

**UNIVERSIDAD NACIONAL AGRARIA  
LA MOLINA**

**FACULTAD DE AGRONOMÍA**



**“ANÁLISIS DE LA DIVERSIDAD GENÉTICA DE CAFETOS DE  
LAS PRINCIPALES REGIONES PRODUCTORAS DE CAFÉ  
MEDIANTE EL USO DE MARCADORES MICROSATÉLITES”**

**TESIS PARA OPTAR EL TÍTULO DE  
INGENIERO AGRÓNOMO**

**RALPH RENEÉ RAMOS RETAMOZO**

**LIMA - PERU**

**2023**

# Tesis\_Agronomia Ralph\_Ramos

---

## INFORME DE ORIGINALIDAD

---

3%

INDICE DE SIMILITUD

3%

FUENTES DE INTERNET

1%

PUBLICACIONES

1%

TRABAJOS DEL  
ESTUDIANTE

---

## FUENTES PRIMARIAS

---

1

[hdl.handle.net](https://hdl.handle.net)

Fuente de Internet

2%

2

[repositorio.lamolina.edu.pe](https://repositorio.lamolina.edu.pe)

Fuente de Internet

1%

3

Submitted to Universidad Nacional Mayor de  
San Marcos

Trabajo del estudiante

1%

4

[pdffox.com](https://pdffox.com)

Fuente de Internet

1%

---

Excluir citas

Activo

Excluir coincidencias < 1%

Excluir bibliografía

Activo

**UNIVERSIDAD NACIONAL AGRARIA LA MOLINA**

**FACULTAD DE AGRONOMÍA**

**“ANÁLISIS DE LA DIVERSIDAD GENÉTICA DE CAFETOS DE LAS  
PRINCIPALES REGIONES PRODUCTORAS DE CAFÉ MEDIANTE  
EL USO DE MARCADORES MICROSATÉLITES”**

**RALPH RENEÉ RAMOS RETAMOZO**

Tesis para optar el Título de:

**INGENIERO AGRÓNOMO**

Sustentada y aprobada ante el siguiente jurado:

.....  
Ph. D. Julio C. Alegre Orihuela  
**PRESIDENTE**

.....  
Dr. Raúl Blas Sevillano  
**ASESOR**

.....  
DR. Alberto M. Julca Otiniano  
**MIEMBRO**

.....  
Dr. Jorge Eduardo Jiménez Dávalos  
**MIEMBRO**

.....  
Ph. D. Dina Lida Gutiérrez Reynoso  
**Co-Asesor**

Lima – Perú

2023

## **DEDICATORIA**

Para Ronald Renato Ramos Retamozo, pronto nos volveremos a encontrar hermano.

## **AGRADECIMIENTOS**

A Dios por permitirme realizar este trabajo de investigación.

A mis padres Oscar Ramos y María Retamozo por su amor, continuo apoyo durante toda mi vida y ser mis ejemplos a seguir.

A mis hermanos Ronald y Raiza por todo su apoyo y todo lo vivido.

A mi tía Bertha por su cariño y consejos desde pequeño.

A mi primo Víctor por ser como un hermano mas.

A Isamar por su apoyo incondicional durante la realización de la investigación.

A mi asesor de tesis el Dr. Raúl Blas por sus consejos y a mi co-asesora la Dra. Dina Gutiérrez.

A los miembros del jurado: Dr. Julio Alegre, Dr. Alberto Julca y Dr. Jorge Jiménez.

A los chicos del laboratorio de Biotecnología: Almendra, Andrea y Jossimar, a los asistentes de investigación que me guiaron: Cecilia, Savina Roger y Jose.

A los chicos del laboratorio IBT: Lucho, Joel, Maribel, Fabiola, Inés, Juan.

Agradezco a cada persona que con palabras de aliento apoyo me permitieron seguir realizando este trabajo.

A la UNALM por la formación y al INIA por el financiamiento de la investigación.

A los agricultores del Perú especialmente a los caficultores por tan valiosa labor.

# ÍNDICE GENERAL

I.	INTRODUCCIÓN .....	1
II.	REVISIÓN DE LITERATURA.....	3
2.1.	El café.....	3
2.1.1.	Aspectos generales.....	3
2.1.2.	Aspectos botánicos .....	6
2.1.3.	Variedades de café.....	7
2.1.4.	Diversidad Genética molecular.....	10
2.1.5.	Estimación de la Diversidad genética.....	15
III.	METODOLOGÍA .....	18
3.1.	Material vegetal.....	18
3.2.	Extracción de ADN de tejido foliar.....	19
3.2.1.	Protocolo de extracción .....	19
3.2.2.	Evaluación de calidad de ADN.....	20
3.2.3.	Cuantificación de ADN .....	21
3.2.4.	Dilución de muestras .....	22
3.3.	Amplificación por PCR.....	23
3.4.	Electroforesis capilar .....	26
3.5.	Generación de electroferogramas.....	26
3.6.	Análisis de diversidad .....	27
3.6.1.	Número de alelos .....	27
3.6.2.	Frecuencias alélicas de los SSR.....	27
3.6.3.	Heterocigosidad .....	27
3.6.4.	Índice de contenido Polimórfico (PIC).....	28
3.6.5.	Alelos Nulos .....	28
3.7.	Análisis de estructura poblacional.....	28
3.7.1.	Índices de Fijación.....	28
3.7.2.	Análisis de varianza molecular (AMOVA) .....	28
3.7.3.	Estructura Poblacional .....	29
3.7.4.	Análisis discriminante de componentes principales (DAPC).....	29
IV.	RESULTADOS Y DISCUSIÓN.....	31
4.1.	Material vegetal.....	31

4.2. Extracción de ADN .....	32
4.3. Amplificación por PCR.....	34
4.4. Análisis de marcadores SSR empleados .....	35
4.5. Diversidad Alélica.....	35
4.5.1. Numero de alelos .....	35
4.5.2. Frecuencias alélicas .....	39
4.5.3. Loci polimórficos.....	43
4.5.4. Alelos Nulos .....	45
4.6. Estructura genética .....	46
4.6.1. Heterocigosidad .....	46
4.6.2. Índice de Fijación - Fst .....	47
4.6.3. Análisis de conglomerados .....	48
4.6.4. Análisis de componentes principales - PCA.....	53
4.6.5. Análisis Discriminante de Componentes Principales - DAPC.....	56
4.6.6. Análisis de varianza molecular - AMOVA .....	57
4.6.7. Análisis de estructura poblacional .....	60
V. CONCLUSIONES .....	64
VI. RECOMENDACIONES .....	65
VII. BIBLIOGRAFÍA.....	66
VIII. ANEXOS.....	73

## ÍNDICE DE TABLAS

Tabla 1: Valores de absorbancia y concentración de 20 muestras de ADN, obtenidos mediante espectrofotometría.....	22
Tabla 2: Cantidades de ADN stock de café y NFW para diluciones a una concentración de 12 ng/μL en un volumen final de 200 uL .....	23
Tabla 3: Lista de Primers empleados con la cola M13 añadida en la secuencia forward ...	24
Tabla 4: Componentes y cantidades para PCR multiplex con tres pares de primers .....	25
Tabla 5: Condiciones de amplificación .....	25
Tabla 6: Rangos alélicos según literatura y rangos obtenidos en el presente estudio .....	35
Tabla 7: Número de alelos por locus y rango de tamaño en pares de base .....	37
Tabla 8: Número de genotipos y alelos por grupos .....	37
Tabla 9: Loci polimórficos en la población evaluados mediante la función isPoly del paquete estadístico Adegnet.....	43
Tabla 10: Valores de índice de contenido polimórfico (PIC) por loci, por región y promedios, hallados mediante el paquete estadístico Polysat y el programa PolyGene .....	43
Tabla 11: Valores de frecuencia de alelos nulos con los marcadores empleados .....	45
Tabla 12: Valores de Heterocigosidad observada y esperada. ....	46
Tabla 13: Valores de heterocigosidad sin marcadores con alta frecuencia de alelos nulos.....	47
Tabla 14: Tabla de índices Fst entre las regiones trabajadas.....	48
Tabla 15: Análisis de varianza molecular (AMOVA) de las regiones de origen de las accesiones de café, usando modelo IAM.....	58
Tabla 16: Análisis de varianza molecular (AMOVA) de las regiones de origen de las accesiones de café, usando modelo SMM .....	58
Tabla 17: Análisis de varianza molecular (AMOVA) de las denominaciones varietales de las accesiones de café. Usando modelo IAM .....	59
Tabla 18: Análisis de varianza molecular (AMOVA) de las denominaciones varietales de las accesiones de café. Usando modelo SMM .....	59

## ÍNDICE DE FIGURAS

Figura 1: Colecta de tejido foliar de café en la colección de café del INIA.....	18
Figura 2: Estado de hojas frescas al retornar a Lima.....	18
Figura 3: Extracción de ADN de café en el Laboratorio de Biología Molecular y Genómica – INIA .....	19
Figura 4: ADN extraído de 38 muestras de café, pertenecientes a la colección de café del INIA en gel de agarosa al 1% (m/v) (MP 1KB: Marcador de peso molecular GeneRuler™ 1kb (Thermo Scientific) .....	21
Figura 5: Agrupamiento de los primers según los productos esperados y sus respectivas temperaturas de <i>annealing</i> .....	25
Figura 6: Electroferograma con primer CaM16, con dos alelos de tamaño: 202 y 209 pb. 26	
Figura 7: Electroferograma con primer SSR03, con tres alelos de 152, 158 y 164 pb .....	27
Figura 8: Cantidad de individuos de la colección de germoplasma de café del INIA por región.....	31
Figura 9: Puntos de colecta de café en cada región, correspondiente a la data pasaporte de las accesiones trabajadas.....	32
Figura 10: Gel de agarosa al 1% con ADN genómico de café con presencia de “smear”, indicativo de degradación.....	33
Figura 11: Gel de agarosa al 1% con ADN de café diluido a 10 ng/μl.....	34
Figura 12: Gel de agarosa al 1.5% para la evaluación de amplificación de los productos de PCR.....	34
Figura 13: Electroferograma generado con el marcador SSR03 mostrando dos alelos de distinto tamaño .....	36
Figura 14: Electroferograma generado con el marcador M20.....	36
Figura 15: Gráfico representado el número de alelos y la cantidad de individuos por región.....	39
Figura 16: Estimación del PIC promedio en población total incluido el outgroup .....	44
Figura 17: Dendrograma de regiones productoras de café evaluadas .....	49
Figura 18: Dendrograma de variedades con distancia de Nei y agrupamiento UPGMA....	49
Figura 19: Dendrograma a nivel de individuos, con distancia euclidiana y agrupamiento UPGMA.....	52

Figura 20: Análisis de componentes principales en base a las frecuencias alélicas de los individuos, mostrando los dos primeros componentes .....	54
Figura 21: Gráfico de sedimentación mostrando el aporte de cada componente .....	55
Figura 22: Análisis discriminante de componentes principales DAPC con 13 loci .....	56
Figura 23: Análisis discriminante de componentes principales DAPC retirando cinco loci con frecuencia de alelos nulos elevada .....	57
Figura 24: Curva de Delta de K para la determinación del óptimo valor de K en la población estudiada .....	60
Figura 25: Curva de Delta de K determinando el óptimo valor de K en la población estudiada, correspondiente a $K=5$ .....	61
B Figura 26: Estructura poblacional de las muestras de café evaluadas. A (3K), B (5K)...	62

## ÍNDICE DE ANEXOS

Anexo 1: Graficas de frecuencias alélicas por cada marcador en la población estudiada...	73
Anexo 2: Cantidad de componentes y concentraciones para reacciones de PCR multiplex .....	78
Anexo 3: Gráficas de análisis de componentes principales PCA .....	79
Anexo 4: Gráficas correspondientes al Análisis de coordenadas principales – PcoA.....	81
Anexo 5: Tablas de análisis discriminante de componentes principales – DAPC .....	83
Anexo 6: Gráficos de análisis de estructura genética con enfoque bayesiano .....	85
Anexo 7: Datos de colecta de las muestras de la colección de café del INIA.....	95

## RESUMEN

El café es fundamental para muchas familias agricultoras del Perú y sirve como principal fuente de ingresos y desarrollo, en los últimos años ha ido obteniendo mayor reconocimiento a nivel internacional; sin embargo, existen algunos problemas por atender como baja rentabilidad, problemas fitosanitarios, entre otros. Estos problemas pueden ser mitigados con el uso del material vegetal adecuado. Este estudio buscó evaluar la diversidad y estructura genética del café proveniente de las principales regiones productoras como Cajamarca, Amazonas, Junín, Pasco, Ucayali y Huánuco. Se trabajó con 169 muestras de la colección de germoplasma de café del INIA y con marcadores moleculares tipo SSR marcados con cola M13, además se emplearon tres muestras de ADN de *Coffea liberica*, *Coffea canephora* y el híbrido S288 como outgroup. Se identificó un total de 70 alelos, con una media de 5.3 alelos por locus y rangos desde 117 hasta 488 pb. Cuatro de los marcadores empleados reportaron un PIC alto, demostrando su utilidad para este tipo de estudios. El promedio de la heterocigosidad observada fue 0.537 y la esperada fue 0.517. Los valores de  $F_{st}$  obtenidos, indicaron un bajo grado de diferenciación genética entre las regiones evaluadas, el mayor valor de  $F_{st}$  fue de 0.151 correspondiente a la región Ucayali y el outgroup. El análisis AMOVA permitió determinar una alta variación genética dentro de los grupos evaluados, regiones y variedades, y una baja variación entre estas. En el análisis de estructura genética con enfoque bayesiano, se encontró un  $K=5$  empleando todos los loci, pero al retirar los loci con una frecuencia elevada de alelos nulos se obtuvo un  $K=3$ , ambos resultados no muestran una asociación correspondiente a las regiones de origen de las accesiones ni a las denominaciones varietales; sin embargo, permitió determinar una marcada diferenciación entre las muestras de arábica y no arábica.

**Palabras clave:** *Coffea arabica*, marcador, microsatélite, diversidad, población, variedades.

## ABSTRACT

Coffee is essential crop for many farmers in Peru and acts as main source of income and development, over the last years coffee has been gaining more recognition worldwide. However, there are still some remaining problems that need attention, such as low revenue, phytosanitary issues, among others. Those problems can be mitigated by the use of proper germplasm. The genetic diversity and genetic structure of coffee from some of the main peruvian coffee producer's regions (Cajamarca, Amazonas, Junín, Pasco, Ucayali y Huánuco) was evaluated, 169 individuals from the INIA's collection were amplified via PCR with a set of 15 pairs of microsatellite markers labelled with a M13-tag and three DNA samples from *C. canephora*, *C. liberica* and the hybrid S.288 (*C. arabica* x *C. liberica*) were used as outgroup. A total of 70 different alleles were found, with an average of 5.3 alleles per locus and ranges from 117 to 488 bp. M32, SSR03, SSRCa18 and M20 markers showed PIC values above 0.5 showing their utility for his kind of studies, the average observed heterozygosity was 0.537 and the average expected heterozygosity was 0.517, the  $F_{st}$  indexes showed low degree of genetic differentiation among the evaluated regions, the highest  $F_{st}$  value was 0.151 obtained between Ucayali region and the outgroup. The analysis of molecular variance (AMOVA) showed a high variation within groups formed by regions and varieties, and low variation between those groups. In the Bayesian analysis of genetic structure, a value of  $K=5$  was found using all the loci, but, when retiring from the analysis loci with high null allele frequencies, the value of  $K$  reduced to  $K=3$ , both results do not show association between the origin regions of the samples nor to varietal denominations, however, it showed separation between Arabica and non-Arabica individuals.

**Keywords:** *Coffea arabica*, microsatellite marker, diversity, population, varieties.

## I. INTRODUCCIÓN

El café es uno de los cultivos comerciales más relevantes a nivel mundial, no solo por los volúmenes de venta, sino también por la gran cantidad de personas involucradas que dependen de este cultivo alrededor del mundo (Syafaruddin *et al.* 2014). A nivel global los principales países productores de café son: Brasil, Vietnam, Colombia y Etiopía (AMECAFE, 2017).

Perú en el año 2017 ocupó el octavo lugar a nivel mundial como país exportador de este producto, con 3.946 millones de sacos de 60 kg (ICO, 2018). A nivel nacional, el café se ubica en los primeros lugares de las exportaciones agrícolas y también dentro de los 10 primeros productos de exportación después de algunos minerales, petróleo, gas natural, harina de pescado, entre otros. Además, de acuerdo con el Plan Nacional de acción del café (MINAGRI, 2017) el Perú se encuentra en el segundo lugar como productor y exportador de café orgánico a nivel mundial.

La calidad del café peruano cuenta con buen reconocimiento y gran aceptación a nivel internacional, se estima que en el Perú el cultivo de café es el sustento económico de más de 223 mil familias involucradas (MINAGRI, 2017) y se reportan ingresos de exportaciones de \$ 703,408.3 (Dólares americanos) como valor FOB (*Free on Board*) para el año 2017, el precio FOB se refiere al valor de la mercancía cuando esta se encuentra en el puerto de embarque (Cámara Peruana de café y cacao, 2019). Si bien el café es un cultivo de gran importancia social y económica, el escenario nacional actual presenta baja rentabilidad; problemas fitosanitarios sin atención; aumento de escasez hídrica; carencia de un adecuado manejo agronómico (JNC, 2016). Sumado a estas problemáticas se encuentra la denominación incorrecta de variedades y/o mezcla de éstas, causada posiblemente por el ingreso informal de germoplasma; transferencia de material de diferentes zonas cafetaleras y en menor medida por segregación y polinización cruzada, posibilitada por el uso de diferentes variedades en áreas pequeñas (Blas *et al.*, 2011).

Ante la gran importancia del café en el Perú, es necesario determinar la diversidad genética presente en los cafetos provenientes de las principales regiones productoras del país a través del empleo de marcadores moleculares y con la información generada servir de base para poder hacer uso del material vegetal en programas de mejoramiento que puedan mitigar los efectos que causan principalmente los problemas fitosanitarios y de estrés abiótico. Con el uso de material genético con buenas características agronómicas y el manejo adecuado, se podría contribuir a garantizar la sostenibilidad económica en el sector cafetalero peruano.

En el presente trabajo de investigación se emplearon muestras provenientes de algunas de las principales regiones cafetaleras en Perú como: Cajamarca (21.04%), Amazonas (12.94%), Junín (18.77%), Pasco (6.58%), Ucayali (3.47%) y Huánuco (3.40%), estas regiones en conjunto representaron 66.21% del total de la producción de café pergamino para el año 2021 (JNC, 2022).

Para el fin del estudio, se requirió obtener ADN genómico de café de buena calidad para generar un proceso de amplificación *multiplex* con marcadores moleculares tipo SSRs (*Simple Sequence Repeats*) basados en secuencias simples repetidas, que permitan reducir tiempo y costos en contra parte a la amplificación individual.

El uso de marcadores genéticos basados en polimorfismos de ADN (ácido desoxirribonucleico) son una alternativa para la determinación de la diversidad genética debido a que estos marcadores no se ven influenciados por el ambiente (Blas, 2010; Sánchez, 2017). Sin embargo, el empleo complementario de marcadores morfológicos, a pesar de verse influenciados por el ambiente, enriquecen en gran medida la data generada en la búsqueda de caracterizar y obtener la mayor información del material vegetal para su potencial uso.

El objetivo del presente estudio consistió en analizar la diversidad y estructura genética del cultivo de café de algunas de las principales regiones productoras del país, mediante el uso de marcadores moleculares tipo SSRs o microsatélites.

## II. REVISIÓN DE LITERATURA

### 2.1. El café

#### 2.1.1. Aspectos generales

El café es uno de los productos agrícolas más valiosos, generador de importantes fuentes de divisas para muchos países y se comercia en los principales mercados de materias primas y de futuros en donde se realizan contratos de compra o venta en una fecha futura con un precio ya acordado, como los de Londres y Nueva York (International Coffee Organization, s.f.). El café comercial se origina principalmente de *Coffea arabica*, en menor medida de *C. canephora* y en cantidades muy pequeñas de *C. liberica* (León, 2000). *C. arabica* pertenece a la familia Rubiácea y debido a que resalta ampliamente desde el punto de vista mercantil, es la especie más importante de esta familia a nivel económico.

Gran parte de las especies de café son diploides, pero *C. arabica* (arábica) es tetraploide y posee  $4 \times 11 = 44$  cromosomas, el doble que *C. canephora* (robusta). Las plantas de arábica son autógamas, aunque hay la posibilidad de alogamia en un porcentaje mínimo, estas plantas se propagan por semillas usualmente en almácigos. Mientras que las plantas de robusta son alógamas es decir son autoestériles y requieren polinización cruzada para el desarrollo de semilla, debido a esto dichas plantas usualmente se propagan por esquejes. A diferencia de las plantas de arábica, las robustas se desarrollan mejor a menores altitudes, resiste elevadas temperaturas y precipitaciones. Además, son más resistentes a plagas y enfermedades, de ahí viene el nombre común de “robusta” (Clarke, 2003).

#### a. Origen del Café

El café tiene como lugar de origen a Etiopía, donde aún es posible encontrarlo en estado silvestre. No se conoce la fecha exacta del inicio del cultivo, pero se presume que podría haberse dado a comienzos del siglo VIII. El cultivo se mantuvo como exclusivo para los árabes cercanos al Mar Rojo y posteriormente se expandió a la actual Yemen (Anthony *et al.*, 1999).

El descubrimiento de las propiedades estimulantes de las semillas de café puede deberse al antiguo hábito de masticarlas, para posteriormente conllevar a la preparación de bebidas y luego la extensión del uso moderno del café como estimulante, gracias al contenido de cafeína en sus semillas (León, 1987).

León (1987) menciona que la expansión del café hacia Arabia desde Etiopía se dio durante la edad media y es a fines del siglo XVIII que se introduce a Java de Yemen. Mientras que Chevalier y Dagrón (1928), citado por Anthony *et al.* (1999) indican que la introducción del café al continente americano se inició en el siglo XVIII, con plantas enviadas desde los jardines botánicos de Ámsterdam y París hacia lo que hoy es Surinam y Las Antillas. Rápidamente fue migrando hacia la Guyana Francesa y la Guyana Holandesa, para finalmente llegar a Brasil en 1727.

El café llegó a la selva central del Perú traído por migrantes europeos, especialmente italianos, esto debido a la colonización. También comenzaron a cultivar otras especies tropicales como la caña de azúcar, el cacao y la coca (Julca *et al.*, 2010). Asimismo, se menciona que el café fue introducido en el año 1760 desde Guayaquil hacia Huánuco (JNC, 2016)

#### **b. Producción de café en el Perú**

En el año 2019 se reportaron 370,000 hectáreas de café de las 425,000 registradas en el censo Nacional Agropecuario del 2012, distribuidas en 210 distritos rurales ubicados en 10 departamentos. Junín es la región cafetalera tradicional, pero actualmente Amazonas y San Martín conforman el 42% de la producción nacional mientras que Junín representa un 18.8% (JNC, 2021). En el Perú, se produce casi exclusivamente café arábica y la variedad Típica correspondería al 70% de la producción, seguida por Caturra con 20% y un 10% correspondiente a otras variedades. La densidad de siembra promedio corresponde a unas 2,000 plantas por hectárea y el 70% de las áreas de cultivo se sitúan a una altura comprendida entre los 1,000 y 1,800 m.s.n.m. El rendimiento promedio de café en el Perú en campaña comercial 2019/2020 se estimó en unos 752 kilogramos por hectárea, aproximadamente 16 quintales (JNC, 2020).

La producción de café en el Perú es manual, es decir, no tecnificada y la cosecha se realiza de forma selectiva o escalonada conforme a la maduración del fruto, por lo cual solo cerezos maduros serán cosechados, luego como parte del proceso poscosecha se puede optar por un beneficio “natural” o seco donde se procede a secar los cerezos para su posterior trillado o remoción de la pulpa, mientras que en el beneficio húmedo se realiza un despulpado inmediatamente después de la cosecha, luego pasa a un proceso de fermentación donde se produce la descomposición del mucilago que cubre el pergamino del café, luego se lava para separar el mucilago y finalmente se realiza el secado, el beneficio húmedo es el más empleado en el Perú. Para el proceso de secado, algunos caficultores optan por secarlo directamente al sol en pisos de cemento o lozas cubiertas de plástico, otros hacen uso de secadores solares. Finalmente, existe otro método de secado de manera artificial en secadores de café, empleando temperaturas que no sobrepasen los 50°C para evitar la cristalización de los granos, la humedad del grano que se busca con el proceso de secado se encuentra entre 10% y 12% (UNALM, 2013).

**c. Clasificación Taxonómica APG IV (The Angiosperm Phylogeny Group, 2016)**

Clado: angiospermas (Angiospermae)

Clado: mesangiospermas (Mesangiospermae)

Clado: eudicotiledóneas (Eudicotyledonae)

Clado: gunnéridas (Gunneridae)

Clado: pentapétalas (Pentapetalae)

Clado: superastéridas

Clado: astéridas

Clado: euastéridas (geniácidas)

Clado: lámidas

Orden: Gentianales

Familia: Rubiaceae

Género: *Coffea*

Especie: *Coffea arabica* L.

## 2.1.2. Aspectos botánicos

### a. Familia Rubiaceae

La familia Rubiácea incluye individuos de distintos portes, comprendiendo desde herbáceas hasta arbóreas. Si bien el café es el cultivo más importante de esta familia, también podemos encontrar a la quina, especie utilizada como planta medicinal. Los individuos de esta familia poseen cáliz y corola regulares con piezas unidas, los pétalos forman una estrella, la parte inferior de la corola forma un tubo y de ella salen los estambres. El ovario generalmente es bilocular, aunque pueden presentar más lóculos. Entre los frutos podemos encontrar bayas, cápsulas y drupas (León, 2000).

### b. *Coffea arabica*

El porte de *C. arabica* se caracteriza por presentar dimorfismo de ejes, de ahí viene el característico eje vertical u ortotrópico de donde se desprenden los ejes laterales o plagiotrópicos, este dimorfismo se puede encontrar en las especies del género *Coffea*, las ramas ortotrópicas muy raramente forman inflorescencias, mientras que las ramas plagiotrópicas son las que llevan las yemas florales. Debido a este dimorfismo en las ramas, la planta al crecer toma una forma piramidal o cónica, este porte no correspondería al de un arbusto, pero generalmente se le denomina así. No es común encontrar plantas con su porte natural, ya que estas son podadas como medidas de manejo agronómico del cultivo, facilitando así también la recolección de semillas (León, 2000).

Las hojas tienen disposición opuesta en las ramas plagiotrópicas y la lámina levemente coriácea y los márgenes un poco ondulados, poseen forma variada de elíptica a lanceolada, con tamaños que van desde 12 a 24 cm de largo por cinco a 12 cm de ancho. En el envés de las hojas, en el ángulo formado por la nervadura central y las secundarias, se forma una pequeña cavidad con entradas de forma irregular denominadas domacias, se desconoce la función de estas cavidades y en ellas se alojan ácaros muy pequeños (León, 1987).

La formación de las flores se da en las yemas localizadas en las axilas foliares, en los nudos de las ramas. Cada nudo de una rama posee dos axilas foliares opuestas en las cuales se forman de tres a cuatro yemas y de cuatro a cinco flores en cada una (Arcila, 2007). Las

flores se abren en las primeras horas de la mañana, las anteras desprenden polen antes de la antesis y como en *C. arabica* no hay mecanismos conocidos de autoesterilidad, se asume que la autofecundación ocurre normalmente (León, 1987).

El fruto del café es una drupa de forma esférica la cual está compuesta de dos semillas planoconvexas, en el caso una aborta se forma el grano de café “caracolillo” el cual pierde su lado plano y pasa a ser redondo (Nosti, 1963; citado en Vargas, 2017). El epicarpio y mesocarpio conforman lo que se conoce como “pulpa”, y el endocarpo es el “pergamino” que una vez alcanzada la maduración se separa y recubre la semilla; también podemos encontrar una película plateada delgada (tegumento seminal) que recubre el grano, la cual se desprende al momento del tostado. La epidermis del fruto toma coloraciones verdosas, amarillas, rojizas e incluso violáceas o negras dependiendo de las variedades y el grado de madurez (León, 1987; Arcila, 2007).

### **2.1.3. Variedades de café**

Dos de las principales variedades de *C. arabica* son: a) Típica, esta variedad fue cultivada en jardines botánicos holandeses y de allí fue llevada por franceses al Caribe, posteriormente a países centroamericanos y sudamericanos. b) Bourbon, esta variedad fue cultivada y mantenida en las islas Bourbon o Reunión (Puerta, 1988). Ambas variedades han dado origen a muchas variedades comerciales, ya sea por mutación, cruzamientos espontáneos o inducidos.

- Típica: Considerada como una de las bases del desarrollo de la caficultura en América central, esta variedad posee forma cónica y un porte alto pudiendo llegar hasta los cuatro metros. Tiene un solo tronco vertical y ramas laterales que forman un ángulo de 50 a 70 grados. Su fruto es alargado, grande y de maduración tardía. Es considerada como una variedad de bajo rendimiento, tiene un comportamiento bienal y es susceptible a la roya. A pesar de su bajo rendimiento es una variedad muy considerada por su alta calidad de taza (ANACAFE, 2016).
- Bourbon: Posee una forma cónica menos acentuada que la variedad Típica y presenta un rendimiento superior en 20% a 30% comparado con esta, sus ramas primarias forman un ángulo de 45 grados con el eje principal. Es una variedad precoz en maduración por lo cual las cerezas tienen riesgo de caerse frente a fuertes vientos o

lluvias, tiene un porte alto llegando a los tres metros y es susceptible a roya. Posee una excelente calidad de taza (ANACAFE, 2016).

- Caturra: Es una mutación natural de la variedad Bourbon descubierta en Brasil, posee un porte bajo de aproximadamente 1.80 metros, las ramas jóvenes forman un ángulo de 45 grados con el eje principal y posee entrenudos cortos, tiene múltiples ramas secundarias dándole una apariencia compacta. Sus frutos tienen una maduración precoz, es demandante en manejo y fertilización. Posee una excelente calidad de taza (ANACAFE, 2016).
- Maragogipe: Mutación natural de típica descubierta en Brasil, posee un porte más alto que el de Típica y Bourbon, sus ramas laterales forman un ángulo de 75 grados con el eje central. Tiene bajo rendimiento, pero una muy buena calidad de taza. Presenta susceptibilidad a roya. Se adapta mejor a zonas comprendidas entre los 1,300 a 1,800 m.s.n.m (ANACAFE, 2016).
- Catimor: Es el nombre que se les da a las variedades provenientes del cruce de Caturra e Híbrido de Timor CIFC 832/1. En general son variedades muy productivas y demandantes en el manejo, principalmente en la fertilización. Poseen resistencia a la roya, conferida gracias al Híbrido del Timor. Algunas variedades como CR95 o Lempira han reportado pérdida de la resistencia a la roya en los países donde han sido generados (Costa Rica y Honduras). Los catimores muestran susceptibilidad a la enfermedad denominada como ojo de gallo, la cual tiene como agente causal a *Mycena citricolor* (ANACAFE, 2016).
- Catuaí: Proviene del cruce artificial realizado en Brasil entre la variedad Mundo novo y Caturra. Tiene un porte bajo con un promedio de 2.25 metros, sus ramas laterales forman un ángulo de 45 grados con el eje principal. Es muy vigorosa y sus frutos poseen maduración tardía, pero estos no se desprenden con facilidad lo cual es ventajoso en zonas lluviosas. Posee excelente calidad de taza, alto rendimiento y es susceptible a la roya (WCR, s.f.).

- **Mundo Novo:** Originada en Brasil por un cruce natural entre la variedad Sumatra (Típica) y la variedad Bourbon. Posee un porte alto de hasta tres metros, sus ramas laterales forman un ángulo de 45 grados con el eje principal. Es muy vigorosa y sus frutos poseen maduración tardía, los frutos son medianos con coloración roja o amarilla. Tiene buen comportamiento en alturas entre 1,070 a 1,680 m.s.n.m. Tiene una excelente calidad de taza y es susceptible a la roya (ANACAFE, 2016).
- **Pache:** Es una mutación natural de Típica, que le confiere enanismo. Tiene buena calidad de taza y un rendimiento medio. Se adapta bien a alturas mayores a los 1,200 m.s.n.m., es susceptible a la roya (WCR, s.f.). Existe una variedad denominada Pache Colís, la cual se presume proviene del cruce entre las variedades Pache y Caturra, posee una altura menor que Pache, presenta cierta tolerancia a la enfermedad denominada como Phoma, pero es susceptible a roya y al ojo de gallo (ANACAFE, 2016).
- **Pacamara:** Proviene del cruce de Pacas y Maragogipe, realizado por el Instituto Salvadoreño de Investigación del Café (ISIC). Tiene un porte bajo como el de Pacas, entrenudos cortos y buen rendimiento. Posee excelente calidad de taza y tiene potencial de muy buen rendimiento, es susceptible a roya. También posee frutos grandes y alargados similares a los de Maragogipe. Se adapta muy bien en alturas entre 1,000 a 1,500 m.s.n.m (WCR, s.f.).
- **Colombia:** Es una variedad de café resistente a la roya, producida por el Cenicafe en 1980 proveniente del cruce de Caturra con el Híbrido de Timor, lo cual lo hace pertenecer al grupo de los catimores, tiene mucha similitud en tamaño y forma con la variedad Caturra. Es una variedad compuesta o multilínea seleccionada en base a su resistencia a roya, buena adaptación, alto rendimiento y excelente calidad de taza. Posee resistencia vertical y horizontal frente a la roya, por lo cual esta es más duradera en el tiempo (Cenicafe, 2000).
- **Castillo:** Fue liberada en el año 2005 por la Federación Nacional de cafeteros de Colombia, es de porte bajo, posee alto rendimiento y excelente calidad de taza. Para

su generación se empleó la misma estrategia que la utilizada para dar origen a la variedad Colombia, por lo cual posee una resistencia a roya más duradera. También se generaron variedades regionales, seleccionadas según su adaptación y comportamiento sobresaliente en regiones específicas, por lo que también se pueden encontrar siete variedades regionales: Castillo Naranjal, Castillo Paraguaicito, Castillo Santa Bárbara, Castillo Pueblo Bello, Castillo El Rosario, Castillo La Trinidad y Castillo El Tambo (ANACAFE, 2016).

- Villa Sarchí: Originada a partir de una mutación natural de Bourbon, fue seleccionada a través de generaciones sucesivas por el Instituto de café de Costa Rica (ICAFFE). Es de bajo porte, posee buena calidad de taza y rendimiento. Es susceptible a la roya (WCR, s.f.).
- Costa Rica 95: Es un catimor proveniente del cruce de Híbrido de Timor 832/1 y Caturra, posee bajo porte, alta productividad y buena adaptabilidad en alturas entre 800 a 1,400 m.s.n.m. Actualmente se ha reportado la pérdida de resistencia a la roya en Costa Rica, así como el caso de pérdida de resistencia de la variedad Lempira en Honduras (WCR, s.f.).

#### **2.1.4. Diversidad Genética molecular**

La diversidad genética está considerada como el componente más básico de la biodiversidad y está conformada por las variaciones genéticas; que son heredadas y que suceden en cada organismo; entre los individuos que conforman una población y entre las poblaciones dentro de una especie en condiciones más o menos estables. Además, la diversidad genética está influenciada por factores ambientales que son responsables de que los individuos de una especie difieran en características fenotípicas (Rimieri, 2017; Sevilla y Holle, 2004).

La información sobre la diversidad genética de los cultivos es fundamental y hay distintas maneras de evaluarlo. A lo largo del tiempo se han usado como descriptores a los rasgos morfológicos con el fin de cuantificar la variabilidad genética. Si bien es una herramienta muy útil para la selección de genotipos para mejoramiento, pueden mostrar ciertas limitantes debido a su número y a que pueden ser modificadas por factores ambientales y ser

controladas por genes con efectos pleiotropicos o epistaticos (Jiménez, 2006; como se citó en Allende, 2017).

#### **a. Marcadores Moleculares**

Los marcadores moleculares son secciones del ADN que se pueden localizar e identificar con precisión. Se puede analizar de ADN proveniente del núcleo, del cloroplasto o de las mitocondrias. Para que estos marcadores sean de utilidad, deben encontrarse distribuidos en todos los cromosomas. Los marcadores moleculares se fundamentan en dos técnicas básicas: las enzimas de restricción y la reacción en cadena de la polimerasa (PCR). (Sevilla y Holle, 2004).

Estos marcadores posibilitan la identificación o diferenciación de genotipos, tienen una locación cromosómica fija y pueden presentar diversas variantes o alelos que determinan su elevado polimorfismo. Son una herramienta ampliamente utilizada en estudios filogenéticos, variabilidad genética de poblaciones, mapas genómicos, entre otros (Paredes, 2013). Los marcadores moleculares tipo RFLPs (*Restriction Fragment Length Polymorphisms*), RAPDs (*Random Amplified Polymorphic DNA*) y AFLPs (*Amplified fragment length polymorphisms*), no necesitan de conocimiento previo del genoma a analizar, mientras que los SNPs (*Single Nucleotide Polymorphisms*) y SSRs (*Simple Sequence Repeats*) si lo requieren.

El uso eficiente de los recursos fitogenéticos está enteramente ligados a su adecuada caracterización. Históricamente estas se han realizado utilizando descriptores fenotípicos, principalmente de tipo morfológico o agronómico. Esto ha permitido diferenciar muchos materiales, desafortunadamente en algunos casos estas diferenciaciones no han sido muy precisas, debido a la interacción del genotipo con el ambiente (Brown, 1979 y Glotieb, 1977. citado en CATIE, 1995).

Como menciona Becerra y Paredes, (2000) citado por Allende (2017), los marcadores moleculares se han convertido, en los últimos años, en herramientas complementarias muy importantes para los estudios de diversidad y de mejoramiento, además de permitir alcanzar los objetivos de mejora de manera más rápida.

La identificación genética con el uso de marcadores moleculares permite un uso óptimo de los recursos de los bancos de germoplasma y de colecciones nucleares para fines de mejoramiento. El uso de *softwares* adecuados para estos análisis facilita al mejorador tener una base idónea para la selección de individuos como progenitores, ya que se analiza la proximidad o divergencia genética entre estos. Optimizando así los cruzamientos y la obtención de híbridos con las características de interés en menor tiempo (Blas *et al.*, 2011).

#### **b. Polimorfismos**

Se considera que hay polimorfismos cuando están presentes formas alternativas de un mismo gen para una población determinada y de haber un mayor número de alelos, mayor es el polimorfismo. Cuando un alelo en un locus para una determinada población se presenta en una frecuencia mayor al uno por ciento, ésta se denomina polimórfico. Un polimorfismo es considerado neutro si su presencia o ausencia no confiere una ventaja o desventaja al individuo (Salazar *et al.*, 2013). Los polimorfismos confieren variación alélica entre los individuos. Existen varios tipos de polimorfismos como las inserciones, las deleciones y los cambios de números en secuencias repetidas, donde se encuentran los microsatélites.

#### **c. Reacción en cadena de la polimerasa (PCR)**

Blas (2010) describe la reacción en cadena de la polimerasa (PCR) como una reacción cíclica dirigida enzimáticamente, que permite la amplificación *in vitro* de un segmento de ADN, el cual está ubicado dentro de dos regiones cuyas secuencias ya se conocen. Se hace uso de dos oligonucleótidos (10-30 nucleótidos), los cuales flanquean el segmento de interés. Estos oligonucleótidos o *primers*, se hibridan a las partes terminales de una secuencia específica de la cadena de ADN y se efectúa la amplificación por medio de la ADN polimerasa. Los componentes para la realización de la PCR son: ADN polimerasa, ADN molde, Desoxinucleótidos, amortiguador o *buffer*, Cloruro de magnesio y oligonucleótidos (*primers*).

Los ajustes en cantidades de componentes y tiempo han permitido que se desarrollen diversas variantes de este método, como la PCR-multiplex, donde con una sola reacción se puede amplificar diferentes regiones de interés (Bolívar *et al.*, 2013).

#### **d. Marcadores Moleculares empleados en café**

Entre los trabajos desarrollados en café, Maluf *et al.*, (2005) realizaron un estudio de diversidad con el uso de marcadores RAPDs y con cultivares de café del IAC (Instituto Agronómico de Campinas), donde reportaron un bajo polimorfismo entre el material de *C. arabica* trabajado. En Perú, se realizó un estudio de diversidad de café, basado en marcadores moleculares tipo RAPDs, con material de Villa Rica realizado por Palomino *et al.*, (2014), en el cual reportaron un bajo grado de diversidad genética atribuido a la biología reproductiva autógena y al proceso de establecimiento del cultivo en el país a través de pocos individuos.

Adicionalmente, Steiger *et al.*, (2002) evaluaron la diversidad genética de 61 accesiones, compuestas por *C. arabica* y dos especies diploides del género *Coffea*, con marcadores moleculares tipo AFLP, estas accesiones estaban compuestas por seis cultivares, los cuales reportaron un promedio de similaridad genética de 0.93 entre los cultivares de *C. arabica* y un promedio de 0.54 con *C. canephora*. En otros estudios, Anthony *et al.*, (2002), Cubry *et al.*, (2008) y Geleta *et al.*, (2012), reportaron un bajo polimorfismo en las muestras de café cultivado analizadas, usando marcadores moleculares tipo microsatélites SSRs. Estos resultados coincidieron con el trabajo reportado por Blas *et al.*, (2011), que obtuvieron con los marcadores moleculares SSR y un polimorfismo ligeramente mayor con el uso de marcadores tipo AFLP. Ante lo mencionado, la baja variabilidad podría atribuirse a la característica autógena de *C. arabica* y que las variedades cultivadas en el Perú tienen un origen geográfico común. Entre los trabajos más recientes, como el realizado en Costa Rica por Sánchez (2017), reportó un PIC (*Polymorphism Information Content*) de 0.5183 con el uso de marcadores tipo SSR y muestras de café arábica.

Diversos estudios han demostrado la baja diversidad genética de arábica comparado con robusta, esto es debido a su comportamiento autógeno, a su evolución histórica y a su domesticación, esto se ve representado en la falta de genotipos resistentes a distintas plagas y enfermedades. En América la base genética es aún más estrecha ya que esta representa solo a un subconjunto total de arábica debido a la introducción de pocos individuos. Por ello es de mucha importancia la obtención de individuos con resistencia a plagas y enfermedades para asegurar una producción sostenible. (Geleta *et al.*, 2012).

#### **e. Marcadores microsatélites (SSR)**

Los marcadores moleculares microsatélites, también llamados *simple sequence repeats* (SSRs) o *short tandem repeats* (STRs), son secuencias de ADN las cuales están conformadas por uno a seis nucleótidos que se repiten en tándem y generalmente las repeticiones de dos nucleótidos son las más comunes (Hamad, *et al.*, 1982, citado por Blas 2010). Los SSRs han sido reportados en muchas especies vegetales, incluyendo el café (Combes *et al.*, 2000; Baruah *et al.*, 2003; Maluf *et al.*, 2005; Vieira *et al.*, 2010; Sánchez 2017). Estos son utilizados como marcadores para múltiples aplicaciones debido a su naturaleza codominante; extensiva cobertura del genoma; alta reproducibilidad, costo accesible y baja demanda técnica (Onysk y Boczkowska, 2017). Además, otra de las ventajas de los SSRs es la automatización, como la que se puede realizar por medio de un analizador genético; donde se realiza la electroforesis capilar que permite la separación de fragmentos amplificados por su tamaño (pb) y por los fluoróforos utilizados; de este modo facilitando la detección de los alelos.

#### **f. Electroforesis capilar**

La electroforesis es la separación de moléculas en un campo eléctrico, esta se da en un medio líquido y una matriz porosa usualmente un gel como los de agarosa. Esta separación depende del tamaño de la partícula o fragmento de ADN, por lo cual fragmentos más grandes migraran más lentamente y viceversa. Esta técnica es usada para separar e incluso purificar macromoléculas, la migración de ácidos nucleicos se da del electrodo negativo (cátodo) al positivo (ánodo) (Blas *et al.*, 2011). Por su parte la electroforesis capilar se basa en el mismo principio de la electroforesis en agarosa, donde se separan moléculas grandes y pequeñas con diferentes velocidades de migración bajo la acción de un campo eléctrico, pero en este caso, se realiza a través de un capilar de un diámetro bastante pequeño (10-200  $\mu\text{m}$ ). Dentro de las ventajas que posee, destaca su gran capacidad de resolución con lo cual permite diferenciar con mayor precisión el tamaño de los productos o fragmentos a migrar, elimina la necesidad de la utilización de un gel de soporte como medio, tiene una disipación de calor muy efectiva y es viable el uso de altos voltajes para conseguir una reducción en el tiempo del proceso (Osatinsky, 2007).

Durante la migración de los fragmentos al electrodo positivo, un láser incide en ellos, ocasionando la emisión de fluorescencia por parte de los fluoróforos. Las señales son

detectadas por una cámara CCD (Charge-coupled device), las cuales se basan en la conversión de señales luminosas a señales eléctricas. Estas señales se pasan a formato digital para su visualización en forma de picos a través de un electroferograma (Vía y Rada, 2015).

#### **g. Cola M13**

Para analizar el tamaño de los productos de PCR mediante electroforesis capilar, es necesario el marcaje de los primers con tintes fluorescentes. Pero, esto puede incrementar considerablemente los costos del análisis ya que se necesitan probar muchos primers y no todos serán polimórficos en los individuos evaluados, además se debe considerar que dichos primers serán usados solo para algunas reacciones y luego al no darles más uso serán refrigerados hasta su pérdida (Schuelke, 2000). Esto se puede superar con la adición de una cola M13 la cual es como menciona David (1993), una secuencia derivada de un vector bacterial.

El uso de la cola M13 añadido en el extremo 5' del primer forward, el primer reverse correspondiente y la misma cola M13 marcada con fluorescencia, nos permite obtener el producto amplificado marcado con fluorescencia, para esto se debe hacer uso de menor cantidad de primer forward con cola M13 ya que en los primeros ciclos de la PCR se generará el producto esperado y posteriormente al agotarse el primer forward, la cola M13 fluoromarcada trabajará como dicho primer para los siguientes ciclos de amplificación, de esta manera el producto amplificado contará con fluorescencia (Onysk y Boczkowska, 2017).

### **2.1.5. Estimación de la Diversidad genética**

#### **a. Variabilidad genética**

La variabilidad genética alude a la variación en los genotipos de una población o especie. Esta variación es la base de la evolución ya sea guiada por selección natural o por deriva génica. Los estudios de variación genética han permitido conocer mejor la diversidad biológica, las relaciones filogenéticas, historia evolutiva, entre otros (Tomasco y Lessa, 2015).

Frecuencia alélica: también conocida como frecuencia génica, determina las frecuencias de los genotipos individuales en una población mediante la proporción relativa de alelos y se expresa mediante el porcentaje o la unidad (Hernández y Trejo, 2014).

$$f(A) = \frac{\#alelos\ observados}{\#total\ de\ alelos}$$

Heterocigosidad observada: se refiere a la proporción de individuos heterocigotos observados entre el total de individuos evaluados en un locus determinado (Frankham *et al.*, 2002).

$$H(obs) = 1 - \Sigma frecuencias$$

Heterocigosidad esperada: es la probabilidad de que dos alelos tomados al azar de la población sean diferentes entre sí (Hernández y Trejo, 2014). Se estima a partir de las sumatorias de las frecuencias obtenidas para cada uno de los alelos; bajo la suposición de estar en equilibrio Hardy – Weinberg (H-W), que postula que la composición genética de una población natural se mantiene en equilibrio mientras no actúe la selección natural, ni otros factores como la deriva génica o la mutación (Paredes, 2013).

$$H(esp) = 1 - \Sigma frecuencias^2$$

Contenido de información polimórfica (PIC): es un valor que puede estar entre 0 y 1, el cual refleja el poder informativo de un marcador en la población. El carácter de informativo está relacionado a la probabilidad de que la progenie cargue un alelo que permita deducir el genotipo de su progenitor en un determinado locus. Se considera a un locus como altamente informativo si el PIC obtenido es mayor a 0.5, razonablemente informativo si el valor se encuentra entre 0.5-0.25 y ligeramente informativo si es menor a 0.25 (Botstein *et al.*, 1980).

## **b. Estructura genética de las poblaciones**

Las diferencias en las proporciones de los alelos para un gen determinado o frecuencias alélicas entre poblaciones constituyen la estructura genética. Cuando se encuentran marcadas diferencias en las frecuencias alélicas entre poblaciones de una especie, se puede considerar como una especie con alta estructura genética. Por otro lado, cuando las poblaciones son casi idénticas, es decir poseen muy pocas o no poseen diferencias en las frecuencias alélicas, se puede considerar como una especie con baja estructura genética (Eguiarte *et al.*, 2010).

(Fst): Índice que mide el grado de diferenciación genética entre subpoblaciones (Paredes, 2013). Si hay flujo génico, es decir si hay una transferencia de alelos dentro de las subpoblaciones, esta probabilidad es baja; mientras que, si hay una baja tasa de flujo génico, las poblaciones divergen, ocurre mayor endogamia y el Fst incrementa. Toma valor de 0 cuando no hay diferenciación y de 1 si las poblaciones son completamente distintas (Frankham *et al.*, 2002).

(Fis): Describe la distribución de los genotipos dentro de las poblaciones e indica que tan distanciada se encuentra una población del equilibrio H-W, que se daría con las frecuencias genotípicas obtenidas de un apareamiento al azar. Si la Fis es 0 indica un apareamiento al azar, si es 1 indica que los individuos son homocigotos, si es negativo hay un exceso de heterocigosis (Eguiarte *et al.*, 2010).

(Fit): Mide la falta de heterocigosidad total y la desviación en las frecuencias genotípicas observadas en la población con relación a las esperadas si estuviera en equilibrio H-W. Va de -1 a 1, indicando exceso o deficiencia de heterocigotos respectivamente (Paredes, 2013).

### III. METODOLOGÍA

#### 3.1. Material vegetal

Se utilizaron 166 muestras de *C. arabica*, las cuales fueron colectadas en el mes de abril de 2019 (Figura 1). Dichas muestras pertenecen a la colección de café del INIA, la cual se encuentra establecida en Pichanaki, Junín (Anexo 7). En la colecta se obtuvo material fresco, el cual fue transportado hacia Lima, para esto se empleó papel toalla humedecido recubriendo cada muestra foliar y fueron colocados en cajas de tecnopor, periódicamente se volvían a humedecer los papeles toalla de ser requerido (Figura 2). También se hizo uso de silica gel con el fin de secar otro juego de muestras que servirían como contra muestras. Las esferas de silica gel fueron constantemente cambiadas cuando éstas iban tomando la humedad de las muestras, y colocadas en estufas para secar las esferas y recuperar su poder higroscópico. Además de ello se emplearon tres muestras adicionales de ADN, correspondientes a *C. canephora* la cual fue colectada en Junín, y *C. liberica* y el híbrido S288 cuyos ADN fue proporcionado por el Laboratorio de Biología Molecular y Genómica del INIA.



**Figura 1: Colecta de tejido foliar de café en la colección de café del INIA**



**Figura 2: Estado de hojas frescas al retornar a Lima**

### 3.2. Extracción de ADN de tejido foliar

Las extracciones de ADN de café se dieron en el laboratorio de Biología Molecular y Genómica del INIA (Figura 3). Debido a que se obtienen mejores resultados en la obtención de ADN al usar material fresco, se priorizó el uso de este material para las extracciones y se realizaron seguidamente al retorno del viaje de colecta, asegurando el buen estado de las hojas mantenidas en fresco. En los casos en los que el ADN obtenido no fuera de la calidad esperada y de no contar con más tejido fresco, se hizo uso del tejido seco almacenado. El ADN fue extraído usando un protocolo que se estandarizó para café, el cual se basa en el método de CTAB (Doyle y Doyle, 1989), con modificaciones como el uso de un buffer de lavado (Inglis, *et al.*, 2018). Este protocolo fue probado en diversas muestras de café, obteniendo resultados de calidad y cantidad que superaron los parámetros requeridos.



**Figura 3: Extracción de ADN de café en el Laboratorio de Biología Molecular y Genómica – INIA**

#### 3.2.1. Protocolo de extracción

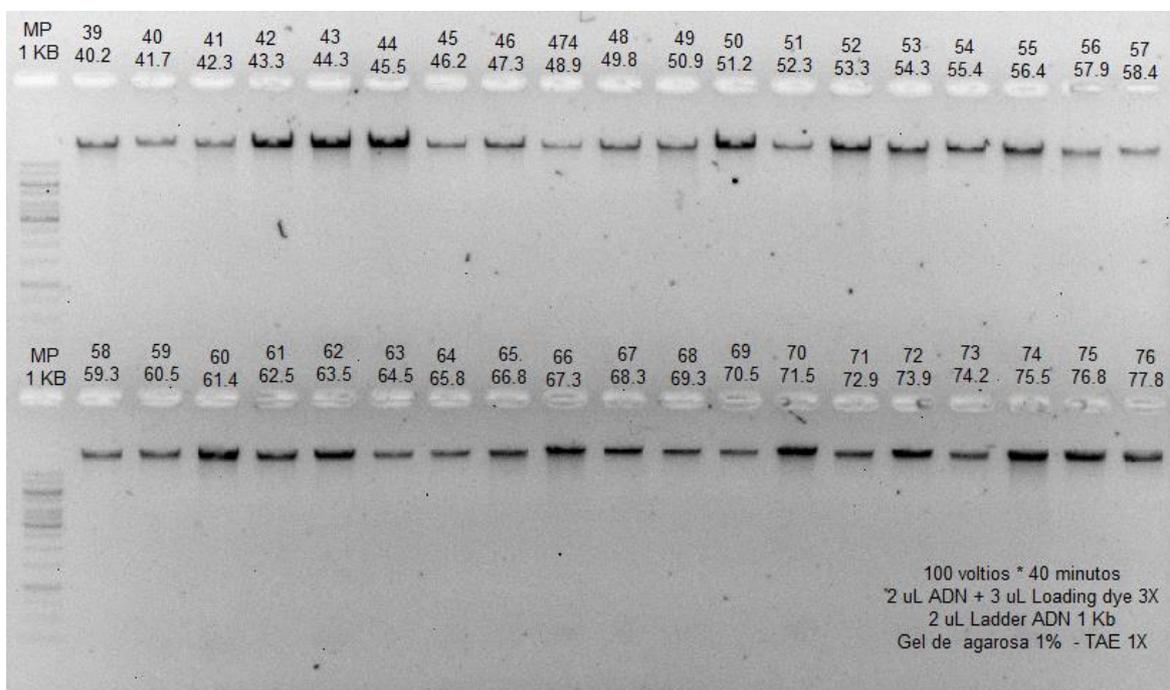
1. Elegir hojas frescas de la planta de café, trozarlas y depositarlas en un mortero. Añadir nitrógeno líquido para su molienda, con ayuda de un pilón.
2. Etiquetar y enfriar los microtubos de 2 mL en un mortero con nitrógeno líquido.
3. Colocar el tejido pulverizado en los microtubos de 2 mL y añadir 700  $\mu$ L del buffer de pre-lavado con sorbitol y 2  $\mu$ L de  $\beta$ -mercaptoetanol.
4. Llevar al vórtex durante 20 segundos.
5. Centrifugar a 14 000 rpm por 10 minutos.

6. Descartar el sobrenadante que contiene polisacáridos y polifenoles y añadir 700  $\mu$ L de buffer de extracción CTAB 2X y 2 $\mu$ L de  $\beta$ -mercaptoetanol por muestra.
7. Incubar las muestras a 65°C por 45 minutos a 1 hora. Agitando los tubos constantemente cada 20 minutos (Baño María o Termomixer).
8. Retirar las muestras y dejar enfriar a temperatura ambiente por 2 min.
9. Agregar 700  $\mu$ L de Cloroformo: Alcohol Isoamílico (24:1), agitar los tubos ligeramente hasta homogenizar el contenido (verde lechoso).
10. Llevar al vórtex durante 20 segundos.
11. Centrifugar a 14 000 rpm por 10 minutos.
12. Transferir 600  $\mu$ L del sobrenadante a nuevos microtubos de 2 mL, previamente etiquetados.
13. Agregar 60  $\mu$ L de CTAB 10X (precalentado a 65°C) y 700  $\mu$ L de Cloroformo: Alcohol Isoamílico (24:1) y agitar hasta homogenizar completamente.
14. Llevar al vórtex durante 20 segundos.
15. Centrifugar a 14 000 rpm por 10 minutos.
16. Transferir 500  $\mu$ L del sobrenadante a microtubos de 1.5 mL, previamente etiquetados.
17. Agregar 2/3 volumen de Isopropanol (330  $\mu$ L), agitar suavemente e incubar a -20°C por 30 minutos.
18. Centrifugar a 14 000 rpm por 10 minutos.
19. Decantar el Isopropanol con cuidado de no perder el pellet.
20. Agregar 500  $\mu$ L de Etanol al 70%. Despegar el pellet.
21. Centrifugar a 14 000 rpm por 10 minutos.
22. Decantar el etanol 70% con cuidado de no perder el pellet.
23. Repetir los pasos 20, 21 y 22.
24. Llevar al concentrador (Speed Vac) a 60°C por 10 min (V-AL)
25. Agregar 50  $\mu$ L T<sub>10</sub>E<sub>1</sub>. Homogenizar suavemente la muestra de ADN.
26. Añadir 2  $\mu$ L de ARNasa A e incubar en el termomixer a 37°C. por 60 minutos.  
Almacenar a 4°C y posteriormente a -20°C.

### **3.2.2. Evaluación de calidad de ADN**

Seguidamente a la extracción, se evaluó la calidad de las muestras de ADN stock de las accesiones extraídas, esto fue determinado a través de electroforesis horizontal en geles de agarosa al 1 %. Se mezclaron 2  $\mu$ L de ADN con 1  $\mu$ L de 6X dye loading (buffer de carga)

el cual otorga mayor densidad al ADN e indicadores de migración mediante coloración. Se utilizó un marcador de peso 1 Kb (MP 1 kb). La corrida electroforética se dio a 100-120 voltios en buffer 1X TAE por 30-40 minutos y fue fotografiado bajo luz UV en un sistema de fotodocumentación Chemidoc™ MP (BIO-RAD). La edición de las fotos se realizó con el programa Quantity One v. 4.6.3 (Biorad, Hercules, CA, USA) (Figura 4).



**Figura 4: ADN extraído de 38 muestras de café, pertenecientes a la colección de café del INIA en gel de agarosa al 1% (m/v) (MP 1KB: Marcador de peso molecular GeneRuler™ 1kb (Thermo Scientific))**

### 3.2.3. Cuantificación de ADN

Posterior a la extracción y evaluación de calidad de ADN de las hojas de café, se realizó la cuantificación mediante el espectrofotómetro NanoDrop™ 8000c.

La relación de las absorbancias A260/A280 permite conocer la pureza del ADN obtenido. Se considera que el ADN es de calidad óptima cuando la relación 260/280 es mayor de 1.8. Una relación A260/280 > 2.1 indica la presencia de ARN en la muestra. Si esta relación es baja (A260/280 < 1.6) indica contaminación por proteínas o fenoles.

Una segunda valoración de la pureza de ácidos nucleicos es la proporción 260/230. Los valores aceptados se encuentran en el rango de 1.5 a 2.2, si la relación es menor de 1.5 indican la presencia de contaminantes en la muestra, tales como sales y carbohidratos.

Las cantidades obtenidas de las muestras trabajadas fueron variables, éstas se situaron en un rango aproximado que va desde 250 ng/μL hasta más de 2000 ng/μL, siendo estos valores adecuados para los procesos requeridos (Tabla 1).

**Tabla 1: Valores de absorbancia y concentración de 20 muestras de ADN, obtenidos mediante espectrofotometría**

<b>Cod. Trabajo</b>	<b>Cod. Accesoión</b>	<b>N° de Réplica extraída (1-10)</b>	<b>260/280</b>	<b>260/230</b>	<b>Conc. (ng/ul)</b>
1	1	4	1,94	1,91	2070
2	2	1	1,91	1,81	585,6
3	3	6	1,92	1,86	477,7
4	4	6	1,97	1,88	692,7
5	5	5	1,96	1,87	1204
6	6	5	1,93	1,83	532,2
7	7	4	1,91	1,71	402,5
8	8	5	1,94	1,78	1989
9	9	3	1,92	1,68	425
10	10	10	1,96	1,82	530,9
11	11	3	1,88	1,57	628
12	12	3	1,8	0,96	274,3
13	13	6	1,91	1,58	348,6
14	14	3	1,93	1,9	566,5
15	15	8	1,99	2,01	1369
16	16	3	1,9	1,72	261
17	17	3	2,04	1,91	657,9
18	18	5	1,97	1,91	1241
19	19	5	1,97	1,97	781
20	20	4	1,87	1,33	339,8

### 3.2.4. Dilución de muestras

Posterior a las evaluaciones de calidad en geles de agarosa, cuantificaciones por espectrofotometría y repeticiones de extracción de muestras que no cumplieron con los estándares requeridos, se procedió a realizar las diluciones de todas las muestras a una concentración de 12 ng/μL para lo cual se usó agua libre de nucleasas (NFW), la cual está libre de enzimas que puedan degradar el ADN, alícuotas de las diluciones fueron colocadas en placas de PCR para facilitar su uso. Para los cálculos de concentraciones finales y cantidades, se hizo uso de las concentraciones iniciales del ADN stock, obtenidas mediante el espectrofotómetro (Tabla 2). Además, también se hizo uso de NFW estéril para la re

suspensión y dilución de los primers hasta una concentración de 10 mM que fue la concentración requerida para la amplificación por PCR.

**Tabla 2: Cantidades de ADN stock de café y NFW para diluciones a una concentración de 12 ng/μL en un volumen final de 200 uL**

Cod. Trabajo	Cod. Accesoión	260/280	260/230	Conc. (ng/μl)	Vol. ADN <sup>1</sup>	Vol. NFW <sup>2</sup>
1	1	1,94	1,91	2070	1,16	198,84
2	2	1,91	1,81	585,6	4,10	195,90
3	3	1,92	1,86	477,7	5,02	194,98
4	4	1,97	1,88	692,7	3,46	196,54
5	5	1,96	1,87	1204	1,99	198,01
6	6	1,93	1,83	532,2	4,51	195,49
7	7	1,91	1,71	402,5	5,96	194,04
8	8	1,94	1,78	1989	1,21	198,79
9	9	1,92	1,68	425	5,65	194,35
10	10	1,96	1,82	530,9	4,52	195,48
11	11	1,88	1,57	628	3,82	196,18
12	12	1,8	0,96	274,3	8,75	191,25
13	13	1,91	1,58	348,6	6,88	193,12
14	14	1,93	1,9	566,5	4,24	195,76
15	15	1,99	2,01	1369	1,75	198,25
16	16	1,9	1,72	261	9,20	190,80
17	17	2,04	1,91	657,9	3,65	196,35
18	18	1,97	1,91	1241	1,93	198,07
19	19	1,97	1,97	781	3,07	196,93
20	20	1,87	1,33	339,8	7,06	192,94

1. Volumen de ADN stock para un volumen de dilución de 200 μl a una concentración de 12 ng/μL.

2. Volumen de NFW para un volumen de dilución de 200 μl a una concentración de 12 ng/μL.

### 3.3. Amplificación por PCR

Para las amplificaciones se utilizó un grupo de primers seleccionados (Tabla 3), los cuales fueron reportados con altos índices de contenido polimórfico (PIC) en investigaciones en café (Baruah *et al.*, 2003; Hendre *et al.*, 2008; Missio *et al.*, 2010; Sánchez, 2017; Syafaruddin *et al.*, 2014; Vieira *et al.*, 2010). Se utilizaron cuatro distintos fluoróforos (FAM, VIC, PET y NED), estos fueron asignados a grupos determinados de primers que poseían temperaturas de *annealing* similares y distintos tamaños de productos amplificados, estos fluoróforos fueron agregados al mix de PCR para su anclaje en los primers y la posterior detección del producto amplificado mediante fluorescencia (Figura 5). Las cantidades y concentraciones finales de los componentes para el mix de PCR fueron ajustadas con el fin de permitir la realización de PCR multiplex, es decir reacciones con más de un par de primers en un mismo tubo, reduciendo así el tiempo y los costos, por la menor cantidad de componentes empleados (Tabla 6, 7 y Anexo 2).

**Tabla 3: Lista de Primers empleados con la cola M13 añadida en la secuencia forward**

<b>Nº</b>	<b>Nombre</b>	<b>Forward Primer</b>	<b>Reverse Primer</b>	<b>Fuente</b>
1	SSR03	CACGACGTTGTAAAACGACGACAAAACACCGCCCAAATA	AGCGAGACAGAGGAAGGGAATATT	Sánchez. 2017
2	CaM46	CACGACGTTGTAAAACGACTGGTGCGGTGTTTTTCAGTTTGGAGA	AACCACCCACGCCACCAATTAAT	Hendre et al., 2008
3	Car M105	CACGACGTTGTAAAACGACTGCTCCTACTAAATACCCAAACA	ATATGCCCAAGAAAATTAGATGAAA	Vieira et al., 2010
4	Car M092	CACGACGTTGTAAAACGACAGGCCAGACTTGTTTGATTTTG	GGCCCTTCTCGCTTTAGTTG	Vieira et al., 2010
5	SSRCa18	CACGACGTTGTAAAACGACGTCTCGTTTCACGCTCTCTC	ATTTTTGGCACGGTATGTTC	Sánchez. 2017
6	CaM16	CACGACGTTGTAAAACGACAAGGCAGCTGAAGCGGGACAAA	TGGGGAGAGCTGCAGTTGGAGG	Sánchez. 2017
7	CaM41	CACGACGTTGTAAAACGACCATCGTCTCCATCGTTGCTCTATC	CCCTCCCCCTCTTTCCTATCTAAT	Hendre et al., 2008
8	M20	CACGACGTTGTAAAACGACCTTGTTTGAGTCTGTCTGCTG	TTCCCTCCCAATGTCTGTA	Vieira et al., 2010
9	M32	CACGACGTTGTAAAACGACAACCTCTCCATTCCCGCATTC	CTGGGTTTTCTGTGTTCTCG	Sánchez. 2017
10	CaM49	CACGACGTTGTAAAACGACCCGGTTAATACATTGGTCTTT	ATGACATTGTTGACTTTGCTATAA	Hendre et al., 2008
11	CaM40	CACGACGTTGTAAAACGACTTGACACGAAACAGGAAATAAATATAG	CCCTCCCCTCATAGCCCTTT	Hendre et al., 2008
12	M753	CACGACGTTGTAAAACGACGGAGACGCAGGTGGTAGAAG	TCGAGAAGTCTGGGGTGT	Sánchez. 2017
13	M24	CACGACGTTGTAAAACGACGGCTCGAGATATCTGTTTAG	TTTAATGGGCATAGGGTCC	Sánchez. 2017
14	CaM38	CACGACGTTGTAAAACGACGAAGCTGAAGCGGGAGGGTAGTAATT	CCCATCCACCCAACCTTCATTTC	Hendre et al., 2008
15	Car M050	CACGACGTTGTAAAACGACATCCCTCCACGGCAACCCAAAATA	ATCCGCAGCCCTCACCATCCA	Vieira et al., 2010

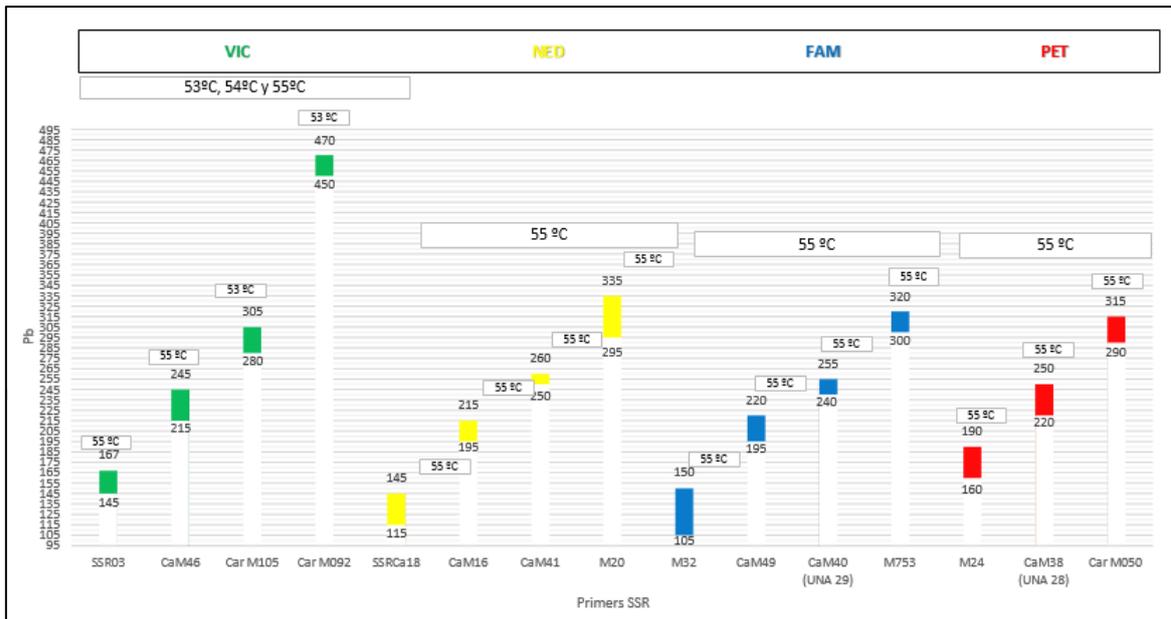


Figura 5: Agrupamiento de los primers según los productos esperados y sus respectivas temperaturas de *annealing*

Tabla 4: Componentes y cantidades para PCR multiplex con tres pares de primers

Componentes	Concentración Inicial	Concentración Final	Vol. Final para 1 Rx.	Vol. Final para 45 Rx.
H <sub>2</sub> O	-	-	2.867 µL	129.015 µL
MgCl <sub>2</sub>	50 mM	2.5 mM	0.5 µL	22.5 µL
Buffer	10 X	1 X	1 µL	45 µL
DNTPs	10 mM	0.2 mM	0.2 µL	9 µL
1er Primer F	10 µM	0.03 µM	0.03 µL	1.35 µL
1er Primer R	10 µM	0.06 µM	0.06 µL	2.70 µL
2do Primer F	10 µM	0.03 µM	0.03 µL	1.35 µL
2do Primer R	10 µM	0.06 µM	0.06 µL	2.70 µL
3er Primer F	10 µM	0.03 µM	0.03 µL	1.35 µL
3er Primer R	10 µM	0.06 µM	0.06 µL	2.70 µL
Fluoróforo	10 µM	0.113 µM	0.113 µL	5.085 µL
Taq Pol	5 U	0.25 U	0.05 µL	2.25 µL
ADN	10 ng	50 ng	5.0	
<b>Volumen Final</b>	-	-	<b>10</b>	<b>450</b>

Tabla 5: Condiciones de amplificación

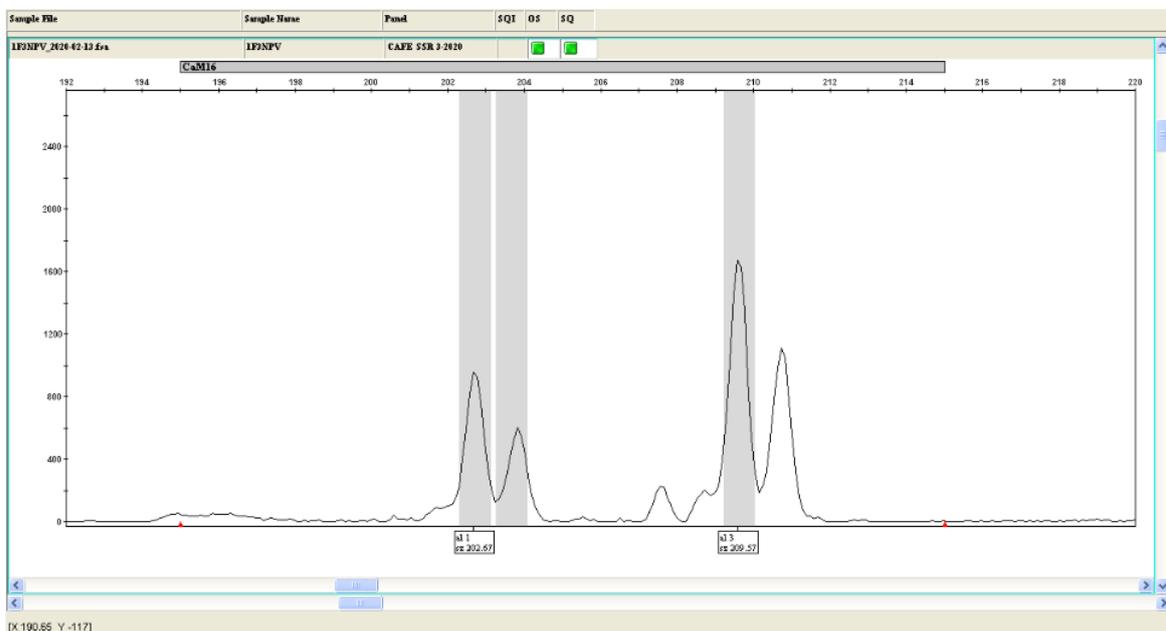
Pasos	Temperatura (°C)	Tiempo (min)	Ciclos
Calentamiento inicial	94	5	1
Desnaturalización	94	1	30
Annealing	Ta	1	30
Extensión	72	1	30
Extensión Final	72	5	1
Almacenamiento Temporal	4	∞	-

### 3.4. Electroforesis capilar

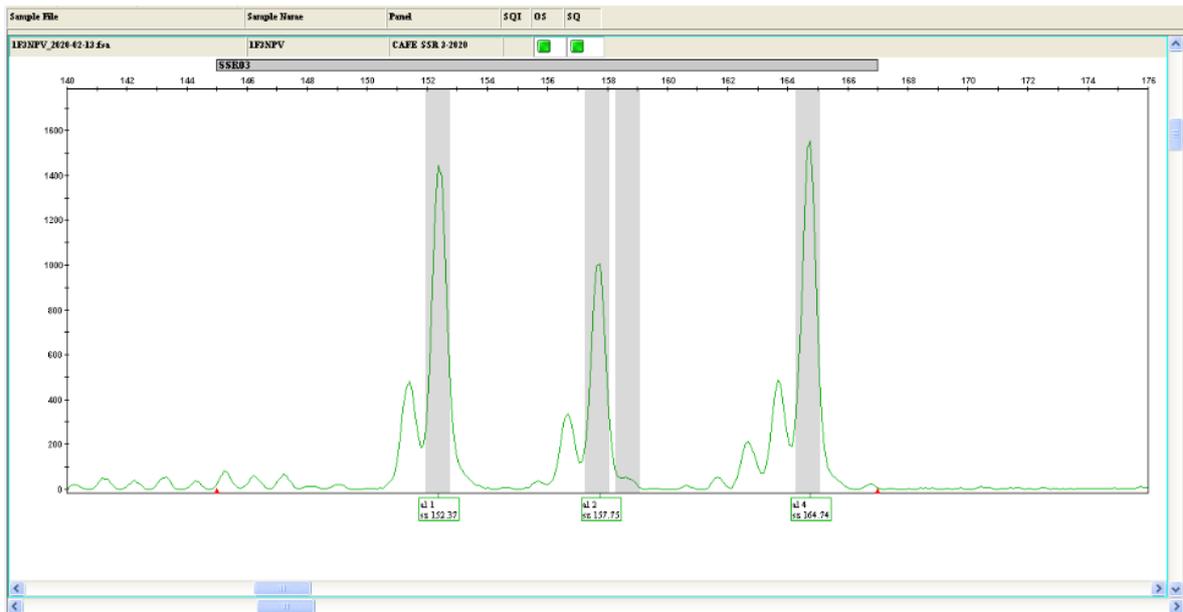
Se realizó electroforesis capilar para evaluar los productos obtenidos por PCR multiplex, mediante la fluorescencia emitida, en un analizador genético ABI 3130XL (Applied Biosystems, USA). Para esto, se mezclaron 9.7  $\mu\text{L}$  de Formamida Hi-Di (Highly Deionized) y 0.3  $\mu\text{L}$  de un estándar de tamaño Liz 600, a esto se le agregó 1  $\mu\text{L}$  de cada uno de los productos de PCR para cada fluoróforo, dando un total de 4  $\mu\text{L}$  de mezclas de productos. Posteriormente la mezcla final fue sometida a un shock térmico, el cual consistió en someter a las muestras a una temperatura de 95  $^{\circ}\text{C}$  por un periodo de 3 minutos e inmediatamente fueron llevadas a -20  $^{\circ}\text{C}$  por tres minutos, luego fue migrado mediante electroforesis capilar en el analizador genético.

### 3.5. Generación de electroferogramas

La data obtenida luego de la electroforesis capilar, fue trabajada con el *software* GeneMapper™ V.4.0 (Applied BioSystems), que permitió generar electroferogramas donde se muestran los tamaños de los alelos en pares de bases (pb), estos alelos se encuentran representados en forma de picos y su altura depende de la intensidad de fluorescencia de cada producto amplificado, la cual se mide en unidades de fluorescencia relativa (RFU). Con el análisis de los electroferogramas se detectaron y definieron como alelos verdaderos a aquellos alelos que tuvieron un RFU mayor a 100 unidades (Figura 6 y 7).



**Figura 6: Electroferograma con primer CaM16, con dos alelos de tamaño: 202 y 209 pb**



**Figura 7: Electroferograma con primer SSR03, con tres alelos de 152, 158 y 164 pb**

### 3.6. Análisis de diversidad

#### 3.6.1. Número de alelos

Con el programa GeneMapper™ V.4.0, se identificó individualmente a cada alelo presente por cada locus en los distintos individuos. De esta manera por observación simple, se logró identificar el número de alelos distintos totales, por locus y por subgrupos. Esto a su vez fue confirmado mediante la función *alleleDiversity* del paquete estadístico Polysat.

#### 3.6.2. Frecuencias alélicas de los SSR

Las frecuencias alélicas fueron estimadas empleando la función *deSilvaFreq* del paquete estadístico Polysat que puede trabajar con datos de loci con dosaje alélico ambiguo, es decir que no se conoce el número de copias alélicas. Con las frecuencias se determinó la proporción en la que están presentes distintos alelos para un locus específico bajo condiciones de herencia polisómica y una tasa de autogamia conocida, para el caso de café arábica se empleó 0.9 (Souza, *et al.*, 2004).

#### 3.6.3. Heterocigosidad

La estimación de la Heterocigosidad Esperada y la Observada en las muestras analizadas, fueron obtenidas mediante la función *summary* del paquete estadístico Adegnet, con la cual a partir de un objeto *genind* se obtuvieron los valores de heterocigosidad observada y

esperada para todos los loci. Estos índices representarían una mejor medida la variación genética al no ser arbitraria (Hernández y Trejo, 2014).

#### **3.6.4. Índice de contenido Polimórfico (PIC)**

Inicialmente se realizó una comprobación del polimorfismo con la función *isPoly* del paquete estadístico Adegnet, con esta función se pueden evaluar los loci polimórficos en la población estudiada. Para el uso de esta función se tuvo una consideración de frecuencia alélica mínima de 0.1 para dar lugar a un polimorfismo.

Posteriormente, se realizó una estimación del polimorfismo, esto se realizó calculando el índice de contenido polimórfico para cada primer empleado en el presente análisis. De esta manera se pudo conocer el poder informativo de cada marcador para la población estudiada. Este índice fue calculado con el uso del paquete *Polysat* 1.7 de uso libre, disponible para el programa R (Clark y Jasieniuk, 2011).

#### **3.6.5. Alelos Nulos**

Se evaluó la presencia de alelos nulos con el uso de la función *deSilvaFreq* del paquete estadístico *Polysat*. Esta función sirve para hallar las frecuencias alélicas considerando herencia polisómica, con un índice de autogamia conocido. Para el caso de café se empleó un valor de 0.9 (Souza, *et al.*, 2004).

### **3.7. Análisis de estructura poblacional**

#### **3.7.1. Índices de Fijación**

Se estimó el  $F_{st}$  que mide el grado de diferenciación genética entre subpoblaciones, para este cálculo, se empleó el paquete estadístico *Polysat* del programa *Rstudio*, con el fin de determinar algún grado de diferenciación genética entre esos grupos. Estos índices  $F_{st}$  (Tabla 16) fueron hallados en base a las frecuencias alélicas y para este fin se empleó la frecuencia simple hallada con la función *simplefreq*.

#### **3.7.2. Análisis de varianza molecular (AMOVA)**

El AMOVA se realizó con la ayuda del programa *Polygene*. Esta herramienta es ampliamente usada para estimar la contribución de los distintos niveles de estructura de la

población a los patrones de variación genética (Fitzpatrick, 2009; citado en Pacheco, 2015). El AMOVA posibilita la división jerárquica entre poblaciones y regiones, también permite estimar estadísticos F. Su base es una matriz de distancias genéticas (Acuña, 2016). Para este análisis se empleó un número de permutaciones de 9 999, la variación genética total se dividió en dos componentes: la variación genética entre las regiones y las variaciones dentro de las regiones. De la misma manera también se evaluó la variación genética entre las variedades y dentro de las variedades.

### **3.7.3. Estructura Poblacional**

Para la evaluación de la estructura poblacional, se realizó un análisis de agrupamiento empleando el programa STRUCTURE (Pritchard *et al.*, 2000), este programa emplea un método de agrupación bayesiana para inferir la estructura genética de la población. Se asignan probabilidades de pertenencia a K grupos a los individuos, sin requerir conocer previamente los límites poblacionales. Para este método se emplean las frecuencias alélicas de cada locus y con ello realizar la estimación de los clusters genéticamente diferentes (K). Además de ello, el programa también emplea MCMC (Markov Chain Monte Carlo) para la ejecución de iteraciones que permitan distribución de probabilidad, dentro de este proceso de iteraciones, la etapa inicial toma el nombre de burn-in o pre corrida. Un método alternativo empleado para identificar el mejor valor de K se hace mediante el método de Delta de K de Evanno *et al.*, (2005), donde se realiza una prueba de tasas de varianzas con las probabilidades generadas por el programa.

Para ese análisis se emplearon los siguientes parámetros: longitud del burn-in (10 000), número de MCMC (100 000) y el modelo de ancestralidad (Admixture). Se evaluó desde K=1 hasta K=10 y el número de iteraciones fue de 10. Con ayuda del paquete estadístico Pophelper del programa Rstudio, se evaluaron los resultados.

### **3.7.4. Análisis discriminante de componentes principales (DAPC)**

Se utilizó el paquete estadístico Adegenet 2.0, disponible para el programa R cuyos *scripts* fueron desarrollados por Jombart (2015). El DAPC es un análisis multivariado para identificar cluster genéticos. A diferencia del PCA, donde se describe la diversidad global, en el DAPC se da mayor peso a las varianzas entre los cluster formados y se minimiza la varianza dentro de éstos, con el fin de demostrar su relación (Jombart, 2013). Con este

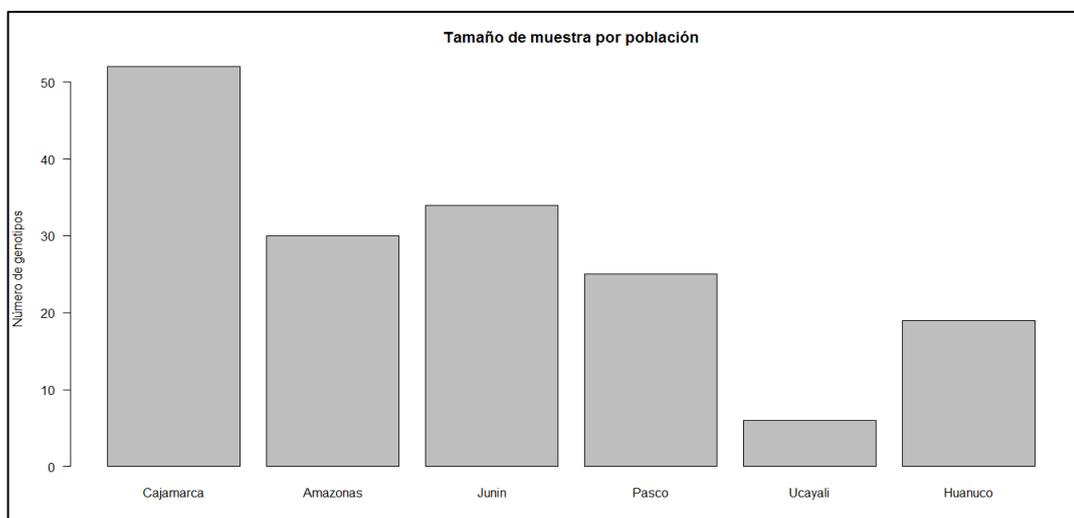
método, inicialmente se realiza un Análisis de Componentes Principales – PCA y luego se identifican grupos mediante un análisis discriminante (DA). Para este análisis se empleó el criterio “goodfit” para la generación de grupos, además se tomaron en consideración seis componentes principales que acumulaban más del 80% de la varianza de la data y cuatro funciones discriminantes.

## IV. RESULTADOS Y DISCUSIÓN

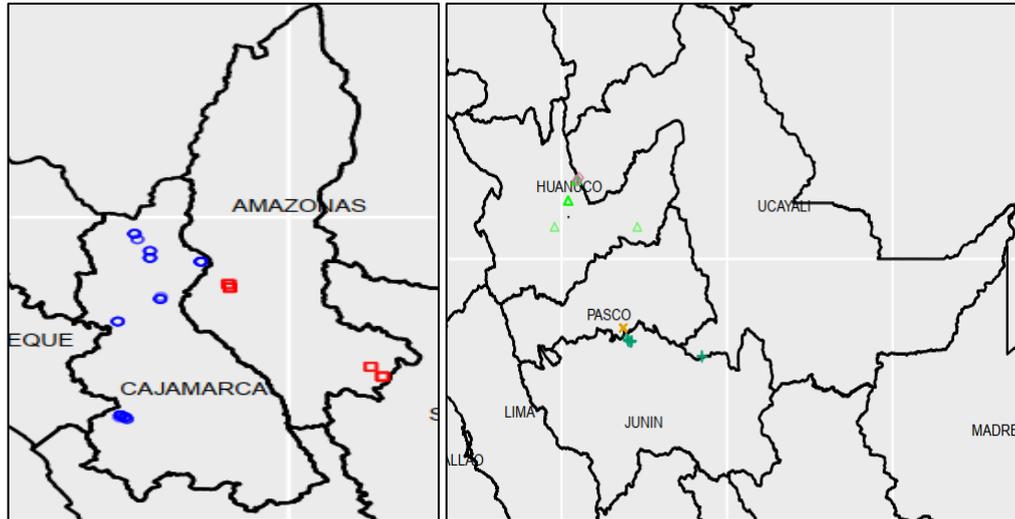
### 4.1. Material vegetal

Se evaluaron un total de 169 muestras, de las cuales 166 corresponden a la colección de germoplasma de café del INIA (Anexo 7), siendo está conformada por individuos de las regiones de Cajamarca, Amazonas, Junín, Pasco, Ucayali y Huánuco (Figura 8). La distribución a nivel de regiones se detalla en la Figura 9. Las tres muestras restantes se emplearon como *outgroup* y corresponden a muestras de: *Coffea liberica*, *Coffea canephora* (ambos de naturaleza diploide) y el híbrido S.288 que correspondería a un individuo tetraploide, resultante del cruce entre *C. arabica* y *c. liberica* (Alkimim *et al.*, 2017).

Se realizó la evaluación y estandarización de los primers de manera individual, lo que inicialmente permitió encontrar las temperaturas óptimas para su amplificación con el ADN de las muestras trabajadas. Posteriormente, gracias a la identificación de las temperaturas óptimas de hibridación en la cual los primers se alinean al templado previamente separado, se organizaron los grupos para la amplificación multiplex agrupados con los distintos fluoróforos (Tabla 8).



**Figura 8: Cantidad de individuos de la colección de germoplasma de café del INIA por región**



**Figura 9: Puntos de colecta de café en cada región, correspondiente a la data pasaporte de las accesiones trabajadas**

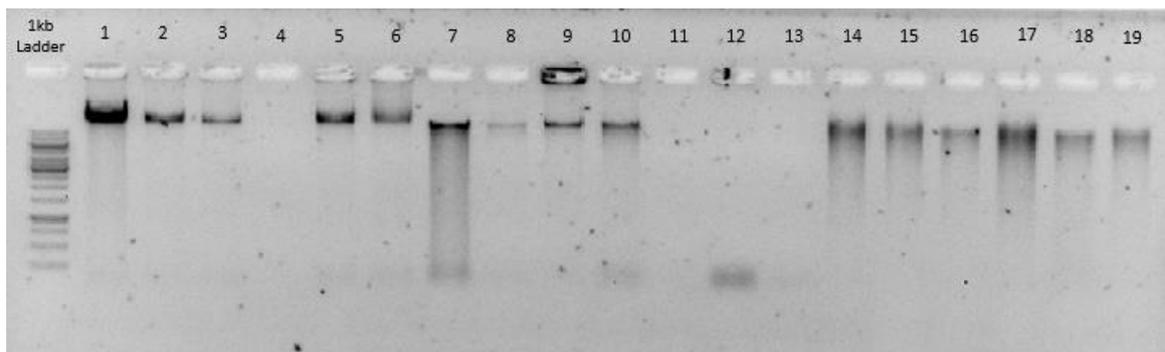
#### 4.2. Extracción de ADN

Con el uso del protocolo anteriormente descrito (Materiales y métodos) con base en CTAB (Doyle y Doyle, 1989) y con modificaciones (Inglis *et al.*, 2018), se pudo obtener ADN de buena calidad con concentraciones desde 250 ng/μL a 2000 ng/μL, la cual fue medida mediante espectrofotometría y los valores de absorbancia 260/280 y 260/230 estuvieron dentro de rangos esperados, adecuados para el posterior procesamiento de las muestras.

Además de la cuantificación se realizó electroforesis horizontal en geles de agarosa al 1%, esto con el fin de obtener una adecuada resolución para los ácidos nucleicos, los resultados de estas evaluaciones mostraron una buena migración de ADN, pocillos de carga limpios lo cual es indicador de ausencia de agentes contaminantes, y no se encontró presencia de “smear” lo cual sería indicativo de una degradación del ADN.

Dentro del proceso de adecuación del protocolo para extracción de ADN de café con una calidad idónea, se realizó la implementación de un lavado previo con un buffer a base de Sorbitol, PVP, Tris-HCl y EDTA (Inglis *et al.*, 2018), ya que el cultivo de café posee metabolitos secundarios (polisacáridos) y presencia de polifenoles que interfieren con la calidad de ADN extraído.

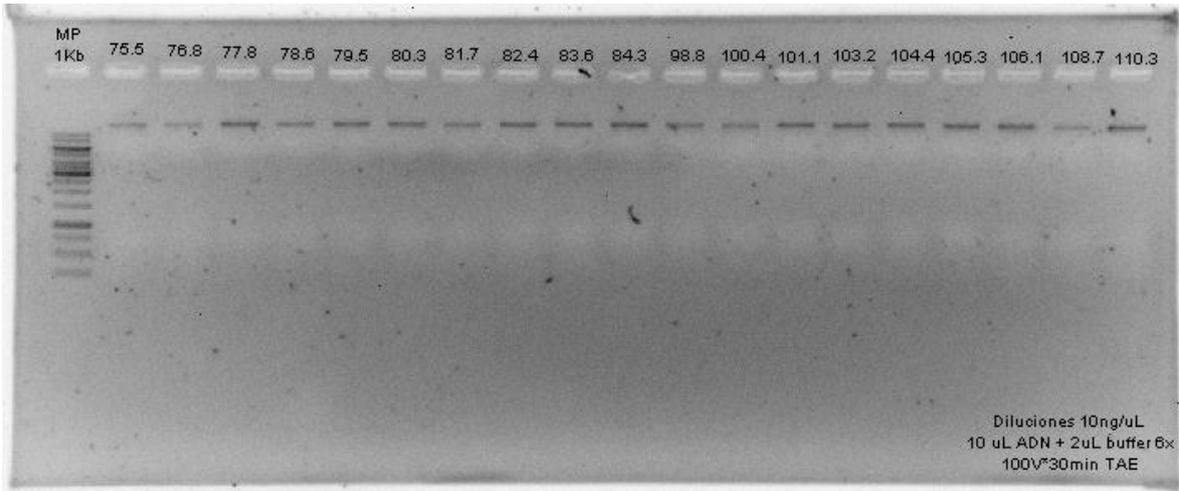
Inicialmente empleando el método original de Doyle y Doyle *et al.*, (1989), se obtuvo una calidad de ADN no adecuada con presencia de compuestos distintos a ácidos nucleicos y valores de absorbancia fuera de los límites permisibles (Figura 10). Posteriormente, con la implementación del lavado previo con el buffer de Sorbitol y el ajuste de los demás pasos de extracción, se obtuvieron valores de absorbancia adecuados, incluso para técnicas de genotipado más modernas como el análisis de SNPs y el genotipado por secuenciamiento. De esta manera se pudo contar con una metodología más accesible y económica en comparación con el uso de kits de extracción de ADN genómico vegetal.



**Figura 10: Gel de agarosa al 1% con ADN genómico de café con presencia de “smear”, indicativo de degradación**

La más alta calidad de ADN obtenido se consiguió con el uso del método de extracción modificado, a partir de hojas frescas, adicionalmente, se emplearon hojas deshidratadas con silica gel, con las cuales se consiguió extraer ADN con buena calidad, pero menor comparada con las hojas frescas.

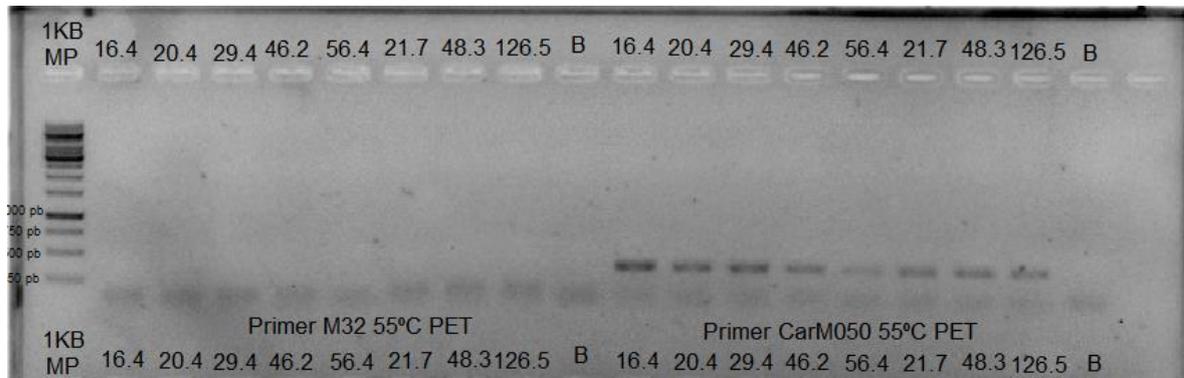
Finalmente, se realizaron las diluciones del ADN a los valores de concentración apropiados para las reacciones de PCR, las pruebas de cuantificación y calidad de diluciones mostraron lo esperado para continuar con los procedimientos posteriores (Figura 11).



**Figura 11: Gel de agarosa al 1% con ADN de café diluido a 10 ng/μl**

### 4.3. Amplificación por PCR

Se realizó la amplificación individual de los marcadores empleando gradientes de temperatura tomando como punto medio las temperaturas de *annealing* reportadas por los autores que hicieron mención a dichos marcadores. Posteriormente, mediante evaluación en geles de agarosa al 1.5% se confirmó la mejor temperatura de *annealing*, considerando una adecuada visualización de los productos esperados y la generación de productos inespecíficos (Figura 12). Con la elección de las temperaturas de *annealing* se procedió a realizar electroforesis capilar y evaluar los electroferogramas obtenidos para conocer con exactitud los rangos alélicos correspondientes a los marcadores empleados en la población estudiada, esto posibilitó el agrupamiento de marcadores cuyos productos esperados tengan rangos distintos para evitar que se superpongan.



**Figura 12: Gel de agarosa al 1.5% para la evaluación de amplificación de los productos de PCR**

#### 4.4. Análisis de marcadores SSR empleados

Luego de evaluar los electroferogramas de las reacciones de PCR multiplex se tomaron en consideración 13 marcadores considerando una adecuada intensidad, la forma de los alelos, su polimorfismo y que posean más de un alelo por locus.

Existen diferencias entre los rangos alélicos obtenidos por otros autores y los obtenidos en el presente estudio, estas diferencias pueden tener explicación en primer lugar al grupo de muestras empleadas, segundo lugar al hecho de agregar una secuencia universal (M13) para la realización de electroforesis capilar, y finalmente a la presencia de los individuos del *outgroup*, los cuales tienen alelos exclusivos y en la mayoría de casos de un tamaño mayor a los encontrados en arábica (Tabla 8).

**Tabla 6: Rangos alélicos según literatura y rangos obtenidos en el presente estudio**

N°	Fluoróforo	Nombre	AT (°C) lit.	AT (°C)	Rango literatura		Rango obtenido		Fuente
1	VIC	SSR03	55	55	134	156	153	175	Sánchez. 2017
2		CaM46	55	55	208	228	224	241	Hendre <i>et al.</i> , 2008
3		Car M105	53	55	230	270	290	306	Vieira <i>et al.</i> , 2010
4		Car M092	53	55	200	300	450	488	Vieira <i>et al.</i> , 2010
5	NED	SSRCa18	55	55	110	126	117	143	Sánchez. 2017
6		CaM16	55	55	184	200	204	220	Sánchez. 2017
7		M20	55	55	200	270	289	330	Vieira <i>et al.</i> , 2010
8	FAM	M32	55	55	99	131	118	149	Sánchez. 2017
9		CaM41	55	55	232	243	249	252	Hendre <i>et al.</i> , 2008
10		M753	55	55	283	296	307	317	Sánchez., 2017
11	PET	M24	55	55	150	180	167	190	Sánchez., 2017
12		CaM38	55	55	214	226	242	262	Hendre <i>et al.</i> , 2008
13		Car M050	55	55	150	190	294	320	Vieira <i>et al.</i> , 2010

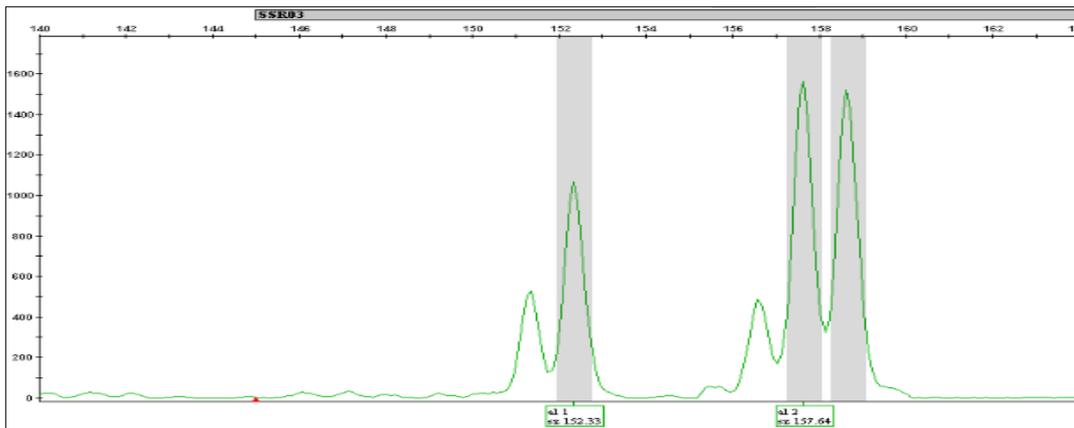
AT lit: Temperatura de annealing correspondiente a la literatura

#### 4.5. Diversidad Alélica

##### 4.5.1. Numero de alelos

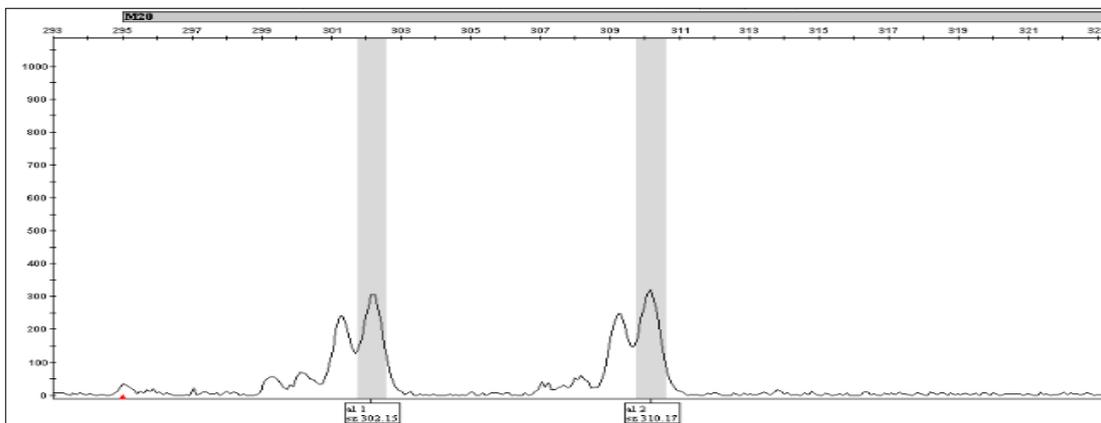
Con el programa GeneMapper™ V.4.0, se identificaron los alelos para las muestras trabajadas con los 13 primers seleccionados (M32, CaM40, M753, SSR03, CaM46, CarM105, CarM092, SSRCa18, CaM16, M20, M24, CaM38 y CarM050). Se evaluaron todos los electroferogramas generados, como el resultante de la amplificación con el primer SSR03 y el fluoróforo VIC, donde se observó la presencia de dos alelos que poseen tamaños

de 152 y 158 pares de bases, además de presentar valores entre 1000 y 1500 unidades de RFU respectivamente (Figura 13).



**Figura 13: Electroferograma generado con el marcador SSR03 mostrando dos alelos de distinto tamaño**

En otros electroferogramas como el resultante de la amplificación del primer M20 y el fluoróforo NED, se observó la presencia de dos alelos con tamaños de 302 y 310 pares de bases, además poseen un valor de RFU mayor a 300 unidades. Adicionalmente se puede observar que ambos alelos presentan la misma forma (Figura 14).



**Figura 14: Electroferograma generado con el marcador M20**

En los 169 individuos de café analizados, el número total de alelos (A) encontrados fue de 70, con una media de 5.308 por locus; además, se identificaron los rangos en pares de bases (pb) en los cuales se encuentran dichos alelos. La mayor diversidad de alelos se encontró en los loci: M20 con ocho alelos, los cuales van desde 289 hasta 330 pb, M32 con siete alelos

que van desde 118 hasta 149 pb, y M24 con siete alelos que van desde 167 hasta 190 pb. Por otro lado, el menor número de alelos se registró en el locus CaM40 en el cual solo se identificaron tres alelos distintos, los cuales van desde 249 hasta 252 (Tabla 9).

**Tabla 7: Número de alelos por locus y rango de tamaño en pares de base**

Nº	Nombre	Número de alelos	Rango mínimo	Rango máximo
1	M32	7	118	149
2	CaM40	3	249	252
3	M753	5	307	317
4	SSR03	5	153	175
5	CaM46	4	224	241
6	CarM105	6	290	306
7	CarM092	6	450	488
8	SSRCa18	6	117	143
9	CaM16	4	204	220
10	M20	8	289	330
11	M24	7	167	190
12	CaM38	4	242	262
13	CarM050	4	294	320

A nivel de regiones, la mayor cantidad de alelos se presentó en Cajamarca y Junín con un total de 39 alelos para cada una, la menor cantidad se dio en la región de Ucayali con un total de 28 alelos. Dentro del *outgroup* se encontró 17 alelos para *C. canephora*, 19 para *C. liberica* y 23 para el híbrido S.288. Huánuco y Ucayali son las regiones con menor número de alelos con 30 y 28 alelos respectivamente, esto puede deberse a la baja cantidad de individuos que contienen estas regiones, las cuales están conformadas por 19 y seis individuos respectivamente. La región Junín cuenta con 34 genotipos que corresponde al 20.19% de la población total y que posee la misma cantidad de alelos que la región de Cajamarca, la cual está conformada por un total de 52 genotipos y corresponden al 30.77% de la población total (Tabla 10).

**Tabla 8: Número de genotipos y alelos por grupos**

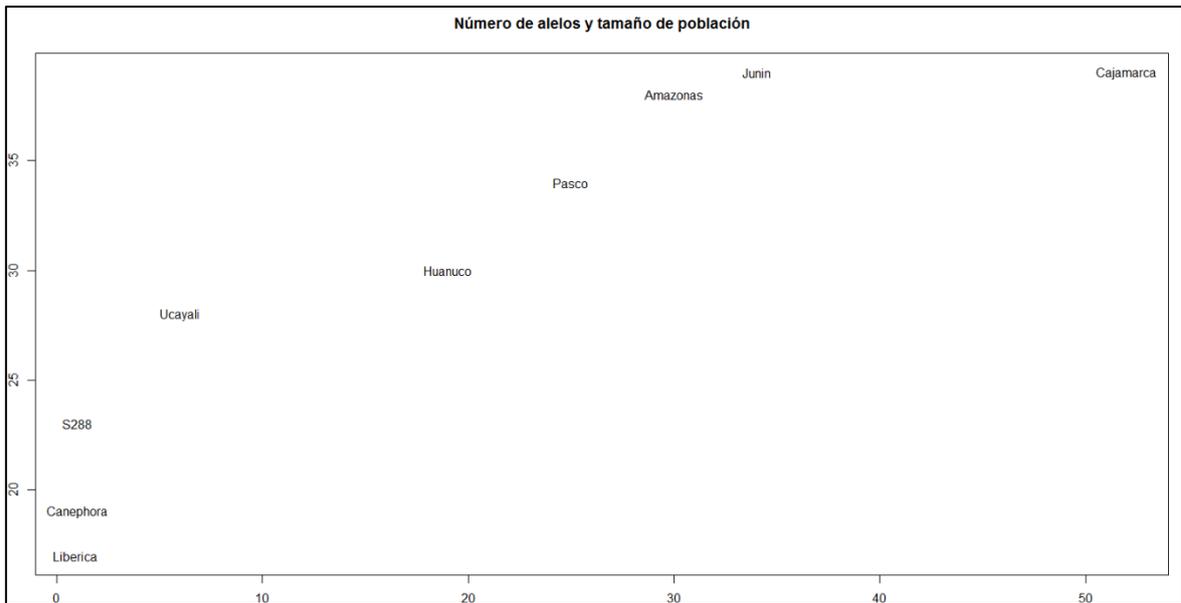
Grupos	Número de genotipos	Número de alelos
Cajamarca	52	39
Amazonas	30	38
Junín	34	39
Pasco	25	34
Ucayali	6	28
Huánuco	19	30
Canephora	1	19
S288	1	23
Liberica	1	17

Según lo reportado por Vieira (2010), los loci CarM105, CarM092 y CarM050 tendrían dos, tres y dos alelos por locus respectivamente. Estos resultados difieren a los obtenidos en el presente estudio donde se han encontrado seis, seis y cuatro alelos por cada locus y podría atribuirse a los alelos exclusivos encontrados en los individuos del *outgroup*.

El estudio realizado por Sánchez (2017) en Costa Rica, indicó que algunos de los marcadores que empleó en su investigación tuvieron altos números de alelos por locus a pesar de la cantidad de individuos empleados. Algunos de los marcadores con mayor cantidad de alelos son: M24 con ocho alelos, M32 con ocho alelos, SSR03 con ocho alelos. Por otro lado, en el presente estudio esos marcadores reportaron cantidades de siete, siete y cinco. La principal explicación respecto a la gran cantidad de alelos por locus encontrada por Sánchez (2017) en solo 54 muestras, se basa en que la mayor parte de las muestras empleadas en dicho estudio corresponden a híbridos.

Otro estudio sobre diversidad de café en la región peruana, específicamente con genotipos provenientes del distrito de Villa Rica y empleando marcadores tipo SSR, realizado por Whiter (2019) reportó seis alelos para el marcador M24, tres para el marcador CaM16 y tres para el marcador SSR03, estos valores son cercanos a los obtenidos en la presente investigación, ya que con el marcador M24 se observaron siete alelos, con el marcador CaM16 se observaron cuatro alelos, y con el marcador SSR03 se observaron cinco alelos.

Se realizó un gráfico correspondiente al número de alelos en el eje “X” y el tamaño de población en el eje “Y”, en dicho gráfico se puede observar a las regiones con mayor cantidad de alelos, las cuales son: Junín y Cajamarca con 39 alelos cada una, Amazonas con 38 alelos, Pasco con 34 alelos, Huánuco 30 alelos y Ucayali con 28 alelos (Figura 15).



**Figura 15: Gráfico representado el número de alelos y la cantidad de individuos por región**

#### 4.5.2. Frecuencias alélicas

A diferencia de los individuos diploides donde las frecuencias de alelos codominantes se pueden obtener por el recuento de los distintos genotipos; en individuos poliploides, la posibilidad de más de dos alelos por locus y distintos patrones de herencia, ocasionan inconvenientes al estimar las frecuencias alélicas, por lo cual se requiere emplear algoritmos para inferirlas, y de acuerdo a lo planteado por De Silva *et al.*, (2005), se obtuvieron los siguientes resultados para cada locus evaluado (los gráficos de estas se ubican en el Anexo 1):

##### a. CaM16

Para el locus CaM16, se obtuvieron los alelos de tamaños: 204, 206, 210 y 220 pb. La frecuencia más alta la tiene el alelo 204 con 0.46 y la más baja el 206 con 0.04. Vale recalcar que, para este locus, se obtuvo un alelo exclusivo para *C. liberica* con tamaño 220 pb.

Sánchez (2017) reportó alelos de tamaño 184, 188, 191, 193, 197 y 200. La mayor frecuencia fue de 0.54, mientras que la más baja fue de 0.0202 dicha frecuencia no causaría que sea catalogado como alelo raro.

**b. CaM38**

Para el locus CaM38, se obtuvieron los alelos de tamaños: 242, 244, 248 y 262. La frecuencia más alta la tiene el alelo 242 con 0.69 y la más baja, el alelo 244 con 0.04. El genotipo *C. canephora* mostró el alelo 248 y el *C. liberica* el 262.

**c. CaM40**

Para el locus CaM40, se obtuvieron los alelos de tamaños: 249, 250 y 252. La frecuencia más alta la tiene el alelo 249 con 0.71. El genotipo *C. canephora* mostró el alelo 252 y el *C. liberica* el 250.

**d. CaM46**

Para el locus CaM46, se obtuvieron los alelos de tamaños: 224, 238, 239 y 241. La frecuencia más alta la tiene el alelo 241 con 0.64. El genotipo *C. canephora* mostró el alelo 238 y el *C. liberica* no mostró alelo alguno.

**e. CarM050**

Para el locus CarM50, se obtuvieron los alelos de tamaños: 294, 300, 308 y 320. La frecuencia más alta la tiene el alelo 300 y 308 con 0.45. El genotipo el *C. liberica* mostró los alelos 294 y 320.

**f. CarM092**

Para el locus CarM092, se obtuvieron los alelos de tamaños: 450, 458, 462, 463, 468 y 488. La frecuencia más alta la tiene el alelo 468 con 0.53. El genotipo *C. canephora* mostró el alelo 458 y el *C. liberica* mostró el alelo 463.

**g. CarM105**

Para el locus CarM105, se obtuvieron los alelos de tamaños: 290, 291, 292, 296, 303 y 306. La frecuencia más alta la tiene el alelo 292 con 0.45. El genotipo *C. canephora* mostró los alelos 291 y 303, *C. liberica* mostró el alelo 290.

#### **h. M20**

Para el locus M20, se obtuvieron los alelos de tamaños: 289, 292, 296, 302, 306, 310, 316 y 330. La frecuencia más alta la tiene el alelo 292 con 0.25. El genotipo *C. canephora* mostró el alelo 316 y el *C. liberica* el alelo 289.

#### **i. M24**

Para el locus M24, se obtuvieron los alelos de tamaños: 167, 170, 172, 176, 180, 182 y 190. La frecuencia más alta la tiene el alelo 182 con 0.31. El genotipo *C. canephora* mostró el alelo 176 y el *C. liberica* mostró el alelo 172.

Sánchez (2017) reportó alelos de tamaño 149, 151, 153, 188, 160, 161, 164, 166, 167, 169. La mayor frecuencia fue de 0.40, mientras que la más baja fue de 0.0050 considerado como alelo raro ( $<0.01$ ).

#### **j. M32**

Para el locus M32, se obtuvieron los alelos de tamaños: 118, 119, 122, 125, 126, 143, 145 y 149. La frecuencia más alta la tiene el alelo 119 con 0.48. El genotipo *C. canephora* mostró el alelo 118 y el *C. liberica* mostró los alelos 122 y 126.

Sánchez (2017) reportó alelos de tamaño 99, 102, 106, 114, 116, 120, 123, 127, 131. La mayor frecuencia fue de 0.61, mientras que la más baja fue de 0.0052 considerado como alelo raro ( $<0.01$ ).

#### **k. M753**

Para el locus M753, se obtuvieron los alelos de tamaños: 307, 309, 311, 315, y 317. La frecuencia más alta la tiene el alelo 307 con 0.45. El genotipo *C. liberica* mostró el alelo 309.

Sánchez (2017) reportó alelos de tamaño 283, 289, 290, 292, 296. La mayor frecuencia fue de 0.467, mientras que la más baja fue de 0.0457 dicha frecuencia no causaría que sea catalogado como alelo raro.

### **l. SSR03**

Para el locus SSR03, se obtuvieron los alelos de tamaños: 153, 158, 160, 165, y 175. La frecuencia más alta la tiene el alelo 153 con 0.35. El genotipo *C. liberica* mostró el alelo 175.

Sánchez (2017) reportó alelos de tamaño 134, 139, 141, 143, 145, 146, 148, 149, 153, 154, 156. La mayor frecuencia fue de 0.326, mientras que la más baja fue de 0.0049 considerado como alelo raro (<0.01).

### **m. SSRCa18**

Para el locus SSRCa18, se obtuvieron los alelos de tamaños: 117, 130, 132, 137, 138 y 143. La frecuencia más alta la tiene el alelo 138 con 0.48. El genotipo *C. canephora* mostró los alelos 117 y 137, *C. liberica* mostró el alelo 132.

Sánchez (2017) reportó alelos de tamaño 110, 111, 112, 118, 119, 124, 126. La mayor frecuencia fue de 0.495, mientras que la más baja fue de 0.0102 dicha frecuencia no causaría que sea catalogado como alelo raro.

Los valores de frecuencias alélicas obtenidos en el presente estudio, muestran diferencias con algunos de los valores obtenidos por Sánchez (2017) con el mismo marcador, esto puede deberse a distintos factores siendo uno de los más relevantes la diferencia del origen de las muestras, así como también el empleo de una cola M13, el protocolo de electroforesis capilar, entre otros.

Dentro del grupo total de muestras empleadas en este estudio, se encontraron 11 alelos que son considerados como raros, debido a que cuentan con una frecuencia menor a 0.05; asimismo, se observó la presencia de 32 alelos considerados muy raros, estos con una frecuencia menor a 0.01. Una de las principales causas para la considerable cantidad de alelos con muy baja frecuencia, se debe a que son exclusivos de los individuos del *outgroup*, por lo tanto, muchos de estos no van a estar presentes en los individuos de arábica.

### 4.5.3. Loci polimórficos

Se comprobó que de los 13 loci estudiados, cuatro de ellos no serían polimórficos y esos corresponden a los loci: CaM40, CaM46, CarM092 y CaM38 (Tabla 11). Los nueve loci restantes corresponderían al grupo de loci polimórficos en los cuales se evaluó el nivel informativo de los mismos determinando en cada locus el índice de contenido polimórfico (PIC) y asimismo de estos en cada región de muestreo y en el *outgroup* (Tabla 12).

**Tabla 9: Loci polimórficos en la población evaluados mediante la función isPoly del paquete estadístico Adegenet**

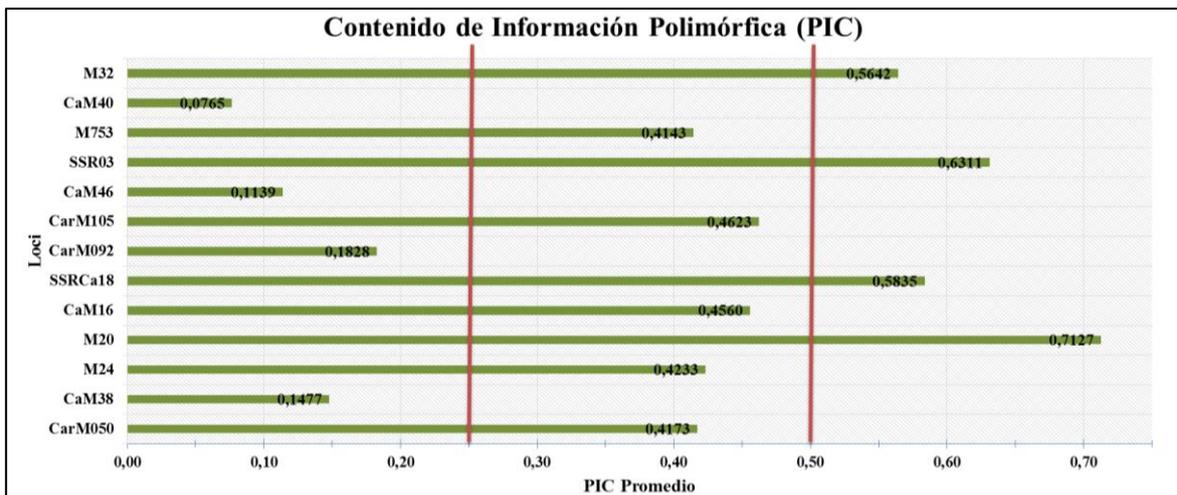
Nº	Locus	Polimorfismo
1	M32	Verdadero
2	CaM40	Falso
3	M753	Verdadero
4	SSR03	Verdadero
5	CaM46	Falso
6	CarM105	Verdadero
7	CarM092	Falso
8	SSRCa18	Verdadero
9	CaM16	Verdadero
10	M20	Verdadero
11	M24	Verdadero
12	CaM38	Falso
13	CarM050	Verdadero

**Tabla 10: Valores de índice de contenido polimórfico (PIC) por loci, por región y promedios, hallados mediante el paquete estadístico Polysat y el programa PolyGene**

Nº	Loci	Cajamarca	Amazonas	Junín	Pasco	Ucayali	Huánuco	Outgroup	PIC Promedio Polysat	PIC PolyGene	Nivel de Inf.
1	M32	0,5451	0,5355	0,5241	0,5260	0,5355	0,5394	0,7438	0,5642	0,5	Alto
2	CaM40	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,5355	0,0765	0,117	Bajo
3	M753	0,4290	0,4141	0,3963	0,3750	0,3750	0,3750	0,5355	0,4143	0,396	Medio
4	SSR03	0,6251	0,6078	0,6707	0,6594	0,5926	0,5926	0,6695	0,6311	0,623	Alto
5	CaM46	0,0370	0,1638	0,1480	0,0739	0,0000	0,0000	0,3750	0,1139	0,362	Bajo
6	CarM105	0,4276	0,4415	0,3963	0,3750	0,4768	0,3750	0,7438	0,4623	0,415	Medio
7	CarM092	0,2243	0,2941	0,1094	0,0000	0,0000	0,2772	0,3750	0,1828	0,436	Bajo
8	SSRCa18	0,5565	0,5478	0,5576	0,5576	0,5786	0,5427	0,7438	0,5835	0,524	Alto
9	CaM16	0,4276	0,3990	0,5185	0,4918	0,3750	0,4445	0,5355	0,4560	0,426	Medio
10	M20	0,7216	0,7337	0,7031	0,7031	0,7031	0,7117	0,7124	0,7127	0,724	Alto
11	M24	0,5036	0,4478	0,5431	0,3784	0,2392	0,2306	0,6204	0,4233	0,531	Medio
12	CaM38	0,0725	0,0322	0,1676	0,1692	0,0000	0,0000	0,5926	0,1477	0,313	Bajo
13	CarM050	0,3750	0,3750	0,3750	0,3750	0,3750	0,3750	0,6713	0,4173	0,384	Medio

Outgroup = Grupo conformado por *C. canephora*, *C. liberica* y el híbrido S288.

Nivel de Inf. = Nivel de Información



**Figura 16: Estimación del PIC promedio en población total incluido el outgroup**

*Nota.* Las líneas de corte separan los valores considerados bajos, medios y altos.

Con los resultados de los índices de contenido polimórfico (PIC) se determinó que dentro del grupo de loci empleados, los que cuentan con un nivel bajo de polimorfismo son: CaM40, CaM46, CarM092 y CaM38, ya que tienen valores menores a 0.25. Los que cuentan con un nivel de polimorfismo medio son M753, CarM105, CaM16, M24 y CarM050 con valores en el rango de 0.25 a 0.50. Finalmente, los loci con un PIC alto son M32, SSR03, SSRCa18 y M20 ya que cuentan con valores superiores a 0.50, esto demuestra que este último grupo de primers mencionados posee una muy buena capacidad de discriminación con el grupo de muestras empleadas (Figura 16).

Los valores estimados de PIC para el locus M753 es de 0,4143 y para el locus SSRCa18 de 0,5835, cercanos a los reportados por Sánchez (2017) donde menciona valores de 0,4635 y 0,5758 respectivamente para una población de arábicas. Asimismo, el valor del PIC para el locus SSR03 es de 0,6311, cercano al reportado por Sánchez (2017) con 0,6712.

Con respecto a los valores de PIC para el locus CaM16 es de 0,4560 y se acerca a lo obtenido por Hendre *et al.* (2008) con 0,49. Por otro lado, los valores de PIC de los loci CaM40, CaM46 y CaM38 contrastan con lo reportado por Hendre *et al.* (2008) ya que en su trabajo indica que estos loci poseen valores de PIC que los harían ser considerados altamente polimórficos en la población estudiada; sin embargo, en el presente estudio se reporta que para la población trabajada los valores del PIC para dichos loci serían considerados con un bajo nivel de polimorfismo. Este contraste significativo en los valores del PIC para este

grupo de loci podría atribuirse al grupo de individuos con el cual se ha trabajado debido a que las muestras evaluadas por Hendre *et al.* 2008 corresponden a individuos pertenecientes al banco de germoplasma de café del Central Coffee Research Institute (CCRI - India), mientras que las muestras empleadas en el presente estudio corresponden al banco de germoplasma de café del INIA.

#### 4.5.4. Alelos Nulos

Tomando en consideración valores de frecuencia mayores a 0.19, se encontró que, de los 13 locus trabajados, cinco de ellos poseían una frecuencia superior al umbral, este grupo de loci son CaM38, CaM40, CaM46, CarM092 y M24 (Tabla 13). (Razafinarivo *et al.*, 2013). Los alelos nulos son aquellos que no amplifican durante la PCR ya que pueden presentar mutaciones en la región que flanquea la secuencia. Su presencia en la población puede ocasionar un déficit de heterocigotos (Tripp, 2009).

La frecuencia de alelos nulos puede ser considerada como un indicador de garantía sobre los resultados de análisis de microsatélites. Esta frecuencia se calcula en base a algoritmos que emplean frecuencias esperadas y observadas de los genotipos evaluados para sus cálculos numéricos (Van Oosterhout *et al.*, 2004 citado por Acuña, 2016). Además, Barahona (2014) postula que los rangos de probabilidad de alelos nulos se encuentran en rangos que van desde: Prob. A.N < 0.05 (insignificante),  $0.05 \leq \text{Prob. A.N} \leq 0.20$  (moderada), Prob. A.N > 0.2 (alta).

**Tabla 11: Valores de frecuencia de alelos nulos con los marcadores empleados**

Nº	Loci	Frecuencia <sup>a</sup>	Frecuencia <sup>b</sup>
1	CaM16.null	0.1094572	0.0
2	CaM38.null	0.2722633	0.178
3	CaM40.null	0.2832113	0.056
4	CaM46.null	0.3351903	0.301
5	CarM050.null	0.1059015	0.0
6	CarM092.null	0.4264495	0.456
7	CarM105.null	0.113516	0.0
8	M20.null	0.00684396	0.001
9	M24.null	0.5351034	0.560
10	M32.null	0.1217752	0.0
11	M753.null	0.1062501	0.0
12	SSR03.null	0.04614497	0.0
13	SSRCa18.null	0.1224814	0.019

Frecuencia<sup>a</sup>: Hallada con el paquete estadístico Polysat

Frecuencia<sup>b</sup>: Hallada con el programa PolyGene

## 4.6. Estructura genética

### 4.6.1. Heterocigosidad

El cálculo de la heterocigosidad esperada y observada fue estimado para cada locus y en el total de la población estudiada (Tabla 14).

Los valores de heterocigosidad observada obtenidos, se muestran muy cercanos a 1 cuando fueron evaluados mediante el paquete estadístico Adegnet, esto posiblemente se debe a que el modelo empleado no sería el más adecuado para el caso de un organismo poliploide como lo es el café, sin embargo, al determinar los valores de heterocigosidad mediante el paquete estadístico PolyGene, el cual es un programa para análisis de datos en especies poliploides y con genotipos ambiguos, es decir, cuando no se puede determinar el número de copias alélicas, los valores de heterocigosidad observada se encontraron en valores más acordes a lo que se esperaría para un organismo poliploide.

**Tabla 12: Valores de Heterocigosidad observada y esperada.**

Nº	Loci	Ho - 1	He - 1	Ho - 2	He - 2
1	M32	0,9941	0,6222	0,564	0,566
2	CaM40	0,0059	0,0176	0,11	0,122
3	M753	0,9941	0,5217	0,568	0,513
4	SSR03	0,9941	0,7085	0,799	0,686
5	CaM46	0,0000	0,1037	0,35	0,446
6	CarM105	0,9941	0,5366	0,565	0,525
7	CarM092	0,0000	0,2301	0,48	0,54
8	SSRCa18	0,9882	0,6352	0,58	0,581
9	CaM16	0,9941	0,5567	0,565	0,528
10	M20	0,9882	0,7650	0,989	0,765
11	M24	0,1012	0,5245	0,56	0,592
12	CaM38	0,1018	0,1191	0,283	0,353
13	CarM050	1,0000	0,5059	0,571	0,506
	<b>Prom.</b>	<b>0,6274</b>	<b>0,4498</b>	<b>0,5372</b>	<b>0,5172</b>

Ho - 1 y He - 1: Hallada con el paquete estadístico Adegnet

Ho - 2 y He - 2: Hallada con el programa PolyGene

Se realizó un segundo análisis de la heterocigosidad retirando los loci cuya frecuencia de alelos nulos sea mayor a 0.15 según los datos mostrados en la tabla 15, para esto se retiraron los loci CaM38, CaM40, CaM46, CarM092 y M24. Con estos ajustes se obtuvo un valor de He promedio de 0,595 en la población total y la Ho presento un valor mayor que la He (Tabla 15).

La menor heterocigosidad observada fue obtenida con el marcador M24 con el que se encuentran menor número de individuos heterocigotos, y M20 es el marcador con mayor heterocigosidad observada pues con este marcador la vasta mayoría de individuos presentaron 4 alelos distintos.

**Tabla 13: Valores de heterocigosidad sin marcadores con alta frecuencia de alelos nulos**

Nº	Loci	Heterocigosidad Observada	Heterocigosidad Esperada
1	M32	0,564	0,5660
2	M753	0,568	0,513
3	SSR03	0,799	0,686
4	CarM105	0,565	0,525
5	SSRCa18	0,58	0,581
6	CaM16	0,565	0,528
7	M20	0,989	0,765
8	M24	0,56	0,592
	Promedio	0,649	<b>0,595</b>

En el análisis de heterocigosidad sin los loci con un índice de alelos nulos mayor a 0.15, el promedio de la He es mayor que cuando se halla con todos los loci trabajados. Los valores de heterocigosidad esperada son altos y se encontraron en un rango entre 0,513 y 0,765, con un promedio de 0,595. Esto muestra que la probabilidad de que dos alelos seleccionados al azar sean diferentes es mayor al 50%. Este valor promedio de heterocigosidad esperada es similar al reportado por el estudio en café en Costa Rica realizado por Sánchez (2017) (He= 0,5885).

#### 4.6.2. Índice de Fijación - Fst

Se realizó la estimación del índice de fijación Fst que mide el grado de diferenciación genética, tomando como subpoblaciones a las regiones de origen de las muestras de café (Cajamarca, Amazonas, Junín, Pasco, Ucayali, Huánuco) y un *outgroup* conformado por *Coffea liberica*, *Coffea canephora* y el híbrido S288 (Tabla 16). Se tomó en cuenta la interpretación de los rangos de Fst determinados por de Vicente *et al.* (2004) quienes mencionan que valores entre 0 – 0.05 representan una pequeña diferenciación, valores de 0.05 – 0.15 una diferenciación moderada, de 0.15 – 0.25 una diferenciación grande y valores mayores a 0.25 una diferenciación muy grande.

**Tabla 14: Tabla de índices Fst entre las regiones trabajadas**

Fst	Cajamarca	Amazonas	Junín	Pasco	Ucayali	Huánuco	Outgroup
Cajamarca	0	0,0014	0,0048	0,0041	0,0032	0,0065	0,0389
Amazonas	0,0014	0	0,0078	0,0068	0,0057	0,0078	0,0577
Junín	0,0048	0,0078	0	0,0038	0,0071	0,0092	0,0533
Pasco	0,0041	0,0068	0,0038	0	0,0066	0,0090	0,0730
Ucayali	0,0032	0,0057	0,0071	0,0066	0	0,0070	0,1509
Huánuco	0,0065	0,0078	0,0092	0,0090	0,0070	0	0,0877
Outgroup	0,0389	0,0577	0,0533	0,0730	0,1509	0,0877	0

Se puede observar que los valores de Fst menores a 0.05 corresponderían a una poca diferenciación entre las regiones. Sin embargo, los valores Fst de la mayoría de las regiones con el *outgroup* muestran valores mayores a 0.05, lo cual correspondería a una diferenciación moderada, esto puede ser explicado debido a que el *outgroup* posee genotipos correspondientes a otras especies del género *Coffea* y un híbrido.

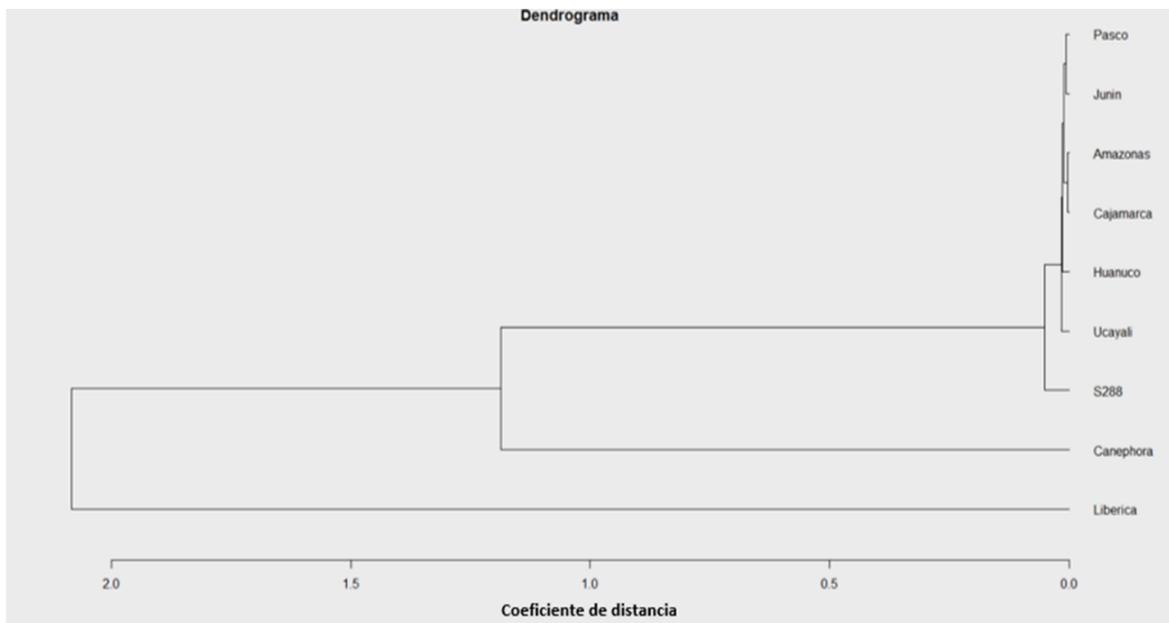
Por otro lado, el Fst de la región de Cajamarca con el *outgroup* toma un valor de 0,039 lo cual significaría poca diferenciación, esto puede ser explicado porque esta región es la que posee mayor cantidad de genotipos y también una de las que poseen mayor número de alelos. Finalmente, el valor de Fst entre la región de Ucayali y el *outgroup* es de 0.151, lo cual evidencia una diferenciación grande, y podría comprenderse por la poca cantidad de genotipos que conforman esta región y ser la que posee menor cantidad de alelos. Los bajos valores de Fst obtenidos podrían ser un indicativo de una alta tasa de flujo génico.

#### 4.6.3. Análisis de conglomerados

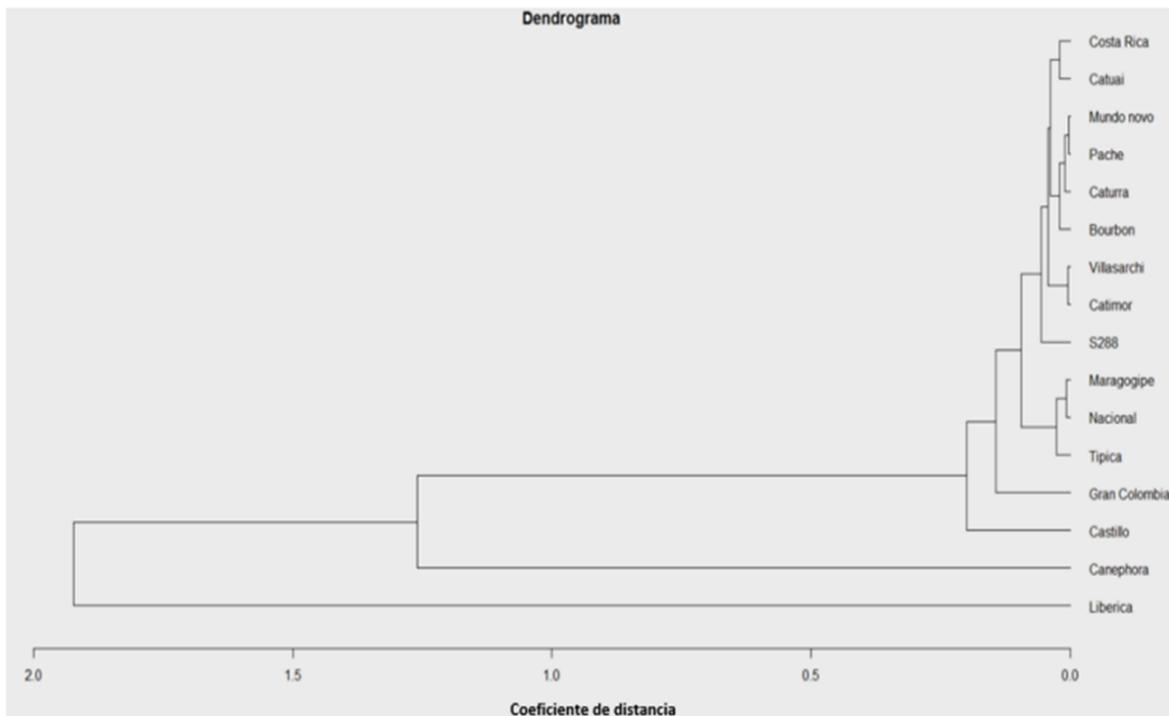
Se realizó un dendrograma empleando la distancia de Nei y el método de agrupamiento UPGMA, correspondiente a las regiones de origen de las muestras trabajadas, éste mostró una cercanía entre Pasco con Junín y entre Amazonas con Cajamarca (Figura 17), se debe considerar que estas regiones son colindantes geográficamente, además los puntos de colecta de las muestras para las regiones de Pasco y Junín son bastante cercanos (Figura 9).

Adicionalmente, se realizó un dendrograma con la distancia de Nei y el método de agrupamiento UPGMA correspondiente a las denominaciones de variedades de los genotipos evaluados. En este caso se tuvo que retirar a los individuos que no poseían

información correspondiente a la denominación de variedad dado que no podrían agruparse en un solo grupo por ser posiblemente genotipos distintos y de esta manera el agruparlos podría causar un sesgo en el análisis de los otros grupos. Debido al retiro de los NN para este análisis, los individuos de las regiones de Pasco y Junín no formaron parte del análisis (Figura 18).



**Figura 17: Dendrograma de regiones productoras de café evaluadas**



**Figura 18: Dendrograma de variedades con distancia de Nei y agrupamiento UPGMA**

Se observó cercanía entre los grupos denominados como Costa Rica y Catuai, se debe considerar que la variedad Costa Rica 95 corresponde a un catimor el cual se generó a partir de un Híbrido de Timor x Caturra, mientras que la variedad Catuai proviene del cruce Mundo novo x Caturra, compartiendo así a uno de los parentales y siendo ésta una de las posibles explicaciones para su cercanía.

También se observó que no hubo distancia entre los grupos denominados como Mundo novo con Pache. Mundo novo proviene del cruce de Típica x Bourbon, mientras que Pache es una mutación natural de Típica. Por lo cual ambas variedades, Mundo novo y Pache, tendría como parte de su origen a Típica lo cual podría ser la razón de mostrar tanta cercanía.

El grupo denominado Maragogipe denotó cercanía con el grupo denominado como Nacional, el cual no corresponde a variedad alguna, sin embargo, CENICAFE s.f menciona que en Colombia a la variedad Típica también se le llama arábico, pajarito o nacional. La variedad Maragogipe tiene origen en una mutación natural de Típica. Considerando que el grupo denominado como Nacional podría estar compuesto por individuos de la variedad Típica, esta sería la razón para la cercanía.

Los grupos denominados como Castillo y Gran Colombia se muestran separados, pero considerando que las variedades de Castillo y Colombia tienen como origen cruces entre Caturra e Híbrido de Timor, es decir son catimores, es de esperar que tengan mayor cercanía entre ellos y con el grupo denominado como Catimor. Esto puede deberse a que la denominación de variedad empleada no es la correcta o el set de marcadores empleados no permite generar mayor cercanía entre los grupos mencionados.

Un tercer dendrograma fue generado en base la distancia euclidiana y agrupamiento UPGMA correspondiente a un análisis a nivel de individuos. Esta muestra baja diversidad genética entre las variedades de *C. arabica*. No se logra observar agrupamiento entre individuos de la misma variedad, lo que se observó fue agrupación de mezclas de individuos de distintas regiones y variedades los cuales tienen una distancia de cero ya que presentaron el mismo perfil genético con el set de marcadores empleados en este estudio. Esto puede tener una explicación por una falta de pureza varietal o porque los marcadores empleados no

son suficientes para realizar dicha discriminación (Figura 19). En este análisis, la mayor parte de los individuos más distanciados corresponden a genotipos provenientes principalmente de Junín, Amazonas y Cajamarca con denominaciones varietales como: Bourbon, Caturra, Catimor, Gran Colombia y NN (sin denominación varietal).

Muchas veces la denominación varietal otorgada por el agricultor no corresponde a la correcta. Esto se observa en estudios de caracterización morfológica como el realizado por Blas *et al.*, 2011 donde se reportó que en el 26% de las entradas la denominación otorgada por el agricultor no sería la correcta y/o se podrían tratar de mezclas de variedades. Por otro lado, los individuos *C. canephora* y *C. liberica* presentaron mayor diferenciación genética con los individuos de arábica. Esto debido a que ambos poseen perfiles genéticos distintos a los de arábica gracias a la presencia de alelos exclusivos para estos individuos.

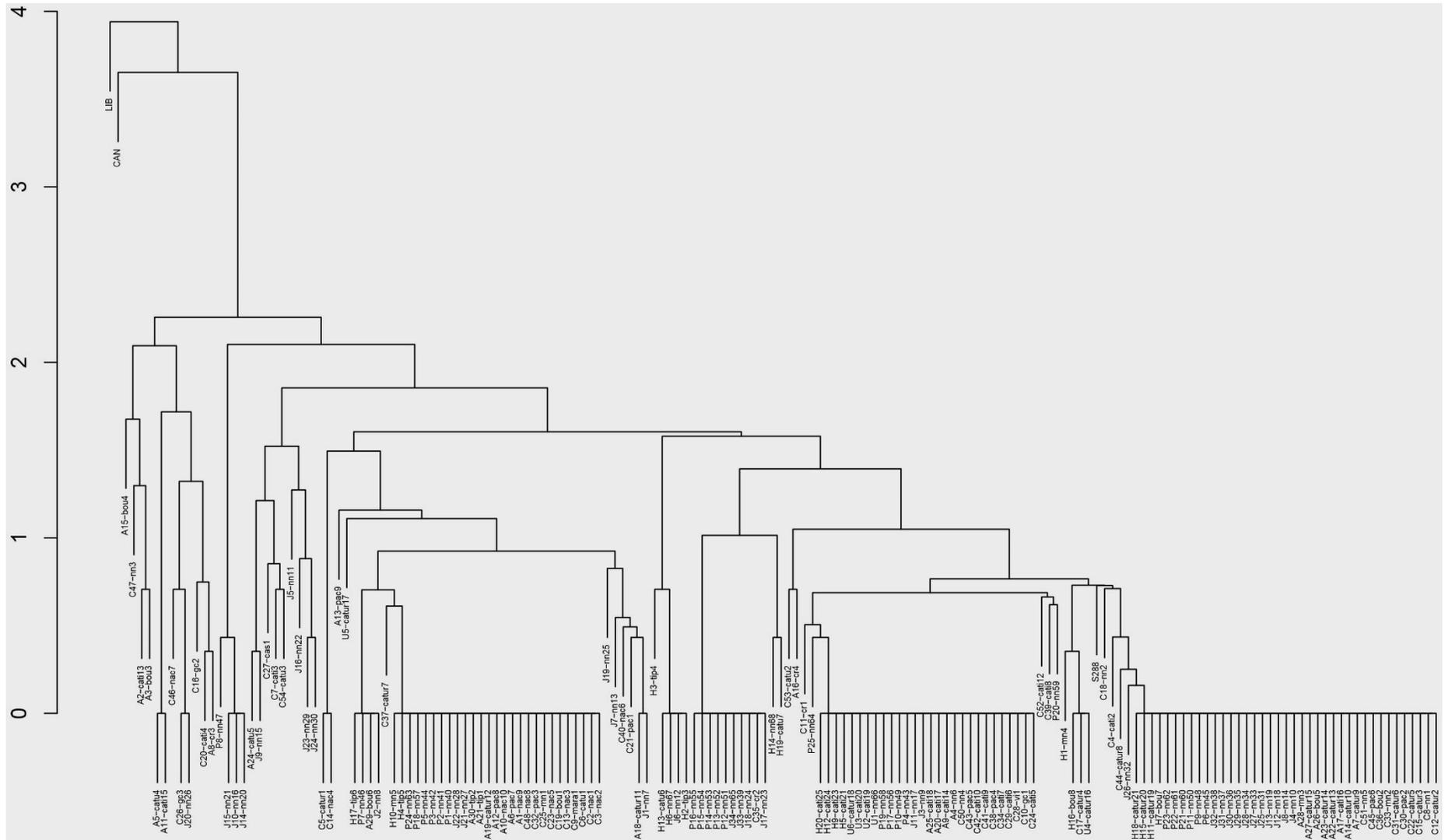


Figura 19: Dendrograma a nivel de individuos, con distancia euclidiana y agrupamiento UPGMA

Resultados de baja diversidad genética en café también han sido reportados como el de Lashermes *et al.* (1996), donde menciona que las causas principales serían su reciente origen, su biología reproductiva autógama y la forma de dispersión desde su centro de origen. En otro trabajo realizado por Steiger *et al.* (2002), reportó una muy alta similitud genética en cultivares de *C. arabica*, aunque el marcador empleado en este caso fue AFLP. Por otro lado, Palomino *et al.* (2014), reportaron un bajo grado de diversidad genética y la falta de agrupamiento en base a variedades de café arábica con marcadores moleculares tipo RAPD.

#### **4.6.4. Análisis de componentes principales - PCA**

Se realizó un análisis de componentes principales a partir de una matriz de frecuencias alélicas individuales. Los resultados obtenidos a partir de este análisis no muestran agrupamiento o estructuración con respecto a las regiones de origen ni a la denominación de variedades (Figura 20). El gráfico de sedimentación muestra que los dos primeros componentes principales acumulan una varianza de 34,5% (Figura 21). Adicional al gráfico correspondiente a los dos primeros componentes, también se evaluaron los gráficos formados en base al componente uno con el tres y otro con el componente dos y el tres (Anexo 3). Determinando así la no agrupación específica de estos individuos.

Se observó la presencia de algunos genotipos que divergen del grupo más grande, entre ellos tenemos una agrupación de individuos sin denominación varietal conocida (NN) que pertenecen a las regiones de Junín y Ucayali: nn23, nn24, nn39, nn52, nn53, nn54, nn55 y nn65. También se presentan alejados algunos individuos catalogados como Castillo, Caturra, Catimor y Costa Rica 95, sería interesante evaluar si a nivel morfológico estos genotipos de café arábica también son los más alejados de la población.

Debido a la naturaleza autógama de *C. arabica*, es de esperar que se genere un bajo patrón de agrupamiento en base a las variedades. Además, el proceso por el cual las variedades han sido generadas puede en parte explicar la falta de una clara diferenciación entre grupos de variedades, como ejemplo se puede tomar a la variedad Caturra, la cual proviene de una mutación natural de la variedad Bourbon, por lo cual una diferenciación significativa entre ambas variedades podría no generarse debido al relativamente corto tiempo transcurrido que tiene la variedad Bourbon desde su generación. Factores como cierto grado de flujo génico entre variedades podrían contribuir en la estructuración de las mismas (Geleta *et al.*, 2012).

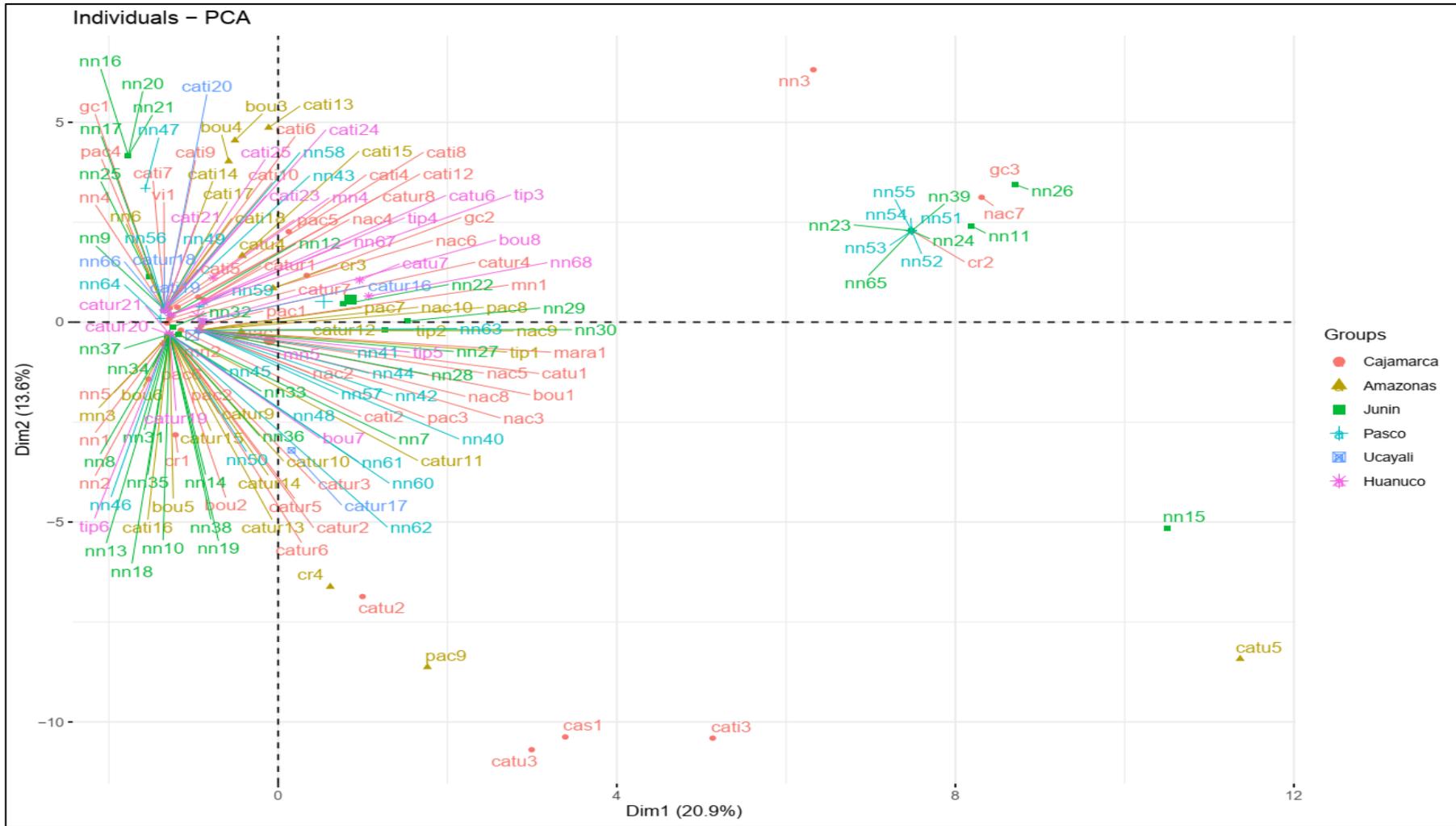
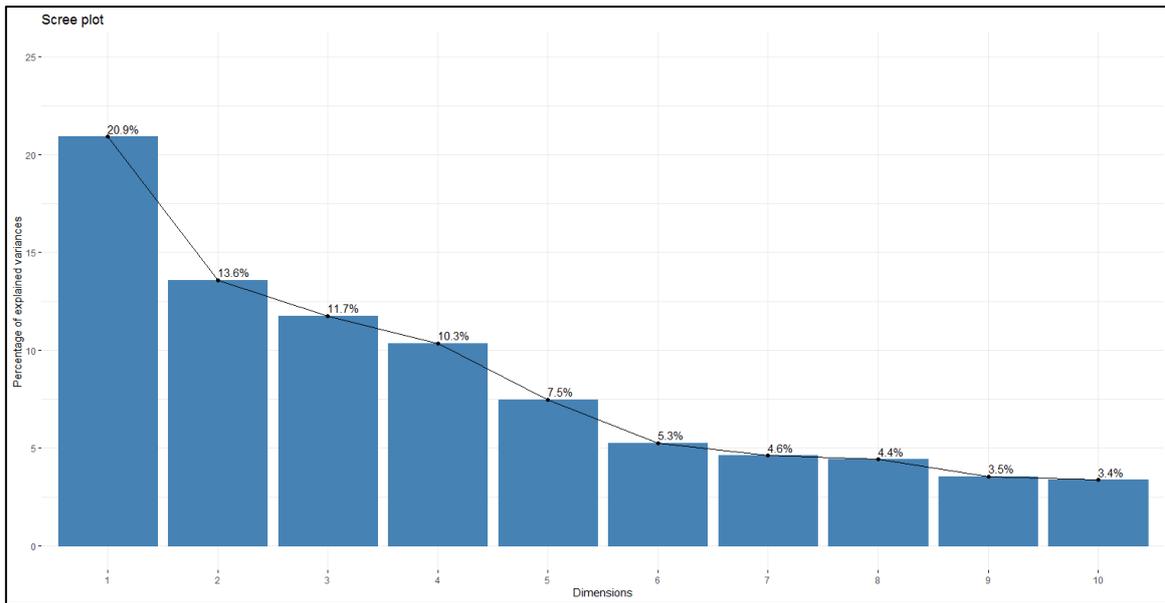


Figura 20: Análisis de componentes principales en base a las frecuencias alélicas de los individuos, mostrando los dos primeros componentes



**Figura 21: Gráfico de sedimentación mostrando el aporte de cada componente**

Otro enfoque de análisis multidimensional empleado con datos moleculares es el realizado con el uso del Análisis de Coordenadas Principales – PCoA, con el cual se puede realizar ordenamiento de datos originados por marcadores moleculares (Balzarini *et al.*, 2006).

Se realizó un diagrama de dispersión con las dos primeras coordenadas empleando una matriz de dispersión en base a la distancia genética de Lynch, la cual no se ve afectada por el nivel de ploidía de los individuos, genotipos ambiguos ni frecuencias alélicas (Vía y Rada, 2015), siendo todo lo anterior característico de las muestras que pertenecen al presente estudio (Anexo 4). Sin embargo, no se pudo identificar ningún tipo de agrupamiento consistente al origen de muestreo, ni tampoco a la denominación varietal (brindada por el agricultor). Asimismo, se corroboró lo obtenido con el Análisis de componente Principales – PCA.

El análisis de coordenadas principales (PCoA) evidenció una agrupación de las muestras correspondientes a café arábico y se pudo observar un aislamiento de *C. canephora* y *C. liberica* pudiendo así discriminar adecuadamente entre muestras de café arábico y no arábico. De la misma manera que el dendrograma de variedades y el de individuos, los resultados gráficos del PCoA respaldan una separación evidente entre las muestras de café arábico y no arábico.

#### 4.6.5. Análisis Discriminante de Componentes Principales - DAPC

En el análisis DAPC (Anexo 6), se encontró un valor de  $K=5$ , empleando el criterio “goodfit” en la búsqueda de clusters para determinar la mejor agrupación con el mínimo número de grupos formados. Se tomaron ocho componentes principales y cuatro funciones discriminantes. En la gráfica realizada se observó que los cluster cinco y el dos mostraron mayor distanciamiento de los otros cluster (uno, tres y cuatro).

En el cluster cinco se encuentran individuos de distintas denominaciones varietales, pero básicamente estos pertenecen a las regiones de Cajamarca y de Junín, finalmente en este cluster también se encontró al individuo *C. liberica*.

El cluster dos es uno de los más grandes, ya que está conformado por 44 individuos, de la misma manera que el cluster cinco, este grupo está conformado por individuos con distintas denominaciones de variedad, sin embargo, principalmente se encuentran individuos pertenecientes a las regiones de Cajamarca, Amazonas, Junín y Pasco.

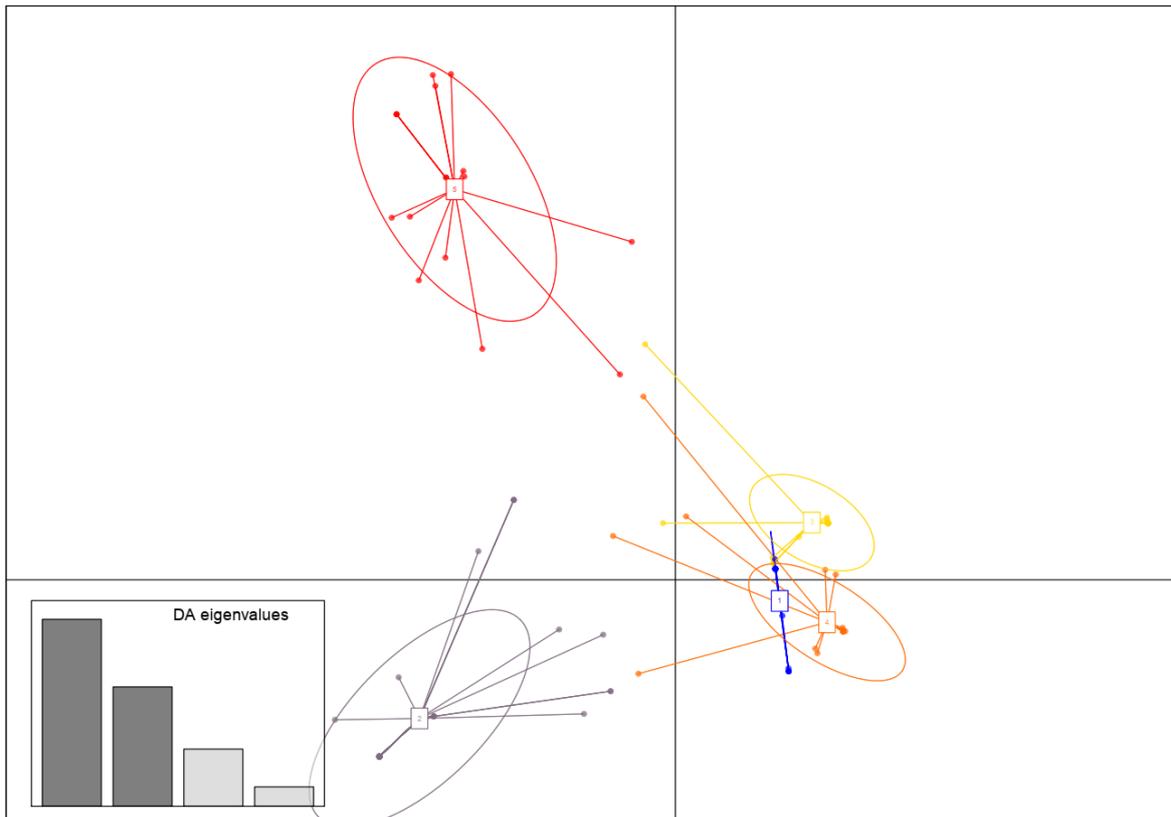
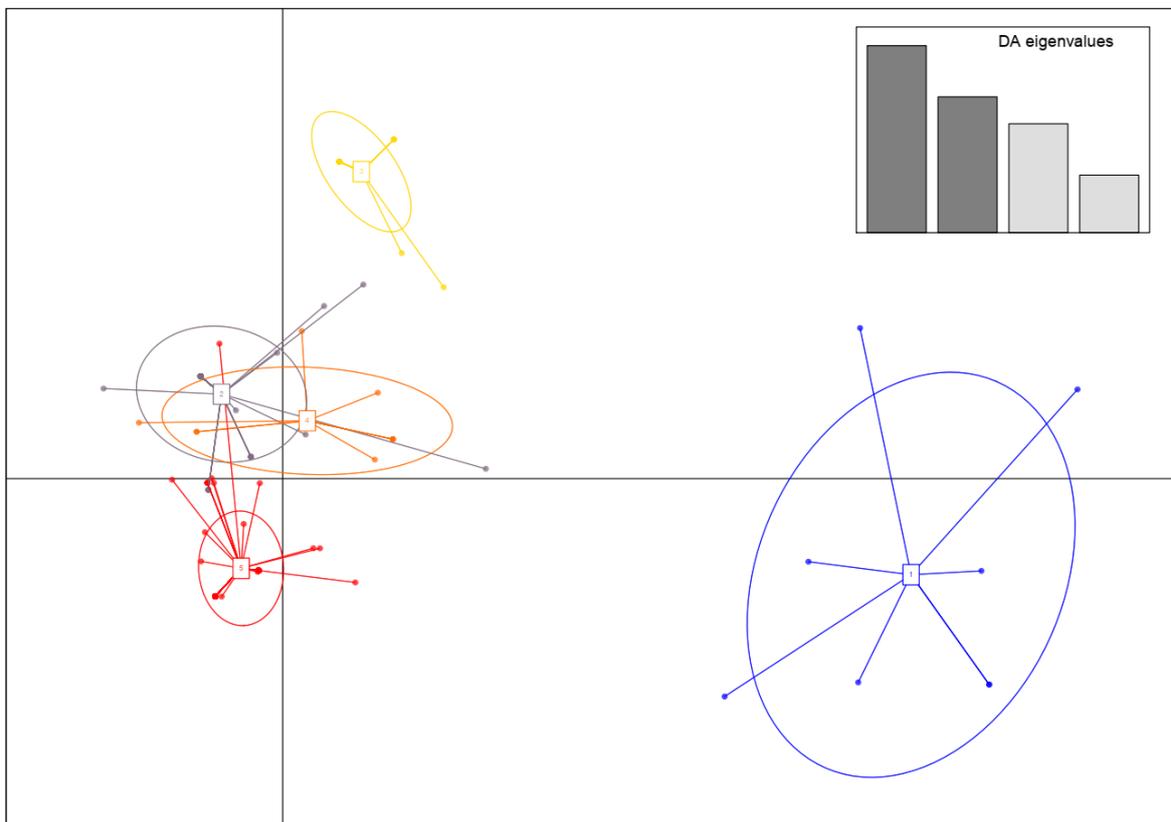


Figura 22: Análisis discriminante de componentes principales DAPC con 13 loci

Realizando un nuevo análisis DAPC retirando los locus con una elevada frecuencia de alelos nulos (Figura 23) y los mismos parámetros ya empleados en el anterior DAPC, se formaron cinco cluster, de ellos el más alejado es el cluster uno donde se encuentran la menor cantidad de individuos con solo siete y la mayoría de ellos pertenecen a las regiones de Cajamarca y Amazonas.

Otro grupo distanciado, pero en menor medida es el tres, donde se encuentran individuos de las regiones de Cajamarca, Junín y Pasco. Vale recalcar que, en este grupo, la gran mayoría de individuos no contaban con denominación de variedad alguna, es decir están etiquetados como NN.



**Figura 23: Análisis discriminante de componentes principales DAPC retirando cinco loci con frecuencia de alelos nulos elevada**

#### 4.6.6. Análisis de varianza molecular - AMOVA

El análisis de varianza molecular (AMOVA) realizado en base a poblaciones según la procedencia de los individuos, mostró que gran parte de la varianza genética perteneciente a los grupos evaluados se encontró dentro de las regiones, con valores superiores al 94% de la

varianza total, lo que indica una alta variación entre los individuos dentro de cada grupo (Tabla 17 y 18). Mientras que la varianza genética entre los grupos se presentó en valores entre 3.3% y 5.6%. Estos resultados nos permitieron observar que hay muy poca variación entre las regiones, esto podría tener explicación en la forma de distribución del germoplasma de una zona cafetalera hacia otra (Blas *et al.*, 2011).

IAM hace referencia a *infinite allele model* donde se asume que todos los alelos posibles comparten la misma probabilidad de ser formados en base a una mutación, mientras que SMM *stepwise mutation model* describe a los microsatélites como un proceso de mutación generado paso a paso añadiendo y retirando unidades de repetición individuales, también llamados motivos (Kosman & Jokela, 2019). En este análisis bajo ambos modelos (IAM - SMM) se pudo observar que la mayor parte de la totalidad de varianza estaría presente dentro de las poblaciones, es decir dentro de cada grupo de individuos según la región de procedencia.

**Tabla 15: Análisis de varianza molecular (AMOVA) de las regiones de origen de las accesiones de café, usando modelo IAM**

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Componentes de varianza	%
Dentro de Poblaciones	667	2125,723	3,187	3,187	94,397
Entre Poblaciones	8	127,469	15,934	0,189	5,603
Total	675	2253,193	3,338	3,376	100

**Tabla 16: Análisis de varianza molecular (AMOVA) de las regiones de origen de las accesiones de café, usando modelo SMM**

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Componentes de varianza	%
Dentro de Poblaciones	667	1423262,592	2133,827	2133,827	96,733
Entre Poblaciones	8	55916,144	6989,518	72,059	3,267
Total	675	1479178,736	2191,376	2205,886	100

Otro AMOVA fue realizado, y para este análisis se evaluó la varianza genética presente entre los grupos conformados por las variedades. Similarmente a lo obtenido en el anterior análisis, se mostró un valor muy alto dentro las poblaciones (variedades), en este caso valores de variabilidad superiores a 91% fueron los registrados (Tabla 19 y 20). Por otro

lado, los valores de variabilidad genética entre estas poblaciones fueron 7.2% y 8.4%, lo cual vuelve a demostrar una baja variabilidad entre los grupos formados por las denominaciones varietales. Los resultados obtenidos en el presente estudio difieren con los reportados por Whiter (2019), donde encontró un porcentaje de variabilidad entre variedades de 66.58%, se debe tener en cuenta que las muestras usadas en ese estudio solo corresponden a individuos del distrito de Villa Rica, mientras que en este estudio se emplearon muestras correspondientes a diferentes zonas cafetaleras del país, algunas de ellas corresponden a caficultores de pequeña escala los cuales comparten semillas de unas parcelas a otras.

Por lo tanto, el análisis AMOVA indicó que no existe una variabilidad genética considerable entre las variedades, posiblemente porque los marcadores seleccionados no permitirían encontrar diferencias moleculares entre las variedades o porque podría existir una considerable mezcla entre las variedades. Además de ello, se debe considerar que la denominación de variedad se dio en base a información proporcionada por el agricultor al momento de colecta y en algunos casos podría no corresponder a la correcta denominación. Por lo tanto, sería muy interesante el realizar una nueva evaluación de la data con información previa de una caracterización morfológica.

**Tabla 17: Análisis de varianza molecular (AMOVA) de las denominaciones varietales de las accesiones de café. Usando modelo IAM**

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Componentes de varianza	%
Dentro de Poblaciones	389	1174,85	3,02	3,02	91,648
Entre Poblaciones	14	137,876	9,848	0,275	8,352
Total	403	1312,726	3,257	3,295	100

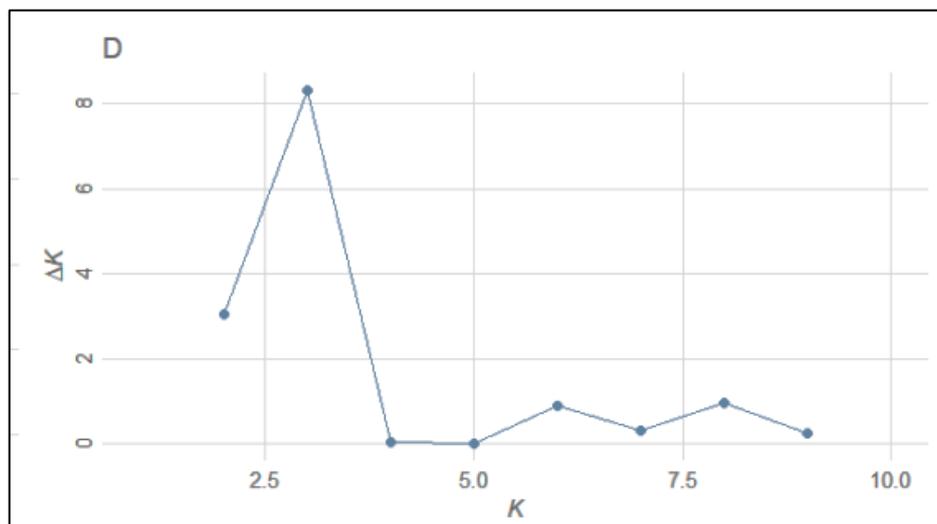
**Tabla 18: Análisis de varianza molecular (AMOVA) de las denominaciones varietales de las accesiones de café. Usando modelo SMM**

Fuente de variación	Grados de libertad	Suma de cuadrados	Cuadrados medios	Componentes de varianza	%
Dentro de Poblaciones	389	206177,165	530,018	530,018	92,804
Entre Poblaciones	14	21694,27	1549,591	41,097	7,196
Total	403	227871,435	565,438	571,115	100

#### 4.6.7. Análisis de estructura poblacional

En el análisis con el programa STRUCTURE, el número de poblaciones asumidas (K) varió de 1 a 10 (Anexo 6). Se pueden observar múltiples barras correspondientes a cada individuo, al tener un solo color en la barra esto nos indicaría una probabilidad de pertenencia de 100% al cluster designado, si se presentaría más de un color, este individuo compartiría genoma con otro cluster (Martinez, 2008).

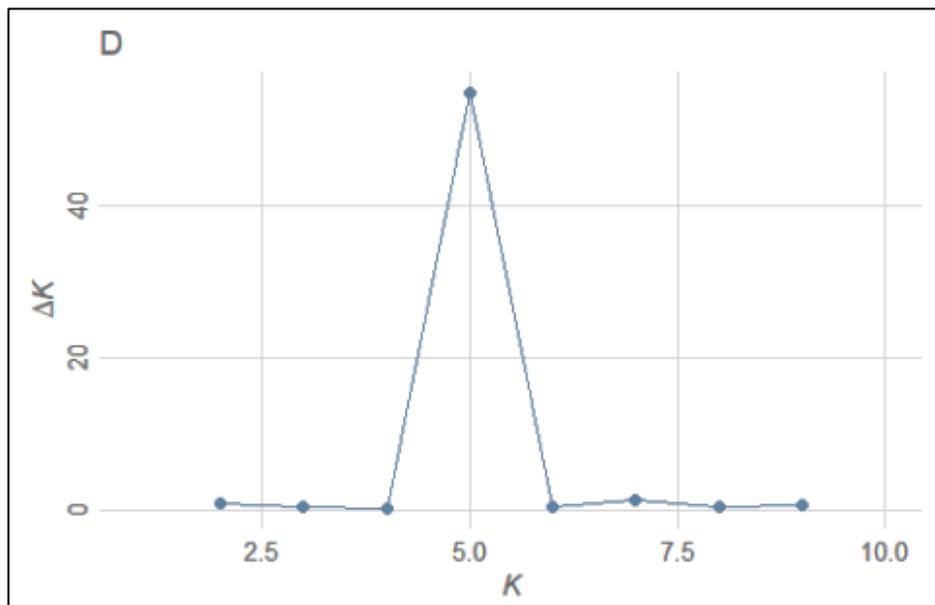
Un primer análisis fue realizado con los 13 loci trabajados, empleando una longitud de burn-in de 10,000 y un MCMC 100,000, con valores de K desde uno hasta diez y fueron repetidas diez veces cada una. El método Delta de K (Evanno *et al.*, 2005), permite identificar el valor óptimo de K que representaría la estructura poblacional bajo dichas condiciones. Para este caso el valor óptimo de K correspondería a tres (Figura 24).



**Figura 24: Curva de Delta de K para la determinación del óptimo valor de K en la población estudiada**

En este análisis se observó que la estructura de la población de café estudiada correspondería a un valor de  $K=3$  (Figura 26A). La mayor parte de los individuos del segundo cluster poseen una pertenencia cercana a 100%, se puede observar que los individuos correspondientes a *C. liberica* y *C. canephora* poseen una contribución de más de un clúster en su estructura genética bajo este análisis. Correspondiente a este enfoque bayesiano de análisis de estructura no se estaría observando una clara diferenciación genética entre las muestras de café arábico y no arábico, estos resultados difieren de los obtenidos por Whiter (2019), donde

con un  $k = 2$  y  $k = 3$  muestra una marcada diferenciación genética entre *C. arabica* y *C. canephora*. Este resultado de baja diferenciación genética, entre las muestras de café arábico y no arábico, obtenido en el presente análisis de estructura genética con un enfoque bayesiano, puede deberse a la cantidad de alelos nulos presentes en los loci trabajados. Por tal motivo, se planteó realizar un nuevo análisis tratando de mitigar el efecto que podrían estar produciendo los alelos nulos en este análisis. Un segundo análisis de estructura fue realizado, empleando los mismos parámetros que el primero, pero con variaciones en la cantidad de loci, retirando los loci con altas frecuencias de alelos nulos. Para este caso el método delta de K mostró que el valor óptimo de K correspondería a cinco (Figura 25).



**Figura 25: Curva de Delta de K determinando el óptimo valor de K en la población estudiada, correspondiente a K=5**

Los resultados obtenidos de este otro análisis de estructura genética con métodos bayesianos, pero esta vez retirando los loci con una frecuencia de alelos nulos mayor a 0.15, mostraron que al evaluar un  $K = 5$  (Figura 26B), los individuos pertenecientes al outgroup se encuentran casi exclusivamente en un solo cluster a excepción del individuo S288, el cual posee probabilidad de pertenencia a los 5 cluster, este resultado podría tener sustento en el origen del híbrido S288 el cual proviene de un cruce de *C. arabica* y *C. liberica*. Este resultado demuestra una clara diferenciación genética entre los individuos de *C. liberica* y *C. canephora* con los individuos de *C. arabica*.



Los resultados de diversidad y estructura genética en café peruano obtenidos por distintos autores (Blas et al. 2010, Mansilla et al. 2021, Palomino et al. 2014 y Wither et al. 2019) y comparados con los obtenidos en el presente estudio, pueden diferir por los distintos orígenes de las muestras empleadas, los marcadores y los enfoques utilizados para el análisis de la data; sin embargo, se ha observado que es similar en cuanto a la estrecha diversidad genética encontrada en el café arábica, esto de acuerdo a su naturaleza reproductiva autógama así como también a su proceso de domesticación.

Adicionalmente, de acuerdo a lo mencionado por Palomino et al. (2014), se debe tener en consideración que las muestras obtenidas de las colectas, muchas veces son el producto de transferencia de semillas entre agricultores y que la variedad catalogada como tal por el agricultor, no corresponde al genotipo o son mezclas, esto también puede darse debido al tipo de introducción del café al país. Tomando en consideración lo mencionado anteriormente, se puede tener una explicación probable hacia la falta de agrupamiento y de estructura de las muestras en base a su denominación varietal.

## V. CONCLUSIONES

1. Se obtuvo ADN con calidad y cantidad suficiente para los procedimientos de PCR.
2. Se establecieron procedimientos para la amplificación vía PCR multiplex logrando amplificar 15 pares de primers en solo 4 reacciones, reduciendo costos y tiempos de ejecución.
3. Se encontraron 70 alelos distintos y una heterocigosidad ( $e$ ) 0,595 indicando una baja variabilidad genética para el café peruano. Índice de fijación ( $F_{st}$ ) indica poca diferenciación entre las poblaciones de café cultivado en el Perú. Los marcadores M32, SSR03, SSRCa018, y M20 mostraron un muy alto contenido de información polimórfica (PIC), por lo cual muestran gran potencial y pueden llegar a ser de mucha utilidad en este tipo de estudios moleculares en café peruano.
4. La estructura genética poblacional muestra 3 poblaciones diferenciadas ( $K=3$ ). Sin embargo, eliminando alelos nulos se generan 5 grupos o estructuras poblacionales de café peruano ( $K=5$ ). Además, las variedades cultivadas en distintas regiones no forman necesariamente un mismo cluster, lo que indica diferenciación por adaptación o problemas de denominación.

## **VI. RECOMENDACIONES**

- Debido a la baja diversidad genética, se recomienda la adición de otros marcadores a los encontrados con alta tasa de polimorfismo en el presente estudio. Además, podrían usarse controles de variedades provenientes de productores certificados.
- Realizar una caracterización morfológica que permita acercarnos a las correctas denominaciones de la variedad y evaluar los perfiles genéticos de los individuos con las características agronómicas de mayor interés. Además, poder conocer que variedades podrían ser los individuos de las regiones de Junín y Pasco cuyo total de individuos están etiquetados como NN.
- Debido a la ambigüedad que generan los genotipos poliploides, se sugiere desarrollar una metodología que permita inferir el dosaje alélico para obtener resultados más precisos.
- Realizar nuevos muestreos de otras regiones de importancia para la caficultura peruana para lograr enriquecer la colección de germoplasma de café del INIA.
- Continuar con las investigaciones que permitan encontrar individuos que sirvan de base para programas de mejoramiento genético de café en el Perú.

## VII. BIBLIOGRAFÍA

- Acuña, W. (2016). *Determinación de la diversidad y estructura genética de patos criollos (Cairina moschata L. 1758) de los departamentos de Lambayeque y san Martín mediante el uso de microsatélites* (Tesis de pregrado). Universidad Nacional Mayor de San Marcos, Lima, Perú.
- Alkimim, E.R., Caixeta, E.T., Sousa, T.V., Pereira, A.A., de Oliveira, A.C.B., Zambolim, L., & Sakiyama, N.S. (2017). Marker-assisted selection provides arabica coffee with genes from other Coffea species targeting on multiple resistance to rust and coffee berry disease. *Molecular Breeding*, 37, 6.
- Allende, M. (2017). *Caracterización morfológica y molecular de accesiones de Quinoa (Chenopodium quinoa Willd.)* (Tesis de pregrado). Universidad Nacional Agraria la Molina, Lima, Perú. Recuperado de [http://biblio.unvm.edu.ar/opac\\_css/doc\\_num.php?explnum\\_id=1395](http://biblio.unvm.edu.ar/opac_css/doc_num.php?explnum_id=1395)
- Asociación Mexicana de la Cadena Productiva del Café (AMECAFE). (2017). *Panorama Internacional Café*. Recuperado de [https://amecafe.org.mx/wp-content/uploads/2017/09/Panorama\\_Internaciona\\_Caf%C3%A9\\_2017.pdf](https://amecafe.org.mx/wp-content/uploads/2017/09/Panorama_Internaciona_Caf%C3%A9_2017.pdf)
- Asociación Nacional del Café (ANACAFE). (2016). *Guía de variedades de café* (2da ed.). Guatemala. 48 p.
- Angiosperm Phylogeny Group (APG). (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*. 181 (1): 1–20
- Anthony, F., Astorga, F., y Berthaud, J. (1999). Los recursos genéticos: las bases de una solución genética a los problemas de la caficultura latinoamericana. In Bertrand, B; Rapidel, B. eds. *Desafíos de la caficultura en Centroamérica*. San José, Costa Rica. IICA. p. 369-405.
- Anthony, F., Bertrand, B., Quiros, O., Lashermes, P., Berthaud, J., & Charrier A. (2001). Genetic diversity of wild coffee (*Coffea arabica* L.) using molecular markers. *Euphytica*. 118: 53-65.

- Anthony, F., Combes, M.C., Astorga, C., Bertrand, B., Graziosi, G., Lashermes, P. (2002). The origin of cultivated *Coffea arabica* L. varieties revealed by AFLP and SSR markers. *Theoretical and Applied Genetics*, 104: 894-900.
- Arcila, J., Farfán, F., Moreno, A.M., Salazar, L.F., y Hincapié, E. (2007). *Sistemas de producción de café en Colombia*. Chinchiná, Colombia: Cenicafé. 309 p.
- Balzarini, M., Arroyo, A., Bruno, C., y Di Renzo, J. (2006). Análisis de datos de marcadores con Info-Gen. In XXXV Congreso Argentino de Genética, San Luis, Argentina.
- Barahona, S. (2014). *Utilización de loci microsatélites y ADN mitocondrial para evaluar la estructuración genético-poblacional de la Caballa (*Scomber japonicus houttuyn*, 1782) en el mar peruano* (Tesis de pregrado). Universidad Nacional Mayor de San Marcos, Lima, Perú.
- Baruah, A., Naik, V., Hendre, P.S., Rajkumar, R., Rajendrakumar, P., y Aggarwal, R.K. (2003). Isolation and characterization of nine microsatellite markers from *Coffea arabica* L., showing wide cross-species amplifications. *Mol Ecol Notes* 3:647–650.
- Blas, R. (2010). *Certificación de semillas y mejoramiento genético de plantas mediante marcadores moleculares*. UNALM, Lima, Perú. Dpto. Fitotecnia, Fac. Agronomía.
- Blas, R., Jananpa, R.C., y Bello, S. (2011). *Informe de Proyecto “Estudio de la variabilidad genética del café y establecimiento de un banco de germoplasma en la selva peruana”*. Caracterización del germoplasma peruano del café. UNALM - FDA – Café Perú – FINCYT. Lima. 58 p.
- Bolivar, A., Rojas, A., Garcia-Lugo, P. (2013). PCR y PCR-Múltiple: parámetros críticos y protocolo de estandarización. Instituto de Inmunología Clínica. Mérida-Venezuela. *Avances en Biomedicina*, 3(1): 25-33.
- Botstein, D., White, R., Skolmick, H. y Davis, R. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphism. *American Journal of Human Genetics*, 32, 314-331.
- Cámara Peruana de Café y Cacao. (2019). *Café-Datos. Exportaciones*. Consultado el 11 de julio de 2019. Disponible en <https://camcafeperu.com.pe/ES/cafe-datos.php>
- CENICAFE. (s.f.). *Cartilla de Variedades de café sembradas en Colombia*. Recuperado de: <https://www.cenicafe.org/es/publications/C1.pdf>
- CENICAFE. (2000). *La variedad Colombia: Veinte años de adopción y comportamiento frente a nuevas razas de la roya del cafeto*. Recuperado de <https://www.cenicafe.org/es/publications/bot022.pdf>
- Centro Agronómico Tropical de Investigación y Enseñanza (CATIE). (1995). *Programa de*

investigación científica. Recuperado de <https://books.google.com.pe/books?id=rdIOAQAAIAAJ&printsec=frontcover#v=onepage&q&f=false>

- Clark, L. y Jasieniuk, M. (2011). POLYSAT: An R package for polyploid microsatellite analysis. *Molecular Ecology Resources* (11): 562–566.
- Clarke, R. (2003). Green coffee. *Encyclopedia of Food Sciences and Nutrition*, 2: 1481-1487. <https://doi.org/10.1016/B0-12-227055-X/00268-6>
- Combes, M., Andrzejewski, S., Anthony, F., Bertrand, B., Rovelli, P., Graziosi, G. y Lashermes, P. (2000). Characterization of microsatellite loci in *Coffea arabica* and related coffee species. *Rev. Molecular Ecology*, 9 (8), 1178-1180.
- Cubry, P., Musoli, P., Legnaté, H., Pot, D., Bellis, F., Poncet, V., Anthony, F., Dufour, M., Leroy, T. (2008). Diversity in coffee assessed with SSR markers: structure of the genus *Coffea* and perspectives for breeding. *Genome* 51(1):50-63.
- De Silva, H., Hall, A., Rikkerink, E., McNeilage, M., Fraser, L. (2005). Estimation of allele frequencies in polyploids under certain patterns of inheritance. *Heredity* 95, 327–334.
- de Vicente, M.C., López, C., y Fulton, T. (2004). *Análisis de la Diversidad Genética Utilizando Datos de Marcadores Moleculares: Módulo de Aprendizaje Volumen 2*. Instituto Internacional de Recursos Fitogenéticos (IPGRI), Roma Italia. Consultado 15 de agosto de 2021. Disponible en [https://www.biodiversityinternational.org/fileadmin/user\\_upload/online\\_library/publications/pdfs/Molecular\\_Markers\\_Volume\\_2\\_es.pdf](https://www.biodiversityinternational.org/fileadmin/user_upload/online_library/publications/pdfs/Molecular_Markers_Volume_2_es.pdf)
- Doyle, J.J. y Doyle, J.L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11-15.
- Eguiarte, L., Aguirre, E., Sheinvar, E., Gonzáles, A., y Souza, V. (2010). *Flujo génico, diferenciación y estructura genética de las poblaciones, con ejemplos en especies de plantas mexicanas*. Universidad Autónoma de México
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology*, 14(8), 2611-2620.
- Geleta, M., Herrera, I., Monzón, A. & Bryngelsson, T. (2012). Genetic Diversity of Arabica Coffee (*Coffea arabica* L.) in Nicaragua as Estimated by Simple Sequence Repeat Markers. *The Scientific World Journal*, vol. 2012, Article ID 939820, 11 p. Recuperado de <https://doi.org/10.1100/2012/939820> .

- Goudet, J. (2002). *FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3)*. University of Lausanne, Department of Ecology and Evolution, Lausanne, Switzerland
- Hendre, P., Phanindranath, R., Annapurna, V., Lalremruata, A., & Aggarwal, R.K. (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): 5
- Hernández, A., Trejo, F. (2014). Estudio genético poblacional de frecuencias alélicas para 15 marcadores STR presentes en la población del Estado de Zacatecas aplicado a la práctica forense. *iMedPub Journals*, 10, (1),1-24.
- ICO (International Coffee Organization). (s.f.). *Aspectos botánicos*. Consultado el 17 de julio de 2019. Disponible: [http://www.ico.org/es/botanical\\_c.asp](http://www.ico.org/es/botanical_c.asp)
- Inglis, P., Pappas, M., Resende, L., & Grattapaglia, D. (2018). Fast and inexpensive protocols for consistent extraction of high quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLOS ONE* 13(10): e0206085.
- Jombart, T. (2013). *A tutorial for Discriminant Analysis of Principal Components (DAPC) using Adegenet 1.3-6*. Recuperado de <https://dokumen.tips/documents/a-tutorial-for-discriminant-analysis-of-principal-adegenetr-forger-a-tutorial.html?page=1>
- Jombart, T. y Collins, C. (2015). *A tutorial for discriminant analysis of principal components (DAPC) using adegenet 2.0.0*. London: Imperial College London, MRC Centre for Outbreak Analysis and Modelling.
- Junta Nacional del Café JNC. (2016). *Café peruano: Desafíos, retos y compromisos*. Recuperado de [http://www.congreso.gob.pe/Docs/comisiones2016/ComercioTurismo/files/informe\\_s\\_invitados/junta\\_nacional\\_del\\_cafe.pdf](http://www.congreso.gob.pe/Docs/comisiones2016/ComercioTurismo/files/informe_s_invitados/junta_nacional_del_cafe.pdf)
- Junta Nacional del Café JNC. (2020). *El café del Perú*. Recuperado de <https://juntadelcafe.org.pe/el-cafe-de-peru/>
- Junta Nacional del Café JNC. (2020). *Estadísticas*. Recuperado de <https://juntadelcafe.org.pe/estadisticas/>
- Julca, A., Blas, R., Borjas, R., Bello, S., Anahui, J., Talaverano, D., Crespo, R., y Fundes, G. (2010). *Informe de colecta de germoplasma de café en el Perú*. UNALM - FDA - Café Perú – FINCyT. Lima. 38 p.
- Kalinowski, T., Taper, L., & Marshall, C. (2007). Revising how the computer program

- CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099-1106.
- Kosman, E., & Jokela, J. (2019). Dissimilarity of individual microsatellite profiles under different mutation models: Empirical approach. *Ecology and evolution*, 9(7), 4038–4054. <https://doi.org/10.1002/ece3.5032>
- Lashermes, P., Trouslot, P., Anthony, F., Combes, M.C., & Charrier, A. (1996). Genetic diversity for RAPD markers between cultivated and wild accessions of *Coffea arabica*. *Euphytica* 87: 59- 64
- León, J. (1987). *Botánica de los cultivos Tropicales*. IICA (Instituto Interamericano de Cooperación para la Agricultura). San José, Costa Rica.
- León, J. (2000). *Botánica de los cultivos tropicales*. San José, Costa Rica: Editorial Agroamérica.
- Maluf, M., Silvestrini, M., Ruggiero, L., Guerreiro, F., & Colombo, C. (2005). Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. *Scientia Agricola*, 62(4), 366-373. <https://dx.doi.org/10.1590/S0103-90162005000400010>
- Martínez, R.D. (2008). *Caracterización genética y morfológica del bovino criollo argentino de origen patagónico* (Tesis de Doctorado). Universidad Politécnica de Valencia, Valencia, España. 244 p.
- MINAGRI. (2017). *MINAGRI presenta Plan Nacional de Acción del Café en beneficio de 233 mil familias dedicadas a este cultivo*. Consultado el 10 de julio de 2019. Disponible en <https://www.gob.pe/institucion/minagri/noticias/20065-minagri-presenta-plan-nacional-de-accion-del-cafe-en-beneficio-de-233-mil-familias-dedicadas-a-este-cultivo>
- Missio, R., Caixeta, E., Zambolin, E., Zambolin, L., Cruz, C. & Sakiyama, N. (2010). Polymorphic information content of SSR markers for *Coffea* spp. *Crop Breeding and Applied Biotechnology* 10: 89-94
- Onyśk, A., Boczkowska, M. (2017). M13-Tailed Simple Sequence Repeat (SSR) Markers in Studies of Genetic Diversity and Population Structure of Common Oat Germplasm. *Methods Mol Biol*. Vol. 1536:159-168. DOI: 10.1007/978-1-4939-6682-0\_12.
- Osatinsky, R. (2007). ¿Qué es la electroforesis capilar?. *Bioquímica y Patología Clínica*, 71(2): 60-66. Asociación Bioquímica Argentina. Argentina.
- Pacheco, E. (2015). *Caracterización morfológica y molecular de mashua (Tropaeolum*

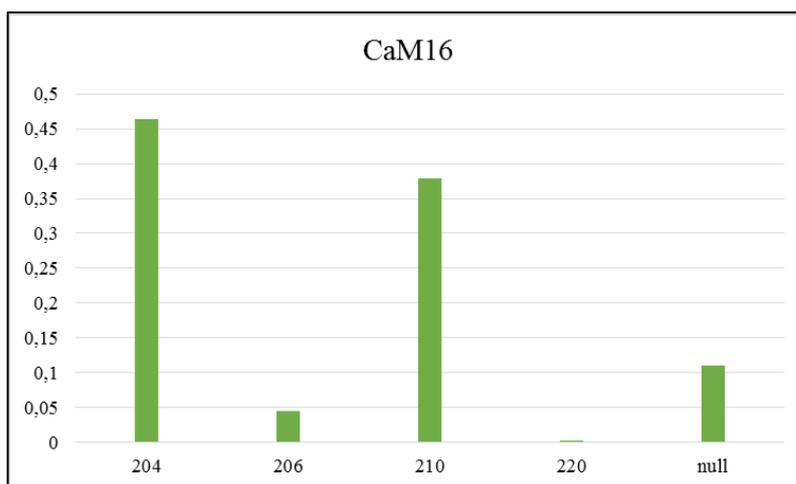
- tuberosum* Ruiz & Pavón) de los departamentos de Cusco y Cajamarca (Tesis de Maestría). Universidad Nacional Agraria La Molina, Lima, Perú.
- Palomino, C., López, C., Espejo, R., Mansilla, R., y Quispe, J. (2014). Evaluación de la diversidad genética de café (*Coffea arabica* L.) en Villa Rica (Perú). *Ecología aplicada*, 13(2):129-134. ISSN 1726-2216.
- Paredes, G. (2013). *Caracterización molecular de las llamas (Lama glama) Ch'aku y Ccara del banco de germoplasma de alpacas de color y llamas del centro experimental Illpa-INIA Anexo Quimsachata, usando marcadores microsatélites* (Tesis de pregrado). Universidad Nacional Agraria La Molina, Lima, Perú.
- Peakall, R., Smouse, P. (2012). GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research- an update. *Bioinformatics Application Note*, 28, 2537-2539. <http://dx.doi.org/10.1093/bioinformatics/bts460>
- Pritchard, J., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945-959
- Puerta, G. (1998). *Calidad en taza de las variedades de Coffea arabica L cultivadas en Colombia*. Biblioteca Cenicafe. Consultado el 15 de agosto de 2021. Disponible en <http://biblioteca.cenicafe.org/handle/10778/64>
- Razafinarivo, N., Guyot, R., Davis, A., Couturon, C., Hamon, S., Cruzillat, D., Rigoreau, M., Dubreuil-Tranchant, C., Poncet, V., DeKochko, A., Rakotomalala, J-J., Hamon, & P. (2013). Genetic structure and diversity of coffee (*Coffea*) across Africa and the Indian Ocean islands revealed using microsatellites. *Annals of Botany* 111:229–248
- Rimieri, P. (2017). La diversidad genética y la variabilidad genética: dos conceptos diferentes asociados al germoplasma y al mejoramiento genético vegetal. BAG. *Journal of basic and applied genetics*, 28(2), 7-13.
- Salazar, A., Sandoval, A., y Armendáriz, J. (2013). *Biología molecular: Fundamentos y aplicaciones en las ciencias de la salud*. México D.F.: McGraw Hill.
- Sanchez, E. (2017). *Evaluación de microsatélites para la identificación del germoplasma de especies del género Coffea en Costa Rica*. Recuperado de <http://repositorio.sibdi.ucr.ac.cr:8080/jspui/handle/123456789/6183>
- Sevilla, R., Holle, M. (2004). *Recursos Genéticos Vegetales*. Lima, Perú: Ed. Torre Azul.
- Souza, F., Santos, J., Costa, J., & Santos, M. (2004). *Características das principais variedades de café cultivadas em Rondônia*. Porto Velho: Embrapa Rondônia informação tecnológica, 21 p.
- Steiger, D., Nagai, C., Moore, P., Morden, W., Osgood, V., Ming, R. (2002). AFLP analysis

- of genetic diversity within and among *Coffea arabica* cultivars. *Theor Appl Genet* 105: 209. <https://doi.org/10.1007/s00122-002-0939-8>
- Syafaruddin, S., Randriani, E., Dani, D., Sulistyorini, I., & Pabendon, M. (2014). Genetic Variability of 15 Robusta Coffee Genotypes Selected by Farmer Based on SSRs Markers. *Journal Tanaman Industri Dan Penyegar*, 1(2), 87-94. doi: <http://dx.doi.org/10.21082/jtidp.v1n2.2014.p87-94>
- Tomasco, I. y Lessa, E. (2015). *Variación genética. Equilibrio Hardy-Weinberg y mutación*. Universidad Nacional Autónoma de México.
- Tripp, M. (2009). *Análisis de la estructura genética poblacional del dorado (Coryphaena hippurus; Linnaeus, 1758) en el noroeste del pacífico mexicano y golfo de California mediante el uso de microsatélites* (Tesis de Maestría). La Paz, B.C.S – México. Centro de Investigaciones Biológicas del Noroeste, S.C.
- Universidad Nacional Agraria La Molina. (2013). *Asistencia técnica dirigida en “Cosecha y postcosecha en el cultivo de café”*. Recuperado de <https://www.agrobanco.com.pe/data/uploads/ctecnica/011-r-cafe.pdf>
- Vargas, G.S. (2017). Tolerancia a la desecación de semillas de café catimor (*Coffea arabica* L. x *Coffea canephora* Pierre ex A. Froehner) (Tesis de pregrado). Universidad Nacional Agraria la Molina, Lima, Perú. Recuperado de <http://repositorio.lamolina.edu.pe/bitstream/handle/UNALM/3721/quintana-vassallo-vania-carmenza.pdf?sequence=1&isAllowed=y>
- Vía y Rada, R. (2015). *Determinación de la diversidad genética de 172 accesiones de la colección nacional de *Chenopodium quinoa* Willd. “Quinoa” mediante marcadores microsatélites* (Tesis de pregrado). Universidad Ricardo Palma, Lima, Perú.
- Vieira, E., Von Pinho, E., Carvalho, M., Esselink, D. y Vosman, B. (2010). Development of microsatellite markers for identifying Brazilian *Coffea arabica* varieties. *Rev. Genetics and Molecular Biology*, 33,3, 507-514.
- World Coffee Research – WCR. (2019). Las variedades de café arábica. Portland. USA. 75 p.

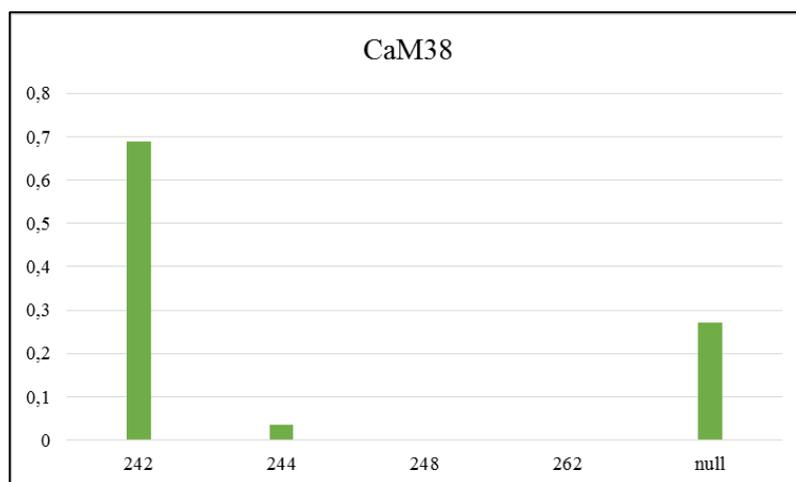
## VIII. ANEXOS

Anexo 1: Graficas de frecuencias alélicas por cada marcador en la población estudiada.

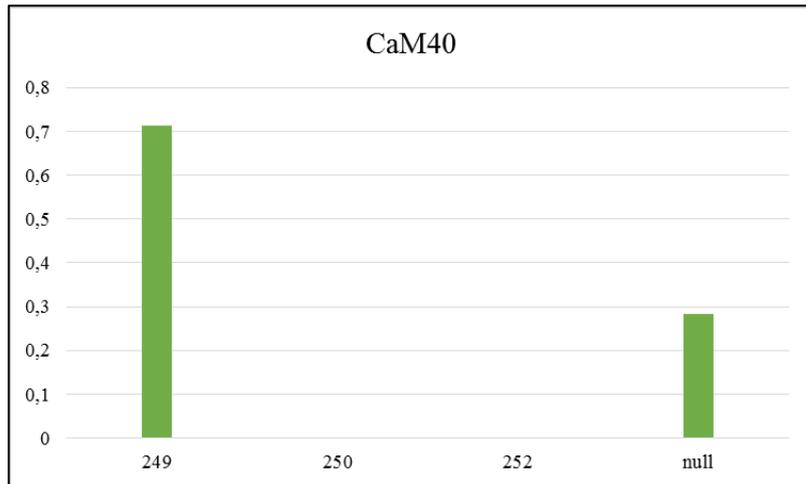
### CaM16



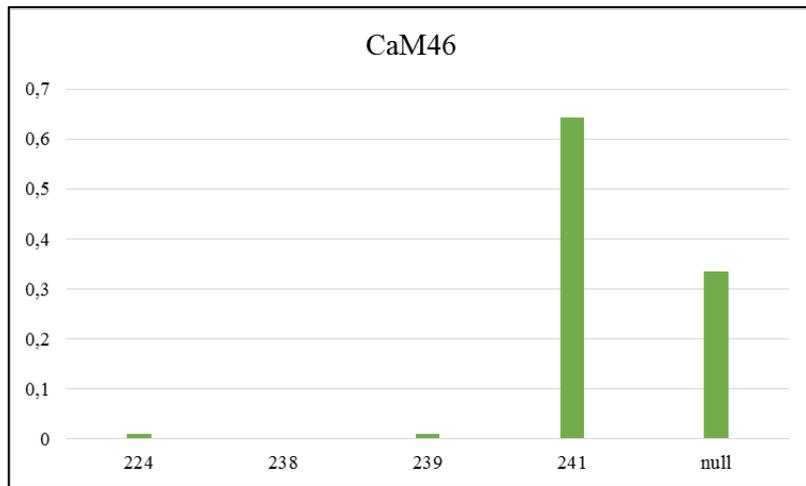
### CaM38



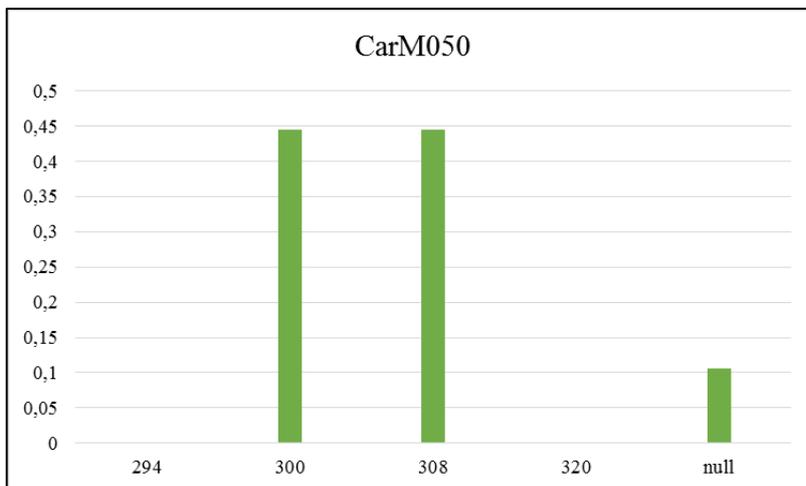
### CaM40



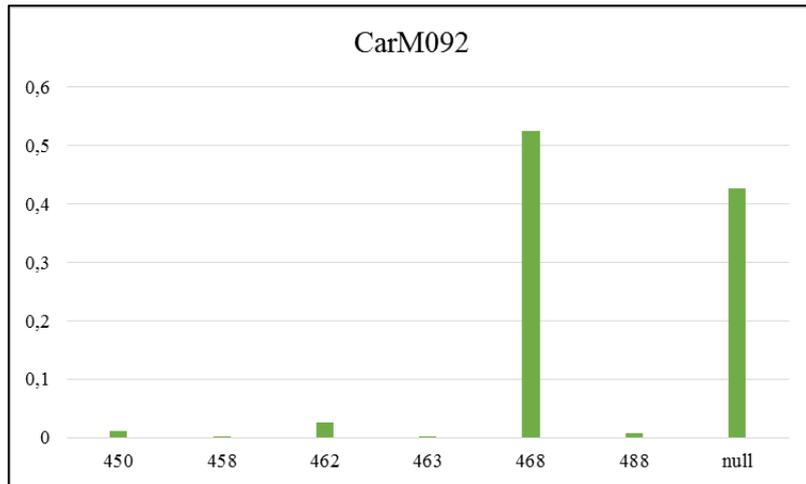
### CaM46



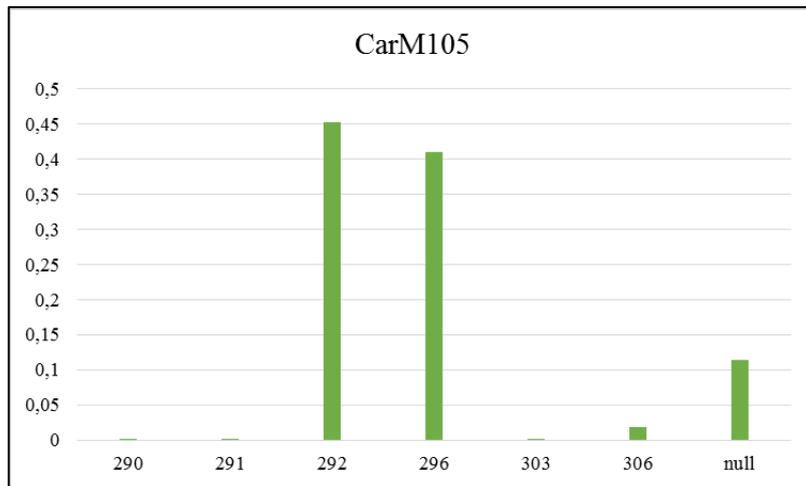
### CarM050



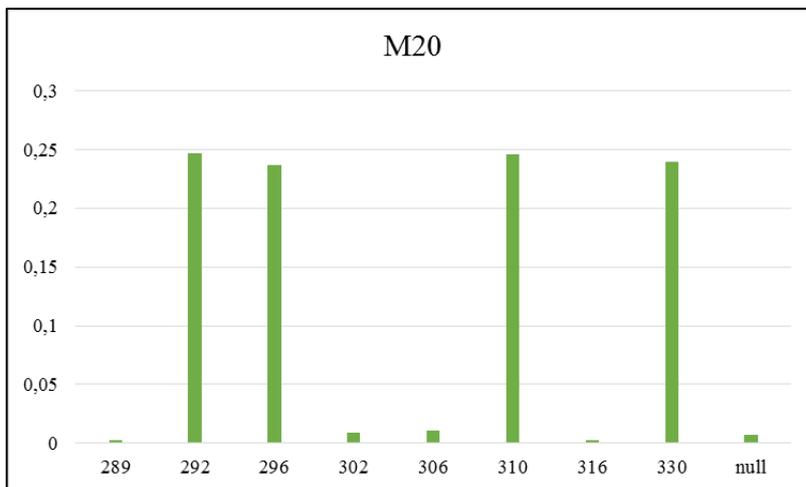
### CarM092



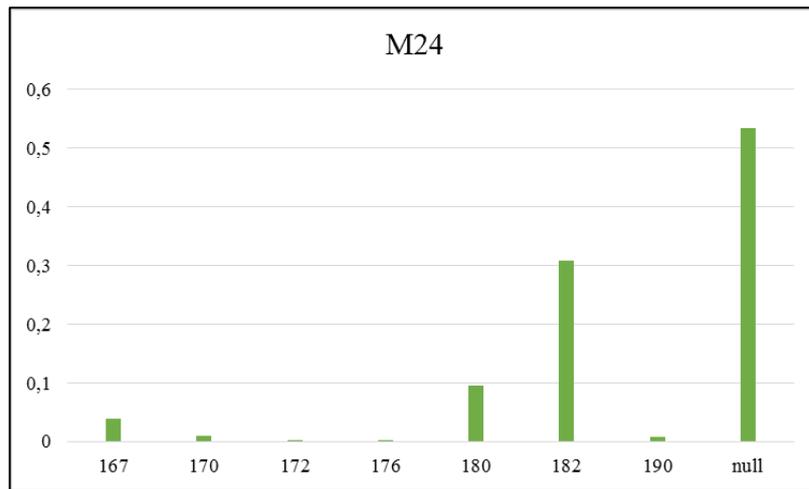
### CarM105



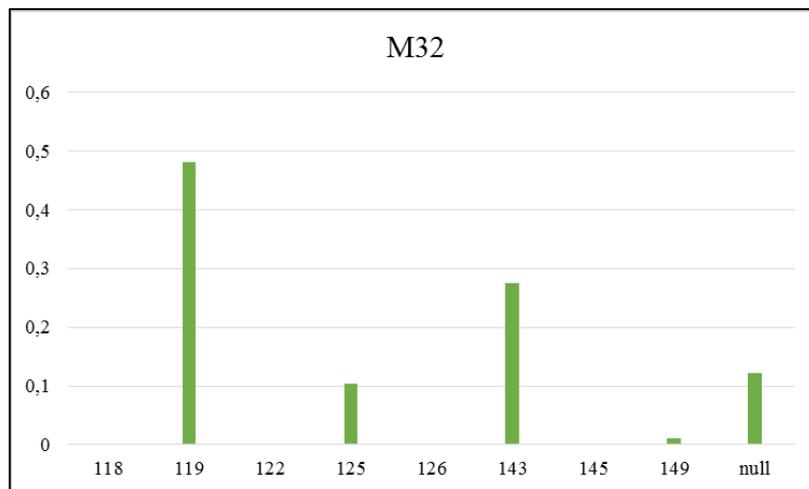
### M20



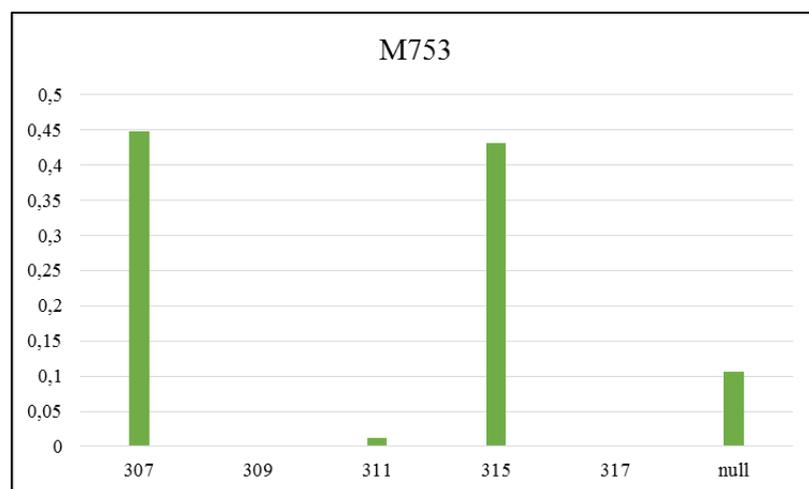
### M24



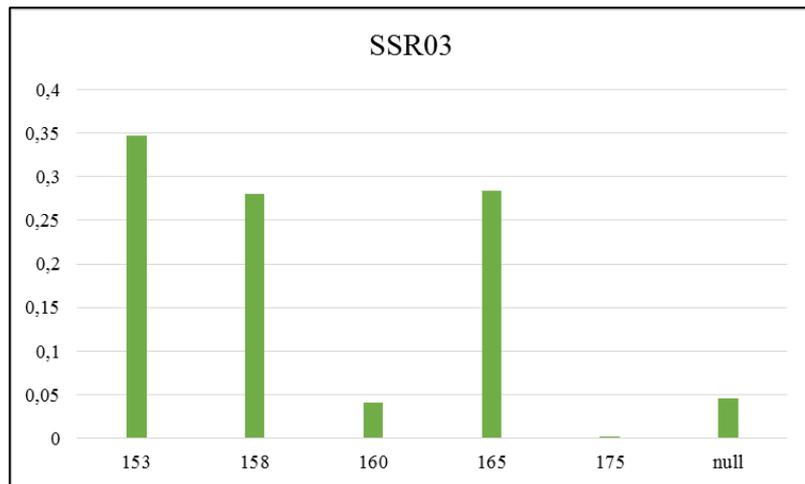
### M32



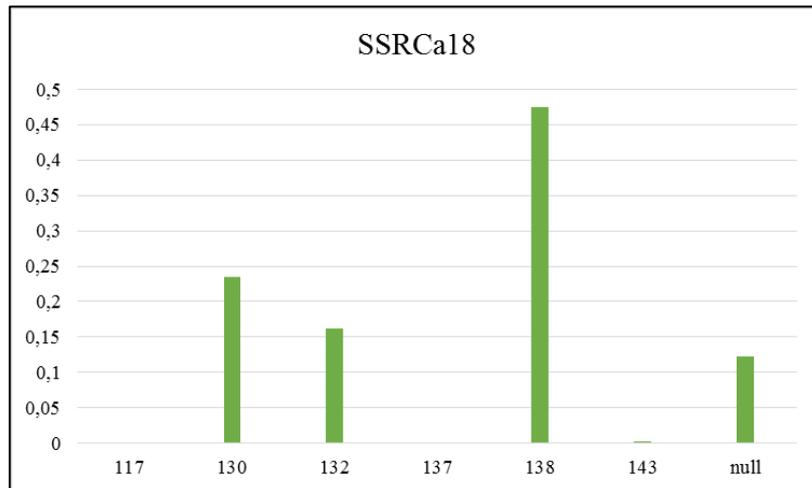
### M753



### SSR03



### SSRCa18



## Anexo 2: Cantidad de componentes y concentraciones para reacciones de PCR multiplex

### Componentes y volúmenes utilizados para 4 pares de cebadores

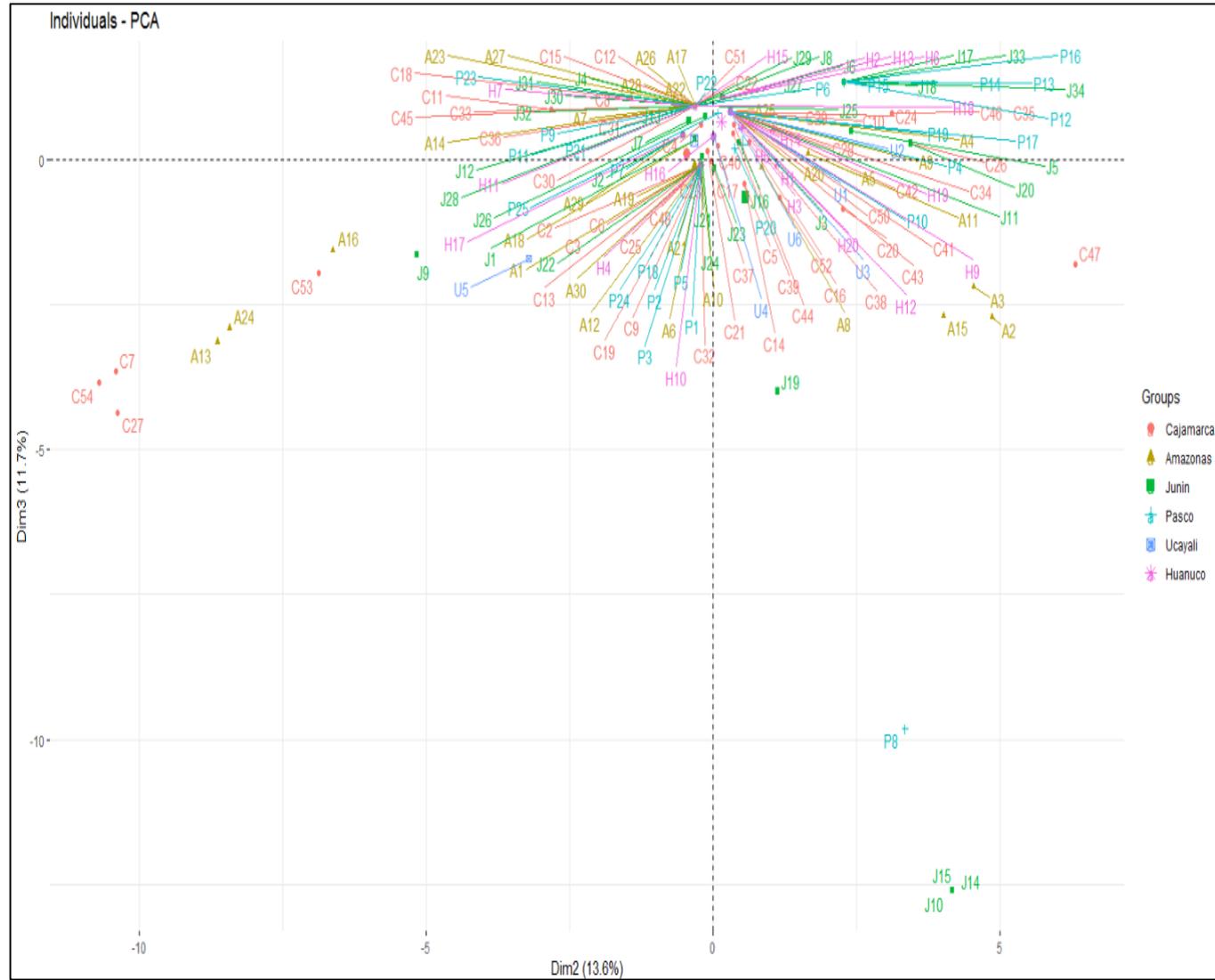
Componentes	Concentración Inicial	Concentración Final	Vol Final para 1 Rx	Vol Final para 45 Rx
H <sub>2</sub> O	-	-	2.729 µL	122.805 µL
MgCl <sub>2</sub>	50 mM	2.5 mM	0.5 µL	22.5 µL
Buffer	10 X	1 X	1 µL	45 µL
DNTPs	10 mM	0.2 mM	0.2 µL	9 µL
1er Primer F	10 µM	0.03 µM	0.03 µL	1.35 µL
1er Primer R	10 µM	0.06 µM	0.06 µL	2.70 µL
2do Primer F	10 µM	0.03 µM	0.03 µL	1.35 µL
2do Primer R	10 µM	0.06 µM	0.06 µL	2.70 µL
3er Primer F	10 µM	0.03 µM	0.03 µL	1.35 µL
3er Primer R	10 µM	0.06 µM	0.06 µL	2.70 µL
4to Primer F	10 µM	0.03 µM	0.03 µL	1.35 µL
4to Primer R	10 µM	0.06 µM	0.06 µL	2.70 µL
Fluoróforo	10 µM	0.151 µM	0.151 µL	6.795 µL
Taq Pol	5 U	0.3 U	0.06 µL	2.7 µL
ADN	10 ng	50 ng	5.0	
<b>Volumen</b>	-	-	<b>10</b>	<b>450</b>

### Componentes y volúmenes utilizados para 4 pares de cebadores (modificado)

Componentes	Concentración Inicial	Concentración Final	Vol Final para 1 Rx	Vol Final para 45 Rx
H <sub>2</sub> O	-	-	2.729 µL	122.805 µL
MgCl <sub>2</sub>	50 mM	2.5 mM	0.5 µL	22.5 µL
Buffer	10 X	1 X	1 µL	45 µL
DNTPs	10 mM	0.2 mM	0.2 µL	9 µL
1er Primer F	10 µM	0.0225 µM	0.0225 µL	1.0125 µL
1er Primer R	10 µM	0.045 µM	0.045 µL	2.025 µL
2do Primer F	10 µM	0.0375 µM	0.0375 µL	1.6875 µL
2do Primer R	10 µM	0.075 µM	0.075 µL	3.375 µL
3er Primer F	10 µM	0.0225 µM	0.0225 µL	1.0125 µL
3er Primer R	10 µM	0.045 µM	0.045 µL	2.025 µL
4to Primer F	10 µM	0.0375 µM	0.0375 µL	1.6875 µL
4to Primer R	10 µM	0.075 µM	0.075 µL	3.375 µL
Fluoróforo	10 µM	0.151 µM	0.151 µL	6.795 µL
Taq Pol	5 U	0.3 U	0.06 µL	2.7 µL
ADN	10 ng	50 ng	5.0	
<b>Volumen</b>	-	-	<b>10</b>	<b>450</b>

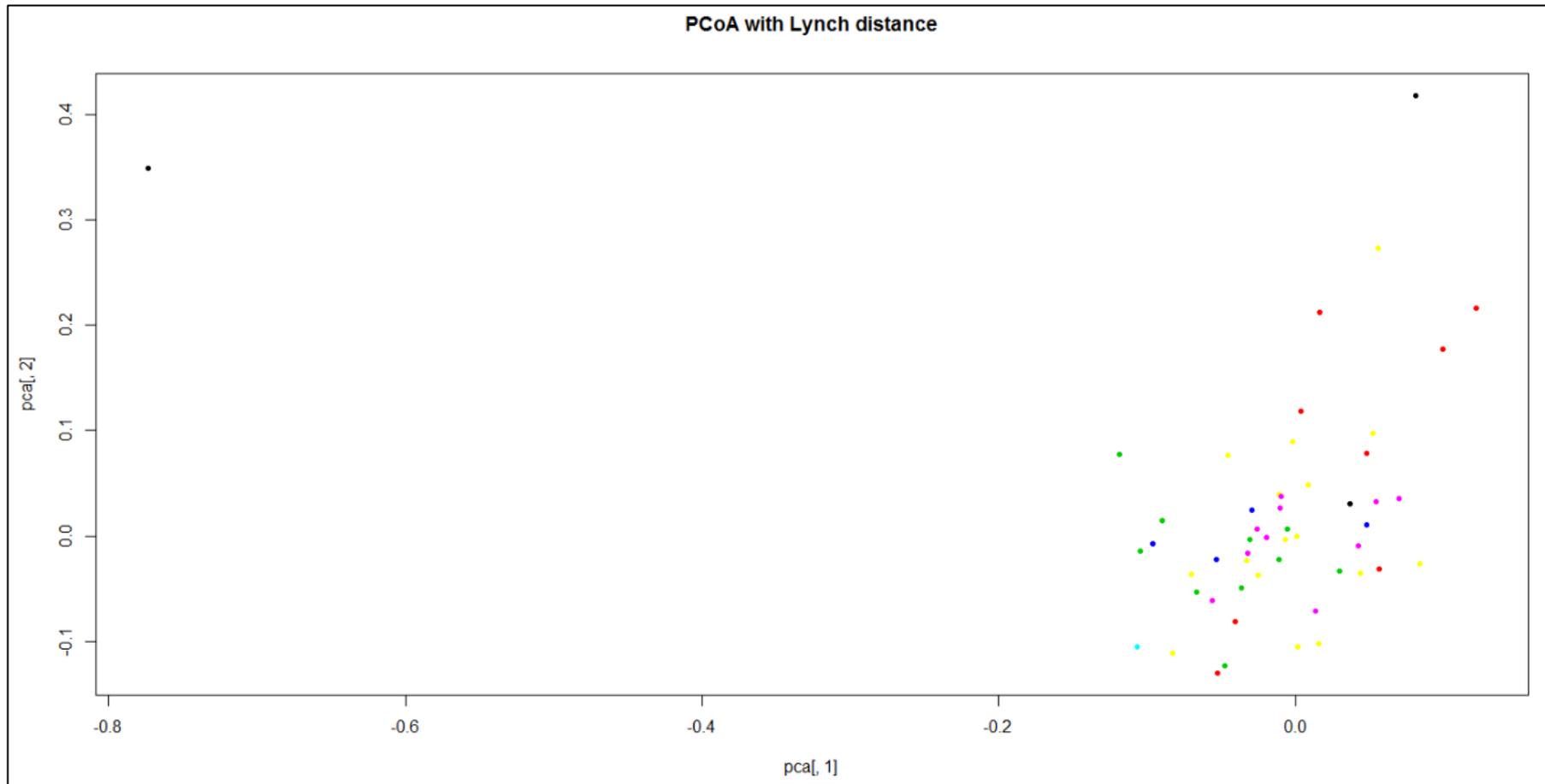


### Gráfica de segundo y tercer componente



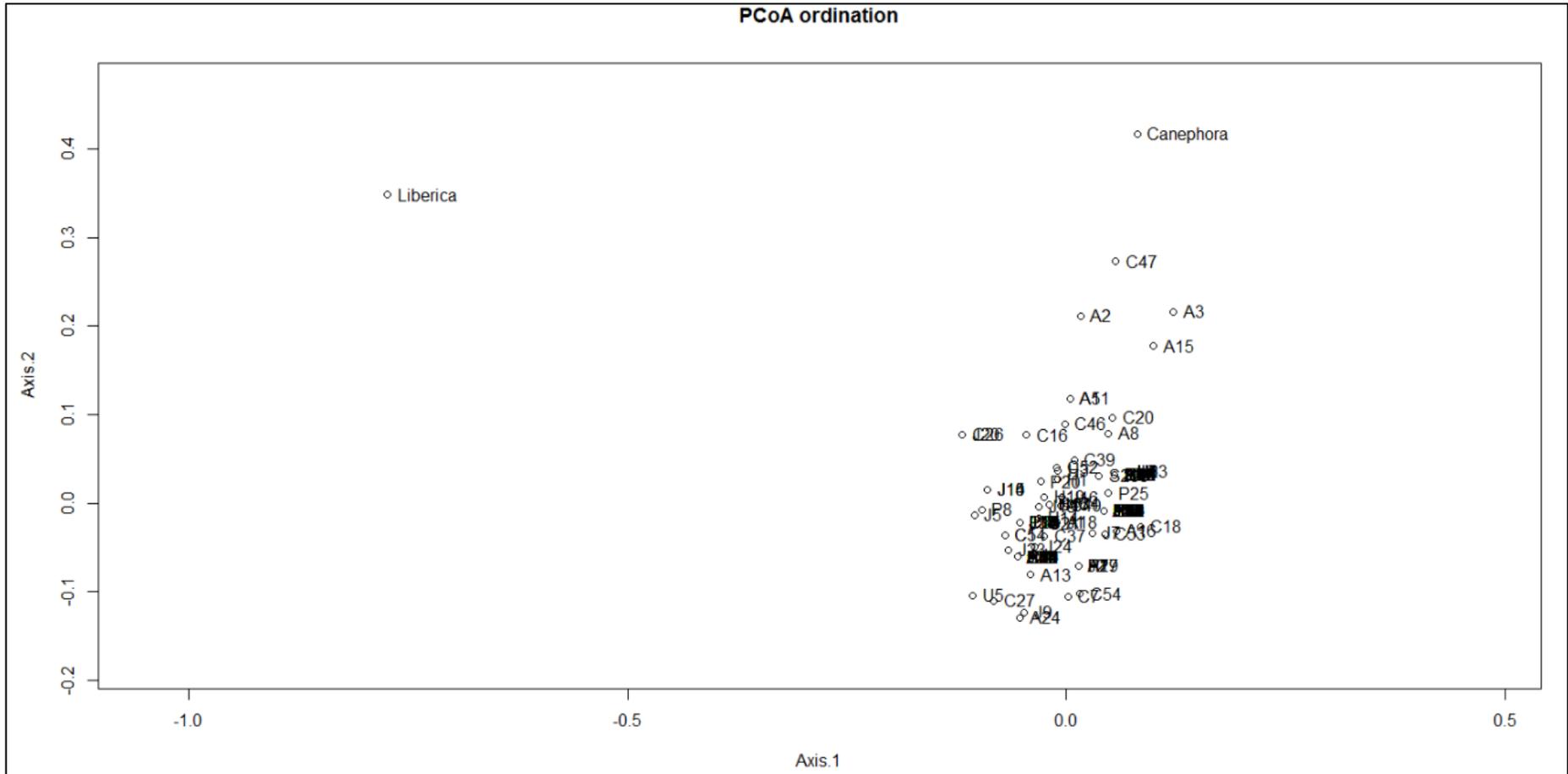
## Anexo 4: Gráficas correspondientes al Análisis de coordenadas principales – PcoA

### PCoA con distancia de Lynch



```
plot(pca[,1], pca[,2], col= rep(c("7", "2", "3", "4", "5", "6", "1"), c(52,30,34,25,6,19,3)) , main = "PCoA with Lynch distance", type = "p", pch = 20)
```

### PCoA biplot



```
ll<-pcoa(testmat, correction="none", rn=NULL) biplot(ll)
```

## Anexo 5: Tablas de análisis discriminante de componentes principales – DAPC

### DAPC 169 ind, 13 loci, 6 componentes principales y 4 funciones discriminantes

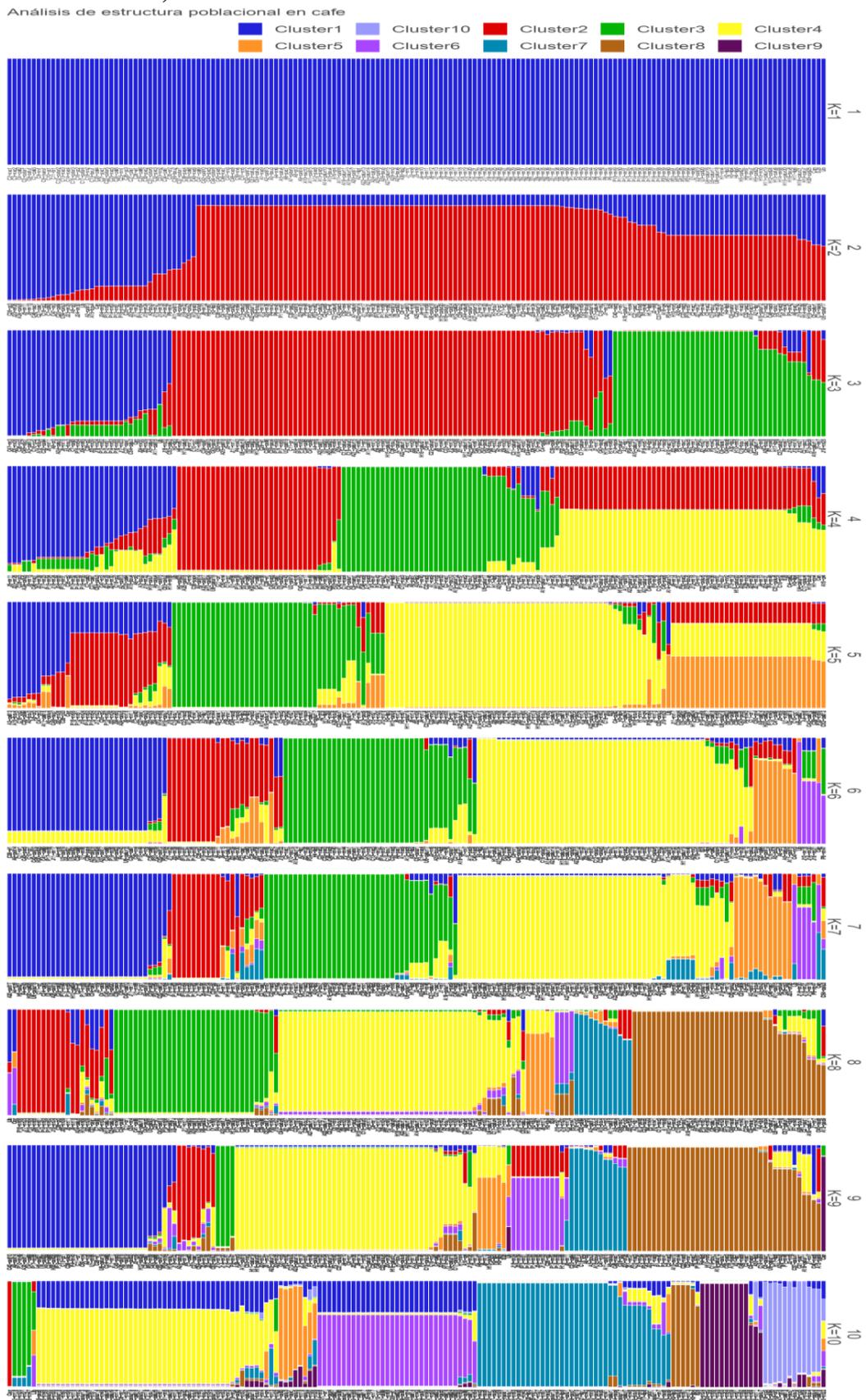
Ind	Cluster	Ind	Cluster	Ind	Cluster	Ind	Cluster
C17-catur4	1	J1-nn7	2	J11-nn17	3	J13-nn19	4
C35-cr2	1	J2-nn8	2	P4-nn43	3	J19-nn25	4
J17-nn23	1	J10-nn16	2	P10-nn49	3	J25-nn31	4
J18-nn24	1	J14-nn20	2	P17-nn56	3	J26-nn32	4
J33-nn39	1	J15-nn21	2	P19-nn58	3	J27-nn33	4
J34-nn65	1	J21-nn27	2	P20-nn59	3	J28-nn34	4
P12-nn51	1	J22-nn28	2	U1-nn66	3	J29-nn35	4
P13-nn52	1	P1-nn40	2	U2-cati19	3	J30-nn36	4
P14-nn53	1	P2-nn41	2	U3-cati20	3	J31-nn37	4
P15-nn54	1	P3-nn42	2	U6-catur18	3	J32-nn38	4
P16-nn55	1	P5-nn44	2	H5-cati21	3	P6-nn45	4
U4-catur16	1	P7-nn46	2	H9-cati23	3	P9-nn48	4
H1-mn4	1	P8-nn47	2	H12-cati24	3	P11-nn50	4
H3-tip4	1	P18-nn57	2	H20-cati25	3	P21-nn60	4
H14-nn68	1	P24-nn63	2	C4-cati2	4	P22-nn61	4
H16-bou8	1	U5-catur17	2	C8-nn1	4	P23-nn62	4
H19-catu7	1	H4-tip5	2	C12-catur2	4	P25-nn64	4
C2-nac1	2	H10-mn5	2	C15-catur3	4	H2-tip3	4
C3-nac2	2	H17-tip6	2	C18-nn2	4	H6-nn67	4
C5-catur1	2	C10-gc1	3	C22-catur5	4	H7-bou7	4
C6-catu1	2	C11-cr1	3	C30-pac2	4	H11-catur19	4
C9-mara1	2	C24-cati5	3	C31-catur6	4	H13-catu6	4
C13-nac3	2	C28-vi1	3	C33-mn2	4	H15-catur20	4
C14-nac4	2	C29-cati6	3	C36-bou2	4	H18-catur21	4
C19-bou1	2	C34-cati7	3	C44-catur8	4	CAN	4
C21-pac1	2	C38-pac4	3	C45-pac6	4	S288	4
C23-nac5	2	C39-cati8	3	C51-nn5	4	C7-cati3	5
C25-mn1	2	C41-cati9	3	C53-catu2	4	C16-gc2	5
C32-pac3	2	C42-cati10	3	A7-catur9	4	C20-cati4	5
C37-catur7	2	C43-pac5	3	A14-catur10	4	C26-gc3	5
C40-nac6	2	C50-nn4	3	A16-cr4	4	C27-cas1	5
C48-nac8	2	C52-cati12	3	A17-cati16	4	C46-nac7	5
A1-nac9	2	A2-cati13	3	A22-catur13	4	C47-nn3	5
A6-pac7	2	A3-bou3	3	A23-catur14	4	C54-catu3	5
A10-nac10	2	A4-nn6	3	A26-bou5	4	A8-cr3	5
A12-pac8	2	A5-catu4	3	A27-catur15	4	A24-catu5	5
A13-pac9	2	A9-cati14	3	A28-mn3	4	J5-nn11	5
A18-catur11	2	A11-cati15	3	J4-nn10	4	J9-nn15	5
A19-catur12	2	A15-bou4	3	J6-nn12	4	J16-nn22	5
A21-tip1	2	A20-cati17	3	J7-nn13	4	J20-nn26	5
A29-bou6	2	A25-cati18	3	J8-nn14	4	J23-nn29	5
A30-tip2	2	J3-nn9	3	J12-nn18	4	J24-nn30	5
						LIB	5

### DAPC 166 ind, 8 loci, 6 componentes principales y 4 funciones discriminantes

Ind	Cluster	Ind	Cluster	Ind	Cluster	Ind	Cluster
C7-cati3	1	P18-nn57	2	C22-catur5	5	J11-nn17	5
C27-cas1	1	P24-nn63	2	C24-cati5	5	J12-nn18	5
C53-catu2	1	U4-catur16	2	C28-vi1	5	J13-nn19	5
C54-catu3	1	U5-catur17	2	C29-cati6	5	J25-nn31	5
A13-pac9	1	H3-tip4	2	C30-pac2	5	J26-nn32	5
A16-cr4	1	H4-tip5	2	C31-catur6	5	J27-nn33	5
A24-catu5	1	H10-mn5	2	C33-mn2	5	J28-nn34	5
J9-nn15	1	H14-nn68	2	C34-cati7	5	J29-nn35	5
C2-nac1	2	H16-bou8	2	C36-bou2	5	J30-nn36	5
C3-nac2	2	H17-tip6	2	C38-pac4	5	J31-nn37	5
C5-catur1	2	C26-gc3	3	C39-cati8	5	J32-nn38	5
C6-catu1	2	C35-cr2	3	C41-cati9	5	P4-nn43	5
C9-mara1	2	C46-nac7	3	C42-cati10	5	P6-nn45	5
C13-nac3	2	C47-nn3	3	C43-pac5	5	P9-nn48	5
C14-nac4	2	J5-nn11	3	C44-catur8	5	P10-nn49	5
C16-gc2	2	J17-nn23	3	C45-pac6	5	P11-nn50	5
C17-catur4	2	J18-nn24	3	C50-nn4	5	P17-nn56	5
C19-bou1	2	J20-nn26	3	C51-nn5	5	P19-nn58	5
C21-pac1	2	J33-nn39	3	C52-cati12	5	P20-nn59	5
C23-nac5	2	J34-nn65	3	A2-cati13	5	P21-nn60	5
C25-mn1	2	P12-nn51	3	A3-bou3	5	P22-nn61	5
C32-pac3	2	P13-nn52	3	A4-nn6	5	P23-nn62	5
C37-catur7	2	P14-nn53	3	A5-catu4	5	P25-nn64	5
C48-nac8	2	P15-nn54	3	A7-catur9	5	U1-nn66	5
A1-nac9	2	P16-nn55	3	A8-cr3	5	U2-cati19	5
A6-pac7	2	C40-nac6	4	A9-cati14	5	U3-cati20	5
A10-nac10	2	A18-catur11	4	A11-cati15	5	U6-catur18	5
A12-pac8	2	J1-nn7	4	A14-catur10	5	H1-mn4	5
A19-catur12	2	J10-nn16	4	A15-bou4	5	H2-tip3	5
A21-tip1	2	J14-nn20	4	A17-cati16	5	H5-cati21	5
A29-bou6	2	J15-nn21	4	A20-cati17	5	H6-nn67	5
A30-tip2	2	J16-nn22	4	A22-catur13	5	H7-bou7	5
J2-nn8	2	J19-nn25	4	A23-catur14	5	H9-cati23	5
J21-nn27	2	P8-nn47	4	A25-cati18	5	H11-catur19	5
J22-nn28	2	C4-cati2	5	A26-bou5	5	H12-cati24	5
J23-nn29	2	C8-nn1	5	A27-catur15	5	H13-catu6	5
J24-nn30	2	C10-gc1	5	A28-mn3	5	H15-catur20	5
P1-nn40	2	C11-cr1	5	J3-nn9	5	H18-catur21	5
P2-nn41	2	C12-catur2	5	J4-nn10	5	H19-catu7	5
P3-nn42	2	C15-catur3	5	J6-nn12	5	H20-cati25	5
P5-nn44	2	C18-nn2	5	J7-nn13	5		
P7-nn46	2	C20-cati4	5	J8-nn14	5		

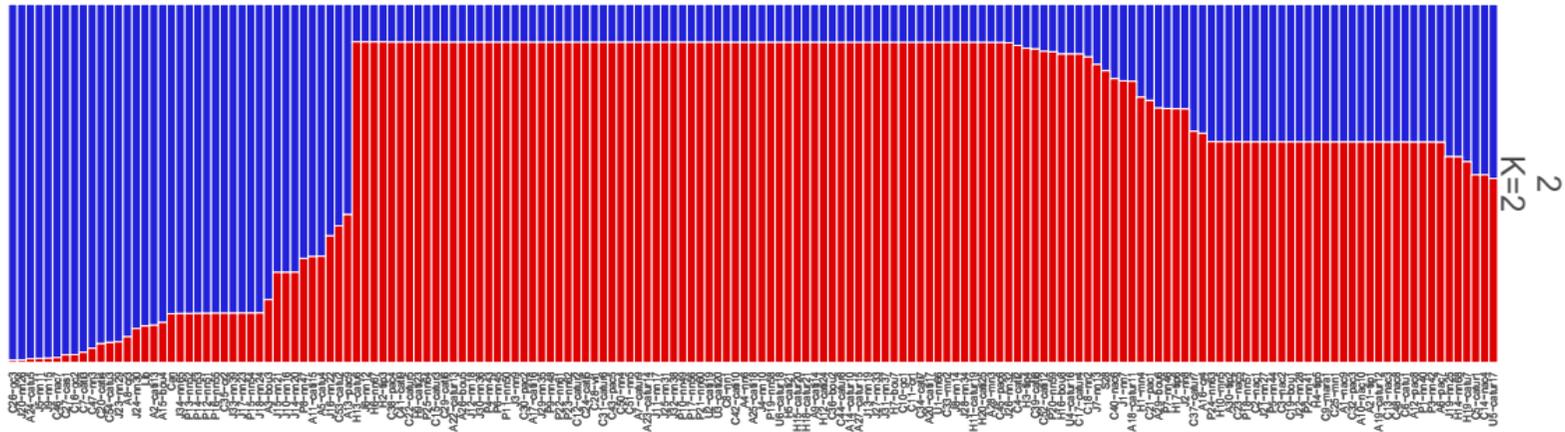
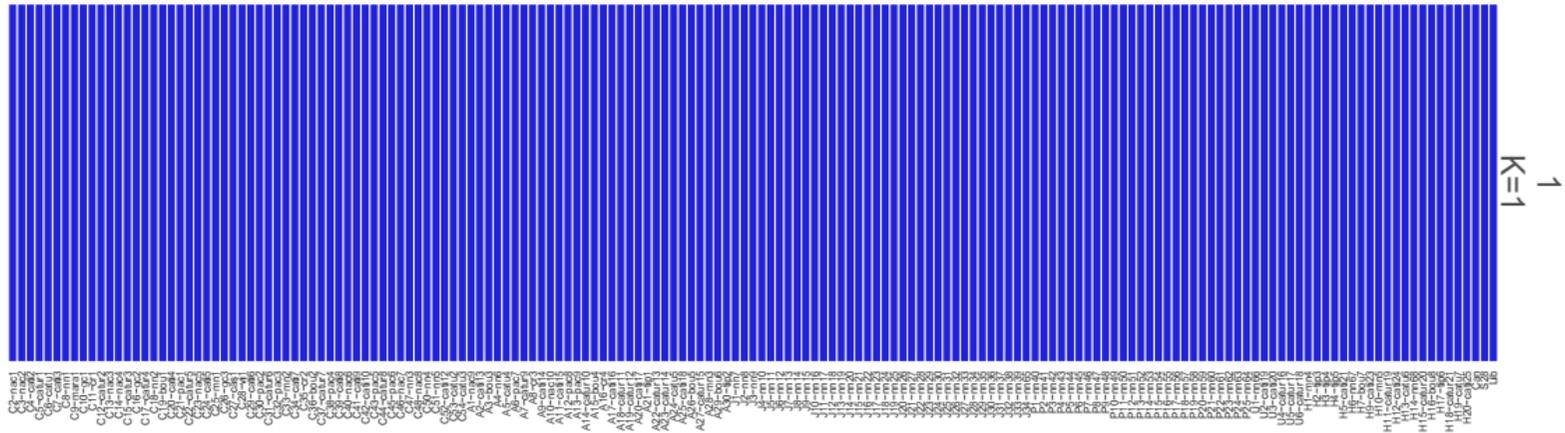
## Anexo 6: Gráficos de análisis de estructura genética con enfoque bayesiano

**K = 1 hasta K = 10, 13 loci**

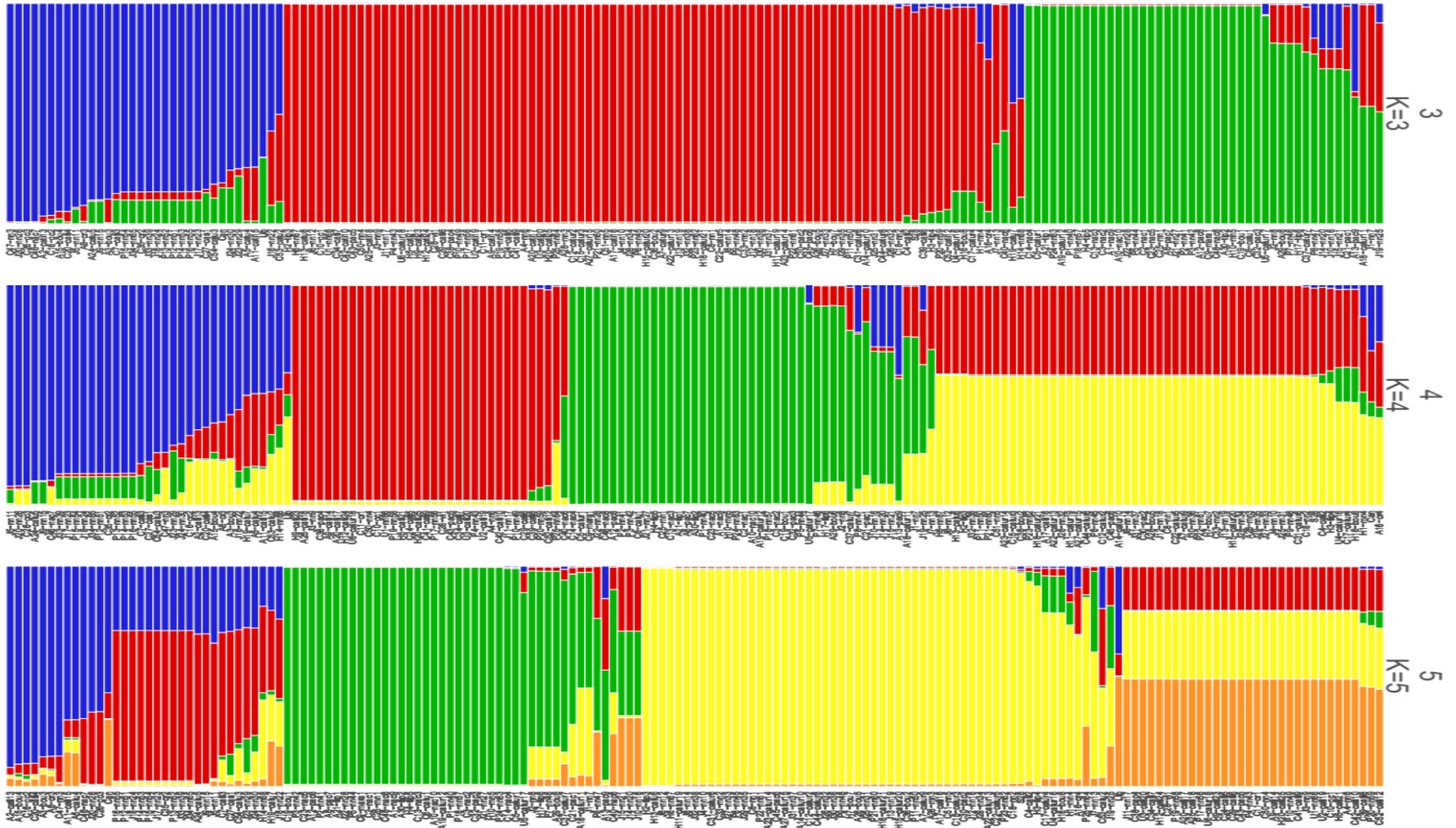


**K = 1 y K = 2, 13 loci**

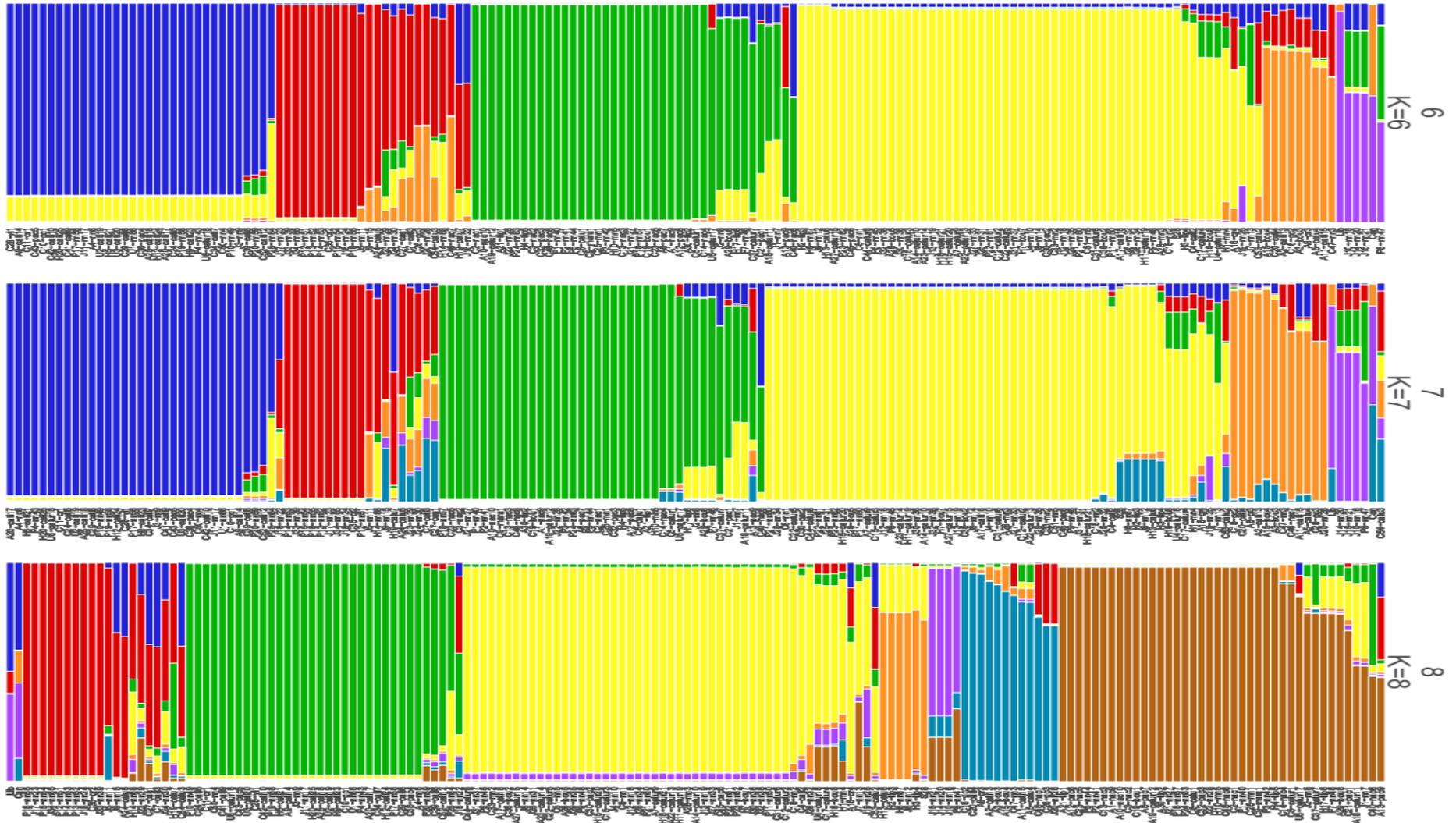
Análisis de estructura poblacional en café



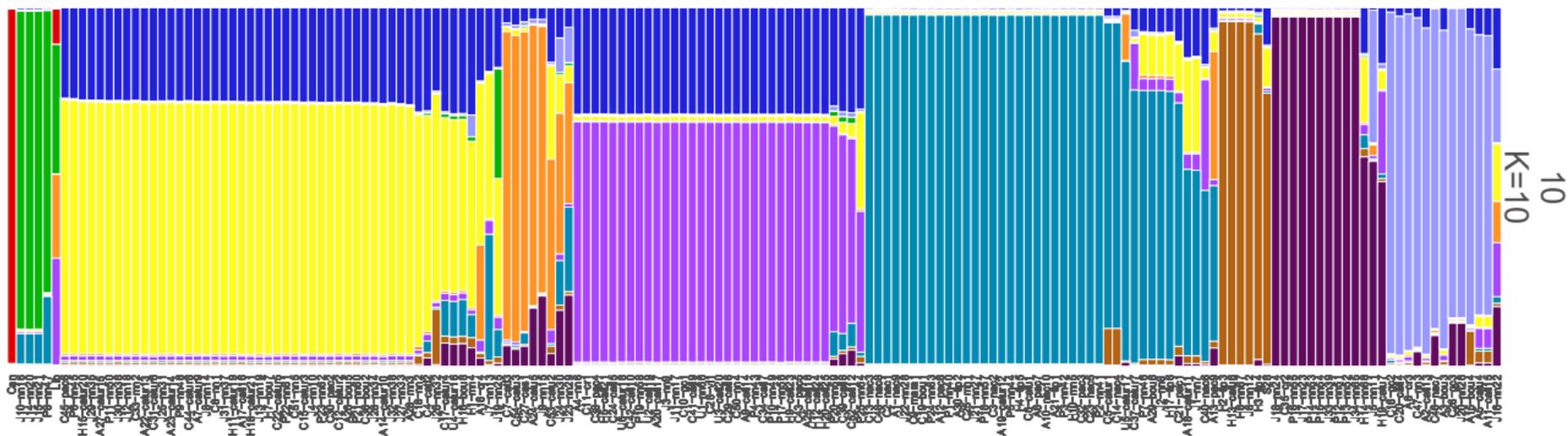
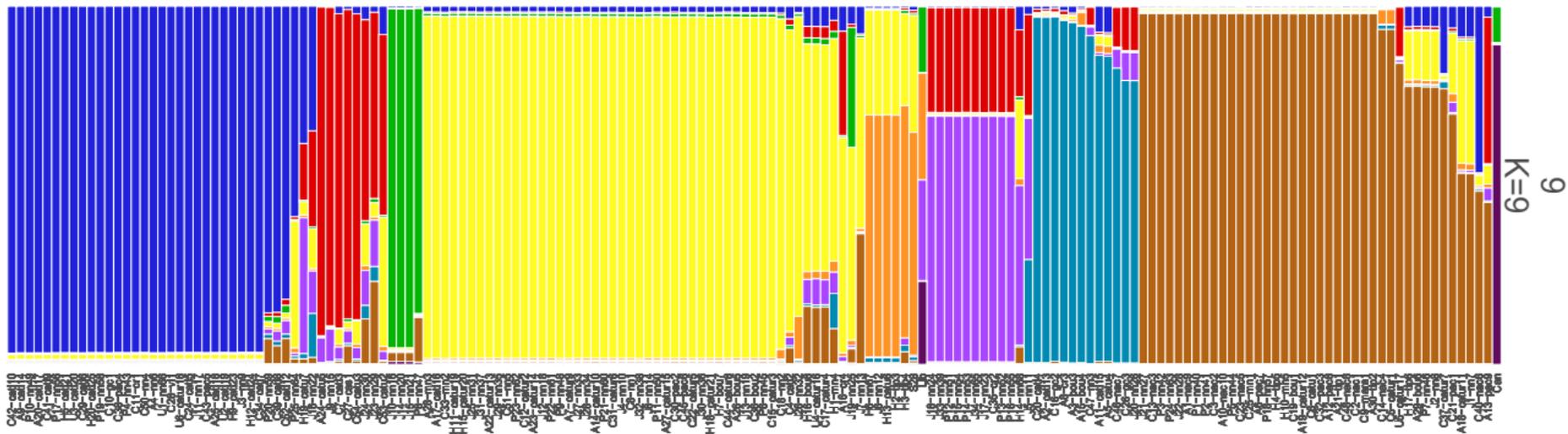
K = 3, K = 4 y K = 5, 13 loci



**K = 6, K = 7 y K = 8, 13 loci**

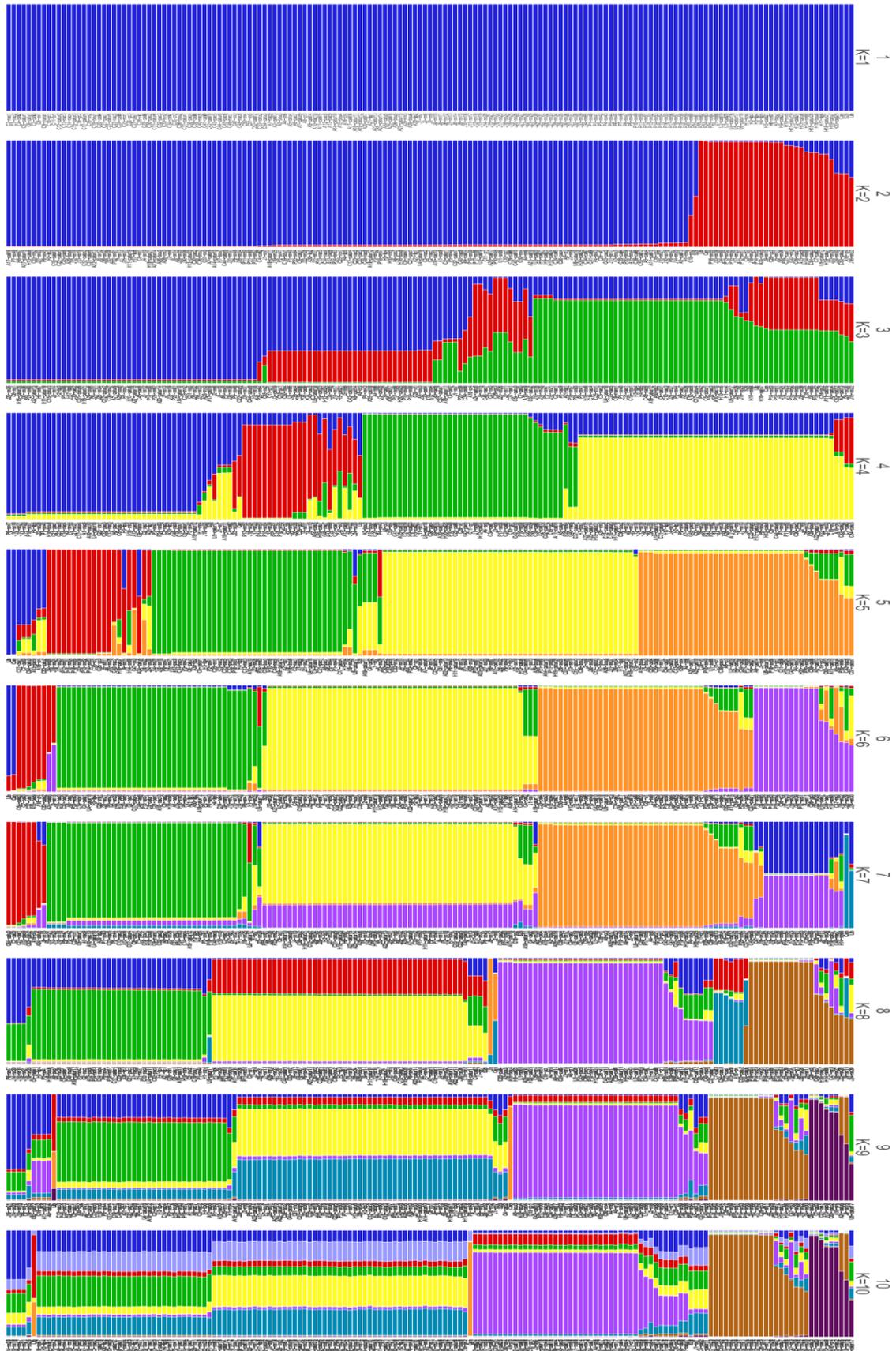


K = 9 y K = 10, 13 loci



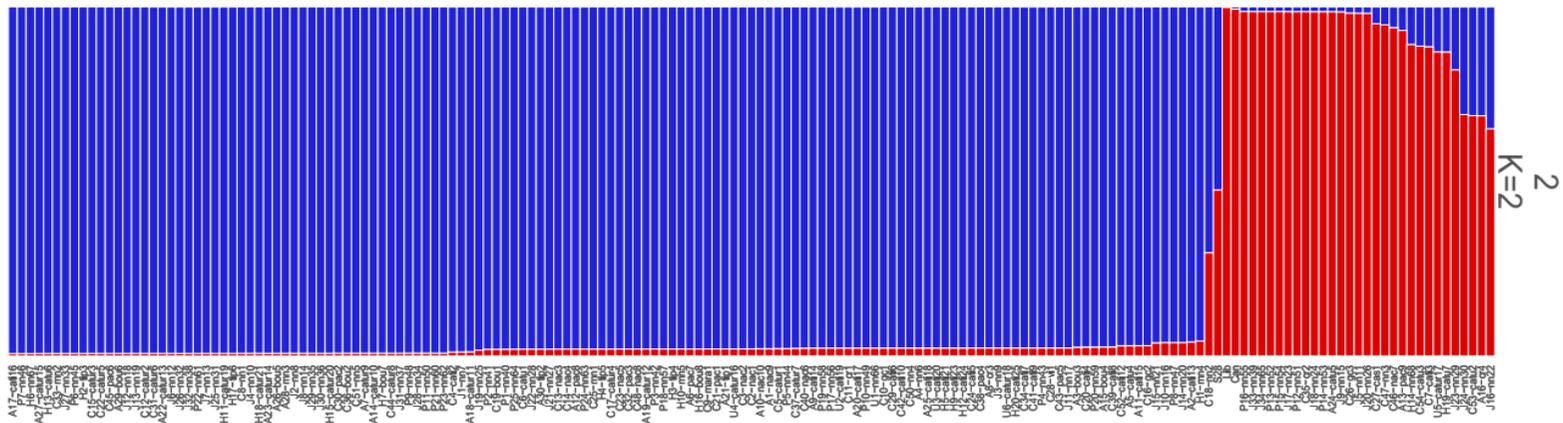
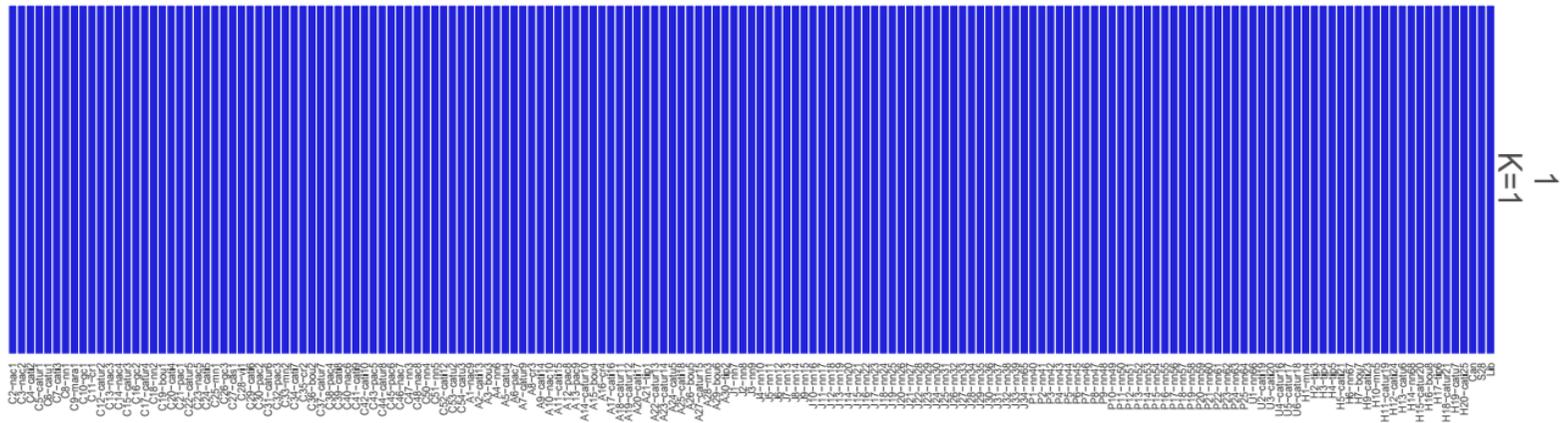
# K = 1 hasta K = 10, 8 loci

Análisis de estructura poblacional en café

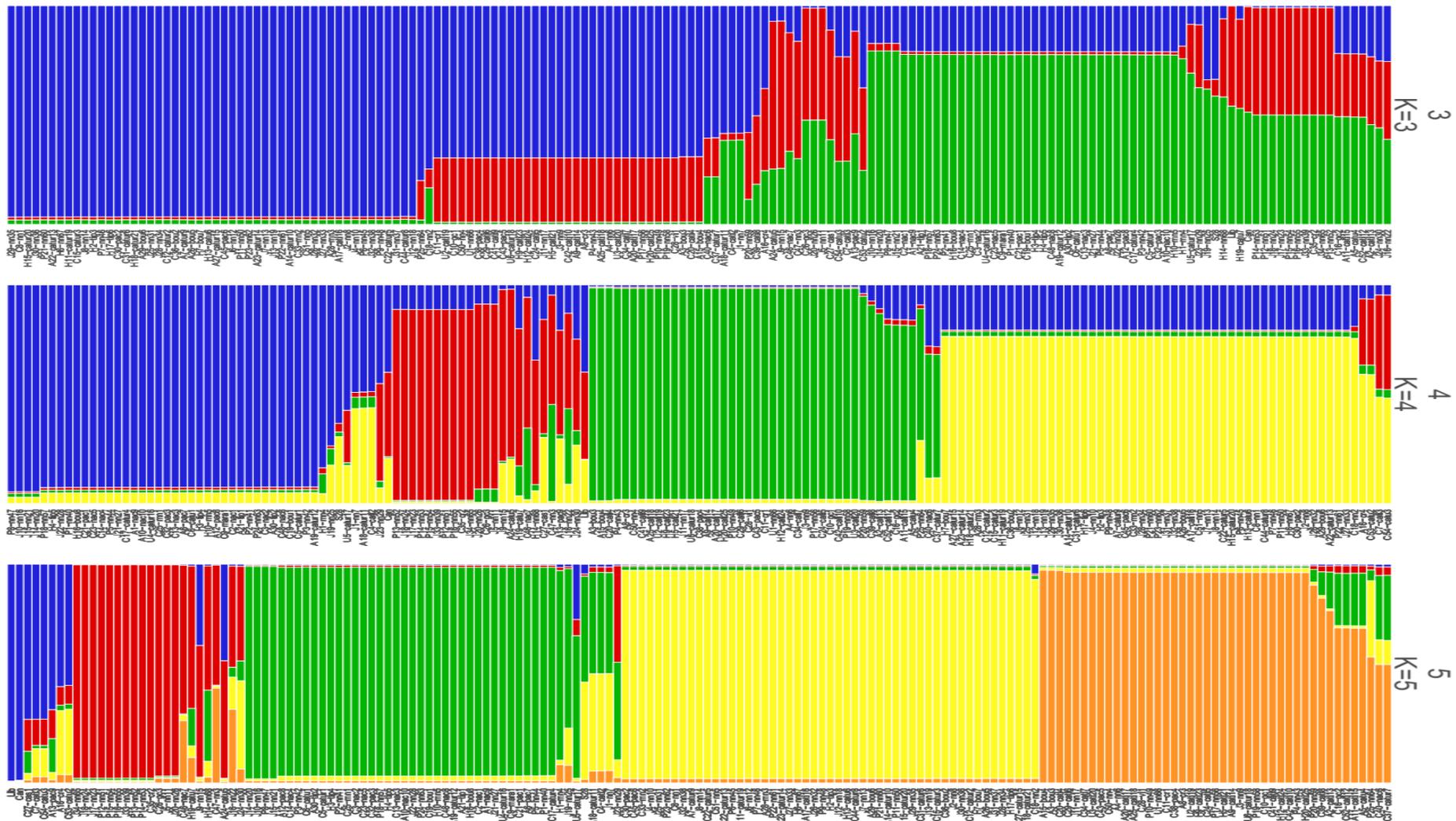


**K = 1 y K = 2, 08 loci**

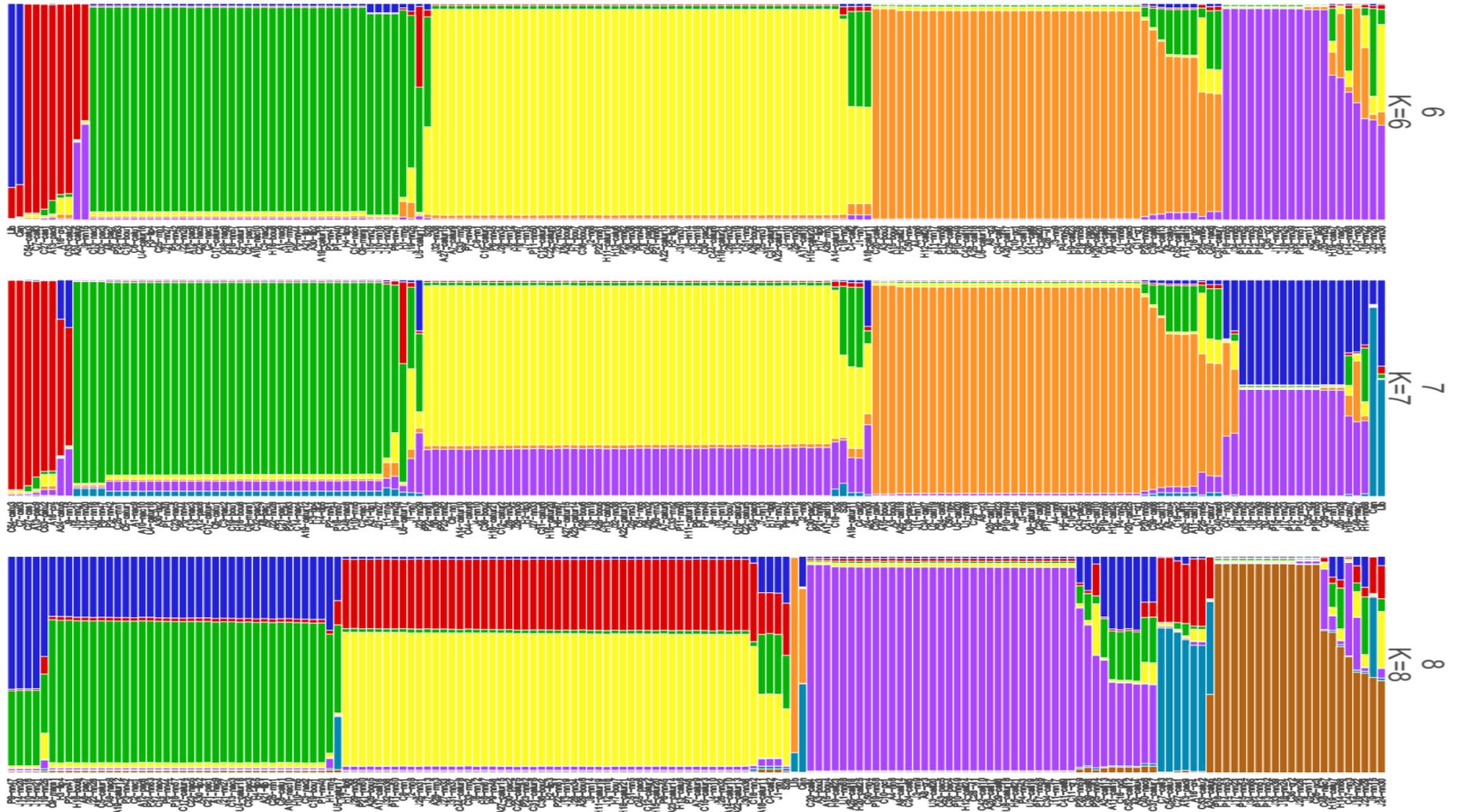
Análisis de estructura poblacional en café



K = 3, K = 4 y K = 5, 08 loci



**K = 6, K = 7 y K = 8, 08 loci**





## Anexo 7: Datos de colecta de las muestras de la colección de café del INIA

Cód. Accesoión	Cód. Para análisis	Variedad*	Departamento	Provincia	Latitud	Longitud	Altitud
2	C2-nac1	Nacional	Cajamarca	San Miguel	-6,893239	-79,065549	1137
3	C3-nac2	Nacional	Cajamarca	San Miguel	-6,893239	-79,065549	1137
4	C4-cati2	Catimor	Cajamarca	Santa Cruz	-6,857848	-79,107576	1293
5	C5-catur1	Caturra	Cajamarca	Santa Cruz	-6,857848	-79,107576	1293
6	C6-catu1	Catuai	Cajamarca	Santa Cruz	-6,857848	-79,107576	1293
7	C7-cati3	Catimor	Cajamarca	Santa Cruz	-6,857848	-79,107576	1293
8	C8-nn1	NN	Cajamarca	Santa Cruz	-6,858167	-79,108117	1305
9	C9-mara1	Maragogipe	Cajamarca	Santa Cruz	-6,858167	-79,108117	1305
10	C10-gc1	Gran Colombia	Cajamarca	Santa Cruz	-6,857233	-79,107533	1314
11	C11-cr1	Costa Rica	Cajamarca	Santa Cruz	-6,857926	-79,106752	1282
12	C12-catur2	Caturra	Cajamarca	Santa Cruz	-6,857926	-79,106752	1282
13	C13-nac3	Nacional	Cajamarca	San Miguel	-6,880955	-79,113736	1205
14	C14-nac4	Nacional	Cajamarca	San Miguel	-6,874023	-79,100736	1252
15	C15-catur3	Caturra	Cajamarca	San Miguel	-6,876770	-79,080072	1402
16	C16-gc2	Gran Colombia	Cajamarca	San Miguel	-6,876770	-79,080072	1402
17	C17-catur4	Caturra	Cajamarca	Santa Cruz	-6,874057	-79,077604	1495
18	C18-nn2	NN	Cajamarca	Santa Cruz	-6,874057	-79,077604	1495
19	C19-bou1	Bourbon	Cajamarca	Santa Cruz	-6,874057	-79,077604	1495
20	C20-cati4	Catimor	Cajamarca	Jaén	-5,977916	-79,125343	1882
21	C21-pac1	Pache	Cajamarca	Jaén	-5,977916	-79,125343	1882
22	C22-catur5	Caturra	Cajamarca	Jaén	-5,977916	-79,125343	1882
23	C23-nac5	Nacional	Cajamarca	Jaén	-5,977916	-79,125343	1882
24	C24-cati5	Catimor	Cajamarca	Jaén	-5,977916	-79,125343	1882
25	C25-mn1	Mundo novo	Cajamarca	San Ignacio	-5,155069	-79,014954	1333
26	C26-gc3	Gran Colombia	Cajamarca	San Ignacio	-5,155069	-79,014954	1333
27	C27-cas1	Castillo	Cajamarca	San Ignacio	-5,155069	-79,014954	1333
28	C28-vi1	Villasarchí	Cajamarca	San Ignacio	-5,155069	-79,014954	1333
29	C29-cati6	Catimor	Cajamarca	San Ignacio	-5,155069	-79,014954	1337
30	C30-pac2	Pache	Cajamarca	San Ignacio	-5,207294	-78,996069	1350
31	C31-catur6	Caturra	Cajamarca	San Ignacio	-5,207294	-78,996069	1350
32	C32-pac3	Pache	Cajamarca	San Ignacio	-5,315929	-78,913598	1348
33	C33-mn2	Mundo novo	Cajamarca	San Ignacio	-5,315929	-78,913598	1348
34	C34-cati7	Catimor	Cajamarca	San Ignacio	-5,315929	-78,913598	1348
35	C35-cr2	Costa Rica	Cajamarca	San Ignacio	-5,315929	-78,913598	1348
36	C36-bou2	Bourbon	Cajamarca	San Ignacio	-5,380434	-78,913551	1588
37	C37-catur7	Caturra	Cajamarca	San Ignacio	-5,380434	-78,913551	1588
38	C38-pac4	Pache	Cajamarca	San Ignacio	-5,380434	-78,913551	1588
39	C39-cati8	Catimor	Cajamarca	San Ignacio	-5,380434	-78,913551	1588
40	C40-nac6	Nacional	Cajamarca	San Ignacio	-5,380434	-78,913551	1588
41	C41-cati9	Catimor	Cajamarca	Jaén	-5,742862	-78,841070	1218

«Continuación»

42	C42-cati10	Catimor	Cajamarca	Jaén	-5,766499	-78,845051	1688
43	C43-pac5	Pache	Cajamarca	Jaén	-5,766499	-78,845051	1688
44	C44-catur8	Caturra	Cajamarca	Jaén	-5,766499	-78,845051	1688
45	C45-pac6	Pache	Cajamarca	Jaén	-5,766499	-78,845051	1688
46	C46-nac7	Nacional	Cajamarca	Jaén	-5,766499	-78,845051	1688
47	C47-nn3	NN	Cajamarca	Jaén	-5,415697	-78,577386	1688
48	C48-nac8	Nacional	Cajamarca	Jaén	-5,415697	-78,577386	1688
50	C50-nn4	NN	Cajamarca	Jaén	-5,415697	-78,577386	1688
51	C51-nn5	NN	Cajamarca	Jaén	-5,415697	-78,577386	1688
52	C52-cati12	Catimor	Cajamarca	Jaén	-5,415697	-78,577386	1688
53	C53-catu2	Catuai	Cajamarca	Jaén	-5,415697	-78,577386	1688
54	C54-catu3	Catuai	Cajamarca	Jaén	-5,415697	-78,577386	1688
55	A1-nac9	Nacional	Amazonas	Bagua	-5,624908	-78,400403	1221
56	A2-cati13	Catimor	Amazonas	Bagua	-5,627491	-78,400085	1199
57	A3-bou3	Bourbon	Amazonas	Bagua	-5,627491	-78,400085	1199
58	A4-nn6	NN	Amazonas	Bagua	-5,628032	-78,399920	1181
59	A5-catu4	Catuai	Amazonas	Bagua	-5,627491	-78,400085	1199
60	A6-pac7	Pache	Amazonas	Bagua	-5,629586	-78,387654	1410
61	A7-catur9	Caturra	Amazonas	Bagua	-5,629586	-78,387654	1410
62	A8-cr3	Costa Rica	Amazonas	Bagua	-5,628635	-78,387126	1415
63	A9-cati14	Catimor	Amazonas	Bagua	-5,661479	-78,387126	1442
64	A10-nac10	Nacional	Amazonas	Bagua	-5,661479	-78,387126	1442
65	A11-cati15	Catimor	Amazonas	Bagua	-5,661479	-78,387126	1442
66	A12-pac8	Pache	Amazonas	Bagua	-5,661479	-78,387126	1442
67	A13-pac9	Pache	Amazonas	Bagua	-5,661479	-78,387126	1442
68	A14-catur10	Caturra	Amazonas	Bagua	-5,661479	-78,387126	1442
69	A15-bou4	Bourbon	Amazonas	Bagua	-5,661479	-78,387126	1442
70	A16-cr4	Costa Rica	Amazonas	Bagua	-5,661479	-78,387126	1442
71	A17-cati16	Catimor	Amazonas	Bagua	-5,661479	-78,387126	1442
72	A18-catur11	Caturra	Amazonas	Rodríguez de Mendoza	-6,400636	-77,462606	1614
73	A19-catur12	Caturra	Amazonas	Rodríguez de Mendoza	-6,400981	-77,462346	1615
74	A20-cati17	Catimor	Amazonas	Rodríguez de Mendoza	-6,405139	-77,454641	1622
75	A21-tip1	Típica	Amazonas	Rodríguez de Mendoza	-6,405139	-77,454641	1622
76	A22-catur13	Caturra	Amazonas	Rodríguez de Mendoza	-6,403632	-77,454137	1640
77	A23-catur14	Caturra	Amazonas	Rodríguez de Mendoza	-6,494957	-77,381227	1445
78	A24-catu5	Catuai	Amazonas	Rodríguez de Mendoza	-6,494957	-77,381227	1445
79	A25-cati18	Catimor	Amazonas	Rodríguez de Mendoza	-6,494957	-77,381227	1445
80	A26-bou5	Bourbon	Amazonas	Rodríguez de Mendoza	-6,494957	-77,381227	1445
81	A27-catur15	Caturra	Amazonas	Rodríguez de Mendoza	-6,494957	-77,381227	1445

«Continuación»

82	A28-mn3	Mundo novo	Amazonas	Rodríguez de Mendoza	-6,494957	-77,381227	1445
83	A29-bou6	Bourbon	Amazonas	Rodríguez de Mendoza	-6,494957	-77,381227	1445
84	A30-tip2	Típica	Amazonas	Rodríguez de Mendoza	-6,487746	-77,378952	1316
98	J1-nn7	NN	Junín	Satipo	-11,021171	-74,304642	1408
100	J2-nn8	NN	Junín	Satipo	-11,021171	-74,304642	1408
101	J3-nn9	NN	Junín	Satipo	-11,021171	-74,304642	1408
103	J4-nn10	NN	Junín	Satipo	-11,021171	-74,304642	1408
104	J5-nn11	NN	Junín	Satipo	-11,021171	-74,304642	1408
105	J6-nn12	NN	Junín	Satipo	-11,022690	-74,301891	1383
106	J7-nn13	NN	Junín	Satipo	-11,022690	-74,301891	1383
108	J8-nn14	NN	Junín	Satipo	-11,034052	-74,303226	1399
110	J9-nn15	NN	Junín	Chanchamayo	-10,862954	-75,152605	1366
112	J10-nn16	NN	Junín	Chanchamayo	-10,862954	-75,152605	1366
113	J11-nn17	NN	Junín	Chanchamayo	-10,862954	-75,152605	1366
114	J12-nn18	NN	Junín	Chanchamayo	-10,862954	-75,152605	1366
117	J13-nn19	NN	Junín	Chanchamayo	-10,862954	-75,152605	1366
119	J14-nn20	NN	Junín	Chanchamayo	-10,862954	-75,152605	1366
120	J15-nn21	NN	Junín	Chanchamayo	-10,872511	-75,169375	1357
122	J16-nn22	NN	Junín	Chanchamayo	-10,872511	-75,169375	1357
123	J17-nn23	NN	Junín	Chanchamayo	-10,872511	-75,169375	1357
124	J18-nn24	NN	Junín	Chanchamayo	-10,875350	-75,161271	1311
125	J19-nn25	NN	Junín	Chanchamayo	-10,875350	-75,161271	1311
130	J20-nn26	NN	Junín	Chanchamayo	-10,862860	-75,193945	1600
131	J21-nn27	NN	Junín	Chanchamayo	-10,862860	-75,193945	1600
133	J22-nn28	NN	Junín	Chanchamayo	-10,862860	-75,193945	1600
134	J23-nn29	NN	Junín	Chanchamayo	-10,858325	-75,191076	1486
135	J24-nn30	NN	Junín	Chanchamayo	-10,858325	-75,191076	1486
136	J25-nn31	NN	Junín	Chanchamayo	-10,851412	-75,197764	1458
137	J26-nn32	NN	Junín	Chanchamayo	-10,851412	-75,197764	1458
138	J27-nn33	NN	Junín	Chanchamayo	-10,851412	-75,197764	1458
139	J28-nn34	NN	Junín	Chanchamayo	-10,851412	-75,197764	1458
140	J29-nn35	NN	Junín	Chanchamayo	-10,851412	-75,197764	1458
141	J30-nn36	NN	Junín	Chanchamayo	-10,851412	-75,197764	1458
142	J31-nn37	NN	Junín	Chanchamayo	-10,841381	-75,205385	1337
143	J32-nn38	NN	Junín	Chanchamayo	-10,841381	-75,205385	1337
144	J33-nn39	NN	Junín	Chanchamayo	-10,813484	-75,218724	1390
172	J34-nn65	NN	Pasco	Oxapampa	-10,722395	-75,249478	1546
145	P1-nn40	NN	Pasco	Oxapampa	-10,726009	-75,257364	1607
146	P2-nn41	NN	Pasco	Oxapampa	-10,726009	-75,257364	1607
147	P3-nn42	NN	Pasco	Oxapampa	-10,726009	-75,257364	1607
149	P4-nn43	NN	Pasco	Oxapampa	-10,726009	-75,257364	1607
150	P5-nn44	NN	Pasco	Oxapampa	-10,726009	-75,257364	1607
152	P6-nn45	NN	Pasco	Oxapampa	-10,726009	-75,257364	1607

«Continuación»

153	P7-nn46	NN	Pasco	Oxapampa	-10,726009	-75,257364	1607
154	P8-nn47	NN	Pasco	Oxapampa	-10,726009	-75,257364	1607
155	P9-nn48	NN	Pasco	Oxapampa	-10,726009	-75,257364	1607
156	P10-nn49	NN	Pasco	Oxapampa	-10,726009	-75,257364	1607
157	P11-nn50	NN	Pasco	Oxapampa	-10,726009	-75,257364	1607
158	P12-nn51	NN	Pasco	Oxapampa	-10,726009	-75,257364	1607
159	P13-nn52	NN	Pasco	Oxapampa	-10,726009	-75,257364	1607
160	P14-nn53	NN	Pasco	Oxapampa	-10,726009	-75,257364	1607
161	P15-nn54	NN	Pasco	Oxapampa	-10,726009	-75,257364	1607
162	P16-nn55	NN	Pasco	Oxapampa	-10,726009	-75,257364	1607
163	P17-nn56	NN	Pasco	Oxapampa	-10,722395	-75,249478	1546
164	P18-nn57	NN	Pasco	Oxapampa	-10,722395	-75,249478	1546
165	P19-nn58	NN	Pasco	Oxapampa	-10,722395	-75,249478	1546
166	P20-nn59	NN	Pasco	Oxapampa	-10,722395	-75,249478	1546
167	P21-nn60	NN	Pasco	Oxapampa	-10,722395	-75,249478	1546
168	P22-nn61	NN	Pasco	Oxapampa	-10,722395	-75,249478	1546
169	P23-nn62	NN	Pasco	Oxapampa	-10,722395	-75,249478	1546
171	P24-nn63	NN	Pasco	Oxapampa	-10,722395	-75,249478	1546
173	P25-nn64	NN	Junín	Chanchamayo	-10,872511	-75,169375	1357
174	U1-nn66	NN	Ucayali	Padre Abad	-9,174139	-75,807250	1483
183	U2-cati19	Catimor	Ucayali	Padre Abad	-9,174583	-75,807167	1493
199	U3-cati20	Catimor	Ucayali	Padre Abad	-9,174222	-75,810139	1373
200	U4-catur16	Caturra	Ucayali	Padre Abad	-9,173667	-75,809528	1399
203	U5-catur17	Caturra	Ucayali	Padre Abad	-9,147889	-75,788000	1245
205	U6-catur18	Caturra	Ucayali	Padre Abad	-9,147806	-75,788139	1249
212	H1-mn4	Mundo novo	Huánuco	Leoncio Prado	-9,192389	-75,815778	1428
213	H2-tip3	Típica	Huánuco	Leoncio Prado	-9,186167	-75,817500	1357
214	H3-tip4	Típica	Huánuco	Leoncio Prado	-9,186111	-75,817444	1360
216	H4-tip5	Típica	Huánuco	Leoncio Prado	-9,189000	-75,819972	1409
218	H5-cati21	Catimor	Huánuco	Leoncio Prado	-9,189000	-75,820083	1410
219	H6-nn67	NN	Huánuco	Leoncio Prado	-9,189056	-75,820139	1417
220	H7-bou7	Bourbon	Huánuco	Leoncio Prado	-9,185111	-75,825000	1410
226	H9-cati23	Catimor	Huánuco	Leoncio Prado	-9,401944	-75,918917	1591
228	H10-mn5	Mundo novo	Huánuco	Leoncio Prado	-9,401833	-75,918972	1587
229	H11-catur19	Caturra	Huánuco	Leoncio Prado	-9,401833	-75,919056	1585
230	H12-cati24	Catimor	Huánuco	Leoncio Prado	-9,401750	-75,919167	1579
232	H13-catu6	Catuai	Huánuco	Leoncio Prado	-9,401806	-75,919000	1583
233	H14-nn68	NN	Huánuco	Leoncio Prado	-9,402750	-75,918528	1581

«Continuación»

236	H15-catur20	Caturra	Huánuco	Leoncio Prado	-9,402750	-75,918472	1577
237	H16-bou8	Bourbon	Huánuco	Leoncio Prado	-9,401083	-75,921278	1509
242	H17-tip6	Típica	Huánuco	Leoncio Prado	-9,400833	-75,921111	1506
245	H18-catur21	Caturra	Huánuco	Leoncio Prado	-9,399333	-75,923389	1479
248	H19-catu7	Catuai	Huánuco	Huánuco	-9,677333	-75,084306	2138
249	H20-cati25	Catimor	Huánuco	Huánuco	-9,677306	-76,084250	2131