermo Scientific™)), (-): control negativo (H₂O).



Anexo III. Fragmentos amplificados de los 3 pooles genéticos con los marcadores M32, M24, CaM03, SSRCa87, SSR04 y M774 en gel desnaturalizante de poliacrilamida al 10% (MM: marcador molecular GeneRuler™ 100 pb (Thermo Scientific™)). Las flechas indican algunas bandas polimórficas.



Anexo IV Frecuencia de alelos nulos en las muestras de café arábico y café no arábico.



Anexo V Número y frecuencia de alelos obtenidos en las muestras de arábicas y no arábicas con 14 marcadores microsatélites

Marcador		N _A	Distribución de alelos											
			1	2	3	4	5	6	7	8	9	10	11	12
CaM03	Arábicas	10	0,0048	0,4219	0,0524	0,0361	-	0,0148	0,0048	0,1906	0,0048	0,0097	-	0,1227
	No arábicas	4	-	-	-	-	0,2736	0,0998	-	-	-	0,0972	0,0998	-
CaM16	Arábicas	4	0,5407	-	0,3596	0,0473	0,0202	-	-	-	-	-	-	-
	No arábicas	4	0,1139	0,0970	0,2313	-	-	0,0970	-	-	-	-	-	-
M24	Arábicas	8	0,0302	-	-	0,0255	0,4048	0,0100	0,0052	0,0050	0,0050	0,0318	-	-
	No arábicas	4	-	0,0998	0,0972	0,2736	0,0998	-	-	-	-	-	-	-
M32	Arábicas	8	0,0164	0,0614	-	0,0052	0,0101	0,2202	0,0586	0,0052	0,4770	-	-	-
	No arábicas	2	0,4879	-	0,4879	-	-	-	-	-	-	-	-	-
M753	Arábicas	3	-	0,4672	-	0,0457	0,4672	-	-	-	-	-	-	-
	No arábicas	5	0,2529	0,1037	0,1002	0,1037	0,1002	-	-	-	-	-	-	-
M764	Arábicas	2	-	0,4879	0,4879	-	-	-	-	-	-	-	-	-
	No arábicas	4	0,0972	0,2736	0,0998	0,0998	-	-	-	-	-	-	-	-
M774	Arábicas	2	0,4879	-	-	0,4879	-	-	-	-	-	-	-	-
	No arábicas	2	-	0,7521	0,0972	-	-	-	-	-	-	-	-	-
SSR03	Arábicas	10	0,3255	0,3159	0,0051	0,0056	0,0056	0,2642	0,0056	0,0252	0,0049	0,0056	-	-
	No arábicas	6	0,0944	0,0944	-	0,1051	0,2918	0,0944	-	-	-	-	0,1051	-
SSR04	Arábicas	2	-	-	0,4879	0,4879	-	-	-	-	-	-	-	-
	No arábicas	2	0,1591	0,7256	-	-	-	-	-	-	-	-	-	-
SSR073	Arábicas	4	0,0109	0,0109	0,4416	0,4416	-	-	-	-	-	-	-	-
	No arábicas	2	-	0,1741	0,1741	-	-	-	-	-	-	-	-	-
SSR09	Arábicas	3	0,4988	0,0053	-	-	0,4694	-	-	-	-	-	-	-
	No arábicas	4	0,1018	-	0,1018	0,1018	0,6258	-	-	-	-	-	-	-
SSRCa18	Arábicas	4	0,3177	-	0,1153	-	0,4945	0,0102	-	-	-	-	-	-
	No arábicas	4	-	0,1062	-	0,1062	0,0955	-	0,0955	-	-	-	-	-
SSRCa87	Arábicas	4	-	0,1114	0,2586	-	0,5291	0,0051	-	-	-	-	-	-
	No arábicas	3	0,1029	0,1029	-	0,1029	-	-	-	-	-	-	-	-
SSRCa88	Arábicas	2	0,4879	0,4879	-	-	-	-	-	-	-	-	-	-
	No arábicas	3	0,2896	0,0972	0,0972	-	-	-	-	-	-	-	-	-

N_A: Número de alelos, Valores en negrita: Alelos raros (<0,01)

Anexo VI Errores estándar de AU (p-valores) para cada bootstrap generado en el dendrograma de distancias genéticas. El 19 pertenece al único bootstrap que obtuvo un error estándar mayor a 0,01.

