

PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR

**FACULTAD DE HÁBITAT, INFRAESTRUCTURA Y
CREATIVIDAD**

MAESTRÍA EN BIOLOGÍA COMPUTACIONAL

Análisis de la genética poblacional de *Coffea canephora*

Autor: Patricio Macas-Pogo

Director del trabajo de titulación: Dra. González García Laura Natalia

Quito, 2025

AGRADECIMIENTOS

Agradezco a la Pontificia Universidad Católica del Ecuador y a los docentes de la maestría por haberme brindado los recursos académicos y las enseñanzas que hicieron posible la realización de este trabajo de titulación.

Agradezco a mi tutora, Laura Natalia González García, por sus clases, recomendaciones y paciencia a lo largo de todo el proceso de titulación. Sus conocimientos fueron fundamentales para el desarrollo de este trabajo.

Agradezco infinitamente la oportunidad de habitar este planeta, a mis padres, Carmen y Franco, por su amor, respeto y ternura, por la fuerza y la valentía que me inspiran cada día para seguir caminando las montañas.

DEDICATORIA

A quienes son mi historia de vida, futuras aventuras...

*A **Ramona** y **Logan**.*

ÍNDICE GENERAL

RESUMEN	9
ABSTRACT	10
1. INTRODUCCIÓN	11
1.1. Contexto y relevancia de <i>Coffea canephora</i> en la industria del café.....	11
1.2. Descripción de <i>Coffea canephora</i>	11
1.3. Importancia de la genética poblacional	12
1.4. Alcances y limitaciones del estudio	14
2. METODOLOGÍA	16
2.1. Recopilación de datos genómicos	16
2.2. Obtención de datos	17
2.3. Flujo de procesamiento de datos: control de calidad, alineamiento y llamado de variantes	18
2.4. Análisis de Diversidad genética y estructura poblacional: heterocigosidad, agrupamiento y diferenciación	20
2.5. Dendrograma Neighbor-Joining y matriz a partir de la ASD.....	22
3. RESULTADOS	23
3.1. Diversidad genética y estructura poblacional.....	23
3.1.1. Distribución de la heterocigosidad observada (H_o)	23
3.1.2. Heterocigosidad observada (H_o) por país	24
3.1.3. Varianza por cada componente principal	25
3.1.4. Agrupamiento genético y diferenciación poblacional (PCA general)	26
3.1.5. PCA de <i>C. canephora</i> por país.....	27

3.2. Estructura poblacional inferida a partir de distancias genéticas.....	28
3.2.1. Agrupamiento mediante Neighbor-Joining (ASD)	28
3.2.2. Análisis de estructura/afinidad genética.....	29
4. DISCUSIÓN.....	31
4.1. Interpretación de los resultados en el contexto de la genética poblacional..	31
4.2. Implicaciones para la conservación de <i>Coffea canephora</i>.....	33
4.3. Posibles aplicaciones en programas de mejoramiento genético	34
4.4. Limitaciones del estudio y recomendaciones para investigaciones futuras..	34
4.5. Avances genómicos y su aplicación en estudios poblacionales de Coffea	35
5. CONCLUSIONES Y RECOMENDACIONES.....	37
6. BIBLIOGRAFÍA.....	39
7. APÉNDICES	43

ÍNDICE DE FIGURAS

Figura 1. Detalles de hojas, flores y frutos de <i>Coffea canephora</i> . Fotografías: Lisa Hester e Irfan Nurarifin (iNaturalist, CC BY-NC).	12
Figura 2. Diagrama circular de sintenia que compara el genoma de <i>Coffea arabica</i> (CA, izquierda) con sus especies progenitoras <i>C. canephora</i> (CC, arriba derecha) y <i>C. eugenioides</i> (CE, abajo derecha). Las líneas internas indican bloques de genes conservados entre los cromosomas. Tomado de Salojärvi et al., 2024.	14
Figura 3. Pipeline de la obtención y preparación de datos en Galaxy para el análisis genómico de las secuencias.	20
Figura 4. Pipeline de la obtención de la Heterocigosidad Observada (Ho) y Componentes Principales (PCA).....	21
Figura 5. Pipeline para obtener el Dendrograma Neighbor-Joining y matriz de distancia genética (ASD).	23
Figura 6. Distribución de la heterocigosidad observada (Ho) en <i>C. canephora</i>	24
Figura 7. Heterocigosidad observada (Ho) en 20 muestras de <i>C. canephora</i> agrupadas por país..	25
Figura 8. Gráfico de varianza explicada por los diez primeros componentes principales..	26
Figura 9. Análisis de componentes principales (PCA) en 20 muestras de <i>C. canephora</i> . Distribución de las muestras en función de los dos primeros componentes principales (PC1 y PC2).....	27
Figura 10. Análisis de componentes principales (PCA) en 20 muestras de <i>C. canephora</i> coloreadas por país..	28

Figura 11. Dendrograma de agrupamiento obtenido con el método Neighbor-Joining (NJ) a partir de la distancia genética ASD en 20 muestras de *C. canephora*. Se observa una separación marcada entre las muestras de México (rojo) y Brasil (verde). Soporte de ramas por bootstrap con b=200 réplicas. 29

Figura 12. Heatmap de la matriz de distancias genéticas calculada con la métrica ASD en 20 muestras de *C. canephora*. La escala de color indica la distancia genética (0 = mayor similitud, 1 = mayor distancia). Los colores de la barra superior y lateral representan el país de origen (verde = Brasil, rojo = México)..... 30

INDICE DE TABLAS

Tabla 1. Muestras de <i>C. canephora</i> obtenidas en NCBI.	16
Tabla 2. Descripción de las muestras de <i>C. canephora</i> utilizadas en el análisis.....	18

RESUMEN

Coffea canephora, conocido como café robusta, es una de las principales especies cultivadas de café a nivel mundial, valorada por su resistencia a enfermedades, alto contenido de cafeína y adaptabilidad a condiciones ambientales adversas. Sin embargo, su diversidad genética en América Latina aún no ha sido caracterizada en profundidad, lo que limita las estrategias de conservación y mejoramiento. El presente estudio analizó la variabilidad genética de 20 muestras de *C. canephora* procedentes de México y Brasil mediante secuenciación Illumina, usando como referencia el genoma de la especie y un flujo bioinformático basado en Galaxy, PLINK y RStudio. Las muestras se seleccionaron aleatoriamente por tratarse de un estudio exploratorio y comparativo, orientado a una primera aproximación de la variabilidad genética. El número de muestras fue limitado debido a la alta capacidad computacional requerida para analizar un conjunto de datos más amplio. Se evaluaron métricas de heterocigosidad, componentes principales (PCA), distancias genéticas (ASD) y agrupamiento mediante Neighbor-Joining. Los resultados revelaron una distribución bimodal de la heterocigosidad observada, con genotipos altamente heterocigotos ($H_o \approx 0.80-0.90$) y otros con valores reducidos ($H_o \approx 0.10-0.15$). El PCA explicó el 82 % de la variación genética ($PC1 = 75.2\%$), mostrando una clara diferenciación entre muestras mexicanas (más homogéneas, ASD promedio ≈ 0.12) y brasileñas (más heterogéneas, ASD promedio ≈ 0.18). Además, se identificaron outliers genéticos, como el par M11–M16, que sugieren posibles eventos de flujo génico o introducciones cruzadas. Los resultados aportan información clave sobre la estructura poblacional de *C. canephora* y constituyen un primer paso hacia la integración de la diversidad regional en estrategias globales de conservación y programas de mejoramiento genético orientados a enfrentar los desafíos del cambio climático y fortalecer la sostenibilidad de la caficultura.

ABSTRACT

Coffea canephora, commonly known as Robusta coffee, is one of the main cultivated coffee species worldwide, valued for its disease resistance, high caffeine content, and adaptability to adverse environmental conditions. However, its genetic diversity in Latin America has not yet been thoroughly characterized, which limits conservation and breeding strategies. The present study analyzed the genetic variability of 20 *C. canephora* samples from Mexico and Brazil using Illumina sequencing, with the species genome as reference and a bioinformatic workflow based on Galaxy, PLINK, and RStudio. Samples were selected randomly, as this was an exploratory and comparative study aimed at providing an initial assessment of genetic variability. The number of samples was limited due to the high computational capacity required to analyze a larger dataset. Heterozygosity metrics, principal component analysis (PCA), allele-sharing distances (ASD), and clustering through Neighbor-Joining were evaluated. The results revealed a bimodal distribution of observed heterozygosity, with highly heterozygous genotypes ($H_o \approx 0.80\text{--}0.90$) and others with reduced values ($H_o \approx 0.10\text{--}0.15$). The PCA explained 82% of the genetic variation (PC1 = 75.2%), showing a clear differentiation between Mexican samples (more homogeneous, ASD mean ≈ 0.12) and Brazilian samples (more heterogeneous, ASD mean ≈ 0.18). In addition, genetic outliers such as the M11–M16 pair were identified, suggesting possible events of gene flow or cross-introductions. Overall, these results provide key insights into the population structure of *C. canephora* and represent an initial step toward integrating regional diversity into global conservation strategies and breeding programs aimed at addressing the challenges of climate change and strengthening the sustainability of coffee cultivation.

1. INTRODUCCIÓN

1.1. Contexto y relevancia de *Coffea canephora* en la industria del café

Coffea canephora, comúnmente conocida como café robusta, es una de las dos especies principales cultivadas comercialmente, junto con *Coffea arabica*. Originaria de África occidental, fue descrita por primera vez por el botánico Louis Pierre en 1895. Esta especie se caracteriza por su resistencia a enfermedades y su adaptabilidad a climas cálidos y húmedos, lo que la hace adecuada para regiones de baja altitud (Zambrano *et al.*, 2018).

En la industria del café, *C. canephora* es valorada por su alto rendimiento y mayor contenido de cafeína en comparación con *C. arabica*. Especialistas en química del café han estimado que los granos de Robusta contienen aproximadamente 2.09 % de cafeína, frente a aproximadamente 1.09 % en Arabica (Zainuri *et al.*, 2023). Se utiliza frecuentemente en la producción de café instantáneo y en mezclas para aportar cuerpo y amargor. Además, su cultivo en altitudes más bajas (200–800 m s. n. m.) y su resistencia a plagas y enfermedades la hacen especialmente adecuada para zonas donde Arabica no prospera (Sahachairungrueng *et al.*, 2022). Estas características contribuyen significativamente a la diversificación y sostenibilidad de la producción cafetera.

1.2. Descripción de *Coffea canephora*

Comúnmente conocida como café robusta, es una especie del género *Coffea* originaria de África occidental y central. Es un arbusto perenne que puede alcanzar hasta 10 metros de altura en su estado natural, aunque en cultivo generalmente se mantiene podado para facilitar la cosecha. Las hojas son elípticas, de color verde oscuro y brillantes, con márgenes ondulados. Las flores son blancas y fragantes, agrupadas en inflorescencias axilares. Los frutos son drupas que pasan de verdes a rojas o púrpuras al madurar, cada uno conteniendo generalmente dos semillas conocidas como granos de café (De Paula *et al.*, 2025).



Figura 1. Detalles de hojas, flores y frutos de *C. canephora*. Fotografías: Lisa Hester e Irfan Nurarifin (iNaturalist, CC BY-NC).

El café robusta es apreciado por su alta resistencia a enfermedades y su capacidad de adaptarse a climas cálidos y húmedos, especialmente en altitudes bajas (200–800 metros sobre el nivel del mar). Estudios morfológicos y genéticos han demostrado que esta especie presenta una notable plasticidad fenotípica, lo que le permite crecer en ambientes diversos, desde bosques húmedos en África central hasta plantaciones comerciales en Asia y América Latina (Loor-Solórzano et al., 2017). A diferencia de *C. arabica*, contiene casi el doble de cafeína, lo que contribuye a su sabor más fuerte y amargo. Este mayor contenido de cafeína, cercano al 2.2–2.7 %, ha sido confirmado en múltiples análisis químicos comparativos (Spinoso-Castillo et al., 2022; Zaidan et al., 2025). Esta característica lo hace popular en la producción de café instantáneo y en mezclas para aportar cuerpo y amargor.

1.3. Importancia de la genética poblacional

La genética poblacional de *C. canephora* es clave para comprender su diversidad genética, estructura y adaptación; esta información es esencial para identificar rasgos como resistencia a enfermedades, tolerancia a estreses abióticos y calidad de bebida (Zaidan et al., 2025; Sant’Ana et al., 2018).

El primer genoma de *C. canephora* fue publicado por Denoeud et al. (2014), ensamblado a partir de un acceso doble haploide con tecnologías Roche 454, Sanger e Illumina. Este

genoma abarcó ~568 Mb, permitió anotar más de 25 000 genes y anclar cerca del 64 % de la secuencia a los 11 cromosomas. A partir de esta referencia se han desarrollado estudios de diversidad y estructura poblacional (Depecker et al., 2023). Estudios en la República Democrática del Congo han evidenciado que las poblaciones silvestres conservan altos niveles de variabilidad y un exceso de heterocigotos, lo que las convierte en reservorios genéticos fundamentales para la caficultura (Vanden-Abeelee et al., 2021; Depecker et al., 2023).

En Ecuador, investigaciones de Loor-Solórzano et al. (2017) demostraron que el germoplasma de *C. canephora* introducido en el país proviene mayoritariamente de los grupos genéticos congolese (SG1 y SG2), mostrando una notable diversidad fenotípica y genética en las colecciones vivas del INIAP. Este acervo representa una base estratégica para programas de mejoramiento y conservación adaptados a las condiciones locales.

En la Amazonía ecuatoriana, se ha demostrado que *C. canephora* puede cultivarse tanto en sistemas agroforestales como en monocultivo, con rendimientos similares entre manejos orgánicos y convencionales, aunque los primeros ofrecen mayores beneficios ambientales y de reciclaje de nutrientes (Tinoco-Jaramillo et al., 2024). Otros trabajos de GWAS, como el de Zaidan et al. (2025), han identificado SNPs asociados a resistencia a plagas (minador de la hoja, roya) y rasgos de desarrollo en *C. canephora*, lo que confirma el potencial de esta especie para programas de selección asistida por marcadores.

Un estudio en México caracterizó genotipos elite de *C. canephora* por su tolerancia a la sequía, identificando aquellos con mayor potencial de adaptación a condiciones climáticas adversas. Además, se han desarrollado nuevos cultivares híbridos en la Amazonia brasileña, con el potencial de triplicar la productividad en esa región (Quintana-Escobar et al., 2017; Toapanta et al., 2023).

Además, los avances genómicos brindan un soporte imprescindible: el genoma de *C. canephora* (Robusta) se ha ensamblado en 672 Mb, con 11 pseudo-cromosomas que representan el 82.7 % del tamaño proyectado, alrededor de 28 880 genes identificados, elementos transponibles que ocupan el 67.5 % del genoma y una integridad del genoma del 96 % segúnfi (Salojärvi et al., 2024). Este nuevo ensamblado, generado mediante tecnologías de lecturas largas (PacBio HiFi) y scaffolding con Hi-C, ofrece resolución haplotípica y mayor continuidad estructural, superando las limitaciones del genoma publicado por Denoeud et al. (2014) basado en lecturas cortas. Por ello, repetir estudios

poblacionales con las muestras ya publicadas, pero ahora alineadas contra esta nueva referencia, constituye una oportunidad para detectar variación estructural y reevaluar la diversidad genética con mayor precisión. Los valiosos datos que se obtienen con las tecnologías actuales facilitan la integración de herramientas como GWAS y la selección asistida por marcadores, lo que acelera el desarrollo de variedades más resilientes y sostenibles frente al cambio climático (Ferrão et al., 2023).

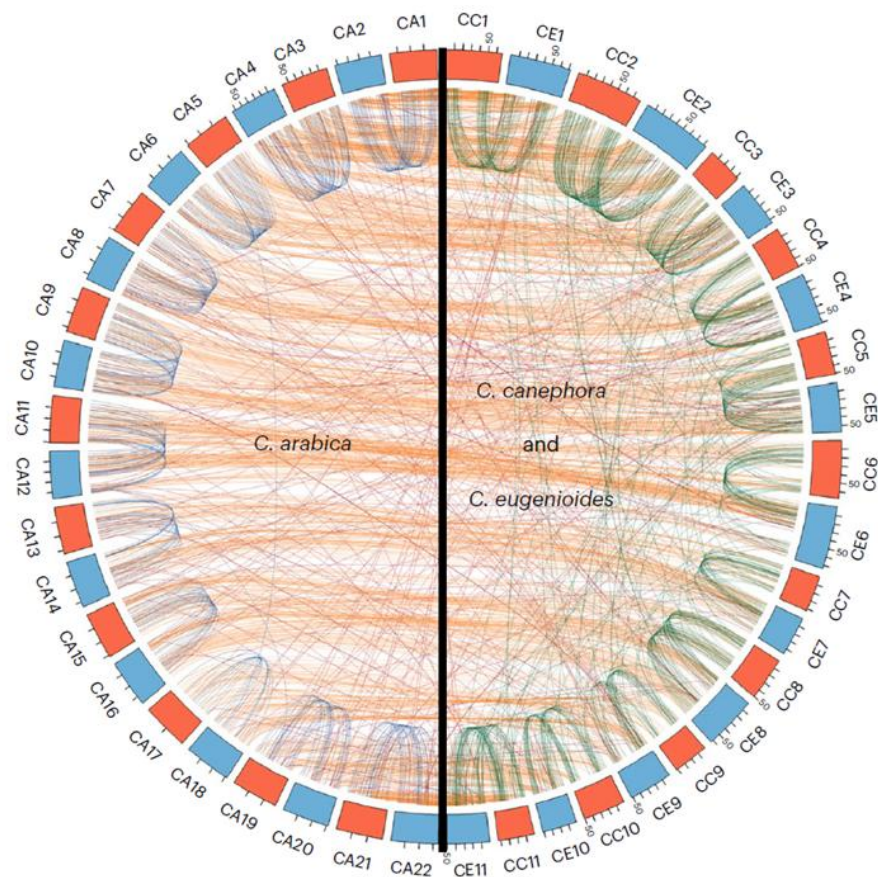


Figura 2. Diagrama circular de sintenia que compara el genoma de *C. arabica* (CA, izquierda) con sus especies progenitoras *C. canephora* (CC, arriba derecha) y *C. eugenioides* (CE, abajo derecha). Las líneas internas indican bloques de genes conservados entre los cromosomas. Tomado de Salojärvi et al., 2024.

1.4. Alcances y limitaciones del estudio

La investigación en genética poblacional de *C. canephora* se vuelve fundamental, ya que permite identificar rasgos de resistencia y adaptación, así como optimizar la gestión de recursos genéticos. El análisis de la diversidad y de la estructura poblacional también

proporciona herramientas para orientar programas de conservación dinámica, que integren tanto colecciones ex situ como poblaciones en paisajes productivos, garantizando la permanencia de la variabilidad genética y el flujo génico en el tiempo (Vanden Abeele et al., 2021; Depecker et al., 2023).

Este tipo de estudios se pueden realizar en la plataforma bioinformática Galaxy, que integra diversas herramientas para análisis genómicos. Entre ellas se encuentra HISAT2, que permite la alineación de lecturas cortas o fragmentadas contra genomas de referencia, ofreciendo un mapeo rápido y confiable incluso entre especies no idénticas (Afgan et al., 2018; Kim et al., 2019). A partir de los alineamientos obtenidos, se pueden identificar variantes genéticas como SNPs y genes que podrían estar conservados, lo que permite explorar la distribución de dichas variantes en el contexto del genoma de referencia (Li et al., 2009).

No obstante, estudios como este pueden verse afectados por ciertas limitaciones. Un tamaño de muestra reducido puede generar estimaciones sesgadas de la diversidad genética, al subrepresentar alelos raros y limitar la precisión en el cálculo de parámetros como la heterocigosidad y la diferenciación poblacional. Esto puede conducir a una subestimación de la variación real dentro y entre poblaciones, así como afectar la detección de señales sutiles de estructura genética o selección natural. En consecuencia, los patrones observados podrían reflejar más el efecto del muestreo que la verdadera distribución genética de la especie (Hale et al., 2012). También debemos tener en cuenta que la caracterización fenotípica se concentró en dos variables principales: resistencia a enfermedades y contenido de cafeína, dejando fuera otros metabolitos de calidad y rasgos asociados a la tolerancia a estreses abióticos. Finalmente, la gran cantidad de información generada y procesada en este tipo de análisis representa un desafío técnico importante, que puede limitar la profundidad y la rapidez de los resultados, dada la necesidad de infraestructura computacional robusta y métodos bioinformáticos especializados (Mardis, 2017; Schatz et al., 2010; Afgan et al., 2018).

Los resultados de este estudio ofrecen una aproximación inicial a la variabilidad genética de *C. canephora*, constituyendo una referencia para futuros trabajos. Dado su carácter exploratorio y comparativo, el estudio se centra en describir patrones de diversidad genética y posibles diferenciaciones entre poblaciones, más que en proponer aplicaciones directas de manejo o selección genética (Holderegger & Wagner, 2008).

2. METODOLOGÍA

2.1. Recopilación de datos genómicos

Se realizó una revisión en la base de datos The National Center for Biotechnology Information (NCBI) para identificar las secuencias cortas de *C. canephora* obtenidas mediante tecnología Illumina, así como el genoma completo de referencia de esta especie. En la Tabla 1 se presentan las muestras seleccionadas de forma aleatoria, que sirvieron como base para los análisis posteriores.

Tabla 1. Muestras de *C. canephora* obtenidas en NCBI.

Nº	Código SRR	Código SRA	Plataforma	País/Institución	Estudio/Referencia
Muestra 1	SRR5074356	SRX2393403	Illumina HiSeq 1000	Brasil – Univ. Federal de Río de Janeiro	PRJNA353111
Muestra 2	SRR5074358	SRX2393405	Illumina HiSeq 1000	Brasil – Univ. Federal de Río de Janeiro	PRJNA353111
Muestra 3	SRR5074359	SRX2393406	Illumina HiSeq 1000	Brasil – Univ. Federal de Río de Janeiro	PRJNA353111
Muestra 4	SRR5074360	SRX2393407	Illumina HiSeq 1000	Brasil – Univ. Federal de Río de Janeiro	PRJNA353111
Muestra 5	SRR5074361	SRX2393408	Illumina HiSeq 1000	Brasil – Univ. Federal de Río de Janeiro	PRJNA353111
Muestra 6	SRR5074362	SRX2393409	Illumina HiSeq 1000	Brasil – Univ. Federal de Río de Janeiro	PRJNA353111
Muestra 7	SRR5074363	SRX2393410	Illumina HiSeq 1000	Brasil – Univ. Federal de Río de Janeiro	PRJNA353111
Muestra 8	SRR5074365	SRX2393412	Illumina HiSeq 1000	Brasil – Univ. Federal de Río de Janeiro	PRJNA353111
Muestra 9	SRR5074366	SRX2393413	Illumina HiSeq 1000	Brasil – Univ. Federal de Río de Janeiro	PRJNA353111
Muestra 10	SRR5074368	SRX2393415	Illumina HiSeq 1000	Brasil – Univ. Federal de Río de Janeiro	PRJNA353111
Muestra 11	SRR5074370	SRX2393417	Illumina HiSeq 1000	Brasil – Univ. Federal de Río de Janeiro	PRJNA353111

Muestra 12	SRR5074371	SRX2393418	Illumina HiSeq 1000	Brasil – Univ. Federal de Río de Janeiro	PRJNA353111
Muestra 13	SRR5417819	SRX2709373	Illumina MiSeq	Brasil – Universidad Federal de Viçosa	PRJNA353233
Muestra 14	SRR32606963	SRX27912512	Illumina HiSeq 2000	Mexico – Centro de Investigación Científica de Yucatán	PRJNA1230212
Muestra 15	SRR32606964	SRX27912511	Illumina HiSeq 2000	Mexico – Centro de Investigación Científica de Yucatán	PRJNA1230212
Muestra 16	SRR32606965	SRX27912510	Illumina HiSeq 2000	Mexico – Centro de Investigación Científica de Yucatán	PRJNA1230212
Muestra 17	SRR32606966	SRX27912509	Illumina HiSeq 2000	Mexico – Centro de Investigación Científica de Yucatán	PRJNA1230212
Muestra 18	SRR32606969	SRX27912506	Illumina HiSeq 2000	Mexico – Centro de Investigación Científica de Yucatán	PRJNA1230212
Muestra 19	SRR32606970	SRX27912505	Illumina HiSeq 2000	Mexico – Centro de Investigación Científica de Yucatán	PRJNA1230212
Muestra 20	SRR32606976	SRX27912499	Illumina HiSeq 2000	Mexico – Centro de Investigación Científica de Yucatán	PRJNA1230212

2.2. Obtención de datos

Debido a las limitaciones técnicas y de capacidad de cómputo, se optó por trabajar con 20 muestras, ya que procesar un volumen mayor habría requerido recursos computacionales considerablemente superiores. Este número resultó manejable y permitió realizar los análisis de manera eficiente dentro de las condiciones técnicas disponibles.

Una vez que se identificó de manera aleatoria 20 secuencias pareadas ILLUMINA (Illumina HiSeq) de *C. canephora*, se empleó un código en Google Colab que permitió obtener archivos .fastq del formato original del NBCI (.SRA) con el fin de continuar con los análisis. En la Tabla 2 se puede observar la descripción de las muestras obtenidas.

Tabla 2. Descripción de las muestras de *C. canephora* utilizadas en el análisis

Muestras	Código SRR	Tejido	Grupo	País
M1	SRR5074356	Hoja	Cultivado	Brasil
M2	SRR5074358	Hoja	Cultivado	Brasil
M3	SRR5074359	Hoja	Cultivado	Brasil
M4	SRR5074360	Hoja	Cultivado	Brasil
M5	SRR5074361	Hoja	Cultivado	Brasil
M6	SRR5074362	Hoja	Cultivado	Brasil
M7	SRR5074363	Hoja	Cultivado	Brasil
M8	SRR5074365	Hoja	Cultivado	Brasil
M9	SRR5074366	Hoja	Cultivado	Brasil
M10	SRR5074368	Hoja	Cultivado	Brasil
M11	SRR5074370	Hoja	Cultivado	Brasil
M12	SRR5074371	Hoja	Cultivado	Brasil
M13	SRR5417819	Hoja	Cultivado	Brasil
M14	SRR32606963	Hoja	Cultivado	México
M15	SRR32606964	Hoja	Cultivado	México
M16	SRR32606965	Hoja	Cultivado	México
M17	SRR32606966	Hoja	Cultivado	México
M18	SRR32606969	Hoja	Cultivado	México
M19	SRR32606970	Hoja	Cultivado	México
M20	SRR32606976	Hoja	Cultivado	México

Además, se descargó el genoma de referencia ([JAZHGG000000000.1](https://www.ncbi.nlm.nih.gov/assembly/JAZHGG000000000.1)) de *C. canephora* (*Coffea canephora* cultivar DH200-94, whole genome shotgun sequencing Project) (Salojärvi et al. 2024).

2.3. Flujo de procesamiento de datos: control de calidad, alineamiento y llamado de variantes

El Análisis se realizó con Galaxy Australia (Entorno bioinformático interactivo de código abierto).

Control de calidad:

- Se realizó una prueba de calidad mediante la herramienta FastQC de cada una de las lecturas pareadas, con el fin de asegurar la confianza e integridad de los datos (Andrews, 2010).
- Se limpió y mejoró la calidad empleando la herramienta Trimmomatic (Herramienta de recorte de lectura flexible para datos NGS de Illumina en Galaxy Versión 0.36.6). Se aplicaron los parámetros por defecto del wrapper, equivalentes a: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36.
- Se volvió a realizar la prueba de calidad mediante la herramienta FastQC.

Alineamiento y evaluación:

- Se alinearon las lecturas mediante la herramienta HISAT2 (Un programa de alineación rápido y sensible en Galaxy Versión 2.2.1+galaxy1)
- Se evaluaron los resultados del alineamiento con la herramienta QualiMap BamQC en Galaxy Version 2.3+galaxy0).
- Con el fin de identificar y marcar lecturas duplicadas empleamos la herramienta MarkDuplicates (Examina registros alineados en conjuntos de datos BAM para localizar moléculas duplicadas en Galaxy Versión 3.1.1.0))
- Aplicamos un filtro de calidad de mapeo con la herramienta Samtools view en Galaxy Versión 1.21+galaxy0). Se establecieron los valores por defecto con una calidad mínima de mapeo de $MAPQ \geq 20$ y la exclusión de lecturas no mapeadas (-F 4).
- De nuevo se volvió a usar QualiMap BamQC.

Llamado de variantes:

- Con los archivos .bam, llamamos variantes con la herramienta FreeBayes (Detector de variantes genéticas bayesianas en Galaxy Versión 1.3.10+galaxy0), Con esto se obtuvieron 20 archivos .vcf que han servido para el análisis posterior.

Antes de los análisis posteriores se unió VCF en Google Colab, para lo cual se comprimió e indexó los .vcf con bgzip y tabix. Finalmente, para unir los .vcf en un solo archivo: merged.vcf.gz., se usó bcftools merge.

En la figura 3 se muestra de manera simplificada la obtención y preparación de datos en Galaxy para el análisis genómico de las secuencias.



Figura 3. Pipeline de la obtención y preparación de datos en Galaxy para el análisis genómico de las secuencias.

2.4. Análisis de Diversidad genética y estructura poblacional: heterocigosidad, agrupamiento y diferenciación

A partir del archivo merged.vcf.gz (resultado de la unión de los 20 archivos .vcf), obtenido del procesamiento de datos, se ejecutó en WSL (Windows Subsystem for Linux) un conjunto de comandos utilizando los programas bcftools y PLINK v1.9.

- Con bcftools se filtraron únicamente los SNPs bialélicos, eliminando aquellos con más del 10 % de datos faltantes y conservando solo los que presentaron una

frecuencia alélica mínima ($MAF \geq 0.01$). Este umbral se aplicó para descartar alelos muy raros que podrían deberse a errores de secuenciación, asegurando la calidad del conjunto de datos y la confiabilidad de las estimaciones de diversidad genética.

- El VCF filtrado se transformó a los archivos binarios (.bed/.bim/.fam) que usa PLINK para los análisis poblacionales.
- Con PLINK se calculó, para cada individuo, el número de genotipos homocigotos y heterocigotos, lo que permitió estimar la Heterocigosidad Observada (H_o).
- Se seleccionó un subconjunto de SNPs independientes, eliminando marcadores redundantes, para tener un set más limpio para el análisis de estructura (Pruning por desequilibrio de ligamiento).
- Con el set cortado se realizó un Análisis de Componentes Principales (PCA) en PLINK, obteniendo coordenadas de cada individuo (PC1, PC2, etc.) y la varianza explicada por cada componente.

La visualización de los resultados se llevó a cabo en RStudio, empleando los paquetes ggplot2 (Wickham, 2016) y dplyr (Wickham et al., 2023).

La figura 4 muestra de forma simplificada el pipeline utilizado para calcular la heterocigosidad observada (H_o) y los componentes principales (PCA)



Figura 4. Pipeline de la obtención de la Heterocigosidad Observada (H_o) y Componentes Principales (PCA)

2.5. Dendrograma Neighbor-Joining y matriz a partir de la ASD

- Se preparó los datos genómicos, para esto se utilizó un archivo .raw (genotipos) obtenido anteriormente del análisis con PLINK.
- Se preparó un CSV con IID, País (y, para contexto, tejido = hoja; grupo = cultivado).
- En R se calculó la ASD obteniendo una Matriz de distancia genética y se exportó en TSV.
- Con ape (nj) y phangorn (enraizado por midpoint) se construyó el dendrograma Neighbor-Joining a partir de as.dist(ASD), coloreando las puntas por País.
- Bootstrap: Se estimó el soporte de ramas con boot.phylo (B réplicas) y se rotularon los nodos con sus valores de soporte.
- Con pheatmap se realizó un heatmap de distancias para visualizar la matriz ASD completa, ordenando filas/columnas con UPGMA (hclust “average”) y anotación por País.

Se empleó la métrica ASD (Average Square Distance) para calcular las distancias genéticas, dado que este enfoque resulta más adecuado cuando los datos consisten en marcadores tipo SNP o microsatélites, ya que permite cuantificar de forma continua las diferencias alélicas acumuladas entre individuos frente a métricas basadas únicamente en frecuencias alélicas (Ramakrishnan & Mountain, 2004).

En la figura 5 se muestra de manera simplificada los pasos para la obtención del Dendrograma Neighbor-Joining y matriz de distancia genética.

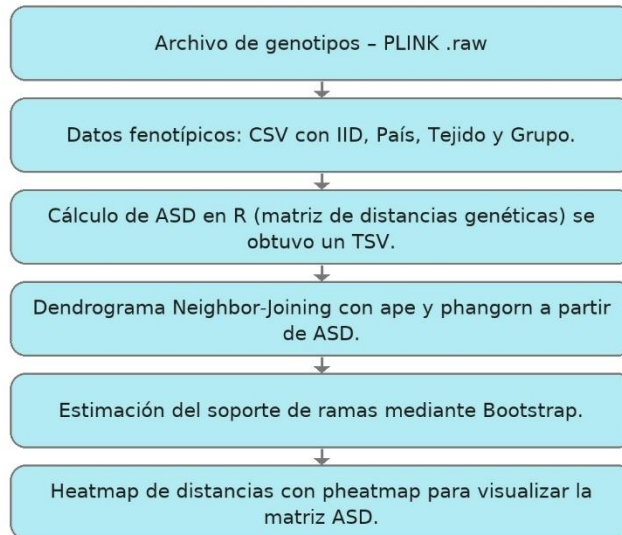


Figura 5. Pipeline para obtener el Dendrograma Neighbor-Joining y matriz de distancia genética (ASD).

3. RESULTADOS

3.1. Diversidad genética y estructura poblacional

3.1.1. Distribución de la heterocigosidad observada (H_o)

En las 20 muestras de *C. canephora* la distribución de la heterocigosidad observada (H_o) mostró un patrón marcadamente bimodal. Un primer grupo de muestras concentró valores bajos de H_o (≈ 0.10 – 0.15 ; $n = 12$), mientras que un segundo grupo presentó valores altos (≈ 0.80 – 0.90 ; $n = 7$). Solo una muestra presentó un valor intermedio de ≈ 0.55 (Figura 6). Este patrón sugiere la coexistencia de genotipos con baja variabilidad (compatible con material clonal, consanguinidad o cuellos de botella locales) y genotipos con alta variabilidad (consistentes con cruzamiento exógamo y/o antecedentes de mezcla/hibridación).

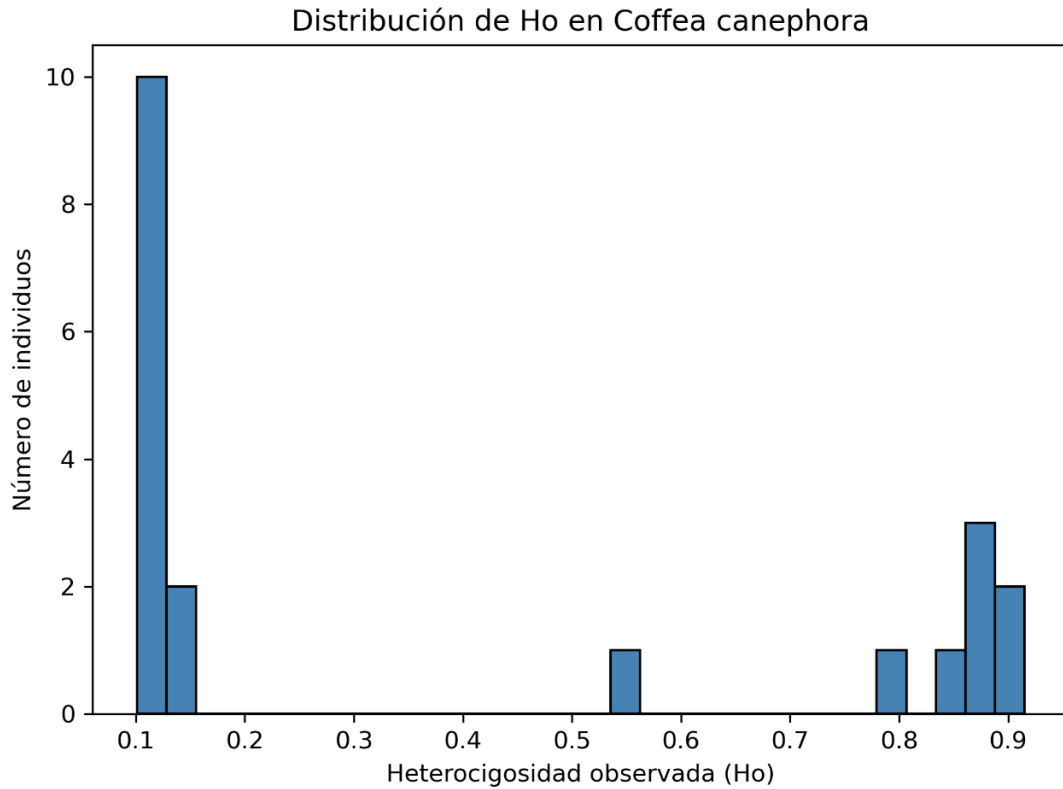


Figura 6. Distribución de la heterocigosidad observada (H_o) en *C. canephora*.

3.1.2. Heterocigosidad observada (H_o) por país

La comparación de la heterocigosidad observada (H_o) entre países mostró diferencias claras entre las muestras de México y Brasil. En México, las muestras presentaron valores consistentemente altos de H_o , con una mediana de 0.85 (rango: 0.77–0.90), mientras que en Brasil las muestras exhibieron valores muy bajos, con una mediana de 0.11 (rango: 0.10–0.14), a excepción de una muestra que presentó un valor de ≈ 0.90 . Solo una muestra de México presentó un valor intermedio de ≈ 0.55 (Figura 7). En conjunto, estos resultados evidencian una marcada diferencia en la variabilidad genética entre ambos países. Las muestras de México presentan una heterocigosidad alta, mientras que las de Brasil muestran valores bajos. Esta diferencia puede deberse al tipo de manejo, reproducción y posible origen de las poblaciones de *C. canephora* en ambos países.

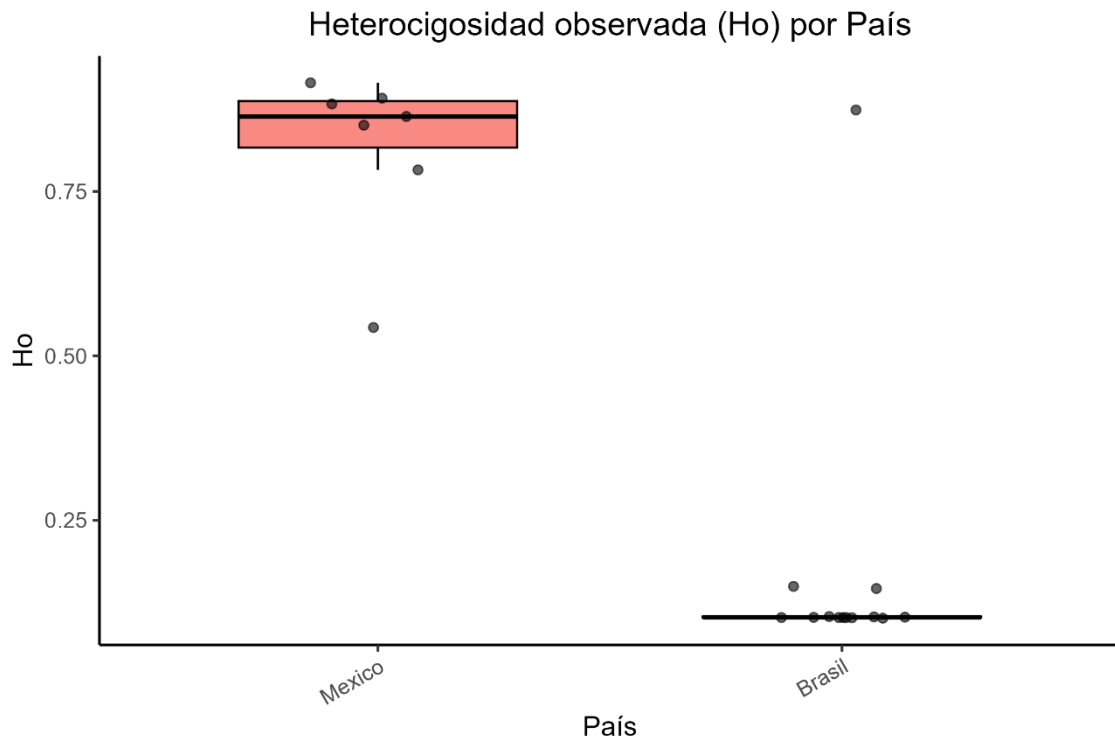


Figura 7. Heterocigosidad observada (Ho) en 20 muestras de *C. canephora* agrupadas por país.

3.1.3. Varianza por cada componente principal

El gráfico de varianza explicada por los componentes principales muestra que el PC1 concentra la mayor parte de la variación genética (75.2%), seguido por el PC2 con 6.8%. A partir del PC3, cada componente explica proporciones muy reducidas de la variación (< 5%), con una contribución acumulada poco significativa en comparación con los dos primeros. Este patrón confirma que la estructura genética de las muestras de *C. canephora* puede describirse de manera robusta con los dos primeros componentes principales, lo que justifica su uso en los análisis de agrupamiento e interpretación de la diferenciación poblacional. (Figura 8).

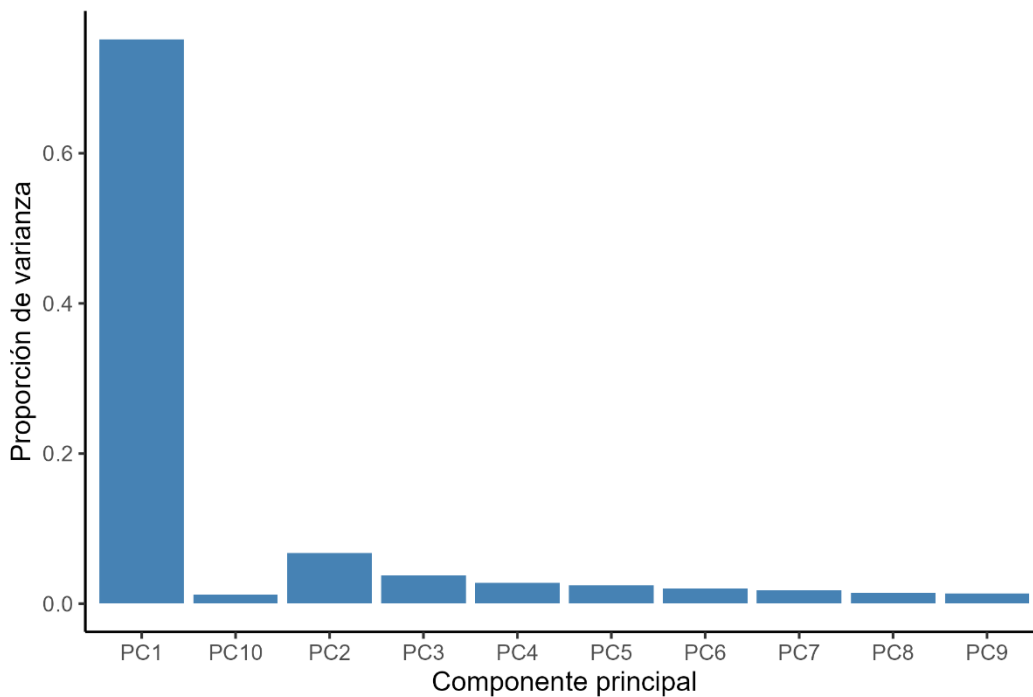


Figura 8. Gráfico de varianza explicada por los diez primeros componentes principales.

3.1.4. Agrupamiento genético y diferenciación poblacional (PCA general)

El análisis de componentes principales (PCA) de las 20 muestras de *C. canephora* reveló una fuerte estructuración genética. Los dos primeros componentes explicaron conjuntamente el 82% de la variación total (PC1 = 75.2%; PC2 = 6.8%). A lo largo del eje PC1 se distinguen tres agrupamientos, un grupo de 12 muestras en el extremo izquierdo (valores de PC1 entre -0.35 y -0.20), un segundo grupo de 7 muestras en el extremo derecho (valores de PC1 entre 0.12 y 0.18) y una muestra aislada que se proyecta hacia valores intermedios (PC1 \approx -0.10; PC2 \approx 0.95). El eje PC2, aunque explica una menor proporción de la varianza, resalta la posición atípica de esta última muestra, que se separa marcadamente del resto. Este patrón confirma la existencia de subestructuras genéticas bien diferenciadas, con un predominio de la variabilidad explicada por PC1, lo cual indica que gran parte de la divergencia genética entre individuos puede resumirse en un solo eje (Figura 9).

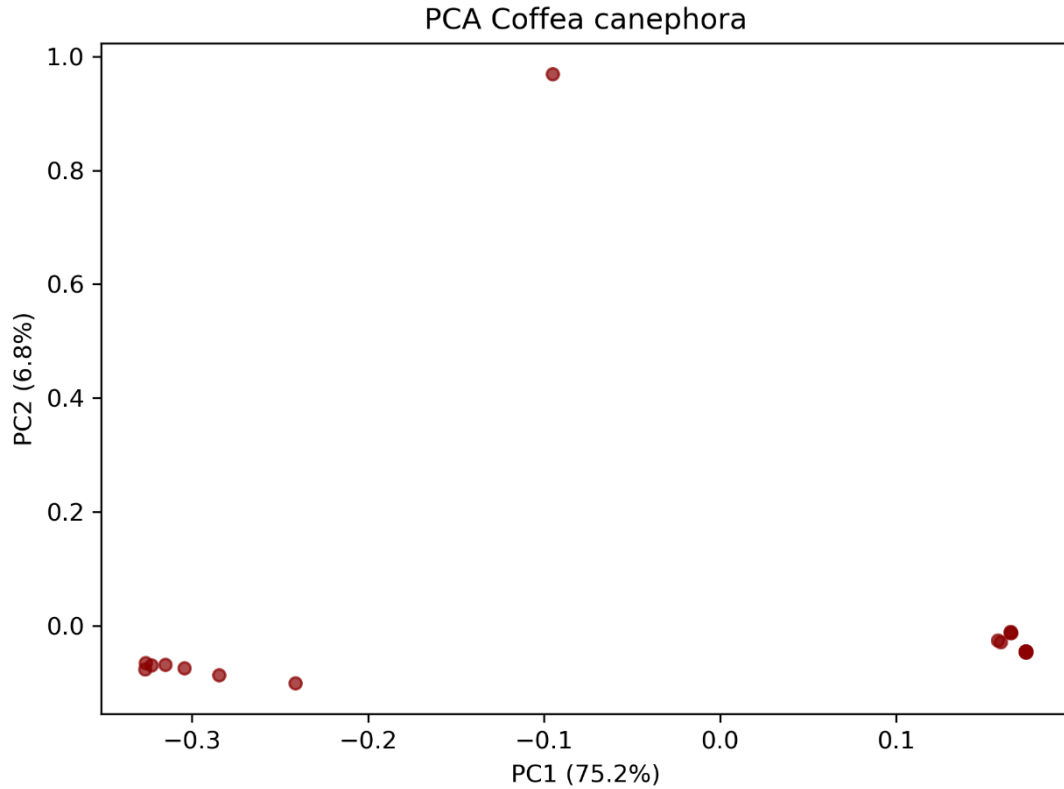


Figura 9. Análisis de componentes principales (PCA) en 20 muestras de *C. canephora*. Distribución de las muestras en función de los dos primeros componentes principales (PC1 y PC2).

3.1.5. PCA de *Coffea canephora*, categorizado por país

El PCA global con categorización por país mostró una separación entre las muestras de México y Brasil. Las muestras de México se agrupan en el extremo izquierdo del eje PC1, con valores entre -0.35 y -0.20, mientras que las de Brasil se concentran en el extremo derecho, con valores entre 0.12 y 0.18. Una muestra de México se proyecta de manera atípica hacia valores intermedios en PC1 y elevados en PC2 (≈ 0.95), lo que lo distingue del resto de las muestras mexicanas. Este patrón evidencia que la mayor parte de la variación genética (75.2%) se explica por diferencias entre países, lo que sugiere una diferenciación poblacional marcada entre las muestras de *C. canephora* procedentes de México y Brasil (Figura 10).

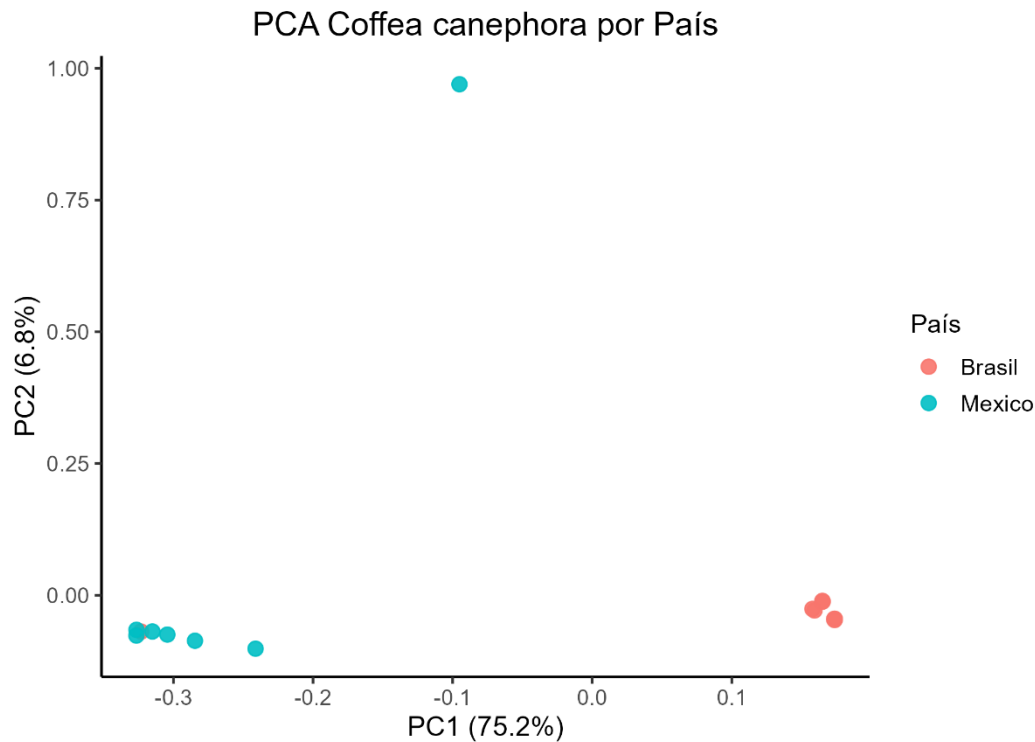


Figura 10. Análisis de componentes principales (PCA) en 20 muestras de *C. canephora* categorizado por país.

3.2. Estructura poblacional inferida a partir de distancias genéticas

3.2.1. Agrupamiento mediante Neighbor-Joining (ASD)

El análisis de agrupamiento mediante Neighbor-Joining (ASD), con soporte de ramas estimado por Bootstrap, mostró una clara separación entre las muestras de Brasil y México. Las muestras mexicanas se agruparon en un clado bien definido, con valores relativamente altos en varios nodos (M14 y M15 con soporte 200; M17–M20 con soportes entre 102 y 195), lo que indica consistencia en su diferenciación genética. En contraste, las muestras brasileñas formaron un segundo clado, subdividido en tres grupos, el primero compuesto por M7 y M10; el segundo con las muestras M9, M4 y M2, con soportes de 151 y 199, y el tercero formado por M8, M3, M1, M5, M13, M12 y M6.

La muestra brasileña M11, es un caso particular ya que se agrupa con la muestra mexicana M16, conformando un subgrupo con soporte de 175, esto contrasta con el patrón de agrupamiento por origen geográfico (Figura 11).

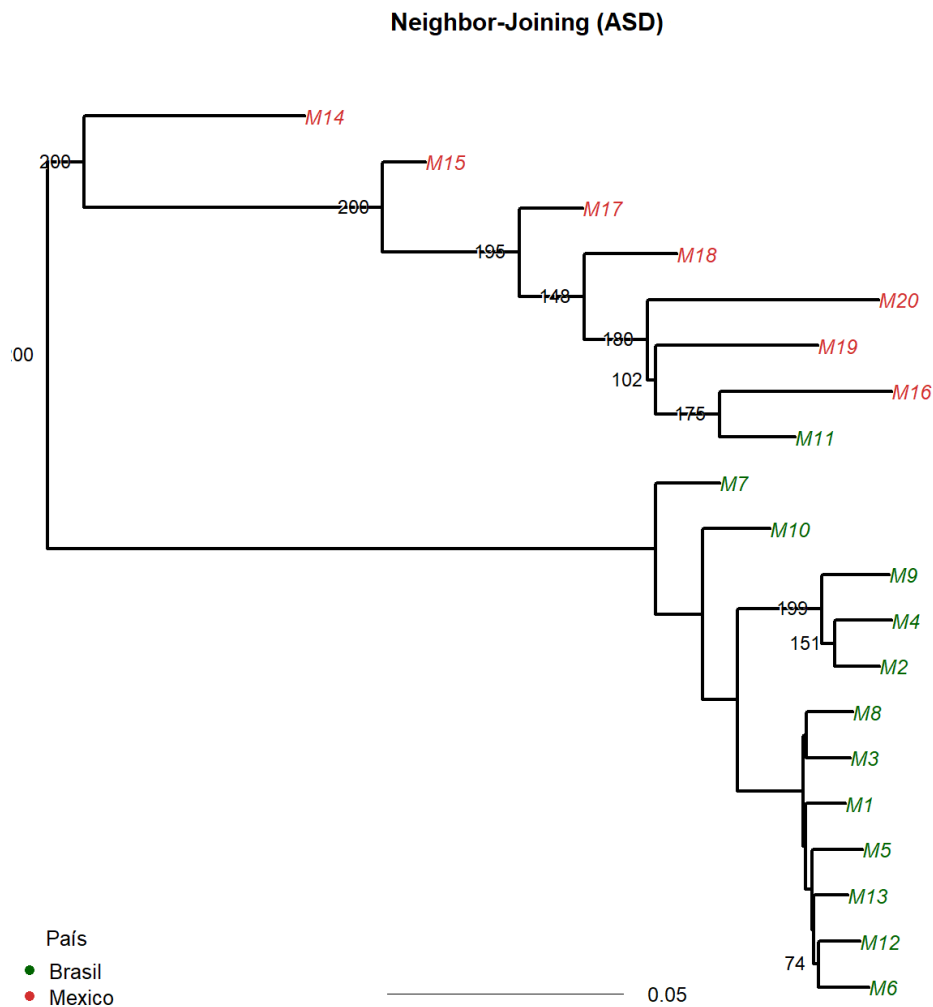


Figura 11. Dendrograma de agrupamiento obtenido con el método Neighbor-Joining (NJ) a partir de la distancia genética ASD en 20 muestras de *C. canephora*. Se observa una separación marcada entre las muestras de México (rojo) y Brasil (verde). Soporte de ramas por bootstrap con $b=200$ réplicas.

3.2.2. Análisis de estructura/afinidad genética

La matriz de distancias genéticas calculada a partir de la métrica ASD mostró un patrón de agrupamiento consistente con los análisis previos de PCA y Neighbor-Joining. En el heatmap se observa una separación clara entre las muestras de México y Brasil, con bloques bien diferenciados de similitud genética. Las muestras mexicanas presentan valores de distancia relativamente bajos entre sí (ASD promedio ≈ 0.12 ; rango 0.10–0.15) (colores más oscuros, indicando mayor cercanía genética), mientras que las brasileñas

también forman un clúster compacto, aunque con menor homogeneidad interna en comparación con las mexicanas (ASD promedio ≈ 0.18 ; rango 0.14–0.22).

Se observa que la muestra M16 se relaciona más con las brasileñas, en particular con M11, con un valor de distancia de ≈ 0.14 , menor al promedio de M16 con el resto de muestras mexicanas (≈ 0.21), lo que coincide con lo observado en el dendrograma NJ y sugiere una relación genética atípica dentro del conjunto mexicano. En general, los patrones de color reflejan la marcada diferenciación poblacional entre países. (Figura 12).

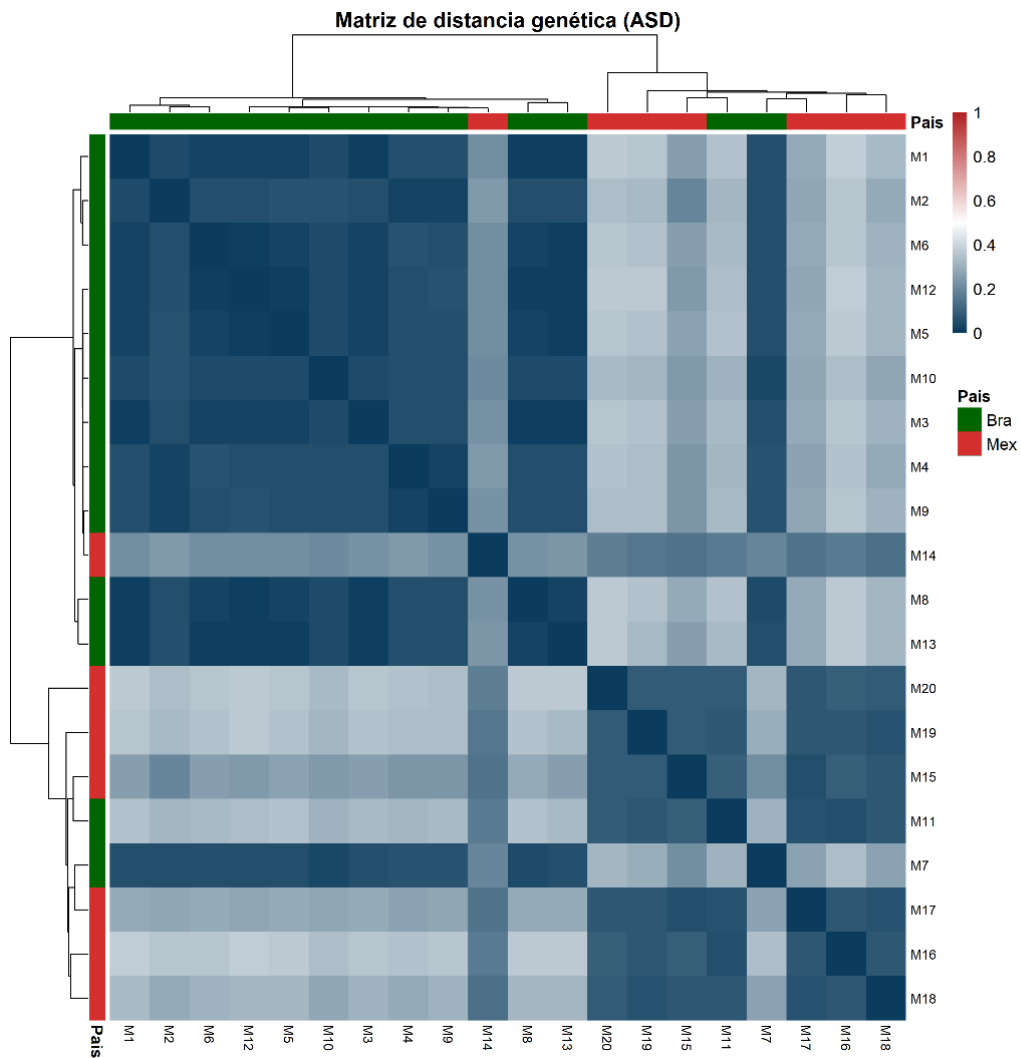


Figura 12. Heatmap de la matriz de distancias genéticas calculada con la métrica ASD en 20 muestras de *C. canephora*. La escala de color indica la distancia genética (0 = mayor similitud, 1 = mayor distancia). Los colores de la barra superior y lateral representan el país de origen (verde = Brasil, rojo = México).

4. DISCUSIÓN

Si bien los análisis realizados (heterocigosidad, PCA, distancias genéticas y agrupamiento mediante Neighbor-Joining) se aplicaron correctamente, es importante reconocer que el número reducido de muestras analizadas ($n = 20$) puede limitar la representatividad de la variabilidad genética real. Un tamaño muestral pequeño tiende a incrementar la varianza de los estimadores, como la heterocigosidad observada (H_o) o las distancias genéticas (ASD), por lo que puede restringir la detección de patrones sutiles de estructura poblacional o de loci con efectos pequeños (Hale et al., 2012; Nazareno et al., 2017). Por tanto, los resultados obtenidos en este estudio deben tomarse con cautela y considerarse de carácter exploratorio, orientados principalmente a validar la metodología bioinformática y estadística empleada, más que a establecer conclusiones definitivas sobre la diversidad genética entre poblaciones de *C. canephora*.

4.1. Interpretación de los resultados en el contexto de la genética poblacional

Los resultados de este estudio indican una distribución bimodal de la heterocigosidad observada (H_o) en las 20 muestras, con un grupo de baja $H_o \approx 0.10$ – 0.15 ($n=12$), un grupo de alta $H_o \approx 0.80$ – 0.90 ($n=7$) y una muestra intermedia ≈ 0.55 , lo cual revela la coexistencia de genotipos con historias de reproducción contrastantes. Los valores altos de H_o son coherentes con el sistema de autoincompatibilidad gametofítica y la alogamia de *C. canephora*, que favorecen el cruzamiento y mantienen diversidad, mientras que los valores muy bajos sugieren clonación prolongada, cuellos de botella o fundación reciente de linajes con base genética estrecha. Este patrón concuerda con la biología reproductiva descrita para la especie, caracterizada por la autoincompatibilidad que obliga al cruzamiento (Lashermes et al., 1996, 2018; Silva et al., 2024).

En la comparación por país, México mostró H_o alta y consistente (mediana ≈ 0.85 , rango 0.77 – 0.90), mientras que Brasil presentó H_o baja (mediana ≈ 0.11 , rango 0.10 – 0.14), con dos excepciones destacables: una muestra brasileña con $H_o \approx 0.90$ y una mexicana con $H_o \approx 0.55$. Este contraste sugiere modos de manejo divergentes y posibles eventos de introducción distintos entre países. En términos poblacionales, según Looor-Solórzano et al. (2017) la persistencia de H_o alta es esperable cuando hay flujo génico por polinización cruzada; en cambio, H_o baja reiterada suele emerger de clonación o deriva intensa en poblaciones pequeñas, ambos escenarios aceptables en materiales introducidos.

La estructura por PCA es marcada: PC1 = 75.2 % y PC2 = 6.8 % (acumulado 82 %). A lo largo de PC1 se separan tres grupos, uno de 12 muestras con PC1 entre -0.35 y -0.20 , el segundo de 7 con PC1 entre 0.12 y 0.18 , y el tercero fue una muestra atípica proyectada en PC2 ≈ 0.95 . Al colorear por país, México se ubica en el extremo izquierdo de PC1 y Brasil en el derecho (ver figura 10), con una muestra mexicana desplazada y una brasileña que no sigue el patrón. Que PC1 explique $\sim 75\%$ de la variación significa que casi toda la diferencia genética entre las muestras se organiza a lo largo de un eje principal, separando grupos con historias de propagación distintas. El PC2 capta señales más finas que sugieren mezcla genética entre grupos (admixture), es decir, cruces o intercambio de semillas o plantas entre orígenes, algo ya reportado en otras colecciones de Robusta (Cubry et al., 2013; Hamon et al., 2015, 2021).

El análisis Neighbor-Joining con ASD, confirmó la separación por país, con soportes bootstrap altos en nodos mexicanos (M14–M15 con 200; M17–M20 con 102–195) y una subdivisión del clado brasileño en tres subgrupos (M7–M10; M9–M4–M2 con 151 y 199; y M8–M3–M1–M5–M13–M12–M6). Destaca el pareamiento de M11 de Brasil con la M16 de México con un soporte de 175, que rompe el patrón geográfico y sugiere afinidad genética no prevista por el origen declarado.

La matriz de distancia genética (ASD) respalda lo anterior y permite cuantificar la cohesión interna. México mostró ASD promedio ≈ 0.12 (rango 0.10 – 0.15), indicando cercanía genética intrapaís, por su parte Brasil presentó ASD promedio ≈ 0.18 (rango 0.14 – 0.22), es decir, mayor heterogeneidad interna. El par de muestras M16–M11 tuvo ASD ≈ 0.14 , menor que el promedio de M16 con el resto de México (≈ 0.21), consistente con el bootstrap del dendrograma (NJ) y con una historia compartida reciente (flujo génico, intercambio de material o introducción común). Metodológicamente, el uso de ASD es apropiado para sintetizar distancias a partir de marcadores polimórficos, ampliamente empleado en genética de poblaciones para captar grupos de similitud y saltos de diferenciación (Goldstein et al., 1995).

En general, la H_o es bimodal (0.10 – 0.15 vs. 0.80 – 0.90) y, por país, México tiene valores altos (mediana ≈ 0.85) y Brasil bajos (mediana ≈ 0.11). El PCA separa a México y Brasil en PC1 (75.2% de la variación) y PC2 solo marca unas pocas muestras atípicas. El dendrograma NJ muestra nodos con alto soporte y la matriz de distancias (ASD) indica que las muestras son más parecidas dentro de cada país que entre países. Todo apunta a una estructura poblacional moldeada por modos de propagación e historias de

introducción distintas, con un caso puntual de mezcla entre países (M11–M16). Este patrón coincide con lo descrito para el germoplasma introducido de *C. canephora* en Latinoamérica (Loor-Solórzano et al., 2017; Dereeper et al., 2015).

Por implicaciones aplicadas, los genotipos de H_o alta y baja ASD intragrupo representan reservorios valiosos para cruzamientos, en tanto que los outliers ameritan verificación de trazabilidad y caracterización adicional antes de decisiones de conservación o mejoramiento. Dado el sistema de autoincompatibilidad y la naturaleza alógama de la especie, estrategias de manejo que maximicen compatibilidades y eviten cuellos de botella ayudarán a sostener la diversidad observada (Denoeud et al., 2014; Silva et al., 2024).

4.2. Implicaciones para la conservación de *Coffea canephora*

Los resultados obtenidos muestran que las poblaciones analizadas de *C. canephora* presentan un patrón heterogéneo de diversidad genética, con grupos de alta heterocigosidad ($H_o \approx 0.80–0.90$) y otros de heterocigosidad reducida ($H_o \approx 0.10–0.15$). Desde la perspectiva de la conservación, este hallazgo implica que el germoplasma disponible no es homogéneo y que se deben aplicar estrategias diferenciadas para cada grupo. Los individuos con H_o elevada constituyen una fuente genética valiosa que puede contribuir a sostener la variabilidad a largo plazo, por otro lado, aquellos con H_o baja podrían estar en riesgo de erosión genética si no se promueve su renovación mediante cruzamientos controlados o la introducción de nuevos materiales compatibles (Silva et al., 2024; Lashermes et al., 1996).

La estructura poblacional observada en el PCA y el NJ, con separación clara entre las muestras de México (más homogéneas, ASD promedio ≈ 0.12) y Brasil (más heterogéneas, ASD promedio ≈ 0.18), refuerza la necesidad de conservar materiales de ambos orígenes. Los genotipos mexicanos, al mostrar alta cohesión genética, podrían representar una población compacta pero vulnerable a eventos de pérdida de diversidad; mientras que los genotipos brasileños, más variables, constituyen una reserva importante de diversidad intraespecífica. Esta diferenciación coincide con los reportes que identifican a Brasil como un centro secundario de diversificación del café Robusta y a México como receptor de introducciones más limitadas (Loor-Solórzano et al., 2017).

La detección de outliers genéticos, como el par M11 de Brasil–M16 de México con ASD bajo (≈ 0.14) y soporte alto en el dendrograma, evidencia posibles flujos génicos recientes

o intercambios de germoplasma entre países. Este tipo de casos resalta la importancia de verificar la trazabilidad de accesiones en bancos de germoplasma y de fomentar programas de caracterización molecular sistemática. La integridad de la información de origen es crucial, ya que errores de clasificación pueden comprometer decisiones de conservación y mejoramiento (Dereeper et al., 2015).

La conservación de *C. canephora* debe adoptar un enfoque complementario entre in situ y ex situ. Mientras los bancos de germoplasma resguardan la diversidad conocida, las poblaciones cultivadas y silvestres en los países productores mantienen procesos dinámicos de flujo génico y adaptación local, indispensables frente al cambio climático y la presión de plagas emergentes (Depecker et al., 2023). Así, la gestión de la variabilidad detectada en este estudio no solo es relevante para programas de mejoramiento, sino también para la resiliencia de la caficultura global.

4.3. Posibles aplicaciones en programas de mejoramiento genético

Las muestras de *C. canephora* presentan niveles contrastantes de heterocigosidad, con grupos de alta y baja heterocigosidad. Esta variabilidad puede aprovecharse en programas de cruzamiento, utilizando genotipos altamente heterocigotos como donadores de diversidad genética y los de baja heterocigosidad como receptores para ampliar bases genéticas (Silva et al., 2024).

La clara estructuración entre muestras mexicanas y brasileñas observada en PCA y el dendrograma NJ, junto con las diferencias en distancias ASD (México: promedio 0.12; Brasil: 0.18), ofrece la oportunidad de diseñar cruzamientos entre orígenes divergentes, generando combinaciones híbridas con mayor variabilidad y posible heterosis (Cubry et al., 2013).

El hallazgo de outliers genéticos, como la relación entre las muestras M11 y M16 de Brasil y México respectivamente, sugiere la presencia de linajes con combinaciones únicas, útiles como puentes genéticos para integrar diversidad de distintos grupos. Estos individuos deben priorizarse en selección asistida y programas de hibridación controlada, pues los outliers con rendimiento o características extremas pueden aportar alelos raros favorables que no están presentes en la mayoría del germoplasma (Bernardo, 2022).

4.4. Limitaciones del estudio y recomendaciones para investigaciones futuras

Aunque los resultados aportan información valiosa sobre la diversidad genética de *C. canephora*, se deben considerar ciertas limitaciones. Una primera limitación del presente estudio radica en el tamaño de muestra analizado ($n=20$), el cual, si bien permitió identificar patrones claros de heterocigosidad, diferenciación genética y agrupamiento, restringe la posibilidad de capturar de manera exhaustiva la variabilidad total de la especie estudiada en la región (Depecker et al. 2023)). La presencia de outliers genéticos como el par M11–M16 refuerza que existen linajes no representados en la muestra y que podrían modificar las interpretaciones si se incluyeran conjuntos de muestras genómicas más grandes y diversas.

En segundo lugar, el análisis se basó en métricas globales de diversidad (H_o , ASD, PCA, NJ) que permiten una caracterización general de la estructura poblacional, pero no en herramientas genómicas de mayor resolución, como GWAS, estimaciones de desequilibrio de ligamiento o haplotipos asociados a caracteres de interés agronómico. Ello limita la capacidad de establecer asociaciones directas entre marcadores y rasgos fenotípicos, como resistencia a enfermedades o contenido de cafeína, que en este estudio solo pudieron discutirse de manera general.

También ha sido limitante que los análisis se centraron en poblaciones introducidas en México y Brasil, sin incluir muestras africanas de origen silvestre o de bancos de germoplasma internacionales. La literatura muestra que estas poblaciones mantienen niveles elevados de variabilidad genética y representan el reservorio primario de diversidad para la especie (Depecker et al., 2023; Vanden Abeele et al., 2021). Por tanto, la ausencia de estas comparaciones restringe la posibilidad de ubicar con precisión los materiales analizados dentro del espectro global de diversidad de *C. canephora*.

4.5. Avances genómicos y su aplicación en estudios poblacionales de *Coffea*

El presente estudio se apoyó inicialmente en el primer genoma de *C. canephora* publicado por Deneud et al. (2014), que constituyó un hito al aportar la primera visión integral de la evolución convergente de la cafeína y de la organización cromosómica de la especie a partir de un ensamblaje de cobertura media ($\sim 30\times$) con más de 25.000 contigs. Posteriormente, el desarrollo del Coffee Genome Hub (Dereeper et al., 2015) permitió consolidar y explotar estos datos mediante la integración de mapas genéticos, anotaciones funcionales y recursos transcriptómicos, lo que facilitó la exploración de la diversidad

genética y la identificación de polimorfismos a escala genómica. Más recientemente, se han generado ensamblajes de referencia de alta calidad, como el de *C. arabica* descrito por Salojärvi et al. (2024), basado en tecnologías de secuenciación de última generación y estrategias híbridas que mejoran la continuidad de los pseudocromosomas, resuelven regiones repetitivas y optimizan la anotación de genes duplicados. En este estudio se empleó dicho genoma de referencia, el cual ofrece un marco más robusto y preciso para realizar análisis de diversidad genética, inferencia poblacional y estudios comparativos de evolución a escala genómica.

Este avance representa un cambio significativo para los estudios de diversidad genética y mejoramiento, ya que ofrece un marco más robusto para el análisis de variantes y la identificación de loci adaptativos. En ese sentido, el presente trabajo puede considerarse un piloto, al evaluar únicamente 20 muestras bajo el contexto del genoma previo, pero al mismo tiempo abre la puerta a que análisis futuros, apoyados en el nuevo ensamblaje, permitan profundizar en la caracterización genética y fenotípica de *C. canephora*. De este modo, los resultados aquí obtenidos se constituyen en un primer paso hacia evaluaciones más integrales, que integren la diversidad regional dentro del marco global de la especie.

5. CONCLUSIONES Y RECOMENDACIONES

Los hallazgos de este estudio constituyen una aproximación inicial a la variabilidad genética de *C. canephora* en las poblaciones analizadas. No obstante, el tamaño reducido de la muestra ($n = 20$) representa una limitación que podría afectar la solidez y generalización de los resultados. En este sentido, las presentes conclusiones deben considerarse preliminares y orientadas a demostrar la aplicabilidad del flujo de análisis empleado. Futuros estudios con un número mayor de individuos y cobertura genómica más amplia permitirán validar y ampliar las tendencias observadas, fortaleciendo la comprensión de la estructura genética y la diversidad poblacional de la especie.

Los resultados de este estudio indican que las muestras analizadas de *C. canephora* presentan una marcada heterogeneidad genética, reflejada en una distribución bimodal de la heterocigosidad, con genotipos altamente heterocigotos ($H_o \approx 0.80\text{--}0.90$) y otros con valores muy reducidos ($H_o \approx 0.10\text{--}0.15$). Esta diferencia posiblemente refleja la influencia del sistema reproductivo de la especie, basado en la alogamia y la autoincompatibilidad, como el impacto de los distintos modos de propagación y manejo en las poblaciones introducidas en México y Brasil.

La clara separación observada en los análisis de PCA y dendrogramas, junto con las diferencias en la cohesión genética intrapaís, pone de manifiesto que México conserva grupos más homogéneos y compactos (ASD promedio ≈ 0.12), mientras que Brasil mantiene una diversidad interna más amplia (ASD promedio ≈ 0.18). Este contraste sugiere trayectorias históricas distintas en el establecimiento y uso del germoplasma. Además, la identificación de outliers genéticos, como el par M11–M16, refuerza la existencia de flujos de germoplasma o procesos de intercambio que han contribuido a la mezcla de linajes.

Estos hallazgos resaltan el valor de los genotipos altamente heterocigotos como reservorios de variabilidad y de los individuos atípicos como posibles puentes para integrar diversidad entre poblaciones. A su vez, confirman la importancia de preservar tanto los linajes homogéneos, que pueden ser más vulnerables a la pérdida de variación, como los heterogéneos, que aportan resiliencia y plasticidad. En conjunto, este trabajo constituye un primer paso en la caracterización genética de *C. canephora*, y aunque se desarrolló con un número limitado de muestras, sienta bases sólidas para estudios posteriores que busquen integrar la diversidad local en el marco global de la especie.

A partir de los resultados obtenidos, se recomienda ampliar el alcance de futuras investigaciones hacia un mayor número de muestras, incluyendo no solo materiales cultivados de diferentes países latinoamericanos, sino también muestras silvestres africanas que representan el reservorio primario de diversidad de la especie. Esto permitirá ubicar con mayor precisión a las muestras analizadas en el contexto global y garantizar estrategias de conservación más completas.

También, resulta necesario incorporar tecnologías de genotipado de alta resolución que permitan analizar la variación genética con mayor detalle, identificar haplotipos asociados a caracteres agronómicos y descubrir la arquitectura de rasgos complejos como resistencia a enfermedades o calidad de taza. La integración de análisis poblacionales más finos, complementarios al PCA, facilitará detectar dinámicas de flujo génico, eventos de hibridación y procesos de mezcla entre linajes.

De igual forma, es fundamental vincular la información genética con datos fenotípicos estandarizados, de manera que la diversidad identificada pueda traducirse en aplicaciones prácticas en programas de mejoramiento. El trabajo coordinado entre bancos de germoplasma, poblaciones cultivadas y colecciones silvestres permitirá no solo conservar la variabilidad existente, sino también aprovecharla activamente para el desarrollo de variedades más resilientes y adaptadas a los retos que impone el cambio climático.

6. BIBLIOGRAFÍA

- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., ... Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, *46*(W1), W537–W544. <https://doi.org/10.1093/nar/gky379>
- Bernardo, R. (2022). Outliers and their distribution in breeding populations. *Crop Science*, *62*(1), 38–48. <https://doi.org/10.1002/csc2.20742>
- Cubry, P., Musoli, P., Legnaté, H., Pot, D., de Bellis, F., Poncet, V., & Leroy, T. (2013). An initial assessment of linkage disequilibrium in coffee trees: LD patterns in groups of *Coffea canephora* using microsatellite analysis. *BMC Genomics*, *14*, 10. <https://doi.org/10.1186/1471-2164-14-10>
- Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M., Zheng, C., Alberti, A., Anthony, F., Aprea, G., Aury, J. M., Bento, P., Bernard, M., Bocs, S., Campa, C., Cenci, A., Combes, M. C., Crouzillat, D., Da Silva, C., ... Wincker, P. (2014). The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science*, *345*(6201), 1181–1184. <https://doi.org/10.1126/science.1255274>
- DePaula, J., Cunha, S. C., Partelli, F. L., Fernandes, J. O., & Farah, A. (2025). Major bioactive compounds, volatile and sensory profiles of *Coffea canephora* flowers and infusions for waste management in coffee production. *Foods*, *14*(6), 911. <https://doi.org/10.3390/foods14060911>
- Depecker, J., Verleysen, L., Asimonyio, J. A., Hatangi, Y., Kambale, J.-L., Mwangi, I. M., ... Honnay, O. (2023). Genetic diversity and structure in wild Robusta coffee (*Coffea canephora*) populations in Yangambi (DR Congo) and their relation to forest disturbance. *Heredity*, *131*, 145–157. <https://doi.org/10.1038/s41437-022-00588-0>
- Dereeper, A., Bocs, S., Rouard, M., Guignon, V., Ravel, S., Tranchant-Dubreuil, C., Poncet, V., Garsmeur, O., Lashermes, P., & Droc, G. (2015). The coffee genome hub: A resource for coffee genomes. *Nucleic Acids Research*, *43*(D1), D1028–D1035. <https://doi.org/10.1093/nar/gku1108>
- Ferrão, M. A. G., da Fonseca, A. F. A., Volpi, P. S., Souza, L. C., Comério, M., Verdin Filho, A. C., ... Ferrão, L. F. V. (2023). Genomic-assisted breeding for climate-smart coffee. *The Plant Genome*, *16*(1), e20321. <https://doi.org/10.1002/tpg2.20321>
- Goldstein, D. B., Ruiz Linares, A., Cavalli-Sforza, L. L., & Feldman, M. W. (1995). Genetic absolute dating based on microsatellites and the origin of modern humans. *Proceedings of the National Academy of Sciences*, *92*(15), 6723–6727. <https://doi.org/10.1073/pnas.92.15.6723>
- Hale, M. L., Burg, T. M., & Steeves, T. E. (2012). Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to

accurately estimate allele frequencies. *PLoS ONE*, 7(9), e45170. <https://doi.org/10.1371/journal.pone.0045170>

Hamon, P., Hamon, S., Razafinarivo, N. J., Guyot, R., Siljak-Yakovlev, S., Couturon, E., ... de Kochko, A. (2015). *Coffea* genome organization and evolution. In V. Preedy (Ed.), *Coffee in health and disease prevention* (pp. 29–39). Academic Press. <https://doi.org/10.1016/B978-0-12-409517-5.00004-8>

Hamon, P., Ducrocq, H., & Schluter, P. (2021). Germplasm banks and genetic resources conservation of *Coffea canephora*. *Journal of Agricultural Science*, 13(2), 115–126.

Holderegger, R., & Wagner, H. H. (2008). Landscape genetics. *BioScience*, 58(3), 199–207. <https://doi.org/10.1641/B580306>

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8), 907–915. <https://doi.org/10.1038/s41587-019-0201-4>

Lashermes, P., Andrzejewski, S., Bertrand, B., Combes, M. C., Dussert, S., & Anthony, F. (2018). Molecular breeding of coffee. *Plant Breeding Reviews*, 24, 421–447.

Lashermes, P., Couturon, E., Moreau, N., Paillard, M., & Louarn, J. (1996). Inheritance and genetic mapping of self-incompatibility in *Coffea canephora* Pierre. *Theoretical and Applied Genetics*, 93(3), 458–462. <https://doi.org/10.1007/BF00223190>

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>

Loor-Solórzano, R. G., De Bellis, F., & Leroy, T. (2017). Revealing the diversity of introduced *Coffea canephora* germplasm in Ecuador: Towards a national strategy to improve Robusta. *The Scientific World Journal*, 2017, 1–12. <https://doi.org/10.1155/2017/1248954>

Mardis, E. R. (2017). Next-generation sequencing platforms. *Annual Review of Analytical Chemistry*, 10, 387–404. <https://doi.org/10.1146/annurev-anchem-061516-045720>

Nazareno, A. G., Bemmels, J. B., Dick, C. W., & Lohmann, L. G. (2017). Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. *Molecular Ecology Resources*, 17(6), 1136–1147. <https://doi.org/10.1111/1755-0998.12654>

Quintana-Escobar, A. O., Iracheta-Donjuan, L., & Jiménez-Hernández, J. (2017). Caracterización de genotipos élite de *Coffea canephora* por su tolerancia a sequía. *Agronomía Mesoamericana*, 28(1), 183–195.

- Ramakrishnan, U., & Mountain, J. L. (2004). Precision and accuracy of divergence time estimates from STR and SNPSTR variation. *Molecular Biology and Evolution*, *21*(10), 1960–1971. <https://doi.org/10.1093/molbev/msh212>
- Sahachairungrueng, W., Meechan, C., Veerachat, N., Thompson, A. K., & Teerachaichayut, S. (2022). Assessing the levels of Robusta and Arabica in roasted ground coffee using NIR hyperspectral imaging and FTIR spectroscopy. *Foods*, *11*(19), 3122. <https://doi.org/10.3390/foods11193122>
- Salojärvi, J., Rambani, A., Yu, Z., Guyot, R., Strickler, S., Lepelley, M., Wang, C., Rajaraman, S., Rastas, P., Zheng, C., Santos Muñoz, D., Meidanis, J., Paschoal, A. R., ... Descombes, P. (2024). The genome and population genomics of allopolyploid *Coffea arabica* reveal the diversification history of modern coffee cultivars. *Nature Genetics*, *56*(4), 721–731. <https://doi.org/10.1038/s41588-024-01695-w>
- Sant’Ana, G. C., Pereira, L. F. P., Pot, D., Ivamoto, S. T., Domingues, D. S., Ferreira, R. V., ... Leroy, T. (2018). Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in *Coffea arabica*. *Scientific Reports*, *8*, 465. <https://doi.org/10.1038/s41598-017-18800-1>
- Silva, A. N. R., Rocha, R. B., Teixeira, A. L., Espindula, M. C., Partelli, F. L., & Caixeta, E. T. (2024). Self-incompatibility and pollination efficiency in *Coffea canephora* using fluorescence microscopy. *Agronomy*, *14*(7), 1564. <https://doi.org/10.3390/agronomy14071564>
- Schatz, M. C., Langmead, B., & Salzberg, S. L. (2010). Cloud computing and the DNA data race. *Nature Biotechnology*, *28*(7), 691–693. <https://doi.org/10.1038/nbt0710-691>
- Spinoso-Castillo, J. L., Pérez-Rodríguez, P., Bello-Bello, J. J., Escamilla-Prado, E., Aguilar-Rincón, V. H., Corona-Torres, T., García-de los Santos, G., & Morales-Ramos, V. (2022). SNP markers identification by genome-wide association study for chemical quality traits of coffee (*Coffea spp.*) germplasm. *Molecular Biology Reports*, *49*(7), 4849–4859. <https://doi.org/10.1007/s11033-022-07339-8>
- Tinoco-Jaramillo, J., Garcia, M., Gunter, S., Maza, D., & Günter, S. (2024). Productivity and nutrient cycling in agroforestry and monoculture systems of *Coffea canephora* under conventional and organic management in the Ecuadorian Amazon. *Forests*, *15*(5), 807. <https://doi.org/10.3390/f15050807>
- Toapanta A., J. R., Quinatoa Lozada, E. F., Luna Murillo, R. A., & Macias Pettao, R. K. (2023). Producción de café (*Coffea canephora* P.) en el subtrópico ecuatoriano en respuesta a diferentes niveles de fertilización inorgánica-orgánica. *Ciencia Latina Revista Científica Multidisciplinar*, *7*(1), 2750-2761.
- Vanden-Abeelee, S., Janssens, S. B., Bawin, Y., Mwanga Mwanga, I., Kambale, J.-L., Stoffelen, P., ... Vandeloos, F. (2021). Genetic diversity of wild and cultivated *Coffea canephora* in northeastern DR Congo and the implications for

conservation. *American Journal of Botany*, 108(12), 2425–2434.
<https://doi.org/10.1002/ajb2.1769>

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2nd ed.). Springer.
<https://doi.org/10.1007/978-3-319-24277-4>

Wickham, H., François, R., Henry, L., & Müller, K. (2023). *dplyr: A grammar of data manipulation* (Version 1.1.3) [R package]. Comprehensive R Archive Network (CRAN). <https://CRAN.R-project.org/package=dplyr>

Zaidan, I. R., Ferreira, M. F. da S., do Couto, D. P., Santos, J. G., Silva, M. A., Canal, G. B., Bernardes, C. O., Azevedo, C. F., & Ferreira, A. (2025). Genome-wide association analysis of traits related to development, abiotic and biotic stress resistance in *Coffea canephora*. *Scientia Horticulturae*, 341, 114004.
<https://doi.org/10.1016/j.scienta.2025.114004>

Zainuri, D. N. A. P., Fatimah, A., Nofrida, R., Rahayu, N., Anggraini, I. M. D., & Utama, Q. D. (2023). The chemical characteristics of Arabica and Robusta green coffee beans from Geopark Rinjani, Indonesia. *BIOTROPIA*, 30(3), 318–328.
<https://doi.org/10.11598/btb.2023.30.3.1940>

Zambrano-Flores, F. G., Loor-Solorzano, R. G., Plaza-Avellán, L. F., Jaimez-Arellano, R. E., Guerrero-Castillo, H. E., Casanova-Mendoza, T. D. J., López, D. A., & Rodríguez-Zamora, G. A. (2018). Relación entre productividad y calidad integral del grano en selecciones avanzadas de café robusta (*Coffea canephora*) en Ecuador. *Agrociencia*, 52(4), 593-607.

7. APÉNDICES

- Heterocigosidad (Ho) por cada muestra.

FID	IID	O_HOM	E_HOM	N_non_missing	F_inbreeding	Ho
M1	M1	25244	19790	28146	0.6527	0.103105
M2	M2	25281	19780	28127	0.6592	0.101184
M3	M3	25303	19820	28191	0.6549	0.102444
M4	M4	25281	19800	28155	0.6561	0.102078
M5	M5	25310	19810	28180	0.657	0.101845
M6	M6	25318	19830	28199	0.6559	0.102167
M7	M7	24039	19800	28159	0.507	0.146312
M8	M8	25240	19770	28113	0.6558	0.102195
M9	M9	25210	19760	28098	0.6538	0.102783
M10	M10	23948	19800	28159	0.4961	0.149544
M11	M11	3495	19490	27709	-1.945	0.873868
M12	M12	25288	19800	28166	0.6559	0.10218
M13	M13	25269	19820	28191	0.6509	0.10365
M14	M14	11607	17900	25408	-0.8376	0.543175
M15	M15	3650	11680	16801	-1.567	0.782751
M16	M16	3718	19230	27319	-1.919	0.863904
M17	M17	3394	15890	22732	-1.826	0.850695
M18	M18	2993	17960	25604	-1.958	0.883104
M19	M19	2342	19390	27602	-2.077	0.915151
M20	M20	3007	19520	27770	-2.002	0.891718

- Varianza de componentes principales (PCA)

PC	eigenvalue	prop_variance	cum_variance
1	11.8958	0.751896204	0.751896204
2	1.06809	0.067510619	0.819406823
3	0.596367	0.037694487	0.85710131
4	0.43632	0.027578419	0.884679729
5	0.385266	0.024351455	0.909031184
6	0.320603	0.02026431	0.929295494
7	0.282865	0.017879009	0.947174504
8	0.225891	0.014277862	0.961452365
9	0.210943	0.013333045	0.974785411
10	0.193865	0.012253599	0.987039009
11	0.115351	0.007290975	0.994329984
12	0.0240884	0.001522552	0.995852536
13	0.0234391	0.001481512	0.997334048
14	0.0221449	0.00139971	0.998733758
15	0.0200141	0.001265028	0.999998786
16	0.019601	0.001238918	1.001237704
17	0.0177292	0.001120607	1.002358311
18	0.0162892	0.001029589	1.003387901
19	0.0154277	0.000975137	1.004363037
20	-0.0690279	-0.004363037	1

- Coordenadas del análisis de Componentes Principales (PCA)

FID	IID	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15	PC16	PC17	PC18	PC19	PC20
M1	M1	0.17327	-0.04548	-0.0406	0.191172	0.005156	0.001126	0.012544	-0.00994	-0.01504	0.138582	-0.00497	0.362957	-0.3136	-0.31433	0.499661	-0.02288	0.509129	-0.07436	0.111192	-0.23196
M2	M2	0.164749	-0.01219	-0.01453	-0.49464	0.014241	0.011985	0.010446	-0.00461	-0.00727	0.099432	-0.00428	0.120359	-0.05081	0.040574	0.439627	0.452658	-0.45612	0.149869	-0.13449	-0.22784
M3	M3	0.173414	-0.04471	-0.03733	0.195416	0.006768	0.000524	0.012122	-0.00914	-0.00645	0.136346	-0.01599	0.102748	-0.01645	0.052257	-0.13882	-0.1622	-0.24864	0.600401	0.613942	-0.22543
M4	M4	0.164819	-0.01066	-0.01248	-0.49132	0.012999	0.007737	0.01303	-0.00544	-0.0094	0.09898	-0.01513	-0.18999	0.190097	0.271121	0.19416	-0.67028	0.190719	-0.03767	-0.00216	-0.22388
M5	M5	0.173663	-0.0453	-0.0404	0.195507	0.007087	-0.00075	0.011509	-0.00641	-0.00875	0.130752	-0.01186	0.005888	-0.21799	0.790551	-0.11997	0.275848	0.147741	-0.26357	0.058043	-0.2296
M6	M6	0.17366	-0.04471	-0.03848	0.191204	0.007633	-0.00116	0.010012	-0.00665	-0.00945	0.136295	-0.01317	0.059727	-0.0434	0.002847	-0.22104	-0.08803	0.09971	0.482025	-0.74522	-0.22596
M7	M7	0.159147	-0.02889	-0.04498	0.03513	0.008088	0.001792	0.006415	-0.00334	0.033871	-0.58144	0.760649	-0.00221	0.011149	0.001506	0.002611	-0.00828	-0.00465	0.001673	-0.00059	-0.22896
M8	M8	0.173606	-0.04703	-0.04214	0.195412	0.007398	0.001653	0.013285	-0.00441	-0.00783	0.144307	-0.01859	-0.00102	0.859246	-0.10116	0.031849	0.258786	0.133385	-0.12104	0.043096	-0.23533
M9	M9	0.165024	-0.01198	-0.01279	-0.49616	0.016885	0.013362	0.012745	-0.00642	-0.00363	0.096034	0.000691	0.066736	-0.13997	-0.30275	-0.63492	0.211541	0.26212	-0.10943	0.13997	-0.22919
M10	M10	0.157676	-0.02596	-0.04219	0.026357	-0.00293	0.002253	0.012501	-0.00928	0.043885	-0.70605	-0.64774	0.003488	-0.00146	-0.00516	0.002808	0.005228	0.005361	0.003392	-8.2E-05	-0.2272
M11	M11	-0.3231	-0.06905	-0.28741	-0.01841	-0.4422	0.075394	-0.1012	-0.74739	0.066428	0.015263	0.004818	-0.00263	0.00159	0.002513	-0.00125	0.000968	0.001297	0.002971	-0.00125	-0.18206
M12	M12	0.17341	-0.0463	-0.03749	0.192315	0.005627	0.001815	0.013914	-0.01332	-0.00731	0.133623	-0.019	0.307544	-0.06068	-0.16421	-0.15386	-0.34341	-0.55454	-0.52194	-0.10918	-0.22811
M13	M13	0.173409	-0.04615	-0.04057	0.192548	0.006024	-0.00063	0.011212	-0.00661	-0.01095	0.134055	-0.01392	-0.83931	-0.22701	-0.2635	0.094691	0.073265	-0.09321	-0.09599	0.02291	-0.22894
M14	M14	-0.09512	0.969794	0.01433	0.048254	0.001052	-0.00073	0.001167	-0.0033	-0.0081	0.013542	0.000322	-0.0004	-8E-05	-0.00052	-0.00029	0.000107	-0.00124	-0.00045	-0.0003	-0.2183
M15	M15	-0.24116	-0.1014	0.758102	0.016232	-0.28986	-0.42889	0.020769	0.012356	-0.0602	-0.00835	0.003574	-2.4E-05	0.000445	-0.00036	-0.00049	0.002361	-8.6E-05	0.000243	-2.1E-05	-0.29037
M16	M16	-0.31504	-0.06883	-0.37806	-0.01874	-0.50525	0.016777	-0.12408	0.660314	0.032015	0.016317	-0.0013	0.0026	-0.00335	-0.00082	-0.00302	-0.00416	-0.00233	0.000148	0.000746	-0.21076
M17	M17	-0.28453	-0.08639	0.340236	0.026727	0.146136	0.844599	-0.04016	0.059513	-0.01372	-0.00383	-0.00076	-0.00237	-0.00091	0.00282	0.001567	-0.00314	0.00048	0.001412	-0.00136	-0.23454
M18	M18	-0.30432	-0.07467	-0.08236	-0.00146	0.521974	-0.24902	-0.58062	0.005722	0.418097	0.028598	-0.00653	0.000584	-0.00165	-0.00064	0.002886	-0.00353	0.000427	-0.00117	0.000604	-0.21844
M19	M19	-0.32649	-0.07652	-0.20692	-0.00725	0.339414	-0.15253	0.05918	-0.02531	-0.8164	-0.04644	-0.00449	0.000969	0.001028	0.000668	-0.00501	-0.00045	-0.00408	0.00143	0.001299	-0.18258
M20	M20	-0.32644	-0.06558	-0.13735	-0.00372	0.234752	-0.10519	0.79374	0.017994	0.381413	0.025689	0.001685	-0.00093	-0.00313	0.000739	0.00268	-0.00106	0.000218	0.002119	-0.00191	-0.16604