

**ESTUDIO DE LA COMPOSICIÓN GENÉTICA DE SAN
BASILIO DE PALENQUE, COLOMBIA, MEDIANTE
MARCADORES DE LINAJE Y AUTOSÓMICOS: UNA
CORRELACIÓN DE LA HERENCIA PATRILINEAL Y
MATRILINEAL**

TESIS DOCTORAL

BEATRIZ MARTÍNEZ, LB, MSc

Director de tesis

LEONOR GUSMÃO, PhD

Professor de la Universidade do Estado do Rio de Janeiro (UERJ),
Brasil

Investigador invitado del Instituto de Patología e Inmunología
Molecular de la Universidad do Porto, Porto, Portugal



Universidad de Cartagena
Doctorado en Ciencias Biomédicas
Instituto de Investigaciones Inmunológicas
Cartagena de Indias D. T. y C, 2022

El presente documento es requisito de grado en el programa de Doctorado en Ciencias Biomédicas, coordinado por el Dr. Luis Caraballo. Este trabajo fue realizado en las instalaciones del Instituto de Investigaciones Inmunológicas de la Universidad de Cartagena, bajo la dirección de la Dra. Leonor Gusmão y con la participación de los tutores

Opta por el grado de Doctor en Ciencias Biomédicas:

Beatriz Martinez

Dra. Leonor Gusmão

Evaluator 1

Evaluator 2

*A mis hijos
Eduardo e Isabella*

“Probablemente seamos una de las últimas
generaciones de Homo sapiens”.

Yuval Noah Harari

Los trabajos realizados como parte de esta tesis, tuvieron el aporte de las siguientes instituciones: Instituto de Investigaciones Inmunológicas de la Universidad de Cartagena; Fundación para la promoción de la investigación y la tecnología (FPIT) del Banco de la República de Colombia; Instituto de Investigação e Inovação em Saúde, i3S, Porto, Portugal; Instituto de Patologia e Imunología Molecular (IPATIMUP), Porto, Portugal; Laboratorio de Diagnósticos por DNA (LDD), Universidade do Estado do Rio de Janeiro (UERJ), Brasil; Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore MD, USA; HRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, Georgia, USA; School of Biology, Georgia Institute of Technology, Atlanta, Georgia, USA, CAAPA Project, Baltimore, USA. Agradezco a mi tutora Leonor Gusmão por impulsarme a terminar esta tarea iniciada muchos años atrás, a mis colegas y amigos del cuerpo docente del Instituto de Investigaciones Inmunológicas: Luis, Javier, Leonardo, Josefina y Nathalie por su apoyo incondicional, y a todos los co-investigadores pertenecientes a las anteriores instituciones por su aporte en los análisis experimentales y participación como coautores en los artículos científicos publicados.

Agradecemos de manera muy especial la colaboración de Regina Miranda, Profesora de la Escuela Benkos Biojó quien ayudó en la selección de las muestras y a todos los niños voluntarios de la escuela de niños Benkos Biojó de San Basilio de Palenque y de Cartagena.

NOTA DE APROBACIÓN DEL JURADO

PRESIDENTE DEL JURADO

JURADO

JURADO

JURADO

Cartagena de Indias D. T. y C., 2022.

La Universidad de Cartagena ni el jurado examinador se hacen responsables de los conceptos emitidos en el presente trabajo

CONTENIDO

RESUMEN	21
ABSTRACT	23
1. INTRODUCCIÓN Y ALCANCES DE LA INVESTIGACIÓN....	25
CAPÍTULO I. Descripción de la Tesis.....	31
2. FUNDAMENTACIÓN TEÓRICA.....	33
2.1. Polimorfismos de ADN en la historia de las poblaciones humanas	35
2.2. Marcadores uniparentales	40
2.2.1. Cromosoma Y	41
2.2.2. La mitocondria.....	52
2.3. Marcadores de ancestría en cromosomas autosómicos	63
2.4. Marcadores de identificación humana de tipo InDel.....	69
2.5. La Diáspora Africana.....	72
2.6. Estructura Poblacional y Diversidad Genética entre las Poblaciones Africanas	75
2.7. Diversidad Etnolingüística en África.....	77
2.8. La expansión bantú.....	79
2.9. Perfil Genómico de las Poblaciones Contemporáneas en América de la Diáspora Africana	82
2.10. Palenque y sus Orígenes Africanos	84
2.11. Ancestría de la población colombiana: desde la colonización europea hasta los días de hoy	88
3. OBJETIVOS.....	91
3.1. Objetivo general.....	91
3.2. Objetivos específicos.....	91
4. METODOLOGÍA.....	93
4.1. Obtención de linajes paternos y maternos con marcadores uniparentales (Objetivos 1 y 2).....	93
4.1.1. Muestra poblacional de SBP.....	93
4.1.2. Extracción del ADN para muestras de San Basilio de Palenque, Bolívar, Colombia.....	94
4.1.3. Genotipificación de marcadores del tipo microsatélite del cromosoma Y (Y-STR).....	94
4.1.4. Genotipificación de SNPs del cromosoma Y	96
4.1.5. Secuenciación de la región control del ADN mitocondrial	97
4.1.6. Construcción de redes filogenéticas para la identificación y determinación del origen de los principales linajes africanos existente en SBP	99

4.1.7. Análisis estadístico	99
4.2. Inferencia de la ancestría continental y subcontinental en SBP y análisis de mezcla sesgada por sexo (objetivos 3, 4 y 5).....	100
4.2.1. Muestra poblacional de SBP y Chocó	100
4.2.2. Secuenciación de genoma completo	100
4.2.3. Análisis de marcadores informativos de ancestría (AIMs)	101
4.2.4. Análisis de marcadores uniparentales.....	101
4.3. Muestras de referencia global para los estudios de GWS	102
4.4. Control de calidad y armonización de SNPs	103
4.5. Inferencia de ancestría genética.....	104
4.6. Relación entre la diversidad lingüística y genética en Palenque	105
4.7. Caracterización de marcadores HID-InDels y comparación con otras poblaciones mezcladas y nativas de Colombia (objetivo 6). 106	
4.7.1. Muestreo	106
4.7.2. Genotipado de los 38 HID-InDels	107
4.7.3. Análisis estadístico	108
CAPÍTULO II. Caracterización de los haplotipos de ChrY y de ADNmt en San Basilio de Palenque	113
5. CARACTERIZACIÓN DE LOS HAPLOTIPOS DE ChrY Y DE ADNmt EN SBP.....	115
5.1. Diversidad genética en SBP	116
5.2. Los orígenes africanos de los linajes de San Basilio de Palenque	118
5.3. Análisis filogeográfico de los linajes africanos de ADNmt y ChrY en Palenque.....	120
5.3.1. Linajes africanos de ADNmt	120
5.3.2. Linajes africanos de ChrY	131
CAPÍTULO III. Identificación de los Componentes Principales de Ancestría Continental y Subcontinental en SBP por GWAS	145
6. IDENTIFICACIÓN DE LOS COMPONENTES PRINCIPALES DE ANCESTRÍA CONTINENTAL Y SUBCONTINENTAL EN SBP POR GWAS	147
6.1. Análisis de Ancestría Continental	147
6.2. Análisis de mezcla con sesgo sexual mediante marcadores uniparentales de ADNmt (femenino) y de ChrY (masculino) junto con comparaciones de los niveles de ancestría del cromosoma X frente a los autosómicos.	151
6.3. Ancestría genética africana y grupos lingüísticos	153

6.3.1. Orígenes genéticos y lingüísticos africanos de Palenque.	157
6.3.2. Análisis de marcadores de haplotipos de ChrY para interrogar los orígenes paternos genéticos y lingüísticos africanos de Palenque.	162
6.3.3. Diversidad genética y efectos fundadores.	164
6.4. Ancestría nativa americana y europea	165
CAPÍTULO IV. Diferenciación poblacional de SBP con poblaciones colombianas mezcladas y nativas mediante estudios con HID-InDels	169
7. DIFERENCIACIÓN POBLACIONAL DE SBP CON OTRAS POBLACIONES COLOMBIANAS MEZCLADAS Y NATIVAS MEDIANTE ESTUDIOS CON HID-InDels	171
7.1. Comparación poblacional.	171
7.2. Representación MDS de la matriz F_{ST} obtenida en comparaciones entre las muestras de 10 poblaciones colombianas y dos poblaciones de referencia de África y Europa.	172
7.3. Frecuencias alélicas, equilibrio de Hardy-Weinberg y parámetros forenses	174
7.4. Comparación del poder de exclusión de 38 HID-InDels y 46 AIMS-InDels en poblaciones colombianas	177
CAPÍTULO V. Discusión	179
8. RAÍCES AFRICANAS EN SBP POR MARCADORES UNIPARENTALES	181
8.1. Ancestría y raíces africanas de San Basilio de Palenque encontradas con marcadores uniparentales.	181
8.2. Ancestría continental y subcontinental, mezcla por sesgo sexual, efecto fundador en SBP con los análisis de WGS	185
8.3. Ancestría africana y grupos lingüísticos en SBP	186
8.4. Palenque en comparación con otras poblaciones colombianas mezcladas y nativas	188
9. REFERENCIAS	191
10. ANEXOS	229
11. FIGURAS COMPLEMENTARIAS	263
12. TABLAS COMPLEMENTARIAS	273

LISTA DE FIGURAS

Figura 1. Patrones de herencia de los segmentos genéticos no-recombinantes (cromosoma Y y ADNmt) y recombinantes (autosomas) del genoma.	41
Figura 2. El Cromosoma Y	43
Figura 3. Versión reducida del árbol filogenético del cromosoma Y ..	45
Figura 4. Distribución global de los haplogrupos de cromosoma Y...46	
Figura 5. Ilustración esquemática de la replicación STR-slippage.	51
Figura 6. Mapa del genoma mitocondrial humano.....	54
Figura 7. Detalle de la región control o D-Loop.	56
Figura 8. Clases de Heteroplasmia.	60
Figura 9. Proporción de ancestría africana, europea y nativa americana estimada con ADNmt, cromosoma Y, cromosoma X y marcadores autosómicos en trece muestras de población latinoamericana.	67
Figura 10. Rutas de la trata de esclavos desde 1500 hasta 1900.	73
Figura 11. La distribución de las cuatro principales familias lingüísticas africanas y la representación esquemática de la expansión bantú.	81
Figura 12. Localización geográfica de SBP, Bolívar, Colombia.	86
Figura 13. Representación gráfica de la genotipificación de Y-STRs ..	95
Figura 14. Resumen de la metodología para obtener los linajes en marcadores uniparentales	95
Figura 15. Árbol de Y-SNPs analizados en este estudio y métodos utilizados.	96
Figura 16. Resumen de la metodología para ADNmt: A) Reacción de PCR. B) Secuenciación	98
Figura 17. Lugares de muestreo y grupos de ancestría continental de las poblaciones de referencia global, codificadas por colores.	103
Figura 18. Número total de muestras analizadas de cada región natural colombiana, incluyendo las muestras de este estudio, las de Pereira et al. (2018) y de Martínez et al. (2013) (Palenque).	107
Figura 19. Representación en red de los haplotipos CR completos (incluidos los polimorfismos InDel) dentro de los haplogrupos africanos.	116
Figura 20. Representación en red de los haplotipos basados en 27 Y-STRs encontrados en Palenque dentro de los clados E y R.	118
Figura 21. Gráfico MDS de distancias genéticas F_{ST} por pares entre el subconjunto de linajes africanos en Palenque y diferentes poblaciones africanas. Tomado de Martínez B, 2020	120

Figura 22. Representación en red del haplogrupo L0a1a + 200.....	121
Figura 23. Representación en red del haplogrupo L0a1a + 200.....	122
Figura 24. Representación en red de haplogrupos dentro de L1c. ...	124
Figura 25. Representación en red del haplogrupo L2a1.....	125
Figura 26. Representación en red del haplogrupo L2a1 + 16189 (16192).	126
Figura 27. Representación en red del haplogrupo L2a1 + 143 + 16309.	127
Figura 28. Representación en red del haplogrupo L2b1a.....	128
Figura 29. Representación en red del haplogrupo L2d + 16129.	129
Figura 30. Representación en red de haplogrupos dentro de L3.	130
Figura 31. Representación en red de los haplotipos seleccionados que podrían pertenecer a YMRCA * (xM13, SRY10831.1).....	132
Figura 32. Representación en red de haplotipos seleccionados dentro del clado B-M60 (excepto aquellos que se sabe que portan el alelo derivado de M109 / M152).....	134
Figura 33. Representación en red de los haplotipos seleccionados dentro del clado E1b1b-M35 (excepto los que portan los alelos derivados de M78, M81, M123, V6, M293).	136
Figura 34. Representación en red de haplotipos seleccionados en el interior del haplogrupo E1a-M33.	138
Figura 35. Representación en red de haplotipos seleccionados dentro del clado R1b-P25 (excluidos los asignados a sublinajes fuera de la rama R1b-V88).....	143
Figura 36. Ancestría continental y mezcla	148
Figura 37. Correlación de estimaciones de ancestría continental. ...	150
Figura 38. Análisis de marcadores informativos de ancestría (AIM).	151
Figura 39. Ancestría y mezcla con sesgo sexual.	152
Figura 40. Reanálisis de la ancestría y mezcla con sesgo sexual.	153
Figura 41. Ancestría genética africana y grupos lingüísticos.....	155
Figura 42. Ancestría y estructura genética para poblaciones africanas de referencia y poblaciones afrodescendientes del Nuevo Mundo. ..	156
Figura 43. Ancestría y estructura genética para poblaciones africanas de referencia y poblaciones afrodescendientes del Nuevo Mundo. ..	157
Figura 44. Ancestría genética, geografía y diversidad lingüística de las lenguas bantú.....	158
Figura 45. Estructura genética africana y ancestría de poblaciones afrodescendientes en las Américas.....	160

Figura 46. La prueba F3 muestra similitud entre las poblaciones africanas de referencia y las poblaciones americanas mezcladas.	160
Figura 47. Distribuciones de tamaño de haplotipo (cM) específicas de ancestría africana.....	161
Figura 48. Correlación de estimaciones de ancestría continental.	162
Figura 49. Análisis de haplotipos de ChrY	163
Figura 50. Gráfico que relaciona los grupos lingüísticos de las poblaciones africanas de referencia para los haplotipos de cromosoma Y.	164
Figura 51. Diversidad genética.....	165
Figura 52. Ancestría de los nativos americanos	167
Figura 53. La prueba F3 muestra similitud entre las poblaciones de referencia nativas americanas y las poblaciones americanas mezcladas.	168
Figura 54. Representación MDS de la matriz de FST obtenida en las comparaciones por pares entre 10 muestras poblacionales de Colombia, y dos poblaciones de referencia de África y Europa.	174
Figura 55. Origen africano más probable de los linajes del ADNmt y del cromosoma Y detectados en SBP.....	183

LISTA DE TABLAS

Tabla 1. Lista de primers utilizados para análisis de ADNmt.	98
Tabla 2. Listado de marcadores HID-InDels utilizados en esta investigación y diseñados para ser usados en identificación humana.	109
Tabla 3. Frecuencias de los haplogrupos del ADNmt y del ChrY detectados en una muestra de la población de Palenque.	117
Tabla 4. Resultados de los análisis de coincidencia de haplotipos entre las muestras E1b1a-M2* (xM154, M191) de Palenque y de diferentes países africanos (consulte la Tabla S5 para obtener referencias a los datos consultados).	140
Tabla 5. Resultados de los análisis de coincidencia de haplotipos entre las muestras E1b1a-M191 de Palenque y de diferentes países africanos (consulte la Tabla S5 para obtener referencias a los datos consultados).	141
Tabla 6. Porcentajes de ancestría continental para las poblaciones afrodescendientes en las Américas.	150
Tabla 7. Porcentajes de ancestría africana para las poblaciones afrodescendientes en las Américas.	159
Tabla 8. Distancias genéticas por pares (FST) entre las poblaciones colombianas, europea y africana (diagonal inferior) y los correspondientes valores de probabilidad de no diferenciación (diagonal superior).	173
Tabla 9. Parámetros de relevancia forense en todas las muestras estudiadas para los 38 HID-InDels.	176
Tabla 10. Parámetros de relevancia forense calculados para los 38 HID-InDels, 46 AIMInDels y el conjunto completo de InDels en muestras de las cinco regiones naturales de Colombia.	178

LISTA DE FIGURAS COMPLEMENTARIAS

Figura Complementaria 1. Localización geográfica de las muestras en las distintas regiones de Colombia.	263
---	-----

LISTA DE TABLAS COMPLEMENTARIAS

Tabla Complementaria 1. Detalles sobre todas las poblaciones de referencia analizadas para los estudios de WGS.	273
Tabla Complementaria 2. Lista de haplotipos de mtDNA y haplogrupos en SBP.	284
Tabla Complementaria 3. Lista de haplotipos de cromosoma Y en SBP.	292
Tabla Complementaria 4. Fuente de los datos utilizados en análisis de distancias genéticas por pares (FST) y filogeográficos (Phyl) de linajes africanos de ADNmt.	299
Tabla Complementaria 5. Fuente de los datos utilizados en los análisis de distancias genéticas por pares (FST) y filogeográficos (Phyl) de los linajes del cromosoma Y africano.	300
Tabla Complementaria 6. Valores de FST por pares entre Palenque (linajes africanos) y poblaciones africanas para las regiones HVI y HVII (debajo de la diagonal) y los valores p de no diferenciación correspondiente (arriba de la diagonal).	302
Tabla Complementaria 7. Valores de FST por pares entre Palenque (linajes africanos) y poblaciones africanas para 10 Y-STRs (debajo de la diagonal) y los valores p de diferenciación correspondientes (arriba de la diagonal).	303
Tabla Complementaria 8. Frecuencias alélicas, heterocigocidad esperada y observada y equilibrio de HW en 10 poblaciones colombianas.	306
Tabla Complementaria 9. Tabla Complementaria 9. Distancias genéticas por pares (FST) entre las poblaciones colombianas, europeas y africanas (diagonal inferior) y los valores P de no diferenciación correspondientes (diagonal superior).	307

LISTA DE ANEXOS

Anexo A. Consentimiento informado de toma de muestras en SBP..	229
Anexo B	230
Anexo C. (extendido en anexo K).....	230
Anexo D	230
Anexo E	230
Anexo F	231
Anexo G	231
Anexo H. Otros trabajos realizados dentro del marco del proyecto ..	231
Anexo I. Otros trabajos realizados dentro del marco del proyecto....	231
Anexo J. Otros trabajos realizados dentro del marco del proyecto ...	232
Anexo K. Genetic ancestry, language, and African origins of San Basilio de Palenque: the first free town in the Americas	232

ABREVIATURAS

ADNmt: ADN mitocondrial.

AGVP: Proyecto de Variación del Genoma Africano.

AIM-InDels: Marcadores informativos de ancestría tipo inserción/delección.

AIMs: Marcadores informativos de ancestría.

ARNr: ARN ribosomal.

ARNt: ARN transferente.

ASS: África subsahariana.

ChrY: el cromosoma Y humano.

CNVs: Variantes de número de copias.

CR: Región control de ADNmt.

CRS: Secuencia de Referencia de Cambridge revisada rCRS: Secuencia de Referencia de Cambridge D-Loop: Asa de desplazamiento.

DNApol: ADN polimerasa.

EHW: Equilibrio de Hardy-Weinberg.

FIS: coeficiente de consanguinidad.

FIS: coeficiente de consanguinidad de Wright.

FROH: las tasas de homocigosidad.

F_{ST} : índice de fijación.

GHEP: Grupo de habla española y portuguesa de la ISFG.

HGDP: Proyecto de Diversidad del Genoma Humano.

HID-InDels: Marcadores de inserción-delección de uso en identificación humana.

HVI: región hipervariable I del ADNmt

HVII: región hipervariable II del ADNmt

HVIII: región hipervariable III del ADNmt

IHGSC: Consorcio Internacional de Secuenciación Genómica Humana

InDel: Marcadores de inserción-delección.

IP: sigla que significa Índice de Paternidad.

ISFG: sociedad internacional de genética forense.

LD: desequilibrio de ligamiento.

Ma: Millones de años.

MEGA: Infinium Multi-Ethnic Genotyping Array.

MPS: Secuenciación masiva en paralelo.

MSY: Región específica masculina del cromosoma Y.

NGS: Secuenciación de próxima generación.
NRY: Región no recombinante del cromosoma Y.
PanAFR: Affymetrix® Axiom® Genome-Wide Pan-African.
PAR: Región pseudo autosómica.
PCA: Análisis de componentes principales.
PCR: Reacción en cadena de la polimerasa.
PD: Poder de Discriminación.
PE: Probabilidad de Exclusión.
PM: Probabilidad de Coincidencia (en inglés probability of match).
PIC: Contenido de Información Polimórfica.
RFHG: Grupos cazadores recolectores de la selva tropical ROH: fracción del genoma en series de homocigosidad.
SBE: extensión de base única.
SBP: San Basilio de Palenque.
SMM: modelo de mutación paso a paso.
SNP: polimorfismo de un solo nucleótido.
STR: Repetición corta en tándem.
UEP: Polimorfismo de único evento.
VNTR: Repeticiones en tándem de número variable.
WGS: secuenciación de genoma completo.
YAP: polimorfismos bialélicos como las inserciones Alu o marcadores de tipo Alu.
YHRD: Base de datos de referencia de haplotipos de cromosoma Y.
Y-STR: Repetición corta en tándem, en la región específica de cromosoma Y Y-SNP: polimorfismo de un solo nucleótido, en la región específica de cromosoma Y 1KGP: Proyecto de los 1000 genomas.



*Todos los derechos reservados
Copyright © 2019 por Armin*

RESUMEN

En la costa Caribe Colombiana, en el departamento de Bolívar, se encuentra la población de San Basilio de Palenque. Es una comunidad fundada como refugio para los esclavos africanos fugitivos desde hace más de 400 años. El reconocimiento de la importancia histórica y cultural de Palenque ha impulsado varios estudios, principalmente sobre las raíces africanas de sus primeros habitantes. Además, los estudios lingüísticos que demuestran un sustrato africano único para el “Palenquero”, la lengua bantú Kikongo, han llevado a la hipótesis de un origen genético africano único para la población de Palenque. Para profundizar en el conocimiento del origen y la diversidad de los linajes parentales de Palenque, analizamos una muestra de 81 individuos para toda la Región Control del ADNmt, así como 92 individuos para 27 Y-STRs y 95 para 51 SNPs de cromosoma Y. Además, realizamos un análisis integrado de secuencias del genoma completo, con marcadores informativos de ancestría (AIMs) y marcadores uniparentales, en 371 individuos de Palenque, Chocó y África. Se consideraron patrones a gran escala de ancestría genética africana en el contexto de grupos lingüísticos africanos en un esfuerzo por resolver la relación entre la diversidad genética y lingüística en Palenque. Además 114 muestras fueron caracterizadas para marcadores HID-InDels con el fin de establecer parámetros de uso forense y comparar SBP con otras poblaciones mezcladas y nativas de Colombia. Los resultados demostraron que su ancestría genética continental es predominantemente africana (80%), seguida de una mezcla de partes iguales de europeos (10%) y nativos americanos (10%) y confirmaron

el fuerte aislamiento de Palenque, con cierto grado de afluencia de linajes maternos nativos americanos y una mezcla europea mediada exclusivamente por hombres. A pesar de su aislamiento, Palenque no muestra ninguna evidencia de diversidad genética reducida debida a efectos fundadores o endogamia. El enfoque filogenético permitió inferir el origen geográfico de algunos linajes de distintas regiones africanas. Con el análisis de los marcadores forenses se observó además que esta población presenta diferencias con las poblaciones mezcladas vecinas de la costa caribe colombiana, lo que permite concluir que para casos forenses y de paternidad debe usarse una base de datos de marcadores genéticos específica de la población. En conclusión, San Basilio de Palenque posee una ancestría genética africana diversa y heterogénea, con contribuciones de numerosas poblaciones africanas de regiones geográficas amplias y muchos grupos lingüísticos distintos de la familia Níger-Congo. Los orígenes genéticos africanos diversos son claramente distintos de los orígenes lingüísticos uniformes de la lengua “Palenquero”. En contradicción con estudios previos que indicaban un solo origen africano, nuestros resultados evidencian contribuciones genéticas paternas de regiones africanas muy diferentes y postulamos que la lengua “kikongo” se estableció temprano como lengua franca y luego fue adoptada por todos los miembros de la población de San Basilio de Palenque, independientemente de sus orígenes africanos.

Palabras claves: Linaje, África, Cromosoma Y, ADN mitocondrial, Palenque, historia oral, autosomas, GWS, ancestría subcontinental, HID-InDels, Kikongo, Colombia.



ABSTRACT

On the Colombian Caribbean coast, in Bolívar, is the town of San Basilio de Palenque. It is a community founded as a refuge for runaway African slaves for more than 400 years. The recognition of the historical and cultural importance of Palenque has prompted several studies, mainly on the African roots of its first inhabitants. Furthermore, linguistic studies showing a unique African substrate for “Palenquero”, the Bantu Kikongo language, have led to the hypothesis of a unique African genetic origin for the Palenque population. To deepen our understanding of the origin and diversity of Palenque parental lineages, we analyzed a sample of 81 individuals for the entire mtDNA Control Region, as well as 92 individuals for 27 Y-STRs and 95 for 51 Y chromosome SNPs. In addition, we performed an integrated analysis of whole genome sequences, with ancestry informative markers (AIMs) and uniparental markers in 371 individuals. Large-scale patterns of African genetic ancestry were considered in the context of African language groups in an effort to resolve the relationship between genetic and linguistic diversity in Palenque. In addition, 114 samples were characterized for HID-InDel markers to establish parameters for forensic use and compare with other admixed and native populations of Colombia. The results showed that their continental genetic ancestry is predominantly

African (80%), followed by a mixture of equal parts European (10%) and Native American (10%). The results confirmed the strong isolation of Palenque, with some influx of Native American maternal lineages and an exclusively male-mediated European admixture.

Despite its isolation, Palenque does not show any evidence of reduced genetic diversity due to founder events or endogamy. A phylogenetic approach allowed inferring the phylogeographic origin of some lineages in Palenque from different African regions. With the analysis of the forensic markers, it was also observed that it presents differences with the neighboring admixed populations of the Colombian Caribbean coast, which allows us to conclude that for forensic and paternity cases a database of genetic markers specific to the population should be used. In conclusion, San Basilio de Palenque possesses a diverse and heterogeneous African genetic ancestry, with contributions from numerous African populations from wide geographic regions and many different language groups from the Niger-Congo region. The diverse African genetic origins are clearly distinct from the uniform linguistic origins of the "*Palenquero*" language. In contradiction to previous studies indicating a single African origin, our results show paternal genetic contributions from very different African regions, and we postulate that the "Kikongo" language was established early as a lingua franca and was later adopted by all members of the population from San Basilio de Palenque, regardless of their African origins.

Keywords: Lineage, Africa, Y chromosome, mitochondrial DNA, Palenque, oral history, autosomes, subcontinental ancestry, HID-InDels, Kikongo, human identification, Colombia.



1. INTRODUCCIÓN Y ALCANCES DE LA INVESTIGACIÓN

En el territorio colombiano el mayor porcentaje de la población afrodescendiente se encuentra localizada en las costas caribes y pacífica. Existe el imaginario social de que la población negra afrocolombiana llegó a los territorios de la actual Colombia como un grupo homogéneo y que aún lo es, pero esta es una creencia equivocada. La población afrocolombiana posee una gran diversidad no solo cultural y regional, sino también genética, con una marcada ancestría lingüística, étnica y cultural africana. Esta población se formó a partir de individuos provenientes de diversas regiones y etnias de África que llegaron al continente en calidad de esclavos. Estudios de la diáspora africana hacia las Américas, entre los siglos XVI – XIX, confirman lo anterior y han permitido localizar diferentes puntos de salida de África, principalmente a lo largo de las costas sur y occidente. Adicionalmente a mediados del año 1700, también fueron traídos esclavos de la costa este de África, particularmente de Mozambique (Arrazola, 1970; Friedemann & Patiño Rosselli, 1983; Friedemann, 1998; Girón, 1621; Mincultura, 2020).

En la costa caribe colombiana, departamento de Bolívar, se encuentra la población de San Basilio de Palenque (SBP), una comunidad fundada por esclavos afrodescendientes que se daban a la fuga desde el siglo XVI estableciéndose en unas fortificaciones llamadas “Palenques”. La comunidad de Palenque de San Basilio conserva una conciencia étnica que le permite entenderse como un pueblo específico, con la única lengua criolla en América llamada

“*Palenquero*”, resultante de la diáspora africana en el continente americano y que tiene como base el léxico español.

Publicaciones del área de genética, del año 1996, realizadas por nuestro grupo de investigación en el Instituto de Investigaciones Inmunológicas de la Universidad de Cartagena, usando marcadores HLA, reportaron que la composición genética de esta población provenía de las costas del occidente y sur de África, encontrándose las distancias genéticas más pequeñas con individuos de la región de Mali (Jiménez et al., 1996). Posteriormente, en el 2009 y 2018, Arnaiz-Villena y col. reportaron que SBP era genéticamente cercano a otras poblaciones afrodescendientes del norte y sur de América y a los grupos de habla bantú de África Occidental (senegaleses; bubi, golfo de Guinea). Además, reportan un flujo genético muy limitado de amerindios y europeos, de acuerdo a reportado por los registros históricos, y encuentran algunos haplotipos HLA extendidos únicos de esta población (Arnaiz-Villena et al., 2018).

Existen solo dos trabajos previos a este en donde se analizan marcadores de linaje en SBP, incluyendo un número limitado de marcadores genéticos en cromosoma Y y, además, estudiando solo las regiones hipervariables HVI y HVII en el ADNmt (Ansari-Pour et al., 2016; Noguera et al., 2014). Ansari-Pour et al. (2016) concluyen que la herencia paterna africana de SBP proviene de la región del Congo de África, con sus raíces en la etnia Yombe. En este último estudio de Ansari y col., 2016, no se logró evidenciar con exactitud el origen de la ancestría africana materna en SBP. En particular, estos datos no fueron suficientes para respaldar una evaluación exhaustiva de la hipótesis del origen bantú único para SBP y de la lengua palenquera.

Por otro lado, genotipando marcadores de ancestría autosómicos tipo AIM-InDels, en el año 2017, nuestro grupo demostró que SBP posee un 80% de ancestría africana, un 10% europea y 10% amerindia (Martínez et al., 2017).

Con el objetivo de obtener un conocimiento más profundo de la diversidad y del origen de los linajes parentales de SBP, en el presente trabajo analizamos toda la región control (CR) del ADNmt, así como una batería de 51 marcadores tipo SNP específicos del cromosoma Y. Basándonos en comparaciones con bases de datos de poblaciones

africanas, pretendimos contrastar la hipótesis de un origen africano único contra la de múltiples orígenes en esta población. Por primera vez, para esta población afrodescendiente hicimos un análisis de redes filogenéticas de cada uno de los haplogrupos encontrados en ambos marcadores.

Adicionalmente realizamos un análisis integrado de marcadores uniparentales y marcadores informativos de ancestría (AIMs), a partir de secuencias del genoma completo (WGS), en un esfuerzo por explorar más profundamente no solo los orígenes genético africano sino también analizar fenómenos de mezcla sesgada por sexo, efecto fundador y endogamia. Este último enfoque se basó en comparaciones de secuencias de todo el genoma entre las muestras de SBP y poblaciones de referencia que cubren las tres principales regiones africanas que participaron en la trata transatlántica de esclavos a Colombia. Investigamos la ancestría genética en el contexto de los grupos lingüísticos africanos en un esfuerzo por resolver la relación entre la diversidad lingüística y genética en Palenque. Además, comparamos los resultados observados entre Palenque y otras poblaciones afrodescendientes en las Américas, con énfasis en la población del Chocó de la Costa Pacífica de Colombia. También comparamos la estructura genética de SBP con la de poblaciones colombianas mezcladas y nativas mediante el análisis de marcadores de uso en aplicaciones forenses como los HID-InDels. La alta divergencia entre las poblaciones colombianas tiene un impacto importante también en estudios de identificación humana.

Los anteriores polimorfismos genéticos se han utilizado cada vez más para complementar las evidencias lingüísticas, arqueológicas y antropológicas de los orígenes y patrones de migración de las poblaciones humanas modernas. Estos muestran diferencias sutiles en las frecuencias alélicas a lo largo de todo el genoma, lo que conlleva a una estructuración geográfica de las poblaciones. Mediante la información genética de los individuos, las poblaciones se pueden agrupar en distintas regiones geográficas. Los estudios realizados con genoma completo permitieron clasificar cinco grupos continentales a nivel mundial divididos así: africanos, un grupo extenso que incluye los

Europeos, asiáticos de Oriente y del sur, asiáticos del este, Oceanía y nativos autóctonos.

En el caso de las poblaciones que conforman el continente americano, los pueblos nativos en todo el continente experimentaron un cambio demográfico sorprendente provocado por la llegada de los europeos y la posterior entrada de los esclavos africanos. Son muchos los estudios que hasta el momento y con distintas metodologías moleculares, han permitido avanzar en la caracterización de los patrones genéticos de estos tres componentes ancestrales continentales en las poblaciones mezcladas. Adicionalmente, son pocos los estudios realizados en poblaciones afrodescendientes y nativas de los diferentes países, en donde todavía quedan preguntas por resolver acerca de su reciente historia demográfica y estructura poblacional. Además, sigue sin esclarecerse el inicio de la mezcla europea en el Caribe y el impacto de las numerosas olas de inmigrantes europeos en siglos posteriores a la colonización, así como los orígenes subcontinentales de África que contribuyeron a los genomas del Caribe actuales.

En muchos lugares de América Latina y en el Caribe, el comercio de esclavos africanos provocó la formación de aldeas fortificadas cuyos habitantes eran fugitivos que ofrecían resistencia al régimen colonizador. En Colombia, estos pueblos fueron conocidos como palenques y fueron famosos por su resistencia a la conquista militar española. San Basilio de Palenque (SBP) es uno de estos ejemplos, localizado en la costa caribe colombiana.

Actualmente es un corregimiento del municipio de Mahates, en el departamento de Bolívar, a 50 km al suroriente de la ciudad de Cartagena. El casco urbano de San Basilio de Palenque está dividido en dos barrios, Barrio Arriba y Barrio Abajo, separados por la iglesia y el puesto de salud.

El uso de marcadores polimórficos, no sujetos a recombinación y en estado haploide, como son los ubicados en el ADN mitocondrial (ADNmt) y en el cromosoma Y humano (ChrY), permiten un análisis más exacto y detallado de los movimientos migratorios que originaron las poblaciones actuales en los diferentes continentes. Estos marcadores son transmitidos casi intactos de padres a hijos e identifican relaciones diferentes referentes a cada progenitor. El segmento exclusivo del ChrY

es transmitido de padre a hijo y el ADNmt se transmite de madre a toda la descendencia, a través del ovulo en la fecundación. La información transmitida se mantiene en la descendencia y se presentan modificaciones posteriores o mutaciones. Por causa de sus patrones de herencia, estos marcadores también llamados uniparentales permiten complementar información independiente acerca de la ancestría materna (ADNmt) y paterna (ChrY) de los pueblos.

Entre las ventajas que tiene el estudio de linajes paternos, es que estos suelen presentar una mayor diferenciación interpoblacional en relación a los linajes maternos o biparentales. Esta elevada diferenciación, manifestada por las elevadas distancias genéticas entre poblaciones, especialmente de distintos continentes, y la marcada especificidad geográfica, hace que el estudio de linajes del cromosoma Y sea una herramienta extremadamente útil para inferir sobre la ancestralidad patrilínea.

Los polimorfismos localizados en la región no recombinante del cromosoma Y (NRY, del inglés Non-Recombining Y) son transmitidos en bloque de una generación a otra, siendo modificados solamente por la mutación. Tanto los microsatélites (o STRs) como los polimorfismos de una sola base (SNPs) del cromosoma Y han sido ampliamente utilizados en la investigación de la ancestría patrilínea de las poblaciones humanas. Los SNPs presentan una tasa de mutación mucho menor que los STRs y, un conjunto de estos, permiten establecer los denominados haplogrupos (o linajes masculinos), cuya distribución geográfica y origen poblacional están bien definidos. En vista de esto, fueron usados para ampliar el conocimiento de la herencia paterna y materna africana en una muestra poblacional reciente de San Basilio de Palenque. Los resultados de este trabajo destacan la importancia del conocimiento no solo a nivel local sino en todo el país de la estratificación de la población actual, su historia, así como la relación entre sus linajes ancestrales y la lengua palenquera. Por otro lado, la caracterización adicional de marcadores de uso en identificación humana y forense, como es el caso de los AIMInDels y de los HID-InDels, deja a disposición para la comunidad nacional e internacional bases de datos útiles tanto para desarrollar estudios de asociación con enfermedades como para ser utilizadas en análisis forense y de

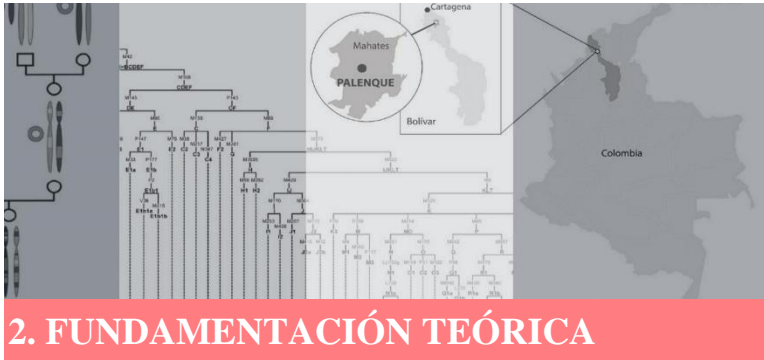
identificación humana. Dentro del contexto de la población actual de Colombia, se pudo corroborar que poblaciones pequeñas aisladas con un fondo genético diferente al de las poblaciones urbanas vecinas y mezcladas deben ser tratadas de manera particular en estudios de asociación con enfermedades, en el análisis forense y pruebas de identificación humana, por lo que este trabajo genera un aporte importante para la comunidad científica de Colombia.

Los resultados en conjunto destacan un conocimiento más detallado de la historia y orígenes de una población afrodescendiente colombiana, no solo a nivel local sino de todo el país, así como la relación entre sus raíces genéticas ancestrales y la lingüística. Además, se describen por primera vez bases de datos de frecuencias alélicas de marcadores genéticos de uso en identificación humana y forense (AIM-InDels y HID-InDels).



CAPÍTULO I.
Descripción de la Tesis

*Copyright © 2019 por Armin Schwegler
Todos los derechos reservados.*



Las poblaciones humanas se caracterizan por una amplia variación genética entre los individuos. Esta variabilidad tiene su origen en las mutaciones y está determinada por fuerzas evolutivas como la selección natural, la deriva genética, la migración (o flujo génico) y la recombinación. Como resultado, todos los individuos tienen un genoma único con variantes características, a excepción de los gemelos monocigotos.

El estudio inicial de la secuenciación del genoma humano publicado hace más de 20 años marcó el inicio de una nueva era que trajo nuevos conocimientos en el estudio de las poblaciones humanas. El genoma haploide humano contiene alrededor de $3,2 \times 10^9$ bases, de las cuales solamente una fracción pequeña es ADN codificante, alrededor del 2%, mientras que la gran mayoría es ADN no codificante que incluye regiones intrónicas de genes, heterocromatina y un número extenso de secuencias repetitivas. De otro lado se obtuvo información relevante sobre los distintos polimorfismos existentes a lo largo de todo el genoma, entre los cuales se encuentran las repeticiones de secuencias simples como minisatelites (VNTR), repeticiones en tándem como microsátelites (STRs) y polimorfismos de un solo nucleótido (SNPs) (Lander et al., 2001; Sachidanandam et al., 2001).

La publicación posterior de una secuencia de referencia de alta calidad por el *International Human Genomic Sequencing Consortium* (IHGSC, 2004), y el proyecto de los *1000 Genomas* (Consortium, 2015), evidenciaron otras variantes genéticas con un papel importante en la patogénesis de las enfermedades hereditarias, tales como variantes

estructurales que comprenden segmentos mayores de alteración genómica de más de 1kb de longitud (Conrad & Hurles, 2007; Feuk et al., 2006). Las variantes estructurales constituyen un grupo heterogéneo de polimorfismos incluyendo deleciones, duplicaciones, variantes de número de copias (CNVs) así como inserciones, inversiones y translocaciones (Conrad & Hurles, 2007).

En los últimos años, varios estudios han dedicado la atención a este tipo de variantes estructurales con un enfoque especial en las variantes en número de copias (CNVs) (Conrad et al., 2010; Redon et al., 2006), y también a los polimorfismos de inserción-delección (InDels), extendiéndose su estudio en las áreas de genética de poblaciones y forense (Mills et al., 2011).

Por otro lado, el estudio de diferentes tipos de marcadores en las porciones no recombinante del genoma (ADNmt y ChrY), desde los años ochenta hasta la actualidad, se convirtió en una herramienta de mucha utilidad para conocer la diversidad y origen genético de las poblaciones humanas. En el caso del ADNmt, que se transmite solo por herencia materna, ha sido posible rastrear el origen y dispersión de los linajes femeninos principales existentes actualmente en los diferentes continentes. De igual manera, el creciente número de marcadores polimórficos identificados en la región no recombinante del cromosoma Y (NRY) añadió conocimiento sobre el origen de los linajes masculinos en las poblaciones continentales, complementando los análisis obtenidos con marcadores autosómicos en la demografía de las poblaciones. De manera que los estudios dirigidos a conocer la contribución paterna y materna de la población de SBP mediante el uso de estos marcadores permitirán ampliar el conocimiento de su estructura genética y además conocer el origen de sus poblaciones ancestrales.

Seguidamente se describirán los polimorfismos genéticos estudiados en este trabajo: STRs y SNPs específicos de ChrY (Y-STRs e Y-SNPs, respectivamente), Región control del ADNmt y marcadores autosómicos informativos de ancestría (AIMs).

2.1. Polimorfismos de ADN en la historia de las poblaciones humanas

Las características generales de un polimorfismo de ADN dependen en gran medida del mecanismo de mutación subyacente a su origen. Con base en la tasa a la que tales eventos mutacionales ocurren, es posible determinar la evolución a lo largo del tiempo. Consideramos como marcadores genéticos aquellos loci que presentan características específicas en términos de variabilidad intra e interpoblacional, que definen su utilidad en diferentes campos de la investigación genética (Weber & May, 1989).

Los microsatélites, también llamados STRs (del inglés “Short Tandem Repeats”), están entre los loci más variables en el genoma. Los STRs son altamente polimórficos y predominantemente multi-alélicos. Han sido encontrados en todas las especies estudiadas y con una distribución casi uniforme, pudiendo estar tanto dentro de las regiones codificantes, como en las no codificantes (Hancock, 1999; Schlötterer & Wiehe, 1999). La tasa mutacional varía entre marcadores, siendo el promedio por locus por generación para los Y-STRs del orden de 3×10^{-3} , pudiendo algunos loci, denominados de “Rapid Mutating”, presentar tasas de mutación de 1 en 100 (50% de posibilidades de encontrar 1 alelo mutado en F2 si se analizan al menos 50 loci) (Ballantyne et al., 2014).

Otra característica de los sistemas STR es que se los considera selectivamente neutros (Schlötterer & Wiehe, 1999), lo que los hace más aptas para estudios de diversidad y comparación de poblaciones, ya que se estaría minimizando el efecto de la selección. Por otro lado, estos marcadores pueden ser detectados mediante técnicas muy sencillas, con un rápido análisis por PCR y electroforesis capilar automatizada, y la posibilidad adicional de ser amplificados en sistemas multiplex que permiten el estudio simultáneo de un gran número de loci en una sola reacción (Gymrek et al., 2016).

Los microsatélites son marcadores ampliamente usados en los estudios asociados con diversidad genética, mapeo genético, estudios de algunas enfermedades hereditarias y en casos forenses tales como identificación de individuos y pruebas de paternidad o de otro tipo de relaciones familiares. El uso combinado de haplogrupos definidos por Y-SNPs y haplotipos definidos por Y-STRs, permite determinar linajes

paternos, a partir de los cuales es posible identificar condiciones de ancestralidad común o correlación filogenética (Bailliet et al., 2011).

Los SNPs son la forma más simple de variación de secuencia originada por una sustitución de un solo nucleótido. Suelen ocurrir con una frecuencia muy baja (2.3×10^8) de modo que es muy poco frecuente observar mutaciones recurrentes en la escala de tiempo de la evolución de los humanos modernos. Por su misma naturaleza, estas mutaciones originan polimorfismos con dos estados alélicos (ancestral y derivado) y son por ello denominados polimorfismos binarios o bialélicos. Puesto que son más estables que los STRs, los SNPs son útiles para estudiar la historia evolutiva de las poblaciones humanas en escalas de tiempo más profundas.

A pesar de presentar una menor diversidad genética que los STRs, los SNPs tienen varias características que los tornan útiles en el campo de la genética forense, tales como: (i) pueden ser analizados en fragmentos muy cortos, lo cual es muy importante para mejorar el éxito de amplificación de ADN altamente degradado; (ii) tienen baja tasa de mutación, lo cual es ventajoso en pruebas de parentesco y, por último; (iii) son adecuados para la automatización y análisis con tecnologías de alto rendimiento (Sobrino et al., 2005; Winchester et al., 2009).

Los polimorfismos de inserción/delección (InDels) son variaciones en la longitud de la cadena de ADN creados por la inserción o delección de uno o más nucleótidos. Después de los SNPs, los InDels se presentan como las variaciones más abundantes en el genoma, comprendiendo alrededor del 20% de todos los polimorfismos existentes en el ADN humano (Mullaney et al., 2010; Weber et al., 2002).

Son muchos los estudios que se encuentran publicados utilizando los marcadores tipo InDel para una variedad de propósitos como, por ejemplo, estudio de la estructura genética de las poblaciones humanas, inferir proporciones de origen ancestral individual o poblacional y en identificación de especies. En 2006, Mills y cols. reportaron un primer mapa de variación de inserciones y delecciones en el genoma humano incluyendo más de 415.000 polimorfismos únicos. De acuerdo con este estudio exhaustivo, los InDels representan alrededor del 16% de los polimorfismos en el ADN humano y se

encuentran ampliamente distribuidos a lo largo de todo el genoma con una densidad media de un InDel por cada 7,2 kb (Mills et al., 2006).

Existe una clase particular de marcadores InDel que incluye inserciones-delecciones de secuencias aparentemente aleatorias y que representan alrededor del 41% de todos los InDels, con un amplio rango de variación de longitud que va desde 2 pares de bases hasta alrededor de 10kb. Curiosamente, más del 99% de todos estos InDels están por debajo de 100 pb de longitud (Mills et al., 2006). En estudios más recientes, esta clase particular de InDels se denominan pequeños InDels, en contraposición a las grandes formas de inserción y delección o expansiones de unidades repetitivas de múltiples pares de bases (Mullaney et al., 2010). Son muchos los grupos de investigación en el área de identificación humana y forense que dedicaron especial atención a la identificación de InDels (Pereira et al., 2012); siendo el estudio de Mills et al.

(2006) pionero en reportar la existencia de casi dos millones de pequeños InDels en el rango de 1 pb a cada 10.000 pb (Bhangale et al., 2005; Kidd et al., 2008; Korbelt et al., 2007).

En los últimos años también se han usado InDels para estudiar ancestría poblacional. En el caso de SBP reportamos, en un estudio previo utilizando un panel de 46 InDels, que tenía aproximadamente un 80% de mezcla africana (Martínez et al., 2017). De igual modo, se ha utilizado este tipo de marcadores en poblaciones colombianas, como en el estudio de Ossa y colaboradores (Ossa et al., 2016), donde se reporta que el pool genético de la población colombiana actual es el resultado de la mezcla de nativos americanos, europeos y africanos. Esta mezcla ocurrió de manera diferente en cada región del país, produciendo una población claramente estratificada.

A nivel de los cromosomas autosómicos, se han desarrollado diferentes conjuntos de AIMs (marcadores informativos de ancestría), más frecuentemente SNPs, para estimar diferentes proporciones de contribución intercontinental en individuos que conforman las poblaciones mezcladas, con el objetivo de controlar la ancestría heterogénea en los estudios de asociación clínica con enfermedades y en aplicaciones en el campo de genética forense.

Su caracterización se inició con distintas metodologías de baja resolución y, actualmente, ya se dispone de técnicas más refinadas que utilizan métodos de secuenciación masiva en paralelo (MPS; del inglés “Massive parallel sequencing”), entre otros (Amirisetty et al., 2012; Bauchet et al., 2007; Collins-Schramm et al., 2003; Enoch et al., 2006; Lao et al., 2006; Mao et al., 2007; Mogensen et al., 2020; Paschou et al., 2007; Phillips et al., 2013; Phillips et al., 2007; Price et al., 2007; Shriver et al., 2005; Shriver et al., 2003).

Las formas de variación genética que hemos detallado ocurren de forma ubicua por todo el genoma nuclear y algunas también en el genoma mitocondrial. Debido a los diferentes modos de herencia de los autosomas, cromosomas sexuales y del ADNmt, la localización de un marcador genético en el genoma tiene un impacto en sus características y consecuentemente en su aplicabilidad. Naturalmente, la gran mayoría de la variación genética se encuentra en los autosomas, los cuales representan aproximadamente un 93% de todo el genoma. Los marcadores genéticos en los autosomas se transmiten de padres a hijos por modo típico de herencia mendeliana, con igual contribución de la madre y del padre. La mezcla de la información genética está garantizada por la recombinación que ocurre en cada generación, lo que lleva a una ruptura progresiva del desequilibrio de ligamiento. Por lo anterior, la mayor proporción de la variación genética está presente en el genoma autosómico comparado con los heterocromosomas o las mitocondrias, siendo entonces los autosomas la principal fuente de marcadores genéticos utilizados en diversas aplicaciones. En genética de poblaciones, por ejemplo, constituyen la única información genética que representa igualmente ambas historias, la masculina y la femenina. Así que los marcadores autosómicos son los más útiles en identificación humana.

Por último, es importante mencionar el cromosoma X que presenta un modo de herencia peculiar ligada al sexo. Las mujeres tienen dos cromosomas X mostrando un modo de transmisión similar a los autosomas, mientras que los hombres poseen solo una copia del X que transmiten casi de manera inalterada a sus hijas; solo las regiones pseudo autosómicas (PAR) mantiene homología con las PAR del cromosoma Y y recombinan durante la meiosis masculina. En cuanto a sus

aplicaciones en identificación humana y forense, el cromosoma X presenta una eficiencia más elevada que los autosomas en los casos de parentesco relacionados con descendencia femenina. También resulta útil en la reconstrucción de haplotipos en los llamados casos complejos de paternidad donde el presunto padre no está disponible. En estos casos, puede servir para complementar la información de los STR autosómicos siempre que estos se revelen insuficientes o no informativos (Schaffner, 2004; Szibor et al., 2003).

La selección de marcadores en el cromosoma X, con fines de investigación poblacional y forense, debe cumplir ciertos requisitos. Concretamente, cuando se pretende seleccionar un conjunto de X-STRs es importante tener en cuenta el grado de polimorfismo que presentan, su conformidad respecto al equilibrio de Hardy-Weinberg y el desequilibrio de ligamiento entre loci. También, siempre que sea posible, deben seleccionarse marcadores neutros (sin relación conocida con enfermedades u otros fenotipos). La combinación de las propiedades de los STRs conjuntamente con las características genéticas del cromosoma X justifica el potencial de la aplicación de marcadores X-STRs en genética de poblaciones y forense. Ha sido claro el incremento de las publicaciones que utilizan polimorfismos X-STRs en estas áreas, durante los últimos años (Santos-Lopes et al., 2007; Gomes et al., 2020). Es de resaltar un estudio realizado en poblaciones colombianas en donde se evaluó la diversidad genética específica de cromosoma X, encontrándose que en la costa caribe colombiana la contribución africana fue principalmente mediada por mujeres, mientras que la contraparte europea fue principalmente impulsada por hombres y la contribución de la ancestría nativa americana no tuvo sesgos de género (Ibarra et al., 2014).

A este punto hay que concluir que los estudios genéticos a nivel del ChrY y del ADNmt son importantes para profundizar en el conocimiento acerca de la historia de las poblaciones mundiales, ya que permiten un análisis más exacto y detallado de los movimientos migratorios que originaron las poblaciones actuales. Los polimorfismos localizados en la región NRY son transmitidos en bloque de una generación a otra, siendo modificados apenas por mutaciones. Tanto los Y-STRs como los Y-SNPs han sido ampliamente utilizados para

determinar la ancestría de las poblaciones humanas. Los Y-SNPs presentan una tasa de mutación mucho menor y, debido a esto, una baja recurrencia y una elevada especificidad geográfica, permitiendo esclarecer las relaciones entre linajes masculinos separadas por muchas generaciones, añadiendo la información contenida en autosomas. Seguidamente hablaremos en detalle de los marcadores que fueron analizados en la primera etapa de este trabajo.

2.2. Marcadores uniparentales

Como se había mencionado, en los seres humanos la mayor parte del genoma se hereda de ambos padres y sufre recombinación, lo que resulta en una mezcla de la información genética en generaciones sucesivas. Sin embargo, hay dos excepciones a esta regla, la región NRY (también conocida por región específica masculina del cromosoma Y - MSY) y el ácido desoxirribonucleico mitocondrial (ADNmt). Dado que estos dos segmentos están presentes en una sola copia / tipo clon por célula, no sufren recombinación durante la transmisión. En consecuencia, a menos que se produzca un evento mutacional, la información genética se transmite de generación en generación sin ninguna alteración. El ADNmt se hereda de forma materna (se transmite de la madre a la descendencia) y el ChrY se hereda de forma paterna (se transmite de padre a hijo). Ver Figura 1. Dado que el haplotipo (grupo de alelos en el mismo cromosoma o en la misma región cromosómica) no se altera en individuos relacionados maternal / paternalmente, los marcadores uniparentales no pueden permitir la discriminación individual. Sin embargo, estas características permiten definir linajes femeninos y masculinos, respectivamente.

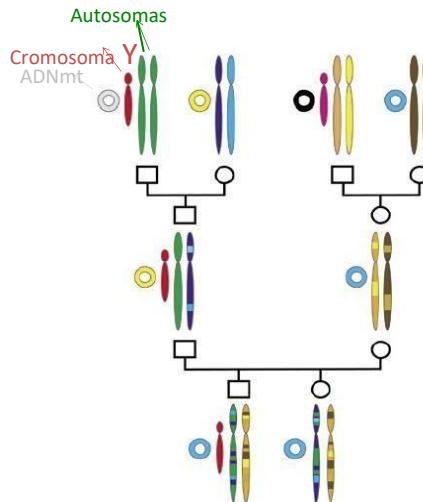


Figura 1. Patrones de herencia de los segmentos genéticos no-recombinantes (cromosoma Y y ADNmt) y recombinantes (autosomas) del genoma.

La región NRY y el ADNmt son útiles en estudios de genética de poblaciones, debido a sus patrones de transmisión específicos, que promueven una gran diferenciación y estructura geográfica entre las poblaciones humanas. Estos segmentos de ADN también son ventajosos en la genética forense, permitiendo la resolución de muchos casos difíciles. Por ejemplo, los marcadores del cromosoma Y son muy útiles para determinar perfiles masculinos en escenarios de muestras resultantes de mezcla entre material biológico de una víctima del sexo femenino e de un agresor del sexo masculino (Awise et al., 1987; Rosser et al., 2000). Por otro lado, el ADNmt es útil por permitir generar perfiles genéticos en muestras forenses degradadas, en las cuales no haya sido posible obtener resultados para marcadores autosómicos.

Seguidamente, describimos en detalle estas dos regiones del genoma y los marcadores genéticos que se analizan en este trabajo.

2.2.1. Cromosoma Y

Hoy en día, los cromosomas sexuales humanos (X e Y) son muy diferentes entre sí. Sin embargo, evolucionaron a partir de un par

de autosomas homólogos. Esta divergencia comenzó hace unos 240-320 millones de años (Ma), con la adquisición de una función determinante del sexo masculino por parte de uno de los autosomas. Después de este evento de divergencia, el cromosoma proto-Y se sometió a una represión de la recombinación que causó la expansión de la región no recombinante (hoy en día esta región encierra alrededor del 95% del cromosoma). Esto llevó a una falta de recombinación que se convirtió en una degeneración del cromosoma, tanto en contenido de genes como en tamaño, hasta aproximadamente 60 millones de pares de bases (Mb), siendo uno de los cromosomas más pequeños en los seres humanos. Este evento, conocido como el trinquete de Muller, puede explicarse por la falta de eliminación de mutaciones perjudiciales causada por la ausencia de recombinación, que conduce a la acumulación de dichas mutaciones y, en última instancia, a la degeneración del cromosoma Y (Skaletsky et al., 2003).

En la porción no recombinante del cromosoma Y se han reportado polimorfismo STR, que mutan con relativa frecuencia, y polimorfismos bialélicos como las inserciones Alu (YAP) y SNPs. Por su baja tasa de mutación, es poco probable que los marcadores bialélicos hayan surgido más de una vez en la evolución humana. Por eso, este último grupo se ha llamado

“Polimorfismos de único evento” (UEP) (Thomas et al., 1999), y por lo general permiten dividir los cromosomas Y en distintos grupos genealógicos, denominados haplogrupos. La capacidad de identificar haplogrupos de alta resolución hacen que sea un instrumento poderoso para la investigación de las relaciones entre poblaciones distantes (Jobling & Tyler-Smith, 1995).

El cromosoma Y determina el género masculino. Inicialmente, la región sin cruce con el cromosoma X se designó como región no recombinante o NRY, pero más recientemente se cambió su nombre a región específica masculina o MSY. Incluso, si el mayor porcentaje del cromosoma Y no es recombinante, los dos cromosomas sexuales Y y X sufren recombinación durante la meiosis masculina, en regiones especializadas donde la identidad de secuencia todavía se mantiene. En estas regiones del cromosoma Y, la recombinación ocurre y son heredados alelos de ambos padres, de la misma manera que para loci en

los autosomas, recibiendo la denominación de regiones pseudo autosómicas. Hay dos áreas con estas características, en ambas puntas del cromosoma Y: PAR 1, en el brazo corto, con alrededor de 2,6 Mb; y PAR 2, en el brazo largo, con alrededor de 0,32 Mb³ (Figura 2).

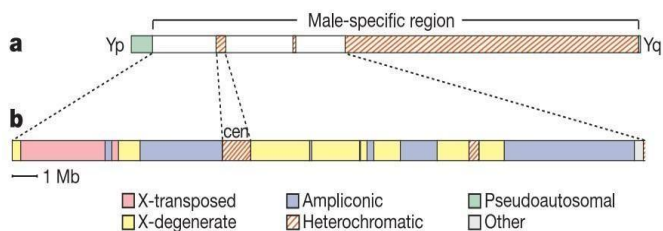


Figura 2. El Cromosoma Y

Nota. La gráfica muestra (a) estructura completa; (b) región eucromática de la región específica masculina. Adaptado de Skaletsky et al. (2003).

En genética de poblaciones, para diferenciar los linajes de cromosoma Y, los marcadores más comúnmente utilizados son los STRs y los SNPs (Gomes et al., 2010), los cuales se describirán en detalle seguidamente ya que son los utilizados en este trabajo.

2.2.2.1. Y-SNPs

Los polimorfismos de un solo nucleótido representan eventos únicos en la evolución humana debido a sus bajas tasas de mutación, como ya se había mencionado (alrededor de 10^{-8} por base por generación). Sus frecuencias alélicas varían entre las poblaciones humanas y tienen muchas aplicaciones en áreas como: genética forense, genética de poblaciones, evolución, antropología, inmunogenética, farmacogenética y farmacogenómica (Brion et al., 2005; Kwon et al., 2015; Solé-Morata et al., 2015).

Son la clase más abundantes de polimorfismos en la especie humana. El foco de las aplicaciones forenses de ADN genético todavía está en los Y-STRs, debido a su alto poder de discriminación. Sin embargo, los Y-SNPs permiten la identificación de haplogrupos (grupo de haplotipos similares que comparten un ancestro común) (Arora et al., 2015). Estos últimos son más estables que los haplotipos de Y-STRs.

Debido a un más bajo efectivo poblacional, los procesos de deriva génica tienen mayor efecto en marcadores del ChrY que en marcadores autosómicos, por lo que estos presentan una mayor capacidad para diferenciar poblaciones humanas, permitiendo reconstruir filogenias y usarlas para estudios antropológicos.

2.2.1.1.1. Construcción de la filogenia a partir de los Y-SNP

La filogeografía es una disciplina que utiliza datos genéticos para desentrañar la historia de las poblaciones, asumiendo que la distribución geográfica actual de los individuos se determina no solo por los factores biológicos y ecológicos actuales, sino también por factores históricos (Van Oven et al., 2014).

Debido a la ausencia de recombinación, los STRs y los SNPs presentes en el cromosoma Y llevan información geográfica poderosa, por lo que posibilitan inferir la representatividad regional de una muestra poblacional específica (Larmuseau et al., 2015). Esto permite inferir la historia evolutiva de una población, en este caso, la historia de los linajes paternos, desde el primer cromosoma Y (*el Adán de los linajes paternos*) hasta los pobladores actuales. La reconstrucción de la filogenia del cromosoma Y trata de inferir las relaciones de parentesco, ancestría y descendencia de las poblaciones. Estas relaciones genéticas han sido estudiadas desde mediados de 1980 mediante el estudio de STRs y SNPs en la región NRY, lo que permitió conocer mejor la secuencia de cambios que se han generado a través de la historia de cada linaje (Casanova et al., 1985; Consortium, 2002; Jobling & Tyler-Smith, 2003; Underhill et al., 1996).

En el año 2008, a partir de 600 Y-SNPs, Karafet y su grupo de investigación llevaron a cabo una revisión de la filogenia del cromosoma Y, reconfiguraron algunos linajes y redenombraron otros. Partiendo de esta filogenia de referencia e implementando nuevas tecnologías de NGS (por sus siglas en inglés “*Next Generation Sequencing*”), en el año 2014, el grupo de investigación de Mannis van Oven propuso una nueva filogenia reducida y concisa, con el fin de optimizar la capacidad de resolución y de esta manera presentar un nuevo estándar de nomenclatura jerárquico para los haplogrupos del cromosoma Y, como se muestra en la Figura 3 (Van Oven et al., 2014).

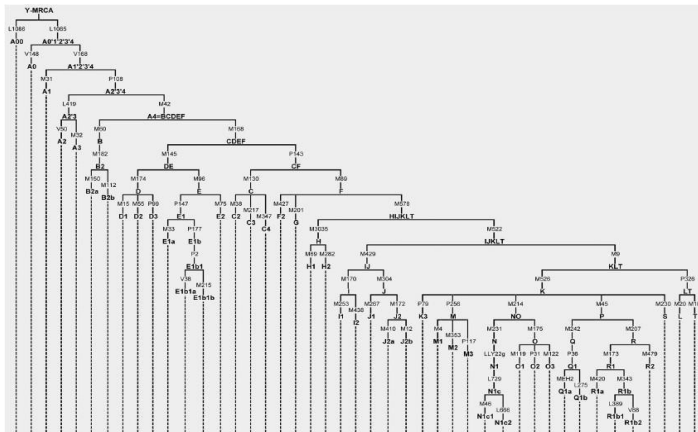


Figura 3. Versión reducida del árbol filogenético del cromosoma Y

Nota. Tomado de Van Oven et al. (2014).

Junto con las diferentes reconfiguraciones filogenéticas de los haplogrupos de cromosoma Y, también se han hecho estimaciones de la antigüedad de los clados más basales, partiendo de las mutaciones acumuladas en cada uno de ellos. Para ello, se han empleado diferentes modelos y las estimaciones han presentado rangos similares de variación temporal. La datación de los linajes es clave para poder establecer la profundidad temporal de las rutas migratorias y detectar los cambios que dieron lugar a nuevos linajes.

La comparación del árbol filogenético con los descubrimientos arqueológicos permitió inferir rutas migratorias de los primeros pobladores americanos y su dispersión a lo largo del continente, en estudios descritos en varias publicaciones (Bortolini et al., 2003; Capodiferro et al., 2021; Reich et al., 2012; Tarazona-Santos et al., 2001) (Figura 4).

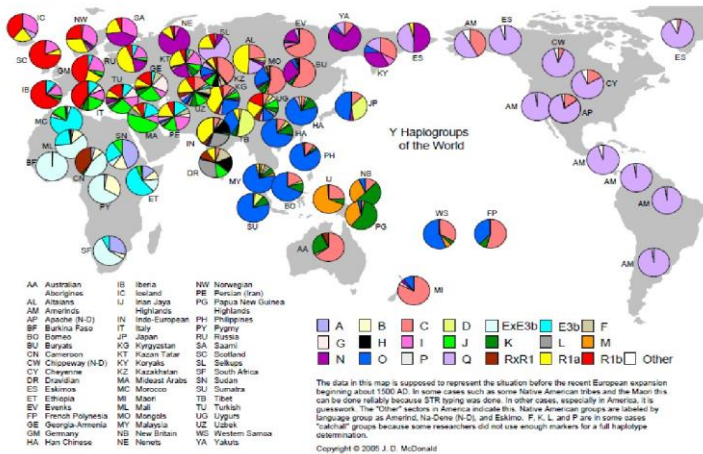


Figura 4. Distribución global de los haplogrupos de cromosoma Y

Nota. Cada círculo representa la distribución de las frecuencias de haplogrupos de cromosoma Y en diez regiones geográficas. Adaptado de Jobling and Tyler-Smith (2003).

En la Figura 3, se puede observar desde la rama más ancestral hasta los clados más derivados. El clado más ancestral está definido por el haplogrupo A. Las mutaciones características de este grupo se han encontrado únicamente en poblaciones nativas del continente africano, como los bantús, etíopes y sudaneses, y con altas frecuencias en poblaciones nilóticas y khoisan (ver Figura 4).

El clado B, al igual que el A, también se ha encontrado exclusivamente en poblaciones del África subsahariana y los haplogrupos dentro de este clado son frecuentes entre los pigmeos y los khoisan, definidos por la mutación M112, y también en bantú, definidos por la mutación M150 (Cruciani et al., 2002; Hammer et al., 2001; Jobling & Tyler-Smith, 2003; Larmuseau et al., 2015; Poznik et al., 2013; Scozzari et al., 2014; Semino et al., 2002; Underhill et al., 2001; Wood et al., 2005).

Con respecto al haplogrupo E, es el más diverso de todos los clados principales del cromosoma Y, con las frecuencias más altas en África, pero también está presente en Oriente Medio, Este y Sur de Europa, aunque con una menor frecuencia, y con una muy baja

frecuencia en poblaciones del Centro y Sur de Asia Central y Asia Meridional (Cruciani et al., 2002; Hammer & Zegura, 1996; Karafet et al., 2008; Sims et al., 2007). Teniendo en cuenta lo anterior, y como se observa en la Figura 4, el árbol filogenético en África muestra que los haplogrupos más extendidos, pertenecen a los clados principales A, B y E.

Con respecto al clado C se le sugiere un origen asiático que se produjo después que las primeras poblaciones migraron fuera de África (Figura 4). Dentro de los subhaplogrupos del haplogrupo C se encuentran C4, C5 y C6, exclusivos de poblaciones de Australia, del Sur y Centro de Asia y de Nueva Guinea, respectivamente (Karafet et al., 2008). El subhaplogrupo C3b y C3* también se encontró en poblaciones nativas americanas (Geppert et al., 2011; Mezzavilla et al., 2015; Zegura et al., 2004).

El clado D solo se ha encontrado dentro de Asia, con altas frecuencias en las poblaciones del Tíbet y Japón (Hammer et al., 2006; Karafet et al., 2001; Su et al., 2000; Thangaraj et al., 2003).

El clado F presenta el estado derivado para el marcador M89. Este alelo derivado es común para todas las demás ramas más derivadas del árbol. El haplogrupo F es cercano al haplogrupo C y presenta una datación que lo asocia con una distribución típicamente fuera de África. La colonización de Eurasia podría haber consistido en un solo proceso migratorio datado de hace 70.000 años, según evidencias genéticas y arqueológicas. Conforme a ello y al número de mutaciones, se estima que F se habría originado hace unos 48.000 años. El paragrupo F(xK) es decir los clados que presentan el estado derivado para M89 pero son ancestrales para M9, se encuentran distribuidos en poblaciones del Este de Asia (Sengupta et al., 2006). Dentro de este grupo están los clados G, H I y J.

El clado G está presente en poblaciones de Medio Oriente y del Mediterráneo (Behar et al., 2004; Cinnioğlu et al., 2004; Jobling & Tyler-Smith, 2003; Nasidze et al., 2005; Regueiro et al., 2006). Una única mutación define la rama más ancestral del clado H; su distribución geográfica está restringida a poblaciones de la India (Jobling & Tyler-Smith, 2003; Karafet et al., 2005; Sengupta et al., 2006) (ver Figura 2).

Los subhaplogrupos del clado I están divididos en dos subclados mayoritarios: I1 e I2 (Hammer et al., 2001; Jobling & Tyler-Smith, 2003; Karafet et al., 2008; Rootsi et al., 2004). El subclado I1 se ha encontrado mayoritariamente en poblaciones del Norte de Europa mientras que I2 es más frecuente en el Este de Europa (Rootsi et al., 2004). Al igual que el clado I, los haplogrupos dentro del clado J se dividen en dos subgrupos mayoritarios: J1 y J2, los cuales presentan una amplia distribución geográfica que va desde Medio Oriente, Norte de África, Europa, Asia Central, Pakistán e India (Behar et al., 2004; Cinnioglu et al., 2004; Hammer et al., 2001; Semino et al., 2002; Sengupta et al., 2006; Underhill et al., 2001).

El clado K está definido por el estado derivado del marcador M9; contiene a los haplogrupos L, M, N, O, P, Q, R, S y T (Karafet et al., 2008). Además, presenta 4 polimorfismos que definen 4 paragrupos K que se han encontrado en poblaciones de Oceanía, Indonesia y

Australia. El clado L tiene una amplia distribución geográfica: se ha encontrado en Medio Oriente, Asia Central, el norte de África y en Europa, a lo largo de la costa mediterránea (Behar et al., 2004; Cinnioglu et al., 2004; Cruciani et al., 2002; Jobling & Tyler-Smith, 2003; Karafet et al., 2005; Sengupta et al., 2006; Underhill et al., 2001).

El clado M tiene una distribución geográfica restringida a las poblaciones de Oceanía y del Este de Indonesia (Capelli et al., 2001; Hammer et al., 2001; Kayser et al., 2006; Su et al., 2000). Los subhaplogrupos del clado N se encuentran distribuidos en las poblaciones del norte de Eurasia (Karafet et al., 2001; Rootsi et al., 2004). El clado O está presente en las poblaciones del Asia Central y del Este y de Oceanía (Karafet et al., 2001; Su et al., 2000; Underhill et al., 2001). Uno de sus subhaplogrupos (O3*) se ha encontrado también en bajas frecuencias en China, Taiwán e Indonesia (Karafet et al., 2008).

El clado S fue redefinido; anteriormente estaba asignado al subhaplogrupo K, definido por la mutación M230, y predominantemente se encuentra en Oceanía e Indonesia (Kayser et al., 2006; Scheinfeldt et al., 2010). El clado T, al igual que el S, fue redefinido de un subhaplogrupo K, el K2. Es característico de poblaciones del Medio Oriente, África y Europa (King et al., 2007; Underhill et al., 2001).

El clado P contiene los dos clados más frecuentes en poblaciones americanas: el R y el clado Q (Karafet et al., 2008). El clado Q está definido por la mutación M242, presenta una amplia distribución en el norte de Eurasia y en algunos grupos siberianos (Karafet et al., 2002).

Con una frecuencia más baja, este haplogrupo se encuentra también en Europa, Este de Asia y Medio Oriente. El subhaplogrupo Q-M3 es propio de poblaciones nativas americanas (Underhill et al., 1996; Zegura et al., 2004). Este clado presenta también un paraguero (Q*) que ha sido encontrado en poblaciones de India, Pakistán y algunas poblaciones del continente americano (Grugni et al., 2012; Karafet et al., 2008; Regueiro et al., 2006; Schurr et al., 2012). El clado R tiene una amplia distribución pero se considera como el clado propio del continente europeo (Cortellini et al., 2013; Karafet et al., 2008; Myres et al., 2011; Sims et al., 2007).

Son muchos los estudios en distintas poblaciones modernas que han caracterizados todos los anteriores haplogrupos de cromosoma Y, reportándose nuevas mutaciones que dan origen a nuevos subclados. En este estudio se utilizaron descripciones de haplogrupos de cromosoma Y encontrados en poblaciones de África subsahariana y de otras regiones, teniendo en cuenta los registros de origen de los africanos que llegaron a las Américas durante la colonización y que se describen a continuación.

2.2.1.1.2. Y-SNP Multiplexes

La genotipificación de SNPs es un procedimiento que cada vez se hace más sencillo y económico. Muchas técnicas permiten su tipificación, como la secuenciación, MALDI-TOF MS o Microarrays, y la PCR en tiempo real asociada a la metodología de sondas FRET o la metodología de sondas TaqMan®. El análisis de Y-SNPs en material altamente degradado proporciona mejores resultados que el análisis de Y-STRs, ya que casi todos los SNPs se tipifican a partir de amplicones menores a 150 bp (Romanini et al., 2012). El Genotipo Multiplex (múltiples sitios de SNP tipificados simultáneamente), entre los ensayos de genotipado de SNPs disponibles, ayuda a minimizar el número de veces que se realiza el procedimiento de genotipado en un estudio

determinado (Aumann et al., 2003). La amplificación multiplex por PCR seguida de la tecnología de extensión de base única (SBE) con el kit SNaPshot™ (Applied Biosystems) es uno de los métodos más populares en los laboratorios de Identificación humana y forenses.

Las reacciones de extensión de base en multiplex por SNaPshot utilizan ddNTPs fluorescentes. En esta reacción, los cebadores se diseñan con el extremo 3' en el sitio inmediatamente corriente arriba de cada SNP y se extienden por un solo ddNTP. Como cada ddNTP está marcado con un fluoróforo diferente, el alelo puede asignarse por la fluorescencia presentada después del análisis en una secuencia de ADN. Se añade una secuencia aleatoria de longitud variable, no homóloga con el ser humano, al extremo 5' de los cebadores utilizados, para obtener diferentes tamaños para cada marcador, lo que permite la identificación de varios SNPs en una sola reacción (Caratti et al., 2009).

2.2.2.2. Y-STRs. Buscando variabilidad dentro de los linajes

Las repeticiones cortas en tándem o STRs son loci multialélicos con repeticiones de 16 pb que poseen una alta tasa de mutación (del orden de 10^{-3} a 10^{-4} por base por generación) (Kayser et al., 2004; Solé-Morata et al., 2015). Su poder informativo puede aumentarse cuando se combinan diferentes Y-STRs para obtener los haplotipos. Hoy en día, el mecanismo más aceptado de mutación en los STR es *strand-slippage replication*, cadena de replicación de deslizamiento, a través del modelo de mutación paso a paso (SMM). De acuerdo con este modelo, la longitud de estos polimorfismos cambia a través de la adición o eliminación de una unidad de repetición (un paso). El deslizamiento ocurre durante la replicación del ADN, causado por la falta de pareo (por una o más unidades de repetición) entre la cadena naciente y la plantilla. Luego, el fragmento repetido de ADN se obliga a "salir" en el sitio de desajuste. El número de repeticiones del STR cambiará si la síntesis de ADN continúa en la molécula (Fan & Chu, 2007) (Ver Figura 5).

Las tasas de mutación en STRs tienden a aumentar a medida que aumenta el tamaño del alelo y con un mayor número de repeticiones ininterrumpidas, por lo que el ritmo de estas mutaciones no es el mismo en todos los loci y ni siquiera en el mismo locus (Fan & Chu, 2007). Dado que los STRs presentes en la región MSY se transmiten de padres

a hijos, son muy útiles en pruebas de paternidad, estudios de estructura genética, identificación de víctimas masculinas de desastres, identificación de linajes masculinos con fines de antropología e identificación de perpetradores masculinos en casos criminales de agresión sexual. De igual forma, la información obtenida con los Y-STRs solo nos proporciona datos sobre el linaje paterno y no sobre el perfil individual, no permitiendo diferenciar a aquellos individuos que son parientes paternos y tienen haplotipos idénticos (Hadi, 2016).

Existen bases de datos con información masiva de haplotipos Y-STR, entre ellas la base de datos de referencia de haplotipos del cromosoma Y (YHRD; www.yhrd.org), que ayudan a comparar los resultados de diferentes estudios y diferentes poblaciones. Por lo general, hay muchas opciones de comparación en estas bases de datos que cubren los diferentes estuches comerciales disponibles y que van desde el haplotipo mínimo (haplotipos de 8 Y-STRs) hasta el haplotipo máximo (haplotipos de 27 Y-STRs) (Romanini et al., 2012).

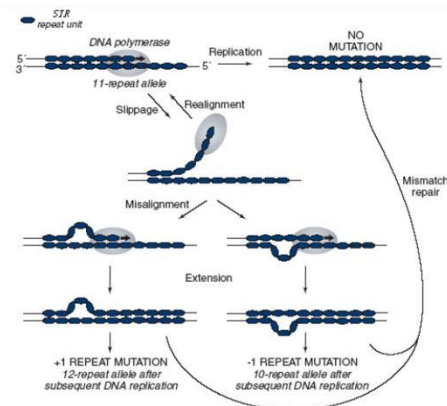


Figura 5. Ilustración esquemática de la replicación STR-slippage.

Nota. Fuente: Fan and Chu (2007).

El uso combinado de haplogrupos definidos por Y-SNPs y haplotipos definidos por YSTRs permite determinar linajes paternos, a partir de los cuales es posible identificar condiciones de ancestralidad común o correlación filogenética de los individuos de una población. Empleando los microsatélites es posible observar diferenciación entre

las poblaciones dentro de un mismo haplogrupo y encontrar posible efecto fundador en los linajes paternos (Altuna et al., 2006; Bailliet et al., 2011; Bolnick et al., 2006). Adicionalmente, mediante métodos de máxima parsimonia, es posible calcular el tiempo de divergencia de los mismos, como ha sido publicado por varios autores para poblaciones americanas, datando linajes con edades que van desde 13.500 – 58.700 años (Bolnick et al., 2006; Bortolini et al., 2003; Ruiz-Linares et al., 1999).

En este trabajo se determinaron los anteriores polimorfismos en la muestra poblacional de SBP.

2.2.2. La mitocondria

2.2.2.1. Estructura y función de la mitocondria

Las mitocondrias son organelas subcelulares de los organismos eucariotas, que varían en tamaño. En los hepatocitos y fibroblastos, poseen un tamaño aproximado de 3 a 4 micras de longitud y 1 micra de diámetro. Su principal característica morfológica es la gran cantidad de membranas internas que la componen, separadas por el espacio intermembranal. La membrana interna forma numerosos pliegues llamados crestas, las cuales se extienden hasta la matriz de la organela. La matriz contiene el genoma mitocondrial y las enzimas responsables del metabolismo oxidativo (Cooper et al., 2007; Roberts et al., 2002).

La mitocondria tiene una función muy importante en la generación de la energía metabólica; es la responsable de producir la mayoría de la energía derivada de la degradación de carbohidratos y ácidos grasos, los cuales son convertidos a ATP por el proceso de fosforilación oxidativa. Un grupo de complejos enzimáticos, designados como complejo I al V, llevan a cabo este proceso dentro de la mitocondria. Adicionalmente a la producción de energía, la mitocondria participa en la regulación de la apoptosis celular y también juega un papel importante en la producción del colesterol y el grupo hemo (componente de la hemoglobina) (Galloway et al., 2012).

2.2.2.2. Origen de la mitocondria como organelo subcelular

Hace dos millones de años, la tensión de oxígeno en la atmósfera terrestre se incrementó rápidamente (pasó de 1 a 15% en 200 millones de años). Varios autores argumentan que el origen de la mitocondria como organelo de los organismos primitivos eucariotas pudo deberse a este cambio en las condiciones del medio ambiente (Kurland & Andersson, 2000). Debido a que el genoma mitocondrial muestra grandes similitudes con el genoma de las células procariotas, se ha sugerido que un precursor de las células eucariotas (protoeucarionte) endocitó algún tipo de célula procariota (simbionte). Este proceso confirió ventajas evolutivas a la nueva célula. Esta teoría es apoyada por algunas características del genoma mitocondrial que se asemejan a las de los genomas bacterianos, tales como su pequeño tamaño, circularidad, la ausencia de intrones, un alto porcentaje de ADN codificante, entre otras (Strachan & Read, 2018). Sin embargo, si se compara el genoma mitocondrial de hoy en día con los genomas bacterianos, se puede observar que el genoma mitocondrial es mucho más pequeño que el bacteriano. Mientras que el mitocondrial humano consta de 16Kb y 37 genes, el genoma de eubacterias contiene de una a varias Mb de ADN con cientos a miles de genes. Las teorías del origen del genoma mitocondrial sugieren que hubo transferencia de genes que originalmente hacían parte de la célula procariota endocitada hacia el núcleo de la célula hospedera. Actualmente se han propuesto dos hipótesis para explicar el origen de la mitocondria como organelo de las células eucarióticas: la hipótesis de la endosimbiosis serial que supone que las células eucarióticas evolucionaron a partir de una serie de endocitosis, lo cual está sustentado en la presencia de regiones del genoma eucariótico similares a las de arqueobacterias y otras similares a las de eubacterias. Por otro lado, está la hipótesis del hidrógeno o de la sintropía, la cual sugiere que los eucariotas aparecieron de una arqueobacteria estrictamente anaeróbica autotrófica la cual requería hidrógeno, con una eubacteria productora de hidrógeno como la apteobacteria. La fuerza que dirigió esta asociación fue el metabolismo simbiótico (la sintropía).

(2 ARNr, 14 ARNt y 12 polipéptidos) se encuentran en una de las cadenas (cadena H o pesada), mientras que los 9 genes restantes (1 polipéptido y 8 ARNt) están en la cadena complementaria (cadena L o ligera). La única zona del ADNmt que no codifica ningún gen es la región del bucle de desplazamiento (bucle-D), localizada alrededor del origen de replicación de la cadena H. Esta región contiene también los promotores de la transcripción y los elementos reguladores de la expresión génica. Otra de las peculiaridades de la organización genética del ADNmt es que los genes de los ARNt se distribuyen entre los genes de los ARNr y los codificantes de proteínas. Esta disposición tiene consecuencias muy importantes para el procesamiento del ARN. Para la replicación del ADNmt hacen falta dos orígenes diferentes, uno para cada cadena (OH y OL). Ambos orígenes de replicación están muy separados, haciendo que el proceso sea unidireccional y asimétrico. La síntesis del ADN se inicia en OH y es realizada por una polimerasa específica de la mitocondria, la DNAPol, que alarga un ARN iniciador fruto del procesamiento de un transcrito primario que se sintetiza a partir del promotor L. La replicación continúa de modo unidireccional hasta alcanzar OL, momento en el cual comienza la síntesis de la segunda cadena del ADN, alargando también un pequeño iniciador de ARN (Yan et al., 2019).

Además de la región codificante, se encuentra la región no codificante del genoma, la región control, que es la responsable de la regulación del ADNmt. Por su aspecto en microscopía electrónica durante la replicación del ADNmt, ha sido también llamada Asa de desplazamiento

(D-Loop). El D-Loop es la región del genoma donde tiene lugar la separación inicial o desplazamiento de las dos hebras de ADN durante la replicación. La nomenclatura empleada para referirse a las posiciones en el genoma fue escogida arbitrariamente. Así, la posición 1 se encuentra en una región intermedia de la región control y se continúa en sentido 5' – 3' hasta la posición 16.569 (Alvarez et al., 2007; Goodwin et al., 2011).

La región control se divide en regiones según su grado de diversidad. De esta manera, se determinan la región hipervariable I (HVI), que abarca desde la base 16.024 hasta la 16.365, la región

hipervariable II (HVII), entre las bases 73-340, y la región hipervariable 3 (HVIII) desde la posición 438 hasta la posición 574. Como comentan Holland y Parsons (1999), estos límites no son exactamente rígidos y pueden variar entre laboratorios (Figura 7).

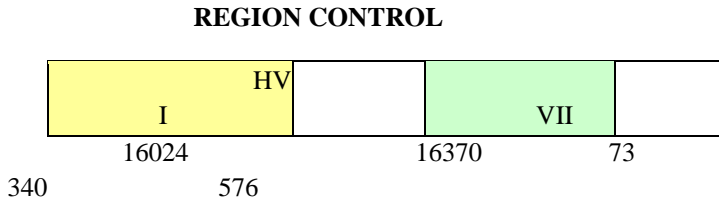


Figura 7. Detalle de la región control o D-Loop.

Nota. Adaptado de Holland and Parsons (1999).

Tradicionalmente se ha considerado que el ADN mitocondrial humano se hereda sólo por vía materna. Según esta concepción, cuando un espermatozoide fecunda un óvulo, el núcleo penetra y su cola, junto con sus mitocondrias, son destruidas en el óvulo materno. Por lo tanto, en el desarrollo del cigoto sólo intervendrían las mitocondrias contenidas en el óvulo. Sin embargo, se ha demostrado que las mitocondrias del espermatozoide pueden ingresar al óvulo. Según algunos autores el ADN mitocondrial del padre puede perdurar en algunos tejidos, como los músculos. Según otros, no llega a heredarse al ser marcado por ubiquitinación y degradado.

Entre las causas biológicas que se han mencionado para explicar este tipo de herencia, se encuentran: (1) Escasa cantidad de mitocondrias en los espermatozoides, comparadas con las miles de mitocondrias del óvulo; (2) la cabeza del espermatozoide es la que penetra el óvulo, las mitocondrias están localizadas en el cuello; (3) un posible mecanismo de reconocimiento que elimina las mitocondrias paternas que atraviesan la membrana del óvulo. Aunque el mecanismo de eliminación del ADNmt no ha sido completamente dilucidado, ni tampoco es absoluto, pues se conocen persistencias del ADNmt paterno extremadamente bajos, es claro que, desde el punto de vista práctico para las pruebas forenses, el mtDNA se comporta como heredado por línea materna.

2.2.2.4. Tasa de mutación y polimorfismos

La región D-loop del genoma mitocondrial es altamente polimórfica y posee una alta frecuencia de mutaciones de 5 a 10 veces mayor que el ADN nuclear, lo cual es útil pues exhibe una gran cantidad de variabilidad en regiones cortas y fácilmente secuenciadas, como las regiones hipervariables I y II (Salas et al., 2007).

Basado en comparaciones entre las secuencias de humanos y de primates, se han estimado tasas de mutación entre 1,5 y 4,95 por sitio, por millón de generaciones (Endicott & Ho, 2008). Basados en divergencias dentro de las poblaciones humanas, Torroni et al. (1998) indican una tasa de 5,4 por sitio, por millón de generaciones. Esta característica de la región control o D-Loop permite diferenciar, por ejemplo, en casos de criminalística, si dos muestras biológicas provienen del mismo origen - tienen la misma secuencia (la excepción se da si se presenta el fenómeno conocido como heteroplasmia, que se explicará más adelante). Así, mismo, en casos de identificación humana es posible descartar con exactitud familiares si éstos no comparten la misma secuencia mitocondrial. Si los familiares y los restos humanos a identificar poseen la misma secuencia, existe una alta probabilidad de que los restos provengan del mismo linaje materno de los supuestos familiares (esta probabilidad depende de la frecuencia de la secuencia en la población).

Los polimorfismos presentes en las regiones hipervariables del genoma mitocondrial se dan por transiciones o transversiones que se presentan a lo largo de la secuencia. Si bien, se pueden presentar en cualquier posición, se han detectado posiciones que mutan más frecuentemente que otras, estas son consideradas puntos calientes de mutación o “hot spots” (Endicott & Ho, 2008; Salas et al., 2007). Salas y colaboradores (2007) mencionan que las 10 substituciones más frecuentes son transiciones presentes en los sitios 16093, 16129, 16189, 16192, 16311, 16362, 146, 150, 152 y 195.

2.2.2.5. Número de copias y herencia materna

El ADNmt tiene un alto número de copias por célula, a diferencia del ADN nuclear que se encuentra en dos copias por célula en organismos diploides. El número de moléculas de ADNmt por célula

varía entre unos pocos cientos en los espermatozoides a más de 100.000 copias en el oocito maduro. En células somáticas esta cantidad varía de 200 a 1.700 copias de ADNmt dependiendo del tipo de tejido (Robin & Wong, 1988).

Otra característica que hace útil el estudio del ADNmt en genética forense es su tipo de herencia por vía exclusivamente materna. Esto quiere decir que todas las personas del mismo linaje materno tendrán, salvo pequeñas mutaciones, el mismo linaje de ADNmt. Por otro lado, el uso del ADNmt, desde los años ochenta hasta la actualidad, se ha convertido en una herramienta de mucha utilidad, ya que no se recombina y se transmite solo por las mujeres.

Tiene características únicas que lo hacen muy apropiado para estudios microevolutivos. La tasa de mutación del ADN mitocondrial es más alta comparada con la del genoma nuclear, lo que se debe en parte a la exposición del ADNmt a radicales de oxígeno que se producen en la fosforilación oxidativa (Goodwin et al., 2011).

El análisis de la secuencia genómica mitocondrial se usa para estudiar las relaciones filogenéticas, no sólo en humanos sino también en muchos otros organismos (Varela T. A., 2009).

2.2.2.6. Heteroplasmia

El ADNmt es esencialmente no recombinante, sin embargo, la tasa de error de la polimerasa y la aparente carencia de mecanismos de reparación colaboran con la alta tasa mutacional implicando una alta variabilidad entre las poblaciones humanas. Esta tasa mutacional permite que en un mismo individuo se encuentre más de un haplotipo mitocondrial, fenómeno conocido como heteroplasmia. La heteroplasmia se puede dar a nivel de la mitocondria, de las células o de los individuos (Danielson et al., 2007; Paneto et al., 2007).

Se conocen dos tipos: la de punto o secuencia, que se da cuando en una misma posición se encuentra más de una base, y la heteroplasmia de longitud, cuando existen inserciones o deleciones en la secuencia. Este último tipo de heteroplasmia se presenta con mucha frecuencia en el fragmento homopolimérico de poli C en la región HVI (Ver Figura 8) (Tully et al., 2001).

La heteroplasmia es un fenómeno que se debe tener en consideración a la hora de la interpretación de resultados, donde se trata de determinar si las secuencias de ADNmt que se comparan provienen del mismo linaje. En el caso de que una de las muestras que se desea comparar presente la heteroplasmia, aunque no existan otras diferencias en la secuencia, es importante preguntarse si esta diferencia se debe a que las dos muestras provienen de linajes diferentes. Al respecto, Paneto et al. (2007) recopilan información de la literatura e indican: *“La ocurrencia de la heteroplasmia puede mejorar el valor de los análisis de ADNmt para propósitos forenses. Si un sospechoso y una muestra coinciden en todas las posiciones y comparten una heteroplasmia en la misma posición, la evidencia del ADNmt se refuerza. Sin embargo, si ambas muestras son homoplásmicas y difieren en un sitio, el resultado es inconcluyente y se deben realizar nuevas investigaciones”*, aunque también hacen notar que esta interpretación ha sido cuestionada recientemente debido a que no es improbable, precisamente por el fenómeno de la heteroplasmia que un individuo pueda contener una diferencia de secuencia entre dos tipos de tejidos, por ejemplo.

La frecuencia de heteroplasmas no es muy alta, ocurre en porcentajes entre el 2 y el 8% de la población, según la descripción de Holland y Parson (1999). Entre los puntos calientes de mutación, se mencionó la 16.189 en la región HVI del D-loop, donde una C es reemplazada por una T. Cuando en esta posición se presenta la citosina, es muy difícil o casi imposible obtener la secuencia a partir de este punto. Por el contrario, cuando la timina está presente es posible obtener la secuencia completa del fragmento. Entre los puntos calientes se encuentran las posiciones 16.093 y 16.129 de la región HVI, así como el tracto de poliC de la región HVII que normalmente presenta heteroplasmia de longitud (Holland & Parsons, 1999).

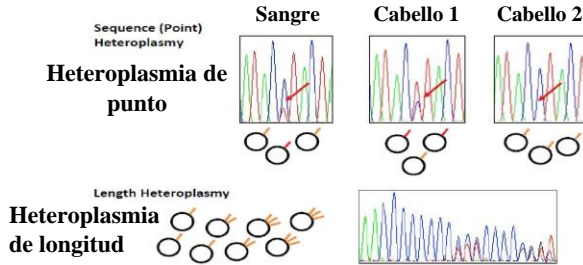


Figura 8. Clases de Heteroplasmia.

Nota. Representación esquemática de las heteroplasmas.
Tomado de Parson et al. (2014).

Se ha observado que la heteroplasma en individuos sanos es un evento cada vez más recurrente, lo cual se puede deber a la sensibilidad de los métodos de detección que se emplean hoy en día. Se han detectado diferentes niveles de heteroplasma entre diferentes tejidos y dentro del mismo tejido del mismo individuo. Así mismo, se ha encontrado evidencia de incremento de la tasa de heteroplasma con la edad (Calloway et al., 2000), razón por la cual este fenómeno es ampliamente discutido a nivel forense pues debe ser interpretado cuidadosamente para no llegar a falsas conclusiones (Bini & Pappalardo, 2005). Actualmente es aceptado para estudios de poblaciones humanas y de forense el análisis de genoma completo o secuenciación completa de la región control con el fin de observar dicho fenómeno (Andréasson et al., 2007; Nilsson et al., 2008).

Por convención, una vez realizada la secuenciación de las muestras de interés, estas son alineadas y comparadas con una secuencia de referencia de ADNmt descrita y se reporta el conjunto de cambios entre las muestras estudiadas y la secuencia de referencia (Budowle et al., 2008). La secuencia de referencia ha sido inicialmente descrita por Anderson et al. (1981) y es conocida por Secuencia de Referencia de Cambridge (CRS, del inglés “Cambridge Reference

Sequence”). El mismo genoma mitocondrial ha sido posteriormente reanalizado por Andrews y colaboradores (1999) que detectan y corrigen 11 errores en la CRS. La nueva secuencia se denomina Secuencia de Referencia de Cambridge revisada (rCRS, del

inglés “revised Cambridge Reference Sequence”) y es la secuencia que se utiliza actualmente como referencia en genética poblacional y forense.

2.2.2.7. Definición de haplogrupos de ADNmt

El análisis molecular de las secuencias de ADN mitocondrial ha mostrado que la variación se ha acumulado secuencialmente a partir de algunos linajes fundadores, durante el proceso de colonización humana de diferentes regiones geográficas (Torroni et al., 1998). Por lo tanto, es posible establecer grupos de secuencias que se pueden asociar geográfica o étnicamente. A partir del análisis filogenético de los linajes mitocondriales se han identificado haplogrupos que son específicos para los africanos, europeos, asiáticos o amerindios (AlvesSilva et al., 2000; Parson et al., 1998).

Los primeros estudios para el establecimiento de haplogrupos se llevaron a cabo con RFLPs de alta resolución y mostraron la existencia de cuatro haplogrupos específicos europeos, denominados H, I, J, K. (Torroni et al., 1998). En 1998, Torroni y colaboradores muestran que existe una correlación entre los haplogrupos definidos por RFLPs y las secuencias para HVI y HVII. Empleando este tipo de análisis conjunto, el autor confirmó la presencia de los cuatro haplogrupos inicialmente definidos para Europa y se adicionaron los haplogrupos T, U, V, W, X. Los haplogrupos T, V y W resultaron asociados específicamente a un origen europeo, mientras que el U se comparte entre europeos y africanos; así mismo, el X se comparte entre europeos y amerindios del norte. También se han definido el M como asiático y los L1 y L2 como africanos (Torroni et al., 1998).

En 1993, Torroni y colaboradores analizan 326 individuos provenientes de 17 poblaciones nativo-americanas, por RFLPs y secuencia de la región control del ADN mitocondrial. En este estudio, muestran que existe una correlación directa entre los cuatro haplogrupos definidos por RFLPs y la variación de secuencia de la región control. Se definen así cuatro haplogrupos fundadores americanos denominados A, B, C, D (Torroni et al., 1993). En 1998, Brown y colaboradores confirman la presencia del haplogrupo X en poblaciones nativas americanas; este se ha observado también en poblaciones europeas

(Brown et al., 1998). La identificación inicial por RFLPs y la distribución geográfica de los haplogrupos a nivel mundial está establecida en la base de datos MITOMAP (<http://www.MITOMAP.org>; Brandon et al., 2005).

Con el fin de conocer la estructura de las poblaciones respecto al ADN mitocondrial, se construyeron bases de datos con las secuencias de cientos de muestras clasificadas por grupos étnicos o geográficos. Esto permite determinar, entre otros, las frecuencias de los haplotipos y haplogrupos en las diferentes poblaciones.

Estas bases de datos son empleadas para la resolución de casos de identificación forense, estableciendo que tan frecuente es en la base de datos el haplotipo encontrado en el caso en cuestión (esto es, contando el número de veces que un haplotipo en particular es observado en la base de datos). Amplias bases de datos, representativas de las poblaciones, son críticas para estimar las frecuencias esperadas; entre mayor sea el número de individuos no relacionados en la base de datos, se pueden hacer mejores estimaciones estadísticas. Goodwin et al. (2011) indican que cuando se emplean bases de datos de bajo número de muestras (por ejemplo, de 100) existe la posibilidad de que los haplotipos menos frecuentes en la población no se vean representados.

A nivel internacional, existen varias bases de datos como MITOMAP, en la cual se encuentra la secuencia completa del ADNmt humano, los loci determinados hasta el momento, referencias bibliográficas y polimorfismos encontrados, entre otros aspectos.

En la base de datos de poblaciones de ADN mitocondrial del FBI (*mtDNA Population Database*) se almacenan, con propósitos forenses, diferencias en las secuencias de la región control con respecto a la secuencia de referencia de Cambridge para las diversas poblaciones afroamericanas (Anderson et al., 1981).

Se encuentra además la *mtDB Human Mitochondrial Genome Data Base*, soportada por la Universidad de Upsala (Suecia), en la cual se reportan las secuencias completas de 1865 individuos y la secuencia de la región codificante en 839 individuos, publicada hasta 29 de julio del 2009. Así mismo, se encuentran disponibles secuencias de ADNmt en la base de datos EMPOP (“EDNAP forensic mtDNA population database”; creada por Walther Parson y mantenida por el “Institute of

Legal Medicine, Innsbruck Medical University”, Austria; disponible en www.empop.org). En la EMPOP, al igual que en la base de datos de ADN mitocondrial del FBI, es posible buscar un determinado haplotipo con el fin de encontrar coincidencias con secuencias de la región control almacenadas en la base de datos. De igual forma, se pueden complementar las búsquedas en HaploGrep (Weissensteiner et al., 2016). En este trabajo se consulta el origen de los haplogrupos y mutaciones en EMPOP (Parson et al., 2004).

En conclusión, estos marcadores moleculares, junto con los hallazgos arqueológicos permiten elaborar teorías sobre cómo llegaron los primeros pobladores al continente americano, cómo se dispersaron y cómo su estructura y conformación genética fue cambiando después de la llegada de los conquistadores europeos (Greenberg, 1987; Greenberg et al., 1986; Reich et al., 2012; Wang et al., 2008; Zegura et al., 2004). Cuando se estudia cuáles fueron las posibles rutas de poblamiento y cuáles fueron los ancestros comunes más recientes que ocuparon por vez primera el continente americano, por ejemplo, es de suma importancia complementar el conocimiento investigando la representatividad de dichos marcadores de linaje en las poblaciones actuales y cómo esa configuración cambió después de la llegada de los conquistadores del Viejo Mundo y la posterior llegada de los esclavos africanos.

En el caso de América Latina y Colombia, gran parte de los linajes extra continentales pertenecen al clado R, evidencia de la llegada de los europeos desde 1492 y, posteriormente, la fuerte migración europea entre el siglo 19 y comienzos del siglo 20 (Bodner et al., 2012; CasasVargas et al., 2017; Simão et al., 2018; Simão et al., 2019).

2.3. Marcadores de ancestría en cromosomas autosómicos

Existe una variedad de polimorfismos de ADN que pueden utilizarse como marcadores informativos de ancestría (AIMs), como son los InDels, los STRs y los SNPs (Frudakis et al., 2003; Lowe et al., 2001; Rosenberg et al., 2003). Los STRs presentan limitaciones ya que no exhiben contrastes lo suficientemente grandes en las frecuencias alélicas entre poblaciones, principalmente por su inestabilidad mutacional (Jorde et al., 2000).

Los AIMS más comunes son los SNPs que presentan marcadas diferencias en la frecuencia de los alelos entre poblaciones de diferentes regiones geográficas. Los marcadores autosómicos pueden proporcionar información adicional sobre la mezcla de los individuos en una población debido a su herencia biparental y recombinación. También son útiles para determinar la ancestría biogeográfica probable o la población de origen de un individuo. Esto es particularmente útil en genética forense, cuando el origen de un individuo no es conocido o no se puede declarar su ancestría (Phillips et al., 2007). Debido a las continuas migraciones humanas, los alelos AIMS se comparten entre todos los grupos humanos. Por lo tanto, no es la presencia / ausencia absoluta de un alelo, sino su frecuencia en la población lo que generalmente se analiza cuando se infiere la ancestría (Phillips et al., 2007).

Además, la escogencia de los SNPs autosómicos como los mejores marcadores de ancestría se debe, en parte, a su estabilidad, densidad de distribución y rango completo de patrones de frecuencia de alelos entre las poblaciones. Dado que la mayor parte de la diversidad genética humana en todo el mundo toma la forma de clines geográficos en lugar de clados (Manica et al., 2005; Rosenberg et al., 2005; Serre & Pääbo, 2004), es esencial encontrar el pequeño número de SNP que muestran las discontinuidades de frecuencia de alelos más pronunciadas entre regiones continentales para crear conjuntos de marcadores con genotipos de “Diagnóstico” de la población (Rosenberg et al., 2002). Para ayudar a localizar tales SNP, un enfoque es examinar la variación genética que ha sido sometida a una fuerte selección regional positiva en el pasado reciente creando adaptaciones localizadas (Akey et al., 2004; Marth et al., 2004).

Los marcadores de inserción - deleción, también presentan una alta variabilidad en sus frecuencias alélicas entre poblaciones ancestrales, de ahí su utilidad en los estudios de ancestría (Pereira, 2011). Como se había mencionado, los InDels representan aproximadamente el 1620% de todos los polimorfismos de ADN humano, pero, a pesar de su abundancia en el genoma y de su importancia en algunas enfermedades humanas, estos marcadores han recibido mucha menos atención que los SNPs autosómicos.

Son muchos los estudios poblacionales mediante rastreos genómicos que aumentaron el conocimiento sobre ancestría poblacional. Un estudio genotipando todo el genoma de seis poblaciones diferentes de ancestría caribeña, junto con la disponibilidad de paneles de referencia a gran escala, permitió abordar preguntas sobre el origen y la historia de la mezcla de la cuenca del Caribe (Moreno-Estrada et al., 2013), encontrándose que las diferencias entre las poblaciones del Caribe insular y continental subrayan la importancia de caracterizar las poblaciones mezcladas a escalas más finas. También mediante estos estudios se ha podido evidenciar la estructura poblacional y fenómenos de deriva génica y cuellos de botella recientes, específicos de ancestría, que afectan a grupos latinos particulares, pero no a otros, que pueden tener implicaciones importantes en estudios de asociación con enfermedades en estas poblaciones.

Por otro lado, la estratificación poblacional dentro de los componentes subcontinentales implica que las variantes genéticas médicamente relevantes pueden estar restringidas geográficamente, lo que refuerza la necesidad de secuenciar las poblaciones objetivo para descubrir variantes locales que solo pueden ser relevantes en estudios de asociación con enfermedades, específicos de latinos (Homburger et al., 2015). En poblaciones colombianas de Antioquia y Choco se demostró la existencia de ancestros subcontinentales que parecen, a primera vista, ser claramente distintos, lo que contrasta con las nociones en competencia de identidad nacional que enfatizan la mezcla étnica (mestizaje) versus las identidades específicas de grupo (multiculturalismo) (Conley et al., 2017).

En poblaciones suramericanas, estos marcadores AIMs han permitido reconstruir su historia demográfica y origen de sus poblaciones ancestrales, como lo demuestran varios estudios (Fortes-Lima et al., 2017; García et al., 2015; Gravel et al., 2013).

Barbieri et al. (2019), comparando poblaciones de las regiones de los Andes, Amazonia y costa del Pacífico de América del Sur, reportó que las poblaciones amazónicas presentan un reciente flujo de genes entre sí y entre dominios eco geográficos, lo que no concuerda con la visión tradicional de grupos pequeños y aislados. Las conexiones genéticas a larga distancia entre hablantes de la misma familia

lingüística sugieren que las lenguas indígenas aquí no se propagaron solo por contacto cultural. Además, las poblaciones nativas americanas se mezclaron con fuentes europeas y africanas post colombinas en diferentes momentos, con pocos casos de aislamiento prolongado. Por el contrario, un análisis también realizado con poblaciones andinas reveló la existencia de ancestros genéticos nativos americanos altamente específicos y una gran homogeneidad intra grupal, junto con trazas limitadas de flujo genético principalmente desde los Andes hacia la Amazonia peruana. Una cantidad sustancial de deriva genética subrayó patrones distintos de endogamia reciente o aislamiento prolongado. Estos hallazgos sugieren extensas migraciones hacia Centroamérica desde fuentes no andinas u ocultan distintos eventos de poblamiento en el Continente Sur (Gnecchi-Ruscone et al., 2019). En la Figura 9, podemos observar un resumen de las proporciones de ancestría latinoamericana encontradas con marcadores genéticos uniparentales y autosómicos.

Por otro lado, la historia de la mezcla en Latinoamérica, con una diversidad elevada resultante en la región, evidenciada por estudios con marcadores autosómicos, ofreció un marco ventajoso para los estudios de asociación genética con enfermedades (Adhikari et al., 2016; Conomos et al., 2016).

También se evidenció en estas poblaciones la existencia de mezcla sesgada por sexo como un fenómeno muy común (Goldberg et al., 2014; Norris et al., 2019). En la Figura 9, por ejemplo, se observa que la proporción de ancestría europea estimada a partir del cromosoma Y es consistentemente mayor que la estimada a partir de marcadores de ADNmt. Por el contrario, las ancestrías nativa americana y africana son mayores en el ADNmt que en el cromosoma Y. Este patrón apunta a un sesgo sexual durante la mezcla (principalmente hombres europeos y mujeres nativas / africanas). De acuerdo con este patrón, la ancestría europea en los autosomas es mayor de lo que se estima con los marcadores del cromosoma X (dado que los hombres y las mujeres aportan el mismo número de autosomas, pero los hombres solo un cromosoma X y las mujeres dos).

Los estudios de mapeo de mezclas también han permitido explorar la posibilidad de que la selección haya dado forma a la

variación en la ancestría genética regional a lo largo del genoma, lo que contradice los modelos estadísticos iniciales en genética médica y de poblaciones, que asumían que los individuos se apareaban al azar y refuerzan el concepto de apareamiento no aleatorio en las poblaciones humanas (Norris et al., 2020; Norris et al., 2019; Zaitlen et al., 2017).

En contraste, para poblaciones de diversas etnias africanas, hay todavía una caracterización insuficiente de estudios genómicos con plataformas de secuenciación de SNPs autosómicos debido a limitaciones de financiación por parte de los gobiernos nacionales; poca infraestructura de investigación genómica y falta de herramientas genómicas óptimas (por ejemplo, matrices GWAS) y metodologías para el interrogatorio de genomas africanos (Rotimi et al., 2017). Pero los resultados de algunos estudios sugieren que los grupos de ancestría africana se encuentran distribuidos más homogéneamente entre las poblaciones mezcladas del continente americano que entre las poblaciones africanas que contribuyeron a la trata transatlántica de esclavos (Gouveia et al., 2020; Rotimi et al., 2016). Lo anterior enfatiza la importancia de incluir regiones poco estudiadas del continente americano en estudios genéticos de alta resolución, con matrices de chips de SNPs para análisis informativos a escala regional.

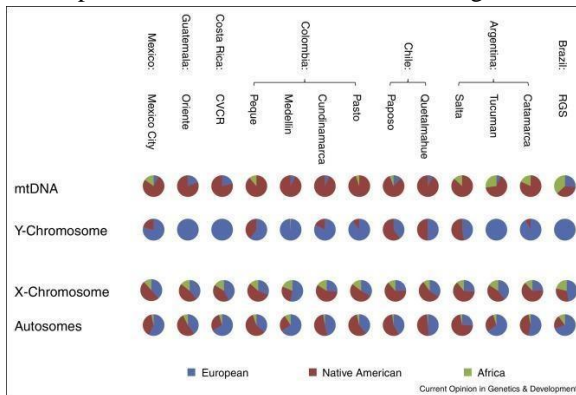


Figura 9. Proporción de ancestría africana, europea y nativa americana estimada con ADNmt, cromosoma Y, cromosoma X y marcadores autosómicos en trece muestras de población latinoamericana.

Nota. Está modificada de (Ruiz-Linares, 2015; Wang et al., 2008).

En contraste, para poblaciones de diversas etnias africanas, hay todavía una caracterización insuficiente de estudios genómicos con plataformas de secuenciación de SNPs autosómicos debido a limitaciones de financiación por parte de los gobiernos nacionales; poca infraestructura de investigación genómica y falta de herramientas genómicas óptimas (por ejemplo, matrices GWAS) y metodologías para el interrogatorio de genomas africanos (Rotimi et al., 2017). Pero los resultados de algunos estudios sugieren que los grupos de ancestría africana se encuentran distribuidos más homogéneamente entre las poblaciones mezcladas del continente americano que entre las poblaciones africanas que contribuyeron a la trata transatlántica de esclavos (Gouveia et al., 2020; Rotimi et al., 2016). Lo anterior enfatiza la importancia de incluir regiones poco estudiadas del continente americano en estudios genéticos de alta resolución, con matrices de chips de SNPs para análisis informativos a escala regional.

En las últimas tres décadas, los estudios genéticos han jugado un papel cada vez más importante en la exploración de la historia humana. Han ayudado a establecer de manera concluyente que los humanos anatómicamente modernos aparecieron por primera vez en África, aproximadamente 250.000-350.000 años antes del presente, y posteriormente emigraron a otras partes del mundo. La historia de los seres humanos en África es compleja e incluye eventos demográficos que influyeron en los patrones de variación genética en todo el continente. A través de estudios genéticos, se ha hecho evidente que la historia profunda de la población africana se refleja en las relaciones entre los cazadores-recolectores, ya que las divergencias de población más profundas del mundo ocurren entre estos grupos, remontándose a 300.000 años antes del presente. Sin embargo, la expansión del pastoreo y la agricultura en los últimos miles de años ha moldeado la distribución geográfica de los africanos actuales y su diversidad genética. Las tecnologías de secuenciación de tercera generación han arrojado más datos de secuencias genómicas completas de diversos grupos poblacionales africanos existentes y prehistóricos, pero aún falta conocimiento de la historia evolutiva humana en África (Campbell et al., 2014).

2.4. Marcadores de identificación humana de tipo InDel

El trabajo de Weber et al. (2002) permitió identificar unos 2000 InDels, con diferencias en la estructura del polimorfismo (bialélicos o multi alélicos), diferencias en la longitud de los alelos, la posición cromosómica, así como las frecuencias alélicas de los mismos en los principales grupos de la población mundial (Mills et al., 2006). Aproximadamente un tercio de los InDels reportados fueron identificados dentro de genes conocidos, y alrededor del 3,7% estaban situados en los exones y en las regiones promotoras. Además, se clasificaron en 5 grupos:

- Inserción o delección de una base
- Expansiones monoméricas de una base
- Expansiones de 2-15 pb
- Inserciones de transposones
- InDels que contienen aparentemente secuencias de ADN al azar.

Tereba (1999) reportó que esta última clase incluye inserciones-delecciones de secuencias de ADN aparentemente aleatorias, que representan aproximadamente el 41% de todos los InDels, y poseen polimorfismos con una amplia gama de variación de la longitud, desde 1 pb hasta aproximadamente de 10 kb.

Casi todos los InDels (más de 99%) generan productos de PCR menores de 100 pb de longitud. Diferentes estudios han sido publicados usando InDels para una variedad de propósitos, por ejemplo, estudiar la estructura genética de las poblaciones humanas, así como para inferir proporciones de origen ancestral individual o poblacional o como marcadores genéticos útiles en el análisis de poblaciones naturales y en identificación de especies. Los InDels poseen tasas de mutación muy baja, del orden de 10^{-8} por base en cada generación, por lo que son particularmente apropiados en investigación de paternidad, especialmente en casos donde se encontraron pocas mutaciones en los marcadores STR.

Li et al. (2011) precisan que los HID-InDels son bialélicos, tienen rangos de tamaño de amplicón entre 50-150 pb, lo que los hace muy útiles en el análisis forense de muestras degradadas o complejas. Mientras que Zidkova et al. (2013) señalan que también pueden ser

analizados utilizando las mismas tecnologías que para los STRs mediante multiplex.

Dado lo anterior se pueden utilizar con éxito para estimar la ancestría, tanto a nivel individual como de población, proporcionando información importante para las investigaciones forenses. Generalmente esto se evalúa comparando con los datos generados por GWAS - Genome Wide Association Studies, como los publicados por Galanter et al. (2012) y Daya et al. (2013). La determinación de la ancestría en hermanos también demostró ser una buena estrategia para evaluar el desempeño de estos marcadores, teniendo en cuenta que un conjunto eficiente de marcadores proporcionará resultados idénticos de ascendencia entre ellos (Pereira et al. (2020).

Comercialmente se encuentra disponible para estudios de identificación forense el kit investigador DIPplex® kit (Qiagen, Hilden, Germany). Este último kit ha sido utilizado en un estudio realizado en Corea por Kim et al. (2014), que demuestran que el conjunto de 30 marcadores es más informativo que los seis STRs del AmpFISTR® COfiler® PCR Amplification Kit (Applied Biosystems), pero menos discriminativo que los nueve marcadores genéticos del AmpFLSTR® Profiler Plus® ID PCR Amplification Kit (Applied Biosystems). Además, cuatro de estos marcadores de tipo InDel (D111, D118, D81, D99) pueden ser utilizados como marcadores de ancestría en asiáticos. Trabajos realizados en otras poblaciones muestran resultados idénticos en términos de la menor capacidad de discriminación de los InDels con respecto a los STRs, y señalan que esto se debe a la menor heterocigosidad presentada por los InDels (Ibarra et al., 2014; Kim et al., 2014; Lander et al., 2001).

En conclusión, los HID-InDels permiten resolver situaciones forenses peculiares cuando los STR autosómicos son insuficientes. Aunque se predijeron sus limitaciones desde la implementación forense de marcadores bialélicos, la evaluación formal de estas restricciones es escasa. Particularmente es útil definir su informatividad en el análisis de parentesco, para evitar desperdiciar trabajo y recursos económicos.

En seguida se describen los parámetros estadísticos relevantes para determinar la informatividad de un conjunto de marcadores

genéticos en casos forenses, sea para identificación individual o para determinación de vínculos de parentesco biológico:

- Probabilidad de discriminación (PD): se define como la probabilidad a priori de que dos individuos no relacionados, seleccionados al azar, puedan distinguirse genéticamente mediante el análisis de un conjunto de marcadores, siendo estos estadísticamente independientes. Este parámetro depende del número de loci analizados y del polimorfismo de cada uno de estos, pues la probabilidad de discriminación es una medida relativa de la eficacia de los sistemas utilizados. Esta probabilidad depende de las frecuencias alélicas de la población a la cual pertenece el individuo analizado. Si la probabilidad está cercana a uno (1), sería ideal para identificar al individuo.
- Probabilidad de coincidencia (PM): probabilidad de que dos individuos tomados al azar en una población tengan genotipos idénticos, es decir, es la probabilidad de que dos individuos no relacionados compartan un mismo perfil de ADN. El PM es todo lo opuesto al concepto de poder de discriminación ($PD = 1 - PM$).
- Poder o Probabilidad de exclusión (PE): se define como la probabilidad de excluir como padre biológico a un individuo al azar de la población acusado falsamente de ser el padre. Es un valor estadístico porcentual, como función directa del polimorfismo de un marcador dado, así que entre más polimórfico sea el sistema y más equilibradas se encuentren las frecuencias de sus alelos, tanto mayor será su probabilidad de exclusión a priori. El cálculo de este parámetro es sencillo cuando los sistemas analizados son bialélicos y con tipo de herencia codominante.
- Heterocigosidad: La heterocigosidad observada (Hobs) determina la frecuencia de individuos heterocigotos encontrados en la población en estudio, mientras que la

heterocigosidad esperada (Hesp) es dada por la frecuencia esperada de heterocigotos asumiendo que la población está en equilibrio de Hardy-Weinberg. La definición clásica de heterocigosidad es la probabilidad de que dos alelos del mismo locus tomados al azar de la población sean diferentes. Además, la Hesp también se conoce como diversidad genética.

- Equilibrio de Hardy-Weinberg (EHW): esta ley fue introducida por George Hardy y Wilhelm Weinberg en 1908. Esta ley se aplica a toda población suficientemente grande, donde los apareamientos se dan al azar, no opera ninguna fuerza evolutiva: mutación, selección, migración y deriva, entre otras. Con estas condiciones ideales las frecuencias génicas y genotípicas se mantienen constantes de generación en generación, cumpliendo con el EHW. Cuando este equilibrio no se cumple, es necesario inferir las causas, siendo que la más frecuente en poblaciones mezcladas es la presencia de subestructura poblacional. En estos casos, es necesario tener en cuenta el grado de subestructura poblacional, en los cálculos estadísticos aplicados a la valoración de evidencias forenses.

2.5. La Diáspora Africana

El término “Diáspora africana” apareció por primera vez en la literatura en la década de 1950 y se ha definido de manera amplia para incluir a todas las comunidades globales descendientes de las migraciones históricas de pueblos de África desde el siglo XV (Palmer, 2000). La diáspora africana también se ha definido de manera más estricta para incluir solo la trata transatlántica de esclavos. Esta definición más restringida, que enfatiza los importantes roles que desempeñaron la negrura, la esclavitud, el colonialismo, el racismo y la geografía en el sostenimiento de la trata transatlántica de esclavos, es la razón por la que algunos se refieren a la “Diáspora africana” como la “Diáspora negra”. El comercio transatlántico de esclavos fue la mayor inmigración de africanos del Viejo Mundo al Nuevo Mundo. La primera

etapa del comercio triangular involucró a barcos de Europa que transportaban bienes (por ejemplo, hierro, brandy, armas y pólvora) que se intercambiaban por esclavos en África. La segunda etapa, denominada Paso Medio, involucró el envío de entre 12 y 14 millones de africanos esclavizados a través del Océano Atlántico hacia las Américas. El último tramo fue el transporte de mercancías (por ejemplo, azúcar, algodón, tabaco, ron y melaza) de América a Europa.

La mayoría de los africanos esclavizados fueron llevados a las colonias europeas en América Latina, mientras que entre el 3 y el 5% fueron llevados a los Estados Unidos de América (EE. UU.). Más de 10 millones de africanos esclavizados fueron transportados a las Américas entre 1500 y 1900 (Jordan, 2020). Figura 10.

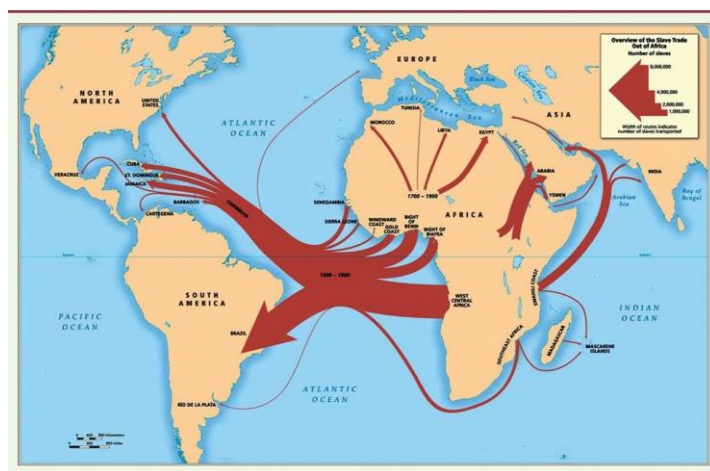


Figura 10. Rutas de la trata de esclavos desde 1500 hasta 1900.

Nota. Se indican los principales puntos de salida y llegada. El grosor de las flechas indica el número de personas involucradas. El tráfico intra-africano, relativamente minoritario durante este período, fue más importante durante los siglos anteriores (mapa presentado por el sitio Slave Voyages, slavevoyages.org).

La Figura 10 ilustra las principales características de este tráfico e indica las regiones de salida y llegada, así como el número aproximado de personas involucradas. Observamos que casi la mitad de los esclavos provienen de África central occidental (Angola y Congo en la actualidad), y que América del Norte es un destino muy pequeño

(alrededor de 0,5 millones) en comparación con el Caribe y América del Sur. Este vergonzoso comercio, que hizo fortuna a varios puertos franceses (Burdeos, La Rochelle, Nantes, etc.), británicos y holandeses, ha dejado muchas huellas escritas que se pueden consultar en particular gracias a una base de datos, la Trata Transatlántica de Esclavos (Micheletti et al., 2020), que contiene información detallada sobre más de 36.000 viajes realizados entre 1500 y 1900. Se puede leer, por ejemplo, que el

Marie-Gertrude, que salió de Saint-Malo el 5 de septiembre de 1719, embarcó 190 esclavos en Whydah (Ouaidah, Benín) y desembarcaron los 154 supervivientes en Léogâne (Haití) el 11 de junio de 1720. En algunos casos, incluso podemos acceder a la lista de esclavos a bordo, conocer su nombre africano, su edad y su estatura.

Estos africanos esclavizados y sus descendientes en los Estados Unidos representan el grupo “afroamericanos”, al igual que otras poblaciones de la diáspora en Latinoamérica (Borucki et al., 2015; Shriner et al., 2014).

Si bien el panorama genético actual de la ancestría africana en las Américas concuerda en gran medida con los registros de viajes transatlánticos, los relatos históricos adicionales permiten aclarar algunas inconsistencias como por ejemplo, la sobrerrepresentación de la ancestría nigeriana en partes de las Américas, que se explica por el comercio intraestadounidense de personas esclavizadas del Caribe británico, la subrepresentación de la ancestría senegambiana que está respaldada por relatos de comercio temprano y alta mortalidad en esta región, los patrones de variación en las Américas, como la baja ancestría africana y un mayor sesgo sexual femenino africano, que pueden atribuirse a factores socioeconómicos y a la mezcla de hombres no africanos entre otros (Micheletti et al., 2020).

Por otro lado, publicaciones en el caso de los afroamericanos (Norte América) reportan múltiples ancestrías con linajes de África, Europa, Asia y América nativa, entre otros. Por lo tanto, el término “afroamericano” no es una entidad genéticamente homogénea, como se refleja en el hecho de que los afroamericanos autoidentificados incluyen individuos que van desde casi ningún antepasado africano hasta casi ningún antepasado europeo (Bryc et al., 2010; Shriner et al., 2011).

La diversidad cultural abunda en los descendientes de la diáspora africana como lo ejemplifican muchas poblaciones afrodescendientes de América, por ejemplo, los brasileños. Con más de 4 millones de esclavos, Brasil ha sido durante mucho tiempo un crisol de ancestros y culturas, como lo demuestra la mezcla de religiones africanas y europeas (candomblé, catolicismo), comidas (feijoada), música y danza (samba). Alrededor del 50% de los ~ 200 millones de brasileños se identifican a sí mismos como personas de ascendencia africana o mezclada (Phillips, 2011). Estas mezclas de ancestros, adquiridas a lo largo de cientos de años, tienen serias implicaciones para las iniciativas biomédicas nacionales y mundiales, como la medicina de precisión, porque es poco probable que las apariencias fenotípicas y la identificación personal y grupal capturen adecuadamente los antecedentes ancestrales de los individuos que componen la diáspora africana.

Si bien los registros transatlánticos nos han contado gran parte de la historia, los conocimientos de los estudios genéticos poblacionales, que investigan los componentes regionales de la ascendencia africana en los afroamericanos actuales, en combinación con otros relatos históricos, aportan conocimientos adicionales del impacto genético de la trata transatlántica de esclavos en las poblaciones actuales de las Américas.

2.6. Estructura Poblacional y Diversidad Genética entre las Poblaciones Africanas

El panorama de la estructura poblacional y la diversidad genética de los africanos subsaharianos (SSA) fue enriquecido por los reportes de dos grandes proyectos internacionales: el Proyecto de Variación del Genoma Africano (AGVP) (Gurdasani et al., 2015) y el Proyecto 1000 Genomas (Consortium, 2015). Los datos de AGVP, generados a partir de 20 grupos etnolingüísticos africanos, revelaron una estructura de la población que antes no se apreciaba, incluidos patrones de mezcla regionalmente distintos. Utilizando el análisis de componentes principales, el análisis de conglomerados no supervisado y la prueba f_3 para la mezcla, se observó evidencia de una mezcla sustancial “euroasiática” y de cazadores-recolectores en toda África

subsahariana (ASS). El momento y las fuentes de mezcla fueron regionalmente distintos, con la mezcla en los africanos occidentales que datan de hace ~ 9000 años y una fuente similar a las poblaciones actuales de Khoe-San y con una mezcla en los africanos del este que datan de ~ 3000 años atrás y una fuente similar a Poblaciones de cazadores-recolectores de la selva tropical Mbuti (Gurdasani et al., 2015).

Las poblaciones africanas están muy subdivididas, con una estructura poblacional de toda África actualmente reconocida como 11 ancestrías que corresponden a la siguiente separación geográfica y lingüística: khoisan en el sur de África; Centroatricano, predominante en pigmeos; hadza en Tanzania; África occidental, predominante en los pueblos de habla mande; África centro-occidental, predominante tanto en los pueblos de habla bantú como en los que no hablan bantú en el área desde Ghana hasta Camerún; y ancestros correspondientes a hablantes de lenguas bereberes, cusitas, bantú oriental, omóticas, nilo-sahariana y bantú meridional (Shriner et al., 2014; Tishkoff et al., 2009).

En comparación, se han detectado 12 ancestrías en el resto del mundo, incluidas dos que definen la diferenciación del norte y sur en Europa y una que abarca la ancestría de los nativos americanos (Shriner et al., 2014). La diferenciación genética, medida por distancias genéticas (F_{ST}), entre ancestros de ASS puede exceder la entre pares de ancestros no africanos. Por ejemplo, el F_{ST} es 0.054 entre ancestros khoisan y omotic, comparado con 0.024 entre ancestros del sur y norte de Europa o 0.042 entre ancestros árabes e indios.

Se observan niveles más altos de diversidad genética entre ASS. El número de variantes por individuo de ancestría ASS es ~ 5 millones, en comparación con ~ 4,0-4,2 millones de variantes por individuo de ancestría de Asia oriental, europea o del sur de Asia (Consortium, 2015). La tasa promedio de diferencias de nucleótidos de 1.2 por kilobase entre un par de individuos Khoe-San excede la de 1.0 por kilobase entre un individuo asiático y un individuo europeo (Schuster et al., 2010). En función de la distancia física, el desequilibrio de ligamiento decae más rápidamente en las poblaciones de SSA que en las poblaciones no africanas (Consortium, 2015), de modo que los haplotipos son más cortos en SSA.

Por último, la disponibilidad de genomas antiguos de África proporciona conocimientos históricos de una profundidad sin precedentes (Choudhury et al., 2018).

2.7. Diversidad Etnolingüística en África

Ahora podemos estudiar los genomas de los africanos existentes y descubrir rastros de migración, mezcla, lenguas y selección de poblaciones mediante la aplicación de sofisticados algoritmos computacionales. Hay cuatro divisiones etnolingüísticas principales entre los africanos actuales: poblaciones de cazadores-recolectores en el sur y centro de África; hablantes nilo-saharianos del norte y noreste de África; hablantes afroasiáticos del norte y este de África; y hablantes de Níger-Congo que son el grupo etnolingüístico predominante distribuido en la mayor parte del África subsahariana. La enorme diversidad etnolingüística en las poblaciones de África subsahariana tiene un paralelo en gran medida con una amplia diversidad genética y, hasta hace una década, se sabía poco sobre los orígenes detallados y la divergencia de estos grupos.

Los resultados de estudios genéticos de poblaciones a gran escala, y más recientemente los datos de la secuencia del genoma completo, están desentrañando el papel crítico de eventos como la migración y la mezcla y factores ambientales, incluida la dieta, las enfermedades infecciosas y las condiciones climáticas, en la configuración de la diversidad de la población actual. Ahora es posible comenzar a proporcionar estimaciones cuantitativas de los tiempos de divergencia, el tamaño de la población y los procesos dinámicos que han afectado a las poblaciones y su riesgo genético de enfermedad.

África fue el lugar de nacimiento de los humanos modernos, y desde este continente se produjo la propagación del *Homo sapiens* al resto del mundo. El estudio de África es esencial para la genética de poblaciones humanas, debido a la gran diversidad genética, lingüística, cultural y fenotípica presente en las poblaciones africanas y entre ellas. En África están presentes más de 2000 grupos etnolingüísticos distintos (Campbell & Tishkoff, 2008). Un grupo etnolingüístico desarrolla una identidad basada en las costumbres y tradiciones en una variedad de

formas, incluidos los hábitos alimentarios y los tipos de ropa (Falola, 2001).

Aproximadamente un tercio de las lenguas del mundo están presentes en África, y pueden incluirse en cuatro de las principales familias lingüísticas: Níger-Congo, afroasiática, nilo-sahariana y khoisan. La lengua Níger-Congo es hablada por poblaciones agrícolas en una vasta región de África (excepto el norte de África). La lengua afroasiática se habla principalmente en el norte y el este de África, por los pastores agrícolas y las poblaciones de pastores. El idioma nilo-sahariano es hablado por poblaciones de pastores en África central y oriental. La familia de la lengua khoisan es hablada por poblaciones de cazadores-recolectores presentes en el este y sur de África (Gomez et al., 2014; Scheinfeldt et al., 2010).

Las poblaciones africanas viven en diversos climas, desde la segunda selva tropical más grande (la cuenca del Congo) hasta el desierto más grande del mundo (el Sahara), las montañas, la sabana y los pantanos. Por lo tanto, las poblaciones africanas poseen un amplio conjunto de medios de subsistencia, tales como: el pastoreo, diversos modos de agricultura, y la caza y la recolección (Lambert & Tishkoff, 2009).

El modelo Fuera de África (OOA, del inglés “Out of Africa”) defiende que los humanos modernos evolucionaron en África hace aproximadamente 200 mil años, y luego, entre 50 y 100 mil años, emigraron de África y colonizaron la mayoría de las otras áreas del mundo. La ubicación exacta del origen de los humanos modernos todavía está en debate con algunos científicos que proponen diferentes regiones africanas como posible origen, como África Central y Noroeste (Campbell & Tishkoff, 2008; Cruciani et al., 2011; Henn et al., 2011; Prugnolle et al., 2005). Así que, los humanos modernos han existido en África por más tiempo que en otros continentes, con muchas poblaciones diferentes, cada una con un gran número de individuos.

Estas poblaciones fueron sometidas a una amplia y diferente variedad de factores de cultura, dieta y geografía, que contribuyeron a altos niveles de diversidad genética y fenotípica dentro de ellos y, en consecuencia, resultando en una historia compleja en todo el continente africano (Tishkoff et al., 2009). Dado que África es la patria del Homo

sapiens su diversidad genética todavía se sigue estudiando ya que, en comparación con las poblaciones euroasiáticas y americanas, el estudio genético del continente africano es de suma importancia para conocer el pasado de las poblaciones afrodescendientes modernas. En el continente africano, los eventos migratorios y las fluctuaciones de la población son fuerzas demográficas que influyen en la variación a escala genómica, lo que dio forma a los patrones de variación genética de las poblaciones africanas modernas. Un ejemplo que ilustra este impacto es la Expansión Bantú (Lambert & Tishkoff, 2009).

2.8. La expansión bantú

La expansión bantú fue la dispersión más significativa que se produjo en África y produjo una transmisión y mezcla de culturas, idiomas y genes durante los últimos tres a cuatro milenios. Esta expansión se asoció con la expansión de la agricultura a través del África subsahariana, que comenzó en la frontera entre Nigeria y los países de Camerún hace aproximadamente 5000 años y fue responsable por un reemplazo casi total de las poblaciones anteriores en esta región. El término “Bantú” se refiere a una familia lingüística que abarca alrededor de 500 idiomas pertenecientes a la rama Benue Congo del supra grupo etnolingüístico Níger-Congo (Berniell-Lee et al., 2009; Rowold et al., 2016). Los hablantes bantúes, que son agricultores, se dispersaron para sostener a su creciente población.

Esta expansión aparece como un evento único masivo que culmina en el movimiento y reasentamiento del pueblo Bantú a lo largo de dos rutas principales: (1) al sur de su tierra natal de Camerún a través de pastizales y bosques abiertos y luego girando hacia el este a través de la selva ecuatorial hacia las ricas tierras cultivables del Este de África y; (2) un curso suroeste a lo largo de las llanuras costeras para aprovechar la abundante fauna marina, que luego terminó en la región seca del suroeste del continente (Newman, 1995; Vansina, 1995). Sin embargo, esta diáspora fue un fenómeno complejo formado por múltiples episodios migratorios de diferentes ritmos y direcciones que ocurrían en una línea de tiempo prolongada. Durante este tiempo, la dispersión bantú implicó una difusión de la lengua y la cultura bantú, así como material genético a los habitantes autóctonos de la nueva frontera

(Figura 11). La convergencia entre estas dos rutas podría haber ocurrido en múltiples ocasiones durante diferentes períodos de tiempo (Plaza et al., 2004).

En general, los estudios sobre la expansión bantú han sido muy desiguales en cuanto a las zonas geográficas estudiadas, con un número muy limitado de poblaciones muestreadas y de alcance limitado. Varias líneas de investigación, que incluyen biogeografía, palinología, geología, lingüística histórica y arqueología, han abordado diferentes aspectos de la expansión bantú (para una lúcida revisión reciente, (Bostoen et al., 2015). Además, recientemente, una serie de estudios basados en SNPs de todo el genoma, así como en marcadores específicos de ChrY e ADNmt, se han centrado en la distribución y la historia de los haplogrupos del ChrY (Batini et al., 2011) así como su correlación con factores lingüísticos y geográficos (de Filippo et al., 2011). Además, los estudios basados en la herencia materna han descubierto una homogeneidad genética entre los grupos de habla bantú de África oriental y occidental (Barbieri et al., 2016; Barbieri et al., 2014), con poblaciones en áreas de expansión que exhiben una contribución pronunciada de cazadores-recolectores (Marks et al., 2015).

Las investigaciones de Barbieri et al. (2013) reportaron una antigua subestructura de los primeros linajes de ADNmt en los khoisan y la diferenciación genética entre la población de habla bantú derivada del flujo de genes específicos de los khoisan (Barbieri, Vicente, et al., 2013). Los datos con estudios de todo el genoma, por otro lado, sugieren solo un flujo de genes limitado entre los inmigrantes de habla bantú y los khoisanos en el sudeste de África (GonzálezSantos et al., 2015). También son particularmente pertinentes los estudios recientes sobre el haplogrupo de cromosoma Y, B2a, que han arrojado dudas sobre su origen específico bantú (Barbieri et al., 2016; Scozzari et al., 2014).

2.9. Perfil Genómico de las Poblaciones Contemporáneas en América de la Diáspora Africana

En los afroamericanos, las diferencias a nivel continental entre africanos y europeos pueden explicar hasta un 8% de la variación fenotípica en una variedad de rasgos antropométricos y cardio-metabólicos (Shriner et al., 2015). Dado el progreso en la delimitación de la estructura de la población a nivel subcontinental descrito anteriormente, se necesita una comprensión más profunda de la estructura genética a escala fina en las poblaciones de la diáspora africana.

La ancestría de África occidental es la predominante en el pueblo Mende de Sierra Leona y en los pueblos Jola, Mandinka y Wolof de Gambia (Consortium, 2015; Gurdasani et al., 2015).

La ancestría centro-occidental africana es predominante en los pueblos que no hablan bantú, como los pueblos Ga-Adangbe de Ghana y los pueblos Esan, Igbo y Yoruba de Nigeria.

Entre los pueblos de habla bantú, la diferenciación genética que siguió a la expansión bantú dio lugar a ancestros distintos en África centro-occidental (que se encuentra en los pueblos Bamum y M'fang en Camerún y en los pueblos Kongo de la República Democrática del Congo), África Oriental (que se encuentra en Los pueblos Baganda, Barundi y Banyarwanda de Uganda y los pueblos Luhya y Kikuyu de Kenia), y el sur de África (que se encuentra en los pueblos Sotho y Zulu de Sudáfrica) (Bryc et al., 2010; Gurdasani et al., 2015; Shriner et al., 2015). En el cromosoma Y, los africanos occidentales tienen comparativamente más E1b1a1a1f, mientras que los africanos centro-occidentales tienen comparativamente más E1b1a1a1g. Estos hallazgos de ancestría regional dentro de África continental indican que es posible un cierto grado de localización de origen africano para los afroamericanos y otros africanos de la diáspora (Consortium, 2015).

El perfil genómico de las poblaciones contemporáneas de las Américas refleja la mezcla que ocurrió entre europeos, nativos americanos y africanos esclavizados en el Nuevo Mundo. Por varias razones, incluido el lugar donde desembarcaron los esclavos y las prácticas socioculturales como el apareamiento selectivo con ancestría positiva, el porcentaje promedio de ancestría africana en las Américas

varía ampliamente. Estimaciones recientes oscilan entre el 87% de los africanos caribeños de Barbados, el 75% de los afroamericanos del suroeste de los EE. UU., el 19% de los puertorriqueños, el 12% de los colombianos en Medellín, el 7% de los de ascendencia mexicana en Los Ángeles y el 4% de ambos peruanos en Lima y argentinos de toda Argentina (Avena et al., 2012; Consortium, 2015; Mathias et al., 2016).

La ascendencia africana en el Caribe parece coherente con dos oleadas, la primera de África occidental seguida de la segunda de África centro-occidental (Moreno-Estrada et al., 2013). En América del Sur, la ascendencia africana promedia el 5% (Homburger et al., 2015). En Brasil, la ascendencia africana varía del 51% en Salvador, en el noreste, al 15% en Bambuí, en el sureste, y el 16% en Pelotas, en el sur (Kehdy et al., 2015).

Las ascendencias de África occidental y central son más frecuentes en el noreste de Brasil, debido a una mayor proporción de desembarques de África occidental y centro-occidental en Salvador, mientras que la ascendencia de África oriental es más prevalente en el sureste y el sur debido a una mayor proporción de desembarques de Mozambique en Río de Janeiro (Rodrigues de Moura et al., 2015). La estructura de la población en las Américas muestra una mayor complejidad por el porcentaje promedio ampliamente variado de ascendencia nativa americana en estas poblaciones, con alrededor del 1% en los caribeños africanos de Barbados, el 4% en los afroamericanos en el suroeste de EE. UU., el 15% en los puertorriqueños, el 28% en colombianos en Medellín, 49% en personas con ascendencia mexicana en Los Ángeles y 78% en peruanos en Lima (Consortium, 2015).

Las matrices de genotipificación iniciales fueron menos eficientes para analizar los genomas de poblaciones de ascendencia africana debido al hecho de que se observa una mayor variación genética en las poblaciones actuales de África que en las poblaciones fuera del continente, lo que resulta en un mayor número de haplotipos, menores niveles de desequilibrio de ligamiento (LD), patrones más divergentes de LD y patrones más complejos de subestructura de la población. Por ejemplo, nuestro reciente esfuerzo de secuenciación del genoma completo en 320 SSA identificó alrededor de 30 millones de variantes, de las cuales hasta una cuarta parte no fueron observadas en otras poblaciones del proyecto de secuenciación de 1000 Genomas, lo que

indica la necesidad de secuenciación profunda de diversas poblaciones a través de SSA y la necesidad de una mejor matriz de genotipado (Consortium, 2007).

Aunque se han hecho algunos esfuerzos para abordar estas preocupaciones, incluido el desarrollo reciente de *Infinium Multi-Ethnic Genotyping Array* (MEGA) y *Affymetrix® Axiom® Genome-Wide Pan-African* (PanAFR) Array, sigue siendo necesario el desarrollo de una matriz de genotipado panafricano que capture una mayor proporción de variación genética común en diversas poblaciones africanas. Para lograr este objetivo, el consorcio H3Africa se ha unido al *Wellcome Trust Sanger Institute*, *Illumina* y otros para desarrollar una matriz de genotipado interrogando secuencias de genoma completo de más de 4.000 africanos muestreados en todo el continente. También existe la necesidad de mejorar los algoritmos para la inferencia de ancestría a partir de los datos de la secuencia del genoma completo. Las cuestiones metodológicas incluyen la contabilidad de genotipos inciertos, desequilibrio y fase de ligamiento, lecturas de secuencia de baja cobertura y poblaciones de origen inadecuadas o faltantes (Bansal & Libiger, 2015; Guan, 2014; Skotte et al., 2013; Wang et al., 2015).

2.10. Palenque y sus Orígenes Africanos

SBP es una pequeña aldea rural de casi 4.000 habitantes, situada a unos 60 kilómetros de Cartagena de Indias (Figura 12), Bolívar, Colombia. SBP cuenta con una posición extraordinaria en la historia social y lingüística latinoamericana, ya que alberga la población genotípicamente más negra de Latinoamérica, y es el único lugar en todo el territorio continental suramericano que posee un criollo de base española (conocido localmente como *lengua*, pero que adopta el nombre de *palenquero* en la literatura académica).

Debido en parte a su prolongado aislamiento del resto de la sociedad colombiana hasta la década de 1990, SBP ha mantenido un conjunto de tradiciones culturales únicas con profundas raíces africanas. Por ejemplo, los palenqueros han estado utilizando cánticos fúnebres conocidos como “lumbalú” durante siglos, una tradición que pasó de generación en generación hasta que los propios participantes olvidaron

el significado literal de algunos de sus célebres textos litúrgicos africanos (Schwegler, 1999, 2000, 2011).

Colombia recibió una cantidad numerosa de esclavos transportados del continente africano, especialmente entre 1595 y 1640, y la ciudad de Cartagena, Bolívar, era no sólo el mercado de esclavos más activo de la región, sino también el principal puerto de entrada legal de esclavos para todos los destinos en Suramérica (Newson y Minchin 2007^a: 136-137). Los ancestros de los palenqueros consiguieron escapar de Cartagena hacia las densas y profundas (en la época) selvas tropicales del interior de la costa caribeña. En 1713, la población estaba constituida por 234 cimarrones, repartidos a partes iguales entre criollos y bozales (Escalante, 1979; Navarrete Peláez, 2008).

A lo largo del tiempo, por relatos históricos se reportó que llegaron africanos provenientes principalmente de la costa sur – occidental y, a mediados del año de 1700, también fueron traídos esclavos de la costa este, particularmente de Mozambique. Se fue observando una variación relacionada con los puntos de captura y también con el envío de los esclavos hacia las colonias. Sin embargo, los registros que generaban datos acerca del origen étnico de una parte de los esclavos traficados podían no ser verdaderos puesto que muchos podrían ser provenientes de regiones distantes de los puertos de salida en África. Después de la llegada, había una distribución selectiva de los esclavos africanos hacia diferentes regiones, la cual, basándose en las características físicas de cada individuo, se determinaba el envío a localidades en donde desarrollaban actividades de acuerdo a lo que se consideraba más apto. Probablemente, los esclavos permanecían en la capital, o eran encaminados para otras regiones, principalmente los individuos seleccionados para trabajar en cultivos de caña de azúcar, mientras que otros eran enviados para trabajar en las minas de extracción de oro y diamantes (Granda Gutiérrez, 1971).

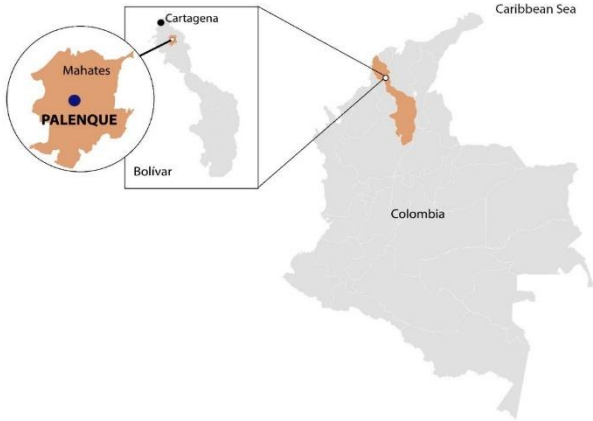


Figura 12. Localización geográfica de SBP, Bolívar, Colombia.

El término palenque se define como aquel lugar poblado por cimarrones o esclavos africanos fugados del régimen esclavista durante el período colonial. De ahí que se convirtió en sinónimo de libertad ya que toda persona que llegaba a formar parte de un palenque era automáticamente libre (Arrazola, 1970; Friedemann & Patiño Rosselli, 1983; Girón, 1621). La comunidad del Palenque de San Basilio conserva una conciencia étnica que le permite entenderse como un pueblo específico en la diáspora africana en el continente americano, con la única lengua criolla con base léxica española, una organización social “sui generis” basada en los ma-kuagro (grupos de edad), así como con complejos rituales fúnebres como el lumbalú o prácticas médicas tradicionales que evidencian un sistema cultural y espiritual excepcional sobre la vida y la muerte en la comunidad de Palenque (Schwegler, 2017).

Después de la Unión de las Coronas Ibéricas en 1580, la mayor parte de la riqueza de Cartagena provino del comercio legal y del contrabando de africanos cautivos. Además, su ubicación geográfica, especialmente la gran extensión de bahía, aumentó los desarrollos en su infraestructura y posicionó a Cartagena como el único puerto de llegada autorizado para africanos esclavizados, incluso de aquellos destinados eventualmente a otros puertos en la parte sur del continente. Entre 1595 y 1640, el comercio de africanos cautivos en los dominios españoles

estuvo monopolizado por comerciantes portugueses a través del asiento, un contrato en el que la Corona concedía derecho de exclusividad en el comercio a los empresarios individuales y sus empresas (Ortega, 2002; Vilar, 1977). Algunos números comienzan a abarcar la escala de esta actividad. Solo en Cartagena, los comerciantes desembarcaron más de 29.000 hombres, mujeres y niños cautivos de África Occidental, legalmente, entre 1585 y 1600 (Wheat, 2016).

Entre 1570 y 1640, cerca de quinientos barcos desembarcaron cautivos legalmente en el puerto de la ciudad. Muchos fueron transbordados al interior, pero con el tiempo, a través de la huida y la manumisión, otros cautivos, como sus descendientes libres y esclavizados, formaron comunidades en Cartagena y sus alrededores (Wheat, 2013, 2016). El comercio legal de cautivos también abrió posibilidades para el contrabando, oportunidades en que los funcionarios del asiento y de la corona, así como los capitanes de barco, buscaban evitar impuestos y vender directamente a los compradores con un beneficio mayor. Si bien el número de cautivos africanos introducidos ilegalmente en Cartagena es imposible de medir, los registros impositivos y portuarios han permitido a los historiadores estimar que el contrabando representó al menos el doble de cautivos que los incluidos en registros oficiales de entrada al puerto (García, 1982).

Para 1602 y 1603, las ganancias generadas por el comercio de cautivos africanos se habían convertido en la fuente de ingresos más importante, no solo para la tesorería real de Cartagena, a través de impuestos, sino también para los comerciantes portugueses, sus aliados locales y funcionarios de todos los rangos (Mancall & Shammas, 2015; Roulston, 2020; Wheat, 2013, 2016).

Por registros históricos como los anteriormente descritos, los esclavos del Caribe procedían de Camerún, Nigeria y Ghana, pero había muchas dudas respecto a su verdadero origen africano. El análisis del genoma completo de tres individuos enterrados en la isla de San Martín en el Caribe en el siglo XVII, en una época marcada por el tráfico de esclavos entre África y América, demuestra que estos esclavos procedían de grupos de habla bantú del norte de Camerún y de comunidades no bantúes de Nigeria y Ghana. Entre 1500 y 1850, más de doce millones de africanos esclavizados fueron transportados al

‘Nuevo Mundo’. La gran mayoría fue trasladada desde África occidental y central. A pesar de lo reportado por los registros históricos y los estudios de genética descritos anteriormente, hasta ahora se desconoce todavía el origen exacto de las poblaciones afrodescendientes de Colombia como es el caso de SBP. En este trabajo nos planteamos la siguiente pregunta: ¿Los antepasados de SBP tienen un origen genético africano único o provienen de múltiples sitios de África?

2.11. Ancestría de la población colombiana: desde la colonización europea hasta los días de hoy

Aunque no es el tema central de este trabajo, es importante dedicar unos párrafos acerca de lo que fue la conquista española en Colombia dado que por este proceso este país presenta actualmente una etnicidad compleja. Los inmigrantes de la península ibérica fueron los primeros colonizadores no amerindios que arribaron a Colombia. Llegaron en el siglo XVI poco antes de que los africanos fueran traídos como esclavos. Esto hizo que se estableciera una población mezclada conocida como “*población criolla*” en la cual el componente amerindio se encontraba poco representado. El proceso de mezcla durante la primera parte de la conquista se dio exclusivamente entre varones europeos y mujeres nativas, al menos durante las tres primeras generaciones y fue disminuyendo a medida que los criollos tomaron su lugar en estas tierras. La expansión de esta población en Colombia estuvo fuertemente influenciada por la geografía, dado que el país presenta unas marcadas diferencias ecológicas: el Este, cubierto por la selva amazónica y las llanuras del Orinoco; el Oeste, dividido por tres cordilleras; la Costa pacífica, cubierta por densas selvas lluviosas; el litoral y las llanuras costeras del Caribe. Esta diversidad ecológica hace que Colombia sea uno de los países en el mundo con mayor diversidad biológica.

La composición racial de las poblaciones colombianas como ya se ha dicho es variable. En Colombia existen tres grupos étnicos bien definidos: nativos americanos, afrocolombianos, (donde se integran los raizales y palenqueros) y los gitanos. Cualquiera de ellos se identifica, ya sea por sus rasgos físicos y culturales, por su pertenencia a un

territorio o por hablar lengua nativa o debido a que ellos mismos reconocen pertenecer a cualquiera de estos grupos étnicos.

El resto de los colombianos son mezclados que representan la inmensa mayoría, pues en esta categoría se agrupa aproximadamente el 83% de los colombianos.

El aporte de ancestría caucásica es similar para las regiones andinas como el viejo Caldas, los Santanderes y la zona Cundiboyacense, pero dentro de cada una de estas regiones hay variaciones en la contribución de mezcla nativa americana y africana siendo el menor componente amerindio para la región de Caldas. La zona del Chocó presenta una alta ancestría africana mientras que las regiones de Tolima y Huila y la región Cundiboyacense presentan el mayor componente amerindio. El crecimiento demográfico de varias regiones del país llevó a un relativo aislamiento durante el siglo XIX. Además, después de la independencia, a principios del año 1800, a Colombia no había llegado la cantidad de inmigrantes que si se dio en otros países latinoamericanos. Así que fuera de las migraciones internas (hacia los centros urbanos), en Colombia existen muchas áreas que proveen un panorama de distribución poblacional establecido desde la época colonial.

Por ejemplo, la población de la costa caribe creció de la unión de un pequeño número de islotes del caribe conectados con la tierra firme. Los españoles que vinieron a tierras colombianas y se establecieron en lo que daría asiento a la ciudad de Cartagena, capital de Bolívar, no llegaron directamente de España, lo hicieron desde la isla Santo Domingo. En el año 1530 don Pedro de Heredia y otros españoles eran ya parte del vecindario de las islas que sirvió de puerto de entrada para las expediciones a los territorios de la América española. En la costa cartagenera entraron desde el 1506 y solo pudieron establecerse hasta 1533, después de una guerra brutal con los indígenas que la habitaron y la cultivaron por siglos (Borucki et al

2015; Múnera 2021). Durante todo el siglo XX, principalmente durante la segunda mitad, el territorio colombiano presenció una profunda transformación a partir de la urbanización y concentración de la mayoría de la población en las áreas urbanas y en especial en las grandes ciudades, proceso generado especialmente por el desarrollo

progresivo de redes de comunicación y transporte cada vez más eficientes, la integración de los mercados regionales, la conformación de un mercado nacional y los procesos de industrialización urbana. Según estadísticas del DANE 2018, Colombia pasó de ser un país rural a uno predominantemente urbano en aproximadamente 50 años. Actualmente tenemos una población campesina del 32% del total de la población colombiana. Dado que existen diferencias marcadas en los porcentajes ancestrales en las distintas regiones del país, descritas arriba, se han presentado dificultades en los análisis de identificación humana y pruebas de forenses en el país, estableciendo que es importante caracterizar bases de datos poblacionales para cada región. En este trabajo también se hace un aporte para la población de SBP. Seguidamente se describirán, para finalizar, algunos parámetros de interés forense y que se describirán en el capítulo IV y ANEXOS D y F.

Este trabajo de tesis doctoral aporta una visión general sobre la ancestralidad de una población afrodescendiente de Colombia, como es SBP. De este conocimiento se nutrirán las ciencias sociales, antropológicas y biomédicas en general y, en particular la genética de las poblaciones humanas.

3. OBJETIVOS

3.1. Objetivo general

Determinar el patrón de la herencia africana y sus orígenes en la población de San Basilio de Palenque, Bolívar, Colombia.

3.2. Objetivos específicos

1. Identificar la diversidad y el origen de los linajes parentales de Palenque, mediante la caracterización de los principales haplotipos y haplogrupos de origen africano, utilizando marcadores de tipo STR y SNP en cromosoma Y e región control del ADN mitocondrial (ADNmt).

2. Construir redes filogenéticas con el fin de verificar el grado de divergencia entre haplotipos e identificar un posible haplotipo ancestral, y determinación del origen de los principales linajes africanos existente en SBP.

3. Identificar los componentes principales de ancestría continental y subcontinental en SBP mediante el análisis integrado de marcadores uniparentales, marcadores informativos de ancestría (AIM) y secuencias del genoma completo (WGS).

4. Resolver la relación entre la diversidad lingüística y genética en Palenque, considerando la ancestría genética en el contexto de los grupos lingüísticos africanos y comparando los datos observados para otras poblaciones afrodescendientes en las Américas, con énfasis en la población del Chocó, de la Costa Pacífica de Colombia.

5. Investigar los eventos de aislamiento, efecto fundador y de mezcla no africana (europea y amerindia) en SBP.

6. Identificar y comparar patrones de diversidad de 38 marcadores HID-InDels, en SBP y otras poblaciones colombianas, de uso en identificación humana.

Para cumplir con los objetivos propuestos en esta investigación fue necesario centralizar toda la información en una página dinámica en Excel, donde fueron depositados todos los datos, desde el principio hasta el final, con el fin de alcanzar la trazabilidad en el proceso de la investigación.

4. METODOLOGÍA

Se describirá teniendo en cuenta el orden de los objetivos señalados encima y que va acorde con la descrita en los artículos científicos resultantes de este trabajo.

4.1. Obtención de linajes paternos y maternos con marcadores uniparentales (Objetivos 1 y 2)

4.1.1. Muestra poblacional de SBP

Se seleccionaron 200 muestras de individuos afrodescendientes entre 5 y 18 años, de San Basilio de Palenque. Estas muestras fueron recolectadas dentro del marco del proyecto de investigación “Epidemiología de las enfermedades alérgicas y factores de riesgo asociados en la población de san Basilio de palenque”, financiado por la fundación del Banco de la República y aprobado al profesor Javier Marrugo del Instituto de Investigaciones Inmunológicas de la universidad de Cartagena. A las personas participantes se les extrajo sangre periférica por punción venosa en tubos con EDTA para la extracción del ADN. Las muestras fueron recolectadas bajo consentimiento informado por escrito de los tutores de los participantes incluidos en el estudio (Anexo A). Con base en la información genealógica, solo se seleccionaron individuos no relacionados (que no compartían abuelos) durante al menos tres generaciones. El proyecto y consentimiento informado fueron aprobados por la Ley Núm. 40 del comité de ética de la Universidad de Cartagena, Colombia; y se siguieron los principios éticos de la Declaración de Helsinki de 2000 de la Asociación Médica Mundial (<http://www.uma.net/e/policy/b3.htm>). La recolección de muestras y el consentimiento informado se realizaron bajo la guía de Regina Miranda y los guardias indios de los Zenúes (Palenque, Colombia).

Para los análisis de marcadores en cromosoma Y y de secuenciación de la región control de ADNmt se seleccionó una submuestra de un total de 95 niños varones (de entre 5 y 18 años).

Los voluntarios se identificaron como naturales de Palenque durante al menos 3 generaciones (todos los padres y abuelos nacidos en Palenque). Un total de 48 niños fueron reclutados en la escuela Rural Benkos Biojó en Palenque (PR), y 47 niños residían en el área urbana de la ciudad de Cartagena (PU). Los niños de la PU fueron reclutados en escuelas que participaron en un programa de etnoeducación creado para preservar aspectos de la cultura palenquera como la danza, la música, la religión y especialmente el idioma Palenquero (Schwegler, 2016). La tipificación de los marcadores InDel se realizó en la totalidad de las muestras, obteniéndose resultados de 187 para AIM-InDels y de 114 para HID-InDels.

4.1.2. Extracción del ADN para muestras de San Basilio de Palenque, Bolívar, Colombia

El ADN fue extraído por el método de *Salting – Out* descrito por Miller et al. (1988), a partir del *buffy coat* separado en tubos *ependorf* de 1,5 ml, después de centrifugación y obteniendo un volumen aproximadamente de 500 μ l. Se cuantificó por espectrofotometría utilizando el NanoDrop modelo 2000 (Thermo Scientific), siendo posteriormente mantenido en la concentración patrón de 10 ng/ μ l. Los ADN fueron conservados en freezer de -70°C hasta su uso. Mediante esta metodología se extrajo el ADN de las muestras que se usaron para el análisis de marcadores en cromosoma Y e ADNmt, así como también las muestras a las que se les hizo la secuenciación de genoma completo.

4.1.3. Genotipificación de marcadores del tipo microsatélite del cromosoma Y (Y-STR)

Este análisis se hizo mediante la técnica de PCR/electroforesis capilar. Se realizó una PCR múltiple usando el estuche comercial de amplificación Yfiler™ Plus (Thermo Fisher Scientific, Waltham, MA, EE. UU.), que permite detectar 27 STRs. Los productos amplificados se separaron mediante electroforesis capilar en secuenciador automático ABI 3500 (Applied Biosystems). Los resultados fueron analizados mediante el programa GeneMapper ® software v4.0. (Thermo Fisher

4.1.4. Genotipificación de SNPs del cromosoma Y

La tipificación de los Y-SNPs se realizó mediante PCR multiple y análisis de SBE (single base extensión) utilizando el estuche comercial de SNaPshot (Thermo Fisher Scientific

Inc.), como se describe en Gomes et al. (2010), Noguera et al. (2014) y Resque et al. (2016). Brevemente, fueron elegidos un grupo de 51 Y-SNPs (Figura 15) para discriminar los principales haplogrupos africanos que se esperaba encontrar en San Basilio de Palenque, así como los haplogrupos característicos de poblaciones europeas y nativo-americanas.

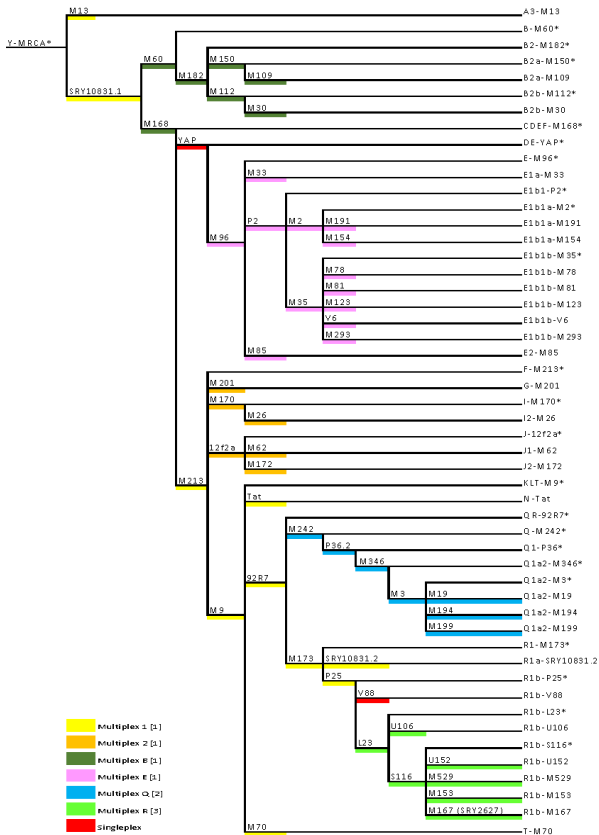


Figura 15. Árbol de Y-SNPs analizados en este estudio y métodos utilizados.

Nota. Los haplogrupos se nombran de acuerdo con Van Oven et al. (2014). Los Y-SNP se eligieron para discriminar los principales haplogrupos africanos que se esperaba encontrar en San Basilio de Palenque, así como los resultantes de la mezcla de europeos y nativos americanos. [1] Gomes et al. (2010); [2] Noguera et al. (2014); [3] Resque et al. (2016).

Para la detección de los productos resultantes de la mini secuenciación se utilizó el secuenciador automático ABI 3500 (Applied Biosystems, EE. UU.). El análisis de los electroferogramas se realizó con la ayuda del software GeneMapper V.3.2 (Applied Biosystems). Los haplogrupos se nombraron de acuerdo con Van Oven et al. (2014).

Se realizó un análisis jerárquico preliminar utilizando los marcadores de tipo Alu (YAP) y M269. El primero se utilizó para determinar la presencia de la inserción de 300pb (YAP +), que se presenta con frecuencia de aproximadamente 80% en la mayoría de las poblaciones del África subsahariana, o su ausencia (YAP -).

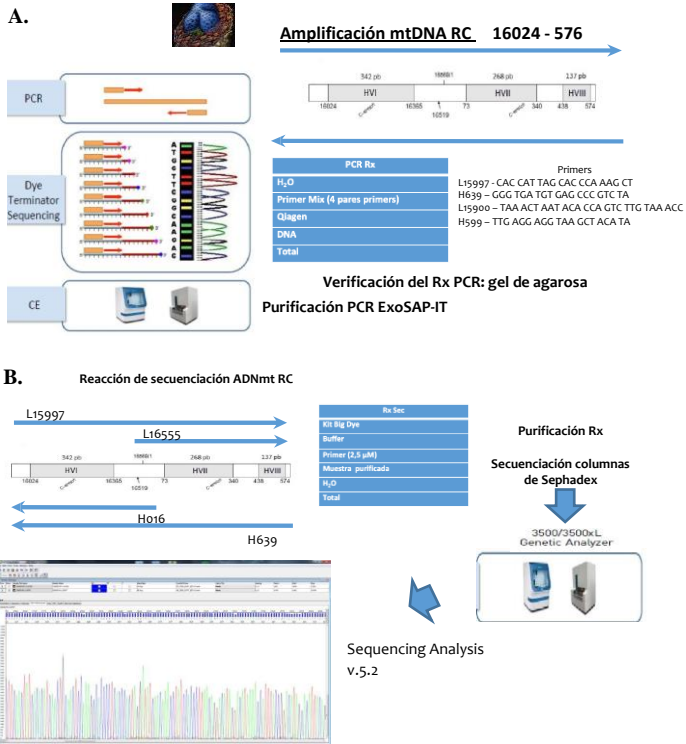
4.1.5. Secuenciación de la región control del ADN mitocondrial

La secuenciación de la región control (los fragmentos entre las posiciones 16024 y 576) se realizó siguiendo las instrucciones de la sociedad internacional de genética forense (ISFG), descritas previamente (Parson et al., 2004).

Brevemente, se realizó una reacción de PCR utilizando los primers descritos en la Tabla 1. Los productos amplificados se purificaron con enzima ExoSap-IT y luego se preparó la reacción de secuenciación con el estuche comercial BigDye1 Terminator v.3.1 Cycle Sequence (AB) usando termociclador PCR 9700. Se purificaron las reacciones mediante columnas de sephadex para luego ser leídas en el secuenciador genético ABI 3500 resuspendida en formamida. Se analizaron los resultados brutos mediante el software de análisis de secuenciación v.5.2 y SeqScape v.3 (Thermo Fisher Scientific). Figura 16.

Tabla 1. Lista de primers utilizados para análisis de ADNmt.

Primer	Secuencia
L15900	5'-TAAACTAATACACCAGTCTTGTAACC-3'
L19997	5'-CACCATAGCACCCAAAGCT-3'
L16268	5'-CACTAGGATACCAACAAACC-3'
L16555	5'-CCCACACGTTCCCTTAAAT-3'
L00314	5'-CCGCTTCTGGCCACAGCACT-3'
L00369	5'-CCCTAACACCAGCCTAACCA-3'
H00016	5'-CCCGTGAGTGGTTAATAGGGT-3'
H00159	5'-AAATAATAGGATGAGGCAGGAATC-3'
H00388	5'-TGGTTAGGCTGGTGTAGGG-3'
H00599	5'-TTGAGGAGGTAAGCTACATA-3'
H00639	5'-GGGTGATGTGAGCCCGTCTA-3'

**Figura 16.** Resumen de la metodología para ADNmt:

A) Reacción de PCR. B) Secuenciación

4.1.6. Construcción de redes filogenéticas para la identificación y determinación del origen de los principales linajes africanos existente en SBP

Las redes filogenéticas se construyeron utilizando el software Network v10.1.0.0 (<http://www.fluxusengineering.com>). El número de Y-STRs utilizados para construir las redes dependía del conjunto común disponible para maximizar la representación de las poblaciones africanas para cada haplogrupo.

En la mayoría de los casos, usamos un conjunto de 11 loci, a saber, DYS389I, DYS389II, DYS19, DYS390, DYS438, DYS392, DYS437, DYS385a / b, DYS393 y DYS439. Sin embargo, para los clados Y-MRCA* (xM13, SRY10831.1) y E1b1b-M35, también se incluyó DYS391, ya que fue genotipado en la mayoría de los estudios que informaron muestras de estos haplogrupos.

4.1.7. Análisis estadístico

Las frecuencias de haplogrupos de cromosoma Y, y ADNmt se calcularon mediante el método de conteo directo. La diversidad genética de haplotipos y haplogrupos, las distancias genéticas por pares de poblaciones y la probabilidad de no diferenciación se calcularon utilizando el programa Arlequin ver. 3.5.1.2 (Excoffier & Lischer, 2010).

Adicionalmente, las distancias genéticas por pares de poblaciones se visualizaron mediante la herramienta de escalamiento multidimensional (MDS), utilizando el programa STATISTICA ver.8.0 (www.statsoft.com).

Los haplotipos de ADNmt se clasificaron utilizando el software SeqScape v2.7 (Thermo

Fisher Scientific). Los haplogrupos se asignaron utilizando la base de datos EMPOP v4 / R1214 y la herramienta Haplogrep y se confirmaron en Phylotree (Van Oven & Kayser, 2009; Weissensteiner et al., 2016)

Los datos se enviaron a la base de datos EMPOP para control de calidad y están disponibles para fines de investigación, con el número de acceso EMP00749. Las secuencias de ADN mitocondrial se depositaron en GenBank: PopSet 1782793150

(<https://www.ncbi.nlm.nih.gov/popset/?term=1782793150>), números de acceso: MK930265MK930345.

4.2. Inferencia de la ancestría continental y subcontinental en SBP y análisis de mezcla sesgada por sexo (objetivos 3, 4 y 5)

4.2.1. Muestra poblacional de SBP y Chocó

Para los análisis de la secuenciación del genoma completo se utilizaron 34 muestras de individuos adultos no relacionados y que habitaban en el área rural de SBP y no compartían abuelos. La extracción del ADN se realizó como se describió previamente para la muestra de niños (4.1.2. Extracción del ADN para muestras de San Basilio de Palenque, Bolívar, Colombia). Todas las personas firmaron consentimiento informado.

La Universidad Tecnológica del Chocó inscribió a donantes de muestra de Chocó, Colombia, como parte del proyecto de investigación ChocoGen (Chocó, 2019; Medina-Rivas et al., 2016). Se seleccionaron donantes de ChocoGen y sus muestras se caracterizaron utilizando matrices de SNPs de todo el genoma como se describió anteriormente (Conley et al., 2017). Todos los donantes firmaron documentos de consentimiento informado que indicaban su comprensión de los riesgos potenciales del proyecto, junto con cómo se manejarían sus datos y cómo se protegería su identidad.

Los datos del genotipo de todo el genoma de ChocoGen están disponibles para fines de investigación, por el Comité de Acceso a Datos de ChocoGen. La aprobación de la junta de revisión institucional para la recolección, genotipificación y análisis comparativo de muestras de ADN humano se obtuvo de la Universidad de Cartagena (Colombia), la Universidad Tecnológica del Chocó (Colombia) y el Instituto de Tecnología de Georgia (Estados Unidos).

4.2.2. Secuenciación de genoma completo

La secuenciación del genoma completo de las 34 muestras de Palenque se realizó a una cobertura de 30x, con plataforma de Illumina HiSeq 2000, como parte del proyecto Consorcio sobre el asma entre poblaciones de ancestría africana en las Américas (CAAPA, 2011), y las

variantes se denominaron como se describió anteriormente (Mathias et al., 2016).

Los AIMS se caracterizaron para una muestra separada de 187 donantes de Palenque Rural como se describió anteriormente (Martínez et al., 2013; Martínez et al., 2017; Pereira et al., 2012).

4.2.3. Análisis de marcadores informativos de ancestría (AIMs)

4.2.3.1. Genotipificación de marcadores de ancestría tipo InDels (AIM-InDels)

Un panel de 46 AIM-InDels se genotipificó mediante PCR seguido por electroforesis capilar, según el protocolo descrito por Pereira et al. (2012). Se realizaron dos reacciones de PCR multiplex utilizando la plataforma FMBIO Iie (Hitachi Co). Los fragmentos fueron separados y detectados usando un ABI 3500 Genetic Analyzer (Life Technologies) y genotipados utilizando el programa GeneMapper v.4.1 (Life Tecnologías). La nomenclatura de alelos fue de acuerdo al artículo de Pereira et al. (2012). Descrito en el Anexo E y en Martínez et al. (2013); Martínez et al. (2017); Pereira et al. (2012).

Las fracciones de ancestría continental se infirieron utilizando el programa

ADMIXTURE, para las muestras de SBP y de poblaciones de referencia de África, América y Europa del *Human Genome Diversity Project* (HGDP) (Tabla complementaria 1). Se realizó un análisis de componentes principales (PCA) para determinar la relación entre las muestras de SBP y las poblaciones de referencia globales del HGDP. Los niveles de diversidad genética para SBP y las poblaciones de referencia globales se midieron utilizando F_{IS} calculado para los AIMS.

4.2.4. Análisis de marcadores uniparentales

Los haplotipos uniparentales se caracterizaron a partir de las variantes obtenidas del WGS (*GWS variant calls*) utilizando los programas HaploGrep 251 (ADNmt) y SNAPPY (YDNA) (Severson et al., 2018). Las variantes de llamado (*call variants*) se pre procesaron para el cambio de hebra de alelos alternativos / de referencia, según fuera necesario, antes de la caracterización del haplotipo. Los orígenes

de los grupos de ancestría continental (africanos, europeos y nativos americanos) para los haplotipos de ADNmt y ChrY fueron caracterizados por WGS, mediante la comparación con las muestras de referencia del proyecto de los 1000 genomas (1KGP) y el proyecto de diversidad del genoma Humano HGDP, como se describió previamente (Rishishwar & Jordan, 2017).

Un conjunto adicional de haplotipos, 151 para ADNmt y 150 para cromosoma Y, previamente caracterizados por Ansari-Pour et al. (2016), se volvieron a analizar mediante la comparación con poblaciones de referencia africanas (Pour et al., 2013; Thomas et al., 2000; Veeramah et al., 2010). Para este segundo conjunto de datos, los orígenes de los grupos de población continentales se caracterizaron para los haplotipos de ADNmt y ChrY, pero solo los haplotipos de cromosoma Y de este conjunto proporcionaron una resolución suficiente para el análisis de ancestría africana subcontinental.

Los orígenes africanos subcontinentales de los haplotipos de ChrY se caracterizaron midiendo la fracción de haplotipos africanos bantú y no bantú y comparando las frecuencias de los haplogrupos de ChrY africanos entre poblaciones, utilizando valores de F_{ST} por pares.

4.3. Muestras de referencia global para los estudios de GWS

Para los estudios de WGS, se recopilaron genotipos de todo el genoma (datos de matriz de SNPs) y marcadores de ancestría AIMs de poblaciones de referencia africanas, europeas y nativas americanas, de fuentes disponibles públicamente, incluido el Proyecto 1000 Genomas (1KGP), el Proyecto de Diversidad del Genoma Humano (HGDP) y el Proyecto *Simons Genome Diversity Project* (Consortium, 2015; Li et al., 2008), el programa PAGE (Mallick et al., 2016; Wojcik et al., 2019) (Figura 17).

Las variantes de WGS (*WGS variant calls*) para poblaciones mezcladas de estadounidenses y afrodescendientes de las Américas se tomaron del proyecto 1KGP (Consortium, 2015). También se tomaron los datos de muestras de referencia de genotipos adicionales de todo el genoma de los estudios a gran escala sobre ancestría africana (Patin et al., 2017) y de ancestría nativa americana (Reich et al., 2012). Las muestras afrocolombianas, afrodescendientes y estadounidenses

mezclas se fusionaron con las poblaciones de referencia globales como se describe a continuación.

Se caracterizaron 584 muestras de referencia de 40 poblaciones del proyecto HGDP para el análisis de AIMs. Una total de 2,328 haplotipos de ADNmt y 2,649 haplotipos de cromosoma Y pertenecientes a 42 y 38 poblaciones africanas de referencia, respectivamente, de varios estudios previos (Pour et al., 2013; Thomas et al., 2000; Veeramah et al., 2010). Los detalles sobre todas las poblaciones de referencia analizadas aquí se proporcionan en la Tabla complementaria 1.

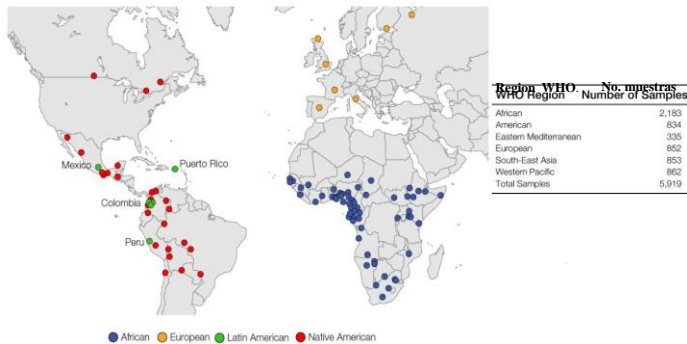


Figura 17. Lugares de muestreo y grupos de ascendencia continental de las poblaciones de referencia global, codificadas por colores.

Nota. El mapa que se muestra aquí se obtuvo del mapa (contributors, 2020), que se distribuye en el dominio público sin restricciones de uso bajo “Creative Commons CC0, opción sin derechos de autor reservados”. REGION WHO: Los Estados Miembros de la OMS se agrupan en 6 regiones de la OMS: Región de África, Región de las Américas, Región de Asia Sudoriental, Región de Europa, Región del Mediterráneo Oriental y Región del Pacífico Occidental.

4.4. Control de calidad y armonización de SNPs

Los genotipos WGS y de todo el genoma recopilados de diferentes fuentes se armonizaron utilizando el programa PLINK versión 1.9 (Chang et al., 2015), junto con un script personalizado. Las variantes genómicas se fusionaron identificando variantes comunes a todos los conjuntos de datos, mientras se corregían los cambios de cadena y las inconsistencias con los identificadores de variantes. Se

eliminaron las variantes fusionadas con $> 1\%$ de ausencia y $< 1\%$ de frecuencia de alelos menores entre las muestras.

Los datos de WGS y del genotipo de todo el genoma se fusionaron en dos conjuntos de variantes separados para evitar la pérdida de sitios genómicos para el análisis posterior, sí: (1) un conjunto de variantes de ancestría africana de 37 poblaciones y 1235 muestras) y (2) un conjunto de ancestría nativa americana de 33 poblaciones y 314 muestras.

Los conjuntos de variantes específicas de ancestría se superponen ampliamente para las muestras de WGS y difieren principalmente con respecto a los datos del genoma de todo el genoma africano y nativo americano (Patin et al., 2017; Reich et al., 2012). Ambos conjuntos de variantes específicas de ancestría incluyen todos los afrocolombianos, otros afrodescendientes y otras muestras estadounidenses mezcladas.

Los dos conjuntos de datos de variantes armonizados se escalonaron utilizando el software ShapeIT versión 2.r837 (Delaneau et al., 2013), con cromosomas individuales escalonados por separado. El conjunto de datos de variantes en fases finales que cubrieron sitios genómicos incluyó 690,739 africanos y 364,470 nativos americanos. Los conjuntos de datos armonizados se analizaron para desequilibrio de ligamiento (LD), usando PLINK con la configuración `-indep -pairwise 50 10 0.1`. Los conjuntos de datos de variantes finales escalonados y reducidos por LD de sitios genómicos incluyeron 275.598 africanos y 287.846 nativos americanos.

4.5. Inferencia de ancestría genética

Se utilizaron los conjuntos de datos de variantes armonizados de todo el genoma para calcular las distancias genéticas por pares utilizando el programa PLINK con la función `-distance`, y se realizaron análisis de componentes principales (PCA) utilizando la función PLINK `-PCA`.

El programa ADMIXTURE versión 1.3.0 se ejecutó en los datos de las variantes podados por LD en modo no supervisado con un $K = 3$, para inferir fracciones de ancestría continental para poblaciones afrodescendientes en las Américas utilizando las muestras de referencia

europeas, africanas y nativas americanas (Alexander et al., 2009). Se utilizó una versión modificada del programa RFMix para inferir la ancestría continental local (Maples et al., 2013). El RFMix se ejecutó durante 12 generaciones utilizando el modo PopPhased con un tamaño de nodo mínimo de 5. Se ejecutaron dos rondas de maximización de expectativas (EM) con la opción “*use-reference-panels-in-EM*” para generar llamadas de ancestría local para muestras de población afrodescendiente y de referencia. El PCA se volvió a ejecutar en datos enmascarados de RFMix para realizar el PCA específico de ancestría africana. Se utilizaron haplotipos específicos de ancestría de la salida de RFMix para caracterizar la ancestría con sesgo sexual como se describió anteriormente (Jordan et al., 2019; Rishishwar et al., 2015).

El programa Chrompainter versión 2 se usó para inferir la ancestría subcontinental mezclada de los genomas americanos, utilizando haplotipos específicos de ancestría en comparación con una muestra de referencia específica del grupo ancestral (Lawson et al., 2012).

El Chrompainter se ejecutó con los datos genómicos enmascarados por RFMix, para caracterizar cada componente de ancestría continental de genomas mezclados (africanos, europeos y nativos americanos) por separado. Se utilizó mínimos cuadrados no negativos (NNLS) para convertir los vectores de pintura de salida de Chrompainter en estimaciones de porcentaje de ancestría, utilizando el paquete R nls versión 1.4 (Mullen & van Stokkum, 2015). Las pruebas F3 se realizaron con el programa qp3pop versión 650 (Patterson et al., 2012). Los valores del índice de fijación por pares (F_{ST}) entre las poblaciones nativo-americanas se calcularon por RFMix, usando el paquete EIGENSOFT versión 7.2.1 para determinar los haplotipos (Patterson et al. 2006). Los valores de F_{ST} entre pares de poblaciones fueron usados para la construcción del árbol filogenético (neighbor joining phylogeny) usando el programa MEGA versión 10.0. (Saitou & Nei, 1987; Tamura et al., 2013).

4.6. Relación entre la diversidad lingüística y genética en Palenque

Los niveles de diversidad genética para muestras afrodescendientes y de referencia se midieron utilizando el coeficiente

de consanguinidad (FIS) de Wright. Las corridas de homocigocidad (FROH) se calcularon con el conjunto de datos de las variantes africanas finales, utilizando el programa PINK versión 1.9 con las funciones heterocigota y homocigota, respectivamente. El coeficiente de FIS se cuantificó como la diferencia normalizada entre los recuentos observados frente a los esperados de sitios autosómicos homocigóticos. FROH se cuantificó como la fracción del genoma en las corridas de homocigocidad (ROH) > 1,5 kb. El coeficiente de FIS también se calculó para los datos de los 46 AIM-InDels.

4.7. Caracterización de marcadores HID-InDels y comparación con otras poblaciones mezcladas y nativas de Colombia (objetivo 6).

4.7.1. Muestreo

Se tomó una submuestra de 114 individuos de SBP no emparentados, como se describe encima, para tipificar 38 HID-InDels de uso en análisis forense (Tabla 2). Con estas muestras se hicieron las comparaciones con otras poblaciones colombianas. En la Figura 18 se puede ver el total de muestras comparadas, distribuidas por las 5 regiones naturales de Colombia. Información más detallada sobre el muestreo está contenida en los mapas de la Figura suplementaria 1. Todas las muestras del estudio fueron obtenidas de individuos sanos y no relacionadas. Se recogieron muestras de sangre mediante consentimiento informado, por escrito. El estudio cumple con los principios éticos de la Declaración de Helsinki de 2000 de la Asociación Médica Mundial y, junto con el consentimiento informado, fue aprobado por los Comités de Ética de la Pontificia Universidad Javeriana, de la Universidad Industrial de Santander y de la Universidad de Cartagena, Colombia.

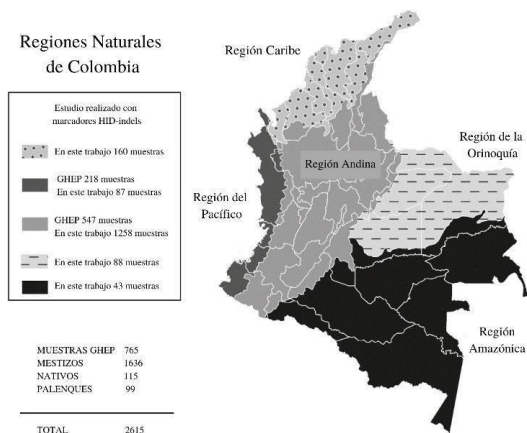


Figura 18. Número total de muestras analizadas de cada región natural colombiana, incluyendo las muestras de este estudio, las de Pereira et al. (2018) y de Martínez et al. (2013) (Palenque).

4.7.2. Genotipado de los 38 HID-InDels

Se caracterizaron los 38 HID-InDels en un único multiplex de PCR, siguiendo el protocolo descrito por Pereira et al. (2009), tras ajustar el volumen total de la reacción a 5 μ L.

Para la amplificación de cada uno de los ADN aislados se utilizó 2,5 μ l de master mix del kit Qiagen, 0,5 μ l de primer mix (todos los primers a una concentración de 2 μ M), 1 μ l de ADN a una concentración final de 2,5 ng/ μ l, para un volumen final de reacción de 5 μ l. La PCR se llevó a cabo en un termociclador Veriti™ 96 Well Fast Thermo Cycler (Applied Biosystems), con las siguientes condiciones de termociclaje: un paso inicial de 15 min a 95°C, seguido de 27 ciclos a 94°C durante 30 s, 60°C durante 1,5 min, 72°C durante 45 s y una extensión final a 72°C durante 60 min.

Los fragmentos amplificados por PCR, marcados con un fluorocromo, se separaron por electroforesis capilar y se detectaron usando un analizador genético Applied Biosystems® serie 3500 (Thermo Fisher Scientific Inc., Waltham, Massachusetts, EE. UU).

En un vial de 2 ml se prepararon 900 μ l de una mezcla de 880 μ l de formamida Hi-Di™ y 20 μ l del marcador de peso molecular GeneScan™ LIZ 500. Se depositaron 9 microlitros en un plato de 96

pozos, de acuerdo con el número de muestras amplificadas a correr en el ABI 3500. Se agregó 1,0 μ l de cada ADN amplificado a su respectivo pozo y se colocó en el secuenciador ABI 3500. En el secuenciador ABI 3500, la separación de los fragmentos amplificados se realizó con polímero POP-7TM (Applied Biosystems), y se usó la matriz estándar DS-33 y un filtro G-5. Una vez finalizada la electroforesis capilar, se analizaron los resultados con el programa GeneMapper v4.0. En este programa se depuran los datos obtenidos, se imprimen los respectivos electroferogramas y los genotipos de cada individuo han sido agrupados en una matriz compatible con el programa STRUCTURE.

4.7.3. Análisis estadístico

Los cálculos de las frecuencias alélicas de los 38 HID-InDels, así como el análisis de diversidad genética (Heterocigosidad esperada = H_e), distancias genéticas (F_{ST}) y el test del equilibrio de Hardy-Weinberg, fueron realizados a través del software Arlequín v3.7 (Excoffier & Lischer, 2010).

Para obtener una representación gráfica de las afinidades entre las distancias (F_{ST}) de las diferentes poblaciones fue realizado un análisis de Escalamiento Multidimensional (MSD) de las distancias génicas (F_{ST}) entre las diferentes poblaciones, utilizando el programa STATISTICA v8.0. (Statsoft, Tulsa, Oklahoma; <http://www.statsoft.com/>).

Para los cálculos estadísticos de interés forense fueron utilizadas las bases de datos de frecuencias alélicas obtenidas a partir de la tipificación con el sistema Indelplex HID. Las frecuencias alélicas de las poblaciones analizadas fueron utilizadas para calcular los parámetros de informatividad forense: poder de discriminación (PD), poder de exclusión (PE), la probabilidad de coincidencia (MP: del inglés Maching Probability) y el Índice de paternidad típico (TPI), usando los programas PowerStats v1.2 (Promega Corporation, Madison, WI. USA) y Familias, versión 3.2.5 (Kling et al., 2014).

Tabla 2. Listado de marcadores *HID-InDels* utilizados en esta investigación y diseñados para ser usados en identificación humana.

	Código rs	Cromosoma	Posición (bp)	Código y posición (bp)	Alelos	Tamaño esperado de amplificación (S-L)
1	rs3047269	1	161077452	NT_004487.18 pos. 13301183	-/CTGA	126-130
2	rs2307579	1	245878706	NT_004836.17 pos. 12569872	-/ATG	104-107
3	rs16624	2	234681130	NT_005120.15 pos. 949145	-/GT	65-67
4	rs2308242	3	8591709	NT_022517.17 pos. 8556709	-/CT	106-108
5	rs2308026	4	119404855	NT_016354.18 pos. 43733552	-/CA	83-85
6	rs2307526	5	5178112	NT_006576.15 pos. 5115112	-/ACAC	95-99
7	rs1160956	5	65414216	NT_006713.14 pos. 15972818	-/AGA	128-131
8	rs1610871	5	171020572	NT_023133.12 pos. 15897552	-/TAGG	61-65
9	rs2307710	6	47929222	NT_007592.14 pos. 38679494	-/AGGA	92-96
10	rs2307839	6	117200251	NT_025741.14 pos. 21262987	-/GA	152-154
11	rs2308137	6	149655891	NT_025741.14 pos. 53718627	-/GA	61-63
12	rs2307978	7	83121850	NT_007933.14 pos. 8518190	-/GA	156-158
13	rs35769550	8	76681235	NT_008183.18 pos. 28372034	-/TGAC	89-93
14	rs5895447	8	138489776	NT_008046.15 pos. 51638773	-/CA	128-130
15	rs16402	9	38396788	NT_008413.17 pos. 38396788	-/TTAT	150-154
16	rs2067294	9	70504241	NT_023935.17 pos. 478953	-/CTT	80-83
17	rs2307580	9	104626014	NT_008470.18 pos. 12907398	-/AATT	120-124
18	rs140809	10	6027167	NT_077569.2 pos. 350057	-/CAA	115-118
19	rs1160886	10	54112392	NT_008583.16 pos. 2993541	-/ACT	75-78

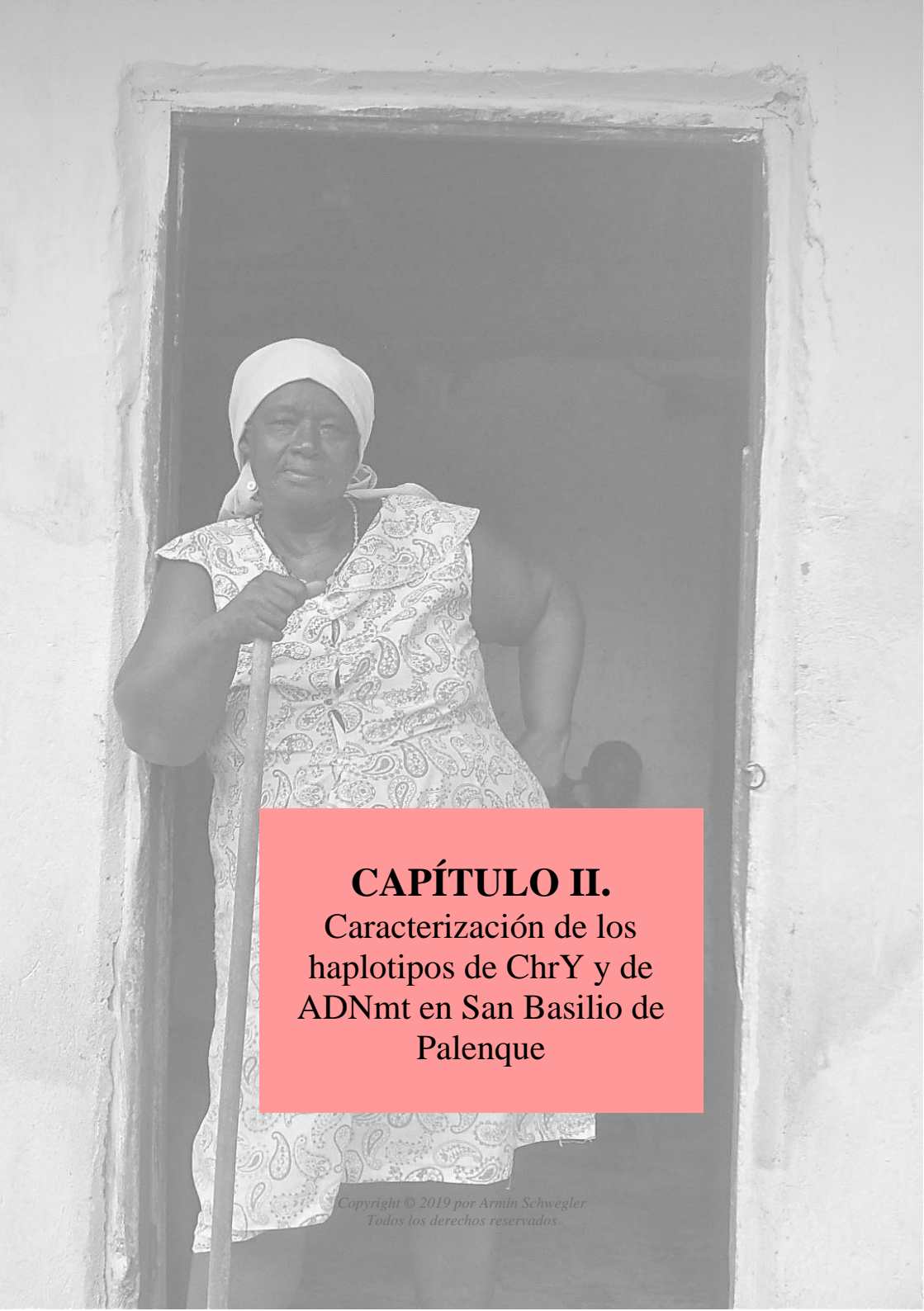
20	rs10688868	11	258180	NT_035113.6 pos. 208180	-/CT	81-83
21	rs34811743	11	30134266	NT_009237.17 pos. 28964931	-/TG	108-110

Continuación Tabla 2. Listado de marcadores *HID-InDels* utilizados en esta investigación y diseñados para ser usados en identificación humana.

	Código rs	Cromosoma	Posición (bp)	Código y posición (bp)	Alelos	Tamaño esperado de amplificación (S-L)
22	rs33972805	11	125794082	NT_033899.7 pos. 29851288	-/CT	135-137
23	rs1610919	12	14801263	NT_009714.16 pos. 7668970	-/AT	142-144
24	rs2067238	12	113772931	NT_009775.16 pos. 5858057	-/GCT	71-74
25	rs2308171	13	43778155	NT_024524.13 pos. 25860155	-/TCTG	135-139
26	rs2308189	14	28106508	NT_026437.11 pos. 10036508	-/AACTA	119-124
27	rs2308020	15	51268809	NT_010194.16 pos. 24272074	-/TT	127-129
28	rs2067208	16	83139788	NT_010498.15 pos. 38196486	-/GCCAG	93-98
29	rs3051300	17	10076666	NT_010718.15 pos. 9733290	-/GTAT	63-67
30	rs3080855	18	21507205	NT_010966.13 pos. 4742309	-/AATT	133-137
31	rs34511541	18	34677042	NT_010966.13 pos. 17912146	-/CTCTT	143-148
32	rs36040336	19	1353662	NT_011255.14 pos. 1342662	-/AT	65-67
33	rs2307689	19	48896180	NT_011109.15 pos. 16472558	-/TTC	74-77
34	rs33917182	20	11643625	NT_011387.8 pos. 11635625	-/CA	142-144
35	rs34541393	20	30165066	NT_028392.5 pos. 897497	-/AACT	57-61
36	rs35605984	21	14556736	NT_011512.10 pos. 1296736	-/TAAAG	151-156
37	rs10629077	21	30294208	NT_011512.10 pos. 17034208	-/AT	74-76

38	rs2307700	22	25120901	NT_011520.11 pos. 6181470	-/TCAC	101-105
----	-----------	----	----------	---------------------------	--------	---------

Nota: Código rs: Lista de marcadores según dbSNP; S = alelo corto y L= alelo lar



CAPÍTULO II.
Caracterización de los
haplotipos de ChrY y de
ADNmt en San Basilio de
Palenque

5. CARACTERIZACIÓN DE LOS HAPLOTIPOS DE ChrY Y DE ADNmt EN SBP

Los resultados presentados en este capítulo se encuentran publicados, ver ANEXO B: (Martínez et al., 2020) y se describirán a continuación.

Las hipótesis planteadas para este capítulo fueron:

- El origen de los linajes parentales de Palenque proviene de sitios distintos de África
- El origen continental de los linajes maternos (ADNmt) proviene de sitios distintos de Africa Occidental
- La combinación de los análisis de distancia genética por pares y redes filogenéticas son útiles para delinear el origen de los haplogrupos fundadores en SBP
- La deriva genética contribuye a la separación de Palenque de sus poblaciones de origen
- El sustrato de la lengua palenquera tiene sus raíces no solamente del kikongo

Para la caracterización de haplogrupos de ChrY, y de ADNmt con el fin de usar el mayor número posible de individuos, las dos submuestras de Palenque (PR y PU) se compararon mediante distancias genéticas F_{ST} y probabilidad de no diferenciación después de 10100 permutaciones. Para los haplotipos de ADNmt no se encontró una diferencia estadísticamente significativa ($F_{ST} = -0.018$; $P = 0.9855 \pm 0.001$) y para el cromosoma Y, la submuestra PR mostró una menor proporción de haplogrupos Y-SNP africanos que la PU (54,2% y 68,1%, respectivamente). Sin embargo, las diferencias entre las submuestras no fueron estadísticamente significativas para las distribuciones de haplotipos Y-STR ($F_{ST} = 0.0041$; $P = 0.1119 \pm 0.0014$) o para haplogrupos Y-SNP ($F_{ST} = 0.0163$; $P = 0.0792 \pm 0.0012$). Sobre la base de estos resultados, las muestras se combinaron para los análisis subsiguientes.

5.1. Diversidad genética en SBP

Los haplotipos de ADNmt y cromosoma Y encontrados en este estudio se describen en las tablas complementarias 2 y 3) junto con las clasificaciones de haplogrupos correspondientes. Con respecto a los haplotipos de ADNmt, se observaron diferencias en las clasificaciones de las bases de datos EMPOP y Haplogrep en una sola muestra (PU063), que fue clasificada como L2a1 + 16189 + (16192) por Haplogrep y L2a1 por EMPOP. La clasificación de EMPOP está respaldada por Phylotree, ya que PU063 carece de 16189C y tiene una heteroplasmia en la posición 16192. Con respecto al ADNmt, se encontró una alta diversidad de haplogrupos (0.8895 ± 0.0193) para un total de 22 haplogrupos diferentes presentes en nuestra muestra. Sin embargo, esta alta diversidad de haplogrupos está asociada con una baja diversidad de haplotipos ($0,9225 \pm 0,0162$). Una gran proporción de haplotipos se compartió dentro de los haplogrupos, y solo 33 haplotipos diferentes estaban presentes en la muestra estudiada. Se encontró una amplia separación de los diferentes haplogrupos africanos, cada uno de los cuales está representado por pocos haplotipos, como se ilustra en el diagrama de red de la Figura 19.

La gran mayoría de las muestras (91%) pertenecen al macrohaplogrupo L africano (Tabla 3). El 9% restante pertenece a haplogrupos nativos americanos (A2, A2af1a1, A2a1, B2d y C1c3). No se observaron linajes maternos europeos.

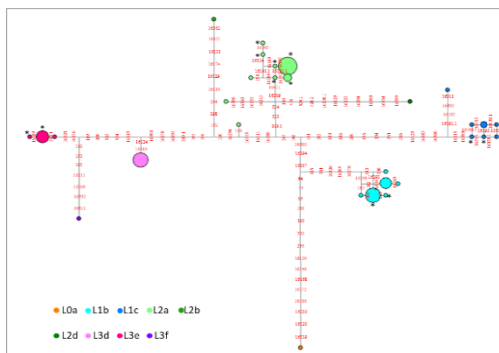


Figura 19. Representación en red de los haplotipos CR completos (incluidos los polimorfismos InDel) dentro de los haplogrupos africanos.

Nota. Las siete muestras de haplogrupos nativos americanos, que representan 5 haplogrupos / haplotipos diferentes, no se incluyeron en la red. Las posiciones polimórficas que separan los haplotipos se describen en rojo. * haplotipos que incluyen una secuencia alternativa, debido a heteroplasmica. Las dos posibles variantes para PR024, PR052, PR081, PU055, PU063 y PU103 (ver haplotipos en la tabla S2) se incluyeron en la red.

Tabla 3. Frecuencias de los haplogrupos del ADNmt y del ChrY detectados en una muestra de la población de Palenque.

ADNmt			Y-SNPs		
Haplogrupo	n	%	Haplogrupo	n	%
L0a1a+200	1	1.23	Y-MRCA*(xM13, SRY10831.1)	3	3.16
L1b1a+189	10	12.35	B2a-M150* (xM109)	3	3.16
L1b1a1'4	7	8.64	E1a-M33	2	2.11
L1b1a18	1	1.23	E1b1a-M191	7	7.37
L1b1a7a	1	1.23	E1b1a-M2* (xM154, M191)	36	37.89
L1c3	1	1.23	E1b1b-M35* (xM78, M81, M123, V6, M293)	3	3.16
L1c3a	2	2.47	R1b-V88	4	4.21
L1c3a1b	4	4.94	AFRICANO		61.05
L2a1	2	2.47	E1b1b-M81	2	2.11
L2a1+143+@16309	1	1.23	E1b1b-M123	8	8.42
L2a1+16189+(16192)	20	24.69	G-M201	1	1.05
L2a1c3b2	1	1.23	I2-M26	1	1.05
L2b1a	1	1.23	J2-M172	1	1.05
L2d+16129	1	1.23	R1a-SRY1831.2	9	9.47
L3d1a1a	11	13.58	R1b-M529	2	2.11
L3e1d	9	11.11	R1b-S116* (xU152, M529, M153, M167)	9	9.47
L3f1b+16292	1	1.23	R1b-U152	1	1.05
AFRICANO		91.36	EUROPEO		35.79
A2+(64)	1	1.23	Q1a2-M3* (xM19, M194, M199)	3	3.16
A2af1a1	2	2.47	NATIVO AMERICANO		3.16
A2a1	1	1.23			
B2d	1	1.23			
C1c3	2	2.47			
NATIVO AMERICANO		8.64			

Nota. La presencia de E1b1b-M81 en Palenque se interpreta en este estudio como resultado de la mezcla europea, aunque no se puede descartar que provenga del norte de África vía África occidental. Aunque se consideró que un origen africano era más probable para el linaje E1b1bM35* (xM78, M81, M123, V6, M293) en Palenque, no podemos excluir la hipótesis de que proceda de Europa, ya que están reportados subclados raros de E1b1b-M35 en poblaciones europeas.

También se encontró una alta diversidad para los haplogrupos de Y-SNPs (0.8185 ± 0.0033), observándose 17 haplogrupos diferentes (Tabla 3). Similar al ADNmt, se observó una baja diversidad de

haplotipos de Y-STRs ($0,9881 \pm 0,0038$), con muchos haplotipos compartidos dentro de los haplogrupos (Figura 20). En toda la muestra, se puede atribuir un origen africano al 61% de los haplogrupos de ChrY, el 36% representa la mezcla europea y tres muestras (3%) pertenecen a un solo haplogrupo nativo americano (Tabla 3).

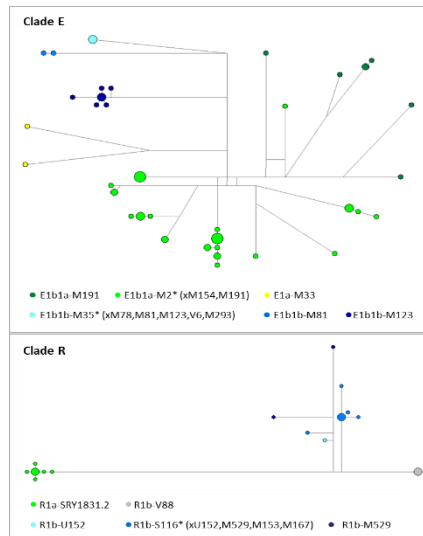


Figura 20. Representación en red de los haplotipos basados en 27 Y-STRs encontrados en Palenque dentro de los clados E y R.

Nota. Además de los clados representados, también se detectaron los siguientes haplogrupos: B2a-M150 * (xM109), con tres muestras que compartan un solo haplotipo; G-M201, I2-M26 y J2-M172, una muestra de cada haplogrupo; Q1a2-M3 * (xM19, M194, M199), tres muestras que solo difieren en DYF387S1; e Y-MRCA * (xM13, SRY10831.1), dos muestras que solo difieren en DYF387S1.

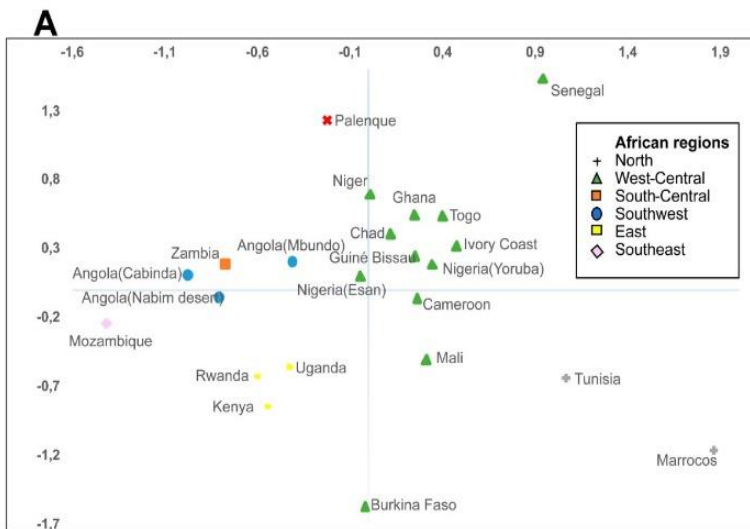
5.2. Los orígenes africanos de los linajes de San Basilio de Palenque

Para investigar los orígenes de los linajes africanos en Palenque, seleccionamos submuestras de haplogrupos africanos encontrados tanto para el ADNmt como para el ChrY. Estas submuestras se compararon con poblaciones de diferentes regiones africanas (ver Tablas complementarias 4 y 5). Se realizó una comparación por pares entre todas las poblaciones seleccionadas. Además, se llevaron a cabo análisis filogeográficos de los haplogrupos africanos en Palenque

basados en frecuencias de haplogrupos y distribuciones en todo el continente africano y considerando el perfil haplotípico de cada linaje (estos resultados se describirán en el CAPÍTULO IV).

En el análisis de diferenciación entre pares de poblaciones para el ADNmt (HVS1 y HVS2), se calcularon las distancias genéticas F_{ST} entre el subgrupo de linajes L en Palenque y las distribuciones de haplotipos en poblaciones africanas (ver Tabla 4). Se observaron valores de F_{ST} relativamente altos y bajas probabilidades de no diferenciación entre Palenque y todas las poblaciones africanas utilizadas para la comparación (ver Tabla complementaria 6). Como se muestra en la gráfica del MDS (Figura 21A), Palenque está lejos de todas las poblaciones, aunque más cerca de las occidentales.

Una imagen similar surge del análisis comparativo entre el grupo de linajes africanos de ChrY de Palenque y de poblaciones de referencias de África (ver Tabla complementaria 5). Se encontraron grandes distancias F_{ST} en todas las comparaciones con Palenque (Tabla complementaria 7), y no se observaron agrupamientos poblacionales, ver en la gráfica MDS (Figura 21B).



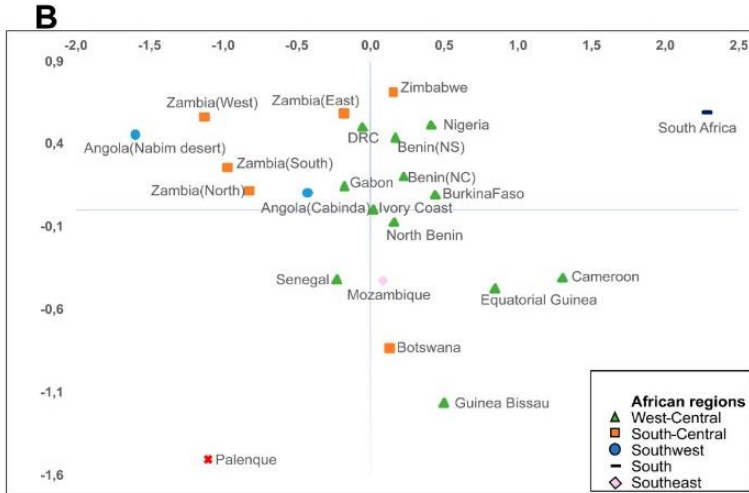


Figura 21. Gráfico MDS de distancias genéticas F_{ST} por pares entre el subconjunto de linajes africanos en Palenque y diferentes poblaciones africanas. Tomado de Martínez B, 2020

Nota. según los haplotipos (A) ADNmt (HVSI y HVSII) y (B) 10 Y-STRs (DYS389I, DYS389II, DYS19, DYS391, DYS390, DYS438, DYS392, DYS437, DYS393 y DYS439).

5.3. Análisis filogeográfico de los linajes africanos de ADNmt y ChrY en Palenque

5.3.1. Linajes africanos de ADNmt

Los linajes africanos en Palenque se dividen en cuatro clados, L0, L1, L2 y L3, que incluyen 17 haplogrupos diferentes. El origen africano de cada haplotipo en Palenque fue investigado por comparación con bases de datos de poblaciones africanas (Tabla complementaria 4). Las regiones sureste y este de África han sido identificadas como el origen probable de los linajes L0 más antiguos como L0k y L0d (Cerezo et al., 2016; Chan et al., 2019). Estudios previos muestran que L0 alcanza las frecuencias más altas y la mayor diversidad en los subhaplogrupos en estas áreas (Cerezo et al., 2016). Por el contrario, debido a una migración posterior al noreste de África, los subclados de L0 más recientes, como el haplogrupo L0a1a, muestran las frecuencias

más altas en África central (Chan et al., 2019; Rito et al., 2013; Soares et al., 2016).

Construimos una red filogenética con datos disponibles en la literatura para el L0a1a + 200, la cual mostro que ninguno de los haplotipos africanos coincide con el haplotipo encontrado en Palenque (Figura 22). Los haplotipos más cercanos fueron de las poblaciones del este (Uganda) (Gomes et al., 2015) y del oeste (Guinea Bissau y Togo) (Carvalho et al., 2011; Göbel et al., 2020). Teniendo en cuenta los registros históricos, es más probable un origen de este linaje en la región Centro-Oeste que en el Este de África.

Un total de 19 muestras de Palenque tenían haplotipos clasificados dentro del haplogrupo L1b (Tabla 2). Las ramas dentro del L1b se extienden por toda África, siendo más frecuentes en la región centro-oeste (Salas et al., 2002). Las ramas L1b1a + 189 y L1b1a18 se encontraron distribuidas en varias regiones de África, como se muestra en la red de la Figura 23. Debido a su amplia distribución, no fue posible señalar el origen de estos linajes en Palenque. En cuanto a L1b1a7a, no se encontraron haplotipos idénticos en los datos de poblaciones recopilados para el análisis de red.

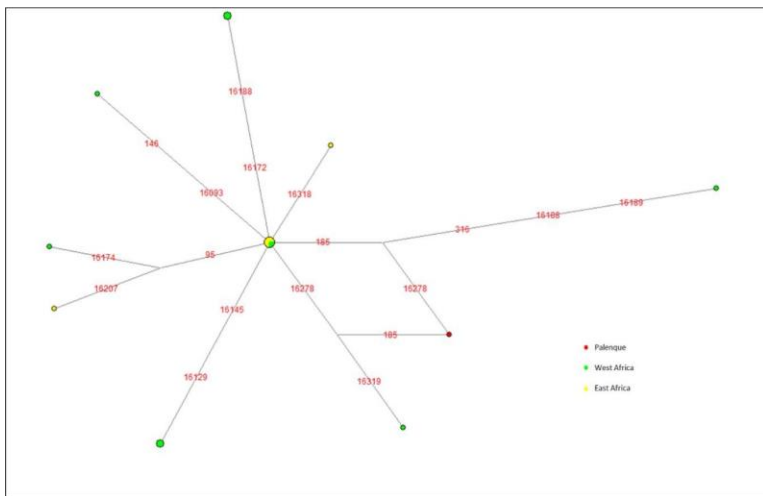


Figura 22. Representación en red del haplogrupo L0a1a + 200.

Nota. La red se construyó utilizando el método de unión de la mediana (Bandelt et al., 1999). Todos los haplotipos se enmarcaron entre las posiciones 16024-16365 (HVS1) y 73-340 (HVSII) y no se consideraron los indeles en las posiciones 16193.xC, 309 xC y 315 xC. En muestras con heteroplasmias, se utilizaron todas las variantes. Las posiciones polimórficas que separan los haplotipos se describen en rojo. Se incluyeron las siguientes muestras: Palenque (n = 1), Guinea Bissau (n = 5), Kenia (n = 1), Nigeria (n = 1), Senegal (n = 2), Togo (n = 1) y Uganda. (n = 4). Las muestras se seleccionaron de las publicaciones incluidas en la Tabla complementaria S4.

El L1b1a1'4 se encontró en una muestra de Guinea Bissau y una de Senegal (Carvalho et al., 2011; Graven et al., 1995). Los dos haplotipos de L1b1a1'4 encontrados en Palenque difieren por una sola mutación, y el más frecuente está separado de los haplotipos de Guinea Bissau y Senegal por un solo paso mutacional (Figura 23). Una búsqueda en la base de datos de EMPOP (empop.online, release v4 / R13) mostró la presencia de 13 y 25 muestras de L1b1a1'4 y L1b1a7a, respectivamente, la mayoría en poblaciones mestizas de América. Aparte de la presencia de L1b1a1'4 en Guinea Bissau, estos haplotipos no se pudieron encontrar en otras poblaciones africanas en EMPOP.

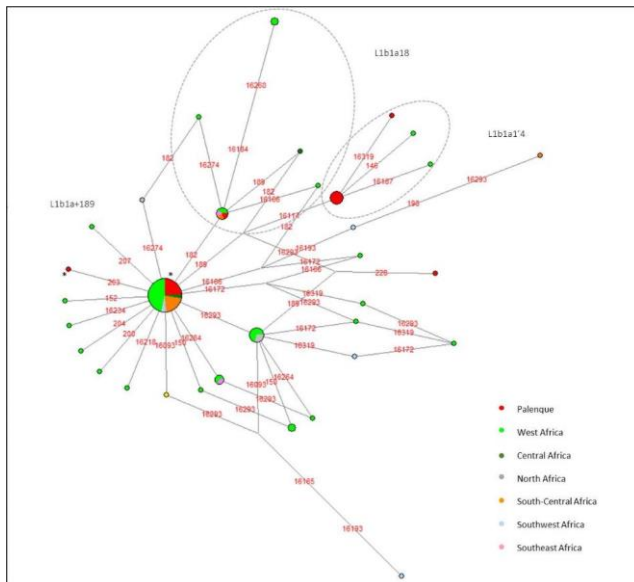


Figura 23. Representación en red del haplogrupo L0a1a + 200.

Nota. La red se construyó utilizando el método de unión de la mediana (Bandelt et al., 1999). Todos los haplotipos se enmarcaron entre las posiciones 16024-16365 (HVS1) y 73-340 (HVSII). No se consideraron los indeles en las posiciones 16193.xC, 309.xC y 315.xC. Se consideraron ambas variantes de posiciones heteroplasmicas. Las posiciones polimórficas que separan los haplotipos se describen en rojo. Se incluyeron las siguientes muestras: Palenque (n = 19), Angola (n = 5), Cameron (n = 2), Chad (n = 4), Ghana (n = 9), Guinea Bissau (= 1), Costa de Marfil (n = 4), Kenia (n = 1), Malí (n = 1), Marruecos (n = 4), Mozambique (n = 2), Níger (n = 2), Nigeria (n = 22), Senegal (n = 2), Togo (n = 2) y Zambia (n = 10). Las muestras se seleccionaron de las publicaciones incluidas en la Tabla complementaria S4. * haplotipos que incluyen una secuencia alternativa de muestras con posición heteroplasmica. Las dos posibles variantes de PU055 (ver haplotipo en la tabla complementaria 2) se incluyeron en la red.

El territorio entre Camerún y Angola ha sido identificado como el origen más probable del haplogrupo L1c27 (Salas et al., 2004). Además, este haplogrupo tiene actualmente las frecuencias más altas en esta región. L1c se observó en 7 muestras de Palenque, distribuidas entre tres subhaplogrupos: una muestra es L1c3, una es L1c3a y cuatro son L1c3a1b. Además, un haplotipo mostró una heteroplasmia de punto en la posición 16355, necesaria para la clasificación del haplogrupo; así: una C en la posición 16355 corresponde a L1c3; una variante T coloca la muestra en L1c3a1b (Tabla complementaria 2). Se construyó la red para los haplogrupos L1c3, L1c3a y L1c3a1b (Figura 23) con datos publicados de las muestras africanas, las cuales aproximadamente el 94% eran de poblaciones de la región centro-sur, suroeste y centro-este. Solo un pequeño número era del este o sureste africano. La red de L1c reveló una alta diversidad haplotípica interna impidiendo una investigación detallada del origen de dichos linajes en Palenque, lo que se había demostrado por estudios previos (Batini et al., 2007; Salas et al., 2002).

No obstante, tres de las cinco muestras de L1c3a1b de Palenque compartieron un haplotipo con una muestra de Ghana (Fendt et al., 2012) (Figura 24). Este haplotipo está a un paso mutacional aparte de dos muestras bantúes de Angola. (Coelho et al., 2009) y dos pasos mutacionales de una muestra de Camerún (Cerezo et al., 2016) (Figura 24). En consecuencia, parece más probable que el origen de estos linajes de Palenque sea la región occidental (central y / o sur) que en las regiones este o norte.

El haplogrupo L2, que se ha asociado con la expansión bantú, tiene su origen más probable en África occidental. Actualmente, L2 se encuentra distribuido por todo el continente africano, presentando altas frecuencias en las regiones centro oeste y sureste de África (Salas et al., 2002; Silva et al., 2015). Tres (L2a, L2b y L2d) de las cinco ramas principales del árbol L2 (L2a-L2e) se encontraron en Palenque (Tabla 1). L2a es el clado más común y extendido en África subsahariana, y fue descrito con las frecuencias más altas en Ghana, Sudán y Mozambique (Silva et al., 2015). Aproximadamente el 92% (n = 24) de las muestras L2 encontradas en Palenque se atribuyeron a ramas dentro de L2a.

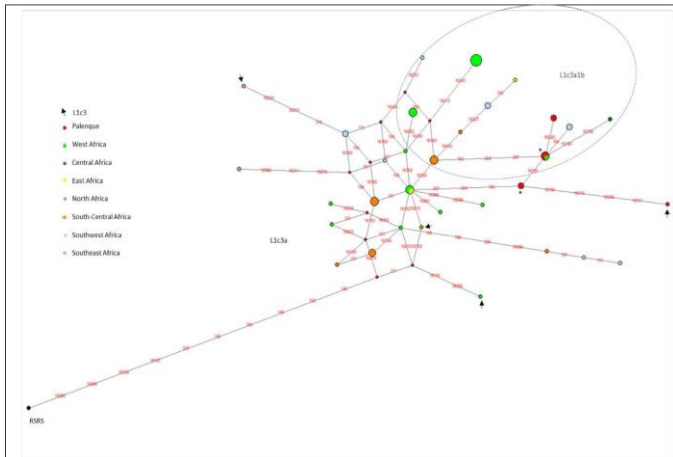


Figura 24. Representación en red de haplogrupos dentro de L1c.

Nota. La red se construyó utilizando el método de unión de la mediana (Bandelt et al., 1999). Todos los haplotipos se enmarcaron entre las posiciones 16024-16365 (HVS1) y 73-340 (HVSII) y no se consideraron los indeles en las posiciones 16193.xC, 309.xC y 315.xC. Se consideraron ambas variantes de posiciones heteroplasmicas. Las posiciones polimórficas que separan los haplotipos se describen en rojo. Se incluyeron las siguientes muestras: Palenque (n = 7), Angola (n = 10), Camerón (n = 1), Gambia (n = 1), Ghana (n = 16), Costa de Marfil (n = 3), Mozambique (n = 2), Níger (n = 1), Nigeria (n = 1), Ruanda (n = 1), Uganda (n = 1) y Zambia (n = 15). Las muestras se seleccionaron de las publicaciones incluidas en la Tabla complementaria S4. * haplotipos que incluyen una secuencia alternativa de muestras con posición heteroplasmica. Las dos posibles variantes de PU103 (ver haplotipo en la tabla complementaria 2) se incluyeron en la red.

En las redes de L2a1 (Figura 25) y L2a1 + 16189 (16192) (Figura 26), es posible ver una amplia dispersión geográfica, típica de los linajes L2a. En ambos casos, una distribución en forma de estrella surge desde un grupo central de haplotipos compartidos por varias regiones africanas. Las dos muestras de L2a1 de Palenque están a uno y / o dos pasos mutacionales del grupo central. Por tanto, todos los haplotipos de L2a1 + 16189 (16192) están asociados con un único fundador que comparte su secuencia con muestras de varias regiones africanas.

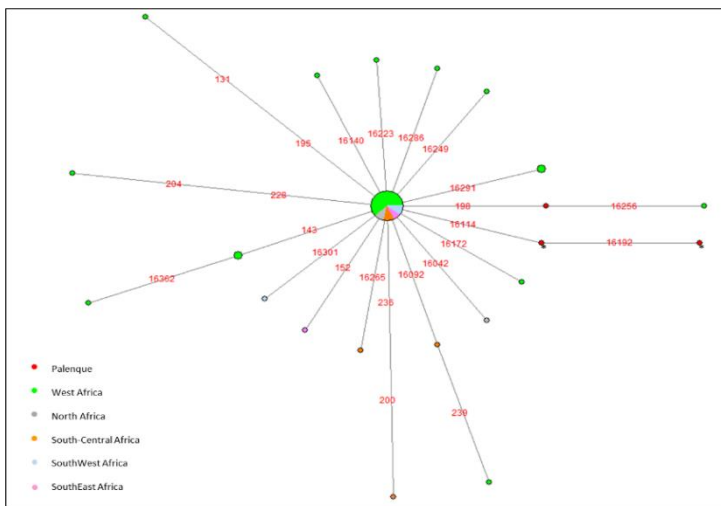


Figura 25. Representación en red del haplogrupo L2a1.

Nota. La red se construyó utilizando el método de unión de la mediana (Bandelt et al., 1999). Todos los haplotipos se enmarcaron entre las posiciones 16024-16365 (HVS1) y 73-340 (HVSII) y no se consideraron los indeles en las posiciones 16193.xC, 309.xC y 315.xC. Se consideraron ambas variantes de posiciones heteroplasmicas. Las posiciones polimórficas que separan los haplotipos se describen en rojo. Se incluyeron las siguientes muestras: Palenque (n = 2), Angola (n = 4), Burkina Faso (n = 1), Chad (n = 1), Ghana (n = 11), Guinea Bissau (n = 3), Costa de Marfil (n = 6), Malí (n = 1), Marruecos (n = 3), Mozambique (n = 3), Nigeria (n = 6), Togo (n = 2), Túnez (n = 1) y Zambia (n = 6). Las muestras se seleccionaron de las publicaciones incluidas en la Tabla complementaria 4. * haplotipos que incluyen una secuencia alternativa de muestras con posición heteroplasmica. Las dos posibles variantes de PU063 (ver haplotipo en la tabla S2) se incluyeron en la red.

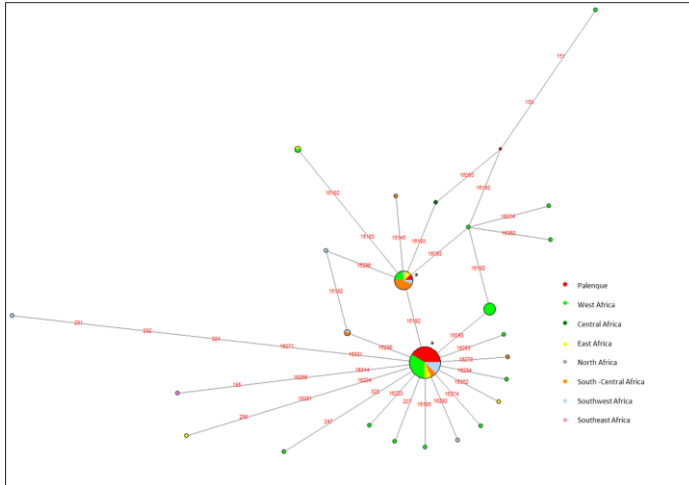


Figura 26. Representación en red del haplogrupo L2a1 + 16189 (16192).

Nota. La red se construyó utilizando el método de unión de la mediana (Bandelt et al., 1999). Todos los haplotipos se enmarcaron entre las posiciones 16024-16365 (HVS1) y 73-340 (HVSII) y no se consideraron los indeles en las posiciones 16193.xC, 309.xC y 315.xC. Se consideraron ambas variantes de posiciones heteroplasmicas. Las posiciones polimórficas que separan los haplotipos se describen en rojo. Se incluyeron las siguientes muestras: Palenque (n = 20), Angola (n = 12), Benín (n = 1), Camerún (n = 1), Gambia (n = 1), Ghana (n = 14), Guinea Bissau (n = 1), Costa de Marfil (n = 4), Kenia (n = 2), Mozambique (n = 1), Níger (n = 6), Nigeria (n = 11), Togo (n = 1), Túnez (n = 1), Uganda (n = 5) y Zambia (n = 15). Las muestras se seleccionaron de las publicaciones incluidas en la Tabla complementaria 4.

El haplotipo L2a1 + 143 + 16309 encontrado en Palenque solamente está separado por una posición mutacional de haplotipos del oeste, del sudoeste y este de África (Figura 27). Una muestra de Palenque para el L2a1c3b2 se diferenció en 5 sitios polimórficos de la única muestra africana disponible para este haplogrupo en la base de datos (Coelho et al., 2009). En resumen, las inferencias sobre el origen africano de los haplotipos de ADNmt de Palenque dentro del clado L2a no fueron posibles debido a su amplia distribución en África, así como debido a la incertidumbre en la clasificación de las subramas asociadas con polimorfismos de mutación rápida (Salas et al., 2002; Silva et al., 2015).

Las subramas L2b y L2d son menos frecuentes en África que L1a (Salas et al., 2002; Salas et al., 2004; Silva et al., 2015). Ambos haplogrupos se originaron en África occidental, con frecuencias más altas en las poblaciones de las regiones centro-oeste y suroeste. Aunque L2b1a es caracterizado por los polimorfismos 146T y 16355T, la muestra L2b1a de Palenque carece de la mutación 146T. No se encontró ninguna muestra en la literatura con este perfil exacto. En la red, el haplotipo de Palenque difiere del grupo central en dos pasos mutacionales (el primero es 146) (Figura 28). De manera comparable, una muestra de Camerún (Göbel et al., 2020) también diverge del grupo central por un polimorfismo en la posición 146.

L2d es el subclado más antiguo del haplogrupo L2 (Salas et al., 2002) y muestra una ocurrencia esporádica en el continente africano. El haplotipo L2d + 16129 de Palenque coincide con una única muestra L2d + 16129 de Angola (Bantú) (Coelho et al., 2009) (Figura 29).

La amplia distribución previamente reportada para otras ramas L se observa una vez más en los sublinajes L2b y L2d. Sin embargo, un claro origen suroeste para el haplotipo L2d en Palenque pudo demostrarse. En cuanto al haplotipo L2b en Palenque, los datos ubican el origen más probable en la región occidental entre la Ensenada de Benín y Angola, excluyéndose sus orígenes de las regiones del este y sureste posiblemente.

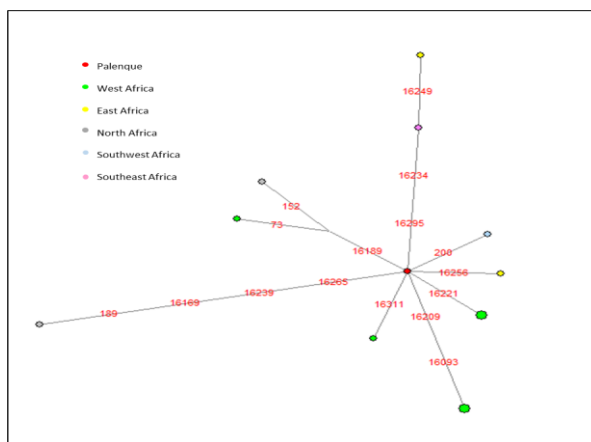


Figura 27. Representación en red del haplogrupo L2a1 + 143 + 16309.

Nota. La red se construyó utilizando el método de unión de la mediana (Bandelt et al., 1999). Todos los haplotipos se enmarcaron entre las posiciones 16024-16365 (HVS1) y 73-340 (HVSII) y no se consideraron los indeles en las posiciones 16193.xC, 309.xC y 315.xC. Se consideraron ambas variantes de posiciones heteroplasmicas. Las posiciones polimórficas que separan los haplotipos se describen en rojo. Se incluyeron las siguientes muestras: Palenque (n = 1), Angola (n = 1), Burkina Faso (n = 1), Chad (n = 2), Ghana (n = 2), Kenia (n = 1), Marruecos (n = 2), Mozambique (n = 1), Nigeria (n = 1) y Uganda (n = 1). Las muestras se seleccionaron de las publicaciones incluidas en la Tabla complementaria 4.

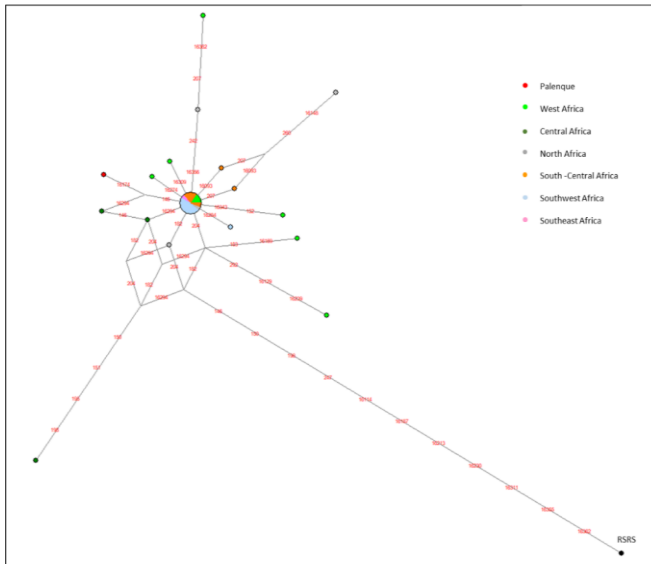


Figura 28. Representación en red del haplogrupo L2b1a.

Nota. La red se construyó utilizando el método de unión de la mediana (Bandelt et al., 1999). Los haplotipos se enmarcaron entre las posiciones 16024-16365 (HVS1) y 73-340 (HVSII) y no se consideraron los indeles en las posiciones 16193.xC, 309.xC y 315.xC.

Se consideraron ambas variantes de posiciones heteroplasmicas. Las posiciones polimórficas que separan los haplotipos se describen en rojo. Se incluyeron las siguientes muestras: Palenque (n = 1), Angola (n = 10), Camerún (n = 3), Chana (n = 1), Guinea Bissau (n = 1), Costa de Marfil (n = 1), Marruecos (n = 4), Mozambique (n = 1), Nigeria (n = 5), Senegal (n = 1) y Zambia (n = 7). Las muestras se seleccionaron de las publicaciones incluidas en la Tabla complementaria 4.

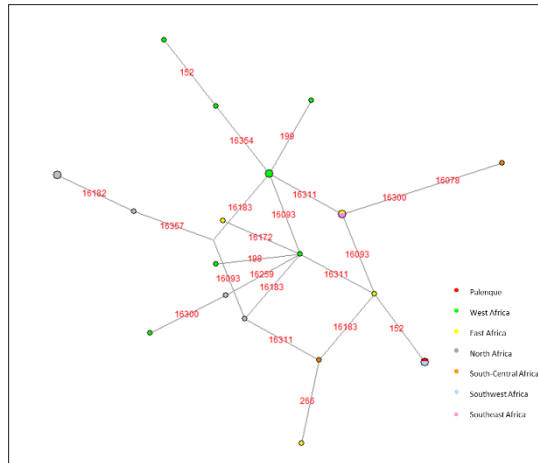


Figura 29. Representación en red del haplogrupo L2d + 16129.

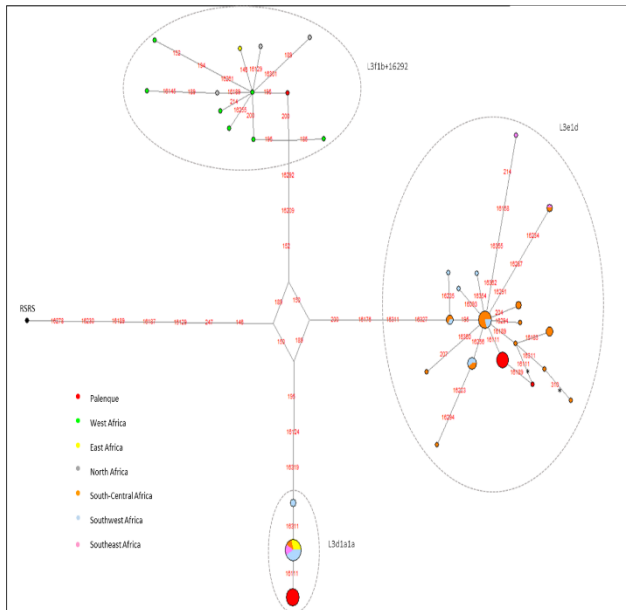
Nota. La red se construyó utilizando el método de unión de la mediana (Bandelt et al., 1999). Todos los haplotipos se enmarcaron entre las posiciones 16024-16365 (HVS1) y 73-340 (HVSII) y no se consideraron los indels en las posiciones 16193.xC, 309.xC y 315.xC. Se consideraron ambas variantes de posiciones heteroplasmicas. Las posiciones polimórficas que separan los haplotipos se describen en rojo. Se incluyeron las siguientes muestras: Palenque (n = 1), Angola (n = 1), Guinea Bissau (n = 2), Costa de Marfil (n = 4), Kenia (n = 1), Marruecos (n = 5), Mozambique (n = 1), Nigeria (n = 2), Ruanda (n = 3) y Zambia (n = 2). Las muestras se seleccionaron de las publicaciones incluidas en la Tabla complementaria 4.

Se cree que L3 se origina en África oriental, la patria de la migración y diversificación fuera de África, donde alcanza la frecuencia más alta (Soares et al., 2016). No obstante, las subramas L3 también se pueden encontrar en otras regiones africanas. Aproximadamente el 28% de los linajes L encontrados en Palenque se distribuyen en tres ramas L3: L3d, L3e y L3f. La rama L3d es muy común en África centro-occidental (Salas et al., 2002), aunque también está presente en otras regiones como el este y el sureste. Aunque está muy extendido, la rama L3d1a1a muestra una baja diversidad, con solo dos haplotipos descritos en la literatura (Plaza et al., 2004) (Figura 30). Las once muestras de L3d1a1a de Palenque tienen el mismo haplotipo. Sin embargo, no fue imposible inferir la región exacta de África donde se originó, ya que este haplotipo (i) no se encontró en las muestras africanas, y (ii) difiere en

un paso del grupo central que incluye muestras de varias regiones (Figura 30). A diferencia de los otros subclados de L3, L3e es el más frecuente y ampliamente distribuido en el continente africano. Es especialmente común en los grupos bantúes.

La red para L3e1d muestra una distribución casi exclusiva en el suroeste africano (Cerezo et al., 2016; Coelho et al., 2009) y África centro-sur (Barbieri, Butthof, et al., 2013; Barbieri et al., 2014) (Figura 30), lo que indica un origen meridional para este linaje en Palenque. Todas las muestras de Palenque comparten el mismo haplotipo (a excepción de una variante de una muestra heteroplasmica), lo que sugiere un único fundador.

El haplogrupo L3f es raro en África (Salas et al., 2002). Se originó en la región oriental y luego se expandió a África central. Esta rama también está presente en África centrooccidental (Salas et al., 2002). Aunque no se encontraron haplotipos coincidentes con el L3f1b + 16365 de la muestra de Palenque, la más cercana fue de Ghana (Figura 30).



Nota. La red se construyó utilizando el método de unión de la mediana (Bandelt et al., 1999). Todos los haplotipos se enmarcaron entre las posiciones 16024-16365 (HVS1) y 73-340 (HVSII) y no se consideraron los indeles en las posiciones 16193.xC, 309.xC y 315.xC. Las posiciones polimórficas que separan los haplotipos se describen en rojo. Se incluyeron las siguientes muestras: Palenque (n = 20), Angola (n = 18), Chad (n = 3), Ghana (n = 2), Kenia (n = 1), Marruecos (n = 3), Mozambique (n = 5), Níger (n = 2), Ruanda (n = 2), Uganda (n = 3) y Zambia (n = 26). Las muestras se seleccionaron de las publicaciones incluidas en la Tabla complementaria S4. * haplotipos que incluyen una secuencia alternativa de muestras con posición heteroplasmica. Las dos posibles variantes de PR052 (ver haplotipo en la tabla complementaria 2) se incluyeron en la red.

5.3.2. Linajes africanos de ChrY

Los linajes paternos africanos encontrados en Palenque se dividen en cuatro clados diferentes: Y-MRCA* (xM13, SRY10831.1), B, E y R, que comprenden siete haplogrupos. Los orígenes más probables de estos haplogrupos se investigaron comparando el perfil de los haplotipos de Y-STRs en las muestras de Palenque y poblaciones africanas (tabla complementaria 5). En este estudio, detectamos tres cromosomas clasificados como Y-MRCA* (xM13, SRY10831.1), que incluye muestras del haplogrupo A. El haplogrupo A está prácticamente restringido al continente africano, alcanzando el nivel más alto de frecuencias en poblaciones de habla khoisan. También es frecuente en los grupos nilóticos del este y noreste de África (Carvalho et al., 2011; Cruciani et al., 2002; Iacovacci et al., 2017; Semino et al., 2000), y se ha observado esporádicamente en las poblaciones del sureste y suroeste africano (Pour et al., 2013; Rowold et al., 2014). Algunos linajes dentro del clado A tienen especificidad geográfica. Por ejemplo, los linajes pertenecientes a los haplogrupos A1-M31 fueron descritos en África occidental (Rosa et al., 2007), diferenciándose de los de la parte oriental del continente que pertenecen a haplogrupos A3-M13 u otros linajes dentro de A3.

Para inferir el origen más probable de las muestras del haplogrupo A encontradas en Palenque, se construyó una red, que incluía los perfiles Y-STR de dos muestras de este estudio (no teníamos información para Y-STRs en una muestra) y los de 109 muestras del haplogrupo A compiladas a partir de la bibliografía (Tabla complementaria 5). La red construida muestra una clara separación entre las muestras pertenecientes a los haplogrupos A1-M31 (de Guinea

Bissau, en África occidental), A3-M13 (incluidas todas las muestras de África oriental, algunas del noreste de África y una muestra de Benín y uno de Guinea Ecuatorial) y A3-M28 (incluidas cuatro muestras del noreste de África) (Figura 31). Por otro lado, las muestras de bantúes de África central y pigmeos de Ghana y Camerún, que no tuvieron Y-SNPs tipados dentro del haplogrupo A [Y-MRCA 9x4A4)], fueron separadas en dos grupos sin superposición en los grupos A1-M31, A3-M28 y A3-M13. Las muestras de bantúes y pigmeos de África Central procedentes de Ghana y Camerún [Y-MRCA (xA4)] se separaron en dos clusters, distintos a los clusters A1-M31, A3-M28 y A3-M13.

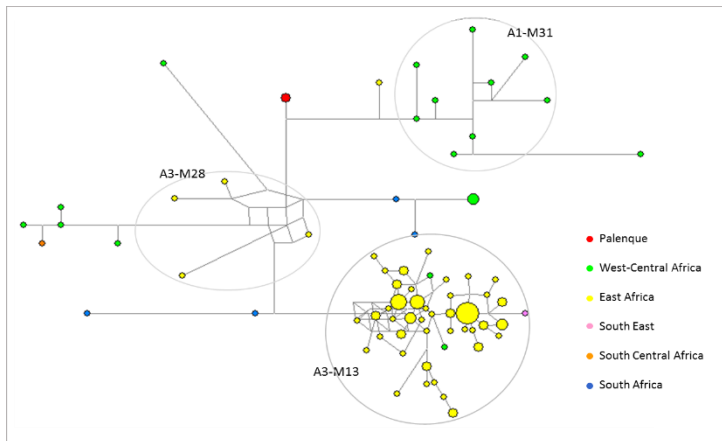


Figura 31. Representación en red de los haplotipos seleccionados que podrían pertenecer a YMRCA * (*xM13, SRY10831.1*).

Nota. La red se construyó utilizando 12 loci Y-STR y se resolvió aplicando los métodos *Reduced Median* (Bandelt et al., 1995) y *Median Joining* (Bandelt et al., 1999) secuencialmente. Se incluyeron las siguientes muestras: Palenque (n = 2) Benín (n = 1), Botswana (n = 1), Burkina Faso (n = 1), Guinea Ecuatorial (n = 1), Eritrea (n = 9), Etiopía (n = 22), Gabón (n = 4), Gabón / Pigmeo de Camerún (n = 3), Guinea Bissau (n = 8), Costa de Marfil (n = 1), Kenia (n = 11), Mozambique (n = 1), Namibia (n = 3), Nigeria (n = 1), Uganda (n = 39), Zambia (n = 1). Las muestras se seleccionaron de las publicaciones incluidas en la Tabla complementaria 5.

Las dos muestras de Palenque (que compartían el perfil completo de Yfiler Plus) se encontraron entre el A1-M31 y A3-M28. Aunque es difícil señalar un origen del haplogrupo encontrado en

Palenque, podemos al menos excluir que vinieron de regiones del este, sureste o sur de África. Por razones históricas, es más probable asumir un origen en el oeste que en el noreste de África. Las ramas de África occidental más cercanas a los haplotipos de Palenque corresponden a muestras de Guinea Bissau, del haplogrupo A1-M31.

El haplogrupo B está presente exclusivamente en las poblaciones de África subsahariana, a excepción de la migración reciente a otros continentes. Los cromosomas del haplogrupo B2b-M112 se encuentran en su frecuencia más alta entre los pigmeos y son también frecuente en khoisan, con algunos linajes prácticamente restringidos a estos grupos de cazadores-recolectores (Batini et al., 2011; Berniell-Lee et al., 2009).

Por el contrario, el subclado B2a está ampliamente disperso en África subsahariana (Barbieri et al., 2016; Scozzari et al., 2012). La presencia de este haplogrupo en diferentes regiones africanas se atribuyó en primer lugar a la dispersión de hablantes de bantú. Sin embargo, estudios recientes mostraron que el haplogrupo B2a ya estaba presente en los grupos khoisan antes de su contacto con poblaciones de habla bantú (Batini et al., 2011).

Identificamos tres muestras pertenecientes a B2a-M150* en Palenque (5,17% de los linajes africanos), todas sin el alelo derivado de M109 que suele estar presente en poblaciones bantúes del suroeste, sureste y sur de África (Beleza et al., 2005; de Filippo et al., 2011; Rowold et al., 2014).

La red muestra una amplia dispersión de los haplotipos (Figura 32), que puede deberse a la baja resolución de algunos linajes. Con el objetivo de una amplia cobertura de la población, hemos incluido todos los posibles cromosomas B2a-M150 * (xM109), incluso aquellos que se tipificaron solo para el SNP de diagnóstico del haplogrupo B. Solo 33 de las 134 muestras de referencia africanas incluidas en la red fueron tipificadas para M109: 3 de Eritreia, 25 de Uganda, 4 de Nigeria y 1 de Ghana. Las muestras de Palenque se sitúan fuera de los dos grupos principales en una rama enraizada en una muestra de Gabón. Dos muestras comparten sus haplotipos con una muestra de Cabinda (Angola), y la tercera muestra difiere por una mutación de un solo paso. Los resultados obtenidos son, por tanto, compatibles con un único

fundador, que podría haber sido traído de los puertos de Loango (en el actual Gabón) o Cabinda (Angola), en el antiguo Reino de Congo.

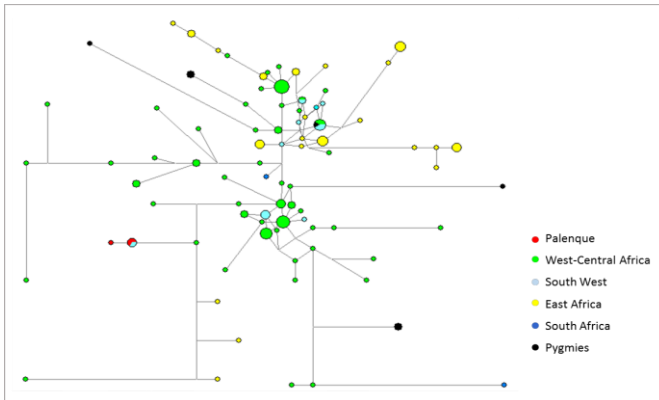


Figura 32. Representación en red de haplotipos seleccionados dentro del clado B-M60 (excepto aquellos que se sabe que portan el alelo derivado de M109 / M152).

Nota. El número de Y-STRs utilizados para construir la red se redujo a un conjunto común de 11 loci: DYS389I, DYS389II, DYS19, DYS390, DYS438, DYS392, DYS437, DYS385a/b, DYS393 y DYS439. La red se resolvió aplicando los métodos *Reduced Median* (Bandelt et al., 1995) y *Median Joining* (Bandelt et al., 1999) secuencialmente. Se incluyeron las siguientes 134 muestras africanas: Palenque (n = 3), Angola (n = 12), Benín (n = 4), Burkina Faso (n = 7), Camerún (n = 9), Gabón / Camerún Pigmeo (n = 6), República Democrática del Congo (n = 1), Eritrea (n = 3), Gabón (n = 48), Ghana (n = 1), Guinea Bissau (n = 1), Namibia (n = 1), Níger (n = 11), Nigeria (n = 4), Sudáfrica (n = 1) y Uganda (n = 25). Las muestras se seleccionaron de las publicaciones incluidas en la Tabla complementaria 5.

El clado E es el más común en África y también está presente en las poblaciones de Europa y Asia occidental. De acuerdo con el origen africano de Palenque, este clado representa el 61,05% de los linajes de ChrY, el 82,76% de los asignados a un origen africano.

El haplogrupo E1b1b-M81 es frecuente en el norte de África, particularmente en los grupos bereberes y también está presente en Iberia y el sur de Italia como legado del noroeste de África. (Capelli et al., 2009; Cruciani et al., 2004). Dado que ningún flujo genético directo se ha reportado desde el norte de África, la presencia de este haplogrupo en la población de Palenque más probablemente resultó de la mezcla

européa. No obstante, un origen norte africano de la E-M81 en Palenque, vía África occidental, no se puede descartar. También se atribuyó un origen europeo a los linajes del haplogrupo E1b1b-M123 encontrados en Palenque. Este haplogrupo está ausente en el África subsahariana y es frecuente en las poblaciones árabes del noreste de África y Oriente Medio (Luis et al., 2004). Este haplogrupo también se diseminó en Europa, aunque con bajas frecuencias. En Iberia, se ha asociado con ancestría judía, norte africana y de Oriente Medio. Un origen directo en África se puede atribuir a los restantes haplogrupos E en Palenque [E1a-M33, E1b1a-M2 * (xM154, M191), E1b1a-M191 y E1b1bM35 * (xM78, M81, M123, V6, M293)], que no están presentes en las poblaciones europeas y probablemente fueron traídos a América durante el comercio de esclavos.

Aparte de E1b1b-M81 y E1b1b-M123 (atribuidos a europeos), dentro del clado E1b1b, encontramos tres muestras pertenecientes a E1b1b-M35 * (xM78, M81, M123, V6, M293).

Aunque es más probable un origen africano de este linaje, no podemos excluir la hipótesis de que procedan de Europa, ya que se detectaron subclados raros de E1b1b-M35 (que albergan mutaciones V1515, V257 o V2009) en poblaciones del Mediterráneo occidental a saber, Portugal, España, Francia e Italia (Trombetta et al., 2015).

El haplogrupo E1b1b-M35* (xM78, M81, M123, V6, M293) está presente en las poblaciones de África oriental y meridional (Cruciani et al., 2004; de Filippo et al., 2011; Iacovacci et al., 2017) y no se ha detectado en las regiones central y suroeste, a saber, Níger,

Nigeria, Camerún, Guinea Ecuatorial, Gabón y Angola (Beleza et al., 2005; Berniell-Lee et al., 2009; Coelho et al., 2009; Cruciani et al., 2004; de Filippo et al., 2011; Oliveira et al., 2019). Fue descrito en África occidental con bajas frecuencias. La red E1b1b-M35 muestra muchas ramas separadas (Figura 33), algunas de las cuales están restringidas a una sola región africana.

La alta diversidad y la amplia dispersión en las poblaciones africanas se explican por la baja resolución de los datos disponibles y muy probablemente representa diversos sublinajes E1b1b-M35. Desde la red, podemos ver que las muestras de Palenque (representado por un solo haplotipo) se colocan en una rama que tiene sus raíces en un

haplotipo compartido entre una muestra de Burkina Faso y Etiopía. Aparte de las muestras de Palenque, esta rama incluía una muestra de Senegal, una de Benín, tres de Guinea Bissau y dos de Burkina Faso. Los resultados así obtenidos son compatibles con un único fundador que es más probable que haya venido de una región entre Senegambia y el golfo de Benín que de otras regiones africanas.

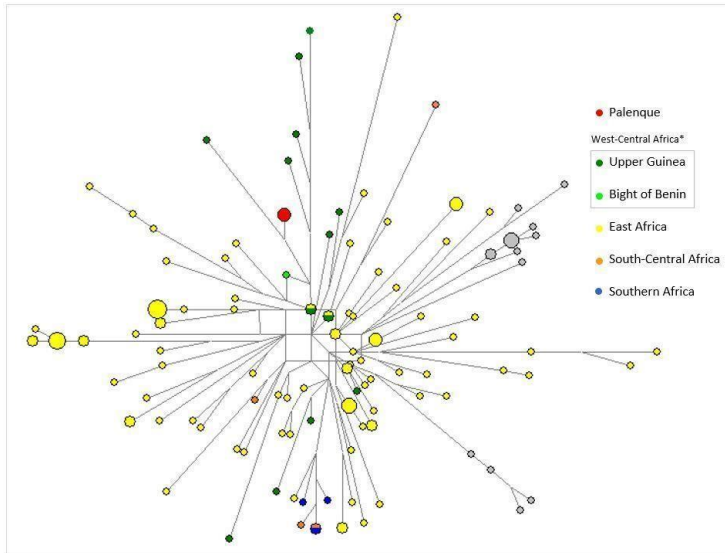


Figura 33. Representación en red de los haplotipos seleccionados dentro del clado *E1b1b-M35* (excepto los que portan los alelos derivados de *M78, M81, M123, V6, M293*).

Nota. La red se construyó utilizando la información disponible sobre 142 muestras africanas del haplogrupo *E1b1b-M35* (no se incluyeron las que portaban el alelo derivado de *M78, M81, M123, V6, M293*). El número de *Y-STR* utilizados para construir la red se redujo a un conjunto común de 12 loci: *DYS389I, DYS389II, DYS19, DYS390, DYS391, DYS438, DYS392, DYS437, DYS385a / b, DYS393* y *DYS439*. La red se resolvió aplicando los métodos *Reduced Median* (Bandelt et al., 1995) y *Median Joining* (Bandelt et al., 1999) secuencialmente. Se incluyeron las siguientes muestras: Palenque ($n = 3$), Argelia ($n = 16$), Benín ($n = 1$), Botswana ($n = 2$), Burkina Faso ($n = 7$), Eritrea ($n = 23$), Etiopía ($n = 45$), Guinea Bissau ($n = 6$), Kenia ($n = 24$), Senegal ($n = 1$), Sudáfrica ($n = 1$), Uganda ($n = 7$) y Zambia ($n = 4$). Las muestras se seleccionaron de las publicaciones incluidas en la Tabla complementaria S5. * Las muestras de la región centro-oeste se discriminaron entre la Alta Guinea y la ensenada de Benín, para una inferencia más clara del origen de las muestras de Palenque.

En Palenque se encontraron dos muestras del haplogrupo E1a-M33, correspondientes al 3,45% de la población africana. Este haplogrupo está presente exclusivamente en África occidental. Las frecuencias más altas reportadas fueron en la región de Mali y Burkina Faso, con una paulatina disminución de la frecuencia hacia el sur (Cruciani et al., 2002; de Filippo et al., 2011). Tampoco ha sido descrito en poblaciones del este y sur (de Filippo et al., 2011; González et al., 2013; Rosa et al., 2007; Soares et al., 2016).

Se construyó una red utilizando la información disponible en 77 muestras de los haplogrupos E1a-M33 (Figura 34). Las dos muestras de Palenque están bien separadas. Una de ellas está próxima de muestras de Guinea Bissau, en una rama que tiene su raíz en un clúster compuesto principalmente por muestras de Burkina Faso. La otra muestra se coloca en la red cerca de un grupo de muestras de Benín y Nigeria. Los resultados obtenidos son, por tanto, compatibles con dos orígenes diferentes en occidente África. Es más probable que uno de los cromosomas Y proceda de la región de la Alta Guinea. El origen más probable del otro cromosoma Y se encuentre en algún lugar a lo largo de la ensenada de Benín.

El haplogrupo E1b1a-M2 (y sus sublinajes) está muy extendido en África y es muy prevalente en todas las poblaciones subsaharianas bantúes, con frecuencias superiores al 80% en la mayoría de las poblaciones (Beleza et al., 2005; de Filippo et al., 2011; Pour et al., 2013; Rowold et al., 2014). En Palenque, el 45,26% de todas las muestras (74,14% de las muestras con haplogrupos africanos) pertenecen al haplogrupo E1b1a-M2* (xM154, M191) y a su subclado E1b1a-M191. El número extremadamente alto y la diversidad de muestras que pueden incluirse en estos haplogrupos no nos permitió resolver la reticulación obtenida en las redes (datos no mostrados). Por lo tanto, para estos dos haplogrupos realizamos un análisis de coincidencia entre los haplotipos encontrados en Palenque y los reportados en Poblaciones africanas de los mismos haplogrupos.

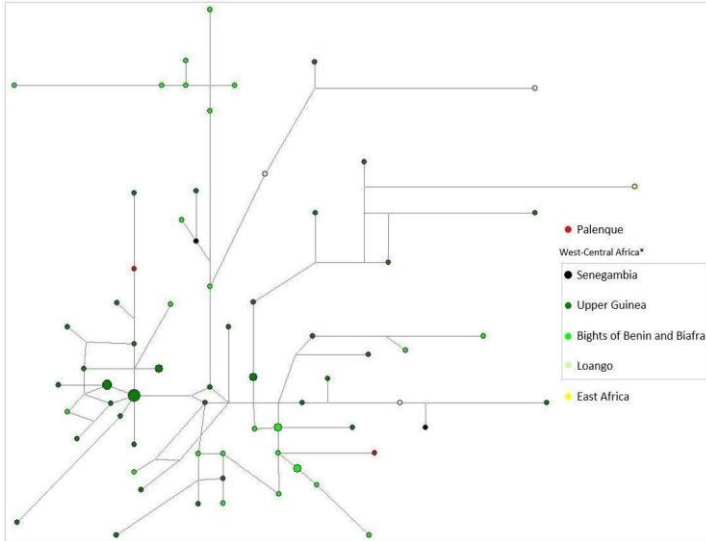


Figura 34. Representación en red de haplotipos seleccionados en el interior del haplogrupo E1a-M33.

Nota. La red se construyó utilizando 11 loci Y-STR y se resolvió aplicando los métodos *Reduced Median* (Bandelt et al., 1995) y *Median Joining* (Bandelt et al., 1999) secuencialmente. Se incluyeron las siguientes muestras: Palenque (n = 4), Benín (n = 7), Burkina Faso (n = 26), Guinea Ecuatorial (n = 23), Etiopía (n = 1), Gabón (n = 3), Guinea Bissau (n = 17), Costa de Marfil (n = 1), Nigeria (n = 7), Senegal (n = 2) y Togo (n = 2). Las muestras se seleccionaron de las publicaciones incluidas en la Tabla complementaria 5. * Las muestras de la región centrooeste se discriminaron entre Senegambia, Alta Guinea, Bights of Benín y Biafra y Loango, para una inferencia más clara del origen de las muestras de Palenque.

Un total de 2204 muestras, seleccionadas del haplogrupo E1b1a-M2 (excluidas las que llevan alelos derivados para M154 o M191), se compararon con las 35 muestras E1b1a-M2* (xM154, M191) de Palenque (Tabla 4). Estas 35 muestras dieron como resultado 9 grupos diferentes, incluidos haplotipos idénticos y vecinos (Figura 20). No se encontraron coincidencias para cinco de los 9 grupos (Tabla 4); por lo tanto, no fue posible inferir su origen. Para dos grupos, solo se encontraron haplotipos coincidentes en Benín, con siete y con una muestra, respectivamente (Tabla 4), colocando el origen más probable de estos linajes en la región de la ensenada de Benín. Para los dos grupos

restantes, se encontraron haplotipos idénticos principalmente en poblaciones al sur de Camerún; por lo tanto, estos linajes probablemente vinieron de la costa occidental africana entre Loango y Angola.

De E1b1a-M191, seleccionamos 934 muestras de 20 países del África subsahariana para compararlas con las siete muestras E1b1a-M191 encontradas en Palenque (Tabla 5). El árbol de estas muestras compartió el mismo haplotipo (Figura 20). Se encontraron coincidencias exactas para solo dos de los cinco haplotipos diferentes. Para un haplotipo se encontró una coincidencia en Gabón. El segundo haplotipo tiene cuatro coincidencias, una en Gabón, dos en Angola y una en Zambia, situando el origen más probable de estos linajes en una región al sur de Camerún, posiblemente entre Loango y Angola.

Tabla 4. Resultados de los análisis de coincidencia de haplotipos entre las muestras E1b1a-M2* (xM154, M191) de Palenque y de diferentes países africanos (consulte la Tabla S5 para obtener referencias a los datos consultados).

Haplotipo*	N												
	F1	13	18	15	21	11	10	14	16	18	13	13	6
F1	13	18	15	21	11	10	14	16	18	13	13	6	-
F2	13	18	15	21	11	10	15	16	18	13	13	6	-
F3	13	17	16	21	11	11	14	17	19	13	13	2	-
F4	13	17	16	21	11	11	14	17	19	13	12	1	-
F5	13	18	15	21	11	11	14	16	17	12	11	5	-
F6	12	18	15	21	11	11	14	16	16	13	12	5	-
F7	12	19	15	21	11	12	14	16	17	13	11	4	-
F8	12	19	15	21	11	12	14	16	16	13	12	1	-
F9	13	18	15	21	11	11	14	16	17	13	11	2	-
F10	14	17	15	21	11	11	14	16	17	13	13	1	-
F11	14	17	17	21	11	11	14	16	17	14	12	1	-
F12	14	18	16	21	11	11	14	16	17	13	13	1	-

	Algeria	Senegal	Gambia	Guinea Bissau	Guinea	Liberia	Costa de Marfil	Burkina-Faso	Ghana	Togo	Benin	Nigeria	Níger	Camerin	Guinea Ecuatorial	Gabón/Camerin_Pigmeo	Gabón	DRC	Angola	Zambia	Botswana	South África	Mozambique	Kenia	Uganda
F1	2	14	3	146	1	1	74	203	9	8	140	54	4	204	47	17	588	22	237	352	14	4	35	22	8
F2	-	-	-	-	-	-	-	-	-	-	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F6	-	-	-	-	-	-	-	-	-	-	6	-	2	-	-	-	-	-	-	-	-	-	-	-	-
F7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F9	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
F12	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Nota. *DYS389I, DYS389II, DYS19, DYS390, DYS438, DYS392, DYS437, DYS385a, DYS385b, DYS393 and DYS439

Tabla 5. Resultados de los análisis de coincidencia de haplotipos entre las muestras E1b1a-M191 de Palenque y de diferentes países africanos (consulte la Tabla S5 para obtener referencias a los datos consultados).

Haplótipo*	N	
F1 14 17 15 10 21 11 11 14 17 17 14 11 1	3	Guinea Bissau
F2 14 17 16 10 21 11 11 14 18 19 14 12 1	1	Sierra León
F3 13 17 16 10 21 11 11 14 16 18 14 12 1	12	Costa de Marfil
F4 12 17 17 10 21 11 11 14 19 19 14 13 1	63	Burkina Faso
F5 13 17 17 10 21 11 11 14 16 17 14 13 3	6	Ghana
	8	Togo
	102	Benín
	73	Nigeria
	1	Níger
	57	Camerún
	35	Guinea Ecuatorial
	268	Gabón
	16	DRC
	71	Angola
	125	Zambia
	10	Botsuana
	1	Sudáfrica
	19	Mozambique
	57	Kenia
	6	Uganda

El haplogrupo R1b-V88 se encontró en el 6,15% de los linajes africanos en Palenque. El Clado R se originó fuera de África y tiene una alta frecuencia en las poblaciones europeas. Sin embargo, el subclado R1b-V88 está esencialmente restringido al continente africano, con altas frecuencias en poblaciones del Sahel central (D'Atanasio et al., 2018). En África subsahariana, este haplogrupo tiene las frecuencias más altas en los grupos Chádicos del norte de Camerún (Cruciani et al., 2010).

En África occidental, este haplogrupo también se encontró en Gabón y Guinea Ecuatorial y es ausente en las poblaciones del sur (Berniell-Lee et al., 2009; Cruciani et al., 2010; González et al., 2013). Teniendo en cuenta su distribución geográfica y niveles de diversidad en diferentes regiones, el haplogrupo R1b-V88 se ha asociado con la propagación trans sahariana de hablantes afroasiáticos proto-chádicos durante el periodo medio/temprano del Holoceno (Cruciani et al., 2010; D'Atanasio et al., 2018).

Construimos una red utilizando la información disponible sobre 56 muestras africanas del haplogrupo R, excluyendo las muestras que fueron asignadas a sublinajes fuera de la rama R1b-V88 y que eran el resultado de la migración europea reciente (Figura 35). Las cuatro muestras de Palenque comparten un haplotipo Y-STR que se encuentra cerca de cuatro de los siete cromosomas R1b-V88 que se encuentran en el grupo étnico punu de Gabón (dos están separados por una mutación de un solo paso y dos por dos pasos). En el mismo clúster está también una muestra de Guinea Ecuatorial y otra de Benín (ambas están a dos pasos de la Palenque muestras). Los resultados obtenidos son compatibles con la presencia de un fundador de África occidental. Los haplotipos cercanos fueron encontrados repartidos a lo largo de las bahías de Benín y Biafra y a lo largo de la costa de Loango. La proximidad genética más cercana fue encontrada con dos muestras de hablantes de punu, poniendo su origen más probable en la región actual de Gabón.

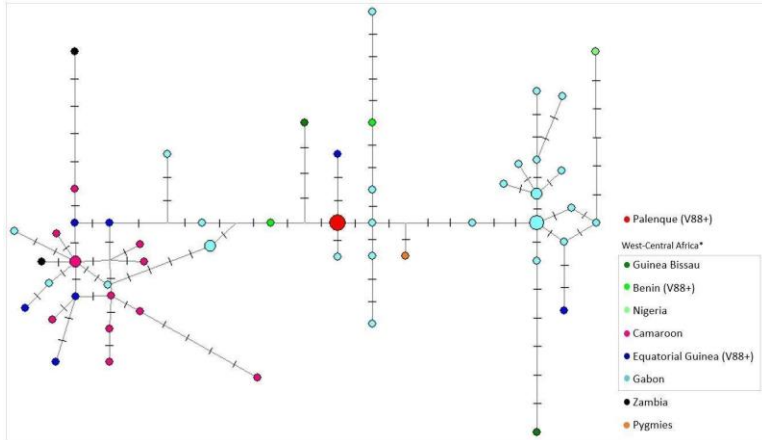
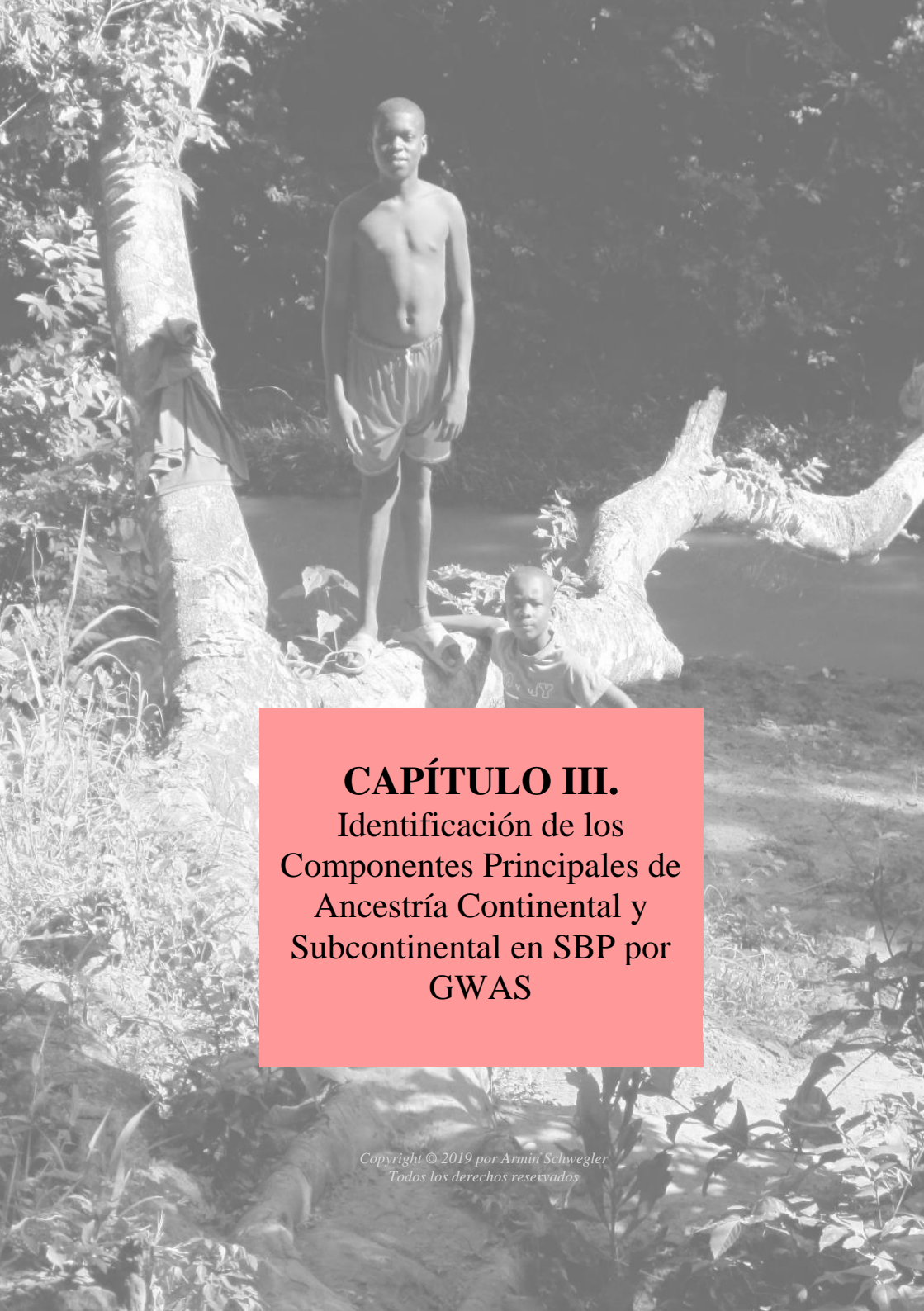


Figura 35. Representación en red de haplotipos seleccionados dentro del clado R1b-P25 (excluidos los asignados a sublinajes fuera de la rama R1b-V88).

Nota. La red se construyó aplicando los métodos *Reduced Median* (Bandelt et al., 1995) y *Median Joining* (Bandelt et al., 1999) secuencialmente. Las muestras se seleccionaron de las publicaciones incluidas en la Tabla complementaria 5. Las muestras seleccionadas son en su mayoría de África occidental, excepto dos muestras de pigmeos y una de Zambia. Dos muestras de Benín y 7 de Guinea Ecuatorial tienen el alelo derivado de V88. Las muestras restantes no se tipificaron para el SNP V88, aunque pertenecen a clados superiores que pueden incluir esta mutación. El número de Y-STR utilizados para construir la red se redujo a un conjunto común de 11 loci: DYS389I, DYS389II, DYS19, DYS390, DYS438, DYS392, DYS437, DYS385a/b, DYS393 y DYS439. La presencia de un alelo intermedio en DYS385 también se consideró en la construcción de la red, codificándola como marcador bialélico. * Las muestras de la región Centro-Oeste fueron discriminadas por país, para una inferencia más clara del origen de las muestras de Palenque.

Se presentaron resultados parciales obtenidos con marcadores de linaje en el simposio “Haploid Markers”, 2018: ANEXO F, (Martínez et al., 2017) (ANEXO E) y la publicación científica (Martínez et al., 2020) (ANEXO B).



CAPÍTULO III.
Identificación de los
Componentes Principales de
Ancestría Continental y
Subcontinental en SBP por
GWAS

6. IDENTIFICACIÓN DE LOS COMPONENTES PRINCIPALES DE ANCESTRÍA CONTINENTAL Y SUBCONTINENTAL EN SBP POR GWAS

Los resultados presentados en este capítulo corresponden a la segunda parte del análisis de ancestría de SBP usando la información obtenida por GWAS, usando herramientas estadísticas de alta resolución autosómica. El manuscrito fue sometido a publicación en la revista científica *Nature Genetics* en febrero del 2021 y debido a la demanda de artículos relacionados con la pandemia COVID 19, se direccionó para la revista *Communications Biology* y recibimos una decisión de revisión a finales del mismo año, y está en espera de publicación (Anexo K).

Las hipótesis planteadas para este capítulo fueron:

- Palenque tiene mayor ancestría africana con menos de europea / nativa que la observada en otras poblaciones afrodescendientes en las Américas.
- Palenque muestra mezcla con sesgo sexual para los linajes masculino y femeninos, dada la dinámica social que caracterizó la conquista y colonización de las Américas.
- Con respecto a la demografía, Palenque tiene una diversidad genética reducida debido al tamaño poblacional pequeño o la presencia de consanguinidad, particularmente por su aislamiento a lo largo de los siglos.
- Los orígenes africanos subcontinentales de Palenque son de distintas regiones de África con altos niveles de diversidad genética y sin evidencia de efecto fundador (no la teoría del sustrato africano único así como un efecto fundador con diversidad genética reducida).

6.1. *Análisis de Ancestría Continental*

Caracterizamos los patrones de ancestría continental de Palenque con base en el análisis de WGS ($n = 34$) y AIMs ($n = 187$). El

WGS de Palenque y Chocó (Figura 36A), junto con otras poblaciones afrodescendientes, se compararon con poblaciones de referencia de África, las Américas y Europa (Tabla complementaria 1 y Figura 17). La proyección del análisis de componentes principales (PCA) de las distancias genómicas muestra que Palenque y Chocó se agrupan cerca del polo africano de la diversidad genética global con evidencia de mezcla de europeos y nativos americanos (Figura 36B). Las poblaciones afroamericanas (Estados Unidos) y de afrocaribeños (Barbados) muestran un patrón algo distinto, con una proximidad similar al polo africano de la PCA, pero extendiéndose principalmente hacia el polo europeo con poca evidencia de ancestría nativa americana. El análisis de ADMIXTURE confirma diferencias.

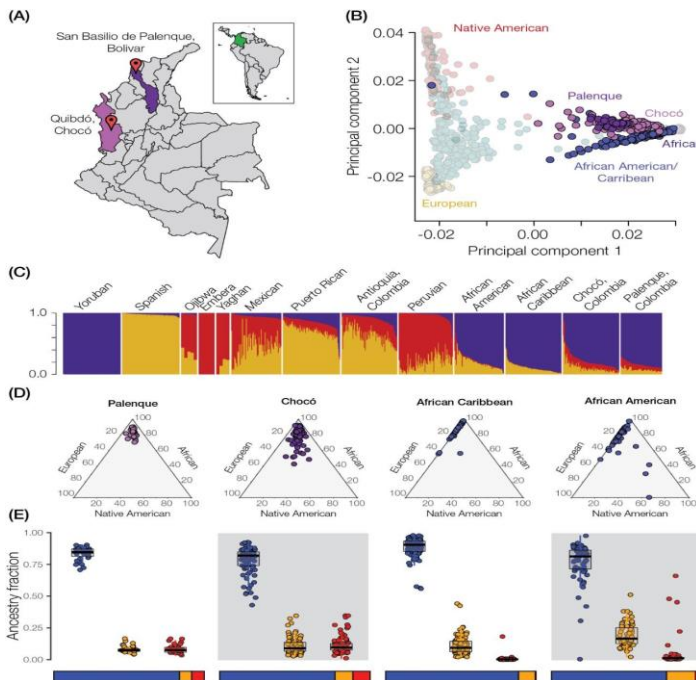


Figura 36. Ancestría continental y mezcla

Nota. (A) Ubicaciones de poblaciones afrocolombianas muestreadas en San Basilio de Palenque y Quibdó, Chocó. (B) PCA que muestra relaciones genéticas entre poblaciones americanas mezcladas junto con poblaciones de referencia africanas, europeas y nativas americanas. Se destacan las poblaciones afrodescendientes de

Colombia, el Caribe (Barbados) y Estados Unidos. (C) Gráfico de ADMIXTURE que muestra las fracciones de ancestría continental: africana (azul), europea (amarillo) y nativa americana (rojo), para las poblaciones de referencia continentales y las poblaciones americanas mezcladas. (D) Gráficos ternarios que muestran las contribuciones relativas de la ancestría continental de los individuos de las cuatro poblaciones afrodescendientes. (E) Distribuciones de ancestría continental y promedios de población para las cuatro poblaciones afrodescendientes.

Las cuatro poblaciones afrodescendientes muestran niveles similares de ancestría africana, con el componente de ancestría no africana que se distingue por partes aproximadamente iguales de la mezcla europea y nativa para las poblaciones colombianas, en comparación con la ancestría principalmente europea para el componente no africano de las poblaciones afroamericanas y afrocaribeñas (Figura 36C y D). Las estimaciones de ancestría continental a partir del análisis de ancestría local realizado con el programa RFMix son casi idénticas a las observadas para el análisis con ADMIXTURE (Figura 37).

El análisis de ancestría continental de Palenque a través de AIMS arrojó niveles similares de ancestría africana, europea y nativa americana (Tabla 6 y Figura 38). Palenque muestra un alto nivel de ancestría africana (WGS = 84%, AIMS = 76%) en comparación con otras poblaciones afrodescendientes en las Américas, pero no excepcionalmente. Las poblaciones cosmopolitas afrodescendientes de Barbados, Jamaica y los Estados Unidos muestran niveles similares o superiores de ancestría africana en comparación con Palenque y las poblaciones de Noir Maroon de la Guayana Francesa y Surinam, que también fueron establecidas por esclavos fugitivos y experimentaron largos períodos de aislamiento posterior, muestran niveles mucho más altos de ancestría africana con poca o ninguna mezcla de europeos y nativos americanos (Consortium, 2015; Fortes-Lima et al., 2017; Mathias et al., 2016).

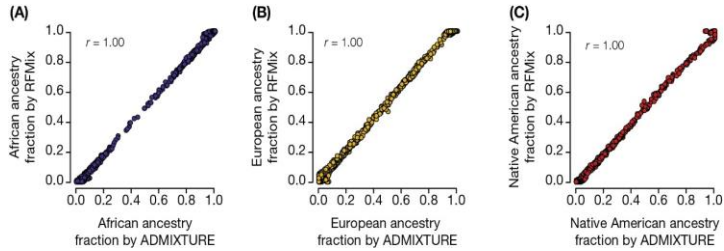


Figura 37. Correlación de estimaciones de ancestría continental.

Nota. Correlación de estimaciones de ancestría continental - africana (azul), europea (amarillo) y nativa americana (rojo) - inferida con RFMix (eje y) y ADMIXTURE (eje x).

Tabla 6. Porcentajes de ancestría continental para las poblaciones afrodescendientes en las Américas.

Población	Referencia	Africana %	Europea %	Nativa americana %
Poblaciones colombianas Afrodescendientes				
Palenque (WGS)	Este estudio	84	8	8
Palenque (AIMs)	Este estudio	76	12	12
Chocó	28855283	77	12	11
Otras poblaciones Afrodescendientes				
Marrones Oscuros	29100086	97	2	1
Jamaica	27725671	89	11	1
Afrocaribeños	26432245	88	11	1
Afroamericanos	26432245	77	19	4
Garifuna	27725671	81	2	17
Brasil	29100086	71	23	5
Brasil	27725671	49	41	10
República Dominicana	27725671	38	52	9
Puerto Rico	26432245	14	72	14

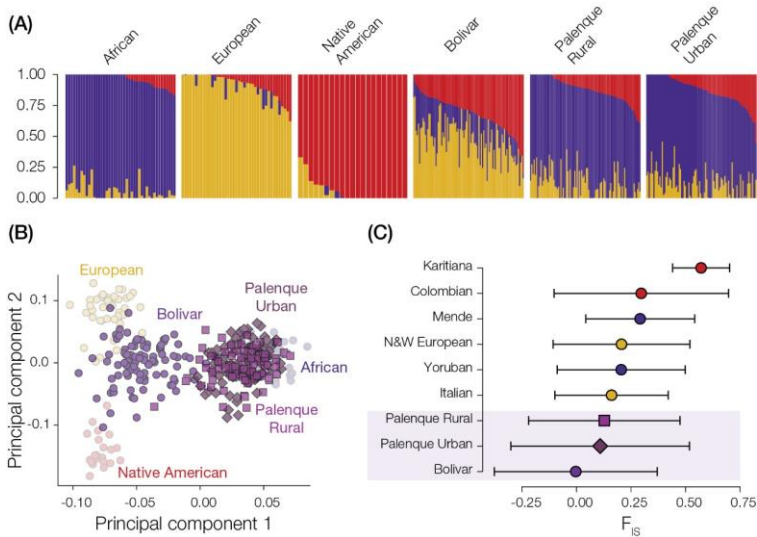


Figura 38. Análisis de marcadores informativos de ancestría (AIM).

Nota. (A) Gráfica ADMIXTURE que muestra estimaciones de ancestría continental - africana (azul), europea (amarillo) y nativa americana (rojo) - para las poblaciones colombianas de Bolívar y Palenque. Las muestras de Palenque se dividen en grupos rurales y urbanos, con base en el muestreo en Palenque o Cartagena, respectivamente. (B) PCA que muestra las relaciones genéticas entre las poblaciones colombianas de Bolívar y Palenque junto con las poblaciones de referencia africanas, europeas y nativas americanas. (C) Valores medios $\pm 2sd$ del coeficiente de consanguinidad (F_{IS}) de Wright para poblaciones americanas mezcladas y poblaciones de referencia globales.

6.2. Análisis de mezcla con sesgo sexual mediante marcadores uniparentales de ADNmt (femenino) y de ChrY (masculino) junto con comparaciones de los niveles de ancestría del cromosoma X frente a los autosómicos.

Un exceso relativo de ancestría de cualquiera de los tres grupos de ancestría para los cromosomas X indica una ancestría con prejuicios femeninos, mientras que un exceso de ancestría autosómica indica una ancestría con prejuicios masculinos. Tanto los haplogrupos uniparentales femeninos (ADNmt, 88%) como masculinos (ChrY, 72%) de Palenque son predominantemente africanos, de acuerdo con los resultados observados para la ancestría del ADN autosómico (Figura 39A). Hay un 12% de haplogrupos de ADNmt de nativos americanos y

no se observan haplogrupos de ADNmt europeos para Palenque, lo que indica una mezcla sustancial con sesgo femenino. El componente no africano del linajes masculinos en Palenque muestra un patrón aún más pronunciado de mezcla sesgada por el sexo con un 28% de haplogrupos europeos en comparación con los haplogrupos nativos americanos. El reanálisis de un conjunto de datos previamente publicados de haplogrupos de ADNmt y de ChrY arrojó resultados similares, con una ancestría casi completamente africana a lo largo del linaje femenino (ADNmt) y un sesgo hacia la ancestría masculina europea para los haplogrupos de ChrY no africanos (Figura 40).

La comparación del cromosoma X frente a la ancestría autosómica también respalda la ancestría femenina con sesgo nativo americano y la ancestría masculina con sesgo europeo en Palenque (Figura 37B). Sin embargo, con mucho, la señal más fuerte de este análisis se observa en la ancestría africana de Palenque con prejuicios masculinos. Este patrón es distinto de lo que se observa en las poblaciones de Chocó, afroamericanos y africanos caribeños, que muestran una combinación de ancestría europea con prejuicios masculinos y ancestría africana con prejuicios femeninos.

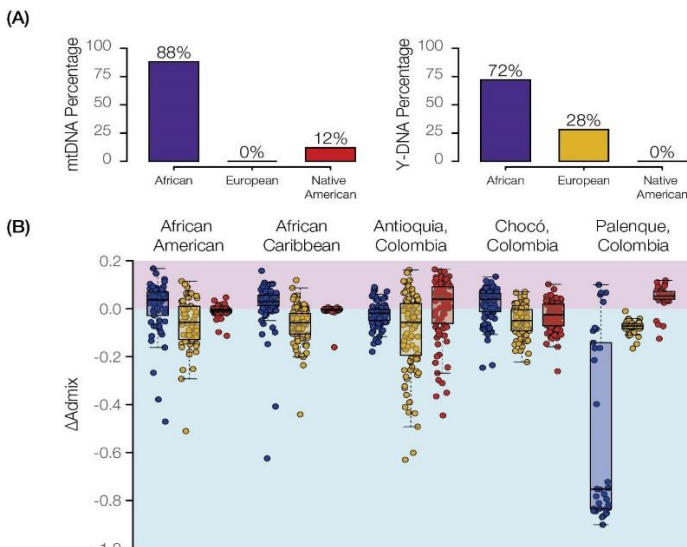


Figura 39. Ancestría y mezcla con sesgo sexual.

Nota. (A) Porcentaje de individuos de Palenque con haplotipos uniparentales africanos, europeos o nativos americanos: ADNmt y cromosoma Y. (B) Distribuciones poblacionales de los valores de Δ admix, que miden las diferencias en la ancestría entre el cromosoma X y los autosomas, para cada ancestría continental: africana (azul), europea (amarillo) y nativa americana (rojo). Los valores por encima de cero indican una mezcla con sesgo femenino y los valores por debajo de cero indican una mezcla con sesgo masculino.

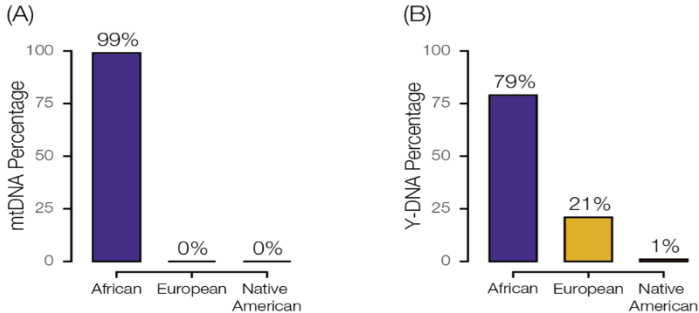


Figura 40. Reanálisis de la ancestría y mezcla con sesgo sexual.

Nota. (A) Porcentaje de individuos de Palenque con haplotipos uniparentales africanos, europeos o nativos americanos: mtDNA (hembra) e Y-DNA (macho). Resultados basados en un nuevo análisis de datos de Ansari-Pour et al. (2016).

6.3. Ancestría genética africana y grupos lingüísticos

Caracterizamos la ancestría genética africana subcontinental en el contexto de la distribución geográfica de los grupos lingüísticos africanos en un esfuerzo por comprender la relación entre la ancestría genética y la lengua criolla palenquera hablada en Palenque. Primero, caracterizamos la ancestría genética a escala fina de 35 poblaciones africanas provenientes de las tres regiones principales que participaron en la trata transatlántica de esclavos hacia Colombia: Senegambia (África Occidental), la Bahía de Benín (África Central Occidental) y la Costa de Loango (Sudoeste de África) (Figura 41A). Se incluyeron poblaciones adicionales de cazadores recolectores de la selva tropical (*Rain Forest Hunter Gatherer*, RFHG) y bantúes del este, para comparar. El eje principal de diversidad genética revelado por PCA divide a los grupos de población de Senegambia y Golfo de Benín del grupo de la costa de Loango, que corresponde a la distinción entre grupos de ancestría africana no bantú y bantú, respectivamente (Figura 41B). Los resultados del análisis de Chromopainter muestran patrones

en gran medida coherentes de ancestría genética dentro de las poblaciones no bantúes de Senegambia y la ensenada de Benín; mientras que las poblaciones de la costa de Loango muestran una estructura poblacional sustancial con cinco subgrupos bantús distintos (Figuras 41C y 42).

Los grupos RFHG y East Bantú muestran sus propios patrones distintos de diversidad genética. Hay dos poblaciones particularmente mezcladas: los Yaundé muestran una mezcla de ancestría no bantú y bantú, consistente con su ubicación geográfica en Camerún entre la Ensenada de Benín y la costa de Loango, y los babongo muestran una mezcla de ancestría RFGH y bantú (Figura 42). El patrón de mezcla visto para babongo es consistente con el hecho de que son considerados como un pueblo RFHG (Mbenga Pygmies); aunque ellos no son particularmente pequeños y hablan una mezcla de varios idiomas bantú diferentes. El análisis de ancestría africana subcontinental con ADMIXTURE arroja patrones muy similares de diversidad genética y estructura de las poblaciones (Figura 43).

A continuación, comparamos los patrones observados de ancestría subcontinental africana con la distribución de grupos de idiomas utilizando el catálogo “Glottolog” de idiomas del mundo (Hammarström, 2021). Los nombres de los idiomas para las poblaciones estudiadas aquí y sus correspondientes ‘Glottocodes’, que pueden usarse para colocarlos en el contexto de idiomas relacionados en el Glottolog, se muestran en la Tabla complementaria 1. Todos los idiomas nativos de las áreas de África consideradas aquí pertenecen al grupo lingüístico NígerCongo, el grupo lingüístico más grande de África en términos de área geográfica, así como en número de hablantes e de idiomas. Los patrones de ancestría en escala fina observados para las poblaciones africanas estudiadas aquí siguen de cerca la distribución de las lenguas Níger-Congo (Figura 41D).

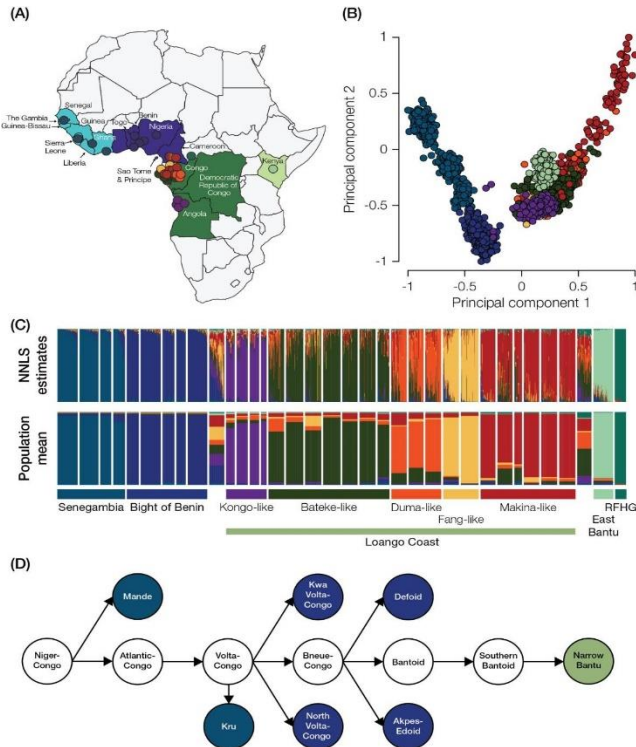


Figura 41. Ancestría genética africana y grupos lingüísticos

Nota. (A) Ubicaciones de las poblaciones africanas de referencia entre las tres principales regiones de muestreo: África occidental (Senegambia, azul claro), África central occidental (Ensenada de Benín, azul oscuro) y África sudoccidental (Costa de Loango, verde). (B) PCA que muestra relaciones genéticas entre poblaciones africanas de referencia (coloreadas como se muestra en el panel A). (C) Análisis cromógrafico que muestra la estructura genética entre las poblaciones africanas de referencia, agrupadas según amplias regiones geográficas y grupos de población. Las estimaciones de ancestría NNLS a nivel individual se muestran junto con las estimaciones promedio de la población. En la Figura 41 se muestra una versión más detallada de este gráfico.

(D) Gráfico que relaciona los grupos lingüísticos de las poblaciones africanas de referencia, codificados por colores según las tres amplias regiones geográficas.

Las poblaciones mandinka, mende y yacouba de Senegambia hablan mande, una rama del idioma básico de la familia Níger-Congo, mientras que la población ahizi habla el idioma kru del Volta-Congo. Las poblaciones del golfo de Benín, bariba y fon de Benín, hablan los

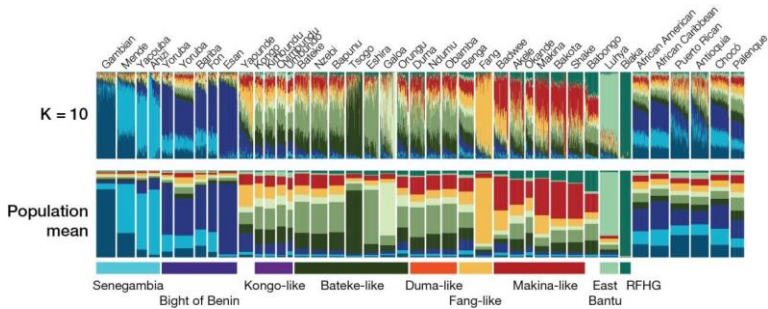


Figura 43. Ancestría y estructura genética para poblaciones africanas de referencia y poblaciones afrodescendientes del Nuevo Mundo.

Nota. Los resultados del análisis ADMIXTURE en $K = 10$ componentes de ancestría se muestran aquí con las etiquetas de población individual y grupal que se muestran en la figura suplementaria 4.

6.3.1. Orígenes genéticos y lingüísticos africanos de Palenque.

Anteriormente se descubrió que la lengua criolla Palenquero de SBP tenía un solo sustrato africano: la lengua yombe (Glottocode: yomb1244) de la región de Pointe-Noire en la República del Congo (Schwegler, 2016, 2017). El yombe es una lengua kikongo, que se ha clasificado en la zona geográfica H de las lenguas bantúes angostas (Figura 44).

Caracterizamos los orígenes subcontinentales africanos de los individuos de Palenque, a través del análisis a escala fina de los datos de WGS de los donantes de Palenque junto con muestras de referencia africanas utilizando *Chromopainter*, para evaluar si la hipótesis del sustrato africano único para el Palenquero también respalda un origen africano único (o estrecho) para el pueblo de Palenque, como se ha sugerido anteriormente (Ansari-Pour et al., 2016). Si bien los individuos de Palenque muestran los niveles más altos de ancestría bantú (costa de Loango), también está claro que tienen contribuciones sustanciales de las tres regiones africanas que participaron en el comercio transatlántico de esclavos hacia Colombia (Figura 32A).

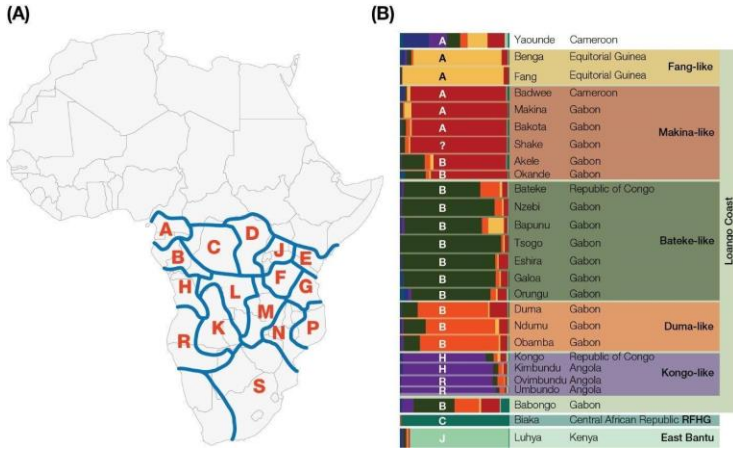


Figura 44. Ancestría genética, geografía y diversidad lingüística de las lenguas bantú.

Nota. (A) Clasificación de Guthrie de las lenguas bantúes. Se muestran las ubicaciones de las dieciséis zonas lingüísticas del bantú estrecho. El mapa se derivó del mapa https://commons.wikimedia.org/wiki/File:World_Map_Blank_-_with_blue_sea.svg, que se distribuye en el dominio público sin restricciones de uso bajo la opción Creative Commons CC, opción “no copyright reserved”. (B) Relación entre la ancestría genética y las zonas lingüísticas Bantú estrechas. Las fracciones de la población se muestran en comparación con sus correspondientes designaciones de zona de idioma bantú (A, B, C, H, J y R). Los grupos de población bantú están organizados y etiquetados como se muestra en la Figura 41C.

Palenque muestra un 51% de ancestría bantú (costa de Loango), seguido de un 35% de ancestría de golfo de Benín y un 14% de ancestría de Senegambia (Tabla 7, Figura 45A). También muestra una diversidad genética sustancial dentro del componente de ancestría genética bantú (Figura 45B). El idioma Yombe corresponde al grupo de población similar al Kongo analizado aquí, que representa el 25% de la ancestría de Palenque, mientras que los grupos de población bantú restantes representan el 26% de la ancestría de Palenque. El análisis de los orígenes genéticos africanos de Palenque utilizando ADMIXTURE y la prueba F3 muestra resultados muy similares, con diversas contribuciones ancestrales de una variedad de poblaciones no bantúes y bantúes (Figuras 43 y 46).

La tradición oral sostiene que Palenque fue establecido por un grupo de esclavos escapados liderados por Benkos Biohó, quien nació en el reino africano de Kongo (Escalante, 1979). Considerado junto con

el sustrato lingüístico único yombe (kikongo) para el Palenquero, es posible que los primeros fundadores de Palenque tuvieran un origen principalmente bantú. Las llegadas posteriores, escapadas o extraídas de Cartagena y las regiones circundantes, pueden haber tenido su origen en otras poblaciones africanas. Para probar esta hipótesis comparamos las distribuciones de los tamaños de los haplotipos de ancestría africana de la costa de Loango (Bantú) en comparación con Senegambia y la Ensenada de Benín. Si la ancestría bantú en Palenque es anterior a la ancestría de otras poblaciones africanas, esperaríamos ver un cambio hacia haplotipos bantú más pequeños, ya que los haplotipos se reducen progresivamente en tamaño con el tiempo debido a la recombinación. Sin embargo, no observamos ninguna diferencia en las distribuciones de los tamaños de los haplotipos africanos de la costa de Loango en comparación con los de Senegambia o de la Ensenada de Benín (Figura 47).

Tabla 7. *Porcentajes de ancestría africana para las poblaciones afrodescendientes en las Américas.*

Poblaciones	Occidente África (Senegambia) %	Centro occidente África (Golfo de Benín) %	Sur occidente África (Costa de Loango) %
<i>Poblaciones afrodescendientes colombianas</i>			
Palenque	14	35	51
Chocó	21	61	18
Africanos del Caribe	18	71	12
Africanos Americanos	24	50	26

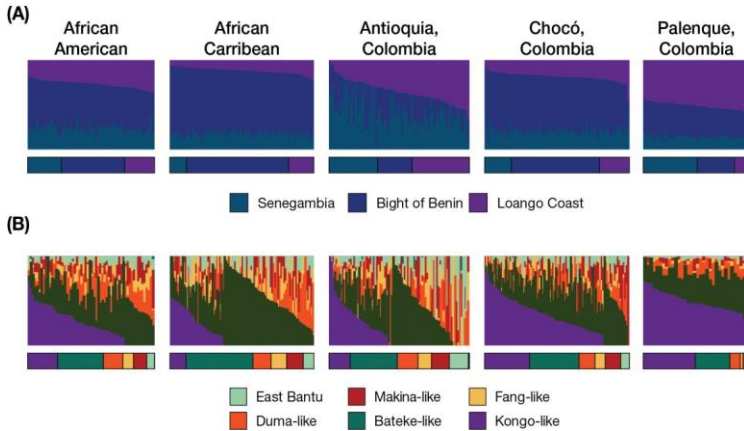


Figura 45. Estructura genética africana y ancestría de poblaciones afrodescendientes en las Américas.

Nota. (A) Estimaciones de ancestría NNLS a nivel individual y promedio de la población para las tres principales regiones de muestreo africanas: Senegambia (azul claro), Ensenada de Benín (azul oscuro), Costa de Loango (violeta). (B) Estimaciones de ancestría NNLS a nivel individual y promedio de la población para el componente de ancestría genética africana de la costa de Loango (Bantú). Los grupos de población bantú específicos (ver Figura 30C) están codificados por colores como se muestra.

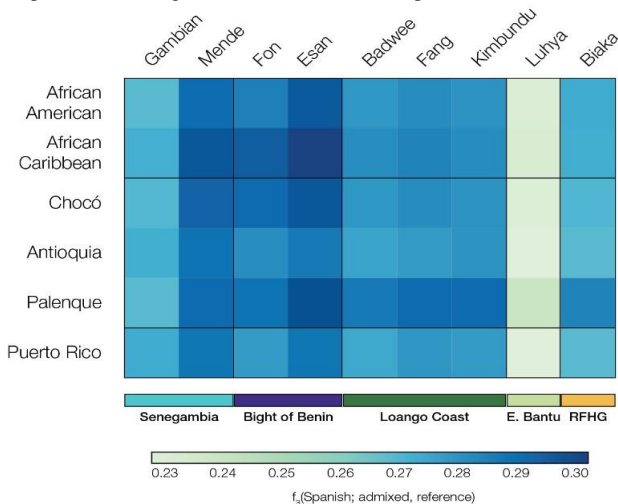


Figura 46. La prueba F_3 muestra similitud entre las poblaciones africanas de referencia y las poblaciones americanas mezcladas.

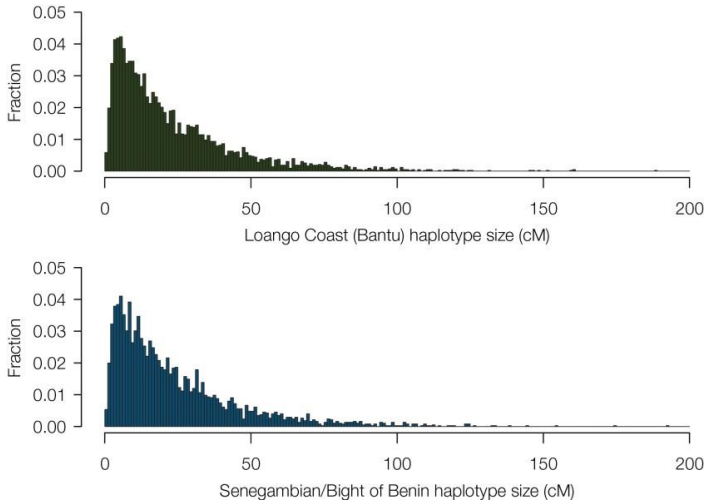


Figura 47. Distribuciones de tamaño de haplotipo (cM) específicas de ancestría africana

Nota. Las distribuciones se comparan para los haplotipos con orígenes en las poblaciones de Loango Coast (Bantu) (arriba, verde) y las poblaciones de Senegambia / Bight of Benín (abajo, azul).

Como se describe en los Métodos, el análisis de ancestría africana a escala fina se realizó utilizando un conjunto de variantes de ancestría africana para evitar la pérdida de sitios genómicos al tiempo que se maximizó el número de poblaciones africanas de referencia. Este conjunto de variantes de ancestría específica se superpone ampliamente con el segundo conjunto de variantes de ancestría nativa americana construido de la misma manera, con genotipos de nativos americanos armonizados con un trasfondo de muestras de referencia global caracterizadas mediante WGS. La consistencia general entre estos dos conjuntos de datos está respaldada por las muy altas correlaciones observadas para las estimaciones de ancestría africana, europea y nativa americana calculadas utilizando ambos conjuntos (Figura 48).

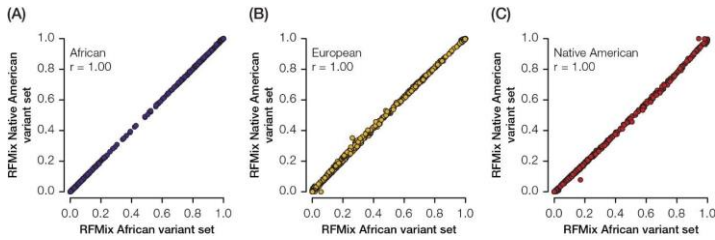


Figura 48. Correlación de estimaciones de ascendencia continental.

Nota. Africana (azul), europea (amarillo) y nativa americana (rojo), inferida con RFMix utilizando el conjunto de variantes de nativos americanos (eje y) y el conjunto de variantes africanas (eje x).

6.3.2. Análisis de marcadores de haplotipos de ChrY para interrogar los orígenes paternos genéticos y lingüísticos africanos de Palenque.

Mediante la comparación con 42 poblaciones africanas de referencia y sus correspondientes grupos de idiomas (Tabla complementaria 1), los hombres de Palenque muestran los haplotipos de ChrY más diversos para cualquiera de estas poblaciones, debido a la presencia de una minoría de haplotipos no africanos juntos con una mezcla de haplotipos característicos de bantúes y no bantúes (Figura 49A). La comparación de las frecuencias poblacionales de los diez haplotipos de cromosoma Y más abundantes, muestra que SBP está igualmente relacionado con la población yombe del Congo y la población chewa de Malawi (Figura 49B). Las poblaciones yombe y chewa hablan lenguas bantúes, aunque bastante distintas: la lengua yombe de la zona costera suroccidental H (Glottocode: yomb1244) y la lengua Nyanga de la zona costera sureste (Glottocode: nyan1308). El análisis de F_{ST} muestra que las poblaciones de Fouban y Wum de Camerún son las más estrechamente relacionadas con Palenque (Figura 49C). Estas dos poblaciones hablan los idiomas Bamoun (Glottocode: bamo1241) y Aghem (Glottocode: aghe1239), respectivamente, de la rama “Wide Grassfields” del grupo de los Bantoides del sur, una rama hermana distinta del “Narrow Bantú” (Figura 50).

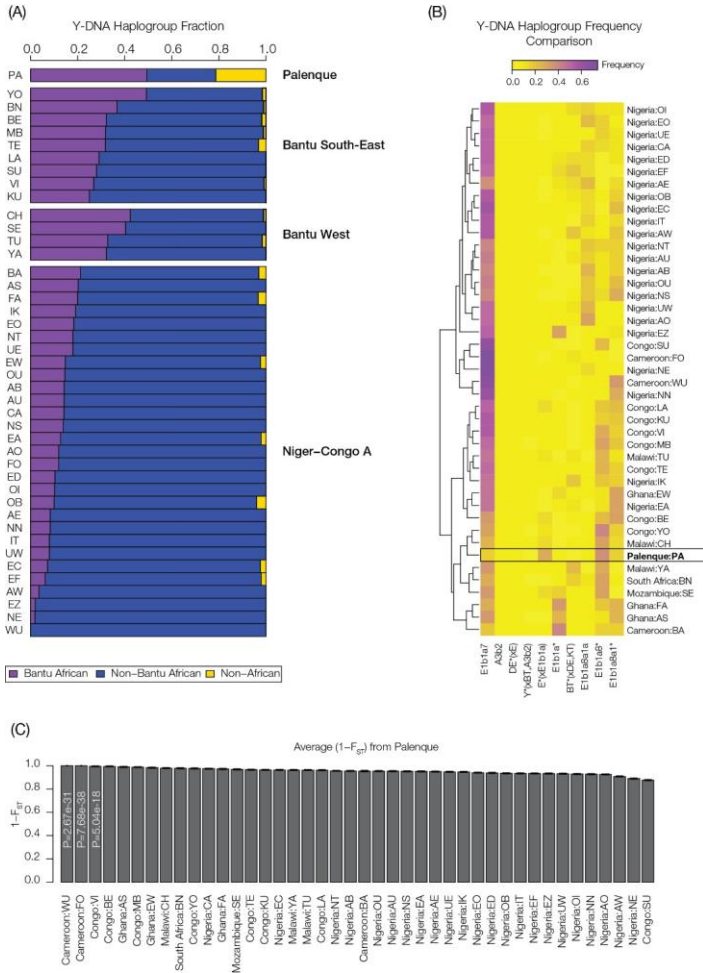


Figura 49. Análisis de haplotipos de *ChrY*

Nota. (A) Fracción de haplotipos de ADN-Y africanos bantús, africanos no bantús y no africanos. (B) Relaciones de población basadas en frecuencias de haplotipos de ADN-Y, codificadas por colores como se muestra. Las abreviaturas de las poblaciones se definen en la Tabla complementaria 1. (C) Similitud entre las poblaciones de referencia africanas y de Palenque según el análisis F_{ST} .

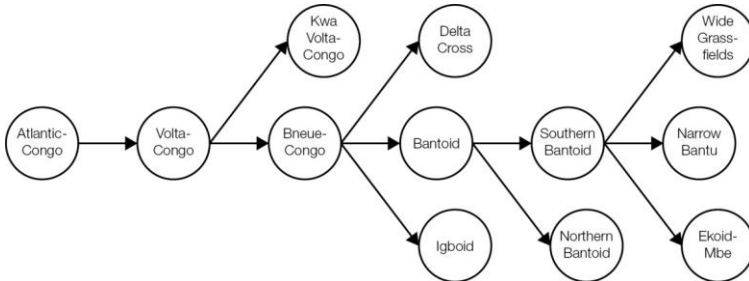


Figura 50. Gráfico que relaciona los grupos lingüísticos de las poblaciones africanas de referencia para los haplotipos de cromosoma Y.

6.3.3. Diversidad genética y efectos fundadores.

SBP fue fundado por un pequeño número de esclavos fugitivos y permaneció aislada como una pequeña población durante siglos, aumentando la posibilidad de efectos fundadores debido a un cuello de botella genético. Comparamos los niveles de diversidad genética en Palenque con otras poblaciones latinoamericanas afrodescendientes y mezcladas para buscar evidencia de un cuello de botella genético. Analizamos tanto el coeficiente de consanguinidad de Wright (FIS) como las tasas de homocigosidad (FROH) como medidas de diversidad genética, o el agotamiento de la misma. FIS mide el agotamiento de la diversidad genética en la generación actual, mientras que FROH mide el agotamiento de la diversidad hace decenas de generaciones. Palenque no muestra niveles elevados de FIS o FROH en comparación con otras poblaciones americanas mezcladas (Figura 51). Esto también se observó utilizando los datos de AIMs para el análisis que mostraron los niveles más bajos de FIS (Figura 38C). En general, la población de Palenque no muestra evidencia de un agotamiento de la diversidad genética consistente con efecto fundador.

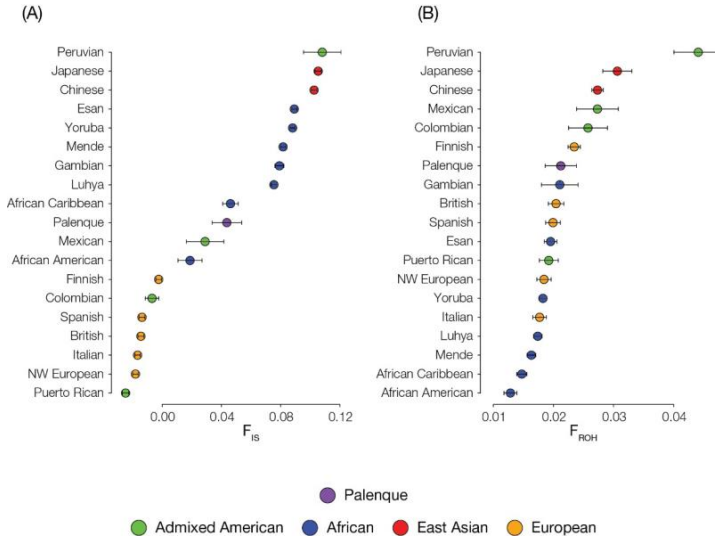


Figura 51. Diversidad genética

Nota. Los valores medios \pm 2sd de (A) el coeficiente de consanguinidad de Wright (FIS) y (B) las rachas de homocigosidad (FROH) se muestran para poblaciones americanas mezcladas y poblaciones de referencia globales, codificadas por colores como se muestra.

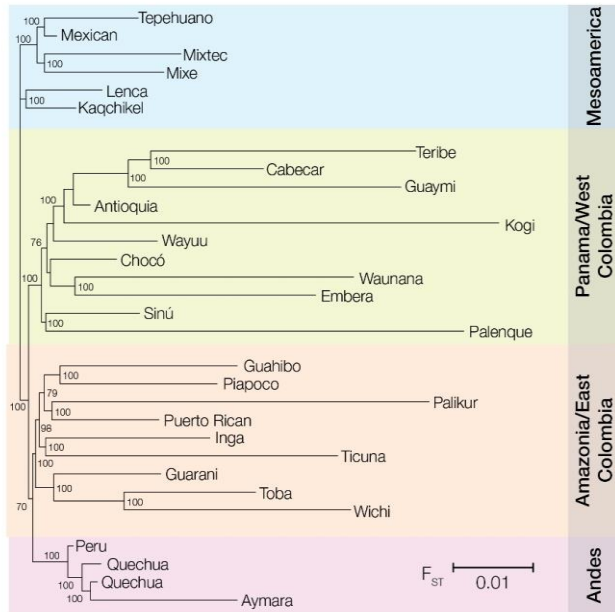
6.4. Ancestría nativa americana y europea

Caracterizamos los componentes subcontinentales de ancestría nativa americana para Palenque utilizando el conjunto de variantes de ancestría nativa americana. La ancestría nativa americana de Palenque se comparó con otras poblaciones latinoamericanas modernas de Colombia, México, Perú y Puerto Rico utilizando un panel de poblaciones indígenas americanas de Mesoamérica, incluidos México, Centroamérica y Sudamérica.

Las poblaciones latinoamericanas modernas muestran patrones de ancestría nativa americana que son concordantes con sus ubicaciones geográficas, consistentes con resultados anteriores (Figura 52) (Homburger et al., 2015; Jordan et al., 2019; Moreno-Estrada et al., 2014; Moreno-Estrada et al., 2013). Grupos de México con poblaciones indígenas mesoamericanas, grupos puertorriqueños con poblaciones amazónicas y grupos de Perú con poblaciones andinas. Las tres

poblaciones colombianas estudiadas aquí, Antioquia, Chocó y Palenque, se agrupan más de cerca con las tribus indígenas de Colombia. Chocó está más estrechamente relacionado con las poblaciones Embera y Waunana de la costa del Pacífico, como se informó anteriormente (Conley et al., 2017), y Palenque está más estrechamente relacionado con la población Sinú de la región de la costa atlántica. La estructura genética del componente nativo americano de las poblaciones colombianas revelada por el análisis ADMIXTURE muestra un patrón similar (Figura 52B). Hay poca diferencia aparente entre las poblaciones colombianas en los componentes de ancestría para $K = 2$ y $K = 4$, pero en $K = 8$ Palenque muestra una clara distinción con más ancestría Sinú y menos ancestría amazónica. El análisis a escala fina de la ancestría de los nativos americanos utilizando la prueba F_3 muestra resultados similares, con poblaciones latinoamericanas modernas que se agrupan con poblaciones de origen indígena según la geografía y Palenque se distingue de las otras poblaciones colombianas por la ancestría Sinú (Figura 53).

(A)



(B)

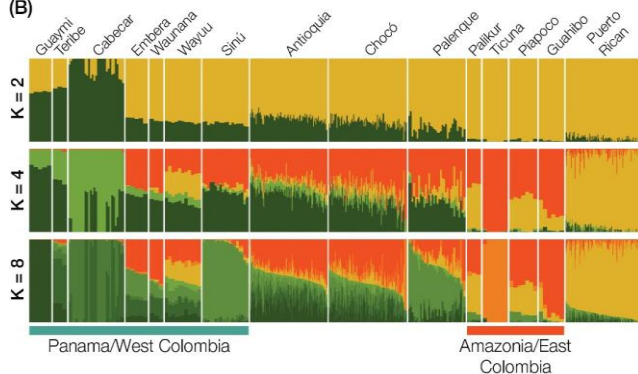


Figura 52. Ancestría de los nativos americanos

Nota. (A) Filogenia que compara el componente de ancestría nativa americana en poblaciones latinoamericanas mezcladas y en poblaciones de referencia nativas americanas. Se indican distintos grupos geográficos de población de referencia de nativos americanos. (B) Análisis ADMIXTURE de poblaciones latinoamericanas mezcladas de Colombia y Puerto Rico en comparación con poblaciones de referencia de nativos americanos colombianos y amazónicos. Los resultados se muestran para diferentes números de componentes de ancestría estimados: $K=2$, $K=4$ y $K=8$.

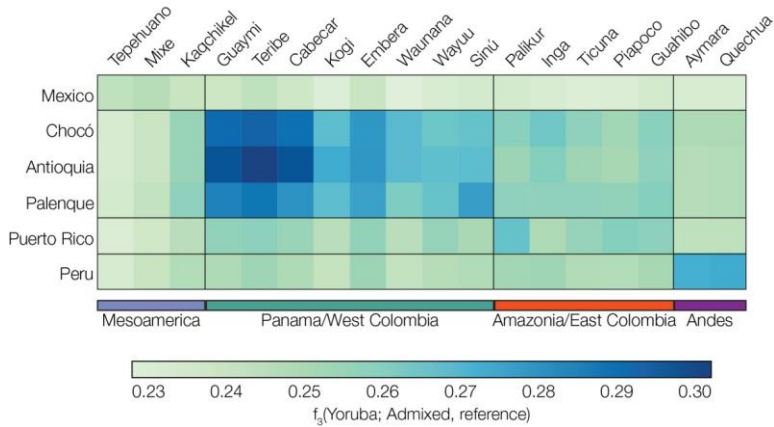



Figura 53. La prueba F_3 muestra similitud entre las poblaciones de referencia nativas americanas y las poblaciones americanas mezcladas.

Todas las poblaciones latinoamericanas modernas estudiadas aquí, incluidas las poblaciones afrocolombianas de Palenque y Chocó, muestran ascendencia europea exclusivamente española, en contraste con la ascendencia europea principalmente británica que se observa en las poblaciones afroamericanas y africanas del Caribe. Si bien este resultado no es en absoluto sorprendente, dada la historia de la colonización de las Américas, sí sirve como un control positivo para nuestro enfoque del análisis de ascendencia subcontinental.



CAPÍTULO IV.
Diferenciación poblacional de
SBP con poblaciones
colombianas mezcladas y
nativas mediante estudios con
HID-InDels

*Copyright © 2019 por Armin Schwegler
Todos los derechos reservados*

7. DIFERENCIACIÓN POBLACIONAL DE SBP CON OTRAS POBLACIONES COLOMBIANAS MEZCLADAS Y NATIVAS MEDIANTE ESTUDIOS CON HID-InDels

Los resultados presentados en este capítulo se encuentran publicados, ver ANEXOS D y F. Con el fin de obtener la información de otros marcadores genéticos en SBP, reportamos inicialmente las frecuencias alélicas de 38 HID-InDels autosómicos en un resumen extendido publicado en la revista científica *FSI Genetics: Supplement Series* (anexo F), en donde también se pueden observar los parámetros estadísticos de eficiencia forense para estos marcadores en SBP. No se encontraron desviaciones significativas a las expectativas de Hardy-Weinberg, excepto en el marcador rs10629077 ($p = 0,0002$).

Además, los datos resultantes fueron comparados con los obtenidos para otras poblaciones mezcladas colombianas (anexo D). Los resultados de la genotipificación para todas las muestras incluidas en este capítulo son listados en la tabla complementaria 8.

La hipótesis para este capítulo fue:

Palenque tiene frecuencias alélicas del 38 InDels cercanas a las de la Región Caribe de Colombia.

7.1. Comparación poblacional.

Se realizó un primer análisis comparativo entre las muestras de este estudio para los departamentos de Antioquia, Cundinamarca, Norte de Santander y Chocó y las publicadas por Pereira et al. (2018) para los mismos departamentos. Como no se encontraron diferencias significativas ($FST \leq 0,00105$; $p \geq 0,13271$), se agruparon las muestras del mismo departamento para el resto de los análisis (tabla complementaria 9).

7.2. Representación MDS de la matriz F_{ST} obtenida en comparaciones entre las muestras de 10 poblaciones colombianas y dos poblaciones de referencia de África y Europa.

Se calcularon los valores de F_{ST} por pares y las probabilidades de no diferenciación entre todos los departamentos con más de 20 muestras.

Entre las poblaciones andinas, aunque se obtuvieron diferencias estadísticamente significativas en algunas comparaciones, éstas mostraron valores de F_{ST} bajos (por debajo de 0,006), excepto en las comparaciones que involucran a Nariño (entre 0,0064 y 0,0147). El AMOVA realizado para las poblaciones andinas en un solo grupo dio un F_{ST} de 0,0032, que disminuyó cuando no se consideró Nariño ($F_{ST} = 0,0027$). Aunque, las diferencias obtenidas son bajas entre Nariño y las restantes poblaciones andinas esto puede deberse a una mayor ancestría nativa americana en esta población, ya que fue el departamento andino que mostró menores distancias genéticas con los grupos nativos colombianos incluidos en el análisis.

Las muestras de los distintos departamentos del Caribe no revelaron diferencias significativas para los marcadores estudiados. Tampoco se encontraron diferencias significativas entre los dos departamentos de la región de la Orinoquia. Se realizó un nuevo análisis tras agrupar las muestras de cada una de las 5 regiones naturales de Colombia, a saber, las regiones andina, caribeña, pacífica, Orinoquia y amazónica, y utilizando Europa y África como poblaciones de referencia (Tabla 8). Los resultados se representaron en un gráfico MDS (Figura 54)

Tabla 8. Distancias genéticas por pares (FST) entre las poblaciones colombianas, europea y africana (diagonal inferior) y los correspondientes valores de probabilidad de no diferenciación (diagonal superior).

	Referencias	Poblaciones mezcladas						Grupos nativos			Afro-desc.
	Europe Africa	Andes	Caribe	Pacific	Orinoquía	Amazonia	Barí-Motilón	Nariño	Guainía	Pijao	Palenque
Europe	* <SE-06	<SE-06	<SE-06	<SE-06	<SE-06	<SE-06	<SE-06	0,1048	* <SE-06	<SE-06	<SE-06
Africa	<SE-06	<SE-06	<SE-06	<SE-06	<SE-06	0,0613	0,1185	* <SE-06	<SE-06	<SE-06	<SE-06
Andes	0,0614	0,0939	0,0045	*	<SE-06	<SE-06	<SE-06	<SE-06	<SE-06	<SE-06	<SE-06
Caribe	0,0764	0,0606	0,0438	0,0233	*	<SE-06	<SE-06	<SE-06	<SE-06	<SE-06	<SE-06
Pacific	0,0648	0,1183	0,0018	0,0053	0,0415	* <SE-06	<SE-06	<SE-06	<SE-06	0,0004	0,0962
Orinoquia	0,0128	* <SE-06	0,1895	0,0053	0,0790	0,1989	0,2400	0,0951	0,0994	0,1560	0,0973
Amazonia	<SE-06							0,0881	* <SE-06	<SE-06	<SE-06
Barí-Motilón	0,1223	0,1818	0,0240	0,0322	0,0787	0,0181	0,0048	0,0743	*	0,0638	0,0347
Nariño	0,1385	0,2003	0,0395	0,0475	0,0994	0,0304	0,0172	0,0978	0,0099	* 0,0007	0,0783
Guainía										0,1465	0,0103
Pijao	0,0114	0,0066	0,0772	0,0106	0,0241	*				0,0189	0,0669
Palenque	0,1031	0,0631	0,0717	0,0434	0,0141	0,0692	0,0866	0,1844	0,1098	0,1289	0,0983
											*

Nota: En rojo se indican los valores de p significativos (P = 0,00022; después de aplicar la corrección de Bonferroni), después de 10100 permutaciones (s.e. <0.0039).

Aunque los 38 HID-InDels fueron seleccionados para maximizar la variación genética dentro de las poblaciones, se observó una clara separación de las diferentes poblaciones continentales (europeos, africanos y nativos americanos). Las muestras de Colombia están muy dispersas en el MDS, a excepción de Andes y Orinoquia que presentan una distancia genética baja (no significativa).

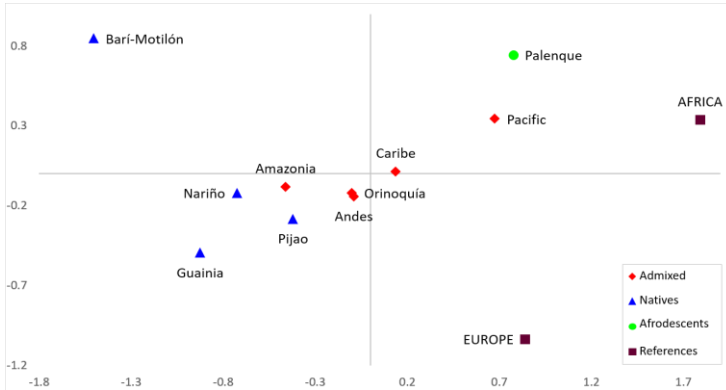


Figura 54. Representación MDS de la matriz de F_{ST} obtenida en las comparaciones por pares entre 10 muestras poblacionales de Colombia, y dos poblaciones de referencia de África y Europa.

Las muestras del Pacífico y Palenque están más cerca de África que de las poblaciones nativas o europeas. Sin embargo, Palenque muestra un cambio a la muestra de referencia africana, que puede explicarse por la deriva genética. Los signos de deriva genética también se percibieron en los nativos Barí-Motilón. La población de la Amazonía está más cerca de los grupos nativos que de las otras poblaciones mezcladas, lo que se espera teniendo en cuenta que ha sido reportada para esta región una mayor ancestría nativa (62%) en comparación con las otras regiones (entre 14% y 26%) (Ossa et al., 2016).

7.3. Frecuencias alélicas, equilibrio de Hardy-Weinberg y parámetros forenses

Las frecuencias alélicas se calcularon para todas las poblaciones estudiadas, tras agrupar los datos de los individuos

mezclados de las 5 regiones de Colombia (Tabla complementaria 8). Se detectaron desviaciones a las expectativas de equilibrio de Hardy-Weinberg para tres loci en la población andina, dos loci en el Chocó y uno en la Amazonia (Tabla complementaria 8). En la mayoría de los casos, las desviaciones se asociaron con un exceso de homocigotos, lo que puede indicar la presencia de cierta heterogeneidad dentro de las poblaciones de estas tres regiones.

Se calcularon los parámetros de relevancia forense en todas las muestras estudiadas (Tabla 9). Los valores medios de PD entre las poblaciones variaron entre 0,370 para mid-1383 y 0,589 para mid-2719 y mid-1493. Para el PE, los valores medios entre poblaciones variaron entre 0,065 para mid-1643 y 0,214 para mid-785. La población de la región del Pacífico mostró la mayor probabilidad de coincidencia acumulada para los 38 HID-InDels (MP=6,67E-10), y el valor más bajo se encontró en Caribe (MP=2,43E-15).

En cuanto a los valores acumulados de PE, estos variaron entre 0,98 en los indígenas Barí-Motilón y 0,9986 en los indígenas del Guainía. Los resultados obtenidos en todos los grupos poblacionales mostraron que este conjunto de InDels es una herramienta útil para la identificación de individuos en Colombia. Sin embargo, tiene un poder de exclusión limitado y debe ser utilizado para complementar la información de otros marcadores en casos de paternidad.

Tabla 9. Parámetros de relevancia forense en todas las muestras estudiadas para los 38 HID-InDels.

Locus	Aha1 (c=1793)			Ccr3 (c=170)			Pde1c (Ctact) (c=103)			Ctawgata (c=86)			Aumata1 (c=73)			Eras-Mobile (c=72)			Ntains_Vardis (c=71)			Ntains_Grains (c=67)			Rise1 (c=59)			Phaema1 (c=59)					
	FD	FE	TFH	FD	FE	TFH	FD	FE	TFH	FD	FE	TFH	FD	FE	TFH	FD	FE	TFH	FD	FE	TFH	FD	FE	TFH	FD	FE	TFH						
B1/ msd-2719	0.622	0.181	0.986	0.391	0.192	1.010	0.630	0.178	0.977	0.619	0.188	1.000	0.200	0.011	0.565	0.549	0.132	0.875	0.623	0.041	0.656	0.583	0.248	1.143	0.624	0.233	1.083	0.561	0.099	0.798			
B2/ msd-185	0.616	0.165	0.949	0.624	0.139	1.027	0.391	0.213	1.059	0.600	0.168	0.956	0.623	0.168	0.956	0.374	0.094	0.785	0.624	0.179	1.000	0.549	0.232	1.133	0.590	0.188	1.000	0.623	0.183	0.990			
B3/ msd-1463	0.624	0.165	0.930	0.569	0.102	0.866	0.459	0.102	0.807	0.571	0.256	1.162	0.395	0.141	0.895	0.450	0.039	0.650	0.474	0.153	0.808	0.566	0.043	0.667	0.568	0.188	1.000	0.462	0.070	0.627			
B4/ msd-1504	0.623	0.173	0.967	0.565	0.136	0.882	0.447	0.097	0.794	0.649	0.126	0.860	0.395	0.148	0.851	0.107	0.000	0.511	0.449	0.132	0.875	0.616	0.248	1.143	0.575	0.146	0.929	0.337	0.034	0.659			
B5/ msd-320	0.574	0.096	0.792	0.500	0.106	0.813	0.500	0.126	0.860	0.465	0.088	0.773	0.302	0.028	0.619	0.400	0.027	0.618	0.297	0.028	0.615	0.385	0.028	0.619	0.618	0.183	0.990	0.618	0.030	0.733	0.853		
B6/ msd-3205	0.625	0.130	0.920	0.625	0.165	0.949	0.423	0.136	0.860	0.621	0.277	1.234	0.617	0.248	1.143	0.617	0.248	1.143	0.616	0.248	1.143	0.615	0.248	1.144	0.623	0.141	1.000	0.623	0.141	1.003	0.623	0.141	1.003
B7/ msd-3221	0.625	0.130	0.920	0.619	0.140	0.893	0.404	0.172	0.865	0.621	0.163	0.940	0.380	0.223	1.083	0.610	0.238	1.167	0.616	0.238	1.167	0.610	0.238	1.167	0.610	0.238	1.167	0.610	0.238	1.167	0.610	0.238	1.167
B8/ msd-3668	0.586	0.130	0.968	0.624	0.132	1.056	0.396	0.242	1.023	0.434	0.198	0.743	0.410	0.183	0.843	0.440	0.078	0.750	0.214	0.031	0.771	0.300	0.028	0.619	0.622	0.143	0.900	0.622	0.143	0.900	0.622	0.143	0.900
B9/ msd-2010	0.574	0.132	0.875	0.497	0.131	0.875	0.415	0.102	0.807	0.434	0.118	0.843	0.395	0.141	0.895	0.470	0.039	0.687	0.619	0.105	0.813	0.619	0.105	0.813	0.619	0.105	0.813	0.619	0.105	0.813	0.619	0.105	0.813
B10/ msd-1116	0.623	0.130	0.982	0.597	0.136	0.882	0.448	0.135	0.882	0.434	0.298	1.265	0.623	0.244	1.133	0.624	0.200	1.040	0.381	0.132	0.875	0.621	0.188	1.000	0.621	0.188	1.000	0.621	0.188	1.000	0.621	0.188	1.000
B11/ msd-783	0.615	0.166	0.951	0.495	0.072	0.735	0.399	0.140	0.892	0.376	0.188	1.000	0.621	0.163	0.944	0.605	0.163	0.929	0.624	0.179	1.000	0.624	0.179	1.000	0.616	0.129	0.867	0.723	0.116	0.859	0.723	0.116	0.859
B12/ msd-1111	0.602	0.139	0.935	0.545	0.127	0.862	0.352	0.052	0.684	0.398	0.159	0.935	0.623	0.168	0.944	0.385	0.094	0.785	0.615	0.209	1.000	0.616	0.248	1.143	0.598	0.097	0.722	0.448	0.085	0.717	0.448	0.085	0.717
B13/ msd-1209	0.489	0.071	0.732	0.397	0.155	0.926	0.408	0.133	0.871	0.497	0.092	0.782	0.312	0.074	0.739	0.073	0.001	0.200	0.440	0.041	0.656	0.116	0.003	0.333	0.339	0.129	0.867	0.620	0.120	0.853	0.620	0.120	0.853
B14/ msd-1514	0.549	0.087	0.771	0.310	0.091	0.781	0.407	0.213	1.059	0.440	0.104	0.811	0.312	0.074	0.739	0.073	0.001	0.200	0.440	0.041	0.656	0.116	0.003	0.333	0.339	0.129	0.867	0.620	0.120	0.853	0.620	0.120	0.853
B15/ msd-2010	0.622	0.161	0.939	0.555	0.095	0.789	0.408	0.180	0.884	0.469	0.244	1.132	0.621	0.061	0.708	0.620	0.200	1.040	0.602	0.209	1.000	0.624	0.248	1.143	0.622	0.052	0.684	0.451	0.049	0.678	0.451	0.049	0.678
B16/ msd-1384	0.600	0.131	0.871	0.375	0.114	0.833	0.397	0.216	1.066	0.391	0.178	0.977	0.623	0.188	1.000	0.602	0.168	0.955	0.593	0.168	0.955	0.582	0.232	1.133	0.580	0.105	0.813	0.623	0.241	1.125	0.623	0.241	1.125
B17/ msd-743	0.625	0.191	1.008	0.616	0.193	1.014	0.438	0.275	1.210	0.374	0.178	0.977	0.606	0.121	0.850	0.380	0.129	0.867	0.602	0.132	0.875	0.508	0.069	0.727	0.619	0.146	0.929	0.574	0.123	0.853	0.574	0.123	0.853
B18/ msd-3097	0.539	0.092	0.782	0.313	0.110	0.824	0.405	0.164	0.847	0.459	0.092	0.782	0.606	0.141	0.895	0.385	0.052	0.684	0.503	0.132	0.875	0.508	0.138	0.889	0.380	0.146	0.929	0.615	0.143	0.900	0.615	0.143	0.900
B19/ msd-1782	0.591	0.137	0.886	0.371	0.131	0.872	0.461	0.095	0.790	0.420	0.159	0.935	0.318	0.051	0.650	0.311	0.075	0.743	0.474	0.058	0.700	0.583	0.138	0.889	0.371	0.105	0.813	0.588	0.136	0.884	0.588	0.136	0.884
B20/ msd-3648	0.542	0.102	0.805	0.623	0.260	1.172	0.479	0.102	0.807	0.429	0.159	0.935	0.362	0.141	0.895	0.419	0.039	0.650	0.474	0.058	0.700	0.597	0.138	0.889	0.605	0.094	0.765	0.723	0.116	0.859	0.723	0.116	0.859
B21/ msd-3890	0.578	0.120	0.847	0.609	0.218	1.071	0.434	0.118	0.843	0.386	0.159	0.935	0.62	0.188	1.000	0.547	0.116	0.839	0.602	0.132	0.875	0.623	0.210	2.000	0.339	0.094	0.765	0.380	0.061	0.707	0.380	0.061	0.707
B22/ msd-3277	0.521	0.083	0.762	0.387	0.127	0.862	0.301	0.068	0.726	0.475	0.126	0.860	0.310	0.074	0.739	0.394	0.143	0.897	0.528	0.103	0.808	0.423	0.069	0.727	0.450	0.052	0.684	0.178	0.069	0.756	0.178	0.069	0.756
B23/ msd-1383	0.586	0.120	0.847	0.608	0.136	0.882	0.392	0.210	1.021	0.478	0.178	0.977	0.620	0.030	0.680	0.400	0.041	0.200	0.620	0.088	0.553	0.000	0.000	0.300	0.346	0.039	0.650	0.604	0.175	0.971	0.604	0.175	0.971
B24/ msd-1824	0.591	0.145	0.903	0.371	0.122	0.832	0.431	0.187	0.953	0.413	0.168	0.956	0.620	0.313	1.308	0.477	0.097	0.722	0.620	0.238	1.167	0.621	0.188	1.000	0.611	0.146	0.929	0.395	0.166	0.952	0.395	0.166	0.952
B25/ msd-3114	0.589	0.145	0.904	0.424	0.232	1.103	0.429	0.124	0.857	0.391	0.118	0.843	0.494	0.088	0.773	0.621	0.171	0.863	0.581	0.132	0.875	0.566	0.043	0.667	0.598	0.146	0.929	0.395	0.166	0.952	0.395	0.166	0.952
B26/ msd-834	0.615	0.142	0.897	0.375	0.171	0.862	0.401	0.201	1.030	0.466	0.118	0.843	0.380	0.074	0.739	0.602	0.143	0.897	0.528	0.103	0.808	0.508	0.138	0.889	0.622	0.188	1.000	0.614	0.146	0.929	0.614	0.146	0.929
B27/ msd-1643	0.522	0.081	0.703	0.353	0.182	0.867	0.378	0.186	0.987	0.453	0.086	0.768	0.351	0.088	0.773	0.444	0.001	0.200	0.400	0.008	0.553	0.397	0.028	0.615	0.450	0.094	0.765	0.432	0.037	0.687	0.432	0.037	0.687
B28/ msd-1945	0.552	0.101	0.803	0.623	0.182	0.987	0.429	0.102	0.807	0.429	0.159	0.935	0.621	0.163	0.944	0.380	0.129	0.867	0.602	0.132	0.875	0.566	0.028	0.615	0.366	0.031	0.711	0.555	0.105	0.813	0.555	0.105	0.813
B29/ msd-1643	0.600	0.142	0.897	0.623	0.140	0.893	0.419	0.177	1.214	0.378	0.111	0.827	0.619	0.074	0.739	0.594	0.171	0.863	0.566	0.088	0.700	0.623	0.222	1.133	0.624	0.310	1.300	0.614	0.183	0.990	0.614	0.183	0.990
B30/ msd-292	0.621	0.197	1.021	0.624	0.150	0.915	0.497	0.147	0.868	0.437	0.147	0.868	0.619	0.074	0.739	0.594	0.171	0.863	0.566	0.088	0.700	0.623	0.222	1.133	0.624	0.310	1.300	0.614	0.183	0.990	0.614	0.183	0.990
B31/ msd-449	0.611	0.172	0.964	0.614	0.160	0.938	0.557	0.037	0.646	0.416	0.220	1.075	0.623	0.104	0.810	0.624	0.248	1.143	0.597	0.132	0.875	0.597	0.148	1.143	0.568	0.129	0.867	0.415	0.049	0.678	0.415	0.049	0.678
B32/ msd-1330	0.594	0.134	0.875	0.623	0.118	0.701	0.407	0.144	0.802	0.405	0.159	0.935	0.620	0.163	0.944	0.624	0.248	1.143	0.549	0.078	0.750	0.597	0.248	1.143	0.568	0.129	0.867	0.415	0.049	0.678	0.415	0.049	0.678
B33/ msd-1830	0.622	0.066	0.769	0.621	0.199	1.027	0.394	0.172	0.865	0.516	0.075	0.741	0.494	0.088	0.773	0.073	0.001	0.200	0.474	0.038	0.608	0.116	0.003	0.333	0.346	0.039	0.650	0.615	0.132	0.854	0.615	0.132	0.854
B34/ msd-3220	0.522	0.083	0.706	0.386	0.131	0.872	0.377	0.216	1.066	0.391	0.178	0.977	0.620	0.030	0.680	0.400	0.041	0.200	0.620	0.088	0.553	0.000	0.000	0.300	0.346	0.039	0.650	0.604	0.175	0.971	0.604	0.175	0.971
B35/ msd-1997	0.559	0.101	0.804	0.359	0.101	0.804																											

7.4. Comparación del poder de exclusión de 38 HID-InDels y 46 AIMs-InDels en poblaciones colombianas

Un total de 1194 individuos incluidos en el presente trabajo fueron analizados previamente para 46 AIM-InDels (Ossa et al., 2016). Los 46 AIM-InDels fueron seleccionados por tener una alta variación entre los diferentes continentes y no entre los individuos de un solo continente. Sin embargo, debido a la naturaleza mezclada de las poblaciones de las cinco regiones de Colombia, esperamos que estos loci tengan valores de diversidad más altos que en poblaciones no mezcladas. Por lo tanto, para comparar la informatividad de los dos conjuntos de InDels, se calcularon los parámetros forenses utilizando las distribuciones de frecuencias alélicas en la submuestra de 1194 individuos con genotipo para los dos conjuntos de InDels (Tabla 10). También se calcularon los mismos parámetros para el conjunto completo de 83 loci InDel (un locus está incluido en ambos multiplex).

Los resultados obtenidos muestran que tanto la PD como la PE acumuladas son más altas para los 46 AIM-InDels que para los 38 HID-InDels, excepto para la PE en las poblaciones de las regiones de la Amazonia y del Pacífico. No obstante, los valores medios de PD y PE por locus fueron siempre más altos para los 38 HID-InDels, lo que demuestra que el número de loci en los 46 AIM-InDel es el único factor que contribuye al mayor poder informativo. Al considerar los resultados de los dos conjuntos de InDels juntos, se observan altos valores de PD y PE acumulados, con probabilidades de coincidencia inferiores a $2,69E-28$.



CAPÍTULO V. Discusión

*Copyright © 2019 por Armin Schwegler
Todos los derechos reservados*

8. RAÍCES AFRICANAS EN SBP POR MARCADORES UNIPARENTALES

8.1. Ancestría y raíces africanas de San Basilio de Palenque encontradas con marcadores uniparentales.

Mediante este trabajo, analizamos numerosos marcadores específicos de cromosoma Y y secuenciamos toda la RC del ADNmt para obtener una información precisa del origen continental de los linajes en Palenque. Aunque ya existían estudios previos con estimaciones ancestrales, estas eran aproximadas debido a una pobre resolución de los datos (los reportados por Arnaiz-Villena et al, 2018 con alelos HLA, Noguera et al, 2014 con ChrY y Ansari Pour et al, 2016 con ADNmt y ChrY).

Este estudio encontró una alta ancestría materna africana, que está de acuerdo con lo reportado previamente por (Ansari-Pour et al., 2016). Sin embargo, no fue posible realizar una comparación simple entre los dos estudios, ya que los datos anteriores solo incluían la información de la región hipervariable HVSI del ADNmt y muchas muestras no pudieron asignarse a un haplogrupo específico, es decir, dentro de algunas ramas africanas y nativas americanas. Aun así, una comparación entre los haplotipos en HVSI no mostró diferencias estadísticamente significativas entre los dos estudios ($F_{ST} = 0,0062$; $P = 0,1278 \pm 0,004$).

A diferencia del ADNmt, en el cromosoma Y se observó una gran afluencia de hombres europeos y solo tres con un origen nativo americano. Aunque en el trabajo de (Noguera et al., 2014) se reportó una entrada de linajes masculinos europeos aún mayor (38,5%), debido a que se atribuyó un origen europeo a todos los linajes R1b, en el presente estudio encontramos muestras en Palenque que pertenecen a un subhaplogrupo africano R1b, caracterizado por el alelo derivado de V88, que no fue analizado en el estudio previo de Noguera et al. (2014). Debido a las resoluciones tan diferentes en la definición de los haplogrupos, no pudimos hacer una comparación significativa entre nuestros datos y los de Ansari-Pour et al. (2016). A diferencia de este estudio, Ansari-Pour et al. (2016) realizaron una caracterización más

detallada de los sublinajes E1b1a-M2 (incluidos los marcadores U175, U181 y U290) pero no diferenciaron los haplogrupos africanos, europeos o nativos americanos dentro de los clados BF-M213* (xM9) y QR-92R7* (xSRY10831.2).

Los resultados generales muestran que la mezcla no africana fue mediada principalmente por hombres europeos, responsables de la introducción de al menos 9 haplogrupos diferentes presentes en nuestra muestra, y no se detectó mezcla con mujeres europeas. La afluencia de nativos americano fue mayor para los linajes maternos que para los paternos. Los tres cromosomas Y nativos americanos detectados podrían explicarse por una sola entrada en SBP, ya que sus haplotipos difieren solo en DYF387S1, que se sabe que tiene una tasa de mutación superior a la media observada en Y-STR ($> 1 \times 10^{-2}$) (Ballantyne et al., 2010).

Uno de los temas más debatidos sobre la historia de SBP ha sido el origen de sus habitantes, y de allí la pregunta de si sus antepasados tenían un origen único o de múltiples sitios en África. La hipótesis de un solo origen bakongo de los fundadores de Palenque fue propuesta en el estudio de Ansari-Pour et al. (2016), basada en evidencia lingüística y apoyada en la diversidad observada en ChrY. Hicieron un análisis comparativo entre SBP y poblaciones que representaban diferentes grupos de África subsahariana que mostró una distancia genética F_{ST} estrecha con el grupo yombe de la República del Congo para ChrY. Sin embargo, con respecto a los linajes de ADNmt, no encontraron evidencia sobre las raíces africanas maternas. Por el contrario, nuestros resultados basados en distancias genéticas por pares, no nos permitieron ubicar el origen de los haplotipos de Y de Palenque en una región africana específica. El análisis de AMOVA realizado entre Palenque en un grupo y todas las poblaciones africanas en un segundo grupo, reveló una variación mayor entre Palenque (submuestra de linajes africanos) y el grupo africano (1,76%) que entre poblaciones de diferentes regiones africanas (0,88%). Esta mayor distancia F_{ST} encontrada entre el sustrato africano de Palenque y todas las poblaciones africanas solo puede explicarse por efecto de deriva génica, que está de acuerdo con la baja diversidad de haplotipos dentro de los haplogrupos de ChrY.

Con respecto a los linajes de ADNmt, similar a las observaciones de Ansari-Pour et al. (2016), no fue posible establecer raíces africanas maternas con base en análisis de distancia genética por pares. En general, los resultados de diversidad y distancia genética de los datos del ADNmt y del ChrY, subrayan la importancia de la deriva génica en la separación de SBP de sus poblaciones de origen. Por tanto, el análisis de distancias genéticas basado en distribuciones de frecuencia de alelos no es la mejor estrategia para buscar afinidades en África. En esta situación, la distancia genética F_{ST} o el análisis de componentes principales reflejarán una deriva aleatoria en lugar de afinidades genéticas distantes. Por tanto, consideramos que un abordaje filogeográfico de cada uno de los linajes presentes en Palenque es más adecuado para la búsqueda de sus orígenes entre las diferentes regiones africanas.

En 79 muestras de Palenque, se pudieron identificar al menos 17 fundadores posibles en el ADNmt (Figura 17). La comparación con los datos reportados de África nos permitió ubicar el origen más probable de siete de estos fundadores (Figura 55). Para casi el 70% de las secuencias de ADNmt africanas en Palenque, no fue posible inferir el origen. Sin embargo, en base a los linajes restantes, diferentes regiones parecen haber contribuido al actual trasfondo materno de Palenque.

	Haplogroup	n	Freq.	Most likely African origin	
mtDNA	L0a1a+200	1	1.3%	Between Upper Guinea and Bight of Benin	
	L1b1a1'4	7	8.9%	Between Senegambia and Upper Guinea	
	L1c3a1b	5	6.3%	Between Gold Coast and Angola	
	L2b1a	1	1.3%	Bight of Benin and Angola	
	L2d+16129	1	1.3%	Angola	
	L3e1d	10	12.7%	Angola	
	L3f1b+I6365	1	1.3%	Gold Coast	
	Other samples	53	67.1%	Unknown	
	Y chromosome	Y-MRCA* (xM13,SRY10831.1)	2	3.4%	Upper Guinea
		B2a-M150* (xM109)	3	5.2%	Loango
E1a-M33		1	1.7%	Upper Guinea	
E1a-M33		1	1.7%	Between Bight of Benin and Bight of Biafra	
E1b1a-M191		1	1.7%	Between Loango and Angola	
E1b1a-M191		1	1.7%	Loango	
E1b1a-M2* (xM154,M191)		13	22.4%	Bight of Benin	
E1b1a-M2* (xM154,M191)		7	12.1%	Between Loango and Angola	
E1b1b-M35* (xM78,M81,M123,V6,M293)		3	5.2%	Between Senegambia and Bight of Benin	
R1b-V88		4	6.9%	Between Bight of Benin and Loango	
Other samples		22	37.9%	Unknown	

Figura 55. Origen africano más probable de los linajes del ADNmt y del cromosoma Y detectados en SBP.

Entre las 58 muestras de Palenque atribuidas a haplogrupos Y africanos, se pudieron identificar al menos 20 fundadores potenciales basados en la diversidad de los haplotipos. La comparación con datos de poblaciones africanas permitió ubicar el origen más probable de 12 de ellos. Los resultados muestran múltiples orígenes en la región de África occidental, en un territorio que se extiende desde la Alta Guinea hasta Angola (Figura 55). Aunque no fue posible asignar el origen de un 38% de los linajes masculinos, los linajes restantes parecen haber sido traídos a América durante la diáspora desde diversos puertos a lo largo de la costa occidental de África.

La detección de los linajes originarios principalmente de la costa occidental de África, así como su variedad de orígenes, está de acuerdo con la información existente sobre la llegada de barcos esclavistas a Cartagena. Según la base de datos *Slave Voyages* (Consortium, 2021), los barcos que llegaron a Cartagena durante la trata de esclavos venían de distintos puertos en África occidental, desde Senegambia hasta Angola. Aunque en este trabajo se analizó toda la región control del ADNmt, para los análisis comparativos, solo pudimos usar la información de las regiones HVI y HVII y así poder incluir las regiones africanas más relevantes. Por la misma razón, solo se pudo usar un pequeño conjunto de todos los Y-STRs genotipados. Además, hay poca superposición entre las publicaciones relacionadas con los Y-SNPs utilizados para la determinación de haplogrupos. Si bien se encontraron afinidades entre Palenque y algunas regiones africanas, no fue posible precisar el origen de muchos linajes, y algunos se remontaron a una vasta región, lo que impidió estimar la proporción exacta en la que diferentes regiones africanas contribuyeron a Palenque. El enfoque filogeográfico utilizado está fuertemente influenciado por el conocimiento de la diversidad genética de las poblaciones originales, estando condicionado por la existencia de datos con alta resolución. Sin embargo, el pool genético de las poblaciones de África subsahariana todavía está poco estudiado, y algunas áreas relevantes en la historia de la trata transatlántica de esclavos han sido poco investigadas hasta ahora.

En conclusión, hasta aquí, los resultados encontrados mostraron que los linajes africanos en Palenque resultaron de un número restringido de fundadores de múltiples orígenes geográficos. Los

resultados no contradicen la importante influencia de los hablantes de kikongo en el Palenque temprano, como está ampliamente documentado y está presente en el idioma palenquero. Sin embargo, estos resultados muestran que aún existen lagunas en la historia de los palenqueros, quienes parecen portar en sus genes otras raíces más allá del kikongo. Los resultados son compatibles con la presencia de varios sustratos africanos en los primeros habitantes de Palenque, aunque uno ha sido dominante en términos de cultura e idioma. En cuanto al legado paterno europeo, queda por investigar en qué momento de la historia de Palenque se habrían introducido estos linajes.

8.2. Ancestría continental y subcontinental, mezcla por sesgo sexual, efecto fundador en SBP con los análisis de WGS

En este estudio encontramos que SBP muestra niveles altos de ancestría africana, pero no de una manera excepcional cuando se compara con otras poblaciones afrodescendientes de las Américas (Tabla 4). El porcentaje de ancestría africana de SBP fue muy similar al reportado con los marcadores tipo InDels y que fueron publicados previamente (Martínez et al., 2017). De igual forma están de acuerdo con lo observado con los análisis de cromosoma Y y de ADNmt (ansari)(Tabla 3). Además, el análisis de WGS tanto de marcadores autosómicos como de haplotipos uniparentales confirmó que los palenqueros muestran evidencia de mezcla con poblaciones europeas e indígenas, similar a otras poblaciones latinoamericanas. Sin embargo, los patrones de mezcla con sesgo sexual observados en Palenque son únicos (Figura 35). A diferencia de todas las demás poblaciones latinoamericanas estudiadas hasta la fecha, muestra una mezcla muy fuerte con sesgo masculino africano y un patrón mucho menos pronunciado de mezcla con sesgo masculino europeo y femenino nativo americano. La ancestría africana con sesgo masculino africano también es diferente a la observada en la población del Chocó, costa pacífica colombiana, y las poblaciones afrodescendientes en los Estados Unidos y el Caribe, que muestran una ancestría africana con sesgo femenino. Finalmente, hay una diferencia aparente en el patrón de ancestría sesgada por sexo revelado por el análisis de marcadores uniparentales en comparación con el patrón mostrado en los autosomas frente al

cromosoma X. En conjunto con los registros históricos y la tradición oral en Palenque, se puede considerar que estos resultados reflejan la dinámica social de su asentamiento desde el periodo de la colonización y periodos de aislamientos que explicarían un efecto fundador y la deriva génica que fue observada con los análisis de cromosoma Y y de ADNmt (Navarrete, 2017). Puede que haya habido un pulso temprano de ancestría masculina con sesgo masculino europeo y con sesgo femenino nativo americano, antes del período colonial o alrededor del momento del asentamiento del Palenque, seguido de un período mucho más largo de aislamiento marcado por una ancestría africana con sesgo masculino. Esto último es consistente con los registros históricos que relatan cómo el contacto con las aldeas indígenas cercanas agregó continuamente población femenina a la comunidad.

Por otro lado, a pesar de su historia de aislamiento, SBP no muestra evidencia de un agotamiento de la diversidad genética, lo que es consistente con la presencia de un cierto grado de deriva génica (Figura 51 y Figura 38C). En la población existe un sistema social descrito como “Kuardo”, que está configurado en grupos de redes familiares y se desalienta a las personas a casarse dentro del mismo Kuardo (R. Miranda, comunicación personal, 5 de julio de 2019). Esto ayuda a que las parejas con parentesco genético críptico no se casen ni se reproduzcan, pero esta conducta social no explica totalmente el hallazgo, por tal un análisis de trectos más profundo al que realizamos podría arrojar más información en estudios posteriores.

8.3. Ancestría africana y grupos lingüísticos en SBP

Múltiples líneas de evidencia, WGS, AIMs y marcadores uniparentales analizados con una variedad de métodos, apuntan a una ancestría africana diversa y heterogénea para SBP como lo muestra este trabajo. Si bien los palenqueros tienen en su mayoría una ancestría africana bantú, también tienen contribuciones sustanciales de ancestría de Senegambia y de la Ensenada de Benín (Tabla 5 y Figura 45). Además, dentro del componente de ancestría bantú hay evidencia de contribuciones de numerosos grupos distintos al Kongo que corresponde a la fuente del sustrato de la lengua africana, descrito por Ansari-Pour et al. (2016). En total, la población de Palenque muestra afinidades

ancestrales con las poblaciones africanas que hablan una amplia variedad de idiomas de la diversa familia lingüística Níger-Congo. Estos resultados destacaron una sorprendente disonancia entre los orígenes genéticos africanos de los palenqueros y el dominio cultural de un solo sustrato de lengua africana para el idioma criollo palenquero.

El idioma kikongo que forma el sustrato africano del palenquero se hablaba en el reino africano de Kongo, que dominó gran parte del suroeste de África desde finales del siglo XIV hasta principios del siglo XX (Thornton & Thornton, 1998). Con mucho, el mayor número de africanos esclavizados, cerca de 5,9 millones, fueron llevados a América desde esta región (que comprendía la costa de Loango) durante la trata transatlántica de esclavos (Consortium, 2021). El kikongo fue, y sigue siendo hasta el día de hoy, la base principal de varios dialectos criollos que se hablan en las regiones africanas adyacentes (Mufwene, 2008). La influencia de la lengua kikongo en la región, considerada junto con la gran cantidad de individuos originarios de esta región, sugiere la posibilidad de que el kikongo haya servido como lengua franca para los africanos esclavizados y sus descendientes en el Nuevo Mundo. De hecho, el kikongo contribuye a una serie de lenguas criollas habladas por comunidades de africanos esclavizados y sus descendientes en Brasil, Colombia, Cuba, República Dominicana, Haití, Puerto Rico y Estados Unidos.

Una vez establecido como el sustrato de la lengua africana dominante por los fundadores de Palenque, las llegadas posteriores pueden haber sido obligadas a adoptar Palenquero, incluidos sus estratos africanos y españoles, independientemente de su lengua africana nativa (Byrd, 2012; Di Paolo, 2014; Dianteill, 2002; Lipski, 2001; Pollitzer, 1999).

El lenguaje y la genética a menudo coevolucionan, como puede verse claramente en las lenguas Níger-Congo en África (Dianteill, 2002) (Figura 39). Este no parece haber sido el caso del “*palenquero criollo*”, donde persistió una sola lengua africana frente a los diversos orígenes genéticos africanos de sus hablantes. La discordancia entre evolución cultural y genética en Palenque puede tomarse como un ejemplo de la diferencia entre evolución memética y genética (Dawkins, 1976). Los memes evolucionan más rápidamente que los genes, debido tanto a la

rápida tasa de cambio como a la combinación de la evolución horizontal y vertical, en oposición a la evolución estrictamente vertical de los genes. El dominio evolutivo de los memes lingüísticos puede ser particularmente pronunciado para las lenguas criollas formadas a través del contacto entre grupos previamente aislados.

Los resultados anteriores podrían sugerir que el sustrato africano del Palenquero es en realidad más complejo de lo que se imaginaba anteriormente, con aportes de múltiples lenguas africanas, aunque años de estudios detallados sugieran que no es así. No obstante, nuestros resultados también podrían tomarse para motivar estudios y potencialmente nuevos descubrimientos sobre el sustrato africano de Palenquero (Granda Gutiérrez, 1971; Schwegler, 1999, 2000, 2011, 2016, 2017).

Considerados en conjunto, los análisis que realizamos con los marcadores autosómicos, así como de los haplogrupos uniparentales, indican que la población de SBP tiene contribuciones ancestrales de un grupo diverso de poblaciones africanas, correspondientes a un conjunto igualmente diverso de grupos lingüísticos de Níger-Congo, en contraste con la hipótesis de un único sustrato africano propuesto para la lengua *palenquero*. Es el estudio más completo descrito hasta ahora para esta población afrodescendiente de la costa caribe colombiana.

8.4. Palenque en comparación con otras poblaciones colombianas mezcladas y nativas

Durante las últimas décadas, los STRs autosómicos fueron los marcadores de elección para fines de identificación humana (HID, por sus siglas en inglés) (Poulsen et al., 2014). Sin embargo, las muestras de baja calidad que se encuentran comúnmente en casos forenses generan perfiles de STR parciales lo que impide llegar a conclusiones (por ejemplo, ADN antiguo o degradado). De manera similar, la tasa de mutación relativamente alta de los STRs genera inconsistencias esporádicas en el análisis de parentesco (García-Aceves et al., 2018; MartínezCortés et al., 2015).

Aunque los polimorfismos SNP también se han implementado para propósitos de HID, su análisis requiere instrumentos que no están disponibles en la mayoría de los laboratorios de genética forense y/o no

involucran metodologías convencionales (G. Martínez-Cortés et al., 2015). Por el contrario, los polimorfismos InDel se han incorporado fácilmente en el trabajo de casos forenses por medio de técnicas de amplificación de PCR multiplex seguidos de electroforesis capilar tradicional (Pinto et al., 2013; Romanini et al., 2012).

En particular, los loci bialélicos, tanto de SNPs como de InDels, han sido útiles para el análisis de restos óseos altamente degradados (Jobling & Gill, 2004) y para tratar casos de STRs con mutaciones ocasionales en pruebas de paternidad, debido a su baja tasa de mutación (Kayser & De Knijff, 2011; Sanchez et al., 2006). En consecuencia, sistemas genéticos basados en HIDInDels se han estudiado en poblaciones de todo el mundo para estimar frecuencias alélicas y parámetros forenses a priori, como heterocigosidad, poder de discriminación (PD), poder de exclusión (PE) e índice típico de paternidad (IP), principalmente (Fondevila et al., 2012). Sin embargo, se han descrito debilidades en la informatividad por su condición bialélica, para el análisis de parentesco, particularmente en casos en los que no se dispone de la madre o cuando el presunto padre es pariente cercano del verdadero padre biológico (Pereira et al., 2009). Curiosamente, algunos parámetros *a posteriori* como el Índice de paternidad (PI) han demostrado ser útiles para evaluar la contribución real de marcadores genéticos en análisis de parentesco convencionales, lo cual sería valioso para evitar desperdiciar trabajo, recursos y, finalmente, resultados inconclusivos.

En este estudio, hemos comparado diferentes grupos poblacionales de Colombia para marcadores de uso en forense como son los 38 HID-InDels. Los resultados mostraron diferencias en la composición genética de la mayoría de los grupos estudiados, excepto entre las poblaciones de los Andes y la Orinoquia. La alta heterogeneidad encontrada entre las poblaciones indica que son necesarias diferentes bases de datos de frecuencias alélicas, para los loci estudiados, para las investigaciones forenses en Colombia. Los parámetros de relevancia forense muestran que el conjunto completo de 38 HID-InDels tiene un valor alto de poder de discriminación en las poblaciones estudiadas, aunque en los casos de paternidad debe ser utilizado en conjunto con otros marcadores. El uso conjunto de datos de

38 HID-InDels y de 46 AIM-InDels demostró ser útil para aumentar la poca capacidad informativa que es inherente a los marcadores bialélicos en comparación con los STRs. Sin embargo, el uso de un número elevado de loci aumenta la probabilidad de que se produzcan asociaciones entre marcadores de un mismo cromosoma, por lo que aún son necesarios más estudios para evaluar este efecto.

Por otro lado, los resultados obtenidos con estos marcadores forenses (anexo D) indicaron que la población de SBP está más cerca de África que de otras poblaciones nativas o europeas. Sin embargo, en este trabajo se observa que SBP muestra diferencias con respecto a las muestras de referencia africana, que pueden explicarse por deriva genética, ya encontrada en los análisis realizados con marcadores de linaje anteriormente descritos.

De igual manera, el estudio de los orígenes de SBP, acompañado de la publicación de bases de datos de marcadores de linaje y en autosómicos, añade la posibilidad de su uso en estudios de epidemiología y de asociación con enfermedades, para una población afrodescendiente de la costa caribe colombiana.

9. REFERENCIAS

Adhikari, K., Mendoza-Revilla, J., Chacón-Duque, J. C., Fuentes-Guajardo, M., & RuizLinares, A. (2016). Admixture in Latin America. *Current opinion in genetics & development*, *41*, 106-114.

Akey, J. M., Eberle, M. A., Rieder, M. J., Carlson, C. S., Shriver, M. D., Nickerson, D. A., & Kruglyak, L. (2004). Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol*, *2*(10), e286.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research*, *19*(9), 1655-1664.

Altuna, M. E., Modesti, N. M., & Demarchi, D. A. (2006). Y-chromosomal evidence for a founder effect in Mbyá-guaraní Amerindians from northeast Argentina. *Human biology*, 635-639.

Alvarez, J. C., Johnson, D. L., Lorente, J. A., Martinez-Espin, E., Martinez-Gonzalez, L. J., Allard, M., Wilson, M. R., & Budowle, B. (2007). Characterization of human control region sequences for Spanish individuals in a forensic mtDNA data set. *Legal Medicine*, *9*(6), 293-304.

Alves-Silva, J., da Silva Santos, M., Guimarães, P. E., Ferreira, A. C., Bandelt, H.-J., Pena, S. D., & Prado, V. F. (2000). The ancestry of Brazilian mtDNA lineages. *The American Journal of Human Genetics*, *67*(2), 444-461.

Amirisetty, S., Hershey, G. K. K., & Baye, T. M. (2012). AncestrySNPminer: a bioinformatics tool to retrieve and develop ancestry informative SNP panels. *Genomics*, *100*(1), 57-63. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., & Sanger, F. (1981). Sequence and organization of the human mitochondrial genome. *Nature*, *290*(5806), 457-465.

Andréasson, H., Nilsson, M., Styrman, H., Pettersson, U., & Allen, M. (2007). Forensic mitochondrial coding region analysis for increased discrimination using pyrosequencing technology. *Forensic Science International: Genetics*, *1*(1), 35-43.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics*, 23, 147.

Ansari-Pour, N., Moñino, Y., Duque, C., Gallego, N., Bedoya, G., Thomas, M. G., & Bradman, N. (2016). Palenque de San Basilio in Colombia: genetic data support an oral history of a paternal ancestry in Congo. *Proceedings of the Royal Society B: Biological Sciences*, 283(1827), 20152980.

Arnaiz-Villena, A., Juarez, I., Palacio-Gruber, J., Muñoz, E., Campos, C., Martinez-Laso, J., Nieto, J., Lopez-Nares, A., Martin-Villa, J. M., & Silvera, C. (2018). The first free Africans in America: HLA study in San Basilio de Palenque (Colombia). *Human immunology*, 79(8), 585-586. <https://doi.org/10.1016/j.humimm.2018.05.010>. Epub 2018 Jun 1. PMID: 29864459

Arora, D., Singh, A., Sharma, V., Bhaduria, H. S., & Patel, R. B. (2015). HgsDb: Haplogroups

Database to understand migration and molecular risk assessment. *Bioinformatics*, 11(6), 272.

Arrazola, R. (1970). Palenque, primer pueblo libre de América: historia de las sublevaciones de los esclavos de Cartagena, Cartagena, Hernández. *Du terrain à la théorie. Les*, 40.

Aumann, Y., Manisterski, E., & Yakhini, Z. (2003). Designing optimally multiplexed SNP genotyping assays. International Workshop on Algorithms in Bioinformatics,

Avena, S., Via, M., Ziv, E., Pérez-Stable, E. J., Gignoux, C. R., Dejean, C., Huntsman, S., Torres-Mejía, G., Dutil, J., & Matta, J. L. (2012). Heterogeneity in genetic admixture across different regions of Argentina. *PLoS one*, 7(4), e34695.

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A., & Saunders, N. C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual review of ecology and systematics*, 18(1), 489-522.

Bailliet, G., Ramallo, V., Muzzio, M., Santos, M. R., Motti, J. M. B., Bianchi, N. O., & Bravi, C. M. (2011). Antecedentes y nuevos

aportes en el estudio del Cromosoma Y en poblaciones humanas sudamericanas. *Journal of Basic & Applied Genetics*, 22.

Ballantyne, K. N., Goedbloed, M., Fang, R., Schaap, O., Lao, O., Wollstein, A., Choi, Y., van Duijn, K., Vermeulen, M., & Brauer, S. (2010). Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *The American Journal of Human Genetics*, 87(3), 341-353.

Ballantyne, K. N., Ralf, A., Aboukhalid, R., Achakzai, N. M., Anjos, M. J., Ayub, Q., Balažic, J., Ballantyne, J., Ballard, D. J., & Berger, B. (2014). Toward male individualization with rapidly mutating y-chromosomal short tandem repeats. *Human mutation*, 35(8), 1021-1032.

Bandelt, H.-J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16(1), 37-48.

Bandelt, H.-J., Forster, P., Sykes, B. C., & Richards, M. B. (1995). Mitochondrial portraits of human populations using median networks. *Genetics*, 141(2), 743-753.

Bansal, V., & Libiger, O. (2015). Fast individual ancestry inference from DNA sequence data leveraging allele frequencies for multiple populations. *BMC bioinformatics*, 16(1), 111.

Barbieri, C., Butthof, A., Bostoen, K., & Pakendorf, B. (2013). Genetic perspectives on the origin of clicks in Bantu languages from southwestern Zambia. *European Journal of Human Genetics*, 21(4), 430-436.

Barbieri, C., Hübner, A., Macholdt, E., Ni, S., Lippold, S., Schröder, R., Mpoloka, S. W., Purps, J., Roewer, L., & Stoneking, M. (2016). Refining the Y chromosome phylogeny with southern African sequences. *Human genetics*, 135(5), 541-553.

Barbieri, C., Vicente, M., Oliveira, S., Bostoen, K., Rocha, J., Stoneking, M., & Pakendorf, B. (2014). Migration and interaction in a contact zone: mtDNA variation among Bantuspeakers in Southern Africa. *PloS one*, 9(6), e99117.

Barbieri, C., Vicente, M., Rocha, J., Mpoloka, S. W., Stoneking, M., & Pakendorf, B. (2013). Ancient substructure in early

mtDNA lineages of southern Africa. *The American Journal of Human Genetics*, 92(2), 285-292.

Batini, C., Coia, V., Battaglia, C., Rocha, J., Pilkington, M. M., Spedini, G., Comas, D., DestroBisol, G., & Calafell, F. (2007). Phylogeography of the human mitochondrial L1c haplogroup: genetic signatures of the prehistory of Central Africa. *Molecular phylogenetics and evolution*, 43(2), 635-644.

Batini, C., Ferri, G., Destro-Bisol, G., Brisighelli, F., Luiselli, D., Sánchez-Diz, P., Rocha, J., Simonson, T., Brehm, A., & Montano, V. (2011). Signatures of the preagricultural peopling processes in sub-Saharan Africa as revealed by the phylogeography of early Y chromosome lineages. *Molecular Biology and Evolution*, 28(9), 2603-2613.

Bauchet, M., McEvoy, B., Pearson, L. N., Quillen, E. E., Sarkisian, T., Hovhannesian, K., Deka, R., Bradley, D. G., & Shriver, M. D. (2007). Measuring European population stratification with microarray genotype data. *The American Journal of Human Genetics*, 80(5), 948-956.

Behar, D. M., Garrigan, D., Kaplan, M. E., Mobasher, Z., Rosengarten, D., Karafet, T. M., Quintana-Murci, L., Ostrer, H., Skorecki, K., & Hammer, M. F. (2004). Contrasting patterns of Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations. *Human genetics*, 114(4), 354-365.

Beleza, S., Gusmao, L., Amorim, A., Carracedo, A., & Salas, A. (2005). The genetic legacy of western Bantu migrations. *Human genetics*, 117(4), 366-375.

Berniell-Lee, G., Calafell, F., Bosch, E., Heyer, E., Sica, L., Mougouiana-Daouda, P., Van der Veen, L., Hombert, J.-M., Quintana-Murci, L., & Comas, D. (2009). Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Molecular Biology and Evolution*, 26(7), 1581-1589.

Bhangale, T. R., Rieder, M. J., Livingston, R. J., & Nickerson, D. A. (2005). Comprehensive identification and characterization of diallelic insertion-deletion polymorphisms in 330 human candidate genes. *Human molecular genetics*, 14(1), 59-69.

Bini, C., & Pappalardo, G. (2005). mtDNA HVI length heteroplasmic profile in different tissues of maternally related members. *Forensic science international*, 152(1), 35-38.

Bodner, M., Perego, U. A., Huber, G., Fendt, L., Röck, A. W., Zimmermann, B., Olivieri, A., Gómez-Carballa, A., Lancioni, H., & Angerhofer, N. (2012). Rapid coastal spread of First Americans: Novel insights from South America's Southern Cone mitochondrial genomes. *Genome research*, 22(5), 811-820.

Bolnick, D. A., Bolnick, D. I., & Smith, D. G. (2006). Asymmetric male and female genetic histories among Native Americans from Eastern North America. *Molecular Biology and Evolution*, 23(11), 2161-2174.

Bortolini, M.-C., Salzano, F. M., Thomas, M. G., Stuart, S., Nasanen, S. P., Bau, C. H., Hutz, M. H., Layrisse, Z., Petzl-Erler, M. L., & Tsuneto, L. T. (2003). Y-chromosome evidence for differing ancient demographic histories in the Americas. *The American Journal of Human Genetics*, 73(3), 524-539.

Borucki, A., Eltis, D., & Wheat, D. (2015). Atlantic history and the slave trade to Spanish America. *The American Historical Review*, 120(2), 433-461.

Bostoen, K., Clist, B., Doumenge, C., Grollemund, R., Hombert, J.-M., Muluwa, J. K., Maley, J., Blench, R., Di Carlo, P., & Good, J. (2015). Middle to late Holocene Paleoclimatic change and the early Bantu expansion in the rain forests of Western Central Africa. *Current Anthropology*, 56(3), 367-368.

Brandon, M. C., Lott, M. T., Nguyen, K. C., Spolim, S., Navathe, S. B., Baldi, P., & Wallace, D. C. (2005). MITOMAP: a human mitochondrial genome database—2004 update.

Nucleic acids research, 33(suppl_1), D611-D613.

Brion, M., Sobrino, B., Blanco-Verea, A., Lareu, M., & Carracedo, A. (2005). Hierarchical analysis of 30 Y-chromosome SNPs in European populations. *International Journal of Legal Medicine*, 119(1), 10-15.

Brown, M. D., Hosseini, S. H., Torroni, A., Bandelt, H.-J., Allen, J. C., Schurr, T. G., Scozzari, R., Cruciani, F., & Wallace, D. C. (1998). mtDNA haplogroup X: an ancient link between Europe/Western

Asia and North America? *The American Journal of Human Genetics*, 63(6), 1852-1861.

Bryc, K., Auton, A., Nelson, M. R., Oksenberg, J. R., Hauser, S. L., Williams, S., Froment, A., Bodo, J.-M., Wambebe, C., & Tishkoff, S. A. (2010). Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proceedings of the national Academy of sciences*, 107(2), 786-791.

Budowle, B., Fisher, C. L., Polansky, D., Den Hartog, B. K., Kepler, R. B., & Elling, J. W. (2008). Stabilizing mtDNA sequence nomenclature with an operationally efficient approach. *Forensic Science International: Genetics Supplement Series*, 1(1), 671-673.

Byrd, S. (2012). *Calunga and the legacy of an African language in Brazil*. UNM Press.

CAAPA. (2011). *Consortium on Asthma among African-ancestry Populations in the Americas CAAPA*. <https://www.caapa-project.org/>

Calloway, C. D., Reynolds, R. L., Herrin Jr, G. L., & Anderson, W. W. (2000). The frequency of heteroplasmy in the HVII region of mtDNA differs across tissue types and increases with age. *The American Journal of Human Genetics*, 66(4), 1384-1397.

Campbell, M. C., Hirbo, J. B., Townsend, J. P., & Tishkoff, S. A. (2014). The peopling of the African continent and the diaspora into the new world. *Current opinion in genetics & development*, 29, 120-132.

Campbell, M. C., & Tishkoff, S. A. (2008). African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu. Rev. Genomics Hum. Genet.*, 9, 403-433.

Capelli, C., Onofri, V., Brisighelli, F., Boschi, I., Scarnicci, F., Masullo, M., Ferri, G., Tofanelli, S., Tagliabracci, A., & Gusmao, L. (2009). Moors and Saracens in Europe: estimating the medieval North African male legacy in southern Europe. *European Journal of Human Genetics*, 17(6), 848-852.

Capelli, C., Wilson, J. F., Richards, M., Stumpf, M. P., Gratrix, F., Oppenheimer, S., Underhill, P., Pascali, V. L., Ko, T.-M., & Goldstein, D. B. (2001). A predominantly indigenous paternal heritage

for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *The American Journal of Human Genetics*, 68(2), 432-443.

Capodiferro, M. R., Aram, B., Raveane, A., Migliore, N. R., Colombo, G., Ongaro, L., Rivera, J., Mendizábal, T., Hernández-Mora, I., & Tribaldos, M. (2021). Archaeogenomic distinctiveness of the Isthmo-Colombian area. *Cell*, 184(7), 1706-1723. e1724.

Caratti, S., Gino, S., Torre, C., & Robino, C. (2009). Subtyping of Y-chromosomal haplogroup E-M78 (E1b1b1a) by SNP assay and its forensic application. *International Journal of Legal Medicine*, 123(4), 357-360.

Carvalho, M., Brito, P., Bento, A., Gomes, V., Antunes, H., Costa, H. A., Lopes, V., Serra, A., Balsa, F., & Andrade, L. (2011). Paternal and maternal lineages in Guinea-Bissau population. *Forensic Science International: Genetics*, 5(2), 114-116.

Casanova, M., Leroy, P., Boucekkine, C., Weissenbach, J., Bishop, C., Fellous, M., Purrello,

M., Fiori, G., & Siniscalco, M. (1985). A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science*, 230(4732), 1403-1406.

Casas-Vargas, A., Romero, L. M., Usaquén, W., Zea, S., Silva, M., Briceño, I., Gómez, A., & Rodríguez, J. V. (2017). Mitochondrial DNA diversity in prehispanic bone remains on the eastern Colombian Andes. *Biomedica*, 37(4), 548-560.

Cerezo, M., Gusmao, L., Černý, V., Uddin, N., Syndercombe-Court, D., Gomez-Carballa, A.,

Goebel, T., Schneider, P. M., & Salas, A. (2016). Comprehensive analysis of PanAfrican mitochondrial DNA variation provides new insights into continental variation and demography. *Journal of Genetics and Genomics*, 43(3), 133-143.

Cinnioğlu, C., King, R., Kivisild, T., Kalfoglu, E., Atasoy, S., Cavalleri, G. L., Lillie, A. S., Roseman, C. C., Lin, A. A., & Prince, K. (2004). Excavating Y-chromosome haplotype strata in Anatolia. *Human genetics*, 114(2), 127-148.

Coelho, M., Sequeira, F., Luiselli, D., Belez, S., & Rocha, J. (2009). On the edge of Bantu expansions: mtDNA, Y chromosome and

lactase persistence genetic variation in southwestern Angola. *BMC evolutionary biology*, 9(1), 1-18.

Collins-Schramm, H. E., Chima, B., Operario, D. J., Criswell, L. A., & Seldin, M. F. (2003). Markers informative for ancestry demonstrate consistent megabase-length linkage disequilibrium in the African American population. *Human genetics*, 113(3), 211-219.

Conley, A. B., Rishishwar, L., Norris, E. T., Valderrama-Aguirre, A., Mariño-Ramírez, L., Medina-Rivas, M. A., & Jordan, I. K. (2017). A comparative analysis of genetic ancestry and admixture in the Colombian populations of Choco and Medellín. *G3: Genes, Genomes, Genetics*, 7(10), 3435-3447.

Conomos, M. P., Laurie, C. A., Stip, A. M., Gogarten, S. M., McHugh, C. P., Nelson, S. C., Sofer, T., Fernández-Rhodes, L., Justice, A. E., & Graff, M. (2016). Genetic diversity and association studies in US Hispanic/Latino populations: applications in the Hispanic Community Health Study/Study of Latinos. *The American Journal of Human Genetics*, 98(1), 165-184.

Conrad, D. F., & Hurler, M. E. (2007). The population genetics of structural variation. *Nature genetics*, 39(7), S30-S36.

Conrad, D. F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T. D., Barnes, C., & Campbell, P. (2010). Origins and functional impact of copy number variation in the human genome. *Nature*, 464(7289), 704-712.

Consortium, G. P. (2015). A global reference for human genetic variation. *Nature*, 526(7571), 68.

Consortium, I. H. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature*, 449(7164), 851.

Consortium, T. S. V. (2021). *Slavevoyages v2.2.13*. Creative Commons <https://www.slavevoyages.org/>

Consortium, Y. C. (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome research*, 12(2), 339-348. contributors, W. C. (2020). *File:World Map Blank - with blue sea.svg*. Wikimedia commons, the free media repository.

Retrieved ID: 520912411 from

https://commons.wikimedia.org/w/index.php?title=File:World_Map_Blank_with_blue_sea.svg&oldid=520912411.

Cooper, G. M., Hausman, R. E., & Hausman, R. E. (2007). *The cell: a molecular approach* (Vol. 4). ASM press Washington, DC.

Cortellini, V., Verzeletti, A., Cerri, N., Marino, A., & De Ferrari, F. (2013). Y-chromosome polymorphisms and ethnic group—a combined STR and SNP approach in a population sample from northern Italy. *Croatian medical journal*, 54(3), 279-285.

Cruciani, F., La Fratta, R., Santolamazza, P., Sellitto, D., Pascone, R., Moral, P., Watson, E., Guida, V., Colomb, E. B., & Zaharova, B. (2004). Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. *The American Journal of Human Genetics*, 74(5), 1014-1022.

Cruciani, F., Santolamazza, P., Shen, P., Macaulay, V., Moral, P., Olckers, A., Modiano, D., Holmes, S., Destro-Bisol, G., & Coia, V. (2002). A back migration from Asia to subSaharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *The American Journal of Human Genetics*, 70(5), 1197-1214.

Cruciani, F., Trombetta, B., Massaia, A., Destro-Bisol, G., Sellitto, D., & Scozzari, R. (2011). A revised root for the human Y chromosomal phylogenetic tree: the origin of patrilineal diversity in Africa. *The American Journal of Human Genetics*, 88(6), 814-818.

Cruciani, F., Trombetta, B., Sellitto, D., Massaia, A., Destro-Bisol, G., Watson, E., Colomb, E. B., Dugoujon, J.-M., Moral, P., & Scozzari, R. (2010). Human Y chromosome haplogroup R-V88: a paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages. *European Journal of Human Genetics*, 18(7), 800-807.

Chan, E. K., Timmermann, A., Baldi, B. F., Moore, A. E., Lyons, R. J., Lee, S.-S., Kalsbeek, A. M., Petersen, D. C., Rautenbach, H., & Förtsch, H. E. (2019). Human origins in a southern African palaeo-wetland and first migrations. *Nature*, 575(7781), 185-189.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the

challenge of larger and richer datasets. *Gigascience*, 4(1), s13742-13015-10047-13748.

Chocó, U. T. d. (2019). *ChocoGen. Patrimonio genético del pacífico colombiano*. <https://www.chocogen.com/>

Choudhury, A., Aron, S., Sengupta, D., Hazelhurst, S., & Ramsay, M. (2018). African genetic diversity provides novel insights into evolutionary history and local adaptations. *Human molecular genetics*, 27(R2), R209-R218.

D'Atanasio, E., Trombetta, B., Bonito, M., Finocchio, A., Di Vito, G., Seghizzi, M., Romano, R., Russo, G., Paganotti, G. M., & Watson, E. (2018). The peopling of the last Green Sahara revealed by high-coverage resequencing of trans-Saharan patrilineages. *Genome biology*, 19(1), 1-15.

Danielson, P. B., Sun, H.-Y., Melton, T., & Kristinsson, R. (2007). Resolving mtDNA mixtures by denaturing high-performance liquid chromatography and linkage phase determination. *Forensic Science International: Genetics*, 1(2), 148-153.

Dawkins, R. (1976). *The Selfish gene*.

de Filippo, C., Barbieri, C., Whitten, M., Mpoloka, S. W., Gunnarsdóttir, E. D., Bostoen, K., Nyambe, T., Beyer, K., Schreiber, H., & de Knijff, P. (2011). Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups. *Molecular Biology and Evolution*, 28(3), 1255-1269.

Delaneau, O., Zagury, J.-F., & Marchini, J. (2013). Improved whole-chromosome phasing for disease and population genetic studies. *Nature methods*, 10(1), 5-6.

Di Paolo, M. A. K. S. (2014). Haitian Creole. In *Languages and Dialects in the US: Focus on Diversity and Linguistics* (pp. 180-195).

Dianteill, E. (2002). Kongo in Cuba: the Transformations of an African Religion. *Archives de sciences sociales des religions*, 117(1), 59-80.

Endicott, P., & Ho, S. Y. (2008). A Bayesian evaluation of human mitochondrial substitution rates. *The American Journal of Human Genetics*, 82(4), 895-902.

Enoch, M.-A., Shen, P.-H., Xu, K., Hodgkinson, C., & Goldman, D. (2006). Using ancestryinformative markers to define populations and detect population stratification. *Journal of Psychopharmacology*, 20(4_suppl), 19-26.

Escalante, A. (1979). *El Palenque de San Basilio: una comunidad de descendientes de negros cimarrones*. Editorial Mejoras.

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources*, 10(3), 564-567.

Falola, T. (2001). *Culture and customs of Nigeria*. Greenwood Publishing Group.

Fan, H., & Chu, J.-Y. (2007). A brief review of short tandem repeat mutation. *Genomics, proteomics & bioinformatics*, 5(1), 7-14.

Fendt, L., Röck, A., Zimmermann, B., Bodner, M., Thye, T., Tschentscher, F., Owusu-Dabo, E., Göbel, T. M., Schneider, P. M., & Parson, W. (2012). MtDNA diversity of Ghana: a forensic and phylogeographic view. *Forensic Science International: Genetics*, 6(2), 244-249.

Feuk, L., Carson, A. R., & Scherer, S. W. (2006). Structural variation in the human genome. *Nature Reviews Genetics*, 7(2), 85-97.

Fondevila, M., Phillips, C., Santos, C., Pereira, R., Gusmao, L., Carracedo, A., Butler, J., Lareu, M., & Vallone, P. (2012). Forensic performance of two insertion–deletion marker assays. *International Journal of Legal Medicine*, 126(5), 725-737.

Fortes-Lima, C., Gessain, A., Ruiz-Linares, A., Bortolini, M.-C., Migot-Nabias, F., Bellis, G., Moreno-Mayar, J. V., Restrepo, B. N., Rojas, W., & Avendaño-Tamayo, E. (2017). Genome-wide ancestry and demographic history of African-descendant maroon communities from French Guiana and Suriname. *The American Journal of Human Genetics*, 101(5), 725-736.

Friedemann, N., & Patiño Rosselli, C. (1983). *Lengua y sociedad en el Palenque de San Basilio*.

Friedemann, N. S. (1998). San Basilio en el universo kilombo-África y Palenque-América. In A. M. Editores (Ed.), *Geografía humana de Colombia* (pp. 79-102). 1998.

Frudakis, T., Venkateswarlu, K., Thomas, M., Gaskin, Z., Ginjupalli, S., Gunturi, S., Ponnuswamy, V., Natarajan, S., & Nachimuthu, P. K. (2003). A classifier for the SNPbased inference of ancestry. *Journal of forensic sciences*, 48(4), 771-782.

Galloway, C. A., Lee, H., & Yoon, Y. (2012). Mitochondrial morphology—emerging role in bioenergetics. *Free Radical Biology and Medicine*, 53(12), 2218-2228.

García-Aceves, M., Rentería, O. R., Díaz-Navarro, X., & Rangel-Villalobos, H. (2018). Paternity tests in Mexico: results obtained in 3005 cases. *Journal of Forensic and Legal Medicine*, 55, 1-7.

García, A., Dermarchi, D. A., Tovo-Rodrigues, L., Pauro, M., Callegari-Jacques, S. M., Salzano, F. M., & Hutz, M. H. (2015). High interpopulation homogeneity in Central Argentina as assessed by ancestry informative markers (AIMs). *Genetics and molecular biology*, 38(3), 324-331.

García, J. M. (1982). VILA VILAR, E.: Hispanoamérica y el comercio de esclavos, Sevilla, Escuela de Estudios Hispanoamericanos, 1977, 306 Pp. *Cuadernos de historia moderna y contemporánea* (3), 319-321.

Geppert, M., Baeta, M., Núñez, C., Martínez-Jarreta, B., Zweynert, S., Cruz, O. W. V., González-Andrade, F., González-Solorzano, J., Nagy, M., & Roewer, L. (2011). Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America. *Forensic Science International: Genetics*, 5(2), 100-104.

Girón, D. G. (1621). Comunicación personal. In Rey (Ed.), (Carta al Rey del Gobernador de Caratagena de Indias, Don Garcia Girón, fechada el 28de marzo de 1621, donde da cuenta de la muerte por ahorcamiento de Benkos, transcrita por Valtierra P.S. J., Angel: Pedro Claver. El santo redentor de los negros. ed., pp. 459-460). Bogotá: Banco de la República.

Gnecchi-Ruscone, G. A., Sarno, S., De Fanti, S., Gianvincenzo, L., Giuliani, C., Boattini, A., Bortolini, E., Di Corcia, T., Sanchez Mellado, C., & Dàvila Francia, T. J. (2019). Dissecting the pre-Columbian genomic ancestry of Native Americans along the Andes–Amazonia divide. *Molecular Biology and Evolution*, 36(6), 1254-1269.

Göbel, T. M., Bodner, M., Robino, C., Augustin, C., Huber, G. E., Marra, M., Mutesa, L., Pasino, S., Santovito, A., & Zimmermann, B. (2020). Mitochondrial DNA variation in Sub-Saharan Africa: Forensic data from a mixed West African sample, Côte d'Ivoire (Ivory Coast), and Rwanda. *Forensic Science International: Genetics*, *44*, 102202.

Goldberg, A., Verdu, P., & Rosenberg, N. A. (2014). Autosomal admixture levels are informative about sex bias in admixed populations. *Genetics*, *198*(3), 1209-1229.

Gomes, I., Pinto, N., Antão-Sousa, S., Gomes, V., Gusmão, L., Amorim, A. (2020). Twenty Years Later: A Comprehensive Review of the X Chromosome Use in Forensic Genetics. *Frontiers in Genetics*. *11*, 926.

Gomes, V., Pala, M., Salas, A., Álvarez-Iglesias, V., Amorim, A., Gómez-Carballa, A., Carracedo, Á., Clarke, D. J., Hill, C., & Mormina, M. (2015). Mosaic maternal ancestry in the Great Lakes region of East Africa. *Human genetics*, *134*(9), 1013-1027.

Gomes, V., Sánchez-Diz, P., Amorim, A., Carracedo, Á., & Gusmão, L. (2010). Digging deeper into East African human Y chromosome lineages. *Human genetics*, *127*(5), 603-613.

Gomez, F., Hirbo, J., & Tishkoff, S. A. (2014). Genetic variation and adaptation in Africa:

implications for human evolution and disease. *Cold Spring Harbor perspectives in biology*, *6*(7), a008524.

González-Santos, M., Montinaro, F., Oosthuizen, O., Oosthuizen, E., Busby, G. B., Anagnostou, P., Destro-Bisol, G., Pascali, V., & Capelli, C. (2015). Genome-wide SNP analysis of southern African populations provides new insights into the dispersal of Bantu-speaking groups. *Genome biology and evolution*, *7*(9), 2560-2568.

González, M., Gomes, V., López-Parra, A. M., Amorim, A., Carracedo, A., Sánchez-Diz, P.,

Arroyo-Pardo, E., & Gusmao, L. (2013). The genetic landscape of Equatorial Guinea and the origin and migration routes of the Y chromosome haplogroup R-V88. *European Journal of Human Genetics*, *21*(3), 324-331.

Goodwin, W., Linacre, A., & Hadi, S. (2011). *An introduction to forensic genetics* (Vol. 2). John Wiley & Sons.

Gouveia, M. H., Borda, V., Leal, T. P., Moreira, R. G., Bergen, A. W., Kehdy, F. S., Alvim, I., Aquino, M. M., Araujo, G. S., & Araujo, N. M. (2020). Origins, admixture dynamics, and homogenization of the African gene pool in the Americas. *Molecular Biology and Evolution*, *37*(6), 1647-1656.

Granda Gutiérrez, G. d. (1971). Sobre la procedencia africana del habla'criolla'de San Basilio de Palenque (Bolívar, Colombia). *Thesaurus: boletín del Instituto Caro y Cuervo*, *26*(1), 84-94.

Gravel, S., Zakharia, F., Moreno-Estrada, A., Byrnes, J. K., Muzzio, M., Rodriguez-Flores, J. L., Kenny, E. E., Gignoux, C. R., Maples, B. K., & Guiblet, W. (2013). Reconstructing Native American migrations from whole-genome and whole-exome data. *PLoS Genet*, *9*(12), e1004023.

Graven, L., Passarino, G., Semino, O., Boursot, P., Santachiara-Benerecetti, S., Langaney, A., & Excoffier, L. (1995). Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. *Molecular Biology and Evolution*, *12*(2), 334-345.

Greenberg, J. H. (1987). *Language in the Americas*. Stanford University Press.

Greenberg, J. H., Turner, C. G., Zegura, S. L., Campbell, L., Fox, J. A., Laughlin, W., Szathmary, E. J., Weiss, K. M., & Woolford, E. (1986). The settlement of the Americas: A comparison of the linguistic, dental, and genetic evidence [and comments and reply]. *Current Anthropology*, *27*(5), 477-497.

Grugni, V., Battaglia, V., Kashani, B. H., Parolo, S., Al-Zahery, N., Achilli, A., Olivieri, A., Gandini, F., Houshmand, M., & Sanati, M. H. (2012). Ancient migratory events in the Middle East: new clues from the Y-chromosome variation of modern Iranians. *PLoS one*, *7*(7), e41252.

Guan, Y. (2014). Detecting structure of haplotypes and local ancestry. *Genetics*, *196*(3), 625642.

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard,

M. O., & Choudhury, A. (2015). The African genome variation project shapes medical genetics in Africa. *Nature*, *517*(7534), 327332.

Gymrek, M., Willems, T., Guilmatre, A., Zeng, H., Markus, B., Georgiev, S., Daly, M. J., Price, A. L., Pritchard, J. K., & Sharp, A. J. (2016). Abundant contribution of short tandem repeats to gene expression variation in humans. *Nature genetics*, *48*(1), 22-29.

Hadi, S. (2016). Analysis of Rapidly Mutating Y Chromosome Short Tandem Repeats (RM YSTRs). In *Forensic DNA Typing Protocols* (pp. 201-211). Springer.

Hammarström, H. F., Robert & Haspelmath, Martin & Bank, Sebastian. (2021). *Glottolog 4.4*. Hammer, M. F., Karafet, T. M., Park, H., Omoto, K., Harihara, S., Stoneking, M., & Horai, S. (2006). Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes. *Journal of human genetics*, *51*(1), 47-58.

Hammer, M. F., Karafet, T. M., Redd, A. J., Jarjanazi, H., Santachiara-Benerecetti, S., Soodyall, H., & Zegura, S. L. (2001). Hierarchical patterns of global human Y-chromosome diversity. *Molecular Biology and Evolution*, *18*(7), 1189-1203.

Hammer, M. F., & Zegura, S. L. (1996). The role of the Y chromosome in human evolutionary studies. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, *5*(4), 116-134.

Hancock, J. M. (1999). Microsatellites and other simple sequences: genomic context and mutational mechanisms. *Microsatellites: Evolution and applications*.

Henn, B. M., Gignoux, C. R., Jobin, M., Granka, J. M., Macpherson, J., Kidd, J. M., RodríguezBotigué, L., Ramachandran, S., Hon, L., & Brisbin, A. (2011). Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proceedings of the national Academy of sciences*, *108*(13), 5154-5162.

Holland, M. M., & Parsons, T. J. (1999). Mitochondrial DNA Sequence Analysis-Validation and Use for Forensic Casework. *Forensic science review*, *11*(1), 21-50.

Homburger, J. R., Moreno-Estrada, A., Gignoux, C. R., Nelson, D., Sanchez, E., Ortiz-Tello, P., Pons-Estel, B. A., Acevedo-Vasquez, E., Miranda, P., & Langefeld, C. D. (2015). Genomic insights

into the ancestry and demographic history of South America. *PLoS genetics*, 11(12), e1005602.

Iacovacci, G., D'Atanasio, E., Marini, O., Coppa, A., Sellitto, D., Trombetta, B., Berti, A., & Cruciani, F. (2017). Forensic data and microvariant sequence characterization of 27 YSTR loci analyzed in four Eastern African countries. *Forensic Science International: Genetics*, 27, 123-131.

Ibarra, A., Restrepo, T., Rojas, W., Castillo, A., Amorim, A., Martínez, B., Burgos, G., Ostos, H., Álvarez, K., & Camacho, M. (2014). Evaluating the X chromosome-specific diversity of Colombian populations using insertion/deletion polymorphisms. *PLoS one*, 9(1), e87202.

IHGSC, I. H. G. S. C. (2004). <https://www.genome.gov/12513430/2004-release-ihgscdescribes-finished-human-sequence>

Jiménez, S., Martínez, B., Hernández, M., Múnera, A., & Caraballo, L. (1996). Análisis inmunogenético y antropológico de la población del Palenque de San Basilio (Colombia). *Polimorfismo génico (HLA) en poblaciones hispanoamericanas. Real Academia de Ciencias Exactas, Físicas y Naturales, Madrid, España*, 245-269.

Jobling, M. A., & Gill, P. (2004). Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics*, 5(10), 739-751.

Jobling, M. A., & Tyler-Smith, C. (1995). Fathers and sons: the Y chromosome and human evolution. *Trends in Genetics*, 11(11), 449-456.

Jobling, M. A., & Tyler-Smith, C. (2003). The human Y chromosome: an evolutionary marker comes of age. *Nature Reviews Genetics*, 4(8), 598-612.

Jordan, B. (2020). Les traces génétiques de la traite des esclaves-Chroniques génomiques. In: EDP Sciences.

Jordan, I. K., Rishishwar, L., & Conley, A. B. (2019). Native American admixture recapitulates population-specific migration and settlement of the continental United States. *PLoS genetics*, 15(9), e1008225.

Jorde, L. B., Watkins, W. S., Bamshad, M. J., Dixon, M., Ricker, C., Seielstad, M., & Batzer,

M. A. (2000). The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *The American Journal of Human Genetics*, 66(3), 979-988.

Karafet, T., Xu, L., Du, R., Wang, W., Feng, S., Wells, R., Redd, A. J., Zegura, S. L., & Hammer, M. F. (2001). Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *The American Journal of Human Genetics*, 69(3), 615-628.

Karafet, T. M., Lansing, J. S., Redd, A. J., Reznikova, S., Watkins, J. C., Surata, S., Arthawiguna, W., Mayer, L., Bamshad, M., & Jorde, L. B. (2005). Balinese Ychromosome perspective on the peopling of Indonesia: genetic contributions from preneolithic hunter-gatherers, Austronesian farmers, and Indian traders. *Human biology*, 93-114.

Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L., & Hammer, M. F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome research*, 18(5), 830-838.

Karafet, T. M., Osipova, L. P., Gubina, M. A., Posukh, O. L., Zegura, S. L., & Hammer, M. F. (2002). High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter-gatherer way of life. *Human biology*, 761-789.

Kayser, M., Brauer, S., Cordaux, R., Casto, A., Lao, O., Zhivotovskiy, L. A., Moysé-Faurie, C., Rutledge, R. B., Schiefenhoefel, W., & Gil, D. (2006). Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Molecular Biology and Evolution*, 23(11), 2234-2244.

Kayser, M., & De Knijff, P. (2011). Improving human forensics through advances in genetics, genomics and molecular biology. *Nature Reviews Genetics*, 12(3), 179-192.

Kayser, M., Kittler, R., Erler, A., Hedman, M., Lee, A. C., Mohyuddin, A., Mehdi, S. Q., Rosser, Z., Stoneking, M., & Jobling, M. A. (2004). A comprehensive survey of human Ychromosomal microsatellites. *The American Journal of Human Genetics*, 74(6), 1183-1197.

Kehdy, F. S., Gouveia, M. H., Machado, M., Magalhães, W. C., Horimoto, A. R., Horta, B. L., Moreira, R. G., Leal, T. P., Scliar, M. O., & Soares-Souza, G. B. (2015). Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proceedings of the national Academy of sciences*, *112*(28), 8696-8701.

Kidd, J. M., Cooper, G. M., Donahue, W. F., Hayden, H. S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., & Antonacci, F. (2008). Mapping and sequencing of structural variation from eight human genomes. *Nature*, *453*(7191), 56-64.

Kim, E.H., Lee, H.Y., Yang, I.S., Yang, W.I., Shin, K.J. (2014). Population data for 30 insertion-deletion markers in a Korean population. *International Journal of Legal Medicine*, *128*(1), 51-52.

King, T. E., Bowden, G. R., Balaresque, P. L., Adams, S. M., Shanks, M. E., & Jobling, M. A. (2007). Thomas Jefferson's Y chromosome belongs to a rare European lineage. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, *132*(4), 584-589.

Kling, D., Tillmar, A. O., & Egeland, T. (2014). Familias 3–extensions and new functionality.

Forensic Science International: Genetics, *13*, 121-127.

Korbel, J. O., Urban, A. E., Affourtit, J. P., Godwin, B., Grubert, F., Simons, J. F., Kim, P. M., Palejev, D., Carriero, N. J., & Du, L. (2007). Paired-end mapping reveals extensive structural variation in the human genome. *Science*, *318*(5849), 420-426.

Kurland, C., & Andersson, S. (2000). Origin and evolution of the mitochondrial proteome. *Microbiology and Molecular Biology Reviews*, *64*(4), 786-820.

Kwon, S. Y., Lee, H. Y., Lee, E. Y., Yang, W. I., & Shin, K.-J. (2015). Confirmation of Y haplogroup tree topologies with newly suggested Y-SNPs for the C2, O2b and O3a subhaplogroups. *Forensic Science International: Genetics*, *19*, 42-46.

Lambert, C. A., & Tishkoff, S. (2009). Genetic structure in African populations: implications for human demographic history. Cold Spring Harbor symposia on quantitative biology, Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K.,

Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissole, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.-F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Smith, D. R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H. M., Dubois, J., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Roe, B. A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglu, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H.-C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G. R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent,

W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J., Slater, G., Smit, A. F. A., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, J. H., Yang, S.-P., Yeh, R.-F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Patrinos, A., Morgan, M. J., International Human Genome Sequencing, C., Whitehead Institute for Biomedical Research, C. f. G. R., The Sanger, C., Washington University Genome Sequencing, C., Institute, U. D. J. G., Baylor College of Medicine Human Genome Sequencing, C., Center, R. G. S., Genoscope, Cnrs, U. M. R., Department of Genome Analysis, I. o. M. B., Center, G. T. C. S., Beijing Genomics Institute/Human Genome, C., Multimegabase Sequencing Center, T. I. f. S. B., Stanford Genome Technology, C., University of Oklahoma's Advanced Center for Genome, T., Max Planck Institute for Molecular, G., Cold Spring Harbor Laboratory, L. A. H. G. C., Biotechnology, G. B. G. R. C. f., *Genome Analysis, G., Scientific management: National Human Genome Research Institute, U. S. N. I. o. H., Stanford Human Genome, C., University of Washington Genome, C., Department of Molecular Biology, K. U. S. o. M., University of Texas Southwestern Medical Center at, D., Office of Science, U. S. D. o. E., & The Wellcome, T. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860-921. <https://doi.org/10.1038/35057062>

Lao, O., van Duijn, K., Kersbergen, P., de Knijff, P., & Kayser, M. (2006). Proportioning wholegenome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry. *The American Journal of Human Genetics*, 78(4), 680-690.

Larmuseau, M. H., Vessi, A., Jobling, M. A., Van Geystelen, A., Primitivo, G., Biondi, G., Martínez-Labarga, C., Ottoni, C., Decorte, R., & Rickards, O. (2015). The paternal landscape along the Bight of

Benín—Testing regional representativeness of West-African population samples using Y-chromosomal markers. *PLoS one*, 10(11), e0141510.

Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. *PLoS Genet*, 8(1), e1002453.

Li, C., Zhao, S., Zhang, S., Li, L., Liu, Y., Chen, J., & Xue, J. (2011). Genetic polymorphism of 29 highly informative InDel markers for forensic use in the Chinese Han population. *Forensic Science International: Genetics*, 5(1), e27-e30.

Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., Cann, H. M., Barsh, G. S., Feldman, M., & Cavalli-Sforza, L. L. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. *Science*, 319(5866), 1100-1104.

Lipski, J. M. (2001). From bozal to boricua: Implications of Afro-Puerto Rican Language in Literature. *Hispania*, 850-859.

Lowe, A. L., Urquhart, A., Foreman, L. A., & Evett, I. W. (2001). Inferring ethnic origin by means of an STR profile. *Forensic science international*, 119(1), 17-22.

Luis, J. R., Rowold, D. J., Regueiro, M., Caeiro, B., Cinnioglu, C., Roseman, C., Underhill, P. A., Cavalli-Sforza, L. L., & Herrera, R. J. (2004). The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *The American Journal of Human Genetics*, 74(3), 532-544.

Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N., Nordenfelt, S., & Tandon, A. (2016). The Simons genome diversity project: 300 genomes from 142 diverse populations. *Nature*, 538(7624), 201-206.

Mancall, P. C., & Shammas, C. (2015). Governing the Sea in the Early Modern Era: Essays in Honor of Robert C. Ritchie (*San Marino: The Huntington Library*, 2015).

Manica, A., Prugnolle, F., & Balloux, F. (2005). Geography is a better determinant of human genetic differentiation than ethnicity. *Human genetics*, 118(3), 366-371.

Mao, X., Bigham, A. W., Mei, R., Gutierrez, G., Weiss, K. M., Brutsaert, T. D., Leon-Velarde, F., Moore, L. G., Vargas, E., &

McKeigue, P. M. (2007). A genomewide admixture mapping panel for Hispanic/Latino populations. *The American Journal of Human Genetics*, 80(6), 1171-1178.

Maples, B. K., Gravel, S., Kenny, E. E., & Bustamante, C. D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *The American Journal of Human Genetics*, 93(2), 278-288.

Marks, S. J., Montinaro, F., Levy, H., Brisighelli, F., Ferri, G., Bertocini, S., Batini, C., Busby, G. B., Arthur, C., & Mitchell, P. (2015). Static and moving frontiers: the genetic landscape of Southern African Bantu-speaking populations. *Molecular Biology and Evolution*, 32(1), 29-43.

Marth, G. T., Czabarka, E., Murvai, J., & Sherry, S. T. (2004). The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics*, 166(1), 351-372.

Martínez-Cortés, G., Gusmão, L., Pereira, R., Salcido, V., Favela-Mendoza, A., Muñoz-Valle, J., Inclán-Sánchez, A., López-Hernández, L., & Rangel-Villalobos, H. (2015). Genetic structure and forensic parameters of 38 Indels for human identification purposes in eight Mexican populations. *Forensic Science International: Genetics*, 17, 149-152.

Martínez, B., Builes, J., Gaviria, A., Burgos, G., Manrique, A., Aguirre, D., Mendoza, L., Bravo, M., Pereira, R., & Gusmão, L. (2013). Population genetic data of 38 autosomal InDels in San Basilio de Palenque, the first free town in America. *Forensic Science International: Genetics Supplement Series*, 4(1), e73-e74.

Martínez, B., Pereira, R., Meza, K., Hernández, L., Amorim, A., Marrugo, J., & Gusmão, L. (2017). Ancestry estimates in afrodescendant population from San Basilio de Palenque, Colombia. *Forensic Science International: Genetics Supplement Series*, 6, e224-e225.

Martínez, B., Simão, F., Gomes, V., Nguidi, M., Amorim, A., Carvalho, E. F., Marrugo, J., & Gusmão, L. (2020). Searching for the roots of the first free African American community. *Scientific reports*, 10(1), 1-11.

Mathias, R. A., Taub, M. A., Gignoux, C. R., Fu, W., Musharoff, S., O'Connor, T. D., Vergara, C., Torgerson, D. G., Pino-Yanes, M., & Shringarpure, S. S. (2016). A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. *Nature communications*, 7(1), 1-10.

Medina-Rivas, M. A., Norris, E. T., Rishishwar, L., Conley, A. B., Medrano-Trochez, C., Valderrama-Aguirre, A., Vannberg, F. O., Mariño-Ramírez, L., & Jordan, I. K. (2016). Choco, Colombia: a hotspot of human biodiversity. *Revista biodiversidad neotropical*, 6(1), 45.

Mezzavilla, M., Geppert, M., Tyler-Smith, C., Roewer, L., & Xue, Y. (2015). Insights into the origin of rare haplogroup C3* Y chromosomes in South America from high-density autosomal SNP genotyping. *Forensic Science International: Genetics*, 15, 115-120.

Micheletti, S. J., Bryc, K., Esselmann, S. G. A., Freyman, W. A., Moreno, M. E., Poznik, G. D., Shastri, A. J., Agee, M., Aslibekyan, S., & Auton, A. (2020). Genetic consequences of the transatlantic slave trade in the Americas. *The American Journal of Human Genetics*, 107(2), 265-277.

Miller, S., Dykes, D., & Polesky, H. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic acids res*, 16(3), 1215.

Mills, R. E., Luttig, C. T., Larkins, C. E., Beauchamp, A., Tsui, C., Pittard, W. S., & Devine, S. E. (2006). An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome research*, 16(9), 1182-1190.

Mills, R. E., Pittard, W. S., Mullaney, J. M., Farooq, U., Creasy, T. H., Mahurkar, A. A., Kemeza, D. M., Strassler, D. S., Ponting, C. P., & Webber, C. (2011). Natural genetic variation caused by small insertions and deletions in the human genome. *Genome research*, 21(6), 830-839.

Mincultura, M. d. C. d. C. (2020). <http://www.mincultura.gov.co/?idcategoria=21356>

Mogensen, H. S., Tvedebrink, T., Børsting, C., Pereira, V., & Morling, N. (2020). Ancestry prediction efficiency of the software GenoGeographer using a z-score method and the ancestry informative

markers in the Precision ID Ancestry Panel. *Forensic Science International: Genetics*, 44, 102154.

Moreno-Estrada, A., Gignoux, C. R., Fernández-López, J. C., Zakharia, F., Sikora, M., Contreras, A. V., Acuña-Alonzo, V., Sandoval, K., Eng, C., & Romero-Hidalgo, S. (2014). The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science*, 344(6189), 1280-1285.

Moreno-Estrada, A., Gravel, S., Zakharia, F., McCauley, J. L., Byrnes, J. K., Gignoux, C. R., Ortiz-Tello, P. A., Martínez, R. J., Hedges, D. J., & Morris, R. W. (2013). Reconstructing the population genetic history of the Caribbean. *PLoS Genet*, 9(11), e1003925.

Mufwene, S. S. (2008). Kikongo-Kituba. In E. Britannica (Ed.), *Encyclopedia Britannica*. Edinburgh.

Mullaney, J. M., Mills, R. E., Pittard, W. S., & Devine, S. E. (2010). Small insertions and deletions (INDELs) in human genomes. *Human molecular genetics*, 19(R2), R131R136.

Mullen, K. M., & van Stokkum, I. H. (2015). The Lawson-Hanson algorithm for non-negative least squares (NNLS). *R package*.

Múnera, A. Cartagena afro: la gran ciudad del Caribe, 1580-1640. En: Múnera, A. La independencia de Colombia: olvidos y ficciones. Cartagena de Indias (1580-1821). Bogotá: Editorial Planeta, 2021.

Myres, N. M., Rootsi, S., Lin, A. A., Järve, M., King, R. J., Kutuev, I., Cabrera, V. M., Khusnutdinova, E. K., Pshenichnov, A., & Yunusbayev, B. (2011). A major Ychromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *European Journal of Human Genetics*, 19(1), 95-101.

Nasidze, I., Quinque, D., Ozturk, M., Bendukidze, N., & Stoneking, M. (2005). MtDNA and Ychromosome variation in Kurdish groups. *Annals of human genetics*, 69(4), 401-412.

Navarrete, M. C. (2017). Palenques: Maroons and Castas in Colombia's Caribbean Regions. *Orality, Identity, and Resistance in Palenque (Colombia): An interdisciplinary approach*, 54, 269.

Navarrete Peláez, M. C. (2008). *San Basilio de Palenque: memoria y tradición: Surgimiento y avatares de las gestas cimarronas en el Caribe colombiano*. Programa Editorial UNIVALLE.

Newman, J. L. (1995). *The peopling of Africa: a geographic interpretation*. Yale University Press.

Nilsson, M., Andréasson-Jansson, H., Ingman, M., & Allen, M. (2008). Evaluation of mitochondrial DNA coding region assays for increased discrimination in forensic analysis. *Forensic Science International: Genetics*, 2(1), 1-8.

Noguera, M. C., Schwegler, A., Gomes, V., Briceño, I., Alvarez, L., Uricoechea, D., Amorim,

A., Benavides, E., Silvera, C., & Charris, M. (2014). Colombia's racial crucible: Y chromosome evidence from six admixed communities in the Department of Bolivar. *Annals of human biology*, 41(5), 453-459.

Norris, E. T., Rishishwar, L., Chande, A. T., Conley, A. B., Ye, K., Valderrama-Aguirre, A., & Jordan, I. K. (2020). Admixture-enabled selection for rapid adaptive evolution in the Americas. *Genome biology*, 21(1), 1-12.

Norris, E. T., Rishishwar, L., Wang, L., Conley, A. B., Chande, A. T., Dabrowski, A. M., Valderrama-Aguirre, A., & Jordan, I. K. (2019). Assortative mating on ancestry-variant traits in admixed Latin American populations. *Frontiers in genetics*, 10, 359.

Oliveira, S., Hübner, A., Fehn, A.-M., Aço, T., Lages, F., Pakendorf, B., Stoneking, M., & Rocha, J. (2019). The role of matrilineality in shaping patterns of Y chromosome and mtDNA sequence variation in southwestern Angola. *European Journal of Human Genetics*, 27(3), 475-483.

Ortega, A. V. (2002). *Cartagena de Indias y la región histórica del Caribe, 1580-1640* (Vol. 81). Universidad de Sevilla.

Ossa, H., Aquino, J., Pereira, R., Ibarra, A., Ossa, R. H., Pérez, L. A., Granda, J. D., Lattig, M. C., Groot, H., & Fagundes de Carvalho, E. (2016). Outlining the ancestry landscape of Colombian admixed populations. *PloS one*, 11(10), e0164414.

Palmer, C. A. (2000). Defining and studying the modern African diaspora. *The Journal of Negro History*, 85(1-2), 27-32.

Paneto, G. G., Martins, J. A., Longo, L. V., Pereira, G. A., Freschi, A., Alvarenga, V. L., Chen, B., Oliveira, R. N., Hirata, M. H., & Cicarelli, R. M. (2007). Heteroplasmy in hair: differences among hair

and blood from the same individuals are still a matter of debate. *Forensic science international*, 173(2-3), 117-121.

Parson, W., Brandstätter, A., Pircher, M., Steinlechner, M., & Scheithauer, R. (2004). EMPOP—the EDNAP mtDNA population database concept for a new generation, highquality mtDNA database. International Congress Series,

Parson, W., Parsons, T. J., Scheithauer, R., & Holland, M. M. (1998). Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. *International Journal of Legal Medicine*, 111(3), 124-132.

Parson., L. P. W. (2014). Curso Bioinformática y nuevas técnicas (NGS) para ADN mitocondrial y sus aplicaciones forenses. XIX Jornadas GEHP-ISFG,

Paschou, P., Ziv, E., Burchard, E. G., Choudhry, S., Rodriguez-Cintron, W., Mahoney, M. W., & Drineas, P. (2007). PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genet*, 3(9), e160.

Patin, E., Lopez, M., Grollemund, R., Verdu, P., Harmant, C., Quach, H., Laval, G., Perry, G. H., Barreiro, L. B., & Froment, A. (2017). Dispersals and genetic adaptation of Bantuspeaking populations in Africa and North America. *Science*, 356(6337), 543-546.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genshoreck, T., Webster, T., & Reich, D. (2012). Ancient admixture in human history. *Genetics*, 192(3), 1065-1093.

Pereira, R., Alves, C., Aler, M., Amorim, A., Arévalo, C., Betancor, E., Braganholi, D., Bravo, M., Brito, P., & Builes, J. (2018). A GHEP-ISFG collaborative study on the genetic variation of 38 autosomal indels for human identification in different continental populations. *Forensic Science International: Genetics*, 32, 18-25.

Pereira, R., Phillips, C., Alves, C., Amorim, A., Carracedo, Á., & Gusmão, L. (2009). A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis*, 30(21), 3682-3690.

Pereira, R., Phillips, C., Pinto, N., Santos, C., dos Santos, S. E. B., Amorim, A., Carracedo, Á., & Gusmão, L. (2012). Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS one*, 7(1), e29684.

Pereira, R. M. L. (2011). *Bridging the gap between SNPs and STRs: insertion deletion polymorphisms in forensic genetics, principles and applications* Universidade de Santiago de Compostela].

Pereira, V., Santangelo, R., Børsting, C., Tvedebrink, T., Almeida, A.P.F., Carvalho, E.F., Morling, N., Gusmão, L. (2020). Evaluation of the Precision of Ancestry Inferences in South American Admixed Populations. *Frontiers in Genetics, 11*:966.

Phillips, C., Fernandez-Formoso, L., Gelabert-Besada, M., Garcia-Magariños, M., Santos, C., Fondevila, M., Carracedo, Á., & Lareu, M. V. (2013). Development of a novel forensic STR multiplex for ancestry analysis and extended identity testing. *Electrophoresis, 34*(8), 1151-1162.

Phillips, C., Salas, A., Sanchez, J., Fondevila, M., Gomez-Tato, A., Alvarez-Dios, J., Calaza, M., de Cal, M. C., Ballard, D., & Lareu, M. (2007). Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Science International: Genetics, 1*(3-4), 273-280.

Phillips, T. (2011). Brazil census shows African-Brazilians in the majority for the first time. *The Guardian, 17*.

Phillipson, D. W. (2005). *African archaeology*. Cambridge University Press.

Pinto, N., Magalhães, M., Conde-Sousa, E., Gomes, C., Pereira, R., Alves, C., Gusmão, L., & Amorim, A. (2013). Assessing paternities with inconclusive STR results: the suitability of bi-allelic markers. *Forensic Science International: Genetics, 7*(1), 16-21.

Plaza, S., Salas, A., Calafell, F., Corte-Real, F., Bertranpetit, J., Carracedo, Á., & Comas, D. (2004). Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. *Human genetics, 115*(5), 439-447.

Pollitzer, W. S. (1999). *The Gullah people and their African heritage*. University of Georgia Press.

Poulsen, L., Friis, S. L., Hallenberg, C., Simonsen, B. T., & Morling, N. (2014). A report of the 2009–2011 paternity and relationship testing workshops of the English Speaking Working Group of the International Society For Forensic Genetics. *Forensic Science International: Genetics, 9*, e1-e2.

Pour, N. A., Plaster, C. A., & Bradman, N. (2013). Evidence from Y-chromosome analysis for a late exclusively eastern expansion of the Bantu-speaking people. *European Journal of Human Genetics*, *21*(4), 423-429.

Poznik, G. D., Henn, B. M., Yee, M.-C., Sliwerska, E., Euskirchen, G. M., Lin, A. A., Snyder, M., Quintana-Murci, L., Kidd, J. M., & Underhill, P. A. (2013). Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science*, *341*(6145), 562-565.

Price, A. L., Patterson, N., Yu, F., Cox, D. R., Waliszewska, A., McDonald, G. J., Tandon, A., Schirmer, C., Neubauer, J., & Bedoya, G. (2007). A genomewide admixture map for Latino populations. *The American Journal of Human Genetics*, *80*(6), 1024-1036.

Prugnolle, F., Manica, A., & Balloux, F. (2005). Geography predicts neutral genetic diversity of human populations. *Current Biology*, *15*(5), R159-R160.

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., Fiegler, H., Shapero, M. H., Carson, A. R., & Chen, W. (2006). Global variation in copy number in the human genome. *Nature*, *444*(7118), 444-454.

Regueiro, M., Cadenas, A. M., Gayden, T., Underhill, P. A., & Herrera, R. J. (2006). Iran: tricontinental nexus for Y-chromosome driven migration. *Human heredity*, *61*(3), 132-143.

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M. V., Rojas, W., Duque, C., & Mesa, N. (2012). Reconstructing native American population history. *Nature*, *488*(7411), 370-374.

Resque, R., Gusmão, L., Geppert, M., Roewer, L., Palha, T., Alvarez, L., Ribeiro-dos-Santos, Â., & Santos, S. (2016). Male lineages in Brazil: Intercontinental admixture and stratification of the European background. *PLoS one*, *11*(4), e0152573.

Rishishwar, L., Conley, A. B., Wigington, C. H., Wang, L., Valderrama-Aguirre, A., & Jordan, I. K. (2015). Ancestry, admixture and fitness in Colombian genomes. *Scientific reports*, *5*(1), 1-16.

Rishishwar, L., & Jordan, I. K. (2017). Implications of human evolution and admixture for mitochondrial replacement therapy. *BMC genomics*, *18*(1), 1-11.

Rito, T., Richards, M. B., Fernandes, V., Alshamali, F., Cerny, V., Pereira, L., & Soares, P. (2013). The first modern human dispersals across Africa. *PLoS one*, *8*(11), e80031.

Roberts, K., Alberts, B., Johnson, A., Walter, P., & Hunt, T. (2002). *Molecular biology of the cell*. New York: Garland Science.

Robin, E. D., & Wong, R. (1988). Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *Journal of cellular physiology*, *136*(3), 507-513.

Rodrigues de Moura, R., Coelho, A. V. C., de Queiroz Balbino, V., Crovella, S., & Brandão, L. A. C. (2015). Meta-analysis of Brazilian genetic admixture and comparison with other Latin America countries. *American Journal of Human Biology*, *27*(5), 674-680.

Romanini, C., Catelli, M., Borosky, A., Pereira, R., Romero, M., Puerto, M. S., Phillips, C.,

Fondevila, M., Freire, A., & Santos, C. (2012). Typing short amplicon binary polymorphisms: supplementary SNP and Indel genetic information in the analysis of highly degraded skeletal remains. *Forensic Science International: Genetics*, *6*(4), 469-476.

Rootsi, S., Kivisild, T., Benuzzi, G., Help, H., Bermisheva, M., Kutuev, I., Barac, L., Peričić, M., Balanovsky, O., & Pshenichnov, A. (2004). Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *The American Journal of Human Genetics*, *75*(1), 128-137.

Rosa, A., Ornelas, C., Jobling, M. A., Brehm, A., & Villems, R. (2007). Y-chromosomal diversity in the population of Guinea-Bissau: a multiethnic perspective. *BMC evolutionary biology*, *7*(1), 1-11.

Rosenberg, N. A., Li, L. M., Ward, R., & Pritchard, J. K. (2003). Informativeness of genetic markers for inference of ancestry. *The American Journal of Human Genetics*, *73*(6), 1402-1422.

Rosenberg, N. A., Mahajan, S., Ramachandran, S., Zhao, C., Pritchard, J. K., & Feldman, M. W. (2005). Clines, clusters, and the

effect of study design on the inference of human population structure. *PLoS Genet*, 1(6), e70.

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., & Feldman, M. W. (2002). Genetic structure of human populations. *Science*, 298(5602), 2381-2385.

Rosser, Z. H., Zerjal, T., Hurler, M. E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., & Barbujani, G. (2000). Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *The American Journal of Human Genetics*, 67(6), 1526-1543.

Rotimi, C. N., Bentley, A. R., Doumatey, A. P., Chen, G., Shriner, D., & Adeyemo, A. (2017). The genomic landscape of African populations in health and disease. *Human molecular genetics*, 26(R2), R225-R236.

Rotimi, C. N., Tekola-Ayele, F., Baker, J. L., & Shriner, D. (2016). The African diaspora: history, adaptation and health. *Current opinion in genetics & development*, 41, 77-84.

Roulston, K. (2020). Examining the Past to Consider the Present: Long Past Slavery:

Representing Race in the Federal Writers' Project. Catherine A. Stewart (Ed.), 2016, Chapel Hill, NC: The University of North Carolina Press. 372 pp. . In: Taylor & Francis.

Rowold, D., Garcia-Bertrand, R., Calderon, S., Rivera, L., Benedico, D. P., Sanchez, M. A. A., Chennakrishnaiah, S., Varela, M., & Herrera, R. J. (2014). At the southeast fringe of the bantu expansion: genetic diversity and phylogenetic relationships to other sub-Saharan tribes. *Meta gene*, 2, 670-685.

Rowold, D. J., Perez-Benedico, D., Stojkovic, O., Garcia-Bertrand, R., & Herrera, R. J. (2016). On the Bantu expansion. *Gene*, 593(1), 48-57.

Ruiz-Linares, A. (2015). How genes have illuminated the history of early Americans and Latino Americans. *Cold Spring Harbor perspectives in biology*, 7(6), a008557.

Ruiz-Linares, A., Ortíz-Barrientos, D., Figueroa, M., Mesa, N., Múnera, J. G., Bedoya, G., VElez, I. D., García, L. F., Pérez-Lezaun, A., & Bertranpetit, J. (1999). Microsatellites provide evidence for Y

chromosome diversity among the founders of the New World. *Proceedings of the national Academy of sciences*, 96(11), 6312-6317.

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry,

S., Mullikin, J. C., Mortimore, B. J., & Willey, D. L. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409(6822), 928-934.

Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4), 406-425.

Salas, A., Bandelt, H.-J., Macaulay, V., & Richards, M. B. (2007). Phylogeographic investigations: the role of trees in forensic genetics. *Forensic science international*, 168(1), 1-13.

Salas, A., Richards, M., De la Fe, T., Lareu, M.-V., Sobrino, B., Sánchez-Diz, P., Macaulay, V., & Carracedo, Á. (2002). The making of the African mtDNA landscape. *The American Journal of Human Genetics*, 71(5), 1082-1111.

Salas, A., Richards, M., Lareu, M.-V., Scozzari, R., Coppa, A., Torroni, A., Macaulay, V., & Carracedo, Á. (2004). The African diaspora: mitochondrial DNA and the Atlantic slave trade. *The American Journal of Human Genetics*, 74(3), 454-465.

Sanchez, J. J., Phillips, C., Børsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C. D., Musgrave-Brown, E., Salas, A., & Syndercombe-Court, D. (2006). A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis*, 27(9), 1713-1724.

Santos-Lopes, S. S., Pereira, R. W., Wilson, I. J., & Pena, S. D. (2007). A worldwide phylogeography for the human X chromosome. *PloS one*, 2(6), e557.

Scozzari, R., Massaia, A., D'Atanasio, E., Myres, N. M., Perego, U. A., Trombetta, B., & Cruciani, F. (2012). Molecular dissection of the basal clades in the human Y chromosome phylogenetic tree. *PloS one*, 7(11), e49170.

Scozzari, R., Massaia, A., Trombetta, B., Bellusci, G., Myres, N. M., Novelletto, A., & Cruciani, F. (2014). An unbiased resource of novel SNP markers provides a new chronology for the human Y

chromosome and reveals a deep phylogenetic structure in Africa. *Genome research*, 24(3), 535-544.

Schaffner, S. F. (2004). The X chromosome in population genetics. *Nature Reviews Genetics*, 5(1), 43-51.

Scheinfeldt, L. B., Soi, S., & Tishkoff, S. A. (2010). Working toward a synthesis of archaeological, linguistic, and genetic data for inferring African population history. *Proceedings of the national Academy of sciences*, 107(Supplement 2), 8931-8938.

Schlötterer, C., & Wiehe, T. (1999). Microsatellites, a neutral marker to infer selective sweeps. In *Microsatellites* (pp. 238-248). Oxford University Press.

Schurr, T. G., Dulik, M. C., Owings, A. C., Zhadanov, S. I., Gaiieski, J. B., Vilar, M. G., Ramos, J., Moss, M. B., Natkong, F., & Consortium, G. (2012). Clan, language, and migration history has shaped genetic diversity in Haida and Tlingit populations from Southeast Alaska. *American journal of physical anthropology*, 148(3), 422-435.

Schuster, S. C., Miller, W., Ratan, A., Tomsho, L. P., Giardine, B., Kasson, L. R., Harris, R. S., Petersen, D. C., Zhao, F., & Qi, J. (2010). Complete Khoisan and Bantu genomes from southern Africa. *Nature*, 463(7283), 943-947.

Schwegler, A. (1999). El vocabulario africano de Palenque (Colombia): Segunda parte: compendio alfabético de palabras (con etimologías). *El vocabulario africano de Palenque (Colombia): segunda parte: compendio alfabético de palabras (con etimologías)*, 171-253.

Schwegler, A. (2000). The African Vocabulary of Palenque (Colombia). *Journal of Pidgin and Creole languages*, 15(2), 241-312.

Schwegler, A. (2011). Palenque (ro): the search for its African substrate. *Creoles, their substrates, and language typology*, 225-249.

Schwegler, A. (2016). Combining population genetics (DNA) with historical linguistics. *Spanish language and sociolinguistic analysis*, 8, 33.

Schwegler, A. (2017). On the African origin (s) of Palenquero. *Orality, identity, and resistance in Palenque (Colombia)*, 51-119.

Semino, O., Passarino, G., Oefner, P. J., Lin, A. A., Arbuzova, S., Beckman, L. E., De Benedictis, G., Francalacci, P., Kouvatsi, A., & Limborska, S. (2000). The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: AY chromosome perspective. *Science*, 290(5494), 1155-1159.

Semino, O., Santachiara-Benerecetti, A. S., Falaschi, F., Cavalli-Sforza, L. L., & Underhill, P. A. (2002). Ethiopians and Khoisan share the deepest clades of the human Y chromosome phylogeny. *The American Journal of Human Genetics*, 70(1), 265-268.

Sengupta, S., Zhivotovsky, L. A., King, R., Mehdi, S. Q., Edmonds, C. A., Chow, C.-E. T., Lin, A. A., Mitra, M., Sil, S. K., & Ramesh, A. (2006). Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *The American Journal of Human Genetics*, 78(2), 202-221.

Serre, D., & Pääbo, S. (2004). Evidence for gradients of human genetic diversity within and among continents. *Genome research*, 14(9), 1679-1685.

Severson, A. L., Shortt, J. A., Mendez, F. L., Wojcik, G. L., Bustamante, C. D., & Gignoux, C. R. (2018). SNAPPY: Single Nucleotide Assignment of Phylogenetic Parameters on the Y chromosome. *BioRxiv*, 454736.

Shriner, D., Adeyemo, A., Ramos, E., Chen, G., & Rotimi, C. N. (2011). Mapping of disease-associated variants in admixed populations. *Genome biology*, 12(5), 1-8.

Shriner, D., Bentley, A. R., Doumatey, A. P., Chen, G., Zhou, J., Adeyemo, A., & Rotimi, C. N. (2015). Phenotypic variance explained by local ancestry in admixed African Americans. *Frontiers in genetics*, 6, 324.

Shriner, D., Tekola-Ayele, F., Adeyemo, A., & Rotimi, C. N. (2014). Genome-wide genotype and sequence-based reconstruction of the 140,000 year history of modern human ancestry. *Scientific reports*, 4(1), 1-9.

Shriver, M. D., Mei, R., Parra, E. J., Sonpar, V., Halder, I., Tishkoff, S. A., Schurr, T. G., Zhadanov, S. I., Osipova, L. P., &

Brutsaert, T. D. (2005). Large-scale SNP analysis reveals clustered and continuous patterns of human genetic variation. *Human genomics*, 2(2), 1-9.

Shriver, M. D., Parra, E. J., Dios, S., Bonilla, C., Norton, H., Jovel, C., Pfaff, C., Jones, C., Massac, A., & Cameron, N. (2003). Skin pigmentation, biogeographical ancestry and admixture mapping. *Human genetics*, 112(4), 387-399.

Silva, M., Alshamali, F., Silva, P., Carrilho, C., Mandlate, F., Trovoada, M. J., Černý, V., Pereira, L., & Soares, P. (2015). 60,000 years of interactions between Central and Eastern Africa documented by major African mitochondrial haplogroup L2. *Scientific reports*, 5(1), 1-13.

Simão, F., Ferreira, A. P., de Carvalho, E. F., Parson, W., & Gusmão, L. (2018). Defining mtDNA origins and population stratification in Rio de Janeiro. *Forensic Science International: Genetics*, 34, 97-104.

Simão, F., Strobl, C., Vullo, C., Catelli, L., Machado, P., Huber, N., Schnaller, L., Huber, G., Xavier, C., & Carvalho, E. F. (2019). The maternal inheritance of Alto Paraná revealed by full mitogenome sequences. *Forensic Science International: Genetics*, 39, 66-72.

Sims, L. M., Garvey, D., & Ballantyne, J. (2007). Sub-populations within the major European and African derived haplogroups R1b3 and E3a are differentiated by previously phylogenetically undefined Y-SNPs. *Human mutation*, 28(1), 97-97.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P. J., Cordum, H. S., Hillier, L., Brown, L. G., Repping, S., Pyntikova, T., Ali, J., & Bieri, T. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, 423(6942), 825837.

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, 195(3), 693-702.

Soares, P., Rito, T., Pereira, L., & Richards, M. B. (2016). A genetic perspective on African prehistory. In *Africa from MIS 6-2* (pp. 383-405). Springer.

Sobrinho, B., Brion, M., & Carracedo, A. (2005). SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic science international*, *154*(2-3), 181-194.

Solé-Morata, N., Bertranpetit, J., Comas, D., & Calafell, F. (2015). Y-chromosome diversity in Catalan surname samples: insights into surname origin and frequency. *European Journal of Human Genetics*, *23*(11), 1549-1557.

Strachan, T., & Read, A. P. (2018). Human molecular genetics.

Su, B., Jin, L., Underhill, P., Martinson, J., Saha, N., McGarvey, S. T., Shriver, M. D., Chu, J., Oefner, P., & Chakraborty, R. (2000). Polynesian origins: insights from the Y chromosome. *Proceedings of the national Academy of sciences*, *97*(15), 8225-8228.

Szibor, R., Krawczak, M., Hering, S., Edelmann, J., Kuhlisch, E., & Krause, D. (2003). Use of X-linked markers for forensic purposes. *International Journal of Legal Medicine*, *117*(2), 67-74.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, *30*(12), 2725-2729.

Tarazona-Santos, E., Carvalho-Silva, D. R., Pettener, D., Luiselli, D., De Stefano, G. F., Labarga, C. M., Rickards, O., Tyler-Smith, C., Pena, S. D., & Santos, F. R. (2001). Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *The American Journal of Human Genetics*, *68*(6), 1485-1496.

Tereba, A. (1999). Profiles in DNA. In P. Corporation (Ed.), *Tools for Analysis of Population Statistics*. (Vol. 2).

Thangaraj, K., Singh, L., Reddy, A. G., Rao, V. R., Sehgal, S. C., Underhill, P. A., Pierson, M., Frame, I. G., & Hagelberg, E. (2003). Genetic affinities of the Andaman Islanders, a vanishing human population. *Current Biology*, *13*(2), 86-93.

Thomas, M. G., Bradman, N., & Flinn, H. M. (1999). High throughput analysis of 10 microsatellite and 11 diallelic polymorphisms on the human Y-chromosome. *Human genetics*, *105*(6), 577-581.

Thomas, M. G., Parfitt, T., Weiss, D. A., Skorecki, K., Wilson, J. F., Le Roux, M., Bradman, N., & Goldstein, D. B. (2000). Y chromosomes traveling south: the Cohen modal haplotype and the

origins of the Lemba—the “Black Jews of Southern Africa”. *The American Journal of Human Genetics*, 66(2), 674-686.

Thornton, J., & Thornton, J. K. T. (1998). *Africa and Africans in the Making of the Atlantic World, 1400-1800*. Cambridge University Press.

Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J. B., Awomoyi, A. A., Bodo, J.-M., & Doumbo, O. (2009). The genetic structure and history of Africans and African Americans. *Science*, 324(5930), 1035-1044.

Torroni, A., Bandelt, H.-J., D'urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M.-L., & Bonn -Tamir, B. (1998). mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *The American Journal of Human Genetics*, 62(5), 1137-1152.

Torroni, A., Schurr, T. G., Cabell, M. F., Brown, M. D., Neel, J. V., Larsen, M., Smith, D. G., Vullo, C. M., & Wallace, D. C. (1993). Asian affinities and continental radiation of the four founding Native American mtDNAs. *American journal of human genetics*, 53(3), 563.

Trombetta, B., D'Atanasio, E., Massaia, A., Ippoliti, M., Coppa, A., Candilio, F., Coia, V., Russo, G., Dugoujon, J.-M., & Moral, P. (2015). Phylogeographic refinement and large scale genotyping of human Y chromosome haplogroup E provide new insights into the dispersal of early pastoralists in the African continent. *Genome biology and evolution*, 7(7), 1940-1950.

Tully, G., B r, W., Brinkmann, B., Carracedo, A., Gill, P., Morling, N., Parson, W., & Schneider, P. (2001). Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. *Forensic science international*, 124(1), 83-91.

Underhill, P. A., Jin, L., Zemans, R., Oefner, P. J., & Cavalli-Sforza, L. L. (1996). A preColumbian Y chromosome-specific transition and its implications for human evolutionary history. *Proceedings of the national Academy of sciences*, 93(1), 196-200. Underhill, P. A., Passarino, G., Lin, A. A., Shen, P., Mirazon Lahr, M., Foley, R. A., Oefner, P. J., & Cavalli-Sforza, L. L. (2001). The phylogeography of Y

chromosome binary haplotypes and the origins of modern human populations. *Annals of human genetics*, 65(1), 43-62.

Van Oven, M., Van Geystelen, A., Kayser, M., Decorte, R., & Larmuseau, M. H. (2014). Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. *Human mutation*, 35(2), 187-191.

Vansina, J. (1995). New linguistic evidence and 'The Bantu expansion'. *Journal of African history*, 173-195.

Varela T. A., A. R. L., Faroña, J. (2009). El ADN mitocondrial y las relaciones filogenéticas de los últimos estadios del género homo. *Revista de la Real Academia Galega de Ciencias* 103-118.

Veeramah, K. R., Connell, B. A., Pour, N. A., Powell, A., Plaster, C. A., Zeitlyn, D., Mendell, N. R., Weale, M. E., Bradman, N., & Thomas, M. G. (2010). Little genetic differentiation as assessed by uniparental markers in the presence of substantial language variation in peoples of the Cross River region of Nigeria. *BMC evolutionary biology*, 10(1), 1-17.

Vilar, E. V. (1977). *Hispanoamérica y el comercio de esclavos* (Vol. 239). Escuela de estudios hispano-americanos Seville.

Wang, C., Zhan, X., Liang, L., Abecasis, G. R., & Lin, X. (2015). Improved ancestry estimation for both genotyping and sequencing data using projection procrustes analysis and genotype imputation. *The American Journal of Human Genetics*, 96(6), 926-937.

Wang, S., Ray, N., Rojas, W., Parra, M. V., Bedoya, G., Gallo, C., Poletti, G., Mazzotti, G., Hill, K., & Hurtado, A. M. (2008). Geographic patterns of genome admixture in Latin American Mestizos. *PLoS Genet*, 4(3), e1000037.

Weber, J. L., David, D., Heil, J., Fan, Y., Zhao, C., & Marth, G. (2002). Human diallelic insertion/deletion polymorphisms. *The American Journal of Human Genetics*, 71(4), 854-862.

Weber, J. L., & May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American journal of human genetics*, 44(3), 388.

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., & Schönherr, S. (2016). HaploGrep 2: mitochondrial haplogroup classification in the

era of high-throughput sequencing. *Nucleic acids research*, 44(W1), W58-W63.

Wheat, D. (2013). Global Transit Points and Travel in the Iberian Maritime World, 1580–1640. *Governing the Sea in the Early Modern Era*.

Wheat, D. (2016). *Atlantic Africa and the Spanish Caribbean, 1570-1640*. UNC Press Books.

Winchester, L., Yau, C., & Ragoussis, J. (2009). Comparing CNV detection methods for SNP arrays. *Briefings in functional genomics and proteomics*, 8(5), 353-366.

Wojcik, G. L., Graff, M., Nishimura, K. K., Tao, R., Haessler, J., Gignoux, C. R., Highland, H. M., Patel, Y. M., Sorokin, E. P., & Avery, C. L. (2019). Genetic analyses of diverse populations improve discovery for complex traits. *Nature*, 570(7762), 514-518.

Wood, E. T., Stover, D. A., Ehret, C., Destro-Bisol, G., Spedini, G., McLeod, H., Louie, L., Bamshad, M., Strassmann, B. I., & Soodyall, H. (2005). Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. *European Journal of Human Genetics*, 13(7), 867-876.

Yan, C., Duanmu, X., Zeng, L., Liu, B., & Song, Z. (2019). Mitochondrial DNA: distribution, mutations, and elimination. *Cells*, 8(4), 379.

Zaitlen, N., Huntsman, S., Hu, D., Spear, M., Eng, C., Oh, S. S., White, M. J., Mak, A., Davis, A., & Meade, K. (2017). The effects of migration and assortative mating on admixture linkage disequilibrium. *Genetics*, 205(1), 375-383.

Zegura, S. L., Karafet, T. M., Zhivotovsky, L. A., & Hammer, M. F. (2004). High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. *Molecular Biology and Evolution*, 21(1), 164-175.

Zidkova, A., Horinek, A., Kebrdlova, V., & Korabecna, M. (2013). Application of the new insertion–deletion polymorphism kit for forensic identification and parentage testing on the Czech population. *International Journal of Legal Medicine*, 127(1), 7-10.

10. ANEXOS

Anexo A. Consentimiento informado de toma de muestras en SBP

Universidad de Cartagena

Vicerrectoría de Investigaciones

Instituto de investigaciones Inmunológicas

CONSENTIMIENTO INFORMADO

Me han explicado con toda claridad las condiciones del estudio: “” ha realizar por el grupo de Alergología experimental e Inmunogenética. He leído o me han leído este formato de información y consentimiento y he tenido oportunidad de hacer todas las preguntas necesarias para entender el estudio; estas han sido respondidas con toda satisfacción. Tengo claro que el estudio no representará ningún tipo de riesgo para mí, que mi participación en este es voluntaria y que por esta razón me puedo retirar del estudio cuando yo lo desee. Tengo claro que el resultado de este estudio no me traerá ningún tipo de beneficio económico. Tengo presente que los resultados encontrados en el estudio son confidenciales y su uso será sólo con fines científicos. He recibido una copia de este formato de consentimiento para futura referencia. Acepto participar libremente en este estudio.

Firma del individuo o representante y documento de identidad

Declaración del encargado de este estudio: He explicado al participante o su representante la naturaleza, las necesidades y los riesgos previsibles del estudio y he sido testigo de la realización del consentimiento.

Firma del encargado del estudio Teléfono Fecha

Firma del Testigo y documento de identidad

Número de Inclusión en el estudio: _____

Anexo B

MARTÍNEZ, Beatriz, et al. Searching for the roots of the first free African American community. *Scientific Reports*, 2020, vol. 10, no 1, p. 1-11. https://www.nature.com/articles/s41598-020-77608-8.epdf?sharing_token=rBB40PDPo-Af3https://www.nature.com/articles/s41598-020-77608-8.epdf?sharing_token=rBB40PDPo-Af3-jmKcwlz9RgN0jAjWeI9jnR3ZoTv0PhRuc0cMQoghVIMhILfDfR1dxNJs0cr0vSJxUXucuVu2fsocEWujIgxhj0wdr7pFL3U0YLQ7jHmujhnwYwtyIkivwgSnQItOauPJK2kppSKpZmoTOwzpfVPq5htqNQW5A%3DjmKcwlz9RgN0jAjWeI9jnR3ZoTv0PhRuc0cMQoghVIMhILfDfR1dxNJs0cr0vSJxUXucuVu2fsocEWujIgxhj0wdr7pFL3U0YLQ7jHmujhnwYwtyIkivwgSnQItOauPJK2kppSKpZmoTOwzpfVPq5htqNQW5A%3D

Anexo C. *(extendido en anexo K)*

Beatriz Martínez, Andy Conley, Shashwat Nagar, Lavanya Rishishwar, Javier Marrugo, King Jordan. Genetic ancestry, language, and African origins of San Basilio de Palenque. Sometido en *Communication Biology*, julio 2021.

Anexo D

OSSA, Humberto, et al. Patterns of genetic diversity in Colombia for 38 indels used in human identification. *Forensic Science International: Genetics*, 2021, vol. 53, p. 102495. https://www.sciencedirect.com/science/article/pii/S187249732100034X?casa_token=uyNwx0x5E6kAAAAA:pE5RKQMsD8RdKEEdUsVxz_UOsdC7fQc5dQwG0vFQfS8zlUdIQ4ogAr-2fgCsmUO1ulP4xnxMAA

Anexo E

MARTÍNEZ, B., et al. Ancestry estimates in afrodescendant population from San Basilio de Palenque, Colombia. *Forensic Science International: Genetics Supplement Series*, 2017, vol. 6, p. e224-e225. https://www.sciencedirect.com/science/article/pii/S1875176817301865?casa_token=WT0https://www.sciencedirect.com/science/article/pii/S1875176817301865?casa_token=WT0-Bb5pAc0AAAAA:tVA153X4Oyo7tMX0ys7t8iDYc_ARUWblBTrX

[mvM4ZWO_o3yusH4fmr8Fdw1ZteiVW2yoV0rbzABb5pAc0AAAAA:tVA153X4Oyo7tMX0ys7t8iDYc ARUWbIBTrXmvM4ZWO_o3yusH4f mr8Fdw1ZteiVW2yoV0rbzA](https://www.fsigeneticssup.com/article/S1875-1768(13)00038-3/fulltext)

Anexo F

MARTÍNEZ, B., et al. Population genetic data of 38 autosomal InDels in San Basilio de Palenque, the first free town in America. *Forensic Science International: Genetics Supplement Series*, 2013, vol. 4, no 1, p. e73-e74.

[https://www.fsigeneticssup.com/article/S1875-1768\(13\)00038-3/fulltext](https://www.fsigeneticssup.com/article/S1875-1768(13)00038-3/fulltext)

Anexo G

B. Martínez, Gómes V, Carvalho EF, Marrugo J, Gusmao L. The paternal legacy of San Basilio de Palenque, Colombia. *Memorias del Simposio de la ISFG: 11th Haploid markers conference 2018, Polonia*. <https://haploidmarkers2018.umk.pl/>

Anexo H. Otros trabajos realizados dentro del marco del proyecto

B. Martinez, M. Nguidi, L. Catelli, C. Vullo, V.O. Okolie, S.O. Keshinro, E.F. Carvalho, L. Gusmão, F. Simão. Mitochondrial genetic profile of the Yoruba population from Nigeria. *Forensic Science International: Genetics Supplement Series* 2019 https://www.sciencedirect.com/science/article/pii/S1875176819301489?casa_token=gEKvhyAzVb8AAAAA:1nYoYgBqN7ykmytb0sltrtG3w5cFWmMPRLTWTF12GzgpGL1ecu9IJYq0l ad5es6lkyFP0Z9LTA

Anexo I. Otros trabajos realizados dentro del marco del proyecto

B. Martinez, L. Catelli, M. Romero, V.O. Okolie, S.O. Keshinro, E.F. Carvalho, C. Vullo, L.

Gusmão. Forensic evaluation of 27 y-str haplotypes in a population sample from Nigeria. *Forensic Science International: Genetics Supplement Series* 6 (2017) e289–e291

https://www.sciencedirect.com/science/article/pii/S1875176817301853?casa_token=6ZnNqZ1R3c0AAAAA:rqtg0oW64xpee2eWPBuzxv0e2Rfscp9oJo7zoZdKYGmU8fToV_eihuZcNAi eP-hByXOK2NEMg

Anexo J. Otros trabajos realizados dentro del marco del proyecto

A. Campos, B. Martinez, L. Catelli, M. Rotondo, V.O. Okolie, S.O. Keshinro, E.F. Carvalho, C. Vullo, L. Gusmão, V. Gomes. Genetic characterization of the male lineages present in Nigeria population. 2018, Portugalíea Genética memorias, Porto. <https://repositorio-aberto.up.pt/handle/10216/119120>

Anexo K. Genetic ancestry, language, and African origins of San Basilio de Palenque: the first free town in the Americas

Beatriz Martínez^{1,*}, Andrew B. Conley^{2,3*}, Lavanya Rishishwar^{2,3}, Shashwat Deepali Nagar^{3,4}, Leonor Gusmão⁵, Michelle Daya⁶, Christopher R. Gignoux⁶, Simon Gravel^{7,8}, Timothy D. O'Connor^{9,10,11}, Leonardo Mariño-Ramírez^{3,12}, Augusto E. Valderrama Aguirre^{3,13}, Miguel A. Medina-Rivas^{3,14}, Rasika A. Mathias¹⁵, Kathleen C. Barnes on behalf of CAAPA^{16,17}, Javier Marrugol^{1,**}, and I. King Jordan^{2,3,4,**}

*co-first authors

**co-corresponding authors

- 1) Molecular Genetics Lab, Institute for Immunological Research, University of Cartagena, Cartagena, Colombia
- 2) IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, Georgia, USA
- 3) PanAmerican Bioinformatics Institute, Cali, Valle del Cauca, Colombia
- 4) School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA
- 5) DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), 20550-900 Rio de Janeiro, Brazil
- 6) Department of Medicine, University of Colorado Denver, Aurora, CO, 80045, USA
- 7) McGill Genome Centre, McGill University, Montreal, Canadá
- 8) Department of Human Genetics, McGill University, Montreal, Canadá
- 9) Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore MD, 21201 USA
- 10) Program in Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore MD, 21201 USA
- 11) Department of Medicine, University of Maryland School of Medicine, Baltimore MD, 21201 USA

- 12) National Institute of Minority Health and Health Disparities, National Institutes of Health, Bethesda, Maryland, USA
- 13) Department of Biological Sciences, Universidad de los Andes, Bogotá Colombia
- 14) Centro de Investigación en Biodiversidad y Hábitat, Universidad Tecnológica del Chocó, Quibdó, Colombia
- 15) GeneSTAR Research Center, Division of General Internal Medicine, School of Medicine, Johns Hopkins University, Baltimore, Maryland
- 16) Department of Medicine, Anschutz Medical Campus, University of Colorado, Aurora, CO 80045, USA
- 17) Colorado Center for Personalized Medicine, School of Medicine, University of Colorado, Aurora, Colorado

Abstract

Established as a refuge by enslaved Africans more than 400 years ago, the Colombian town of San Basilio de Palenque has maintained distinctly African culture and traditions, including the African-Spanish creole language Palenquero. Linguistic studies demonstrating a single African substrate for Palenquero, the Bantu language Kikongo, have led to the single African genetic origin hypothesis for the people of Palenque. We performed an integrated analysis of whole genome sequences, ancestry informative markers, and uniparental markers on a sample of 371 individuals to interrogate the genetic origins of Palenque. Fine-scale patterns of African genetic ancestry were considered in the context of African language groups in an effort to resolve the relationship between Palenque's genetic and linguistic heritage. The population of Palenque has predominantly African continental genetic ancestry (80%), followed by equal parts European (10%) and Native American (10%) admixture. Palenque shows a pattern of highly male-biased African ancestry not seen for any other Afro-descendant population in the Americas. Despite its isolation and small population, Palenque does not show any evidence of reduced genetic diversity owing to population size or consanguinity. Most notably, Palenque has a diverse, heterogeneous African genetic ancestry, with contributions from numerous source populations covering several broad geographic regions and many distinct Niger-Congo language groups. The diverse African genetic origins of Palenque are clearly distinct from the uniform linguistic origins of Palenquero. We posit that Kikongo was

established early as a lingua franca in Palenque and was later adopted by all members of the population, irrespective of their African origins.

Introduction

The Spanish word *cimarronaje* refers to the sustained resistance to slavery by Africans and their descendants in the New World¹. Widespread *cimarronaje* during the colonial period led to the formation of autonomous communities of enslaved Africans and their descendants (*maroons*) throughout the Americas². Maroon communities took refuge in fortified villages known as *palenques*, which were strategically located in inhospitable and hard to reach regions. More than twenty *palenques* were established in the Caribbean coastal region of Colombia during the 16th and 17th centuries³. The town of San Basilio de Palenque, referred to locally as Palenque, is the only such community that exists to this day, more than 400 years after it was established^{4,5}.

Palenque is located in the Montes de Maria region, just over 40 miles southeast of the port city of Cartagena in Colombia's Bolivar department. Oral history holds that Palenque was established by a group of 37 enslaved Africans from Cartagena in 1603, led by Benkos Biohó who was born to a royal family in the African kingdom of Kongo⁶. Biohó went on to lead a rebellion against Cartagena, and notwithstanding a peace agreement signed with the Spanish governor in 1612, he was ultimately recaptured and executed in 1621⁷. Despite the loss of their leader, the inhabitants of Palenque continued to resist colonial authorities, disrupting commerce and undermining the slave economy through raids on nearby haciendas. Unable to put down the resistance and facing pressure to resolve the situation from wealthy landowners, the Spanish crown was forced to grant Palenque full independence in 1713. Palenque is accordingly recognized as the "first free town in the Americas"⁸.

Independent and physically isolated, Palenque maintained distinctly African cultural traditions over the centuries, including aspects of its gastronomy, music and dance, and language. To this day, the residents of Palenque speak a creole language known as *Palenquero*, which combines Spanish and African linguistic elements^{6,9}. *Palenquero* has been documented to contain more than 200 words of African origin, and it is the only Spanish-based creole in Latin America^{10,11}. Linguists have characterized a single African substrate for *Palenquero* – the Kikongo language, which is a member of the Bantu language

family spoken in the Republic of the Congo, the Democratic Republic of Congo, and Angola^{6,9,12}. The Yombe language, from the PointeNoire area of the Republic of the Congo, is considered to be the specific source of Palenquero^{13,14}.

The finding of a single African substrate for the Palenquero language, considered together with the legacy of Benkos Biohó, was taken to suggest the possibility of a single Bantu origin for the people of Palenque (Palenqueros). It was posited that “The Palenqueros were one type of people when they ran away, coming from a single place, having a single language”¹⁵. Analysis of uniparental Y-DNA markers confirmed that the Yombe people, among 41 other African tribes, were the most likely ancestral group for Palenque¹⁶. Nevertheless, the single African substrate theory for the Palenquero language, and its corollary of a single African (Bantu) genetic origin for Palenque, are difficult to reconcile with the historical record of the transatlantic slave trade.

Africans were brought to Colombia from many different regions, including West Africa (Senegambia), West Central Africa (Bight of Benín), and Southwest Africa (Loango Coast) (Supplementary Figure 1)^{17,18}. Indeed, more than 70 distinct African languages were spoken in Cartagena in the 16th and 17th centuries¹⁹, and palenque communities were known to be ethnically diverse, populated by maroons born in various parts of Africa and the Americas³. It is reasonable to expect that the Palenqueros, who originally escaped from Cartagena and haciendas in the surrounding region, would represent a heterogeneous mix of African genetic ancestry. Indeed, a recent study of uniparental markers suggested genetic contributions to Palenque from widely different African regions²⁰.

Previous studies on the genetic ancestry of Palenque have focused almost exclusively on uniparental markers, or relatively small numbers of autosomal markers (haplotypes), affording limited resolution on the African ancestry of Palenque^{16,20-25}. In particular, these data were not sufficient to support a thorough evaluation of the single Bantu origin hypothesis for Palenqueros. For this study, we performed integrated analysis of uniparental markers, ancestry informative markers (AIMs), and whole genome sequences (WGS) in an effort to more deeply interrogate the African genetic origins of Palenque. Our approach is powered by genome-wide sequence comparisons between Palenqueros and reference populations covering the three main African regions that participated in the transatlantic slave trade to Colombia.

We considered genetic ancestry in the context of African language groups in an effort to resolve the relationship between linguistic heritage and genetic diversity in Palenque, and we compared results on the genetic ancestry of Palenque to the patterns seen for other Afro-descendant populations in the Americas, with an emphasis on the population of Chocó from the Pacific Coast of Colombia.

Materials and Methods

Colombian sample donors

DNA sample donors from Palenque, Colombia were enrolled by the Institute of Immunological Research of the University of Cartagena. All donors provided written informed consent for the use of their DNA in genetic studies. Donors identified as having both parents and all four grandparents born in Palenque. Sample collection and informed consent were conducted under the guidance of Regina Miranda and the Indians of the Zenues guards (Palenque, Colombia). Whole genome sequencing of 34 Palenque samples to 30x coverage was performed on the Illumina HiSeq 2000, as part of the Consortium on Asthma among African Ancestry Populations in the Americas (CAAPA) project (<https://www.caapa-project.org/>), and variants were called as previously described²⁶. CAAPA WGS data are publicly available at the NCBI dbGaP database (phs001123.v2.p1). Ancestry informative markers (AIMs) were characterized for a separate sample of 187 Palenque donors as described previously^{22,23,27}.

Sample donors from Chocó, Colombia were enrolled by the Universidad Tecnológica del Chocó as part of the ChocoGen research project (<https://www.chocogen.com/>)²⁸. ChocoGen donors were selected and their samples were characterized using genome-wide SNP arrays as previously described²⁹. All donors signed informed consent documents indicating their understanding of the potential risks of the project, along with how their data would be handled and how their identity would be protected. ChocoGen genome-wide genotype data are made available for research purposes by the ChocoGen Data Access Committee.

Institutional review board approval for the collection, genotyping, and comparative analyses of human DNA samples was obtained from the Universidad de Cartagena (Colombia), the Universidad Tecnológica del Chocó (Colombia), and the Georgia Institute of Technology (United States).

Global reference samples

WGS, genome-wide genotypes (SNP array data), and AIMs from African, European, and Native American reference populations were collected from publicly available sources, including the 1000 Genomes Project (1KGP)³⁰, the Human Genome Diversity Project³¹ (HGDP), the PAGE program³², and the Simons Genome Diversity Project³³ (Supplementary Figure 2). WGS variant calls for Admixed American and Afro-descendant populations from the Americas were taken from the 1KGP. Additional genome-wide genotype reference samples were taken from large-scale studies on African ancestry³⁴ and Native American ancestry³⁵. Afro-Colombian, Afro-descendant, and admixed American samples were merged with global reference populations as described in the following section. 584 reference samples from 40 populations from the HGDP were characterized for the AIMs analysis. A collection of 2,328 mtDNA haplotypes and 2,649 Y-DNA haplotypes from 42 and 38 African reference populations, respectively, were taken from a number of previous studies^{16,36-38}. Details on all of the reference populations analyzed here are provided in Supplementary Table 1.

SNP quality control and harmonization

WGS and genome-wide genotypes collected from different sources were harmonized using PLINK version 1.9³⁹ together with a custom script. Genomic variants were merged by identifying variants common to all datasets, while correcting for strand flips and inconsistencies with variant identifiers. Merged variants with >1% missingness and <1% minor allele frequency among samples were removed. WGS and genome-wide genotype data were merged into two separate variant sets to avoid the loss of genomic sites for subsequent analysis: (1) an African ancestry variant set (37 populations and 1,235 samples) and (2) a Native American ancestry set (33 populations and 314 samples). The two variant sets overlap extensively for WGS samples and differ mainly with respect to African³⁴ and Native American genome-wide genotype data³⁵. Both variant sets include all of the Afro-Colombian, other Afro-descendant, and other admixed American samples. The two harmonized variant data sets were phased using ShapeIT version 2.0⁴⁰, with individual chromosomes phased separately. The final phased variant data sets cover include 690,739 (African) and 364,470 (Native American) genomic sites, respectively. The harmonized data sets were linkage disequilibrium (LD) pruned using PLINK with the --

indep --pairwise 50 10 0.1 settings. The program KING version 2.2.4 was used to test for kinship among individuals from the harmonized datasets⁴¹. All but one representative individual was removed for each set of related individuals, yielding a final set of unrelated individuals. The final phased and pruned variant data sets include 275,598 (African) and 287,846 (Native American) genomic sites, respectively.

Genetic ancestry inference

Harmonized genome-wide variant data sets were used to calculate all-against-all pairwise allele sharing distances using PLINK with the --distance function, and principal component analyses (PCA) were performed using the PLINK --PCA function. ADMIXTURE version 1.3.0 was run on the LD pruned variant data in unsupervised mode with $K=3$ to infer continental ancestry fractions for Afro-descendant populations in the Americas⁴² using European, African, and Native American reference samples. A modified version of the program RFMix was used to infer local continental ancestry^{43,44}. RFMix was run for 12 generations using the PopPhased mode with a minimum node size of 5. Two rounds of expectation maximization (EM) were run with the “use-reference-panels-in-EM” option to generate local ancestry calls for Afrodescendant and reference population samples. PCA was re-run on RFMix masked data to perform African ancestry-specific PCA. Ancestry-specific haplotypes from the RFMix output were used to characterize sex-biased ancestry as previously described^{44,45}. The program Chromopainter version 2 was used to infer subcontinental ancestry admixed American genomes using ancestry-specific haplotypes compared to ancestry group-specific reference samples⁴⁶. Chromopainter was run on RFMix masked genomic data to characterize each continental ancestry component of admixed genomes – African, European, and Native American – separately. Non-negative least squares (NNLS) was used to convert Chromopainter output painting vectors to percent ancestry estimates using the R package nnls version 1.447. F3 tests were performed with the qp3pop program version 65048. Pairwise fixation index (F_{ST}) values between Native American populations were computed using RFMix determined Native American haplotypes with the EIGENSOFT package version 7.2.149. Population pairwise F_{ST} values were used to reconstruct a neighbor joining phylogeny using the program MEGA version 10.050,51.

Genetic diversity and demography

Levels of genetic diversity for Afro-descendant and reference samples were measured using Wright's inbreeding coefficient (FIS) and runs of homozygosity (FROH) calculated on the final African variant dataset, using PLINK version 1.9 with the --het and --homozyg functions, respectively. FIS is quantified as the normalized difference between the observed versus the expected counts of homozygous autosomal sites. FROH is quantified as the fraction of the genome in runs of homozygosity (ROH) >1.5kb. FIS was also computed for the insertion/deletion AIM data described in the following section.

The timing of admixture events was inferred from the distributions of RFMix defined ancestriespecific haplotypes using the program TRACTS version 0.352,53. All three possible orderings of African, European, and Native American admixture were considered, and for each ordering TRACTS was used to evaluate admixture timing from 14 to 6 generations ago. The most likely admixture scenario was chosen based on 1,000 bootstrap replicates for each of the three orderings.

Ancestry informative marker (AIM) analysis

Palenque donor samples and global reference population samples from the HGDP were characterized using a panel of 46 ancestry-informative insertion-deletion polymorphisms (indel AIMs) as previously described^{22,23,27}. Indel AIMs were scored as homozygous deletion (0 copies), heterozygous indel (1 copy), or homozygous insertion (2 copies). Continental ancestry fractions for Palenque samples were inferred using ADMIXTURE analysis of indel AIMs with HGDP reference samples from Africa, the Americas and Europe (Supplementary Table 1). PCA analysis of indel AIMs was used to analyze the relationship between Palenque samples and global reference populations from the HGDP. Levels of genetic diversity for Palenque and global reference populations were measured using FIS calculated for AIMs.

Uniparental marker analysis

Uniparental haplotypes were characterized from WGS variant calls using the programs HaploGrep 254 (mtDNA) and SNAPPY (Y-DNA)⁵⁵. Variant calls were pre-processed for reference/alternate allele strand flipping, as needed, prior to haplotype characterization. Continental ancestry group origins – African, European, and Native American – for WGS characterized mtDNA and Y-DNA haplotypes

were characterized via comparison with 1KGP and HGDP reference samples, as previously described⁵⁶. An additional set of previously characterized 151 mtDNA and 150 Y-DNA haplotypes¹⁶, together with haplotypes from African reference populations, were re-analyzed via comparison with African reference populations³⁶⁻³⁸. Continental population group origins were characterized for both mtDNA and Y-DNA haplotypes for this second data set, but only the Y-DNA haplotypes from this set provided sufficient resolution for subcontinental African ancestry analysis. Y-DNA haplotype African subcontinental origins were characterized by measuring the fraction of Bantu and nonBantu African haplotypes and by comparing African Y-DNA haplogroup frequencies among populations using pairwise F_{ST} values.

Results

Continental ancestry

We characterized the continental ancestry patterns for Palenque based on the analysis of WGS (n=34) and AIMs (n=187). WGS from Palenque and Chocó (Figure 1A), along with other Afrodescendant populations, were compared to reference populations from Africa, the Americas, and Europe (Supplementary Table 1 and Supplementary Figure 2). Principal component analysis (PCA) projection of pairwise genomic distances shows that Palenque and Chocó group closely to the African pole of global genetic diversity with evidence of both European and Native American admixture (Figure 1B). African American (United States) and African Caribbean (Barbados) populations show a somewhat distinct pattern, with similar proximity to the African pole of the PCA, but extending mainly towards the European pole with little evidence of Native American ancestry. ADMIXTURE analysis confirms the distinction between the two Afro-Colombian populations and the African American and African Caribbean populations (Figure 1C). All four Afro-descendant populations show similar levels of African ancestry, with the non-African ancestry component distinguished by approximately even parts of European and Native admixture for the Colombian populations, compared to primarily European ancestry for the non-African component of the African American and African Caribbean populations (Figures 1D & E). Continental ancestry estimates from local ancestry analysis performed with the program RFMix are almost identical to those seen for the ADMIXTURE analysis (Supplementary Figure 3).

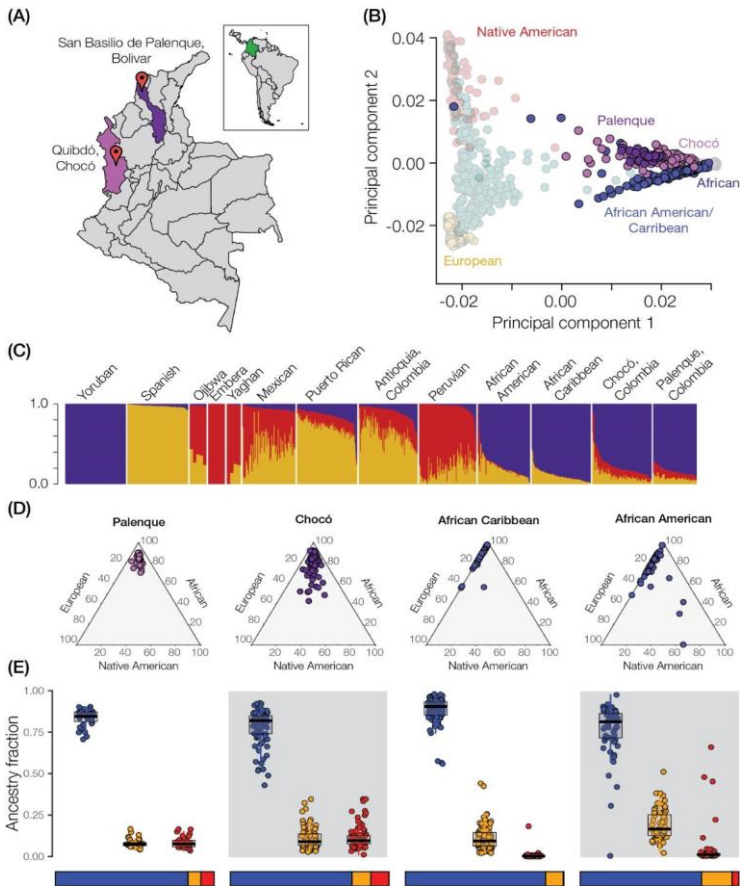


Figure 1. Continental ancestry and admixture. (A) Locations of Afro-Colombian populations sampled from San Basilio de Palenque and Quibdó, Chocó. (B) PCA showing genetic relationships among admixed American populations together with African, European, Native American reference populations. Afro-descendant populations from Colombia, the Caribbean (Barbados), and the United States are highlighted. (C) ADMIXTURE plot showing continental ancestry fractions – African (blue), European (yellow), and Native American (red) – for continental reference populations and admixed American populations. (D) Ternary plots showing the relative continental ancestry contributions for individuals from the four Afro-descendant populations. (E) Continental ancestry distributions and population averages for the four Afro-descendant populations.

Continental ancestry analysis of Palenque via AIMs yielded similar levels of African, European, and Native American ancestry (Table 1 and Supplementary Figure 4). Palenque shows a high level of African ancestry (WGS=84%, AIMs=76%) compared to other Afro-descendant populations in the Americas but not exceptionally so. The cosmopolitan Afro-descendant populations of Barbados, Jamaica, and the United States show similar or higher levels of African ancestry compared to Palenque, and the Noir Maroon populations of French Guiana and Suriname, which were also established by enslaved Africans and experienced long periods of subsequent isolation, show far higher levels of African ancestry with little or no European and Native American admixture^{26,30,57}.

Table 1. Continental ancestry percentages for Afro-descendant populations in the Americas. Standard error values are shown in parentheses for the Palenque population analyzed here.

Population	Source ^a	African %	European %	Native American %
Colombian Afro-descendant populations				
Palenque (WGS)	This study	83.5 (0.9)	8.1 (0.5)	8.4 (0.6)
Palenque (AIMs)	This study	76.0 (0.9)	11.5 (0.8)	12.5 (0.8)
Chocó	28855283	77	12	11
Other Afro-descendant populations				
Noir Maroon	29100086	97	2	1
Jamaica	27725671	89	11	1
African Caribbean	26432245	88	11	1
African American	26432245	77	19	4
Garifuna	27725671	81	2	17
Brazil (Rio de Janeiro)	29100086	71	23	5
Brazil (Conde)	27725671	49	41	10
Dominican Republic	27725671	38	52	9

aPubMed identifier (PMID) for the studies from which the ancestry estimates were obtained.

Sex-biased admixture was analyzed via the analysis of uniparental mtDNA (female) and YDNA (male) markers along with comparisons of X-chromosome versus autosomal ancestry levels. A relative excess of ancestry from any of the three ancestry groups for X-chromosomes indicates female-biased ancestry, whereas excess autosomal ancestry

indicates male-biased ancestry. Both the female (mtDNA, 88%) and male (Y-DNA, 72%) uniparental haplogroups from Palenque are predominantly African, consistent with the results seen for autosomal DNA ancestry (Figure 2A). There are 12% Native American mtDNA haplogroups and no European mtDNA haplogroups seen for Palenque, indicative of substantial female-biased admixture. The non-African component of the male lineage in Palenque shows an even more pronounced pattern of sex-biased admixture with 28% European Y-DNA haplogroups compared to no Native American Y haplogroups. Re-analysis of a previously published set of mtDNA and YDNA haplogroups from Palenque yielded similar results, with almost entirely African ancestry along the female lineage (mtDNA) and a bias towards European male ancestry for the nonAfrican Y-DNA haplogroups (Supplementary Figure 5). Comparison of X-chromosome versus autosomal ancestry also supports Native American-biased female ancestry and European-biased male ancestry in Palenque (Figure 2B). However, by far the strongest signal from this analysis is seen for male-biased African ancestry in Palenque. This pattern is distinct from what is observed for the Chocó, African American, and African Caribbean populations, all of which show a combination of male-biased European ancestry and female-biased African ancestry. TRACTS analysis shows relatively smaller European and Native American haplotypes, compared to African haplotypes, pointing to early admixture between European males and Indigenous females followed by later admixture with Afro-descendants in Palenque (Figure 2C and Supplementary Figure 6).

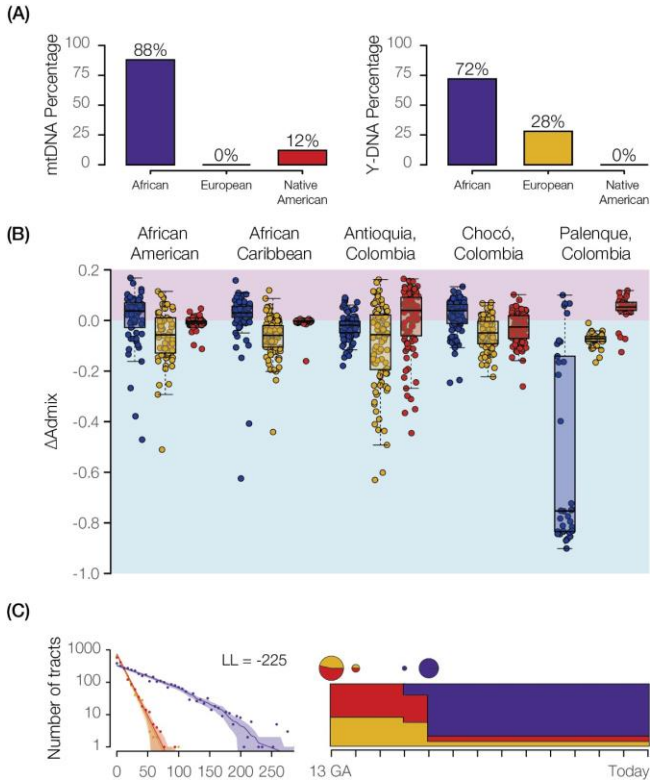


Figure 2. Sex-biased ancestry and admixture timing. (A) Percentage of individuals from Palenque with African, European, or Native American uniparental haplotypes: mtDNA (female) and Y-DNA (male). (B) Population distributions of ΔAdmix values, which measure differences in ancestry between the X chromosome and the autosomes, for each continental ancestry: African (blue), European (yellow), and Native American (red). Values above zero indicate female-biased admixture, and values below zero indicate male-biased admixture. (C) Ancestry-specific haplotype (tract) size distributions (left panel) and estimated admixture timings (right panel) for each of the three ancestral groups. Left: observed (points) and predicted (solid lines) ancestry tract size distributions with shaded 95% confidence intervals. Right: admixture event timings in generations ago and admixture proportions. Each admixture event is indicated by a circle scaled according to the size of the founding population and the relative ancestry proportion.

African genetic ancestry and language groups

We characterized subcontinental African genetic ancestry in the context of the geographic distribution of African language groups in an effort to understand the relationship between genetic ancestry and the creole Palenquero language spoken in Palenque. First, we characterized the fine-scale genetic ancestry of 35 African populations coming from the three main regions that participated in the transatlantic slave trade to Colombia: Senegambia (West Africa), the Bight of Benín (West Central Africa), and the Loango Coast (Southwest Africa) (Figure 3A). Additional Rain Forest Hunter Gatherer (RFHG) and East Bantu populations were included for comparison. The major axis of genetic diversity revealed by PCA divides the Senegambia and the Bight of Benín population groups from the Loango Coast group, which corresponds to the distinction between non-Bantu and Bantu African language groups, respectively (Figure 3B). The results of Chromopainter analysis show largely coherent patterns of genetic ancestry within the non-Bantu populations from Senegambia and the Bight of Benín; whereas the Loango Coast populations show substantial population structure with five distinct Bantu sub-groups (Figure 3C and Supplementary Figure 7). The RFHG and East Bantu groups show their own distinct patterns of genetic diversity. There are two particularly admixed populations: the Yaounde show a mix of non-Bantu and Bantu ancestry, consistent with their geographical location in Cameroon between the Bight of Benín and the Loango Coast, and the Babongo show a mix of RFHG and Bantu ancestry (Supplementary Figure 7). The pattern of admixture seen for Babongo is consistent the fact that they are considered as a RFHG people (Mbenga Pygmies); although, they are not particularly short and speak a mix of several different Bantu languages. Subcontinental African ancestry analysis with ADMIXTURE yields very similar patterns of genetic diversity and population structure (Supplementary Figure 8).

Next, we considered the observed patterns of African subcontinental ancestry to the distribution of language groups using the Glottolog catalog of world languages (<https://glottolog.org/>).

Language names for the populations studies here and their corresponding ‘Glottocodes’, which can be used to place them in the context of related languages in the Glottolog, are shown in Supplementary Table 1. Native languages from the areas of Africa considered here all fall into the Niger-Congo language group, Africa’s largest language group in terms of geographical area as well as the

number of speakers and languages. The observed fine-scale ancestry patterns for the African populations studied here closely follow the distribution of Niger-Congo languages (Figure 3D). The Mandinka, Mende, and Yacouba populations from Senegambia all speak Mande, a basal language branch of the Niger-Congo family, whereas the Ahizi population speaks the Volta-Congo language Kru. The Bight of Benín populations Bariba and Fon, from Benín, speak North Volta-Kongo and Kwa Volta-Kongo languages, respectively, whereas the Yoruba and Esan populations from Nigeria speak the Benue-Congo languages Defoid and Akpes-Edoid, respectively. All of the Bantu populations and the single RFGH population all speak languages from the Narrow Bantu group of Bantoid languages. The Narrow Bantu languages can be further sub-divided into geographic zones, the distributions of which are also largely concordant with the African population genetic structure observed here (Supplementary Table 1 and Supplementary Figure 9).

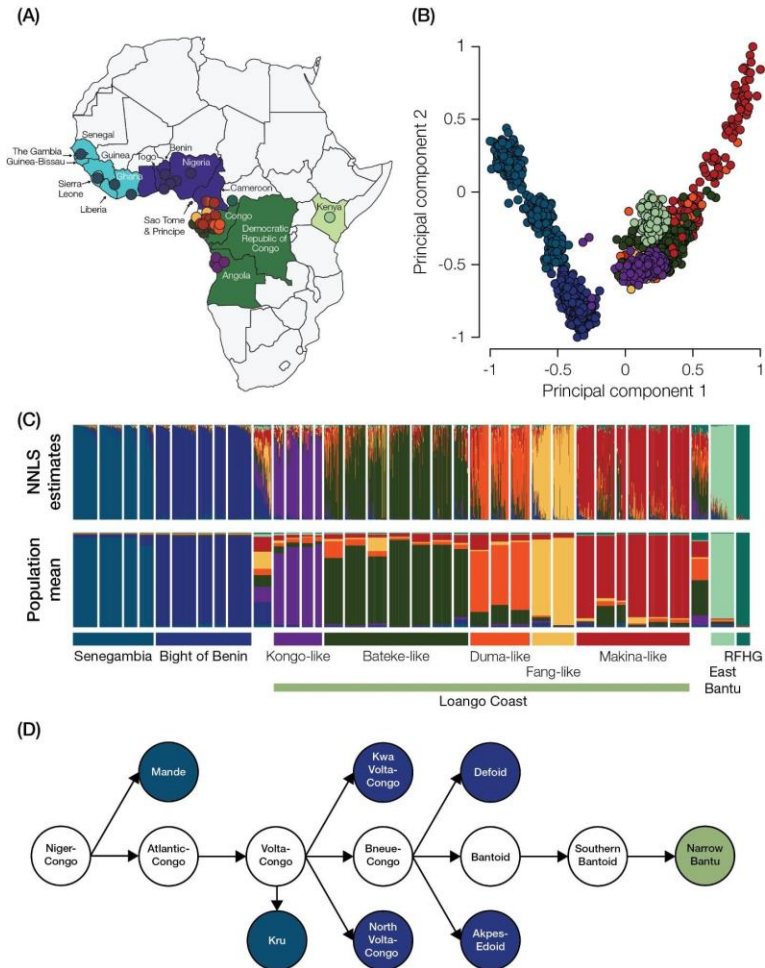


Figure 3. African genetic ancestry and language groups. (A) Locations of African reference populations among the three main sampling regions: West African (Senegambia, light blue), West Central Africa (Bight of Benin, dark blue), and Southwest Africa (Loango Coast, green). (B) PCA showing genetic relationships among African reference populations (colored as shown in panel A). (C) Chromopainter analysis showing genetic structure among African reference populations, grouped according to broad geographic regions and population groups. Individual level NNLS ancestry estimates are shown along with population average estimates. A more detailed version of this plot is shown in Supplementary Figure 7. (D) Graph relating the language groups of the African reference populations, color-coded according to the three broad geographic regions.

African genetic and linguistic origins of Palenque

The creole Palenquero language of Palenque was previously found to have a single African substrate – the Yombe language (Glottocode: yomb1244) from the Pointe-Noire region in the Republic of the Congo 13,14. Yombe is a Kikongo language, which has been classified into the geographic zone H of the Narrow Bantu languages (Supplementary Figure 9). We characterized the African subcontinental origins of individuals from Palenque, via fine-scale analysis of WGS data from Palenque donors together with African reference samples using Chromopainter, to evaluate whether the single African substrate hypothesis for Palenquero also supports a single (or narrow) African origin for the people of Palenque, as has been previously suggested¹⁶. While individuals from Palenque show the highest levels of Bantu (Loango Coast) ancestry, it is also clear that they have substantial contributions from all three of the African regions that participated in the transatlantic slave trade to Colombia (Figure 4A). Palenque shows 51% Bantu (Loango Coast) ancestry, followed by 35% Bight of Benín ancestry, and 14% Senegambia ancestry (Table 2). Palenque also shows substantial genetic diversity within the Bantu genetic ancestry component (Figure 4B). The Yombe language corresponds to the Kongo-like population group analyzed here, which makes up 25% of Palenque ancestry, while the remaining Bantu population groups make up 26% of Palenque ancestry. Analysis of the African genetic origins of Palenque using ADMIXTURE and the f3 test show very similar results, with diverse ancestral contributions from a variety of non-Bantu and Bantu populations (Supplementary Figures 8 and 10).

Oral tradition holds that Palenque was established by a group of escaped slaves led by Benkos Biohó, who was born in the African kingdom of Kongo⁶. Considered together with the single Yombe (Kikongo) language substrate for Palenquero, it could be possible that the early founders of Palenque had primarily Bantu origins. Subsequent arrivals, escaping or taken from Cartagena and surrounding regions, may have had origins from other African populations. We compared the distributions of the sizes of African ancestry haplotypes from the Loango Coast (Bantu) compared to Senegambia and the Bight of Benín to test this hypothesis. If Bantu ancestry in Palenque pre-dates ancestry from other African populations, we would expect to see a shift towards smaller Bantu haplotypes, since haplotypes are progressively reduced in size over time owing to recombination. However, we did not observe any difference in the distributions of the sizes of African haplotypes

from the Loango Coast compared to those from Senegambia or the Bight of Benín (Supplementary Figure 11).

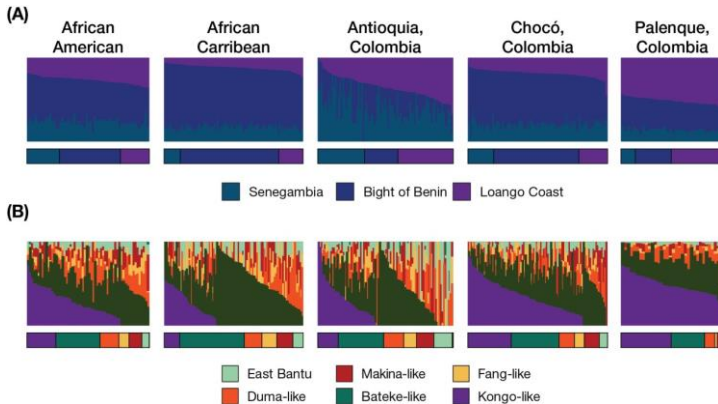


Figure 4. African genetic structure and ancestry for Afro-descendant populations in the Americas. (A) Individual level and population average NLS ancestry estimates for the three main African sampling regions: Senegambia (light blue), Bight of Benín (dark blue), Loango Coast (purple). (B) Individual level and population average NLS ancestry estimates for the Loango Coast (Bantu) component of African genetic ancestry. Specific Bantu population groups (see Figure 3C) are color coded as shown.

Table 2. African ancestry percentages for Afro-descendant populations in the Americas.

Population	West Africa (Senegambia) %	West Africa (Bight of Benín) %	Central Africa (Loango Coast) %
<i>Colombian Afro-descendant populations</i>			
Palenque	14	35	51
Chocó	21	61	18
African Caribbean	18	71	12
African American	24	50	26

As described in the Methods, the African fine-scale ancestry analysis was done using an African ancestry variant set to avoid the loss of genomic sites while maximizing the number of African reference populations. This ancestry-specific variant set overlaps extensively with the second Native American ancestry variant set constructed the same way, with Native American genotypes harmonized to a background of global reference samples characterized via WGS. The overall consistency between these two variant datasets is supported by

the very high correlations seen for African, European, and Native American ancestry estimates calculated using both sets (Supplementary Figure 12).

We also performed analysis of uniparental (male) Y-DNA haplotype markers to interrogate the African genetic and linguistic origins of Palenque, via comparison with 42 African reference populations and their corresponding language groups (Supplementary Table 1). Males from Palenque show the most diverse Y-DNA haplotypes for any of these populations, owing to the presence of a minority of non-African Y-DNA haplotypes along with a mix of both Bantu and non-Bantu characteristic haplotypes (Supplementary Figure 13A). Comparison of the population frequencies of the ten most abundant Y-DNA haplotypes shows Palenque to be equally related to the Yombe population from the Congo and the Chewa population from Malawi (Supplementary Figure 13B). The Yombe and Chewa populations both speak Narrow Bantu languages, albeit ones that are quite distinct: the southwestern coastal zone H Yombe language (Glottocode: yomb1244) and the southeastern coast zone N Nyanga language (Glottocode: nyan1308). F_{ST} analysis shows the Fouban and Wum populations of Cameroon to be the most closely related to Palenque (Supplementary Figure 13C). These two populations speak the Bamoun (Glottocode: bamo1241) and Aghem (Glottocode: aghe1239) languages, respectively, from the Wide Grassfields branch of the Southern Bantoid group, a distinct sisterbranch to the Narrow Bantu (Supplementary Figure 14). The differences in the results obtained via comparison of haplotype frequencies compared to the F_{ST} analysis likely reflect the substantial African admixture seen for Palenque, as both methods assess different aspects of genetic diversity.

Considered together, our analysis of both dense autosomal markers and uniparental Y-DNA haplotypes indicate that the population of Palenque has ancestral contributions from a diverse group of African populations, corresponding to an equally diverse set of Niger-Congo language groups, in contrast to the single African substrate posited for the Palenquero language.

Genetic diversity and founder effects

Palenque was established by a small number of enslaved Africans and their descendants, and remained isolated with a small population for centuries, raising the possibility of founder effects owing to a genetic bottleneck. We compared the levels of genetic diversity in Palenque to

other Afro-descendant and other admixed Latin American populations to search for evidence of reduced genetic diversity. We analyzed both Wright's inbreeding coefficient (FIS) and runs of homozygosity (FROH) as measures of genetic diversity, or the depletion thereof. FIS measures a depletion of genetic diversity in the current generation, whereas FROH measures a depletion of diversity tens of generations ago. Palenque does not show elevated levels of either FIS or FROH compared to other admixed American populations (Figure 5). Palenque also showed among the lowest levels of FIS when AIMs data were used for analysis (Supplementary Figure 4C). Overall, the population of Palenque does not show evidence for any depletion of genetic diversity consistent with founder effects or consanguinity.

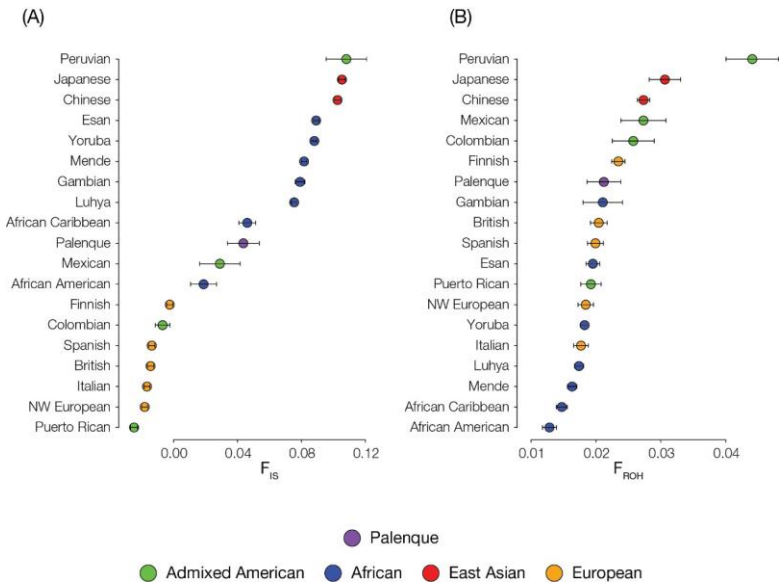


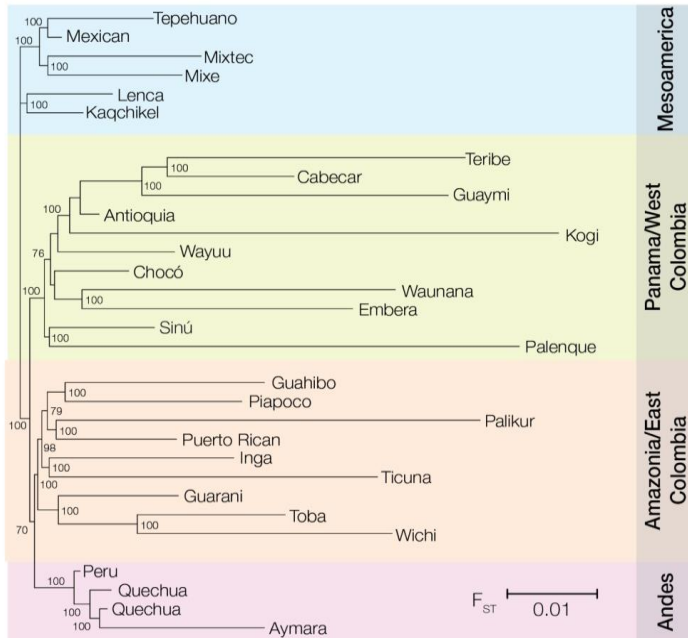
Figure 5. Genetic diversity. Mean $\pm 2sd$ values of (A) Wright's inbreeding coefficient (F_{IS}) and (B) runs of homozygosity (F_{ROH}) are shown for admixed American populations and global reference populations, color-coded as shown.

Native American and European ancestry

We characterized the subcontinental Native American ancestry components for Palenque using the Native American ancestry variant set. Native American ancestry for Palenque was compared to other modern Latin American populations from Colombia, Mexico, Peru, and

Puerto Rico using a panel of Indigenous American populations from Mesoamerica, including Mexico and Central America, and South America. Modern Latin American populations show patterns of Native American ancestry that are concordant with their geographic locations, consistent with prior results (Figure 6A)44,58-60. Mexico groups with Mesoamerican Indigenous populations, Puerto Rico groups with Amazonian populations, and Peru groups with Andean populations. All three Colombian populations studied here – Antioquia, Chocó, and Palenque – group most closely with Indigenous tribes from Colombia. Chocó is most closely related to the Embera and Waunana populations from the Pacific coast, as previously reported 29, and Palenque is most closely related to the Sinú population from the Atlantic coastal region. The genetic structure of the Native American component of the Colombian populations revealed by ADMIXTURE analysis shows a similar pattern (Figure 6B). There is little apparent difference among the Colombian populations at K=2 and K=4 ancestry components, but at K=8 Palenque shows a clear distinction with more Sinú ancestry and less Amazonian ancestry. Finescale analysis of Native American ancestry using the f3 test shows similar results, with modern Latin American populations grouping with Indigenous source populations according to geography and Palenque distinguished from the other Colombian populations by Sinú ancestry (Supplementary Figure 15). All of the modern Latin American populations studied here, including both Afro-Colombian populations Palenque and Chocó, show exclusively Spanish European ancestry, in contrast to the mainly British European ancestry seen for the African American and African Caribbean populations. While this result is not at all surprising given the history of the colonization of the Americas, it does serve as a positive control for our approach to subcontinental ancestry analysis.

(A)



(B)

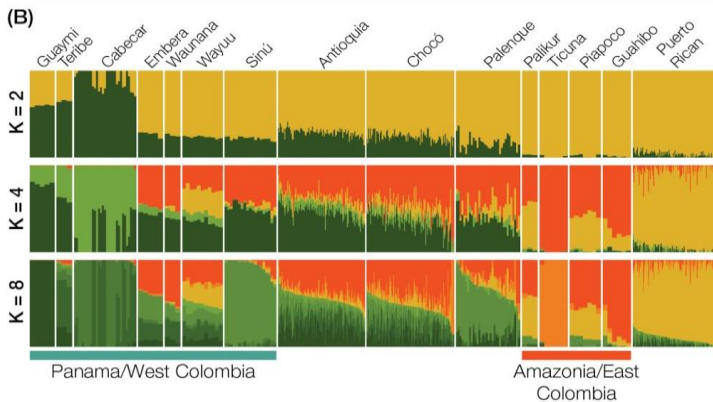


Figure 6. Native American ancestry. (A) Phylogeny comparing the Native American ancestry component for admixed Latin American populations to Native American reference populations. Distinct Native American reference population geographical groups are indicated. (B) ADMIXTURE analysis of admixed Latin American populations from Colombian and Puerto Rico compared to Colombian and Amazonian Native American reference populations. Results are shown for different numbers of estimated ancestry components: $K=2$, $K=4$, and $K=8$.

Discussion

For this study, we tested a number of hypotheses about the origins of San Basilio de Palenque, including whether Palenqueros (i) have more African ancestry than other Afro-descendant populations in the Americas and (ii) show evidence of admixture with European and Indigenous populations. Given the unique history of Palenque, particularly its isolation over the centuries, we hypothesized that Palenqueros would show more African and less European/Indigenous ancestry than seen for other Afro-descendant populations in the Americas. We also tested whether (iii) there are different levels of continental ancestry, i.e. sex-biased admixture, for the male and female lineages in Palenque. Given the social dynamics that characterized the conquest and colonization of the Americas, and if there is evidence of European and/or Indigenous admixture among Palenqueros, we expected to observe sex-biased patterns with an excess of European (Spanish) ancestry on the male lineage and an excess of Native American ancestry on the female lineage. With respect to demography, we asked whether (iv) Palenqueros show evidence of reduced genetic diversity owing to small effective population size or consanguinity. Finally, we were most interested to ask about (v) the sub-continental African origins of Palenqueros. The single African substrate theory for Palenquero, along with previous genetic ancestry results, support a single African origin for Palenqueros as well as a founder effects with reduced genetic diversity. However, the broader historical record of the transatlantic slave trade, and studies of other maroon communities suggests otherwise, and we hypothesized that Palenque should show a mix of genetic ancestry from different regions of Africa with high levels of genetic diversity and no evidence of founder effects or reduced genetic diversity.

We found that Palenque shows levels of African ancestry that are high, but by no means exceptionally so, compared to other Afro-descendant populations in the Americas (Table 1). Analysis of both dense autosomal markers and uniparental haplotypes confirms that Palenqueros show evidence of admixture with both European and Indigenous populations, similar to other Latin American populations. However, the patterns of sex-biased ancestry seen for Palenque are unique (Figure 2). Unlike all other Latin American populations studied to date, Palenque shows very strong African male-biased admixture and a far less pronounced pattern of European male-biased and Native American female-biased admixture. The African male-biased ancestry

distinguishes Palenque from Chocó in Colombia and Afro-descendant populations in the United States and the Caribbean, which all show female-biased African ancestry. Finally, there is an apparent difference in the pattern of sex-biased ancestry revealed by the analysis of uniparental markers compared to the pattern shown via comparison of ancestry on the autosomes versus the X-chromosome, with shorter European and Native American haplotypes compared to longer African haplotypes. Taken together with historical records and oral tradition in Palenque, these results can be considered to reflect the social dynamics of its establishment and subsequent isolation 3-6,8. There appears to have been an early pulse of European male-biased and Native American female-biased ancestry, in the colonial period prior to or around the time of the establishment of Palenque, followed by a much longer period of isolation marked by African male-biased ancestry. African male-biased ancestry following the establishment and isolation of Palenque is consistent with historical records relating how contact with nearby Indigenous villages continuously added to the female population of the community.

Despite its century long history of isolation, the population of Palenque does not show any evidence of a depletion of genetic diversity consistent with small effective population size or consanguinity (Figure 5 and Supplementary Figure 4C). This may be attributed in part to the “Kuadro” social system, wherein the village is divided into zones of family networks and individuals are discouraged from marrying within the same Kuadro. This helps to ensure that couples with cryptic genetic relatedness do not marry and reproduce.

Finally, multiple lines of evidence – WGS, AIMS, and uniparental markers analyzed with a variety of methods – all point to a diverse, heterogeneous African ancestry for Palenque. While Palenqueros do show a majority of Bantu African ancestry, they also have substantial ancestry contributions from Senegambia and the Bight of Benín (Table 2 and Figure 4). Furthermore, within the Bantu ancestry component, Palenque shows evidence of contributions from numerous groups beyond the single Kongo-like group that corresponds to the source of the African language substrate. Altogether, the population of Palenque shows ancestral affinities to African populations that speak a wide variety of languages from the diverse Niger-Congo language family. These results highlight a striking dissonance between the African genetic origins of Palenqueros and the cultural dominance of a single African language substrate for the creole Palenquero language.

The Kikongo language that forms the African substrate for Palenquero was spoken in the African kingdom of Kongo, which dominated much of southwestern Africa from the late 14th to the early 20th centuries⁶¹. By far, the largest number of enslaved Africans, close to 5.9 million, were taken to the Americas from this region (Loango Coast) during the transatlantic slave trade (<https://www.slavevoyages.org/>). Kikongo was, and remains to this day, the main base of a number of creole dialects spoken in adjacent African regions⁶². The influence of the Kikongo language in the region, considered together with the large number of individuals originating from this region, suggests the possibility that Kikongo may have served as a lingua franca for enslaved Africans and their descendants in the New World. Indeed, Kikongo contributes to a number of creole languages spoken by communities of enslaved Africans and their descendants in Brazil, Colombia, Cuba, the Dominican Republic, Haiti, Puerto Rico, and United States⁶³⁻⁶⁷. Once established as the dominant African language substrate by the founders of Palenque, later arrivals may have been compelled to adopt Palenquero, including both its African and Spanish strata, irrespective of their native African tongue.

Language and genetics often coevolve, as can clearly be seen for the Niger-Congo languages in Africa (Figure 3)⁶⁸. This does not seem to have been the case for Palenquero creole, where a single African language persisted in the face of the diverse African genetic origins of its speakers. The discordance between cultural and genetic evolution in Palenque can be taken as an example of the difference between memetic and genetic evolution⁶⁹. Memes evolve more rapidly than genes owing to both the rapid rate of change and the combination of horizontal and vertical evolution, as opposed to strictly vertical evolution for genes. The evolutionary dominance of linguistic memes may be particularly pronounced for creole languages formed via contact between previously isolated groups.

Lastly, a more prosaic explanation for our results could be that the African substrate for Palenquero is actually more complex than previously imagined, with contributions from multiple African languages, although years of detailed studies suggest that this is not the case⁹¹⁴. Nevertheless, our results could also be taken to motivate further studies and potentially new discoveries regarding the African substrate of Palenquero.

References

- Navarrete, M.C. El cimarroneo: una alternativa de libertad para los esclavos negros. *Historia Caribe* 2, 89-98 (2001).
- Price, R. *Maroon societies: Rebel slave communities in the Americas*, (Johns Hopkins University Press, Baltimore, MD, 1996).
- Navarrete, M.C. Palenques: Maroons and Castas in Colombia's Caribbean Regions. in *Orality, Identity, and Resistance in Palenque (Colombia): An interdisciplinary approach* (eds. Schwegler, A., Kirschen, B. & Maglia, G.) 265-293 (Johns Benjamins, Amsterdam, Netherlands, 2017).
- Friedemann, N. & Patiño Rosselli, C. *Lengua y sociedad en el Palenque de San Basilio*, (Instituto Caro y Cuervo, Bogotá, Colombia, 1983).
- Navarrete Peláez, M.C. *San Basilio de Palenque: memoria y tradición: surgimiento y avatares de las gestas cimarronas en el Caribe colombiano*, (Universidad del Valle, Cali, Colombia, 2008).
- Escalante, A. *El Palenque de San Basilio: una comunidad de descendientes de negros cimarrones*, (Editorial Mejoras, Barranquilla, Colombia, 1979).
- Tomaselli, W. *The colonial betrayal that haunts Colombian peace*. Vol. 2020 (Ozy, Mountain View, California, 2017).
- Caicedo, R.A. *Palenque, primer pueblo libre de América*, (Todo Impresores, Bogotá, Colombia, 1986).
- Granda Gutiérrez, G.d. Sobre la procedencia africana del habla criolla de San Basilio de Palenque (Bolívar, Colombia). *Thesaurus: Boletín del Instituto Caro y Cuervo* 26, 84-94 (1971). 10.
- Schwegler, A. The African Vocabulary of Palenque (Colombia). *Journal of Pidgin and Creole languages* 15, 241-312 (2000).
- Schwegler, A. El vocabulario africano de Palenque (Colombia): Segunda parte: compendio alfabético de palabras (con etimologías). in *Palenque, Cartagena y Afro-Caribe: historia y lengua* (eds. Moñino, Y. & Schwegler, A.) 171-226 (Max Niemeyer Verlag, Tubingen, Germany, 2002).
- Schwegler, A. Palenque (ro): the search for its African substrate. in *Creoles, their substrates, and language typology* (ed. Lefebvre, C.) 225-249 (Johns Benjamins, Amsterdam, Netherlands, 2011).
- Schwegler, A. Combining population genetics (DNA) with historical linguistics. in *Spanish language and sociolinguistic analysis*, Vol. 8 (eds. Sessarego, S. & Tejedo-Herrero, F.) 33-88 (Johns Benjamins, Amsterdam, Netherlands, 2016).

- Schwegler, A. On the African origin (s) of Palenquero. in *Orality, Identity, and Resistance in Palenque (Colombia): An interdisciplinary approach* (eds. Schwegler, A., Kirschen, B. & Maglia, G.) 51-119 (Johns Benjamins, Amsterdam, Netherlands, 2017).
- Schwegler, A. African roots in Latin America: Palenque (Colombia). Vol. 2020 (2017). 16. Ansari-Pour, N. et al. Palenque de San Basilio in Colombia: genetic data support an oral history of a paternal ancestry in Congo. *Proc Biol Sci* 283, 20152980 (2016).
- Rodriguez, J. Bantúes y otros Africanos en Colombia in *Velorios Y Santos Vivos. Comunidades Negras, Afrocolombianas, Raizales Y Palenqueras*, (Museo Nacional De Colombia, Bogotá, Colombia, 2008).
- Maya Restrepo, L.A. Brujería y reconstrucción de identidades entre los africanos y sus descendientes en la Nueva Granada Siglo XVII, (Ministerio de Cultura, Bogotá, Colombia, 2005).
- Stunt, V. Reviving Colombia's 'language of resistance'. Vol. 2020 (2019).
- Martinez, B. et al. Searching for the roots of the first free African American community. *Sci Rep* 10, 20634 (2020).
- Arnaiz-Villena, A. et al. The first free Africans in America: HLA study in San Basilio de Palenque (Colombia). *Hum Immunol* 79, 585-586 (2018).
- Martínez, B. et al. Population genetic data of 38 autosomal InDels in San Basilio de Palenque, the first free town in America. *Forensic Science International: Genetics Supplement Series* 4, e73-e74 (2013).
- Martínez, B. et al. Ancestry estimates in afrodescendant population from San Basilio de Palenque, Colombia. *Forensic Science International: Genetics Supplement Series* 6, e224-e225 (2017).
- Noguera, M.C. et al. Colombia's racial crucible: Y chromosome evidence from six admixed communities in the Department of Bolivar. *Ann Hum Biol* 41, 453-9 (2014).
- Solé-Llussà, A. et al. Exploring the African genetic influence in the first settlement founded by African slaves in America. *Anthropological Science* 123, 121-128 (2015).
- Mathias, R.A. et al. A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. *Nat Commun* 7, 12522 (2016).

- Pereira, R. et al. Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS One* 7, e29684 (2012).
- Medina-Rivas, M.A. et al. Choco, Colombia: a hotspot of human biodiversity. *Rev Biodivers Neotrop* 6, 45-54 (2016).
- Conley, A.B. et al. A Comparative Analysis of Genetic Ancestry and Admixture in the Colombian Populations of Choco and Medellin. *G3 (Bethesda)* 7, 3435-3447 (2017).
- Genomes Project, C. et al. A global reference for human genetic variation. *Nature* 526, 68-74 (2015).
- Li, J.Z. et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319, 1100-4 (2008).
- Wojcik, G.L. et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature* 570, 514-518 (2019).
- Mallick, S. et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* 538, 201-206 (2016).
- Patin, E. et al. Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. *Science* 356, 543-546 (2017).
- Reich, D. et al. Reconstructing Native American population history. *Nature* 488, 370-4 (2012).
- Ansari Pour, N., Plaster, C.A. & Bradman, N. Evidence from Y-chromosome analysis for a late exclusively eastern expansion of the Bantu-speaking people. *Eur J Hum Genet* 21, 423-9 (2013).
- Thomas, M.G. et al. Y chromosomes traveling south: the cohen modal haplotype and the origins of the Lemba--the "Black Jews of Southern Africa". *Am J Hum Genet* 66, 674-86 (2000).
- Veeramah, K.R. et al. Little genetic differentiation as assessed by uniparental markers in the presence of substantial language variation in peoples of the Cross River region of Nigeria. *BMC Evol Biol* 10, 92 (2010).
- Chang, C.C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7 (2015).
- Delaneau, O., Zagury, J.F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods* 10, 5-6 (2013).
- Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* 26, 2867-73 (2010).

- Alexander, D.H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19, 1655-64 (2009).
- Maples, B.K., Gravel, S., Kenny, E.E. & Bustamante, C.D. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet* 93, 278-88 (2013).
- Jordan, I.K., Rishishwar, L. & Conley, A.B. Native American admixture recapitulates population-specific migration and settlement of the continental United States. *PLoS Genet* 15, e1008225 (2019).
- Rishishwar, L. et al. Ancestry, admixture and fitness in Colombian genomes. *Sci Rep* 5, 12376 (2015).
- Lawson, D.J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet* 8, e1002453 (2012).
- Mullen, K.M. & van Stokkum, H.M. The Lawson-Hanson algorithm for non-negative least squares (NNLS). 1.4 edn (R Foundation for Statistical Computing, Vienna, Austria, 2012).
- Patterson, N. et al. Ancient admixture in human history. *Genetics* 192, 1065-93 (2012).
49. Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet* 2, e190 (2006).
- Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406-25 (1987).
- Tamura, K., Stecher, G., Peterson, D., Filipiński, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30, 2725-9 (2013).
- Gravel, S. et al. Reconstructing Native American migrations from whole-genome and whole-exome data. *PLoS Genet* 9, e1004023 (2013).
- Gravel, S. Population genetics models of local ancestry. *Genetics* 191, 607-19 (2012).
- Weissensteiner, H. et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res* 44, W58-63 (2016).
- Severson, A.L. et al. SNAPPY: Single Nucleotide Assignment of Phylogenetic Parameters on the Y chromosome. *BioRxiv* (2018).
- Rishishwar, L. & Jordan, I.K. Implications of human evolution and admixture for mitochondrial replacement therapy. *BMC Genomics* 18, 140 (2017).

- Fortes-Lima, C. et al. Genome-wide Ancestry and Demographic History of African-Descendant Maroon Communities from French Guiana and Suriname. *Am J Hum Genet* 101, 725-736 (2017).
- Homburger, J.R. et al. Genomic Insights into the Ancestry and Demographic History of South America. *PLoS Genet* 11, e1005602 (2015).
- Moreno-Estrada, A. et al. Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science* 344, 1280-5 (2014).
- Moreno-Estrada, A. et al. Reconstructing the population genetic history of the Caribbean. *PLoS Genet* 9, e1003925 (2013).
- Thornton, J.K. *Africa and Africans in the Making of the Atlantic World, 1400-1800*, (Cambridge University Press, Cambridge, 1998).
- Mufwene, S.S. Kikongo-Kituba. in *Encyclopedia Britannica* (Encyclopedia Britannica, Edinburgh, 2008).
- Pollitzer, W.S. *The Gullah people and their African heritage*, (University of Georgia Press, 1999).
- Sessarego, S. Calunga and the Legacy of an African Language in Brazil by Steven Byrd. *Luso-Brazilian Review* 53, E53-E56 (2016).
- Spears, A.K. Haitian Creole. *Languages and Dialects in the US: Focus on Diversity and Linguistics*, 180-195 (2014).
- Lipski, J.M. From bozal to boricua: Implications of Afro-Puerto Rican Language in Literature. *Hispania*, 850-859 (2001).
- Dianteill, E. Kongo in Cuba: the Transformations of an African Religion. *Archives de sciences sociales des religions* 117, 59-80 (2002).
- Busby, G.B. et al. Admixture into and within sub-Saharan Africa. *Elife* 5(2016).
- Dawkins, R. *The Selfish Gene*, (Oxford university press, Oxford, 1976).

11. FIGURAS COMPLEMENTARIAS

Figura Complementaria 1. Localización geográfica de las muestras en las distintas regiones de Colombia.

Sitios de muestreo por departamentos



Risaralda



Región Andina



Región Caribe



Santa Marta

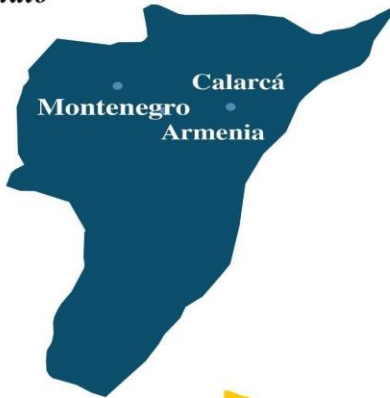


Ciénaga

Fundación

Magdalena

Quindío



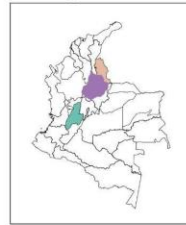
Región Andina



Caldas



Región Andina



Norte de Santander

● Admixed ★ Natives Bari Motilón



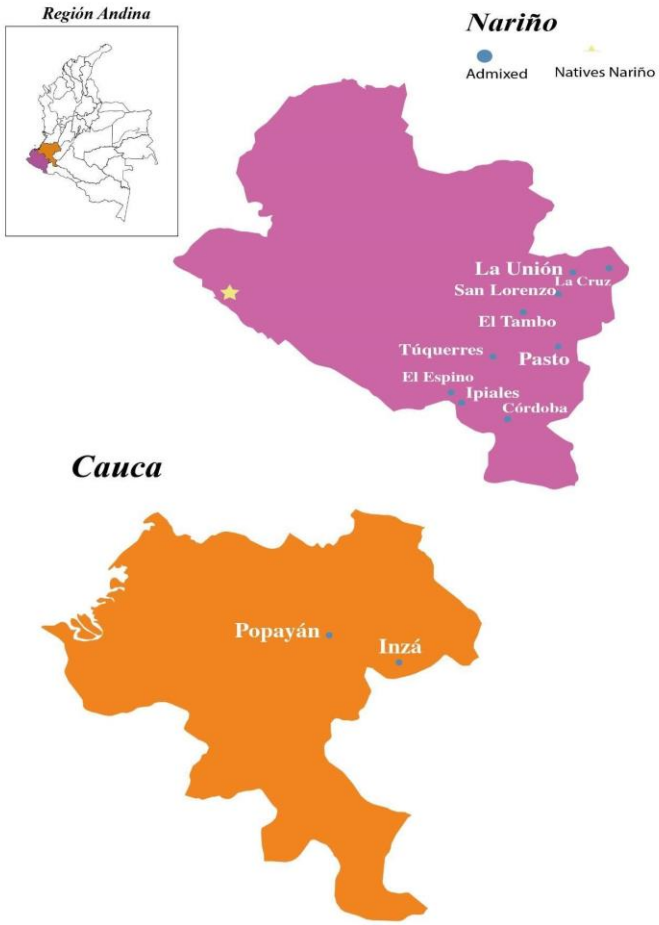
Tolima

● Admixed ★ Natives Pijaos



Santander



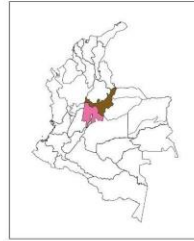




Cundinamarca



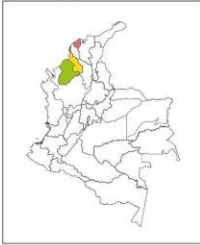
Región Andina



Boyacá



Región Caribe



Atlántico



Barranquilla

Córdoba



Sahagún
Montería

Sucre



Sincelejo
Corozal



12. TABLAS COMPLEMENTARIAS

Tabla Complementaria 1. Detalles sobre todas las poblaciones de referencia analizadas para los estudios de WGS.

African aut. DNA											
Region	Subregion	Country	Population	Language	Glotto Language Subgroup	Guthrie Bantu Subclassification (Narrow Bantu)	Glottocode	Latitude	Longitude	n	Source (PMID)
Senegambia	Senegambia	Gambia	Mandinka	Mandinka	Mande	na	mand1436	19,00	-91,00	22	26432245
Senegambia	Senegambia	Sierra Leone	Mende	Mende	Mande	na	mend1266	8,56	-11,79	128	26432245
Senegambia	Senegambia	Ivory Coast	Yacouba	Dan	Mande	na	west2878	7,27	-8,17	17	28473590
Senegambia	Senegambia	Ivory Coast	Ahizi	Aizi	Kru	na	aizi1248	5,25	-4,61	20	28473590
Bight of Benin	Bight of Benin	Nigeria	Yoruba	Yoruba	Defoid	na	yoru1245	9,59	8,09	108	26432245
Bight of Benin	Bight of Benin	Nigeria	Yoruba	Yoruba	Defoid	na	yoru1245	7,36	2,60	19	28473590
Bight of Benin	Bight of Benin	Benin	Bariba	Baatonum	North Volta-Congo	na	baat1238	9,35	2,62	20	28473590
Bight of Benin	Bight of Benin	Benin	Fon	Fon	Kwa Volta-Congo	na	foan1241	6,35	2,41	12	28473590
Bight of Benin	Bight of Benin	Nigeria	Esan	Esan	Akpes-Edoid		esan1238	9,59	8,09	99	26432245
Bight of Benin/Loango Coast	Admixed	Cameroon	Yaounde	Yaunde-Fang	Southern Bantoid: Narrow Bantu	A	yaun1239	3,85	11,50	39	28473590
Loango Coast	Kongo-like	Republic of the Congo	Kongo	Kikongo	Southern Bantoid: Narrow Bantu	H	core1256	-8,83	13,22	10	28473590

Loango Coast	Kongo-like	Angola	Kimbundu	Kimbundu	Southern Bantoid: Narrow Bantu	H	kimb1241	-8,83	13,22	17	28473590
Loango Coast	Kongo-like	Angola	Ovimbundu	Umbundu	Southern Bantoid: Narrow Bantu	R	umbu1257	-8,83	13,22	15	28473590
Loango Coast	Kongo-like	Angola	Umbundo	Umbundu	Southern Bantoid: Narrow Bantu	R	umbu1257	-8,83	13,22	5	28473590
Loango Coast	Batekee-like	Republic of the Congo	Bateke	Teke	Southern Bantoid: Narrow Bantu	B	teke1283	-0,82	12,70	44	28473590
Loango Coast	Batekee-like	Gabon	Nzebi	Njebi	Southern Bantoid: Narrow Bantu	B	njeb1242	-1,57	13,20	60	28473590
Loango Coast	Batekee-like	Gabon	Bapunu	Punu	Southern Bantoid: Narrow Bantu	B	punu1239	-1,87	11,02	49	28473590
Region	Subregion	Country	Population	Language	Glotto Language Subgroup	Guthrie Bantu Subclassification (Narrow Bantu)	Glottocode	Latitude	Longitude	n	Source (PMID)
Loango Coast	Batekee-like	Gabon	Tsogo	Tsogo	Southern Bantoid: Narrow Bantu	B	tsog1243	-1,03	10,67	61	28473590
Loango Coast	Batekee-like	Gabon	Eshira	Punu	Southern Bantoid: Narrow Bantu	B	punu1239	-1,22	10,60	41	28473590
Loango Coast	Batekee-like	Gabon	Galoa	Myene	Southern Bantoid: Narrow Bantu	B	myen1241	-0,70	10,22	49	28473590
Loango Coast	Batekee-like	Gabon	Orungu	Myene	Southern Bantoid: Narrow Bantu	B	myen1241	-0,72	8,78	20	28473590

Loango Coast	Duma-like	Gabon	Duma	Duma	Southern Bantoid: Narrow Bantu	B	duma1253	-0,82	12,70	43	28473590
Loango Coast	Duma-like	Gabon	Ndumu	Ndumu	Southern Bantoid: Narrow Bantu	B	ndum1239	-1,63	13,58	37	28473590
Loango Coast	Duma-like	Gabon	Obamba	Ombamba	Southern Bantoid: Narrow Bantu	B	omba1241	-0,68	13,78	46	28473590
Loango Coast	Fang-like	Equatorial Guinea	Benga	Benga	Southern Bantoid: Narrow Bantu	A	beng1282	0,58	9,33	46	28473590
Loango Coast	Fang-like	Equatorial Guinea	Fan	Fang	Southern Bantoid: Narrow Bantu	A	fang1246	1,60	11,58	68	28473590
Loango Coast	Makina-like	Cameroon	Badwe'e	Koonzime	Southern Bantoid: Narrow Bantu	A	koon1245	3,56	13,12	39	28473590
Loango Coast	Makina-like	Gabon	Akele	Kele	Southern Bantoid: Narrow Bantu	B	kele1257	-0,70	10,22	41	28473590
Loango Coast	Makina-like	Gabon	Okande	Kande	Southern Bantoid: Narrow Bantu	B	kand1300	-0,05	11,62	8	28473590
Loango Coast	Makina-like	Gabon	Makina	Fang	Southern Bantoid: Narrow Bantu	A	fang1246	-0,10	11,93	41	28473590
Loango Coast	Makina-like	Gabon	Bakota	Kota	Southern Bantoid: Narrow Bantu	A	kota1274	0,57	12,87	50	28473590
Loango Coast	Admixed	Gabon	Babongo	Teke/Njebi/Tsogo	Southern Bantoid: Narrow Bantu	B	na			33	28473590

East Bantu		Kenya	Luhya	Luyia	Southern Bantoid: Narrow Bantu	J	cent2288	0,60	37,80	99	26432245
Rain Forest Hunter Gatherer (RFHG)		Central African Republic	Biaka	Yaka	Southern Bantoid: Narrow Bantu	C	yaka1272	30,50	66,50	23	28473590

African Y-DNA										
Country	Population		Language	Glotto Language Subgroup	Guthrie Bantu Subclassification (Narrow Bantu)	Glottocode	Latitude	Longitude	n	Source (PMID)
	Name	Abbreviation								
Cameroon	Bankim	BA	Tikar	Northern Bantoid	na	tika1246	6,08	11,48	33	27030413
Cameroon	Foumban	FO	Bamoun	Southern Bantoid: Wide Grassfields	na	bamo1241	5,73	10,90	117	27030413
Cameroon	Wum	WU	Aghem	Southern Bantoid: Wide Grassfields	na	aghe1239	6,39	10,07	116	27030413

Congo	Bembe	BE	Bembe	Southern Bantoid: Narrow Bantu	H	bemb1255	-4,80	11,85	109	27030413
Congo	Kuni	KU	Kuni	Southern Bantoid: Narrow Bantu	H	na	-4,80	11,85	68	27030413
Congo	Lari	LA	Lari	Southern Bantoid: Narrow Bantu	H	na	-4,26	15,29	62	27030413
Congo	Mboshi	MB	Mboshi	Southern Bantoid: Narrow Bantu	C	koyo1244	-4,26	15,29	91	27030413
Congo	Sundi	SU	Sundi	Southern Bantoid: Narrow Bantu	H	na	-4,26	15,29	25	27030413
Congo	Teke	TE	Teke	Southern Bantoid: Narrow Bantu	B	teke1283	-4,26	15,29	63	27030413
Congo	Vili	VI	Vili	Southern Bantoid: Narrow Bantu	H	vili1238	-4,80	11,85	108	27030413
Congo	Yombe	YO	Yombe	Southern Bantoid: Narrow Bantu	H	yomb1244	-4,43	12,11	65	27030413
										27030413
Ghana	Asante	AS	Akan	Kwa Volta- Congo	na	akan1250	6,11	-1,88	94	27030413

Ghana	Ewe	EW	Ewe	Kwa Volta-Congo	na	ewee1241	6,60	0,47	88	27030413
Ghana	Fante	FA	Akan	Kwa Volta-Congo	na	fant1241	5,82	-2,82	60	27030413
Malawi	Chewa	CH	Nyanja	Southern Bantoid: Narrow Bantu	N	nyan1308	-13,61	33,92	92	27030413
Malawi	Tumbuka	TU	Tumbuka	Southern Bantoid: Narrow Bantu	N	tumb1250	-14,27	34,79	61	27030413
Malawi	Yao	YA	Yao	Southern Bantoid: Narrow Bantu	P	yaoo1241	-12,77	33,87	56	27030413
										27030413
Mozambique	Sena	SE	Sena	Southern Bantoid: Narrow Bantu	N	nucl1396	-17.442	35,03	62	27030413
										27030413
South Africa	Bantu speakers Pretoria	BN	Narrow Bantu	Southern Bantoid: Narrow Bantu	S	na	-25,75	28,19	98	27030413
Nigeria	Abak	AB	Anaang	Delta Cross	na	anaa1238	5,05	7,72	56	27030413
	Afaha Eket	AE	Ibibio	Delta Cross	na	ibib1240	4,72	7,87	48	27030413

Country	Name	Abbreviation	Language	Glotto Language Subgroup	Guthrie Bantu Subclassification (Narrow Bantu)	Glottocode	Latitude	Longitude	n	Source (PMID)
	Afaha Okpo	AO	Oron	Delta Cross	na	na	4,83	8,23	50	27030413
	Afaha Ukwong	AU	Oron	Delta Cross	na	na	4,75	8,25	49	27030413
	Awa-Onna	AW	Ibibio	Delta Cross	na	ibib1240	4,69	7,82	28	27030413
	Calabar	CA	Igbo	Igboid	na	nucl1417	4,95	8,32	99	27030413
	Ediene Ikono	ED	Ibibio	Delta Cross	na	ibib1240	4,78	7,88	48	27030413
	Efut Akpabuyo	EF	Efik	Delta Cross	na	efik1245	4,91	8,44	48	27030413
	Efut Odukpani	EO	Efik	Delta Cross	na	efik1245	5,17	7,98	49	27030413
	Ejagham-Akamkpa	EA	Ejagham	Southern Bantoid: Ekoid-Mbe	na	ejag1239	5,35	8,35	47	27030413
	EjaghamCalabar	EC	Ejagham	Southern Bantoid: Ekoid-Mbe	na	ejag1239	4,95	8,32	83	27030413
	Eziagu Nenwe	Igbo	Igbo	Igboid	na	nucl1417	6,12	7,52	49	27030413
	Ikono	IK	Anaang	Delta Cross	na	anaa1238	4,99	7,76	42	27030413
	Itam	IT	Ibibio	Delta Cross	na	ibib1240	5,04	7,84	50	27030413
	Nike Enugu	NE	Igbo	Igboid	na	nucl1417	6,43	7,48	54	27030413
	Nnung NdemOnna	NN	Ibibio	Delta Cross	na	ibib1240	4,63	7,85	48	27030413

	Nsit	NS	Ibibio	Delta Cross	na	ibib1240	4,83	7,90	36	27030413
	Ntan Ibiono	NT	Ibibio	Delta Cross	na	ibib1240	5,23	7,93	50	27030413
	Obong Itam	OB	Ibibio	Delta Cross	na	ibib1240	5,13	7,97	50	27030413
	Oku-Itu	OI	Ibibio	Delta Cross	na	ibib1240	5,13	7,93	49	27030413
	Oku-Uyo	OU	Ibibio	Delta Cross	na	ibib1240	5,10	7,97	48	27030413
	Ukpom Ete	UE	Ibibio	Delta Cross	na	ibib1240	4,62	7,65	50	27030413
	Uwanse	UW	Efik	Delta Cross	na	efik1245	4,95	8,32	50	27030413

European aut. DNA

Region	Country	Population	Latitude	Longitude	n	Source (PMID)
Northern Europe	Finland	Finnish in Finland	60,17	24,94	99	26432245
Southern Europe	Spain	Iberian populations in Spain	40,24	-3,65	107	26432245
Southern Europe	Italy	Toscani in Italy	42,80	12,07	107	26432245
Western and Central Europe	United Kingdom	British in England and Scotland	56,50	-6,03	91	26432245
Western and Central Europe	United States	Utah residents with Northern and Western European ancestry	42,17	-2,76	99	26432245

Native American aut. DNA

Region	Country	Population	Latitude	Longitude	n	Source (PMID)
Amazonia/East Colombia	Argentina	Toba	-26,50	-59,30	4	22801491
Amazonia/East Colombia	Argentina	Wichi	-22,50	-63,80	5	22801491
Amazonia/East Colombia	Colombia	Ticuna	-3,81	-70,01	6	22801491
Amazonia/East Colombia	Colombia	Inga	1,00	-77,00	9	22801491
Amazonia/East Colombia	Colombia	Piapoco	3,00	-68,00	7	22801491
Amazonia/East Colombia	Guiana	Palikur	4,00	-51,80	3	22801491
Amazonia/East Colombia	Colombia	Guahibo	5,80	-69,50	6	22801491
Amazonia/East Colombia	Paraguay & Argentina	Guarani	-23.00/-22.50	-54.00/-63.80	6	22801491
Andes	Bolivia & Peru	Quechua	-14.50/-14.00	-69.00/-74.00	40	22801491
Andes	Bolivia & Chile	Aymara	-16.50/-22.00	-68.20/-70.00	23	22801491
Mesoamerica	Honduras	Lenca	13,53	-87,68		31217584
Mesoamerica	Guatemala	Kaqchikel	14,00	-91,00	13	22801491
Mesoamerica	Mexico	Mixe	17,00	-96,00	17	22801491
Mesoamerica	Mexico	Mixtec	17,00	-97,00	5	22801491
Mesoamerica	Mexico	Tepehuano	23,20	-104,50	25	22801491
Panamanian/West Colombian	Colombia	Waunana	5,00	-77,00	3	22801491
Panamanian/West Colombian	Colombia	Embera	7,00	-76,00	5	22801491
Panamanian/West Colombian	Costa Rica	Guaymi	8,50	-82,00	5	22801491
Panamanian/West Colombian	Costa Rica	Teribe	9,00	-83,20	3	22801491
Panamanian/West Colombian	Colombia	Sinú	9,24	-75,82	19	31217584
Panamanian/West Colombian	Costa Rica	Cabecar	9,50	-84,00	31	22801491
Panamanian/West Colombian	Costa Rica	Maleku	10,60	-84,80	3	22801491
Panamanian/West Colombian	Colombia	Kogi	11,00	-74,00	4	22801491
Panamanian/West Colombian	Colombia	Wayuu	11,00	-73,00	11	22801491

AIMs						
Continental Group	Country	Population	Latitude	Longitude	n	Source (PMID)
European	Russia-Caucasus	Adygei	44,0	39,0	17	17044859
African	Kenya	BantuKenya	-3,0	37,0	11	17044859
African	Lesotho	BantuSouthEast	-28,4	27,6	1	17044859
African	Angola	BantuSouthWest	-15,8	18,7	1	17044859
African	Botswana Or Namibia	BantuSouthWest	-21,0	18,7	4	17044859
European	France	Basque	43,0	0,0	24	17044859
African	CAR	BiakaPygmy	4,0	17,0	23	17044859
East Asian	Cambodia	Cambodian	12,0	105,0	10	17044859
The Americas	Colombia	Colombian	3,0	-68,0	7	17044859
East Asian	China	Dai	21,0	100,0	10	17044859
East Asian	China	Daur	48,5	124,0	9	17044859
European	France	French	46,0	2,0	28	17044859
East Asian	China	Han	32,5	114,0	34	17044859
East Asian	China	Han-NChina	39,0	114,0	10	17044859
East Asian	China	Hezhen	47,5	133,5	9	17044859
European	Italy	Italian	46,0	10,0	13	17044859
Continental Group	Country	Population	Latitude	Longitude	n	Source (PMID)
East Asian	Japan	Japanese	38,0	138,0	29	17044859
The Americas	Brazil	Karitiana	-10,0	-63,0	14	17044859
East Asian	China	Lahu	22,0	100,0	8	17044859
African	Senegal	Mandenka	12,0	-12,0	22	17044859
The Americas	Mexico	Maya	19,0	-91,0	21	17044859

African	Congo	MbutiPygmy	1,0	29,0	13	17044859
Oceania	Bougainville	Melanesian	-6,0	155,0	11	17044859
East Asian	China	Miao	28,0	109,0	10	17044859
East Asian	China	Mongola	48,5	119,0	10	17044859
East Asian	China	Naxi	26,0	100,0	7	17044859
European	Scotland	Orcadian	59,0	-3,0	15	17044859
East Asian	China	Oroqen	50,5	126,5	9	17044859
Oceania	New Guinea	Papuan	-4,0	143,0	17	17044859
The Americas	Mexico	Pima	29,0	-108,0	14	17044859
European	Russia	Russian	61,0	40,0	25	17044859
African	Namibia	San	-21,0	20,0	6	17044859
European	Italy	Sardinian	40,0	9,0	28	17044859
East Asian	China	She	27,0	119,0	10	17044859
America	Brazil	Surui	-11,0	-62,0	8	17044859
East Asian	China	Tu	36,0	101,0	10	17044859
East Asian	China	Tujia	29,0	109,0	10	17044859
European	Italy	Tuscan	43,0	11,0	8	17044859
East Asian	China	Xibo	43,5	81,5	9	17044859
East Asian	Siberia	Yakut	63,0	129,5	25	17044859
East Asian	China	Yi	28,0	103,0	10	17044859
African	Nigeria	Yoruba	8,0	5,0	22	17044859

Tabla Complementaria 2. Lista de haplotipos de mtDNA y haplogrupos en SBP.

ID	ADNmt	Haplogrupo		
	Haplotipo (16024-576)	EMPOP	Haplogrep	Score Haplogrep
PR004	16114A 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 195C 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a1'4	L1b1a1'4	1,000
PR005	16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000
PR007	16189C 16215G 16223T 16278T 16294T 16311C 16355T 16360T 16519C 73G 146C 152C 195C 263G 315.1C	L2a1c3b2	L2a1c3b2	0,768
PR009	16129A 16183C 16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16360T 16390A 16519C 73G 151T 152C 182T 183G 186A 189C 204C 247A 257G 263G 309.1C 315.1C 316A 523del 524del	L1c3	L1c3	0,821
PR010	16111T 16126C 16223T 16290T 16319A 16362C 64T 73G 146C 153G 235G 263G 309.1C 309.2C 315.1C 523del 524del	A2a1	A2a1	1,000
PR011	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a+189	L1b1a+189	1,000
PR013	16114A 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 195C 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a1'4	L1b1a1'4	1,000

PR018	16114A 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 195C 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a1'4	L1b1a1'4	1,000
PR024	16189C 16192Y 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,049
PR026	16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000
PR027	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a+189	L1b1a+189	1,000
PR033	16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000
PR038	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16311C 16519C 73G 152C 182T 185T 195C 228A 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a7a	L1b1a7a	0,971
PR039	16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del	L3d1a1a	L3d1a1a	0,942
PR040	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a+189	L1b1a+189	1,000
PR041	16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del	L3d1a1a	L3d1a1a	0,942
PR051	16223T 16265G 16298C 16319A 16325C 16327T 16519C 73G 249del 263G 290del 291del 309.1C 315.1C 489C	C1c3	C1c3	1,000
PR052	16111T 16176T 16189Y 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 309.1C 315.1C	L3e1d	L3e1d	0,954
PR058	16093C 16129A 16189C 16278T 16300G 16311C 16354T 16390A 16399G 16519C 73G 146C 150T 152C 195C 263G 315.1C 456T 524.1A 524.2C	L2d+16129	L2d+16129	0,953

PR061	16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del	L3d1a1a	L3d1a1a	0,942
-------	---	---------	---------	-------

ID	ADNmt	Haplogrupo		
	Haplotipo (16024-576)	EMPOP	Haplogrep	Score Haplogrep
PR062	16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del	L3d1a1a	L3d1a1a	0,942
PR068	16223T 16278T 16294T 16390A 73G 143A 146C 152C 195C 263G 309.1C 315.1C 523del 524del	L2a1+143+@16309	L2a1+143+@16309	1,000
PR071	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a+189	L1b1a+189	1,000
PR072	16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000
PR073	16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del	L3d1a1a	L3d1a1a	0,942
PR077	16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000
PR081	16189C 16191.1C 16192Y 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,049
PR082	16111T 16176T 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 309.1C 315.1C	L3e1d	L3e1d	0,954
PR085	16038.1A 16129A 16183C 16189C 16193.1C 16215G 16223T 16278T 16294T 16311C 16355T 16360T 16519C	L1c3a1b	L1c3a1b	0,925

	73G 151T 152C 182T 183G 186A 189C 204C 247A 257G 263G 309.1C 315.1C 316A 523del 524del			
PR086	16189C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000
PR087	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a+189	L1b1a+189	1,000
PR088	16111T 16176T 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 309.1C 315.1C	L3e1d	L3e1d	0,954
PR090	16114A 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 195C 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a1'4	L1b1a1'4	1,000
PR091	16189C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000
PR093	16111T 16176T 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 309.1C 315.1C	L3e1d	L3e1d	0,954
PR095	16111T 16176T 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 315.1C	L3e1d	L3e1d	0,954
PR097	16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000
PR099	16092C 16111T 16223T 16290T 16319A 16362C 64T 73G 146C 153G 235G 263G 315.1C 482C 523del 524del	A2+(64)	A2+(64)	0,911
PR100	16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del	L3d1a1a	L3d1a1a	0,942
PR102	16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000

PR103	16038.1A 16129A 16183C 16189C 16215G 16223T 16278T 16294T 16311C 16355T 16360T 16519C 73G 151T 152C 182T 183G 186A 189C 204C 247A 257G 263G 309.1C 315.1C 316A 523del 524del	L1c3a1b	L1c3a1b	0,925
PU004	16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del	L3d1a1a	L3d1a1a	0,942
PU005	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a+189	L1b1a+189	1,000

ID	ADNmt	Haplogrupo		
	Haplotipo (16024-576)	EMPOP	Haplogrep	Score Haplogrep
PU006	16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del	L3d1a1a	L3d1a1a	0,942
PU009	16114A 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 195C 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a1'4	L1b1a1'4	1,000
PU011	16183C 16189C 16217C 16519C 73G 263G 309.1C 309.2C 315.1C 498del 499A	B2d	B2d	1,000
PU015	16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000
PU016	16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000

PU019	16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000
PU027	16129A 16183C 16189C 16193.1C 16215G 16223T 16278T 16294T 16311C 16360T 16519C 73G 151T 152C 182T 183G 186A 189C 204C 247A 257G 263G 309.1C 315.1C 316A 523del 524del	L1c3a	L1c3a	0,922
PU029	16114A 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16319A 16519C 73G 152C 182T 185T 195C 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a1'4	L1b1a1'4	0,982
PU030	16111T 16223T 16290T 16319A 16360T 16362C 89C 97A 106del 107del 108del 109del 110del 111del 146C 198T 235G 263G 309.1C 315.1C	A2af1a1	A2af1a1	0,916
PU037	16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000
PU040	16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del	L3d1a1a	L3d1a1a	0,942
PU041	16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000
PU042	16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000
PU043	16111T 16223T 16290T 16319A 16360T 16362C 89C 97A 106del 107del 108del 109del 110del 111del 146C 198T 235G 263G 309.1C 315.1C	A2af1a1	A2af1a1	0,916
PU044	16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000

PU045	16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 198T 263G 315.1C	L2a1	L2a1	0,952
PU047	16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000
PU050	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a+189	L1b1a+189	1,000
PU053	16111T 16176T 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 309.1C 315.1C	L3e1d	L3e1d	0,954
PU055	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263R 309.1C 315.1C 357G 523del 524del	L1b1a+189	L1b1a+189	1,050
PU057	16223T 16265G 16298C 16319A 16325C 16327T 16519C 73G 249del 263G 290del 291del 309.1C 315.1C 489C	C1c3	C1c3	1,000
PU062	16114A 16129A 16174T 16213A 16223T 16278T 16355T 16362C 16390A 73G 146C 150T 152C 182T 195C 198T 204C 263G 309.1C 315.1C 418T 523del 524del	L2b1a	L2b1a	0,948
ID	ADNmt	Haplogrupo		
	Haplotype (16024-576)	EMPOP	Haplogrep	Score Haplogrep
PU063	16114A 16191.1C 16192Y 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1	L2a1+16189+(16192)	0,955
PU065	16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del	L3d1a1a	L3d1a1a	0,942
PU067	16111T 16176T 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 309.1C 315.1C	L3e1d	L3e1d	0,954

PU069	16111T 16176T 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 309.1C 315.1C	L3e1d	L3e1d	0,954
PU073	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a+189	L1b1a+189	1,000
PU076	16111T 16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523del 524del	L3d1a1a	L3d1a1a	0,942
PU079	16114A 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 195C 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a1'4	L1b1a1'4	1,000
PU080	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 182T 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a+189	L1b1a+189	1,000
PU081	16209C 16223T 16292T 16311C 16519C 73G 189G 195C 200G 263G 315.1C	L3f1b+16292	L3f1b+16292	0,976
PU092	16189C 16191.1C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 315.1C	L2a1+16189+(16192)	L2a1+16189+(16192)	1,000
PU094	16129A 16183C 16189C 16193.1C 16215G 16223T 16278T 16294T 16311C 16355T 16360T 16519C 73G 151T 152C 182T 183G 186A 189C 204C 247A 257G 263G 309.1C 315.1C 316A 523del 524del	L1c3a1b	L1c3a1b	0,925
PU095	16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16311C 16320T 64T 93G 152C 189G 200G 236C 247A 263G 309.1C 315.1C 523del 524del	L0a1a+200	L0a1a+200	0,928
PU097	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16293G 16311C 16519C 73G 152C 185T 189G 195C 247A 263G 309.1C 315.1C 357G 523del 524del	L1b1a18	L1b1a18	1,000
PU098	16111T 16176T 16223T 16327T 16519C 73G 150T 152C 189G 200G 263G 309.1C 315.1C	L3e1d	L3e1d	0,954
PU100	16129A 16183C 16189C 16193.1C 16215G 16223T 16278T 16294T 16311C 16355T 16360T 16519C 73G	L1c3a1b	L1c3a1b	0,925

	151T 152C 182T 183G 186A 189C 204C 247A 257G 263G 309.1C 315.1C 316A 523del 524del			
PU103	16129A 16183C 16189C 16215G 16223T 16278T 16294T 16311C 16355Y 16360T 16519C 73G 151T 152C 182T 183G 186A 189C 204C 247A 257G 263G 309.1C 315.1C 316A 523del 524del	L1c3a1b	L1c3a1b	0,943

Tabla Complementaria 3. Lista de haplotipos de cromosoma Y en SBP.

Muestra ID	Cromosoma Y																							Haplogrupo			
	D Y 57 6	D Y 38 9I	D Y 63 5	D Y 38 9I	D Y S6 27	D Y S 46 0	D Y S 45 8	D Y S 19	Y T A H 4	D Y S4 48	D Y S 39 1	D Y S 45 6	D Y S 39 0	D Y S 43 8	D Y S 39 2	D Y S 51 8	D Y S 57 0	D Y S 43 7	D Y S 3 8	D Y S 44 9	D Y S 39 3	D Y S 43 9	D Y S 48 1		D Y S 387 S1	D Y S 53 3	
PR004	19	14	23	30	22	11	17	14	12	19	11	16	24	12	13	37	17	15	11, 14	29	13	12	22	35, 37	12	R1b-U152	
PR005	18	13	22	29	19	11	17	13	12	21	11	13	23	11	14	37	17	14	14, 14	29	13	13	25	35, 36, 41	12	Q1a2-M3* (xM19,M194,M199)	
PR006	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	Y-MRCA* (xM13,SRY10831.1)
PR007	16	13	21	31	18	10	17	15	12	21	10	15	21	11	10	40	21	14	16, 18	29	13	13	28	35, 39	11	E1b1a-M2* (xM154,M191)	
PR009	18	12	21	29	19	11	18	13	11	19	10	15	24	10	11	40	20	14	16, 20	30	13	12	26	39, 39	10	E1b1b-M123	
PR010	18	14	23	29	21	11	16	14	12	19	11	15	26	12	14	36	17	15	11, 14	30	13	12	22	35, 35	13	R1b-M529	
PR011	18	13	21	30	19	11	16	17	11	20	10	17	21	11	11	42	17	14	16, 17	32	14	13	24	39, 39	11	E1b1a-M191	
PR013	19	13	23	32	17	12	14	16	12	20	11	16	23	11	11	42	10	14	11, 13	32	13	10	24	37, 41	13	R1a-SRY1831.2	
PR018	17	14	24	31	18	11	16	15	11	20	10	15	21	11	11	40	16	14	17, 17	30	14	11	25	37, 40	11	E1b1a-M191	

PR024	18	13	23	29	21	10	17	14	12	19	11	16	24	12	13	38	17	15	11, 14	25	13	12	21	37, 37	12	R1b-M529
PR026	18	12	21	29	19	11	18	13	11	19	10	15	24	10	11	39	20	14	16, 20	30	13	12	26	39, 39	10	E1b1b-M123
PR027	15	13	21	30	19	10	16	16	12	21	10	14	21	11	11	40	18	14	17, 19	28	13	13	28	36, 39	11	E1b1a-M2* (xM154,M191)
PR033	15	13	21	31	21	10	16	15	12	21	10	16	21	11	11	38	19	14	16, 17	28	12	11	28	37, 39	11	E1b1a-M2* (xM154,M191)
PR038	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	<i>nd</i>	E1b1a-M2* (xM154,M191)
PR039	18	12	21	29	19	11	18	13	11	19	10	15	24	10	11	40	20	14	16, 20	30	13	13	26	39, 39	10	E1b1b-M123
PR040	16	12	23	31	20	10	16	15	11	21	10	15	21	11	12	38	19	14	16, 16	32	13	12	30	35, 36	11	E1b1a-M2* (xM154,M191)
PR041	17	12	20	30	20	10	17	15	12	21	10	15	21	11	11	40	18	14	16, 16	29	13	12	27	35, 38	10	E1b1a-M2* (xM154,M191)
PR043	18	14	22	30	18	11	18	13	12	20	9	16	24	10	11	40	22	14	13, 16	32	13	10	29	36, 37	11	E1b1b-M81
PR048	19	14	23	30	23	10	17	14	11	18	12	17	24	12	13	39	17	14	12, 14	30	13	12	22	35, 36	12	R1b-S116* (xU152,M529,M153,M167)
PR051	16	12	23	31	20	10	16	15	11	21	10	15	21	11	12	39	19	14	16, 17	32	13	11	30	35, 35	11	E1b1a-M2* (xM154,M191)
PR052	15	13	21	31	21	10	16	15	12	21	10	16	21	11	11	38	18	14	16, 17	28	12	11	28	37, 39	11	E1b1a-M2* (xM154,M191)
PR058	15	13	21	31	21	10	17	15	12	20	11	15	21	11	11	39	20	14	16, 17	28	13	11	28	37, 39	11	E1b1a-M2* (xM154,M191)
PR061	16	13	21	31	18	10	17	15	12	21	10	15	21	11	10	40	21	14	16, 18	29	13	13	28	35, 39	11	E1b1a-M2* (xM154,M191)
PR062	15	14	23	31	18	10	17	15	11	21	10	15	21	11	11	42	20	14	16, 17	31	13	13	28	38, 39	11	E1b1a-M2* (xM154,M191)
PR065	18	12	21	29	20	11	18	13	11	19	10	15	24	10	11	40	19	14	16, 20	30	13	12	26	39, 39	10	E1b1b-M123
PR068	19	14	23	30	23	10	17	14	11	18	12	17	24	12	13	39	17	14	12, 14	30	13	12	22	35, 37	12	R1b-S116* (xU152,M529,M153,M167)
PR071	16	13	21	31	17, 2	10	17	15	12	21	10	15	21	11	10	41	21	15	16, 18	29	13	13	28	35, 39	11	E1b1a-M2* (xM154,M191)
PR072	15	13	21	30	19	10	16	16	12	20	10	14	21	11	11	40	18	14	17, 19	28	13	12	28	36, 39	11	E1b1a-M2* (xM154,M191)
PR073	17	13	17	32	22	11	18	15	13	22	11	13	20	10	12	36	19	16	17, 17	28	13	10	24	35, 42	12	Y-MRCA* (xM13,SRY10831.1)

PR075	18	13	22	29	19	11	17	13	12	21	11	13	23	11	14	37	17	14	14, 14	29	13	13	25	35, 41	12	Q1a2-M3* (xM19,M194,M199)
PR077	18	12	21	29	19	11	18	13	11	19	10	15	24	10	11	40	20	14	16, 20	30	13	12	26	39, 39	10	E1b1b-M123
PR079	17	14	21	31	19	11	17	16	11	22	10	16	21	11	11	41	18	14	18, 19	33	14	12	26	39, 40	11	E1b1a-M191
PR081	15	13	21	31	21	10	16	15	12	21	10	16	21	11	11	38	19	14	16, 17	28	12	11	28	37, 39	11	E1b1a-M2* (xM154,M191)
PR082	18	14	22	30	18	11	18	13	12	20	9	16	24	10	11	40	22	14	13, 15	32	13	10	29	36, 37	11	E1b1b-M81
PR083	16	13	23	29	23	11	14	15	12	21	9	15	23	9	11	39	18	15	13, 17	31	12	11	22	37, 38	12	J2-M172
PR085	19	13	23	32	17	12	14	16	12	20	11	16	23	11	11	42	10	14	11, 13	32	13	10	24	37, 39	12	R1a-SRY1831.2
PR086	18	12	21	29	19	11	18	13	11	19	10	15	24	10	11	40	20	14	17, 20	30	13	12	26	39, 39	10	E1b1b-M123
PR087	15	13	21	31	21	10	17	15	12	20	11	15	21	11	11	39	20	14	16, 17	28	13	11	28	37, 39	11	E1b1a-M2* (xM154,M191)
PR088	18	13	21	30	19	11	16	17	11	20	10	17	21	11	11	42	17	14	16, 17	31	14	13	24	39, 39	11	E1b1a-M191
PR090	18	12	21	29	19	11	15	14	11	19	10	15	22	10	11	37	19	17	17, 17	29	12	12	24	37, 39	11	E1a-M33
PR091	17	12	20	30	20	10	17	15	12	21	10	15	21	11	11	40	18	14	16, 16	29	13	12	27	35, 38	10	E1b1a-M2* (xM154,M191)
PR093	18	12	21	29	19	12	18	14	11	19	10	15	24	10	11	40	20	14	16, 20	30	13	12	25	39, 39	10	E1b1b-M123
PR095	19	14	23	30	23	10	17	14	11	18	12	17	24	12	13	39	17	14	12, 14	30	13	12	22	35, 36	12	R1b-S116* (xU152,M529,M153,M167)
PR097	19	13	23	32	17	11	15	16	12	20	11	16	23	11	11	42	10	14	11, 13	32	13	10	24	37, 40	12	R1a-SRY1831.2
PR099	17	12	20	30	20	10	17	15	12	21	10	15	21	11	11	40	18	14	16, 16	29	13	12	27	35, 38	10	E1b1a-M2* (xM154,M191)
PR100	16	13	21	31	18	10	17	15	12	21	10	15	21	11	10	40	21	14	16, 18	29	13	13	28	35, 39	11	E1b1a-M2* (xM154,M191)
PR102	18	12	21	29	19	11	18	13	11	19	10	15	24	10	11	40	20	14	16, 20	30	13	12	26	39, 39	10	E1b1b-M123
PR103	17	13	17	32	22	11	18	15	13	22	11	13	20	10	12	36	19	16	17, 17	28	13	10	24	35, 40	12	Y-MRCA* (xM13,SRY10831.1)
PU004	16	12	23	31	20	10	16	15	11	21	10	15	21	11	12	39	19	14	16, 17	32	13	11	30	35, 35	11	E1b1a-M2* (xM154,M191)
PU005	16	12	21	29	19	10	16	17	11	21	10	15	21	11	11	38	18	14	19, 19	33	14	13	25	38, 38	12	E1b1a-M191

PU006	16	13	21	31	18	10	17	15	12	21	10	15	21	11	10	41	21	15	16, 18	29	13	13	28	35, 39	11	E1b1a-M2* (xM154,M191)
PU009	15	12	21	29	21	10	17	16	11	23	10	15	21	10	11	36	18	16	13, 16	28	15	11	22	37, 39	10	G-M201
PU011	17	12	20	30	20	10	17	15	12	21	10	15	21	11	11	40	18	14	16, 16	29	13	12	27	35, 38	10	E1b1a-M2* (xM154,M191)
PU014	15	13	21	31	21	10	16	15	12	21	10	16	21	11	11	38	19	14	16, 17	28	12	11	28	37, 39	11	E1b1a-M2* (xM154,M191)
PU015	16	13	21	31	18	10	18	15	12	21	10	15	21	11	10	40	21	15	16, 18	29	13	13	28	35, 39	11	E1b1a-M2* (xM154,M191)
PU016	18	14	23	30	21	11	17	15	12	19	11	15	24	12	13	40	19	14	13, 15	29	13	12	22	36, 37	12	R1b-V88
PU019	15	13	21	31	21	10	17	15	12	21	10	16	21	11	11	38	19	14	16, 17	28	12	11	28	37, 39	11	E1b1a-M2* (xM154,M191)
PU027	19	14	23	30	23	10	17	14	11	18	12	17	24	12	13	38	16	14	12, 14	30	13	10	22	35, 36	12	R1b-S116* (xU152,M529,M153,M167)
PU029	16	13	17	33	21	10	18	15	11	19	11	14	24	10	11	38	18	13	11, 15	31	14	14	30	35, 41.2	10	B2a-M150* (xM109)
PU030	18	13	21	30	19	11	16	17	11	20	10	17	21	11	11	42	17	14	16, 17	31	14	13	24	39, 39	11	E1b1a-M191
PU037	20	13	23	32	17	12	14	16	12	20	11	16	23	11	11	42	10	14	11, 13	32	13	10	24	37, 41	12	R1a-SRY1831.2
PU040	18	13	21	30	18	10	16	16	11	22	10	17	21	11	11	41	17	14	16, 18	33	14	12	26	38, 39	11	E1b1a-M191
PU041	16	13	21	30	19	10	17	17	11	19, 2	10	16	22	8	11	36	17	17	15, 16	28	13	11	28	36, 38	11	E1a-M33
PU042	18	14	23	30	23	10	17	14	11	18	12	17	24	12	13	39	17	14	12, 14	30	13	12	23	35, 36	12	R1b-S116* (xU152,M529,M153,M167)
PU043	19	14	23	30	23	10	17	14	11	18	12	17	24	12	13	39	17	14	12, 14	30	13	13	22	35, 36	12	R1b-S116* (xU152,M529,M153,M167)
PU044	15	14	21	31	19	9	18	17	11	21	10	15	21	11	11	40	18	14	16, 17	31	14	12	24	36, 40	12	E1b1a-M2* (xM154,M191)
PU045	16	13	21	31	18	10	17	15	12	21	10	15	21	11	10	40	21	14	16, 18	29	13	13	28	35, 39	11	E1b1a-M2* (xM154,M191)
PU047	18	14	24	32	21	10	18	13	12	20	10	15	24	10	11	37	19	14	15, 17	29	12	13	24	37, 37	12	E1b1b-M35* (xM78,M81,M123,V6,M293)

PU050	16	12	23	31	20	10	16	15	11	21	10	15	21	11	12	39	19	14	16, 17	32	13	11	30	35, 35	11	E1b1a-M2* (xM154,M191)
PU053	16	13	21	31	18	10	18	15	12	21	10	15	21	11	10	40	21	15	16, 18	29	13	13	28	35, 39	11	E1b1a-M2* (xM154,M191)
PU055	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	R1b-S116* (xU152,M529,M153,M167)
PU057	18	14	23	31	22	11	17	14	12	19	11	15	24	12	13	38	17	15	11, 14	29	13	13	22	35, 36	12	R1b-S116* (xU152,M529,M153,M167)
PU061	19	14	23	30	23	10	17	14	11	18	12	17	24	12	13	39	17	14	12, 14	30	13	12	22	35, 36	12	R1b-S116* (xU152,M529,M153,M167)
PU062	18	14	24	32	21	10	18	13	12	20	10	15	24	10	11	37	19	14	15, 17	29	12	13	24	37, 37	12	E1b1b-M35* (xM78,M81,M123,V6,M293)
PU063	18	13	22	29	19	11	17	13	12	21	11	13	23	11	14	37	17	14	14, 14	29	13	13	25	35, 40	12	Q1a2-M3* (xM19,M194,M199)
PU064	18	14	24	32	21	10	18	13	12	20	10	15	24	10	11	37	19	14	15, 17	29	12	13	24	37, 37	12	E1b1b-M35* (xM78,M81,M123,V6,M293)
PU065	18	14	23	30	21	11	17	15	12	19	11	15	24	12	13	40	19	14	13, 15	29	13	12	22	36, 37	12	R1b-V88
PU066	16	13	21	31	18	10	17	15	12	21	10	15	21	11	10	40	21	15	16, 18	29	13	13	28	35, 39	11	E1b1a-M2* (xM154,M191)
PU067	16	14	22	32	20	10	18	16	11	21	10	15	21	11	11	38	19	14	16, 17	32	13	13	27	38, 40	13	E1b1a-M2* (xM154,M191)
PU069	16	13	21	31	19	10	17	15	12	21	10	15	21	11	10	40	21	14	16, 18	29	13	13	28	35, 39	11	E1b1a-M2* (xM154,M191)
PU073	16	13	21	31	18	10	17	15	12	21	10	15	21	11	10	41	21	15	16, 18	29	13	13	28	35, 39	11	E1b1a-M2* (xM154,M191)
PU076	16	12	23	31	20	10	16	15	11	21	10	15	21	11	12	38	19	14	16, 17	32	13	11	30	35, 35	11	E1b1a-M2* (xM154,M191)
PU077	16	13	17	32	21	10	18	15	11	19	11	14	24	10	11	38	18	13	11, 15	31	14	14	30	35, 41.2	10	B2a-M150* (xM109)
PU079	17	12	20	30	20	10	17	15	12	21	10	15	21	11	11	40	18	14	16, 16	29	13	12	27	35, 38	10	E1b1a-M2* (xM154,M191)
PU080	18	13	23	32	17	12	14	16	12	20	11	16	23	11	11	42	10	14	11, 13	32	13	10	24	37, 40	12	R1a-SRY1831.2
PU081	16	13	21	28	23	11	18	17	11	20	10	14	23	10	11	36	19	15	11, 12	29	13	12	22	35, 36	12	I2-M26

PU082	16	13	17	32	21	10	18	15	11	19	11	14	24	10	11	38	18	13	11, 15	31	14	14	30	35, 41.2	10	B2a-M150* (xM109)
PU091	19	13	23	32	17	12	14	16	12	20	11	16	23	11	11	42	10	14	11, 13	32	13	10	24	38, 40	12	R1a-SRY1831.2
PU092	19	13	23	32	17	12	14	16	12	20	11	16	23	11	11	42	10	14	11, 13	32	13	10	24	37, 40	12	R1a-SRY1831.2
PU094	18	14	23	30	21	11	17	15	12	19	11	15	24	12	13	40	19	14	13, 15	29	13	12	22	36, 37	12	R1b-V88
PU095	15	13	21	30	19	10	16	16	12	21	10	14	21	11	11	40	18	14	17, 19	28	13	13	28	36, 39	11	E1b1a-M2* (xM154,M191)
PU097	19	13	23	32	17	11	14	16	12	20	11	16	23	11	11	42	10	14	11, 13	32	13	10	24	37, 40	12	R1a-SRY1831.2
PU098	16	13	21	31	18	10	17	15	12	21	10	15	21	11	10	40	21	14	16, 18	29	13	13	28	35, 39	11	E1b1a-M2* (xM154,M191)
PU100	19	13	23	32	17	12	14	16	12	20	11	16	23	11	11	42	10	14	11, 13	32	13	10	24	37, 40	12	R1a-SRY1831.2
PU103	18	14	23	30	21	11	17	15	12	19	11	15	24	12	13	40	19	14	13, 15	29	13	12	22	36, 37	12	R1b-V88

Tabla Complementaria 4. Fuente de los datos utilizados en análisis de distancias genéticas por pares (*FST*) y filogeográficos (*Phyl*) de linajes africanos de ADNmt.

Región africana	País	grupo étnico	Familia lingüística	N
North	Marrocos [1]	General population*	Afro-Asiatic	56
	Marrocos [2]	Arab/Berber/Saharwi	Afro-Asiatic	164
	Marrocos [2]	General population*	Afro-Asiatic	344
	Tunisia [1]	General population*	Afro-Asiatic	64
East	Kenya, Nairobi [3]	General population*	unknown	100
	Rwanda [4]	General population*	unknown	153
	Uganda [5]	Nilotic	Nilo-Saharan	261
West-Central	Burkina Faso [6]	Tuareg	Afro-Asiatic	20
	Cameroon [7,8]	Tcheboua Fulani	Niger Congo	40
	Cameroon [4]	General population*	unknown	11
	Chad [6]	Kanembu	Nilo Saharan	50
	Ghana [9]	Akan	Niger-Congo	192
	Ghana [4]	General population*	Niger-Congo	22
	Guinea Bissau [10]	General population*	Niger-Congo	79
	Ivory Coast [4]	General population*	Niger-Congo	27
	Ivory Coast [4]	Several groups**	Niger-Congo	100
	Mali [6]	Tuareg	Afro-Asiatic	14
	Niger [6]	Tuareg	Afro-Asiatic	27
	Niger [4]	General population*	unknown	4
	Nigeria [11]	Esan	Niger-Congo	99
	Nigeria [11]	Yoruba	Niger-Congo	108
	Senegal [12]	Mandeka	Niger-Congo	88
Togo [4]	General population*	Niger-Congo	34	
Southwest	Angola [13]	Several groups**	Niger-Congo	365
	Angola [14]	Mbundo/Bakongo	Niger-Congo	42
	Angola, Cabinda [6]	General population*	Niger-Congo	105
Southeast	Mozambique [6]	Several groups**	Niger-Congo	147
South-Central	Zambia [15]	Several groups**	Niger-congo	169
	Zambia [16]	Several groups**	Niger-congo	446

Nota. * muestras de la población general sin indicación sobre la(s) afiliación(es) étnica(s); **incluye cuatro o más grupos étnicos de la misma familia lingüística.

Tabla Complementaria 5. Fuente de los datos utilizados en los análisis de distancias genéticas por pares (*FST*) y filogeográficos (*Phyl*) de los linajes del cromosoma Y africano.

African Region	Country	Ethnic group	Linguistic family	N	<i>FST</i>	<i>Phyl</i>
North	Algeria [1]	Mozabites	Afro-Asiatic	20		√
East	Djibouti [2]	Afar/Somali	Afro-Asiatic	54		√
	Eritrea [2]	Saho/Tigray/Tigre	Afro-Asiatic	127		√
	Eritrea [2]	Cunama/Nara	Nilo-Saharan	34		√
	Ethiopia [1,2]	Several groups**	Afro-Asiatic	188		√
	Kenya [2]	Borana/Somali	Afro-Asiatic	12		√
	Kenya [1,2]	Maasai/other	Nilo-Saharan	105		√
	Kenya [1,2]	Luhya/other	Niger-Congo	72		√
	Tanzania [3]	Several groups**	Niger-Congo	64		√
	Tanzania [3]	Datog	Nilo-Saharan	31		√
	Tanzania [3]	Burunge	Afro-Asiatic	23		√
	Tanzania [3]	Hadza/Sandawe/other	khoisan	121		√
	Uganda [4]	Karimojong	Nilo-Saharan	118		√
West-Central	Equatorial Guinea [5]	General population*	Niger-Congo	84	√	√
	Benín [6,7]	Several groups**	Niger-Congo	281	√	√
	Benín [6]	Natitingou, Dendi	Nilo-Saharan	37	√	√
	Benín [8]	General population*	unknown	1		√
	Benín [9]	North-Benín	unknown	51	√	
	Burkina Faso [1]	Several groups**	Niger-Congo	335	√	√
	Cameroon [10]	General population*	Niger-Congo	87	√	√
	Cameroon [11]	General population*	unknown	27	√	
	Cameroon [1]	General population*	unknown	70		√
	DRC [11]	General population*	unknown	7	√	
	DRC [1]	Several groups**	Niger-Congo	58	√	√
	CAR/DRC [1]	Pygmy (Biaka, Mbuti)	unknown	34		√
	Cameroon/Gabon [10]	Pygmy (Bakola)	Niger-Congo	60		√
Gabon [10]	Akele/Benga/Duma/other	Niger-Congo	733	√	√	

	Gambia [8]	General population*	unknown	3		√
	Ghana [8]	General population*	Niger-Congo	17		√
	Guinea [8]	General population*	Niger-Congo	1		√
	Guinea Bissau [12,13]	Several groups**	Niger-Congo	201	√	√
	Ivory Coast [8]	General population*	Niger-Congo	15		√
	Ivory Coast [11]	General population*	Niger-Congo	18	√	
	Ivory Coast [7]	Ahizi/Yacouba	Niger-Congo	90		√
	Liberia [8]	General population*	Niger-Congo	1		√
	Niger [8]	General population*	unknown	6		√
	Nigeria [8]	General population*	unknown	18		√
	Nigeria, Ibadan [9]	General population*	unknown	81	√	
	Nigeria [1,14]	Yoruba/Igbo	Niger-Congo	122	√	√
	Nigeria [14]	Hausa	Afro-Asiatic	4		√
	Senegal [8]	General population*	unknown	2		√
	Senegal [1]	Mandenka	Niger-Congo	15	√	√
	Sierra Leone [8]	General population*	Niger-Congo	1		√
	Togo [8]	General population*	Niger-Congo	19		√
Southwest	Angola [15]	Namibe Province	Niger-Congo	236	√	√
	Angola [11]	General population*	Niger-Congo	61	√	
	Angola, Cabinda [16]	General population*	Niger-Congo	75	√	√
Southeast	Mozambique [11]	General population*	Niger-Congo	35	√	
	Mozambique, Maputo [17]	General population*	Niger-Congo	78		√
South-Central	Zambia_East [1,18]	Several groups**	Niger-Congo	93	√	√
	Zambia_North [1]	Several groups**	Niger-Congo	37	√	√
	Zambia_South [1]	Several groups**	Niger-Congo	36	√	√
	Zambia_West [1]	Several groups**	Niger-Congo	377	√	√
	Zimbabwe [9]	General population*	unknown	55	√	
South	Botswana [1]	Kalanga/Tswana	Niger-Congo	39	√	√
	Namibia [1]	San	khoisan	5		√
	South Africa [3]	Khwe	khoisan	8		√
	South Africa [9]	General population*	unknown	113	√	
	SouthAfrica [1]	Bantu South	Niger-Congo	8	√	√

Nota: *muestras de la población general sin indicación sobre la(s) afiliación(es) étnica(s); **incluye cuatro o más grupos étnicos de la misma familia lingüística.

Tabla Complementaria 6. Valores de FST por pares entre Palenque (linajes africanos) y poblaciones africanas para las regiones HVI y HVII (debajo de la diagonal) y los valores p de no diferenciación correspondiente (arriba de la diagonal).

	Palenque	Marrocos	Tunisia	Senegal	Guiné Bissau	Cameroun	Togo	Ivory Coast	Ghana	Nigeria (Yoruba)	Nigeria (Esan)	Chad	Niger	Burkina Faso	Mali	Uganda	Kenya	Rwanda	Zambia	Angola (Cabin da)	Angola (Mbundo)	Angola (Namib desert)	Mozambique	
Palenque	*	<5E-06	<5E-06	<5E-06	0.0003	0.0002	0.00723	<5E-06	<5E-06	<5E-06	0.0002	0.00149	0.0694	0.0002	0.01178	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	0.00386	<5E-06	<5E-06
Marrocos	0.17969	*	0.0593	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	0.00139	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06
Tunisia	0.10586	0.00494	*	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	0.03584	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06
Senegal	0.09688	0.19346	0.1355	*	0.0001	<5E-06	0.00426	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	0.00139	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06
Guiné Bissau	0.03877	0.10177	0.04526	0.0361	*	0.09643	0.24483	0.28364	0.0199	0.12791	0.07752	0.02099	0.0198	0.0001	0.05722	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	0.03881	<5E-06	<5E-06
Cameroon	0.05792	0.10021	0.04992	0.06578	0.00852	*	0.00802	0.00634	0.0005	0.02831	0.02416	0.00119	0.00099	<5E-06	0.04079	0.0004	<5E-06	0.0003	0.0001	0.00059	0.00416	0.0002	<5E-06	<5E-06
Togo	0.04238	0.12465	0.06487	0.04281	0.00392	0.03092	*	0.62974	0.2183	0.31947	0.03495	0.25473	0.21295	0.00129	0.02505	<5E-06	<5E-06	0.0002	<5E-06	0.0002	0.01376	0.0002	<5E-06	<5E-06
Ivory Coast	0.04482	0.10201	0.04932	0.05232	0.00152	0.01948	-	*	0.06584	0.16969	0.00366	0.01752	0.04316	<5E-06	0.0295	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	0.00149	<5E-06	<5E-06
Ghana	0.0299	0.12201	0.06173	0.065	0.00975	0.03008	0.0044	0.00439	*	0.0293	0.00178	0.04336	0.22582	0.00089	0.03841	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	0.01861	<5E-06	<5E-06
Nigeria (Yoruba)	0.04318	0.10214	0.04681	0.06321	0.00453	0.0132	0.0019	0.00276	0.00682	*	0.32195	0.02297	0.0197	0.0001	0.0492	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	0.00604	<5E-06	<5E-06
Nigeria (Esan)	0.03912	0.10883	0.04804	0.0738	0.00703	0.01475	0.01717	0.01338	0.01366	0.00098	*	0.0097	0.01178	<5E-06	0.08653	<5E-06	<5E-06	0.0001	0.0001	0.00069	0.16464	0.0001	<5E-06	<5E-06
Chad	0.0387	0.13154	0.06331	0.07432	0.01632	0.03341	0.00442	0.01434	0.01078	0.01497	0.01808	*	0.41194	0.00218	0.06494	<5E-06	<5E-06	<5E-06	0.0001	<5E-06	0.03396	<5E-06	<5E-06	<5E-06
Niger	0.02136	0.14652	0.07566	0.10055	0.02321	0.05092	0.0072	0.01575	0.00448	0.02293	0.02739	-	*	0.01594	0.03346	0.0003	<5E-06	<5E-06	0.0003	0.0003	0.09068	0.0002	0.0001	<5E-06
Burkina Faso	0.11112	0.17113	0.12082	0.1505	0.07277	0.11104	0.06865	0.06835	0.06152	0.0724	0.07742	0.05014	0.04411	*	0.0396	0.0001	0.0001	0.0001	<5E-06	<5E-06	0.00069	0.0001	<5E-06	<5E-06
Mali	0.07714	0.06992	0.03207	0.10403	0.02788	0.03777	0.04802	0.0348	0.03538	0.03035	0.02329	0.02704	0.04689	0.05868	*	0.01535	0.01742	0.03168	0.02039	0.01485	0.06831	0.01733	0.00436	<5E-06
Uganda	0.08652	0.10652	0.0567	0.11513	0.03974	0.03279	0.05897	0.05381	0.06356	0.04426	0.03332	0.04691	0.05804	0.08287	0.04065	*	0.01762	0.01792	<5E-06	<5E-06	0.00158	<5E-06	<5E-06	<5E-06
Kenya	0.10751	0.10657	0.0561	0.14092	0.05302	0.04357	0.0686	0.0662	0.07562	0.05237	0.04002	0.05461	0.06816	0.0967	0.04011	0.00801	*	0.11524	<5E-06	<5E-06	0.0001	<5E-06	<5E-06	<5E-06
Rwanda	0.09955	0.11369	0.06191	0.12182	0.0434	0.03508	0.06108	0.05856	0.06903	0.04797	0.03544	0.0512	0.06577	0.08906	0.03685	0.00586	0.00371	*	<5E-06	<5E-06	0.00208	<5E-06	<5E-06	<5E-06

Zambia	0.05651	0.14036	0.0815	0.09705	0.03741	0.04051	0.0575	0.05352	0.04524	0.04553	0.02297	0.0445	0.04838	0.09062	0.04519	0.03956	0.04959	0.03826	*	0.05762	0.26611	0.0001	<SE-06
Angola (Cabinda)	0.07491	0.16634	0.08786	0.11922	0.04331	0.03825	0.07093	0.06724	0.05885	0.05398	0.02456	0.05564	0.06587	0.10856	0.05413	0.03895	0.04814	0.03623	0.00463	*	0.10385	0.05425	<SE-06
Angola (Mbundo)	0.03695	0.14931	0.07053	0.09182	0.01579	0.03156	0.03055	0.0288	0.01687	0.0231	0.00584	0.01823	0.01568	0.06969	0.03111	0.02834	0.04047	0.03006	0.00216	0.00962	*	0.08069	0.00515
Angola (Namib desert)	0.06651	0.14414	0.0801	0.10335	0.03753	0.0387	0.05988	0.05741	0.05172	0.04435	0.02124	0.04565	0.0563	0.09948	0.04676	0.03012	0.03802	0.02817	0.00814	0.00534	0.0086	*	<SE-06
Mozambique	0.10389	0.19169	0.11114	0.15794	0.07381	0.07924	0.09326	0.10101	0.09525	0.08166	0.05939	0.07737	0.07864	0.1272	0.08262	0.03653	0.04224	0.03685	0.03457	0.04053	0.03194	0.02068	*

Nota: Los valores significativos después de la corrección de Bonferroni se indican en rojo ($p < 0.0002$). s.e. ≤ 0.0055 . 10100 permutaciones.

Tabla Complementaria 7. Valores de FST por pares entre Palenque (linajes africanos) y poblaciones africanas para 10 Y-STRs (debajo de la diagonal) y los valores p de diferenciación correspondientes (arriba de la diagonal).

	Palenque	Angola	Namibe Desert	Benín_NS	Benín_NC	North Benín	Botswana	Burkina Faso	Cameron	DRC	Equatorial Guinea	Gabon	Guinea Bissau	Ivory Coast	Mozambique	Nigeria	Senegal	South Africa	Zambia - East	Zambia - North	Zambia - South	Zambia - West	Zimbabwe
Palenque	*	<SE-06	<SE-06	0.00099	<SE-06	0.00059	<SE-06	<SE-06	<SE-06	<SE-06	<SE-06	<SE-06	<SE-06	0.01673	0.02445	<SE-06	0.06118	<SE-06	<SE-06	0.00079	0.00010	<SE-06	<SE-06
Angola	0.05179	*	<SE-06	0.04851	0.00277	0.06098	0.0093	<SE-06	<SE-06	0.0276	0.01505	0.03643	<SE-06	0.14692	0.01554	<SE-06	0.17434	<SE-06	0.0132	0.22661	0.05277	<SE-06	0.0133
Namibe Desert	0.07486	0.02529	*	<SE-06	<SE-06	<SE-06	<SE-06	<SE-06	<SE-06	0.0001	<SE-06	<SE-06	<SE-06	0.01089	0.00020	<SE-06	0.00812	<SE-06	<SE-06	0.42887	0.38768	0.0232	<SE-06
Benín_NS	0.04946	0.01461	0.04889	*	0.19355	0.80784	0.0090	0.01673	0.01445	0.1869	0.00861	0.05534	<SE-06	0.51609	0.09752	0.0692	0.11217	<SE-06	0.1226	0.03970	0.01208	0.0005	0.2547
Benín_NC	0.04559	0.01716	0.05191	0.00625	*	0.99772	0.0088	0.00198	0.00030	0.3033	0.00099	0.04089	<SE-06	0.41580	0.14038	0.5149	0.28116	<SE-06	0.0723	0.00495	0.01247	<SE-06	0.6283
North Benín	0.03307	0.00898	0.04655	0.00835	0.01150	*	0.0331	0.04683	0.01238	0.4404	0.03020	0.23839	0.0003	0.70241	0.36759	0.2294	0.40204	<SE-06	0.1437	0.01663	0.02208	0.0002	0.5963

Botswana	0.05646	0.02181	0.06000	0.03433	0.02415	0.01801	*	0.00030	0.00139	0.00644	0.00921	0.00812	0.00020	0.37244	0.04455	0.00010	0.20414	0.00149	0.00238	0.00871	0.00802	0.00030	0.00624
BurkinaFaso	0.06175	0.02241	0.05130	0.01768	0.01339	0.00796	0.03025	*	<5E-06	0.01653	<5E-06	<5E-06	<5E-06	0.54054	0.02723	0.00020	0.19355	<5E-06	0.00307	0.00416	0.00069	<5E-06	0.00446
Cameroon	0.08979	0.03304	0.10967	0.02444	0.02431	0.01691	0.03549	0.03568	*	0.00019	0.08138	<5E-06	<5E-06	0.02990	<5E-06	<5E-06	0.12167	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	0.00010
DRC	0.05841	0.01329	0.03440	0.00756	0.00223	0.00004	0.03269	0.01192	0.04222	*	0.00228	0.04326	0.00000	0.50282	0.04198	0.03346	0.07088	<5E-06	0.70369	0.04792	0.02980	0.00317	0.32274
Equatorial Guinea	0.06783	0.01227	0.07705	0.02703	0.02107	0.01324	0.02440	0.02606	0.00666	0.02820	*	<5E-06	<5E-06	0.05306	0.00347	<5E-06	0.24522	<5E-06	0.00020	0.00079	<5E-06	<5E-06	0.00020
Gabon	0.05689	0.00373	0.02911	0.01131	0.00556	0.00238	0.01726	0.01563	0.02994	0.00832	0.02058	*	<5E-06	0.36274	0.03039	0.00030	0.29789	<5E-06	0.02396	0.07019	0.03792	<5E-06	0.05732
GuineaBissau	0.05363	0.04147	0.08373	0.04088	0.03109	0.02450	0.03667	0.02177	0.04014	0.04220	0.02617	0.04033	*	0.03168	0.01970	<5E-06	0.64221	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06
Ivory Coast	0.04188	0.01273	0.03929	-0.00291	0.00064	-0.00815	0.0024	-0.00327	0.03141	-0.00246	0.02341	0.00188	0.02229	*	0.60202	0.15929	0.29185	0.01089	0.43045	0.12524	0.07623	0.03821	0.50569
Mozambique	0.01988	0.02063	0.03991	0.01365	0.00782	0.00169	0.01982	0.01368	0.05253	0.01785	0.03108	0.01248	0.01430	-0.00531	*	0.02594	0.43362	0.00069	0.00723	0.10098	0.04099	0.00069	0.02030
Nigeria	0.06964	0.02692	0.06198	0.01202	-0.00082	0.00333	0.04221	0.01504	0.03329	0.01181	0.03208	0.01244	0.03061	0.01135	0.01646	*	0.24285	<5E-06	0.03336	0.00198	0.00079	<5E-06	0.22701
Senegal	0.02944	0.01247	0.05277	0.02232	0.00648	0.00146	0.01279	0.01007	0.01750	0.02611	0.00777	0.00470	-0.00589	0.00992	0.00031	0.00844	*	0.01099	0.06128	0.27730	0.21146	0.03722	0.02940
southAfrica	0.11192	0.07965	0.10929	0.09494	0.06470	0.06631	0.04829	0.06783	0.07689	0.07245	0.05943	0.06142	0.07777	0.05217	0.05132	0.08880	0.06060	*	<5E-06	<5E-06	<5E-06	<5E-06	<5E-06
Zambia_East	0.07829	0.01099	0.03332	0.00917	0.00657	0.00579	0.03270	0.01045	0.04010	-0.00359	0.02869	0.00574	0.04751	0.00008	0.02428	0.00825	0.02535	0.07731	*	0.06059	0.02584	0.00030	0.63875
Zambia_North	0.04339	0.00421	-0.00012	0.02303	0.02843	0.02306	0.03330	0.02254	0.07032	0.01756	0.04263	0.00898	0.04365	0.01763	0.01327	0.03071	0.00759	0.08895	0.01220	*	0.68488	0.78873	0.01139

Zambia_South	0.05305	0.01341	0.00036	0.03687	0.02555	0.02384	0.0383	0.03149	0.07654	0.0227	0.05188	0.01205	0.0555	0.02764	0.02215	0.0352	0.01302	0.09206	0.0186	-	*	0.7875	0.0124
						1				9		8			0			8		0.00673		5	7
Zambia_West	0.07005	0.01982	0.00377	0.03762	0.03588	0.03264	0.0429	0.03667	0.09195	0.0212	0.06654	0.01858	0.0717	0.02565	0.03226	0.0411	0.03121	0.10020	0.0177	-	-	*	0.0002
						6				4		8			9			4		0.00501	0.00518		0
Zimbabwe	0.07098	0.01602	0.04564	0.00471	-	-	0.0323	0.01554	0.03783	0.0020	0.03568	0.00663	0.0517	-	0.02321	0.0033	0.03943	0.08844	-	0.02831	0.03057	0.0295	*
					0.00283	0.00254	9			3		4		0.00269		6			0.0027			9	

Tabla Complementaria 8. *Frecuencias alélicas, heterocigocidad esperada y observada y equilibrio de HW en 10 poblaciones colombianas.*

Locus	Andes (n=1293)			Caribe (n=153)			Pacífico (n=303)			Orinoquia (n=83)			Amazonia (n=34)			Barrizón (n=52)			Nubres_Nariño (n=21)			Nubres_Guainía (n=16)			Pisao (n=26)			Páramo (n=39)		
	FD	FE	TIH	FD	FE	TIH	FD	FE	TIH	FD	FE	TIH	FD	FE	TIH	FD	FE	TIH	FD	FE	TIH	FD	FE	TIH	FD	FE	TIH	FD	FE	TIH
B1/mid2719	0.622	0.011	0.965	0.575	0.140	0.948	0.391	0.192	1.010	0.593	0.256	1.162	0.621	0.121	0.950	0.594	0.094	0.768	0.623	0.041	0.656	0.625	0.039	1.000	0.624	0.223	1.003	0.617	0.280	1.100
B2/mid1805	0.622	0.199	1.005	0.622	0.212	1.058	0.391	0.175	0.971	0.680	0.178	0.977	0.639	0.188	1.000	0.200	0.011	0.955	0.599	0.132	0.875	0.563	0.248	1.143	0.624	0.223	1.003	0.561	0.299	0.759
B3/mid1498	0.616	0.015	0.949	0.624	0.199	1.027	0.391	0.213	1.059	0.600	0.168	0.955	0.628	0.188	1.000	0.574	0.094	0.789	0.624	0.379	1.500	0.621	0.522	1.358	0.590	0.188	1.000	0.625	0.383	0.950
B4/mid2946	0.624	0.015	0.950	0.599	0.102	0.906	0.497	0.097	0.794	0.697	0.125	0.980	0.595	0.141	0.895	0.450	0.399	0.670	0.474	0.103	0.808	0.395	0.245	0.657	0.598	0.188	1.000	0.462	0.074	0.739
B5/mid1504	0.623	0.173	0.957	0.590	0.106	0.915	0.390	0.149	0.913	0.590	0.125	0.980	0.539	0.141	0.895	0.107	0.003	0.951	0.599	0.132	0.875	0.615	0.248	1.143	0.595	0.156	0.929	0.957	0.090	0.637
B6/mid500	0.534	0.096	0.782	0.595	0.136	0.952	0.447	0.097	0.794	0.697	0.125	0.980	0.495	0.088	0.773	0.522	0.028	0.619	0.400	0.027	0.618	0.297	0.026	0.615	0.395	0.028	0.619	0.618	0.883	0.930
B7/mid2005	0.625	0.152	0.950	0.625	0.165	0.946	0.462	0.110	0.924	0.395	0.125	0.980	0.621	0.277	1.214	0.617	0.243	1.190	0.615	0.209	1.050	0.616	0.240	1.143	0.615	0.254	1.102	0.620	0.123	0.859
B8/mid3221	0.623	0.190	1.005	0.619	0.140	0.938	0.404	0.172	0.985	0.429	0.249	1.049	0.621	0.163	0.944	0.590	0.223	1.093	0.610	0.259	1.167	0.621	0.099	0.600	0.685	0.361	1.444	0.622	0.211	1.059
B9/mid2698	0.596	0.134	0.878	0.624	0.212	1.058	0.396	0.210	1.052	0.434	0.159	1.034	0.432	0.050	0.680	0.595	0.015	0.813	0.440	0.078	0.750	0.214	0.013	0.571	0.503	0.028	0.619	0.625	0.143	0.900
B10/mid1116	0.574	0.132	0.875	0.547	0.131	0.877	0.415	0.102	0.907	0.434	0.138	0.943	0.595	0.141	0.895	0.605	0.129	0.857	0.591	0.209	1.050	0.620	0.248	1.143	0.598	0.129	0.867	0.495	0.091	0.728
G1/mid785	0.623	0.100	0.965	0.597	0.136	0.982	0.348	0.155	0.882	0.484	0.235	1.265	0.625	0.244	1.133	0.624	0.205	1.040	0.591	0.132	0.875	0.621	0.088	1.000	0.619	0.416	1.625	0.623	0.211	1.059
G2/mid1151	0.615	0.166	0.951	0.495	0.072	0.735	0.399	0.140	0.902	0.376	0.188	1.000	0.621	0.163	0.944	0.605	0.156	0.929	0.624	0.359	1.500	0.624	0.248	1.143	0.598	0.129	0.867	0.523	0.116	0.889
G3/mid1109	0.602	0.159	0.935	0.545	0.127	0.962	0.522	0.052	0.694	0.398	0.159	0.935	0.625	0.163	0.944	0.595	0.094	0.788	0.615	0.209	1.050	0.616	0.248	1.143	0.598	0.067	0.722	0.448	0.095	0.717
G4/mid1514	0.547	0.087	0.771	0.510	0.091	0.781	0.407	0.213	1.059	0.440	0.104	0.811	0.432	0.074	0.739	0.673	0.001	0.520	0.440	0.041	0.656	0.116	0.003	0.953	0.539	0.129	0.867	0.620	0.123	0.889
G5/mid2050	0.489	0.071	0.732	0.597	0.185	0.926	0.406	0.131	0.871	0.497	0.092	0.782	0.591	0.088	0.773	0.620	0.142	0.897	0.593	0.132	0.875	0.470	0.045	0.657	0.477	0.067	0.722	0.495	0.078	0.750
G6/mid1394	0.622	0.161	0.959	0.595	0.095	0.789	0.408	0.180	0.984	0.409	0.244	1.132	0.621	0.061	0.708	0.620	0.205	1.040	0.602	0.209	1.050	0.624	0.248	1.143	0.622	0.052	0.684	0.415	0.049	0.716
G7/mid743	0.600	0.131	0.871	0.575	0.114	0.833	0.397	0.216	1.066	0.391	0.178	0.977	0.625	0.188	1.000	0.602	0.142	0.897	0.598	0.168	0.955	0.582	0.322	1.383	0.590	0.165	0.813	0.625	0.241	1.128
G8/mid3097	0.625	0.191	1.009	0.616	0.198	1.034	0.488	0.275	1.210	0.374	0.178	0.977	0.621	0.121	0.890	0.580	0.129	0.857	0.602	0.132	0.875	0.508	0.069	0.727	0.619	0.156	0.929	0.574	0.123	0.889
G9/mid1782	0.539	0.092	0.782	0.513	0.110	0.834	0.465	0.164	0.947	0.459	0.092	0.782	0.605	0.141	0.885	0.395	0.052	0.684	0.593	0.132	0.875	0.598	0.138	0.889	0.580	0.156	0.929	0.615	0.143	0.900
Y1/mid2648	0.591	0.137	0.895	0.621	0.131	0.921	0.461	0.095	0.790	0.420	0.159	0.935	0.388	0.031	0.630	0.511	0.075	0.743	0.474	0.089	0.700	0.563	0.138	0.889	0.521	0.105	0.813	0.528	0.136	0.884
Y2/mid2890	0.542	0.102	0.835	0.575	0.240	1.172	0.479	0.102	0.807	0.429	0.159	0.935	0.562	0.141	0.885	0.474	0.039	0.680	0.474	0.089	0.700	0.597	0.138	0.889	0.605	0.084	0.765	0.593	0.116	0.889
Y3/mid3277	0.578	0.120	0.847	0.609	0.238	1.071	0.434	0.118	0.943	0.386	0.159	0.935	0.562	0.188	1.000	0.547	0.116	0.889	0.602	0.132	0.875	0.625	0.010	2.000	0.539	0.084	0.765	0.580	0.081	0.707
Y4/mid1110	0.521	0.083	0.762	0.597	0.127	0.962	0.501	0.068	0.726	0.475	0.125	0.980	0.510	0.074	0.739	0.594	0.142	0.897	0.528	0.103	0.808	0.423	0.069	0.727	0.440	0.052	0.684	0.178	0.091	0.585
Y5/mid1983	0.596	0.120	0.847	0.606	0.136	0.982	0.392	0.210	1.052	0.478	0.178	0.977	0.380	0.050	0.680	0.140	0.001	0.520	0.169	0.009	0.593	0.600	0.000	0.930	0.346	0.039	0.650	0.504	0.175	0.971
Y6/mid1604	0.599	0.145	0.903	0.571	0.122	0.952	0.431	0.167	0.953	0.413	0.168	0.956	0.625	0.131	0.308	0.477	0.067	0.722	0.620	0.259	1.167	0.621	0.088	1.000	0.611	0.156	0.929	0.541	0.116	0.889
Y7/mid3814	0.589	0.145	0.904	0.624	0.282	1.103	0.429	0.124	0.957	0.391	0.138	0.943	0.494	0.088	0.773	0.623	0.171	0.953	0.591	0.132	0.875	0.366	0.045	0.657	0.598	0.156	0.929	0.595	0.166	0.952
Y8/mid894	0.615	0.142	0.897	0.575	0.171	0.992	0.401	0.201	1.030	0.466	0.138	0.943	0.580	0.074	0.739	0.622	0.142	0.897	0.528	0.103	0.808	0.508	0.138	0.889	0.622	0.188	1.000	0.614	0.123	0.889
Y9/mid1643	0.522	0.081	0.757	0.596	0.140	0.959	0.526	0.054	0.690	0.469	0.104	0.811	0.494	0.088	0.773	0.107	0.003	0.951	0.400	0.009	0.593	0.257	0.026	0.625	0.493	0.084	0.765	0.452	0.057	0.697
R0/mid1945	0.592	0.011	0.893	0.613	0.182	0.989	0.378	0.186	0.957	0.463	0.068	0.769	0.591	0.088	0.773	0.144	0.001	0.920	0.440	0.026	0.698	0.366	0.013	0.571	0.395	0.135	0.813	0.625	0.136	0.884
R1/mid2592	0.600	0.142	0.897	0.623	0.140	0.991	0.419	0.257	1.154	0.379	0.111	0.827	0.621	0.277	1.214	0.624	0.105	0.813	0.595	0.093	0.700	0.625	0.322	1.328	0.624	0.310	1.300	0.614	0.303	0.930
R2/mid649	0.621	0.017	1.021	0.624	0.150	0.935	0.447	0.047	0.698	0.383	0.138	0.943	0.639	0.074	0.739	0.594	0.171	0.953	0.602	0.169	0.955	0.610	0.138	0.889	0.625	0.510	1.300	0.606	0.383	0.950
R3/mid1880	0.611	0.172	0.944	0.614	0.160	0.959	0.357	0.057	0.646	0.405	0.220	1.075	0.625	0.104	0.810	0.624	0.264	1.182	0.602	0.132	0.875	0.597	0.138	0.889	0.622	0.129	0.867	0.415	0.091	0.678
R4/mid1330	0.594	0.134	0.879	0.625	0.218	1.071	0.407	0.144	0.902	0.405	0.159	0.935	0.623	0.141	0.895	0.449	0.075	0.743	0.549	0.078	0.750	0.597	0.248	1.143	0.598	0.129	0.867	0.595	0.183	0.950
R5/mid3220	0.522	0.096	0.789	0.612	0.199	1.027	0.404	0.172	0.985	0.516	0.075	0.741	0.494	0.088	0.773	0.073	0.001	0.520	0.474	0.103	0.808	0.116	0.003	0.953	0.346	0.039	0.650	0.615	0.158	0.934
R6/mid1997	0.625	0.209	1.006	0.596	0.131	0.921	0.385	0.104	1.037	0.377	0.125	0.980	0.595	0.141	0.895	0.140	0.005	0.942	0.596	0.108	0.808	0.621	0.322	1.383	0.580	0.105	0.813	0.621	0.252	1.116
R7/mid3081	0.659	0.011	0.804	0.625	0.171	0.962	0.464	0.090	0.755	0.400	0.138	0.976	0.562	0.031	0.630	0.624	0.156	0.929	0.581	0.078	0.750	0.214	0.013	0.571	0.483	0.084	0.765	0.564	0.091	0.765
R8/mid1579	0.505	0.071	0.728	0.624	0.246	1.135	0.373	0.183	0.990	0.489	0.085	0.785	0.475	0.074	0.739	0.073	0.001	0.920	0.474	0.089	0.700	0.366	0.013	0.571	0.302	0.028	0.619	0.603	0.136	0.884
R9/mid2895	0.638	0.184	0.958	0.595	0.171	0.962	0.400	0.210	1.052	0.443	0.288	1.229	0.619																	

Tabla Complementaria 9. Tabla Complementaria 9. Distancias genéticas por pares (FST) entre las poblaciones colombianas, europeas y africanas (diagonal inferior) y los valores P de no diferenciación correspondientes (diagonal superior).

	Andes								Caribe			Pacific	Orinoquía	Amazonia	Native Groups			Atrodesc				
	Antioquia	Cundinamarca	Norte de Santander	Santander	Boyacá	Tolima	Valle de Cauca	Huila	Nariño	Guajira	Cesar	Magdalena	Bolivar	Chocó	Araucario	Meta	Putumayo	Barí-	Nariño	Guainaque	Pijaño	Palestina
			<5E+06	<5E+06	<5E+06	0,03117	0,0738	<5E+06	<5E+06	0,49723	0,77342	0,11828	<5E+06	<5E+06	0,00147	0,00175	<5E+06	<5E+06	<5E+06	<5E+06	<5E+06	<5E+06
Antioquia	*	<5E+06								0,03723	0,77342	0,25028	<5E+06	<5E+06	0,00041	<5E+06	<5E+06	<5E+06	<5E+06	0,00000	<5E+06	<5E+06
Cundinamarca	0,0031	*	0,22007	<5E+06	0,01335	0,02372	0,13704	0,02212	<5E+06	0,03055	0,77848	0,25082	<5E+06	<5E+06	0,00012	<5E+06	<5E+06	<5E+06	<5E+06	0,00000	<5E+06	<5E+06
Norte de Santander	0,00332	0,00032	*	0,00006	0,01963	0,03619	0,09672	0,0014	<5E+06	0,26132	0,94115	0,1543	<5E+06	<5E+06	0,00036	<5E+06	<5E+06	<5E+06	<5E+06	0,00000	<5E+06	<5E+06
Santander	0,00371	0,00209	0,00225	*	<5E+06	0,00604	0,43188	<5E+06	<5E+06	0,15306	0,91663	0,24743	<5E+06	<5E+06	0,01026	<5E+06	<5E+06	<5E+06	<5E+06	0,00000	<5E+06	<5E+06
Boyacá	0,00439	0,001148	0,0027	0,00358	*	0,00303	0,00498	0,26971	<5E+06	0,28045	0,83408	0,04054	<5E+06	<5E+06	0,20006	<5E+06	<5E+06	<5E+06	<5E+06	0,00000	<5E+06	<5E+06
Tolima	0,00254	0,00232	0,00231	0,00335	0,00444	*	0,07131	0,00128	<5E+06	0,08683	0,81061	0,15685	<5E+06	<5E+06	0,02000	<5E+06	<5E+06	<5E+06	<5E+06	0,00000	<5E+06	<5E+06
Valle de Cauca	0,0028	0,00169	0,00224	0,00446	0,0057	0,00369	*	0,00811	0,00337	0,11284	0,84788	0,58562	<5E+06	<5E+06	0,02000	<5E+06	<5E+06	<5E+06	<5E+06	0,10010	<5E+06	<5E+06
Huila	0,00523	0,00133	0,00272	0,00472	0,00042	0,00485	0,00501	*	<5E+06	0,04719	0,54301	0,04319	<5E+06	<5E+06	0,00017	<5E+06	<5E+06	<5E+06	<5E+06	0,00000	<5E+06	<5E+06
Nariño	0,01473	0,00639	0,01027	0,00868	0,00904	0,01188	0,00768	0,00961	*	0,00078	0,04658	0,00038	<5E+06	<5E+06	0,00009	<5E+06	<5E+06	<5E+06	<5E+06	0,01001	<5E+06	<5E+06

Guajira	0,000 04	0,00635	0,001 7	0,003 99	0,00 156	0,00 518	0,00 578	0,00 574	0,01 409	*	0,92 574	0,174 33	0,69 888	0,00 004	0,72 682	0,12 438	0,000 04	<SE+ 6	<SE +06	<SE +06	0,00 002	<SE+ 06	
Cesar	0,002 9	0,00326	0,006 18	0,004 66	0,00 452	0,00 472	0,00 537	0,00 13	0,00 684	0,01 026	*	0,814 79	0,88 12	<SE +06	0,76 676	0,57 011	0,084	<SE+ 6	0,00 566	0,00 002	0,12 988	<SE+ 06	
Magdalena	0,002 9	0,00124	0,002 13	0,001 76	0,00 431	0,00 269	0,00 087	0,00 413	0,01 104	0,00 417	0,00 6	*	0,04 42	<SE +06	0,02 063	0,02 212	<SE+ 06	<SE+ 6	0,00 002	<SE +06	0,00 068	<SE+ 06	
Bolivar	0,005 8	0,00953	0,007 05	0,008 74	0,00 811	0,01 115	0,00 672	0,01 153	0,02 117	0,00 238	0,00 547	0,004 *	<SE +06	0,00 498	0,00 008	<SE+ 06	<SE+ 6	<SE +06	<SE +06	<SE +06	<SE +06	<SE+ 06	
Chocó	0,031 17	0,04923	0,046 03	0,044 26	0,04 514	0,04 559	0,04 066	0,05 198	0,06 012	0,01 788	0,02 588	0,038 36	0,01 883	<SE +06	<SE +06	<SE+ 06	<SE+ 6	<SE +06	<SE +06	<SE +06	<SE +06	<SE+ 06	
Arauca	0,005 14	0,00348	0,003 95	0,003 54	0,00 1	0,00 415	0,00 491	0,00 407	0,00 555	0,00 4	0,00 54	0,006 24	0,00 496	0,03 266	*	0,21 056	0,000 22	<SE+ 6	0,00 014	<SE +06	0,00 046	<SE+ 06	
Meta	0,007 21	0,00027	0,000 26	0,004 93	0,00 034	0,00 462	0,00 995	0,00 187	0,00 837	0,00 608	0,00 103	0,008 94	0,01 212	0,05 504	0,00 177	*	0,001 02	<SE+ 6	0,00 026	0,00 022	0,01 291	<SE+ 06	
Putumayo	0,026 44	0,01497	0,016 41	0,018 29	0,02 029	0,01 993	0,01 711	0,01 778	0,00 639	0,02 453	0,00 845	0,021 69	0,02 082	0,06 149	0,01 257	0,01 465	*	<SE+ 6	0,18 236	0,00 245	0,08 06	<SE+ 06	
Barí-	0,112 74	0,09616	0,100 44	0,096 22	0,10 141	0,11 707	0,10 605	0,10 203	0,08 828	0,12 217	0,10 5	0,109 568	0,10 603	0,15 66	0,09 267	0,11 14	0,088	*	<SE +06	<SE +06	<SE +06	<SE+ 06	
Nativos_N ariño	0,035 8	0,02065	0,028 53	0,026 74	0,02 352	0,03 294	0,02 297	0,02 102	0,00 437	0,03 564	0,02 008	0,026 03	0,03 758	0,07 87	0,01 801	0,02 08	0,004 82	0,0742 5	*	0,06 523	0,03 426	<SE+ 06	
Nativos_G uainia	0,053 76	0,03785	0,038 7	0,043 24	0,03 125	0,04 849	0,04 295	0,03 621	0,02 824	0,04 877	0,03 658	0,050 86	0,05 185	0,09 942	0,03 602	0,02 569	0,017 16	0,0978 0,0771	0,00 991	*	0,00 058	<SE+ 06	
Pijao	0,019 08	0,00828	0,010 32	0,011 62	0,01 3	0,01 609	0,00 513	0,00 191	0,00 658	0,01 371	0,02 635	0,00 47	0,006 175	0,06 69	0,01 293	0,01 078	0,006 64	0,0771 8	0,01 059	0,02 *	<SE+ 06		
Palenque	0,059 67	0,07816	0,073 44	0,072 29	0,07 646	0,07 718	0,06 826	0,08 277	0,09 182	0,04 358	0,05 897	0,066 64	0,03 452	0,01 408	0,05 873	0,08 717	0,1843 8	0,1843 8	0,10 982	0,12 885	0,09 826	*	
																		0,086 55					*

En rojo están los valores p de F_{ST} significativos (nivel de significación = 0,00022, después de aplicar la corrección de Bonferroni), después de 10100 permutaciones (s.e. <0,0023).