

Identificación de linajes maternos presentes en individuos nacidos en la Región Andina colombiana. Una aproximación a su uso en Genética Forense

Adriana Castillo Pico

Tesis para optar al título de Doctora en Ciencias Biomédicas

Director

Fernando Rondón González

Doctor en Ciencias Biología

Codirectora

Leonor Gusmão

Doctora en Biología

Universidad Industrial de Santander

Facultad de Salud

Escuela de Medicina

Posgrado en Ciencias Biomédicas

Bucaramanga

2023

Dedicatoria

A mis padres que me enseñaron
que con amor, disciplina y dedicación
se logra todo lo que uno se proponga.
A Felipito que siempre ha confiado en mis
capacidades y me ha apoyado en todo
impulsándome a asumir los retos académicos
y profesionales que Dios ha puesto en mi camino.
Y a mi princesita Marianita, el motor que impulsa mi vida.

Agradecimientos

A Dios que siempre me guía y me ha permitido alcanzar todas las metas que me he propuesto en mi vida.

A mi director de tesis, doctor Fernando Rondón G. y a mi codirectora, la doctora Leonor Gusmão, por todas sus enseñanzas, paciencia, dedicación y esmero con que me dirigieron durante estos años para llevar a feliz término este proyecto.

A la doctora Filipa Simão por su acompañamiento y enseñanzas durante el desarrollo del trabajo de laboratorio y por compartir conmigo su amplia experiencia de trabajo con el ADN mitocondrial.

Al doctor Humberto Ossa que de manera generosa e invaluable aceptar participar en este proyecto.

A todos los miembros del laboratorio LDD de la UERJ y del laboratorio de Genética UIS por toda su colaboración para llevar a cabo el procesamiento y análisis de las muestras.

A todas las personas que participaron aportando sus muestras porque gracias a ellas podemos conocer nuestra genética y determinar nuestra ancestría materna.

A mis compañeros del departamento de Ciencias Básicas, a mis familiares, a mis amigos y a todas las personas que de una u otra manera han colaborado para llevar a cabo este ambicioso proyecto.

Tabla de Contenido

Introducción	17
1. Justificación.....	23
2. Hipótesis	26
3. Objetivos	27
3.1 Objetivo general	27
3.2 Objetivos específicos.....	27
4. Marco Teórico	28
4.1 Historia del poblamiento de la región andina colombiana.....	28
4.1.1 Las regiones de Colombia	28
4.1.2 La Región Andina	34
4.1.3 El departamento de Santander.....	39
4.2 Desaparición forzada y proceso de paz en Colombia	42
4.3 La prueba de ADN en la administración de justicia.....	49
4.4 Marcadores genéticos uniparentales.....	51
4.5 ADN mitocondrial (ADNmt)	53
4.5.1 Estructura	55
4.5.2 Características	56
4.5.2.1 Número de moléculas.....	57
4.5.2.2 Resistencia a la degradación..	57
4.5.2.3 Tasa de mutación.....	57
4.5.2.4 Herencia.	58
4.5.2.5 Heteroplasmia.....	59

4.5.3 Nomenclatura	61
4.5.4 Análisis del ADNmt en genética forense	62
4.5.4.1 Métodos de análisis..	63
4.5.4.2 Secuencia de referencia..	66
4.5.4.3 Interpretación de resultados.	66
4.5.4.4 Valoración estadística de resultados.	67
4.5.4.5 Bases de datos de ADNmt.....	69
4.5.5 Ancestría de los linajes maternos	72
4.6 Revisión sistemática de alcance	75
4.6.1 Información del autor	77
4.6.2 Título, objetivo y pregunta	77
4.6.3 Antecedentes	78
4.6.4 Criterios de inclusión	78
4.6.5 Estrategia de búsqueda	79
4.6.6 Extracción de los resultados	81
4.6.7 Presentación de los resultados.....	82
5. Materiales y métodos.....	83
5.1 Revisión de alcance	83
5.1.1 Criterios de inclusión:	83
5.1.2 Estrategia de búsqueda	84
5.1.2.1 Fuentes de información. Se exploraron las siguientes fuentes de información:.....	85
5.1.2.2 Selección de los estudios. Todas las citas de los estudios identificados con la.....	86
5.1.3 Extracción de datos	86
5.1.4 Presentación de datos	86

5.2 Trabajo experimental.....	87
5.2.1 Muestras	87
5.2.2 Extracción de ADN	90
5.2.3 Cuantificación de ADN	92
5.2.4 Genotipificación del ADNmt	93
5.2.4.1 Amplificación de la RC de ADNmt.	93
5.2.4.2 Verificación del producto amplificado.....	94
5.2.4.3 Purificación de productos amplificados.. ..	95
5.2.4.4 Secuenciación de la RC del ADNmt.	95
5.2.4.5 Purificación del producto secuenciado.....	97
5.2.5 Detección de las secuencias por Electroforesis Capilar (EC)	98
5.2.6 Análisis de secuencias.....	99
5.2.7 Asignación de haplotipos	103
5.2.8 Asignación de haplogrupos	104
5.3 Análisis estadístico.....	105
5.3.1 Análisis estadísticos y cálculo de parámetros de importancia forense.....	105
5.3.2 Construcción de redes	106
6. Resultados y Discusión	107
6.1 Revisión de alcance	107
6.2 Linajes de ADNmt en el departamento de Santander	111
6.2.1 Haplotipos hallados	112
6.2.2 Diversidad genética poblacional a partir de los haplotipos.....	114
6.2.3 Linajes maternos en Santander.....	119
6.2.4 Haplogrupos nativos americanos en Santander.....	124

6.2.5 Evaluación de la subestructura poblacional en Santander	127
6.2.5.1 Diferenciación genética entre provincias.	127
6.2.5.2 Análisis de Componentes Principales (PCA).....	131
6.2.5.3 Análisis de Varianza Molecular (AMOVA).	133
6.3 Linajes de ADNmt en los departamentos de la región andina	135
6.3.1 Haplotipos hallados	135
6.3.2 Diversidad genética poblacional a partir de los haplotipos.....	138
6.3.3 Linajes Maternos en la Región Andina	143
6.3.4 Linajes nativos americanos en la región andina.....	146
6.3.5 Evaluación de la subestructura poblacional en los departamentos de la región andina	152
6.3.5.1 Diferenciación genética entre departamentos.	152
6.3.5.2 Análisis de componentes principales (PCA) a partir de haplogrupos nativos.	157
6.3.5.3 PCA a partir de la frecuencia de haplogrupos integrando Antioquia 10000 genomas. ...	163
7. Conclusiones	166
8. Recomendaciones	169
Referencias Bibliográficas	170
Apéndices	181

Lista de Tablas

Tabla 1. Registro de datos extraídos	86
Tabla 2. Muestras tomadas de las provincias de Santander	89
Tabla 3. Protocolo para el montaje de la PCR	93
Tabla 4. Programa de Amplificación de la RC del ADNmt.....	93
Tabla 5. Protocolo para el montaje de la purificación de los productos de PCR.	95
Tabla 6. Protocolo para secuenciación de la RC del ADNmt	95
Tabla 7. Programa de secuenciamiento de la RC del ADNmt.	96
Tabla 8. Primers usados para secuenciar la RC del ADNmt.....	97
Tabla 9. Aplicaciones de los estudios seleccionados	108
Tabla 10. Región del ADNmt analizada	109
Tabla 11. Metodología de genotipado.....	109
Tabla 12. Heteroplasmias de posición en muestras de Santander.....	112
Tabla 13. Parámetros de diversidad genética y forenses estimados para el departamento de Santander.....	114
Tabla 14. Parámetros de diversidad genética y forenses para poblaciones de Suramérica.....	116
Tabla 15. Frecuencia de los haplogrupos mitocondriales en el departamento de Santander	119
Tabla 16. Frecuencias de haplogrupos mitocondriales en poblaciones de Suramérica	121
Tabla 17. Valores de F_{ST} por pares entre provincias de Santander a partir de haplotipos	127
Tabla 18. Valores de F_{ST} por pares entre provincias de Santander a partir de haplotipos	129
Tabla 19. Valores de F_{ST} por pares entre provincias de Santander a partir de haplogrupos	129
Tabla 20. Valores de F_{ST} por pares entre provincias de Santander a partir de los macro-haplogrupos.	130
Tabla 21. Resultados de AMOVA para los tres agrupamientos evaluados	134

Tabla 22. Heteroplasmas de posición detectadas.....	136
Tabla 23. Parámetros de diversidad genética calculados para los departamentos de la Región Andina.....	138
Tabla 24. Frecuencias publicadas para los haplogrupos, en departamentos de la región andina.	145
Tabla 25. Valores de F_{ST} por pares entre los departamentos de la región andina.....	152
Tabla 26. Valores de F_{ST} por pares para los haplogrupos nativos entre los departamentos de la región andina.....	155
Tabla 27. Valores de F_{ST} por pares entre los departamentos de la región andina a partir de RC sin Indel.....	162

Lista de Figuras

Figura 1. Mapa de las regiones naturales de Colombia.....	28
Figura 2. Mapa del relieve colombiano.....	31
Figura 3. Departamentos que conforman de la región andina colombiana	34
Figura 4. Mapa en relieve de Santander.....	39
Figura 5. Mapa de Santander con sus provincias y municipios	40
Figura 6. Eventos victimizantes según el hecho	47
Figura 7. Estructura de la mitocondria.....	54
Figura 8. ADN mitocondrial (ADNmt).....	55
Figura 9. Heteroplasmas en el ADNmt	59
Figura 10. Métodos aplicados por EMPOP para el control de calidad de muestras de ADNmt....	71
Figura 11. Distribución geográfica de los haplogrupos de ADNmt.....	74
Figura 12. Muestreo realizado en los departamentos de la región andina.	88
Figura 13. Secuencias alineadas con respecto a la rCRS.	99
Figura 14. Deleción 498del y sustitución de base 499A	100
Figura 15. Inserción de las bases 524.1A y 524.2C	100
Figura 16. Heteroplasma de posición (146Y)	101
Figura 17. Heteroplasma de longitud en la posición 16193 de la región HVSI.....	102
Figura 18. Heteroplasma de longitud en la posición 309 de la región HVSII.	102
Figura 19. Heteroplasma de longitud en la posición 573 de la región HVSIII.	103
Figura 20. Código IUPAC.....	104
Figura 21. Diagrama de flujo (PRISMA) del proceso de selección e inclusión de los estudios..	107
Figura 22. Estudios de ADNmt publicados en América	110

Figura 23. País de procedencia de los estudios seleccionados	110
Figura 24. Distribuciones del número de diferencias entre pares de secuencias para los haplotipos en las provincias de Santander	117
Figura 25. Representación en red de las ramas dentro del haplogrupo B.	124
Figura 26. Representación en red de los haplogrupos A2+64, A2+64+153 y A2+64+16129.....	125
Figura 27. Representación en red del haplogrupo A2ac.	126
Figura 28. PCA de las frecuencias de macrohaplogrupos de ADNmt en Santander	131
Figura 29. Distribuciones del número de diferencias entre pares de secuencias para los haplotipos en los departamentos de la Región Andina.	140
Figura 30. Ancestría de los haplogrupos por departamento.....	143
Figura 31. Distribución de los macrohaplogrupos nativos americanos por departamento	147
Figura 32. Representación en red de los haplogrupos A.....	148
Figura 33. Representación en red de los haplogrupos B.....	149
Figura 34. Representación en red de los haplogrupos C.....	151
Figura 35. Representación en red de los haplogrupos D.....	152
Figura 36. Valores de F_{ST} por pares entre los departamentos de la región andina.....	153
Figura 37. MDS de los F_{ST} obtenidos para haplogrupos nativos de los departamentos de la región andina.	156
Figura 38. PCA a partir de las frecuencias de haplogrupos nativos en los departamentos de la región andina.	157
Figura 39. PCA a partir de haplogrupos nativos integrando la muestra del proyecto 1000 genomas.....	163

Lista de Apéndices

Apéndice A. Consentimiento informado participantes de la región andina.....	181
Apéndice B. Consentimiento informado participantes de Santander.....	184
Apéndice C. Ponencia en el 29th Congress of the International Society for Forensic Genetics. 188	
Apéndice D. Artículo “Sample collection strategies when building mitochondrial DNA forensic databases”.....	189
Apéndice E. Ponencia en el LV congreso nacional & VI internacional de ciencias biológicas .1900	
Apéndice F. Ponencia en el 29th Congress of the International Society for Forensic Genetics.	1922
Apéndice G. Certificado EMPOP QC EMP00862 (Santander, 204 datos)	1933
Apéndice H. Carta de aceptación de artículo: Maternal ancestry and lineages diversity of the Santander population from Colombia	1955
Apéndice I. Haplotipos y haplogrupos región andina	1966
Apéndice J. Ponencia en el 28th Congress of the International Society for Forensic Genetics. .2300	
Apéndice K. Artículo “Maternal genetic characterization of a colombian andean population” 2311	
Apéndice L. Certificado EMPOP QC EMP00871 (Departamentos Región Andina, 458 datos) 2322	
Apéndice M. Frecuencia de los haplogrupos en los departamentos de la región andina.....	2344

Resumen

Título: Identificación de linajes maternos presentes en individuos nacidos en la Región Andina colombiana. Una aproximación a su uso en Genética Forense*

Autor: Adriana Castillo Pico**

Palabras claves: ADN mitocondrial, ascendencia nativa, haplotipo, haplogrupo, genética forense, subestructura poblacional.

Descripción:

La región andina de Colombia está conformada por los departamentos de Antioquia, Boyacá, Caldas, Cundinamarca, Huila, Norte de Santander, Quindío, Risaralda, Santander y Tolima. Su población actual es el resultado de la mezcla intercontinental de nativos americanos autóctonos, colonos europeos y esclavos africanos, que en diferentes momentos de su historia se asentaron allí. Para crear una base de datos forense con las frecuencias de los haplotipos de ADN mitocondrial (ADNmt) es crucial identificar la diversidad genética de la población y evaluar la presencia de subestructura. Por ello se caracterizaron los haplotipos de la región control (RC) total de ADNmt para 662 individuos nacidos en estos diez departamentos, mediante secuenciación por Sanger seguida de electroforesis capilar, y a partir de ellos se asignaron los respectivos haplogrupos.

Se encontró que la herencia nativa materna está altamente preservada en el fondo genético de estos departamentos, con un promedio de 89,3% de haplotipos pertenecientes a los macrohaplogrupos A2, B4, C1 y D, seguida de los haplogrupos euroasiáticos (6,6%) y africanos (4,1%). Comparaciones por parejas exhibieron en algunos casos valores de F_{ST} correspondientes a una diferenciación genética moderada y estadísticamente no significativos, excepto los pares que incluyen a Risaralda y algunos de Cundinamarca. Un Análisis de Componentes Principales permitió inferir que las diferencias fueron debidas, principalmente, a contribuciones disimiles de los haplogrupos nativos. La base de datos diseñada en este estudio con las frecuencias de haplotipos de la RC del ADNmt puede ser utilizada con fines forenses.

* Tesis doctoral

**Facultad de Salud. Escuela de Medicina. Director: Fernando Rondón G. Codirectora: Leonor Gusmão.

Abstrac

Title: Identification of maternal lineages present in individuals born in the Colombian Andean Region. An approach to its use in Forensic Genetics*

Author: Adriana Castillo Pico**

Keywords: Mitochondrial DNA, native ancestry, haplotype, haplogroup, forensic genetics, population substructure.

Description:

The Andean region of Colombia comprises the departments of Antioquia, Boyacá, Caldas, Cundinamarca, Huila, Norte de Santander, Quindío, Risaralda, Santander, and Tolima. Its current population results from the intercontinental admixture of autochthonous Native Americans, European settlers, and enslaved Africans who settled there at different times in their history. In order to create a forensic database with the frequencies of mitochondrial DNA (mtDNA) haplotypes, it is crucial to identify the population's genetic diversity and evaluate the presence of substructure. For this reason, haplotypes of the entire mtDNA control region (RC) for 662 individuals born in these ten departments were characterized by Sanger sequencing followed by capillary electrophoresis, and the corresponding haplogroups were assigned.

It was found that the maternal native inheritance is highly preserved in the genetic background of these departments, with an average of 89.3% of haplotypes belonging to the A2, B4, C1, and D macro haplogroups, followed by the Eurasian haplogroups (6.6%) and Africans (4.1%). In some cases, pairwise comparisons between departments exhibited moderate genetic differentiation but non-statistically significant F_{ST} values, except for the pairs that include Risaralda and some involving Cundinamarca. The Principal Components Analysis allows us to infer that the differences were mainly due to differential contributions of the native haplogroups. The database designed in this study with the frequencies of mtDNA CR haplotypes can be used for forensic purposes.

* Doctoral thesis

**Facultad de Salud. Escuela de Medicina. Director: Fernando Rondón G. Codirectora: Leonor Gusmão.

Introducción

La genética forense es una especialidad dedicada a la evaluación de las evidencias biológicas a partir del análisis de ADN, con el fin de establecer los perfiles genéticos de los individuos implicados en litigios y situaciones de conflicto. Para llevar a cabo esto se han desarrollado herramientas moleculares que permiten dilucidar de forma objetiva e imparcial el reconocimiento de relaciones de parentesco, y la identificación de personas desaparecidas, víctimas de crímenes, ataques terroristas, conflictos armados, así como la identificación de restos humanos antiguos (1).

Mediante la llamada *prueba de ADN*, se analiza el material genético de un individuo y la genética forense colabora con los tribunales de justicia de todo el mundo para esclarecer y solventar las situaciones judiciales tanto de ámbito civil (filiación) como penal (identificación y criminalística) (2). Para la justicia, el derecho a conocer la verdadera filiación es el derecho a la identidad de la persona humana, y hoy gracias a la ciencia y los enormes desarrollos de la genética forense, este se encuentra ligado a una única realidad biológica (3).

En la actualidad, la mayoría de los casos forenses se resuelven mediante la genotipificación de marcadores genéticos tipo STR (*Short Tandem Repeat*) ubicados en los autosomas, que, por su herencia mendeliana codominante y su elevado polimorfismo derivado de su alta tasa de mutación poseen en las poblaciones un gran número de alelos y por tanto, son extraordinariamente

discriminativos y alcanzan con el genotipado de 20 marcadores una probabilidad de identidad $PI > 10^{29}$ (4).

La llamada prueba de ADN tiene gran importancia en el campo judicial debido principalmente a los avances tecnológicos de la biología molecular que ha hecho posible la obtención de información genética a partir de cantidades muy pequeñas de material biológico y a la inclusión de una valoración estadística de estos resultados obtenidos en el laboratorio, lo que permite que la prueba tenga un carácter científico y sea objetiva (5).

Para realizar el análisis estadístico de la evidencia biológica hallada, se requiere de la valoración previa de las frecuencias de los marcadores genéticos en la población donde ocurrió el caso y para ello se deben analizar muestras de un grupo representativo de individuos de esta población y así determinar la distribución de las frecuencias de los alelos o los haplotipos de cada marcador. Una vez se establezca el valor de estas frecuencias en la población de referencia, estos marcadores podrán ser introducidos en la resolución de los casos (6).

El estudio de las frecuencias alélicas o haplotípicas debe realizarse en la población particular donde ocurrió el hecho a valorar, porque la distribución de las frecuencias está relacionada con las diferencias intra e inter poblacionales que se pueden encontrar en las diferentes poblaciones, dado que las poblaciones son distintas desde el punto de vista genético (7). Esta realidad se ve reflejada aún más en países con una alta diversidad genética como lo es Colombia, derivada del aporte de los individuos que conforman actualmente su población que es mayoritariamente mezclada y que se ha originado a partir de tres grandes grupos con orígenes etnográficos muy diferentes: los indígenas nativos que habitaban nuestro territorio al

descubrimiento de América, los conquistadores europeos y los africanos que fueron traídos como esclavos después de la conquista europea (8). La mezcla genética de nuestra población ha sido detectada en los estudios realizados por investigadores que han evaluado las poblaciones de las distintas regiones del país, con diferentes tipos de marcadores autosómicos biparentales (9,10,11).

Las diferencias en las frecuencias de los marcadores genéticos en las distintas poblaciones tienen impacto en el informe de resultados porque generan diferencias en los resultados estadísticos y, por tanto, la objetividad de un informe pericial puede modificarse cuando no se utilizan las frecuencias adecuadas para resolver el caso (12). Los laboratorios forenses del país han realizado estudios poblacionales de individuos nacidos en la mayoría de los departamentos del país y se cuenta con bases de datos de frecuencias de STRs autosómicos (11,13).

En casos complejos como la investigación biológica de filiación, cuando no se cuenta con familiares directos del hijo/a por reconocer y/o en las pericias de criminalística que deban realizarse a partir de vestigios biológicos como manchas de sangre, restos óseos, pelos, entre otros; puede requerirse genotipar marcadores genéticos adicionales como los localizados en el cromosoma Y o en el ADN mitocondrial (ADNmt) (14,15). Estos marcadores de herencia uniparental, el cromosoma Y de herencia patrilineal y el ADNmt de herencia matrilineal, no recombinan y los cambios que se presenten entre individuos son debidos a mutaciones, se heredan en conjunto conformando un haplotipo que es compartido por todos los individuos de un mismo linaje paterno (cromosoma Y) o materno (ADNmt) y no permiten el mismo grado de individualización que los marcadores autosómicos, solo permiten discriminar entre linajes de individuos (14,15,16).

En Colombia, el conflicto armado interno al que se ha visto enfrentado el país durante los últimos 50 años ha llevado a la desaparición forzada de cientos de personas cuyos cadáveres se encuentran enterrados en fosas comunes, en alto grado de descomposición o son ya restos óseos y requieren su identificación (17,18). Para estos casos, los laboratorios forenses deberían contar con el análisis de ADNmt para obtener o complementar los resultados (15,19), especialmente para identificar los desaparecidos dentro del acuerdo de paz firmado entre el gobierno y la guerrilla de las FARC (20).

A pesar de ser reconocida la importancia del análisis del ADNmt por la comunidad científica que trabaja en laboratorios forenses de la mayoría de los países del mundo (21,22,23,24), en Colombia pocos laboratorios han validado e incorporado este análisis para resolver sus casos complejos. De acuerdo con la revisión sistemática de alcance realizada como uno de los objetivos de esta tesis para recopilar la evidencia disponible de publicaciones que analicen ADN mitocondrial humano en poblaciones americanas; solo se encontraron 21 estudios realizados en población colombiana actual, que utilizaron el análisis de ADNmt en la literatura publicada desde 1982 (año en que se secuenció el primer genoma mitocondrial) hasta septiembre de 2022 (25).

De estas publicaciones originales que presentaron resultados del análisis de ADNmt en muestras de individuos no relacionados nacidos y residenciados actualmente en diferentes lugares del país, el 57% analizaron poblaciones aisladas nativo-americanas o afro-descendientes (26,27,28,29,30,31,32,33,34,35,36,37) el 29% analizaron muestras de poblaciones aisladas nativo americanas y de población mezclada (38,39,40,41,42,43) y solo el 14% estudiaron muestras de individuos mezclados que vivían en centros urbanos del país (44,45,46) y una tesis de maestría que evaluó ADNmt en población actual residente en los departamentos de la región caribe colombiana

(47). De estas publicaciones la mayoría tuvo un enfoque de genética de poblaciones para el análisis de los datos y solo tres incluyeron parámetros de utilidad forense como parte de su análisis (43,45,46), evidenciando así la deficiencia en el estudio del ADNmt con fines forenses en el país.

Adicionalmente, en una búsqueda realizada en la base de datos internacional EMPOP (*EDNAP Mitochondrial DNA Population Database*), (www.EMPOP.org) (48), donde los laboratorios forenses pueden consultar las bases de datos de su respectivo país para resolver sus casos, solo se encontraron dos estudios en población colombiana realizados en poblaciones aisladas (nativos y afroamericanos), (consulta realizada el 10/01/23).

De acuerdo con esta revisión existe un vacío de conocimiento de los linajes de ADNmt del país y esto impide que el análisis de ADNmt sea incorporado a la resolución de casos forenses. Por ello los resultados obtenidos de esta tesis son un aporte para conocer la composición genética del ADNmt de una muestra poblacional de 662 individuos nacidos y residenciados en los 10 departamentos que conforman la región andina colombiana, seleccionada para este estudio teniendo en cuenta que allí vive más del 60% de la población colombiana, de acuerdo con el censo de 2018 realizado por el Departamento Administrativo Nacional de Estadística (DANE) (49) y que esta región registra el 50% de los desaparecidos reportados en el Sistema de Información Red de Desaparecidos y Cadáveres (SIRDEC) del Instituto de Medicina Legal (50).

En este estudio se establecieron los linajes maternos presentes en las muestras analizadas, las frecuencias de los haplotipos, se comparó la composición de estos linajes entre los departamentos y se evaluó su ancestría a partir de la estrategia metodológica utilizada de secuenciación de la región control completa del ADNmt (HVS I, II y III). Adicionalmente, se

conformó una base de datos con los haplotipos y haplogrupos de las muestras genotipadas para la RC del ADNmt en cada uno de los departamentos, que permitirá estimar los índices forenses incluidos en un informe pericial y que estará disponible para ser consultada por los laboratorios forenses del país, cumpliendo con los requisitos necesarios para que quedase bien conformada y fuese confiable tanto en sus perfiles genéticos como en la organización y agrupación de los departamentos evaluados.

En este documento se presentan los resultados obtenidos en el orden en que fueron procesadas y analizadas las muestras. Primero se presentan, analizan y discuten los resultados obtenidos para el departamento de Santander, y luego se presentan, analizan y discuten los resultados obtenidos para los demás departamentos de la región andina (Antioquia, Boyacá, Caldas, Cundinamarca, Huila, Norte de Santander, Quindío, Risaralda y Tolima).

1. Justificación

La población actual de Colombia es mayoritariamente mezclada, producto del aporte diferencial de distintos orígenes, lo que presume importante diversidad genética. Esta diversidad es esencial para resolver los casos forenses (51,52) y ha sido evaluada en Colombia a partir de marcadores genéticos biparentales autosómicos como los Short Tandem Repeats (STRs), los Single Nucleotide Polimorfisms (SNPs) o las Inserciones – Delecciones (Indel) (8,11,13,53,54). A partir de estos análisis se han calculado y publicado los datos de las frecuencias poblacionales de la mayoría del país y son utilizadas al resolver los casos tanto de criminalística como de filiación, en los laboratorios forenses (8,9,13).

Si se cuenta con buena cantidad y calidad de muestra de ADN y/o con muestras del grupo familiar directamente implicado en un caso de filiación, estos son resueltos con marcadores autosómicos (54,55). En los casos que no pueden resolverse únicamente a partir del análisis de estos marcadores, porque las muestras no son óptimas, están degradadas, contaminadas o en mínima cantidad, como ocurre en criminalística o en los casos complejos de filiación que requieren reconocer un individuo teniendo disponibles solo familiares en segunda o tercera línea generacional emparentados por vía materna, es de importancia primordial incluir la evaluación del ADN mitocondrial (ADNmt) de la muestra (15,24).

En el país el uso del ADNmt para resolver casos forenses es muy limitado, solo Medicina Legal lo utiliza en casos de criminalística. Los demás laboratorios forenses, a pesar de estar

certificados y acreditados para analizar el ADN con fines forenses, solo resuelven casos de filiación y no tienen participación en la resolución de casos de criminalística e identificación, donde tiene mayor utilidad valorar el ADNmt, porque esto son competencia de instituciones del gobierno como el Instituto de Medicina Legal, la fiscalía general de la Nación y la Policía Nacional.

Esta es una gran limitación en el aporte que estos laboratorios podrían hacer en la coyuntura actual del proceso de paz que se desarrolla en el país y representa una subutilización del recurso humano formado y competente en Genética Forense que podría participar en la resolución de estos casos, contribuyendo además en disminuir el tiempo de entrega de los informes de resultados a las familias de las víctimas y logrando que el proceso de identificación de desaparecidos por el conflicto armado, bastante lento al día de hoy, fuese más eficiente (20,50).

Como todo marcador genético que se desee usar en forense, se requiere conocer las frecuencias haplotípicas en la población donde ocurrió el hecho, para realizar los cálculos estadísticos, una vez se han establecido los perfiles genéticos, y darle el peso matemático a la evidencia analizada (23). Debido a que la distribución de los haplotipos de ADNmt presentes en los individuos que conforman una población es diferente a la de individuos de otras poblaciones, y teniendo en cuenta que hay muy pocos estudios realizados y publicados que analicen el ADNmt en población colombiana actual mezclada (38,39,40,42,43), y estos a su vez analizaron las muestras con la metodología PCR-RFLP, solo se han identificado los haplogrupos presentes en algunos departamentos para establecer su ancestría y no con fines forenses, porque no es posible discriminar uno a uno los haplotipos de las muestras analizadas y calcular el poder de discriminación de este marcador.

Adicionalmente, mediante el uso de la secuenciación como estrategia metodológica se podría determinar la ascendencia de los linajes maternos: nativa americana, euroasiática o africana en los departamentos que conforman la región andina. Estos concentran más del 60% de la población del país y hacen parte de la región donde están reportados más del 50% de los casos de desaparecidos (49,50). Los hallazgos permitirán implementar una base de datos que contenga las frecuencias de los haplotipos para la región control de ADNmt en la base EMPOP, donde en la actualidad no están reportados datos de ninguna población mezclada de los centros urbanos colombianos (48,58). Esta base de datos de frecuencias haplotípicas estará disponible para los laboratorios forenses que la requieran.

Por todas las razones expuestas es perentorio identificar los haplotipos para la RC de ADNmt presentes en la población mezclada colombiana, establecer las frecuencias de cada haplotipo, la diversidad genética, la ascendencia de los linajes maternos presentes y comparar los haplotipos hallados en los diferentes departamentos considerados para diseñar la(s) base(s) de datos necesaria(s) a ser aplicadas en forma correcta al momento de requerirse cálculos estadísticos de cada caso que se deba resolver (15). De esta manera esperamos que esta tesis contribuya al conocimiento de nuestra genética, así como a la identificación de desaparecidos dentro del proceso de paz en el cual está inmerso nuestro país actualmente.

2. Hipótesis

Sí se asume que la heterogeneidad genética de la región Andina colombiana, debida a patrones de asentamiento humano precolombino disímiles, así como a diferencias en los procesos de mestizaje, se ha desvanecido en las poblaciones actuales por efecto de la migración entre regiones y, entre todos sus departamentos, se esperaría que dentro y entre los aislados poblacionales en esta región no existan diferencias significativas y por lo tanto, exhiban un patrón genético matrilineal similar.

3. Objetivos

3.1 Objetivo general

Determinar la composición, diversidad y grado de diferenciación genética en una muestra de individuos nacidos en la región Andina colombiana a partir de información presente en el ADN mitocondrial, para su uso en genética forense.

3.2 Objetivos específicos

- 3.2.1 Mapear las aplicaciones que tiene el análisis de las variaciones del ADN mitocondrial en poblaciones americanas por medio de una revisión sistemática de alcance.
- 3.2.2 Establecer la composición genética de los individuos nacidos y residentes en los departamentos que conforman la región andina colombiana, a partir de secuencias completas de la región control del ADN mitocondrial.
- 3.2.3 Estimar la diversidad genética, el grado de diferenciación poblacional y los patrones de flujo genético entre los departamentos analizados, a partir de los haplotipos encontrados.
- 3.2.4 Evaluar y comparar la contribución matrilineal de diferentes orígenes etnogeográficos, en las muestras de cada departamento considerado.
- 3.2.5 Crear una base de datos con los haplotipos de ADN mitocondrial identificados en el estudio, delineada con base en la estructura poblacional hallada, que sea adecuada para la resolución de casos forenses que involucren muestras de la región andina colombiana.

4. Marco Teórico

4.1 Historia del poblamiento de la región andina colombiana

4.1.1 *Las regiones de Colombia*

Colombia es un país ubicado en el extremo Noroccidental de Sur América. De acuerdo con el DANE la cifra estimada de habitantes del país, según el último censo es de 48.258.494 personas (estadísticas para 10/01/23) (49). El poblamiento de Colombia está muy asociado a su espacio geográfico debido a que el país posee un área continental y una marítima. (Figura 1) El área continental está conformada geográficamente por 5 regiones naturales (Caribe, Pacífica, Andina, Amazonía y Orinoquía) y el área marítima está conformada por el archipiélago de San Andrés y Providencia; cada una de ellas está asociada a las características específicas de sus entornos naturales (59). Estas regiones naturales se consideran unidades territoriales que tienen características comunes en aspectos físicos, como el relieve y el clima, en las costumbres y tradiciones culturales, o también en actividades económicas o políticas. (60).

Figura 1. Mapa de las regiones naturales de Colombia.

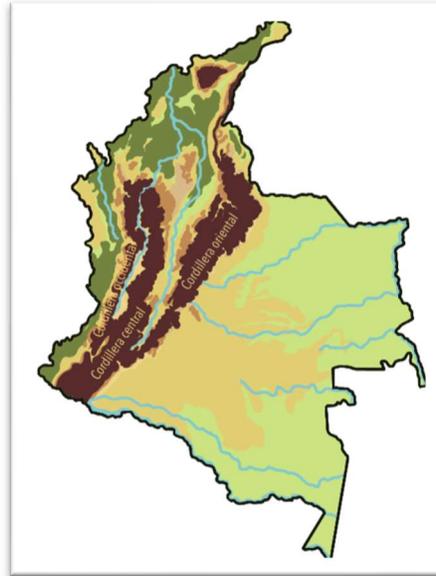


Históricamente la mayoría de las actividades del país han tenido lugar en la región andina y la del caribe, siendo las demás regiones secundarias en cuanto a su papel en la historia del país. La descripción de las características y fenómenos relacionados con el espacio geográfico es útil para entender donde ha transcurrido la historia y también cómo este aspecto y sus relaciones han jugado un papel fundamental en la configuración y la construcción de los procesos sociales en distintos momentos de la historia del país (59). Existen muchos factores ambientales y sociales que determinan y construyen su espacio geográfico, entre los de mayor relevancia están las características asociadas a su posición en el planeta (país ecuatorial), disponibilidad de agua (numerosas vertientes de agua), su topografía (país de montañas y llanuras) y su biodiversidad (abundantes formas de vida) y cómo estos factores interactúan entre sí para generar particularidades en los diferentes lugares del espacio geográfico colombiano (59).

Los factores biofísicos como ser un país ecuatorial, (su ubicación es aproximadamente 12° de latitud norte y 4° de latitud sur), hace que tenga una temperatura casi constante a lo largo del

año y que solo se vea modificada por la altitud, esto permite una agricultura productiva en diferentes climas, durante todo el año. Ser un país rico en agua hace que Colombia sea uno de los países del mundo con mayor disponibilidad de ella, debido a que está cerca de tres grandes fuentes de humedad (océanos atlántico y pacífico y cuenca amazónica) y a su posición ecuatorial. Ser un país de montañas y llanuras lo caracteriza por la diversidad de rocas y sustratos, relieve y geoformas que generan una gran diversidad de tipos de suelos, la región andina posee las tres cordilleras que se forman a partir de la cordillera de los Andes que entra al territorio colombiano por el sur y entre estas tres ramas (Cordilleras occidental, central y oriental) hay profundos valles interandinos por los que pasan los ríos Cauca y Magdalena, atravesándolos de sur a norte. Por su parte la zona oriental del país conformada por las regiones de Orinoquía y Amazonas es rica en llanuras y colinas, es casi toda plana y está irrigada por abundantes y caudalosos ríos. Las regiones del caribe y el pacífico también están asociadas con relieve relativamente plano, aunque cuentan con algunas serranías (Figura 2). La interacción entre las montañas y la posición ecuatorial genera pisos bioclimáticos altitudinales bien diferenciados, esto favorece gran variedad de condiciones climáticas, de fauna y flora y de recursos minerales que están disponibles muy cerca unos de otros (59).

Figura 2. Mapa del relieve colombiano.



Desde el inicio del poblamiento del país, los grupos nativos aprovecharon este potencial de los pisos bioclimáticos de la región andina para su producción agropecuaria en los diferentes pisos altitudinales y realizaron diversas actividades de intercambio comercial entre tierras bajas y altas y los ríos Magdalena y Cauca sirvieron como medios para el desplazamiento y penetración dentro de la región; así mismo los conquistadores buscaron establecerse en climas más saludables y menos calurosos llevando a que se diera un proceso histórico de concentración de la población en la región andina, de tal forma que desde el inicio de su poblamiento, durante la conquista, luego en la colonia, el periodo republicano y hasta nuestros días, los andes colombianos concentran la mayoría de la población, así como el poder económico y político del país. Por ello, aunque solo el 18% del área de Colombia está por encima de los 1000 metros de altitud, más del 60% de la población colombiana vive en la región andina (59,60).

Finalmente Colombia tiene una alta variedad de formas de vida, está considerado como el tercer país con más biodiversidad del planeta, cuenta hasta la fecha con 54.871 especies registradas, de las cuales 3.625 especies son exclusivas: 66 aves, 1.500 plantas, 367 anfibios endémicos (únicos en Colombia), cuenta con 30.436 especies de plantas, 32 biomas terrestres y 314 tipos de ecosistemas, los de páramo representan aproximadamente el 1,7% del territorio colombiano que aportan agua al 70% de la población, Minciencias consultado 20/10/22 (61). El hecho de que Colombia sea un país rico en biodiversidad ha sido fundamental para su desarrollo ya que esta permite que la sociedad obtenga la mayoría de los bienes y servicios ambientales esenciales para su bienestar. Este fue uno de los factores que promovieron el inicio de la expedición botánica por parte de los españoles, con la idea de buscar y ubicar especies que permitieran a la corona española aprovechar económicamente esta biodiversidad, evento que favoreció adentrarse en las montañas que forman las tres cordilleras andinas, avanzando a través de los ríos Magdalena, Cauca y de sus numerosos afluentes (59).

Los ecosistemas, especialmente de las regiones andina y caribe han sido transformados por las actividades de la sociedad desde el inicio de la conquista (60,62), esta transformación también se ha relacionado con la tenencia de la tierra y el control territorial, los conquistadores impusieron a su llegada nuevas prácticas agropecuarias basadas en especies del nuevo mundo y la ganadería se consolidó como una de las actividades más ligadas al control territorial, así como la expansión de las grandes haciendas de los narcotraficantes y sus testaferros lo que ha llevado a que la tenencia de la tierra se concentre en pocas personas, por lo que el manejo y dominio del espacio geográfico se ha constituido en una herramienta de exclusión social, de control territorial y en un impulsor del desplazamiento forzado (59).

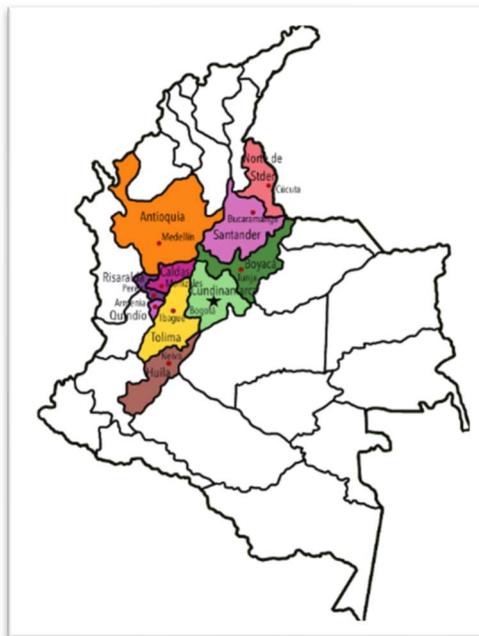
Otra característica muy destacada del espacio geográfico del país es la dificultad de movilización terrestre, por ello las evidencias del intercambio comercial entre los nativos que vivían en la era prehispánica, estaba restringido a zonas muy cercanas entre sí. Desde el inicio de la conquista, el río Magdalena se convirtió en el medio de comunicación más importante entre la región andina y las costas del caribe, las vías terrestres eran muy precarias y por tanto, la comunicación entre los asentamientos y fundaciones españolas eran muy esporádicas y lentas y en general comunicaban los asentamientos de la región andina con los puertos sobre el río Magdalena, patrón que se mantuvo hasta el siglo XX cuando la llegada del ferrocarril conectó las ciudades andinas con los puertos (59).

Las dificultades de comunicación permitieron que desde la conquista los asentamientos se formaran a partir de enclaves fundacionales en los que los conquistadores europeos tenían el control territorial, alrededor de estas zonas se establecieron actividades agropecuarias para sustentar el núcleo urbano y conformaron los pueblos que más adelante conformaron las actuales ciudades que fueron creciendo un poco aislados, no solo por las dificultades de comunicación sino porque a la par algunos grupos indígenas belicosos defendieron sus territorios y surgieron así los llamados pueblos de indios y más adelante los afrodescendientes con sus asentamientos llamados palenques, generando un patrón de enclaves bajo el dominio de diferentes grupos, de los cuales la mayoría creció con la presencia de europeos, criollos (hijos de europeos nacidos en América) y mestizos que se fueron sumando a partir de la mezcla de estos tres grupos: nativos americanos, europeos y afrodescendientes. Este patrón de enclaves ha sido característico en la historia del país y perdura hasta hoy, en algunos de ellos diversos grupos como la guerrilla y los paramilitares luchan por el control de los territorios y del poder, a lo largo del país (59,63).

4.1.2 La Región Andina

Se llama así porque en ella se encuentran las tres cordilleras de los Andes: occidental, central y oriental. Se localiza en el centro del país, su relieve es variado, posee montañas, valles, sabanas, volcanes y nevados. Por esta razón presenta casi todos los climas: templado, frío, páramo. Glaciar, húmedo y semihúmedo (59,64). Esta región está conformada por los departamentos de Antioquia, Boyacá, Caldas, Cundinamarca, Huila, Norte de Santander, Quindío, Risaralda, Santander y Tolima. De acuerdo con el último censo la población estimada de esta región es de 28.987.763 es decir, concentra el 60.07% de la población del país (49). En ella se encuentran las principales ciudades de país: Bogotá (la capital), Medellín, Bucaramanga, Cúcuta, Manizales, Pereira, Armenia, entre otras. (Figura 3).

Figura 3. Departamentos que conforman de la región andina colombiana



De acuerdo con los estudios arqueológicos, los primeros humanos que llegaron a Suramérica lo hicieron hace aproximadamente 15 mil años, procedían de norte y centro América, ingresaron a Colombia por el istmo de Panamá, como paso obligado por su situación geográfica (7), algunos siguieron avanzando hacia el interior de Suramérica y otros se instalaron en Colombia aprovechando su biodiversidad. Los estudios genéticos y lingüísticos sugieren que desde ese momento ingresaron poblaciones muy diversas que mantuvieron esta diversidad por intercambios genéticos y culturales continuos entre estas poblaciones (28).

Inicialmente conformaron aldeas nucleadas a la orilla de los ríos y se estima que en el último milenio antes de Cristo, el desarrollo del cultivo del maíz les permitió extenderse sobre las laderas de las cordilleras andinas en búsqueda de condiciones propicias para sus cultivos. Allí, dadas las condiciones tan accidentadas de las cordilleras andinas, se conformaron unidades sociales más pequeñas, lo que llevó a una manifiesta descentralización y a nuevas formas de adaptación que se caracterizaron por su diversidad, su notable regionalismo y su elaboración de instituciones económicas, sociales y religiosas (66). De esta forma se ubicaron en diversas regiones del país con ambientes diferentes y se adaptaron a modos de vida que difirieron mucho. Se dedicaron a la agricultura, la caza y más adelante varios de ellos se dedicaron a la alfarería y la orfebrería (63). Las culturas más importantes en ese momento en Colombia fueron San Agustín, Muisca, Tairona y Zenú. Con el paso del tiempo se dio un proceso de crecimiento demográfico y de establecimiento de formas cada vez jerarquizadas de organización social (63). Los vestigios culturales más antiguos del país hallados hasta ahora, ha sido en la región andina (66). La mayor parte de la población se ubicaba en la cordillera central y en la región caribe, seguida de los andes orientales y el valle del río Magdalena (63).

Existe controversia acerca de la cantidad de indígenas que habitaban el territorio de la actual región andina colombiana, especialmente porque los documentos que narran este tema son de la época colonial, sin embargo, la mayoría de los historiadores concuerdan en que la región andina del país era la más poblada en la época prehispánica, dado que allí se practicaban las formas de agricultura más intensiva, según los hallazgos arqueológicos (63). Por ello a su llegada a este territorio, los conquistadores españoles encontraron múltiples sociedades nativas con una amplia diversidad cultural y lingüística, siendo tan diversas como el territorio que ocuparon, algunas con líderes de carácter permanente que vivían en asentamientos grandes y otras más pequeñas y sin liderazgos durables (63). Estas sociedades se habrían desarrollado primero en la costa Caribe, luego habrían colonizado las vertientes de ríos para dar origen a sociedades como la muisca en los territorios de Cundinamarca y Boyacá y las de la Sierra Nevada de Santa Marta que fueron estables y crecieron dadas las condiciones del medio (65). Por otra parte, muchas otras comunidades nunca alcanzaron altos niveles de población y se convirtieron en comunidades pequeñas e inestables. Este patrón se conserva hasta hoy con un poblamiento predominantemente mestizo con una abundante población andina campesina que ocupa el antiguo territorio indígena y poblaciones menos densas en las tierras bajas que poseen un importante aporte afrodescendiente (63). Los europeos llegaron al país en 1500, la mayoría de ellos fueron españoles y portugueses, el primer conquistador fue Rodrigo de Bastidas, quien llegó a Colombia por el Cabo de la Vela y junto con otros españoles conquistaron los indígenas de la costa atlántica hasta llegar al Darién, donde se fundó el primer pueblo español San Sebastián de Urabá en 1509 seguido de la fundación de Santa Marta por Rodrigo de Bastidas en 1525 y de Cartagena por Pedro de Heredia en 1533 (68). Hacia 1530 se inició la conquista del interior del país y para ello los conquistadores españoles y algunos portugueses ingresaron a través del Río Magdalena para llegar hasta las tierras habitadas por los

distintos grupos indígenas a lo largo de las cordilleras andinas, lo que se constituyó en una verdadera odisea para los conquistadores por los peligros y el hambre que acompañaron la conquista de estas inhóspitas tierras de geografía quebrada. En 1536 Gonzalo Jiménez de Quezada realizó una expedición que avanzó por el Río Magdalena, pasó por Vélez y el valle de Moniquirá y llegó a la meseta chibcha bien poblada por indígenas, donde fundó a Santa Fe de Bogotá en 1538 (67).

Desde la conquista y después de ella, todas las poblaciones indígenas prehispánicas declinaron su tamaño, unas en forma más elevada que otras debido principalmente a la esclavitud a la que fueron sometidos para la realización de trabajos forzados en la agricultura y la minería, asociado también a la presencia de enfermedades infecciosas, muchas de ellas adquiridas antes de la conquista y otras traídas durante ella (67). Después de la conquista, la población del país se concentró principalmente en la cordillera oriental, seguida del valle de Aburrá, patrón que se conserva hasta la actualidad, donde como ya se expuso vive la mayoría de la población del país, por arriba de los 1500 msnm. Adicionalmente hubo mezcla de los individuos dando lugar al mestizaje y continuó el descenso demográfico de los nativos americanos, asociado a las enfermedades, la violencia de la conquista y la explotación colonial (63) Este patrón observado en esta región fue distinto a lo largo del país, haciendo de la región andina un territorio mestizo a partir de la mezcla de indígenas y conquistadores, mientras que en otras regiones sobrevivieron algunas poblaciones indígenas aisladas que se conservan hasta hoy y en otras, los nativos fueron exterminados (63).

Por otra parte, desde el siglo XVI hasta el siglo XIX, durante la trata transatlántica de esclavos, millones de africanos llegaron a América (<https://www.slavevoyages.org/>). La esclavitud

era una institución corriente especialmente en la costa occidental del África y por ello buena parte del comercio de negros tuvo como fuente la población africana ya esclavizada. Sin embargo, el mayor volumen de esclavos se obtuvo mediante la caza directa violenta, el fraude y promoviendo guerras entre las tribus, por parte de los tratantes portugueses, holandeses, franceses e ingleses (68). De estos esclavos traídos desde África miles llegaron a Colombia por el océano atlántico, siendo Cartagena el principal puerto de llegada y de allí fueron llevados a lo largo de las costas para trabajar en las minas y en plantaciones de caña de azúcar y café. (30). Los primeros esclavos africanos llegaron una vez fundadas Santa Marta y Cartagena y su introducción al interior del país se produjo más adelante, debido a que la colonización española, que estaba fundamentada en la explotación de las minas de oro, plata y esmeraldas, mediante la utilización de grandes concentraciones de indios sedentarios, requirió de los esclavos africanos para reemplazar los indígenas, a medida que estos indígenas esclavizados se fueron diezmado por causa de los trabajos forzados (68). Con el cálculo muy pronto generalizado, de que el trabajo de un negro producía lo que tres indios juntos, se aumentó la introducción de negros esclavos que duró hasta el siglo XVIII; su participación estuvo orientada a la explotación minera durante la época de la conquista y más adelante, durante la colonia, al labrado y cultivo de la tierra, la ganadería y el servicio personal de los dueños de las grandes haciendas (68).

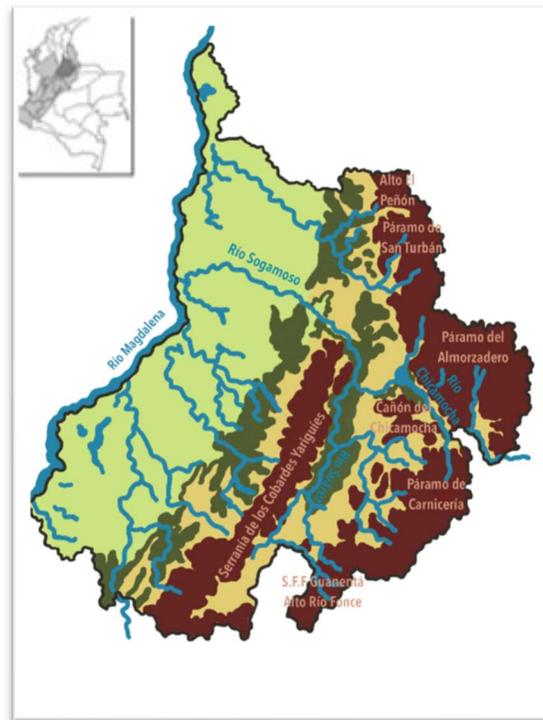
Con el paso del tiempo estos afrodescendientes se cruzaron principalmente con nativos americanos, criollos y mestizos (30,38,41), generando así cruces interétnicos que dieron lugar a poblaciones heterogéneas conformadas por estos tres grupos bien diferenciados: nativos americanos autóctonos, conquistadores europeos y esclavos africanos, a lo largo de todo el país,

aunque con aportes variables de cada uno de estos tres grupos en los diferentes departamentos dada la gran heterogeneidad geográfica (34,42).

4.1.3 El departamento de Santander

El departamento Santander está ubicado en la región andina, al noreste de Colombia. De acuerdo con el último censo tiene una población de aproximadamente 2'184.837 habitantes (49). En la Figura 4 se presenta el mapa geográfico de este departamento, donde se destaca su relieve representado por las tierras altas montañosas de la Cordillera Oriental de los Andes (marrón), que atraviesan el departamento de sur a norte, los ríos (azul) y las tierras bajas planas (diferentes tonos de verde). En la parte superior izquierda de la Figura 4 se observa el mapa de Colombia y en él se destaca la ubicación de la región andina (gris más claro) y del departamento de Santander (gris más oscuro). Este departamento limita por el norte con el departamento de Cesar, por el occidente con Bolívar y Antioquia, por el sur con Cundinamarca y Boyacá y por el oriente con el departamento de Norte de Santander.

Figura 4. Mapa en relieve de Santander.



Santander fue fundado oficialmente en 1857 y está conformado actualmente por 87 municipios distribuidos en siete provincias: Comunera, García Rovira, Guantán, Metropolitana (que incluye la capital Bucaramanga), Soto Norte, Vélez y Yariguíes. (49). Su población actual es mestiza en su mayoría, producto de la mezcla de los indígenas que habitaban el territorio y los conquistadores españoles que llegaron en el siglo XVI (69). En la Figura 5 se pueden observar las provincias del departamento resaltadas en diferentes colores, la capital de cada provincia destacada por estrellas negras y los municipios que lo conforman, representados por puntos negros.

Figura 5. Mapa de Santander con sus provincias y municipios



La topografía del departamento es accidentada, con áreas de tierras altas combinadas con áreas de tierras bajas planas, que dificultan el acceso a algunas provincias (Figura 4) y su población se concentra en las zonas urbanas, ubicadas principalmente sobre la región montañosa (Figura 5). En la época precolombina estaba poblado por varios grupos indígenas de los cuales el más numeroso y representativo era el de los Guanes, que ocupaban la parte media de la cordillera oriental, en la cuenca media y baja del río Suárez, por su parte los Chitareros y los Lache ocupaban las partes altas de las montañas del oriente y los Yariguíes ocupaban la región del Magdalena medio (Figura 4). La mayoría de estos grupos pertenecían a la civilización Chibcha (69). Durante la colonización europea que inició en el siglo XVI, los colonizadores españoles llegaron al territorio que hoy conforma el departamento, por la parte norte a través de las riberas del río Magdalena y del río Lebrija (Figura 4). A medida que avanzaron se conformaron los primeros asentamientos: unos como pueblos de indios, donde los conquistadores realizaban labores de evangelización y

donde se inició el mestizaje especialmente de hombre conquistador con mujer indígena, y otros como pueblos de blancos, en áreas antes no ocupadas, que fueron habitados por conquistadores, por criollos (hijos de españoles nacidos en América) y más adelante por mestizos, en una proporción variable en los diferentes pueblos (37). Más adelante, durante los siglos XVI al XIX, cuando el comercio de esclavos desde África trajo a América a millones de africanos (30), algunos de ellos fueron enviados a Santander, a través del río Magdalena, para reemplazar la mano de obra indígena que estaba siendo diezmada cada vez más por el trabajo pesado al que estaban siendo sometidos (68.69).

Estas nuevas comunidades de criollos (descendientes de españoles nacidos en Colombia) incorporaron gradualmente a otras personas nativo-americanas, africanas convirtiendo así estos asentamientos, en poblaciones mezcladas que fueron creciendo y afianzaron el surgimiento de la mayoría de los municipios que actualmente conforman el departamento; algunos de ellos crecieron en forma aislada, dadas las condiciones geográficas de su territorio que dificultan el acceso a los municipios de algunas provincias manteniéndolas apartadas de otras (70,71).

4.2 Desaparición forzada y proceso de paz en Colombia

La historia de Colombia es muy compleja y una parte importante de ella incluye los conflictos y guerras generados desde los inicios de su creación como nación, hasta nuestros días (72). Desde la conquista y colonización de nuestro territorio, se conformaron enclaves con un patrón de dominio por distintos grupos actores que ha sido característico a lo largo del tiempo y en el cual las guerras civiles del siglo XIX y las disputas posteriores, se han generado por el control de los territorios y del poder para construir espacialidades específicas denominadas “repúblicas independientes” en donde algunos grupos mantienen el monopolio de la fuerza y con ello el control

del territorio (59). Esto ocurre de forma similar en la actualidad, a lo largo y ancho del país, asociado a la lucha entre el estado y los actores ilegales (la guerrilla y los paramilitares) que quieren tener el control de los territorios para tener acceso a vías de comunicación seguras y controladas para la producción y comercialización de drogas, un negocio muy lucrativo que ha marcado al país no solo por el conflicto interno y el desarrollo de la cultura “traqueta” sino por la mala imagen a nivel internacional que se ha expandido y generalizado contra nuestro país (59).

La violencia del siglo XX en Colombia es quizás el asunto más espinoso de la historia contemporánea del país, no solo por los estragos que ha causado en su población, sino por su duración (73). Las distintas fracturas que han llevado a las disputas a lo largo de los años han tenido diversos protagonistas y se han desenvuelto con un trámite y resolución que han dejado siempre algo suspendido sin resolver. Ese algo suspendido está relacionado con fragmentaciones no resueltas debidas a la hegemonía de poder de actores privados por encima del estado y lo público, o debidas a la desagregación de los tejidos sociales, o debidas al surgimiento de antagonismos insuperables entre las partes en conflicto (73). Por ello la violencia vuelve y resurge; aunque las razones y actores hayan ido cambiando a lo largo del tiempo. Inicialmente por la colonización de tierras, luego por el bipartidismo entre liberales y conservadores que nace a mediados del siglo XIX y que dio origen a la ciudadanía en armas, y después en los años 70 del siglo pasado por el surgimiento de las guerrillas, los paramilitares y el narcotráfico que sumergieron al país en una guerra encargada de tratar de despedazarlo (73).

El siglo XX tuvo 4 episodios violentos grandes: a) La guerra de los mil días (1899 a 1902) duró 3 años y se extendió por todo el territorio nacional, sus protagonistas fueron los miembros de los partidos tradicionales de la época: liberales y conservadores, el resultado final fue que los

liberales perdieron y se calcula dejó como saldo 100.000 muertos de los 4'500.000 habitantes que se estima era la población del país (73). b) La violencia liberal (1930-1934) los liberales finalmente llegaron al poder e iniciaron una confrontación persiguiendo, desterrando y aniquilando a los conservadores y la respuesta de los conservadores no se hizo esperar; la confrontación tuvo características de guerra en Boyacá y los Santanderes, finalmente el gobierno envió un destacamento del ejército a la zona y logró dar fin al conflicto armado (73), c) La violencia con mayúscula (1946-1965) nuevamente aparecen las discordias entre conservadores y liberales y se desata una guerra civil no declarada, con el levantamiento popular ante el asesinato el 9 de abril de 1948 del candidato liberal a la presidencia Jorge Eliecer Gaitán, quien se constituía en una esperanza de cambio para los sectores menos favorecidos. La puerta de la confrontación sangrienta quedó abierta, por una parte, había escuadrones conservadores apoyados por el gobierno y por otra parte entre los liberales proliferaban las cuadrillas y los destacamentos de autodefensas que con el tiempo adoptaron la ideología comunista e influyeron en los fundadores de la guerrilla de las FARC (74).

Finalmente, en 1953 sube al poder el general Gustavo Rojas Pinilla quien logra gobernar por 4 años y le sigue una junta militar y el retorno a la democracia en las urnas después de un acuerdo de alternancia en el poder entre liberales y conservadores llamado Frente Nacional (73), y d) La violencia contemporánea (1965-?) Al tiempo que el ejército hacía esfuerzos por aniquilar el bandolerismo, se gestaban la recomposición de la vieja resistencia y la irrupción en el escenario nacional de la guerrilla moderna. A comienzos de los años 70 tras el debate en el parlamento sobre las llamadas repúblicas independientes, (zonas campesinas como Marquetalia al sur del departamento del Tolima, conformadas a partir de la violencia, de influencia comunista y con un

proyecto de gestión autónoma, se decide desde el gobierno enviar al ejército a combatirlos por tierra y aire llevando a la fuga de las familias campesinas hacia el piedemonte de la cordillera oriental resguardadas por sus destacamentos armados y poco después darían origen a una organización insurgente: la guerrilla de las FARC. Por esta misma época y como producto de la influencia de la insurrección cubana en el país, se conformaron dos nuevas guerrillas: el Ejército de Liberación Nacional (ELN) y el Ejército Popular de Liberación (EPL) con el proyecto de sustituir al estado y fundar un orden social alternativo (73).

Más adelante se conformaron más grupos revolucionarios de resistencia armada, a principios de los años 80 el gobierno, presionado ante la realidad guerrillera inició los acercamientos para hacer negociaciones que permitieran llegar a acuerdos con estos grupos al margen de la ley, en búsqueda de una paz duradera. Sin embargo, esto aún no se ha logrado, y posterior a la mitad de los años 70 se identifican 4 periodos: el implante de la guerrilla contemporánea (1965-1982), los procesos de paz con la insurgencia (1982-1991), las oscilaciones de guerra y paz (1991-1997) y la era paramilitar (1991-?). Han sido muchos los factores que han influido para que haya sido tan difícil lograr una paz duradera, entre otros que las guerrillas negociaron con el gobierno por separado e hicieron acuerdos independientes, varios grupos de la guerrilla no negociaron y se declararon en disidencia, y mientras unos negociaban, otras fuerzas de la guerrilla proliferaron. Hacia finales de los 80 el gobierno abrió un amplio proceso de diálogo nacional tendiente a introducir un conjunto de reformas que sirvieran de garantía al desarme y la inserción de sus miembros a la vida civil pero estos procesos abortaron y aparecieron en el escenario otros dos actores: ante la expansión de la guerrilla surgen en el Magdalena Medio los grupos paramilitares y en las zonas rurales despobladas, surgen los cultivos ilícitos y el negocio del

narcotráfico, añadiendo mayor complejidad al panorama del conflicto, a tal punto que los mediados de los años 80 son tiempos memorables para el país por el panorama de anarquía y caos caracterizado por el terror (73).

Se lanzó entonces en 1990 la propuesta de una Asamblea Nacional Constituyente para formular una nueva carta de navegación para el país: la constitución de 1991, con la esperanza de superar la crisis y buscar la paz; durante estos años la guerrilla del M-19, una fracción de la guerrilla del EPL y la guerrilla Quintín Lame, firmaron un pacto de paz, entregaron las armas y se incorporaron a la democracia. Por su parte los cabecillas de los dos carteles del narcotráfico hicieron un acuerdo con el gobierno e ingresaron a la cárcel, sin embargo, el narcotráfico no desapareció, solo cambió de grandes carteles de la droga a pequeños carteles, acompañados del crecimiento exponencial de las áreas de cultivo de coca, marihuana y amapola llevando a Colombia a ser el primer productor de coca en el mundo. Los actores armados (guerrillas y paramilitares) se benefician del negocio de los cultivos ilícitos y apalancan su expansión. Desde este momento en adelante el conflicto armado en el país se convierte en una confrontación a muerte por la coca (73).

En medio de las confrontaciones ha estado la sociedad civil de las zonas rurales, que estando en medio del conflicto ha sufrido sus consecuencias (ataques a poblaciones rurales, secuestro, tortura, desaparición o muerte de sus familiares, reclutamiento de menores, minas antipersonales, despojo o abandono forzado de sus tierras y desplazamiento forzado hacia las ciudades para garantizar su vida despojo y delitos contra la libertad sexual, entre otras realidades que se han derivado del proceso de expansión de los grupos ilegales y la ganancia de poder territorial (73). En la Figura 6 se presentan los datos de eventos victimizantes reportados en el boletín de víctimas del conflicto armado del 9 de abril de 2021 (75).

Figura 6. Eventos victimizantes según el hecho

HECHO	%
Desplazamiento forzado	79,0 %
Homicidio	10,1 %
Amenaza	4,8 %
Desaparición forzada	1,7 %
Pérdida de bienes muebles o inmuebles	1,1 %
Acto terrorista / Atentados / Combates / Enfrentamientos / Hostigamientos	0,8 %
Confinamiento	0,7 %
Secuestro	0,3 %
Delitos contra la libertad y la integridad sexual en desarrollo del conflicto armado	0,3 %
Sin información	0,3 %
Abandono o despojo forzado de tierras	0,3 %
Lesiones personales psicológicas	0,1 %
Lesiones personales físicas	0,1 %
Minas antipersona, Munición sin explotar y Artefacto Explosivo improvisado	0,1 %
Tortura	0,1 %
Vinculación de niños, niñas y adolescentes a actividades relacionadas con grupos armados	0,0 %

Nota. Tomado de: Boletín víctimas del conflicto armado (75)

En este cuadro se puede observar que el desplazamiento forzado constituye el hecho más común entre los eventos victimizantes, seguido del homicidio y en tercer lugar las amenazas (75). Esta realidad ha llevado a que mucha gente del área rural haya tenido que migrar hacia las ciudades huyendo de la violencia y establecer su residencia en estos nuevos lugares.

Las dinámicas más recientes presentan un panorama de transición hacia la fase de desarme y postconflicto, procesos que se derivan de diálogos hacia la paz con algunos grupos guerrilleros, iniciados en 2012 (17). En el 2010 el congreso expidió la ley 1408 reglamentada por el Decreto Nacional 303 de 2015, donde se rinde homenaje a las víctimas del delito de desaparición forzada y se dictan medidas para ubicar e identificar plenamente, así como para brindar asistencia a los familiares de estas víctimas en los actos de entrega de los cuerpos o restos exhumados identificados (76). A partir de allí el Instituto Nacional de Medicina Legal y Ciencias Forenses creó el registro de personas desaparecidas en el SIRDEC y se creó el Banco de Perfiles Genéticos de Desaparecidos por el Estado y bajo la dirección de la Fiscalía General de la Nación para buscar que las demás instituciones del Estado, que trabajan en el área forense, tengan acceso delimitado y controlado a la información del Banco de Perfiles Genéticos de Desaparecidos que se requiera para el desarrollo de los procesos de identificación de víctimas de desaparición forzada a su cargo. Así mismo los laboratorios estatales de Genética Forense deberán procesar, indexar, organizar e ingresar al Banco de Perfiles Genéticos de desaparecidos toda la información obtenida del análisis de los perfiles genéticos de los cuerpos y restos de las víctimas, así como de las muestras biológicas de referencia de los familiares que hayan sido genotipificados, quienes de manera voluntaria, mediante la firma de un consentimiento informado, podrán autorizar la toma de muestra, el procesamiento de estas para obtener los perfiles genéticos y el ingreso de sus datos para ser comparados con los obtenidos de los restos incluidos en el Banco de Perfiles Genéticos de Desaparecidos para lograr así la identificación de los desaparecidos reportados y las víctimas del conflicto armado (76).

En 2016, el gobierno realiza un nuevo acercamiento para alcanzar la paz con la guerrilla de las FARC, la más antigua, la más grande en Colombia y con la cual se habían tenido más de

cincuenta años de conflicto armado. El gobierno negocia con esta guerrilla un Acuerdo de Paz que busca poner fin a la violencia con esa guerrilla (20). El acuerdo busca impedir que haya más víctimas y concentrar todos los esfuerzos en construir una paz estable y duradera, aunque no todos los miembros de esta guerrilla se acogen a este acuerdo, surgiendo así grupos disidentes que a la fecha no han permitido que el país logre alcanzar la tan anhelada paz. En este acuerdo de paz, la guerrilla de las FARC se comprometieron a entregar todas sus armas a las Naciones Unidas, a no incurrir en delitos como el secuestro, la extorsión o el reclutamiento de menores, a romper sus vínculos con el narcotráfico y a cesar los ataques a la fuerza pública y a la población civil. También se estableció que habría verdad, justicia y reparación para las víctimas y que, a partir de su aprobación por los ciudadanos, las FARC harían política sin armas. Adicionalmente el acuerdo incluyó un plan de desarrollo agrario integral con acceso a tierras y servicios y una estrategia de erradicación y sustitución de cultivos ilícitos (20). En el marco de este acuerdo se creó la unidad para la búsqueda de personas reportadas como desaparecidas por el conflicto armado, que es una unidad especial de carácter humanitario y extrajudicial, que busca establecer lo ocurrido a estas personas desaparecidas por el conflicto armado, identificarlas, hacer entrega de sus restos a sus familiares y de esa manera contribuir a satisfacer los derechos de las víctimas a la verdad, la justicia y la reparación integral (20).

4.3 La prueba de ADN en la administración de justicia

La medicina forense ha tenido un gran avance desde la incorporación de la huella de ADN descubierta por Alec Geffrey en 1981, ésta dio entrada a la genética en la administración de justicia con el surgimiento de la llamada *prueba de ADN*, una técnica que cambiaría para siempre el campo de la medicina forense. A partir de allí han sido muchos los marcadores genéticos introducidos

para resolver casos forenses y las técnicas analíticas desarrolladas para su genotipificación (5,52,77).

Los desarrollos de la genética forense utilizando el ADN, han sido enormes. En principio cualquier región del genoma humano puede ser usada para la investigación genética, sin embargo, en la práctica se utilizan marcadores ubicados en las regiones no codificantes porque al no sufrir fuertes efectos de selección suelen presentar valores más altos de diversidad genética, sumado al hecho de que la información obtenida a partir de ellos no está directamente asociada a enfermedades o a la susceptibilidad genética del individuo, lo que hace que sean éticamente recomendados. Adicionalmente entre los criterios de selección de estos marcadores está que sean polimórficos para que presenten una mayor variación entre los individuos de una población y tengan un mayor poder de discriminación, favoreciendo la posibilidad de identificar inequívocamente un individuo cualquiera de entre muchos que conviven en el mismo lugar, donde podría estar el responsable del delito o el padre biológico de un menor que se está interesado en identificar (78,79).

De estos marcadores, son los STRs autosómicos los más ampliamente utilizados por los laboratorios forenses a nivel mundial para resolver los casos de filiación e identificar un individuo, porque los resultados del perfil genético obtenidos con ellos son muy objetivos, no son modificables a lo largo de la vida y son directamente heredables a los descendientes. Adicionalmente por ser los marcadores más utilizados se han desarrollado múltiples paneles o kits comerciales que amplifican simultáneamente hasta 23 marcadores STRs en una sola reacción de PCR, facilitando así su uso, optimizando los tiempos para el análisis de las muestras y aumentando

su poder de discriminación. Son los marcadores de elección para resolver los casos en el trabajo rutinario de los laboratorios al resolver las pericias que realizan (54,55,80,81).

Sin embargo, los laboratorios forenses reciben casos complejos que incluyen situaciones como no disponer de muestras de los miembros del grupo familiar directo y requerir la reconstrucción del perfil del progenitor ausente, o casos de identificación o criminalística donde se desconoce a quien pertenecen unos restos hallados. Algunos de estos casos no se logran resolver solo con la evaluación de los STRs autosómicos porque solo se cuenta con familiares por línea paterna o materna o las muestras recibidas no son de la mejor calidad, no están en buena cantidad, los vestigios encontrados son pelos o lo único que se tiene para el análisis es una mancha con esperma del violador. En estas últimas situaciones es importante valorar otro tipo de marcadores ubicados en el cromosoma Y y/o en el ADN mitocondrial (ADNmt) (14,15).

4.4 Marcadores genéticos uniparentales

Son marcadores que reflejan la herencia de solo uno de los progenitores, debido a su modo de herencia: los del cromosoma Y de herencia paterna y el ADN mitocondrial de herencia materna. Por esta razón no sufren procesos de recombinación, sus variaciones solo pueden darse a través de las mutaciones que sufran y sus marcadores se heredan en bloque, constituyendo el denominado haplotipo. Debido a la ausencia de recombinación no permiten el mismo grado de individualización de los marcadores autosómicos, solo permiten la discriminación entre linajes de individuos (14,15,82,83).

En los casos complejos que se presentan en forenses, estos marcadores uniparentales se constituyen en la herramienta valiosa para el análisis complementario, o en otros casos son los únicos marcadores susceptibles de tipificarse debido a las malas condiciones de las muestras

recibidas, el tipo de vestigio a ser analizado, la cantidad de material biológico disponible, la presencia de mezclas entre víctima y agresor, o la poca disponibilidad de familiares directos de un caso de filiación a resolver (14,15).

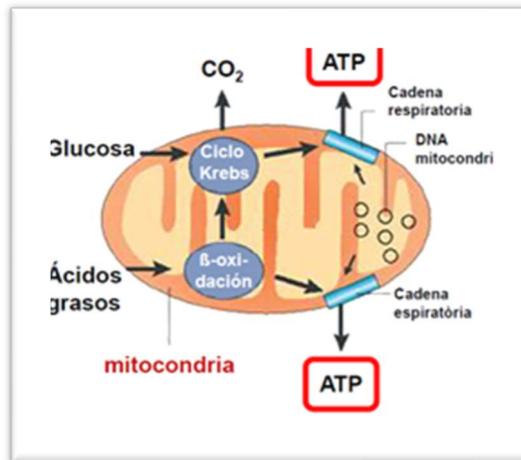
El análisis de marcadores del cromosoma Y es ampliamente utilizado en genética forense para resolver la mayoría de las agresiones sexuales y en crímenes violentos donde los agresores son casi siempre varones. En la rutina forense es común tener fluidos o tejidos biológicos donde el ADN de la víctima y el del agresor(es) están mezclados y aunque existen técnicas para realizar un análisis diferencial en estas mezclas, la extracción diferencial no siempre es posible, no se logra separar los componentes debido a la proporción de los contribuyentes en la mezcla, especialmente cuando el aporte seminal es más bajo, que es lo más común. También es muy útil cuando las agresiones han sido cometidas por más de un varón puesto que de hallar más de un alelo en alguno(s) marcadores del cromosoma Y, establecería el aporte de más de un individuo a la mezcla. En estos casos, así el aporte masculino esté en una proporción baja, es posible realizar el análisis de los marcadores del Y para luego compararlo con el perfil(es) del(os) sospechoso(s) (14).

En las relaciones de parentesco, los marcadores del Y son útiles cuando se está resolviendo un caso complejo y los resultados obtenidos con los marcadores autosómicos no son concluyentes, así mismo cuando el presunto padre está fallecido y se dispone de familiares del presunto padre por línea masculina que sean varones; teniendo siempre en cuenta que el análisis de los marcadores del cromosoma Y solo es útil en filiación, si el caso a resolver involucra un descendiente masculino del presunto padre y que no servirían para realizar exclusión, si hay más de un presunto padre implicado y pertenece al mismo linaje paterno del padre biológico (14,84).

Por su parte el ADNmt tiene un papel protagónico en la genética forense desde que fue secuenciado por primera vez en 1981 por Anderson y colaboradores (85), en cuatro situaciones forenses específicas: a) cuando las muestras a analizar están degradadas (como restos óseos, piezas dentales, etc.) b) su cantidad es mínima (el número de moléculas de la mitocondria es elevado y va de 200 a 2700) lo cual facilita desde el inicio la obtención de resultados en muestras limitantes, c) cuando solo existe éste ADN en la muestra (tallos de los pelos), porque es el único polimorfismo que se puede analizar en cabellos sin bulbo y el pelo es un vestigio que aparece con mucha frecuencia en la escena de delitos y finalmente d) cuando es necesario establecer algún tipo de parentesco por vía materna porque no se dispone de los progenitores y se cuenta con familiares cercanos por vía materna (15). También ha sido de utilidad cuando existe un sospechoso de un hecho delictivo, pero no se cuenta con muestra indubitada del mismo, pudiéndose recurrir al estudio del ADNmt de un familiar del sospechoso relacionado matrilinealmente para excluirlo. En todos estos casos la probabilidad de obtener resultados en estos indicios biológicos es más alta con el análisis del ADNmt que si se hubiesen analizado los marcadores STR's autosómicos convencionales (15).

4.5 ADN mitocondrial (ADNmt)

El ADNmt es una molécula que se encuentra dentro de las mitocondrias, que son organelas celulares especializadas en la producción de energía como producto de realizar la función de la respiración celular y producir ATP para el funcionamiento perfecto del organismo (15). Estas ocupan hasta el 20% del volumen de la célula, contienen dos membranas una externa y una interna que las separan en cuatro compartimentos distintos el externo, el interno, el espacio intermembranal y la matriz mitocondrial (Figura 7).

Figura 7. Estructura de la mitocondria.

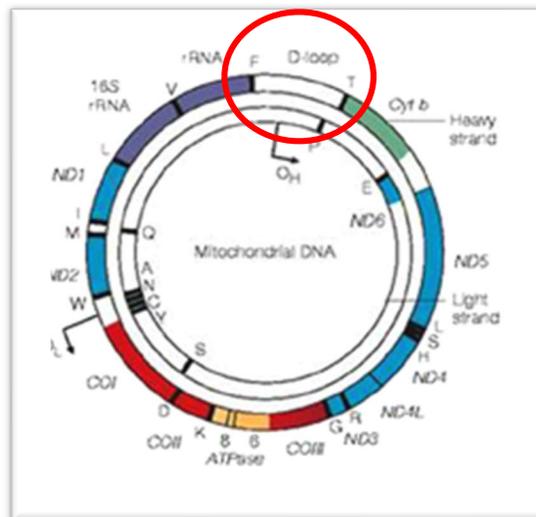
Nota. Tomado de Taylor, R., Turnbull, D. Mitochondrial DNA mutations in human disease. Nat Rev Genet 6, 389–402 (2005). <https://doi.org/10.1038/nrg1606>

La membrana interna es muy plegada y posee numerosas ondulaciones o crestas que son muy importantes porque allí se lleva a cabo el transporte de electrones a través de varios complejos proteicos que dan la fuerza motriz a algunas enzimas para la síntesis de ATP que es la molécula que almacena la energía producida en la fosforilación oxidativa y, que es indispensable para llevar a cabo las tareas básicas del metabolismo de las células que requieren energía para realizarse. Además, poseen ADN como parte de su composición y este constituye entre el 1 al 2% del ADN total de la célula y codifica para varias de las proteínas que conforman los complejos de la cadena respiratoria (15).

4.5.1 Estructura

El ADNmt es una molécula circular de doble cadena conformada por 16.569 pares de bases. Fue secuenciada por Anderson y cols en 1981 y desde entonces la secuencia de este ADNmt se utiliza como secuencia de referencia para los análisis de cualquier ADNmt. Fue secuenciada nuevamente por Andrews y cols en 1999 y se incorporaron algunas modificaciones a la secuencia original, por ello hoy en día se denomina secuencia de referencia revisada de Cambridge (rCRS) (86). Las dos hebras del ADNmt poseen una distribución asimétrica de guaninas y citosinas lo que genera una hebra pesada (H) y una liviana (L). La cadena pesada (H) es más rica en Guaninas y Adeninas y la liviana (L) contiene más Citocinas y Timinas. (Figura 8) (15, 87).

Figura 8. ADN mitocondrial (ADNmt)



Nota. Tomado de Taylor, R., Turnbull, D. Mitochondrial DNA mutations in human disease. *Nat Rev Genet* 6, 389–402 (2005). <https://doi.org/10.1038/nrg1606>

El ADNmt en casi su totalidad, es codificante, posee información para la síntesis de alrededor del 10% de las proteínas encargadas de las funciones de respiración que lleva a cabo la

mitocondria. Tiene 37 genes que codifican para 13 proteínas, 2 rRNAs y 22 tRNAs y solamente una pequeña región de su estructura es no codificante (alrededor de un 10%) allí no posee genes, se denomina región control o D-loop y mide 1100 pb (Figura 8). Esta región a su vez se subdivide en tres regiones, denominadas regiones hipervariables I, II y III. La región hipervariable I (HVS I) se extiende desde el nucleótido 16024 al 16365 (342pb), la región hipervariable II (HVS II) se extiende desde el nucleótido 73 al 340 (268pb) y la región hipervariable III (HVS III) se extiende desde el nucleótido 438 al 574 (137pb) (88,89). Esta región es de utilidad en forense y se analiza mediante la secuenciación de las 1100 pb que conforman la región control completa, incluyendo las tres regiones hipervariables HVS I, II y III, para valorar los polimorfismos ubicados allí y establecer su correspondencia o no con las muestras valoradas en el caso (16,22,23,24).

La mitocondria utiliza esta región control como sitio de inicio del proceso de replicación de la molécula de ADN, específicamente de la cadena H y al presentar una mayor variación en su secuencia, genera mucha variabilidad entre los individuos, por lo que hoy es utilizada no solo en *genética forense*, especialmente para análisis criminalísticos y estudios de identificación humana en muestras degradadas o con escaso contenido de ADN nuclear, sino en *genética de poblacionales* para determinar la ancestría de las muestras analizadas y establecer sus orígenes etno-geográficos y en *antropología*, como un fuente muy importante de información acerca de la estructura genética, edad y origen del hombre moderno (15).

4.5.2 Características

El ADNmt posee una serie de características que le permiten ser evaluado, como ya se explicó, en ciertas situaciones forenses en las que los marcadores de uso rutinario (STRs autosómicos) no arrojan resultados. Entre estas están: la presencia de gran cantidad de moléculas

en cada célula, ser menos susceptible a la degradación por parte de microorganismos, tener una alta tasa de mutación, así como su herencia uniparental materna. (15,16,87,88).

4.5.2.1 Número de moléculas. El número de moléculas por cada célula es elevado (cada célula tiene de cientos a miles de mitocondrias y cada mitocondria tiene entre dos a diez moléculas de ADN mitocondrial), lo que facilita el aislamiento de al menos una molécula en la muestra a analizar, por mínima que sea su cantidad. Esto le confiere una gran ventaja frente al genoma nuclear, del cual solo se cuenta con una copia paterna y una materna por cada célula (15).

4.5.2.2 Resistencia a la degradación. Su forma circular y su tamaño pequeño le confieren una mayor resistencia a la acción de las endonucleasas que actúan degradando el ADN, porque a estas enzimas se les dificulta más acceder al ADNmt y difícilmente logran degradar todas las moléculas presentes en una muestra, esto le confiere una mayor resistencia a la degradación que el ADN nuclear (15).

4.5.2.3 Tasa de mutación. El ADNmt tiene la capacidad de replicarse en forma independiente del ADN nuclear lo que genera una mayor probabilidad de mutar, esto favorece la diversidad. La tasa de mutación estimada para el ADNmt es de 2.06×10^{-6} , es decir hasta 10 veces mayor que la tasa de mutación del ADN nuclear (90). El ADNmt muta más rápido que el ADN nuclear debido a que es muy sensible a la acción de los radicales libres generados durante la fosforilación oxidativa que realizan las mitocondrias, la oxidación generada por estos radicales provocan efectos mutagénicos y actúan con mayor facilidad sobre la molécula porque ésta no está unida a proteínas de tipo histonas y por lo tanto no está protegido y tiene una menor compactación molecular. Aunado a esto, su tasa de mutación es alta porque la ADN polimerasa mitocondrial

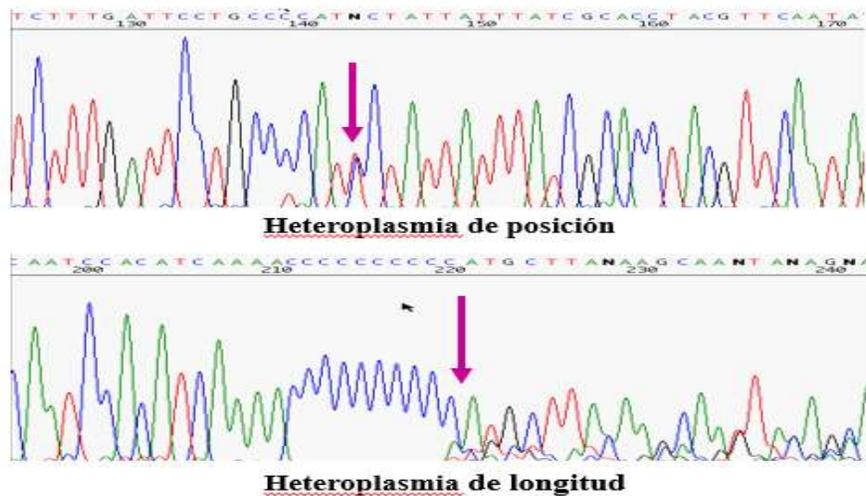
tiene una pobre actividad correctora por lo que se mantienen los errores cometidos durante la replicación del ADNmt (15,90). La región control del ADNmt, por ser una región no codificante, acumula la mayor parte de estas mutaciones porque se pueden fijar allí sin producir ningún efecto deletéreo para la molécula como sí podría ocurrir en las regiones codificantes. Esto genera que la región control sea la más polimórfica del ADNmt y por tanto la de mayor interés para ser analizada en genética forense, mediante metodologías como la secuenciación, incluyendo las tres regiones variables que la conforman obteniéndose la secuencia de 1100 pb comprendidas entre los nucleótidos que ocupan las posiciones 16024 a la 574 (15, 88). La región HVSI posee nucleótidos más propensos a la mutación que otros en las posiciones 16093, 16129, 16153, 16189, 16192, 16293, 16309 y 16337, por su parte la región HVSII presenta más mutaciones de los nucleótidos 72, 153, 189, 207 y 279 y la región HVSIII presenta más mutaciones en una zona repetitiva de dinucleótidos AC entre las posiciones 514 a 524, cuyo tamaño más común es de 5 repeticiones AC, pero se pueden encontrar variaciones entre 3 a 7 repeticiones, lo que permite utilizar esta variabilidad con fines de discriminación (15,88,89).

4.5.2.4 Herencia. Las mitocondrias de un cigoto recién formado provienen exclusivamente del óvulo fecundado debido a que el espermatozoide concentra la mayoría de sus mitocondrias en el cuello para generar la energía que requiere para su desplazamiento en busca del óvulo y al fecundarlo solo ingresa su cabeza, dejando fuera sus mitocondrias (15). Por ello Al analizar el ADNmt en genética forense se requiere tener presente que el genoma de la mitocondria es haploide debido a su herencia estrictamente materna, por lo tanto, como no sufre procesos de recombinación los datos obtenidos de la tipificación del ADNmt no pueden ser aplicados a la identificación de individuos de la misma manera que se aplican los análisis de los marcadores de ADN nuclear que

recombinan. En este caso todas las diferencias se heredan en bloque, conforman un haplotipo y por tanto se asumen como un único marcador. A pesar de ello la región control del ADNmt es uno de los marcadores genéticos más discriminatorios en forense por la alta tasa de mutación de esta región que origina mutaciones puntuales tipo sustituciones, inserciones o deleciones, también por el hecho de ser una región de tamaño pequeño lo que facilita su amplificación mediante la Reacción en Cadena de la Polimerasa (PCR) y posterior secuenciación y por el hecho de que cada célula posee entre 1000 y 10000 moléculas de ADNmt, lo que también favorece que se encuentren moléculas de ADNmt sin degradar, en muestras antiguas o sometidas a la acción de bacterias, humedad, etc (15).

4.5.2.5 Heteroplasmia. Se denomina heteroplasmia a la presencia de más de un genotipo o secuencia de ADNmt en un individuo, puede ser intramitocondrial, intracelular o intratisular y es derivada de las altas tasas de replicación y de mutación del ADNmt que facilitan la aparición de cambios en la secuencia y por tanto encontrar una o más diferencias en las secuencias de ADNmt en la rama materna de una familia, incluso en un mismo individuo, pueden ser de posición o de longitud (Figura 9).

Figura 9. Heteroplasmas en el ADNmt



De esta manera cuando una mutación surge en una molécula de ADNmt de una mitocondria, genera una mezcla intracelular de moléculas mutantes y normales. La proporción de moléculas mutantes en los tejidos de un individuo heteroplásmico puede variar debido a que en la división de una célula heteroplásmica la distribución de mitocondrias en sus descendientes es al azar, esto podría explicar porque se pueden hallar diferentes estados de heteroplasmia entre individuos de un mismo linaje materno o entre tejidos de un mismo individuo. Por esta última razón es importante tener en cuenta esta característica en el análisis forense al momento de realizar comparaciones entre individuos que sean familiares por línea materna o entre vestigios de un mismo individuo pues si bien su hallazgo en una secuencia no invalida su análisis, si requiere la comprobación de que es una heteroplasmia y no una exclusión (15,16,22,24).

Las heteroplasmas pueden ser de posición o puntuales cuando se debe al cambio de un único nucleótido que se sustituye por otro en una determinada posición del genoma mitocondrial, para confirmarla las guías de la ISFG recomiendan realizar la secuenciación de las dos hebras, usando diferentes cebadores para observar en ambos casos la misma heteroplasmia, adicionalmente

tener en cuenta que la lectura de la secuencia sea clara y que la altura del pico de la base menor sea de al menos el 20% del área total (22,24).

Por su parte se denominan heteroplasmia de longitud o de tamaño cuando se ganan o pierden nucleótidos en un tracto homopolimérico, lo que puede ocurrir en las posiciones de la región control donde hay repeticiones mononucleotídicas o dinucleotídicas. Son mucho más frecuentes que las heteroplasmas de posición. Entre las más comunes está en la región HVSI, la que se produce en el segmento homopolimérico entre las posiciones 16183 y 16193 que se genera por la transición de una timina a una citosina en la posición 16189, este tracto de 10 citosinas favorece el deslizamiento de la polimerasa y la aparición de segmentos de distinta longitud en esta región lo que dificulta la lectura e interpretación de la secuencia en el electroferograma y generalmente requiere de la secuenciación de esta región con un cebador contrario para poder realizar la lectura del segmento posterior a la posición 16193. También son frecuentes en la región HVSII entre las posiciones 303 y 310 (15,22,24).

4.5.3 Nomenclatura

La ISFG ha emitido guías en 2007 y 2014 para el trabajo con ADNmt que incluyen recomendaciones para utilizar una nomenclatura unificada que permita establecer criterios únicos e internacionales para la interpretación de los resultados del análisis de ADNmt (22,24). La nomenclatura recomendada por la ISFG se basa en los códigos IUPAC-IUB que utiliza letras mayúsculas para definir cada una de las bases nitrogenadas o la mezcla de dos, tres o las cuatro de ellas según se puedan observar en una misma posición. Por ejemplo, una mezcla de A y G debe ser representado por una R, mientras que una mezcla de C y T debe ser representado por una Y. Junto con la letra asignada, se debe especificar la posición concreta dentro del genoma mitocondrial

donde se halla la heteroplasmia. Por su parte, las discrepancias con respecto a la rCRS se indican siguiendo estas recomendaciones: Las sustituciones se indican con la posición seguida de la base detectada, por ejemplo, 73G, las deleciones se registran con la posición seguida del sufijo del en minúscula, por ejemplo, 523del y las inserciones se registran con la posición seguida de un punto, el número de la inserción y la base insertada, por ejemplo, 16193.1C, teniendo en cuenta la recomendación de la ISFG para análisis forense, de alinear las inserciones y deleciones en la posición 3' en relación con la posición afectada en secciones de secuencia homopoliméricas y en otras regiones de repetición (24). En caso de presentarse heteroplasmas de posición, como se explicó se usa la nomenclatura establecida por el código IUPAC que permite definir múltiples posibilidades de nucleótidos en una posición y asignar la letra correspondiente, de acuerdo con la mezcla heteroplasmática detectada, siguiendo las recomendaciones de la ISFG. (24).

4.5.4 Análisis del ADNmt en genética forense

Las anteriores características del ADNmt le permiten tener un papel protagónico en cuatro situaciones forenses específicas: a) cuando las muestras a analizar están degradadas (como restos óseos, piezas dentales, etc.) b) si la cantidad de muestra disponible para el análisis es mínima (como se expuso, el número de moléculas de la mitocondria es elevado y va de 200 a 2700) lo cual facilita desde el inicio la obtención de resultados en muestras limitantes, c) cuando solo existe éste ADN en la muestra (tallos de los pelos) porque es el único polimorfismo que se puede analizar en cabellos sin bulbo y el pelo es un vestigio que aparece con mucha frecuencia en la escena de delitos y d) cuando es necesario establecer algún tipo de parentesco por vía materna porque no se dispone de los progenitores y se cuenta con familiares cercanos por vía materna (15). También es de utilidad cuando existe un sospechoso en un hecho delictivo, pero no se cuenta con muestra indubitada del

mismo, en esta situación se puede recurrir al estudio del ADNmt de un familiar del sospechoso relacionado matrilinealmente para excluirlo. En todos estos casos la probabilidad de obtener resultados en estos indicios biológicos es más alta que si se hubiesen analizado los marcadores STR's autosómicos convencionales (15).

Es muy importante sin embargo tener presente que las ventajas del análisis de ADNmt especialmente obtener resultados a partir de muestras en mínima cantidad y calidad, requiere de un cuidadoso control de calidad en los laboratorios forenses, pues la capacidad de obtener resultados con muestras mínimas también le confiere la desventaja de ser muy susceptible a sufrir contaminación y de ahí la necesidad de que los laboratorios de genética forense estén acreditados y cuenten con un sistema de calidad que garantice los resultados obtenidos tomando medidas como la separación de áreas de pre y post PCR, la inclusión de controles negativos y positivos en cada una de las etapas del análisis, la participación en ejercicios intercomparación que le permitan verificar la veracidad de sus resultados, la tipificación de los funcionarios para facilitar la detección de contaminaciones, entre otras medidas (15).

4.5.4.1 Métodos de análisis. El análisis de ADNmt con fines forenses puede realizarse a partir de la valoración de la región control completa del ADNmt (1100 pb) o más recientemente de todo su genoma, también denominado mitogenoma (16569 pb) y para ello pueden ser utilizados diferentes abordajes metodológicos. En la actualidad la mayoría de los laboratorios forenses que analizan muestras para ADNmt las evalúan por la metodología de secuenciación por Sanger con marcaje de las bases con fluorocromos para ser detectadas por electroforesis capilar (EC). Sin embargo, en los últimos años se ha incorporado la Secuenciación de Próxima Generación (NGS)

también llamada Secuenciación Masiva en Paralelo (MPS) utilizando plataformas para realizar la secuenciación con diferentes químicas que no se basan en el método Sanger (91,92,93).

En la secuenciación por Sanger se realiza una PCR directa que permite la amplificación del segmento de ADN mitocondrial correspondiente a la región variable completa, cuando las muestras son de óptima calidad, si por el contrario la muestra disponible tiene escaso ADN o está degradado, pueden requerirse varias amplificaciones hasta lograr secuenciar toda la región control, por esto es muy importante en estos casos implementar controles de calidad porque estas muestras tienen la desventaja de ser más susceptibles de contaminarse con un ADN diferente lo que llevaría a obtener resultados erróneos (15).

Una vez realizada la PCR, la secuenciación del fragmento amplificado puede detectarse por varias metodologías de las cuales la más usada es la secuenciación automatizada usando fluorocromos para marcar e identificar cada una de las bases nitrogenadas y así establecer el orden de los nucleótidos en la secuencia al realizar una EC. También puede establecerse la secuencia en el producto amplificado utilizando la metodología NGS, de reciente introducción en algunos laboratorios forenses y aunque el ADNmt fue de los primeros marcadores en analizarse con esta metodología, aun es de poco uso en la mayoría de los laboratorios forenses entre otras razones por los elevados costos de las plataformas o equipos, los reactivos e insumos que se utilizan, así como por la necesidad de implementar un soporte bioinformático para analizar los resultados que se generan por lo que se requiere realizar una amplia validación y evaluación de los resultados obtenidos con esta metodología antes de que pueda ser implementada en un laboratorio forense de manera completamente confiable. No menos importante es la necesidad de desarrollar nuevas bases

de datos para los cálculos estadísticos del informe de resultados forenses ya que las existentes no son informativas para el nivel de resolución molecular obtenido con esta metodología (91,94).

En el futuro muy probablemente la implementación de la metodología de secuenciación masiva en paralelo (MPS) o secuenciación de próxima generación (NGS) será una realidad en los laboratorios forenses porque permite el análisis de los marcadores clásicamente empleados en forense como son los STRs autosómicos, como para analizar marcadores tipo SNPs del cromosoma X, del cromosoma Y, Indel y para el análisis del ADNmt, permitiendo obtener resultados de cientos o miles de marcadores de ADN diferentes en forma simultánea (91). Las tecnologías actualmente usadas en forense y que se están validando son el *Miniseq FGx de Illumina* que se basa en la secuenciación por síntesis en fase sólida (SBS) mediante terminación cíclica reversible y la detección de los dNTPs marcados fluorescentemente y el *Ion GeneStudio S5 de Thermo Fisher* basada en SBS en chip, mediante la adición de un único nucleótido con la detección de los cambios de potencial, derivado de los cambios de pH al incorporarse cada dNTP (92).

También se verá favorecida la implementación de la tecnología MPS en el caso específico del análisis forense del ADNmt porque se pueden analizar una gran cantidad de muestras al mismo tiempo y obtener tanto la secuencia de la región control como la del genoma completo de ADNmt con una buena resolución y en una sola reacción de secuenciación, para ello se han diseñado kits comerciales como el *Forenseq mtDNA Whole Genome*, que usando la plataforma Illumina, permite secuenciar todo el genoma mitocondrial y obtener el llamado mitogenoma con la secuencia de los 16569 nucleótidos generando un nuevo nivel de discriminación al no tener solo el haplotipo de la región control como referente para la comparación de perfiles sino determinar todas las diferencias presentes a lo largo del genoma mitocondrial al comparar las secuencias de dos muestras. Este considerable incremento del poder de discriminación al analizar el mitogenoma, así como la

posibilidad de generar pequeños amplicones solapantes podría mejorar el proceso de identificación en casos de personas desaparecidas o en sucesos de grandes catástrofes con estas plataformas MPS llevando a resolver muchos de los problemas actuales de la casuística forense, especialmente en el caso de muestras con mezcla de ADNs donde la MPS permite la obtención de resultados de los contribuyentes individuales sin ruidos de fondo y revela información genética que no se puede establecer usando secuenciación por CE (91,95).

4.5.4.2 Secuencia de referencia. La secuencia de referencia que se utiliza en todos los laboratorios forenses del mundo para el análisis de resultados de ADNmt fue secuenciada en 1981 por Anderson y cols en la Universidad de Cambridge (85,86) y fue la primera secuencia de ADNmt humano obtenida a partir de una mujer de descendencia europea y por lo tanto representa un linaje filogenético moderno del ADN mitocondrial humano (24,85). En la actualidad se denomina Secuencia de referencia de Cambridge corregida porque fue revisada en 1999 por Andrews y cols y sufrió algunas correcciones (86). Cuando se informan los haplotipos de ADNmt en relación con el rCRS, se debe reportar el rango de la secuencia (excluyendo la información de secuencia del cebador) para que se pueda entender y reconstruir la secuencia completa de la muestra (22,24).

4.5.4.3 Interpretación de resultados. Los resultados obtenidos de la secuenciación de cada muestra se comparan con la secuencia de referencia de Cambridge revisada (rCSR), para establecer las diferentes posiciones de la secuencia en las que difiere la muestra con respecto a la rCSR y determinar su haplotipo (15). Después se comparan los resultados obtenidos de los individuos implicados en un caso de filiación, o en las muestras problema (dubitadas) y las muestras de referencia (indubitadas) si es un caso de criminalística, para establecer si hay coincidencia en toda la secuencia de las muestras analizadas o si por el contrario aparecen discrepancias entre las

secuencias de las muestras. Una vez se realiza la comparación se pueden obtener tres tipos de resultados: *Exclusión*, *no se puede excluir* y *no concluyente*. Se considera *exclusión* cuando se comparan los resultados de las dos secuencias y los haplotipos discrepan en dos o más posiciones entre sí, *no concluyente* cuando al comparar las dos secuencias los haplotipos discrepan en una posición nucleotídica y *no exclusión* cuando los dos haplotipos son idénticos, es decir coinciden en toda su secuencia. En este último caso procede realizar los cálculos matemáticos para obtener el peso estadístico de los resultados genéticos hallados teniendo en cuenta que el resultado se trata como un haplotipo (15).

4.5.4.4 Valoración estadística de resultados. Una vez se tienen los resultados de no exclusión para el análisis del ADNmt, es necesario realizar la valoración matemática en el contexto forense. Para ello se requiere consultar las bases de datos de la población donde ocurrió el caso, determinar la frecuencia de ese perfil en esa población y realizar la valoración estadística. El parámetro matemático-estadístico a aplicar, recomendado por la ISFG, es *la razón de verosimilitud (RV) o likelihood ratio (LR)* que enfrenta dos hipótesis o proposiciones alternativas:

$$\text{Fórmula 1} \quad LR = \frac{\text{Pr}(E / H_0)}{\text{Pr}(E / H_1)}$$

Donde:

H₀: los restos biológicos encontrados proceden de un familiar del individuo por vía materna.

H_1 : los restos biológicos encontrados proceden de un individuo tomado al azar no relacionado genéticamente con el individuo, es decir, tienen linajes maternos diferentes.

Si el valor obtenido de este cociente es $\bar{p} > 1$ indica que la evidencia apoya la primera hipótesis (H_0) y si el valor es < 1 , la evidencia está a favor de la segunda hipótesis (H_1).

Dada la diversidad de los haplotipos del ADNmt, es posible que al evaluar un caso de coincidencia se obtengan haplotipos que no estén representados en la población de referencia usada, o se tenga un número reducido de secuencias registradas para esa población, situación que exige la aplicación de un parámetro estadístico de corrección para el cálculo de la frecuencia haplotípica (15). Uno de los más usados es el método de Balding y Nichols que corrige un error de muestreo sumando 1 a la base de datos consultada (el valor 1 procede de la coincidencia de haplotipos hallada entre las dos muestras que se están comparando) de manera que la probabilidad de encontrarlo sería:

Fórmula 2

$$p = x + 1/n + 1$$

Donde x es el número de observaciones de ese haplotipo en la base de datos y n es el tamaño poblacional de esta base de datos.

El valor de LR obtenido se debe informar en los resultados del dictamen pericial expresando que las secuencias de ADNmt que definen al individuo 1 y al 2 coinciden entre sí. Esa coincidencia

es X veces más probable si el individuo 1 y el 2 comparten una relación por vía materna que si entre ellos no existe ninguna relación de parentesco por dicha vía (15).

Por esta razón es importante contar con bases de datos con un amplio número de haplotipos contribuyentes que evite sobrestimar la valoración estadística del caso ya que la frecuencia del haplotipo en cuestión tiene implicaciones directas en el cálculo del LR que se presenta en el reporte de resultados y por ende en la toma de decisiones por parte del juez a partir del informe pericial recibido, más aún cuando no se dispone de ninguna base de datos con las frecuencias de los haplotipos presentes en la población donde se está realizando la pericia (5,24). Por esta razón, grupos como el de habla española y portuguesa (GHEP) de la Sociedad internacional de Genética Forense (ISFG) han publicado estudios colaborativos para poner al servicio de las pericias judiciales bases de datos amplias de lugares como la península ibérica (España y Portugal) y de varias poblaciones del Brasil, con la intencionalidad de mejorar la disponibilidad de perfiles genéticos de la región variable completa del ADN_m para resolver casos forenses en estos países, que es especialmente útil cuando el haplotipo hallado en los individuos genotipificados, es raro en la población (23).

4.5.4.5 Bases de datos de ADN_{mt}. Las bases de datos forenses son un conjunto de información sobre los perfiles genéticos obtenidos de las muestras analizadas en forense, que se almacenan de forma ordenada y coherente en softwares informáticos para su consulta y que contienen no solo la información de los perfiles genéticos, sino todo dato asociado a las muestras, información que puede ser recuperada y comparada de acuerdo con parámetros previamente establecidos (48). Como ya se expuso, es imprescindible contar con bases de datos poblacionales para poder valorar los resultados de los análisis de ADN_{mt}. Estas desempeñan un papel clave en

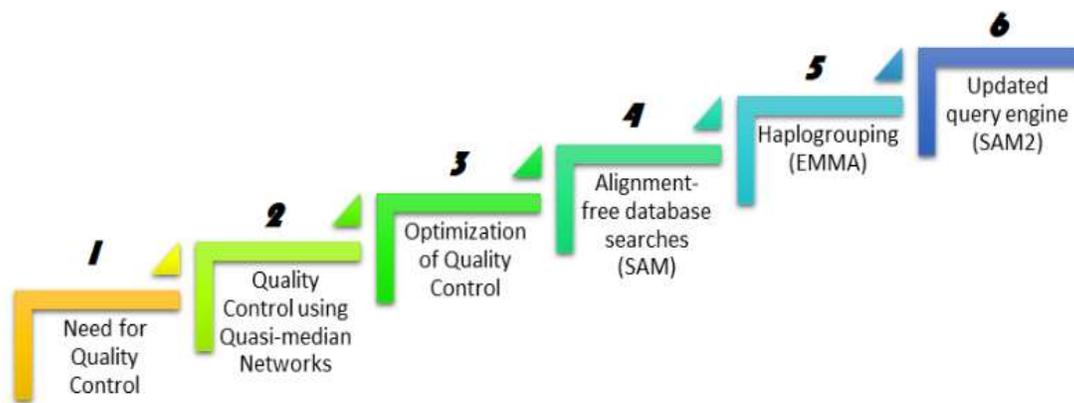
la estimación de las frecuencias de los distintos haplotipos cuando se encuentra coincidencia entre dos muestras analizadas y se debe realizar la valoración estadística (15).

En el 2014 la ISFG publicó las últimas recomendaciones de la comisión de ADN con los lineamientos revisados y ampliados para la tipificación de ADNmt para su uso en identificación humana, incluyendo entre otros aspectos la nomenclatura de las secuencias mitocondriales para ser incluidas en las bases de datos de referencia, así como la definición de los estándares mínimos para la inclusión de haplotipos en las bases de datos de ADNmt, que incluye la secuenciación de la región control completa (HVS I, II y III), no solo para la resolución de casos sino para los estudios genéticos poblacionales que se realicen para la creación de las bases de datos (24).

La base de datos mundial más grande de ADNmt es la EMPOP (EDNAP Mitochondrial DNA Population Database), www.empop.org, fue creada hace 20 años con el objetivo de recopilar los datos obtenidos por los laboratorios forenses de todo el mundo, realizarles control de calidad y presentar los datos de una forma fácil para la búsqueda de haplotipos por parte de los laboratorios forenses que lo requieran, es mantenida por el Instituto de Medicina Legal de Innsbruck, Austria (48). Cuenta con la colaboración de los grupos científicos EDNAP (The European DNA Profiling Group), Ge.F.I. (Gruppo Ematologi Forensi Italiani) y GHEP-ISFG (Grupo Español y Portugués de la ISFG) (48). Su repositorio de datos posee en la actualidad 48.572 haplotipos de ADNmt con control de calidad, con la información de al menos la región hipervariable HVS-I (16024-16365) de la RC del ADNmt, de estos 46.963 haplotipos incluyen las regiones HVS-I y HVS-II (16024-16365 73-340), 38.361 poseen información de la RC completa (16024-576) y 4.289 tienen la información de todo el genoma del ADNmt (mitogenomas), consulta realizada el 10/01/23. Esta base ha podido ser conformada gracias a la contribución de datos de más de 60 instituciones de 25

países alrededor del mundo (48). Esta base de datos realiza un exhaustivo control de calidad a las secuencias que solicitan ser incluidas y solo si pasan estos controles de calidad son aceptadas para ser incluidas allí y para ello desarrolló una serie de métodos y herramientas que permiten evaluar los conjuntos de datos de mtDNA de las muestras sometidas a control de calidad (Figura 10)

Figura 10. Métodos aplicados por EMPOP para el control de calidad de muestras de ADNmt



Nota. Tomado de Base de datos EMPOP.online (48)

Otra base de datos importante es el CODIS (Combined DNA Index System) que pertenece al FBI (Buró Federal de Investigaciones), que es la principal rama de investigación del Departamento de Justicia de los Estados Unidos de Norteamérica. Es el programa creado por el FBI para apoyo a las bases de datos criminales de ADN del Departamento de Justicia, junto con el software utilizado para ejecutar estas bases de datos. Cuenta con secuencias de ADNmt tomadas de referencias bibliográficas y de laboratorios forenses, diferenciando las secuencias de los haplotipos de muestras forenses aportadas por laboratorios que siguen protocolos validados, de las aportadas y obtenidas por laboratorios de investigación o de docencia (15).

Como ya se mencionó, en Colombia no se cuenta con bases de datos publicadas de las regiones hipervariables del ADNmt en poblaciones mezcladas que viven actualmente en los centros urbanos del país. Solo hay tres estudios publicados (43,45,46) con fines forenses, pero ninguno evalúa las muestras por secuenciación, solo hay un trabajo que evaluó parte de la región control, (Regiones HVSI y II) en individuos nacidos en los departamentos de la región caribe colombiana (47), pero sus resultados no están publicados en un artículo sino en una tesis y tampoco están disponibles en bases de datos como la EMPOP que siendo la base de datos mundial más grande no tiene incluido ningún dato de la población colombiana mezclada actual de Colombia (Consulta realizada el 10/01/23) (48). Esto representa una limitación para usar el ADNmt en los laboratorios forenses del país puesto que el abordaje estadístico no puede ser realizado por comparación con haplotipos obtenidos de sus propias poblaciones donde ocurrieron los hechos.

4.5.5 Ancestría de los linajes maternos

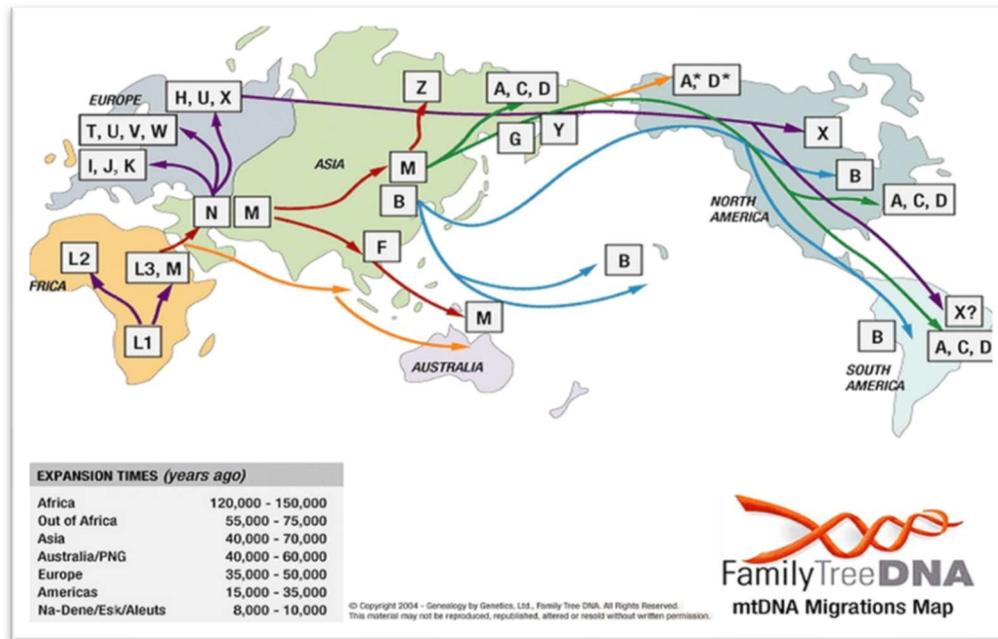
A partir de los haplotipos de ADNmt puede establecerse a qué haplogrupo pertenece cada haplotipo y por tanto su ancestría o linaje materno, el modo de herencia uniparental ha permitido establecer linajes maternos a través de múltiples generaciones a lo largo de la historia de la humanidad y cómo han sido los patrones migratorios humanos durante la colonización de los distintos continentes por el hombre moderno. Un haplogrupo se define como un conjunto de haplotipos que comparten polimorfismos específicos y que tienen un origen común porque están relacionados por descendencia. Los haplogrupos están definidos por polimorfismos específicos que están ubicados tanto en la región control como en la región codificante y que se consideran específicos de ciertas regiones geográficas con lo cual se pueden establecer también relaciones interpopulacionales (15). El conocimiento de la información genética sobre la ancestría de una

muestra es de gran importancia en la persecución de individuos implicados en delitos. La identificación de haplotipos de ADNmt característicos de una determinada zona geográfica junto con el establecimiento de la ancestría del ADN nuclear, permite realizar inferencias que ayudan a orientar la investigación de hechos delictivos (15).

En la actualidad la hipótesis más aceptada acerca del origen y expansión del ADNmt es la conocida como *la Eva mitocondrial* según la cual todos los tipos de ADNmt actuales fueron originados a partir de un ancestro común, este ancestro probablemente vivió en Africa hace alrededor de doscientos mil años (96). El Phylotree build 17, un árbol filogenético preciso, actualizado, detallado y global del ADN mitocondrial humano (ADNmt) que incluye varias aplicaciones forenses de ADNmt, incluido el control de calidad de datos y la asignación de haplogrupos para la inferencia de ascendencia biogeográfica, cuya versión más actualizada es de construido el 18 febrero de 2016 y puede ser consultada en <http://www.phylotree.org>, (97), es ampliamente aceptado como la filogenia de referencia para el ADNmt humano y esencial en forense para el control de calidad de los datos de mtDNA, para estimar las tasas de mutación específicas del sitio y para la asignación de haplogrupos de mtDNA que permitan la inferencia de ascendencia biogeográfica matrilineal (97). PhyloTree utiliza toda la variación de la secuencia disponible a lo largo de todo el genoma de ADNmt de 16,6 kb para construir una filogenia mundial detallada en la que cada rama (haplogrupo) se define por un conjunto específico de mutaciones y ha sido ampliamente adoptado como la filogenia de referencia para el mtDNA humano y sirve como el sistema de clasificación de haplogrupos para EMPOP, HaploGrep y Phy-Mer, entre otros (48, 97).

A partir del conocimiento filogenético se pueden correlacionar la información genética con los hechos históricos, por ejemplo, se han analizado ADNmt de individuos con ascendencia africana reciente en América y en Eurasia para rastrearlos hasta amplias regiones geográficas dentro de África, en gran parte de acuerdo con la evidencia histórica (98). De acuerdo con el árbol filogenético construido por PhyloTree, los haplogrupos mitocondriales más antiguos se encuentran en el África subsahariana y en este continente se encuentra la mayor diversidad mitocondrial formada por los haplogrupos L0 a L6. Se considera que individuos del haplogrupo L3 salieron de África y originaron dos subhaplogrupos denominados macrohaplogrupos euroasiáticos: M y N. A continuación, N divergió hacia el gran subclado R que incluye los haplogrupos B, F, HV, H, J, K, P, T, U y V, y hacia los haplogrupos A, I, S, W, X y Y. Por su parte el macrohaplogrupo M divergió hacia los haplogrupos C, D, E, Q y Z. Posteriormente los haplogrupos A, B, C y D colonizaron Asia y todo el continente americano (Figura 11) (97).

Figura 11. Distribución geográfica de los haplogrupos de ADNmt



Nota. Tomado de: <https://www.familytreedna.com>

4.6 Revisión sistemática de alcance

Una revisión de alcance (*scoping review*) es una forma de síntesis de conocimiento que aborda una pregunta de investigación exploratoria dirigida a mapear conceptos clave, tipos de evidencia, y vacíos en la investigación relacionada con un área o campo definido, para buscar, seleccionar y sintetizar sistemáticamente el conocimiento existente (99,100). Las revisiones de alcance pueden cumplir con varios objetivos: pueden examinar la extensión o tamaño, la variedad y características de la evidencia sobre un tema o pregunta, resumir los hallazgos a partir de un grupo heterogéneo de conocimientos, establecer los métodos empleados para evaluar un tema o identificar lagunas en la literatura para planificar investigaciones futuras (101,102).

Por ello las revisiones sistemáticas de alcance son las más utilizadas en el ámbito de la salud como una herramienta para obtener las síntesis de diferentes áreas dedicadas al cuidado y bienestar

del paciente. Se pueden utilizar para sintetizar diferentes tipos de evidencia de investigación dada la magnitud de conocimiento producto de investigación en salud que se publica diariamente. Desde el 2012 la cantidad de revisiones de alcance publicadas ha crecido exponencialmente, convirtiéndola en la forma de revisión más utilizada (100,101). Las síntesis de conocimiento que se obtienen con las revisiones de alcance son esenciales para avanzar en la práctica y la investigación a través de la consolidación de la evidencia, sus resultados ayudan a los profesionales de la salud, a trabajar más eficientemente y tomar sus decisiones basados en las evidencias (101,102).

Dadas las implicaciones de los resultados de las revisiones de alcance para identificar conocimientos recientes que sean relevantes para la toma de decisiones, estas involucran en su desarrollo un rigor metodológico que permita que los resultados obtenidos respondan de manera correcta y completa a las preguntas formuladas por el investigador (101). Para ello se han diseñado instrumentos que aplican esta metodología siguiendo unos pasos específicos y determinados que permiten a cualquier investigador que la use seguir el mismo protocolo para que el proceso de extracción de datos, análisis y selección de información esté controlado. Así se logra alcanzar una mayor comprensión de la terminología relevante e informar los conceptos y elementos claves como resultado de la revisión de alcance realizada. Esto incluye listas de chequeo con todos los elementos que debe contener la información requerida y la elaboración de diagramas PRISMA (Preferred Reporting Itemsn for Systematic Reviews and Meta-Analyses), como herramientas y modelos para la presentación de los informes. (99,100).

La revisión de alcance realizada como parte de esta tesis siguió el protocolo metodológico propuesto por el Instituto Joanna Briggs (JBI) para la realización de este tipo de revisiones (25).

Como ya se expuso, las revisiones de alcance se pueden utilizar para mapear los conceptos clave que sustentan un área de investigación, así como para aclarar las definiciones de trabajo, y/o los límites conceptuales de un tema y por ello una revisión de alcance puede centrarse en uno de estos objetivos o en todos ellos como un conjunto (100). Para ello el JBI determinó que se debe desarrollar un protocolo a priori, es decir antes de iniciar la revisión, porque es importante definir previamente los objetivos y métodos de la revisión. Este enfoque sistemático para la realización y presentación del informe de la revisión permite la transparencia del proceso y que los lectores vean cómo se llegó a los resultados (25).

El protocolo debe detallar los criterios que los revisores pretenden usar para incluir y excluir estudios, así como para identificar qué datos son relevantes y cómo serán extraídos y mapeados los datos. Para ello incluye:

4.6.1 Información del autor

Todas las revisiones de JBI requieren al menos dos revisores para minimizar el sesgo de informe. Los nombres de todos los revisores y sus afiliaciones institucionales se deben incluir, así como la dirección de correo electrónico del autor de correspondencia (25).

4.6.2 Título, objetivo y pregunta

El título debe ser informativo y dar una indicación clara del tema de la revisión de alcance, debe estar estructurado para reflejar los elementos centrales de la revisión: la población, el concepto y el contexto (PCC). El objetivo debe establecerse claramente y ser congruente con el título y debe indicar lo que busca lograr el proyecto de la revisión de alcance. Por su parte la pregunta guía y dirige el desarrollo de criterios de inclusión específicos para la revisión de alcance. La claridad en

la pregunta de revisión ayuda a desarrollar el protocolo, facilita la eficacia en la búsqueda bibliográfica, y proporciona una estructura clara para el desarrollo del informe final y al igual que con el título, la pregunta debe incorporar los elementos del PCC (25).

4.6.3 Antecedentes

Los antecedentes deben cubrir todos los elementos principales del tema bajo revisión. Dado que las revisiones de alcance son esencialmente exploratorias, no se espera que éste apartado del protocolo cubra el conocimiento existente en el área bajo revisión, sino la razón de emprender la revisión de alcance junto con lo que se pretende lograr con la revisión. Se recomienda que su extensión sea aproximadamente de 1000 palabras y debe incluir información para poner los criterios de inclusión en contexto, incluyendo si existen o no revisiones, revisiones sistemáticas, síntesis de investigación y/o trabajos de investigación primarios disponibles sobre el tema (de haberlas se debe incluir una justificación detallada que especifique en qué se diferenciará la revisión propuesta de las ya realizadas e identificadas). Debe concluir con una declaración de que se ha llevado a cabo una búsqueda preliminar de revisiones de alcance sobre el tema, especificando las bases de datos o plataformas de búsqueda utilizadas (25).

4.6.4 Criterios de inclusión

Los criterios de inclusión del protocolo detallan la base sobre la cual se considerarán las fuentes a incluir en la revisión de alcance y se constituyen en una guía para escogerlas. Debe haber una clara congruencia entre el título, los objetivos, las preguntas y los criterios de inclusión en toda revisión de alcance.

Tipos de participantes

Se deben detallar las características importantes de los participantes, incluida la edad y demás requisitos que deben cumplir.

Concepto

El concepto central examinado por la revisión de alcance debe estar claramente articulado para guiar el alcance y la amplitud de la investigación.

Contexto

El contexto debe estar claramente definido y puede incluir, la consideración de factores culturales y/o intereses de ancestría o de género, estar limitado a un país o región geográfica según el tema y los objetivos.

Tipos de fuentes

Una revisión de alcance puede incluir cualquier literatura como fuente, por ej. estudios de investigación primaria, revisiones sistemáticas, metaanálisis, cartas, guías, etc. Los revisores pueden dejar la fuente de información abierta para incluir cualquier tipo de fuente que le aporte datos importantes o especificar cuáles serán las fuentes que incluirá en la revisión, esto puede hacerse sobre la base de tener conocimiento de los tipos de fuentes que serían más útiles y apropiadas para un tema en particular (25).

4.6.5 Estrategia de búsqueda

La estrategia de búsqueda para una revisión de alcance debe aspirar a ser integral para identificar tanto estudios primarios publicados y no publicados (literatura gris), así como revisiones. El protocolo del JBI recomienda utilizar una estrategia de búsqueda de tres pasos (25).

Cada paso debe ser claramente indicado en esta sección del protocolo. El primer paso es una búsqueda inicial limitada a al menos dos bases de datos en línea, relacionadas con el tema, seguido de un análisis de las palabras del texto contenidas en el título y el resumen de los artículos recuperados, y de los términos del índice utilizados para describir los artículos. El segundo paso es la búsqueda utilizando todas las palabras clave identificadas y los términos del índice en todas las bases de datos incluidas y el tercer paso es la búsqueda de nuevos estudios incluidos en la lista de referencias de todos los informes y artículos identificados. Se debe incluir una declaración de los revisores donde se exprese la intención de los revisores de contactar a los autores de los estudios primarios o revisiones para obtener información adicional, si es relevante. Los revisores también deben incluir los idiomas que se considerarán para la revisión, así como la franja de tiempo en el que se publicó la información que se seleccionará, con una justificación apropiada y clara para estas selecciones.

Como la pregunta de revisión puede ser amplia, los autores pueden encontrar que es apropiado buscar en todas las fuentes de evidencia (por ejemplo, estudios primarios y artículos de texto/opinión) simultáneamente como estrategia de búsqueda. Esto también depende de la relevancia de las fuentes de evidencia para el tema en cuestión. Este enfoque conducirá a una mayor sensibilidad en la búsqueda, que es lo deseable para las revisiones de alcance. También puede irse incorporado en la estrategia de búsqueda nuevos términos potencialmente útiles a medida que se realiza la búsqueda cuando los revisores se van familiarizando con las bases de datos, las palabras clave y las fuentes adicionales (25).

4.6.6 Extracción de los resultados

En las revisiones de alcance, el proceso de extracción de datos se denomina representación gráfica de los resultados. Este proceso proporciona al lector un resumen lógico y descriptivo de los resultados que se alinea con el objetivo y pregunta/s de la revisión de alcance. Se debe desarrollar una Tabla o formulario de registro en esta etapa de protocolo para registrar la información clave que se va tomando de la fuente como: autor, referencia y resultados o hallazgos relevantes para las preguntas de revisión. Esto se puede refinar aún más en la etapa de revisión y actualizar la Tabla de gráficos a medida que se avanza en la extracción de datos. Parte de la información clave que los revisores pueden optar por graficar son:

- a. Autor(es).
- b. Año de publicación.
- c. Origen/país de origen (donde se publicó o realizó el estudio).
- d. Objetivos/propósito.
- e. Población de estudio y tamaño de la muestra (si corresponde).
- f. Metodología/métodos.
- g. Resultados y detalles de estos.
- h. Hallazgos clave que se relacionan con la(s) pregunta(s) de la revisión de alcance.

Para facilitar la referencia y el seguimiento, se sugiere que los revisores mantengan registros cuidadosos para identificar cada fuente. La Tabla de gráficos se va actualizando continuamente y

el IJB sugiere que el equipo de revisión se familiarice con el registro de los resultados en el formulario de extracción con dos o tres estudios de los seleccionados para garantizar todos los revisores extraen los mismos datos (25).

4.6.7 Presentación de los resultados

En el momento del desarrollo del protocolo, los revisores deben definir cómo se van a presentar los resultados, por ejemplo, en un gráfico o una Tabla. Este aspecto se refinará hacia el final de la revisión, cuando los revisores tienen mayor conocimiento del contenido de los estudios incluidos. Los resultados de la revisión de alcance pueden presentarse como un mapa de los datos extraídos de los documentos incluidos, en forma de diagramas o Tablas y/o en un formato descriptivo que se alinee con el objetivo/s y alcance de la revisión. Los elementos de los criterios de inclusión del PCC pueden ser útiles para guiar cómo se deben mapear los datos más apropiadamente (25).

5. Materiales y métodos

5.1 Revisión de alcance

Al iniciar el planteamiento de esta tesis doctoral, se realizó una revisión de la literatura publicada, para buscar las revisiones sistemáticas sobre ADN mitocondrial para usos forenses que hubiesen sido publicadas y que, por tanto, estuviesen disponibles para el planteamiento de la propuesta y la posterior discusión de resultados. Una vez realizada la búsqueda no se encontró ninguna revisión publicada y por ello uno de los objetivos propuesto fue realizar una revisión sistemática de la literatura tipo alcance, para conocer y recopilar los estudios, realizados en población americana actual, que analizaron el ADN mitocondrial con fines forenses o poblacionales y para establecer la distribución de los diferentes haplotipos y haplogrupos en esas poblaciones evaluadas.

La búsqueda incluyó los estudios publicados desde la secuenciación del primer genoma mitocondrial en 1981 hasta septiembre del año 2022. Esta revisión se realizó siguiendo la metodología propuesta por el Instituto Joanna Briggs (JBI) para la realización de revisiones sistemáticas tipo alcance (Ref Ins Johanna Brings).

5.1.1 Criterios de inclusión:

Como criterios de inclusión para la revisión se tuvieron en cuenta los siguientes:

Participantes

Estudios genéticos desarrollados con ADNmt humano proveniente de individuos, vivos o muertos, bancos de muestras de laboratorios, muestras procedentes de catástrofes como desastres naturales, accidentes aéreos, atentados terroristas, identificación humana, entre otros.

Concepto

Para esta revisión se consideraron estudios que proporcionaran información relacionada con: Aplicaciones del ADN mitocondrial, metodologías utilizadas para el análisis molecular de ADN mitocondrial humano, o métodos de análisis de los datos producto de la genotipificación molecular de ADN mitocondrial humano.

Contexto

Se consideraron estudios desarrollados en cualquier lugar del mundo que hubiesen utilizado muestras de individuos nacidos en América.

Tipos de estudio

Se incluyeron estudios de tipo forense o poblacional, así como presentación de casos forenses que utilizaran el análisis de ADN mitocondrial humano.

5.1.2 Estrategia de búsqueda

La estrategia de búsqueda siguió la metodología de tres pasos propuestos por el JBI (25): Se realizó una búsqueda inicial en el portal Pubmed, seguida del análisis de las palabras contenidas en el título y resumen y de los términos utilizados para describir el estudio. A continuación, se realizó una segunda búsqueda en todas las fuentes de información, usando las palabras y términos identificados en la búsqueda inicial y, por último, se realizó una tercera búsqueda de estudios

adicionales a partir de la lista de referencias de los artículos identificados en las búsquedas previas. Las búsquedas fueron realizadas por dos revisores de manera independiente.

Se consideraron las publicaciones realizadas en los idiomas inglés, portugués y español, por ser estos idiomas de dominio de los autores. De cada fuente de información se incluyeron los artículos publicados desde el año 1981, fecha en la que fue secuenciado el primer genoma completo de un ADN mitocondrial por Anderson y cols (Ref Sec Ander), hasta septiembre de 2022. Para la elaboración de los protocolos de búsqueda se usaron operadores booleanos y truncadores de los términos, también se usó un asterisco al final de las palabras claves que representaron un comodín para habilitar alternativas y los sufijos para buscar la palabra. La estrategia de búsqueda fue lo suficientemente específica para cumplir con los objetivos de esta revisión y su alcance.

5.1.2.1 Fuentes de información. Se exploraron las siguientes fuentes de información:

- a. Portales: Pubmed (Medline y NCBI), Ebsco (3 bases de datos) y Scielo.
- b. Bases de datos: Scopus, Science direct, Web of science, CINAHL, Embase y BVS (Biblioteca Virtual en Salud).
- c. La búsqueda para literatura no publicada se realizó en: Repositorio Brasileiro de tesis y disertaciones CAPES, Repositorio DART, Tesis doctorales en red, Teseo y en el repositorio de la producción científica de América Latina.
- d. También se realizó la búsqueda de literatura publicada y no publicada en el motor de búsqueda Google Academic.

5.1.2.2 Selección de los estudios. Todas las citas de los estudios identificados con la estrategia de búsqueda fueron cargadas en el gerenciador de referencias Zotero, donde se eliminaron los duplicados. En este fueron examinados todos los títulos y resúmenes de los estudios en relación con el cumplimiento de los criterios de inclusión. Luego, se recuperaron los textos completos de aquellos estudios que atendieron con los criterios de inclusión por título y resumen para una evaluación detallada del cumplimiento de los criterios de inclusión definidos para la revisión. Las razones por las cuales se excluyeron los estudios después de la lectura por texto completo fueron reportadas en un archivo. de los estudios. Todo el proceso de selección de los estudios fue llevado a cabo, de manera independiente, por dos revisores (estudiante de doctorado y director).

5.1.3 Extracción de datos

Los datos extraídos de los estudios incluidos en la revisión fueron registrados en una Tabla de Excel elaborada para este fin. En la Tabla 1 se discriminan los datos extraídos de cada referencia.

Tabla 1. Registro de datos extraídos

No.	Año	Título	Autor	País	Participantes	Concento	Contexto	Métodos	Aplicaciones	Cumple	Observaciones

5.1.4 Presentación de datos

Los hallazgos de la revisión fueron sintetizados en formato narrativo acompañado de Figuras y Tablas.

5.2 Trabajo experimental

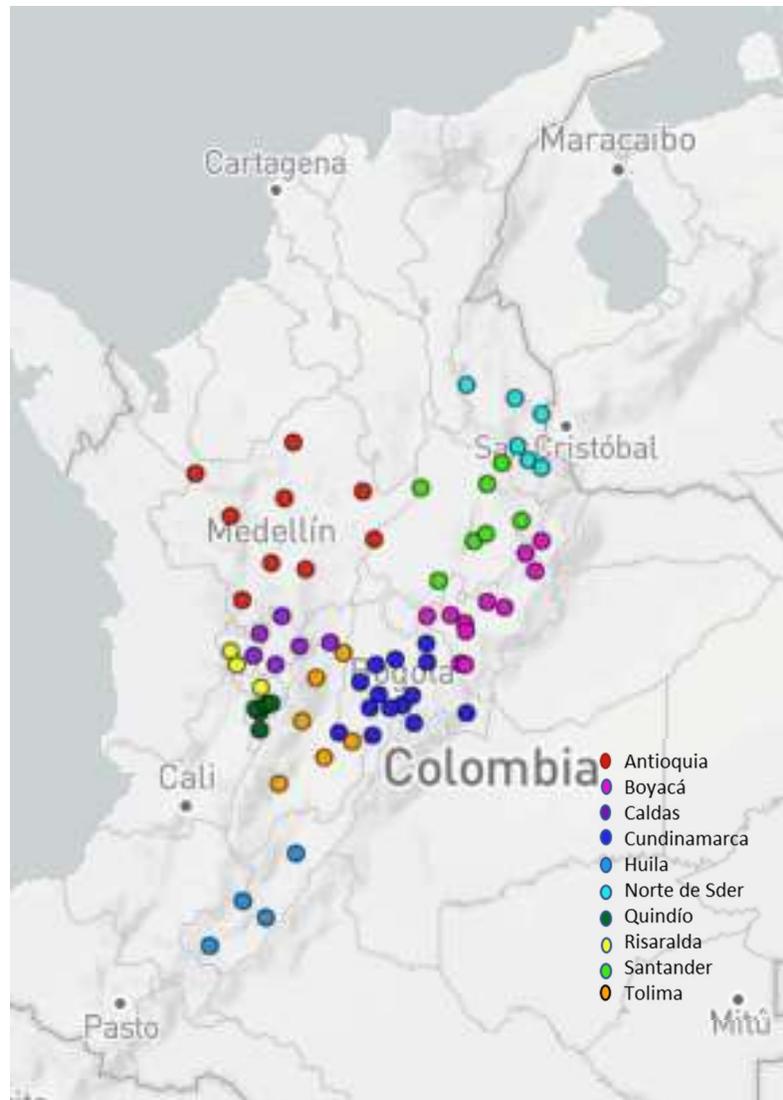
5.2.1 Muestras

Se tomaron y procesaron 662 muestras de individuos adultos no emparentados, nacidos y residenciados en los departamentos que conforman la Región Andina Colombiana: Antioquia, Boyacá, Caldas, Cundinamarca, Huila, Norte de Santander, Quindío, Risaralda, Santander y Tolima. Se tomaron 3 mL de sangre total en un tubo con anticoagulante EDTA, previa lectura, aceptación y firma del consentimiento informado para participar en esta investigación (Apéndice A). Esta propuesta de tesis junto con el consentimiento informado para participantes fue sometida y aprobada por el comité de ética de la UIS (CEINCI) el 21/05/2021. La propuesta presentada y aprobada incluyó el procesamiento de algunas muestras del banco de ADN del laboratorio de Genética y Biología Molecular con sede en Bogotá y dirigido por el doctor Humberto Ossa, quien autorizó el uso de una alícuota de 20 μ L de ADN extraído de muestras de personas que aceptaron participar del estudio, previa firma de consentimiento informado. Estas alícuotas fueron entregadas con un código aleatorio para garantizar la anonimización de estas muestras. Las muestras procedentes del departamento de Santander fueron tomadas y procesadas en el marco de un proyecto de investigación aprobado y financiado por la Vicerrectoría de Investigación y Extensión de la UIS (Proyecto VIE-UIS 2488); el CEINCI aprobó esta propuesta con su respectivo consentimiento informado el 11/10/2019. El consentimiento informado se presenta en el Apéndice B.

De cada departamento, con excepción de Santander, se tomaron aproximadamente 50 muestras. El muestreo se realizó teniendo en cuenta la distribución de la población en cada departamento, de acuerdo con el último censo realizado en el país (49), para que la población de las diferentes

provincias de cada departamento estuviese representada proporcionalmente en la muestra seleccionada (Figura 12).

Figura 12. Muestreo realizado en los departamentos de la región andina.



Diseñada con: <https://microreact.org/>

Todos los departamentos tuvieron una representación en la muestra del 100% de las provincias que los conforman con excepción de Boyacá donde se incluyeron muestras de doce de sus quince provincias (80%). El muestreo del departamento de Santander estuvo representado por 204 muestras de sangre de individuos no emparentados nacidos en las siete provincias que conforman el departamento. La muestra seleccionada se presenta en la Tabla 2.

Tabla 2. Muestras tomadas de las provincias de Santander

SANTANDER	N	% de muestras por provincia
Comunera	30	15
García Rovira	30	15
Guanentá	28	14
Metropolitana	36	17
Soto Norte	26	13
Vélez	23	11
Yariguies	31	15
TOTAL	204	100%

Teniendo en cuenta que la base de datos adecuada en un contexto forense es la que representa a la población actual y sin desconocer que para establecerla correctamente debe capturar la diversidad existente en las poblaciones, es claro que la estrategia de recolección de muestras es un factor determinante. Por ello se presentó una ponencia internacional con los resultados del

análisis de las muestras de los individuos participantes en este estudio y en dos más de Suramérica (Ecuador y Paraguay) clasificándolos de acuerdo con su lugar de nacimiento, después con el lugar de nacimiento de sus madres y finalmente con el de sus abuelas y establecer si existían diferencias de acuerdo con la estrategia de recolección de muestras que pudiese influir al construir una base de datos forenses de ADN mitocondrial. Los resultados se presentan en el Apéndice C. Adicionalmente los resultados fueron publicados en las actas del “29th Congress of the International Society for Forensic Genetics”, publicadas en el Forensic Science International Genetics Supplement Series, en 2022 (Apéndice D).

5.2.2 Extracción de ADN

El ADN se extrajo a partir de la muestra de sangre total, usando el kit PureLink™ Genomic DNA Mini Kit (ThermoFisher Scientific). Para ello, cada muestra se procesó en el área asignada para la extracción de ADN, del laboratorio de genética de la UIS, siguiendo el protocolo que se detalla a continuación:

- a. Preparación del lisado:
 - Encender el bloque seco a 55°C.
 - Adicionar 200 µL de sangre fresca a un tubo de microcentrífuga de 1,5 mL estéril.
 - Adicionar 20 µL de proteinasa K.
 - Adicionar 20 µL de RNasa A.

- Mezclar el tubo en un vortex e incubar a temperatura ambiente por 2 minutos.
- Adicionar 200 μ L de buffer de lisis/binding y mezclar muy bien en vortex hasta obtener una solución homogénea.
- Incubar el tubo en el bloque seco a 55°C por 10 minutos para promover la digestión de proteínas.
- Sacar el tubo y adicionar al lisado 200 μ L de etanol al 96-100%.
- Mezclar muy bien el contenido del tubo en un vortex durante 5 segundos.

b. Purificación:

- Ubicar el lisado preparado en la etapa anterior (640 μ L) en una columna con tubo de colección.
- Centrifugar la columna a 8300 rpm durante 3 minutos a temperatura ambiente.
- Descartar el tubo de colección y ubicar la columna en un nuevo tubo de colección, limpio.

c. Lavado:

- Adicionar a la columna 500 μ L de buffer de lavado 1 preparado en etanol 96 – 100 %
- Centrifugar la columna a 8300 rpm durante 3 minutos a temperatura ambiente.
- Descartar el tubo de colección y ubicar la columna en un nuevo tubo de colección, limpio.
- Adicionar a la columna 500 μ L de buffer de lavado 2 preparado en etanol 96 – 100 %

- Centrifugar la columna a velocidad máxima (14000 rpm) durante 3 minutos a temperatura ambiente.

d. Elución del ADN:

- Descartar el tubo de colección y ubicar la columna en un tubo de microcentrífuga estéril de 1,5 mL.
- Adicionar a la columna entre 25 – 200 μ L de agua bidestilada estéril.
- Incubar a temperatura ambiente por 1 minuto.
- Centrifugar la columna a velocidad máxima (14000 rpm) durante 1 minuto a temperatura ambiente para recuperar el ADN extraído.

e. Almacenamiento del ADN:

- Almacenar el ADN extraído a -20°C .
- El ADN que va a ser usado inmediatamente puede almacenarse a 4°C .

5.2.3 Cuantificación de ADN

El ADN extraído se cuantificó en un equipo Nanodrop de *Thermo Scientific* para establecer su concentración y ésta se ajustó a 5 ng/ μ L para su uso en la amplificación de la región control (RC) del ADNmt.

5.2.4 Genotipificación del ADNmt

La tipificación de las muestras se realizó de acuerdo con los lineamientos propuestos por la Comisión de ADN de la Sociedad Internacional de Genética Forense (ISFG) en las pautas revisadas y extendidas para la tipificación de ADNmt, del 2014 (24). Para ello se llevaron a cabo las siguientes etapas:

5.2.4.1 Amplificación de la RC de ADNmt. La RC del ADNmt, comprendida entre las posiciones 16024 y 576 de la molécula, fue amplificada en una única reacción, usando los primers L15900 (5'-CACCATTAGCACCCAAAGCT-3') y H639 (5'-GGGTGATGTGA GCCCGTCTA-3'). La PCR fue realizada en un volumen final de 5 μ L. (Tabla 3).

Tabla 3. Protocolo para el montaje de la PCR

Reactivo	Volumen por muestra
2x Qiagen Multiplex PCR Master Mix	2,5 μ L
Mix de primers (2 μ M)	0,5 μ L
H ₂ O dd esteril	1 μ L
ADN (1 – 5 ng/ μ L)	1 μ L
TOTAL	5 μL

En cada montaje se incluyeron un blanco de reactivos y un control negativo y uno control positivo de extracción y amplificación, siguiendo las recomendaciones propuestas por la ISFG (24). Las muestras se amplificaron en un termociclador 2420 de *Applied Biosystems* (Tabla 4).

Tabla 4. Programa de Amplificación de la RC del ADNmt

DENATURACIÓN INICIAL	35 CICLOS			EXTENSIÓN FINAL
	DENATURACIÓN	ANILLAMIENTO	EXTENSIÓN	
95°C	94°C	60°C	72°C	72°C
15 minutos	30 segundos	90 segundos	60 segundos	10 minutos

5.2.4.2 Verificación del producto amplificado. Se realizó una electroforesis en gel de agarosa al 1,2%, con el objetivo de verificar la amplificación de la RC del ADNmt, antes de continuar con la secuenciación de las muestras. El montaje del gel se realizó así:

- Pesar 1,2 g de agarosa y mezclar con 100 mL de tampón TAE al 1 %.
- Calentar la mezcla en el microondas hasta que se disuelva.
- Distribuir la mezcla sobre el soporte de gel y dejar solidificar.
- En un parafilm, distribuir 1 μ L de safer (Kasvi) por cada muestra para visualizar el amplificado.
- Mezclar 1 μ L de producto amplificado con el μ L de colorante safer.
- Colocar la mezcla (2 μ L) en cada pozo del gel.
- Montar los controles negativo y positivo de extracción y amplificación.

5.2.4.3 Purificación de productos amplificados. Los productos amplificados fueron purificados con las enzimas exonucleasa y fosfatasa alcalina del camarón (ExoSAP-IT, *Applied Biosystems*) (Tabla 5).

Tabla 5. Protocolo para el montaje de la purificación de los productos de PCR.

Reactivo	Volumen
Amplificado	4,0 μ L
ExoSAP-IT	1,6 μ L
TOTAL	5,6 μ L

Se mezcló en un vortex y se llevó al termociclador por 15 minutos a 37° C, para actuación enzimática, seguido de un paso a 85 °C para inactivación de las enzimas.

5.2.4.4 Secuenciación de la RC del ADNmt. La secuenciación fue realizada usando el kit BigDye Terminator v3.1 cycle Sequencing de *Applied Biosystems*, siguiendo el protocolo propuesto por la casa comercial fabricante. Las muestras amplificadas fueron secuenciadas con cuatro primers: L15900, L16555, H159 y H639 (Tabla 6).

Tabla 6. Protocolo para secuenciación de la RC del ADNmt

Reactivo	Volumen por muestra
BigDye Terminator cycle Sequencing	0,8 μ L
Buffer	0,6 μ L

Primers (2,5 μ M)	0,5 μ L
H ₂ O dd esteril	2,1 μ L
ADN (amplificado y purificado)	1 μ L
TOTAL	5 μL

Y se secuenciaron con el protocolo presentado en la Tabla 7, en un termociclador 2420 de *Applied Biosystems*.

Tabla 7. Programa de secuenciamiento de la RC del ADNmt.

DENATURACIÓN INICIAL	35 CICLOS			EXTENSIÓN FINAL
	96°C 2 minutos	96°C 15 segundos	50°C 9 segundos	60°C 2 minutos

En los casos en que una muestra presentó heteroplasmia de posición ésta se verificó usando otro primer que de ser posible amplificara el mismo segmento en dirección opuesta al primer ya utilizado para amplificar el segmento. Las heteroplasmas ubicadas en la región HVSI se secuenciaron con los primers: L15900 o L15997 y H159 o H016, de acuerdo con la posición de la heteroplasmia. Las heteroplasmas de posición ubicadas en las regiones HVSII o HVSIII se secuenciaron con los primers L16555 o L16268 y H639 o H599.

Las heteroplasmas de longitud (normalmente asociadas a los Indel de las posiciones 309.xC, 315.xC, 523del, 524del, 523.xC o 576.xC) se secuenciaron con los primers L314, H388 y H484. Los primers L amplificaron la secuencia de izquierda a derecha (Forward) y los primers H amplificaron la secuencia de derecha a izquierda (Reverse). En la Tabla 8 se presentan las secuencias de los primers utilizados para secuenciar la RC del ADNmt.

Tabla 8. Primers usados para secuenciar la RC del ADNmt.

Primer	Secuencia
L15900	5'-TAAACTAATACACCAGTCTTGTAACC-3'
L15997	5'-CACCATTAGCACCCAAAGCT-3'
L16268	5'-CACTAGGATACCAACAAACC-3'
L16555	5'-CCCACACGTTCCCCTTAAAT-3'
L00314	5'-CCGCTTCTGGCCACAGCACT-3'
H00016	5'-CCCGTGAGTGGTTAATAGGGT-3'
H00159	5'-AAATAATAGGATGAGGCAGGAATC-3'
H00388	5'-TGGTTAGGCTGGTGTAGGG-3'
H00484	5' TGAGATTAGTAGTATGGGAG3'
H00599	5'-TTGAGGAGGTAAGCTACATA-3'
H00639	5'-GGGTGATGTGAGCCCGTCTA-3'

5.2.4.5 Purificación del producto secuenciado. Los productos de secuenciación fueron purificados utilizando columnas de Sephadex® G-50 de *Sigma-Aldrich*, usando el siguiente protocolo:

- Preparar una solución de sephadex diluyendo 3,3 g de sephadex en 50 mL de agua miliQ

- Depositar las columnas (con filtro) en tubos de centrifuga de 2 mL sin tapa.
- Mezclar bien el sephadex y depositar 750 μ L en la columna.
- Centrifugar durante 3 min. a 5000 rpm.
- Descartar el tubo con el agua y trasladar la columna a un tubo de microcentrifuga nuevo y debidamente marcado.
- Depositar sobre la columna con el sephadex los 5 μ l de producto secuenciado.
- Centrifugar durante 3 min. a 5000 rpm.
- El producto secuenciado y purificado queda depositado en el tubo de microcentrifuga.

5.2.5 Detección de las secuencias por Electroforesis Capilar (EC)

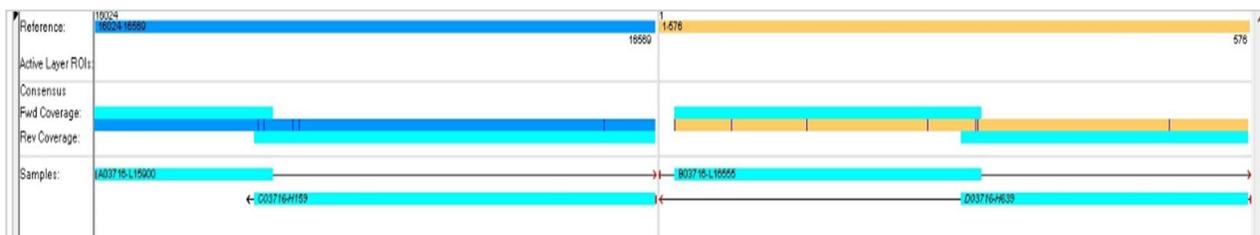
La detección de las secuencias se realizó por EC en un analizador genético 3500 (Applied Biosystems), de acuerdo con el siguiente protocolo:

- Mezclar el producto de secuenciación purificado con 10 μ l de formamida de *Applied Biosystems*. Depositar 10 μ l de la mezcla en un pozo de la placa de 96 pozos. Incluir en cada montaje un control negativo y uno positivo.
- Una vez servidas las muestras en la placa, cubrirla con una septa y colocar sobre ella un retenedor. Montar la placa en el analizador genético 3500 de *Applied Biosystems*.
- Definir las condiciones de corrido en el software Data collection v 3.1 del equipo, indicando la fecha de la corrida, el ensayo a realizar (secuenciación), el polímero usado (Pop 7) y el capilar de secuenciación usado (50 cm).
- Diligenciar la hoja de muestras con la identificación de cada una de ellas, el primer con el que se secuenció y la posición de cada muestra, en la placa montada.
- Poner a correr el equipo.

5.2.6 Análisis de secuencias

Las secuencias de muestras corridas por EC quedaron grabadas en el equipo y se analizaron inicialmente con el software Sequencing Analysis v.5.4 de *Applied Biosystems* para establecer la calidad de la secuenciación. El software determina, según el color asignado a la muestra, si la secuencia obtenida fue óptima (color azul), buena (color verde), aceptable (color amarillo) o no se obtuvo secuencia (color rojo). Posteriormente, se empleó el software SeqScape v2.7 de *Applied Biosystems* para crear un proyecto para cada muestra, incluyendo todas las secuencias corridas de una muestra, con los diferentes primers usados. Las secuencias fueron asignadas al proyecto seleccionando en el software la opción secuenciación de ADNmt. Se verificó la cobertura lograda con cada primer para establecer que la región control estuviese secuenciada en su totalidad, desde la posición 16024 a la 576 (1100 pares de bases) para luego proceder con el análisis de la muestra (Figura 13).

Figura 13. Secuencias alineadas con respecto a la rCRS.



Una vez se verificó la cobertura completa de la RC con la secuenciación, se visualizaron todas las secuencias de una misma muestra en forma simultánea, para analizarlas alineándolas con las regiones hipervariables I, II y III, de la RC de la rCRS (85,86). Se determinaron las posiciones en las que la muestra difería de la secuencia de referencia y el tipo de cambio (sustitución, inserción o deleción), así como la presencia de heteroplasmas de posición o longitud presentes en la muestra.

En las Figuras 14 y 15 se presentan ejemplos de las variaciones encontradas en las muestras. También se presentan ejemplos de heteroplasmas de posición y de longitud halladas en algunas de las muestras analizadas (Figuras 16,17,18,19).

Figura 14. Delección 498del y sustitución de base 499A

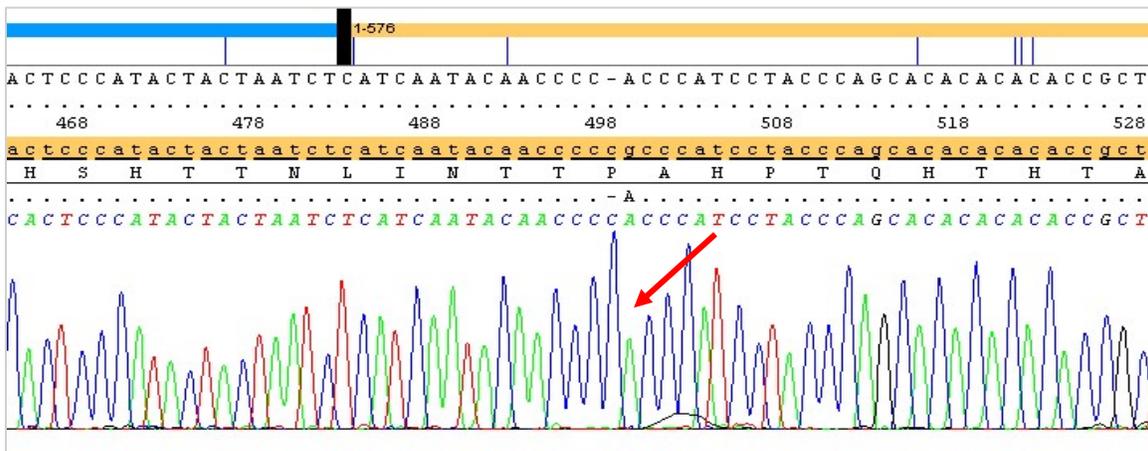


Figura 15. Inserción de las bases 524.1A y 524.2C

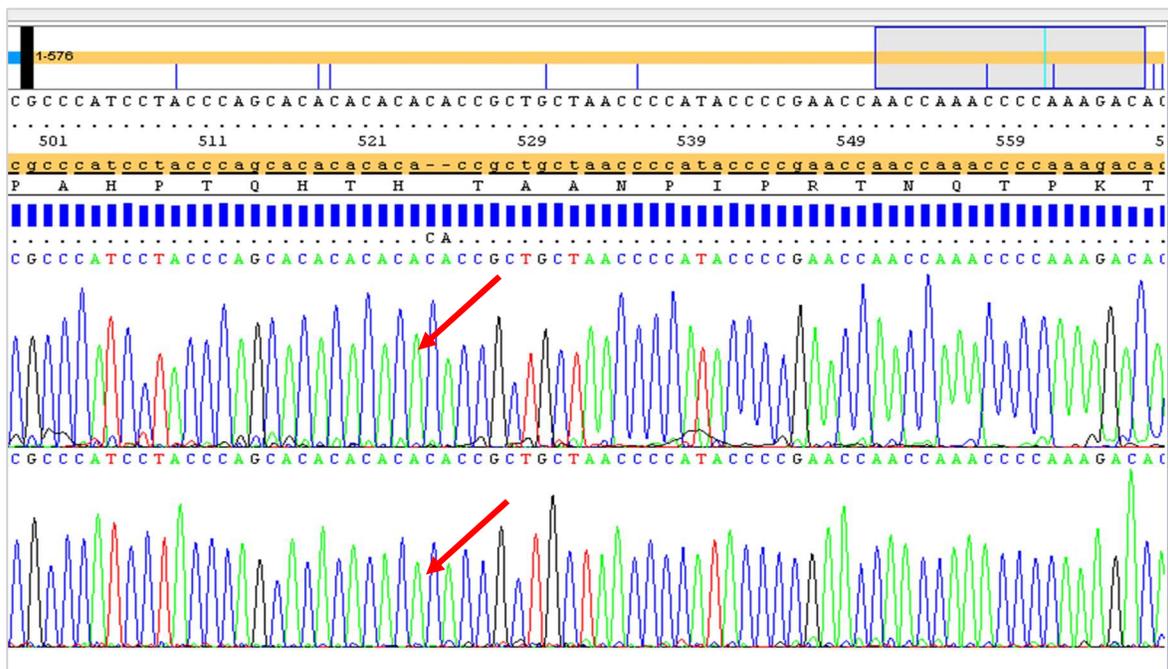


Figura 16. Heteroplasmia de posición (146Y)

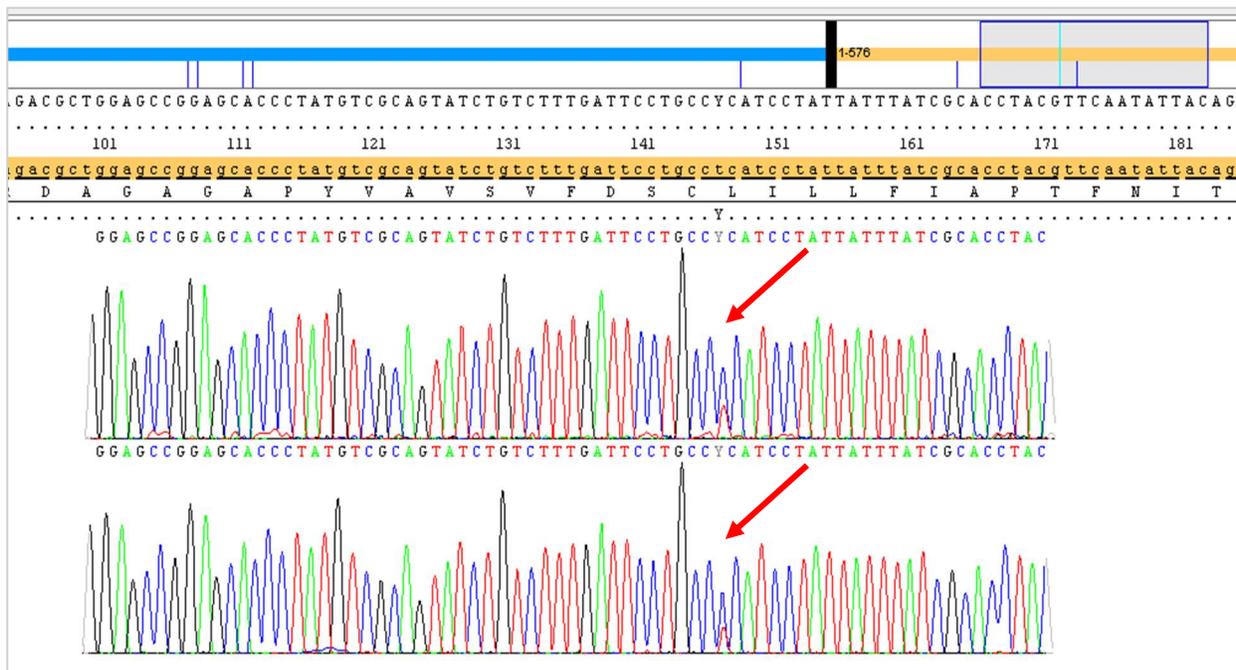


Figura 17. Heteroplasmia de longitud en la posición 16193 de la región HVSI.

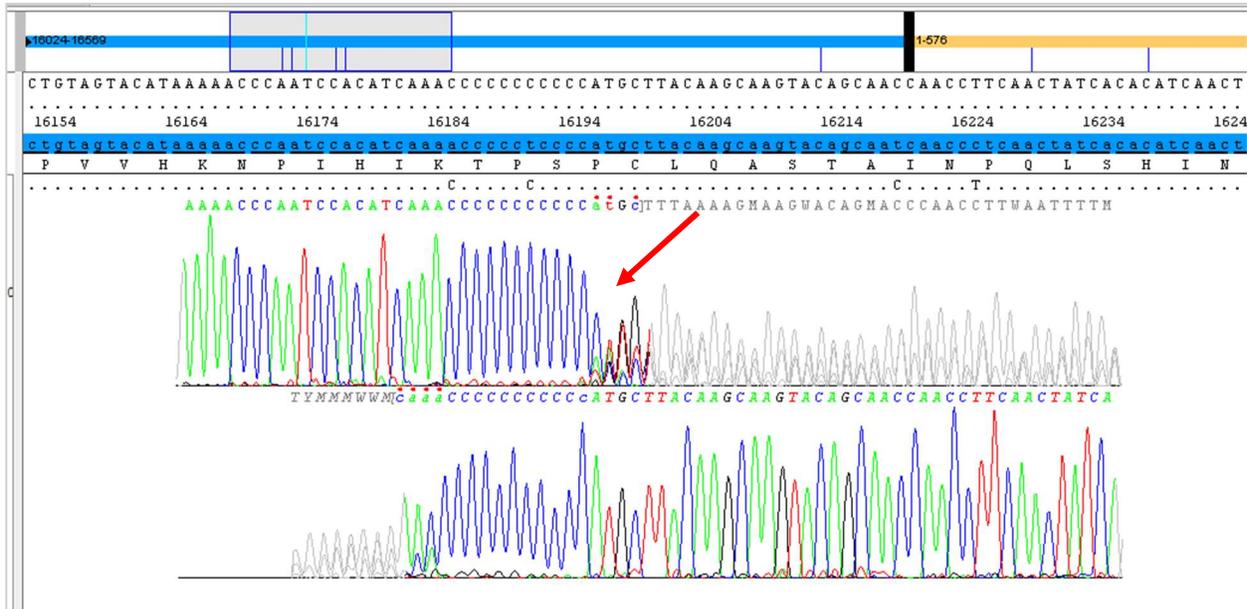


Figura 18. Heteroplasmia de longitud en la posición 309 de la región HVSII.

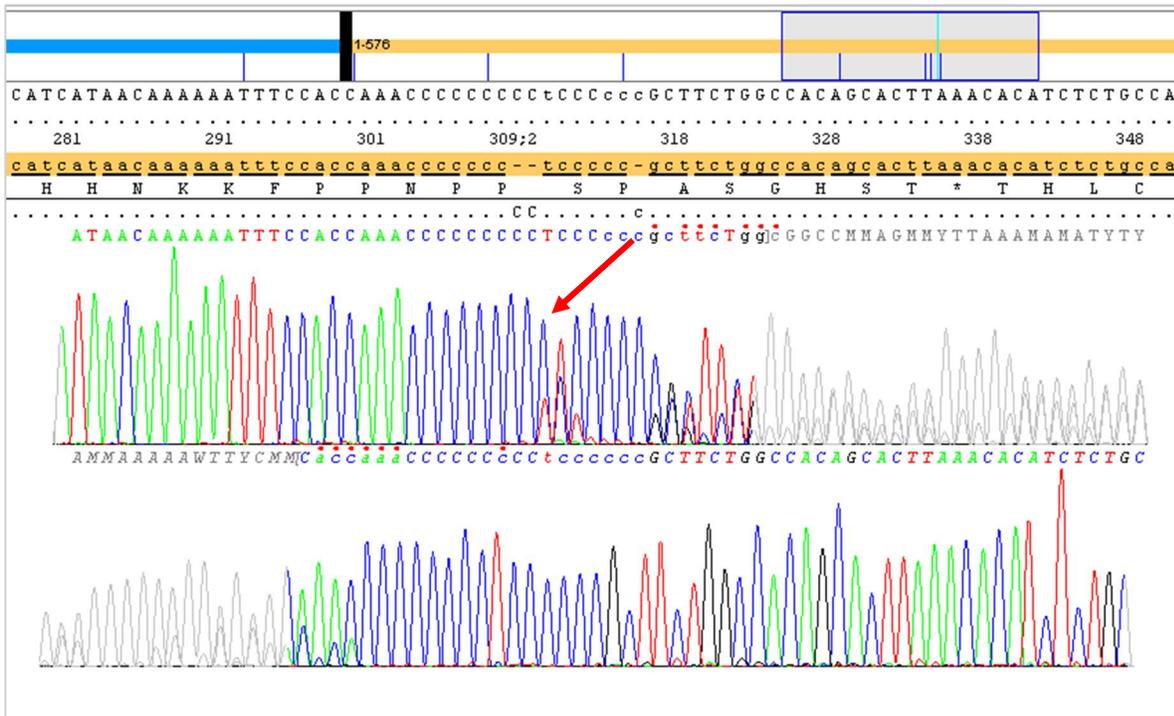
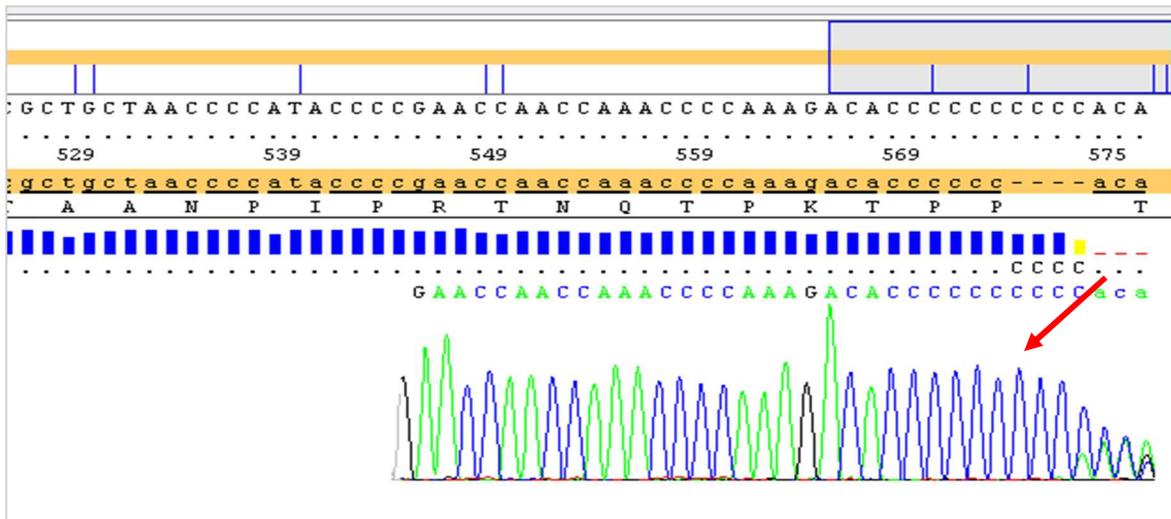
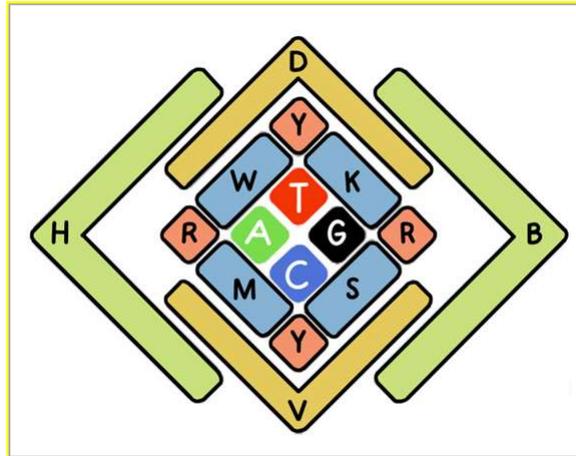


Figura 19. Heteroplasmia de longitud en la posición 573 de la región HVSIIL.

5.2.7 Asignación de haplotipos

Las muestras alineadas con la rCRS usando el software SeqScape v2.7 de *Applied Biosystems*, se analizaron para determinar su haplotipo, identificando todas las posiciones en las que la muestra difirió de la rCRS. Cada posición donde el software asignó una diferencia fue revisada y confirmada mediante la lectura detallada de las secuencias. Una vez establecidas todas las modificaciones se generó el reporte de las mutaciones en el software y se grabó como archivo .txt. Posteriormente este archivo se exportó a un archivo Excel para registrar el haplotipo de cada muestra, utilizando la nomenclatura recomendada por la ISFG (24). Las sustituciones se indicaron con la posición seguida de la base detectada, por ejemplo, 73G, las deleciones se registraron con la posición seguida del sufijo del, por ejemplo, 523del y las inserciones se registraron con la posición seguida de un punto, el número de la inserción y la base insertada, por ejemplo, 16193.1C. En caso de presentarse heteroplasmias de posición, fue usada la nomenclatura establecida por el código IUPAC que permite definir múltiples posibilidades de nucleótidos en una posición y asignar la letra correspondiente, de acuerdo con la mezcla heteroplasmática detectada (Figura 20).

Figura 20. Código IUPAC

Posteriormente cada haplotipo se revisó en EMPOP para confirmar su correcta alineación.

5.2.8 Asignación de haplogrupos

Los haplogrupos se asignaron con los criterios de la EMPOP (48) y de acuerdo con el Phylotree build 17, construido el 18 febrero de 2016 (97). Para ello se introdujo cada haplotipo en la base de datos y se verificó que la asignación hecha del haplotipo fuese correcta, luego se determinó su frecuencia en la base de datos y se estableció el haplogrupo correspondiente. Los haplotipos y haplogrupos obtenidos fueron anonimizados asignándoles un código aleatorio diferente al usado al interior del laboratorio y se sometieron a control de calidad en la EMPOP. Para ello, previo al envío fueron cargados en una hoja para una revisión inicial por el software EMPcheck de EMPOP para confirmar la alineación correcta, así como la calidad del conjunto de datos y detectar potenciales errores en los datos a someter (48,94). Una vez revisadas y confirmadas las posiciones críticas detectadas por el software EMPcheck de EMPOP, los datos se sometieron a esta base de datos. La EMPOP realizó la revisión y solicitó las secuencias de varias muestras para verificar: 1) las heteroplasmias de posición detectadas, 2) las posiciones en que las comprobaciones

filogenéticas sugirieron la omisión de una variante y 3) las posiciones en que las variantes notificadas eran nuevas para esta base de datos.

Se enviaron los electroferogramas de las secuencias de las muestras seleccionadas por la EMPOP, para que fuesen revisadas en forma detallada y para confirmar la correcta asignación de los haplotipos y haplogrupos. Una vez realizada la revisión, la EMPOP emitió un informe de la evaluación realizada, destacando que no se encontró ningún error en los datos revisados ni en la asignación hecha de cada haplotipo y su haplogrupo. Así mismo emitió los certificados como constancia del sometimiento, revisión y aceptación de los datos enviados y asignó un código mediante el cual estos haplotipos ingresarán en esta base de datos mundial, donde estarán disponibles para las búsquedas forenses (48,58). Adicionalmente, este código es requerido para que los resultados sean aceptados para ser publicados en las revistas científicas: *Forensic Science International Genetics* e *International Journal of Legal Medicine*, como una constancia de que fueron sometidos a control de calidad y aceptados después de ser revisados y confirmados (24).

5.3 Análisis estadístico

5.3.1 Análisis estadísticos y cálculo de parámetros de importancia forense

Las frecuencias de los haplotipos y de los haplogrupos se calcularon mediante conteo directo. Los haplotipos fueron convertidos en secuencias alineadas usando el software Haplosearch (104) para posteriormente ser puestas en el formato y ser introducidas en el software Arlequín v3.5.2.2 (105). En este software se calcularon los parámetros estadísticos de diversidad haplotípica (H), diversidad nucleotídica (π), el número medio de diferencias entre pares de secuencias (MNPd, del inglés *Mean Number of Pairwise Differences*), y la distribución del número de diferencias entre pares de secuencias (*mismatch distribution*). Adicionalmente se utilizó este software para realizar

la comparación de las poblaciones y calcular la distancia genética (por medio del estadístico F_{ST}) y el Análisis de varianza molecular (AMOVA, del inglés *Analysis of MOlecular VAriance*).

El poder de exclusión (mtCE) se calculó teniendo en cuenta la presencia de dos o más diferencias entre dos haplotipos, requisito propuesto tanto por ISFG como por SWGDAM para realizar la exclusión de un individuo aportante de una muestra, de pertenecer a un mismo linaje materno de otro individuo cuya muestra ha sido analizada para ADNmt con fines forenses (106,107). El software STATISTICA (TIBCO Software Inc. (2020). Data Science Workbench, version 14. <http://tibco.com>) (TIBCO Software Inc., Palo Alto, CA, USA) se utilizó en la representación de escalamiento multidimensional (MDS) de distancias genéticas por pares y para realizar análisis de componentes principales (PCA) basado en las frecuencias de los haplogrupos.

5.3.2 Construcción de redes

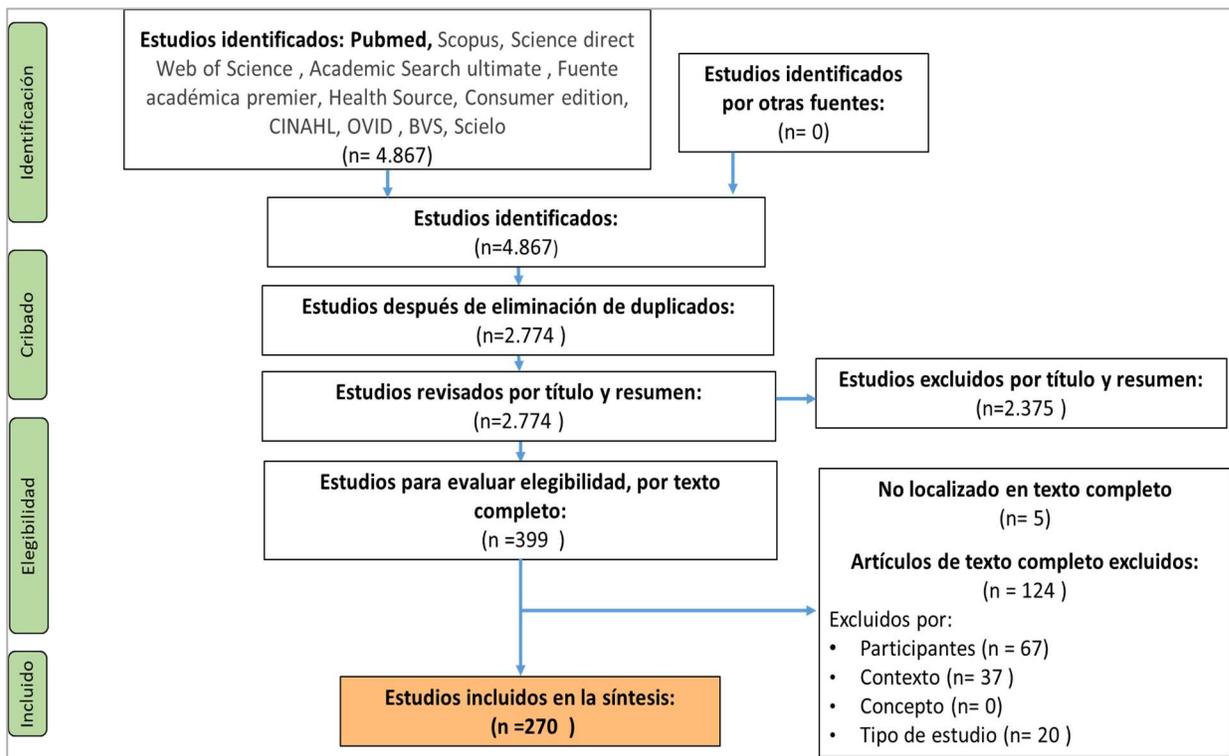
Las redes filogenéticas se diseñaron con el método de unión de la mediana (*median-joining*) en Network 10.0 software (Fluxus Technology Ltd., Colchester, UK), utilizando los haplogrupos de ADNmt sin considerar los Indel en las posiciones 16193, 309, 315, 523-524 y 573 de la RC. Estos Indel están asociados a tractos homopoliméricos y por ello, algunos laboratorios forenses los ignoran en las comparaciones de las pruebas forenses y en las búsquedas en bases de datos (24).

6. Resultados y Discusión

6.1 Revisión de alcance

Se examinó la literatura científica publicada desde 1981 a septiembre de 2022, para mapear las aplicaciones que ha tenido el análisis de las variaciones del ADN mitocondrial en poblaciones americanas. Apéndice. La aplicación de la estrategia de búsqueda permitió identificar 4867 potenciales estudios para la revisión, de estos, 270 fueron incluidos en la revisión (Figura 21).

Figura 21. Diagrama de flujo (PRISMA) del proceso de selección e inclusión de los estudios.



Como aspectos generales de la revisión está que los estudios seleccionados fueron publicados en su mayoría en artículos originales (98,9%) y tres de ellos fueron tesis de maestría o doctorado (1,1%). El idioma más usado para publicar fue el inglés (92,6% de los estudios seleccionados) y mucha menor cantidad se utilizó el español (6,3%) y los restantes (1,1%) fueron publicados en portugués.

Con respecto a los hallazgos de la revisión, los resultados arrojaron para cada una de las preguntas formuladas con los aspectos de interés para establecer con la revisión, que los participantes fueron en su totalidad en individuos vivos y no relacionados. El número de participantes en estos estudios fue muy disímil, estuvo comprendido entre 16 individuos y 1745 individuos. Las aplicaciones dadas a los estudios seleccionados se pueden observar en la Tabla 9.

Tabla 9. Aplicaciones de los estudios seleccionados

Tipo de estudio	N	%
Con fines poblacionales	200	74
Con fines forense y poblacional	49	18
Con fines forenses	21	8

La región de la molécula seleccionada para ser genotipada y analizada, así como las metodologías empleadas para realizar el genotipado del ADNmt, en los estudios seleccionados, se pueden observar en las Tablas 10 y 11.

Tabla 10. Región del ADNmt analizada

Región analizada	N	%
Región control parcial	160	59
Región control completa	92	34
Mitogenoma	18	7

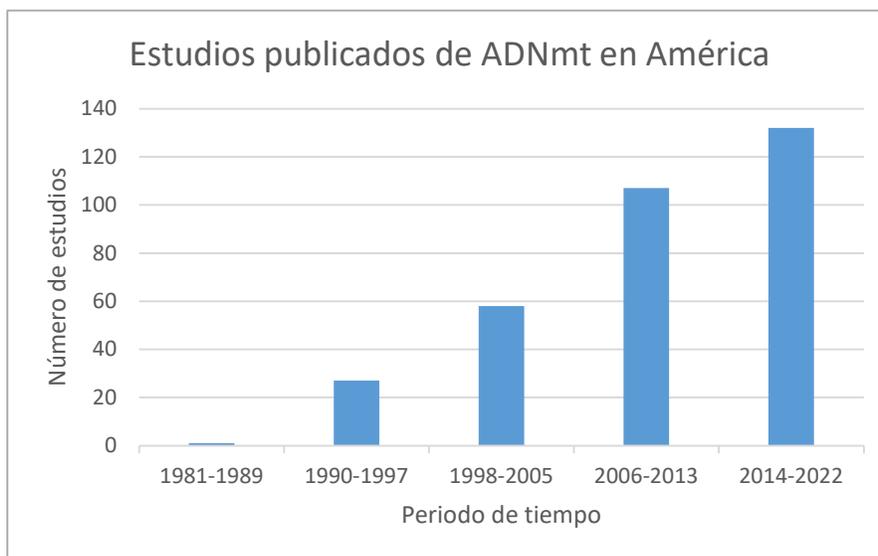
Tabla 11. Metodología de genotipado

Metodología	N	%
RFLP	112	41,48
Secuenciación por Sanger	93	34,44
RFLP + Secuenciación por Sanger	25	9,25
Secuenciación de próxima generación (NGS)	18	6,66
Minisequenciación	10	3,70
Otra	12	4,44

La revisión de alcance permitió establecer el número de estudios publicados por décadas, desde 1981 en adelante, en los países del continente americano. En la Figura 22 se puede observar que en la primera década una vez se secuenció el primer genoma de ADNmt, solo se encontró un estudio publicado en población de los Estados Unidos de América en 1985 (108) y de ahí en adelante, cada década se ha ido incrementando el número de publicaciones que analizan el ADNmt en las poblaciones de los países americanos lo que ha permitido conocer mejor la composición

genética y la diversidad presente en el continente americano, a partir de los linajes maternos así como comparar las semejanzas o diferencias presentes en los haplotipos y haplogrupos presentes en los diferentes países que conforman América.

Figura 22. Estudios de ADNmt publicados en América



Los estudios seleccionados procedieron de diferentes países de América, siendo Estados Unidos y Canadá con 52 publicaciones los países con mayor número de publicaciones, seguidos de Brasil con 42 publicaciones y en tercer lugar Argentina con 35 publicaciones. De Colombia se encontraron 22 publicaciones lo que lo ubica en el cuarto lugar en número de publicaciones (Figura 23).

Figura 23. País de procedencia de los estudios seleccionados



De los 22 estudios realizados en población colombiana que fueron identificados con la revisión de alcance realizada, 12 analizaron comunidades nativas y/o afrodescendientes actuales aisladas (26,27,28,29,30,31,32,33,34,35,36,37), 6 analizaron poblaciones nativas y/o afrodescendientes actuales aisladas junto con poblaciones mezcladas (38,39,40,41,42,43) y 4 analizaron solo poblaciones mezcladas actuales de diversos departamentos (44,45,46,47). Los artículos y tesis seleccionados porque cumplieron los criterios de búsqueda, fueron revisados para la redacción de esta tesis y para la discusión de los resultados hallados en el presente estudio.

6.2 Linajes de ADNmt en el departamento de Santander

En este apartado se presentan los resultados obtenidos de las muestras del departamento de Santander, con el análisis realizado para evaluar la presencia de subestructura en sus provincias. Apéndice Los resultados finales fueron sometidos a control de calidad en la EMPOP. Una vez evaluados, pasaron este control de calidad (QC) sin tener ninguna corrección a la asignación de los haplotipos y haplogrupos. Apéndice Apéndice

6.2.1 Haplotipos hallados

Los haplotipos de la RC del ADNmt encontrados en las muestras de individuos del departamento de Santander, con los correspondientes haplogrupos asignados, se presentan en el Apéndice I (muestras identificadas con los códigos S001 a S204). En varias de las muestras analizadas se detectaron heteroplasmas de posición y de longitud. En diez de ellas (5%), se detectaron heteroplasmas de posición, ubicadas en nueve posiciones diferentes (Tabla 12).

Tabla 12. Heteroplasmas de posición en muestras de Santander

Heteroplasma	Muestra	Haplogrupo
146Y	S111	B2d
152Y	S020	B2d
152Y	S031	B2d
153R	S013	A2af1a1
182Y	S079	B2d
16092Y	S174	D1f
16172Y	S011	A2+(64)
16182M	S175	B2d
16211Y	S141	A2ac
16311Y	S060	B2d

Código IUPAC: Y: mezcla de C y T

R: mezcla de A y G M: mezcla de A y C

Se realizó la búsqueda de estas heteroplasmas en la base de datos EMPOP que cuenta con 38.361 haplotipos de la región control completa de ADNmt de diversos países en el mundo y se encontró que la heteroplasma C/T en la posición 16211, no ha sido reportada allí con anterioridad. Las demás heteroplasmas de posición detectadas en este estudio se informaron previamente en otras poblaciones que están disponibles en esa base de datos (EMPOP v4/R13), consultada el 10/01/2023. Adicionalmente, el porcentaje de heteroplasmas puntuales halladas está acorde con lo encontrado en un estudio realizado en más de 5000 muestras de 20 poblaciones mundiales donde la heteroplasma puntual se detectó en el 6% de las muestras, y en ese estudio están reportadas con heteroplasma casi todas las posiciones detectadas con heteroplasma puntual en el presente estudio con excepción de las heteroplasmas en las posiciones 182, 16172 y 16211 (109).

También se detectó heteroplasma de longitud en el 45% de las muestras analizadas. En 27 de ellas se detectó heteroplasma de longitud en más de un sitio (13%). La mayoría se presentaron en los tramos poli-C de las regiones HVSI y HVSII. El mayor número de heteroplasmas de longitud (37%) se presentó en HVSI por una sustitución T/C en la posición 16189 junto con inserciones C en la posición 16193 y/o sustituciones A/C en las posiciones 16182 y 16183. En HVSII, la heteroplasma de longitud se asoció con inserciones de C en las posiciones 309 y/o 315 (17%) y con menor frecuencia se detectaron heteroplasmas de longitud debidas a inserciones de C en las posiciones 463 y 573 (2 %), y en las repeticiones AC entre las posiciones 515- 524 (2%). Se detectaron Indel raros (sustituciones poco frecuentes) en las posiciones 89 (inserción de una C en las muestras S034 y S133), 106 (inserción de CC en la muestra S010), 116 (delección de una A en las muestras S096 y S104) y 498 (inserción de CC muestra S001). Estos resultados son similares, aunque un poco más bajos en porcentaje que los hallados en el estudio realizado en más de 5000 muestras donde la heteroplasma de longitud estuvo presente en el 52% de las muestras, sin

embargo, en ambos estudios las heteroplasmas de longitud fueron detectadas en las mismas posiciones (109).

6.2.2 Diversidad genética poblacional a partir de los haplotipos

Los resultados hallados para todo el departamento y en cada una de sus provincias se presentan en la Tabla 13 parte A.

Tabla 13. Parámetros de diversidad genética y forenses estimados para el departamento de Santander

Todas las posiciones (A)							
Provincia	n	H	Haplotipos Únicos	Únicos %	mtCE	MNPD	II
Todas	204	0,9925 ± 0,0020	124	61	0,9673	15,1850 ± 6,8146	0,0133 ± 0,0066
Comunera	30	0,9885 ± 0,0126	23	77	0,9379	14,7471 ± 6,7895	0,0131 ± 0,0067
García Rovira	30	0,9885 ± 0,0126	23	77	0,9517	14,3080 ± 6,5966	0,0127 ± 0,0065
Guanentá	28	0,9947 ± 0,0112	24	86	0,9868	14,9339 ± 6,8859	0,0132 ± 0,0068
Metropolitana	36	0,9921 ± 0,0084	26	72	0,9746	15,7270 ± 7,1849	0,0139 ± 0,0071

Soto Norte	26	0,9846 ± 0,0160	19	73	0,9815	16,4585 ± 7,5752	0,0146 ± 0,0075
Vélez	23	0,9881 ± 0,0186	20	87	0,9644	14,3874 ± 6,6912	0,0128 ± 0,0066
Yariguíes	31	0,9914 ± 0,0116	26	84	0,9462	14,4989 ± 6,6742	0,0128 ± 0,0066
Excluyendo los Indel (B)							
Provincia	n	H	Haplotipos Únicos		mtCE	MNPD	π
			Únicos	%			
Todas	204	0,9783 ± 0,0055	95	47	0,9293	13,1026 ± 5,9220	0,0116 ± 0,0058
Comunera	30	0,9724 ± 0,0209	21	70	0,9034	12,6851 ± 5,8837	0,0113 ± 0,0058
García Rovira	30	0,9586 ± 0,0269	21	70	0,9149	12,5678 ± 5,8321	0,0112 ± 0,0058
Guanentá	28	0,9868 ± 0,0141	21	75	0,9127	12,7169 ± 5,9100	0,0113 ± 0,0059
Metropolitana	36	0,9714 ± 0,0190	24	67	0,9476	13,8556 ± 6,3670	0,0123 ± 0,0063
Soto Norte	26	0,9846 ± 0,0160	19	73	0,9723	14,3200 ± 6,6318	0,0128 ± 0,0066
Vélez	23	0,9802 ± 0,0198	16	70	0,9447	12,3794 ± 5,8011	0,0110 ± 0,0058

Yariguíes	31	$0,9548 \pm 0,0311$	24	77	0,8903	$12,3183 \pm 5,7171$	$0,0110 \pm 0,0057$
-----------	----	---------------------	----	----	--------	----------------------	---------------------

Diversidad haplotípica (H), poder de exclusion (mtCE), número medio de diferencias por pares (MNPd) y diversidad nucleotídica (π).

Se detectaron, a partir de la secuenciación de la región control completa de las 204 muestras, un total de 148 haplotipos diferentes, 124 de los cuales fueron únicos (Apéndice I). La diversidad haplotípica (H) fue de $0,9925 \pm 0,0020$ y el poder de exclusión (mtCE) $0,9673$. El número medio de diferencias por pares (MNPd) entre haplotipos fue de $15,1850 \pm 6,8146$, con una diversidad nucleotídica (π) de $0,0133 \pm 0,0066$. Los valores de diversidad genética hallados en Santander están en el rango de los informados anteriormente para otras poblaciones mezcladas de América del Sur como Chile y Paraguay (110,111), incluidos países vecinos como Brasil, Venezuela y Ecuador (23,112 - 118), pero son más elevadas que las encontradas en distintas regiones de Argentina (119) (Tabla 14).

Tabla 14. Parámetros de diversidad genética y forenses para poblaciones de Suramérica

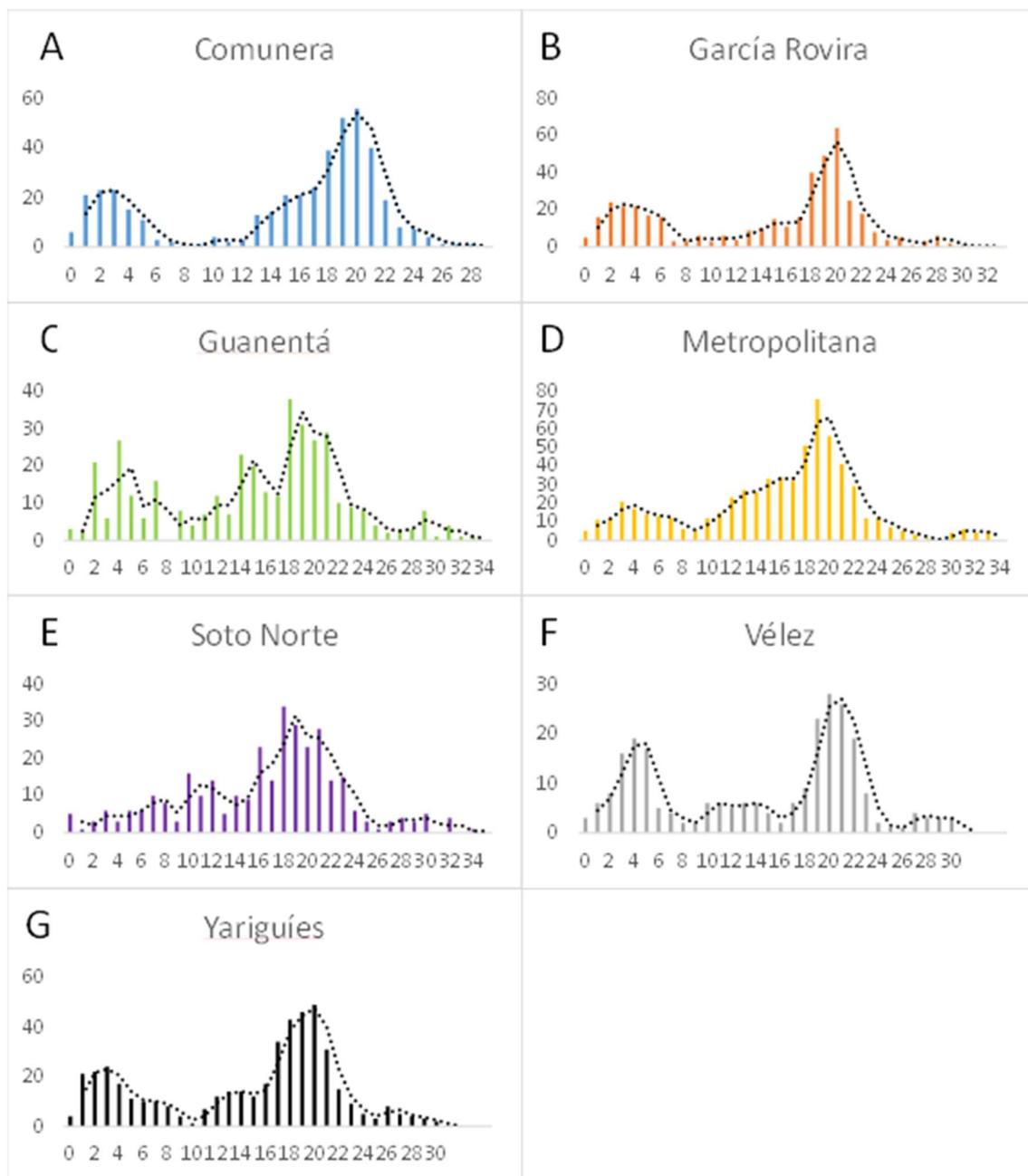
País	Muestra	n	H	mtCE	MNPd	π	Referencia
Colombia	Santander	204	0,9925	0,9673	$15,18 \pm 6,81$	$0,0133 \pm 0,006$	Este estudio
	Norte	98	0,9060	---	$15,53 \pm 6,99$	---	
Argentina	Central	193	0,9370	---	$14,47 \pm 6,51$	---	Bobillo et al. (119)
	Sur	47	0,8780	---	$14,50 \pm 6,61$	---	
Brasil	Rio de Janeiro	205	0,9994	0,9976	18,6130	---	Simão et al. (112)
	Espírito Santo	214	0,9994	---	$16,89 \pm 7,54$	---	Reis et al. (113)

	São Paulo	142	0,9991	---	$17,13 \pm 7,66$	---	Prieto et al. (23)
	Santa Catarina	80	0,9930	---	$15,00 \pm 6,77$	$0,0132 \pm 0,006$	Palencia et al. (114)
	Paraná	108	0,9976	---	$16,17 \pm 7,25$	$0,0143 \pm 0,007$	Poletto et al. (115)
	Distrito Federal	306	0,9988	---	---	---	Freitas et al. (116)
Chile	Varias regiones	995	0,9701	---	$7,59 \pm 3,54$	$0,0222 \pm 0,011$	Gómez et al. (110)
Paraguay	Oriente	299	0,9937	---	---	---	Simão et al. (111)
Ecuador	Varias regiones	54	0,9986	---	$17,58 \pm 7,93$	---	Burgos et al. (118)
Venezuela	Caracas	101	0,9939	---	$16,60 \pm 7,45$	$0,0137 \pm 0,007$	Castro et al. (117)

Nota: Valores de diversidad calculados con la RC de ADNmt completa.

Los parámetros de importancia forense también fueron calculados para cada provincia, incluyendo la diversidad haplotípica (H), el poder de exclusión (mtCE), el número medio de diferencias por pares (MNPd) y la diversidad nucleotídica (π) como se observa en la Tabla 13. Para el rango 16024-576, las diferencias en las diversidades haplotípicas entre grupos fueron inferiores al 1%, mientras que el poder de exclusión varió hasta un 5%. Las distribuciones del número de diferencias entre pares de secuencias se presentan en la Figura 24.

Figura 24. Distribuciones del número de diferencias entre pares de secuencias para los haplotipos en las provincias de Santander



El eje X muestra el número de diferencias por pares y el eje Y muestra el número de observaciones.

Todas las provincias del departamento mostraron perfiles multimodales. En estas distribuciones multimodales se observan haplotipos con un bajo número de diferencias (0-8) así como haplotipos con más de 10 diferencias. Este perfil es común en poblaciones estabilizadas, al

contrario de las distribuciones unimodales que son características de poblaciones en expansión (120,121). Esta distribución también podría ser explicada por los recientes movimientos de inmigración a Santander de un número considerable de individuos con antecedentes genéticos diferentes (49). Teniendo en cuenta la historia y la demografía reciente de la población santandereana, ambos fenómenos podrían haber contribuido a la generación de los perfiles observados.

Parte de la diversidad observada se debió a los Indel ubicados en los trectos homopoliméricos, y debido a que algunos laboratorios forenses descartan estos tramos para los análisis forenses (24), los parámetros de diversidad genética también se calcularon después de excluir estos polimorfismos y se presentan en la Tabla 13 Parte B. Los valores de diversidad genética y forenses disminuyeron considerablemente y las diferencias halladas entre las provincias fueron mayores al descartar estos Indel. En estas condiciones, la diversidad de haplotipos entre provincias varió hasta un 3% y el poder de exclusión hasta un 8%.

6.2.3 Linajes maternos en Santander

La mayoría de las muestras tuvieron ascendencia materna nativo americana (90%) clasificada dentro de los macrohaplogrupos A2, B4, C1 y D. Así mismo, la frecuencia hallada del aporte de linajes maternos no nativo americanos en Santander fue baja, correspondiendo solo al 10% de los haplotipos, distribuidos en un 6% de linajes de origen euroasiático identificados en ramas dentro de los macrohaplogrupos R y M, y un 4% de linajes de origen africano pertenecientes a los haplogrupos L0 a L3. (Tabla 15).

Tabla 15. Frecuencia de los haplogrupos mitocondriales en el departamento de Santander

Haplogrupo	n	%	Haplogrupo	n	%
A2	4	2,0%	H	3	1,5%
A2+(64)	35	17,2%	J1c	3	1,5%
A2+(64)+@153	3	1,5%	K1a	1	0,5%
A2+(64)+16129	3	1,5%	R0	3	1,5%
A2ac	21	10,3%	U5a2	1	0,5%
A2af1a1	2	1,0%	U5b	1	0,5%
A2r1	2	1,0%	Total Euroasiático	12	5,9%
A2v	5	2,5%			
A2y	1	0,5%			
B2b3a	2	1,0%			
B2d	60	29,4%			
B2h	1	0,5%			
B4b	1	0,5%			
B4	1	0,5%			
C1	10	4,9%			
C1b	1	0,5%	L0a1a+200	1	0,5%
C1c5	4	2,0%	L1b	1	0,5%
C1d+194	3	1,5%	L1c1'2'4'5'6	1	0,5%
D	1	0,5%	L2a1	1	0,5%
D1	8	3,9%	L2a1+143	1	0,5%
D1a2	2	1,0%	L3b	1	0,5%
D1f	10	4,9%	L3d5a	1	0,5%
D4h3a	4	2,0%	L3f1b1a1	1	0,5%

Total Nativo	184	90,2%	Total Africano	8	3,9%
---------------------	------------	--------------	-----------------------	----------	-------------

Un valor similar de linajes de ascendencia nativo-americana fue reportado previamente por Rojas et al. (42) para Santander (91,0%), mientras que una menor proporción fue encontrada por Yunis et al. (46) (82,6 %). Esta discrepancia puede deberse a diferencias en la estrategia de muestreo y el tamaño de muestra. Los datos informados para el ADNmt en otras poblaciones colombianas mezcladas muestran una variación en la proporción de ascendencia nativa entre los departamentos, lo que apunta a diferentes patrones de mezcla en todo el país, aunque en todos ellos los linajes nativos americanos son los más frecuentes (40,41,42,46). Proporciones similares de haplogrupos de ADNmt nativos (superiores al 80%) se informaron para poblaciones de los Andes occidentales de América del Sur y en Paraguay, mientras que la ascendencia nativa va disminuyendo en las poblaciones del centro y este de Suramérica, alcanzando los valores más bajos en las poblaciones de la costa atlántica de Brasil (por debajo del 35%) (110-119) (Tabla 16).

Tabla 16. Frecuencias de haplogrupos mitocondriales en poblaciones de Suramérica

País	Población	N	Haplogrupos						Euroasiático	Africano	Referencia
			A2	B4	C1	D	Total Nativo				
	Norte	98	0,2121	0,1111	0,2121	0,1212	0,6565	0,3233	0,0202		
Argentina	Central	193	0,1347	0,0932	0,1295	0,1243	0,4820	0,5076	0,0104	Bobillo et al. (119)	
	Sur	47	0,1087	0,1086	0,1304	0,3478	0,6955	0,2828	0,0217		
	Rio de Janeiro	205	0,0780	0,0975	0,0682	0,0146	0,2586	0,3219	0,4195	Simão et al. (112)	
	Espírito Santo	214	0,0880	0,0650	0,0600	0,0093	0,2242	0,4906	0,2850	Reis et al. (113)	
Brasil	Santa Catarina	80	0,0250	0,0875	0,0875	0,0125	0,2125	0,6375	0,1500	Palencia et al.	
	Paraná	122	0,1390	0,1560	0,1560	0,0410	0,4920	0,4010	0,1070	Poletto et al.	
	Distrito Federal	306	0,1209	0,0948	0,0719	0,0458	0,3334	0,1765	0,4902	Freitas et al. (116)	
	Iquique (Norte)	200	0,09	0,3050	0,3400	0,2100	0,9450	0,0450	0,0100		
	Santiago (Centro)	203	0,074	0,2610	0,2610	0,2270	0,8230	0,1580	0,0200		
Chile	Concepción	201	0,065	0,3030	0,2790	0,2390	0,8860	0,1090	0,0050	Gómez-Carballa et al. (110)	
	Temuco (Centro)	194	0,0460	0,2730	0,3090	0,2470	0,8760	0,1080	0,0150		
	Punta Arenas (Sur)	197	0,0100	0,1470	0,3650	0,3550	0,8780	0,1170	0,0050		
Ecuador	Mestizos	54	0,2590	0,2970	0,2220	0,1670	0,9410	0,0380	0,019	Burgos et al. (118)	
Paraguay	Este	548	0,2005	0,2703	0,2528	0,1482	0,8720	0,0780	0,0500	Simão et al. (111)	
	Caracas	101	0,3700	0,0800	0,1400	0,0800	0,6700	0,0800	0,2600		
Venezuela	Lara	81	0,2200	0,3600	0,0600	0,1000	0,7400	0,0100	0,2500	Castro de Guerra et al. (117)	
	Estados Oriente	125	0,4300	0,1800	0,1000	0,0400	0,7500	0,0700	0,1800		

Para los 12 haplotipos detectados con ascendencia europea (6% de las muestras) se encontraron secuencias idénticas dentro de las ramas H, J1c y R0 en EMPOP. Todos los haplogrupos europeos hallados se encuentran en altas frecuencias en toda Europa (48). Teniendo en cuenta el proceso de asentamiento histórico de Colombia, la mayoría de estos haplotipos muy

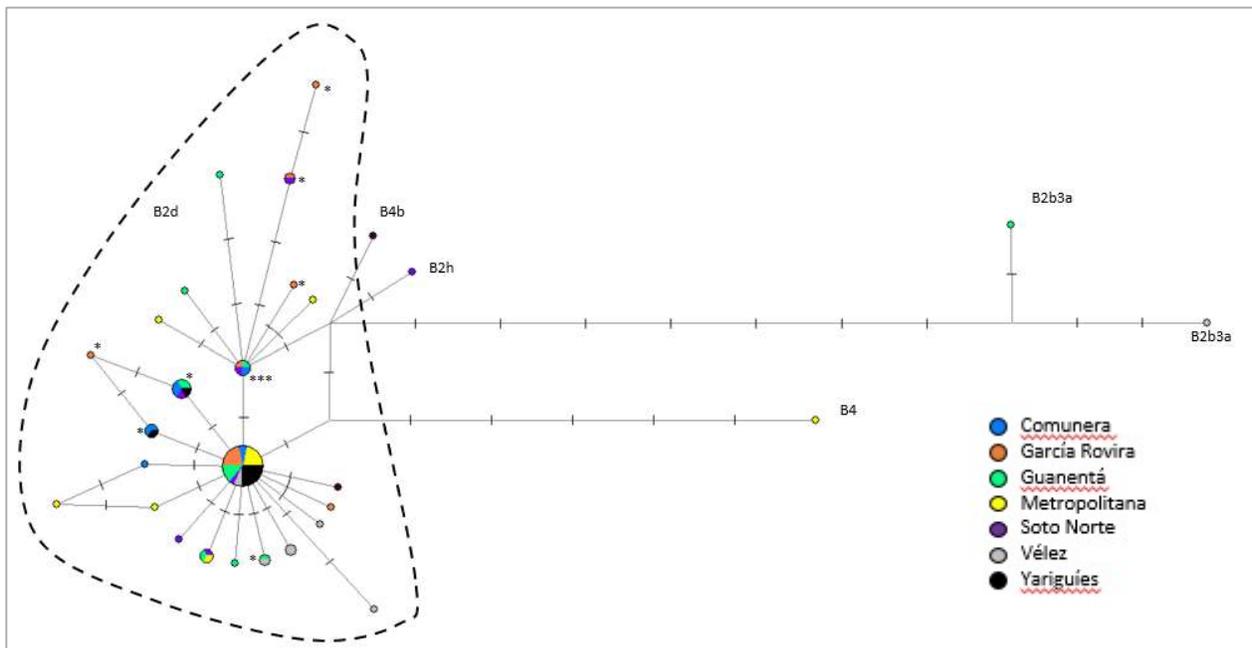
probablemente procedieron de la Península Ibérica (63). De hecho, la búsqueda en EMPOP mostró que los haplotipos J1c y U5b, encontrados en Santander, solo han sido detectados en muestras de Portugal y España. Por otro lado, de los 5623 haplotipos del haplogrupo H depositados en EMPOP, solo tres coincidieron con muestras de Santander tipificadas en este estudio, y las tres están reportadas en el continente americano (119,122).

Finalmente, la ascendencia africana (4% de las muestras) estuvo representada por ocho haplotipos clasificados dentro de ocho sub-haplogrupos L distintos (Tabla 15). Para rastrear el origen en África de los linajes L encontrados en Santander, se buscaron haplotipos compartidos entre muestras de Santander y de varias regiones de África (48). Solo tres haplotipos mostraron coincidencias dentro de África. Dos de los tres haplotipos (clasificados dentro de los haplogrupos L1b y L3f1b1a1) coincidieron con muestras del centro-oeste y centro-sur de África (48). El tercer haplotipo (del haplogrupo L3b) mostró una amplia distribución, desde el oeste hasta el este de África (48). Debido a la escasa representatividad de las poblaciones africanas en las bases de datos existentes, es difícil rastrear el origen de los linajes L tipificados en América hasta regiones africanas específicas (48,98). Por ello, son necesarios estudios más exhaustivos que permitan establecer con precisión la diversidad existente en el continente africano, para poder rastrear el origen geográfico de los linajes africanos que actualmente se encuentran en América. La mayor diversidad detectada en los haplotipos africanos en comparación con los euroasiáticos puede estar relacionada con la diversidad de las poblaciones de origen (15,98). Por un lado, las poblaciones africanas son las más diversas del mundo y, por otro, los africanos fueron llevados a América desde diferentes regiones de África (30,98). Otra explicación podría guardar relación con la traída de un mayor número de mujeres africanas que europeas a Colombia, durante la conquista y el período colonial (98).

6.2.4 Haplogrupos nativos americanos en Santander

Las ramas dentro de los haplogrupos A2 y B4 fueron las más comunes en Santander, en particular los linajes B2d, A2+(64) y A2ac, que en conjunto sumaron el 63% de los haplogrupos nativos encontrados. El haplogrupo B2d, caracterizado por una delección en la posición 498, se encontró en 60 muestras (29,4%) y presentó baja diversidad de secuencias, con un número medio de diferencias por pares de 1,35. La representación de la red de esta rama mostró que las secuencias difieren entre sí en un máximo de cinco posiciones (Figura 25).

Figura 25. Representación en red de las ramas dentro del haplogrupo B.

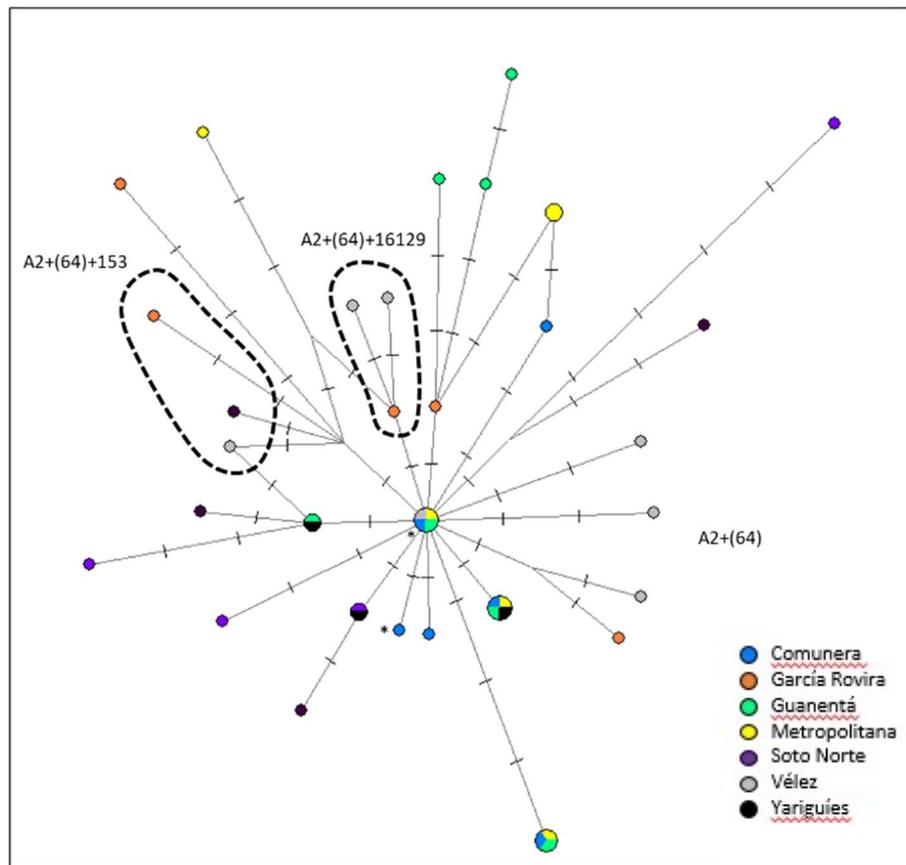


Adicionalmente, denota un elevado número de secuencias compartidas que corroboran la baja variabilidad intra-haplogrupo. La distribución en estrella de esta rama junto con la baja variabilidad detectada, apunta a eventos de cuello de botella o aislamiento. Este linaje es más frecuente en poblaciones de la parte baja de América Central y del norte de América del Sur de

acuerdo con la base de datos EMPOP, v4/R13, consultada el 10/01/2023, lo que podría indicar cierto grado de especificidad geográfica.

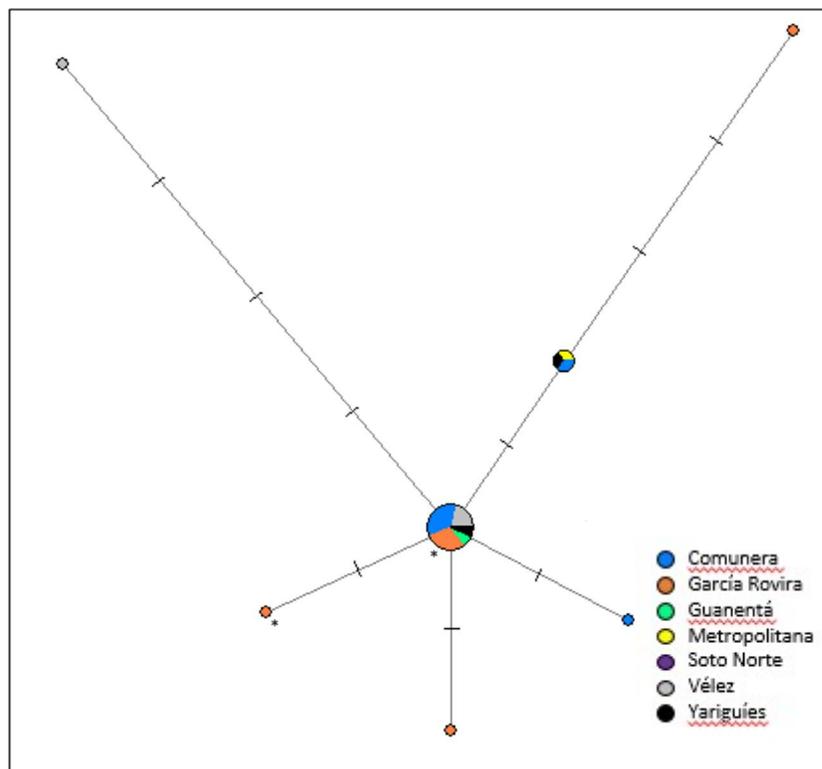
Según Phylotree (compilación 2016), los haplotipos A2+(64) comparten los polimorfismos 146C, 152T, 153G y 16111T con los haplotipos A2 y tienen un polimorfismo extra en la posición 64. Dentro de este haplogrupo, dos de los haplotipos de Santander no incluían el polimorfismo 153G y otros cuatro carecían de la transición en 16111. La distribución de la red de esta rama (Figura 26) mostró una gran variedad de haplotipos que diferían entre sí en 11 posiciones en promedio.

Figura 26. Representación en red de los haplogrupos A2+64, A2+64+153 y A2+64+16129.



El haplogrupo A2+(64) ha sido descrito previamente en poblaciones nativas y mezcladas de toda América, sin una especificidad geográfica o lingüística (48). Los haplotipos dentro de A2+(64) que presentan una transición G/A en la posición 16213 se clasificaron como A2ac. Catorce de las veintiun muestras de A2ac de Santander tenían la misma secuencia en la RC (después de descartar los Indel en trectos homopoliméricos). La mayoría de las muestras de A2ac en EMPOP (v4/R13, consultado el 10/01/2023) han sido descritas en la población colombiana de Antioquia (34). También se detectaron algunas en Ecuador y en los Estados Unidos de América. Al igual que la rama B2d, A2ac también mostró una distribución en forma de estrella (Figura 27).

Figura 27. Representación en red del haplogrupo A2ac.



El linaje D1f fue el siguiente en frecuencia en Santander, después de los anteriormente descritos y se ha encontrado en poblaciones de la parte norte de América del Sur, con frecuencias altas reportadas en Ecuador, Colombia y Venezuela (34,117,118).

6.2.5 Evaluación de la subestructura poblacional en Santander

6.2.5.1 Diferenciación genética entre provincias. Para evaluar la diferenciación intra poblacional dentro del departamento de Santander, se dividieron las 204 muestras según la provincia de nacimiento del individuo en: Comunera (n=30), García Rovira (n=30), Guantán (n=28), Metropolitana (n=36), Soto Norte (n=26), Vélez (n=23) y Yariguíes (n=31) (Figura 5). Los valores de F_{ST} por pares de la población se calcularon utilizando los haplotipos de la RC antes y después de descartar los Indel en los trectos homopoliméricos (Tabla 17).

Tabla 17. Valores de F_{ST} por pares entre provincias de Santander a partir de haplotipos

Haplotipos de la RC de ADNmt completa								
Provincia	n	Vélez	García Rovira	Comunera	Yariguíes	Guantán	Soto Norte	Metropolitana
Vélez	23	*	0,9885	0,5492	0,2150	0,0626	0,0519	0,0468
García Rovira	30	-0,0257	*	0,7096	0,1947	0,0561	0,0402	0,0446
Comunera	30	-0,0097	-0,0129	*	0,3382	0,0986	0,1735	0,2304
Yariguíes	31	0,0109	0,0122	0,0011	*	0,8419	0,3873	0,7945
Guantán	28	0,0420	0,0404	0,0258	-0,0160	*	0,1614	0,7268
Soto Norte	26	0,0422	0,0434	0,0138	-0,0002	0,0154	*	0,4628
Metropolitana	36	0,0389	0,0372	0,0076	-0,0127	-0,0113	-0,0028	*
Haplotipos de la RC de ADNmt sin Indel								
Provincia	n	Vélez	García Rovira	Comunera	Yariguíes	Guantán	Soto Norte	Metropolitana
Vélez	23	*	0,9712	0,5326	0,1985	0,0636	0,0660	0,0610

García Rovira	30	-0,0244	*	0,6383	0,1661	0,0529	0,0520	0,0456
Comunera	30	-0,0085	-0,0111	*	0,2872	0,0827	0,2363	0,2471
Yariguíes	31	0,0131	0,0150	0,0042	*	0,8214	0,6361	0,7908
Guanentá	28	0,0435	0,0426	0,0298	-0,0163	*	0,4274	0,7596
Soto Norte	26	0,0369	0,0390	0,0085	-0,0098	-0,0020	*	0,7450
Metropolitana	36	0,0355	0,0357	0,0070	-0,0128	-0,0125	-0,0112	*

Nivel de significancia después de la corrección de Bonferroni: $P=0,0024$. El valor de P se calculó con 10100 permutaciones. Valores de $s.e. \leq 0,0056$. Valores de F_{ST} (abajo de la diagonal) y valores de probabilidad (P) de no diferenciación (arriba de la diagonal).

Los resultados hallados fueron similares en ambos análisis. Se obtuvieron valores bajos de F_{ST} con valores de P de no diferenciación altos entre la mayoría de las provincias, lo que apoya una homogeneidad genética entre ellas. Como dato interesante, el valor de F_{ST} más bajo se obtuvo entre las poblaciones de García Rovira y Vélez, dos provincias geográficamente distantes entre sí. Estas poblaciones se encuentran en zonas montañosas de difícil acceso, quedando más aisladas que otras provincias. El aislamiento de grupos nativos con acervos genéticos idénticos, en el mismo momento histórico, podría explicar la similitud genética de estas poblaciones.

Aunque todos los valores de P fueron $> 0,0024$ (nivel de significancia aplicando la corrección de Bonferroni), algunas comparaciones mostraron distancias genéticas entre 1 y 5%, lo que indica pequeña diferenciación (124). Este es el caso de la mayoría de las comparaciones por pares que incluyen a García Rovira y Vélez, las dos provincias más aisladas, y algunas comparaciones por pares que incluyen a Soto Norte (Tabla 17). Estos valores de F_{ST} observados en algunas comparaciones por pares podrían deberse a contribuciones no nativas diferenciales en cada provincia. Para probar esta hipótesis se realizó un nuevo análisis después de excluir los haplotipos no nativos (Tabla 18).

Tabla 18. Valores de F_{ST} por pares entre provincias de Santander a partir de haplotipos

Haplotipos nativos americanos para la RC de ADNmt sin Indel								
Provincia	n	Vélez	García Rovira	Comunera	Yariguíes	Guanentá	Soto Norte	Metropolitana
Comunera	30	*	0,68468	0,08811	0,23384	0,43946	0,46817	0,25403
García Rovira	30	-0,01412	*	0,07841	0,06970	0,18869	0,98178	0,24225
Guanentá	28	0,03592	0,0376	*	0,69122	0,49352	0,06752	0,83615
Metropolitana	36	0,00973	0,03109	-0,01342	*	0,80626	0,05425	0,69419
Soto Norte	26	-0,00511	0,01645	-0,00737	-0,0189	*	0,13712	0,61568
Vélez	23	-0,00722	-0,02754	0,04925	0,04437	0,03006	*	0,19018
Yariguíes	31	0,00833	0,0083	-0,02175	-0,01362	-0,01488	0,01643	*

Nivel de significancia después de la corrección de Bonferroni: $P=0,0024$. El valor de P se calculó con 10100 permutaciones. Valores de $s.e. \leq 0,0057$. Valores de F_{ST} (abajo de la diagonal) y valores de probabilidad (P) de no diferenciación (arriba de la diagonal).

Los resultados mostraron que persistieron los mismos niveles de diferenciación, lo que demuestra que las diferencias entre provincias no pueden explicarse únicamente por la proporción de linajes africanos y/o euroasiáticos, sino también por diferencias en los antecedentes de los linajes nativos americanos. Debido a que el ADNmt es altamente polimórfico, los valores de F_{ST} basados en las frecuencias haplotípicas pueden no ser suficientemente sensibles para detectar la diferenciación poblacional, y detectar haplotipos compartidos. De acuerdo con esta perspectiva se estimaron los valores de F_{ST} usando las frecuencias de los haplogrupos, los cuales son menos variables (Tabla 19).

Tabla 19. Valores de F_{ST} por pares entre provincias de Santander a partir de haplogrupos

Frecuencias de haplogrupos de la RC ADNmt								
Provincia	n	Vélez	García Rovira	Comunera	Yariguíes	Guanentá	Soto Norte	Metropolitana
Vélez	23	*	0,98495	0,79269	0,59895	0,29898	0,21127	0,44213
García Rovira	30	-0,0271	*	0,88189	0,60143	0,18810	0,08296	0,27185
Comunera	30	-0,0156	-0,01565	*	0,26958	0,11999	0,19087	0,26938
Yariguíes	31	-0,0082	-0,00746	0,00603	*	0,97990	0,28423	0,93416
Guanentá	28	0,00555	0,01267	0,02232	-0,0222	*	0,19632	0,83239
Soto Norte	26	0,01078	0,02312	0,01234	0,00517	0,01207	*	0,62835
Metropolitana	36	-0,0021	0,00549	0,00578	-0,0152	-0,01348	-0,00688	*

Nivel de significancia después de la corrección de Bonferroni: $P=0,0024$. Valores de $s.e.\leq 0,0056$ para 10.100 permutaciones. Valores de F_{ST} (abajo de la diagonal) y valores de probabilidad (P) de no diferenciación (arriba de la diagonal)

Aunque los valores de F_{ST} disminuyeron, el patrón de diferenciación se mantuvo con algunas distancias genéticas menores al 2,5% asociadas a valores de P no significativos (Tabla 19). Cuando las muestras se clasificaron por los macrohaplogrupos A, B, C, D, R o L, surgió el mismo patrón, con distancias genéticas que alcanzaron el 4,5 % como se observa en la Tabla 20.

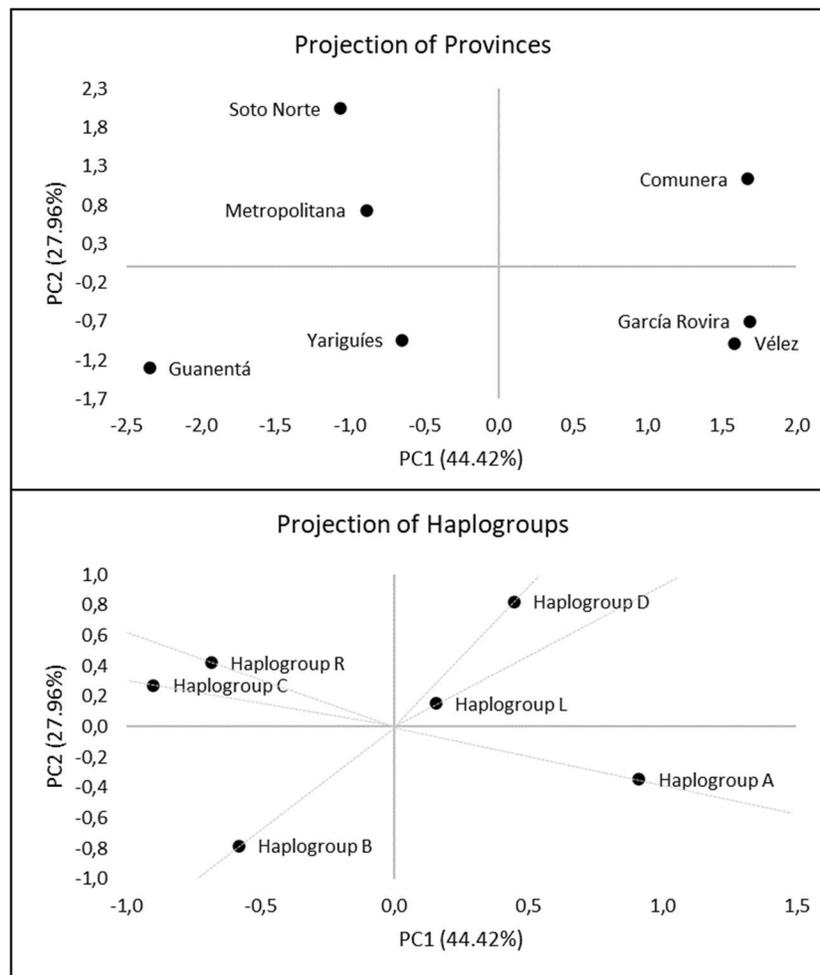
Tabla 20. Valores de F_{ST} por pares entre provincias de Santander a partir de los macrohaplogrupos.

Frecuencias de macro- haplogrupos de la RC ADNmt								
Provincia	n	Comunera	García Rovira	Guanentá	Metropolitana	Soto Norte	Vélez	Yariguíes
Comunera	30	*	0,68063	0,07633	0,40065	0,28898	0,61746	0,30037
García Rovira	30	-0,01561	*	0,08138	0,12355	0,06574	0,96614	0,17652
Guanentá	28	0,0383	0,04534	*	0,63449	0,35709	0,13197	0,79537
Metropolitana	36	-0,00088	0,02665	-0,0123	*	0,69439	0,16345	0,74181
Soto Norte	26	0,00849	0,04521	0,00333	-0,01456	*	0,12128	0,30185
Vélez	23	-0,01536	-0,03388	0,03154	0,02283	0,03107	*	0,32403
Yariguíes	31	0,00577	0,01858	-0,02014	-0,01563	0,00579	0,00412	*

Nivel de significancia después de la corrección de Bonferroni: $P=0,0024$. Valores de $s.e.\leq 0,0056$ para 10.100 permutaciones. Valores de F_{ST} (abajo de la diagonal) y valores de probabilidad (P) de no diferenciación (arriba de la diagonal)

6.2.5.2 Análisis de Componentes Principales (PCA). El PCA confirmó que las frecuencias de los macrohaplogrupos nativos contribuyen de manera importante a la diferenciación entre poblaciones (Figura 28).

Figura 28. PCA de las frecuencias de macrohaplogrupos de ADNmt en Santander



El primer componente principal separó las provincias de García Rovira, Vélez y Comunera de las de Soto Norte, Metropolitana, Yariguíes y Guanentá, debido a una mayor frecuencia de los haplogrupos A/D y B/C, respectivamente. El segundo componente separó la provincia Comunera de García Rovira y Vélez; así mismo separó las provincias Metropolitana y Soto Norte de Yariguíes y Guanentá. García Rovira y Vélez presentaron una frecuencia del haplogrupo A superior al 50%; mientras que Comunera, a pesar de tener una alta frecuencia de este haplogrupo (43%), también tuvo el mayor aporte del haplogrupo D. Guanentá y Yariguíes tuvieron alrededor del 40% de frecuencia del haplogrupo B, mientras que Soto Norte y Metropolitana tuvieron la mayor frecuencia del haplogrupo C.

Además del aporte nativo, los linajes euroasiáticos y africanos también parecieron contribuir para la diferenciación de las provincias. La ubicación del haplogrupo R en el cuadrante superior izquierdo del PCA indica que este linaje contribuyó a la separación de Metropolitana y Soto Norte de las demás provincias. Aunque en menor medida, el haplogrupo L contribuyó a la separación de Comunera de las demás poblaciones. Las diferencias en las proporciones de ascendencia nativa y no nativa entre provincias se evaluaron a través de una prueba de Fisher. Tanto García Rovira como Soto Norte mostraron valores de P significativos frente a todas las provincias ($P \leq 0,0373$). García Rovira no mostró haplogrupos no nativos mientras que Soto Norte tuvo la contribución no nativa más alta (23%).

6.2.5.3 Análisis de Varianza Molecular (AMOVA). Aunque se estimaron valores de F_{ST} entre 1 - 5% en algunas comparaciones por pares entre provincias (Tabla 20), los valores de P de no diferenciación nunca fueron estadísticamente significativos. No obstante, según el PCA, parecían existir diferencias en las frecuencias de haplogrupos entre algunas provincias. Por ello, se realizó el AMOVA considerando diferentes agrupamientos de las provincias con base en los resultados anteriores y de acuerdo con diferentes criterios geográficos y/o demográficos. En el agrupamiento A, las provincias se separaron, teniendo en cuenta la Cordillera Oriental de los Andes, en Oriente (García Rovira, Comunera, Guanentá, Metropolitana y Soto Norte) y Occidente (Yariguíes y Vélez). En el agrupamiento B, las provincias se agruparon considerando el primer componente principal (representando el 44.42% de la variación entre provincias): Grupo 1 (Guanentá, Metropolitana, Soto Norte y Yariguíes) y Grupo 2 (Comunera, García Rovira y Vélez). En el agrupamiento C, se tomaron en cuenta los dos primeros componentes principales y se agruparon las muestras según los cuatro cuadrantes del PCA: Grupo 1 (García Rovira y Vélez),

Grupo 2 (Guanentá y Yariguíes), Grupo 3 (Soto Norte y Metropolitana) y Grupo 4 (Comunera)
Los resultados se presentan en la Tabla 21.

Tabla 21. Resultados de AMOVA para los tres agrupamientos evaluados

	Entre grupos		Entre poblaciones dentro de grupos		Dentro de las poblaciones	
	% de variación	Probabilidad de no diferenciación	% de variación	Probabilidad de no diferenciación	% de variación	Probabilidad de no diferenciación
Agrupamiento A	-0,76	0,6594 ± 0,0046	1,28	0,0905 ± 0,0029	99,48	0,1337 ± 0,0030
Agrupamiento B	3,72	0,0280 ± 0,0017	-1,18	0,9477 ± 0,0022	97,47	0,1379 ± 0,0034
Agrupamiento C	2,97	0,0090 ± 0,0009	-1,62	0,9780 ± 0,0015	98,66	0,1346 ± 0,0034

Los resultados del AMOVA no mostraron una variación estadísticamente significativa entre las poblaciones agrupadas en el Este y el Oeste de la cordillera oriental (agrupamiento A), con una mayor variación dentro de los grupos que entre ellos (Tabla 21). Por el contrario, las estrategias de agrupamiento basadas en los resultados de PCA permitieron obtener los valores más altos de variación entre grupos, asociados a valores de *P* de no diferenciación estadísticamente significativos (Tabla 21). Las diferencias entre los grupos fueron mayores para la estrategia de agrupamiento B que para la C. Estos resultados de AMOVA mostraron que el agrupamiento B reflejaría mejor la estructura genética de la población de Santander.

6.3 Linajes de ADNmt en los departamentos de la región andina

En este apartado se presentan los resultados obtenidos del análisis de los haplotipos de la RC del mtDNA de los departamentos que conforman la región andina colombiana, excepto Santander, porque los resultados de este departamento ya fueron presentados y discutidos en el apartado anterior. Apéndice Apéndice

Los resultados obtenidos fueron sometidos a control de calidad en EMPOP y, una vez evaluados, pasaron este control sin tener ninguna corrección a la asignación de los haplotipos y haplogrupos. EMPOP emitió un certificado con la decisión de que los resultados de los mitotipos evaluados fueron aceptados en el contexto de la genética forense por EMPOP QC. El número de acceso de estos conjuntos de datos en esta base de datos es EMP00871 (Departamentos de la Región Andina, 458 datos) (Apéndice L).

6.3.1 Haplotipos hallados

Los haplotipos hallados y sus correspondientes haplogrupos, se presentan en el Apéndice I. Las muestras están identificadas con un código que inicia con la letra de cada departamento, seguida del número de muestra. Por ejemplo, las muestras de Antioquia corresponden a los códigos A001 a A049, las del departamento de Boyacá se identifican con el código B001 a B052 y así sucesivamente.

En las 458 muestras analizadas se detectó heteroplasmia de posición en 18 de ellas (3,93%), en 17 posiciones diferentes (Tabla 22).

Tabla 22. Heteroplasmias de posición detectadas

Heteroplasmia puntual	Muestra	Haplogrupo
127Y	Q011	C1b
146Y	T049	B2d
151Y	B019	A2+(64)
152Y	CU009	A2+(64)
195Y	B018	B2d
204Y	CU001	L1c1a
204Y	R047	A2ac
235R	N009	A2ac
16154Y	CA031	B2d
16172Y	B052	L3f1b1a1
16183M	B013	L2a1
16189Y	B023	A2+(64)
16223Y	N039	D1a2
16256Y	Q012	B2d
16260Y	N048	B2d
16265R	R041	D1f
16463R	CA008	D4h3a6
16465Y	B030	D4h3a

Código IUPAC: Y: mezcla de C y T

R: mezcla de A y G M: mezcla de A y C

Las heteroplasmas puntuales detectadas en las posiciones 127 C/T, 151 C/T, 235A/G, 16154 C/T y 16465 C/T, no se hallaron reportadas previamente en otras poblaciones en la base de datos EMPOP (v4/R13), consultada el 10/01/2023. El porcentaje de heteroplasmas puntuales hallado fue un poco más bajo que el encontrado en un estudio realizado en más de 5000 muestras de 20 poblaciones mundiales, donde fue estimada en 6%. En ese estudio no se detectaron heteroplasmas en las posiciones 127, 235, 16154, 16172, 16223, 16260, 16265, 16463 y 16465 (109), que si fueron halladas en el presente estudio.

Por otra parte, las heteroplasmas de longitud se detectaron en el 39,3% de las muestras y en 60 de estas (13,10%) se detectó en más de un lugar de la secuencia. La mayoría de éstas fueron en los tramos poli-C en HVSI y HVSII. En HVSI correspondieron a una sustitución T/C en la posición 16189 junto con inserciones C en la posición 16193 y/o sustituciones A/C en las posiciones 16182 y 16183 (34,93%). En HVSII, las heteroplasmas de longitud estuvieron asociadas a inserciones de C en las posiciones 309 y/o 315 (14,84%). También se detectaron heteroplasmas de longitud debidas a inserciones de C en las posiciones 463 y 573 (3,49 %) y con menor frecuencia en las repeticiones AC entre las posiciones 515 a 524 (0,87%). El porcentaje de heteroplasmas de longitud fue más bajo que el hallado en el estudio realizado en más de 5000 muestras; esta se encontró en el 52% de las muestras y coincidieron en las posiciones donde fueron detectadas, esto es en los tractos homopoliméricos de HVSI y HVSII (109).

Adicionalmente, se detectaron variantes nucleotídicas no reportadas previamente en ninguna de las 38.361 muestras genotipadas para la RC completa incluidas en la base de datos EMPOP (48), consultada 10/01/2023. Estas se hallaron en las posiciones 89 (inserción de una A en las muestras CU030 y Q043), 100 (Inserción de una A en la muestra CU030), 116 (delección de

una A en la muestra H007), 455 (inserción TTTY en muestra CA044) y 16038 (inserción de una A en la muestra T027).

6.3.2 Diversidad genética poblacional a partir de los haplotipos

Los resultados de los parámetros de diversidad hallados para cada uno de los departamentos se presentan en la Tabla 23.

Tabla 23. Parámetros de diversidad genética calculados para los departamentos de la Región Andina

Todas las posiciones (A)							
Departamento	n	Únicos	Únicos	H	mtCE	MNPD	π
Antioquia	49	32	65%	0,9889 +/- 0,0068	0,9728	15,8673 +/- 7,2003	0,01401 +/- 0,0070
Boyacá	52	39	75%	0,9894 +/- 0,0078	0,9465	14,4492 +/- 6,5780	0,0127 +/- 0,0064
Caldas	50	31	62%	0,9665 +/- 0,0180	0,938	17,1436 +/- 7,7518	0,0150 +/- 0,0075
Cundinamarca	54	44	81%	0,9965 +/- 0,0043	0,9734	13,7638 +/- 6,2767	0,0121 +/- 0,0061
Huila	50	41	82%	0,9951 +/- 0,0054	0,9845	16,0897 +/- 7,2944	0,0141 +/- 0,0071
Norte de Sder	50	45	90%	0,9967 +/- 0,0051	0,9788	14,7461 +/- 6,7111	0,0130 +/- 0,0065
Quindío	50	30	60%	0,9682 +/- 0,0140	0,9208	14,9967 +/- 6,8199	0,0132 +/- 0,0066
Risaralda	50	26	52%	0,9796 +/- 0,0094	0,9478	12,7355 +/- 5,8381	0,0112 +/- 0,0056
Tolima	53	38	72%	0,9935 +/- 0,0052	0,9775	16,1857 +/- 7,3290	0,0143 +/- 0,0072
Santander	204	124	61%	0,9921 +/- 0,0020	0,9660	15,1654 +/- 6,8062	0,0133 +/- 0,0066
Excluyendo los Indel (B)							
Departamento	n	Únicos	Únicos	H	mtCE	MNPD	π
Antioquia	49	26	53%	0,9813 +/- 0,0091	0,9413	13,5816 +/- 6,2077	0,0121 +/- 0,0061

Boyacá	52	32	62%	0,9661 +/- 0,0151	0,9072	12,0784 +/- 5,5491	0,0107 +/- 0,0054
Caldas	50	27	54%	0,9535 +/- 0,0207	0,9086	14,7559 +/- 6,7154	0,0130 +/- 0,0066
Cundinamarca	54	36	67%	0,9895 +/- 0,0061	0,9453	11,9692 +/- 5,4983	0,0106 +/- 0,0054
Huila	50	34	68%	0,9812 +/- 0,0107	0,9592	13,8563 +/- 6,3248	0,0123 +/- 0,0062
Norte de Sder	50	44	88%	0,9878 +/- 0,0104	0,9478	12,8236 +/- 5,8764	0,0113 +/- 0,0057
Quindío	50	27	54%	0,9445 +/- 0,0199	0,8506	12,9240 +/- 5,9200	0,0115 +/- 0,0058
Risaralda	50	18	36%	0,9322 +/- 0,0240	0,8000	9,8424 +/- 4,5811	0,0087 +/- 0,0045
Tolima	53	34	64%	0,9862 +/- 0,0074	0,9565	14,2162 +/- 6,4748	0,0126 +/- 0,0063
Santander	204	95	47%	0,9783 +/- 0,0055	0,9285	13,0538 +/- 5,9010	0,0115 +/- 0,0057
Antioquia 1000 genomas*	94	36	38%	0,9407 +/- 0,0154	0,8817	11,9414 +/- 5,4499	0,0106 +/- 0,0053

* Corresponde al análisis de los datos de Antioquia 1000 genomas en la tabla sin Indel porque solo están disponibles las secuencias sin los Indel.

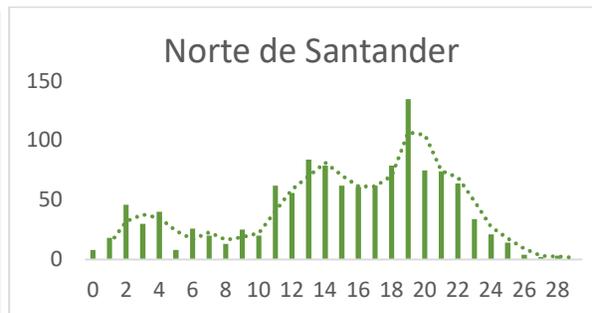
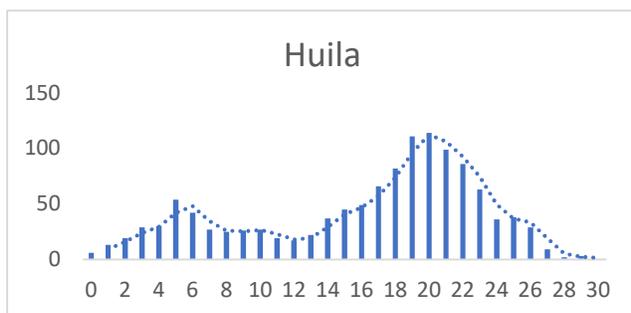
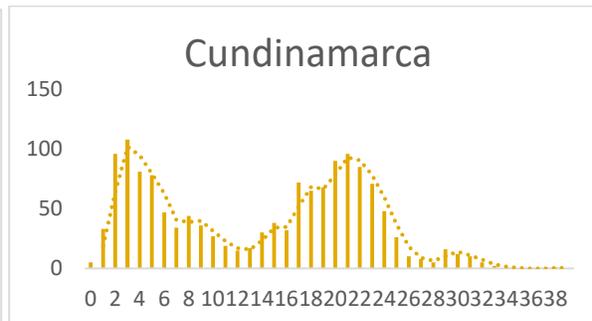
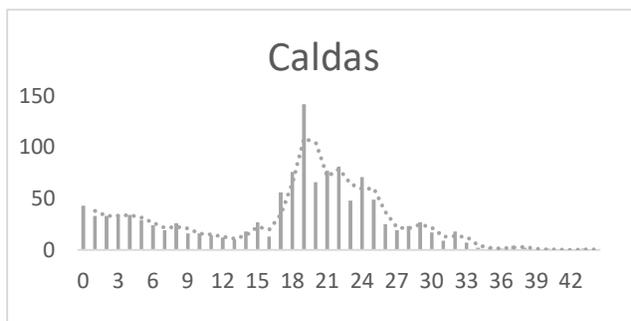
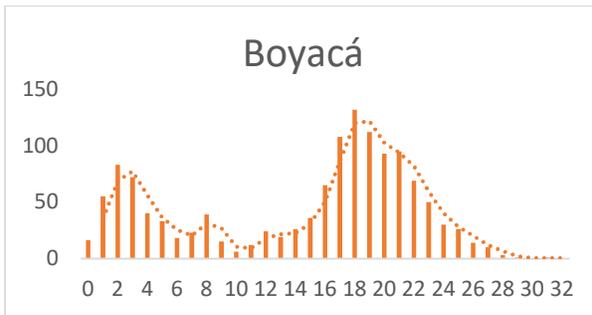
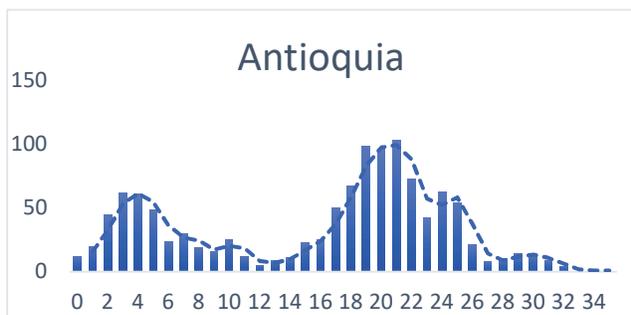
Los resultados encontrados muestran que el departamento que presentó más haplotipos únicos fue el Norte de Santander y el de menor número fue Risaralda. La diversidad haplotípica estuvo entre 0,9665 y 0,9967, siendo alta para la mayoría de los departamentos (>0,992). La diversidad más elevada se halló en Norte de Santander y la más baja en Caldas. El poder de exclusión más alto se halló en el conjunto de muestras de Huila y el más bajo en el Quindío. Por su parte los valores más altos de MNPD y de π se hallaron en Caldas y los más bajos se hallaron en Risaralda.

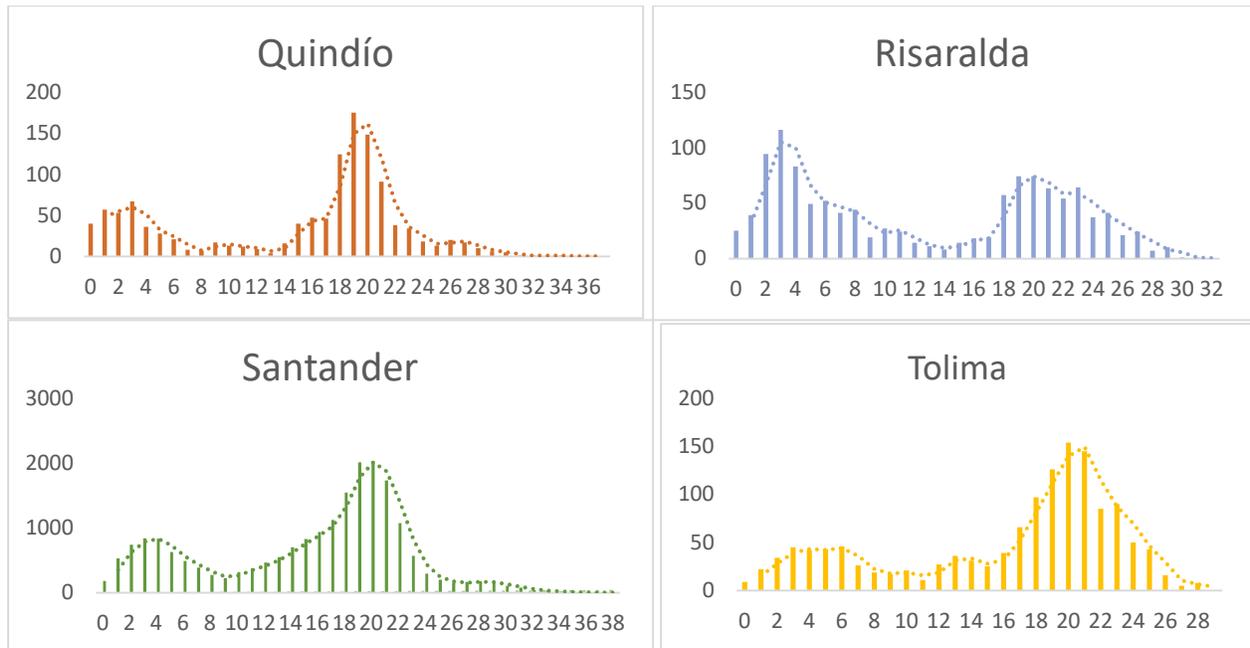
De acuerdo con estos resultados, la diversidad haplotípica hallada estuvo en el rango de diversidades haplotípicas reportadas anteriormente en otras poblaciones mezcladas de América del Sur que cuentan con estudios publicados del análisis de la RC completa del ADNmt, como Brasil,

Chile, Ecuador, Paraguay y Venezuela (112-118) y fueron más altas que las reportadas para todas las regiones de Argentina (119) (Tabla 16). De igual forma, los valores obtenidos de MNPD y de diversidad nucleotídica para los departamentos de la región andina colombiana fueron muy similares a los reportados en estas poblaciones suramericanas (Tabla 14). Finalmente, el poder de exclusión obtenido en las muestras analizadas de los departamentos de la región andina colombiana es ligeramente inferior respecto al reportado para Rio de Janeiro (112), único estudio que reportó este parámetro forense. Comparando las Tablas 16 y 23 se puede observar que la diversidad haplotípica obtenida en la población de Rio de Janeiro es similar a la obtenida en la mayoría de los departamentos de la región andina colombiana. Sin embargo, el poder de exclusión obtenido en los departamentos de la región andina colombiana es inferior en todos los casos al reportado para Rio de Janeiro, indicando que los haplotipos hallados en Rio de Janeiro son más diversos, lo que es soportado por el valor de MNPD más alto por el hecho de que la población de Río de Janeiro está conformada en una alta proporción por afrodescendientes que poseen los linajes más diversos del mundo.

Por otra parte, las distribuciones del número de diferencias entre pares de secuencias mostraron perfiles multimodales en todos los departamentos de la región andina. Los resultados se presentan en la Figura 29.

Figura 29. Distribuciones del número de diferencias entre pares de secuencias para los haplotipos en los departamentos de la Región Andina.





El eje X muestra el número de diferencias por pares y el eje Y muestra el número de observaciones.

Estas distribuciones multimodales mostraron haplotipos con un bajo número de diferencias, así como haplotipos con más de 10 diferencias, perfil que es común en poblaciones estabilizadas (120,121). Esta distribución también podría ser explicada por los movimientos de inmigración en los diferentes departamentos, que en forma más reciente se han dado en Colombia, como consecuencia del desplazamiento forzado y la migración del campo y los pueblos pequeños hacia las grandes ciudades en búsqueda de mejores oportunidades de vida, de un número considerable de individuos con antecedentes genéticos diferentes (49).

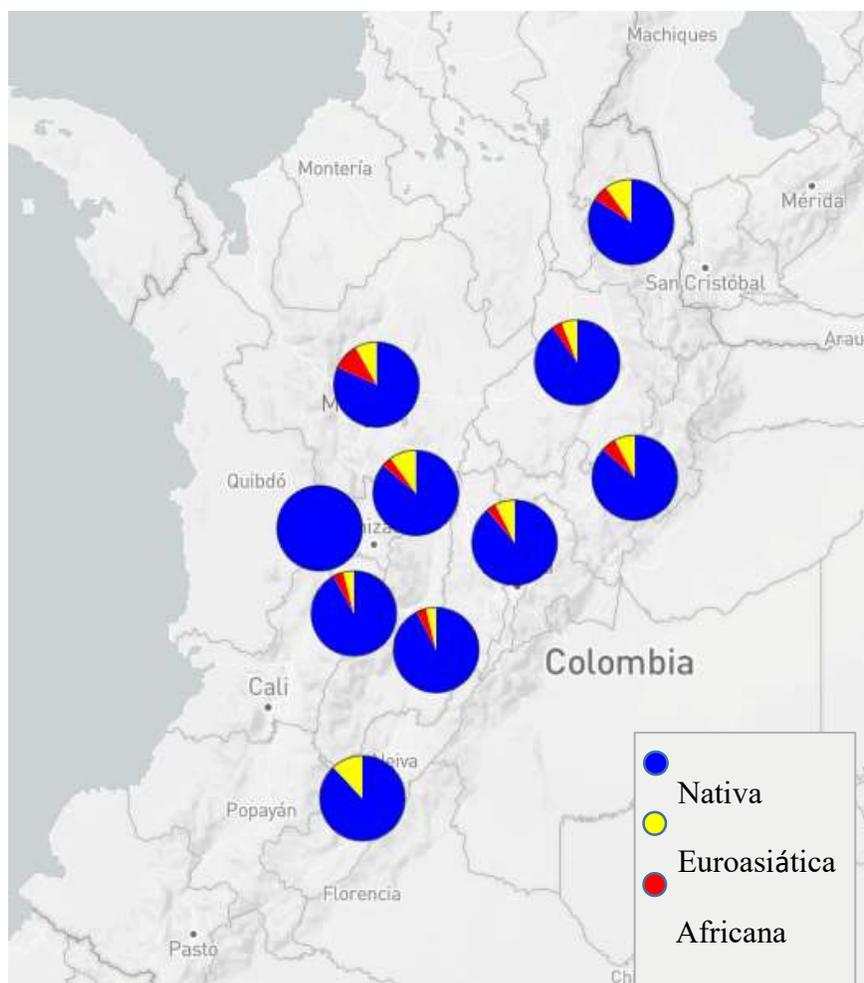
Por otra parte, se calcularon los parámetros de diversidad genética para la RC excluyendo los Indel de los tractos homopoliméricos de las posiciones 16193, 309, 315, 523-524 y 573, debido a que, como ya se indicó, algunos laboratorios forenses descartan estos tramos en sus análisis. Los resultados obtenidos se pueden observar en la Tabla 23 parte B. En todos los departamentos

disminuyeron los valores de los parámetros de diversidad, como era de esperarse, porque parte de la diversidad presente en las secuencias, está ubicada en los sitios polimorfismos de estos tractos homopoliméricos.

6.3.3 Linajes Maternos en la Región Andina

Los linajes maternos hallados en todos los departamentos de la región andina con su respectiva frecuencia se presentan en el Apéndice M. En la Figura 30 se puede observar gráficamente la distribución de la ancestría en cada departamento.

Figura 30. Ancestría de los haplogrupos por departamento



Diseñada con: <https://microreact.org/>

La mayor frecuencia de haplogrupos fue de ascendencia nativa americana, clasificada dentro de los macrohaplogrupos A2, B4, C1 y D. El porcentaje de haplogrupos con ancestría nativa varió entre el 82% (Antioquia) y el 100% (Risaralda), con un promedio de 89,3%. Por su parte, la frecuencia de linajes de origen no nativo fue baja, con una contribución de haplogrupos euroasiáticos pertenecientes a ramas dentro de los macrohaplogrupos R y M que estuvo entre 0% (Risaralda) y el 12% (Huila), con un promedio de 6,6%. Estos haplogrupos correspondieron en su mayoría a ramas del haplogrupo R (87,5%), que se encuentran en una alta frecuencia en Europa, incluida la península ibérica de donde procedieron los europeos que llegaron al país a inicios del siglo XVI (59). En orden de frecuencia, los haplogrupos hallados fueron J (22,5%), H (20%), R (17,5%), U (17,5%), T (5%) y K (5%). El resto de haplogrupos de ascendencia euroasiática pertenecieron al macrohaplogrupo M (12,5%), que se encuentra en altas frecuencias en el sureste asiático y oriente medio y en menores frecuencias en Europa, de acuerdo con las muestras que conforman la base de datos EMPOP consultada el 10/01/2023 (48).

Respecto a los haplogrupos de ascendencia africana, ubicados dentro de los linajes L0 a L3, tuvieron una frecuencia que varió entre 0% (Risaralda) y 10% (Antioquia) con un promedio de 4,1%. Estos fueron encontrados en todos los departamentos con excepción de Huila y Risaralda y estuvieron dentro de las ramas del macrohaplogrupo L incluyendo L0, L1, L2 y L3. En orden de frecuencia, los haplogrupos hallados fueron: L3 (37%), L2 (30%), L1 (26%) y L0 (7%). El 78% de los individuos tuvieron diferentes sub-haplogrupos dentro de estas cuatro ramas del macrohaplogrupo L, mostrando la alta diversidad que presentan en las poblaciones analizadas. Los haplogrupos hallados están registrados en la EMPOP en muestras cuyos orígenes se remontan a varias regiones de África, principalmente el centro-oeste, y centro-este y, en menor proporción, en

el norte de este continente consulta realizada 10/01/23 (48). Como ya se ha mencionado, la escasa representatividad de los haplotipos africanos en las bases de datos existentes dificulta determinar más exactamente el origen de los linajes L en regiones africanas específicas. Sin embargo, al someter a EMPOP los haplotipos pertenecientes a los haplogrupos L, hallados en las muestras de este estudio se encontraron haplotipos iguales que estaban reportados allí de muestras de poblaciones provenientes de diversos puntos del continente africano, lo cual está acorde con el hecho histórico de que los africanos traídos como esclavos a América provenían de diversas partes de ese continente (30,123).

Los resultados obtenidos en el presente estudio se compararon con otros publicados para poblaciones Andinas de Colombia. En la Tabla 24 se presentan las frecuencias reportadas en varios estudios para poblaciones mezcladas de los departamentos de la región andina. En el caso de las poblaciones que fueron evaluadas en más de un estudio, se presentan las frecuencias ponderadas. No fue posible diferenciar la ancestría euroasiática de la africana en la mayoría de esos estudios porque solo reportaron ancestría no nativa, muy probablemente por la dificultad para diferenciar estas dos ancestrías con el nivel de resolución del método molecular empleado (PCR-RFLP).

Tabla 24. Frecuencias publicadas para los haplogrupos, en departamentos de la región andina.

Población	N	Haplogrupos						Referencia
		A	B	C	D	Total Nativo	No nativos	
Bogotá	91	0,3730	0,2640	0,0770	0,0660	0,7800	0,2200	Rodas et al. (38)
Antioquia	131	0,4046	0,3863	0,0733	0,0214	0,8855	0,1145	Rojas et al. (42), Yunis et al. (46)
Boyacá	150	0,5067	0,2733	0,1000	0,0867	0,9667	0,0333	Yunis et al. (46)
Caldas	193	0,5300	0,3700	0,0800	0,0000	0,9800	0,0200	Rojas et al. (42)

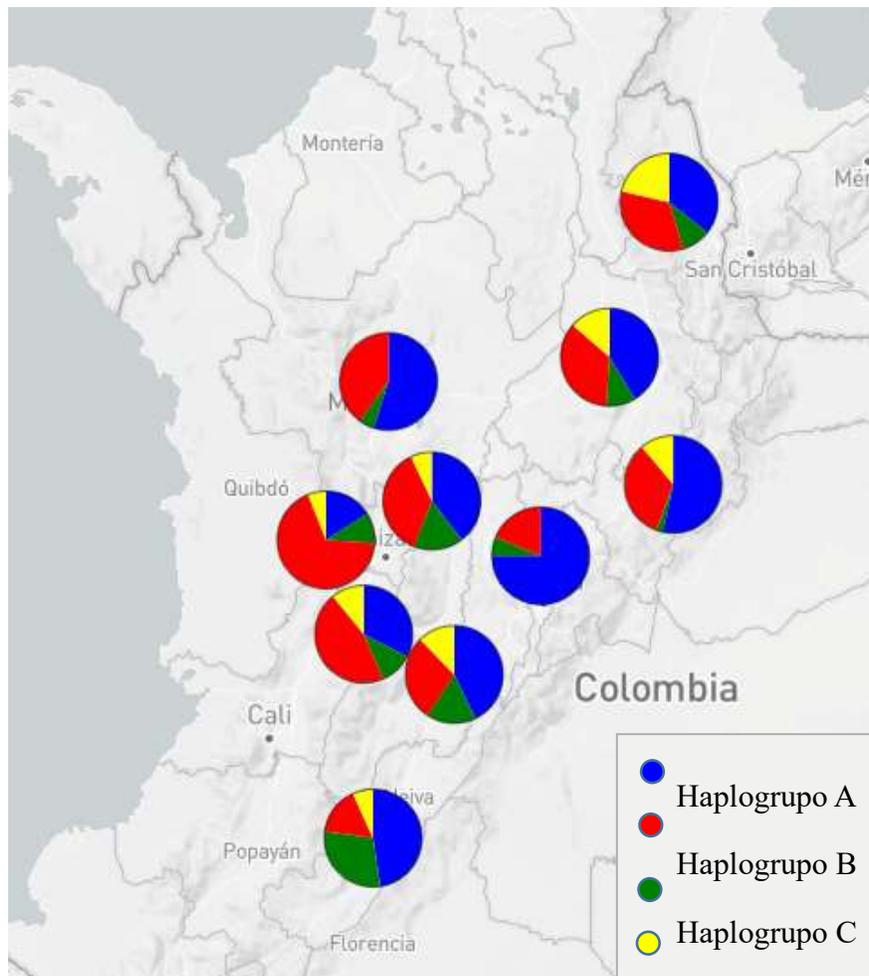
Cundinamarca	175	0,4377	0,3509	0,0286	0,0457	0,8629	0,1371	Rojas et al. (42), Yunis et al. (46)
Huila	24	0,3300	0,3800	0,1300	0,0800	0,9200	0,0800	Rojas et al. (42)
Tolima	353	0,3022	0,3131	0,1971	0,0825	0,8949	0,1051	Criollo et al. (43), Yunis et al. (46)
Norte de Sder	185	0,4422	0,2170	0,0603	0,1676	0,8870	0,1130	Rojas et al. (42), Yunis et al. (46)
Peque	163	0,1700	0,6600	0,1100	0,0000	0,9400	0,0600	Rojas et al. (42)
Quindío	58	0,3600	0,4300	0,1400	0,0200	0,9500	0,0500	Rojas et al. (42)
Santander	232	0,2717	0,3271	0,0635	0,1938	0,8561	0,1439	Rojas et al. (42), Yunis et al. (46)

Los resultados para la ancestría hallada en el presente estudio (Apéndice M) están acordes con los resultados reportados en estos estudios presentados en la Tabla 24, donde se observa una variación en la proporción de ascendencia nativa entre los diferentes departamentos de la región andina colombiana lo que podría deberse a diferentes patrones de mezcla en el país, pero conservando en todos ellos, a los linajes nativos americanos como los más frecuentes.

De igual manera, proporciones similares de haplogrupos de ADNmt nativos (>80%) se informaron en poblaciones andinas de América del Sur como Ecuador, Venezuela, Chile y Paraguay (Tabla 16). Por el contrario, en las poblaciones mezcladas de Brasil la ascendencia nativa disminuyó hasta alcanzar valores muy bajos en las poblaciones de la costa atlántica como Rio de Janeiro (<35%) (Tabla 16).

6.3.4 Linajes nativos americanos en la región andina

Como se mencionó anteriormente, los linajes nativos americanos fueron los más frecuentes en todos los departamentos (>82%). En la Figura 31 se puede observar la distribución de los macrohaplogrupos nativos A, B, C, y D en los diferentes departamentos de la región andina.

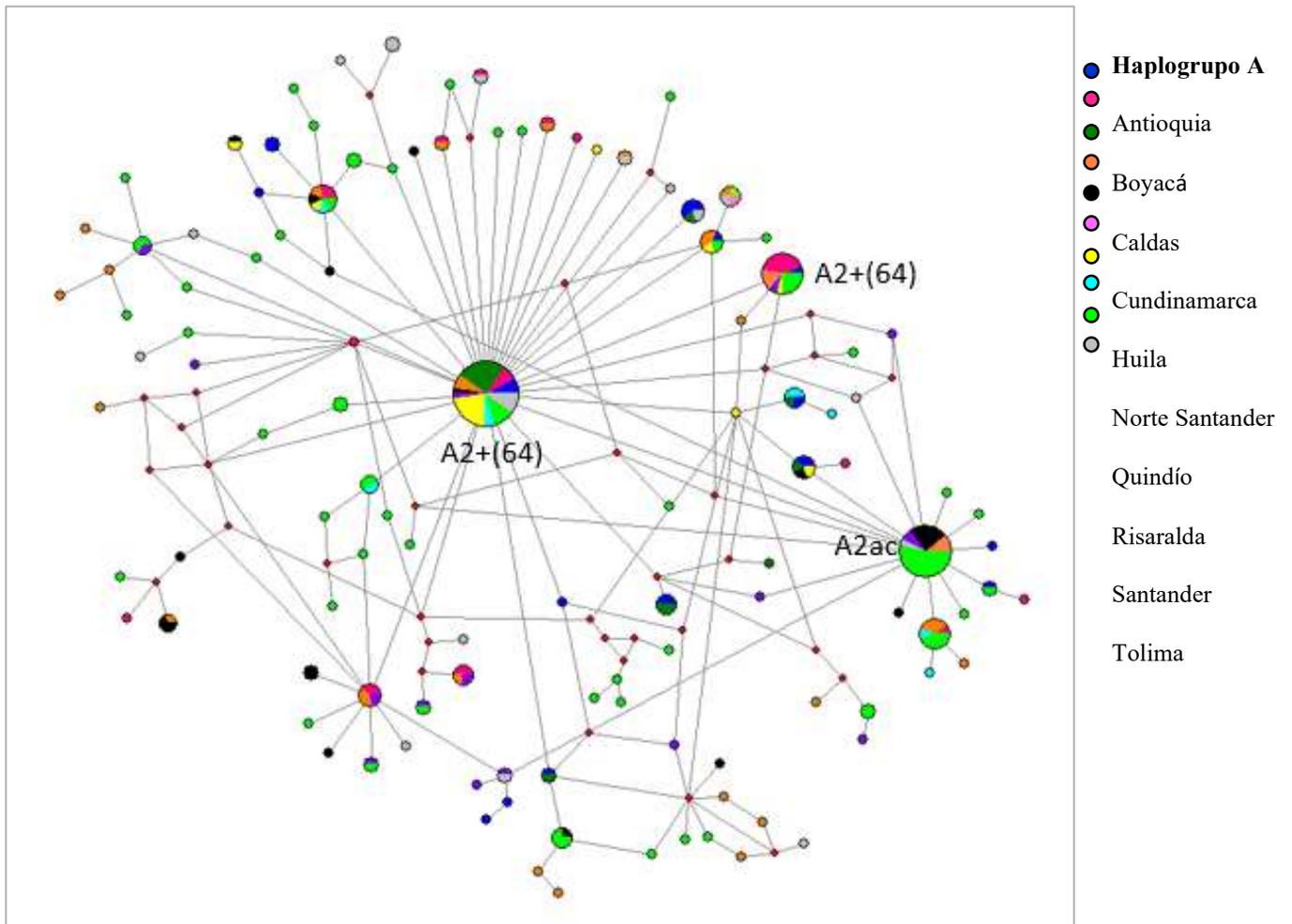
Figura 31. Distribución de los macrohaplogrupos nativos americanos por departamento

Diseñada con: <https://microreact.org/>

Dentro de los haplogrupos nativos hallados en los departamentos de la región andina, en promedio el haplogrupos A fue el más frecuente (38,3%), seguido del haplogrupos B (32,4%), luego el haplogrupos C (10,3%) y con la menor frecuencia el haplogrupos D (7,9%). Las ramas más comunes dentro de cada macrohaplogrupos fueron: dentro del haplogrupos A, el A2+(64) que se halló en 132 muestras (19,93%). De acuerdo con Phylotree (97), compilación de febrero de 2016, este haplogrupos está caracterizado por poseer variaciones en las posiciones 146C, 152T, 153G y 16111T (características del haplogrupos A2), con una variación adicional en la posición 64 donde

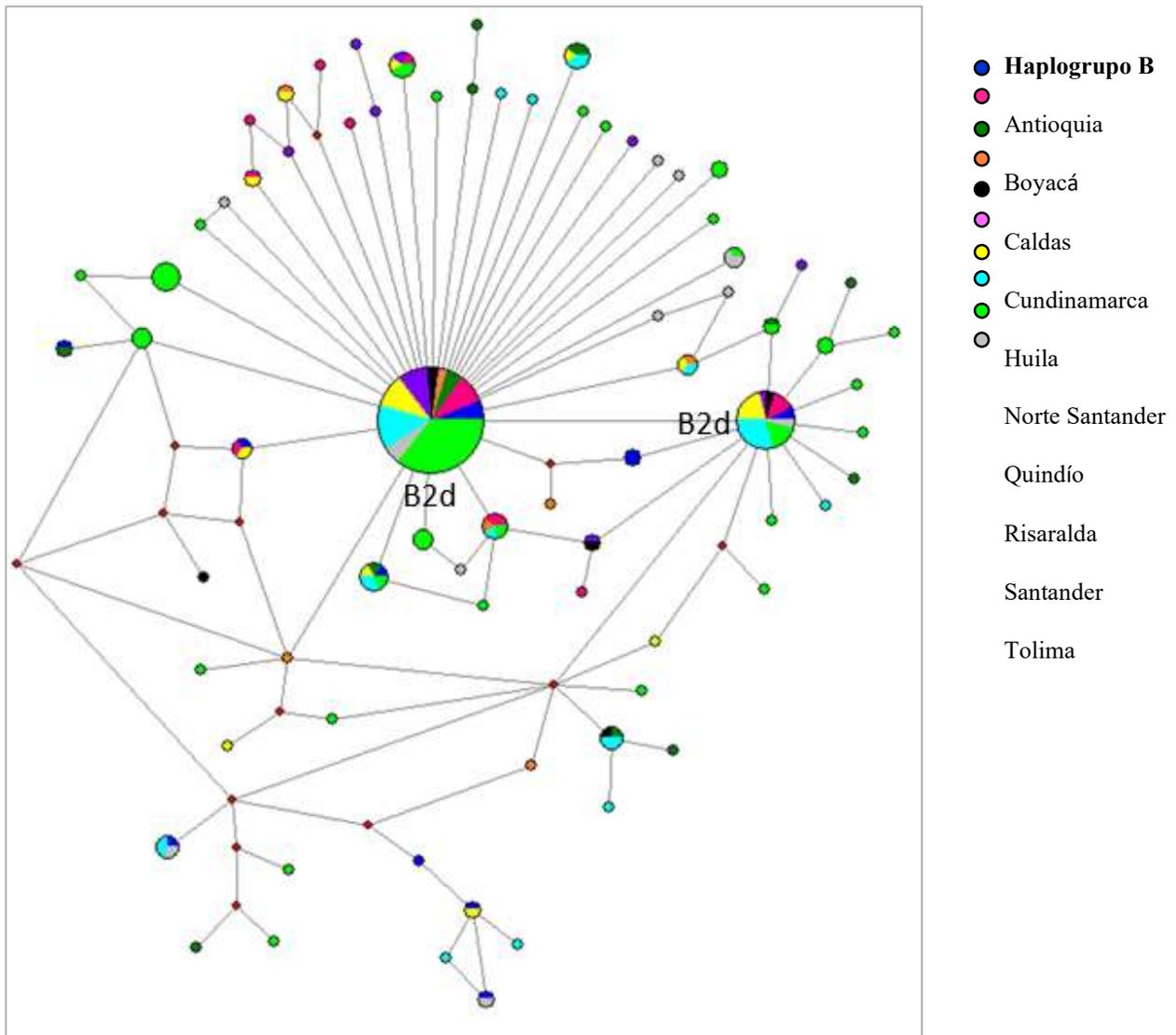
hay una transición de una C por T (64T) con respecto a la rCRS. El haplogrupo A2ac, hallado en 56 muestras (8,45%), fue el segundo en frecuencia dentro del macrohaplogrupo A. El A2ac se caracteriza por poseer una variación adicional en la posición 16213 donde posee una A en vez de G. El resto de las muestras de este haplogrupo fueron clasificadas en subhaplogrupos de A2, con contribuciones mínimas en los distintos departamentos. Estos resultados se pueden observar en la representación en red de esta rama (Figura 32), en cuyo centro se ubica A2+(64) que estuvo presente en todos los departamentos y a partir del cual se distribuyen los demás subhaplogrupos derivados de A2.

Figura 32. Representación en red de los haplogrupos A

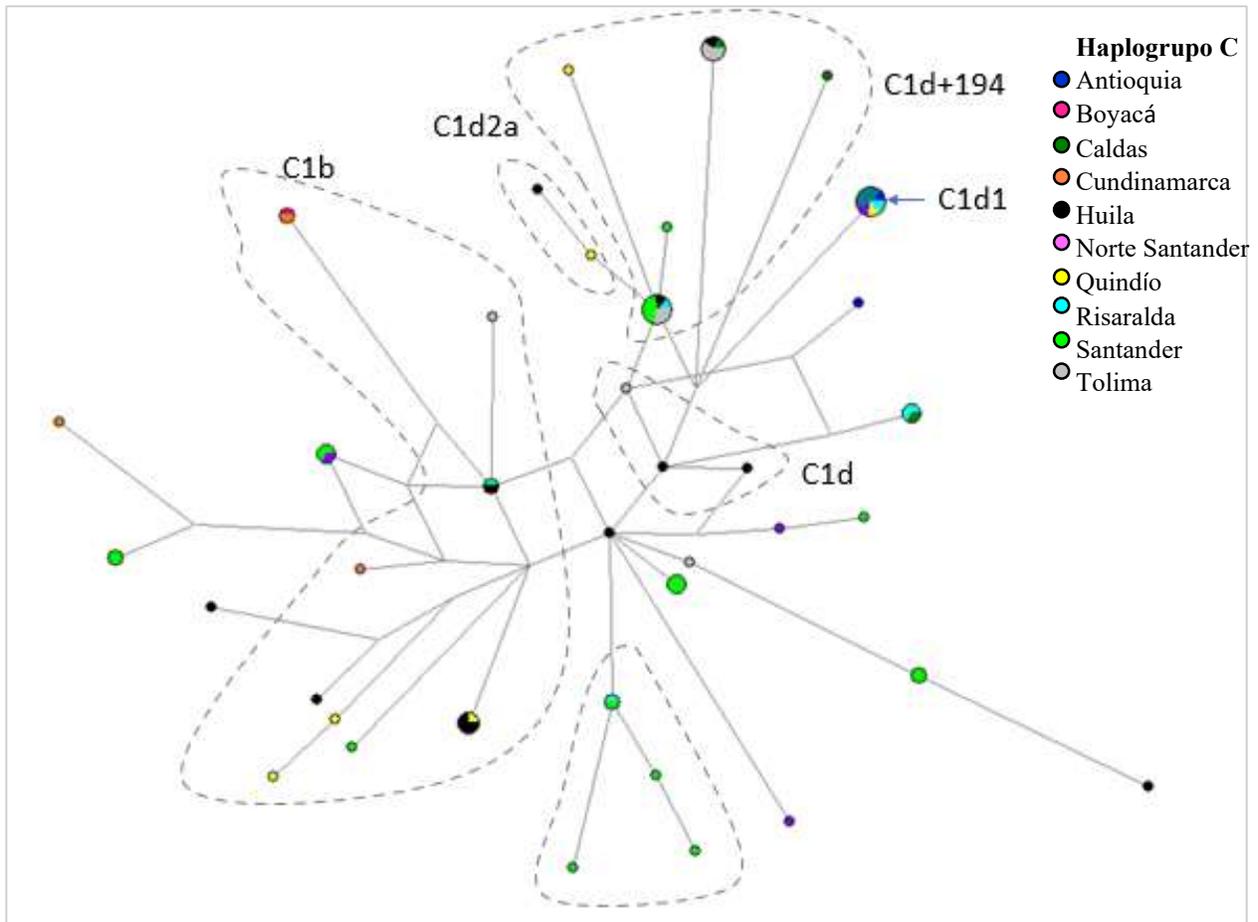


Dentro del haplogrupo B, el B2d fue el haplogrupo más frecuente en todos los departamentos. Este haplogrupo, caracterizado por una delección en la posición 498, se encontró en 180 muestras (27,19%). Al igual que para Santander, en las muestras de los demás departamentos de la región andina se detectó una baja diversidad de secuencias, con un número medio de diferencias bajo (Figura 33).

Figura 33. Representación en red de los haplogrupos B

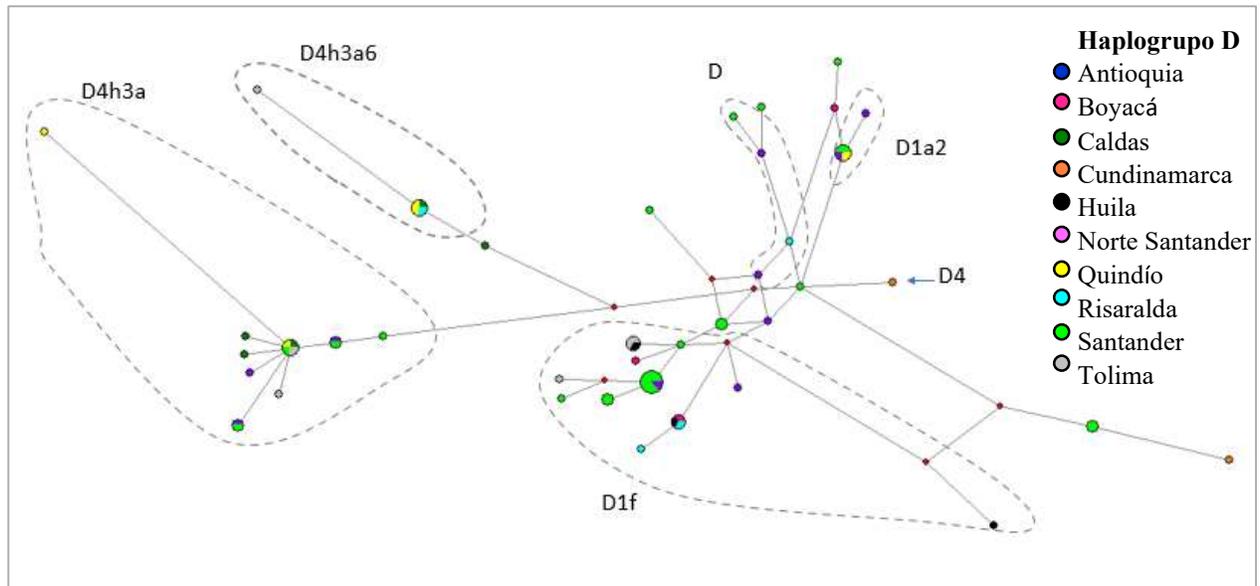


El haplogrupo C estuvo representado por varios subhaplogrupos de C1, con una contribución homogénea pero baja en todos los departamentos. Este haplogrupo se caracteriza por las deleciones en las posiciones 290 y 291 y una transición en la posición 16325 de una T por una C (16325C) (97). El más común fue el C1b que posee adicionalmente una transición en la posición 493 de una A por una G (493G) y se presentó en 13 muestras (2,0%), de las cuales 5 se hallaron en Huila, departamento que presentó la mayor contribución de este haplogrupo entre todos los departamentos. (Figura 34).

Figura 34. Representación en red de los haplogrupos C.

Nota: Los haplogrupos que no están delimitados son C1.

De los haplogrupos nativos americanos, el haplogrupo D fue el menos frecuente en promedio, con una contribución muy baja en todos los departamentos. El subhaplogrupo D1f, caracterizado por una transición en la posición 16142 de una C por una T (16142T), fue el más común y fue asignado a 21 muestras (3,17%), halladas principalmente en Norte de Santander y Santander que fueron los departamentos con mayor frecuencia del haplogrupo D aunque en baja proporción (18% y 12% respectivamente). En la Figura 35 se puede observar la representación de este haplogrupo en la red, los haplogrupos que no están delimitados son D1.

Figura 35. Representación en red de los haplogrupos D

Nota: Los haplogrupos que no están delimitados son D1.

6.3.5 Evaluación de la subestructura poblacional en los departamentos de la región andina

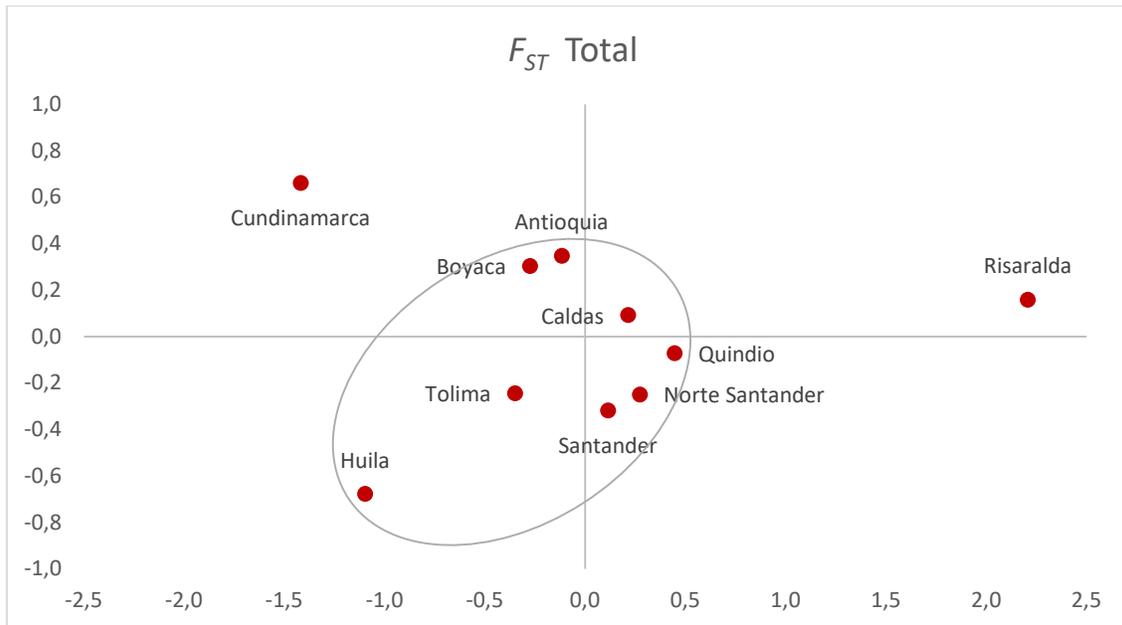
6.3.5.1 Diferenciación genética entre departamentos. Para evaluar la hipótesis nula planteada en este estudio, se realizó una prueba de diferenciación genética entre los diez departamentos que conforman la región andina colombiana. Se calcularon los valores de F_{ST} por pares a partir de los haplotipos de la RC de las muestras de cada departamento, incluido Santander. Los valores de F_{ST} y de probabilidad de no diferenciación (P) obtenidos se presentan en la Tabla 25. Para una mejor visualización de los valores de F_{ST} obtenidos, se realizó una representación gráfica en MDS (Figura 36).

Tabla 25. Valores de F_{ST} por pares entre los departamentos de la región andina

Haplotipos de la RC total de ADNmt											
		Antioquia	Boyacá	Caldas	Cundinamarca	Huila	Norte de Sder	Quindío	Risaralda	Santander	Tolima
Antioquia	49	*	0,5381	0,3594	0,0736	0,0164	0,0948	0,2007	0,0001	0,1590	0,2342
Boyacá	52	-0,0043	*	0,1393	0,0756	0,0132	0,2072	0,1527	0,0001	0,2922	0,3210
Caldas	50	-0,0001	0,0102	*	0,0039	0,0420	0,3810	0,7765	0,0034	0,2865	0,3483
Cundinamarca	54	0,0174	0,0172	0,0503	*	0,0123	0,0013	0,0015	0,0000	0,0009	0,0251
Huila	50	0,0337	0,0353	0,0233	0,0348	*	0,0133	0,0057	0,0000	0,0051	0,1788
Norte de Sder	50	0,0143	0,0057	-0,0003	0,0654	0,0333	*	0,5894	0,0010	0,7756	0,3390
Quindío	50	0,0063	0,0092	-0,0083	0,0696	0,0458	-0,0047	*	0,0167	0,3925	0,2287
Risaralda	50	0,0975	0,1098	0,0595	0,2099	0,1653	0,0709	0,0383	*	0,0000	0,0000
Santander	204	0,0057	0,0014	0,0016	0,0431	0,0303	-0,0050	-0,0006	0,0766	*	0,2770
Tolima	53	0,0047	0,0011	0,0006	0,0277	0,0075	0,0010	0,0051	0,1035	0,0018	*

Nivel de significancia después de la corrección de Bonferroni 0,0011. El valor de P se calculó con 50175 permutaciones. Valores de $s.e. \leq 0,0023$. Valores de F_{ST} (abajo de la diagonal) y valores de probabilidad (P) de no diferenciación (arriba de la diagonal). Los valores de P significativos están resaltados en negro.

Figura 36. Valores de F_{ST} por pares entre los departamentos de la región andina



Al observar los valores de F_{ST} obtenidos (Tabla 25) y graficados en el MDS (Figura 36), se evidenció que los departamentos de Antioquia, Boyacá, Caldas, Norte de Santander, Quindío, Santander y Tolima presentaron distancias genéticas bajas con valores de P de no diferenciación altos. Por su parte el departamento de Huila pese a presentar en algunas comparaciones valores de F_{ST} entre 2 – 4%, todos los valores de P fueron no significativos. Este hallazgo apoya la idea de proximidad genética, a partir de los linajes maternos, entre estos ocho departamentos, por lo cual se presentan agrupados en el MDS (Figura 36). Por el contrario, el departamento de Risaralda presentó diferencias significativas (valores de P de no diferenciación $<0,0011$ después de la corrección de Bonferroni) con la mayoría de los departamentos. Lo anterior permite afirmar que la muestra poblacional de Risaralda tiene una composición genética diferente, a partir de su ADNmt, con respecto a la mayoría de los demás departamentos que conforman la región andina, excepto con Quindío y Caldas (dos departamentos vecinos geográficamente y con los que antes conformaba uno solo denominado “viejo Caldas”), con los que los valores de P de no diferenciación no fueron significativos después de la corrección de Bonferroni aunque estos valores de P estuvieron muy

próximos al valor de significancia y los valores de F_{ST} por parejas mostraron moderada diferenciación genética (124).

También se evidenció que, el departamento de Cundinamarca presentó diferencias significativas con Santander además de las ya mencionadas con Risaralda. Con los demás departamentos, Cundinamarca exhibió valores de P de no diferenciación no significativos, aunque mostró valores de F_{ST} por parejas entre 2,7% y 6,9%, excepto con Antioquia y Boyacá.

Teniendo en cuenta que los resultados mostraron algunas diferencias significativas al comparar los departamentos, se evaluó la posibilidad que estas se debieran a una proporción diferencial de los haplogrupos nativos. Para esto, se realizaron nuevos análisis que incluyeron únicamente las muestras con haplogrupos nativos de cada departamento. Los resultados se presentan en la Tabla 26 y, para una mejor visualización de los valores de F_{ST} obtenidos, se realizó una representación gráfica en MDS (Figura 37).

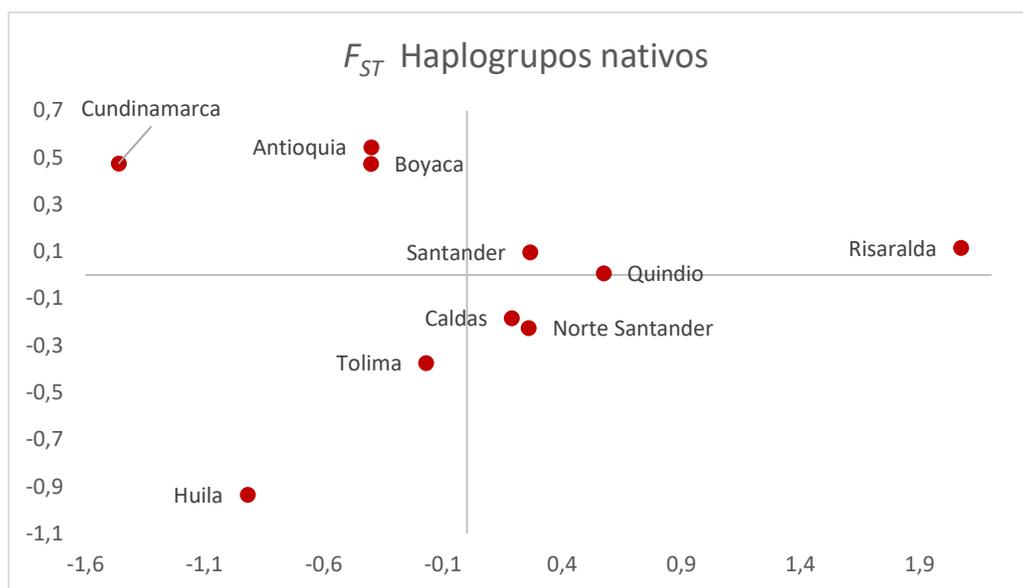
Tabla 26. Valores de F_{ST} por pares para los haplogrupos nativos entre los departamentos de la región andina

Haplogrupos nativos de la Región Control total de ADNmt											
	N	Antioquia	Boyacá	Caldas	Cundinamarca	Huila	Norte de Sder	Quindío	Risaralda	Santander	Tolima
Antioquia	40	*	0,4853	0,2770	0,0827	0,0128	0,0770	0,1009	0,0003	0,0922	0,1385
Boyacá	45	-0,0056	*	0,1597	0,0783	0,0111	0,1990	0,0949	0,0001	0,1988	0,2435
Caldas	43	0,0037	0,0119	*	0,0053	0,0528	0,3719	0,7168	0,0041	0,3116	0,4376
Cundinamarca	48	0,0237	0,0222	0,0642	*	0,0109	0,0022	0,0006	0,0000	0,0006	0,0150
Huila	44	0,0508	0,0472	0,0277	0,0452	*	0,0161	0,0048	0,0000	0,0052	0,1515

Norte de Sder	42	0,0250	0,0085	-0,0002	0,0811	0,0413	*	0,6096	0,0010	0,8142	0,4339
Quindío	46	0,0204	0,0197	-0,0098	0,1001	0,0597	-0,0069	*	0,0219	0,3413	0,1778
Risaralda	50	0,1222	0,1337	0,0660	0,2518	0,1843	0,0741	0,0380	*	0,0000	0,0001
Santander	184	0,0132	0,0050	0,0012	0,0593	0,0388	-0,0071	0,0004	0,0823	*	0,2646
Tolima	49	0,0135	0,0052	-0,0023	0,0417	0,0112	-0,0018	0,0093	0,1130	0,0023	*

Nivel de significancia después de la corrección de Bonferroni: $P=0,0011$. El valor de P se calculó con 50175 permutaciones. Valores de $s.e. \leq 0,0024$. Valores de F_{ST} (abajo de la diagonal) y valores de probabilidad (P) de no diferenciación (arriba de la diagonal). Los valores de P significativos están resaltados en negro.

Figura 37. MDS de los F_{ST} obtenidos para haplogrupos nativos de los departamentos de la región andina.



Los resultados obtenidos en este análisis muestran un comportamiento similar a los anteriormente presentado. Con base en estos hallazgos, se confirma que los niveles de diferenciación entre el departamento de Risaralda y los demás departamentos comparados (excepto

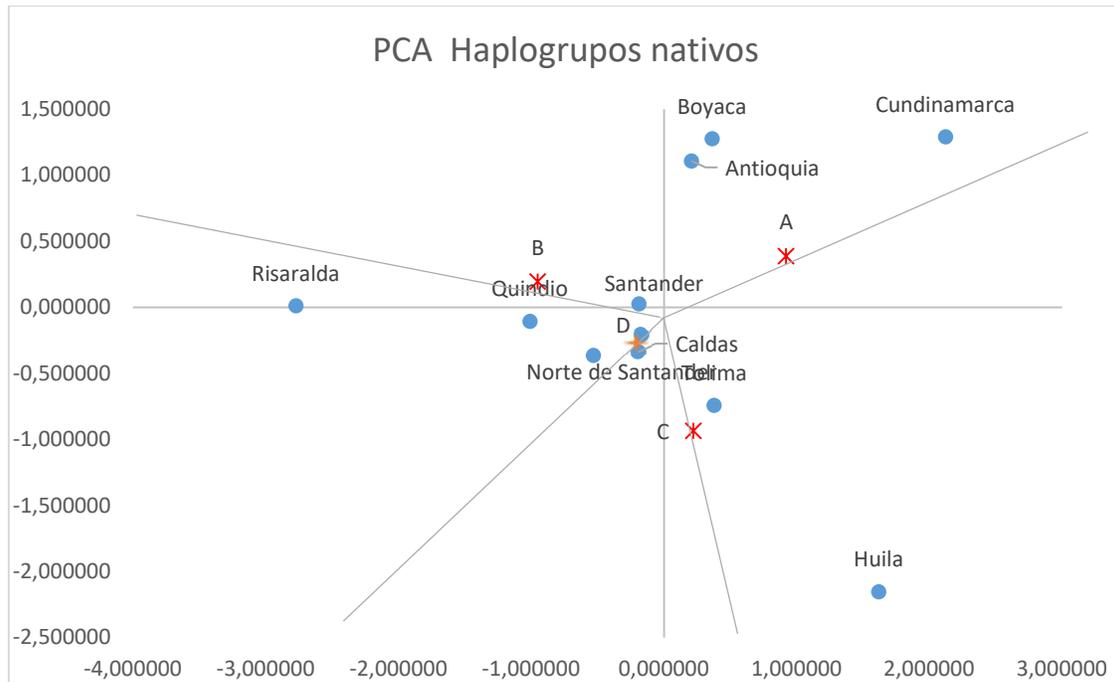
con Quindío), son altos y que éste se estaría comportando como una población distante porque su acervo genético difiere de los demás departamentos andinos.

Adicionalmente, Cundinamarca mostró diferencias significativas ($P < 0,0011$) con el departamento de Quindío, además de las ya observadas con Risaralda y Santander. Teniendo en cuenta que esta comparación fue hecha solo con el componente nativo, es claro que las diferencias halladas entre los departamentos pueden ser atribuidas a contribuciones genéticas nativas diferentes que podrían ser originadas por contribuciones diferenciales de los grupos nativos que las pudieron originar. Así las cosas, Risaralda (Emberas y Quimbayas), Cundinamarca (Muiscas) y Santander (Guanes, Chitareros y Laches).

6.3.5.2 Análisis de componentes principales (PCA) a partir de haplogrupos nativos.

Para interpretar mejor los resultados obtenidos al comparar el componente nativo americano de los departamentos a partir de los valores pareados de F_{ST} , se realizó un PCA basado en las frecuencias de los haplogrupos nativos (Figura 38).

Figura 38. PCA a partir de las frecuencias de haplogrupos nativos en los departamentos de la región andina.



El PCA corroboró que los aportes de los haplogrupos A, B, C y D en cada departamento contribuyeron a la diferenciación entre ellos. De acuerdo con el PCA, el primer componente principal separó a Antioquia, Boyacá, Cundinamarca, Huila y Tolima, de Caldas, Norte de Santander, Santander, Quindío y Risaralda; debido a una mayor frecuencia de los haplogrupos A/C y B/D, respectivamente. El segundo componente separó a Cundinamarca, Boyacá y Antioquia, de Huila y Tolima, debido a que los tres primeros fueron los departamentos con la mayor frecuencia del haplogrupo A (63%, 46% y 45%, respectivamente), mientras que Huila y Tolima, pese a presentar frecuencias altas de haplogrupo A, tuvieron la frecuencia más alta del Haplogrupo C entre todos los departamentos de la región andina (26% y 15%, respectivamente). Adicionalmente, este segundo componente separó a Risaralda y Caldas de Norte de Santander, Santander y Quindío, debido a que Risaralda presentó la frecuencia más alta del haplogrupo B de todos los departamentos (68%) y en Caldas este fue el haplogrupo más frecuente en su población, de los cuatro haplogrupos nativos (36%). Por su parte, Norte de Santander, Santander y Quindío presentaron una frecuencia

alta del haplogrupo B (28%, 31% y 42%, respectivamente), y adicionalmente, tuvieron la frecuencia más alta del haplogrupo D de todos los departamentos de la región andina (18%, 12% y 10%, respectivamente).

Las diferencias halladas entre los departamentos de la región andina, debidas a la disimilitud en los aportes de los cuatro linajes nativos americanos, pueden explicarse por el hecho histórico de que diferentes grupos indígenas habitaron el territorio de los Andes colombianos en la época precolombina. Estos grupos llegaron a Suramérica procedentes de Centroamérica e ingresaron por istmo de Panamá (26,66), por lo cual Colombia fue un paso obligado para todos ellos. Algunos se establecieron en el país y la mayoría de estos grupos nativos escogieron las zonas montañosas de los andes colombianos por el clima y las ventajas que ofrecían sus tierras para la agricultura entre otras razones y conformaron diferentes comunidades nativas que poseían acervos genéticos, lingüísticos y culturales muy disímiles (59,63,65). Esta diversidad genética aportada por los diferentes grupos nativos ha sido corroborada con estudios genéticos realizados a partir de ADNmt en comunidades indígenas aisladas que son descendientes de esas comunidades precolombinas y que viven en la actualidad en el país (33,34,35,36,39,43). Adicionalmente, las condiciones geográficas de la región favorecieron el aislamiento de estas comunidades; como ya se expuso, las comunicaciones solo se mantenían entre áreas geográficas muy cercanas por las dificultades que ofrecía el relieve quebrado para la movilidad (59,63). Este aislamiento continuó en los asentamientos fundados por los conquistadores, donde se dio la mezcla entre mujeres nativas con hombres europeos o afrodescendientes (65,68), práctica que se extendió durante el periodo colonial y republicano hasta finales del siglo XIX (68).

Las diferencias encontradas en las proporciones de los distintos haplogrupos nativos también pueden ser atribuidas al hecho de que algunos departamentos fueron poblados más recientemente por migrantes procedentes de otros departamentos ya existentes, atraídos por las oportunidades económicas que surgían en nuevos asentamientos, como ocurrió con la expansión del cultivo del café hacía el reconocido “viejo Caldas” (Quindío, Risaralda y Caldas), donde llegaron nuevos colonos procedentes principalmente de Antioquia. (125,126). De igual manera, más recientemente los departamentos de la región andina han recibido la contribución de acervos genéticos de migrantes que llegan principalmente a las grandes ciudades procedentes de las áreas rurales, de donde han emigrado las personas huyendo de la violencia (72,73). Por su parte, otras poblaciones ubicadas en las áreas rurales de las provincias que conforman estos departamentos continuaron aisladas y se mantienen así hasta hoy (59,68,70,71).

Algunas de estas características del poblamiento de la región andina podrían explicar los resultados obtenidos para el departamento de Risaralda., el cual presentó diferencias significativas con todos los departamentos, excepto Quindío y Caldas que, cómo se explicó son los más próximos geográficamente a él y con los cuales conformó hasta 1968 un solo departamento (Viejo Caldas) (127), lo que permite inferir que deberían compartir gran parte de su acervo genético, histórico y cultural. De acuerdo con los historiadores, en el actual territorio del departamento de Risaralda ocurrió un proceso de colonización desigual que inició a finales del siglo XIX por terratenientes procedentes de Antioquia. De esta forma este departamento recibió aportes de tres grupos. El primer grupo conformado por campesinos que habitaban ya este territorio, cuyos orígenes se remontaban al occidente de Colombia, descendientes de los grupos indígenas Quimbaya y Emberá que por siglos habitaron allí (126,127). Otro grupo estuvo constituido por los antioqueños provenientes del suroeste que ocuparon los municipios al occidente del río Cauca (127). El tercer

grupo fue conformado por inmigrantes antioqueños provenientes de Salamina y otros distritos del sur del viejo Antioquia y poblaron a Santa Rosa, Pereira y Marsella (127). Este contexto podría contribuir a explicar las diferencias halladas en este departamento con respecto a los restantes de la región andina. Si bien es cierto que la base de la economía cafetera que se impulsó masivamente a inicios del siglo XX en el país trajo pobladores a estos tres departamentos, hoy llamados del “eje cafetero” (128), Risaralda tuvo como fuente diferencial el aporte de campesinos procedentes del occidente colombiano descendientes de los indígenas, sumado al hecho de su característico regionalismo (125,126) que podría haber favorecido los cruces selectivos entre miembros de su población contribuyendo posiblemente con la aparición de apareamientos endógamos que podrían explicar su conformación completamente nativa en la muestra analizada. Otro aspecto que podría explicar las diferencias sería que la colonización desigual ocurrida desde Antioquia trajo migrantes de solo las regiones occidental y sur de ese departamento donde también habitaron los indígenas Emberá (42,126.127).

Por su parte las diferencias halladas en los resultados obtenidos para Cundinamarca con respecto a algunos departamentos podrían ser explicadas en parte, como ya se expuso, por las disimilitudes en los grupos nativos que habitaron en la época precolombina el altiplano cundiboyacense, con respecto de los que habitaron Santander y los del Quindío, y adicionalmente como resultado de los fenómenos migratorios ocurridos en los últimos 200 años hacia la capital del país, centro urbano que concentra aproximadamente el 17% de la población colombiana y que congrega individuos procedentes de diferentes regiones de Colombia que se han establecido allí atraídos por la amplia oferta educativa y las posibilidades de desarrollo laboral que permitan contar con una mejor calidad de vida y cuyo territorio se encuentra dentro de Cundinamarca. (49).

Finalmente, teniendo en cuenta que algunos laboratorios forenses descartan los tractos homopoliméricos de las posiciones 16193, 309, 315, 523-524 y 573 (48), y que uno de los objetivos de este estudio es conformar una base de datos que sea útil para su uso forense, se calcularon nuevamente los valores de F_{ST} por parejas utilizando los haplotipos de la RC de todas las muestras, eliminando los Indel de los tractos homopoliméricos. En este análisis se adicionó una muestra poblacional de Medellín – Colombia, incluida en el proyecto 1000 genomas, consistente de 94 secuencias obtenidas por NGS. Esta es la única población mezclada de Colombia que posee datos publicados de la región control completa (sin los Indel de los tractos homopoliméricos). Los resultados obtenidos se presentan en la Tabla 27.

Tabla 27. Valores de F_{ST} por pares entre los departamentos de la región andina a partir de RC sin Indel.

Haplotipos de la Región Control de ADNmt sin Indel												
	n	Antioquia	Boyacá	Caldas	Cundinam	Huila	Norte de Sder	Quindío	Risaralda	Tolima	Santander	Ant 1000 genomas
Antioquia	49	*	0,5858	0,3127	0,0620	0,0125	0,1260	0,1909	0,0004	0,1861	0,1816	0,7359
Boyacá	52	-0,0054	*	0,1343	0,1082	0,0164	0,2283	0,1474	0,0000	0,2891	0,3096	0,5496
Caldas	50	0,0017	0,0108	*	0,0039	0,0364	0,4531	0,7687	0,0041	0,3197	0,2829	0,3321
Cundinam	54	0,0197	0,0128	0,0522	*	0,0131	0,0025	0,0014	0,0000	0,0246	0,0017	0,0153
Huila	50	0,0381	0,0343	0,0254	0,0361	*	0,0321	0,0077	0,0000	0,2608	0,0100	0,0038
Norte Sder	50	0,0112	0,0050	-0,0021	0,0579	0,0258	*	0,6281	0,0006	0,4271	0,7549	0,1462
Quindío	50	0,0069	0,0102	-0,0084	0,0703	0,0445	-0,0057	*	0,0233	0,2023	0,3809	0,2094
Risaralda	50	0,0970	0,1197	0,0555	0,2154	0,1644	0,0736	0,0364	*	0,0000	0,0000	0,0001
Tolima	53	0,0072	0,0024	0,0016	0,0283	0,0039	-0,0013	0,0069	0,1076	*	0,2732	0,1573

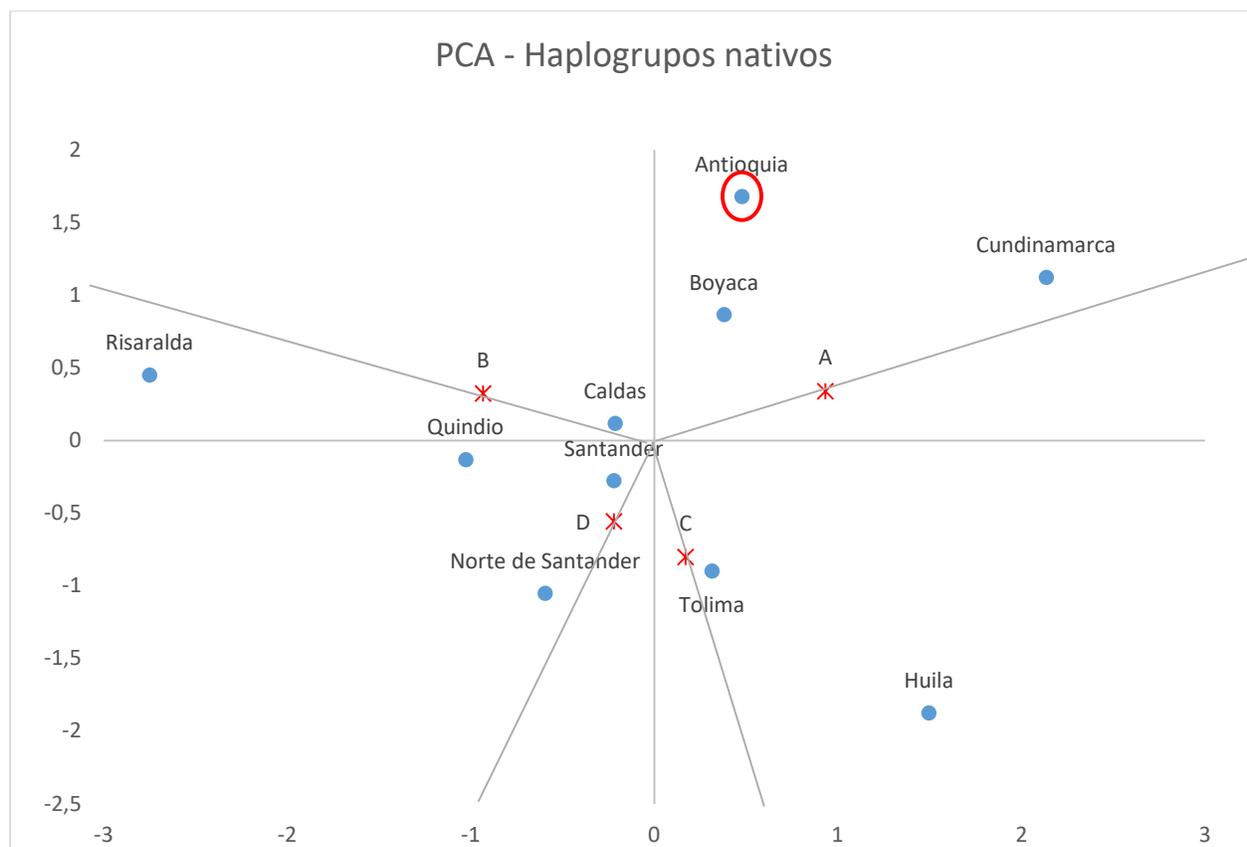
Santander	204	0,0048	0,0010	0,0018	0,0391	0,0263	-0,0049	-0,0003	0,0767	0,0019	*	0,2186
Ant 1000 g	94	-0,0064	-0,0039	0,0007	0,0311	0,0434	0,0079	0,0048	0,0915	0,0072	0,0022	*

Nivel de significancia después de la corrección de Bonferroni: $P=0,0009$. El valor de P se calculó con 50175 permutaciones. Valores de $s.e. \leq 0,0024$. Valores de F_{ST} (abajo de la diagonal) y valores de probabilidad (P) de no diferenciación (arriba de la diagonal). Los valores de P significativos están resaltados en negro.

Los valores de F_{st} y P obtenidos con este nuevo análisis, son muy similares a los hallados cuando se analizaron las secuencias de la región control completa ya presentados (Tabla 27). Adicionalmente, se estableció que la muestra de Medellín (Antioquia 1000) no presenta diferencias significativas con la muestra de Antioquia evaluada en este estudio ($F_{st} = -0,0064$), por lo que podrían unirse en una sola muestra con un $N = 143$.

6.3.5.3 PCA a partir de la frecuencia de haplogrupos integrando Antioquia 10000 genomas. En la Figura 39 se presenta un PCA donde el departamento de Antioquia está representado con las frecuencias de los haplogrupos nativos, calculadas a partir de la muestra conjunta del presente estudio y la analizada en el proyecto 1000 genomas.

Figura 39. PCA a partir de haplogrupos nativos integrando la muestra del proyecto 1000 genomas.



En este PCA se puede observar que Antioquia (con las dos muestras integradas) y señalado en círculo rojo, continúa ubicándose en el cuadrante superior derecho, muy similar a su ubicación en el PCA de la Figura 38, corroborando que el comportamiento de las frecuencias de esta muestra poblacional integrada varía muy poco, porque son muy similares, tal como se observa en los resultados del *Fst* presentados en la Tabla 27.

Los resultados obtenidos del análisis de la estructura genética de los departamentos de la región andina colombiana muestran que existe subestructura en esta región, rechazando la hipótesis planteada inicialmente en este estudio de homogeneidad genética en todos los departamentos de la región andina colombiana. Las diferencias halladas fueron debidas principalmente al aporte

diferencial de haplotipos de ADN mitocondrial pertenecientes a los haplogrupos nativos americanos, en las muestras analizadas.

De acuerdo con los resultados obtenidos y el análisis realizado de las distancias genéticas halladas al comparar mediante F_{ST} por pares los departamentos, se recomienda conformar una base de datos que integre los departamentos de Antioquia, Boyacá, Caldas, Huila, Norte de Santander, Quindío, Santander y Tolima (e incorporar a ella los resultados reportados de Antioquia 1000 genomas), ya que no se evidenció que entre ellos existan diferencias significativas de acuerdo con valores de P de no diferenciación no significativos que fueron obtenidos.

Por otra parte, los datos del departamento de Cundinamarca debido a que presentó diferencias significativas en los valores de P de no diferenciación después de la corrección de Bonferroni, con los de los departamentos de Risaralda, Santander y Quindío deberían ser tratados de manera independiente

De igual forma, los datos del departamento de Risaralda que presentó valores de F_{ST} por pares $>6\%$, con respecto a los demás departamentos con valores P de no diferenciación significativos con la mayoría de ellos, deberían ser usados de forma independiente y por tanto estaría aportando sus datos para la resolución de los casos forenses de Risaralda en particular.

7. Conclusiones

- 7.1 La revisión sistemática de alcance permitió establecer que la mayoría de los estudios científicos publicados del análisis de ADNmt en poblaciones americanas tienen fines poblacionales y muy pocos fines forenses. En su mayoría han estudiado solo alguna de las regiones hipervariables de la región control de la molécula de ADNmt y han utilizado como metodología de genotipado la secuenciación por Sanger. Aplicar esta herramienta permite conocer el estado del arte en forma global, antes de proponer las investigaciones de cualquier tema de interés en el futuro.
- 7.2 Los resultados de este estudio permitieron conocer la composición de la RC completa del ADNmt de muestras de individuos nacidos en los departamentos de la región andina colombiana, datos que no han sido obtenidos ni reportados previamente y que serán de utilidad para los laboratorios forenses del país al momento de resolver casos complejos que requieran del genotipado del ADNmt.
- 7.3 Al evaluar la diversidad genética estimada de los departamentos de la Región Andina se halló que la diversidad a partir de los linajes maternos es alta, en la mayoría de ellos >99%. Al analizar la estructura genética entre los departamentos de la Región Andina, se encontró que hay diferencias estadísticamente significativas entre algunos de ellos, lo que indica que no existe homogeneidad genética; por tanto, se rechaza la hipótesis planteada. Las diferencias halladas podrían ser atribuibles, principalmente, a la distribución diferencial de las frecuencias

de los haplogrupos nativos americanos en los departamentos. Estas diferencias muy probablemente sean consecuencia de la diversidad genética de los grupos nativos que habitaron la región andina en la época precolombina, aunado al aislamiento de algunas regiones al interior de estos departamentos y el favorecimiento de cruces principalmente entre individuos de un mismo departamento, aumentando con ello la representación selectiva de algunos aportes nativos en sus genomas mitocondriales y disminuyendo la diversidad de los mismos en estas poblaciones, a pesar de los movimientos migratorios más recientemente ocurridos en toda la región andina colombiana.

7.4 Los resultados hallados permitieron confirmar que el aporte a los linajes maternos de la población de la región andina colombiana es mayoritariamente nativo americano (>89%), aunque la representación de los haplogrupos nativos americanos es diferente en los departamentos, siendo el haplogrupo A el más frecuente en promedio, seguido del haplogrupo B, el C y finalmente el D. Por su parte el aporte de los haplogrupos no nativos fue del 6% en promedio para los linajes de ascendencia euroasiática y del 4% para los linajes de ascendencia africana.

7.5 La evaluación de la subestructura poblacional, mostró que hay diferencias estadísticamente significativas entre algunos departamentos. Los departamentos que presentaron mayores diferencias genéticas fueron Risaralda y Cundinamarca con valores de $F_{ST} > 6\%$ y P de no diferenciación < 0.0011 , lo que llevó a refutar la hipótesis planteada de homogeneidad genética en todos los departamentos de la Región Andina. Asumir que la región andina colombiana no se comporta como una población única y que se presentan diferencias en la composición genética del ADNmt en algunos departamentos tiene implicaciones importantes en la

conformación de las bases de datos de ADNmt para su uso forense y por ello se recomienda que los datos obtenidos de los departamentos de Antioquia, Boyacá, Caldas, Huila, Norte de Santander, Quindío, Santander y Tolima conformen una sola base de datos que puede ser usada para resolver los casos ocurridos en cualquiera de estos departamentos. Esta similitud puede explicarse por su cercanía geográfica, por los antecedentes de poblamiento por los mismos grupos nativos que habitaron estos departamentos en la época prehispánica y por la organización histórica posterior a la colonia que mantuvo algunos de ellos agrupados inicialmente, por ejemplo, a Quindío y Caldas como parte del viejo Caldas y a Norte de Santander y Santander como el gran Santander. En esta base de datos puede incluirse la población de Antioquia (Medellín) del proyecto 1000 genomas, pues no se hallaron diferencias significativas (F_{ST} -0.00644 y $P= 0.73593$) con respecto a la población de Antioquia evaluada en este estudio. Por otra parte, se recomienda que los datos del departamento de Cundinamarca, así como los del departamento de Risaralda sean utilizados en forma independiente, cada uno de estos debería utilizar su base de datos para resolver sus casos en particular.

8. Recomendaciones

- 8.1 Incorporar a la base de datos diseñada, la población de Antioquia del proyecto 1000 genomas, pues no se hallaron diferencias significativas ($F_{st} = -0.00644$ y $P = 0.73593$), con la población de Antioquia evaluada en este estudio.
- 8.2 Teniendo en cuenta que los resultados obtenidos en este estudio representan un aporte al conocimiento de los haplotipos para la RC completa de ADNmt del país y de su distribución en la población actual mezclada de los departamentos de la región andina, se debería continuar su exploración a través de nuevos estudios que permitan generar un mayor número de datos de la RC del ADNmt de cada departamento, (en especial Risaralda y Cundinamarca) que contribuirán a aumentar las bases de datos para su uso forense y con ello, el poder de discriminación de estas bases de datos.
- 8.3 El alcance de los nuevos estudios debería incluir a las demás regiones del país para aproximarnos a conocer la totalidad de la diversidad genética de nuestras poblaciones, a partir de los linajes de ADNmt que poseemos y establecer las diferencias reales en el aporte de cada uno de los tres grandes grupos continentales en cada una de estas regiones, así como generar bases de datos específicas de los departamentos donde ocurrieron los hechos.

Referencias Bibliográficas

1. Roewer,L. DNA fingerprinting in forensics: past, present, future. *Investig Genet.* 1985: 4:22.
2. Gill P, Jeffreys A, Werrett D. Forensic application of DNA ‘fingerprints’. *Nature.* 1985: 318(6046): 577-79.
3. Mojica L. La prueba técnica ADN en los procesos sobre filiación. *Estudios Socio-Jurídicos.* 2003; 5(1): 250 – 265.
4. Crespillo M. Efectos de la contaminación, transferencia y persistencia del ADN en la interpretación de la prueba genética. *Genética Forense: del laboratorio a los tribunales.* España: Ediciones Díaz de Santos. 2019; 51-77.
5. Prieto L, Carracedo A. El valor de la prueba de ADN. *Genética Forense: del laboratorio a los tribunales.* España: Ediciones Díaz de Santos.2019; 445-70.
6. Luque, J. A. (2019). Estudio de las relaciones de parentesco. *Genética Forense: del laboratorio a los tribunales.* España: Ediciones Díaz de Santos. 351-381.
7. Salzano F, Sans M. Interethnic admixture and the evolution of Latin American population. *Genetics and Molecular Biology.* 2014; 37(1) suppl: 151-170.
8. Castillo A, Hincapie M, Vargas C. et al. Population genetic data for F13A01, FES/FPS, F13B and LPL in Colombia (Department of Santander). *Forensic Science International: Genetics Supplement Series 2.* 2009: 357–358.
9. Castillo A, Gil A, Pico A. et al. Genetic variation for 20 STR loci in a northeast Colombian population (Department of Santander), *Forensic Science International: Genetics Supplement Series 4.* 2013: e298–e299.
10. Ossa H, Aquino J, Pereira R. et al. Outlining the Ancestry Landscape of Colombian Admixed Populations. *Plos One.* 2016; 11(10): 1 – 15.
11. Ossa H, Posada Y, Trujillo N, et al. Patterns of genetic diversity in Colombia for 38 Indel used in human identification. *For Sci Int Genet.* 2021; 53: 102495.

12. Amorin A, et al. Formulation and communication of evaluative forensic science expert opinion – A GHEP-ISFG contribution to the establishment of standards. *For Sci Int Genet.* 2016; 25: 210-213.
13. Castillo A, Gusmao L, Vargas C, et al. Genetic variation of 23 STR loci in a Northeast Colombian population (department of Santander). *Forensic Science International: Genetics Supplement Series* 7. 2019: 33–35.
14. Pereira V, Gusmao L. (2019). Análisis de marcadores STRs del cromosoma Y. En: Crespillo M. *Genética Forense: del laboratorio a los tribunales.* España: Ediciones Díaz de Santos, 217-40.
15. López, M. (2019). Análisis de ADN mitocondrial. *Genética Forense: del laboratorio a los tribunales.* España: Ediciones Díaz de Santos, 241-68.
16. Parson T, et al. A high observed substitution rate in the human mitochondrial DNA control region. *Nature Genetics.* 1997; 15: 363 – 368.
17. Venegas R, Gutierrez E, Caicedo M. Investigaciones y comprensiones del conflicto armado en Colombia. *Salud mental y familia. Universitas Psychologica.* 2017; 16(1): 3-10.
18. Mora-Gómez F. Reconocimiento de víctimas del conflicto armado en Colombia: Sobre tecnologías de representación y configuraciones de Estado. *Universitas Humanística.* 2016; 82: 75-101.
19. Lareu MV, et al. Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing, *Forensic Science International.* 2004; 140(2): 251-257.
20. Presidencia de la República, FARC-EP. (2016). Acuerdo final Gobierno de Colombia FARC-EP para la terminación del conflicto y la construcción de una paz estable y duradera. https://www.cancilleria.gov.co/sites/default/files/Fotos2016/12.11_1.2016nuevoacuerdofinal.pdf
21. Schmitter H, et al. The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results, and perspectives. *Forensic Science International.* 2004; 139(2): 215-226.
22. Parson W, et al. Extended guidelines for mtDNA typing of population data in forensic science. *Forensic Science International: Genetics.* 2007; 1: 13 – 19.

23. Prieto L, et al. The GHEM-EMPOP collaboration on mtDNA population data – A new resource for forensic casework. *Forensic Science International: Genetics*. 2011; 5: 146 – 151.
24. Parson W, et al. DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing- *Forensic Science International: Genetics*. 2014; 13: 134 – 142.
25. Institute TJB. The Joanna Briggs Institute Reviewers: Manual 2015 Methodology for JBI Scoping Reviews. Melbourne. Joanna Briggs Institute Reviewers. South Australia, The University of Adelaide. 2015. 24 p.
26. Horai S, Kondo R, Nakagawa Y, et al. Peopling of the Americas, Founded by Four Major Lineages of Mitochondrial DNA. *Mol. Biol. Evol.* 1993; 10(1): 23-47.
27. Monsalve M, Groot H, Espinel A, et al. Evidence of mitochondrial DNA diversity in South American Aborigines. *Ann. Hum. Genet.* 1994; 58: 265-73.
28. Mesa N, Mondragón C, Soto I, et al. Autosomal, mtDNA, and Y-Chromosome Diversity in Amerinds: Pre- and Post-Columbian Patterns of Gene Flow in South America. *Am. J. Hum. Genet.* 2000; 67: 1277-86.
29. Keyeux G, Rodas C, Gelvez N, et al. Possible Migration Routes into South America Deduced from Mitochondrial DNA Studies in Colombian Amerindian Populations. *Human Biology*. 2002; 74(2): 211-233.
30. Salas A, Pichards M, Lareu MV, et al. Shipwrecks and Founder Effects: Divergent Demographic Histories Reflected in Caribbean mtDNA. *American journal of Physical Anthropology*. 2005; 128: 855–860.
31. Torres M, Bravi C, Bortolini M, et al. A Revertant of the Major Founder Native American Haplogroup C Common in Populations from Northern South America. *American Journal of Human Biology*. 2006; 18: 59–65.
32. Melton P, Briceño I, Gómez A, et al. Biological Relationship Between Central and South American Chibchan Speaking Populations: Evidence From mtDNA. *American journal of Physical Anthropology*. 2007; 133: 753-770.
33. Usme-Romero S, Alonso M, Hernandez-Cuervo H, et al. Genetic differences between Chibcha and Non-Chibcha speaking tribes based on mitochondrial DNA (mtDNA)

- haplogroups from 21 Amerindian tribes from Colombia. *Genetics and Molecular Biology*. 2013; 36(2): 149-157.
34. Xavier C, Builes JJ, Gomes V, et al. Admixture and genetic diversity distribution patterns of non-recombining lineages of native american ancestry in colombian populations. *PLoS One*. 2015; 10: 1–13.
35. Noguera-Santamaría MC, Anderson C, Uricoechea D, et al. Mitochondrial DNA analysis suggests a Chibchan migration into Colombia. *Univ. Sci*. 2015; 20(2): 261-278.
36. Arias L, Barbieri C, Barreto G, et al. High-resolution mitochondrial DNA analysis sheds light on human diversity, cultural interactions, and population mobility in Northwestern Amazonia. *Am J Phys Anthropol*. 2018; 165: 238–255.
37. Martínez B, Simão F, Gomes V, et al. Searching for the roots of the first free African American community. *Sci Rep*. 2020; 10: 1–11.
38. Rodas C, Gelvez N, Keyeux G. Mitochondrial DNA Studies Show Asymmetrical Amerindian Admixture in Afro-Colombian and Mestizo Populations. *Human Biology*. 2003; 75(1):13–30.
39. Rondón F, Orobio R, Braga Y, et al. Estudio de diversidad genética de cuatro poblaciones aisladas del centro y suroccidente colombiano. *Salud UIS*. 2006; 38: 12-20.
40. Salas A, Acosta MA, Alvarez-Iglesias V, et al. The mtDNA Ancestry of Admixed Colombian Populations. *American Journal of Human Biology*. 2008; 20: 584–591.
41. Rondón F, Osorio J, Peña A, et al. Genetic diversity in human populations of two regions from Colombia. *Colombia Medica*. 2008; 39(2): 52-60.
42. Rojas W, Parra MV, Campo O, et al. Genetic Make Up and Structure of Colombian Populations by Means of Uniparental and Biparental DNA Markers. *American journal of Physical Anthropology*. 2010; 143: 13-20.
43. Criollo-Rayó A, Bohórquez M, Prieto R, et al. Native American gene continuity to the modern admixed population from the Colombian Andes: implication for biomedical, population and forensic studies. *Forensic Sci Int Genet*. 2018; 36: e1-e7.
44. Carvajal-Carmona L, Soto I, Pineda N, et al. Strong Amerind/White Sex Bias and a Possible Sephardic Contribution among the Founders of a Population in Northwest Colombia. *Am. J. Hum. Genet*. 2000; 67: 1287-1295.

45. Mora C. Análisis de la variación en las secuencias del ADNmt humano en individuos residentes de la región Cundi-boyacense. *Colombia Forense*. 2009; 1(3): 33-37.
46. Yunis JJ, Yunis E. Mitochondrial DNA (mtDNA) haplogroups in 1526 unrelated individuals, from 11 Departments of Colombia. *Genet Mol Biol*. 2013; 36(3): 329-335.
47. Terreros G. Determinación de la variación de las secuencias de las regiones HVI y HVII de la región control del ADN mitocondrial en una muestra de la población Caribe colombiana. tesis para optar al título de Magister en Biología. Pontificia Universidad Javeriana. 2010. 115 p.
48. Parson W, Dür A. EMPOP-a forensic mtDNA database. *Forensic Sci. Int. Genet*. 2007; 1:88-92.
49. Departamento Administrativo Nacional de Estadística (DANE), Censo Nacional de Población y Vivienda 2018. <http://www.dane.gov.co/censo> (acceso 10/01/23).
50. Sistema de Información Red de Desaparecidos y Cadáveres (SIRDEC). <https://sirdec.medicinalegal.gov.co:38181/mapaDesaparecidos/> (acceso 10/01/23).
51. Martínez, MB. et al. La prueba de ADN en medicina forense. Primera edición. Madrid. Editorial Masson S.A. 1999.
52. Carracedo A. La genética forense y sus aplicaciones en investigación criminal, *Revista de Occidente*. 2013; 391:124-126.
53. Carvalho Gontijo C, Porrás-Hurtado G. L, Freire-Aradas A, et al. PIMA: A population informative multiplex for the Americas. *For Sci Int: Genetics*, 2020; 44:101-108.
54. Urquhart A, Kimpton CR, Gill P. Variation in Short Tandem Repeats sequences: a survey of twelve microsatellites loci for use as forensic identification markers. *Int J Legal Med*. 1994; 107:13-20.
55. Pifarré A. Marcadores STR autosómicos de interés forense: amplificación y artefactos. En: Crespillo M. *Genética Forense: del laboratorio a los tribunales*. España: Ediciones Díaz de Santos; 2019. 161-88.
56. Builes JJ, et al. Y-chromosome STRs in an Antioquian (Colombia) population sample. *Journal of Forensic Science International*. 2005; 208-215.

57. Castillo A, Rueda A, Pico A, et al. Comparing different population groups in Santander – Colombia through Y-STR haplotype análisis. *Forensic Science International: Genetics Supplement Series* 5. 2015; e482-e483.
58. Dür A, Huber N, Parson W. Fine-Tuning Phylogenetic Alignment and Haplogrouping of mtDNA Sequences; *Int J Mol Sci.* 2021; 22(11):5747.
59. Guhl A. La geografía de Colombia como actor histórico. En: *Colombia Preguntas y respuestas sobre su pasado y presente* Ediciones Uniandes. 1ra edición, 2010; 7-25.
60. Sarmiento B, Merchán C. *Colombia, historia y personajes*. Prolibros Tercera edición. 2003. 176 p.
61. Ministerio de Ciencia y tecnología, https://minciencias.gov.co/sala_de_prensa/colombia-el-segundo-pais-mas-biodiverso-del-mundo/.
62. Marquez G. *Mapas de un fracaso: naturaleza y conflicto en Colombia*. Bogotá: Universidad Nacional de Colombia e Instituto de estudios ambientales (IDEA), 2004.
63. Langebaek C. ¿Cuántos eran? ¿Dónde estaban? ¿Qué les pasó? Poblamiento indígena en la Colombia prehispánica y su transformación después de la conquista. En: *Colombia Preguntas y respuestas sobre su pasado y presente* Ediciones Uniandes. 1ra edición, 2010; 27-47.
64. Berdugo L, De León A, Pernet J, et al. *Ciencias sociales regionalizada: Región Andina*. Mundo Latino SA. 1ra edición. 1995. Bogotá D.C. 200 p.
65. Jaramillo J. La población indígena de Colombia en el momento de la conquista y sus transformaciones posteriores en: *Ensayos sobre la historia de Colombia*. Universidad Nacional de Colombia. Bogotá. Imprenta Nacional. 1968. 89-148.
66. Reichel-Dolmatoff G. Colombia indígena, periodo prehispánico. En: *Nueva historia de Colombia*. Ediciones Planeta colombiana SA. Bogotá. Vol 1. 3ra edición, 1984; 27-68.
67. Friede J. Colombia indígena, periodo prehispánico. En: *Nueva historia de Colombia*. Ediciones Planeta colombiana SA. Bogotá. Vol 1. 3ra edición, 1984; 69-116.
68. Palacios J. La esclavitud y la sociedad esclavista. En: *Nueva historia de Colombia*. Ediciones Planeta colombiana SA. Bogotá. Vol 1. 3ra edición, 1984; 153-174.
69. Pita R. Vestigios de la lengua guane: una aproximación al fenómeno del mestizaje idiomático en Santander. *Lingüística y Lit.* 2013; 63: 295–316.

70. Martínez A, Guerrero A. La provincia de Soto: orígenes de sus poblamientos urbanos. 1ra ed. Publicaciones Universidad. Industrial de Santander. Bucaramanga. 1995; 226 p.
71. Martínez A. Historia Básica de Bucaramanga, cuatro siglos de un poblamiento 1622 – 2022. 1st ed Publicaciones Universidad. Industrial de Santander. Bucaramanga. 2022.
72. Bonnett D, La Rosa M, Nieto M. Colombia preguntas y respuestas sobre su pasado y su presente. 1ra edición. U. de los Andes. 2010; 366 p.
73. Perea CM. Colonización, ciudadanía en armas y narcotráfico. La violencia en Colombia, siglo XX. En: Colombia Preguntas y respuestas sobre su pasado y presente Ediciones Uniandes. 1ra edición, 2010; 301-338.
74. Molano A. Trochas y fusiles, Instituto de estudios políticos y relaciones internacionales (Iepri) y Ancora. Universidad de Texas (USA) 1994; 230 p.
75. Boletín de víctimas del conflicto armado, Unidad de servicio público de empleo, Bogotá, 9 de abril de 2021.
76. LEY 1408 DE 2010. Por la cual se rinde homenaje a las víctimas del delito de desaparición forzada y se dictan medidas para su localización e identificación. Congreso de Colombia. Bogotá agosto 20 de 2010.
77. Gill P, Jeffreys AJ, Werrett DJ. Forensic application of DNA ‘fingerprints’. Nature. 1985; 318 (6046) 577-79.
78. Huguet F. Parámetros estadísticos en genética forense. Investigación de la paternidad. En: Martínez MB. La prueba de ADN en medicina forense. España: Editorial Masson S.A.; 1999. 273-76.
79. García O. Parámetros en Identificación forense y Criminalística Biológica. En: Martínez MB. La prueba de ADN en medicina forense. España: Editorial Masson S.A.; 1999. 286-300.
80. Weber J, May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet. 1989; 44: 388-96.
81. Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Ac. Res. 1989; 17(16):6463-71.

82. Ballantyne KN, Goedbloed M, Fang R, et al. Mutability of Y- chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *American journal of human genetics*. 2010; 87(3):341-53.
83. Stoneking M. Hypervariable sites in the mtDNA control region are mutational hotspots. *Am J Hum Genet*. 2000; 113(1-3):97-101.
84. Jobling M.A, Tyler Smith C. The human Y chromosome: an evolutionary marker comes of ages. *Nature Reviews Genetics*. 2003; 4(8):598-612.
85. Anderson S, et al. Sequence and organization of the human mitochondrial genome. *Nature*. 1981;290: 457 – 465.
86. Andrews R, Kubacka I, Chinnery P, et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet*. 1999; 23:147.
87. Robin E.D, Wong R. Mitochondrial DNA molecules and virtual number of mitochondrial per cell in mammalian cells. *J Cell Physiol*. 1988; 136(3):507-13.
88. Calloway C. D, et al. The frequency of heteroplasm in the HVII region of mtDNA differs across tissue types and increases with age. *Am J Hum Genet*. 2000; 66(4):1387-97.
89. Lutz S, et al. Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions? *Forensic Sci Int*. 2000; 113(1-3):97-101.
90. Wallace D, et al. Sequence analysis of cDNAs for the human and bovine ATP synthase b subunit: mitochondrial DNA genes sustain seventeen times more mutations. *Curr Genet*. 1987; 12:81 – 90.
91. Barrio P, et al. Nuevas tecnologías de Secuenciación Masiva en Paralelo (MPS) de uso en genética forense. En: Crespillo M. *Genética Forense: del laboratorio a los tribunales*. España: Ediciones Díaz de Santos; 2019. 319-350.
92. Mardis E. R. Next Generation Sequencing Platforms, *Annu Rev Anal Chem*, 2013; 6:287-303.
93. Zhang W, Cui H, Wong L. Comprehensive One-Step Molecular Analyses of Mitochondrial Genome by Massively Parallel Sequencing. *Clinical Chemistry*. 2012; 58(9):1322-31.
94. Huber N, Parson W, Dür A. Next generation database search algorithm for forensic mitogenome analyses. *Forensic Sci Int Gen*. 2018; 37: 204-214.

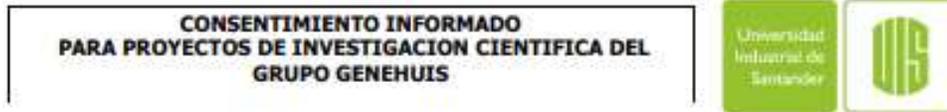
95. de Knij P. From next generation sequencing to now generation sequencing in forensics. *Forensic Sci Int Gen.* 2019; 38: 175-180.
96. Cann RL, Stoneking M, Wilson AC. Mitochondrial DNA and human evolution. *Nature.* 1987; 325: 31-36.
97. van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat.* 2009; 30(2): e386–e394. <http://www.phylotree.org>.
98. Salas A, Martín R, Lareu MV, et al. The African diaspora: Mitochondrial DNA and the Atlantic slave trade. *Am J Hum Genet.* 2004; 74: 454-465.
99. Tricco A, Lillie E, Zarin W, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Annals of Internal Medicine.* 2018; 1-20.
100. Colquhoun, HL. Levac, D. O'Brien, KK. et al. (2014). Scoping reviews: time for clarity in definition, methods, and reporting. *J Clin Epidemiol*, 67(129), 1-4.
101. Tricco AC, Lillie E, Zarin W, et al. A scoping review on the conduct and reporting of scoping reviews. *BMC Med Res Methodol.* 2016; 16: 1-10.
102. Canadian Institutes of Health Research. (2010). A guide to knowledge synthesis: a knowledge synthesis chapter. 2010. Accessed at www.cihr-irsc.gc.ca/e/41382.html on 10 January 2022.
103. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol.* 2005; 8(1): 19-32.
104. Fregel R, Delgado S. HaploSearch: A tool for haplotype-sequence two-way transformation. *Mitochondrion.* 2011; 11: 366–367.
105. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 2010; 10: 564–567.
106. Carracedo A, Bar W, Lincoln P, et al. DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing. *Forensic Sci Int.* 2000; 110(2): 79-85.

107. Scientific Working Group on DNA Analysis Methods (SWGDAM). Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories. 2013. Disponible en <https://www.swgdam.org/publications>.
108. Wallace D, Garrison K, Knowler W. Dramatic Founder Effects in Amerindian Mitochondrial DNAs. *American journal of physical anthropology*. 1985; 68: 149-155.
109. Irwin J, Saunier J, Niederstatter H, et al. Investigation of Heteroplasmy in the Human Mitochondrial DNA Control Region: A Synthesis of Observations from More Than 5000 Global Population Samples. *J Mol Evol*. 2009; 68:516–527.
110. Gómez-Carballa A, Moreno F, Álvarez-Iglesias V, et al. Revealing latitudinal patterns of mitochondrial DNA diversity in Chileans. *Forensic Sci Int Genet*. 2016; 20:81–88.
111. Simão F, Ribeiro J, Vullo C, et al. The ancestry of eastern Paraguay: A typical south american profile with a unique pattern of admixture. *Genes (Basel)*. 2021;12:1-17.
112. Simão F, Ferreira AP, de Carvalho EF, et al. Defining mtDNA origins and population stratification in Rio de Janeiro. *Forensic Sci Int Genet*. 2018;34:97–104.
113. dos Reis RS, Simão F, dos Santos Stange V, et al. A view of the maternal inheritance of Espírito Santo populations: The contrast between the admixed and Pomeranian descent groups. *Forensic Sci Int Genet*. 2019;40:175–181.
114. Palencia L, Valverde L, Álvarez A, et al. Mitochondrial DNA diversity in a population from Santa Catarina (Brazil): Predominance of the European input. *Int J Legal Med*. 2010;124:331–115.
115. Poletto M, Malaghini M, Silva JS, et al. Mitochondrial DNA control region diversity in a population from Parana state—increasing the Brazilian forensic database. *Int J Legal Med*. 2018; 133:347-351.
116. Freitas J, Fassio L, Braganholi D, et al. Mitochondrial DNA control region haplotypes and haplogroup diversity in a sample from Brasília, Federal District, Brazil. *Forensic Sci Int Genet*. 2019; 40:e228-e230.
117. Castro de Guerra D, Figuera C, Bravi CM, et al. Sequence variation of mitochondrial DNA control region in North Central Venezuela. *Forensic Sci Int Genet*. 2012;6:e131-3.

118. Burgos G, Simão F, Flores-Espinoza R, et al. An approach to maternal ancestry in a sample of Ecuadorian “mestizo” population by sequencing the control region of mtDNA. *Forensic Sci Int Genet. Supplement Series* 7. 2019;537-538.
119. Bobillo MC, Zimmermann B, Sala A, et al. Amerindian mitochondrial DNA haplogroups predominate in the population of Argentina: towards a first nationwide forensic mitochondrial DNA sequence database. *Int J Legal Med*. 2010; 124:263-268.
120. Slatkin M, Hudson RR. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*. 1991; 129:555–562.
121. Marjoram P, Donnelly P. Pairwise Comparisons of Mitochondrial DNA Sequences in Subdivided Populations and Implications for Early Human Evolution. *Genetics*. 1994; 136: 673-683.
122. Baeta M, Núñez C, Cardoso Sosa C, et al. Different Evolutionary History for Basque Diaspora Populations in USA and Argentina Unveiled by Mitochondrial DNA Analysis. *PLoS ONE*. 2015; 10(12):e0144919.
123. Friedemann NS. *La saga del negro: Presencia africana en Colombia*. 1. ed. Instituto de Genética Humana, Facultad de Medicina, Pontificia Universidad Javeriana; 1993.
124. Hartl D, Clark A. *Principles of population genetics*, Third edition, Sinauer Associates USA, 1997; 118-119.
125. Vélez F. La colonización y Fundación de Risaralda (Caldas). *Academia caldense de historia*. 2011;20-36.
126. Rivera JA. “Del período precolombino al mito fundacional de Pereira: cien siglos de historia previa”. En: *Revista Virajes*, 2014; Vol. 16, No. 2. Manizales: Universidad de Caldas.
127. Botero CA, Mayorga O. Vista del poblamiento y colonización de Risaralda, *Revista Páginas Universidad Católica de Pereira*. Núm. 14 (Feb - Abr, 1986).
128. Ocampo JH. Los orígenes de la industria cafetera, 1830 – 1929. En: *Nueva historia de Colombia*. Ediciones Planeta colombiana SA. Bogotá. Vol 5. 3ra edición, 1984; 213-232.

Apéndices

Apéndice A. Consentimiento informado participantes de la región andina



Versión 02

Código: _____

Fecha: _____

Proyecto "Identificación de linajes maternos presentes en individuos nacidos en la región Andina colombiana. Una aproximación a su uso en genética forense"

Estimado usuario, lo invitamos a participar en este proyecto que será desarrollado por investigadores del grupo de investigación en Genética Humana de la Universidad Industrial de Santander (GENEHUIS) para determinar la composición, diversidad y grado de diferenciación genética en una muestra de individuos nacidos en la región Andina colombiana, a partir de información presente en el ADN mitocondrial, estos resultados contribuirán a implementar el análisis de ADN mitocondrial en los laboratorios de genética forense del país y de esta forma hacer más expedita la identificación de personas. Los resultados de este trabajo son de gran importancia para nuestro país, pues permitirán conocer los componentes genéticos de origen materno, que caracterizan nuestra población y utilizarlos en la resolución de casos forenses complejos.

Por favor lea con cuidado y haga las preguntas que desee hasta su total comprensión. Si usted acepta participar le será tomada una muestra de sangre con una aguja y un tubo al vacío nuevo y estéril, este procedimiento es riesgo mínimo para usted y con su muestra realizaremos la extracción del material genético llamado Acido desoxirribonucleico (ADN). Usted como participante recibirá el resultado del análisis de su ADN mitocondrial denominado haplotipo y de acuerdo con las variaciones encontradas conocerá a que haplogrupo pertenece, es decir de donde proviene ancestralmente su ADN mitocondrial.

Su muestra será identificada con un código único al procesarla, lo que permitirá el manejo y análisis de los resultados sin incluir su información personal, tal como su nombre, documento de identidad etc., garantizando de esta manera la privacidad y confidencialidad de su información. La muestra tomada solo se utilizará para los fines propuestos, quedará almacenada bajo el código que se le asignará y se mantendrá bajo cadena de custodia. Si usted acepta participar tendrá la libertad de retirar su consentimiento y salir del estudio en el momento que lo desee y por tanto retirar la información que haya aportado a la investigación.

21/5/21

Su participación en este estudio no tiene ningún costo, estos serán asumidos por la investigación científica de la Facultad de Salud UIS. Los resultados solo se utilizarán para el propósito de los proyectos de investigación propuestos por los investigadores de GENEHUIS, y parte de los análisis pueden llevarse a cabo en otros laboratorios de investigación ubicados en las universidades o instituciones de investigación nacionales o internacionales con los que la UIS tenga convenios de cooperación académica y científica. Sus resultados podrán ser presentados en eventos científicos, podrán ser incluidos en publicaciones científicas o incluidos en la base de datos del Laboratorio de Genética humana de la UIS o en otra base de datos que pueda ser consultada por laboratorios de genética forense, asignando códigos aleatorios para ser incluidos.

El proyecto al cual aporta la información recolectada tiene aprobación del comité de ética de la UIS (CEINCI).

Habiendo comprendido lo anterior y una vez que se le aclararon todas las dudas que surgieron con respecto a su participación, ¿usted acepta participar?

Autoriza _____

No autoriza _____

Nombre del participante

Firma de autorización

Lugar de nacimiento: _____

Edad: _____

El grupo de investigadores de GENEHUIS continuará estudiando la población Colombiana, a partir del análisis de marcadores genéticos uni y biparentales con alta variabilidad, para establecer la diversidad genética de las personas nacidas en nuestro país y conocer el componente ancestral que cada uno posee de acuerdo con el/los grupo/s étnicos poblacionales de los que proviene; por ello le invitamos a permitir el uso de la muestra que le tomaremos, en futuros proyectos de investigación que el grupo desarrolle analizando estos marcadores genéticos.

Autoriza _____

No autoriza _____



21/5/21

**CONSENTIMIENTO INFORMADO
PARA PROYECTOS DE INVESTIGACION CIENTIFICA DEL
GRUPO GENEHUIS**



Nombre del testigo 1 _____ Firma _____

Dirección _____ Tel: _____

Nombre del testigo 2 _____ Firma _____

Dirección _____ Tel: _____

Nombre del Investigador principal o
persona que toma la muestra _____ Firma _____

Adriana Castillo Pico

21/5/21

Para cualquier información relacionada con este proyecto usted puede comunicarse con Adriana Castillo Pico, investigadora principal del proyecto, a los teléfonos del Laboratorio de Genética UIS. Directo: 6356641 ó PBX: 6344000 Ext. 3117 ó 3131.

Adicionalmente usted se puede comunicar al Comité de Ética en Investigación Científica de la Universidad Industrial de Santander, ubicado en la sede Bucarica UIS, Carrera 19 No. 35 – 02, centro, oficina 245, teléfono 6344000 ext. 3808, correo comitedetica@uis.edu.co, en el horario de lunes a viernes de 7 a.m. a 12:00 m y de 2:00 pm a 5:00 pm, quien vigilará el cumplimiento de los principios éticos involucrados en esta investigación.

Apéndice B. Consentimiento informado participantes de Santander

**CONSENTIMIENTO INFORMADO
PARA LA UTILIZACION DE MUESTRAS DE USUARIOS
DEL LABORATORIO DE GENETICA UIS
EN PROYECTOS DE INVESTIGACION CIENTIFICA**



Versión 01

Código: _____

De acuerdo con los principios establecidos en la Resolución 008430 de 4 de Octubre de 1993 por la cual se establecen las normas para la investigación en salud, específicamente en el Artículo 15, en lo relacionado con el Consentimiento Informado, usted deberá conocer acerca de esta investigación y aceptar participar en ella si lo considera conveniente. Por favor lea con cuidado y haga las preguntas que desee hasta su total comprensión.

El grupo de investigación en Genética Humana de la Universidad Industrial de Santander se encuentra realizando el estudio de la población nacida en Colombia, a partir del análisis de marcadores genéticos con alta variabilidad, para establecer la diversidad genética de las personas nacidas en nuestro país y conocer el componente ancestral que cada uno posee de acuerdo con el/los grupo/s étnicos poblacionales de los que proviene.

Si usted acepta participar le será tomada una gota de sangre con una lanceta nueva y estéril para ser depositada en un papel de filtro estéril. Este procedimiento no tiene ningún riesgo o complicación para usted. Con su muestra realizaremos la extracción del material genético llamado Acido desoxirribonucleico (ADN) y determinaremos su perfil genético para los marcadores poblacionales analizados. Los resultados de este trabajo son de gran importancia para nuestro país pues permitirán conocer los componentes genéticos que caracterizan nuestra población y que permiten diferenciarnos de otras y darnos autenticidad. Estos resultados quedarán incluidos en la base de datos del Laboratorio de Genética que es una base de datos privada, no es pública. Solo se utilizará para el propósito de los proyectos de investigación propuestos por los investigadores del laboratorio y sus resultados podrán ser presentados en eventos científicos y/o ser incluidos en una publicación científica.

Si usted tiene alguna pregunta sobre la toma de muestra o las características de las investigaciones, por favor infórmenos para poderle brindar una explicación más detallada y si lo desea puede retirarse del estudio en cualquier momento sin que por ello se creen prejuicios para usted.

Su muestra será identificada con un código para asegurar que no se conocerá su información privada, tal como su nombre, documento de identidad etc., y se mantendrá la confidencialidad total de sus datos personales en caso de realizar la publicación científica de los resultados. La muestra tomada solo se utilizará para los fines propuestos, quedará almacenada bajo el código que se la adjudicará a toda su información y se mantendrá bajo cadena de custodia. No se usará para crear líneas celulares permanentes ni inmortalizadas. Los datos que sean tomados en la ficha de recolección diligenciada para cada usuario, serán manejados de acuerdo a lo estipulado en el Título III-Categorías de datos sensibles, artículo 5° y 6° numeral e) de la Ley estatutaria 1581 del 2012 que reglamentó el artículo 15 de la Constitución Política, en la que se establece que -los datos relativos a la salud se consideraran sensibles y se prohíbe su tratamiento excepto cuando el tratamiento de los datos tenga una finalidad histórica, estadística o científica, así como lo estipulados en el decreto 1377 de 2013 y la resolución de Rectoría No. 1227 de agosto 22 de 2013.

UNIDAD DE INVESTIGACION EN GENETICA HUMANA
 COMITÉ DE ÉTICA
 11 OCT 2018

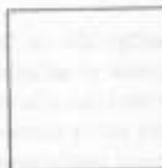

Con tal fin a cada ficha de recolección de datos correspondiente a cada usuario se le asignará un número de código único, con el cual se identificará la muestra a procesar, lo que permitirá el manejo y análisis de los resultados de cada individuo sin incluir su información personal, garantizando de esta manera la privacidad y confidencialidad de la información.

Su participación en este estudio no tiene ningún costo, estos serán asumidos por la investigación Científica de la Facultad de Salud de la UIS para realizar dichos estudios.

Con fecha _____, habiendo comprendido lo anterior y una vez que se le aclararon todas las dudas que surgieron con respecto a su participación, usted acepta participar?

Autoriza _____

No autoriza _____



Nombre del participante _____

Firma de autorización _____

Nombre del testigo 1 _____

Firma _____

Dirección _____

Tel: _____

Relación que guarda con el participante _____

Fecha de la firma _____

Nombre del testigo 2 _____

Firma _____

Dirección _____

Tel: _____

Relación que guarda con el participante _____

Fecha de la firma _____

**CONSENTIMIENTO INFORMADO
PARA LA UTILIZACION DE MUESTRAS DE USUARIOS
DEL LABORATORIO DE GENETICA UIS
EN PROYECTOS DE INVESTIGACION CIENTIFICA**



Nombre del Investigador principal

Firma

Para cualquier información relacionada con este proyecto usted puede comunicarse al Laboratorio de Genética UIS. Teléfonos: Directo; 6356641 PBX: 6344000 Ext. 3117 ó 3131

Adicionalmente usted se puede comunicar al Comité de Ética en Investigación Científica de la Universidad Industrial de Santander, ubicado en la Sede Bucarica UIS, Carrera 19 N° 35 – 02, centro, oficina 245, teléfono 6344000 ext. 3808, correo: comitedetica@uis.edu.co, en el horario de lunes a viernes de 7:00 am a 12:00 m y de 2:00 pm a 5:00 pm, quien vigilará el cumplimiento de los principios éticos involucrados en esta investigación.



COMITÉ DE ÉTICA

OCT 2019

Apéndice C. Ponencia en el 29th Congress of the International Society for Forensic Genetics.

Sample collection strategies when building mitochondrial DNA forensic databases

Filipa Simão¹, Adriana Castillo², Germán Burgos^{1,4}, Leonor Gismondi³.

¹DNA Diagnostic Laboratory, State University of Rio de Janeiro, Rio de Janeiro, Brasil
²Genetics Laboratory, Industrial University of Santander, Bucaramanga, Colombia
³School of Medicine, Faculty of Health Sciences, University of the Americas (UDLA), Quito, Ecuador
⁴Group of Genomic Medicine, University of Santiago de Compostela, Santiago

IBRAG
 UFRJ - Universidade Federal do Rio de Janeiro

P121

Financial support:
 CNPq
 FAPERJ

Background

The current genetic diversity in South America is mainly attributed to the coexistence of native groups from different ethnolinguistic backgrounds during the pre-colonial period, and by the subsequent arrival of European settlers and African slaves from the 16th century on. The more recent immigration from Europe and Asia, and the constant movement of individuals between and within countries is also influencing the genetic composition of the populations from South America.

The high mtDNA haplotype diversity reported in South America highlights the importance of building highly representative forensic databases, that capture the existing variability. Many authors choose to restrict the sample collection to individuals with residence and proven inheritance for, at least, three generations in a specific geographic location, which not necessarily represent the reference database in most forensic scenarios, especially in populations subject to recent migrations. In these circumstances, sample collection strategy becomes a crucial factor in the construction of adequate databases.

AIM

Considering the different patterns of admixture across South America, and the constant migratory fluxes of individuals, to and within the subcontinent: AIM: evaluate to what extent the sample strategies impact on capturing the existing diversity in South America. Accordingly, the maternal genetic background of admixed populations from Paraguay, Ecuador and Colombia was studied, to evaluate if differences have occurred on the genetic composition of these populations during the last three generations.

Results and discussion

- Haplotype diversity and exclusion power did not change significantly over the generations established.
- In some departments/regions, differences between diversities across the generations were higher than 2% (values highlighted in red).
- Lower diversities were detected for some departments/regions when only the individuals with three generations at a specific birthplace were considered (values highlighted in yellow).

Table 1. Values of Haplotype Diversity (H) and exclusion power (mCE) obtained for three generations in several departments of Paraguay, Colombia, and Ecuador.

Department	Paraguay (n=10)					Colombia (n=10)					Ecuador (n=10)				
	Gen 1	Gen 2	Gen 3	Gen 4	Gen 5	Gen 1	Gen 2	Gen 3	Gen 4	Gen 5	Gen 1	Gen 2	Gen 3	Gen 4	Gen 5
Asunción	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
Bogotá	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
Medellín	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
Quito	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
Guayaquil	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
Manizales	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
Neiva	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
Soledad	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
Valledupar	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
Yopal	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

- The AMOVA results showed no variation on the mtDNA haplotype and haplogroup composition between generations.

Final remarks

The absence of statistically significant differences among the subsets from the same department/region can be due to:

- Homogeneity among generations;
- High diversity of mtDNA / small sample size;
- Gene flow among populations that are not significantly different.

Haplotype frequency databases built to disclose population history are often used for forensic purposes. However, such databases do not always represent the current genetic diversity of the population, but the fraction of those individuals with three-generation heritage in a given geographic region. Therefore, to assess the extent to which databases built with different purposes can be interchangeable, it is crucial to consider demographic and historical data that may point to genetic differences between generations. This aspect is particularly important in the construction of databases of South American populations, which have complex population dynamics, with different levels of recent immigration and of isolation.

Materials and Methods

Sampling:

Region	Department	City	Year	n	Gen 1	Gen 2	Gen 3	Gen 4	Gen 5
Paraguay	Asunción	Asunción	2018	10	1	2	3	4	5
					6	7	8	9	10
					11	12	13	14	15
					16	17	18	19	20
					21	22	23	24	25
Colombia	Bogotá	Bogotá	2018	10	1	2	3	4	5
					6	7	8	9	10
					11	12	13	14	15
					16	17	18	19	20
					21	22	23	24	25
Ecuador	Quito	Quito	2018	10	1	2	3	4	5
					6	7	8	9	10
					11	12	13	14	15
					16	17	18	19	20
					21	22	23	24	25

mtDNA typing: Haplotypes from Paraguay were retrieved from Simão et al. [1]. Samples from Ecuador and Colombia were amplified, sequenced, and analyzed for total control region with the same methodologies as in Simão et al. [1].

Statistical analysis: Haplotype diversity (H) and Analysis of molecular variance (AMOVA) were obtained with Arlequin software [2]. The exclusion power (mCE) was calculated according to Simão et al. [3].

References:

[1] Simão F, Ribeiro J, Valle C, et al. The ancestry of eastern Paraguay: A typical south american profile with a unique pattern of admixture. *Genet (Basel)*. 2021;12. [2] Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetic data analysis. *Evol Bioinform Online*. 2009;1:47-50. [3] Simão F, Ferreira AR, de Carvalho EF, et al. Defining mtDNA origins and population stratification in Rio de Janeiro. *Forensic Sci Int Genet*. 2018;34:97-104.

Apéndice D. Artículo “Sample collection strategies when building mitochondrial DNA forensic databases”

Filipa Simão, Adriana Castillo, Germán Burgos, Leonor Gusmão. Sample collection strategies when building mitochondrial DNA forensic databases. *Forensic Science International: Genetics Supplement Series*, Volume 8, 2022, Pages 91-96, <https://doi.org/10.1016/j.fsigss.2022.09.033>.

Apéndice E. Ponencia en el LV congreso nacional & VI internacional de ciencias biológicas

ASOCIACIÓN COLOMBIANA DE CIENCIAS BIOLÓGICAS
LV CONGRESO NACIONAL & VI INTERNACIONAL DE CIENCIAS BIOLÓGICAS
VIRTUAL

Revisión sistemática de la literatura científica para mapear las aplicaciones del análisis de ADN mitocondrial en poblaciones americanas

Autores: Adriana Castillo Pico^{1*}, Leidy Johanna Rueda Díaz², Fernando Rondón González³

Afiliación: ¹Departamento de Ciencias Básicas, Escuela de Medicina, Universidad Industrial de Santander, Bucaramanga – Colombia.

²Escuela de Enfermería, Universidad Industrial de Santander, Bucaramanga – Colombia.

³Escuela de Biología, Universidad Industrial de Santander, Bucaramanga - Colombia.

*e-mail: ccastillo@uis.edu.co

Abstract: (50 palabras) (no estructurado)

With the aim to map the applications, methodologies and haplotypes of mitochondrial DNA in American population, a systematic scoping review was conducted, following the guidelines of the Joanna Briggs Institute. It is concluded that most of the studies are population-based and very few are forensic.

KeyWords: mitochondrial DNA, Haplotypes, Americas, Systematic Review.

Resumen: (50 palabras)

Con el objetivo de mapear las aplicaciones, metodologías y haplotipos del ADN mitocondrial en poblaciones americanas, se realizó una revisión sistemática de alcance siguiendo los lineamientos del Instituto Joanna Briggs. Se concluye que la mayoría de los estudios tienen fines poblacionales, y muy pocos, tienen fines forenses.

Palabras claves: ADN mitocondrial, Haplotipos, América, Revisión Sistemática.

Introducción: (100 palabras)

El uso del ADN mitocondrial en genética forense implica genotipar perfiles genéticos y valorar estadísticamente los resultados [1]. Lo anterior, demanda conocer la distribución de haplotipos en la población donde se analiza el caso [2]. Para esto, es esencial tener información de los estudios realizados hasta la fecha para establecer la valoración de las frecuencias haplotípicas en la población de interés. Las revisiones sistemáticas de alcance constituyen una herramienta para recopilar evidencia disponible al respecto [3]. El objetivo de esta revisión fue examinar la literatura científica existente para mapear las aplicaciones, metodologías y haplotipos del ADNmt en poblaciones Americanas.

Materiales y métodos: (150 palabras)

Esta revisión de alcance se desarrolló según la metodología del Instituto Johanna Briggs. Fueron incluidos estudios poblacionales y forenses desarrollados a partir de ADNmt

humano de individuos nacidos en América. Los estudios debían proporcionar información sobre la aplicación del ADNmt, metodologías utilizadas para su análisis molecular y los métodos de análisis de datos empleados. Las búsquedas de literatura fueron realizadas entre mayo y septiembre de 2019 en el portal Pubmed, las bases de datos Scopus, Science direct, Web of Science; Academic Search ultimate, Fuente académica premier, Health Source – Consumer edition, CINAHL, OVID, BVS y Scielo; los repositorios de tesis o disertaciones CAPES, Dart Europe, tesis doctorales en red y Ramata, y el buscador Google Académico. Las búsquedas y la extracción de datos fueron realizadas por dos revisores de manera independiente. Para la extracción de datos se empleó un formulario elaborado por los autores.

Resultados y discusión: (200 palabras)

Esta revisión identificó inicialmente 4538 estudios, de estos quedaron 2724 una vez se eliminaron los duplicados, estos se revisaron por título y resumen para establecer el cumplimiento de criterios de inclusión y se seleccionaron 364 estudios los cuales fueron examinados en la etapa de lectura del texto completo para determinar el cumplimiento de los criterios de participantes, concepto y contenido. En total, 249 estudios fueron elegibles para ser incluidos en la revisión. De acuerdo con el tipo de estudio, el 93% tuvieron aplicación con fines poblacionales y el 7% con fines forenses, de igual manera el 96% de los estudios analizaron la región control del ADNmt y el 4% analizaron la molécula en su totalidad (Mitogenoma). Así mismo el 50% de los estudios usó secuenciación como metodología de análisis molecular, el 24% RFLPs, el 20% utilizó simultáneamente RFLP's y Secuenciación, un 4% usó Secuenciación de Próxima generación (NGS) y el 4% restante uso minisequenciación. Finalmente para el análisis de datos el 100% de los estudios presentó las frecuencias haplotípicas y de estos un 51% incluyó el análisis de diversidad genética, un 44% presentó adicionalmente el análisis de redes de haplotipos y el 11% presentó los parámetros estadísticos de importancia forense.

Noviembre 20-24 de 2020. Virtual.

www.asociacioncolombianadecienciasbiologicas.org
e-mail:scbb55congresoovirtual@gmail.com

TL: +57 3125883543

ASOCIACIÓN COLOMBIANA DE CIENCIAS BIOLÓGICAS
LV CONGRESO NACIONAL & VI INTERNACIONAL DE CIENCIAS BIOLÓGICAS
VIRTUAL

Conclusión: (50 palabras)

Esta revisión sistemática dilucidó que la mayoría de los estudios tienen fines poblacionales y muy pocos fines forenses, la región control del ADNmt es la más analizada usando la metodología de secuenciación por Sanger y todos los estudios presentan las frecuencias haplotípicas y a partir de ellas, la diversidad genética.

Financiación y agradecimientos:

Los autores agradecen a la Vicerrectoría de Investigación y Extensión por el apoyo del proyecto VIE-URE 2488: "Caracterización de los linajes maternos, basados en ADN mitocondrial, presentes en una muestra poblacional del departamento de Santander", con el cual se adelantó la presente contribución.

Referencias: (solo 3)

[1] Crespiño M, Barrio P. (2019) Genética Forense: del laboratorio a los tribunales. Ediciones Díaz de Santos. 10:241-267.

[2] Parson W. y cols. (2014) DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing. Forensic Science International: Genetics. 13:134-142.

[3] The Joanna Briggs Institute. (2015) Methodology for JBI Scoping Reviews. 1-24.

Seleccione el área:

- Ciencias Biomédicas
- Ciencias Genómicas y Biotecnología
- Ciencias Agropecuarias
- Ciencias Ambientales y desarrollo sostenible
- Biología Animal
- Biología Vegetal
- Enseñanza de las Ciencias Biológicas

Modalidad: Oral Póster _____

Trabajo de: Pregrado _____ Postgrado Otro _____

Presenta trabajo completo a premios Si _____ No
 (ver instructivos extensos para premios)

Apéndice F. Ponencia en el 29th Congress of the International Society for Forensic Genetics.



mtDNA genetic diversity in the population from Santander department (Colombia)

A. Castillo¹, F. Simão², F. Rondón¹, E. F. de Carvalho², L. Gusmão²

¹Genetics Laboratory, Industrial University of Santander (UIS), Bucaramanga, Colombia.
²DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil.



Corresponding author: Adriana Castillo, Tel: 57-7-6356641 email: castillo@uis.edu.co

INTRODUCTION

Santander, one of the 32 departments of Colombia, is located in the Andean region of the country and includes seven provinces: Comunera, García Rovira, Guantotá, Metropolitana (which includes the capital Bucaramanga), Soto Norte, Vélez and Yariquíes (Figure 1).



Fig. 1. Department of Santander with its provinces highlighted in different colors. The municipalities are represented by black dots and the capital of each province is represented by black stars.

The Santander population was shaped by intercontinental admixture between autochthonous Native Americans, European settlers, and African slaves [1]. The admixture took place essentially between European men and Native women [2].

The population of Santander has been previously characterized for autosomal and sex-chromosomal markers of forensic relevance [3, 4, 5].

However, the genetic composition and the distribution of mtDNA haplotypes in Santander is still unknown, and mtDNA databases for forensic purposes are not available. To establish such databases, it is crucial to investigate the genetic diversity and to evaluate population substructure, which is frequent in admixed populations.

This study presents preliminary results of total control region mtDNA haplotypes of 204 individuals born in the seven provinces across the department [1]. Therefore, the objective of this study was to obtain an overview on the maternal lineage composition and diversity of the Santander population.

MATERIAL AND METHODS

Blood samples were collected from 204 unrelated individuals born in Santander department (Colombia). Samples were selected to represent the seven provinces of Santander: Comunera (n=30), García Rovira (n=30), Guantotá (n=28), Metropolitana (n=36), Soto Norte (n=26), Vélez (n=23) and Yariquíes (n=31). Informed Consent, including the details of the study, was signed by the participants prior to donating the sample. This study was approved by the Ethics Committee of the Industrial University of Santander (21/05/21).

DNA was extracted using PureLink™ Genomic DNA Mini Kit (ThermoFisher Scientific). Samples were amplified for the CR using the Qiagen® multiplex PCR kit, 1-5 ng DNA and the primers L15900 and H839 (0.2 μM) in a final volume of 5 μl. PCR products were purified with ExoSAP. Sequencing was performed with the BigDye™ Terminator v3.1 Cycle Sequencing kit, with the primers L15900, L16555, H159 and H839. Sequencing products were purified with Sephadex. Separation and detection was performed in an GA 3500 (Applied Biosystems). Sequences were compared to the rCRS using the software SeqScape v2.7. Haplogrouping was performed via EMPOP, according to Phylotree build 17. Haplogroup frequencies were calculated by direct counting and haplotype diversities were calculated using the Arlequin software v3.5.

REFERENCE 5

[1] Ossa et al., *PLoS ONE* 11 (2016) e0164414; [2] Garrica et al., *Ed. UIS* (1995); [3] Castillo et al., *Forensic Sci. Genetics Supplement Series* 7 (2019) 35-39; [4] Bana et al., *PLoS ONE*, 9 (2014), e017202; [5] Castillo et al., *Forensic Sci. Genetics Supplement Series* 5 (2015) e482-e483; [6] Simão et al., *Forensic Sci. Int. Genet.* 34 (2018) 97-104; [7] Simão et al., *Genes (Basel)*, 2021,12; [8] Gómez-Carballa et al., *Forensic Sci. Int. Genet.* (2016) 20:81-89; [9] Castro de Guerra et al., *Forensic Sci. Int. Genet.* 30 (2019) 68-72; [10] Burgos et al., *Forensic Sci. Genetics Supplement Series* 7 (2019) 537-539.

RESULTS AND DISCUSSION

The Native maternal heritage is highly preserved in the genetic background of the seven provinces across the department. Most samples have native American haplogroups (90%) classified inside macrohaplogroups A2, B4, C1, and D (Table 1). The input of non-native maternal lineages in Santander was low, corresponding to only 10% of the samples. The 12 haplotypes detected with European ancestry (8%) belong to branches inside macrohaplogroup R that can be found at high frequencies across Europe. The African ancestry (4%) was represented by eight haplotypes classified inside eight different L-branches (L0-L3).

Haplogroup	n	%	Haplogroup	n	%
A2	18	8.8	L1	1	0.5
A2a	18	8.8	L2	1	0.5
A2b	0	0	L3	1	0.5
A2c	0	0	L4	1	0.5
A2d	0	0	L5	1	0.5
A2e	0	0	L6	1	0.5
A2f	0	0	L7	1	0.5
A2g	0	0	L8	1	0.5
A2h	0	0	L9	1	0.5
A2i	0	0	L10	1	0.5
A2j	0	0	L11	1	0.5
A2k	0	0	L12	1	0.5
A2l	0	0	L13	1	0.5
A2m	0	0	L14	1	0.5
A2n	0	0	L15	1	0.5
A2o	0	0	L16	1	0.5
A2p	0	0	L17	1	0.5
A2q	0	0	L18	1	0.5
A2r	0	0	L19	1	0.5
A2s	0	0	L20	1	0.5
A2t	0	0	L21	1	0.5
A2u	0	0	L22	1	0.5
A2v	0	0	L23	1	0.5
A2w	0	0	L24	1	0.5
A2x	0	0	L25	1	0.5
A2y	0	0	L26	1	0.5
A2z	0	0	L27	1	0.5
A2aa	0	0	L28	1	0.5
A2ab	0	0	L29	1	0.5
A2ac	0	0	L30	1	0.5
A2ad	0	0	L31	1	0.5
A2ae	0	0	L32	1	0.5
A2af	0	0	L33	1	0.5
A2ag	0	0	L34	1	0.5
A2ah	0	0	L35	1	0.5
A2ai	0	0	L36	1	0.5
A2aj	0	0	L37	1	0.5
A2ak	0	0	L38	1	0.5
A2al	0	0	L39	1	0.5
A2am	0	0	L40	1	0.5
A2an	0	0	L41	1	0.5
A2ao	0	0	L42	1	0.5
A2ap	0	0	L43	1	0.5
A2aq	0	0	L44	1	0.5
A2ar	0	0	L45	1	0.5
A2as	0	0	L46	1	0.5
A2at	0	0	L47	1	0.5
A2au	0	0	L48	1	0.5
A2av	0	0	L49	1	0.5
A2aw	0	0	L50	1	0.5
A2ax	0	0	L51	1	0.5
A2ay	0	0	L52	1	0.5
A2az	0	0	L53	1	0.5
A2ba	0	0	L54	1	0.5
A2bb	0	0	L55	1	0.5
A2bc	0	0	L56	1	0.5
A2bd	0	0	L57	1	0.5
A2be	0	0	L58	1	0.5
A2bf	0	0	L59	1	0.5
A2bg	0	0	L60	1	0.5
A2bh	0	0	L61	1	0.5
A2bi	0	0	L62	1	0.5
A2bj	0	0	L63	1	0.5
A2bk	0	0	L64	1	0.5
A2bl	0	0	L65	1	0.5
A2bm	0	0	L66	1	0.5
A2bn	0	0	L67	1	0.5
A2bo	0	0	L68	1	0.5
A2bp	0	0	L69	1	0.5
A2bq	0	0	L70	1	0.5
A2br	0	0	L71	1	0.5
A2bs	0	0	L72	1	0.5
A2bt	0	0	L73	1	0.5
A2bu	0	0	L74	1	0.5
A2bv	0	0	L75	1	0.5
A2bw	0	0	L76	1	0.5
A2bx	0	0	L77	1	0.5
A2by	0	0	L78	1	0.5
A2bz	0	0	L79	1	0.5
A2ca	0	0	L80	1	0.5
A2cb	0	0	L81	1	0.5
A2cc	0	0	L82	1	0.5
A2cd	0	0	L83	1	0.5
A2ce	0	0	L84	1	0.5
A2cf	0	0	L85	1	0.5
A2cg	0	0	L86	1	0.5
A2ch	0	0	L87	1	0.5
A2ci	0	0	L88	1	0.5
A2cj	0	0	L89	1	0.5
A2ck	0	0	L90	1	0.5
A2cl	0	0	L91	1	0.5
A2cm	0	0	L92	1	0.5
A2cn	0	0	L93	1	0.5
A2co	0	0	L94	1	0.5
A2cp	0	0	L95	1	0.5
A2cq	0	0	L96	1	0.5
A2cr	0	0	L97	1	0.5
A2cs	0	0	L98	1	0.5
A2ct	0	0	L99	1	0.5
A2cu	0	0	L100	1	0.5
A2cv	0	0	L101	1	0.5
A2cw	0	0	L102	1	0.5
A2cx	0	0	L103	1	0.5
A2cy	0	0	L104	1	0.5
A2cz	0	0	L105	1	0.5
A2da	0	0	L106	1	0.5
A2db	0	0	L107	1	0.5
A2dc	0	0	L108	1	0.5
A2dd	0	0	L109	1	0.5
A2de	0	0	L110	1	0.5
A2df	0	0	L111	1	0.5
A2dg	0	0	L112	1	0.5
A2dh	0	0	L113	1	0.5
A2di	0	0	L114	1	0.5
A2dj	0	0	L115	1	0.5
A2dk	0	0	L116	1	0.5
A2dl	0	0	L117	1	0.5
A2dm	0	0	L118	1	0.5
A2dn	0	0	L119	1	0.5
A2do	0	0	L120	1	0.5
A2dp	0	0	L121	1	0.5
A2dq	0	0	L122	1	0.5
A2dr	0	0	L123	1	0.5
A2ds	0	0	L124	1	0.5
A2dt	0	0	L125	1	0.5
A2du	0	0	L126	1	0.5
A2dv	0	0	L127	1	0.5
A2dw	0	0	L128	1	0.5
A2dx	0	0	L129	1	0.5
A2dy	0	0	L130	1	0.5
A2dz	0	0	L131	1	0.5
A2ea	0	0	L132	1	0.5
A2eb	0	0	L133	1	0.5
A2ec	0	0	L134	1	0.5
A2ed	0	0	L135	1	0.5
A2ee	0	0	L136	1	0.5
A2ef	0	0	L137	1	0.5
A2eg	0	0	L138	1	0.5
A2eh	0	0	L139	1	0.5
A2ei	0	0	L140	1	0.5
A2ej	0	0	L141	1	0.5
A2ek	0	0	L142	1	0.5
A2el	0	0	L143	1	0.5
A2em	0	0	L144	1	0.5
A2en	0	0	L145	1	0.5
A2eo	0	0	L146	1	0.5
A2ep	0	0	L147	1	0.5
A2eq	0	0	L148	1	0.5
A2er	0	0	L149	1	0.5
A2es	0	0	L150	1	0.5
A2et	0	0	L151	1	0.5
A2eu	0	0	L152	1	0.5
A2ev	0	0	L153	1	0.5
A2ew	0	0	L154	1	0.5
A2ex	0	0	L155	1	0.5
A2ey	0	0	L156	1	0.5
A2ez	0	0	L157	1	0.5
A2fa	0	0	L158	1	0.5
A2fb	0	0	L159	1	0.5
A2fc	0	0	L160	1	0.5
A2fd	0	0	L161	1	0.5
A2fe	0	0	L162	1	0.5
A2ff	0	0	L163	1	0.5
A2fg	0	0	L164	1	0.5
A2fh	0	0	L165	1	0.5
A2fi	0	0	L166	1	0.5
A2fj	0	0	L167	1	0.5
A2fk	0	0	L168	1	0.5
A2fl	0	0	L169	1	0.5
A2fm	0	0	L170	1	0.5
A2fn	0	0	L171	1	0.5
A2fo	0	0	L172	1	0.5
A2fp	0	0	L173	1	0.5
A2fq	0	0	L174	1	0.5
A2fr	0	0	L175	1	0.5
A2fs	0	0	L176	1	0.5
A2ft	0	0	L177	1	0.5
A2fu	0	0	L178	1	0.5
A2fv	0	0	L179	1	0.5
A2fw	0	0	L180	1	0.5
A2fx	0	0	L181	1	0.5
A2fy	0	0	L182	1	0.5
A2fz	0	0	L183	1	0.5
A2ga	0	0	L184	1	0.5
A2gb	0	0	L185	1	0.5
A2gc	0	0	L186	1	0.5
A2gd	0	0	L187	1	0.5
A2ge	0	0	L188	1	0.5
A2gf	0	0	L189	1	0.5
A2gg	0	0	L190	1	0.5
A2gh	0	0	L191	1	0.5
A2gi	0	0	L192	1	0.5
A2gj	0	0	L193	1	0.5
A2gk	0	0	L194	1	0.5
A2gl	0	0	L195	1	0.5
A2gm	0	0	L196	1	0.5
A2gn	0	0	L197	1	0.5
A2go	0	0	L198	1	0.5
A2gp	0	0	L199	1	0.5
A2gq	0	0	L200	1	0.5
A2gr	0	0	L201	1	0.5
A2gs	0	0	L202	1	0.5
A2gt	0	0	L203	1	0.5
A2gu	0	0	L204	1	0.5
A2gv	0	0	L205	1	0.5
A2gw	0	0	L206	1	0.5
A2gx	0	0	L207	1	0.5
A2gy	0	0	L208	1	0.5
A2gz	0	0	L209	1	0.5
A2ha	0	0	L210	1	0.5
A2hb	0	0	L211	1	0.5
A2hc	0	0	L212	1	0.5
A2hd	0	0	L213	1	0.5
A2he	0	0	L214	1	0.5
A2hf	0	0	L215	1	0.5
A2hg	0	0	L216	1	0.5
A2hh	0	0	L217	1	0.5
A2hi	0	0	L218	1	0.5
A2hj	0	0	L219	1	0.5

Apéndice G.Certificado EMPOP QC EMP00862 (Santander, 204 datos)

Institute of Legal Medicine

Medical University of Innsbruck
Director: Prof. Richard Schweithauer, MD
Muelisenstraße 44, A-6020 Innsbruck, Austria
Tel: +43/51210003-70000, Fax: +43/51210003-70000
e-mail: groa@igi.ti.ac.at



Prof. Dr. Walther Parson
Head of High Throughput DNA Database Unit
Head of Forensic Genomics
walther.parson@ti-med.ac.at

Innsbruck, 02.06.2022

EMPOP QC Report

Dear submitters,

thank you for providing mtDNA data for quality control (QC) to EMPOP!

The haplotypes were evaluated with the EMPOP QC software suite and further manually scrutinized. The final version of the data was established based on your feedback and our interpretation of the data that you kindly provided.

1. Dataset

origin of samples: Santander department, Colombia
metapopulation: Native American, Admixed
contact: Adriana Castillo
no. of samples: 204 (all used)
mtDNA segment: 16024-576
accession number: EMP00862
format: rCRS-coded haplotypes

2. General information

The submitted mitotypes were generated by Sanger Sequencing covering the entire Control Region following the updated ISFG guidelines where applicable (Parson et al 2014).

3. Observations

No observations were made during QC, which is a rare phenomenon. Congratulations!

Institute of Legal Medicine, Medical University of Innsbruck

Haplogrouping

In the course of the EMPDP QC process the haplotypes were turned into fasta-like nucleotide strings, phylogenetically aligned and haplogrouped according to the EMPDP phylogeny (SAM2, Dür et al 2021). There were no discrepancies.

If you want to make use of this information, please cite

Dür A, Huber N, Parson W (2021). Fine-Tuning Phylogenetic Alignment and Haplogrouping of mtDNA Sequences; *Int J Mol Sci* **22**(11):5747; doi: 10.3390/ijms22115747.

4. Decision

The EMPDP QC results indicate that the mitotypes are **acceptable** in the forensic genetic context.

The accession number of this dataset is **EMP00862**. Please use this number for publication of the data and refer to the QC process by citing the following reference:

Parson W and Dür A (2007) EMPDP—a forensic mtDNA database. *Forensic Sci Int Genet* **1**(2): 88-92.

Best wishes,



Prof. Dr. Walther Parson
On behalf of the EMPDP Team

References

1. Parson W, et al. DNA Commission of the International Society for Forensic Genetics: revised and extended guidelines for mitochondrial DNA typing. *Forensic Sci Int Genet*, 2014, **13**: 134-142.

Disclaimer: The applied quality control cannot be regarded as comprehensive independent evaluation of the submitted dataset that would be required in a forensic setting, but constitutes an optimized procedure for the detection of common data idiosyncrasies. The signatories cannot be made liable for correctness, completeness and topicality of the contents.

Apéndice H. Carta de aceptación de artículo: Maternal ancestry and lineages diversity of the Santander population from Colombia

em Forensic Sciences Research Adriana Castillo | Logout

Home Main Menu Submit a Manuscript About Help

← Submissions with an Editorial Office Decision for Author

Page: 1 of 1 (1 total completed submissions) Results per page 10

Action	Manuscript Number	Title	Initial Date Submitted	Status Date	Current Status	Date Final Disposition Set	Final Disposition
View Submission Author Status Author Response View Decision Letter Correspondence Send E-mail	TFSR-2022-0084	Maternal ancestry and lineages diversity of the Santander population from Colombia	Jun 30, 2022	Feb 12, 2023	Accept		

Page: 1 of 1 (1 total completed submissions) Results per page 10

editorialmanager.com/tfsr/ViewLetter.aspx?id=76415&lsid={111BB637-4D2E-42CC-8903-22AB7EF91756}

View Letter

Close

Date: Feb 12, 2023
To: "Adriana Castillo" castillo@uis.edu.co
From: "Forensic Sciences Research" fsr@ssfjd.cn
Subject: 228379291 (Forensic Sciences Research) Your submission has been accepted

Feb 12, 2023

Ref.: Ms. No. TFSR-2022-0084R1
 228379291
 Maternal ancestry and lineages diversity of the Santander population from Colombia
 Forensic Sciences Research

Dear Adriana Castillo,

I am pleased to tell you that your work has now been accepted for publication in Forensic Sciences Research.

It was accepted on Feb 12, 2023

Thank you for submitting your work to this journal.

With kind regards

Shiyong Li
 Section Editor
 Forensic Sciences Research

In compliance with data protection regulations, you may request that we remove your personal registration details at any time. (Use the following URL: <https://www.editorialmanager.com/tfsr/login.asp?a=r>). Please contact the publication office if you have any questions.

A017	B2d	16183C	16189C	16217C	16519C	73G	207A	263G	309.1C	309.2C	315.1C	498del	499A								
A018	A2q	16111T	16131C	16209C	16223T	16290T	16311C	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del		
A019	B2d	16183C	16189C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A	524.1A	524.2C								
A020	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	133C	146C	153G	235G	263G	309.1C	315.1C	523del	524del			
A021	M1a1i	16129A	16182C	16183C	16189C	16223T	16249C	16311C	16359C	16519C	16527T	73G	195C	204C	315.1C	489C					
A022	H1bd	16519C	204C	263G	315.1C																
A023	B2	16182C	16183C	16189C	16217C	16278T	16519C	73G	263G	309.1C	315.1C	498del	499A								
A024	B4b	16182C	16183C	16189C	16217C	16519C	73G	152C	215G	263G	309.1C	309.2C	315.1C	499A							
A025	A2+(64)	16104T	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del				
A026	B2d	16183C	16189C	16217C	16256T	16519C	73G	263G	309.1C	315.1C	498del	499A	523del	524del							
A027	B4b	16086C	16182C	16183C	16189C	16217C	16519C	73G	146C	152C	263G	309.1C	309.2C	315.1C	499A						
A028	A2+(64)+16129	16111T	16129A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del				
A029	B2d	16182C	16183C	16189C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A								
A030	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A								
A031	H1bd	16519C	204C	263G	315.1C																
A032	L2c	16093C	16223T	16278T	16390A	16519C	73G	93G	95C	103A	146C	150T	152C	182T	195C	198T	263G	315.1C	325T	523del	524del
A033	A2+(64)+16189	16111T	16183C	16189C	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del			
A034	L3d1d	16124C	16194T	16223T	16256T	16324C	16368C	73G	152C	263G	309.1C	315.1C	523del	524del							
A035	L2a5	16223T	16224C	16278T	16295T	16390A	16519C	73G	151T	152C	182T	195C	263G	309.1C	315.1C	373G	511T	537T			
A036	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del					

B008	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	152C	153G	214G	235G	263G	309.1C	315.1C	523del	524del		
B009	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A									
B010	B2d	16183C	16189C	16193.1C	16217C	16324C	16519C	73G	263G	309.1C	315.1C	498del	499A								
B011	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	309.3C	315.1C	498del	499A							
B012	B2d	16129A	16183C	16189C	16193.1C	16217C	16519C	73G	234G	263G	309.1C	315.1C	498del	499A							
B013	L2a1	16183M	16189C	16193.1C	16223T	16230G	16278T	16294T	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C				
B014	T1a	16126C	16163G	16186T	16189C	16193del	16294T	16519C	73G	203C	263G	309.1C	315.1C								
B015	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	73G	146C	153G	234G	235G	263G	309.1C	309.2C	315.1C	523del	524del			
B016	A2al	16092C	16111T	16126C	16223T	16290T	16319A	16362C	16518A	64T	73G	146C	153G	235G	263G	315.1C	523del	524del			
B017	A2+(64)	16111T	16223T	16290T	16319A	16362C	16519C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del				
B018	B2d	16172C	16183C	16189C	16217C	16519C	73G	195Y	263G	309.1C	315.1C	498del	499A								
B019	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	151Y	153G	235G	263G	309.1C	315.1C	523del	524del				
B020	B2d	16183C	16189C	16193.1C	16217C	16319A	16519C	73G	153G	210G	263G	309.1C	309.2C	309.3C	315.1C	498del	499A				
B021	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	207A	263G	309.1C	315.1C	498del	499A								
B022	D1f	16142T	16223T	16325C	16362C	16519C	73G	146C	152C	263G	309.1C	315.1C	489C	524.1A	524.2C	573.1C					
B023	A2+(64)	16111T	16189Y	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del				
B024	A2+(64)	16111T	16223T	16240T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del				
B025	D1f	16142T	16147T	16223T	16325C	16362C	16519C	73G	195C	263G	309.1C	315.1C	489C	524.1A	524.2C						
B026	A2+(64)	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del						
B027	A2+(64)	16092C	16111T	16223T	16260T	16290T	16319A	64T	73G	146C	152C	153G	182T	235G	263G	309.1C	309.2C	315.1C	316A	523del	524del

B028	A2+(64)	16092C	16111T	16223T	16260T	16290T	16319A	64T	73G	146C	152C	153G	182T	235G	263G	309.1C	309.2C	315.1C	316A	523del	524del
B029	B2d	16183C	16189C	16193.1C	16217C	16356C	16519C	73G	263G	309.1C	315.1C	498del	499A								
B030	D4h3a	16187T	16203G	16223T	16241G	16245T	16301T	16319A	16342C	16362C	16465Y	16468C	16533C	73G	152C	263G	309.1C	315.1C	489C		
B031	A2+(64)+@153	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	235G	263G	309.1C	315.1C	523del	524del						
B032	A2+(64)	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del					
B033	A2+(64)	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del					
B034	M	16129A	16223T	16274A	16325C	16362C	73G	263G	309.1C	315.1C	489C	573.1C	573.2C	573.3C	573.4C	573.5C	573.6C				
B035	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A									
B036	D4h3a	16187T	16203G	16223T	16241G	16245T	16270T	16301T	16319A	16342C	16362C	16468C	16533C	73G	152C	263G	309.1C	315.1C	489C		
B037	B2d	16183C	16189C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A										
B038	A2+(64)	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del						
B039	C1b	16189A	16223T	16291T	16298C	16325C	16327T	16519C	73G	152C	234G	249del	263G	290del	291del	315.1C	489C	493G	523del	524del	
B040	A2+(64)	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del						
B041	A2+(64)	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del						
B042	B2d	16183C	16189C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A										
B043	A2+(64)	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del						
B044	A2+(64)	16111T	16223T	16290T	16319A	16362C	16519C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del			
B045	D1	16189C	16223T	16272G	16325C	16362C	73G	143A	210G	263G	309.1C	315.1C	489C								
B046	B2d	16183C	16189C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A									
B047	A2+(64)	16111T	16189C	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del				

CU006	A2y	16111T	16258G	16290T	16293G	16304C	16319A	16362C	64T	73G	146C	152C	153G	189G	204C	207A	235G	263G	309.1C	315.1C	523del	524del
CU007	C1b	16189C	16193.1C	16193.2C	16223T	16298C	16311C	16325C	16327T	73G	249del	263G	290del	291del	309.1C	309.2C	315.1C	489C	493G	523del	524del	
CU008	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del					
CU009	A2+(64)	16111T	16290T	16319A	16335G	16356C	16362C	64T	73G	146C	152Y	153G	234G	235G	263G	309.1C	315.1C	523del	524del			
CU010	C1	16183C	16185T	16189C	16193del	16223T	16298C	16325C	16327T	73G	204C	249del	263G	282C	290del	291del	309.1C	315.1C	489C	493G	523del	524del
CU011	A2+(64)+16129	16111T	16129A	16223T	16256T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del			
CU012	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del					
CU013	A2+(64)	16223T	16290T	16319A	16362C	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del								
CU014	A2ac	16111T	16213A	16223T	16256T	16290T	16319A	16362C	64T	73G	146C	153G	182T	235G	263G	309.1C	315.1C	523del	524del			
CU015	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	182T	235G	263G	309.1C	309.2C	315.1C	523del	524del			
CU016	A2+(64)	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del						
CU017	A2+(64)+@153	16111T	16195C	16223T	16290T	16293G	16319A	16362C	16519C	64T	73G	146C	189G	199C	206C	235G	263G	309.1C	315.1C	523del	524del	
CU018	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	182T	235G	263G	309.1C	315.1C	523del	524del				
CU019	A2+(64)	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del							
CU020	B4b	16086C	16182C	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	315.1C	499A									
CU021	A2+(64)	16111T	16223T	16287T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del					
CU022	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del					
CU023	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	146C	263G	309.1C	315.1C	498del	499A									
CU024	U5b2b3	16224C	16270T	16362C	16519C	73G	150T	263G	309.1C	309.2C	315.1C	517T										
CU025	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del						

CU026	A2	16111T	16189C	16223T	16266T	16290T	16319A	16356C	16362C	16519C	73G	146C	153G	195C	235G	263G	315.1C	523del	524del	
CU027	A2+(64)	16111T	16223T	16290T	16319A	16362C	16519C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del		
CU028	B2d	16129A	16183C	16189C	16217C	16519C	73G	195C	263G	315.1C	498del	499A								
CU029	A2r1	16111T	16223T	16261T	16290T	16311C	16319A	16362C	73G	103A	114T	146C	153G	235G	263G	309.1C	315.1C	523del	524del	
CU030	D1	16223T	16254G	16325C	16362C	73G	89.1A	92A	100.1A	106del	107del	108del	109del	110del	111del	263G	309.1C	309.2C	315.1C	489C
CU031	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A							
CU032	T2	16126C	16294T	16296T	16519C	73G	263G	315.1C												
CU033	A2+(64)	16111T	16189C	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del			
CU034	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A							
CU035	A2+(64)	16111T	16223T	16290T	16319A	16362C	16519C	64T	73G	146C	153G	235G	263G	291del	309.1C	315.1C	456T	523del	524del	
CU036	A2	16111T	16223T	16290T	16319A	16356C	16362C	16422C	16519C	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del		
CU037	A2	16111T	16223T	16290T	16319A	16356C	16362C	16422C	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del		
CU038	J2a2b1	16069T	16126C	16241G	16519C	73G	150T	185A	189G	195C	263G	295T	315.1C	489C						
CU039	D4	16223T	16224C	16325C	16362C	73G	194T	263G	315.1C	489C										
CU040	A2f3	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	151T	153G	235G	263G	309.1C	315.1C	523del	524del			
CU041	A2+(64)+16129	16111T	16129A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del			
CU042	B2	16183C	16189C	16193.1C	16217C	16278T	16355T	16519C	73G	103A	263G	309.1C	309.2C	315.1C	498del	499A				
CU043	B4b	16183C	16189C	16217C	16519C	73G	263G	309.1C	315.1C	499A										
CU044	U2e1	16051G	16129C	16183C	16189C	16193.1C	16362C	16519C	73G	152C	217C	263G	309.1C	309.2C	315.1C	340T	508G			
CU045	A2+(64)	16111T	16223T	16290T	16319A	16355T	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del		

CU046	A2+(64)	16092C	16111T	16223T	16260T	16290T	16319A	64T	73G	146C	152C	153G	182T	235G	263G	309.1C	309.2C	315.1C	316A	523del	524del
CU047	C1b	16189A	16223T	16291T	16298C	16325C	16327T	16519C	73G	152C	234G	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del
CU048	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	182T	235G	263G	309.1C	315.1C	523del	524del			
CU049	B2d	16172C	16183C	16189C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A								
CU050	A2h	16111T	16223T	16290T	16319A	16337T	16362C	16526A	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	456T	523del	524del	
CU051	L3b1a8	16145A	16213A	16223T	16278T	16362C	16519C	73G	195C	263G	279C	315.1C	523del	524del							
CU052	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del			
CU053	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	315.1C	523del	524del						
CU054	A2+(64)+16129	16111T	16129A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del				
H001	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A								
H002	C1	16223T	16298C	16325C	16327T	73G	249del	263G	290del	291del	309.1C	315.1C	489C								
H003	C1b	16169T	16223T	16298C	16325C	16327T	73G	195C	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del			
H004	C1b	16169T	16223T	16298C	16325C	16327T	73G	195C	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del			
H005	C1b	16169T	16223T	16298C	16325C	16327T	73G	195C	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del			
H006	B2d	16182C	16183C	16189C	16217C	16519C	73G	263G	309.1C	309.2C	309.3C	315.1C	498del	499A							
H007	C1	16223T	16298C	16325C	16327T	16390A	73G	115C	116del	152C	249del	263G	290del	291del	309.1C	309.2C	315.1C	356.1C	489C		
H008	B4b	16182C	16183C	16189C	16217C	16301T	16304C	16519C	73G	140T	263G	309.1C	309.2C	315.1C	499A						
H009	A2+(64)	16111T	16189C	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del				
H010	B2d	16183C	16189C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A										
H011	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del				

N002	M1a3	16129A	16183C	16189C	16193.1C	16311C	16519C	73G	195C	263G	309.1C	309.2C	315.1C	346C	460C	463.1C	463.2C	463.3C	489C
N003	B2d	16182C	16183C	16189C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A						
N004	B2d	16183C	16189C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A								
N005	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	195C	263G	309.1C	315.1C	498del	499A						
N006	A2	16111T	16223T	16290T	16319A	16356C	16362C	73G	146C	153G	235G	263G	310C	523del	524del				
N007	R	16519C	93G	151T	152C	263G	315.1C	373G											
N008	B2d	16183C	16189C	16217C	16356C	16519C	73G	263G	315.1C	498del	499A								
N009	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235R	263G	309.1C	315.1C	523del	524del		
N010	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A						
N011	L3e2b	16172C	16183G	16189C	16193.1C	16223T	16320T	16519C	73G	150T	195C	263G	282C	309.1C	315.1C				
N012	R0	16042A	16519C	263G	315.1C	523del	524del												
N013	L3d1c	16124C	16166G	16223T	73G	152C	263G	315.1C	523del	524del									
N014	D1f	16092C	16142T	16223T	16325C	16362C	16519C	73G	152C	263G	315.1C	489C							
N015	C1	16223T	16298C	16325C	16327T	73G	85A	188G	200G	228A	249del	263G	290del	291del	315.1C	489C			
N016	C1	16183C	16189C	16223T	16298C	16325C	16327T	16519C	73G	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del
N017	A2+(64)	16111T	16223T	16290T	16319A	16362C	16519C	64T	73G	146C	153G	235G	263G	315.1C	523del	524del			
N018	D4h3a	16203G	16223T	16241G	16245T	16301T	16319A	16342C	16362C	16468C	16533C	73G	152C	263G	309.1C	315.1C	489C		
N019	B2d	16183C	16189C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A								
N020	B2d	16183C	16186T	16189C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A							
N021	D	16189C	16223T	16325C	16362C	16519C	73G	263G	309.1C	309.2C	315.1C	489C							

N022	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del				
N023	A2+(64)+@153	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	235G	263G	266C	309.1C	315.1C	523del	524del				
N024	D	16189C	16223T	16325C	16362C	73G	263G	309.1C	315.1C	454C	489C	574T								
N025	A2+(64)	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del					
N026	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	456T	523del	524del			
N027	A2ac	16111T	16147T	16213A	16223T	16290T	16319A	16362C	16519C	64T	73G	146C	153G	215G	235G	263G	309.1C	315.1C		
N028	L3e4a	16051G	16223T	16264T	16519C	73G	150T	263G	315.1C	523del	524del									
N029	H	16519C	93G	151T	263G	315.1C														
N030	D4h3a	16187T	16203G	16223T	16241G	16245T	16301T	16319A	16342C	16362C	16468C	16533C	73G	152C	185A	263G	309.1C	315.1C	489C	
N031	D4h3a	16092C	16187T	16203G	16223T	16241G	16245T	16301T	16319A	16342C	16362C	16533C	73G	152C	263G	309.1C	315.1C	489C		
N032	A2r1	16111T	16223T	16261T	16287T	16290T	16311C	16319A	16362C	60.1T	73G	103A	146C	153G	235G	263G	309.1C	315.1C	523del	524del
N033	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A								
N034	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del			
N035	B2d	16182C	16183C	16189C	16217C	16311C	16519C	73G	146C	263G	315.1C	498del	499A							
N036	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	16519C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del		
N037	D1	16223T	16325C	16362C	16519C	73G	263G	309.1C	315.1C	489C										
N038	D1f	16142T	16223T	16325C	16362C	16519C	73G	263G	289G	309.1C	315.1C	489C								
N039	D1a2	16223Y	16272G	16325C	16362C	73G	143A	210G	263G	309.1C	315.1C	489C								
N040	A2ac	16111T	16213A	16223T	16290T	16311C	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del		
N041	C1	16223T	16274A	16298C	16325C	16327T	73G	185A	249del	263G	290del	291del	309.1C	315.1C	489C					

N042	A2+(64)	16092C	16111T	16223T	16260T	16290T	16319A	64T	73G	146C	152C	153G	182T	235G	263G	309.1C	309.2C	315.1C	316A	523del	524del
N043	A2+(64)	16111T	16223T	16290T	16319A	16362C	16519C	64T	73G	146C	153G	198T	235G	263G	309.1C	315.1C	523del	524del			
N044	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A								
N045	A2+(64)	16092C	16111T	16155G	16223T	16290T	16319A	64T	73G	146C	153G	182T	235G	263G	309.1C	315.1C	316A	523del	524del		
N046	J1c	16069T	16126C	73G	185A	188G	263G	295T	315.1C	462T	482C	489C									
N047	B2d	16183C	16186T	16189C	16216G	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A								
N048	B2d	16183C	16189C	16193.1C	16217C	16260Y	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A							
N049	C1d1	16051G	16172C	16223T	16298C	16325C	16327T	73G	146C	194T	195C	249del	263G	290del	291del	309.1C	315.1C	489C	523del	524del	
N050	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A								
Q001	A2+(64)	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del						
Q002	D4h3a6	16172C	16185T	16192T	16223T	16301T	16342C	16362C	16463G	73G	143A	152C	195C	263G	309.1C	309.2C	315.1C	489C			
Q003	A2+(64)+16189	16111T	16183C	16189C	16223T	16287T	16290T	16319A	16362C	73G	146C	153G	235G	263G	315.1C	523del	524del				
Q004	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del					
Q005	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del					
Q006	B2d	16183C	16189C	16217C	16519C	73G	263G	315.1C	498del	499A											
Q007	A2+(64)	16111T	16223T	16290T	16319A	16362C	10C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del			
Q008	B2d	16182C	16183C	16189C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A								
Q009	B2d	16179T	16183C	16189C	16217C	16519C	73G	263G	315.1C	498del	499A										
Q010	B2d	16182C	16183C	16189C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A								
Q011	C1b	16223T	16224C	16298C	16325C	16327T	16344T	73G	127Y	152C	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del	

R002	B2d	16183C	16189C	16217C	16519C	73G	263G	315.1C	498del	499A													
R003	C1	16051G	16172C	16189C	16223T	16298C	16325C	16327T	16362C	73G	249del	263G	290del	291del	309.1C	315.1C	489C	523del	524del	573.1C	573.2C	573.3C	573.4C
R004	A2w1	16111T	16187T	16189C	16223T	16290T	16319A	16362C	73G	146C	153G	235G	309.1C	315.1C	485C	573.1C	573.2C	573.3C	573.4C	573.5C			
R005	A2+(64)	16111T	16189C	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del						
R006	B2d	16179T	16183C	16189C	16217C	16519C	73G	263G	315.1C	498del	499A												
R007	B2d	16182C	16183C	16189C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A										
R008	B2d	16183C	16189C	16217C	16519C	73G	146C	263G	309.1C	315.1C	498del	499A											
R009	B2d	16183C	16189C	16193.1C	16217C	16256T	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A									
R010	C1d1	16051G	16172C	16223T	16298C	16325C	16327T	73G	146C	194T	195C	249del	263G	290del	291del	309.1C	315.1C	489C	523del	524del			
R011	B2	16092C	16182C	16183C	16189C	16217C	16519C	73G	94A	263G	315.1C	498del	499A										
R012	A2v	16111T	16223T	16239T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	315.2C	523del	524del					
R013	B2d	16182C	16183C	16189C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A											
R014	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del							
R015	B4b	16086C	16153A	16182C	16183C	16189C	16193.1C	16217C	16519C	73G	114T	146C	152C	263G	309.1C	309.2C	309.3C	315.1C	499A				
R016	A2w1	16111T	16187T	16223T	16290T	16319A	16362C	73G	146C	153G	235G	309.1C	315.1C	485C	523del	524del	573.1C	573.2C	573.3C	573.4C			
R017	B2d	16179T	16183C	16189C	16217C	16519C	73G	263G	315.1C	498del	499A												
R018	B2d	16183C	16189C	16217C	16324C	16519C	73G	263G	309.1C	315.1C	498del	499A											
R019	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A										
R020	B2d	16182C	16183C	16189C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A										
R021	D4h3a6	16172C	16185T	16192T	16223T	16301T	16342C	16362C	16463G	73G	143A	152C	195C	263G	309.1C	309.2C	315.1C	489C					

R022	C1	16051G	16172C	16189C	16223T	16298C	16325C	16327T	16362C	73G	249del	263G	290del	291del	309.1C	315.1C	489C	523del	524del	573.1C	573.2C	573.3C
R023	B2d	16183C	16189C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A											
R024	C1d+194	16051G	16223T	16298C	16325C	16327T	16519C	73G	194T	249del	263G	290del	291del	309.1C	315.1C	489C	523del	524del				
R025	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A									
R026	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A									
R027	B2d	16182C	16183C	16189C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A										
R028	B4b	16182C	16183C	16189C	16217C	16301T	16304C	16519C	73G	140T	263G	309.1C	309.2C	315.1C	499A							
R029	B4b	16182C	16183C	16189C	16217C	16519C	73G	152C	215G	263G	309.1C	309.2C	315.1C	499A								
R030	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del					
R031	C1c5	16223T	16298C	16325C	16327T	16526A	73G	115del	249del	263G	290del	291del	309.1C	315.1C	489C	524.1A	524.2C					
R032	A2w1	16111T	16187T	16223T	16290T	16319A	16362C	73G	146C	153G	235G	309.1C	315.1C	485C	523del	524del	573.1C	573.2C	573.3C	573.4C	573.5C	573.6C
R033	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A									
R034	B4b	16182C	16183C	16189C	16217C	16301T	16304C	16519C	73G	140T	263G	309.1C	315.1C	498.1C	499A							
R035	B4b	16182C	16183C	16189C	16217C	16519C	73G	152C	215G	263G	309.1C	309.2C	315.1C	499A								
R036	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A	524.1A	524.2C							
R037	B4b	16182C	16183C	16189C	16217C	16301T	16304C	16519C	73G	140T	263G	309.1C	309.2C	315.1C	499A							
R038	B2d	16182C	16183C	16189C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A									
R039	D	16189C	16223T	16325C	16362C	73G	263G	309.1C	315.1C	489C												
R040	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A	524.1A	524.2C							
R041	D1f	16142T	16147T	16223T	16265R	16325C	16362C	16519C	73G	195C	263G	309.1C	315.1C	489C	524.1A	524.2C						

S012	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	182T	235G	263G	309.1C	315.1C	523del	524del								
S013	A2af1a1	16111T	16223T	16290T	16319A	16360T	16362C	89C	106del	107del	108del	109del	110del	111del	146C	152C	153R	195C	198T	235G	263G	309.1C	309.2C	315.1C	523del	524del
S014	B2d	16183C	16189C	16193.1C	16217C	16356C	16519C	73G	263G	309.1C	315.1C	498del	499A													
S015	C1	16155G	16223T	16298C	16325C	16327T	73G	249del	263G	290del	291del	315.1C	489C	573.1C	573.2C											
S016	A2+(64)	16111T	16189C	16193.1C	16223T	16259T	16290T	16300G	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del						
S017	A2r1	16111T	16223T	16261T	16290T	16311C	16319A	16362C	60.1T	73G	103A	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del						
S018	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	182T	235G	263G	309.1C	315.1C	523del	524del								
S108	H	16519C	93G	151T	263G	315.1C																				
S019	B2d	16183C	16189C	16217C	16324C	16519C	73G	263G	309.1C	315.1C	498del	499A														
S020	B2d	16183C	16189C	16217C	16519C	73G	152Y	263G	309.1C	315.1C	498del	499A														
S021	D1f	16092C	16142T	16223T	16325C	16362C	16519C	73G	152C	263G	315.1C	489C														
S022	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del										
S023	A2v	16111T	16223T	16239T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del									
S024	B2d	16183C	16189C	16217C	16299G	16519C	73G	263G	309.1C	315.1C	498del	499A														
S025	A2+(64)	16111T	16189C	16223T	16259T	16290T	16300G	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del							
S026	C1	16155G	16223T	16298C	16325C	16327T	73G	249del	263G	290del	291del	315.1C	489C	573.1C												
S027	C1c5	16192T	16223T	16298C	16325C	16327T	16526A	73G	115del	249del	263G	290del	291del	309.1C	309.2C	315.1C	489C									
S028	B2d	16183C	16189C	16217C	16256T	16324C	16519C	73G	263G	315.1C	498del	499A														
S029	B2d	16183C	16189C	16217C	16519C	73G	152C	263G	309.1C	315.1C	498del	499A														
S030	B2b3a	16182C	16183C	16189C	16217C	16249C	16312G	16344T	16519C	73G	152C	263G	271T	309.1C	309.2C	315.1C	455del	460C	463.1C	463.2C	499A					

S031	B2d	16183C	16189C	16217C	16299G	16519C	73G	152Y	263G	309.1C	315.1C	498del	499A							
S032	B2d	16182C	16183C	16189C	16217C	16519C	73G	263G	309.1C	309.2C	309.3C	315.1C	498del	499A						
S033	A2+(64)	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del				
S034	D1	16223T	16325C	16362C	73G	89.1A	106del	107del	108del	109del	110del	111del	263G	315.1C	489C					
S035	C1c5	16223T	16298C	16325C	16327T	16362C	16526A	73G	115del	249del	263G	290del	291del	310C	489C					
S036	C1c5	16192T	16223T	16298C	16325C	16327T	16526A	73G	115del	119C	249del	263G	290del	291del	309.1C	309.2C	315.1C	489C		
S037	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A								
S038	A2+(64)	16111T	16189C	16223T	16224C	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	338T	523del	524del	
S039	B2d	16183C	16189C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A									
S040	D1f	16092C	16142T	16223T	16325C	16362C	16519C	73G	152C	263G	315.1C	489C								
S041	B2d	16183C	16189C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A									
S042	B4b	16129A	16182C	16183C	16189C	16217C	16519C	73G	263G	309.1C	315.1C	499A								
S043	A2+(64)	16111T	16223T	16290T	16319A	16362C	16519C	64T	73G	146C	153G	198T	235G	263G	309.1C	315.1C	523del	524del		
S044	B2b3a	16182C	16183C	16189C	16217C	16249C	16312G	16344T	16519C	73G	152C	204C	263G	271T	309.1C	309.2C	315.1C	460C	463.1C	499A
S045	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del				
S046	A2+(64)	16111T	16134T	16223T	16290T	16292A	16319A	16362C	64T	73G	146C	153G	235G	263G	315.1C	523del	524del			
S047	L2a1	16071T	16189C	16192T	16223T	16278T	16294T	16309G	16390A	16519C	73G	146C	152C	195C	263G	309.1C	315.1C			
S048	A2+(64)	16069T	16111T	16223T	16290T	16319A	16362C	16519C	64T	73G	146C	153G	235G	263G	315.1C	523del	524del			
S049	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A	524.1A	524.2C						
S050	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A							

S071	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del			
S072	D1f	16092C	16142T	16223T	16325C	16362C	16519C	73G	152C	263G	315.1C	489C								
S073	A2+(64)+16129	16111T	16129A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del			
S074	A2+(64)	16111T	16183C	16189C	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del
S075	A2	16111T	16290T	16319A	16356C	16362C	73G	146C	153G	234G	235G	263G	309.1C	309.2C	315.1C	523del	524del			
S076	D1	16223T	16325C	16362C	16519C	73G	152C	263G	309.1C	315.1C	489C									
S077	B2d	16183C	16189C	16217C	16311C	16519C	73G	263G	309.1C	315.1C	498del	499A								
S078	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del			
S079	B2d	16182C	16183C	16189C	16217C	16519C	73G	182Y	263G	308del	309del	315.1C	498del	499A						
S080	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del			
S081	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	152C	263G	309.1C	309.2C	315.1C	498del	499A						
S082	R0	16519C	263G	309.1C	315.1C															
S083	A2+(64)	16093C	16111T	16223T	16290T	16319A	16362C	64T	73G	143A	146C	153G	235G	263G	309.1C	315.1C	523del	524del		
S084	D1	16189C	16223T	16272G	16291T	16325C	16362C	73G	143A	210G	263G	309.1C	315.1C	489C						
S085	R0	16519C	263G	309.1C	315.1C															
S086	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A							
S087	B2d	16183C	16189C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A									
S088	D1	16125A	16169T	16189C	16223T	16325C	16362C	16519C	73G	152C	263G	309.1C	315.1C	489C						
S089	J1c	16069T	16126C	73G	185A	188G	263G	295T	315.1C	462T	482C	489C								
S090	C1	16172C	16223T	16274A	16298C	16325C	16327T	73G	185A	249del	263G	290del	291del	315.1C	489C					

S112	B2d	16182C	16183C	16189C	16193.1C	16217C	16519C	73G	263G	308del	309del	315.1C	498del	499A	523del	524del			
S113	J1c	16069T	16126C	73G	185A	188G	263G	295T	315.1C	462T	482C	489C							
S114	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del		
S115	A2+(64)	16092C	16111T	16155G	16223T	16290T	16319A	64T	73G	146C	153G	182T	235G	263G	309.1C	315.1C	316A	523del	524del
S116	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	456T	523del	524del		
S117	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A						
S118	D4h3a	16203G	16223T	16241G	16245T	16301T	16319A	16342C	16362C	16468C	16533C	73G	152C	263G	309.1C	315.1C	489C		
S119	B2d	16183C	16189C	16193.1C	16217C	16256T	16519C	73G	263G	309.1C	315.1C	498del	499A						
S120	A2+(64)	16111T	16290T	16319A	16356C	16362C	64T	73G	146C	153G	234G	235G	263G	309.1C	309.2C	315.1C	523del	524del	
S121	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del		
S122	A2r1	16111T	16223T	16261T	16290T	16311C	16319A	16362C	60.1T	73G	103A	146C	153G	235G	263G	309.1C	315.1C	523del	524del
S123	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del		
S124	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A							
S125	A2+(64)+16129	16111T	16129A	16223T	16278T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del	
S126	D1a2	16223T	16272G	16325C	16362C	73G	143A	210G	263G	309.1C	309.2C	315.1C	489C						
S127	H	16519C	93G	151T	263G	315.1C													
S128	A2+(64)	16111T	16290T	16319A	16356C	16362C	64T	73G	146C	153G	234G	235G	263G	309.1C	315.1C	523del	524del		
S129	A2+(64)	16111T	16189C	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del		
S130	D1a2	16223T	16272G	16325C	16362C	73G	143A	210G	263G	309.1C	315.1C	489C							
S131	A2	16111T	16223T	16256T	16290T	16319A	16356C	16362C	73G	146C	153G	235G	263G	315.1C	523del	524del			

S132	L1b	16126C	16187T	16189C	16223T	16264T	16270T	16278T	16311C	16519C	73G	152C	182T	185T	195C	247A	263G	315.1C	357G	523del	524del
S133	D1	16223T	16325C	16362C	73G	89.1A	106del	107del	108del	109del	110del	111del	263G	309.1C	309.2C	315.1C	489C				
S134	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del				
S135	A2+(64)	16111T	16213A	16223T	16290T	16319A	16362C	16526A	64T	73G	146C	200G	235G	263G	309.1C	309.2C	315.1C	523del	524del		
S136	A2v	16111T	16223T	16239T	16284G	16290T	16319A	16362C	16508T	16519C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del	
S137	D1	16223T	16325C	16362C	16519C	73G	152C	263G	309.1C	315.1C	489C										
S138	A2+(64)	16111T	16129A	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	199C	235G	263G	309.1C	315.1C	523del	524del			
S139	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A								
S140	B2d	16093C	16183C	16189C	16217C	16519C	73G	263G	315.1C	498del	499A										
S141	A2ac	16111T	16211Y	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	315.1C	523del	524del				
S142	C1d+194	16051G	16223T	16298C	16325C	16327T	16519C	73G	194T	249del	263G	290del	291del	309.1C	315.1C	489C	523del	524del			
S143	B2d	16183C	16189C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A										
S144	R0	16519C	263G	309.1C	309.2C	315.1C															
S145	B2h	16182C	16183C	16189C	16217C	16468C	16519C	73G	263G	309.1C	315.1C	499A	573.1C	573.2C							
S146	B2d	16182C	16183C	16189C	16217C	16519C	73G	151T	263G	309.1C	315.1C	498del	499A								
S147	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	215G	235G	263G	315.1C	523del	524del					
S148	D4h3a	16092C	16187T	16203G	16223T	16241G	16245T	16301T	16319A	16342C	16362C	16533C	73G	152C	263G	309.1C	315.1C	489C			
S149	B2d	16182C	16183C	16189C	16217C	16519C	73G	114T	263G	309.1C	309.2C	315.1C	498del	499A							
S150	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	189G	215G	235G	263G	315.1C	523del	524del				
S151	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del			

S152	B2d	16183C	16189C	16217C	16299G	16519C	73G	263G	309.1C	315.1C	498del	499A											
S153	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del					
S154	C1	16155G	16223T	16298C	16325C	16327T	73G	249del	263G	290del	291del	315.1C	489C	573.1C									
S155	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del						
S156	L0a1a+200	16129A	16148T	16168T	16172C	16187T	16188G	16189C	16223T	16230G	16311C	16320T	16362C	93G	185A	189G	200G	236C	247A	263G	315.1C	523del	524del
S157	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A											
S158	A2y	16111T	16137G	16290T	16293G	16304C	16319A	16362C	64T	73G	146C	152C	153G	189G	198T	207A	235G	263G	309.1C	315.1C	523del	524del	
S159	L3d5a	16093C	16124C	16184T	16223T	16362C	16519C	73G	152C	195C	199C	263G	315.1C	523del	524del								
S160	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	215G	235G	263G	315.1C	523del	524del							
S161	A2+(64)+16129	16111T	16129A	16223T	16256T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del					
S162	B2d	16169T	16182C	16183C	16189C	16217C	16362C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A								
S163	A2v	16111T	16223T	16239T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del						
S164	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A										
S165	A2+(64)	16111T	16290T	16319A	16356C	16362C	64T	73G	146C	153G	234G	235G	263G	309.1C	315.1C	523del	524del						
S166	B2d	16183C	16189C	16193.1C	16217C	16218T	16519C	73G	263G	315.1C	498del	499A	524.1A	524.2C									
S167	A2+(64)+@153	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	152C	235G	263G	315.1C	523del	524del								
S168	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del						
S169	B2d	16183C	16189C	16193.1C	16217C	16343G	16519C	73G	263G	315.1C	498del	499A											
S170	B2d	16183C	16189C	16193.1C	16217C	16299G	16519C	73G	263G	309.1C	315.1C	498del	499A										
S171	A2+(64)	16111T	16189C	16223T	16224C	16290T	16319A	64T	73G	146C	153G	235G	263G	309.1C	315.1C	338T	523del	524del					

S172	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	456T	523del	524del		
S173	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	214G	235G	263G	309.1C	309.2C	315.1C	523del	524del
S174	D1f	16092Y	16142T	16223T	16325C	16362C	16519C	73G	152C	263G	309.1C	315.1C	489C						
S175	B2d	16182M	16183C	16189C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A						
S176	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	210G	235G	263G	309.1C	315.1C	523del	524del		
S177	C1b	16223T	16298C	16309G	16317G	16325C	16327T	73G	249del	263G	290del	291del	315.1C	316A	489C	493G	523del	524del	
S178	C1	16183C	16189C	16223T	16325C	16327T	73G	204C	249del	263G	282C	290del	291del	315.1C	489C	493G	523del	524del	
S179	A2v	16111T	16223T	16239T	16290T	16319A	16362C	16519C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del	
S180	A2+(64)	16111T	16188.1C	16223T	16290T	16319A	64T	73G	146C	153G	263G	309.1C	315.1C	523del	524del				
S181	D1f	16092C	16142T	16223T	16325C	16362C	16519C	73G	152C	263G	315.1C	489C							
S182	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del	
S183	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A							
S184	L2a1+143	16129A	16223T	16278T	16294T	16311C	16390A	73G	143A	146C	152C	182T	195C	263G	309.1C	315.1C	523del	524del	
S185	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A							
S186	D1f	16092C	16142T	16223T	16325C	16362C	16519C	73G	152C	263G	315.1C	489C							
S187	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del		
S188	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	64T	73G	146C	153G	182T	235G	263G	309.1C	315.1C	523del	524del	
S189	D4h3a	16187T	16203G	16223T	16241G	16245T	16301T	16319A	16342C	16362C	16468C	16533C	73G	152C	263G	309.1C	315.1C	489C	
S190	D1f	16092C	16142T	16209C	16223T	16325C	16362C	16519C	73G	152C	263G	315.1C	489C						
S191	C1	16183C	16189C	16223T	16325C	16327T	73G	204C	249del	263G	282C	290del	291del	315.1C	489C	493G	523del	524del	

S192	D1f	16092C	16142T	16223T	16274A	16311C	16325C	16362C	16519C	73G	152C	263G	315.1C	489C							
S193	D1f	16092C	16142T	16209C	16223T	16325C	16362C	16519C	73G	152C	263G	315.1C	489C								
S194	A2+(64)	16111T	16126C	16223T	16290T	16319A	16362C	64T	73G	146C	151T	153G	235G	263G	315.1C	523del	524del				
S195	U5b	16183C	16189C	16193.1C	16256T	16270T	16286T	73G	150T	200G	263G	309.1C	309.2C	315.1C							
S196	C1d+194	16051G	16093Y	16223T	16298C	16325C	16327T	16519C	73G	194T	249del	263G	290del	291del	315.1C	489C	523del	524del			
S197	A2v	16111T	16223T	16239T	16284G	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	315.1C	523del	524del				
S198	B2d	16183C	16189C	16217C	16299G	16519C	73G	263G	309.1C	315.1C	498del	499A									
S199	B2d	16182C	16183C	16189C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A								
S200	B2d	16097C	16098G	16183C	16189C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A								
S201	A2+(64)+@153	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	235G	263G	309.1C	309.2C	315.1C	456T	523del	524del				
S202	A2+(64)	16111T	16223T	16231C	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	456T	523del	524del			
S203	J1c	16069T	16126C	73G	185A	188G	263G	295T	315.1C	462T	482C	489C									
S204	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A								
T001	B2d	16182C	16183C	16189C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A								
T002	A2al	16092C	16111T	16126C	16223T	16290T	16319A	16362C	16518A	64T	73G	146C	153G	235G	263G	315.1C	523del	524del			
T003	J2a2b1	16069T	16126C	16241G	16519C	73G	150T	185A	189G	195C	263G	295T	315.1C	489C							
T004	A2+(64)	16111T	16223T	16290T	16319A	16362C	16519C	64T	73G	119C	125C	127C	146C	153G	235G	263G	309.1C	309.2C	315.1C	498del	499A
T005	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	150T	263G	309.1C	315.1C	498del	499A								
T006	B2d	16183C	16189C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A									
T007	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del				

T008	D4h3a6	16129A	16185T	16192T	16223T	16241G	16301T	16342C	16362C	16463G	73G	143A	152C	263G	315.1C	489C						
T009	A2+(64)	16111T	16223T	16290T	16292T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del					
T010	C1d+194	16051G	16094C	16223T	16261T	16298C	16325C	16327T	16518A	73G	153G	194T	249del	263G	290del	291del	309.1C	315.1C	489C	523del	524del	
T011	A2	16111T	16223T	16290T	16319A	16362C	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del							
T012	A2ac	16111T	16213A	16223T	16290T	16319A	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del	573.1C					
T013	C1d+194	16051G	16223T	16298C	16325C	16327T	16519C	73G	194T	249del	263G	290del	291del	309.1C	315.1C	489C	523del	524del	573.1C			
T014	B2d	16183C	16189C	16193.1C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A										
T015	A2ac	16111T	16213A	16223T	16290T	16319A	16362C	16519C	64T	73G	146C	153G	235G	263G	315.1C	523del	524del					
T016	L3e2b	16172C	16183C	16184T	16189C	16223T	16320T	16519C	16525G	73G	150T	152C	195C	263G	309.1C	315.1C						
T017	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del						
T018	D1f	16142T	16223T	16325C	16362C	16519C	73G	150T	152C	263G	309.1C	315.1C	489C									
T019	D1f	16142T	16223T	16325C	16362C	16519C	73G	150T	152C	263G	309.1C	315.1C	489C									
T020	A2+(64)	16104T	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del					
T021	B4b	16182C	16183C	16189C	16217C	16519C	73G	152C	215G	263G	309.1C	309.2C	315.1C	499A								
T022	C1d+194	16051G	16094C	16223T	16261T	16298C	16325C	16327T	16518A	73G	153G	194T	249del	263G	290del	291del	309.1C	315.1C	489C	523del	524del	
T023	C1d+194	16051G	16223T	16298C	16325C	16327T	16519C	73G	194T	249del	263G	290del	291del	315.1C	489C	523del	524del					
T024	R0	16042A	16519C	263G	315.1C	523del	524del															
T025	B4	16183C	16189C	16194C	16195C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A								
T026	A2+(64)	16111T	16192T	16223T	16248T	16290T	16319A	16362C	59C	62T	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del		
T027	L2a1	16038.1A	16129A	16189C	16192T	16223T	16278T	16294T	16309G	16390A	73G	143A	146C	152C	195C	263G	309.1C	315.1C				

T048	B2d	16097C	16098G	16183C	16189C	16217C	16519C	73G	263G	309.1C	315.1C	498del	499A						
T049	B2d	16183C	16189C	16193.1C	16217C	16223T	16519C	73G	146Y	263G	309.1C	309.2C	315.1C	498del	499A				
T050	C1b	16223T	16298C	16322G	16325C	16327T	16519C	73G	150T	249del	263G	290del	291del	309.1C	315.1C	489C	493G	523del	524del
T051	A2+(64)	16111T	16223T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	315.1C	523del	524del			
T052	A2+(64)+16129	16111T	16129A	16223T	16256T	16290T	16319A	16362C	64T	73G	146C	153G	235G	263G	309.1C	309.2C	315.1C	523del	524del
T053	B2d	16183C	16189C	16217C	16519C	73G	263G	309.1C	309.2C	315.1C	498del	499A							

Apéndice J. Ponencia en el 28th Congress of the International Society for Forensic Genetics.



Laboratorio de Genética

MATERNAL GENETIC CHARACTERIZATION OF A COLOMBIAN ANDEAN POPULATION

A. Castillo¹, Deccache², E. F. de Carvalho², F. Simão², L. Gusmão²

¹Genetics Laboratory, Industrial University of Santander (UIS), Bucaramanga, Colombia.
²DNA Diagnostic Laboratory (LDD), State University of Rio de Janeiro (UERJ), Rio de Janeiro, Brazil.

Corresponding author: Adriana Castillo, Tel: 57-7-6358641 email: castillo@uis.edu.co



INTRODUCTION

As for most American populations, the genetic composition of Colombia results from the admixture between Native Americans, the first inhabitants of the continent, European, that arrived during the colonization period, and Africans, who were brought to America as slave handwork. The admixture among these three well differentiated populations resulted in a high genetic diversity. Previous data obtained for autosomal ancestry informative markers, as well as for X and Y-chromosomal markers, showed that European and African ancestries varied over the country [1,2,3]. In the light of this, the aim of this study was to obtain an overview on the maternal lineages composition and diversity of the Colombian Andean populations.



Population distribution, according to the 2005 Colombian census.
<https://www.dane.gov.co>

Fig. 1. Map of Colombia. The Andean region is subdivided in 10 departments, concentrating about 70% of the population.

RESULTS AND DISCUSSION

For the 67 samples analysed in this study, 54 different haplotypes were found. A total of 43 haplotypes were unique, 9 were shared between two individuals and two were shared between three individuals, resulting in a haplotype diversity of 0.9932 ± 0.0039 . Although high, this value is below those found in other South American admixed populations, for example, Rio de Janeiro (HD=0.9994; [4]).

The majority of the haplotypes (91%) belong to Native American haplogroups included in the branches A2, B2s, B4, C1, D1 and D4. The remaining belong to the Eurasian haplogroups J1c, M and U5a2 and African haplogroups L3b1a8 and L3f1b1a1. Most native American haplogroups were previously reported in populations all over the continent. However, a search in EMPOP revealed that the haplogroups A2ac, A2y, B2b3a, C1d1d, D1f and D1f2 are restricted to South America. In particularly the haplogroups A2ac and D1f2 were only found in Colombia.

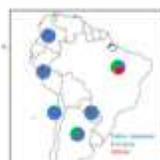


Fig. 2 Continental ancestry in different South American countries

The differences found with Brazil and Argentina can be explained by the lower proportion of non-native haplogroups in Colombia (Fig. 2). The remaining populations, although harbouring similar proportions of admixture with non-natives, distances remain high, most probably due to differences in the Native American background of these populations.

It is worth mentioning that the Colombian Andean sample from this study showed a lower proportion of African maternal ancestry than the sample from the 1000 genomes project. This can be explained by differences in sampling strategies, since about 70% of the samples in this work were collected in Santander department (Northeast Andean region), and the samples from the 1000 genomes project came from Medellín (department of Antioquia, located in the Central West Andean region). Previous study on autosomal markers showed that Native American ancestry is higher in the Northeast than in Central West Andean region [1].

	Colombia	Peru	Paraguay	Paraguay	Paraguay	Colombia	Andean
	Andean	Andean	Andean	Andean	Andean	Andean	Andean
Peru-Andean	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Paraguay-Andean	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Colombia-Andean	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 1. Matrix of pairwise F_{ST} genetic distances South American admixed populations (below diagonal) and the corresponding differentiation p-value (above diagonal) obtained for 1000 permutations, for aDNA control region.

F_{ST} genetic distances were calculated between the Andean sample and published data from other South American admixed populations. No significant differences were observed between our sample and the Colombian population deposited on The 1000genome Project [5]. High F_{ST} genetic distances were obtained in the comparison of our sample with Peru [5], Argentina[6], Paraguay [7] and Brazil [4].

REFERENCES

[1] Ossa et al., PLoS ONE 11 (2016) e0164414; [2] Castillo et al., Forensic Sci. Genetics Supplement Series. V (2015) 482-483; [3] Ibarra et al., PLOS ONE, 9 (2014), e87202; [4] Simão et al., Forensic Sci. Int. Genet. 34 (2018) 97-104; [5] 1000 Genomes Project Consortium, A. Auton, L. Brooks, R. Durbin, et al., Nature, 526 (2015) 68-74; [6] Bobillo et al., Int J Legal Med. 124(2010):263-6; [7] Simão et al., Forensic Sci. Int. Genet. 39 (2019) 66-72.

CONCLUSION

Several studies report a lack of association between Native American mtDNA lineages and ethnic affiliation or geographic origin. This was once again observed in the Andean population from Colombia, since all haplogroups obtained were previously found both in central and south American populations.

MATERIAL AND METHODS

Sixteen-seven blood samples were collected from unrelated individuals from Norte de Santander, Santander and Cundinamarca. DNA was extracted with Chelex. Samples were amplified for the CR using the Qiagen® multiplex PCR kit, 1-5 ng DNA and primers L15900 and H639 (0.2 μM) in 5 μl final volume. PCR products were purified with ExoSAP. Sequencing was performed with the BigDye™ Terminator v3.1 Cycle Sequencing kit, with primers L15900, L16555, L159 and L639. Sequencing products were purified with Sephadex. Separation and detection was performed in an ABI 3500 (Applied Biosystems). Sequences were compared to the rCRS using the software SeqScape v2.7. Haplogrouping was performed via EMPOP, according to Phylotree build 17. Haplogroup frequencies were calculated by direct counting and haplotype diversities and genetic distances were calculated using the Arlequin software v3.5.

Apéndice K. Artículo “Maternal genetic characterization of a colombian andean population”

A. Castillo, L. Deccache, E.F. de Carvalho, F. Simão, L. Gusmão, Maternal genetic characterization of a Colombian Andean population, *Forensic Science International: Genetics Supplement Series*, Volume 7, Issue 1, 2019, Pages 342-344, <https://doi.org/10.1016/j.fsigss.2019.10.004>.

Apéndice L.Certificado EMPOP QC EMP00871 (Departamentos Región Andina, 458 datos)

Institute of Legal Medicine
Medical University of Innsbruck
Director: Prof. Richard Scheithauer, MD
Muehlenstraße 44, A-6020 Innsbruck, Austria
Tel: +43/512/9003-70600, Fax: +43/512/9003-73600
e-mail: gmi@gmi.eu



Prof. Dr. Walther Parson
Head of High Throughput DNA Database Unit
Head of Forensic Genomics
walther.parson@i-med.ac.at

Innsbruck, 31.10.2022

EMPOP QC Report

Dear submitters,

thank you for providing mtDNA data for quality control (QC) to EMPOP!

The haplotypes were evaluated with the EMPOP QC software suite and further manually scrutinized. The final version of the data was established based on your feedback and our interpretation of the data that you kindly provided.

1. Dataset

origin of samples: Colombia, diverse departments
metapopulation: Admixed
contact: Adriana Castillo, Humberto Ossa, Filipa Simão, Leonor Gusmão
no. of samples: 458 (all used)
mtDNA segment: CR
accession number: EMP00871
format: rCRS-coded haplotypes

2. General information

The submitted mitotypes were generated by Sanger Sequencing covering the entire Control Region following the updated ISFG guidelines (Parson et al 2014).

3. Observations

The observations made during QC were discussed with the submitters. **No changes to the dataset were required. Congratulations!**

Institute of Legal Medicine, Medical University of Innsbruck

4. Haplogrouping

In the course of the EMPOP QC process the haplotypes were turned into fasta-like nucleotide strings, phylogenetically aligned and haplogrouped according to the EMPOP phylogeny (SAM2, Dür et al 2021). **No changes to alignment and haplogrouping were required.**

If you make use of this information, please cite

Dür A, Huber N, Parson W (2021). Fine-Tuning Phylogenetic Alignment and Haplogrouping of mtDNA Sequences; Int J Mol Sci **22**(11):5747; doi: 10.3390/ijms22115747.

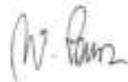
5. Decision

The EMPOP QC results indicate that the mitotypes are **acceptable** in the forensic genetic context **without required changes**.

The accession number of this dataset is EMP00871. Please use this number for publication of the data and refer to the QC process by citing the following reference:

Parson W and Dür A (2007) EMPOP--a forensic mtDNA database. Forensic Sci Int Genet **1**(2): 88-92.

Best wishes,



Prof. Dr. Walther Parson
On behalf of the EMPOP Team

References

1. Parson W. et al. DNA Commission of the International Society for Forensic Genetics: revised and extended guidelines for mitochondrial DNA typing. Forensic Sci Int Genet, 2014. 13: 134-142.

Disclaimer: The applied quality control cannot be regarded as comprehensive independent evaluation of the submitted dataset that would be required in a forensic setting, but constitutes an optimized procedure for the detection of common data idiosyncrasies. The signatories cannot be made liable for correctness, completeness and topicality of the contents.

Apéndice M. Frecuencia de los haplogrupos en los departamentos de la región andina

Haplogrupo	Antioquia		Boyaca		Caldas		Cundinamarca		Huila		Norte de Santander		Quindio		Risaralda		Tolima		Santander	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
	A2	-	-	-	-	-	-	3	5,6%	-	-	1	2,0%	-	-	-	-	2	3,8%	4
A2+(64)	11	22,4%	17	32,7%	13	26,0%	16	29,6%	7	14,0%	7	14,0%	12	24,0%	3	6,0%	11	20,8%	35	17,2%
A2+(64)+@153	-	-	1	1,9%	-	-	1	1,9%	-	-	1	2,0%	-	-	-	-	-	-	3	1,5%
A2+(64)+16129	1	2,0%	1	1,9%	-	-	3	5,6%	-	-	-	-	1	2,0%	-	-	1	1,9%	3	1,5%
A2+(64)+16189	1	2,0%	-	-	-	-	-	-	1	2,0%	-	-	1	2,0%	-	-	-	-	-	-
A2ab	-	-	-	-	-	-	-	-	1	2,0%	-	-	-	-	-	-	-	-	-	-
A2ac	6	12,2%	3	5,8%	1	2,0%	7	13,0%	8	16,0%	5	10,0%	1	2,0%	1	2,0%	3	5,7%	21	10,3%
A2ad2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	5,7%	-	-
A2af1a1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	1,0%
A2al	-	-	1	1,9%	-	-	-	-	1	2,0%	1	2,0%	-	-	-	-	1	1,9%	-	-
A2f3	-	-	-	-	-	-	1	1,9%	-	-	-	-	-	-	-	-	-	-	-	-
A2h	-	-	-	-	-	-	1	1,9%	-	-	-	-	-	-	-	-	-	-	-	-
A2q	2	4,1%	-	-	2	4,0%	-	-	-	-	-	-	-	-	-	-	-	-	-	-
A2r1	-	-	-	-	-	-	1	1,9%	-	-	-	-	-	-	-	-	-	-	2	1,0%
A2v	-	-	-	-	-	-	-	-	-	-	-	-	-	1	2,0%	-	-	-	5	2,5%
A2w1	1	2,0%	-	-	1	2,0%	-	-	-	-	-	-	-	3	6,0%	-	-	-	-	-
A2y	-	-	1	1,9%	-	-	1	1,9%	3	6,0%	-	-	-	-	-	-	-	-	1	0,5%
B2	2	4,1%	-	-	-	-	1	1,9%	-	-	-	-	-	-	1	2,0%	-	-	-	-
B2b1	-	-	-	-	-	-	-	-	1	2,0%	-	-	-	-	-	-	-	-	-	-
B2b3a	-	-	-	-	1	2,0%	-	-	-	-	-	-	-	-	-	-	-	-	2	1,0%
B2d	10	20,4%	15	28,8%	15	30,0%	6	11,1%	5	10,0%	14	28,0%	18	36,0%	26	52,0%	11	20,8%	60	29,4%
B2h	-	-	-	-	-	-	-	-	-	-	-	1	2,0%	-	-	-	-	-	1	0,5%
B4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1,9%	1	0,5%
B4b	4	8,2%	-	-	2	4,0%	2	3,7%	1	2,0%	-	-	2	4,0%	7	14,0%	2	3,8%	1	0,5%
C1	1	2,0%	-	-	1	2,0%	1	1,9%	3	6,0%	3	6,0%	-	-	2	4,0%	1	1,9%	10	4,9%

U2e1	-	-	-	-	-	-	1	1,9%	-	-	-	-	-	-	-	-	-	-	-	
U4a1	-	-	-	-	-	-	-	-	1	2,0%	-	-	-	-	-	-	-	-	-	
U5a1b1	-	-	-	-	-	-	-	-	1	2,0%	-	-	-	-	-	-	-	-	-	
U5a2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0,5%	
U5b	-	-	-	-	-	-	-	-	-	-	-	-	1	2,0%	-	-	-	-	1	0,5%
U5b2b3	-	-	-	-	-	-	1	1,9%	-	-	-	-	-	-	-	-	-	-	-	
Total Eurasian	4	8,2%	4	7,7%	5	10,0%	4	7,4%	6	12,0%	5	10,0%	2	4,0%	0	0,0%	2	3,8%	12	5,9%
L0a1a+200	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0,5%
L0a1b	-	-	-	-	1	2,0%	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L1b	-	-	1	1,9%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0,5%
L1b2	-	-	-	-	-	-	-	0,0%	-	-	-	-	1	2,0%	-	-	-	-	-	-
L1c1'2'4'5'6	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0,5%
L1c1a	-	-	-	-	-	-	1	1,9%	-	-	-	-	-	-	-	-	-	-	-	-
L1c2b2	1	2,0%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L1c3b	-	-	-	-	1	2,0%	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L2a1	-	-	1	1,9%	-	-	-	-	-	-	-	-	-	-	-	1	1,9%	1	0,5%	
L2a1+143	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0,5%
L2a5	1	2,0%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L2c	1	2,0%	-	-	-	-	-	-	-	-	-	-	1	2,0%	-	-	-	-	-	-
L2c3	1	2,0%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L3b	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0,5%
L3b1a8	-	-	-	-	-	-	1	1,9%	-	-	-	-	-	-	-	-	-	-	-	-
L3d1c	-	-	-	-	-	-	-	-	-	-	1	2,0%	-	-	-	-	-	-	-	-
L3d1d	1	2,0%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L3d5a	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0,5%
L3e2b	-	-	-	-	-	-	-	-	-	1	2,0%	-	-	-	-	1	1,9%	-	-	-
L3e4a	-	-	-	-	-	-	-	-	-	1	2,0%	-	-	-	-	-	-	-	-	-
L3f1b1a1	-	-	1	1,9%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0,5%
Total African	5	10,2%	3	5,8%	2	4,0%	2	3,7%	0	0,00%	3	6,0%	2	4,0%	0	0,00%	2	3,8%	8	3,9%

Total	49	100%	52	100%	50	100%	54	100%	50	100%	50	100%	50	100%	50	100%	53	100%	204	100%
-------	----	------	----	------	----	------	----	------	----	------	----	------	----	------	----	------	----	------	-----	------